Exploring Social Phenomena with Complex Systems Tools

Sébastian Grauwin

THÈSE

en vue d'obtenir le grade de

Docteur de l'Ecole Normale Supérieure de Lyon

Spécialité Physique

Université de Lyon

Laboratoire de Physique de l'ENS

Ecole Doctorale de Physique et d'Astrophysique de Lyon

présentée et soutenue publiquement

le 1^{er} Juillet 2011

par

Sébastian GRAUWIN

Exploring Social Phenomena with Complex Systems Tools

The Journey of a Physicist in an Interdisciplinary Playground

---

devant la commission d'examen composée de

Directeur de thèse:	Pablo JENSEN
Rapporteurs:	Jean-Philippe BOUCHAUD
	Jean-Pierre NADAL
Examinateurs:	Eric FLEURY
	Bruno LATOUR
	Jean-François PINTON

To my wife Christine and our daughter Jeanne

Visit my webpages at `http://www.sebastian-grauwin.com/`

Remerciements

Au moment où j'écris ces lignes, je prends conscience de ce que représentent les trois dernières années écoulées: un ensemble de rencontres, de discussions, de lectures, de recherches... toutes plus enrichissantes et stimulantes les unes que les autres. C'est donc tout naturellement que je tiens à adresser ici mes remerciements à tous ceux qui ont pu m'accompagner à un moment ou un autre au cours de ma thèse.

Tout d'abord un immense merci à Pablo pour m'avoir introduit dans le monde des systèmes complexes, pour sa confiance, ses encouragements et ses conseils tout au long de ces années.

Toute ma gratitude va à Jean-Philippe Bouchaud et Jean-Pierre Nadal qui ont bien voulu prendre la peine d'évaluer ce manuscript. Merci à eux ainsi qu'à Eric Fleury, Jean-François Pinton et Bruno Latour de me faire le plaisir et l'honneur de composer ensemble un jury de thèse de choc.

Toute ma reconnaissance va également à l'ensemble de mes collaborateurs et collègues de l'IXXI et d'ailleurs qui ont contribué à rendre cette thèse fructueuse et agréable. Pour toutes nos discussions, échanges et moment de détente, merci donc dans le désordre à Eric Bertin, Guillaume Beslon, Florence Goffette-Nagot, Sara Franceschelli, Céline Robardet, Dominique Boullier, Tommaso Venturini, Paul Girard, Jean-Baptiste Rouquier, John Mc Breen, Remi Lemoy... et pardon à tous ceux que j'oublie!

Merci aux auteurs de tous les bouquins que j'ai pu dévorer depuis mon enfance, vous tous qui avez attisé ma curiosité et qui m'avez donné le goût d'apprendre et de comprendre. Merci également à tous les acteurs d'internet: développeurs de logiciels libres, contributeurs aux encyclopédies en ligne, bloggeurs, ... cette thèse serait bien différente si vous n'aviez pas été là!

À mes amis et à ma famille: merci pour votre soutien, pour votre intérêt et pour votre patience devant mes longues tirades pleines de détails farfelus et de digressions. Quelle idée aussi, de m'interroger sur mon travail en cours!

Un mot enfin pour les dédicataires de cette thèse: Christine, ma compagne depuis plus de 10 ans et Jeanne, notre petit bout de fille. Le soutien et la compréhension de Christine représente tout pour moi. Cette thèse n'existerait pas sans elle. Notre vie ensemble, surtout depuis l'arrivée de notre petite merveille, est ce qui donne sens à mon travail. Merci à vous deux.

Contents

Introduction

This thesis explores the problems raised by the aggregation of entities into a global, collective level. This is certainly an old problem in many fields of science. Already in the 5^{th} century before Christ, Leucippus of Miletus started the atomic tradition, which argued that differences in atomic shape and size determined the various properties of matter. At the Renaissance, thinkers started to think about the articulation between individuals and a global entity called society thanks to quantitative data. In the 19^{th} century, Thomas Buckle published the *History of Civilization in England* which aimed at showing that, while individual destinies seem erratic and unpredictable, statistical laws do govern the course of human progress in a fixed and regular way. Maxwell was influenced by his reading of this very well-known work when he founded statistical physics by abandoning a Newtonian view of deterministic atoms to seek explanations in terms of statistics. This allowed him to explain quantitatively how the macroscopic characteristics of gases depend on its atoms' behaviours.

More generally, fundamental science has striven to reduce the diversity of the world to some stable building blocks such as atoms and genes. To be fruitful, this reductionist approach must be complemented by the reverse step of obtaining the properties of the whole (materials, organisms) by combining the microscopic entities, a notoriously difficult task (Hayden, 2010; Chouard, 2008; Anderson, 1972; Grauwin *et al.*, 2009a; Gannon, 2007). Therefore, linking the microscopic and macroscopic behavior is at the heart of many natural and social sciences. However, this apparent similarity conceals essential differences across disciplines: while physical particles are assumed to optimize the global energy, economic agents maximize their own utility. In sociology, the very distinction of two levels (individuals and social groups) is put into question, as humans cannot properly be defined without taking into account the social groups they need to

live in (e.g. language, an essential part of the individual which is acquired through society).

More specifically, our scope here is to address the aggregation problem in social systems, using tools derived from statistical physics, and more generally quantitative tools. This problem has received new impetus recently, mainly from to the avalanche of digital data on social systems. These cover many different fields (Lazer, 2009): stock markets, demographics, geographical tracking of individual mobility, social networking, scientometrics... The amount and quality of these data open the possibility of a profound renewal of quantitative sociology (Latour, 2010).

The increase of computing power has also brought the possibility of modelling complex socio-economic systems in a new way. Agent-based models simulate virtual societies in which simple rules are used to characterize the actions and interactions of large amount of agents. Simulations have shown that unexpected collective behaviors may emerge from these simple rules (Schelling, 1969). In economics, they have allowed to go beyond the simplistic assumptions of, for example, perfectly rational agents, which are necessary to obtain analytical theories (Grauwin *et al.*, 2009b) but can lead to unrealistic predictions (Bouchaud, 2008).

The situation seems ripe for tighter connections between social and natural scientists. As most social scientists lack the training needed to handle, visualize and model massive amount of data, they could benefit from collaborations with mathematicians, computer scientists or physicists, who are more familiar with that kind of tools. However, for these collaborations to be fruitful, mutual knowledge and respect are essential. From our "modelling" side, one has to be cautious to understand and respect the specificity of social systems, without trying to export our preferred models in a "wild way" (what Bernard Walliser called "brute transfer analogies" (Walliser, 2005)).

In the following, we present three examples of our "respectful" approach to the modelling of social systems. These three parts of the thesis can be read independently.

Individual dynamics and Schelling's segregation model

The first part of this thesis focuses on a paradigmatic model of the emergence of puzzling macroscopic behaviour from simple individual rules, Schelling's segregation model (Schelling, 1969, 1978). As soon as 1969, Thomas C. Schelling proposed a model aiming at formalizing the aggregate consequences of individual preferences regarding the social environment. In his 1971's *Dynamic Models of Segregation* paper, he simulates the evolution of the spatial repartition of two types of agents living in a virtual city. The agents

have preferences over the composition of their neighborhood and are given the opportunity to move in order to satisfy their preferences, or *utilities*. Schelling shows that even "mildly" segregative individual preferences lead to high levels of segregation, even if this global outcome does not correspond to a residential configuration maximizing the collective utility. In other words, the global pattern is not a linear extrapolation of the individual preferences.

Schelling's 1971 paper is widely known thanks to this apparently paradoxical effect: a mild preference for one's neighbors to be of the same color leads to total segregation, even if total segregation does not maximize collective utility. Later research showed that even a peaked utility function, that is, a function achieving its maximum for a 50%- 50% environment, can lead to a fully segregated equilibrium as soon as this function is asymmetric - even in a city where the two groups are equally proportioned (Zhang, 2004b; Pancs & Vriend, 2007; Barr & Tassier, 2008). Schelling's results are also robust to different definitions of individual's environment (Pancs & Vriend, 2007; Fagiolo *et al.*, 2007). Recently, Zhang (2004a,b, 2010) developed analytical solutions to different versions of Schelling's model. Basically, the paradox arises from a lack of coordination among agents, which generate externalities preventing the system to reach the social optimum. Our work on Schelling's model builds on this literature.

The two first chapters focus on economics' aspects of Schelling's model. Chapter one examines the effects of introducing coordination in the moving decisions. This coordination is achieved in two different ways. We first impose different levels of taxes proportional to the externality generated by each move of the agents. We then investigate the effect of the introduction of a local coordination by vote of co-proprietors, who are defined as the closest neighbors of each agent. In both cases, we show that even a small amount of coordination can significantly reduce segregation.

Chapter two proposes an analytical resolution of Schelling segregation model for a wide range of utility functions. Using evolutionary game theory, we provide existence conditions for a potential function, which characterizes the global configuration of the city and is maximized in the stationary state. We use this potential function to analyze the outcome of the model for three utility functions corresponding to different degrees of preference for mixed neighborhoods. The main results are :

- We show that linear utility functions are the only case for which the potential function is proportional to collective utility, the latter being hence maximized in stationary configurations.

- Schelling's original utility function is shown to drive segregation at the expense of

collective utility.

- If agents have a strict preference for mixed neighborhoods but still prefer being in the majority *versus* in the minority, the model converges to perfectly segregated configurations, which clearly diverge from the social optimum.

Departing from earlier literature, these conclusions are based on analytical results. Our model being based on bounded neighborhoods rather than continuous ones (as in Schelling's original model), we discuss the differences between the bounded and continuous definition of neighborhoods and show that, in the continuous neighborhood case, a potential function exists if and only if the utility functions are linear. As a by-product, our analysis builds a bridge between Schelling's model and the Duncan and Duncan segregation index. The collaboration with an economist (Florence Goffette-Nagot) has lead to two papers submitted to economics' journals. One is almost accepted in a prestigious journal (J of Public Economics), showing the relevance of our approach for that community, while the other is also submitted to a very good journal (J of Economic Behavior and Organization).

Chapter 3 switches to a physicist's point of view on Schelling model. We focus on a simplified version of the model, with agents of a single color and vacant sites, which still retains the essential feature of the original model, namely the paradox of macroscopic segregation for microscopic "tolerant" individuals. By generalizing the concept of free energy in order to include dynamics driven by individual optimization, we are able to solve exactly the model for arbitrary utility functions. Specifically, we introduce a parameter which interpolates continuously between cooperative and individual dynamics. We show that increasing the degree of cooperativity induces a qualitative transition from a segregated phase of low utility towards a mixed phase of high utility. This work has been published in PNAS in 2009.

In the last chapter, we present our "linking" approach in a more general way. We show how our approach could pave the way to analytical treatments of a wide range of socio-economic models. As a first example, we derive an analytic solution for a congestion model. This chapter has been accepted for publication in Advances in Complex Systems (2011).

Maps of Science

The second part of this thesis addresses the question of emergence of macroscopic features out of interactions between individuals from a different perspective. Instead of

simulating simple models, it uses data from real social systems and explores the question of aggregation of these data into insightful (social) groups. Specifically, we use the huge existing databases on scientific literature (mainly Web of Science) to investigate the existence and evolution of paradigms or scientific institutions.

Chapter 5 tackles the question of the existence and coherence of a hypothetical field, that of "complex systems science". Picking up again the reductionist theme, it is interesting to note that the science of complex systems tries to obtain the properties of the whole by combining the microscopic entities, albeit from a different perspective. It adds the idea that "universal principles" could exist, which would allow for the prediction of the organization of the whole regardless of the nature of the microscopic entities. Ludwig Von Bertalanffy wrote already in 1968: "It seems legitimate to ask for a theory, not of systems of a more or less special kind, but of universal principles applying to systems in general" (Von Bertallanffy, 1976). This dream of universality is still active: "[Complex networks science] suggests that nature has some universal organizational principles that might finally allow us to formulate a general theory of complex systems" (Solé, 2000). Have such universal principles been discovered? Could they link disciplines such as sociology, biology, physics and computer science, which are very different in both methodology and objects of inquiry (SantaFe, 2010)? Or does the "complexity" field have too high a "mouth to brain ratio" (Scientific American, 1995)?

Thanks to a large database (141 098 records) of relevant articles published between 2000 and 2008, we empirically study the "complex systems" field and its claims to find universal principles applying to systems in general. The study of references shared by the papers allows us to obtain a global point of view on the structure of this highly interdisciplinary field. We show that its overall coherence does not arise from a universal theory but instead from computational techniques and fruitful adaptations of the idea of self-organization to specific systems. We also find that communication between different disciplines goes through specific "trading zones", *ie* sub-communities that create an interface around specific tools (a DNA microchip) or concepts (a network). We have also gathered an exhaustive database of French natural science articles (around 65 000 published in 2000 and 80 000 in 2010) to investigate the place of complex systems science within the whole landscape of science. These maps suggest that there exists (yet?) no conceptual kernel for "complexity science", some unified set of theories or concepts that could give the field enough coherence to be a subfield identifiable by shared references (and therefore by bibliographic coupling). Certainly, *Self-Organization*, *Self-Organized Criticality* or *Complex Networks* do constitute such coherent subfields. However, despite some claims, these subfields seem to have only weak links to the rest of the complex

systems "galaxy" shown above, which prevents the formation of a coherent "complex systems science". This work is currently submitted to Plos One.

In chapter 6, we develop a set of routines that allows to draw, in a few hours, different maps of the research carried out in a scientific institution. Our toolkit uses OpenSource elements to analyze bibliometric data gathered from the Web Of Science. We take the example of our institution, *ENS de Lyon*, to show how different maps, using co-occurrence (of authors, keywords, institutions...) and bibliographic coupling can be built. These maps may become a valuable tool for institutions' directors, as they offer different views on the institution at a global scale. This work is submitted to Scientometrics.

In these two studies, we use an old but quite forgotten approach to measure the relations between articles, i.e. bibliographic coupling (Kessler, 1963). The idea is that papers sharing references are closer than papers with no common reference. By grouping similar papers with standard techniques (such as modularity optimization (Girvan & Newman, 2004)), we define scientific communities or subdisciplines in a "natural" way, which is in any case largely independent of administrative or institutional structures such as disciplines or laboratories. This has the decisive advantage of allowing for a "fresh" look at science in action or research frontiers. For example, we identify a large community working on *complex networks* that has received no institutional support (in France at least) or recognition in terms of "subject categories" from Web of Science and would therefore be invisible had we interrogated the database through these categories. We also identify several distinct communities working on nanostructures in France (labelled as nanoindentation, nanowires, quantum dots ...) in 2010, which questions the use of a single discipline from an institutional point of view. In short, using large databases should not prevent scientists from inquiring about the interpretation of these data : the categories used to extract the data or the meaning of the aggregations used to examine it.

Emergence of Institutions in Social Systems

The meaning of aggregations in the social space is the central question of the last part of the thesis. Various sociological theories try to apprehend social phenomena under different points of view. According to the *holistic paradigm*, mainly based on Emile Durkheim's ideas, individuals are embedded into social structures and institutions that constrain them, shape their actions and emotions. On the contrary, in the atomic (or individualistic) paradigm, individuals should be considered as the central ontological

elements in social systems. Constant interactions between the "social atoms" construct society, *ie* structures, institutions, norms, which are supposed to serve the interests of the individuals. Both the holistic and the atomistic approaches are based on a strong assumption: there is a clear dichotomy between two "levels", namely individuals and society. This distinction is common to many fields and is often summarized by the notion that "the whole is greater than the sum of its parts".

Chapter 7 questions this fundamental distinction. For more than a year, we have been collaborating with a team of sociologists from the MediaLab at Sciences Po. The scope is to use the social theory developed by Gabriel Tarde at the end of the 19th century to imagine different ways of conceptualizing the articulation between "wholes" and "parts". One hypothesis is that the distinction between two levels is an artefact originating in the difficulty of navigating through huge amounts of data. This hinders the possibility to visualizing (and therefore conceptualizing) the evolution of social phenomena without making a distinction between the individual and the aggregated levels. Using scientometrics data, we first explore alternative ways of visualizing wholes and parts. We have also been working in collaboration with the MediaLab sociologists in order to attempt to formalize Tarde's ideas in a model which would lead to a simulation algorithm. We started by defining the features requested by sociologists for the algorithm to be faithful to Tarde's intuitions. The main point was to obtain "wholes" through the simplification of complex individuals. We then developed a first prototype which fulfilled several of these demands but raised many more questions. While this process was positive because it uncovered several fuzzy features in Tarde's social theory, it also turned out that the scope of the model was too ambitious for the time scale of this thesis. Chapter 8 then presents the current state of ongoing work on another - more standard - approach into emergence of social structures. We still focus on a question raised by our colleagues sociologists, namely the existence of lasting structures from non lasting entities. But we use the standard (and safer!) physicist approach : start with simple individuals and build groups with interesting features. Specifically, our model is based on Deffuant's opinion model (Deffuant *et al.*, 2000) with an adaptive network similar to what can be found in Kozma & Barrat (2008). While these papers focus on the stationary properties of the model - the final, stable states reached at the end of the simulations - we chose to build a model to investigate the dynamical properties of social structures which would always be changing. For this, we have added three features. First, noise, which allows to recover the intuitive fact that convergence of opinions cannot be *strictly* limited to persons already sharing similar opinions. Second, we add the possibility for agents to "die", to investigate whether the groups obtained lasted longer

than the typical agent lifetime. Third, we introduce heterogeneity in opinion space by adding the "age" of the agent as a characteristic influencing the exchange of opinions. While the outcomes of the model are not yet fully explored, we already found interesting group dynamics. All along the runs, new groups may emerge, groups may disappear by merging with other groups or by death of all their agents, some groups may last for hundreds of agents' lifetimes... These investigations echo the famous study by Georg Simmel of "the persistence of social groups" Simmel (1898): "it is meaningful to speak of group identity, despite shifting membership and low institutionalization, if there is some membership continuity in contiguous stages [...] The change, the disappearance and entrance of persons, affects in two contiguous moments a number relatively small compared with the number of those who remain constant. The departure of the older and the entrance of the younger elements proceed so gradually and continuously that the group seems as much like a unified self as an organic body in spite of the change of its atoms".

Despite this tempting parallel, we do not think that this kind of simplistic model should aim at being realistic by inclusion of additional ingredients. Rather, our discussions with Sciences Po's sociologists open another direction. We speculate that these models can help them enriching their conceptualizations of the structuration phenomenon. This could arise from a detailed examination of the structurations that happen in this virtual society and their tentative interpretation using the usual sociological theories. Does this confrontation help them in any way in renewing their conceptual repertoire? Do we observe group evolutions that are unexpected and difficult to explain? The future of this stimulating and demanding collaboration will tell!

Part I

Individual dynamics and Schelling's segregation model

Chapter 1

Schelling's Model with Local Coordination

1.1 Introduction

As soon as 1969, Thomas C. Schelling proposed a model aiming at formalizing the aggregate consequences of individual preferences regarding the social environment (Schelling, 1969, 1978). In his 1971's *Dynamic Models of Segregation* paper, Schelling showed that if the preferences considered are such that an environment of more than 50% of own-group agents is highly preferred to a less than 50% of own-group environment, then the equilibrium configuration exhibits high levels of segregation. Schelling's 1971 paper is widely known thanks to this apparently paradoxical effect: a mild preference for one's neighbors to be of the same color leads to total segregation, even if total segregation does not maximize collective utility. Later research showed that even a peaked utility function, that is, a function achieving its maximum for a 50%- 50% environment, can lead to a fully segregated equilibrium as soon as this function is asymmetric - even in a city where the two groups are equally proportioned (Zhang, 2004b; Pancs & Vriend, 2007; Barr & Tassier, 2008). Later literature showed that Schelling's results are also robust to different definitions of individual's environment (Pancs & Vriend, 2007; Fagiolo *et al.*, 2007). A physical analogue of Schelling model (Vinkovic and Kirman, 2006) was proposed. Zhang (2004a,b, 2010) develop analytical solutions to different versions of Schelling's model. Grauwin *et al.* (2009a, 2011) generalize this type of solution and propose an analytical treatment that allows to calculate the global segregation pattern starting from a broad range of individual utility functions.

In Schelling-type models, an assumed condition for integration to occur is that agents

have a preference for mixed environments. But that condition is often not a sufficient one, because mixed configurations are unstable with respect to fluctuations, whereas segregated configurations are very stable. The stability of the city configurations and the gaps between the agents' micro-motives and the emergent macro-behavior have been recognized as the results of a coordination problem (Zhang, 2004a,b, 2010). It is indeed the existence of locational externalities, that is, the fact that moving agents do not take into account their impact on their neighbors' utilities, that moves the city to highly segregated configurations, even if individuals' preferences favor a mixed environment. This is particularly the case when the agents' preferences are given by the asymmetrically peaked utility function, the system ending in highly segregated configurations whereas the agent's main preference is for mixed neighborhoods (Zhang, 2004b; Pancs & Vriend, 2007; Grauwin *et al.*, 2009b,a). Externalities produced by agents' mobility are also encountered in economic geography (see Charlot *et al*, 2006) and little is known about the means allowing to fight them.

A way to reach the welfare maximizing configuration in Schelling model could be a mechanism internalizing the externalities generated by the agents' selfish moves. To do so, a classical idea in economics is to impose a taxation equal to the generated externality. Such a policy implemented by a central authority is known to lead to social optimum, where the collective utility reaches its highest possible value. Recently, some papers have proposed to add some corrective mechanisms in the original Schelling model to reinforce the integrated configurations. Dokumaci & Sandholm (2007) propose to tax the agents proportionally to the density of population in their neighborhood. The tax level depends on the ethnic origin of the agents, as might be the case under various form of affirmative action. Barr & Tassier (2008) introduces additional social interactions into Schelling's model by coupling it with a Prisoner's Dilemma played with neighbors. In both cases however, the effect of the added mechanism could be reformulated in terms of a redefinition of the agents' utility function.

The present chapter, based mainly on simulations, examines the effects of introducing coordination in the moving decisions, and shows that even a small amount of coordination can break segregation. We first verify that introducing a Pigouvian tax leads to the social optimum in a dynamic model of segregation with an asymetrically peaked utility function (section 3). We also investigate the impact of different levels of taxes and we show that a tax equal to only one fifth of the generated externalities is sufficient in certain cases to reduce consequently the gap between the equilibrium and the social optimum, created by the agents' selfish behavior. To the best of our knowledge, this work has never been done in the context of Schelling segregation models. However,

these tax policies are not easy to implement in practice since they require from the central authority a perfect knowledge of the city at the local scale. Such information is rarely perfectly nor freely available. Hence, the policies implemented by central authorities correspond generally to mechanisms that differ from the Pigouvian tax. Therefore, we propose in section 4 a tax based on the share of similar neighbors' in agents' neighborhoods. Finally, we investigate in section 5 the effect of the introduction of a local coordination by vote of co-proprietors. That completely new coordination mechanism has the advantage of remaining in the spirit of Schelling's model, adding only individual decisions based on the same utility as the moving agent, without any need of a central authority. Before developing models with some coordination, we present in section 2 the standard Schelling model which we will use as a benchmark.

1.2 A standard model

1.2.1 Basic setup

The city and neighborhood definition

Our artificial city is a two-dimensional NxN square lattice with periodic boundary conditions, *ie* a torus containing N^2 cells. Each cell corresponds to a dwelling unit, all of equal quality. We suppose that a certain characteristic divides the population of this city in two groups of households that we will refer to as red and green agents. Each location may thus be occupied by a red agent, a green agent, or may be vacant. We denote by N_V the number of vacant cells, and by N_R and N_G the number of respectively red and green agents. All these numbers are kept fixed over a simulation. The parameter N thus controls the size of the city, the parameter $v = N_V/N^2$ its vacancy rate, and the fraction $n_R = N_R/(N_R + N_G)$ its composition.

We define a *state* x of the city as a N^2-vector, each element of this vector labeling a cell of the NxN lattice. Each state x thus represents a specific configuration of the city. We note X the set of all possible configurations, the demographic parameters (N, v, n_R) being fixed.

Since Schelling (1969)'s work, *continuous neighborhood* models describe cities where the neighborhoods are centered on the local perception of each agent. In a continuous neighborhood description, one assumes that the neighborhood of an agent is composed of the H nearest locations surrounding him. The $H = 4$ "Von Neumann neighborhood" and the $H = 8$ "Moore neighborhood" that are displayed among other examples on Fig 2.1 are the most commonly used neighborhoods in agent-based computational models.

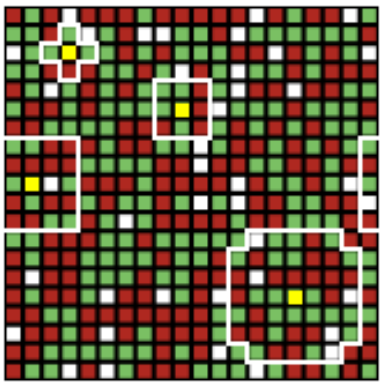

Figure 1.1: **Different forms of neighborhood.** Red, green and white squares denote respectively red agents, green agents and vacant cells. The neighborhood of an agent corresponds to his H nearest cells/locations. Around the agents marked in yellow, we enlightened by the white frontiers a $H = 4$, a $H = 8$, a $H = 24$ and a $H = 44$ continuous neighborhood. [If you printed this document in black and white, the red and green squares should appear respectively in dark grey and soft grey.]

Note that since some locations remain empty, the size H of the neighborhood of an agent can also be interpreted as the maximum number of neighbors an agent can have. The global characteristics of our model remain qualitatively independent of any specific definition of neighborhood, provided its size H is relatively small compared to the size N^2 of the city, in order to maintain the "local" property of neighborhood.

In all the simulations presented in this chapter, the demographic parameters of the city are fixed. The size of the city is set to $N^2 = 400$, a good compromise between the necessity to take a large value of N to avoid small city effects[1] and the convenience to take a small value of N to achieve short computation times. The number of agents of each group is fixed to $N_R = N_G = 180$ and the vacancy rate is fixed to $v = 10\%$. As usually assumed, we use continuous neighborhoods, that is, the neighbors of an agent are the agents living on the H nearest cells surrounding him. Unless otherwise stated, the neighborhood size is fixed to $H = 8$.

Agent's utility function

Each agent computes his own level of satisfaction via a utility function which depends only on his neighborhood composition. Let us consider an agent whose neighborhood is composed of R red agents, G green agents and V vacant cells. Since $R + G + V = H$, one needs two independent parameters to describe the composition of the neighborhood of the agent. In all generality, we can thus write the utility of an agent for example as a

[1] Such as those emphasized by Singh *et al* (2009).

function of R and G or as a function of the fraction s of the agent's similar neighbors and V. Most models of the literature assume for simplification that agents of a same group share the same utility function. Hence, one only needs a utility function u_R to describe the preference of the red agents and a utility function u_G to describe the preference of the green agents.

It is easy to understand that a utility function can be defined up to an additive constant depending on a reference situation but also up to a multiplicative constant depending on the measure scale. A common choice in the literature is to take these constants such that a zero utility level denotes a complete dissatisfaction of the agent and a utility of one denote complete satisfaction. We will stick to that use in the following.

We will suppose that all the agents share the same "asymmetrically peaked utility function" $u_m(s)$, where s is the fraction of one's similar neighbors and $0 \leq m \leq 1$; u_m is defined by:

$$u(s) = \begin{cases} 2s & \text{for } s \leq 0.5 \\ 2 - m - 2(1-m)s & \text{for } s > 0.5 \end{cases} \tag{1.1}$$

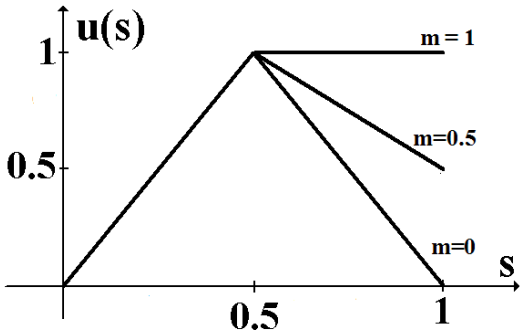

Figure 1.2: In our simulations, the agents all share the same utility function $u_m(s)$. We explore the change in behavior of the agents according to the value of parameter m. Except for $m = 1$, the agents always have a strict preference for perfectly mixed neighborhood. Except for $m = 0$, their utility function presents an asymmetry: they prefer all-similar neighborhoods to all-dissimilar neighborhoods.

By varying the value of the parameter m, we will then be able to explore the responses of our system to a whole family of utility functions. Our choice of working only with the asymmetrically peaked utility function is driven by the fact for $m > 0$, the asymmetry in favour of the all-similar neighborhood in these utility functions leads to segregation patterns at the city scale at the cost of a low collective utility (see Grauwin *et al.*, 2009b; Pancs & Vriend, 2007; Barr & Tassier, 2008). This family of utility functions is thus the perfect candidate for testing whether the introduction of any type of coordination

might allow to break segregative patterns and lead to more integrated patterns in which the collective utility would be higher.

Aggregate measures

In order to characterize a configuration at the global (city) scale, we need to introduce aggregate measures. Let s_k, $k \in \{1, ..N_R + N_G\}$ be the fraction of agent k's similar neighbors. In order to characterize the global level of segregation, we introduce for each configuration x the average fraction of same-type neighbors, or *similarity*:

$$\bar{s}(x) = \frac{1}{N_R + N_G} \sum_k s_k \tag{1.2}$$

Similarity is a well-known measure of segregation that was already used by Schelling (1971). However, the knowledge of $\bar{s}(x)$ may not be sufficient to determine if the level of segregation of a given configuration is significantly high or low compared to a random configuration with the same demographic parameters: a similarity of 0.8 would point out a high degree of segregation in a city with equally proportioned groups ($n_R = 0.5$) but would be insignificant in a city with disproportioned groups ($n_R = 0.9$). To avoid this kind of problem, we define in the spirit of Carrington & Troske (1997) the normalized index of similarity $s^* : X \to [-1, 1]$ by

$$s^*(x) = \begin{cases} \dfrac{\bar{s}(x) - \bar{s}_{random}}{1 - \bar{s}_{random}} \text{ if } \bar{s}(x) \geq \bar{s}_{random} \\ \\ \dfrac{\bar{s}(x) - \bar{s}_{random}}{\bar{s}_{random}} \text{ if } \bar{s}(x) < \bar{s}_{random} \end{cases} \tag{1.3}$$

where $\bar{s}_{random}$ is the expected value of the similarity index $\bar{s}$ implied by a random allocation of the agents in the city. This value depends on the size N, vacancy rate v and composition n_R of the city and on the size H of a neighborhood. However, in the cases studied below, i.e. for $n_R = 0.5$, $v > 0$, $H \geq 1$, $N \gg 1$, the value of $\bar{s}_{random}$ is indistinguishable from 0.5.

Fig 1.3 displays some examples of configurations along with their s^*-values. The reference value $s^* = 0$ corresponds to an average random configuration. Positive values of s^* mean that the agents have more similar neighbors than in the random case, and negative values that their neighborhood contains less similar neighbors than in an average random case. A maximum value $s^* = 1$ corresponds to the case where all the neighbors of all agents belong to their own group. Practically, given the random fluc-

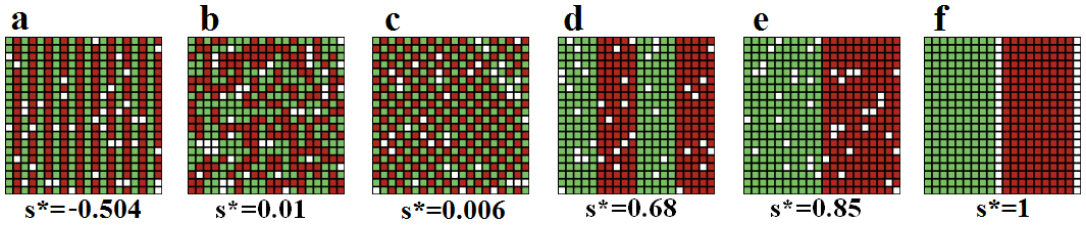

Figure 1.3: Some examples of configurations of the city, along with their s^*-values. For ordered configurations such as **a**,**c**,**d**,**e**, the value of s^* fluctuates with the precise location of the vacant cells. The neighborhood size is $H = 8$ and the demographic parameters are fixed to ($N = 20$, $v = 10\%$, $n_R = 0.5$).

tuations of the configurations, an absolute deviation larger than 0.05 corresponds to configurations which deviate significantly from the random configuration. Notice finally that the normalized similarity index cannot grasp every aspects of the city configurations. Because of the random fluctuations, it can't for example make the difference between a 'checkerboard' (Fig. 1.3 c) ordered configuration and a random configuration (Fig. 1.3 b).[2]

We also introduce notations in order to characterize the level of collective utility:

$$U(x) = \sum_k u_k \tag{1.4}$$

$$U^*(x) = \frac{1}{N_R + N_G} U(x) \tag{1.5}$$

where u_k is the utility of agent k, $U(x)$ denote the collective utility of a configuration x and $U^*(x)$ its normalized value.

1.2.2 The logit dynamic rule

As explained before, in standard Schelling-type models, agents move only to satisfy their own interest. In our simulations, we consider that the dynamics follows a logit behavioral rule: at each iteration, an agent and a vacant cell are randomly chosen and the probability that this picked agent moves in that vacant cell is written as:

$$Pr\{move;\, WC\} = \frac{1}{1 + e^{-\Delta u/T}} \tag{1.6}$$

[2]For a complete and detailed discussion on segregation indices, see for example Massey & Denton (1988), Reardon & Firebaugh (2004), Reardon & O'Sullivan (2004).

where WC stands for "Without Coordination" and with $T > 0$ a fixed parameter and Δu the change in utility if the agent was to move to the chosen cell. At the first iteration, an initial configuration is randomly chosen.

The scalar T can be interpreted as a measure of the level of noise in an agent's decision. Clearly, the probability for an agent to take a utility-decreasing move drops down as $T \to 0$ and the described rule thus converges to the non-strict best-response rule. For any finite $T > 0$, the agents choose non-best replies with a non-zero probability, but actions that yield smaller payoffs are chosen with smaller probability.

This kind of perturbed best-response dynamics has been developed in e.g. Anderson *et al.* (1992) or Young (1998). Taken as a behavioral rule, the underlying logit choice function in Eq. 1.6 is rooted in the psychology literature (Thurston, 1928). From the microeconomic point of view, it can be given a justification in terms of a random-utility model where the random part in the utility function can be interpreted as a way to take into account criteria other than the neighborhood composition such as the quality of the housing, the proximity to one's workplace or any other idiosyncratic amenity. As shown in section 1.4, the logit dynamic rule has the advantage to grasp more aspects of reality (taking into account idiosyncratic amenities) and to lead the system to stationary results which are independent of the initial configuration. This is why we prefer that kind of dynamics to a more standard "best response" dynamic rule.

Beside bringing some degree of realism through the introduction of randomness in the agents' behavior, the logit rule also provides a strong analytical framework to Schelling's model (see Grauwin *et al.*, 2009a,b).

Obviously, it implies that the probability that the state at the t^{th} iteration x^t is equal to a given state x only depends on the state at the previous iteration x^{t-1}:

$$\Pr(x^t = x | x^{t-1}, \ldots, x^1, x^0) = \Pr(x^t = x | x^{t-1}) \tag{1.7}$$

The dynamic rule thus yields a finite Markov process.
It is then easy to figure out that the Markov chain describing our system is irreducible (since $T > 0$ each imaginable move has a non-zero probability to happen and it is thus possible to get to any state from any state), aperiodic (given any state x and any integer k, there is a non-zero probability that we return to state x in a multiple of k iterations) and recurrent (given that we start in state x, the probability that we will never return to x is 0). These three properties ensure that the probability to observe any state x after t iterations starting from a state y converges toward a fixed limit independent of the starting state y as $t \to \infty$.

In other words, for each set of parameters and dynamic rule, there exists a stationary distribution

$$\Pi : x \in X \to \Pi(x) \in [0,\, 1] \;,\; \sum_{x \in X} \Pi(x) = 1 \tag{1.8}$$

which gives the probability with which each state x will be observed in the long run.

Clearly, for $T \to \infty$, the randomness introduced in the dynamical rule prevails and the stationary distribution is just a constant. Similarly, for any finite $T > 0$, our dynamical system (the city) is evolving toward an attractor composed of a subset A of X. It follows that any measure $\mathcal{M}$ performed on the states space X - such as the global utility U - will in the long run fluctuate around a mean value $\mathcal{M}_\infty = \sum_{x \in A} \Pi(x)\mathcal{M}(x)$. These mean values may depend on the amplitude of the noise T, but the amplitude of the fluctuations decreases as $T \to 0$. These intuitions are confirmed later.

1.2.3 Simulations: influence of parameters m and T

We introduce the parameter τ as the average number of moves per agent. Considering the demographic parameters used in our simulations, an increment of 1 in τ corresponds to $(1 - v)N^2 = 360$ performed moves. In the simulations presented below, we use τ as a chronological reference.[3]

We present on Fig. 1.4 a typical evolution of the city in the case where the agents move without coordination, the level of noise being fixed to $T = 0.1$ and the parameter m to 0.5. Starting from a random configuration, we observe the rapid formation of homogeneous areas which slowly melt into one another leading to the emergence of a highly segregated configuration where the city is divided into two uniform areas, each inhabited by only one type of agent. The bottom panel of Fig. 1.4 shows that after a transition time, the city enters a stationary phase in which the segregation index s^* and the (normalized) collective utility U^* fluctuate with rather low amplitudes. Even though the agents' preferences go to mixed configurations, their moves lead to a highly segregated configuration at the city level (the stationary value of s^* is close to 0.8) in which most of the agents are far from being fully satisfied (the stationary value of U^* is close to 0.6). As previously stated in the literature, this is the consequence of locational externalities: agents slightly favor majority status over minority status and do not account for their impact on neighbors' neighborhood composition (Zhang, 2004b; Pancs &

[3] A more obvious choice of chronological reference could be the number t of simple iterations (i.e. the number of attempted moves). Neither choice accounts for the proportion of accepted moves (whose cumulated value is τ/t), whose instantaneous value depends on the values of the various parameters and on the state of the system. Ideally, it may be interesting to follow this rate of "moving iterations" with the dynamic evolution of the city.

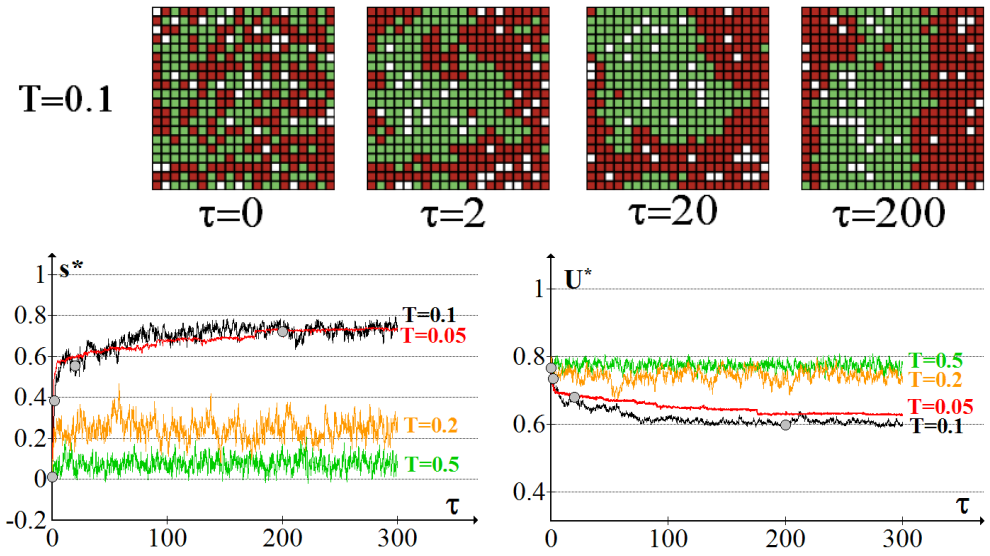

Figure 1.4: Evolution towards a highly segregated configuration starting from a random configuration in the case of the WC rule. **Top panel.** Some snapshots of the evolution of the city for a noise level $T = 0.1$; **Bottom panel.** Evolution with τ of the index of similarity and of the collective utility for different simulations with different noise levels. The grey dots on the $T = 0.1$ curves correspond to the snapshots presented on the top panel. $m = 0.5$.

Vriend, 2007). This makes mixed neighborhoods unstable and segregated configurations very stable. In particular, once the city is divided into homogeneous areas, a red (green) agent will have no incentive to go from the red (green) area to the green (red) one, because his utility would then drop from m to 0.

Fig. 1.4 further shows that the level of noise has an effect on the fluctuations of s^* and U^* in the stationary phase. While these fluctuations are rather low for $T = 0.05$, their amplitude increases with T before reaching a saturation value. Finally, while further studies would be necessary to characterize the influence of T on the time needed to reach the stationary phase, one can infer from fig 1.4 that as T decreases, this transition time increases.

The influence of the parameters T and m on the stationary configurations can be observed on Fig. 1.5 and 1.6. For high values of T ($T \geq 0.5$), the dynamics is essentially governed by the randomness introduced in the logit. In the limit $T \gg 1$, the agents are distributed uniformly in the corresponding stationary configurations, which induces a similarity index equal to zero for all values of m. In this case, the probability $p(s)$ for an agent to have a fraction s of similar neighbors being independent of m and thank to the bilinear form of the asymmetrically peaked utility function, the collective utility can

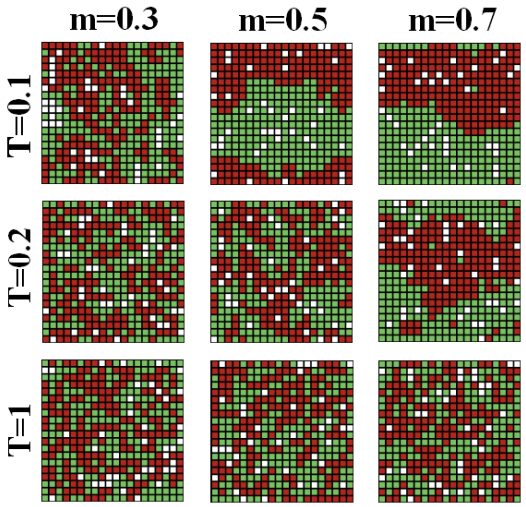

Figure 1.5: **Typical stationary configurations obtained with the WC dynamic rule for different values of m and T.**

be written (the sums being taken on all the possible discrete values of s):

$$U^* = \sum_s p(s)u(s) = U^*_{m=0} + m \sum_{s \geq 0.5} (2s-1)p(s) \tag{1.9}$$

This form explains the linear dependency to m of the stationary mean collective utility observed on Fig. 1.6 for high values of T. Indeed, there are two parts in U^* when T is large and the distribution of agents almost random: because the distribution does not differ from the distribution when $m = 0$, agents have the same "baseline" utility; however, those who have more than half of same-type neighbors have a higher utility level than in the $m = 0$ case, and the gap is linearly increasing in m.

For low values of T (roughly $T \leq 0.1$), the dynamics is governed mainly by the deterministic part of the logit rule, *i.e*, the agents' preferences. For $m < 0.3$, the preference for a mixed neighborhood of the agent prevails: locally mixed configurations are observed and the segregation index is low. The dispersion in the distribution of neighborhoods' composition induces a level of mean utility around 0.75, which is just a bit better than what is obtained with a random allocation of agents (see Fig. 1.6). For $m \geq 0.3$, the asymmetry in the agents' utility function induces a higher stability of highly segregated states to which the system converges. These segregated states are particularly harmful in terms of welfare for m comprised between 0.3 and 0.8. For this range of value of m, preferences for majority status are high enough for segregation

to appear, but not high enough for agents to benefit from the stationary segregated configurations. Therefore, the final outcome is always worse from a welfare point of view than it is in a random distribution of the agents. This observation clearly points out the deficiency with respect to social welfare of the location mechanism, which generates locational externalities.

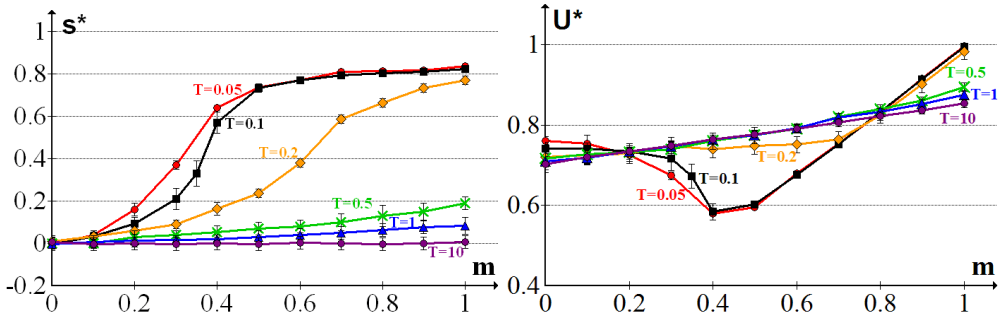

Figure 1.6: **Stationary mean values of the similarity index and of the collective utility** as a function of m, for different level of noise T. The error bars give the standard deviation of the fluctuations of s^* and U^* once the system has reached its stationary phase. The mean value and standard deviation are computed over 10 periods of the stationary phases (when the fluctuation are important, we use larger temporal windows).

The next three sections present three different ways of introducing coordination in the model, in order to explore the robustness of the deficiency of the location mechanism to the introduction of different coordination mechanisms.

1.3 A partial coordination by a Pigouvian tax

1.3.1 Basic setup

The idea of introducing a partial coordination is to make the moving agents take into account the whole or a fraction of the externality generated by their move on all the affected agents, *ie* her past and potentially new neighbors. A mechanism of this kind can results from the intervention of a benevolent planner who taxes negative externalities and rewards positive externalities. To that aim, we introduce a tax which is paid when the move occurs, and does not affect the utility received in the location itself. The amount of the tax is proportional to the externality produced by the move. In our reference case (without coordination), an agent decides to move according solely to the benefit Δu she would achieve if she was to move. As a consequence from that move, she would generate an externality that amounts to $\Delta U - \Delta u$. The tax to be paid by this

agent when moving is therefore:

$$\tau = \alpha(\Delta U - \Delta u) \tag{1.10}$$

where $0 \leq \alpha \leq 1$ is a parameter controlling the tax level, the limit case $\alpha = 0$ corresponding to the without coordination case and the limit case $\alpha = 1$ corresponding to a 'Global Coordination' case where only the collective utility is taken into account.

1.3.2 Dynamic rule

The probability that a move happens depends then on the differential of utility enjoyed by the agent in the initial location and destination and the tax she has to pay, as follows:

$$Pr\{move; PT\} = \frac{1}{1 + e^{-[\Delta u + \alpha(\Delta U - \Delta u)]/T}} \tag{1.11}$$

where PT stands for "Pigouvian tax".

From an analytical point of view, these changes clearly do not affect the main property of the Markov chain theory: there exists one unique stationary distribution and hence the independence of the final configurations on the initial ones is still valid. For $\alpha = 1$, the probability to move involves only the global function U. The stationary distribution can therefore be written as (see Grauwin *et al.*, 2009a):

$$\Pi_{PT,\alpha=1}(x) = \frac{e^{U(x)/T}}{\sum_{z \in X} e^{U(z)/T}} \tag{1.12}$$

It is then easy to figure out that the configurations obtained in the limit $T \to 0$ are those which maximize the collective utility.

For other values of α, it is possible to find an analytical expression of the stationary distribution in the context of bounded neighborhoods (see Grauwin *et al.*, 2009a, 2011). For continuous neighborhoods, an analytical approach is no longer possible. The reason can be stated quite simply: in the bounded neighborhood case, the information used to calculate the utility difference achieved by the moving agent (the initial and final neighborhood compositions) allows to calculate the difference in collective utility produced by the move. This is because the agent's initial neighbors share the same neighborhood as her, and their utility difference can therefore be calculated, the same being true for the final neighbors. Instead, in the continuous neighborhood case, the global utility difference depends on the neighbors of the neighbors of the moving agent. Indeed, the utility difference felt by the neighbors of the moving agent depends on their own neighbors,

most of which are not neighbors of the moving agent. Lacking the analytical approach, one needs to turn to simulations in order to investigate the effects of the introduction of partial coordination.

1.3.3 Simulations

We present on Figs. 1.7 and 1.8 snapshots of typical stationary configurations along with the corresponding values of s^* and U^* for different values of α and m, with $T = 0.1$. As previously stated, the case $\alpha = 0$ corresponds exactly to the without-coordination case that we already commented on in the previous section. With $\alpha = 1$, adding coordination to the $m = 0$ case shifts the random configuration to an ordered one: enhancing the utility level is possible only by achieving $s = 0.5$ in every location. In this case, even if a tax does not affect the value of the similarity index, it allows to enhance welfare by clustering vacancies, thus diminishing the number of agents that do not have $s = 0.5$ exactly. Obviously, for high values of m, due to the form of the agents' utility function, a high utility level is obtained whatever the tax level. Still, the similarity index is lower when coordination is introduced: changing α from 0.5 to 0.8 and then to 1 decreases s^*. For intermediate values of m, that have been identified as situations where equilibrium leads to harmful segregation in the without-coordination case, increasing α breaks segregation patterns.

Economic theory predicts that optimality (here a collective utility of 1) is obtained if a tax equal to the generated externalities is implemented. For $\alpha = 1$, this is what would be obtained in the limit $T = 0$. Here, the collective utility is slightly inferior to 1 because of the finite value of the noise ($T = 0.1$).[4] More interesting, our simulations show that a low level of taxation is able to significantly increase welfare: Fig. 1.8 shows that even a tax equal to one fifth of the generated externalities is enough to change the segregation level and to increase utility for $0.4 \leq m \leq 0.7$.

An explanation for this result relies on an informal comparison of individual's and her neighbors' utility variations. From the segregation viewpoint, harmful individual moves are those in which an agent changes neighborhood to follow her taste for majority, for instance leaving a 40%-60% neighborhood for a 60%-40% neighborhood. The utility gain allowed by these moves are rather low. Still, those moves drive neighborhoods out of mixed situations and create avalanches of further moves that will end in segregated states. When their neighborhood composition changes by one individual, the variation of utility of each affected agent is of the order of $2/H$ for agents in minority and $2(1-m)/H$

[4] Having a strictly positive temperature is however necessary to avoid blocked states.

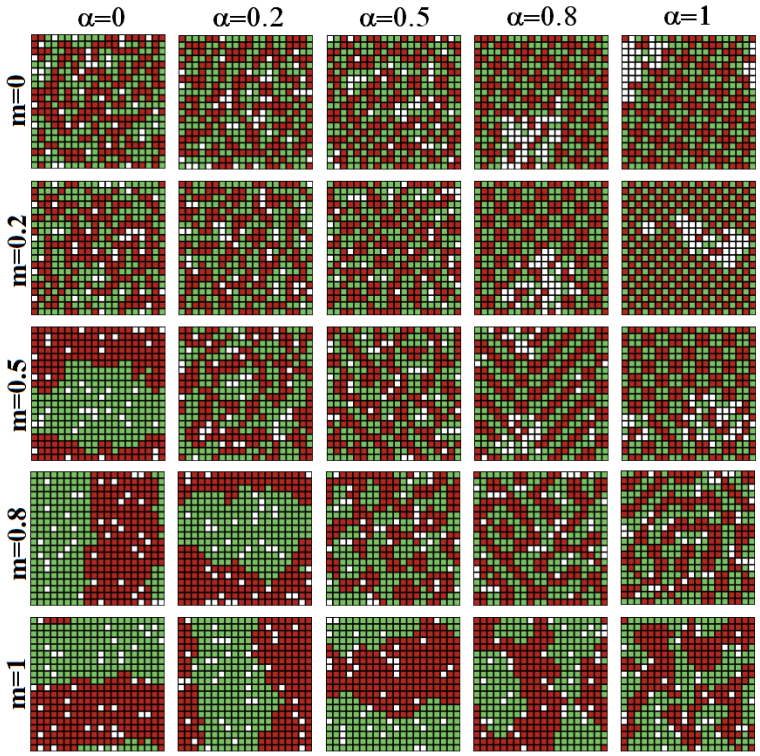

Figure 1.7: **Snapshots of typical stationary configurations** of the city for different values of α and m. The introduction of partial coordination can destabilize the highly segregated configurations and leads to configurations which present structured mixed patterns. $T = 0.1$ and $H = 8$.

for agents in majority in their neighborhood. As the utility gain for the agent is low compared to the externality affecting the other agents, internalizing even only a part of the externality with a tax on moves can be enough to hinder those individual moves that yield a small utility gain for the agent while increasing segregation trends and creating avalanches of segregating moves. In other words, the tax brakes the cumulative effects engendered by individual moves.

1.4 A local coordination by a feasible tax

1.4.1 Basic setup

A Pigouvian tax such as the one presented in the previous section supposes that the virtual central government has a precise knowledge of the neighborhood composition of each moving agent in order to tax all the moves, which is a rather utopian assumption. We propose in this section a new variation of the Schelling model incorporating a differ-

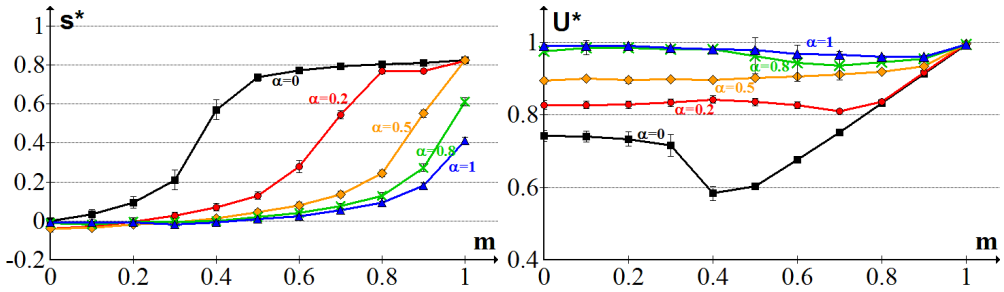

Figure 1.8: **Stationary mean values of the similarity index and of the collective utility** as a function of m, for different tax level α, with the partial coordination dynamic rule. The error bars give the standard deviation of the fluctuations of s^* and U^* once the system has reached its stationary phase. The mean value and standard deviation are computed over 10 periods of the stationary phases. $T = 0.1$ and $H = 8$.

ent tax rule, based on more realistic assumptions regarding the government intervention ability.

We define the *local district* of an agent as the H_g nearest cells surrounding him. We then introduce $s_i^{H_g}$, the fraction of similar agents present in the local district of agent i. The size $H_g \geq H$ of the district takes into account in a stylized way the central government's lack of precision in the knowledge of the agents' locations. It is also the scale on which the central government can act to limit unwanted phenomena such as segregation. We propose here a simple tax definition, which aims at preventing the emergence or maintenance of a dominant group in each district. The central government imposes a *tax* on an agent i, which is defined as:

$$r_i(s_i^{H_g}, \theta) = \begin{cases} \theta |s_i^{H_g} - 0.5| & \text{if } s_i^{H_g} > 0.5 \\ 0 & \text{otherwise} \end{cases} \tag{1.13}$$

where θ is a fixed parameter controlling the tax level. Note that the tax does not penalize agents which are in the minority group inside their own district. The tax is payed by every agent belonging to the majority in the district. The utility of an agent is then redefined as:

$$\hat{u}_m(s_i, s_i^{H_g}, \theta) = u_m(s_i) - r_i(s_i^{H_g}, \theta) \tag{1.14}$$

this new definition taking into account the penalty imposed by the tax.

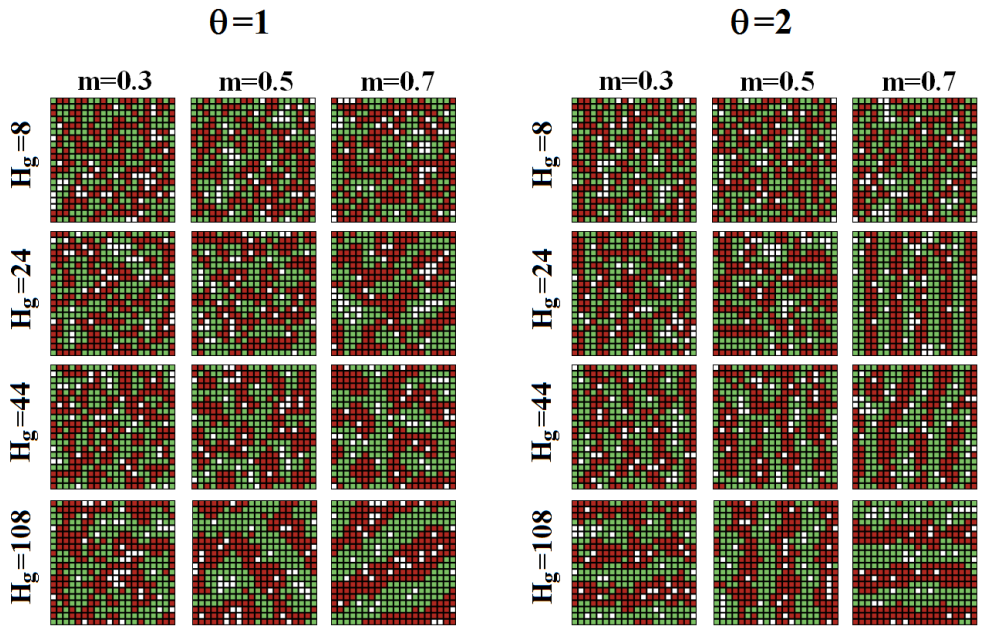

Figure 1.9: **Snapshots of typical stationary configurations** of the city for different values of H_g, θ and m. The introduction of tax proportional to the level of segregation at the district level can destabilize the highly segregated configurations and leads to configurations which present structured mixed patterns. $T = 0.1$ and $H = 8$.

1.4.2 Dynamic rule

As in the WC reference case, an agent decides to move according solely to the benefit in utility he would achieve if he was to move. The only difference here is that her utility incorporates the tax imposed by our virtual central government. We thus write the probability that a move happens as:

$$Pr\{move; FT\} = \frac{1}{1 + e^{-\Delta\hat{u}/T}} \tag{1.15}$$

where FT stands for "feasible tax".

From an analytical point of view, the properties of the WC model resulting from the Markov chain theory remain unchanged: there exists one unique stationary distribution and the final configurations do not depend on the initial ones.

Taking $\theta = 0$ cancels the tax mechanism and the FT rule comes back to the WC rule. For $\theta > 0$, we expect that a rise in θ implies a rise in the probability of having non-segregated configurations in the stationary states.

As to the influence of the size H_g of the districts, two limits can be looked upon:

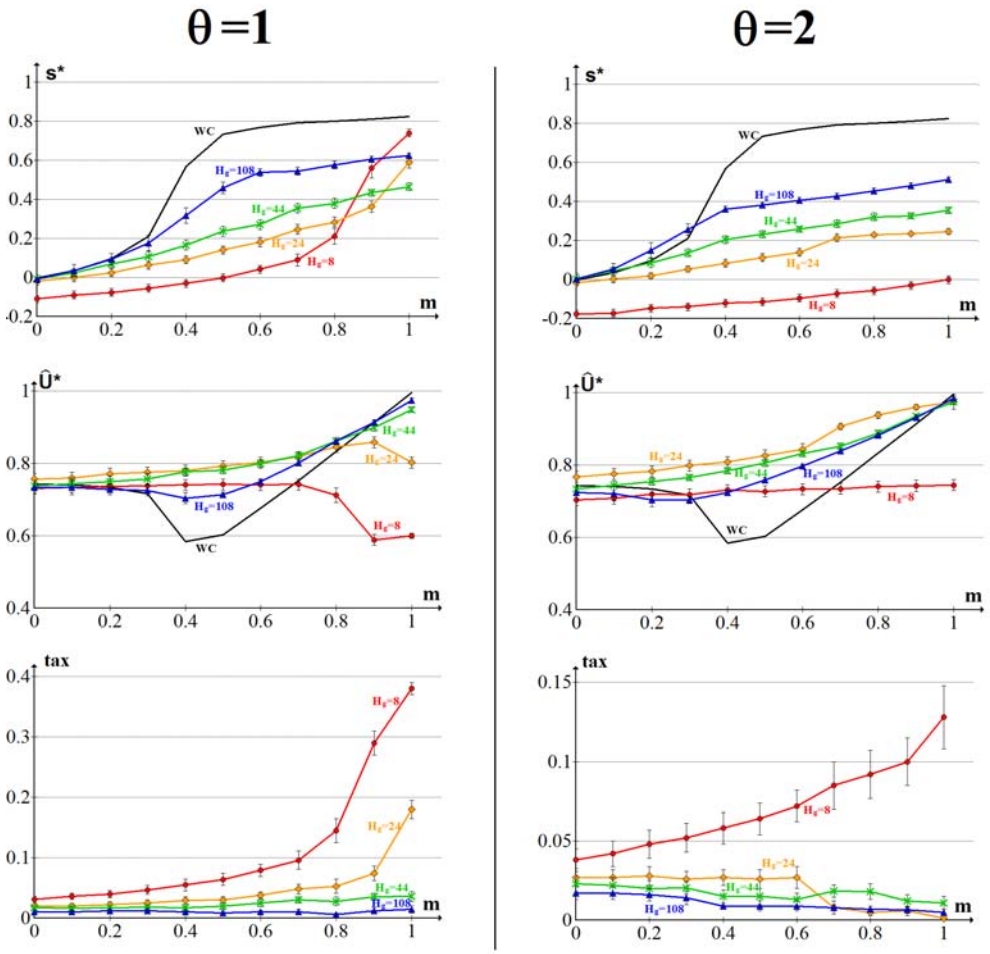

Figure 1.10: **Stationary mean values of the similarity index, collective utility and tax** as a function of m, for different district size H_g, with the feasible tax rule. The error bars give the standard deviation of the fluctuations of s^* and U^* once the system has reached its stationary phase. The mean value and standard deviation are computed over 20 periods of the stationary phases. $T = 0.1$, $H = 8$.

- For $H_g = H$, the fraction $s_i^{H_g}$ of similar agent in agent i's district can be identified with the fraction s_i of similar neighbors. It is straightforward to notice that an

agent utility $\hat{u}$ can thus be rewritten as:

$$\begin{aligned}
\hat{u}_m(s_i, s_i^{H_g}, \theta) &= u_m(s_i) - r_i(s_i, \theta) \\
&= \begin{cases} 2s_i - 0 & \text{if } s_i \leq 0.5 \\ 2 - m - 2(1-m)s_i - \theta|s_i - 0.5| & \text{if } s_i > 0.5 \end{cases} \\
&= \begin{cases} 2s_i & \text{if } s_i \leq 0.5 \\ 2 - (m - \theta/2) - 2(1 - (m - \theta/2))s_i & \text{if } s_i > 0.5 \end{cases} \\
&= u_{m-\theta/2}(s_i) \qquad (1.16)
\end{aligned}$$

Hence our tax mechanism can be interpreted in this case as a direct control by the central government of the asymmetric parameter m in the agents' utility function. According to the results obtained by simulation in the WC case (see Fig. 1.6), the asymmetric parameter must be inferior to $m_c \simeq 0.3$ in order to avoid segregation. The equivalence stated in Eq 1.16 thus allows us to predict that the minimal tax level necessary to break segregated patterns is:

$$\theta_c(m) = 2(m - m_c) \simeq 2m - 0.6 \qquad (1.17)$$

- For $H_g = N^2$, the district corresponds to the whole city, in which we supposed that the two groups shared the same number of agents. Hence, for all the agents $s^{H_g} = 0.5$ and $r = 0$. The tax mechanism has no effect and the outcomes are the same as in the WC reference case.

1.4.3 Simulations

We present on Fig. 1.9 some typical snapshots of stationary configurations obtained by simulating our feasible tax mechanism, while the different panels on Fig. 1.10 present the values of s^*, U^* and the mean tax paid by the agents. As previously stated, the WC case, which we take as reference, corresponds both to the $\theta = 0$ case or the $H_g = N^2 = 400$ case. Putting aside for the moment the $H_g = 8$ case, we observe on Fig. 1.10 that for a fixed value of m and starting with $H_g = 400$, decreasing H_g leads almost always to a decrease in the segregation index, a decrease in the mean tax and an increase in the mean utility of the agents. The raise in utility compared with the WC case is particularly strong for intermediate values of m.

We thus show that the chosen tax, even if it differs from the Pigouvian tax, is able to lessen the segregation mechanism. This tax is governed by two parameters. θ determines the impact of the deviation from having a majority status in the neighborhood; the

higher theta, the stronger the incentive to form mixed neighborhood, which explains the impacts found in the simulations. The impact of m seen in the simulations is straightforward: it determines the strength of the preference for the majority status and therefore the segregation level; it also impacts the average tax, as having a strong preference for the majority status leads agents to move to segregated neighborhoods, even if they have to pay the tax. More interesting, the simulation results show the effect of the local neighborhood size, that is, the consequence of the government lack of precision regarding the measure of segregation and give some hints as to the sufficient values of H_g to avoid segregation. The smallest H_g, the less segregation there will be. However, even with $H_g = 44$, that is five times larger than the neighborhood considered by the agents, the utility level enjoyed by the agents for intermediate values of m is improved compared to the WC case.

The snapshots of the final configurations displayed on Fig. 1.9 give more details as to the city patterns. They show clearly that the tax mechanism produces organized patterns which are more regular for larger values of m and higher values of θ. The typical width of these patterns increases with H_g. These patterns ensure that most of the agents have an equal or higher number of agents of the other group than agents of their own group, inside their district (ie $s^{H_g} \leq 1/2$). In the meantime, the segregation index is always positive, meaning that on average an agent still has more similar neighbors than dissimilar ones.

In the $H_g = H = 8$ case, no clear pattern is visible on Fig. 1.9, but $s^* < 0$, that is, an agent has on average more dislike neighbors than similar ones. The case $H_g = H$ is special because there is a direct competition between the 'standard effect' of Schelling model which favours values of s^H greater than 0.5 and the tax mechanism which favours values of s^{H_g} below 0.5. The results obtained for $H_g = H = 8$ corresponds to what we stated above based on the transformed utility function: for $m \geq m_c + \theta/2$, the segregation patterns that exist in the WC case vanish.

1.5 A local coordination by voting

1.5.1 Basic setup

We define the *co-proprietors* of an agent as the agents living on the h nearest cells surrounding him. Here, h is a fixed integer which verifies $h \leq H$. *Co-proprietors* represent in a stylized way next-door neighbors or the people living in the same residential building whereas the *neighbors* represent the people living in the same street or in the same district. Examples of possible forms of neighborhoods and co-properties are shown

below in Fig 1.11.

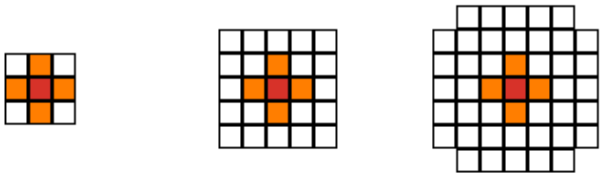

Figure 1.11: **Examples of neighborhood and co-property used in our simulations.** A red agent is located on the central cell. His co-property corresponds to the orange cells and his neighborhood to the orange and white cells. From the left to the right, $(H = 8, h = 4)$, $(H = 24, h = 4)$ and $(H = 44, h = 4)$.

We introduce local coordination by taking into account the potential change of utility of the co-proprietors of the vacant cell considered by the potential mover. In the following, we will denote by $\mathcal{C}$ this set of agents. It seems more logical to introduce local coordination through the potentially new co-proprietors (who have you take an admission exam) than through the current co-proprietors (that you can quit on your free will). For mathematical convenience, we will suppose that the probability that the move happens can be computed as the product of the probability that the potential mover would like to move and the probability that the agents of $\mathcal{C}$ accept him: the two events are independent. We propose here one dynamic rule to counterbalance the wish of the potential mover by the opinion of his potentially new co-proprietors. Other mechanisms can of course be imagined. The 'local' nature of the implied coordination comes from the fact that only a fraction of the agents who might be affected by the potential move are consulted: they don't include previous neighbors and not all the new ones.

1.5.2 Dynamic rule

The simplest local coordination rule is that the potential mover needs the majority of the co-proprietors to endorse his moving in. Let Δu_i be the variation of utility of the co-proprietor $i \in \mathcal{C}$ if the move was to take place. We write the probability that the co-proprietor i votes 'for' the move:

$$Pr\{i, \text{'for'}\} = \frac{1}{1 + e^{-\Delta u_i / T_2}} \tag{1.18}$$

The logit form of this acceptance probability can be justified in the same way as the logit for the moving agent. The parameter $T_2 > 0$ can then be interpreted as the

amplitude of a noise that represents in a stylized way the preferences of the co-proprietors over any characteristics of the potential mover other than and not correlated to the group he belongs to (marital status, number of children, profession, religion, friendship, etc...). Since it is related to other factors, the noise affecting the co-proprietors is different from the noise affecting a moving agent. We moreover argue that since a co-proprietor deciding whether to accept or not a new neighbor is qualitatively subject to less characteristics other than the group membership (*ie* to less noise) than a moving agent, it is realistic to suppose $T_2 \leq T$.

The move takes place with a probability:

$$Pr\{move; LC\} = \frac{1}{1 + e^{-\Delta u/T}} Y\left(\left\{\frac{1}{1 + e^{-\Delta u_i/T_2}}\right\}_{i \in \mathcal{C}}\right) \tag{1.19}$$

where $Y = 1$, $1/2$ or 0 if respectively more than half, exactly half or less than half of the co-proprietors vote 'for' the move, and where LC stands for "Local Coordination".

From an analytical point of view, the introduction of the vote of the co-proprietors does not change the main property of our system: it can always be described as a Markov chain, and since $T_2 > 0$, every move still has a non-zero probability to happen. This ensures the existence of one unique stationary distribution and hence the independence of the stationary states with regards to the initial starting state.

1.5.3 Simulations

We first present on Fig. 1.12 a typical evolution of the city in the case where the agents move according to the local coordination rule by consulting before each move $h = 4$ out of $H = 8$ of their potentially new co-proprietors. We limit our investigations by fixing the noise level T to 0.1. The level of noise attached to the co-proprietors is fixed to $T_2 = 0.1$ in order to be comparable to the chosen value of T and the parameter m of the utility function is fixed to 0.5. Starting from a highly segregated configuration, we observe the disaggregation of the two large homogeneous areas into a much less segregated configuration presenting more local patterns of segregation. This first simulation hence shows that the introduction of a bit of local coordination can be sufficient to break undesired segregated patterns and therefore, *a fortiori*, to prevent segregation to appear starting from a mixed configuration.

The explanation of why local coordination works is quite simple. It corrects the default of the WC mechanism by rendering highly segregated configurations less stable than before since if a red agent goes by mistake into the green area, he will, compared

to the WC case, encourage a second red agent to join him by his vote. Hence the formation of nuclei is encouraged by local coordination. A second kind of mechanism to get out of segregated patterns is the advance of the frontier zone. On the opposite, the locally mixed patterns are more stable. Indeed, once an integrated pattern is reached, the co-proprietors tend to prevent the moves which would increase local segregation.

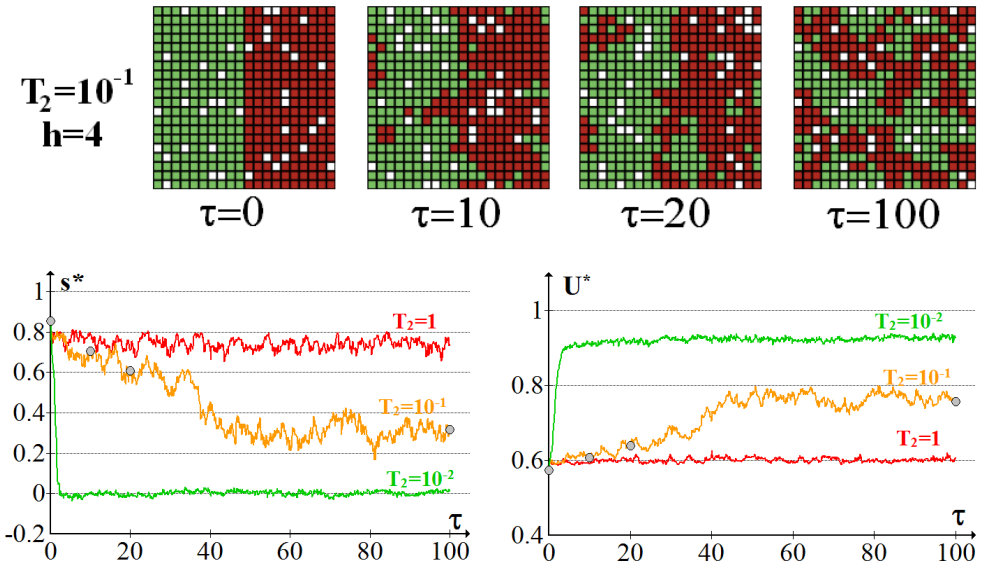

Figure 1.12: The introduction of local coordination destabilizes the highly segregated configurations and leads to configurations which present locally mixed patterns. **Top panel**: some snapshots of the evolution of the city for $T_2 = 0.1$ and $h = 4$. **Bottom panel**: evolution with τ of the index of similarity and of the collective utility, the grey dots on the $T_2 = 0.1$ curves corresponding to the snapshots of the top panel. $T = 0.1$ and $m = 0.5$.

The influence of the parameters T_2 and m can be observed on Figs. 1.12 to 1.15. For high values of T_2, the co-proprietors decision whether to accept or not the moving agent is purely random and the local coordination mechanism has no impact on the dynamics. Hence on Fig. 1.15 for $T_2 \gg 1$, the values of s^* and U^* are similar to their values in the WC case. On the contrary, for lower values of T_2, the local coordination mechanism allows to break the segregation patterns, leading to mixed configurations presenting locally ordered patterns. It happens the more rapidly, the lower the T_2 value (see the bottom panel of Fig. 1.12). Notice on Fig 1.13 that the impact of local coordination in terms of welfare is more important for intermediate values of m (between 0.2 and 0.8), precisely the values for which the WC is the most deficient when compared to random allocations.

The results displayed on the right panel of Fig 1.13 and on Fig 1.15 correspond to

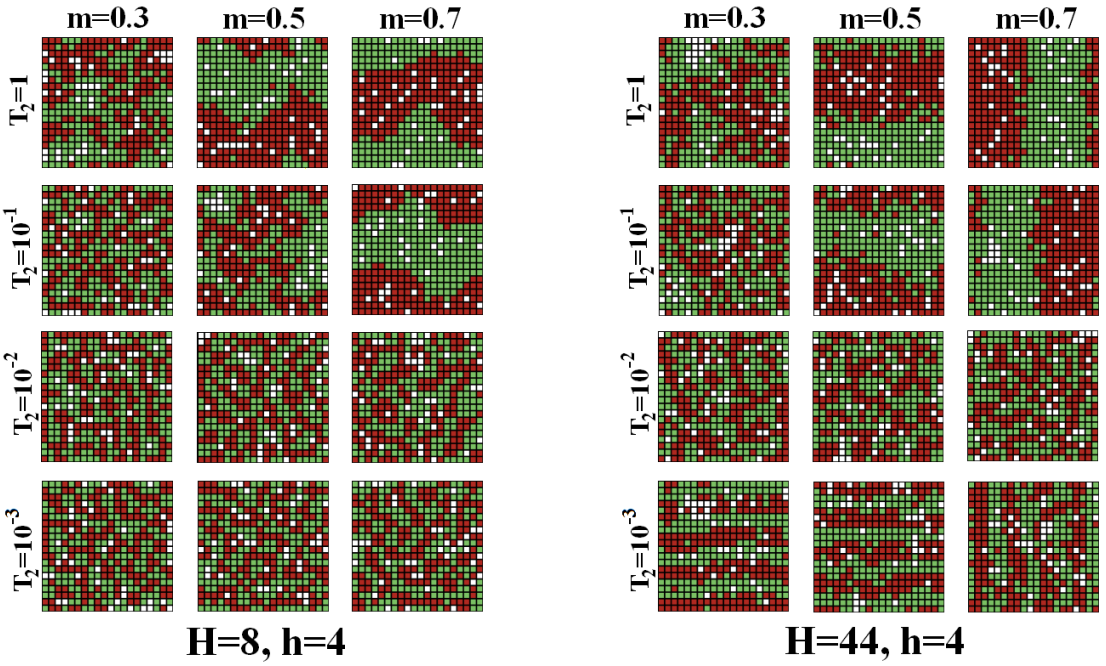

Figure 1.13: **Snapshots of typical stationary configurations obtained with the "qualified vote" dynamic rule** for different values of m and T_2. **Left panel**: the $h = 4$ new co-proprietors of a moving agent out of the $2H = 16$ agents he potentially affects by moving are consulted. **Right panel**: the $h = 4$ new co-proprietors of a moving agent out of the $2H = 88$ agents he affects by moving are consulted. Noise level: $T = 0.1$.

simulations where the moving agents consult $h = 4$ new co-proprietors out of $H = 44$ of their potentially neighbors. Since the move of an agent can affect at most $2H$ agents (neighbors in the departure and arrival locations), the LC mechanism simulated here is only taking account of $h/2H \simeq 5\%$ of the agents affected by the externalities generated by the moving agents.[5]

For $T_2 \geq 10^{-2}$, the results are comparable to the previous ones: even if the size of the co-property is relatively less important, the LC mechanism still allows to break the segregated patterns and lead to locally mixed and ordered patterns. The size H of an agent's neighborhood being greater than previously, the typical size of these ordered patterns is also greater.

For $T_2 = 10^{-3}$ however, one can observe that U^* is lower than in the $T_2 = 10^{-2}$ case (Fig 1.13), and that the normalized similarity s^* is negative for $m \leq 0.9$, meaning that the agents have on average less similar neighbors than dissimilar ones. The correspond-

[5]However, since some of the potentially new neighbors share almost the same neighborhood than voting co-proprietors, there are spatial correlations between the H potentially new neighbors. Hence the LC mechanism takes effectively into account more than $h/2H$ of the affected agents.

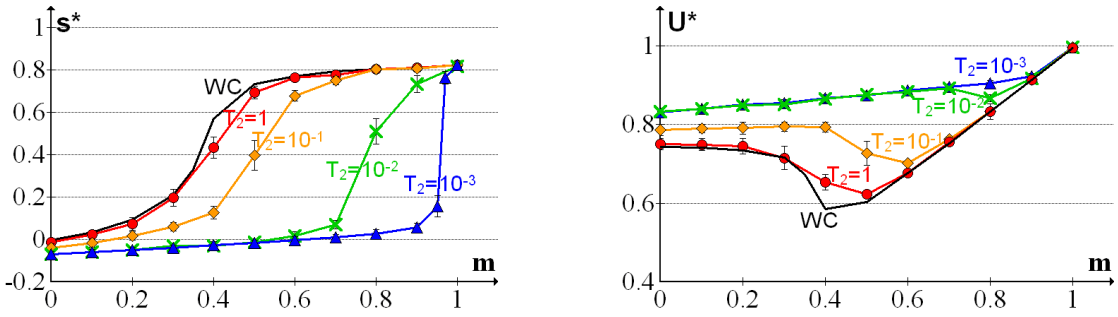

Figure 1.14: **Stationary mean values of the similarity index and of the collective utility** as a function of m, for different level of noise T_2, with the qualified vote dynamic rule. The error bars give the standard deviation of the fluctuations of s^* and U^* once the system has reached its stationary phase. The mean value and standard deviation are computed over 10 periods of the stationary phases. The plots corresponds to a neighborhood size $H = 8$, a co-properties size $h = 4$ and a noise level $T = 0.1$. The black curve corresponds to 'Without Coordination' simulations that have been performed using the same parameters.

ing snapshots of the stationary configurations on Fig 1.15 show that the system ends in stripe-like globally ordered states. This result can be understood through the notion of externalities and lack of coordination. Indeed, when $h/H \ll 1$, we can separate the voting co-proprietors whose neighborhood is close to the vacant cell envisaged by the moving agent and the neighbors living on a further ring. The respective neighborhoods of these two kind of neighbors are not spatially correlated, which means that the interest of these two kind of neighbors are clearly different. There are hence three kind of agents at play: the moving one, the inner ring of new neighbors (to which the voting agents belong) and all the other affected agents (the outer ring of new neighbors and the former neighbors). For finite values of T and $T_2 \to 0$, the interest of the second group is in fact the sole taken into account in the LC mechanism. The interest of the potential movers is therefore practically not taken into account and some utility maximizing moves become impossible.

1.6 Conclusion

Schelling's model is characterized by the paradoxical result that, while the dynamics is governed by agents moving to improve their own utility, their moves lead to highly segregated configurations in which most of the agents are far from being fully satisfied. Actually, as already argued by Zhang (2004b) and Pancs & Vriend (2007), individual preferences for integrated environments may lead to segregated configurations because location choice by an agent affects her neighbors' utility. As stated by Zhang (2004b):

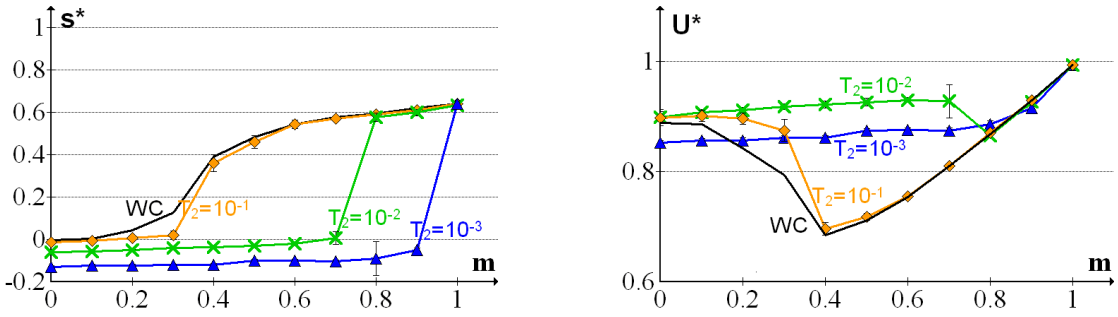

Figure 1.15: **Stationary mean values of the similarity index and of the collective utility** as a function of m, for different level of noise T_2, with the qualified vote dynamic rule. The error bars give the standard deviation of the fluctuations of s^* and U^* once the system has reached its stationary phase. The mean value and standard deviation are computed over 10 periods of the stationary phases. The plots correspond to a neighborhood size $H = 44$, a co-properties size $h = 4$ and a noise level $T = 0.1$. The black curve corresponds to 'Without Coordination' simulations that have been performed using the same parameters.

"*although nobody likes complete segregation, the residential pattern is very stable. Only moving across the color line by a considerable number of agents could disturb the segregation equilibrium, but nobody has incentive to do so because it causes a loss of [individual] utility. [...] Segregation is stable not because people like it, but because any individual who wants to change the situation unilaterally will have to go across the color line, which may not be the desirable thing to do from the individual's perspective. The failure of the system to escape complete segregation is similar to the phenomenon of "coordination failure" studied by economists in many other contexts. It is the agents' inability to move simultaneously that make them stuck in a situation nobody likes...*"

We presented extended versions of Schelling-type models incorporating different kinds of coordination between the agents and examined stationary configurations based on simulations. We showed in a first mechanism that introducing partial coordination through a tax on the externality generated by individual moves is sufficient to break the gap between the agents' micro-motives and the emergent macro-behavior and therefore to break undesired segregative patterns. Moreover, we showed that it is not necessary for the tax to be equal to the externality to reduce segregation significantly. A tax equivalent to one fifth of the externality might well be sufficient. In a second mechanism, we assumed that the government is able to tax agents based on their majority status in the neighborhood. This again is a mechanism that allows to decrease segregation. In a third mechanism, we introduced local coordination through a voting mechanism which involves only individual decisions and, unlike the tax mechanisms, does not require the intervention of a benevolent central authority. This model, remaining in the

"individual decision" spirit of Schelling model, is shown to be sufficient to reach stationary configurations with a significantly higher collective utility than the no-coordination case. We hope that such an individualistic based coordination model can be seen as a valuable alternative to the coordination models presented in the literature, mostly based on affirmative action policies (Dokumaci & Sandholm, 2007). We also hope that these results can shed light on other location externality issues, such as those encountered in economic geography models / agglomeration phenomena. Indeed, we have shown that due to cumulative mechanisms occurring in such situations, the tax level that allows the system to get or stay close to the social optimum are far lower than the level of the pigouvian tax.

Of course, it has to be noticed that in the first and third coordination mechanisms we proposed, the introduction of coordination cannot break the segregative pattern without the agents having a certain preference for mixed neighborhood. Coordination is only a way to reinforce on the large scale the wishes of the agents on the local scale: if they are intolerant, segregation will occur, if they are tolerant, integration may occur.

One of the most interesting tracks for future work would certainly be to explore more thoroughly what drives the agents' preferences regarding their neighbors' attributes. Indeed, in real life preferences regarding mixed environments seem influenced by individual past experiences as well as social norms. Such an analysis could be done by coupling Schelling's model with another model describing the dynamic evolution of preferences. For example, one could introduce heterogeneity in the agents' preferences and allow these preferences to evolve over time, taking into account the individual's past experiences and its neighbors' preferences.

Chapter 2

Dynamic Models of Residential Segregation: An Analytic Solution

2.1 Introduction

Ethnic and immigrant residential segregation is a striking feature of most Western cities. Extensive views of segregation patterns in the U.S. have been provided recently by Cutler *et al.* (2008), Iceland & Scopilliti (2008) and Reardon *et al.* (2008). Cutler *et al.* (2008) examine a range of potential determinants of immigrant segregation, including cultural traits of immigrants and nativist sentiment among U.S. natives. Card *et al.* (2008) results on racial segregation for the 1970-2000 period show evidence of tipping-like behaviors: the rise of the minority share in a neighborhood above a certain threshold leads to a further decrease in the white population. This analysis is one of the first providing clear empirical evidence of non linear dynamic aggregate behaviors, as those predicted by social interaction models. According to these results, whites' utility in a neighborhood seems to exhibit a sharp decrease beyond a certain minority share. A direct link between white attitudes toward minority members and aggregate configurations, as measured by the location of the tipping point, is also shown by the authors. The theoretical relationship between individual preferences and aggregate configurations has however not been fully explored to date.

An early contribution was provided by Schelling, who proposed a model aiming at formalizing the aggregate consequences of individual preferences regarding the social environment (Schelling, 1969, 1971, 1978). The two basic ingredients of Schelling's "Dy-

namic models of segregation" (1971) are an individual utility function that determines entirely the level of satisfaction enjoyed by an agent in a location and a dynamical rule that drives agents' location changes and therefore the evolution of the city configuration. Using an inductive approach, Schelling showed that if the preferences considered are such that an environment of more than 50% of own-group agents is highly preferred to a less than 50% of own-group environment, then the equilibrium configuration exhibits high levels of segregation, although there is no preference for segregation *per se*. Schelling 1971 paper is widely known thanks to this apparently paradoxical effect: mild individual preferences for own-group neighbors lead to a complete segregation at the global scale. However, a moment of reflection suffices to understand that, given the highly asymmetrical utility function used in this model, it could hardly lead to an integrated environment. Yet, later research showed that even a peaked utility function, that is, a function achieving its maximum for a perfectly mixed environment, can lead to a fully segregated equilibrium as soon as this function is asymmetric (Zhang, 2004b; Pancs & Vriend, 2007; Barr & Tassier, 2008).

Criticizing the realism of Schelling model is straightforward: it ignores institutional causes of segregation, income effects, or cities' social structure. Anyway, the model has become a favorite example, in the modelling of social systems, of the unintended macro-level consequences of individual behavior, and Schelling 1971 paper is his most widely cited publication (more than 460 as of 2010, June 10th). After years of relatively low citation records, this paper accrues since 2003 around 40 citations per year, showing the renewed interest in Schelling model. It is interesting to notice that citations arise from widely different fields: economics and sociology represent the two strongest contributors (40% of the total number of citations) but computer science, mathematics and physics gather 24% of the citations. This substantial scientific activity has lead to new insights: the interpretation of the emergence of segregation patterns as the result of a coordination problem (Zhang, 2004a,b); a physical analogue of Schelling's model (Vinkovic and Kirman, 2006); the robustness of Schelling's results with respect to different definitions of individual utilities and/or environment (Pancs & Vriend, 2007; Fagiolo *et al.*, 2007); the impact of heterogeneous agents and public policies (O'Sullivan, 2009); the exploration of tipping behaviors (Zhang, 2010).[1] Most of this work relies on agent-based simulations.

Attempts to solve Schelling model analytically include Dokumaci & Sandholm (2007); Mobius & Rosenblat (2000); Pollicott & Weiss (2001); Pancs & Vriend (2007); Zhang (2004a,b, 2010). Zhang (2004a,b, 2010)'s contributions represent to date the closest

[1] See Clark & Fossett (2008) for a literature review.

achievement in this direction. Zhang proposes variations of Schelling model which he analyzes formally using the concept of potential function developed in evolutionary game theory. Zhang (2004a) considers a model with vacant cells and linear utility functions. Zhang (2004b, 2010) use an asymmetrically peaked utility function in a model with no vacant cells. The latter choice raises the issue of individual rationality, as it is assumed that, for individual moves to occur, two agents have to coordinate and agree on exchanging locations. As such, this analysis departs from Schelling original framework. Furthermore, these three contributions only cover two specific utility functions.

In the present chapter, we build on Zhang (2004a,b, 2010). We place Schelling model in the context of evolutionary game theory aiming at characterizing the equilibrium segregation level by means of a potential function. Compared to Zhang (2004b, 2010), we consider bounded neighborhoods, *ie* blocks where all the agents share the same neighbors. This permits us to formalize the externalities for a relatively broad range of utility functions. In this context, we can predict the global pattern emerging from different utility functions, which, to the best of our knowledge, was never done before. We show that a potential function of the model exists if and only if the utility functions are such that the externalities generated by one type of agents are symmetric to the ones generated by the other type of agents. Under this condition, a general form of the potential function is found. The main property of this potential function is that it reflects both the macro and micro scale. On one hand, this aggregate function only depends on the number of agents of each type in each block. On the other hand, it keeps tracks of the individual level since it corresponds to a sum of individual utility changes generated by individual moves.

We use this potential function to characterize the segregation level of the stationary configurations of the model for different utility functions, representing different degrees of preference for mixed environments. We examine successively (i) linear utility functions, with a continuous preference for segregated environments, (ii) Schelling's original utility function in which there is a mild preference for a mixed environment and (iii) asymmetrically peaked utility functions, according to which agents clearly exhibit a preference for a mixed environment. We show that there is no divergence between individual moves and social welfare with increasing linear utility functions, although segregation prevails in stationary configurations. We also show that even with the strongest preference for mixed environments - in the asymmetrically peaked utility function case - the model yields segregated stationary configurations. This case is also the one in which the divergence between the stationary segregation level and the optimal segregation level is the highest. These results complement those obtained by Zhang (2004b, 2010) based on

more restrictive hypotheses. Compared to this previous work, our analysis presents the advantage of remaining in Schelling's spirit by including vacant cells. While the condition on the utility functions for a potential function to exist with vacant cells restrict somewhat the range of utility functions that can be considered, all utility functions are valid in the limit case with no vacant cells.

In summary, our work provides a very general solution to Schelling's model with bounded neighborhoods, that encompasses previous work on this model and that paves the way to the analysis of many structures of preferences, for instance those based on empirical findings concerning racial preferences. In addition, a few simulation results are shown for illustrative purpose.

This chapter is organized as follows. The model features are presented in section 2.2. Section 2.3 defines the potential function concept and states our main result. The potential function is then used in section 2.4 to study the stationary configurations obtained for three different utility functions. In one of these cases, we are able to build a connection with a commonly used segregation measure. In section 2.5, we demonstrate the supplementary result that a potential function exists with continuous neighborhoods if and only if the utility functions are linear and discuss the differences between the bounded and continuous neighborhood cases. We also draw a parallel with general concepts of coalitional games, consider preferences of the agents for local amenities and analyse a case where a taxation is introduced against segregative behaviors.

2.2 A general dynamic model of segregation

2.2.1 The city and the agents

Our artificial city is a two-dimensional NxN square lattice with periodic boundary conditions, *ie* a torus containing N^2 cells. Each cell corresponds to a dwelling unit, all of equal quality. We suppose that a certain characteristic divides the population of this city into two groups of households that we will refer to as red and green agents. Each location may thus be occupied by a red agent, a green agent, or may be vacant. We denote by N_V the number of vacant cells, and by N_R and N_G the number of respectively red and green agents. The parameter N thus controls the size of the city, the parameter $v = N_V/N^2$ its vacancy rate, and the fraction $n_R = N_R/(N_R + N_G)$ its composition.

We define a *state* x of the city as a N^2-vector, each element of this vector labelling a cell of the NxN lattice. Each state x thus represents a specific configuration of the city. We note X the set of all possible configurations, the demographic parameters (N, v, n_R) being fixed.

2.2.2 Neighborhoods

Since Schelling (1969)'s work, two ways of conceiving the neighborhood of an agent have been developed and used in analytical and simulation models.

Bounded neighborhood models (Fig 2.1a) describe cities divided into geographical units within which all agents are connected. The neighborhood of an agent is thus composed entirely and exclusively of the locations present in the same geographical unit as his own. In the following, when we refer to a bounded neighborhood model, we will implicitly assume that the city is divided in a set $\mathcal{Q}$ of blocks, each of which contains $H+1$ locations, where H is a fixed integer that corresponds to the number of locations in an agent's neighborhood (hence, the relation $|\mathcal{Q}|(H+1) = N^2$ must hold). Obviously, the description of the city as a lattice with periodic boundary conditions is unnecessary in this case. Note that since some locations remain empty, the size H of the neighborhood of an agent can also be interpreted as the maximum number of neighbors an agent can have. For a given configuration $x \in X$ of the city, we denote by $R_q(x)$ and $G_q(x)$ the number of red and green agents that live inside the block $q \in \mathcal{Q}$. Taking into account that some locations of each block may remain empty, the $\{R_q\}$ and the $\{G_q\}$ must thus verify :

$$\sum_q R_q = N_R \tag{2.1}$$

$$\sum_q G_q = N_G \tag{2.2}$$

$$(R_q, G_q) \in E_{H+1} \equiv \{(R,G), 0 \leq R+G \leq H+1\} \tag{2.3}$$

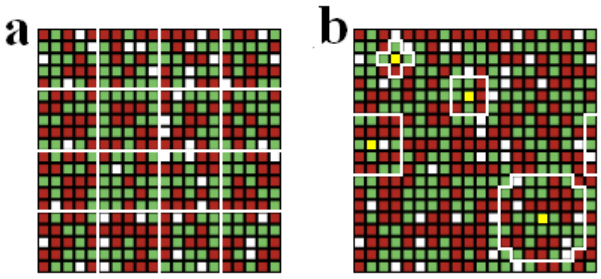

Figure 2.1: **Different forms of neighborhood.** Red, green and white squares denote respectively red agents, green agents and vacant cells. **a.** Example of a bounded neighborhood in which the city is divided in square blocks containing $H+1 = 25$ cells/locations; **b.** In the case of a continuous neighborhood description, the neighborhood of an agent corresponds to his H nearest cells/locations. Around the agents marked in yellow, we enlightened by the white frontiers a $H = 4$, a $H = 8$, a $H = 24$ and a $H = 44$ continuous neighborhood. [If you printed this document in black and white, the red and green squares should appear respectively in dark grey and soft grey.]

Continuous neighborhood models (Fig 2.1b) describe cities where the neighborhoods do not correspond to a zoning at the city level, but are centered on the local perception of each agent. In a continuous neighborhood description, one assumes that the neighborhood of an agent is composed of the H nearest locations surrounding him. The $H = 4$ "Von Neumann neighborhood" and the $H = 8$ "Moore neighborhood" that are displayed among other examples on Fig 2.1.b are the most commonly used in agent-based computational models.

As will be demonstrated in section 2.5.1, a potential function exists in the continuous neighborhood case only with very specific utility functions. In the following, we place ourselves, unless otherwise mentioned, in a bounded neighborhood case that allows us to consider a relatively broad range of utility functions. There is actually no argument in favor of bounded or continuous neighborhoods as far as the realism of the assumptions is concerned. Bounded neighborhoods can be thought of as reproducing the effects of the administrative divisions of real cities such as census areas or school districts. Still, bounded neighorhoods present the drawback that the spatial arrangement of the neighborhoods is not taken into account: each neighborhood is considered independently without any connection to other neighborhoods. This is the checkerboard issue, well known in the literature on residential segregation measures. We present in section 2.5.1 some simulation results illustrating the impact of the neighborhood description on the forms of segregation at the city scale and a formal analysis showing how connections between neighborhoods, that make agents close to neighborhoods boundary also consider the composition of the contiguous neighborhood, can be introduced in our framework.

2.2.3 Agent's utility function

Each agent has a utility level which depends only on his neighborhood composition. Let us consider an agent whose neighborhood is composed of R red agents, G green agents and V vacant cells. Since $R + G + V = H$, one needs two independent parameters to describe the composition of the agent's neighborhood. In all generality, we can thus write the utility of an agent for example as a function of R and G. Like most models of the literature, we assume for simplification that agents of a same group share the same utility function.[2]

[2]See O'Sullivan (2009) for a treatment with heterogeneous agents.

Without any loss of generality, we write the utility of an agent as:

$$u = u_R(R, G) \quad \text{for a red agent with } R \text{ red and } G \text{ green neighbors,}$$
$$u = u_G(R, G) \quad \text{for a green agent with } R \text{ red and } G \text{ green neighbors.}$$

The utility of an agent is thus a function of $E_H \to \mathbb{R}$. More specifically in the case of a bounded neighborhood description, we will have :

$$u = u_R(R_q - 1, G_q) \quad \text{for a red agent living in block } q,$$
$$u = u_G(R_q, G_q - 1) \quad \text{for a green agent living in block } q.$$

In order to facilitate the comparison of different utility functions, utility in the examples presented below is such that a zero utility level denotes a complete dissatisfaction of the agent and a utility of one denotes complete satisfaction.

We also introduce a notation in order to characterize the level of utility on the global (city) scale:

$$U(x) = \sum_k u_k \tag{2.4}$$

where u_k is the utility of agent k and $U(x)$ denote the collective utility of a configuration x.

2.2.4 A behavioral rule: the logit dynamical rule

The core of dynamic segregation models is that agents are given opportunities to move to increase their individual utility. Once the static description of the model is specified, one must add a dynamic rule that governs these moves. In the following, the city configuration evolves according to an iterative process. At the first iteration, an initial configuration is randomly chosen. At each iteration, one agent and one vacant cell are picked at random.[3] The picked agent then chooses to move to that vacant cell with a probability $Pr\{move\}$ that depends on the utility gain Δu he would achieve if he was to move, as follows:

$$Pr\{move\} = \frac{1}{1 + e^{-\Delta u/T}} \tag{2.5}$$

where $T > 0$ is a fixed parameter.

[3]Instead of assuming, as Schelling did, that the agents move to the nearest satisfactory position (the idea being that the cost of moving increases with distance), we suppose here that the distance between the current and envisaged locations of an agent does not intervene in his decision whether to move or not.

Eq. 2.5 represents a logit choice function as developed in McFadden (1974) as the outcome of a random-utility model. In such a model, it may happen that an agent takes a utility-decreasing move, either because he is making a mistake or because of a lack of information. The scalar T is used to determine the relative importance of the random part with respect to the deterministic part of the random utility function. The probability for an agent to take a utility-decreasing move drops down as $T \to 0$ and the described rule thus converges to the non-strict best response rule. For any finite $T > 0$, the agents choose non-best replies with a non-zero probability, but actions that yield smaller payoffs are chosen with smaller probability. This kind of perturbed best-response dynamics has been developed in Young (1998) in the context of evolutionary games. Here, neighborhood's composition is supposed to be the main determinant of agents' actions : we restrict our analysis to the case of low values of T.

Compared to a best-response behavioral rule (where the agents move if and only if they strictly improve their utility), the logit rule allows for some fluidity in the model in the sense that blocked states are avoided (Vinkovic and Kirman, 2006). Besides, it provides a strong analytical framework to Schelling model. Obviously, it implies that the probability that the state at the t^{th} iteration x^t is equal to a given state x only depends on the state at the previous iteration x^{t-1}:

$$\Pr(x^t = x | x^{t-1}, \ldots, x^1, x^0) = \Pr(x^t = x | x^{t-1}) \tag{2.6}$$

The dynamic rule thus yields a finite Markov process.

It is then easy to figure out that the Markov chain describing our system is irreducible (since $T > 0$ each imaginable move has a non-zero probability to happen and it is thus possible to get to any state from any state), aperiodic (given any state x and any integer k, there is a non-zero probability that we return to state x in a multiple of k iterations) and recurrent (given that we start in state x, the probability that we will never return to x is 0). These three properties ensure that the probability to observe any state x after t iterations starting from a state y converges toward a fixed limit independent of the starting state y as $t \to \infty$.

In other words, for each set of parameters and dynamic rule, there exists a stationary distribution

$$\Pi : x \in X \to \Pi(x) \in [0,\, 1] \;,\; \sum_{x \in X} \Pi(x) = 1 \tag{2.7}$$

which gives the probability with which each state x will be observed in the long run.

Clearly, for $T \to \infty$, the randomness introduced in the dynamical rule prevails and

the stationary distribution is just a constant. Similarly, for any finite $T > 0$, our dynamical system (the city) evolves toward an attractor composed of a subset A of X. It follows that any measure $\mathcal{M}$ - such as the global utility U - performed on the state space X will in the long run fluctuate around a mean value $\mathcal{M}_\infty = \sum_{x \in A} \Pi(x)\mathcal{M}(x)$. These mean values may depend on the intensity of the noise T, but the amplitude of the fluctuations decreases as $T \to 0$.

In the following, we refer to two states x and y as *immediately communicating states* (ICS) if we can switch from state x to state y by moving one single agent. We also note $\Delta_{xy}u$ the variation of utility of this agent induced by this particular move and P_{xy}^T the probability to be in state y at a given iteration if the system was in state x at the previous iteration. According to the dynamic rule presented above, one has:

$$P_{xy}^T = \gamma(1 + e^{-\Delta_{xy}u/T})^{-1} \quad \text{if } x \text{ and } y \text{ are ICS} \tag{2.8}$$

$$P_{xy}^T = 0 \quad \text{if } x \text{ and } y \text{ are not ICS} \tag{2.9}$$

where the parameter $\gamma = 1/(N_V(N_R + N_G)) = 1/(v(1-v)N^4)$ takes into account the probability to pick the right agent and the right vacant cell that allow to pass from x to y. P^T thus corresponds to the probability transition matrix for a fixed T and the stationary distribution Π is by definition the unique normalized function defined on X that verifies for all $x \in X$:

$$\sum_y P_{yx}^T \Pi(y) = \Pi(x) \tag{2.10}$$

2.3 Model solving with a potential function

2.3.1 Definitions and properties

Following Zhang (2004a,b), we place our model in the context of evolutionary game theory and use the concept of potential function to solve it. In game theory, the concept of potential function was proposed by Monderer & Shapley (1996). A game is said to be a potential game if the incentive of all players to choose their strategy can be expressed in one global function, which is called the potential function. In our context, the definition of a potential function takes the rather simple following form:

Definition 1. Let $\mathcal{F} : x \in X \to \mathcal{F}(x) \in \mathbb{R}$ be an aggregate function describing each of the potential configurations. By definition, $\mathcal{F}$ will be a (cardinal) potential function of our model if and only if each gain in utility Δu of a moving agent is equal to the

variation $\Delta\mathcal{F}$ that is induced on the global level by the move of this agent.[4] A cardinal potential function will thus verify: $\mathcal{F}(y) - \mathcal{F}(x) = \Delta_{xy}u$ with $\Delta_{xy}u$ previously defined (section 2.2.4).

The main property of a potential function is to link the variation of a purely individual function (the utility of the moving agent) to the variation of a global function defined on the space X of all possible configurations and characterizing the city configuration. The ensuing lemma points out even more the value of the potential function as an analytical tool.

Lemma 1

If $\mathcal{F}$ is a potential function of the system, then the stationary distribution Π is such that for any configuration x:

$$\Pi(x) = \frac{e^{\mathcal{F}(x)/T}}{\sum_{z\in X} e^{\mathcal{F}(z)/T}} \tag{2.11}$$

It follows that for $T \to 0$, the stationary configurations are those that maximize $\mathcal{F}$.

Proof. The following proof follows the classical argument presented in Young (1998).

Let π be the function defined as $\pi : X \to [0, 1]$; $x \to \pi(x) = e^{\mathcal{F}(x)/T} / \sum_z e^{\mathcal{F}(z)/T}$. The first step of the proof consists in checking that π satisfies the *detailed balance condition*:

$$\pi(x)P^T_{xy} = \pi(y)P^T_{yx} \tag{2.12}$$

If x and y are two different and not communicating states, equality 2.12 is trivially satisfied since in this case $P^T_{xy} = P^T_{yx} = 0$. If $x = y$, the detailed balance condition is also trivially verified. In the case where $x \neq y$ and x and y are two communicating states, one has:

$$\begin{aligned}
\pi(x)P^T_{xy} &= \pi(x)\gamma\frac{1}{1+e^{-\Delta_{xy}u/T}} = \pi(x)\gamma\frac{1}{1+e^{-(\mathcal{F}(y)-\mathcal{F}(x))/T}} = \pi(x)\gamma\frac{e^{\mathcal{F}(y)/T}}{e^{\mathcal{F}(x)/T}+e^{\mathcal{F}(y)/T}} \\
&= \pi(y)\gamma\frac{e^{\mathcal{F}(x)/T}}{e^{\mathcal{F}(x)/T}+e^{\mathcal{F}(y)/T}} = \pi(y)\gamma\frac{1}{1+e^{-(\mathcal{F}(x)-\mathcal{F}(y))/T}} = \pi(y)\gamma\frac{1}{1+e^{-\Delta_{yx}u/T}} \\
&= \pi(y)P^T_{yx}
\end{aligned}$$

[4]Games can be either ordinal or cardinal potential games. In cardinal games, the difference in individual payoffs for each player from individually choosing one's strategy *ceteris paribus* has to have the same value as the corresponding difference in value for the potential function. In ordinal games, only the signs of the differences have to be the same.

recalling that $\gamma = 1/(N_V(N_R + N_G)) = 1/(v(1-v)N^4)$.

Hence the detailed balance condition is always verified and

$$\sum_{x\in X} \pi(x)P^T_{xy} = \sum_{x\in X} \pi(y)P^T_{yx} = \pi(y)\sum_{x\in X} P^T_{yx} = \pi(y)\cdot 1 = \pi(y)\,, \tag{2.13}$$

which defines π as a stationary distribution of the process. Because the Markov chain is finite and irreducible, it has a unique stationary distribution. Hence, for each state x, $\Pi(x) = \pi(x) = e^{\mathcal{F}(x)/T}/\sum_z e^{\mathcal{F}(z)/T}$.

Define then X_F as the subset of X of the states that maximize the potential function $\mathcal{F}$:

$$X_F = \{y, \forall x \in X \;\; \mathcal{F}(y) \geq \mathcal{F}(x)\} \tag{2.14}$$

The second part of the lemma can now be proved as follows: for two states x and y of X_F, we will have $\mathcal{F}(x) = \mathcal{F}(y)$ and therefore $\Pi(x)/\Pi(y) = e^{[\mathcal{F}(x)-\mathcal{F}(y)]/T} = 1$, which means that two states that strictly maximize $\mathcal{F}$ are observed in the long run with the same probability; for two states $x \in X \setminus X_F$ and $y \in X_F$, we will have $\mathcal{F}(x) - \mathcal{F}(y) \leq 0$ and therefore $\Pi(x)/\Pi(y) = e^{[\mathcal{F}(x)-\mathcal{F}(y)]/T} \to 0$ as $T \to 0$. This means that for $T \to 0$, the probability to observe a state that does not maximize the potential function $\mathcal{F}$ becomes in the long run infinitesimally small. □

The potential function is hence a very powerful analytical tool. First, it establishes a relation between individual changes in utility and a global characteristic of the city configuration. Second, because stationary configurations can be defined as those maximizing the potential function for low noise levels, the existence of a potential function allows to qualify analytically stationary configurations. The fact that the knowledge of Δu is sufficient to say something on the global level is highly non-trivial since, in particular, there is no way to determine the externalities produced by the move of an agent - *ie* the variation of the utility of his former and new neighbors - only from the knowledge of the utility variation of the moving agent. Note that a low level of T is required for the maximum of the potential function to be achieved at stationary configurations.[5] Still, T has to remain strictly positive to avoid blocked states.

[5]In the case of finite values of the noise level ($T > 0$), it can be demonstrated using standard tools of statistical physics that the states which are the more probable to appear are those which maximize $F(x) + TS(x)$ where $S(x)$ is an entropy-like global function taking into account the number of ways of locating R_q red agents and G_q green agents in each block q of the city (Grauwin *et al.*, 2009a).

2.3.2 Main result: existence of a potential function

It is possible, using the potential function, to examine analytically the outcome of the model for different utility functions, representing different degrees of preference for mixed environments. To do so, two questions are to be answered first: given any pair of utility functions (u_R, u_G), does a potential function exist and can we compute it? Reciprocally, given a potential function, can we find a pair of utility functions (u_R, u_G) that can be translated into this specific potential function?

We show in the following that in the context of bounded neighborhoods, one can achieve an analytical resolution of the model under a rather mild condition. Let us begin with some definitions.

Definition 2. Let $\mathbb{U}$ be the set of pairs of utility functions (u_R, u_G) that verify, for all $(R, G) \in E_H$, the following condition:

$$u_R(R, G) - u_R(R, G + 1) = u_G(R, G) - u_G(R + 1, G) \tag{2.15}$$

Condition 2.15 only imposes that if a block contains $R + 1$ red agents and $G + 1$ green agents, the utility gain a red agent would achieve if a green agent left must be the same as the utility gain a green agent would achieve if a red agent left. The results in the following apply to pairs of utility functions verifying this condition. As we show below, this condition is not strongly restrictive from a theoretical viewpoint.

Definition 3. Let $\mathbb{F}$ be the set of aggregate functions of the form $\mathcal{F}(x) = \sum_{q \in \mathcal{Q}} F(R_q, G_q)$, where F is an intermediate function defined on the set E_{H+1} of all possible numbers of red and green agents that can be present in a block.

The main result of this chapter consists in the following proposition:

Proposition 1

Each aggregate function $\mathcal{F} \in \mathbb{F} : x \to \mathcal{F}(x) = \sum_{q \in \mathcal{Q}} F(R_q, G_q)$ is a potential function to which corresponds at least[a] one pair (u_R, u_G) of utility functions of $\mathbb{U}$ that can be expressed as:

$$\begin{cases} u_R(R,G) = F(R+1,G) - F(R,G) \\ u_G(R,G) = F(R,G+1) - F(R,G) \end{cases} \tag{2.16}$$

Reciprocally, for each pair of utility functions (u_R, u_G) of $\mathbb{U}$, there exists one corresponding potential function $\mathcal{F}_{[u_R,u_G]} \in \mathbb{F}$. This function can be expressed through the functional $F_{u_R,u_G} : E_{H+1} \to \mathbb{R}$ - such that $\mathcal{F}_{[u_R,u_G]}(x) = \sum_{q \in \mathcal{Q}} F_{[u_R,u_G]}(R_q, G_q)$ - which is defined for all $(R,G) \in E_{H+1}$ by:[b]

$$F_{[u_R,u_G]}(R,G) = \sum_{r=1}^{R} u_R(r-1,0) + \sum_{g=1}^{G} u_G(R,g-1) \tag{2.17}$$

$$= \sum_{r=1}^{R} u_R(r-1,G) + \sum_{g=1}^{G} u_G(0,g-1) \tag{2.18}$$

[a] Since the definition of the potential function makes only intervene its variation, $\mathcal{F}$ can be defined up to an additive constant. Similarly, a utility function can also be defined up to a constant. All the formula in this insert are written with the convention $u(0,0) = F(0,0) = F(0,1) = F(1,0) = 0$. For more details, see the proof in A.1.

[b] Our notation assumes that a sum is null whenever its upper bound of summation is inferior to its lower bound of summation.

Proof. See Appendix A.1. □

Proposition 1 states that it is always possible to define a function F defined at the neighborhood level corresponding to the variation of utility of a moving agent, but only a pair of utility functions verifying condition 2.15 allows this function to be path-independent and therefore uniquely defined for any given configuration. Reciprocally, for any pair of utility functions verifying condition 2.15, the game has a potential function that is maximized at stationary configurations, and this function is the sum of neighborhood-level intermediate components. As Eq 2.17 shows, the intermediate component of the potential function corresponds to the sum of utilities of the agents arriving in succession in the block. This sum is calculated starting from an empty block, agents being introduced one by one, first the red ones and then the green ones. As Eq 2.18 shows, the same sum is obtained if green agents are introduced first and red agents after. Before giving a more general interpretation in section 2.3.3, it is useful to give the following corollary, aimed at showing that utility functions verifying condition 2.15 can be given a convenient formulation in which interactions between the two groups are

expressed through the same function in the two utility functions, thus giving a more general formulation for the potential function.

Corollary 1

Any pair of utility functions (u_R, u_G) belonging to $\mathbb{U}$ can be written:

$$u_R(R,G) = \xi_R(R) + \sum_{g=0}^{G-1} \xi(R,g) \tag{2.19}$$

$$u_G(R,G) = \xi_G(G) + \sum_{r=0}^{R-1} \xi(r,G) \tag{2.20}$$

where ξ_R and ξ_G are arbitrary functions of $\{0,1,..,H\} \to \mathbb{R}$ and ξ is an arbitrary function of $E_H \to \mathbb{R}$.

For each pair of utility functions (u_R, u_G) verifying 2.19 and 2.20, thanks to Eq. 2.17, one can rewrite the general form of the potential function $\mathcal{F}_{[u_R,u_G]}$ as:

$$\mathcal{F}(x) = const + \sum_q \Big(\sum_{r=0}^{R_q-1} \xi_R(r) + \sum_{g=0}^{G_q-1} \xi_G(g) + \sum_{r=0}^{R_q-1} \sum_{g=0}^{G_q-1} \xi(r,g) \Big) \tag{2.21}$$

Proof. For any pairs of utility functions (u_R, u_G), one can define ξ_R and ξ_G, two functions of $\{0,1,..,H\} \to \mathbb{R}$ and ξ_{RG} and ξ_{GR}, two functions of $E_H \to \mathbb{R}$ by

$$\begin{cases} \xi_R(r) = u_R(r,0) \\ \xi_G(g) = u_G(0,g) \end{cases} \qquad \begin{cases} \xi_{RG}(r,g) = u_R(r,g+1) - u_R(r,g) \\ \xi_{GR}(r,g) = u_G(r+1,g) - u_G(r,g) \end{cases}$$

for all $0 \leq r \leq H$ and $0 \leq g \leq H$. By definition, one can then write the utility functions as

$$u_R(R,G) = \xi_R(R) + \sum_{g=0}^{G-1} \xi_{RG}(R,g) \tag{2.22}$$

$$u_G(R,G) = \xi_G(G) + \sum_{r=0}^{R-1} \xi_{GR}(r,G) \tag{2.23}$$

for all $(R,G) \in E_H$.

The condition given by Eq. 2.15 is obviously equivalent to $\xi_{RG} = \xi_{RG} \equiv \xi$, which proves Corollary 1.

□

2.3.3 Interpretation

We propose here first an interpretation of condition 2.15 and then of the form of the potential function. Proposition 1 ensures that there exists a potential function for any pair of utility functions verifying condition 2.15. In its original form, this condition says that there is a symmetry in the externalities generated by green agents on red agents and by red agents on green agents: starting from a given neighborhood composition, the variation in utility produced by the departure of an agent of the other type must be the same for the two categories. This can be seen as rather limiting, as some real world situations do not conform to this condition. For instance, well-known surveys on the appreciation by white and black individuals of their preferred residential environment show that blacks are in favor of integrated neighborhoods, whereas whites favor all-white neighborhoods (Farley *et al.*, 1978; see Farley *et al.*, 1997 for recent figures).

However, condition 2.15 covers more general types of preferences when the rate of vacant cells is low. Namely, in the limit of a very low vacancy rate, there is no vacant cells in most of the blocks, *ie* in these blocks the relation $R_q + G_q = H + 1$ holds. Hence, one only needs one parameter among (R_q, G_q, V_q) to define a utility function and considering for instance that an agent's utility only depends on his number of similar neighbors is sufficient to describe all possible cases. This can be done by taking $\xi \equiv 0$ in Eq. 2.19 and Eq. 2.20, while keeping the functions ξ_R and ξ_G independent and free. In other words, in the limit of no vacant cells, each agent arriving in a neighborhood receives a utility that is fully determined by the number of like-neighbors. Therefore, the order in which the agents settle in the neighborhood does not matter and the condition for having a potential function holds. The set $\mathbb{U}$ hence describes all possible pairs of utility functions in the limit $v \to 0$. It follows also that condition 2.15 holds for all pairs of utility functions in situations where vacant cells are considered in the same way as unlike-color neighbors. Note also that condition 2.15 applies to the utility functions considered in Zhang (2004a), where preferences over neighborhoods are determined by the number of like-agents only and the symmetric effect of unlike-color neighbors on each type of agent emerges as the result of the determination of housing prices by densities.

In its original form, the potential function $\mathcal{F}$ can be interpreted as the sum of the incentives the agents had (when they settled) to move into the neighborhood where they are located. Indeed, if $x(t)$ denotes the state of the city at iteration t, then the potential can be rewritten as

$$\mathcal{F}(x(t)) - \mathcal{F}(x(0)) = \sum_{t'=1}^{t} \Delta_{x(t'-1)x(t')} u \tag{2.24}$$

where $\Delta_{x(t'-1)x(t')}u = 0$ by definition if no move happens at iteration t' and where we can take $\mathcal{F}(x(0)) = 0$ since the potential is defined up to a constant. Conversely, the potential function $\mathcal{F}$ can also be viewed as the minimum utility level each agent would require to accept quitting his neighborhood. As such, it represents, in the case $T \to 0$, the stability of the configuration x: the higher the potential function, the smaller the incentives for agents to move.

To interpret further the potential function, it is worth noting that condition 2.15 can also be written as follows:

$$u_R(R, G) + u_G(R + 1, G) = u_G(R, G) + u_R(R, G + 1) \tag{2.25}$$

which means that starting from any initial composition of a block, the sum of utilities of a red agent and a green agent entering successively in this block is the same whatever the order in which they enter. This expression stresses that, under condition 2.15, the value of function F in a given neighborhood q does not depend on the particular path of events that lead to the composition of this neighborhood. This is also particularly clear in the form of condition 2.15 given in corollary 1. It hence follows that the potential function $\mathcal{F}$, which is the sum of the F intermediate functions, is independent of the particular order in which the agents arrived in the neighborhoods.

Hence it is also possible to define $\mathcal{F}$ as the average over all the possible ways of ordering the agents, which will be shown formally in section 2.5.2.

2.4 Segregation for different levels of preference for mixed environments

The main property of the potential function obtained in the previous section is that it reflects both the macro and micro scale. On one hand, $\mathcal{F}$ is an aggregate function defined at the city level which only depends on the number R_q and G_q of red and green agents in each block. On the other hand, $\mathcal{F}$ also keeps tracks of the individual level since it corresponds to a sum of the utility differences generated by individual moves. When the stationary states are reached in the case $T \to 0$, $\mathcal{F}$ is maximized, which means that no agent can strictly improve her utility by moving. The potential function can now be used to assess the outcomes of our location model for different utility functions, representing different degrees of preference for mixed environments, as far as these functions verify condition 2.15. We examine successively (i) linear utility functions, with a continuous preference for segregated environments, (ii) Schelling original utility function

in which there is a mild preference for a mixed environment and (iii) an asymmetrically peaked utility function, according to which agents exhibit a strict preference for a mixed environment.[6]

2.4.1 Linear utility functions

We consider here utility functions that exhibit a monotone effect of the number of same-color neighbors on utility, through linear utility functions. Zhang (2004a) proposes an analytical solution of a dynamic model of segregation with a linear utility function and shows that the halved sum of individual utilities is a potential function of the game thus defined. In this section, we show that Proposition 1 allows to find similar results for all linear utility functions verifying condition 2.15 in the context of bounded neighborhoods.

Suppose that u_R and u_G are expressed as:

$$\begin{aligned} u_R(R,G) &= aR + bG \\ u_G(R,G) &= bR + dG \end{aligned} \tag{2.26}$$

where a, b, d are constant parameters.[7]

One can easily verify that this particular pair of utility functions verifies condition 2.15 and compute the corresponding potential function:

$$\mathcal{F}(x) = \frac{1}{2}\sum_q (aR_q(R_q-1) + dG_q(G_q-1) + 2bR_qG_q) \tag{2.27}$$

One can rewrite this potential function as:

$$\mathcal{F}(x) = \Big(b - \frac{a+d}{2}\Big)\rho_{RG}(x) - \frac{a}{2}\rho_{RV}(x) - \frac{d}{2}\rho_{GV}(x) \tag{2.28}$$

with:

ρ_{RG}	$= \sum_q R_qG_q$	the number of red-green pairs of neighbors,
ρ_{RV}	$= \sum_q R_q(H+1-R_q-G_q)$	the number of red-vacant pairs of neighbors and
ρ_{GV}	$= \sum_q G_q(H+1-R_q-G_q)$	the number of green-vacant pairs of neighbors.

This last form provides a convenient interpretation of the potential function. Putting

[6] Refer to Grauwin *et al.* (2009b) for the study of other utility functions.

[7] Zhang (2004a)'s utility function corresponds to $b = d = -1$ and $a \geq -1$, the utility also including a fixed income term which makes it positive. In Zhang (2004a)'s framework, the impact of unlike neighbors is not due to preferences, but to the impact of density on housing prices.

aside at this point the last two terms, $\mathcal{F}(x)$ is proportional to ρ_{RG}, that gives a measure of the relative contact between the two groups. Hence, the sign of the prefactor $b-(a+d)/2$ indicates whether mixed states (when positive) or segregated states (when negative) are obtained at the global level. Notice that the two groups do not need to both have strong preferences for like neighbors for segregation to emerge. It is the average preference over the two groups that determines the level of segregation.

The terms proportional to ρ_{RV} and ρ_{GV} show that agents avoid the proximity of vacant cells when $a > 0$ and $d > 0$. All these insights gained from the study of the potential function can be checked by means of simulations (Fig 2.2).

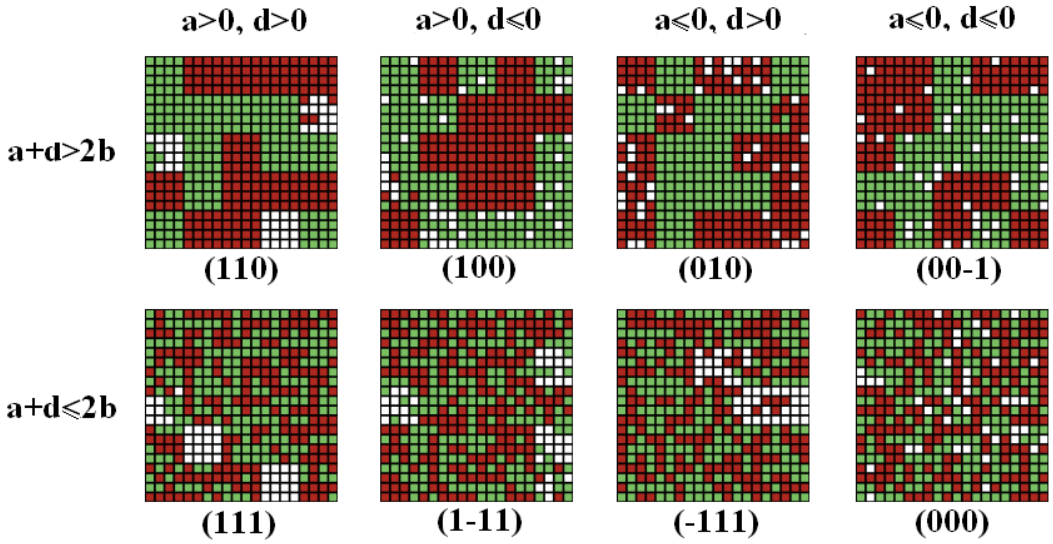

Figure 2.2: **Typical stationary configurations obtained by simulations for different values of (adb). Top panel:** for $2b-(a+d)<0$, the system evolves towards segregated configurations where red and green agents tends to live in different blocks. **Bottom panel:** for $2b-(a+d)>0$, the system evolves towards mixed configurations where the number of red-green pairs of neighbors is maximized. **From left to right:** the sign of a and d controls the tendency of red and green agent to prefer to live in dense or uncrowded areas. The demographic parameters are ($N = 20, v = 10\%, n_R = 0.5$). Neighborhood size is fixed to $H+1 = 16$ and the level of noise is $T = 0.1$

Turning now to the link between segregation of the stationary configurations and collective utility, it is useful, using proposition 1, to write the potential function as

follows:[8]

$$\begin{aligned}\mathcal{F}(x) &= \frac{1}{2}\sum_q (aR_q(R_q-1)+dG_q(G_q-1)+2bR_qG_q)\\ &= \frac{1}{2}\sum_q (R_q u_R(R_q-1,G_q)+G_q u_G(R_q,G_q-1))\\ &= \frac{1}{2}U(x)\end{aligned}$$

With this choice of utility functions, the potential function is thus proportional to collective utility and therefore lemma 1 ensures that, for low values of T, the stationary configuration maximizes collective utility. Reciprocally, one can verify (see proof in A.2) that if we want the potential function to be proportional to the collective utility, so that states that maximize the potential function also maximize collective utility, then the constant of proportionality is necessarily 0.5 and the pair of utility functions must take the form displayed in Eq. 2.26 (up to a constant).

To sum up, the linear utility functions as defined in 2.26 lead to segregated or mixed states depending on the values of the parameters. However, in all cases, there is no divergence between stationary configurations and the optimum: these utility functions are such that utility-improving moves also improve collective utility.

2.4.2 Schelling utility function

Suppose that the agents compute their utility with Schelling utility function, which is equal to 1 if their fraction of similar neighbors is superior or equal to 0.5, and equal to 0 otherwise. This utility function can be expressed in terms of the number of red and green neighbors as follows:

$$\begin{aligned}u_R(R,G) &= \Theta(R-G) = \frac{1}{2}(1+|R+1-G|-|R-G|)\\ u_G(R,G) &= \Theta(G-R) = \frac{1}{2}(1+|R-1-G|-|R-G|)\end{aligned} \tag{2.29}$$

where Θ is the Heaviside function defined by: $\Theta(x)=0$ if $x<0$ and $\Theta(x)=1$ if $x\geq 0$.

It is easy to figure out that this particular pair of utility functions respects condition 2.15, and is therefore in the set $\mathbb{U}$. Indeed, the form of the symmetric externality

[8] Note that this result is similar to the one obtained in the continuous neighborhood case as will be seen in section 2.5.1. See also Grauwin *et al.* (2009b).

produced by a new unlike-color neighbor is the following:

$$u_R(R, G+1) - u_R(R, G) = |R - G| - \frac{1}{2}(|R - G - 1| + |R - G + 1|) \quad (2.30)$$
$$= u_G(R+1, G) - u_G(R, G) \quad (2.31)$$

It is possible to compute the potential function rather directly thanks to its interpretation, as the sum of the utility of the agents being introduced one by one in the city, this sum being independent of the precise order of introduction of the agents. To do so for a given configuration $x \equiv \{R_q, G_q\}$, let us consider that we introduce in each block first the agents in majority (*ie* the red ones if $R_q > G_q$, the green ones if $G_q > R_q$, either the red or the green ones if $R_q = G_q$) and second the agents in minority. Each of the first agents has a utility of 1 as he settles in the city while each of the other minority agents has a zero utility when he settles.[9] Hence it is straightforward to write the potential function as:[10]

$$\begin{aligned}
\mathcal{F}(x) &= const + \sum_{q \in \mathcal{Q}} \max(R_q, G_q) \\
&= const + \sum_{q \in \mathcal{Q}} \frac{1}{2}\Big(R_q + G_q + |R_q - G_q|\Big) \\
&= const' + \frac{1}{2} \sum_{q \in \mathcal{Q}} |R_q - G_q|
\end{aligned}$$

One can verify that the same expression can be found using relation 2.17 (see A.3), the computation being in this case more formal than what we present here.

The reader can recognize an expression well-known to scientists working on residential segregation. This potential function is indeed a linear form of the Duncan and Duncan dissimilarity index, which in the case where the total number of red and green agents in the city are equal ($N_R = N_G = N$), is written as $D(x) = \frac{1}{2}\sum_q |R_q/N_R - G_q/N_G| = \frac{1}{2N}\sum_q |R_q - G_q|$ (Duncan and Duncan, 1955). To the best of our knowledge, an analytical connection between the two "historical" works of Schelling and Duncan and Duncan on segregation has never been found before.

It is worth here investigating the link between the potential function and collective

[9] This example shows that to compute the potential function corresponding to a given pair (u_R, u_G) of utility functions, it may be worth to think ahead of a practical order of introduction of the agents. The computation of $\mathcal{F}$ is indeed easier and bears more meanings with an appropriate order.

[10] Notice that in this particular example, we do not use the convention $u(0, 0) = 0$. See A.3 for details.

utility. The collective utility in a neighborhood q is:

$$\begin{cases} U_q &= \frac{1}{2}\left(R_q + G_q + |R_q - G_q|\right) & \text{if } R_q \neq G_q \\ U_q &= R_q + G_q & \text{if } R_q = G_q \end{cases}$$

The potential function can therefore be written:

$$\mathcal{F}(x) = const + U(x) - \sum_{q|R_q=G_q} \frac{1}{2}\Big(R_q + G_q\Big) \tag{2.32}$$

$$\tag{2.33}$$

This expression shows that the configuration that maximizes $\mathcal{F}$ does not correspond to the maximum collective utility. The divergence is due to the existence of perfectly mixed neighborhoods. To be more specific, let us compare two configurations differing by the existence, in configuration x_1, of two perfectly mixed neighborhoods with $K < (H+1)/2$ agents of each color, that are changed to segregated ones in configuration x_2 due to the exchange of two agents of different color. The difference in the potential function between the two configurations is only due to the change affecting these two neighborhoods. It is written:

$$\Delta\mathcal{F} = \mathcal{F}(x_2) - \mathcal{F}(x_1) = 2 \tag{2.34}$$

because each of the two neighborhoods gained one agent of one color and lost an agent of the other color. The difference in collective utility consists of the loss of utility of the agents who are now in minority in their neighborhood, that is:

$$\Delta U = U(x_2) - U(x_1) = -2(K-1) = 2 - 2K \tag{2.35}$$

Comparing the difference in the potential function and in collective utility between these two configurations, one observes that decreasing the number of perfectly mixed neighborhoods decreases collective utility (due to the loss of those who are in the minority in the new configuration) while increasing the value of the potential function. This is because the two moving agents have still a utility of 1 in their new neighborhood, while they clearly exert negative externalities on the agents of the group which is now in minority in this neighborhood. Stationary configurations will tend therefore to exhibit few perfectly mixed neighborhoods, at the expense of collective utility.

Our analysis based on Schelling original utility function shows thus two points. First, it provides an analytical demonstration of Schelling result, that this pair of utility func-

tions leads to segregated stationary configurations. Second, it sheds light on the source of the discrepancy between the collective utility of stationary configurations and the maximum utility that could be attained with a perfectly mixed environment.

2.4.3 Asymmetrically peaked utility functions

In this section, we apply our analytical framework to asymmetrically peaked utility functions displayed on Fig. 2.3, that have been studied in Pancs & Vriend (2007). We will see that our potential function provides a criterion for global segregation or for integration. These utility functions are particularly appealing for demonstrating Schelling's intuition, that the aggregate outcome of the game can run against individual preferences. Indeed, these functions consider a case where agents strictly prefer perfectly mixed neighborhoods against any level of segregation.

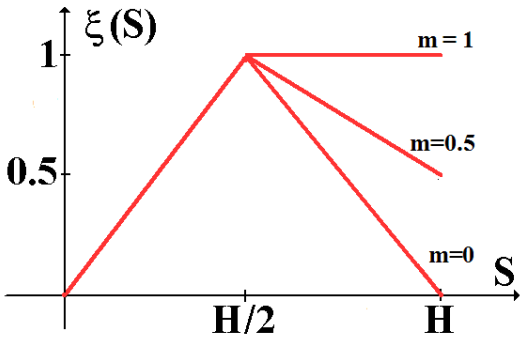

Figure 2.3: **Asymmetrically peaked function for some values of** m**.**

In the following, we place ourselves in the case $\xi \equiv 0$ where the utility of each type of agents can be described entirely by the number of like-color agents. This choice permits us to write analytical results more easily. It can furthermore be understood as the limit case $v \to 0$ For simplicity, we suppose that the number H of possible neighbors of an agent is even. The asymmetrically peaked utility functions can then be written:

$$\begin{cases} u_R(R,G) = \xi_{ap}(R) \\ u_G(R,G) = \xi_{ap}(G) \end{cases}$$

with:

$$\begin{cases} \xi_{ap}(s) & = 2s/H & \text{if } s \leq H/2 \\ \xi_{ap}(s) & = 2 - m - 2(1-m)s/H & \text{if } s > H/2 \end{cases}$$

with m a fixed parameter (see Fig. 2.3).

Using Θ the Heaviside function defined by: $\Theta(x) = 0$ if $x < 0$ and $\Theta(x) = 1$ if $x \geq 0$, this utility function can also be written:

$$\xi_{ap}(s) = 2\frac{s}{H} - (2-m)\frac{2}{H}\left(s - \frac{H}{2}\right)\Theta\left(s - \frac{H}{2}\right)$$

For a given state x of the city, simple calculations show that the collective utility can be written as $U(x) = \sum_q \left(\tilde{U}(R_q) + \tilde{U}(G_q)\right)$, with

$$\tilde{U}(S) = 2\frac{S(S-1)}{H} - (2-m)\frac{2S}{H}\left(S - 1 - \frac{H}{2}\right)\Theta\left(S - 1 - \frac{H}{2}\right), \qquad \forall 0 \leq S \leq H \quad (2.36)$$

Likewise, the corresponding potential function is $\mathcal{F}(x) = const + \sum_q F(R_q, G_q) = const + \sum_q \left(\tilde{F}(R_q) + \tilde{F}(G_q)\right)$, where

$$\begin{aligned}\tilde{F}(S) &= \sum_{s=0}^{S-1} \xi_{ap}(s) \\ &= \frac{(S-1)S}{H} - \frac{2-m}{H}\left(S - 1 - \frac{H}{2}\right)\left(S - \frac{H}{2}\right)\Theta\left(S - \frac{H}{2} - 1\right) \\ &= \frac{1}{2}\left[\tilde{U}(S) + (2-m)\left(S - 1 - \frac{H}{2}\right)\Theta\left(S - 1 - \frac{H}{2}\right)\right] \end{aligned} \quad (2.37)$$

Proof. See A.4. □

This expression implies once again that the potential $\mathcal{F}$ and the collective utility U are linearly related when the individual utility is linear (case $m = 2$). The lower m, the less linear the individual utility and the greater the divergence from the $\mathcal{F} = const + U/2$ relation. Thus, relation 2.37 puts forward the crucial role of the asymmetric parameter m which is the driver of the moves that produce externalities.

To be more specific, let us compare two configurations x_1 and x_2, which differ only in the repartition of $H+1$ red and $H+1$ green agents in two neighborhoods. In configuration x_1, the repartition is rather homogeneous, with $H/2+1$ red and $H/2$ green agents in the first neighborhood and $H/2$ red and $H/2+1$ green agents in the second one. In configuration x_2, the repartition is more segregated, with $H/2+1+K$ red and $H/2-K$ green agents in the first neighborhood and $H/2-K$ red and $H/2+1+K$ green agents in the second one, $K \in \{0, 1, ..., H/2\}$ being an integer determining the level of segregation. The difference in the potential function between the two configurations is only due to

the change affecting these two neighborhoods. It is written:

$$\begin{aligned}\Delta\mathcal{F} &= 2\tilde{F}(H/2+1+K)+2\tilde{F}(H/2-K)-2\tilde{F}(H/2+1)-2\tilde{F}(H/2)\\ &= \frac{2m}{H}K(K+1)\end{aligned} \tag{2.38}$$

Proof. See A.4. □

As could be expected, $\Delta\mathcal{F}$ increases with m, which means that the segregated configuration is more probable and stable than the mixed one as the asymmetry toward like-agents is stronger. It also increases with K, which means that for a given m a highly segregated block is more probable than a slightly segregated one. More importantly, a perfectly segregated block will be more probable than a perfectly mixed one if and only if relation 2.38 is positive, *ie* if and only if $m > 0$.

The corresponding difference in collective utility consists of the loss of all the agents. It can be written:

$$\begin{aligned}\Delta U &= 2\tilde{U}(H/2+1+K)+2\tilde{U}(H/2-K)-2\tilde{U}(H/2+1)-2\tilde{U}(H/2)\\ &= 4K\left(m\left(\frac{1}{2}+\frac{K+1}{H}\right)-1\right)\end{aligned} \tag{2.39}$$

Proof. See A.4. □

It is obvious that increasing segregation also increases collective utility as soon as $m \geq 1$. It is straightforward to verify based on equation 2.39 that $\Delta U \geq 0 \Leftrightarrow m \geq m^* = H(H+1)^{-1}$.[11]

Our analysis provides a microscopic criterion allowing to predict a global outcome. For $0 < m < m^* \simeq 1$, complete segregated configurations will be obtained at the expense of the collective utility and for $m < 0$, perfectly mixed configuration will be obtained. These results hold of course in the limit of a low noise level ($T \to 0$).

The same is observed with simulations. The snapshots presented on the left panel of Fig. 2.4 are typical stationary configurations obtained by simulating an artificial city where the agents' preferences are given by the asymmetrically peaked utility function with bounded neighborhoods. These snapshots allow us to compare the analytical results obtained for $v \to 0$ and $T \to 0$ with a more realistic $v = 5\%$ and $T = 0.1$. We

[11] The limit value is not strictly equal to 1 because of the precise definition of our model: the argument of the utility function is the number of neighbors, which does not include the agent himself. In the limit $H \gg 1$, m^* converges toward 1.

can see that for values of the asymmetry parameter m close to 0, the system converges toward randomly-organized mixed configurations which also maximize the utility of most agents. On the contrary, for higher values of m, completely segregated configurations are obtained. The transition between these two extreme outcomes occurs for values of m included between 0.05 and 0.2. The relative smoothness of this transition is due to the non-zero values of the vacancy rate and of the level of noise.

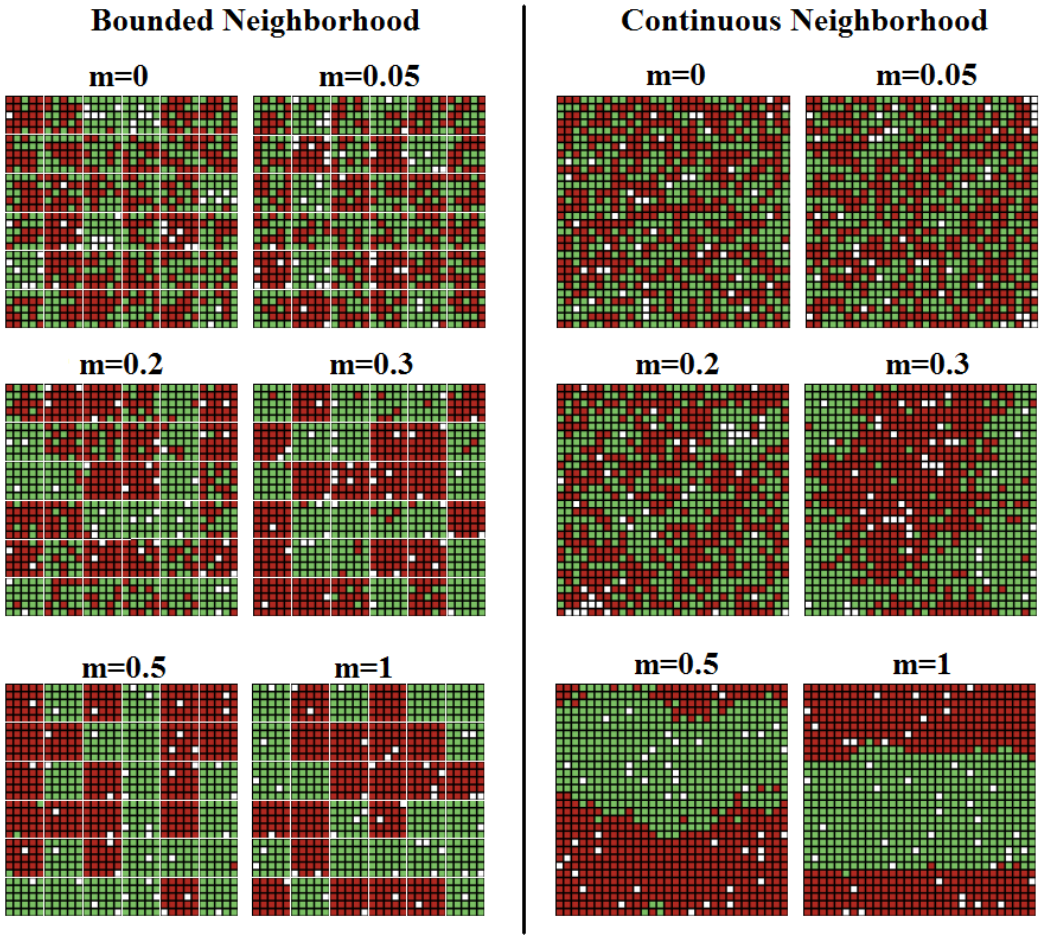

Figure 2.4: **Typical stationary configurations obtained by simulations with the asymmetrically peaked utility function.** The demographic parameters are $(N = 30, v = 5\%, n_R = 0.5)$. Neighborhood sizes are fixed to $H = 24$, and the level of noise is fixed to $T = 0.1$. Left: with a bounded neighborhood description. Right: with a continuous neighborhood description.

The outcome for values of m higher than 0.2 illustrates the paradox of Schelling model: large segregative patterns appear although they absolutely do not maximize the utility of most agents, as most of them are stuck inside an homogeneous area with a

utility of 0.5. In this case, one of the key element driving segregation is the asymmetry of this utility function, ie, the fact that even if the agents have a strict preference for mixed environments, they still prefer to belong to the majority group instead of being in minority, as was already shown in Pancs & Vriend (2007). In particular, with the asymmetrically peaked utility function, a red (green) agent may move for example from a 49% red (green) neighborhood to a 51% red (green) neighborhood because it slightly increases her utility. Meanwhile, this move is likely to decrease the utility of the previous and new neighbors and therefore decrease the collective utility level. Both of these factors imply that a highly-segregated configuration is necessarily very stable. Indeed, once the city is divided into homogeneous areas, a red agent will have no incentive to go from the red area to the green one, because his utility would drop from 0.5 to 0.[12]

2.5 Discussion, limits and extensions

In this section, we first motivate the bounded neighborhood assumption with a supplementary result in the continuous neighborhood case and show how our model can be extended to get closer to a continuous neighborhood definition. We then show that the potential function of our model can be interpreted in terms of the Shapley value of a coalitional game, which allows us to relate our results to broader issues of game theory. In a third development, we use the analytical framework to investigate some extensions of the model, such as different agents' preferences and an analysis of taxation.

2.5.1 Bounded *vs* continuous neighborhoods

Most papers based on simulations as well as those presenting analytical results for particular cases (Pancs & Vriend, 2007; Zhang, 2004a) deal with continuous neighborhoods. In this paragraph, we discuss the differences between bounded and continuous neighborhoods.

Analytic limitations in the continuous neighborhood case

Our previous analysis can be extended by considering continuous neighborhoods. In the following proposition, we establish the important result that, in the context of continuous neigborhoods, a potential function only exists with linear utility functions.

Proposition 2

[12] And even though a red agent goes from time to time into the green area by mistake, he will have a strong incentive to return to the red area because of the asymmetry in the utility function, and he will do so very likely before a second red agent joins him in the green area.

When using a continuous neighborhood description, a potential function exists if the agents' individual utility functions are bilinear functions of the form

$$\begin{aligned} u_R(R,G) &= u_R(0,0) + aR + bG \\ u_G(R,G) &= u_G(0,0) + bR + dG \end{aligned} \tag{2.40}$$

where a, b, d are three real constants. In this case, a potential function $\mathcal{F}$ can moreover be written as $\mathcal{F} : x \to \mathcal{F}(x) = const + U(x)/2$. Reciprocally, no potential function exists for any other form of utility functions.

The first part of this claim - which is that a potential function exists if the utility functions are chosen as in Eqs. 2.40 - is the main subject of Zhang (2004a). The second part of this claim - which is that no potential function exists for any other choice of utility functions (in the context of the model presented in this chapter and with continuous neighborhoods) has to our knowledge never been proved elsewhere.[13] A sketch of the proof of Proposition 2 is given in A.5. The main reason why no potential function exists in general is that when a moving agent generates externalities on his (past and new) neighbors, these externalities depend on the type of his neighbors' own neighbors. A way to handle this issue in the continuous neighborhood case is to consider linear utility functions. For these, the generated externalities do not depend on the neighbor's neighbors, because linearity implies that the newcomer generates the same change in utility on his neighbors, independently of their initial situation; the same holds for the neighbors he leaves. For non linear utilities, a potential function can exist only in the bounded neighborhood case.[14]

Simulations and real segregation measures

Let us now investigate the differences between bounded *versus* continuous neighborhoods in the general case. Based on the simulations in the case of the asymmetrically peaked utility function, the snapshots displayed on the right panel of Fig. 2.4 present stationary configurations obtained using a continuous neighborhood description. The left panel of Fig. 2.4 shows the influence of a bounded neighborhood on the stationary configurations obtained with the same parameters. Mixed random configurations are obtained in both bounded and continuous neighborhood descriptions for low values

[13] Note however that the inexistence of a potential function does not preclude the convergence of the game to a stationary state.

[14] Zhang (2004b) proposes a continuous neighborhood model without vacant cases which allows to derive a potential function for utility functions defined by two linear pieces.

of m and segregative patterns also appear in both descriptions for high values of m. The two descriptions produce a different "transition range" (roughly 0.1 ± 0.05 in the bounded neighborhood case *versus* 0.25 ± 0.05 in the continuous neighborhood case). The strongest difference between the two cases lies in the patterns observed at the city scale: as there are no connections across neighborhood boundaries in the bounded case, red and green neighborhoods are observed side by side. In the continuous neighborhood case on the contrary, segregated patterns appear at the city scale. However, these simulations suggest that the local degree of segregation does not change with the definition of neighborhoods that is used.

The same concern was raised in the residential segregation literature, where the most traditional segregation measures consider a bounded neighborhood definition with no connections between blocks (Duncan and Duncan, 1955). To remedy this issue, spatial measures of segregation have been developed, in which the spatial arrangement of neighborhoods intervenes (Morrill, 1991; Reardon & O'Sullivan, 2004; Wong, 2005). This can be done either by taking the contiguity between neighborhoods into account, by integrating a distance matrix between all neighborhoods with a distance-decay function, thus considering that all the city locations matter when measuring potential contacts between groups in a specific location, or by using point locations instead of areal tracts. Drawing the parallel with the bounded and continuous neighborhood segregation models, the last option would correspond to a continuous neighborhood definition.

From this viewpoint, the main drawback of our segregation model is that an agent is not affected by the composition of the blocks next to his own, which may seem unrealistic and explains why red and green neighborhoods are observed side by side in stationary configurations. Referring to the segregation measure literature, this could be solved by accounting for the contiguity between blocks. In the following, we propose a variation of our bounded neighborhood model which includes transitivity between blocks while maintaining the existence of a potential function.

Bounded neighborhood with spatial transitivity

The bounded neighborhood formulation is based on a given partition (let us call it Q_1) of the city lattice into blocks, as shown on Fig 2.1a. We may define a second partition Q_2 of the city lattice in such a way that blocks of partition Q_2 overlap with blocks of partition Q_1. Fig 2.5b shows an example where partitions are shifted from each other by a half a block's diagonal. Let us assign to each red (resp. green) agent the two blocks $q_1 \in Q_1$ and $q_2 \in Q_2$ he is living in and the corresponding utilities $u_{R,i} = u_R(R_{q_i} - 1, G_{q_i})$ where $(i \in \{1, 2\})$ (resp. $u_{G\,i} = u_G(R_{q_i}, G_{q_i} - 1)$). Two potential functions $\mathcal{F}_1$ and $\mathcal{F}_2$

where $\mathcal{F}_i = \mathcal{F}(\{R_{q_i}, G_{q_i}\})$ can similarly be assigned to the two partitions.

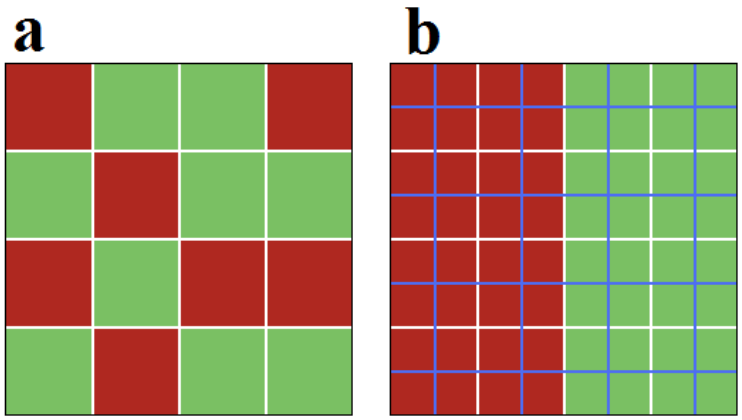

Figure 2.5: **Illustration of the expected effects of a "two partitions" model. a** In the one partition model, homophilic preference leads to a pattern where blocks are exclusively populated by either red or green agents as seen in Fig 2.4. In this case however, agents are not aware of the composition of nearby blocks. **b** In the two partitions framework (in each partition, blocks are delimited by either white or blue lines), nearby blocks of a given partition have a transitive influence on each other through the overlapping blocks of the other partition. The same segregative dynamics implying a minimization of the number of mixed blocks of either partition will lead to a gathering of same-color blocks.

Let us now finally define a model encompassing both partitions. The utility u of an agent is taken as the sum $u_1 + u_2$ corresponding to his utility in both partitions. It is straightforward to verify that the analytical properties of our model hold in this larger description and that the states of the city can be characterized by the potential function $\mathcal{F}_{12} = \mathcal{F}_1 + \mathcal{F}_2$, whose maxima correspond to the stationary states of the city. Moreover, the fact that partitions overlap implies spatial transitivity between blocks of a given partition. Indeed, suppose that agents compute their utility with the asymmetrically peaked utility function, as in section 2.4.3. We showed earlier in the model with one partition that for $m > 0$, stationary states maximizing the potential function correspond to states where each block is either completely red or completely green, while mixed blocks were dynamically unstable. It is possible to show that in the two partitions model, the same result holds because the simultaneous maximization of the two potential functions can be achieved by minimizing the number of mixed blocks, which leads to a gathering of blocks of the same color, as illustrated on Fig 2.5b.

2.5.2 A coalitional game formulation

The formulation of Schelling model in the context of evolutionary game theory can yield other fruitful results. The following corollary gives a general form of the potential

function that can be interpreted in terms of the Shapley value of a non cooperative coalitional game (Shapley, 1953).

Corollary 2

The potential function $\mathcal{F}$ can be written as the sum of the following intermediate functions F:

$$F_{[u_R,u_G]}(R,G) = \sum_{\substack{0\leq r\leq R,\ 0\leq g\leq G \\ (r,g)\neq(0,0)}} \binom{R}{r}\binom{G}{g} \frac{(r+g-1)!\,(R+G-r-g)!}{(R+G)!}\,\nu(r,g) \quad (2.41)$$

$$\text{where} \quad \nu(r,g) = r\,u_R(r-1,g) + g\,u_G(r,g-1) \quad (2.42)$$

is the collective utility in a block having r red and g green agents.

Let us consider a coalitional game defined by the set of the $R_q + G_q$ agents present in block q along with a coalition worth equal to the sum of their utilities $\nu_q(R_q, G_q) = R_q\,u_R(R_q-1, G_q) + G_q\,u_G(R_q, G_q-1)$, that is the collective utility at the neighborhood level. The sum of the potential functions of the $|Q|$ coalitional games defined on the $|Q|$ blocks is equal to the potential function $\mathcal{F}$.

Proof. The form of F given in expression 2.41 is derived in A.1. Then, acknowledging that in a neighborhood with $R_q + G_q$ agents, there are $\binom{R}{r}\binom{G}{g}$ possible coalitions having g green and r red agents and applying the formula of the potential function of the Shapley value derived in Hart & Mas-Colell (1989), one obtains exactly the formula of the potential function given in 2.41. Hence the potential function $\mathcal{F}$ can be written as the sum of these potential functions of the $|Q|$ coalitional games defined on the $|Q|$ blocks. □

The potential function $\mathcal{F}$ of our non cooperative game can thus be written as the sum of the potential functions of $|Q|$ coalitional games defined on the $|Q|$ blocks.[15] Notice furthermore that the corresponding Shapley value can be straightforwardly identified, thanks to Eq. 2.16 and the relationship between the Shapley value and its potential, as the vector of $\mathbb{R}^{R_q+G_q}$ whose components are the utilities enjoyed by the agents inside block q. This also means that condition 2.15, that has to be verified by the utility functions to have a potential function, corresponds to the Balance Contribution property of the Shapley value taken in the particular case of two agents of different colors.

[15] This result illustrates a theorem presented in Ui (2000), which extends the notion of Shapley value to non cooperative games such as ours.

Drawing the parallel a bit further allows also to highlight that the same difference exists between the potential function of the coalitional game and the grand coalition worth as between the potential function $\mathcal{F}$ and the collective utility U. Whereas $\mathcal{F}$ represents the sum of the agents' utilities at the time when they have moved into their current location starting with a totally empty city (or are considered to have done so), U represents the sum of the agents' utilities once they are all settled. Hence, while stationary configurations maximize $\mathcal{F}$ they do not necessarily (and the following examples show that they generally don't) maximize the collective utility.

2.5.3 Segregation by ethnic origin, income, and preferences for public amenities

Up to now, we have always implicitly supposed that the sole characteristic that the agents use to evaluate a location is the composition of its neighborhood. Other determinants of residential location choice however exist that are not necessarily correlated to neighbors' characteristics, such as local public goods. The red and green labelling of our two groups thus may correspond to two different ethnic origins or two groups with different preferences for local public goods. Tiebout (1956)'s analysis of the sorting induced by local public goods is perhaps the main competitor of Schelling (1971) in terms of its influence on later work on neighborhood choice.

It is very easy to write versions of our model which take into account the agents' preferences for public goods while keeping the existence and properties of a potential function. Noting for example A the set of all the public facilities and $d_{i,a}$ the distance between an agent i and an amenity $a \in A$, the utility of an agent i could be rewritten in a general fashion as

$$u_i(R, G) \longrightarrow u_i(R, G) + \tilde{u}_i(\{d_{i,a}\}_{a \in A}) \tag{2.43}$$

and one could then easily derive the more general form of the potential function

$$\mathcal{F}(x) \longrightarrow \mathcal{F}(x) + \sum_i \tilde{u}_i(\{d_{i,a}\}_{a \in A}) \tag{2.44}$$

This generalized approach could provide a means to correct one of the bias of our analytical model, namely the lack of heterogeneity in locations. However, the extraction of the properties of the stationary states from this condensate global function would become quite challenging, as the dimension of the state variable of the system increases with the number of added amenities.

2.5.4 Taxation

In order to illustrate the usefulness of the potential function, we further consider the introduction of a tax against segregation. The basic concept at the center of a model *à la* Schelling is that of an agent deciding where to move according solely to the utility gain she would achieve if she was to move. Her move affecting her past and new neighbors, an implicit consequence is that she could generate externalities that amount to $\Delta U - \Delta u$ while moving.

Suppose now the existence of a benevolent planner who subsidizes positive externalities and taxes negative externalities. A way to model the action of that planner is to write the probability that a move happens as:

$$Pr\{move\} = \frac{1}{1 + e^{-\Big(\Delta u + \alpha(\Delta U - \Delta u)\Big)/T}} \tag{2.45}$$

where $0 \leq \alpha \leq 1$ is a parameter controlling the tax level. The limit case $\alpha = 0$ corresponds to a standard Schelling model and the limit case $\alpha = 1$ corresponds to a case where only the interest of the collectivity as a whole is taken into account.

Following the path of the proofs developed in section 2.2, one can infer the stationary distribution in the bounded neighborhood framework:

$$\Pi(x) = \frac{e^{\Big((1-\alpha)\mathcal{F}(x) + \alpha U(x)\Big)/T}}{\sum_z e^{\Big((1-\alpha)\mathcal{F}(z) + \alpha U(z)\Big)/T}} \tag{2.46}$$

The potential function can thus in this context be generalized to $(1-\alpha)\mathcal{F}(x) + \alpha U(x)$. We already noted that the configurations maximizing $\mathcal{F}$ are not in general maximizing U and could even in certain cases (asymmetrically peaked utility functions) be very unfavorable to U. We show here that the parameter $\alpha = 1$ of the tax indeed allows the planner to obtain a stationary configuration in which the collective utility is maximized.[16]

Such a Pigouvian tax supposes that the central government has a precise knowledge of the neighborhood composition of each moving agent, which is a rather utopian assumption. We propose here a new variation of Schelling model incorporating a different tax rule, based on more realistic assumptions regarding the goverment intervention ability. We define a simple tax, which aims at preventing the emergence or maintenance of a dominant group in each block. The central government imposes on each agent a *tax*

[16]For development on the level of the tax α necessary or sufficient to break undesired stationary configurations, see Grauwin *et al.* (2009a).

which is defined as:

$$r(S,\,\theta) = \begin{cases} \theta|S/H - 0.5| & \text{if } S/H > 0.5 \\ 0 & \text{otherwise} \end{cases} \tag{2.47}$$

where S is the number of agents similar to the taxed agent within his block and θ is a fixed parameter controlling the tax level. Note that the tax is paid only by every agent belonging to the majority in their district and therefore does not penalize agents in the minority. Note also that this definition of the tax is defined at the scale of the block of size H. This assumption implies a direct competition between the segregative effect of a model *la* Schelling, which favors values of s/H greater than 0.5, and the tax mechanism which favors values of S/H below 0.5.

The utility of red and green agents is then redefined, taking into account the penalty imposed by the tax, as:

$$\begin{cases} \hat{u}_R(R, G, \theta) = u_R(R, G) - r(R,\,\theta) \\ \hat{u}_G(R, G, \theta) = u_G(R, G) - r(G,\,\theta) \end{cases} \tag{2.48}$$

The tax is null for $\theta = 0$. For $\theta > 0$, one can expect that a rise in θ implies a rise in the probability of having non-segregated configurations in the stationary states.

From an analytical viewpoint, the properties of the model resulting from the Markov chain theory remain unchanged: there exists one unique stationary distribution and the final configurations do not depend on the initial ones. It is also straightforward to check that if the pair of utility functions (u_r, u_G) verifies condition 2.15, then so does the pair $(\hat{u}_r, \hat{u}_G)$ and it is hence possible to derive a potential function incorporating the tax effects.

To be more specific, suppose that the utility of the agents is given by the asymmetrically peaked function introduced in section 2.4.3, which is a case where the the stationary configuration clearly diverges from the collective optimum. An agent utility $\hat{u}$ can be rewritten in this case as:

$$\begin{aligned} \hat{u}(S,\,\theta) &= u_{ap,\,m}(S) - r(S,\,\theta) \\ &= \begin{cases} 2S/H - 0 & \text{if } S/H \leq 0.5 \\ 2 - m - 2(1-m)S/H - \theta|S/H - 0.5| & \text{if } S/H > 0.5 \end{cases} \\ &= \begin{cases} 2S/H & \text{if } s_i \leq 0.5 \\ 2 - (m - \theta/2) - 2(1 - (m - \theta/2))S/H & \text{if } S/H > 0.5 \end{cases} \\ &= u_{ap,\,m-\theta/2}(S) \end{aligned} \tag{2.49}$$

Hence our tax mechanism can be interpreted in this case as a direct control by the central government of the asymmetry parameter m in the agents' utility function. According to the analytical results obtained in section 2.4.3, the asymmetry parameter must be inferior to $m_c = 0$ in order to avoid segregation. The equivalence stated in Eq. 2.49 thus allows us to predict that the minimal tax level necessary to break segregated patterns is:

$$\theta_c(m) = 2(m - m_c) = 2m \tag{2.50}$$

Our analytical framework thus allows one to consider the consequences of different tax levels in a very simple way. Other public policies against segregation could also probably be analysed.

2.6 Conclusion

In this chapter, we used recent tools from evolutionary game theory to develop an analytical solution of Schelling segregation model for bounded neighborhoods and two homogeneous groups of agents with general utility functions. This represents a major step forward compared to previous work, mostly based on computer simulations or providing analytical results for specific models. We showed that the stationary configurations reached following the selfish individual moves of the agents maximize a potential function under mild conditions on the agents' utility functions. This potential function can be interpreted as the sum of the agents' utilities as they move into their neighborhood, starting from a totally empty city. In other words, the potential function cumulates the incentives the agents had to move into the neighborhood where they are located. Thanks to this potential function, we are able to solve Schelling model with general utility functions.

This step forward was enabled by a partial reduction in the heterogeneity of agents' neighborhoods through the use of bounded neighborhoods. This allows one to keep track of how each individual move affects the global configuration. Instead, when continuous neighborhoods are used, this information is lost because the way a moving agent affects his past and new neighbors depends on factors (the type of their neighbors' neighbors) that are not fully determined by the agent's decision. Therefore, it is generally impossible to know how an individual move affects a function of the global configuration unless the utility functions are linear.

We used the potential function to assess the outcomes of our location model for different utility functions, representing different degrees of preference for mixed environments. We examined successively linear utility functions, Schelling original utility function and

asymmetrically peaked utility functions. The first two utility functions lead to segregated stationary configurations. In the linear utility case (and for meaningful values of the parameters), the segregated configurations, that maximize the potential function, also maximize the collective utility. With Schelling original utility function, a divergence between collective utility and the potential function appears. Asymmetrically peaked utility functions lead to segregated configurations even for a slight asymmetry, because this asymmetry provokes moves to slightly segregated neighborhoods that will never be compensated by reverse moves. Note finally that when the vacancy rate approaches 0, any pair of utility functions gives a potential function that allows to characterize the stationary configurations.

Our analytical approach helps understanding the ingredients that contribute to the paradoxical result that has generated interest for Schelling model. Even if the dynamics is governed by agents moving to improve their own utility, the evolution leads to city configurations in which most of the agents are far from being satisfied. The results presented in this chapter show rigorously what the two main ingredients of segregation are. First, the most important element driving segregation is the *asymmetry* of the utility function. Symmetric functions do not lead to segregation. Once utility functions favor a majority status over a minority status, segregation is found, even if agents have a strict preference for mixed environments, as in the asymmetrically peaked utility function. The second important element is the existence of externalities. As already noted by Zhang (2004b) and Pancs & Vriend (2007), the existence of externalities explains why individual preferences for integrated environments may lead to segregated configurations. Indeed, location choice by an agent is only based on her own utility level, even if it also affects her neighbors' utility levels. This makes mixed neighborhoods unstable and segregated configurations very stable. The unstability of mixed neighborhoods is particularly clear in the block configuration for the asymmetrically peaked function. Starting with the Nash equilibrium where $R_q = G_q = (H+1)/2$ and $T > 0$, the logit rule implies that there is a positive probability that an agent accepts a slight decrease of his utility, and leaves a block with composition $R_q = G_q = (H+1)/2$. The agents of the same colour remaining in his former block now have a lower utility and are even more likely to leave. This creates an avalanche which empties the block of agents of the same color, as each move away further decreases the utility of the remaining agents. Conversely, highly-segregated configurations are very stable. Indeed, once the city is divided into homogeneous areas, a red agent will have no incentive to go from the red area to the green one, his utility dropping from m to 0.

The analytical tool given here will permit one to consider the outcomes of other

types of utility functions, in particular those that emerge from empirical findings on social preferences. It is now conceivable to analyze the theoretical outcomes of these preferences and possibly to test the effect of introducing public policy instruments aimed at decreasing segregation. It is also worth noticing that with the original utility function suggested by Schelling, the potential function happens to be a linear form of the Duncan and Duncan segregation index. We thus built a bridge between theoretical models of segregation and residential segregation measures.

The kind of solution that is developed here has been used in physics in equilibrium statistical mechanics. Equilibrium statistical mechanics has developed powerful tools to link the microscopic and macroscopic levels. These tools are usually limited to physical systems, where dynamics is governed not by a selfish criterion but by a global quantity such as the total energy. Blume (1993) already built a bridge between statistical physics and a coordination game with local interactions. Here, by using the potential function, which is analogous to state functions in thermodynamics, we have extended the analytical framework of statistical mechanics to Schelling model. By doing so, our work paves the way to analytical treatments of a much wider class of social systems, where dynamics is governed by individual strategies.

Chapter 3

Competition between collective and individual dynamics

3.1 Introduction

The intricate relations between the individual and collective levels are at the heart of many natural and social sciences. Different disciplines wonder how atoms combine to form solids (Cotterill, 2008; Goodstein, 1985), neurons give rise to consciousness (Damasio, 1995; Changeux, 2009) or individuals shape societies (Smith, 1776; Latour, 2007). However, scientific fields assume distinct points of view for defining the "normal", or "equilibrium" aggregated state. Physics looks at the collective level, selecting the configurations that minimize the global free energy (Goodstein, 1985). In contrast, economic agents behave in a selfish way, and equilibrium is attained when no agent can increase its own satisfaction (Mas-Colell *et al.*, 1995). Although similar at first sight, the two approaches lead to radically different outcomes.

In this chapter, we illustrate the differences between collective and individual dynamics on an exactly solvable model, similar to Schelling's segregation model (Schelling, 1971). The model considers individual agents which prefer a mixed environment, with dynamics that lead to segregated or mixed patterns at the global level. A "tax" parameter monitors continuously the agents' degree of altruism or cooperativity, i.e., their consideration of the global welfare. At high degrees of cooperativity, the system is in a mixed phase of maximal utility. As the altruism parameter is decreased, a phase transition occurs, leading to segregation. In this phase, the agents' utilities remain low, in spite of continuous efforts to maximize their satisfaction. This paradoxical result of

Schelling's segregation model (Schelling, 1971) has generated an abundant literature. Many papers have simulated how the global state depends on specific individual utility functions, as reviewed by (Clark & Fossett, 2008). There have been attempts at solving Schelling's model analytically, in order to provide more general results concerning the consequences of individual preferences on segregation levels (Pollicott & Weiss, 2001; Zhang, 2004b; Dokumaci & Sandholm, 2007). However, these are limited to specific utility functions. More recently, physicists have tried to use a statistical physics approach to understand the segregation transition (Dall'Asta *et al.*, 2008; Vinkovic and Kirman, 2006; Gauvin *et al.*, 2009). The idea seems promising, since statistical physics has successfully bridged the micro-macro gap for physical systems governed by collective dynamics. However, progress was slow by lack of an appropriate framework allowing for individual dynamics (Dall'Asta *et al.*, 2008). In this chapter, we introduce a rigorous generalization of the physicist's free energy, which includes individual dynamics. By introducing a "link" state function[1] which is maximized in the stationary state, we pave the way to analytical treatments of a much wider class of systems, where dynamics is governed by individual strategies. As an example, we provide a quantitative solution to Schelling's segregation model for very general utility functions.

3.2 Model

Our model represents in a schematic way the dynamics of residential moves in a city. For simplicity, we include one type of agent, but our results can readily be generalized to deal with agents of two "colors", as in the original Schelling model (Schelling, 1971) (see below and appendix B). The city is divided into Q blocks ($Q \gg 1$), each block containing H cells or flats (Fig 3.1). We assume that each cell can contain at most one agent, so that the number n_q of agents in a given block q ($q = 1, \ldots, Q$) satisfies $n_q \leq H$, and we introduce the density of agents $\rho_q = n_q/H$. Each agent has the same utility function $u(\rho_q)$, which describes the degree of satisfaction concerning the density of the block he is living in. The collective utility is defined as the total utility of all the agents in the city: $U(x) = H \sum_q \rho_q u(\rho_q)$, where $x \equiv \{\rho_q\}$ corresponds to the coarse-grained configuration of the city, i.e. the knowledge of the density of each block. For a given x, there is a large number of ways to arrange the agents in the different cells. This number of arrangements is quantified by its logarithm $S(x)$, called the entropy of the configuration x.

[1]Word of caution: in this chapter and in the following, we use the notation L (as link) to designate the potential function previously called $\mathcal{F}$. In these two more 'physics-centered' chapters, we keep the notation F to designate an analogue the notion of free energy introduced in physics.

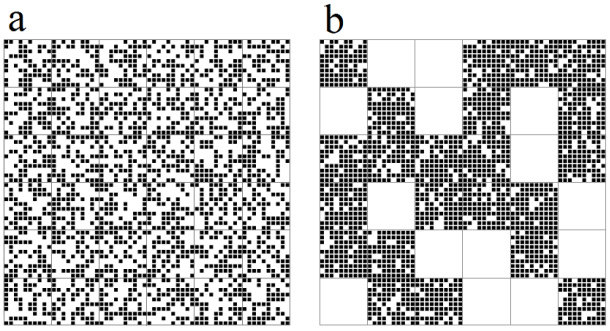

Figure 3.1: Configurations of a city composed of $Q = 36$ blocks containing each $H = 100$ cells, with $\rho_0 = 1/2$. **(a) Mixed** state. Stationary state of the city for $m = 0.5$, $\alpha = 1$ and $T \to 0$. Agents are distributed homogeneously between the blocks, each of them having a density of 0.5. **(b) Segregated** configuration. Stationary state of the city for $m = 0.5$, $\alpha = 0$ and $T \to 0$. Agents are gathered on 20 blocks of mean density 0.9, the other blocks being empty. In the original Schelling model Schelling (1971), each agent has a distinct neighborhood, defined by its 8 nearest neighbors. Here, we only keep the essential ingredient of blocks of distinct densities. Our model shows the same qualitative behavior as Schelling's but can be solved exactly, thanks to the *partial* reduction of agent's heterogeneity.

The dynamical rule allowing the agents to move from one block to another is the following. At each time step, one picks up at random an agent and a vacant cell, within two different blocks. Then the agent moves in that empty cell with probability:

$$P_{xy} = \frac{1}{1 + e^{-\mathcal{C}/T}}, \tag{3.1}$$

where x and y are respectively the configurations before and after the move, and $\mathcal{C}$ is the cost associated to the proposed move. The positive parameter T is a "temperature" which introduces in a standard way (Anderson *et al.*, 1992) some noise on the decision process. It can be interpreted as the effect of features that are not explicitly included in the utility function but still affect the moving decision (urban facilities, friends...). We write the cost $\mathcal{C}$ as :

$$\mathcal{C} \quad = \quad \Delta u + \alpha(\Delta U - \Delta u) \tag{3.2}$$

where Δu is the variation of the agent's own utility upon moving and ΔU is the variation of the total utility of all agents. The parameter $0 \leq \alpha \leq 1$ weights the contribution of the other agents' utility variation in the calculation of the cost $\mathcal{C}$, and it can thus

be interpreted as a degree of cooperativity (or altruism). For $\alpha = 0$, the probability to move only depends on the selfish interest of the chosen agent, which corresponds to the spirit of economic models such as Schelling's. When $\alpha = 1$, the decision to move only depends on the collective utility change, as in physics' models. An economical interpretation could be that individual moves are controlled by a central government, via a tax that internalizes all the externalities (more on this below). Varying α in a continuous way, one can interpolate between the two limiting behaviors of individual and collective dynamics.

3.3 Results

We wish to find the stationary probability distribution $\Pi(x)$ of the microscopic configurations x. If the cost $\mathcal{C}$ can be written as $\mathcal{C} = \Delta V \equiv V(y) - V(x)$, where $V(x)$ is a function of the configuration x, then the dynamics satisfies detailed balance Evans *et al.* (2005) and the distribution $\Pi(x)$ is given by

$$\Pi(x) = \frac{1}{Z} e^{F(x)/T}, \tag{3.3}$$

with $F(x) = V(x) + TS(x)$ and Z a normalization constant. The entropy has for large H the standard expression $S(x) = H \sum_q s(\rho_q)$, with

$$s(\rho) = -\rho \ln \rho - (1-\rho)\ln(1-\rho). \tag{3.4}$$

We now need to find the function $V(x)$, if it exists. Given the form (3.2) of $\mathcal{C}$, finding such a function $V(x)$ amounts to finding a "linking" function $L(x)$, connecting the individual and collective levels, such that $\Delta u = \Delta L$. The function V would thus be given by $V(x) = (1-\alpha)L(x) + \alpha U(x)$. By analogy to the entropy, we assume that $L(x)$ can be written as a sum over the blocks, namely $L(x) = H \sum_q \ell(\rho_q)$. Considering a move from a block at density ρ_1 to a block at density ρ_2, ΔL reduces in the large H limit to $\ell'(\rho_2) - \ell'(\rho_1)$, where ℓ' is the derivative of ℓ. The condition $\Delta u = \Delta L$ then leads to the identification $\ell'(\rho) = u(\rho)$, from which the expression of $\ell(\rho)$ follows:

$$\ell(\rho) = \int_0^\rho u(\rho')d\rho'. \tag{3.5}$$

As a result, the function $F(x)$ can be expressed in the large H limit as $F(x) = H \sum_q f(\rho_q)$, with a block potential $f(\rho)$ given by :

$$\begin{aligned} f(\rho) &= -T\rho\ln\rho - T(1-\rho)\ln(1-\rho) \\ &+ \alpha\rho u(\rho) + (1-\alpha)\int_0^\rho u(\rho')d\rho'. \end{aligned} \tag{3.6}$$

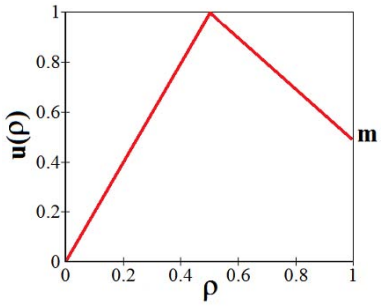

Figure 3.2: Asymmetrically peaked individual utility as a function of block density. The utility is defined as $u(\rho) = 2\rho$ if $\rho \leq 1/2$ and $u(\rho) = m+2(1-m)(1-\rho)$ if $\rho > 1/2$, where $0 < m < 1$ is the asymmetry parameter. Agents strictly prefer half-filled neighborhoods ($\rho = 1/2$). They also prefer overcrowded ($\rho = 1$) neighborhoods to empty ones ($\rho = 0$).

The probability $\Pi(x)$ is dominated by the configurations $x = \{\rho_q\}$ that maximize the sum $\sum_q f(\rho_q)$ under the constraint of a fixed $\rho_0 = 1/Q\sum_{q=1}^{Q}\rho_q$. To perform this maximization procedure, we follow standard physics methods used in the study of phase transitions (like liquid-vapor coexistence (Callen, 1985)), which can be summarized as follows. If $f(\rho)$ coincides with its concave hull at a given density ρ_0, then the state of the city is homogeneous, and all blocks have a density ρ_0. Otherwise, a phase separation occurs: some blocks have a density $\rho_1^* < \rho_0$, while the others have a density $\rho_2^* > \rho_0$ (see Appendix B).

Interestingly, the potential $F = (1-\alpha)L + \alpha U + TS$ appears as a generalization of the notion of free energy introduced in physical systems. Mapping the global utility U onto the opposite of the energy of a physical system, it turns out that for $\alpha = 1$, the maximization of the function $U + TS$ is equivalent to the minimization of the free energy $E - TS$. For $\alpha < 1$, the potential F takes into account individual moves through the link function L. Furthermore, the potential F can be calculated for arbitrary utility functions, allowing to predict analytically the global town state. Such an achievement eluded so far individualistic, Schelling-type models, which had to be solved through numerical simulations (Clark & Fossett, 2008).

To obtain explicitly the equilibrium configurations, one needs to know the specific

	$m < 2/3$	$2/3 \leq m \leq 1$
ρ_s	$\frac{1}{2}\sqrt{(2-m)/(1-m)}$	1
U^*	$\frac{1}{1+\sqrt{(1-m)/(2-m)}}$	m
L^*	$\frac{1}{1+\sqrt{(1-m)/(2-m)}}$	$1/2 + m/4$

Table 3.1: **Characteristics of the segregated equilibrium.** The table displays the density ρ_s in the non-empty blocks, the normalized collective utility U^* and the normalized link L^* of the stationary configurations obtained for $\alpha = 0$. It is straightforward to check that $U^*(m) \leq 1$ and $L^*(m) \geq 1/2$ for $m \leq 1$.

form of the utility function. To illustrate the dramatic influence of the cooperativity parameter α, we use the asymmetrically peaked utility function (Pancs & Vriend, 2007), which indicates that agents prefer mixed blocks (Fig 3.2). The overall town density is fixed at $\rho_0 = 1/2$ to avoid the trivial utility frustration resulting from the impossibility to attain the optimal equilibrium ($\rho_q = 1/2$ for all blocks). We also consider for simplicity the limit $T \to 0$, to avoid entropy effects. The qualitative behaviour of the system is unchanged for $\rho_0 \neq 1/2$ or for low values of the temperature, as shown in the appendix B.

In the collective case ($\alpha = 1$), the optimal state corresponds to the configuration that maximizes the global utility, which can be immediately guessed from Figure 3.2, namely $\rho_q = 1/2$ for all q (Fig 3.1a). On the contrary, in the selfish case ($\alpha = 0$, Fig 3.1b), maximization of the potential $F(x)$ shows that the town settles in a segregated configuration where a fraction of the blocks are empty and the others have a density $\rho_s > 1/2$. Surprisingly, the city settles in this state of low utility in spite of agents' continuous efforts to maximize their own satisfaction. To understand this frustrated configuration, note that the collective equilibrium ($\rho_q = 1/2$ for all q) is now an *unstable* Nash equilibrium at $T > 0$. The instability can be understood by noting that at $T > 0$ there is a positive probability that an agent accepts a slight decrease of its utility, and leaves a block with density $\rho_q = 1/2$. The agents remaining in its former block now have a lower utility and are more likely to leave to another $\rho_q = 1/2$ block. This creates an avalanche which empties the block, as each move away further decreases the utility of the remaining agents. This avalanche stops when the stable (Nash) equilibrium, given by the maximum of the potential, is reached. To understand the transition between mixed and segregated configurations, it is instructive to calculate the values of both the overall utility and the potential, for different values of m (at $\alpha = 0$). In case of

a homogeneous town (an unstable Nash equilibrium for which $\rho_q = 1/2$ for all q), the normalized collective utility is given by $U^* = U/(\rho_0 HQ) = u(\rho_0 = 1/2) = 1$. The normalized link equals, for all m, $L^* = L/(\rho_0 HQ) = \ell(\rho_0)/\rho_0 = 1/2$, where ℓ is given in Eq. 3.5. The values of L^* and U^* displayed in Table 3.1 show that the utility of the segregated equilibrium is lower *but* that its potential is higher, explaining its stability. Note that the gap between the link values of the homogeneous and segregated configurations increases with m.

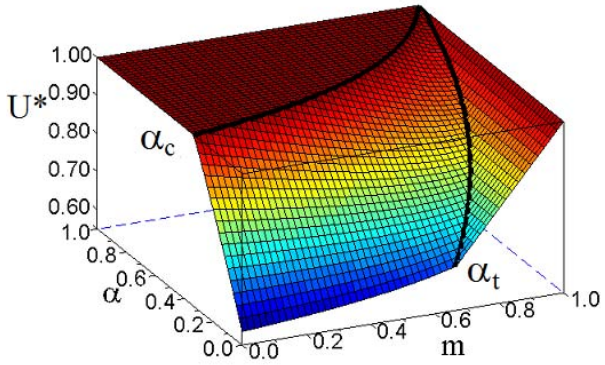

Figure 3.3: Phase diagram of the global utility as a function of the cooperativity α and the asymmetry m, at $T \to 0$ and $\rho_0 = 1/2$. The average utility per agent $U^* = U/(\rho_0 HQ)$ is calculated by maximizing the potential $F(x)$ for the peaked utility shown in Fig. 3.2, see Appendix B. The plateau at high values of α corresponds to the mixed phase of optimal utility, which is separated from the segregated state by a phase transition arising at $\alpha_c = 1/(3-2m)$. The overall picture is qualitatively unchanged for low but finite values of the temperature, see appendix B.

This helps understanding why the greater the m, the greater the value of tax parameter necessary to reach the homogeneous configuration. Indeed, the segregated states are separated from mixed ones by a phase transition at the critical value $\alpha_c = 1/(3-2m)$ - which increases with m (Figure 3.3). This transition differs from standard equilibrium phase transitions known in physics, which are most often driven by the competition between energy and entropy. Here, the transition is driven by a competition between the collective and individual components of the agents' dynamics. The unsatisfactory global state of the city can be interpreted, from the economics' point of view, as an effect of externalities: by moving to increase its utility, an agent may decrease other agents' utilities, without taking this into account. From a standard interpretation in terms of Pigouvian tax (Auerbach, 1985), one expects that $\alpha = 1$ is necessary to reach the optimal state, since by definition this value internalizes all the externalities the agent causes

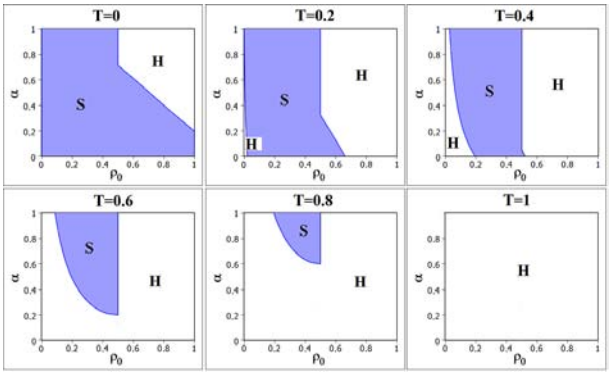

Figure 3.4: **Phase diagrams** for the asymmetrically peaked individual utility (Fig. 3.2, with $m = 0.8$) for different values of T. Increasing the temperature T tends to favour homogeneous states. For small but finite temperatures (roughly $T < 0.2$), the phase diagram is modified only for extremal values of ρ_0, as expected from the entropic term $Ts(\rho) = -T\rho\ln\rho - T(1-\rho)\ln(1-\rho)$. As T is increased, the whole diagram is affected by the entropic term. Compared to the $T = 0$ case, the main change is the appearance of a second homogeneous phase for $\rho_0 < 1/2$. But while for $\rho_0 > 1/2$ homogeneity corresponds to the optimal choice for the agents, for $\rho_0 < 1/2$, collective utility is not maximized in a homogeneous city. The city is homogeneous by noise, not by choice. Note that an increase in α tends to reduce this domain, while it tends to increase the homogeneous domain for $\rho_0 > 1/2$.

to the others when moving. Our results show that the optimal state is maintained until much lower tax values (for example, $\alpha_c = 1/3$ at $m = 0$), a surprising result which deserves further analysis. Another interesting effect is observed for $m > 2/3$ (Figure 3.3). Introducing a small tax has no effect on the overall satisfaction, the utility remaining constant until a threshold level is attained at $\alpha_t = (3m-2)/(6-5m)$.

We focused up to now on the zero temperature limit. For low temperatures, the main qualitative conclusions are not modified, as the phase diagram is modified only for extremal values of ρ_0 by entropic contributions. At higher temperatures the city tends to become homogeneous, as the effect of "noise" (i.e., of the features that are not described in the model) dominates over the utility associated to density of the blocks (see Fig. 3.4).

3.4 Discussion

There are two main differences between our simple model and Schelling's original model (Schelling, 1971) : the existence of agents of two colors and the definition of the agent's neighborhoods. We now show that these additional features do not introduce any essential effect.

Let us start by introducing agents of two "colors" (such as red and green). Simple calculations (see appendix B) show that for two species which only care about the density of neighbors of their own color, the block potential (eq. 3.6) becomes :

$$\begin{aligned} f(\rho_R,\rho_G) &= -T\rho_R \ln \rho_R - T\rho_G \ln \rho_G \\ &- T(1-\rho_R-\rho_G)\ln(1-\rho_R-\rho_G) \\ &+ \alpha\Big[\rho_R\, u_R(\rho_R) + \rho_G\, u_G(\rho_G)\Big] \\ &+ (1-\alpha)\Big[\int_0^{\rho_R} u_R(\rho')d\rho' + \int_0^{\rho_G} u_G(\rho')d\rho'\Big] \end{aligned}$$

with straightforward notations (for example $u_R(\rho_R)$ represents the utility of a red agent in a block with a density ρ_R of red agents). In the more general case of utility functions depending on both the density of similar and dissimilar neighbors, it is also possible to derive a block potential if the utility functions verify a symmetry constraint. This constraint is not very restrictive, in the sense that no qualitative feature of the model is lost when one restrains the study to utilities that verify it (see the appendix B).

Finding the equilibrium configurations amounts to finding the set $\{\rho_{qR}, \rho_{qG}\}$ which maximizes the potential $F(x) = \sum_q f(\rho_{qR}, \rho_{qG})$ with the constraints $\sum_q \rho_{qR} = Q\rho_{0R}$ and $\sum_q \rho_{qG} = Q\rho_{0G}$, where ρ_{0G} and ρ_{0R} represent respectively the overall concentration of green and red agents.

Because of the spatial constraints (the densities of red and green agents in each block q must verify $\rho_{qR} + \rho_{qG} \leq 1$), the 'two populations' model can not formally be reduced to two independent 'one population' models. However, the stationary states can still be easily computed. Let us focus once again on the $T \to 0$ limit and suppose for example that $\rho_{0R} = \rho_{0G} = \rho_0/2$. The stationary states depends once again on the values of ρ_0, m and α. For low values of α, it can be shown that the system settles in segregated states where each block contains only one kind of agent with a density ρ_0 (see Figure 3.5a). For $\alpha \geq \alpha_c$, the system settles in mixed states where the density of a group in a block is either 0 or 1/2 (see Figure 3.5b). The reader is referred to appendix B for more details.

We now turn to the difference in agent's neighborhoods. In Schelling's original model, agents' neighbors are defined as their 8 nearest neighbors. Our model considers instead predefined blocks of common neighbors. First, it should be noted that there is no decisive argument in favor of either neighborhood definition in terms of the realism of the description of real social neighborhoods. Second, we note that introducing blocks allows for an analytical solution for arbitrary utility functions. This contrasts with the nearest neighbor case, where the best analytical approach solves only a modified model which abandons the individual point of view and is limited to a specific utility function (Zhang, 2004b). Finally, the simulations presented on Figures 3.5 show that the transition from segregated to mixed states is not affected by the choice of the neighborhood's definition. We conclude that the block description is more adapted to this kind of simple modelling, which aims at showing stylized facts as segregation transitions.

Our simple model raises a number of interesting questions about collective or individual points of view. In the purely collective case ($\alpha = 1$), the stationary state corresponds to the maximization of the average utility, in analogy to the minimization of energy in physics. In the opposite case ($\alpha = 0$), the stationary state strongly differs from the simple collection of individual optima (Kirman, 1992): the optimization strategy based on purely individual dynamics fails, illustrating the unexpected links between micromotives and macrobehavior (Schelling, 1978). However, the emergent collective state can be efficiently captured by the maximization of the linking function $\ell(\rho)$ given in Eq. (3.5), up to constraints in the overall town density. This function intimately connects the individual and global points of view. First, it depends only on the global town configuration (given by the ρ_q), allowing a relatively simple calculation of the equilibrium. At the same time, it can be interpreted as the sum of the *individual* marginal utilities gained by agents as they progressively fill the city after leaving a reservoir of zero utility. In the stationary state, a maximal value of the potential L is reached. This means that no agent can increase its utility by moving (since $\Delta u = \Delta L$), consistently with the economists' definition of a Nash equilibrium.

Equilibrium statistical mechanics has developed powerful tools to link the microscopic and macroscopic levels. These tools are limited to physical systems, where dynamics is governed by a global quantity such as the total energy. By introducing a link function, analogous to state functions in thermodynamics or potential functions in game theory Monderer & Shapley (1996), we have extended the framework of statistical mechanics to a Schelling-like model. Such an approach paves the way to analytical treatments of a much wider class of systems, where dynamics is governed by individual strategies.

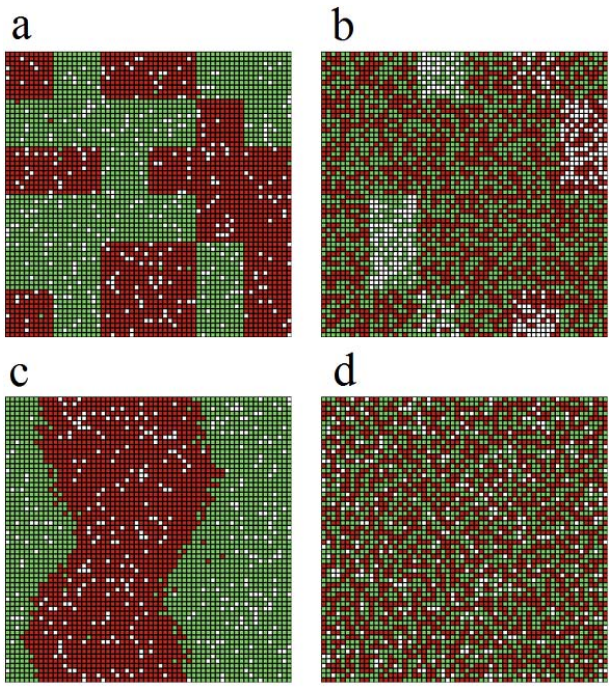

Figure 3.5: Stationary configurations obtained by simulating the evolution of a city inhabited by a equal number of red and green agents whose preference are given by the asymmetrically peaked utility function ($m = 0.5$). The rate of vacant cells (in white) is fixed to 10%. **Top panel.** The city is divided into blocks of size $H = 100$. In accordance with the analytic model, a segregated configuration is obtained when $\alpha = 0$ (snapshot a) and a more homogeneous configuration is obtained for $\alpha = 1$ (snapshot b). **Bottom panel.** The utility of an agent depends on the local density of similar neighbors computed on the $H = 108$ nearest cells. While of different topological nature, a segregated configuration is still obtained for $\alpha = 0$ (snapshot c) and a homogeneous configuration is still obtained for $\alpha = 1$ (snapshot d). In all those simulations, we take $T = 0.1$. The small amount of noise hence generated, while not changing the nature of the stationary states compared to the case $T \to 0$, conveniently reduces the time of convergence of the system.

Chapter 4

Effective Free Energy for Individual Dynamics

4.1 Introduction

The intricate relations between the individual and collective levels are at the heart of many natural and social sciences (Grauwin *et al.*, 2009a). Physics looks at the collective level, selecting the configurations that minimize the global free energy (Goodstein, 1985). In contrast, economic agents behave in a selfish way, and equilibrium is attained when no agent can increase its own satisfaction (Mas-Colell *et al.*, 1995).

Recently, physicists have tried to use statistical physics approaches to understand social phenomena such as the segregation transition (Dall'Asta *et al.*, 2008; Vinkovic and Kirman, 2006). The idea seems promising because statistical physics has successfully bridged the gap between the micro and macroscopic levels for physical systems governed by collective dynamics. However, progress remained slow due to the lack of an appropriate framework allowing to take into account the selfish dynamics typical of socio-economic agents. On the other hand, game theorists have developed in the last decades the notion of Potential Games, in which each player's gain resulting from a change of state is equal to the variation of a potential function (Anderson *et al.*, 1992; Young, 1993; Monderer & Shapley, 1996; Ui, 2000). This potential function hence provides what physicists need: a link between individual and collective levels.

In this chapter, we propose a generic analytical framework that builds on concepts originating both from statistical physics and game theory. We introduce a rigorous generalization of the physicist's free energy, which encompasses individual dynamics. By introducing a "link" state function that is maximized in the stationary state, we

pave the way to analytical treatments of a much wider class of systems, where dynamics is governed by individual strategies. Quantitative solutions of two models are also provided as examples.

4.2 The Model

4.2.1 Generic model

Our model represents in a schematic way the dynamics of agents making individual choices. Throughout this chapter, $\mathcal{N} = \{1, \ldots, N\}$ denotes a finite set of agents, $q_i \in \{1, \ldots, Q\}$ the choice of agent $i \in \mathcal{N}$, the vector $\vec{q} = (q_i)_{i \in \mathcal{N}}$ describing the state of the system and $\mathcal{N}_q$ (resp. n_q) the set (resp. number) of agents following choice $q \in \{1, \ldots, Q\}$. Each agent $i \in \mathcal{N}$ can moreover be characterized by his utility function, which describes the degree of satisfaction concerning his choice. An agent's utility function is supposed to depend only on his own choice and on the set $\mathcal{N}_{q_i}$ of agents making the same choice, namely $u_i(\vec{q}) = u_i(q_i, \mathcal{N}_{q_i})$. We also introduce the collective utility, defined as the total utility of all the agents: $U(\vec{q}) = \sum_i u_i(\vec{q})$.

The dynamical rule allowing the agents to change their choice is the following. At each time step, one picks up at random an agent and a choice $q^* \in \{1, \ldots, Q\}$. Then the agent goes from choice q_i to choice $q_i' = q^*$ with probability:

$$P_{q_i \to q_i'} = \frac{1}{1 + e^{-\Delta u_i / T}}, \tag{4.1}$$

where $\Delta u_i = u_i(\vec{q'}) - u_i(\vec{q})$ is the variation of the agent's own utility upon his change of choice. The parameter $T > 0$ is a "temperature" that introduces in a standard way some noise on the decision process (Anderson *et al.*, 1992). It can be interpreted as the effect of features that are not explicitly included in the utility function but still affect the decision.

We wish to find the stationary probability distribution $\Pi(\vec{q})$ of the microscopic configurations $\vec{q}$. If Δu_i can be written as $\Delta u_i = \Delta L \equiv L(\vec{q'}) - L(\vec{q})$, where $L(\vec{q})$ is a state function of the configuration $\vec{q}$, then the dynamics satisfies detailed balance (Evans *et al.*, 2005) and the distribution $\Pi(\vec{q})$ is given by

$$\Pi(\vec{q}) = \frac{1}{Z} e^{L(\vec{q})/T}, \tag{4.2}$$

with Z a normalization constant.

It can be shown (Grauwin *et al.*, 2009a; Ui, 2000) that a sufficient and necessary

condition[1] for a "linking" function L to exist is

$$u_i(q_i, \mathcal{N}_{q_i} \setminus \{j\}) - u_i(q_i, \mathcal{N}_{q_i}) = u_j(q_j, \mathcal{N}_{q_j} \setminus \{i\}) - u_j(q_j, \mathcal{N}_{q_j}) \tag{4.3}$$

for any $i \in \mathcal{N}$ and $j \in \mathcal{N}$. Note that this relation is automatically satisfied in the case $q_i \neq q_j$. Eq. (2.15) expresses a symmetric condition on the utility variation (or *externality*) an agent produces on another one when he changes his choice. Condition Eq. (4.3) is also rather easy to satisfy in case of homogeneous agents sharing the same utility function (see examples in section 4.3). In contrast, this condition imposes more restriction to models with heterogeneous agents and explicit examples are then more difficult to build.

Interestingly, the linking function L appears as (the opposite of) an effective energy in terms of physical systems (Eq. (4.2) being the analogue of a Gibbs distribution), but also corresponds to the notion of potential function in game theory (Monderer & Shapley, 1996; Ui, 2000).

4.2.2 Homogeneous agents

In the following, we restrict our study to models where the agents' utility functions can be written as $u_i(\vec{q}) = u(q_i, n_{q_i}/H)$, where H is a parameter characterizing the typical number of agents making a given choice (for instance the natural capacity of an infrastructure). This parameter is assumed to scale linearly with N, the ratio $h = H/N$ being fixed. This particular form of utility function implies that the agents share homogeneous properties and that they are sensitive to the relative proportion of agents making the same choice as them. It also implies that Eq (4.3) is verified, meaning that a linking function L always exists. It is straightforward to check that it can be written as :

$$L(\vec{q}) = \sum_{q=1}^{Q} \sum_{m=0}^{n_q(\vec{q})} u(q, m/H) \tag{4.4}$$

In the limit $N \to \infty$ with $\rho_q = n_q/H$ fixed, one finds

$$L(\vec{q}) \quad \to \quad hN \sum_{q=1}^{Q} \int_0^{\rho_q} u(q, \rho)\, d\rho. \tag{4.5}$$

[1] If the state function L exists, the relation $\Delta u_i = \Delta L$ requires only that it be defined up to a constant.

This particular form of the potential function L allows us to interpret it as the sum of the individual marginal utilities gained by agents as they progressively enter the system after leaving a reservoir of zero utility.

Since the agents are supposed to be identical (but still distinguishable), it seems natural to keep track of "mesoscopic" observables such as the coarse-grained states $x \equiv \{\rho_q\}$ rather than the "microscopic" states $\vec{q}$. The number of states $\vec{q}$ corresponding to a given coarse-grained state x is quantified by its logarithm:

$$S(x) = \ln \frac{N!}{\prod_q n_q!} \quad \xrightarrow[N\to\infty]{h,\{\rho_q\} \text{ fixed}} \quad -N \ln h - hN \sum_{q=1}^{Q} \rho_q \ln \rho_q. \tag{4.6}$$

The stationary distribution of the coarse-grained configurations hence takes the form:

$$\Pi_{N,T}(x) = \frac{1}{Z_{N,T}} e^{F(x)/T} = \frac{1}{Z'_{N,T}} e^{(hN/T)\sum_q f_{N,T}(q,\rho_q)} \tag{4.7}$$

where $F(x) \equiv L(x) + TS(x)$ can be seen as (the opposite of) an effective free energy of the system, $Z'_{N,T} = Z_{N,T} h^{-N}$ and where

$$f_{N,T}(q,\rho) = \sum_{m=0}^{\rho H} u(q, m/H) - T\rho \ln \rho \tag{4.8}$$

In the limit $N \to \infty$ with $\rho_q = n_q/H$ fixed, one finds

$$f_{N,T}(q,\rho) \quad \to \quad f_{\infty,T}(q,\rho) \equiv \int_0^{\rho} u(q,\rho')\, d\rho' - T\rho \ln \rho \tag{4.9}$$

According to the form of the distribution $\Pi_{N,T}$ given by Eq (4.7), in the limit of large N the stationary configurations are those that maximize the sum $\sum_q f_{\infty,T}(q,\rho_q)$ under the constraint $h\sum_q \rho_q = 1$. We explore different maximization procedures in examples of applications presented in next section.

4.3 Applications

4.3.1 Road congestion

We apply here our generic model framework to a simplified version of Chu's (Chu, 1995) congestion model. In this model, a number N of identical commuters travel every morning from home to work. All agents travel on the same road and wish to arrive at

time t^*. Since congestion is a collective phenomenon that no single agent can master to arrive at her preferred arrival time t^*, agents have to choose a less optimal arrival time $t \in \mathbb{Z}$ (time is supposed to be discrete) in order to minimize the private trip cost, $c(t)$, which includes two parts. The first part is the travel time cost $\alpha TT(t)$ where α is the unit cost of travel time and $TT(t)$ is the travel time. The second part is the schedule delay cost, which is $\beta(t^* - t)$ if one arrives early and $\nu(t - t^*)$ if one arrives late, β and ν being unit costs of schedule delay. To make analytical calculations possible, the travel time is supposed to depend on the number n_t of commuters arriving at time t through the function $TT(t) = (n_t/H)^\gamma$ where H and γ are fixed parameters. The parameter H can be interpreted as a standard road capacity (the linearity between H and N can thus reflect that bigger roads are built when the traffic is more important) and the parameter γ measures the elasticity of travel time with respect to n_t.

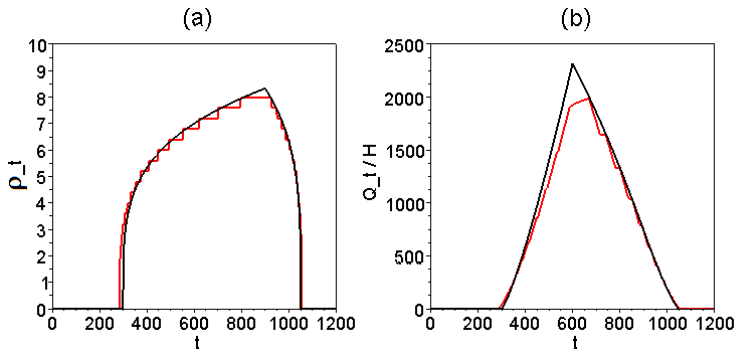

Figure 4.1: **Stationary values** derived by a maximization of L for $N = 5000$ (in red) and $N \to \infty$ (in black) commuters. **(a) Normalized proportion of arriving agents** $\rho_t = n_t/H$. **(b) Normalized proportion of agents on the road** Q_t/H, where Q_t is the number of agent queueing on the road at a given time t. It is computed as the difference between the number of departed and arrived agents $Q_t = \sum_{t'=0}^{t} (d_{t'} - n_{t'})$, where the distribution of departure time $\{d_t\}$ is deduced from the distribution of arrival time $\{n_t\}$ and travel time $\{TT(t)\}$. See Chu (1995) for more details on this procedure. The computations have been realized in the limit $T \to 0$, with the parameters values $t^* = 900$, $\alpha = 2$, $\beta = 1$, $\nu = 4$, $\gamma = 4$, $h = 4.\,10^{-4}$.

This congestion model fits our framework model, the utility of an agent arriving at time t being

$$u(n_t) = -c(t) = -\begin{cases} \alpha(n_t/H)^\gamma + \beta|t^* - t| & \text{if } t < t^* \\ \alpha(n_t/H)^\gamma + \nu|t^* - t| & \text{if } t \geq t^* \end{cases} \tag{4.10}$$

The stationary coarse-grained configurations in the limit of large N are thus those that maximize the sum $\sum_t f_{\infty,T}(t,\rho_t)$, where:

$$f_{\infty,T}(t,\rho) = \begin{cases} \frac{\alpha}{\gamma+1}\rho^{\gamma+1} + \beta\rho|t^* - t| - T\rho\ln\rho & \text{if } t < t^* \\ \frac{\alpha}{\gamma+1}\rho^{\gamma+1} + \nu\rho|t^* - t| - T\rho\ln\rho & \text{if } t \geq t^* \end{cases} \tag{4.11}$$

under the constraint $h\sum_t \rho_t = 1$. Since the sum $\sum_t f_{\infty,T}(t,\rho_t)$ is maximized by the stationary configuration, the stationary values of $\{\rho_t\}$ verify (for any t such that $\rho_t \neq 0$) the relation:

$$\frac{\partial f_{\infty,T}}{\partial\rho}(t,\rho_t) = \frac{\partial f_{\infty,T}}{\partial\rho}(t^*,\rho_{t^*}) \tag{4.12}$$

which can be easily derived using Lagrange multipliers. Eq (4.12) provides for each time step t an implicit relation between ρ_t and ρ_{t^*} (resp. the normalized densities at time t and t^*). All these implicit relations along with the conservation of the number of agents expressed by the condition $h\sum_t \rho_t = 1$ allow one to compute numerically the distribution $\{\rho_t\}$. Fig. 4.1 presents results that have been computed with this method, for $N \to \infty$ and for a finite value ($N = 5000$) of the number of agents obtained by solving numerically an equivalent of Eq. (4.12). These results suggest that finite size effects remain small as soon as N reaches values of a few thousands, and that the main properties of the model can be studied in the limit $N \to \infty$. Notice that our analysis of the congestion model in terms of a potential games allows us to determine (numerically) the stationary state of the systems even for finite size, which is rarely possible with usual economics analysis since, quoting (Otsubo & Rapoport, 2008), *the common practice in transportation science and economics is to use continuous models for analyzing phenomena that are essentially discrete.*

4.3.2 Residential choice

As a second example, we apply our generic framework to a Schelling-like model describing the dynamics of residential moves in a city (Grauwin *et al.*, 2009a; Schelling, 1978). The virtual city is divided into Q blocks ($Q \gg 1$), each block containing H cells or flats (Fig 4.2). We assume that each cell can contain at most one agent, so that the number n_q of agents in a given block satisfies $n_q \leq H$. An agent's utility depends only on the density $\rho_q = n_q/H$ of the block he is living in, ie $u_i(\vec{q}) = u(\rho_q)$, with the convention that $u(\rho) = -\infty$ for $\rho > 1$.

Once again, this residential model can be solved thanks to our framework. With the particular choice of the asymmetrically peaked utility function given Fig 4.2(a), the stationary coarse-grained configurations in the limit of large N are thus those which

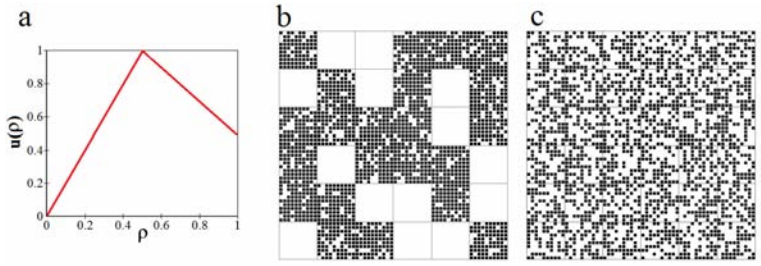

Figure 4.2: **(a)**. Asymmetrically peaked individual utility as a function of block density. The utility is defined as $u(\rho) = 2\rho$ if $\rho \leq 1/2$ and $u(\rho) = 3/2 - \rho$ if $\rho > 1/2$. Agents strictly prefer half-filled neighborhoods ($\rho = 1/2$). They also prefer overcrowded ($\rho = 1$) neighborhoods to empty ones ($\rho = 0$). **(b)** Stationary configurations of a virtual city for $T \to 0$. Blocks are separated in two phases with different densities of agents. **(c)** Stationary configurations of a virtual city for $T \gg 1$. Agents are distributed homogeneously between the blocks. The city is composed of $Q = 36$ blocks containing each $H = 100$ cells, with a mean density $\langle \rho_q \rangle = N/HQ = 1/2$.

maximize the sum $\sum_q f_{\infty,T}(q, \rho_q)$, where[2] :

$$f_{\infty,T}(q,\rho) = \begin{cases} \rho^2 - T\rho\ln\rho & \text{if } \rho < 1/2 \\ -\rho^2/2 + 3\rho/2 - 3/8 - T\rho\ln\rho & \text{if } \rho \geq 1/2 \end{cases} \tag{4.13}$$

To perform the maximization procedure, one can follow standard physics methods used in the study of phase transitions (like liquid-vapor coexistence (Callen, 1985)). In this case, it is a simple exercise to determine that for $N \gg 1$ and in the limit $T \to 0$, a phase separation occurs. A fraction $\min\left(2\sqrt{3}/3hQ,\, 1\right)$ of the blocks have a density $\min\left(\sqrt{3}/2,\, 1/hQ\right)$ while the others are empty. This result is illustrated on Fig. 4.2. For more details on the maximization procedure, the interested reader is referred to (Grauwin *et al.*, 2009a, or appendix B).

4.4 Discussion

We derived in this chapter an effective free energy $F = L + TS$ in a generic framework model. The main property of this function is to intimately connect the individual and global points of view. Our simple model raises a number of interesting questions: in the limit $T \to 0$, the stationary configurations are those maximizing the potential L and not

[2]Caution: the expression of the entropy $S(x)$ depends on the precise definition of the "microscopic" states of the system. If the agents were to choose the *cell* (instead of the *block*) in which they would like to move, one would need to add the term $S'(x) = \sum_q \ln \frac{H!}{(H-n_q)!}$ in the expression of the entropy.

the collective utility U. Hence, they may differ from the simple collection of individual optima (Kirman, 1992), illustrating the unexpected links between micromotives and macrobehavior.

More specifically, we derived a simple expression of this effective free energy in the thermodynamic limit $N \to \infty$ for homogeneous agents. We showed that this simple form allows us to derive in an easy and flexible way quantitative solutions to economics models based on individual dynamics. This approach can be extended to some models with heterogeneous agents. Possible examples range from a two-population residential segregation model (Grauwin *et al.*, 2009a, or chapter 2), Ising-like model with heterogeneous pairs interactions (Nadal & Gordon, 2005), or the Hopfield model (Ui, 2000).

Part II

Maps of Science

Chapter 5

Complex Systems Science: Dreams of Universality, Reality of Interdisciplinarity

5.1 Introduction

Fundamental science has striven to reduce the diversity of the world to some stable building blocks such as atoms and genes. To be fruitful, this reductionist approach must be complemented by the reverse step of obtaining the properties of the whole (materials, organisms) by combining the microscopic entities, a notoriously difficult task (Hayden, 2010; Anderson, 1972; Grauwin *et al.*, 2009a; Gannon, 2007). The science of complex systems tackles this challenge, albeit from a different perspective. It adds the idea that "universal principles" could exist, which would allow for the prediction of the organization of the whole regardless of the nature of the microscopic entities. Ludwig Von Bertalanffy wrote already in 1968: "*It seems legitimate to ask for a theory, not of systems of a more or less special kind, but of universal principles applying to systems in general* " (Von Bertallanffy, 1976). This dream of universality is still active: "*[Complex networks science] suggests that nature has some universal organizational principles that might finally allow us to formulate a general theory of complex systems* " (Solé, 2000). Have such universal principles been discovered? Could they link disciplines such as sociology, biology, physics and computer science, which are very different in both methodology and objects of inquiry (SantaFe, 2010)?

Table 5.1: **The 20 references (including books and articles) and 20 journals most cited by the 141 098 articles of our database**

Reference	Times used	Journal	Times used
Albert R,2002,REV MOD PHYS(74)	2058	NATURE	115648
Watts DJ,1998,NATURE(393)	1735	SCIENCE	104838
Barabasi AL,1999,SCIENCE(286)	1697	J BIOL CHEM	103428
Newman MEJ,2003,SIAM REV(45)	1309	P NATL ACAD SCI USA	102289
Bak P,1987,PHYS REV LETT(147)	1172	PHYS REV LETT	68093
Strogatz SH,2001,NATURE(410)	,909	J AM CHEM SOC	58387
Sambrook J,1989,MOL CLONING LAB MANU(3)	855	CELL	57810
Bak P,1988,PHYS REV A(38)	751	J NEUROSCI	34283
Laemmli UK,1970,NATURE(227)	721	EMBO J	34264
Dorogovtsev SN,2002,ADV PHYS(51)	689	MOL CELL BIOL	33889
Albert R,2000,NATURE(406)	686	PHYS REV B	30212
Kohonen T,1982,BIOL CYBERN(44)	643	GENE DEV	27122
Pastorsatorras R,2001,PHYS REV LETT(87)	621	J CELL BIOL	27096
Goldberg DE,1989,GENETIC ALGORITHMS S	613	JCHEM PHYS	26974
Rumelhart DE,1986,PARALLEL DISTRIBUTED(1)	587	J IMMUNOL	24636
Newman MEJ,2001,PHYS REV E 2(64)	572	ANGEW CHEM INT EDIT	22161
Kohonen T,1995,SELF ORG MAPS	553	MACROMOLECULES	22151
Bradford MM,1976,ANAL BIOCHEM(72)	541	APPL PHYS LETT	19669
Haykin S,1999,NEURAL NETWORKS COMP	509	BIOCHEMISTRY-US	19597
Jeong H,2000,NATURE(407)	508	J NEUROPHYSIOL	18690

5.2 Results

In this chapter, we empirically study the "complex systems" field and its claims to universality. To collect a representative database of articles, we selected from the ISI Web of knowledge (WoS, 2011) all records containing topic keywords relevant for the field of complex systems (refer to Table 5.5 for the precise query. In this chapter, we limit ourselves to papers published after 2000). Table 5.1 contains the 20 most frequent cited references and journals within our dataset. To analyze the data, we build a network (Börner & Schernhost, 2009) in which the $\sim$ 141 000 articles are the nodes. These nodes are linked according to the proportion of shared references (bibliographic coupling Kessler, 1963). For this study, bibliographic coupling offers two advantages over the more usual co-citation link: it offers a faithful representation of the fields, giving equal weight to all published papers (whether cited or not) and it can be applied to recent papers (which have not yet been cited). For more details, the reader is referred to the section "Methods".

Figure 5.1 shows the largest communities (thereafter also called "fields" or "disciplines") obtained by modularity maximization of the network of papers published in the years 2000-2008. We first note that all important complex systems subfields[1] are

[1]In the following, we use italics to refer to the names of the communities.

present[2]. At the center, we find mostly theoretical domains: *self-organized criticality*, *dynamical systems*, *complex networks*, *neural networks*. These fields are connected to more experimental communities lying at the edges (*materials science*, *biology* or *neurosciences*). The links between theoretical and experimental fields suggest that complex systems science models have connections to the "real" world, as claimed by their practitioners.

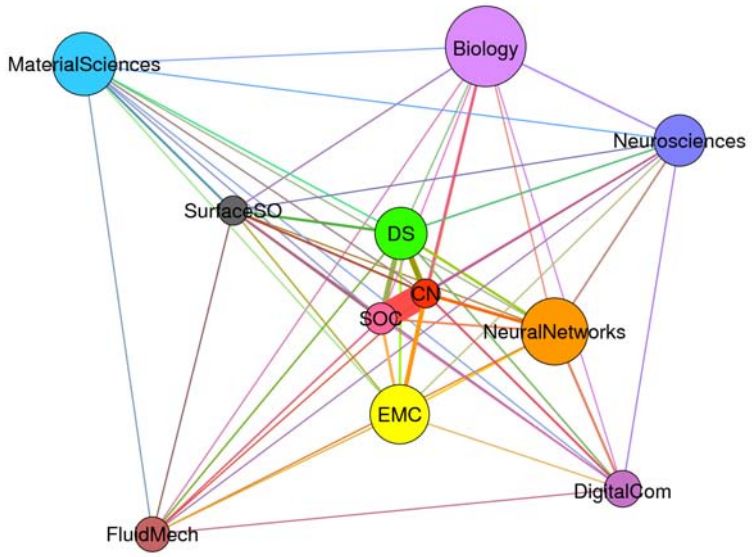

Figure 5.1: **Community structure obtained with a first run of the modularity maximization (Blondel *et al*, 2008) on the 2000-2008 network (141,098 articles).** The surface of a community I is proportional to its number of articles N_I and the width of the link between two communities I and J is proportional to the mean bibliographic coupling $\langle\omega\rangle_{IJ} = \sum_{i\in I, j\in J} \omega_{ij}/N_I\, N_J$. The layout of the graph is obtained thanks to a spring-based algorithm implemented in the Gephi visualization software (Bastian *et al* , 2009). For the sake of clarity, communities with less than 300 articles are not displayed. The label of a community represents the most frequent and/or significant keyword of its articles. *CN* stands for Complex Networks, *SOC* for Self Organized Criticality, *DS* for Dynamical Systems, *DigitCom* for Digital Communication and *SurfaceSO* for Self-organization on Surfaces. *EMC* is a more composite community where the three most representative keywords are *Ecology*, *Management* and *Computational Models*. See Figure 5.2 for details.

[2]As can be checked by consulting authoritative CS web sites such as Santa Fe's and its "Exploring complexity" lectures: `http://www.santafe.edu/news/item/exporing-complexity-science-and-technology-santa-fe-institute-perspective/` (accessed June 1st, 2010).

To understand the inner structure of these large communities, we use recursive modularity optimization (see Fortunato & Barthélemy (2007) and "Methods"). Most fields display a rich inner structure (Figure 5.2) with subcommunities (thereafter also called "subfields" or "subdisciplines") organized around several specific topics and references. The only exceptions are *self-organized criticality* and *complex networks*, where all articles cohere around a few references. Table 5.2 displays some quantitative informations about each subfield. For a more detailed presentation of the subfields, including their main authors, most used journals, references and keywords, see Appendix C. We analyze this complex structure at two levels. First, at the global scale, complex systems science appears to be a densely interconnected network. This is somewhat surprising since sharing references between subdisciplines means that they are able to read and understand these references, and moreover, that they find them useful. Would these shared references point to "universal" principles? Second, we focus on a more local scale, on the links that specifically connect two different disciplines (*ie* two different colors in Figure 5.1) to understand how they manage to exchange knowledge.

5.2.1 Complex systems' science overall coherence

Let us start with the field's overall coherence. We have looked for the references cited by many subfields and form the "glue" that links many subdisciplines and connects the network. More precisely, we define the networking force of a reference $\mathcal{N}(r)$ as the sum, over all pairs of subfields, of the proportion of their links explained by that reference (see Methods). Table 5.3 shows that the references that glue the network are more methodological than theoretical: the most networking reference is "Numerical Recipes" (Press *et al*, 2010), a series of books that gathers many routines for various numerical calculations and their implementation in computers. Most of the other linking references are mathematical handbooks or data analysis tools. If one looks for universality in the complex systems field, the computer – as a tool – seems to be a serious candidate. Among the leading contributors to the glue, we also find several references on self-organization (SO). Self-organization is not a predictive theory, but an approach that focuses on the spontaneous emergence of large-scale structures out of local interactions between the system's subunits (Mehdi *et al*, 2009). Several subdisciplines in Figure 5.2 can be related to this approach, as they use a keyword akin to "self-organization" (SO) in more than 10% of their articles (for a more complete list of the communities using this keyword, see appendix C). Among theses we find *swarm SO*, *molecular SO* linking chemistry to biology, *growth SO* and *pattern formation SO* linking surface science to dynamical systems. This suggests that the field of complex systems focuses on the

cases in which the link from microscopic to macroscopic can be analyzed through self-organization, which gave rise to several fruitful scientific programs, as we discuss below (section Discussion).

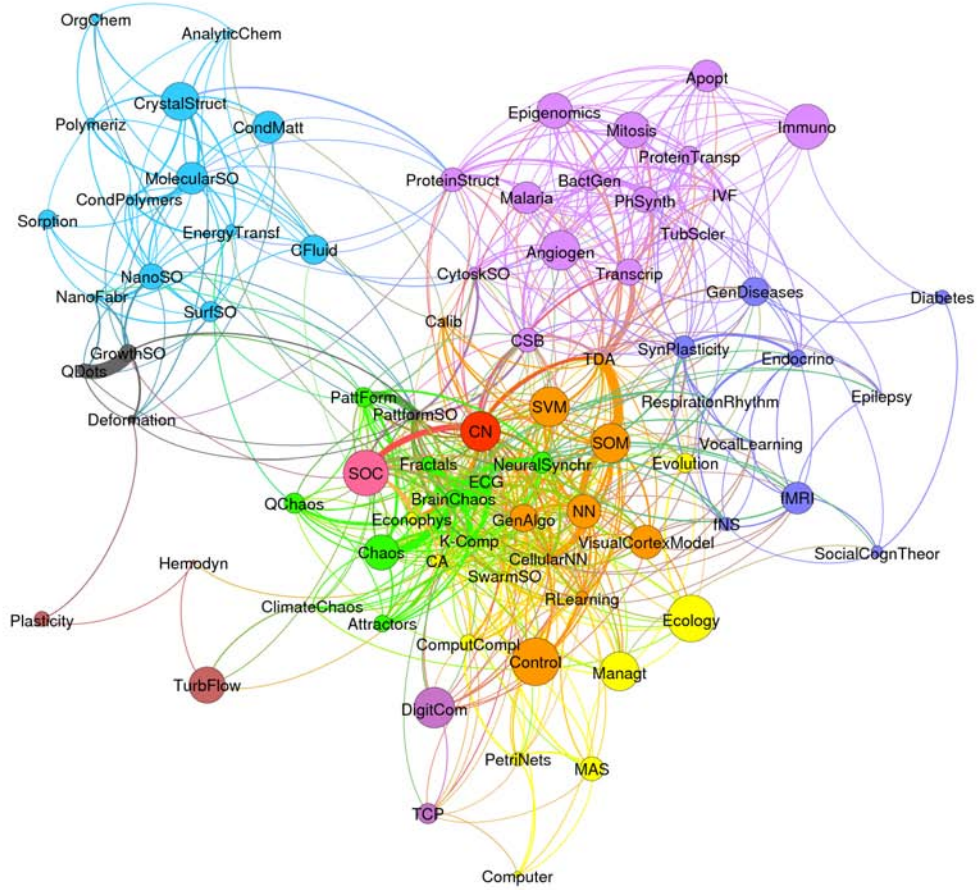

Figure 5.2: **Community structure obtained with a second run of the modularity maximization on the 2000-2008 network.** This community structure is obtained by optimizing the internal modularity Q^i of each community obtained by the first run of the modularity maximization algorithm on the 2000-2008 network, displayed on Figure 5.1 (See Methods for details on the procedure). The surface of each community is proportional to its number of articles and the width of the link between two communities I and J is proportional to the mean weight $\langle\omega\rangle_{IJ}$. For the sake of clarity, communities with less than 300 articles and links with a mean weight $\langle\omega\rangle_{IJ}$ less than 2.10^{-5} are not displayed. The color of a community (see online) corresponds to the color of the field (Fig 5.1) it belongs to. Community labels are based on a frequency analysis of the keywords. For a detailed presentation of the subfields, including their authors, most used journals, references and keywords, see Appendix B.

Table 5.2: **2000-2008 subfields' sizes N and inner coherences $< \omega_{in} >^{-1}$** The acronyms and abbreviations in parenthesis correspond to the label of the subfields displayed on Fig 5.2. The opposite of the average of the inner links's weight of a subfield $< \omega_{in} >^{-1}$ is (the opposite of) an inner coherence measure that support a straightforward interpretation. Indeed, if the weights of the inner links of a subfield were homogeneously distributed between all pairs of articles, two articles randomly chosen of this subfield would share 1 reference over $< \omega >^{-1}$.

Subfield	N	$< \omega >^{-1}$
Analytic Chemistry (AnalyticChem)	419	336.14
Angiogenesis (Angiogen)	3642	1408.78
Apoptosis (Apopt)	2632	1424.23
Attractors	1161	524.42
Bacterial Genomics (BactGen)	1517	120.19
Brain Chaos	1175	70.06
Calibration (Calib)	538	371.17
Cellular Automata(CA)	846	164.26 .0023
Cellular Neural Networks (CellularNN)	620	86.17
Chaos	3134	531.3
ClimateChaos	352	205.31
Complex Fluids (CFluid)	2310	994.14
Complex Networks (CN)	3684	21.87
Computationnal Complexity (ComputCompl)	1134	379.92
Computationnal Systems Biology (CSB)	1799	323.99
Computer	526	437.08
Condensed Matter (CondMatt)	2629	631.99
Condensed Matter - Polymers (CondPolymers)	471	131.85
Control	4772	1086.35
Crystal Structure (CrystalStruct)	3386	350.41
Cytoskeleton Self-Organization(CytoskSO)	651	132.58
Deformation	590	432.48
Diabetes	1015	669.24
Digital Communication (DigitCom)	3811	470.56
Ecology	4751	1846.16
Econophysics (Econophys)	738	96.12
Electrocardiogram (ECG)	987	117.68
Endocrinology (Endocrino)	1223	607.63
Energy Transfert (EnergyTransf)	801	258.27
Epigenomics	3055	677.49
Epilepsy	316	273.67
Evolution	1318	796.76
Fractals	1015	192.02
Functionnal MRI (fMRI)	2634	897.28
Functionnal Neurosciences (fNS)	935	497.44
Genetic Algorithm (GenAlgo)	2177	197.96
Genetic Diseases (GenDiseases)	2273	387.34
Growth Self-Organization (GrowthSO)	1192	113.84
Hemodynamics (Hemodyn)	346	344.37
Immunology (Immuno)	4403	2234.66
In Vitro Fertilization (IVF)	591	281.01
Kolmogorov Complexity (K-Comp)	501	121.94
Malaria	2702	747.44
Management (Managt)	3563	2159.31
Mitosis	3171	564.98
Molecular Self-Organization (MolecularSO)	2684	409.08
Multi-agent System (MAS)	1787	1094.91
Nanofabrication (NanoFabr)	457	45.28
Nanosciences (Nano)	1995	418.01
Neural Networks (NN)	2902	221.15
Neural Synchronization (NeuralSynchr)	1451	453.59
Organic Chemistry (OrgChem)	649	368.67
Pattern Formation (PattForm)	1403	205.82
Pattern Formation & Self-Organization (PattformSO)	691	142.82
Petri Nets	957	275.76
Photosynthesis (PhSynth)	2000	224.48
Plasticity	1066	915.05
Polimerization (Polymeriz)	645	98.54
Protein Structure (ProteinStruct)	1830	237.07
Protein Transport (ProteinTransp)	1305	308.77
Quantum Chaos (QChaos)	1456	636.22
Quantum Dots (QDots)	921	130.93
Reinforcement Learning (RLearning)	891	287.57
Respiration Rhythm	416	57.35
Self-Organized Criticality (SOC)	4447	199.3
Self-Organizing Maps(SOM)	3495	168.85
Social Cognition Therory (SocialCognTheor)	800	680.59
Sorption	1354	925.37
Support Vector Machines(SVM)	3660	867.91
Surface Self-Organization (SurfSO)	1511	468.34
Swarm Intelligence (SwarmIntel)	608	145.94
Synaptic Plasticity (SynPlasticity)	1625	370.25
Transcriptomics (Transcrip)	2043	439.37
Transcriptomics Data Analysis (TDA)	628	43.32
Transmission Control Protocol (TCP)	1473	718.93
Tuberous Sclerosis (TubScler)	766	153.33
Turbulent Flow (TurbFlow)	3172	1212.32
Visual Cortex Model	2851	845.22
VocalLearning	389	162.56

5.2.2 Interdisciplinary trading zones

At a more local scale, let us now look at the links that specifically connect two distinct disciplines. How are those connections established? It is widely accepted that scientific disciplines cannot easily communicate or be linked (in our case, share references) simply because it is difficult for a physicist to understand a biology paper and vice-versa. In addition, different disciplines have different definitions of what counts as a result or as an interesting research topic. For example, physical sciences look for universal laws, while social (Borgatti *et al*, 2009) and biological (Fox-Keller, 2005) sciences emphasize the variations in structure across different groups or contexts and use these differences to explain differences in outcomes. Physicians are interested in practical medical advances while physicists want to know whether physiological rhythms are chaotic or not (Glass, 2001).

Where do the links come from then? In an illuminating analogy, Peter Galison (1997) compares the difficulty of connecting scientific disciplines to the difficulty of communicating between different languages. History of language has shown that when two cultures are strongly motivated to communicate - generally for commercial reasons - they develop simplified languages that allow for simple forms of interaction. At first, a "foreigner talk" develops, which becomes a "pidgin" when social uses consolidate this language. In rare cases, the "trading zone" stabilizes and the expanded pidgin becomes a creole, initiating the development of an original, autonomous culture. Analogously, biologists may create a simplified and partial version of their discipline for interested physicists, which may develop to a full-blown new discipline such as biophysics. Specifically, Galison (1997) has studied how Monte Carlo simulations developed in the postwar period as a trading language between theorists, experimentalists, instrument makers, chemists and mechanical engineers. Our interest in the concept of a trading zone is to allow us to explore the dynamics of the interdisciplinary interaction instead of ending analysis by reference to a "symbiosis" or "collaboration".

Table 5.4 gives a list of the main "trading zones" which connect theoretical and experimental fields in Figure 5.1 and capture a significant fraction of the links between these fields. The clearest example is *transcriptomics data analysis*, a subfield of *neural networks* which connects biologists interested in the interpretation of data retrieved from DNA chips and computer scientists interested in data analysis via methods from the *neural networks* field. The *transcriptomics data analysis* subfield represents 2.3% of *neural networks* papers but accounts for 46.3% of the connections between *neural networks* and *biology* and 16.5% of the links between *neural networks* and *complex networks*. Other trading zones are *computational systems biology*, linking *biology* to many theoret-

Table 5.3: **The 20 most networking references in the 2000-2008 decade.** The references followed by a star correspond to books or papers which appeared in the database under several forms - essentially different publication years for the books - for which the networking power $\mathcal{N}(r)$ have been summed. The complete references of these papers are given in Appendix B.

Reference	Topic	$\mathcal{N}(r)$ (%)
Press et al. (1992)*	Numerical recipes (book)	1.250
Shannon (1948)*	Information theory	0.607
Metropolis et al. (1953)	Monte Carlo integration	0.509
Nicolis et al. (1977)*	Self organization (book)	0.420
Kauffman (1993)*	Self organization (book)	0.309
Hebb (1949)	Neuropsychology and behavior (book)	0.297
Alberts et al. (1994)	Molecular and cellular biology (book)	0.288
Abramowitz et al. (1968)*	Handbook of mathematical functions	0.269
Feller (1958)*	Introduction to probability theory (book)	0.268
Watson & Crick (1953)	Structure of DNA	0.250
Lakowicz (1999)	Fluorescence spectroscopy	0.249
Turing (1952)	Morphogenesis	0.237
Witten et al. (1981)	Diffusion-limited aggregation	0.234
Cohen (1988)	Statistics and behavioral sciences (book)	0.223
Hopfield (1982)	Neural networks	0.217
Stanley (1971)	Phase transition (book)	0.202
Whitesides et al. (2002)	Self-assembly	0.188
Marquardt (1963)	Applied mathematics	0.174
Chomczynski (1987)	RNA isolation	0.167
Venter et al. (2001)	Human genome sequence	0.160

ical fields, among which *dynamical systems*, *self-organized criticality* and *complex networks*, *neural synchronization* linking *dynamical systems* and *neurosciences*, *cytoskeleton self-organization* linking *biology* to *dynamical systems* and *self-organized criticality* and *calibration* linking *neural networks* and *material sciences*. Note that a single trading zone can be used by a fields to exchange with several other fields, as long as these other fields share the same "language". For example, *computational systems biology*, links *biology* to *dynamical systems*, *self-organized criticality* and *complex networks*, three subfields which share the physicists' toolkit. Since our map cannot cover all scientific fields, we may not recognize some subfields as trading zones, such as *electrocardiogram* which is likely to connect *dynamical systems* to medecine, or even miss a trading zone between geosciences and *self-organized criticality*.

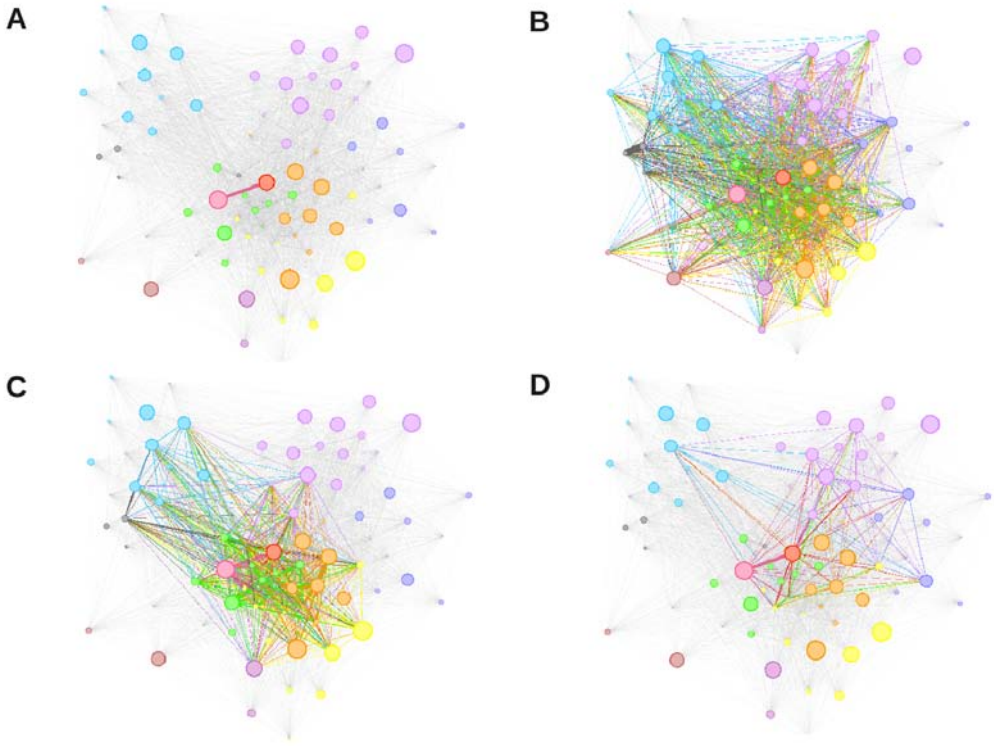

Figure 5.3: **Local "networking" force for four different references on the 2000-2008 network (Fig 5.2).** Links established using the reference are shown in color. The number of citations corresponds to those included in papers of our database published between 2000 and 2008. References used: **a.** Albert & Barabasi (2002) (2058 citations) **b.** Press *et al* (2010), Numerical Recipes - all editions (1267 citations) **c.** Nicolis & Prigogine (1977) (342 citations) **d.** Barábasi & Oltvai (2004) (244 citations).

By analyzing carefully the references used by trading zones and also the references that make the links between the trading zones and their neighbors, we can distinguish two types of trading zones, applicative and speculative. Let us start with *transcriptomics data analysis*, which is a clear example of "applicative" trading zone. The development of new measurement techniques in cellular biology (mainly DNA microarrays) produced huge amounts of data together with the need of new tools to analyze them. Since this new technique promised a better understanding of cell dynamics, a new scientific subdiscipline, able to understand data analysis and its biological interest was built around transcriptomic tools. The two references most used by this subfield stress the applicative side: the purpose of the first paper is "*to describe a system of cluster analysis*

Table 5.4: **Strongest trading zones.** The trading force T of a subcommunity measures the fraction of the links between two fields (in Fig. 5.1) which goes through this subcommunity. More precisely, the trading force of I, a subcommunity of $\bar{I}$, towards any community $\bar{J}$ is the total weight of the article-article links between the subcommunity I and community $\bar{J}$, normalized by the total weight of the the article-article links between $\bar{I}$ and $\bar{J}$: $T_{\bar{I}\bar{J}}(I) = \sum_{i\in I, j\in\bar{J}} \omega_{ij} / \sum_{i\in\bar{I}, j\in\bar{J}} \omega_{ij}$. The expected force T^{exp} is the value of the trading force one would expect if all the links between $\bar{I}$ and outside communities were equally shared among all sub-communities of $\bar{I}$, which is simply the fraction $N_I/N_{\bar{I}}$ of articles of I in $\bar{I}$. The acronyms of the subfiels used here correspond to those explained in Table 5.2.

Subfield	Fields	T (%)	T/T^{exp}
TDA	Biology/Neural Networks	46.355	20.51
CSB	Biology/Dynamical Systems	49.704	8.87
CSB	Biology/SOC	49.255	8.79
ProteinStruct	Biology/Material Sciences	47.442	8.32
CSB	Biology/CN	42.931	7.66
TDA	CN/Neural Networks	16.543	7.32
Hemodyn	Neurosciences/FluidMech	54.552	7.22
NeuralSynchr	Neurosciences/Dynamical Systems	59.788	5.20
Hemodyn	Biology/FluidMech	39.113	5.18
CytoskSO	Biology/SOC	9.561	4.71
CytoskSO	Biology/Dynamical Systems	9.522	4.69
Calib	Material Sciences/Neural Networks	8.717	4.51
CellularNN	CN/Neural Networks	9.605	4.30
Transcrip	Biology/CN	25.806	4.05

for genome-wide expression data from DNA microarray hybridization [...] in a form intuitive for biologists" (Eisen *et al*, 1998) while the second "*describes the application of self-organizing maps for recognizing and classifying features in complex, multidimensional [transcriptomic] data*" (Tamayo *et al*, 1999). The *transcriptomics data analysis* papers are clustered together because they share references presenting this kind of applications. The applicative character of *transcriptomics data analysis* can also be seen in the origin of the references that link them to neighbor subfields (Figure 5.4). The common references between *transcriptomics data analysis* and *biology* (mainly *transcriptomics*) are similar to references used by *transcriptomics data analysis* papers themselves. This means that the link arises from biologists citing results obtained by *transcriptomics data analysis* scientists or techniques they use. On the other hand, the common references between *transcriptomics data analysis* and *self-organizing maps* (a subfield of *neural net-*

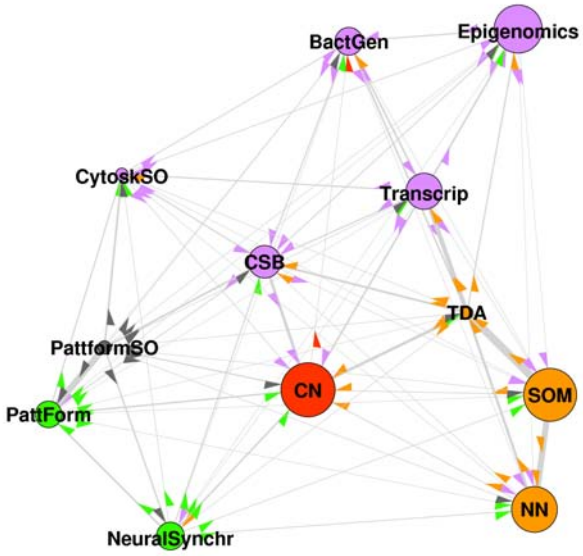

Figure 5.4: **Directed network.** On this subset of the graph presented in Figure 5.2, the arrows are directed to the subfield that uses the other subfield's references to establish the link. More precisely, the common references shared by two linked subfields are more similar to the internal references of the subfield from which the arrow originates than to the internal references of the subfield to which the arrow points (see "Methods" for more details). The figure shows that *transcriptional data analysis* (TDA) feeds from *self-organizing maps* (SOM) and *neural networks* (NN) methodological references, while biology subcommunities (mainly Transcriptomics) use *transcriptional data analysis* references. The orientation of the links is quite different for *computational systems biology* (CSB) and *complex networks* (CN), because these subfields tend to pump their neighbors' references, while the other subfields do not find much use in *computational systems biology* and *complex networks* references.

works) are similar to references used by *self-organizing maps* papers. This means that the link arises from *transcriptomics data analysis* scientists citing classification techniques created by *self-organizing maps* scientists, while these scientists do not often use *transcriptomics data analysis* references. Therefore, *transcriptomics data analysis* allows *self-organizing maps* techniques to be understood and used to interpret biological data, with a relevance certified by biologists' citations. The case of another trading zone, *computational systems biology*, is different. Its most used references point to computational methods - mainly Gillespie's algorithm (Gillespie, 1977) or to experimental papers in which there is no explicit modelling but that show complex cellular dynamics, thus justifying indirectly the need for modelling. The link between experiments and modelling

is still speculative, as summarized by one of the most used references in this subfield (Tyson *et al*, 2010): "*we hope that this review will [...] promote closer collaboration between experimental and computational biologists.*" Moreover, the common references between *computational systems biology* and *biology* are from *biology*, as if *computational systems biology* scientists were eager to quote potentially interesting biological applications for their modelling approach, while many biologists were still unaware of these models. In short, compared to *transcriptomics data analysis*, *computational systems biology* seems a more speculative trading zone, at the frontier of biology and modelling, but presently lacking a specific object or concept to define an operational trading zone.

5.3 Complex Systems's relation to the overall scientific literature

To explore the question of the relation of CSS to the overall science, we present two preliminary maps displaying the communities obtained from a database containing records of the overall French scientific literature of 2000 and 2010 (as obtained from the Web of Science). We try to point out the place of CSS in these overall maps. We have gathered the records of the 63840 articles published in 2000 and the 78058 articles published in 2010 containing a french address from the ISI Web of Knowledge database (WoS, 2011, query performed in March 2011). We present in this section the two preliminary maps of the state of French research in 2000 and 2010, obtained once again by computing the bibliographic weight between articles and detecting the communities thanks to the Louvain algorithm.

These two maps show an overall division of the science literature in less than a dozen main fields (here represented by colors), with a clear separation between biology, medicine and neuroscience (on the left hand side) and chemistry, mathematics, physics and computer science (on the right hand side).

While these maps are very interesting in many ways, it should be noted that they do not contain any "complex systems" subfield, although we can find a *Self-Organized Criticality* (SOC) community on the 2000 map (middle right hand side) and a *Complex Networks* (CN) community on the 2010 map (middle right hand side). These maps suggest that there exists (yet?) no conceptual kernel for "complexity science", some unified theory that could give the field enough coherence to be a subfield identifiable by shared references (and therefore by bibliographic coupling). *Self-Organization*, *Self-Organized Criticality* or *Complex Networks* do constitute such coherent subfields. However, despite some claims, these subfields seem to have only weak links to the rest of the complex

systems "galaxy" shown above, which prevents the formation of a coherent "complex systems science".

5.4 Discussion

Our empirical study of the "complex systems" field shows that its overall coherence does not arise from a universal theory but from computational techniques and successful adaptations of the idea of self-organization. The computer is important for advancing the understanding of complex systems because it allows scientists to play with simple but nonlinear models and to handle large sets of data obtained from complex systems. The essential role of the computer is confirmed by historical studies showing (Fox-Keller, 2009) that the complex systems field is heir of the postwar sciences born around the computer: operational research, game theory and cybernetics. These fields started when physicists, mathematicians and engineers started collaborating to maximize the efficiency of WW II military operations (Pickering, 1995; Bowker, 1993).

Let's now examine the various claims to universality. A "general systems theory" would possess a collection of theoretical books or papers revealing the "universal" explanation. This would be evidenced in Figure 5.2 by a central group to which other groups would connect. Instead, our analysis shows a variety of modelling disciplines in a central position. Certainly, a few theoretical papers, such as Bak's (Bak *et al*, 1987) (in *self-organized criticality*) or Albert and Barabasi's (Albert & Barabasi, 2002) (in *complex networks*) point to "universal" mechanisms and are heavily cited. However, more than 80% of their citations arise from modelers themselves [3], suggesting that they may be universal... for theorists. Our data support the local character of these "universal" laws. First, Albert and Barabasi's (Albert & Barabasi, 2002) paper is the most cited in the 2000-2008 decade but only links *complex networks* and *self-organized criticality* subfields (Figure 5.3a). The contrast with the global networking achieved by methodological references such as Metropolis' algorithm or Numerical Recipes (Figure 5.3b) or self-organization references (Figure 5.3c) is clear. Second, the references that *complex networks* (Figure 5.4) and *self-organized criticality* communities share with experimental

[3] Counted on Web of Science (January 28th 2011) by analyzing the citing papers by discipline. Specifically, "Subject Areas" (Web of Science name for subdisciplines) related to Physics (such as Physics Multidisciplinary or Physics Mathematical) account for 2272 out of 3158 citations (72%) for Bak's paper (Bak *et al*, 1987) while more "applied" subject areas (such as Geosciences and Applied mathematics) account for 676 citations (21%). Subject areas related to Physics or Mathematics account for 4201 out of 5281 citations (80%) of Albert and Barabasi's paper (Albert & Barabasi, 2002), while subject areas related to biology or more applied fields account for 1060 citations (20%). Instead, areas related to biology account for most of citations to Barabási and Oltvai's introduction of network theory to biologists (Barábasi & Oltvai, 2004).

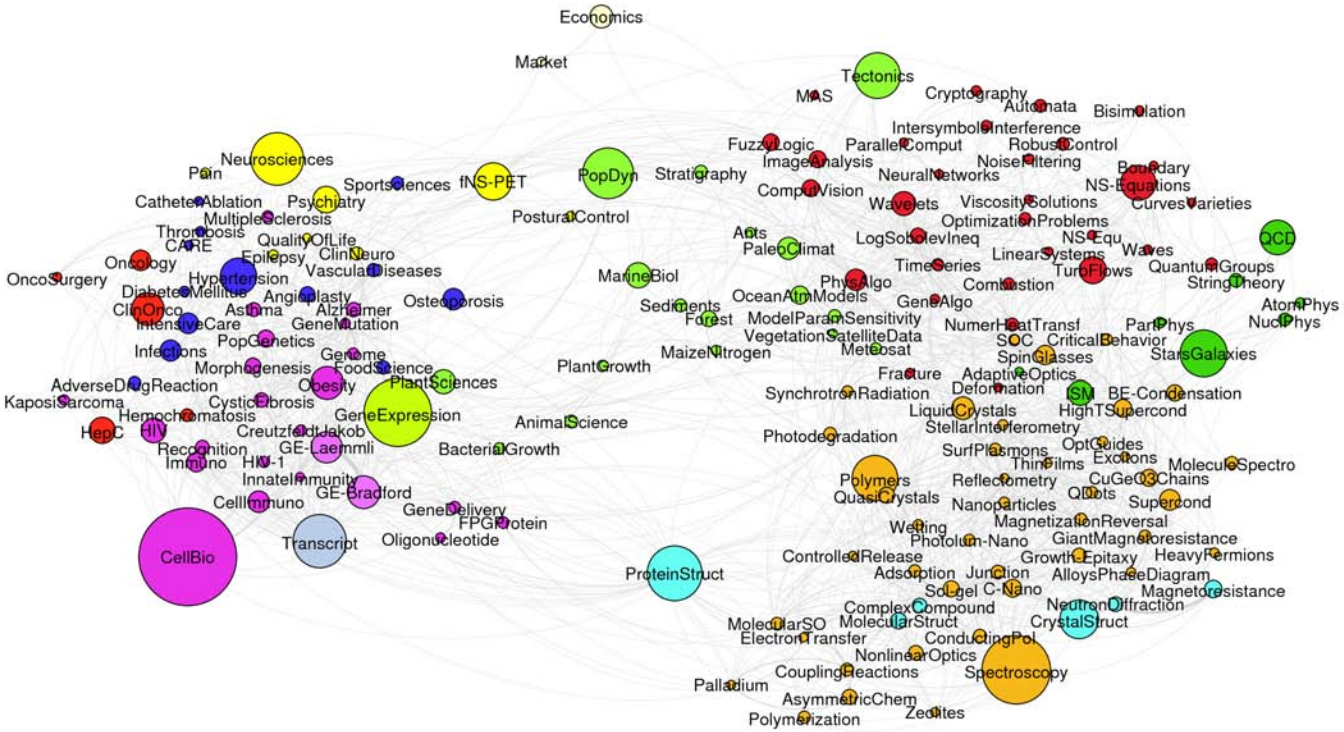

Figure 5.5: **2000's french research subfield community networks.** The surface of a community I is proportional to its number of articles N_I and the width of the link between two communities I and J is proportional to the mean bibliographic coupling $\omega_{IJ} = \sum_{i \in I,\, j \in J} \omega_{ij} / N_I\, N_J$. For the sake of clarity, communities with less than 40 articles and links with a mean weight $\omega < 2.10^{-5}$ are not displayed. Labels are based on a frequency analysis.

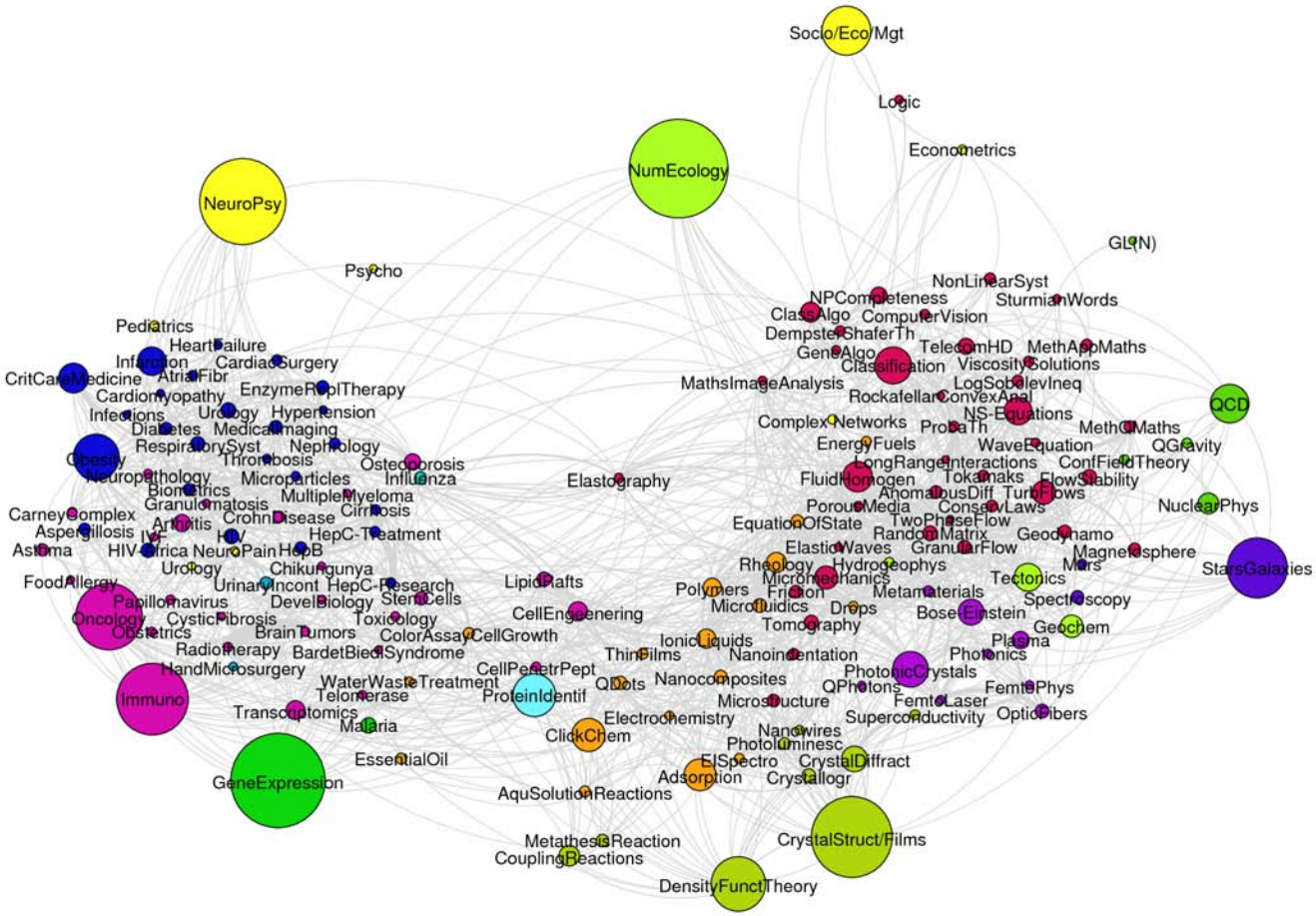

Figure 5.6: **2010's french research subfield community networks.** The surface of a community I is proportional to its number of articles N_I and the width of the link between two communities I and J is proportional to the mean bibliographic coupling $\omega_{IJ} = \sum_{i \in I,\, j \in J} \omega_{ij} / N_I\, N_J$. For the sake of clarity, communities with less than 40 articles and links with a mean weight $\omega < 2.10^{-5}$ are not displayed. Labels are based on a frequency analysis.

fields are similar to those of the experimental fields. This means that the links between these modelling practices and their potential applications are mostly rhetorical: *complex networks* and *self-organized criticality* papers often quote experimental work as legitimating their models, while experimentalists rarely refer to them. To try to become universal, theoretical approaches have to be "translated" into other disciplines. An example of this strategy is shown in Figure 5.3d which shows the links established between network science and biology thanks to Barabasi and Oltvai's introduction of networks for biologists (Barábasi & Oltvai, 2004).

It could be argued that links between these theories and experimental fields take time to establish and will be seen in the future. An interesting insight of the possible evolution of universality claims is given by the history of self-organization, which was considered by many as a universal key to Nature in the 1980's (Fox-Keller, 2009). This idea was fecund in that it gave birth to several active subdisciplines (*cytoskeleton SO*, *growth SO*...) (Figure 5.2). However, it should be noticed that these heirs of self-organization are nowadays almost unrelated. The different *self-organization* subfields are more linked to their own discipline (*biology*, *materials science* ...) than between them. This is shown by the plain fact that community detection puts these SO subfields into different disciplines (different colors in Figure 5.2) instead of creating a single, unified, *self-organization* field. The reason is that these subfields use widely different references, as illustrated by the fact that there is no common reference among the 10 most used references for all the different *self-organization* subcommunities. Self-organization is therefore not a universal explanation but rather a kind of banner, which needs to be associated to references to specific elements (including techniques, microscopic entities and their interactions) to be fruitful.

In summary, we have obtained a *global* point of view on the structure of the "complex systems" field. This has allowed us to test empirically the idea of universality, showing that it remains a dream, albeit one which has lead to interesting but more modest realities. At the global scale, the whole domain is linked by the focus on self-organization and the use of computer-based methods for solving non-linear models. At a more local scale, the links between different disciplines are achieved through the development of "trading zones" (Galison, 1997). These allow for coordination between vastly different scientific cultures, for example theoretical and experimental disciplines, which are only marginally connected. These disciplines may differ on the very conception of what is an interesting topic, but can work together around specific tools (a DNA microchip) or concepts (a network). Today, these interdisciplinary collaborations are a key to

essential scientific challenges such as the analysis of the massive amount of data recently made available on biological and social systems (Lazer, 2009; Microsoft, 2006) and the understanding of the complex intertwining of different levels of organization that is characteristic of these systems.

5.5 Methods

5.5.1 Extraction of the data

Our data have been extracted from the ISI Web of Knowledge database (WoS, 2011) in December 2008. The science of complex systems is particularly challenging as an epistemic object since there exists no consensual definition of the domain, nor any list of disciplines or journals that would gather all the relevant papers. Therefore, we selected all the articles of the database whose title, abstract (for articles published after 1990) or keywords contained at least one of a chosen list of topic keywords (Table 5.5). These keywords were derived from discussions with experts of the field, mainly scientists working at the complex systems institute in Lyon (IXXI). We have retrieved 215 305 articles (141 098 between 2000 and 2008) containing 4 050 318 distinct references. Each record contains: authors, journal, year of publication, title, keywords (given by the authors and/or ISI Web of Science) and the list of references of the article. Any choice of keywords being potentially biased and partial, our strategy was to risk choosing too many of them - thus bypassing the lack of precise definition of the "complex systems" field and retrieving all its important subfields - and to trust the subsequent analysis to eliminate irrelevant articles.

In fact, as shown in table 5.5, around 40% of the articles of the database comes solely from the combination of keywords *"complex*" and "control"*. While most of those articles were close to biology and not directly related to the field of complex systems, we chose to keep them in order to test the robustness of our analysis. As shown below, our strategy was successful, since most of these "irrelevant" articles are grouped into a few communities (such as *Apoptosis* or *Immunology*) that lie at the network's edges and do not bias the results.

5.5.2 Links between articles

Weight of links between articles are calculated through their common references (bibliographic coupling Kessler, 1963). We define a similarity between two articles i and j as the cosine distance:

$$\omega_{ij} = \frac{|\mathcal{R}_i \cap \mathcal{R}_j|}{\sqrt{|\mathcal{R}_i|\,|\mathcal{R}_j|}} \tag{5.1}$$

where $\mathcal{R}_i$ is the set of references of article i. By definition, $\omega_{ij} \in [0, 1]$, is equal to zero when i and j do not share any reference and is equal to 1 when their sets of references are identical. For this study, bibliographic coupling offers two advantages over the more usual co-citation link: it offers a faithful representation of the fields, giving equal weight to all published papers (whether cited or not) and it can be applied to recent papers (which have not yet been cited). Moreover, the links are established on the basis of the author's own decisions (to include or not a given reference) rather than retrospectively from other scientists' citations. Thus, bibliographic coupling can be used to analyze the community of research as it builds itself rather than as it is perceived by later scientists that cite its publications.

5.5.3 Community detection and characterization

In order to structure this network into groups of cohesive articles, we partition the set of papers by maximizing the modularity function. Given a partition of the nodes of the network, the modularity is the number of edges inside clusters (as opposed to crossing between clusters), minus the expected number of such edges if the network was randomly conditioned on the degree of each node. Community structures often maximize the modularity measure. We compute our partition using the algorithm presented in Blondel *et al* (2008), which is designed to efficiently maximize the modularity function in large networks. More precisely, we used the weighted modularity Q Girvan & Newman (2004); Fortunato (2010), which is defined as $Q = \sum_I q_I$, where the *module* q_I of a community I is given by

$$q_I = \frac{\Omega_{II}}{\Omega} - \left(\frac{\sum_{J \neq I} \Omega_{IJ} + 2\Omega_{II}}{2\Omega} \right)^2 \tag{5.2}$$

where

$$\Omega_{II} = \frac{1}{2} \sum_{i \in I,\, j \in I} \omega_{ij} \qquad \text{is the total weight of the links inside community } I,$$

$$\Omega_{IJ} = \sum_{i \in I,\, j \in J} \omega_{ij} \qquad \text{is the total weight of the links between communities } I \text{ and } J \neq I$$

$$\Omega = \frac{1}{2} \sum_{i,j} \omega_{ij} = \sum_{(i,j)} \omega_{ij} \qquad \text{is the total weight of the links of the graph.}$$

Each module q_I compares the relative weight of edges $\frac{\Omega_{II}}{\Omega}$ inside a community I with the expected weight of edges $\left(\frac{\sum_{J\neq I}\Omega_{IJ}+2\Omega_{II}}{2\Omega}\right)^2$ that one would find in community I if the network were a random network with the same number of nodes and where each node keeps its degree, but edges are otherwise randomly attached. See Ref Fortunato & Barthélemy (2007) for a more explicit interpretation of the modularity, its properties and limits.

Applying the Louvain algorithm yields a first partition of the network into communities (also referred to as "fields" or "disciplines", see Figure 5.1). To obtain the substructure of these communities, we apply the Louvain algorithm a second time on each of them. We find that most of these communities display a clear substructure with high values of internal modularity Q^i (typically between 0.4 and 0.8). Only two of them (*self-organized criticality* and *complex networks*) are strongly bound around a few references and present much lower values of Q^i (typically less than or around 0.2). Consequently they were not split into subfields which would not have much scientific relevance.

This recursive modularity optimization leads us to a "subfield" graph (Figure 5.2). We have checked that all the obtained sub-communities satisfy the criterion ($q_I \geq 0$) proposed by Fortunato & Barthélemy (2007) to check their relevance (see Table 5.2).

5.5.4 Links between communities and their orientation

The link between two communities I and J can be quantified by the average distance between an article $i \in I$ and an article $j \in J$:

$$< w >_{IJ}^{-1} = < w_{ij} >_{i\in I, j\in J}^{-1} = (\Omega_{IJ}/N_I\, N_J)^{-1} \tag{5.3}$$

A link between a community I and a community J exists if at least one reference is shared between an article of I and an article of J. To analyze the scientific content conveyed by the link, it is important to know if the shared references are more similar to the references used by community I or to the references used by community J. To take into account this similarity, we define the *orientation* of a community-community link in the following way.

Let $n_{r,I}$ be the number of papers of community I using reference r. Then,

- the number of article-article links inside community I which use reference r is $L_{r,II} = n_{r,I}\,(n_{r,I}-1)/2$
- the number of article-article links between communities I and J which use reference

r is $L_{r,IJ} = n_{r,I} n_{r,J}$

We compare the set of references shared by the two communities I and J to the references used by I and J by computing the cosine similarity measures:

$$\cos_{II,IJ} = \frac{\sum_r L_{r,II} L_{r,IJ}}{\sqrt{\sum_r L_{r,II}^2 \sum_r L_{r,IJ}^2}} \qquad \text{comparing the shared refs to those of } I$$

$$\cos_{JJ,IJ} = \frac{\sum_r L_{r,JJ} L_{r,IJ}}{\sqrt{\sum_r L_{r,JJ}^2 \sum_r L_{r,IJ}^2}} \qquad \text{comparing the shared refs to those of } J$$

For example, if $\cos_{II,IJ} < \cos_{JJ,IJ}$, the shared references are more similar to the references binding community J than to the references binding community I. We then direct the link from community J to community I, as community I "pumps" community J references to establish the link. See Figure 5.4 for examples of link orientation.

5.5.5 Networking power of references

To understand which references link the different subdisciplines to form a connected network, we define the "glue" as the set of references shared between subfields. To give equal weight to all these links, we normalize each link to 1, leading to the normalized networking strength $\mathcal{N}(r)$ of reference r as:

$$\mathcal{N}(r) = \frac{1}{Z} \sum_{I \neq J} f_{IJ}(r) \tag{5.4}$$

where $f_{IJ}(r)$ is the fraction of links between an article of community I and an article of community J in which reference r is used and where Z is a normalization constant such that $\sum_r \mathcal{N}(r) = 1$. The normalization ensures that $\mathcal{N}(r)$ represents the proportion of all the links of the complex systems field that can be assigned to reference r.

Table 5.5: **Topic keywords used in our request in the ISI Web of Knowledge database and number of articles matching independently to each of these topic keywords.** Each topic keywords except the first six where coupled with the topic keywords "complex*". We moreover rejected the articles containing the topic keywords "complex scaling" or "linear search", two terms refering respectively to (heavily used) specific methods of quantum chemistry and computer science.

topic keywords	Results
"self organ*"	32484
"complex network*"	6953
"dynamical system"	8205
"econophysics"	633
"strange attractor"	769
"synergetics"	379
"adaptive system*"	1141
"artificial intelligence"	1812
"attractor"	1034
"bifurcation"	3164
"chaos"	5370
"control"	116017
"criticality"	980
"ecology"	5869
"economics"	2243
"epistemology"	345
"far from equilibrium"	253
"feedback"	12881
"fractal"	3867
"ising"	975
"multi agent"	2032
"multiagent"	665
"multi scale"	779

topic keywords	Results
"multifractal"	390
"multiscale"	1439
"neural network*"	12747
("non linear*" OR "nonlinear*") NOT "equation*"	10240
"non linear dynamic*"	560
"non linear system*"	391
"nonlinear dynamic*"	2285
"nonlinear system*"	1826
"phase transition"	5503
"plasticity"	6667
"random walk"	758
"robustness"	6498
"scaling"	7008
"social system*"	586
"spin glass*"	643
"stability" AND ("lyapunov" OR "non linear*" OR "nonlinear*")	1399
"stochastic"	9184
"synchronization"	4645
"turbulence"	4602
"universality"	861
"cell* automat*"	1659

Chapter 6

Mapping Science Institutions: the case of *ENS de Lyon*

6.1 Introduction

Scientometrics has proved a valuable tool to understand the organization of scientific fields (Small, 1999) and their evolution (Chavalarias & Cointet, 2009; Mogoutov & Kahane, 2007). Global science maps (Small, 1999; Klavans & Boyack, 2009; Small, 1973; Börner & Schernhost, 2009; Leydesdorff & Rafols, 2009; Rafols & Leydesdorff, 2010; Agarwal & Skupin, 2008) have become feasible recently, offering a tentative overall view of scientific fields and fostering dreams of a "science of science" (Börner & Schernhost, 2009). In this article, we propose a more modest but less explored mapping, that of single scientific institutions. The scope is to achieve a global point of view on the institutions that no individual can have, in order to understand their organization, their strong and weak points, the papers or authors that link different Departments or disciplines... Such maps may become important as policy tools as few directors have such a global view of their institution.

Recently, Rafols and Leydesdorff have suggested a simple way to picture the disciplinary weight of an institution (Rafols & Leydesdorff, 2010). This method is rapid and can be carried out online. As it uses Web of Science (WoS, 2011) "subject categories" as relevant subdisciplines to project the data, it has the advantage of enabling a comparison across different institutions or years. The drawback of this rigid projection skeleton is that it preselects, without local information, the relevant communities. As acknowledged by Rafols and Leydesdorf (Rafols & Leydesdorff, 2010): "The two characteristics that make overlay maps so useful for comparison, their fixed positional and

cognitive categories, are also inevitably, their major limitations and a possible source of misreadings. Since the position in the map is only given by the attribution in the disciplinary classification, it does not say anything about the direct linkages between the nodes."

Here we propose different ways of mapping scientific institutions based on the articles published with that address (and not the journals as in Rafols & Leydesdorff (2010)). We do not propose a real methodological innovation, but rather a toolbox that allows to draw, in a few hours, several maps of the chosen scientific institution. More specifically, we show four different ways of mapping our institution, ENS de Lyon, and show how each of these gives different information. Our scope is to display - in an accessible (but not too simplistic) way - the institution's complexity thus helping to generate discussions on its policy among its scientists.

6.2 Methodology

6.2.1 Data Extraction

The "Ecole normale supérieure de Lyon" (ENS de Lyon), focused on Natural sciences, was created in Lyon in 1987 after a move from Saint-Cloud in the suburbs of Paris. In 2010, it merged with the "Social and Human Sciences" Ecole Normale Supérieure. Today, it gathers 350 researchers, 270 professors, 390 administrative and technical personnel and a budget of more than 110 million Euros. A simple query (performed in January 2011) in the ISI Web of Knowledge database (WoS, 2011) yields 7584 papers containing an *ENS de Lyon* address (mostly under the form "Ecole Normal Super Lyon", but also "ENS-LYON" and "ENS de Lyon"). We save the "Full records" of all these articles, the records containing authors, journal, year of publication, title, keywords (given by the authors and/or ISI Web of Science), subjects, addresses (institutions, cities and countries), and the list of references of the articles. It is well-known that Social and Human sciences (especially French ones) are not well represented in Web of Science. Therefore, our maps mainly deal with the natural sciences at ENS de Lyon.

Records are parsed and gathered in MySQL tables, which renders the handling of the data more straightforward. Simple frequency analysis of the records allows to a get a first global representation of the institution. Our method uses the relations present in the data (Börner *et al*, 2003) to display different perspectives on the inner structure of an institution.

6.2.2 Bibliographic coupling

Links between articles are calculated through their common references. The bibliographic coupling similarity between two articles i and j is defined as Kessler (1963):

$$\omega_{ij} = \frac{|\mathcal{R}_i \cap \mathcal{R}_j|}{\sqrt{|\mathcal{R}_i|\,|\mathcal{R}_j|}} \tag{6.1}$$

where $\mathcal{R}_i$ is the set of references of article i.

In comparison to co-citation link (which is the more usually used measure of articles similarity), bibliographic coupling (BC) offers two advantages: it allows to map recent papers (which have not yet been cited) and since it deals with all published papers (whether cited or not), it represents "normal science". The reason why weighted links are used is that they reinforce the dense (in terms of links per article) regions of the BC networks. This reinforcement facilitates the partition of the network into meaningful groups of cohesive articles, or communities. A widely used criterion to measure the quality of a partition is the modularity function Girvan & Newman (2004); Fortunato & Barthélemy (2007), which is roughly is the number of edges inside communities (as opposed to crossing between communities), minus the expected number of such edges if the network were randomly produced. We compute the graph partition using the efficient heuristic algorithm presented in Blondel *et al* (2008).

Applying the Louvain algorithm yields a partition of the network into communities (see Figure 6.2). Simple frequency analysis then allows to characterise each community through its more frequent items (keywords, authors, etc...). The significativity σ of the presence of a given item into a community is computed by comparing its frequency f in the community to its frequency f_0 within the whole database. More precisely, we use the normalized deviation

$$\sigma = \sqrt{N}\frac{f - f_0}{\sqrt{f_0(1 - f_0)}} \tag{6.2}$$

where N is the total number of article in the database. The links between two communities I and J can also be characterized qualitatively by analyzing their shared references and quantitatively by computing the mean weight $\omega_{IJ} =< \omega_{ij} >_{i\in I,\, j\in J}$.

The final step in order to create a representation of the BC communities network is to choose a visualization algorithm. We use the Gephi software Bastian *et al* (2009). Gephi is a intuitive and interactive software allowing, in which force-directed layout algorithms are implemented. These algorithms produce a graph by simulating the dynamics of the network as if it were a physical system (the nodes being charged particles and the edges

springs). The simulation is run until the system comes to an equilibrium state.

Subject	Prop of articles (%)
Biochemistry & Molecular Biology	8.25
Physics, Multidisciplinary	7.85
Computer Science, Theory & Methods	7.5
Mathematics	7.49
Geochemistry & Geophysics	7.34
Chemistry, Physical	6.41
Physics, Mathematical	6.18
Astronomy & Astrophysics	5.47
Mathematics, Applied	4.73
Cell Biology	4.43
Chemistry, Multidisciplinary	4.34
Physics, Condensed Matter	4.17
Physics, Atomic, Molecular & Chemical	3.9
Physics, Fluids & Plasmas	3.57
Genetics & Heredity	3.23

Table 6.1: **Most frequent *ENS de Lyon*'s Subjects.**

6.2.3 Copublication coupling

The data can also be analyzed through more common approaches, such as coauthoring or co-keyword analysis (Börner *et al*, 2003). For this, a list of all items (authors, keywords, addresses) are taken from the records to obtain the nodes of our maps, whose size are proportional to the number of articles in which they appear. Two nodes (items) i and j are linked whenever the number n_{ij} of articles in which they both appear is non-zero. More specifically, we use weighted links, where the co-occurrence normalized weight is chosen as

$$w_{ij} = \frac{n_{ij}}{\sqrt{n_i\, n_j}} \tag{6.3}$$

The visualization step of the produced maps is once again achieved through Gephi and its force-based layout algorithms.

6.2.4 Software available

We have developed a "Biblio Toolbox" which allows to draw the different maps presented here in a few hours. The toolbox needs access to Web of Science database but otherwise

relies on OpenSource software. It is available at our website (http://www.sebastian-grauwin.com/).

6.3 Gaining perspective on the *ENS de Lyon*

6.3.1 Statistical analysis

ENS de Lyon gathers a broad spectrum of scientific subjects (Table 6.1), mostly in the natural sciences as discussed above. The institution has significantly grown over the last 20 years, as shown by its increasing production of papers (Fig 6.1). Our data gathers 12398 distinct authors, among which 952 have authored more than 5 papers. By construction of the database, at least one author of each article is a member of ENS de Lyon but this number also takes into account all the authors of the papers among whom some may not be members of the ENS. *ENS de Lyon* collaborates with a broad range of institutions of different countries as shown below.

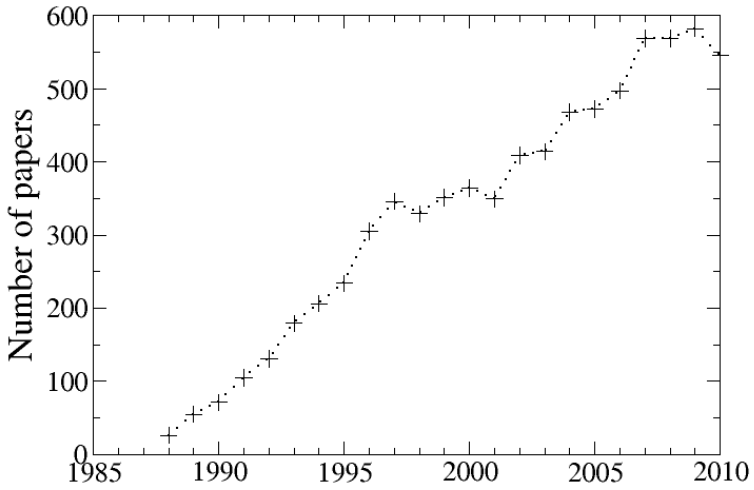

Figure 6.1: **Number of paper with an *ENS de Lyon* address published by year**, according to WoS, January 2011.

6.3.2 Bibliographic Coupling communities map

Figure 6.2 shows the map obtained with bibliographic coupling of articles and their grouping into "natural" subfields through modularity maximization. Each community is characterized by its more frequent author and keyword. Table 6.2 displays an "ID card" for the community labelled *Hansen JP/Molec-Dynamics*. This community gathers physicists interested in the understanding of condensed matter using molecular dynamics simulations. The "ID Cards" of the other communities are available online on http://www.sebastian-grauwin.com/.

What do we learn from this first map? First, note that the spatial organization of the communities fits well with the scientific organization of ENS de Lyon in different departments (different colors in Figure 6.2). This confirms that bibliographic coupling can recover the scientific organization of institutions. Interestingly, the precise community structure does not match the inner administrative/scientific subdivision of departments. For example, the physics lab is administratively divided into four groups, while our map distinguishes seven teams. This raises interesting questions on the structuration of the groups and their interactions. Two physics' communities (Oswald P/LiquidCrystals and Peyrard M/DNA) belong to the "soft-matter and biological systems" group but our map shows that they are quite distant, which means that they do not share many references. The difference between the map and the physics lab organization is one example of the discussions that our work can generate.

Another example is given by the overall spatial structure. Our map clearly places physics at the scientific center of the ENS de Lyon, a fact that was used by its director to suggest the importance of his lab within the institution. The question is then : how much does this central position depend on the precise visualization algorithm used? Is it robust enough to allow for an interpretation and possibly orient governance? The forthcoming maps will comment on this issue, but let us already note that the central position of the Physics Lab *within this representation* is quite robust. The reason is quite simple : the Physics Lab is the only one to have strong links to the other labs. Indeed, different physics' communities are linked to all other labs (for example Mathematics and Computer science (through Livine E/Quantum gravity), to Biology (through Peyrard M/DNA)...). The other labs are strongly linked only to one or two other labs (for example, Biology is only linked to Chemistry, through Pichot C/Adsorption, in addition to its link with the Physics lab), which explains their more peripheral position in the map. Therefore, the central position of the Physics lab can tentatively be interpreted as its central position in terms of modelling tools (molecular simulations tools

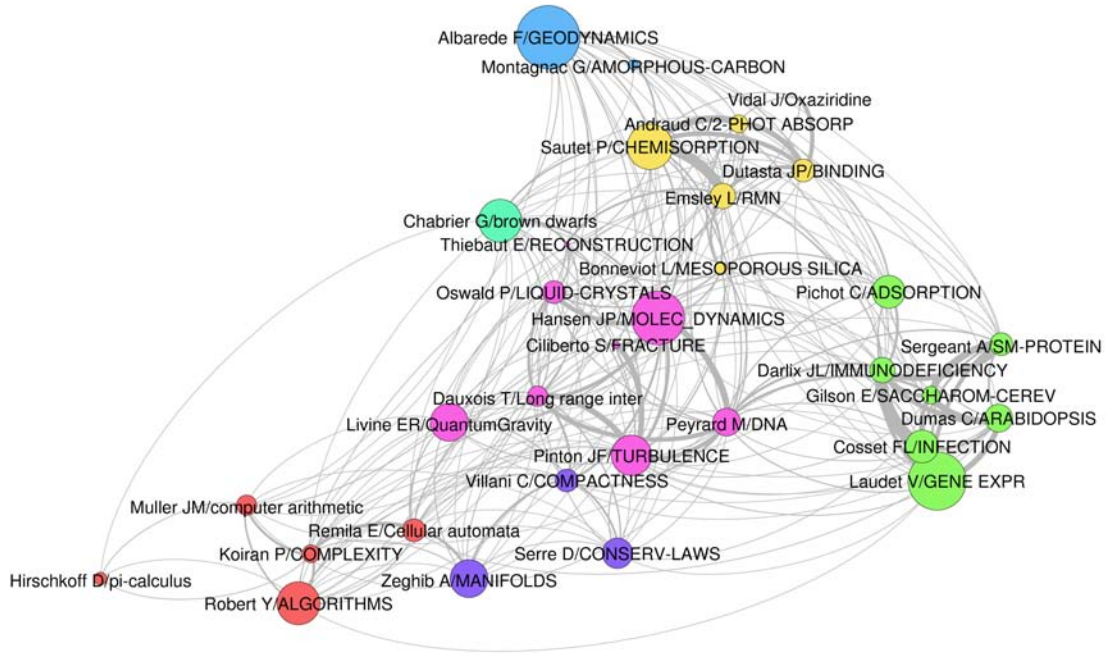

Figure 6.2: **"Bibliographic Coupling" community structure of the *ENS de Lyon*.** The surface of a community I is proportional to its number of articles N_I and the width of the link between two communities I and J is proportional to the mean bibliographic coupling. For the sake of clarity, communities with less than 10 articles are not displayed. Labels are obtained thanks to a frequency analysis of the authors and keywords. Each color corresponds to one of the ENS de Lyon scientific departments : biology (green), chemistry (yellow), physics (pink), computer science (red), mathematics (violet), earth sciences (blue) and astrophysics (turquoise). The belonging of a community to a department is determined through the proportion of community's articles that use the department (for an example, Table 2 shows that more than 50% of "HansenJP/Molec-Dynamics" articles' display the Physics Lab in the address).

shared with chemists for example), experimental tools (on "frictional mechanics" with the geophysics lab for example) or theoretical concepts (spin glass theory also studied by mathematicians). All these shared tools generate common references which lead to the links that structure our map.

6.3.3 International collaborations

It is straightforward to use the communities of the preceding map to include the international collaborations of the different teams (Fig. 6.3). We simply define links as given by the frequency of appearance of a foreign country in the community's articles addresses. For example, the strongest link is obtained for the Astrophysics papers, for which 41% of the papers are written in collaboration with a USA institution. The map shows that some groups rely heavily on many international collaborations (Emsley L/RMN has strong links with England, Italy and USA), while others are strongly linked to a single country (Dauxois T/Long range inter, to Italy) and others have mainly French collaborations (Oswald P/Liquid crystals).

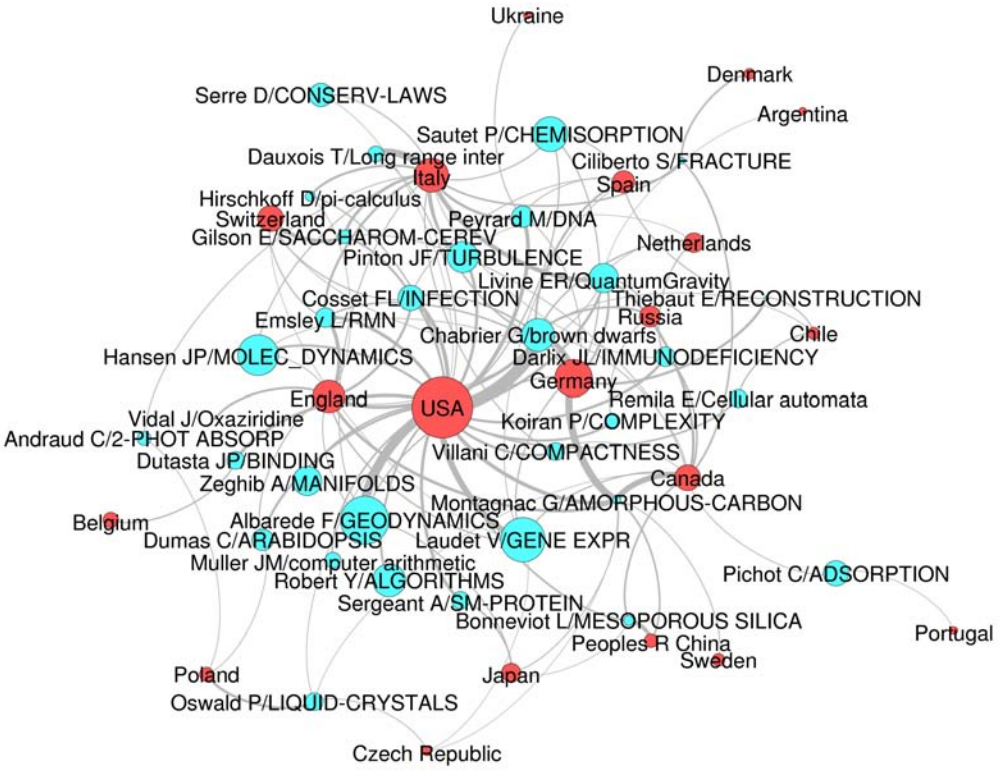

Figure 6.3: **International collaborations of the communities.** The size of the nodes correspond to the number of articles in each community which imply a collaboration with a foreign country. We only keep countries appearing in more than 10 articles and links corresponding to more than 3% of the articles implying a collaboration with the linked country. The width of the links is proportional to the proportion of linked articles.

6.3.4 Co-keywords, co-authors and heterogeneous maps

We now turn to more traditional maps, obtained by co-occurrence of keywords or authors in articles. Figure 6.4 shows the co-keywords map obtained by using Web of Science and authors' keywords. One should be cautious since some terms are clearly polysemic ("evolution", "particles"...) and create links between subdisciplines which are not very relevant. However, it is clear that physics is no longer at the center of the map. Instead, "crystal-structure" links chemistry (top left) with biology (right), "growth" links biology to physics and "transition" and "dynamics" link chemistry to physics (left). Another significant difference : what appeared to be a coherent whole when investigated through bibliographic coupling (the "Albarede F/Geodynamics" community) turns out to split into geochemistry (bottom of Figure 6.4) and geophysics (just up of the latter, close to physics, with keywords as "high-temperatures" or "high-pressures").

Figure 6.5 displays a co-author map. This represents an accessible way of showing data to the institutions' scientists, since names are usually well-known by the community. It also represents a good way to tap into directors' previous knowledge of the institution. However, coauthorship indicates quite a different (and stronger) link from the link established by sharing references (as in bibliographic coupling). This is visible in Figure 6.5 which does not show many links across disciplines (and some of the links are actually homonyms, such as Bertin E). The main co-publication link arises from collaborations between a biophysics lab and computer simulations of biological molecules (Peyrard/Bouvet/Gilson).

To improve over the limitations of both co-keyword and co-author analysis and gather most of the available information in a single map, it is possible to include all the co-occurrences between keywords, authors and institutions. Fig. 6.6 shows the map obtained for the ENS de Lyon. It displays the connecting rôle of a physics-biology interdisciplinary lab (Lab Joliot Curie, center right). One can also see that, while the CNRS plays an important an central rôle, other institutions collaborate on more specialized subfields (for example Univ California, Berkeley, lower left).

6.4 Discussion, Conclusions

Our aim in this chapter was to present a toolbox to make institution mapping easy and rapid and to show on the example of our institution, ENS de Lyon, what kind of insights can be derived from these maps. It should be clear by now that there is not a unique (or a "best") map of a scientific institutions, but rather many possible representations, each map containing a projection from a specific perspective (Leydesdorff & Rafols, 2009;

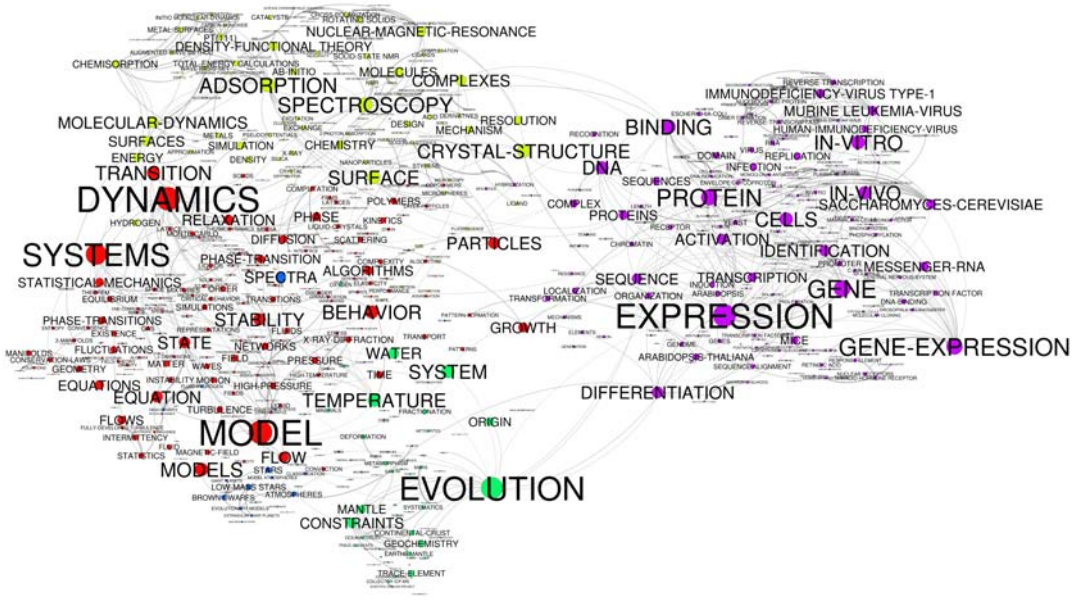

Figure 6.4: **Co-Keywords Network.** The size of the nodes is proportional to the number of times a keyword is used in our database. The width of the links indicates the cooccurrence weight between two keywords in the same article. We keep only keywords used in more than 10 publications. Colors correspond to a community analysis performed by gephi based on the same Louvain algorithm used for the Bibliographic Coupling analysis.

Roessner, 2000; Stirling, 2008).

We are now experimenting with ENS scientists' and direction. ENS heads are enthusiastic about this global vision and five posters representing this chapters' figures are now displayed in the building. We hope that scientists at ENS de Lyon will test these maps against their own knowledge of the institution, will argue with us when what we picture does not fit and join the public discussion by offering alternative interpretations. As Nietzsche said (Nietzsche, 1969): "the more affects we allow to speak about one thing, the more eyes, different eyes, we can use to observe one thing, the more complete will our 'concept' of this thing, our 'objectivity' be". The point is that although everybody acknowledges that maps are only representations and not the real thing, maps

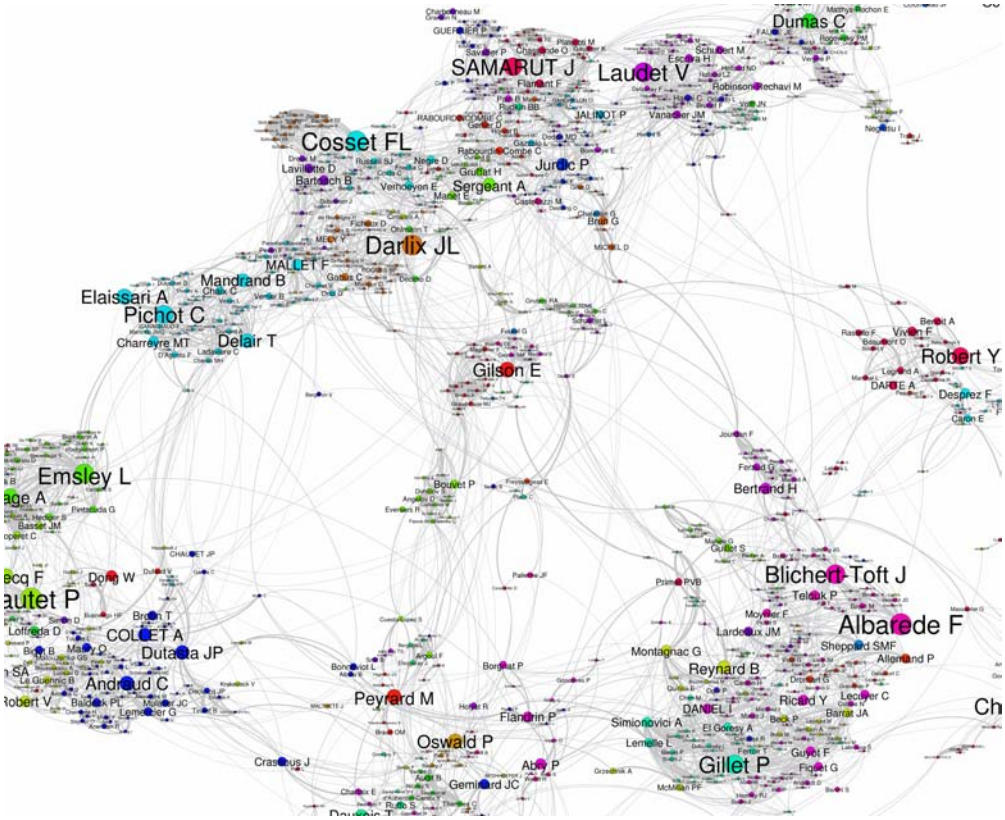

Figure 6.5: **Co-Authors Network (detail).** The size of the nodes is proportional to the number of articles of our database authored by the author. The width of the links indicates the cooccurrence weight between two co-authors. We keep only authors used with more than 5 publications. Colors correspond to a community analysis performed by Gephi based on the same Louvain algorithm we used for the bibliographic coupling analysis.

affect how we think about the institution (Wood & Fels, 2008).

We hope that our toolbox will lead other scientists to build maps of their own institutions, thus fostering ongoing dialogue and praxis in the institution. Future work includes preparing different maps for successive time periods, in order to grasp the evolution of the institution, and collaboration with other institutions (such as CNRS and CEMAGREF) which are interested in such global maps.

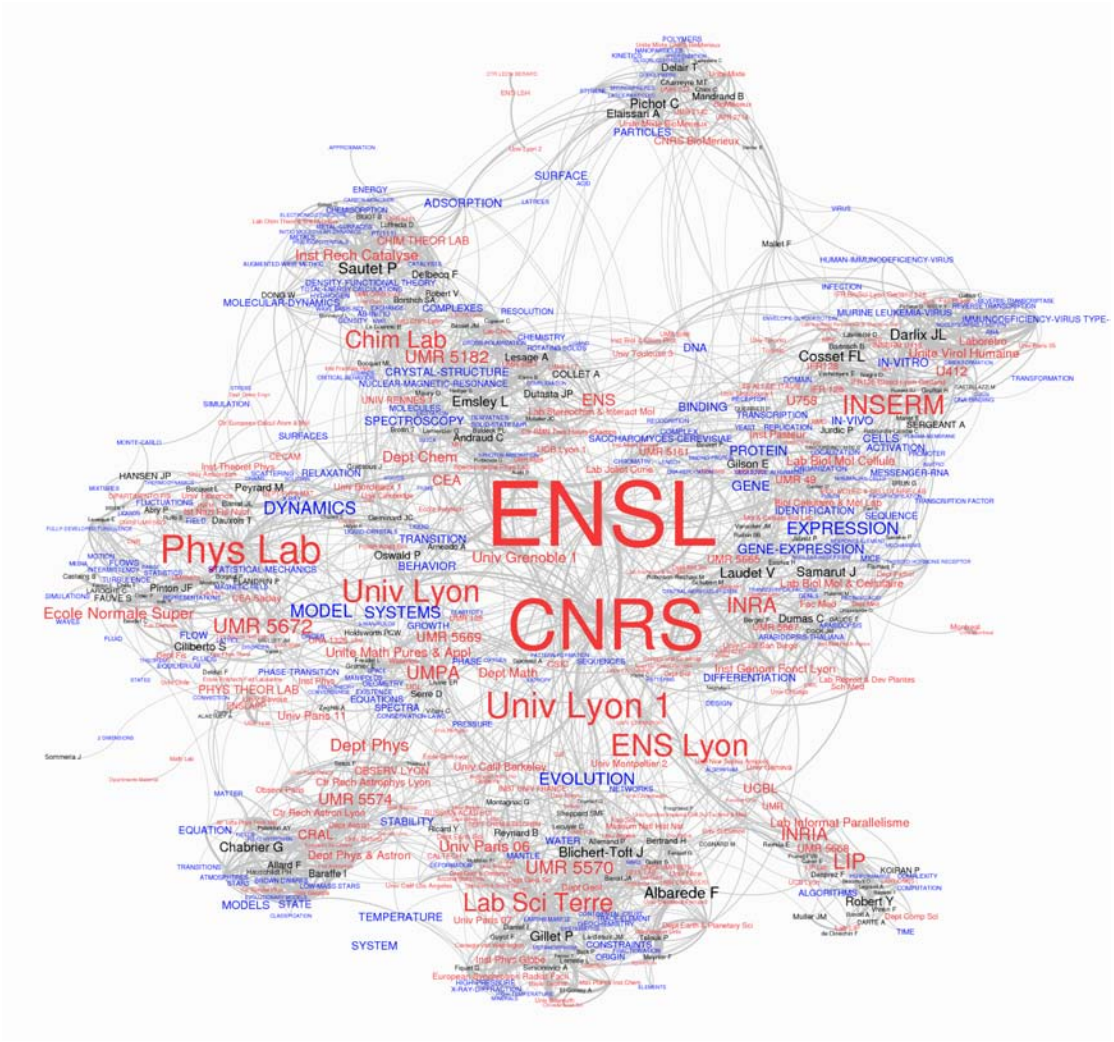

Figure 6.6: **Heterogeneous Network, mixing authors, keywords and institutions.** The size of the labels is proportional to the number of articles of our database in which an item appear (we keep only items used in more than 20 publications). The width of the links indicates the cooccurrence weight between two items (we kept only links with a co-occurrence weight $\omega > 0.1$). Colors correspond to the type of the items (authors in black, keywords in blue and institutions in red).

Table 6.2: **Community "ID Card".** The community *Hansen JP/Molec-Dynamics* contains $N = 547$ articles. Its average internal link weight is $< \omega_{in} > \simeq 1/223$ (roughly, two random articles within the community share 1 reference over 223).

Institution	prop	σ
Ecole Normale Super Lyon	0.766	0.19
Phys Lab	0.543	26.29
CNRS	0.508	1.95
UMR 5672	0.133	10.52
Univ Lyon	0.111	1.33
Dept Phys	0.076	6.73
Univ Lyon 1	0.075	-3.41
ENS Lyon	0.073	-1.47
Phys Theor Lab	0.073	12.21
CECAM	0.065	15.67

Authors	Nb Paper
Hansen JP	55
Barrat JL	40
Bocquet L	38
Ciliberto S	38
Geminard JC	24
Holdsworth PCW	22
Alastuey A	21
Charlaix E	20
Dong W	20
Cornu F	19

Subject	prop	σ
Physics, Multidisciplinary	0.27	16.7
Physics, Mathematical	0.254	18.67
Physics, Fluids & Plasmas	0.16	15.77
Physics, Condensed Matter	0.128	10.09
Physics, Atomic, Molecular & Chemical	0.117	9.42
Chemistry, Physical	0.098	3.3
Materials Science, Multidisciplinary	0.069	6.25
Mechanics	0.065	6.7
Physics, Applied	0.064	9.03
Polymer Science	0.042	3.15

Countries	Nb Paper
France	704
USA	89
Italy	47
England	37
Germany	35
Netherlands	25
Poland	25
Switzerland	23
Japan	20
Chile	15

Keyword	prop	σ
DYNAMICS	0.135	14.03
SYSTEMS	0.117	13.65
MODEL	0.104	9.9
MOLECULAR-DYNAMICS	0.053	10.41
BEHAVIOR	0.043	6.65
TRANSITION	0.042	6.12
FLUIDS	0.04	12.14
RELAXATION	0.038	7.27
FLOW	0.038	6.96
MONTE-CARLO	0.036	12.65

Refs	Times used
Hansen JP, 1986, THEORY SIMPLE LIQUID	60
Cugliandolo LF, 1997, PHYS REV E	37
Cugliandolo LF, 1993, PHYS REV LETT	34
Kosterlitz JM, 1973, J PHYS C SOLID STATE	25
Gotze W, 1992, REP PROG PHYS	22
Bouchaud JP, 1998, SPIN GLASSES RANDOM	22
Jaeger HM, 1996, REV MOD PHYS	22
Alastruey A, 1989, PHYS REV A	21
Frenkel D, 2002, UNDERSTANDING MOL SI	18
Grigera TS, 1999, PHYS REV LETT	18

Journal	prop	σ
PHYSICAL REVIEW E	0.155	19.32
PHYSICAL REVIEW LETTERS	0.095	11.28
JOURNAL OF CHEMICAL PHYSICS	0.067	13.92
EUROPHYSICS LETTERS	0.058	11.24
JOURNAL OF PHYSICS-CONDENSED MATTER	0.043	11.22
JOURNAL OF STATISTICAL PHYSICS	0.038	9.18
EUROPEAN PHYSICAL JOURNAL B	0.029	6.78
JOURNAL OF STATISTICAL MECHANICS-THEORY AND EXPERIMENT	0.027	8.39
PHYSICAL REVIEW B	0.023	2.1
JOURNAL OF PHYSICAL CHEMISTRY B	0.021	4.36

Refs (journals)	Times used
PHYS REV LETT	1709
J CHEM PHYS	1406
PHYS REV B	480
PHYS REV E	408
PHYS REV A	399
EUROPHYS LETT	353
PHYS REV E 1	343
J STAT PHYS	310
NATURE	286
PHYSICA A	276

Part III

Emergence of Institutions in Social Systems

Chapter 7

The Whole is Smaller than the Sum of its Parts: a Tentative Tardean Model

Word of caution: for more than a year, we have been meeting a team of sociologists led by Bruno Latour at Science Po Medialab. In the frequent discussions, we have been trying to define the essential bricks of a sociologically grounded model, focusing on the social theory developed by Gabriel Tarde at the end of the 19*th century. Most of the time, however, was devoted to mutual understanding and to development of a common language. This was the case for me at least! The introduction of this chapter is an attempt to summarize the main ideas that came up in our discussions. I hope that the unavoidable misunderstanding and distortions of sociological ideas will be forgiven.*

7.1 Introduction

The individual / society dichotomy

Various sociological theories try to apprehend social phenomena under different points of view. Hence, according to the *holistic paradigm* which is issued from Emile Durkheim's ideas, individuals are embedded into social structures and institutions that constrain them, shape their actions and emotions. Social facts (such as social classes) are notions pre-existing the individuals that should be taken as primary and most significant to explain social phenomena. On the contrary, in the atomic (or individualistic) paradigm, individuals should be considered as the central ontological elements in social systems. According to this notion, each individual is a social atom, linked to other atoms and

which acts in function of its interests and desires. Constant interactions between the atoms construct society, *ie* structures, institutions, norms, which are supposed to serve the interests of the individuals.

Both the holistic and the atomistic approaches are based on a strong assumption: there is a clear dichotomy between two "levels", namely individuals and society. This clear distinction of two levels is often summarized by the notion that "the whole is greater than the sum of its parts". This assumption is put into question in some recent approaches (see Latour et al, forthcoming). According to these approaches, the distinction between two levels is an artefact originating in the difficulty of navigating through huge amounts of data and in visualizing the evolution of social phenomena without making a distinction between the individual and the aggregated levels.

From KISS to KIDS

A clear example of the "atomistic" approach are agent-based models, which attempt to recreate complex collective phenomena through the interactions of "rational", utility-maximizing agents. Examples of such models include :

- Schelling's segregation model (Schelling, 1971), which uses individual interactions to explain the emergence of segregation on a global scale
- Large-scale models were proposed by Axelrod (1997). They simulate the emergence of global phenomena such as exchange markets, seasonal migrations... starting from individual interactions between simple agents.
- Opinion dynamics, whose main scope is to understand why cultural diversity can persist at the global level even when individual interactions push towards convergence. These models have specifically interested a growing number of physicists (for a recent review, see Science, 2009; Castellano *et al.*, 2009).
- Many additional examples can be found in the Journal of Artificial Societies and Social Simulation (http://jasss.soc.surrey.ac.uk/JASSS.html).

In the last decades, "atomic" agent-based models demonstrated their importance to overpass the severe limitations of current theoretical frameworks, especially in the modelling of economic phenomena. They allow to move beyond equilibrium states to study transient regimes, which show far more complex and interesting behavior. They overpass the drastic simplifications needed to warrant analytical treatments (such as agents' homogeneity, cognitive closure in game theory...). A key principle common to

most of these model is known as K.I.S.S. ("Keep it simple, stupid", Axelrod (1997)). Indeed, in the various effort to model social phenomena, the main tradition has so far mainly consisted in starting from discrete atomic entities endowed with the smallest set of properties and to see whether or not, through their interactions, it would be possible to obtain the emergent phenomena most often associated with the notion of society (For a recent overview of this approach, see Cho, 2005). In this tradition, the success of the model often lies on finding the minimal sufficient rules for atoms and interactions allowing to obtain the maximal number of structural features.

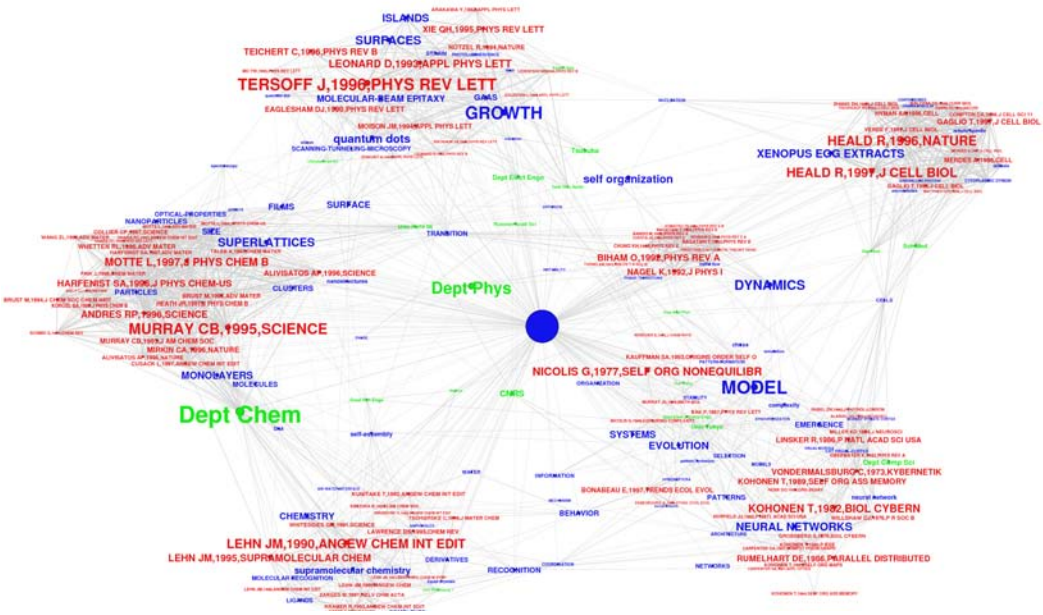

Figure 7.1: **'Profile' of the keyword *self-organization*.** This co-occurrence map shows the keywords, references and addresses of the articles which use the keyword *self-organization* (the central node without label) between 1996 and 2000 according to the Web of Science©. Refer to chapter 6 for the method of construction of such maps.

Although this tradition has been productive in many cases especially because it cleared the way of too many ad hoc assumptions added by social scientists, it had the drawback of appearing too simplistic to account for the whole complexity of social existence: individual can not be reduced to a unique real characterizing their opinion nor do they move solely for finding a "good" neighborhood! Instead of appearing as a great enlightenment, these simplistic models might have appear to the eyes of some social scientists as artefacts. As a consequence, the whole agent-based approach might have in certain cases amplified the traditional divide between quantitative and qualitative

methods in social sciences.

As a reply to this kind of criticisms, some scientists have developed an alternative modelling principle, called K.I.D.S. ("Keep it descriptive, stupid!", Edmonds & Moss (2005)). One starts with a model that is as descriptive as possible, taking into account as many features of agents as possible, generally the characteristics for which data is available, including anecdotal accounts and expert opinion. Simplification is only applied when the model and evidence justifies it.

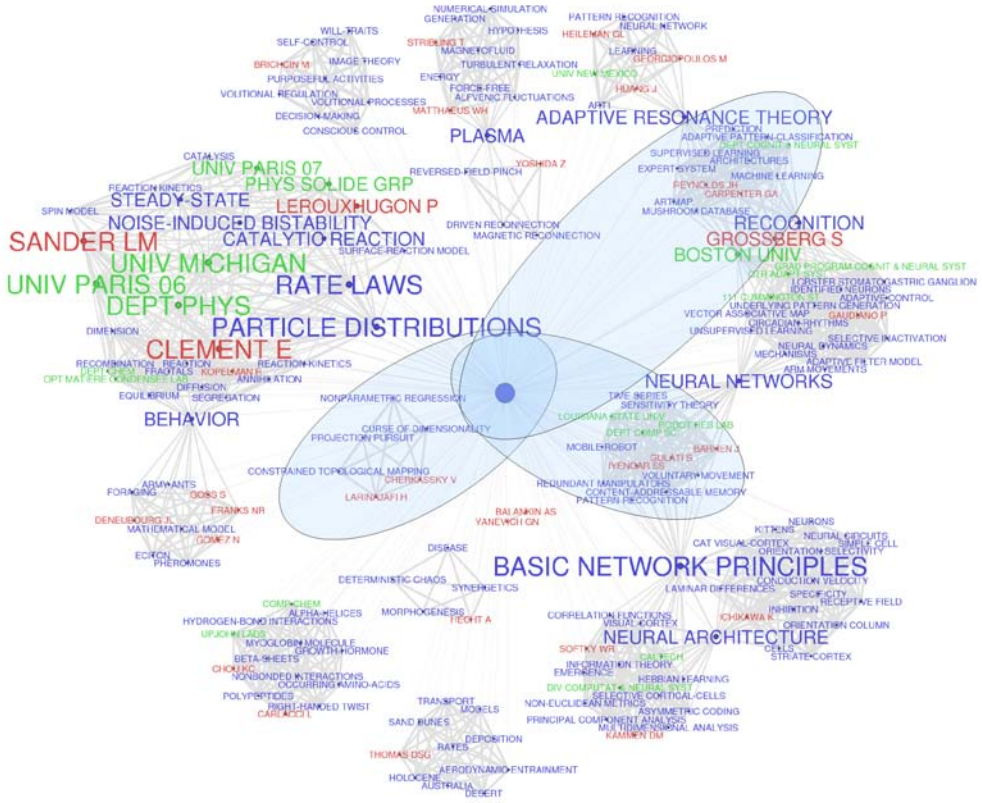

Figure 7.2: **Keyword *self-organization* considered as a 'whole'** produced by the intersection of 'monads' (articles) which are far richer than this single keyword. We use the 18 articles published in 1991 which use the keyword *self-organization* (the central node without label) and built the co-occurrence map of their authors, addresses and keywords as in Figure 7.1. To highlight the idea of 'intersection', the attributes of three 'monads' (articles) are shown surrounded by a circle.

A Tardean approach for understanding social data

Therefore, it seems interesting to build a model able to follow the emergence and evolution of social structures or institutions with two specific starting points :

- there should not be any ontological distinction between "agents" and "structures" *ie* no notion of levels
- following the KIDS principle, agents should be taken as complex as possible

A good starting point to overpass the individual-structure paradigm may well be the work by Gabriel Tarde, a $19th$ century sociologist whose work has been somewhat dormant in the social science during the $20th$ century, but has been recently unearthed and given a new impetus by the development of Actor-Network Theory (Latour, 2005; Barry & Thrift, 2007; Latour, 2010).

The central element in Tarde's theory (see Tarde, 1893, 1890) is the *monad*, a concept originally developed by Leibniz. For Tarde, there is no need for a notion of overarching society and there is no individual to begin with either: everything is a monad, that is *a representation, a reflection, or an internalization of a whole set of other elements borrowed from the world around it* (Latour, 2010). A monad has to be understood as an entity with some asperity which can be linked to other entities. An individual is a monad defined by the set of links towards all the entities that characterizes the individual (in a sense, his "profile", or curriculum vitae, *ie* a list of everything the individual has done in the past, the people he has seen, the clothes he has worn...). A group of people is a monad defined by the the set of links towards entities that defines them (/ that they have in common). A neuron is a monad defined by the set of neurons connected with it. There is not a single entity that can be defined *per se*, every entity is defined by its "profile", i.e. the set of entities to which it is linked (Figure 7.1). Each "individual" entity is then a "whole" by itself. In Tarde's ontology, there are no individuals, no wholes, only monads, monads connected to monads, monads within monads...

Monads perfectly match our conditions: they are complex entities which go beyond the notions of agents and structures. Still, how can social phenomena be understood in terms of monads? Tarde proposed two notions to explain social phenomena: *imitation*[1]

[1] In Tarde (1890)'s words:
Quel est le fait social élémentaire ?
C'est la communication ou la modification d'un état de conscience par l'action d'un être conscient sur un autre. (...) parler à quelqu'un, prier une idole, tisser un vêtement, scier un arbre, donner un coup de couteau à un ennemi, sculpter une pierre, ce sont là des actes sociaux, car il n'y a que l'homme en

and *invention*. Everybody imitates what or who he admires, things that he finds good for him, but creates an original mix from the imitated models chosen at different sources. History can be understood as successive imitative fluxes, successive successful models copied by a large number of individuals. Imitation is a natural counterpart of the notion of monad. Indeed, conceiving individuals as monads means conceiving them as a large set of reflections. We find pieces of each other in one another, and choose to imitate pieces of one another. In Tarde's terminology, our choice of things to imitate and the way we innovate by combining them is driven by *beliefs* and *desires*.

Let us finally stress an important point, which is the relation between monads and the usual notion of "structural features" or wholes. For Tarde, what is usually taken as "structural features" are the intersections of complex individual monads (Figure 7.2). In some sense, for him, the whole is always simpler and smaller than the parts... These structures evolve in time because the monads themselves combine differently the elements of the world but may repeat and keep unchanged some feature (the keyword "self-organization" persist in time while the authors citing it change over time), which Tarde calls *imitative rays*.

In next section, we describe a tentative model describing a virtual society built on Tarde's precepts. Our goal is to emphasize and discuss the harsh constraints imposed by the need for properly defined, complete and coherent algorithms required for simulations. Needless to say, the positive side of these constraints is to force social theorists to provide better definitions of their concepts !

7.2 A Tentative Tardean Model

We present in this section an attempt to formalize Tarde's ideas into a mathematically defined model to allow the construction of a simulation algorithm. To build models of interactions between complex agents, we have greatly benefited from discussions with Guillaume Beslon and we have freely gleaned in a wide literature, mainly (Holland, 2000; Dawkins, 1976; Kauffman, 1993; Gilbert *et al*, 2001). A discussion on the difficulties encountered in this exercise follows in next section.

The model considers a population of N complex agents ($i = 1, 2, ...N$) completely characterized by a sequence $S_i = (s_{i1}, s_{i2}, ...s_{iL})$ of $L \gg 1$ bits. In analogy with what is done in genetic algorithms (see for example Knibbe *et al.*, 2007), we assume the existence

société qui agisse de la sorte, et sans l'exemple des autres hommes qu'il a copiées volontairement ou involontairement depuis le berceau, il n'agirait pas ainsi. Le caractère commun des actes sociaux, en effet, c'est d'être imitatifs. Voilà donc un caractère bien net et, qui plus est, objectif.

of a 'fitness function'[2] f which assign a score $f(S) \in [-1, 1]$ to each sequence S. This fitness function is taken to be very simple for test cases, but we ultimately want to use complicated function presenting several competing optima (see Figure 7.3). While the interpretation of the fitness function is widely accepted in Darwinian evolution, its meaning in a social context is less clear and will be discussed below. This function is supposed to be fixed and identical for all the agents.

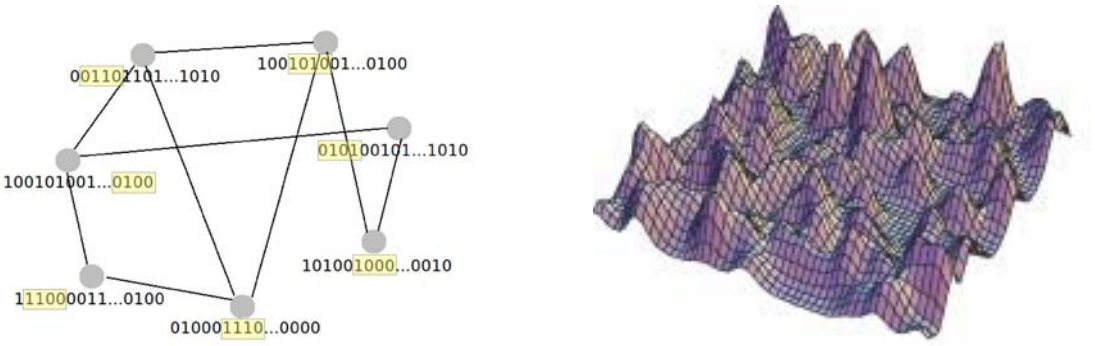

Figure 7.3: **Left. Complex agents.** The agents of our model are completely characterized by a sequence of bits that they can exchange with their neighbors. **Right. Fitness landscape.** Example of a bumpy fitness landscape, here with two input variables. The landscape shows several local maxima corresponding to several optimum sets of input parameters.

The agents can imitate each other by copying parts of each other bits' sequences if they are linked. The initial interaction network is taken as an Erdös-Rényi network of average degree K, *ie* the agents are initially linked to K neighbors.

The dynamics of the systems is defined as follows: at each interaction, an agent i is picked at random. Then, one of two following processes occurs :

- With a probability α_1, imitation occurs.
  A neighbor j of the picked agent i and a random sequence of agent j is chosen at random. Agent i then imitates agent j by copying the sequence of bits in his own sequence (at the same position) with a probability:

$$P_1 = \frac{1}{1 + \exp -\beta_1(f(S'_i) - f(S_i))} \tag{7.1}$$

  where $f(S_i)$ is the fitness of agent i sequence and $f(S'_i)$ the fitness his sequence

[2]In genetic algorithms, fitness functions are used to determine the degree of optimality of a given solution (such as a virtual chromosome). The best solutions are then mixed to produce a new generation of more adapted solutions.

will have if he decides to imitate. The parameter β_1 allows for some tuning of the imitation process. For $\beta_1 \gg 1$, imitation will occur if and only if it implies an increase in agent i fitness. For $\beta_1 \ll 1$, the probability to imitate does not depend on the fitness value and agent i decision is just random.

- With a probability α_2, innovation occurs.
  A bit $l \in (1, ..., L)$ of agent i is chosen at random. Agent i will to choose to 'innovate' by flipping this bit with a probability

$$P_2 = \frac{1}{1 + \exp -\beta_2(f(S_i') - f(S_i))} \tag{7.2}$$

  where $f(S_i)$ is the fitness of agent i's sequence and $f(S_i')$ the fitness his sequence will have if he decides to innovate. Here again, the parameter β_2 allows for some tuning of the influence of the fitness function on the innovation process.

Obviously, the parameters α_1 and α_2 (where $\alpha_1 + \alpha_2 = 1$) define the ratios between the time scales on which each process occurs. The parameters β_1 and β_2 measure the influence of noise in the agents' choices

In order to understand the rationale behind these rules, we discuss the expected qualitative evolution of the model. When agents are completely isolated (*ie* not linked to other agents), the only way to increase their fitness is by a succession of inventions in their sequences, which will happen on a rather long time scale (since only one bit is changed at each time). When the agents are connected, two processes will accelerate the increase in fitness. First, longer sequences may be exchanged in a single iteration. Second, "good" sequences may be imitated and therefore act like social coordinators, forming stable "structures" or "institutions".

Depending on the exact definition of the fitness function, a large phenomenology of collective events can be expected. If the fitness function is constant, each sequence is a priori equivalent. In case there is no innovation ($\alpha_2 = 0$), one expect that the agents will converge in the long run toward the same sequence S^*. With our choice of W and D, one can even expect this sequence to correspond to the sequence of initial majority bits (*ie* for each $l \in (1, ..., L)$, $s_l^* = 1$ if $\sum_i s_{il} > N/2$, and $s_l^* = 0$ otherwise). If invention is added, there will be a competition between imitation and invention, the imitative process promoting convergence of the agents' sequences and the invention process promoting diversity.

Suppose now that the fitness function has a single maximum, *eg* by taking $f(S) = \sum_l s_l$ (meaning that the fitness function increases linearly with the number of 1 in a

sequence). In case there is no innovation, once again we expect a convergence toward a consensus sequence S^*. However, even if the agents are rational imitators ($\beta_1 \gg 1$), the consensus sequence does not necessarily correspond to the optimum sequence in terms of fitness. Indeed, all it takes for imitating a sequence is an increase in fitness. A 'second-best' sequence such as '11110111' would be imitated by most of the agents and could diffuse and impose itself to the general population. In a sense, there is a competition between imitation and optimization that can temper the collective optimization. Turning rational ($\beta_2 \gg 1$) innovation on would allow here the agents to mutate their sequence to obtain a fully optimal sequence. If innovation is more random (β_2 finite), the invention process promote diversity and there is here again a competition between imitation and invention.

In case of more complex fitness functions presenting several local maxima (see Figure 7.3), more complicated collective behaviors are obviously expected. In line with Tarde's intuition, we do not expect a single sequence to dominate all the society. We rather anticipate that these local maxima will affect the dynamics by "recruiting" agents in some "niches" (*ie* the patterns of sequences allowing to reach those maxima). These standardized niches will endure through time before being modified again through agent's innovations. We therefore expect global trend towards complexification of sequences, leading to an overall increase of agent's fitness. An important and tricky part of the work consists in playing with the numerous parameters of the model to obtain these satisfactory outputs.

Obviously, all these speculations are only anticipations whose validity has to be checked. Some unpredictable consequences of the chosen dynamic rules, parameters, fitness function may very well be uncovered through actual simulations.

7.3 Discussion

The different elements present in our model has been chosen to fulfil Tarde's prescriptions. Our agents have been chosen as complex entities : long binary sequences. Moreover, sequences and sequences define at the same time the agents and the structures (diffusing sequences). Hence, in our model there is no ontological distinction between agents and structure, which is exactly what we wanted at the beginning. In agreement with Tarde's sociology, our model offers a complete reversibility in following how agents come to agree by simplifying themselves through the influence of diffusing sequences, and following how these sequences spread among populations of agents by modifying themselves and "recruiting" agents interested in them. In that sense, these diffusing

sequences adopted and sometimes modified by the agents are similar to Tarde's 'imitative rays'. Moreover, once a sequence has been accepted by an agent, it changes the agent's perception of all the other sequences (through the utility function), making these perceptions more similar to the perceptions of all the agents sharing this sequence.

If our model manages to recover important Tardean features, several of the choices we made are debatable. The main problem we encountered was to find a proper explicit criterion for deciding whether an agent should choose to imitate or not, innovate or not, etc... Tarde bases the agents' choices on the notions of *belief* and *desire*. We choose to transcribe these notions in a unique, constant in time fitness function, which already shows complex collective behaviors. From a sociological point of view, this fitness function gives a strong foundation to the social outcomes without completely determining them (for a given complex fitness function, the states of the systems might depend on the initial conditions, some random processes...). This 'monolithic' extrinsic aspect may however appear at odds with the common notion of beliefs and desires. In another version of the model, individual and dynamic fitness functions could emerge as a result of the agents' interactions and choices. Another possibility would be to introduce individual effective fitness functions. In this hypothesis, the existence of a unique objective function is assumed, but the agents are not aware of it. They use their own filtered versions obtained through experience and social interactions. But could we define properly those effective fitness functions?

Our choices of dynamic rules are also debatable. We use logit probabilities mostly by convenience, while the logit form is *a priori* no more justified than any other form taking into account the fitness variations. The same criticism applies to the other (un)explicit choices. For example, we could choose to work in a whole connected society rather than on a random network, or allow the agents to change neighbors instead of assuming a static interaction network. The general question we encountered while trying to formalize Tarde's ideas hence was : what are the good criteria to judge that we have built a 'good' model, or that the simulation outputs could be interesting?

The answer we reach here is that the formalization of a theory is not sufficient to built a complete, ready-for-simulations model. In order to complete this framework, one should start from the study of a specific social phenomenon and derive precise rules from analysis of real data...

Chapter 8

Lasting Structures from Non Lasting Entities

One of the central questions of sociology is the durability of institutions at the human time scale : how can lasting structures emerge from non lasting entities? This is one of the recurrent topics of our discussions with the Medialab team. Here we show a simple model, inspired by models proposed by physicists, which tackles this question. The relevance of such a model for the real social world is discussed at the end of this chapter.

8.1 Introduction, motivation

One of the standard strands of models for physicists interested in social systems are the so-called "opinion" models. As often in this field, one should be cautious about the terms, because "opinion" refers here only to a drastically simplified image of this rich social concept, i.e. a real or integer number supposed to distinguish among different possible opinions. These models introduce interactions between the agents which lead to changes in agents' opinions and the links they establish with other agents. Various models differ on the precise nature of the opinion (whether it is discrete or continuous, uni or multi-dimensional), on the topology and the dynamics of the interaction network (complete graph, regular lattices, random graph, scale-free network, static or adaptive network), on the precise rule of interaction between agents ... (for a recent review, refer to Castellano *et al.*, 2009, and references therein).

A central question investigated by these models is the persistance of pluralism of opinion in a more and more connected world, in which individuals are constantly imi-

tating each other. In Axelrod (1997)'s word: *If people tend to become more alike in their beliefs, attitudes and behaviors when they interact, why do not all differences disappear ?* Most of these models concentrate on the study of the final states of the system. There are several possibilities : either all agents end up into a single cluster (*consensus*), or several opinion clusters form, containing a significant fraction of agents (*polarization*), or alternatively agents are scattered on a large number $O(N)$ of clusters (*fragmentation*). Most papers are focused in obtaining the well-balanced *polarized* state, because the two extremes of consensus or fragmentation are trivial outcomes. Axelrod (1997) and Deffuant *et al.* (2000) are two widely known examples of models with imitating agents which yield polarized final states. However, the stability of the polarized state rests on a - rather unrealistic - strict interaction rule allowing agents to interaction only if their opinions are close enough. If this strict condition is softened (*eg* by introducing some noise), the polarized states become instable and the system reaches consensus (Klemm *et al.*, 2003; Kozma & Barrat, 2008). To improve the robustness of the polarized states, recent models have introduced an "adaptive" rewiring of the network, allowing agents to dynamically rewire their links to agents with similar opinions (Holme & Newman, 2006; Kozma & Barrat, 2008; Iñiguez *et al.*, 2009). This rewiring leads to the creation of more stable polarized states. As Kozma & Barrat (2008) claimed : *the behavior of the model on adaptive networks is in fact more robust than on static networks, since the same global picture* [polarized states] *is observed for strict or probabilistic communication rules.* However, a close inspection of the probabilistic interaction rules they use reveals that these are asymmetric and rather artificial. Actually, the probability for two neighbors with an opinion difference less than d to break their link is strictly 0 (see Kozma & Barrat, 2008). While this asymmetry prevents agents close enough in opinion to break their links and therefore prevents any disaggregation of clusters, it is not clear if the stationary states of their model remain robust when this rule is softened.

In this chapter, we build on a modified version of (Kozma & Barrat, 2008) to investigate this question and try to obtain structures that last from entities that do not last, as our friends sociologists request. We first show that by introducing a more natural, symmetric noise in the interaction rule, the rewiring process is not sufficient to avoid a collapse towards consensus in the long run. Then, we show how robust polarized states can be obtained by introducing random fluctuations in the agents' opinions. Ironically, this polarized state is now too robust to please sociologists. We therefore end this chapter by showing our work in progress on a last model which takes into account the "age" of the agents and leads to interesting "generation" phenomena and constantly changing polarized states. In this last model, opinion clusters can continuously grow by recruit-

ing new members, contract and disappear by losing members, merge with other clusters ... In other words, we obtain a dynamically stable state which is more appealing for sociologists and might stimulate their conceptualization of structure emergence.

8.2 An opinion model with adaptive network

8.2.1 Model

Our model is based on the Deffuant Model (Deffuant *et al.*, 2000) with an adaptive network similar to what can be found in Kozma & Barrat (2008). The model considers a population of N agents ($i = 1, 2, ...N$) which are characterized by a continuous opinion at time t $o(t) \in [0, 1]$. The initial interaction network is taken as an Erdös-Rényi network of average degree K, *ie* agents are initially linked to K neighbors on average and opinions are also chosen at random. Following the spirit of Kozma & Barrat (2008)'s model, we choose to introduce in our model a non-strict update rule ensuring a non-zero probability for neighbours of non-close opinions to communicate. Specifically, we define the probability of convergence between two agents i and j:

$$p_{conv}(i,\, j) = \frac{1}{1 + \exp\left(\omega\left(\Delta o - d\right)/d\right)} \tag{8.1}$$

where $\Delta o = |o_j(t) - o_i(t)|$ is difference of opinion between the agents, d is the usual Deffuant threshold, and ω is a "leakage" parameter. For $\omega \gg 1$, our update rule is very similar to Deffuant's. For $\omega \ll 1$, p_{conv} is finite and does not depend on Δo hence all pairs of neighbors interact in the same way. Convergence means that the agents' opinions are updated according to

$$\begin{aligned} o_i(t+1) &= o_i(t) + \mu(o_j(t) - o_i(t)) \\ o_j(t+1) &= o_j(t) + \mu(o_i(t) - o_j(t)) \end{aligned} \tag{8.2}$$

where $\mu \in [0,\, 1/2]$ is a convergence parameter. We will assume here that $\mu = 1/2$, meaning that both agents i and j adopt the same opinion once they have interact. We also introduce a possibility for link rewiring. To do so, we suppose that at each iteration, with a probability $p_{break} = 1 - p_{conv}$, the link between i and j is broken and a new link between agent i and a randomly chosen agent is created. In our simulation, we typically take $\omega = 5$, meaning that two neighbors sharing the same opinion will break their link with a probability ~ 0.01.

Finally, we introduce the fact that agents are "non lasting" entities. For this, after

the agent and one of its neighbors have been selected and the interaction/rewiring process has been carried out, with a probability ν we change randomly its opinion while keeping the same links. This can be interpreted as the "death" of the agent and the birth of another one, with different characteristics (different opinion) but belonging to the same group (same links). In other words, the agents' average lifetime is $1/\nu$.

To summarize, on each step we do the following:

- 1. Pick an agent i at random. If that agent has no neighbor, do nothing. Otherwise, pick one of its neighbors j and continue the process.
- 2. Generate a random number $r \in [0, 1]$.
- 3. If $r < p_{conv}(i, j)$, then the opinions of i and j converge according to

$$\begin{aligned} o_i(t+1) &= o_i(t) + (o_j(t) - o_i(t))/2 \\ o_j(t+1) &= o_j(t) + (o_i(t) - o_j(t))/2 \end{aligned}$$

- 4. Else (*ie* if $r \geq p_{conv}(i, j)$) update the link between i and j:

  Choose a randomly chosen agent k which is neither i nor i's neighbor

  Connect i to k and break the link between i and j.

- 5. With a probability ν, update agent i: o_i takes a new random value between 0 and 1.

Following Kozma & Barrat (2008), we track the evolution of the system by detecting the *opinion clusters* of agents. Defining *opinion neighbours* as two linked agents whose difference of opinions is smaller than the tolerance threshold d, an opinion cluster is defined as a set of agents such that between any two of these agents there exists a path of opinion neighbours. Compared to topological clusters which track the connected component of the network, opinion clusters tracks the set of agents between which a path of *communication* ($\Delta o < d$) exists.

8.3 Results

The detailed analysis of our model is still in progress. In this section, we only present some simulation outputs illustrating the different behaviors obtained by varying the

different parameters. In all the simulations presented here the opinions of the agents are initially randomly chosen between 0 and 1, the number of agents is fixed to $N = 1000$, the average degree to $K = 10$ and the tolerance threshold is fixed to $d = 0.04$.

8.3.1 Reference case : no noise, no death

In order to investigate the effects of each parameter, we start with the reference case, where $\nu = 0$ (*ie* without opinion noise or, in our interpretation of ν as a typical lifetime, with "immortal" agents), and with no leakage ($\omega \gg 1$). This last condition is fulfilled by using :

$$\begin{cases} p_{conv} = 0 \text{ when } \Delta o > d \\ p_{conv} = 1/2 \text{ when } \Delta o = d \\ p_{conv} = 1 \text{ when } \Delta o < d \end{cases}$$

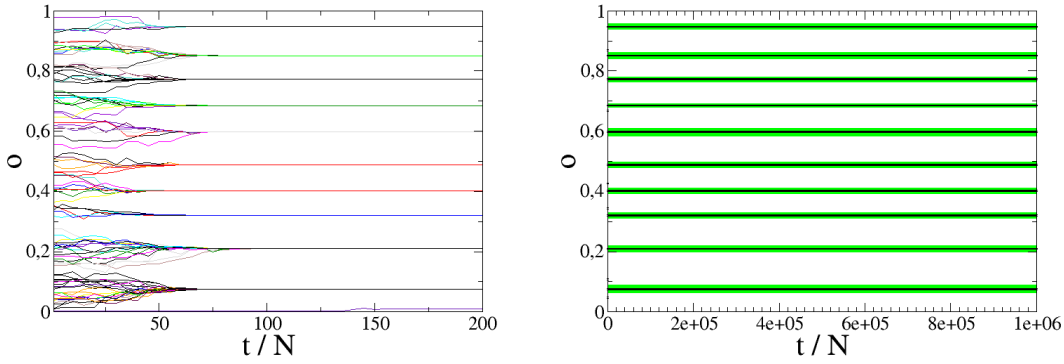

Figure 8.1: **Limit case $\omega \to \infty$: without leakage, convergence towards stable polarized states. Left**: Evolution of the opinions of a sample of 100 out of the $N = 1000$ agents. **Right:** Evolution of the mean opinion of detected opinion clusters with more than 10 agents. The width of the error bars (in green) is proportional to the number of agents within the opinion clusters. Other parameters are fixed to $K = 10$, $d = 0.04$ and $\nu = 0$.

Fig 8.1 displays the temporal evolution of the mean opinion inside the opinion clusters containing more than 10 agents and the temporal evolution of the opinion of a sample of agents. Opinion clusters are detected independently at each normalized iteration (corresponding to N iterations). We also keep track of the number of agents within each opinion cluster. On the left panel of Fig 8.1, we observe that, starting from random opinion, local convergence processes take place and lead to several opinion clusters. On

the right panel of Fig 8.1, we can check that the obtained opinion clusters are indeed stable: in the absence of leakage, an agent can only interact with and never break a link with another member of his opinion cluster.

These polarized states are the typical outcome obtained in the case of no opinion noise and no leakage (at least for small d and reasonable values of N and K), and correspond to the results obtained by Kozma & Barrat (2008).

8.3.2 Introducing opinion leakage

We now investigate the effect of the introduction of leakage. Do we still obtain several stable opinion groups when we introduce a small probability for neighbors with close opinions to break their link and neighbors with very different opinions to interact?

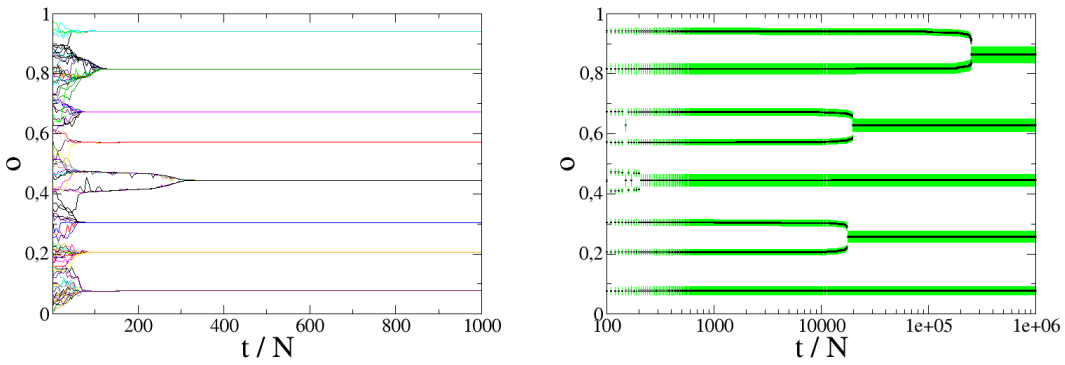

Figure 8.2: **Introducing leakage yield slow convergence towards consensus. Left**: Evolution of the opinions of a sample of 100 out of the $N = 1000$ agents. **Right:** Evolution of the mean opinion of detected opinion clusters with more than 10 agents. The width of the error bars (in green) is proportional to the number of agents within the opinion clusters. Other parameters are fixed to $K = 10$, $d = 0.04$, $\omega = 5$ and $\nu = 0$.

As in the reference case, the left panel of Fig 8.2 shows that on small time scales, local convergence processes lead to the formation of several opinion clusters. However, as emphasized by the right panel of Fig 8.2, on longer time scales, we observe coalescence of opinion clusters. This coalescence is due to the addition of several unlikely but possible steps allowed by the leakage mechanism. First, each agent within an opinion cluster can break a link with another member of the cluster and create a new link with a randomly chosen agent, possibly in another cluster distant in opinion space. This happens with probability $p_{break} = 1/(1 + \exp \omega(0 - d)/d) \simeq \exp(-\omega)$. While this probability is small

(7.10^{-3}) for typical ω values used here ($\omega = 5$), once multiplied by the number of links within a cluster ($\simeq 100 * 10 = 10^3$), it generates many links breaking, leading to the creation of many bonds between members of different clusters. Once a intercluster link is generated and picked again in a subsequent iteration, it may survive and lead to convergence with probability $p_{conv} = 1/(1 + \exp \omega(\Delta O - d)/d) \sim \exp(\omega/d(\Delta O - d))$, where ΔO is the opinion difference between the two clusters (and therefore the agents). Once this step is accomplished both agents can still be easily re-attracted in their original opinion clusters, but they can also attract new agents towards their new opinion. The second step will be easier since the first two agents have created a "bridge" right in the middle of their two original clusters.

This qualitative explanation is in agreement with Fig 8.2's right panel. The coalescence between two groups seems to accelerate once a first bridge has been built and the time needed for two groups to coalesce increases exponentially (Fig 8.3) with the initial difference of opinion between the two clusters considered. Therefore, in the limit of infinite time, all clusters coalesce into a single one, leading to a consensus state. This result, obtained with a symmetric opinion noise, contradicts Kozma & Barrat (2008) and suggests that even adaptive networks are not robust to noise or leakage.

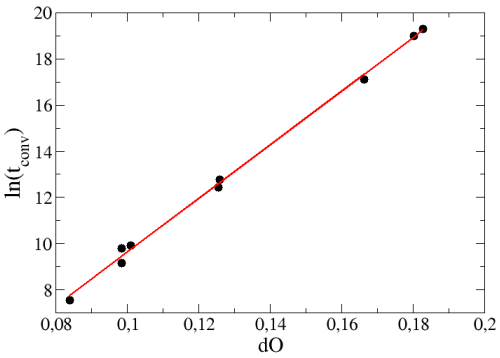

Figure 8.3: **Time for cluster - cluster coalescence as a function of their initial opinion difference**. Data is well fitted by t_{coal}= 0.136 exp($\mathcal{O}$ * 116.03), with $116 \simeq \omega/d = 125$. Estimating the time needed for coalescence into a single cluster (assuming a distance $\Delta O = 0.5$ for the two last clusters) is 10^{26}. Parameter values : $K = 10$, $d = 0.04$, $\omega = 5$ and $\nu = 0$.

8.3.3 Lasting Structures from Non Lasting Entities

We now introduce an additional feature to the model: the opinion noise or death/birth cycle of the agents. Obviously, the introduction of agents born with a random opinion tends to counteract the unavoidable convergence of opinions implied by the leakage. Two time scales are now relevant : a characteristic coalescence time t_{coal} (see above) and the characteristic lifetime $1/\nu$ of the agents. For $\nu t_{coal} \ll 1$, the coalescence process is preponderant and the system converges towards a consensus state with some fluctuations. For $\nu t_{coal} \gg 1$, noise wins and agents constantly change their opinions randomly, leading to a fragmentation state. The interesting question now becomes : is there an intermediary stage?

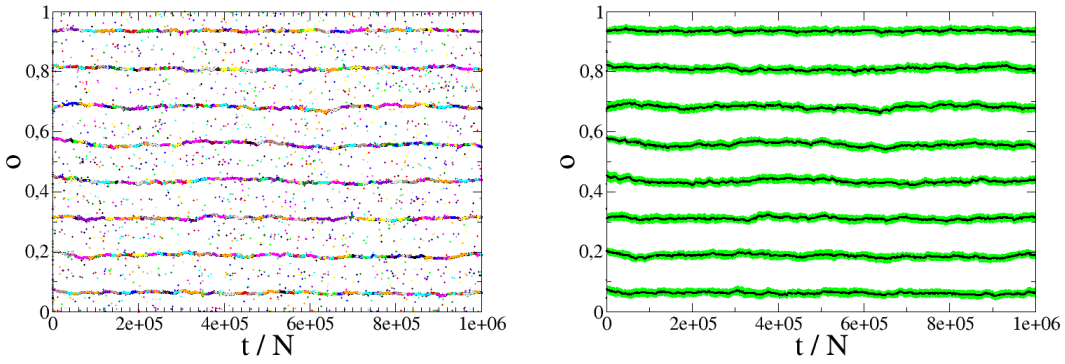

Figure 8.4: **Opinion noise combined to leakage yield slightly fluctuating polarized states. Left**: Evolution of the opinions of a sample of 100 out of the $N = 1000$ agents. **Right:** Evolution of the mean opinion of detected opinion clusters with more than 10 agents. The width of the error bars (in green) is proportional to the number of agents within the opinion clusters. Other parameters are fixed to $K = 10$, $d = 0.04$, $\omega = 5$ and $\nu = 10^{-4}$.

The simulation results presented in Fig 8.4 suggest that an intermediary stage does exist. We observe the existence of opinion clusters with fluctuating mean opinions. Clearly, the number of opinion clusters depend on d, but also on ν and ω (which controls t_{coal}). Preliminary analysis suggest that at short time scales, clusters randomly diffuse both in number of agents and average opinion (Figure 8.4). This brownian motion arises by death of the clusters' agents and capture of free agents that are born close in opinion space. At longer time scales however, this random drift is limited by the cluster-cluster correlations induced by the interactions of the capture zones of neighboring clusters in

opinion space. This competition for newborn agents leads to a strong stability of the total number of clusters, which last for times much longer than the agents' lifetime.

8.4 Addition of Generation Effects

At this point, we have managed to obtain long lasting structures (the opinion clusters) from perishable agents. Actually, from a sociological point of view, the structures might even last too long, since they never seem to decay. The reason of the stability of these structures is the simplicity of the opinion space, which leads to easy assimilation of newborns by existing opinion clusters. In other words, the one-dimensional opinion space is too homogeneous to allow for a complex collective behavior.

We have started to explore the effects of including a last ingredient to the opinion space, namely "generation" effects. Specifically, we keep the same model, but we replace in Eq. 8.1 the simple opinion distance Δo by a combination of opinion and age distance, where the age $a(t)$ of an agent is simply the number of normalized iterations it has lived since its birth. Eq. 8.1 then becomes :

$$p_{conv}(i,\, j) = \frac{1}{1 + \exp\left(\omega\left(\sqrt{|\Delta o|^2 + |\nu \Delta a|^2} - d\right)/d\right)} \tag{8.3}$$

The term $\nu\Delta a$ introduces heterogeneity in the opinion space : for example, two agents with similar opinions but widely different ages (measured in ν units) may easily break their link. A (pseudo)sociological rationale for this ingredient would be that different generations may see the opinion space differently. Fig 8.5 shows an example of the complex phenomena produced by generation effects. On short time scales (compared to $1/\nu$), opinion clusters still exist and are rather stable. On longer time scales, we observe that some opinion clusters disappear by progressive death of their agents, while others can recruit young agents and last for hundreds of agents' lifetimes. We also see merging of clusters close enough in opinion and growth of a new cluster in the space left free by this fusion. We are currently investigating this rich phenomenology, trying to understand what determines clusters' distinct evolutions, to quantify their lifetime, their evolution in opinion space (random walk?), how mergers can be predicted ...

8.5 Discussion

The work presented in this chapter focuses on simple models of "opinion" clustering, of the emergence and persistence of structures that can last longer than individuals that

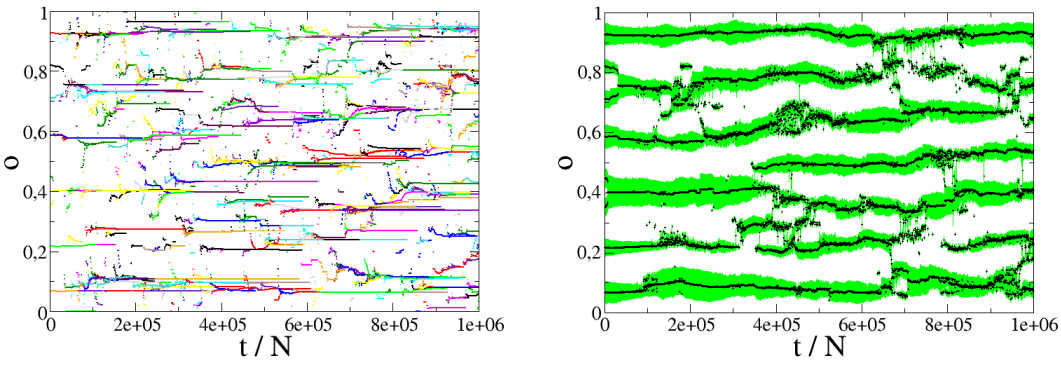

Figure 8.5: **Generation effects produce rich collective behaviors. Left**: Evolution of the opinions of a sample of 50 out of the $N = 500$ agents. **Right:** Evolution of the mean opinion of detected opinion clusters with more than 10 agents. The width of the error bars (in green) is proportional to the number of agents within the opinion clusters. Other parameters are fixed to $K = 10$, $d = 0.04$, $\omega = 5$ and $\nu = 10^{-5}$.

give rise to them. We have developed a series of models that lead to robust polarized states and dynamical equilibria. The last model reminds us of a classic sociology text, Georg Simmel's study of the persistence of social groups (Simmel, 1898) : *it is meaningful to speak of group identity, despite shifting membership and low institutionalization, if there is some membership continuity in contiguous stages [...] The change, the disappearance and entrance of persons, affects in two contiguous moments a number relatively small compared with the number of those who remain constant. The departure of the older and the entrance of the younger elements proceed so gradually and continuously that the group seems as much like a unified self as an organic body in spite of the change of its atoms.*

Despite this parallel, we do not think that this kind of simplistic model should aim at being realistic by inclusion of additional ingredients. Rather, our discussions with Sciences Po's sociologists open another direction. We speculate that these models can help them enriching their conceptualizations of the structuration phenomenon. This could arise from a detailed examination of the structuration that happen in this virtual society and their tentative interpretation using the usual sociological theories. Does this confrontation help them in any way in renewing their conceptual repertoire? Do we observe group evolutions that are unexpected and difficult to explain? The future of this stimulating and demanding collaboration will tell!

Conclusion

Recapitulation

The original motivation for this thesis was to explore social phenomena using tools derived from statistical physics, and more generally quantitative tools. Most of our work was concentrated on three different projects which all share a central question, that of the relation between individual and collective, aggregated entities.

The first part of the thesis focused on Schelling's segregation model. We believe that our work has strongly contributed to the development of a deeper understanding of the mechanisms at stake in this paradigmatic model. We hence investigated in Chapter 1 the robustness of unwanted global segregation patterns in respect with the introduction of coordination between the agents' moving decisions. Our simulations of different forms of taxation or induced collaboration between the agents showed that even a small amount of coordination can significantly reduce unwanted segregation.

We then presented in Chapter 2 a framework for Schelling's model which allowed us to derived general and unprecedented analytical results. Specifically, we were able to characterize the global configurations of a virtual city by a potential function which maximizes the stationary states - at least for a given range of the individual utility functions. We showed in particular that in the context of continuous neighborhood, the only solvable case was the one already presented by Zhang (2004a), *ie* the case of linear utility functions. We hence put forward the concept of bounded neighborhood - an alternative way to define the network of interactions of the agents in the city. In that context, we showed that a potential function exists for a much broader range of utility function. We use this potential function to analyze the outcome of the model

for three families of utility functions, namely linear utility functions, Schelling's original stair-like utility functions and asymmetrically peaked utility functions. In all cases, the potential function allowed us to derive the segregation level and the corresponding cost in the collective utility.

In chapter 3, we used a simplified version of our analytical model - with only one type of agent - to come back to the question of the level of coordination necessary to reduce unwanted segregation patterns. By introducing a 'cooperativity' parameter α allowing to tune continuously between strictly individual-driven and strictly collective-driven dynamics, we characterized a qualitative transition from a segregated phase of low collective utility towards a mixed phase of high collective utility. This study was performed from a physicist's point of view, in which we identify the generalized potential function - encompassing both individual and collective dynamics - as an effective free energy.

In chapter 4, we presented our analytical approach in a more general framework, showing how it could apply to a wider range of socio-economic models.

In the second part of the thesis, we explored huge databases on scientific literature (mostly Web of Science) to investigate the existence and evolution of paradigms or scientific institutions.

Chapter 5 hence tackled the question of the existence and coherence of "complex system science". Starting from a large database (141 098 records) of relevant articles published between 2000 and 2008, we measured a scientific similarity between articles thanks to the bibliographic coupling approach - based on shared references - and we detected the field communities emerging from these interactions. These 'natural communities' presented a far richer structure than the usual partition of science in fixed disciplines or institutions. Our analysis of this structure revealed that the overall coherence of the 'complex systems' field does not arise from a universal theory but rather from computational techniques and fruitful adaptations of the idea of self organization to specific systems. We also put forward the existence of 'trading zones', *ie* small communities creating an interface between disciplines around specific tools or concepts. The place of complex systems science within the whole landscape of science was also investigated in two preliminary maps representing French 2000 and 2010 natural science communities. While *Self-Organization*, *Self-Organized Criticality* and *Complex Network* do constitute coherent subfields in those maps, no unified conceptual kernel for "complexity science" is identified.

In Chapter 6, we presented a set of routines allowing to draw different maps of the

research carried out in a scientific institution. We took the example of our institution, *ENS de Lyon*, to show how different maps, namely co-occurrence (of authors ,keywords, institutions) maps and bibliographic coupling communities maps, can be built and what kind of information they can provide. We emphasized the fact that these different maps, based on the same elements, offered different views on the institutions at the global scale.

Finally, in the third part we presented the work that resulted from our collaboration with a team of sociologists from the MediaLab at Sciences Po.

Chapter 7 hence questioned the assumption of a clear dichotomy between two 'levels', namely individuals and society. Building on the social theory developed by Gabriel Tarde at the end of the 19th century, we explored different possibilities to visualize (and therefore conceptualize) the evolution of social phenomena without making a distinction between two levels. Bibliometric data were used as an example. We also proposed an attempt to formalize Tarde's theory in the scope of an algorithmic model. While our prototype model fulfilled several of Tarde's precept, it raised many more questions.

In Chapter 8, we focused on a single question, namely the existence of lasting structure from non lasting entities. We built on the - more standard - physicist approach developed in opinion model. While most papers focus on the stationary properties of the model, we chose to build a model to investigate the dynamical properties of social structures which are always changing. The key ingredients of the model we presented are the use of adaptive network, the introduction of noise in the agents' interactions, a turnover in the population of agents and the introduction of generation effect, the agents taking into account their opinion *and* age difference in their interaction. The outcomes of our model display a rich phenomenology of group dynamics.

Approach

The research presented in this thesis was performed in a strong interdisciplinary spirit.

Indeed, for each research project we took great care to tackle the questions at the heart of these projects in a mutually beneficial way for both social and natural sciences. Hence, our work on economics's aspects of Schelling model in chapters 1 an 2 is counterbalanced by a the presentation of aspects of this resolution that may be of interest for a physicist in chapters 3 and 4. In the same way, our discussion on the nature of social entities of the chapter 7 is counterbalanced by chapter 8.

If this balance is somehow necessary to ensure the permanence of a collaboration between researchers from different disciplines, it is my firm belief that these interdisci-

plinary collaborations's worth comes from the development of new ways of thinking and of conceptualizing the objects of one's discipline.

Perspectives

Several extensions of the work presented in this thesis could be explored.

As stated in chapter 4, we believe that the concept of 'link' potential function could be applied to other systems. Lemoy *et al* (2011) already used it in a typical model of urban economics dealing with land use and transport. In particular, they extended our comparison of the potential function with an effective free energy to a comparison with an effective chemical potential. A second interesting extension of the work presented in chapter 4 would be a deeper analysis of the individual / collective transition. Going back to simulations and continuous neighborhood, this transition could be analyzed in a more standard physics framework. What is the nature of this transition, what are the critical exponents characterizing it, how does it relate to the usual Ising transition? All these questions are certainly worth to investigate.

The conclusions we drawn in chapter 5 on the nature of complex systems science are based on a particular database issued from the Web of Science, and mostly on one technique: bibliographic coupling. It would hence be interesting to check the robustness of our conclusions with other databases selected with different criteria than ours, or by using other techniques for defining a similarity between articles. A possibility could be to performed co-word analysis, *ie* defining a similarity based on the nature and number of words - for example the word present in the titles and abstracts - shared by the articles.

We hope that the toolbox we presented in chapter 6 will lead other scientists to build maps of their own institutions, thus fostering ongoing dialogue and praxis in the institution. The policy issues raised by the utilisation of bibliometrics indices and such 'maps of sciences' are certainly not going to disappear and that is why we think it is important that these tools may be known and accessible to everyone. A collaboration with other institutions (such as CNRS and CEMAGREF) which are interested in such global maps have already begun. Future improvement of our toolbox will include the possibility to prepare different maps for successive time periods, in order to grasp the evolution of the institution. It might also be interesting to couple the bibliometric data that can be obtained on the Web of Science with data accounting for the other activities of the researchers of an institution (teaching, patents, ...).

Finally, our collaboration with sociologists is still on its way: the main perspective of the third part of this thesis is to continue to develop and explore the kind of models we presented. Is it possible to build a solid bridge between physicists' models starting from simple agents and Tarde's theories demanding to start from complex monads?

Bibliography

Agarwal P, Skupin A, 2008. Self-organising maps: applications in geographic information science, Ed. John Wiley & Sons Inc.

Albert R, Barabási AL, 2002. Statistical mechanics of complex networks. *Rev Mod Phys* 74:47-97.

Anderson PW, 1972. More is Different. *Science* 177:393-396.

Anderson SP, De Palma A, Thisse, JF, 1992. *Discrete Choice Theory of Product Differentiation* (MIT Press).

Auerbach A, 1985. The theory of excess burden and optimal taxation. *Handbook Publ Econ* 1:61-127.

Axelrod R, 1997. The Complexity of Cooperation: Agent-Based Models of Competition and Collaboration. *Princeton: Princeton University Press.*

Axelrod R, 1997. The dissemination of culture: A model with local convergence and global polarization. *Journal of conflict resolution* 41:203-226.

Bak P, Tang C, Wiesenfeld K, 1987. Self-organized criticality: an explanation of 1/f noise. *Phys Rev Lett* 59:381-384.

Bak P, Tang C, Wiesenfeld K, 1988. Self-organized criticality. *Phys Rev A* 38:364-374.

Barabási AL, Albert R, 1999. Emergence of Scaling in Random Networks. *Science* 286:509-512.

Barabási AL, Oltvai Z, 2004. Network biology: understanding the cell's functional organization. *Nature Rev Genet* 5:101-113.

Barr J, Tassier T, 2008. Segregation and strategic neighborhood interaction. *Eastern Economic Journal* 34:480-503.

Barry A, Thrift N, 2007. Gabriel Tarde: Imitation, invention and economics. *Economy and Society* 36(4):509-525.

Bastian M, Heymann S, Jacomy M, 2009. Gephi: An Open Source Software for Exploring and Manipulating Networks. *Proc 3rd Intl ICWSM Conf.*

Von Bertalanffy L, 1976. General System Theory: Foundations, Development, Applications (George Braziller).

Blondel VD, Guillaume J-L, Lambiotte R, Lefebvre E, 2008. Fast unfolding of communities in large networks. *J Stat Mech* P10008.

Blume LE, 1993. The Statistical Mechanics of Best-Response Strategy Revision. *Games and Economic Behavior* 11:111-145.

Borgatti SP, Mehra A, Brass DJ, Labianca G, 2009. Network Analysis in the Social Sciences. *Science* 323:892-895.

Börner K, Chen Ch, Boyack K, 2003. Visualizing Knowledge Domains. *Annual Review of Information Science and Technology* 37:179-255.

Börner K, Scharnhorst A, 2009. Visual conceptualizations and models of science. *Journal of Informetrics* 3:161-172.

Bouchaud JP, 2008. Economics needs a scientific revolution. *Nature* 455:1181.

Bowker G, 1993. How to Be Universal: Some Cybernetic Strategies, 1943-70. *Social Studies of Science* 23:107-127.

Bruch EE, Mare RD, 2006. Neighborhood Choice and Neighborhood Change. *American Journal of Sociology*, 112(3):667-709.

Callen H, 1985. *Thermodynamics and an introduction to thermostatistics*, eds Wiley J and sons (New York).

Card D, Mas A, Rothstein J, 2008. Tipping and the dynamics of segregation. *The Quarterly Journal of Economics*, 123:177-218.

Carrington WJ, Troske KR, 1997. On measuring segregation in samples with small units. *Journal of Business & Economic Statistics*, 402-409.

Castellano C, Fortunato S, Loreto V, 2009. Statistical physics of social dynamics. *Rev Mod Phys* 81(2):591-646.

Changeux JP, 2009. *The Physiology of Truth: Neuroscience and Human Knowledge* (The Belknap Press).

Charlot S, Gaigné C, Robert-Nicoud F, Thisse J-F, 2006. Agglomeration and welfare: The core-periphery model in the light of Bentham, Kaldor, and Rawls. *Journal of Public Economics* 90(1-2):325-347.

Chavalarias D, Cointet J-P, 2008. Bottom-up scientific field detection for dynamical and hierarchical science mapping, methodology and case study. *Scientometrics* 75(1):37-50.

Chavalarias D, Cointet J-P, 2009. The Reconstruction of Science Phylogeny, *arXiv* 0904.3154v3.

Cho A, 2009. Ourselves and our interactions : the ultimate physics problem? *Science* 325:406-408.

Chouard T, 2008. Beneath the surface. *Nature* 456:300-303.

Chu X, 1995. Endogenous trip scheduling: the Henderson approach reformulated and compared with the Vickrey approach. *J Urban Econ* 37:324-343.

Clark WAV, Fossett M, 2008. Understanding the social context of the Schelling segregation model. *Proceedings of the National Academy of Sciences* 105:4109-4114.

Cotterill R, 2008. *The Cambridge Guide to the Material World* (Cambridge University Press).

Clark W, Fossett M, 2008. Understanding the social context of the Schelling segregation model. *Proc Natl Acad Sci USA* 105:4109-4114.

Cutler DM, Glaeser EL, Vigdor JL, 2008. Is the melting pot still hot? Explaining the resurgence of immigrant segregation. *Review of economics and statistics* 90:478-497.

Dall'Asta L, Castellano C, Marsili M, 2008. Statistical physics of the Schelling model of segregation. *Journal of statistical mechanics - theory and experiment*, L07002

Damasio AR, 1995. *The Feeling of What Happens* (Harcourt Brace and Company, New York).

Dawkins R, 1976. The Selfish Gene, New York: Oxford University Press.

Deffuant G, Neau D, Amblard F, Weisbuch G, 2000. Mixing beliefs among interacting agents. *Advances in Complex Systems* 3(1-4):87-98.

Dokumaci E, Sandholm WH, 2007. Schelling Redux: An Evolutionary Dynamic Model of Residential Segregation. *Unpublished manuscript*, University of Wisconsin-Madison.

Duncan OD, Duncan B, 1955. A methodological analysis of segregation indexes. *American Sociological Review* 20:210-217.

Edmonds B, Moss S, 2005. From KISS to KIDS - an 'anti-simplistic' modelling approach. *Multi-Agent and Multi-Agent-Based Simulation* 130-144.

Eisen MB, Spellman PT, Brown PO, Botstein D, 1998. Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci USA* 95:14863-14868.

Epstein JM, Axtell R, 1996. Growing Artificial Societies: Social Science From the Bottom Up. *MIT Press/Brookings Institution.*

Evans M, Hanney T, 2005. Nonequilibrium statistical mechanics of the zero-range process and related models. *J Phys Math Gen* 38:R195-R240.

Fagiolo G, Valente M, Vriend NJ, 2007. Segregation in networks. *Journal of Economic Behavior and Organization* 64:316-336.

Farley R, Schuman H, Bianchi S, Colasanto D, Hatchett S, 1978. Chocolate city, vanilla suburb: will the trend toward racially separate communities continue? *Social Science Research* 7:319-344.

Farley R, Fielding EL, Krysan M, 1997. The residential preferences of balcks and whites: a four-metropolis analysis. *Housing Policy Debate* 8:763-800.

Fortunato S, Barthélemy M, 2007. Resolution limit in community detection. *Proc Natl Acad Sci USA* 104:36-41.

Fortunato S, 2010. Community detection in graphs. *Physics Reports* 486:75-174;

Fox-Keller E, 2005. Revisiting "scale-free" networks. *BioEssays* 27:1060-1068.

Fox-Keller E, 2009. Organisms, Machines, and Thunderstorms, a history of Self-Organization. *Historical studies in the natural sciences* 38-39.

Galison P, 1997. Image and Logic: A Material Culture of Microphysics (University Of Chicago Press).

Gannon F, 2007. Too complex to comprehend? *EMBO Rep* 8:705.

Gauvin L, Vannimenus J, Nadal JP, 2009. Phase diagram of a Schelling segregation model. *Eur Phys J B* 70(2):293-304.

Gilbert N, Pyka A, Ahrweiler P, 2001. Innovation Networks - A Simulation Approach *Journal of Artificial Societies and Social Simulation* 4(3).

Gillespie DT, 1977. Exact Stochastic Simulation of Coupled Chemical Reactions. *J Phys Chem* 81(25):2340-2361.

Girvan M, Newman MEJ, 2004. Finding and evaluating community structure in networks. *Phys Rev E* 69:026113.

Glass L, 2001. Synchronization and rhythmic processes in physiology. *Nature* 410:277-284.

Goodstein D, 1985. *States of Matter* (Dover Publications).

Grauwin S, Bertin E, Lemoy R, Jensen P, 2009a. Competition between collective and individual dynamics. *Proc Natl Acad Sci USA* 106:20622-20626.

Grauwin S, Goffette-Nagot F, Jensen P, 2009b. Dynamic models of residential segregation: brief review, analytical resolution and study of the introduction of coordination. *GATE Working Paper* (available at ftp://ftp.gate.cnrs.fr/RePEc/2009/0914.pdf).

Grauwin S, Goffette-Nagot F, Jensen P, 2011. Dynamic models of residential segregation: an analytical solution. Revised & Resubmitted to J Pub Econ

Grauwin S et al, 2011. Complex systems science: dreams of universality, reality of interdisciplinarity. Submitted.

Hart S, Mas-Colell A, 1989. Potential, value and consistency. *Econometrica* 57:589-614.

Hayden EC, 2010. Life is complicated. *Nature* 464:664-667.

Holland, JH, 2000. Building Blocks, Cohort Genetic Algorithms and Hyperplane-Defined Functions. *Evolutionary Computation* 8(4):373-391.

Holme P, Newman MEJ, 2006. Nonequilibrium phase transition in the coevolution of networks and opinions. *Physical Review E* 74(5):56108.

Iceland J, Scopilliti M, 2008. Immigrant residential segregation in US metropolitan areas. 1990–2000. *Demography* 45:79-94.

Iñiguez G, Kertész J, Kaski KK, Barrio RA, 2009. Opinion and community formation in co-evolving networks. *Physical Review E* 80:066119

Kauffman S, 1993. The Origins of Order: Self-Organization and Selection in Evolution. Oxford University Press.

Kessler MM, 1963. Bibliographic coupling between scientific papers. *American Documentation* 24:123-131.

Kirman AP, 1992. Whom or What Does the Representative Individual Represent? *J Econ Perspect* 6:117-36.

Klavans R, Boyack KW, 2009. Toward a consensus map of science, *J Am Soc Inform Sci Tech* 60(3):455-476.

Klemm K, Eguíluz VM, Toral R, San Miguel M, 2003. Nonequilibrium transitions in complex networks: A model of social interaction. *Physical Review E* 67:026120.

Knibbe C, Coulon A, Mazet 0, Fayard JM, Beslon G, 2007. A long-term Evolutionary Pressure on the Amount of Non-coding DNA. *Molecular Biology and Evolution* 24(10):2344-2353.

Kozma B, Barrat A, 2008. Consensus formation on adaptive networks. *Physical Review E* 77(1):16102.

Lancichinetti A, Kivela, M, Saramaki J, Fortunato S, 2010. Characterizing the Community Structure of Complex Networks. *PLoS ONE* 5(8):e11976.

Latour B, 2005. Reassembling the Social - An Introduction to Actor-Network-Theory *Oxford University Press.*

Latour B, 2007. *Reassembling the Social: An Introduction to Actor-Network-Theory* (Oxford University Press).

Latour B, 2010. Tarde's idea of quantification. *in: Candea M (2010) The Social After Gabriel Tarde: Debates and Assessments, London, Routledge.*

Lazer D *et al* (2009) Computational social science. *Science* 323:721-723.

Lemoy R, Bertin E, Jensen P, 2011. Socio-economic utility and chemical potential *Europhysics Letters* 93(3):38002.

Leydesdorff L, Rafols I, 2009. A Global Map of Science Based on the ISI Subject Categories, *J Am Soc Inform Sci Tech* 60(2).

Mäs M, Flache A, Helbing D (2010. Individualization as Driving Force of Clustering Phenomena in Humans. *PLoS - Comput Biol* 6(10):e1000959.

Mas-Colell A, Whinston MD, Green JR, 1995. *Microeconomic Theory* (Oxford University Press, New York).

Massey DS, Denton NA, 1988. Dimensions of Residential Segregation. *Social Forces* 6:281.

McFadden D, 1974. Conditional Logit Analysis of Qualitative Choice Behavior. *Frontiers in Econometrics* 76:689-708.

Towards 2020 Science (Microsoft, 2006) `http://research.microsoft.com/towards2020science`

Mobius MM, Rosenblat TS, 2000. The Formation of Ghettos as a Local Interaction Phenomenon. *Unpublished manuscript*, Harvard University.

Mogoutov A, Kahane B, 2007. Data search strategy for science and technology emergence: A scalable and evolutionary query for nanotechnology tracking. *Research Policy* 36:893-903.

Mogoutov A, Cambrosio A, Keating P, Mustar P, 2008. Biomedical innovation at the laboratory, clinical and commercial interface: A new method for mapping research projects, publications and patents in the field of microarrays. *Journal of Informetrics* 2(4):341-353.

Monderer D, Shapley LS, 1996. Potential games. *Games and Economic Behavior* 14:124-143.

Morrill RL, 1991. On the Measure of Spatial Segregation. *Geography Research Forum* 11:25-36.

Moussaid M, Garnier S, Theraulaz G, Helbing D, 2009. Collective Information Processing and Pattern Formation in Swarms, Flocks, and Crowds. *Topics in Cognitive Science* 1:469-497.

Nadal JP, Gordon MB, 2005. Physique statistique de phénomènes collectifs en sciences économiques et sociales, *Mathématiques et sciences humaines* 172:67-89.

Newman MEJ, 2003. The structure and function of complex networks. *SIAM Review* 45:167-256.

Nicolis G, Prigogine I, 1977. Self-organization in nonequilibrium systems (Wiley New York)

Nietzsche F, On the genealogy of morals, New York, Vintage Books, 1969. *cited by Flyvbjerg B, Making social science matter, Cambridge University Press, 2001.*

O'Sullivan A, 2009. Schelling's model revisited: Residential sorting with competitive bidding for land. *Regional Science and Urban Economics* 39:397-408.

Otsubo H, Rapoport A, 2008. Vickrey's model of traffic congestion discretized. *Transportation Research Part B* 42:873-889.

Pancs R, Vriend NJ, 2007. Schelling's spatial proximity model of segregation revisited. *Journal of Public Economics* 91:1–24.

Pickering A, 1995. The mangle of practice (University Of Chicago Press).

Pollicott M, Weiss H, 2001. The Dynamics of Schelling-type segregation models and a nonlinear graph laplacian variational problem. *Advances in Applied Mathematics* 27:17-40.

Press WH et al. (2007, 3rd edition) Numerical Recipes. The Art of Scientific Computing (Cambridge University Press, NY)

Rafols I, Leydesdorff L, 2010. Science Overlay Maps: A NewTool for Research Policy and Library Management *J Am Soc Inform Sci Tech* 61(9):1871-1887.

Reardon SF, Firebaugh G, 2002. Measures of Multigroup Segregation. *Sociological Methodology*, 32(1):33-67.

Reardon SF, O'Sullivan D, 2004. Measures of spatial segregation. *Sociological Methodology* 34:121-162.

Reardon SF, Matthews SA, O'Sullivan D, Lee BA, Firebaugh G, Farrell CR, Bischoff K, 2008. The geographic scale of metropolitan racial segregation. *Demography* 45:489-514.

Roessner D, 2000. Quantitative and qualitative methods and measures in the evaluation of research. *Research Evaluation* 9(2):125-132.

The web site of the Santa Fe complex systems institute (`http://www.santafe.edu/about/`, accessed June 1st, 2010) defines its aim as promoting "multidisciplinary collaborations in the physical, biological, computational, and social sciences."

Schelling TC, 1969. Models of Segregation. *American Economic Review, Papers and Proceedings* 59:488-493.

Schelling TC, 1971. Dynamic models of segregation. *Journal of Mathematical Sociology* 1:143-186.

Schelling TC, 1978. *Micromotives and Macrobehavior*. Norton, New York.

Science, 2009. *Special issue on the modeling of complex systems with networks* 325.

Scientific American special issue titled "From Complexity to Perplexity" (June 1995), in which many scientists criticize the high "mouth to brain ratio" common in the field.

Shapley LS, 1953. A Value for n-Person Games. in: *Contributions to the Theory of Games II*, ed. by Kuhn H. W., TuckerA. W. Princeton University Press, Princeton. 307-317.

Simmel G, 1898. The Persistence of Social Groups. *The American Journal of Sociology* 3(5)662-698.

Singh A, Vainchtein H, Weiss D, 2009. Schelling's Segregation Model: Parameters, Scaling and Aggregation. *Demographic Research* 21(12):341-366.

Small H, 1973. Co-citation in the scientific literature: A new measure of the relationship between two documents, *J Am Soc Inform Sci* 24:265-269.

Small H, 1999. Visualizing Science by Citation Mapping, *J Am Soc Inform Sci* 50(9):799-813.

Smith A, 1776. *An Inquiry into the Nature and Causes of the Wealth of Nations*, eds Strahan W, Cadell T (London).

Solé RV, 2000. *quoted in the New York Times*, December 26, First Cells, Then Species, Now the Web (`http://www.nytimes.com/2000/12/26/science/26WEBS.html?printpage=yes?pagewanted=1`, last accessed Sept. 27th, 2010).

Stirling A, 2008 "Opening up" And "Closing down": Power, participation, and pluralism in the social appraisal of technology. *Science, Technology & Human Values* 33(2):262-294.

Tarde G, 1890. Les lois de l'imitation.

Tarde G, 1893. Monadologie et sociologie.

Tamayo P et al, 1999. Interpreting patterns of gene expression with self-organizing maps: Methods and application to hematopoietic differentiation. *Proc Natl Acad Sci USA* 96:2907-2912.

Thurstone LL, 1928. Attitudes Can Be Measured. *American Journal of Sociology* 33(4):529.

Tiebout CM, 1958. A pure theory of local expenditures. *The Journal of Political Economy* 64:416.

Tyson JJ, Chen KC, Novak B, 2003. Sniffers, buzzers, toggles and blinkers: dynamics of regulatory and signaling pathways in the cell. *Curr Opin Cell Biol* 15:221-231.

Ui T, 2000. A Shapley value representation of potential games. *Games and Economic Behavior* 31:121-135.

Vinkovic D, Kirman A, 2006. A physical analogue of the Schelling model. *Proceedings of the National Academy of Sciences* 103:19261-19265.

Walliser B, 2005. les fonctions des modèles économiques *available at www.centrecournot.org/pdf/Walliser.pdf*

Watts DJ, Strogatz SH, 1998. Collective dynamics of small-world networks. *Nature* 398:440-442.

Wong DWS, 2005. Formulating a general spatial segregation measure. *The Professional Geographer* 57:285-294.

Wood D, Fels J, 2008. The Natures of Maps: Cartographic Constructions of the Natural World, Univ Chicago Press.

Web of Science, `http://apps.isiknowledge.com/`

Young HP, 1993. The evolution of conventions, *Econometrica* 61:57-84.

Young HP, 1998. *Individual Strategy and Social Structure: An Evolutionary Theory of Institutions*. Princeton University Press, Princeton.

Zhang J, 2004a. A dynamic model of residential segregation. *The Journal of Mathematical Sociology* 28:147-170.

Zhang J, 2004b. Residential segregation in an all-integrationist world. *Journal of Economic Behavior and Organization* 54:533-550.

Zhang J, 2010. Tipping and residential segregation: a unified Schelling model. *Journal of Regional Science*, forthcoming.

Appendix A

Annex to Chapter 2

A.1 Proof of Proposition 1

In this section, we place ourselves in a bounded neighborhood description.

Let us first prove the first part of 1, that is that any aggregate function $\mathcal{F} = \sum_{q\in\mathcal{Q}} F(R_q, G_q)$ is a potential function that corresponds to (at least) one pair of utility functions (u_R, u_G) of $\mathbb{U}$.

Suppose that $\mathcal{F} = \sum_{q\in\mathcal{Q}} F(R_q, G_q)$ is a potential function of the game, where the intermediate function F is known. Let us assume that an agent is moving from a block 1, characterized by the numbers $(R_1, G_1) \in E_{H+1}$ of red and green agents who live in it, to a block 2 characterized similarly by the numbers $(R_2, G_2) \in E_H$ of red and green agents living in it (since there must be at least one vacant location in block 2 for an agent to move in it, we necessarily have $R_2 + G_2 < H + 1$). By definition, the utility variation of a moving agent must be equal to the variation of $\mathcal{F}$ it induces. Hence :

- to cover the cases when the moving agent is a red one: for all $(R_1, G_1) \in E_{H+1}$ with $R_1 \geq 1$,

$$u_R(R_2, G_2) \quad - \quad u_R(R_1 - 1, G_1) = \\ F(R_2 + 1, G_2) + F(R_1 - 1, G_1) - F(R_2, G_2) - F(R_1, G_1) \quad \text{(A.1)}$$

- to cover the cases when the moving agent is a green one: for all $(R_1, G_1) \in E_{H+1}$

with $G_1 \geq 1$,

$$u_G(R_2, G_2) - u_G(R_1, G_1 - 1) = F(R_2, G_2 + 1) + F(R_1, G_1 - 1) - F(R_2, G_2) - F(R_1, G_1) \quad \text{(A.2)}$$

Taking $R_2 = G_2 = 0$ in equations A.1 and A.2, one finds that the utility functions u_R and u_G verify for all $(R, G) \in E_H$:

$$u_R(R, G) - u_R(0, 0) = F(R + 1, G) - F(R, G) - F(1, 0) + F(0, 0) \quad \text{(A.3)}$$
$$u_G(R, G) - u_G(0, 0) = F(R, G + 1) - F(R, G) - F(0, 1) + F(0, 0) \quad \text{(A.4)}$$

These relations define (up to a constant $u(0, 0)$) the utility functions the agents necessarily have if $\mathcal{F} = \sum_{q \in \mathcal{Q}} F(R_q, G_q)$ is a potential function of the game. It still remains to prove that this pair of utility functions belongs to the set $\mathbb{U}$. According to relations A.3 and A.4, one has for all $(R, G) \in E_H$:

$$\begin{aligned} u_R(R, G) - u_R(R, G + 1) &= \big(F(R + 1, G) - F(R, G)\big) - \big(F(R + 1, G + 1) - F(R, G + 1)\big) \\ &= \big(F(R, G + 1) - F(R, G)\big) - \big(F(R + 1, G + 1) - F(R + 1, G)\big) \\ &= u_G(R, G) - u_G(R + 1, G) \end{aligned}$$

Hence relation 2.15 holds, which means by definition that the pairs of utility functions (u_R, u_G) defined by relations A.3 and A.4 belongs to $\mathbb{U}$. Notice that in our demonstration no particular constraint has to be assumed on the form of function F. As a consequence, any aggregate function $\mathcal{F} = \sum_{q \in \mathcal{Q}} F(R_q, G_q) \in \mathbb{F}$ is a potential function of the game as soon as the pair of agents' utility functions is chosen so that relations A.3 and A.4 hold.

Let us now prove the second part of proposition 1, which is that to any pair of utility functions (u_R, u_G) of $\mathbb{U}$ corresponds a potential function of the form $\mathcal{F} = \sum_{q \in \mathcal{Q}} F(R_q, G_q)$.

Let $(u_R, u_G) \in \mathbb{U}$ be a pair of utility functions verifying condition 2.15. Suppose that $F(0, 0)$, $F(0, 1)$ and $F(1, 0)$ are given and let us define recursively the function F on E_{H+1} by the following equations, verified for all $(R, G) \in E_H$:

$$F(R + 1, G) - F(R, G) = F(1, 0) - F(0, 0) + u_R(R, G) - u_R(0, 0) \quad \text{(A.5)}$$
$$F(R, G + 1) - F(R, G) = F(0, 1) + F(0, 0) + u_G(R, G) - u_G(0, 0) \quad \text{(A.6)}$$

The most important thing to notice is that these two relations are consistent with each other thanks to condition 2.15 that links the two utility functions u_R and u_G. By summing Eq. A.5 on R and then Eq.A.6 on G, one finds the following expression for function F:

$$F(R,G) - F(0,0) = R\Big(F(1,0) - F(0,0)\Big) + \sum_{r=1}^{R}\Big(u_R(r-1,0) - u_R(0,0)\Big) \\ +G\Big(F(0,1) - F(0,0)\Big) + \sum_{g=1}^{G}\Big(u_G(R,g-1) - u_G(0,0)\Big) \quad \text{(A.7)}$$

or conversely by summing Eq. A.6 on G then Eq.A.5 on R,

$$F(R,G) - F(0,0) = R\Big(F(1,0) - F(0,0)\Big) + \sum_{r=1}^{R}\Big(u_R(r-1,G) - u_R(0,0)\Big) \\ +G\Big(F(0,1) - F(0,0)\Big) + \sum_{g=1}^{G}\Big(u_G(0,g-1) - u_G(0,0)\Big) \quad \text{(A.8)}$$

Hence, since $\mathcal{F} = \sum_{q\in\mathcal{Q}} F(R_q, G_q)$ one can obtain, after rearranging the different terms, a symmetric expression of the potential:

$$\begin{aligned}\mathcal{F} &= |\mathcal{Q}|F(0,0) + N_R\big(F(1,0) - F(0,0) - u_R(0,0)\big) + N_G\big(F(0,1) - F(0,0) - u_G(0,0)\big) \\ &\quad +\frac{1}{2}\Big[\sum_{r=1}^{R}\Big(u_R(r-1,0) + u_R(r-1,G)\Big) + \sum_{g=1}^{G}\Big(u_G(0,g-1) + u_G(R,g-1)\Big)\Big] \quad \text{(A.9)}\end{aligned}$$

Since the potential can be chosen up to a constant, it is clear from the previous expression that the choice of $F(0,0)$, $F(0,1)$, $F(1,0)$, $u_R(0,0)$ and $u_G(0,0)$ does not really matter. This justifies our choice to put them to zero to simplify the generic expressions of the potential given in Eq. 2.17 and 2.18.

According to Eq A.7, $F(R,G)$ can be interpreted as the sum of the settling utility of R red and G green agents, these agent settling one by one in an initially empty block, the red agents first and then the green ones. The same goes for Eq. A.8 while the green agents settle first, the red coming next. In fact, relation 2.15 ensures that the sum does not depend on the exact order with which the agents settle (see section 2.3.3 for more precisions). Hence $F(R_q, G_q)$ can be written as the mean of the settling utilities of the agents over all possible orders of settlement:

$$F(R_q, G_q) = \sum_{\substack{0\leq r\leq R_q,\\ 0\leq g\leq G_q\\ (r,g)\neq(0,0)}} \Big(\frac{\alpha_R(r,g)}{\alpha(R_q,G_q)} u_R(r-1,g) + \frac{\alpha_G(r,g)}{\alpha(R_q,G_q)} u_G(r,g-1)\Big)$$

where $\alpha(R_q, G_q) = (R_q + G_q)!$ is the number of settling orders of the $R_q + G_q$ agent living in block q, $\alpha_R(r,g)$ is the number of such orders in which a red agent settle after $r-1$ red and g green, in which case his utility is $u_R(r-1,g)$, and similarly $\alpha_G(r,g)$ is the number of such orders in which a green agent settles after r red and $g-1$ green, in which case the utility of this agent is $u_R(r-1,g)$.

$\alpha_R(r,g)$ is the product of:

- R_q, the number of ways of choosing the r^{th} red settling agent,
- $\binom{R_q-1}{r-1}\binom{G_q}{g}$, the number of ways of sharing out the $R_q - 1 + G_q$ other agents - either before or after this agent,
- $(r+g-1)!$, the number of ways of ordering the agents who settle before and
- $(R_q + G_q - r - g)!$, the number of ways of ordering the agents who settle after.

$\alpha_G(r,g)$ can be similarly computed. We thus end up with a new formula for $F(R_q, G_q)$:

$$F(R_q, G_q) = \sum_{\substack{0\leq r\leq R_q,\, 0\leq g\leq G_q\\ (r,g)\neq(0,0)}} \binom{R_q}{r}\binom{G_q}{g} \frac{(r+g-1)!\,(R_q+G_q-r-g)!}{(R_q+G_q)!}\big(r\; u_R(r-1,g) + g\; u_G(r,g-1)\big) \tag{A.10}$$

A.2 Relation between the potential function $\mathcal{F}$ and the collective utility U

Let us suppose that $(u_R, u_G) \in \mathbb{U}$, and that the potential function of the system can be expressed as a linear function of the collective utility, *ie* $\mathcal{F}(\{R_q, G_q\}) = \lambda U(\{R_q, G_q\}) + \mu$. Since the potential function can be defined up to constant, we can take $\mu = 0$.

Writing the utility functions under the form

$$\begin{aligned} u_R(R,G) &= \xi_R(R) + \sum_{g=0}^{G-1} \xi(R,g) \\ u_G(R,G) &= \xi_G(G) + \sum_{r=0}^{R-1} \xi(r,G) \end{aligned}$$

introduced in Eq. 2.19 and 2.20, the relation of proportionality between the potential and the collective utility can be written as

$$\begin{aligned} & \sum_q \Big(\sum_{r=0}^{R_q-1} \xi_R(r) + \sum_{g=0}^{G_q-1} \xi_G(g) + \sum_{r=0}^{R_q-1} \sum_{g=0}^{G_q-1} \xi(r,g) \Big) \\ = \; & \lambda \sum_q \Big(R_q \xi_R(R_q-1) + R_q \sum_{g=0}^{G_q-1} \xi(R_q-1,g) + G_q \xi_G(G_q-1) + G_q \sum_{r=0}^{R_q-1} \xi(r,G_q-1) \Big) \end{aligned}$$

Since this relation must hold for all $\{R_q, G_q\}$, it follows that that for all $(R,G) \in E_H$, the following holds:

$$\begin{aligned} & \sum_{r=0}^{R-1} \xi_R(r) + \sum_{g=0}^{G-1} \xi_G(g) + \sum_{r=0}^{R-1} \sum_{g=0}^{G-1} \xi(r,g) \\ = \; & \lambda \Big(R \xi_R(R-1) + R \sum_{g=0}^{G-1} \xi(R-1,g) + G \xi_G(G-1) + G \sum_{r=0}^{R-1} \xi(r,G-1) \Big) \end{aligned} \tag{A.11}$$

Taking successively $G = 0$ and $R = 0$ in that last equation provides three independent relations dissociating the three functions ξ_R, ξ_G and ξ:

$$\forall R > 0, \qquad \textstyle\sum_{r=0}^{R-1} \xi_R(r) = \lambda R \xi_R(R-1) \tag{A.12}$$

$$\forall G > 0, \qquad \textstyle\sum_{g=0}^{G-1} \xi_G(g) = \lambda G \xi_G(G-1) \tag{A.13}$$

$$\forall (R,G), \in E_H \quad \textstyle\sum_{r=0}^{R-1} \sum_{g=0}^{G-1} \Big(\lambda \xi(R-1,g) + \lambda \xi(r,G-1) - \xi(r,g) \Big) = 0 \tag{A.14}$$

Notice moreover that the convention $u(0,0) = 0$ implies $\xi_R(0) = \xi_G(0) = 0$. Let us also define $a = \xi_R(1)$, $d = \xi_G(1)$ and $b = \xi(0,0)$. Starting from equations A.12 to A.14,

it is straightforward to prove recursively that

$$\begin{aligned} \lambda &= 1/2 \\ \forall R > 0, \quad \xi_R(R) &= aR \\ \forall G > 0, \quad \xi_G(G) &= dG \\ \forall (R,G) \in E_H \quad \xi(R,G) &= b \end{aligned}$$

Hence the agents' utility functions corresponds exactly to those introduced in Eq. 2.26:

$$\begin{aligned} u_R(R,G) &= aR + bG \\ u_G(R,G) &= bR + dG \end{aligned}$$

The individual utilities are thus necessarily linear in the numbers of similar and dissimilar neighbors in case the potential function $\mathcal{F}$ is proportional to the collective utility U $\square$

A.3 Calculation of a potential function in Schelling case

Suppose that the agents compute their utility with Schelling utility function (which is equal to 1 if their fraction of similar neighbors is superior or equal to 0.5, and equal to 0 otherwise). This utility function can be expressed in terms of the number of red and green neighbors as follows:

$$\begin{aligned} u_R(R,G) &= \Theta(R-G) = \frac{1}{2}(1 + |R+1-G| - |R-G|) \\ u_G(R,G) &= \Theta(G-R) = \frac{1}{2}(1 + |R-1-G| - |R-G|) \end{aligned} \tag{A.15}$$

where Θ is the Heaviside function defined by: $\Theta(x) = 0$ if $x < 0$ and $\Theta(x) = 1$ if $x \geq 0$. Notice that in this example (and in this example only) the convention $u(0,0) + 0$ used in proposition 1 is not respected. The form we choose to write Schelling utility function imposes $u_R(0,0) = u_G(0,0) = 1$. It is easy to figure out that this particular pair of utility functions respect condition 2.15, and is therefore in the set $\mathbb{U}$. Indeed,

$$u_R(R,G) - u_R(R,G+1) = \Theta(R-G) - \Theta(R-G-1) = \begin{cases} 0-0 &= 0 \quad \text{if } R \leq G-1 \\ 1-0 &= 1 \quad \text{if } R = G \\ 1-1 &= 0 \quad \text{if } R \geq G+1 \end{cases}$$

and

$$u_G(R,G) - u_G(R+1,G) = \Theta(G-R) - \Theta(G-R-1) = \begin{cases} 1-1 & = 0 \quad \text{if } R \leq G-1 \\ 1-0 & = 1 \quad \text{if } R = G \\ 0-0 & = 0 \quad \text{if } R \geq G+1 \end{cases}$$

Hence relation $u_R(R,G) - u_R(R,G+1) = u_G(R,G) - u_G(R+1,G)$ is always verified.

To compute a corresponding potential function, one can refer to the general form of Eq. A.9 (since we do not use the convention $u(0,0) = 0$ in this particular example) which can be written here as:

$$\begin{aligned}
\mathcal{F} &= const + \frac{1}{2}\sum_{q\in\mathcal{Q}}\Big[\sum_{r=0}^{R_q-1}\Big(u_R(r,0)+u_R(r,G_q)\Big) + \sum_{g=0}^{G_q-1}\Big(u_G(0,g)+u_G(R_q,g)\Big)\Big] \\
&= const + \frac{1}{4}\sum_{q\in\mathcal{Q}}\Big[\sum_{r=0}^{R_q-1}\Big(3+|r+1-G_q|-|r-G_q|\Big) + \sum_{g=0}^{G_q-1}\Big(3+|R_q-1-g|-|R_q-g|\Big)\Big] \\
&= const + \frac{1}{4}\sum_{q\in\mathcal{Q}}\Big[\Big(3R_q+|R_q-G_q|-G_q\Big) + \Big(3G_q+|R_q-G_q|-R_q\Big)\Big] \\
&= const + \frac{1}{2}\sum_{q\in\mathcal{Q}}\Big(R_q+G_q+|R_q-G_q|\Big) \\
&= const + \frac{1}{2}(N_R+N_G) + \frac{1}{2}\sum_{q\in\mathcal{Q}}|R_q-G_q| \\
&= const' + \frac{1}{2}\sum_{q\in\mathcal{Q}}|R_q-G_q|
\end{aligned}$$

A.4 Potential and collective utility in the case of the asymmetrically peaked utility

Proof of equation 2.37

We have to derive the expression $\tilde{F}(S) = \sum_{s=0}^{S-1}\xi_{ap}(s)$, where

$$\begin{cases} \xi_{ap}(s) & = 2s/H & \text{if } s \leq H/2 \\ \xi_{ap}(s) & = 2-m-2(1-m)s/H & \text{if } s > H/2 \end{cases}$$

For $S-1 \leq H/2$, it is straightforward to write:

$$\tilde{F}(S) = \frac{2}{H}\sum_{s=0}^{S-1} s = \frac{(S-1)S}{H} \tag{A.16}$$

For $S-1 > H/2$, one has

$$\begin{aligned}
\tilde{F}(S) &= \frac{2}{H}\sum_{s=0}^{H/2} s + (2-m)(S-1-H/2) - (1-m)\frac{2}{H}\sum_{s=H/2+1}^{S-1} s \\
&= \frac{2}{H}(2-m)\sum_{s=0}^{H/2} s + (2-m)(S-1-H/2) - (1-m)\frac{2}{H}\sum_{s=0}^{S-1} s \\
&= (2-m)(H/4+1/2+S-1-H/2) - (1-m)\frac{(S-1)S}{H} \\
&= (2-m)\left[S-H/4-1/2-\frac{(S-1)S}{H}\right] + \frac{(S-1)S}{H} \\
&= \frac{(S-1)S}{H} - (2-m)\frac{1}{H}\left[-\left(S-\frac{H}{2}\right)\left(\frac{H}{2}\right) - (S-1)\left(\frac{H}{2}\right) + (S-1)S\right] \\
&= \frac{(S-1)S}{H} - \frac{2-m}{H}\left(S-\frac{H}{2}-1\right)\left(S-\frac{H}{2}\right)
\end{aligned} \tag{A.17}$$

Thanks to the Heaviside function both results can then be written under the general form:

$$\tilde{F}(S) = \frac{(S-1)S}{H} - \frac{2-m}{H}\left(S-\frac{H}{2}-1\right)\left(S-\frac{H}{2}\right)\Theta\left(S-\frac{H}{2}-1\right) \tag{A.18}$$

Proof of equations 2.38 and 2.39

The computation of $\Delta\mathcal{F}$ in relation 2.38 is based on the expression of $\tilde{F}(S)$ (equation A.18), which gives, with $K \in \{0,\, 1, ..., H/2\}$:

$$\begin{aligned}
\Delta\mathcal{F} &= 2\tilde{F}(H/2+1+K) + 2\tilde{F}(H/2-K) - 2\tilde{F}(H/2+1) - 2\tilde{F}(H/2) \\
&= \frac{2}{H}\left[\left(\frac{H}{2}+K\right)\left(\frac{H}{2}+K+1\right) + \left(\frac{H}{2}-K-1\right)\left(\frac{H}{2}-K\right) - \left(\frac{H}{2}\right)\left(\frac{H}{2}+1\right) - \left(\frac{H}{2}-1\right)\left(\frac{H}{2}\right)\right] \\
&\quad - \frac{2(2-m)}{H}K(K+1) \\
&= \frac{2m}{H}K(K+1)
\end{aligned}$$

In the same way,

$$\begin{aligned}
\Delta U &= 2\tilde{U}(H/2+1+K)+2\tilde{U}(H/2-K)-2\tilde{U}(H/2+1)-2\tilde{U}(H/2) \\
&= \frac{4}{H}\left[\left(\frac{H}{2}+K\right)\left(\frac{H}{2}+K+1\right)+\left(\frac{H}{2}-K-1\right)\left(\frac{H}{2}-K\right)-\left(\frac{H}{2}\right)\left(\frac{H}{2}+1\right)-\left(\frac{H}{2}-1\right)\left(\frac{H}{2}\right)\right] \\
&\quad -\frac{4(2-m)}{H}K(H/2+K+1) \\
&= \frac{8}{H}K(K+1)-\frac{4(2-m)}{H}K(H/2+K+1) \\
&= \frac{4K}{H}\big(m(H/2+K+1)-H\big) \\
&= 2\Delta\mathcal{F}-2(2-m)K \qquad \text{(A.19)}
\end{aligned}$$

A.5 Proof of Proposition 2

In this section, we place ourselves in a continuous neighborhood description.

Constraint on the form of the utility functions

Suppose that there exists a potential function $\mathcal{F} : X \to \mathbb{R}$, and let $u_R : E_H \equiv \{(R,G),\, 0 \leq R+G \leq H\} \to \mathbb{R}$ and $u_G : E_H \to \mathbb{R}$ be the red and green agents' utility functions. We want here to investigate if the existence of the potential function imposes any constraint on the form of these utility functions.

Suppose as shown on Fig A.1 that A, B and C are three cells of the lattice such that cells B and C are in each other's neighborhood while cell A is neither in the neighborhood of cell B or C. Let x be a given state in which cells A and C are each occupied by a red agent while cell B is empty. We denote by $(R_A, G_A) \in E_H$, $(R_B, G_B) \in E_H$, $(R_C, G_C) \in E_H$ the number of red and green agents within the neighborhood of cells A, B and C in state x.[1]

Let y be the state obtained when the red agent on cell A moves on cell B and z the state obtained if next the red agent on cell C moves to the now empty cell A. The system goes back from state z to state x by a move of the red agent in cell B to cell C. The function $\mathcal{F}$ being a potential function, the following relations hold (taking into account that the move of the agents change the number of red agents in the neighborhoods of

[1]Beware of the restrictions imposed on the values of (R_B, G_B) and (R_C, G_C) due to the overlapping of the neighborhoods of cells B and C.

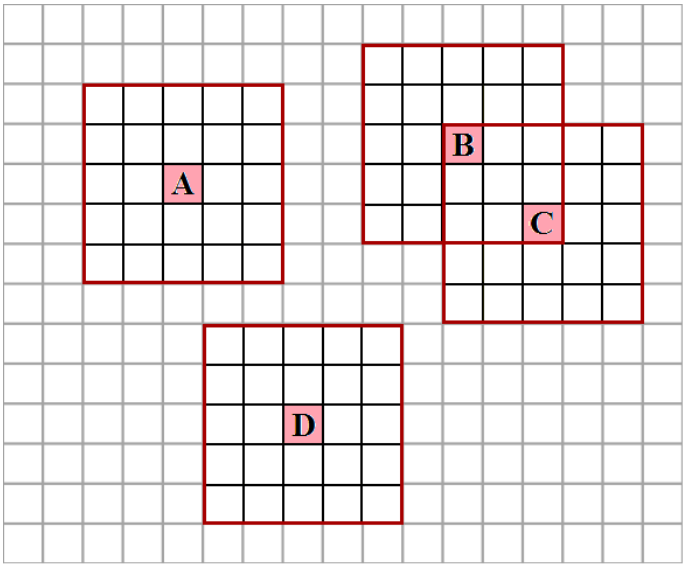

Figure A.1: Example of configuration where the neighborhood of cells B and C overlap while those of cells A and D do not overlap with any of the other enlightened cells. The neighborhood of a cell is composed here by the $H = 24$ nearest cells surrounding it.

cells B and C):

$$\begin{aligned}
\mathcal{F}(y) - \mathcal{F}(x) &= \Delta_{xy}u = u_R(R_B, G_B) - u_R(R_A, G_A) \\
\mathcal{F}(z) - \mathcal{F}(y) &= \Delta_{yz}u = u_R(R_A, G_A) - u_R(R_C + 1, G_C) \\
\mathcal{F}(x) - \mathcal{F}(z) &= \Delta_{zx}u = u_R(R_C, G_C) - u_R(R_B - 1, G_B)
\end{aligned}$$

Summing all these relations yields

$$u_R(R_B, G_B) - u_R(R_B - 1, G_B) = u_R(R_C + 1, G_C) - u_R(R_C, G_C)$$

If the values of (R_B, G_B) and (R_C, G_C) were independent, it would be straightforward to prove that for all values of $(R, G) \in E_H$ such that $R \geq 1$:

$$u_R(R, G) - u_R(R - 1, G) = const$$

We leave to the sagacity of the reader to verify that the restrictions on the values of (R_B, G_B) and (R_C, G_C) due to the overlapping of the neighborhoods of cells B and C

do not change this result. Hence $\forall (R,G) \in E_H$, $\exists\, a \in \mathbb{R}$

$$u_R(R,G) = u_R(0,G) + aR \tag{A.20}$$

The same kind of reasoning can similarly prove that $\forall (R,G) \in E_H$, $\exists\, d \in \mathbb{R}$:

$$u_G(R,G) = u_G(R,0) + dG \tag{A.21}$$

Suppose now as shown on Fig A.1 that A, B, C and D are four cells of the lattice such that cells B and C are in each other's neighborhood while the neighborhoods of cells A and D do not overlap the neighborhood of any of the other three cells. Let x be a given state in which cell A is occupied by a red agent while cell D is occupied by a green agent and cells B and C are empty. We denote by $(R_A,G_A) \in E_H$, $(R_B,G_B) \in E_H$, $(R_C,G_C) \in E_H$ and $(R_D,G_D) \in E_H$ the number of red and green agents within the neighborhood of cells A, B, C and D in state x.

Let y be the state obtained when the red agent on cell A moves to cell B, z the state obtained if next the green agent on cell D moves to cell C, w the state obtained if the red agent on cell B next goes back to cell A. The system goes back from state w to state x by a move of the green agent in cell C to cell D. The function $\mathcal{F}$ being a potential function, the following relations hold (taking into account that the move of the agents change the number of red agents in the neighborhoods of cells B and C):

$$\begin{aligned}
\mathcal{F}(y) - \mathcal{F}(x) &= \Delta_{xy}u = u_R(R_B,G_B) - u_R(R_A,G_A) \\
\mathcal{F}(z) - \mathcal{F}(y) &= \Delta_{yz}u = u_G(R_C+1,G_C) - u_G(R_D,G_D) \\
\mathcal{F}(x) - \mathcal{F}(z) &= \Delta_{zx}u = u_R(R_A,G_A) - u_R(R_B,G_B+1) \\
\mathcal{F}(x) - \mathcal{F}(z) &= \Delta_{zx}u = u_G(R_D,G_D) - u_G(R_C,G_C)
\end{aligned}$$

Summing all these relations yields

$$u_R(R_B,G_B+1) - u_R(R_B,G_B) = u_G(R_C+1,G_C) - u_G(R_C,G_C)$$

Once again, one can check that while the values of (R_B,G_B) and (R_C,G_C) are not independent, the intuitive results holds, which is $\forall (R,G) \in E_{H-1}$, $\exists\, b \in \mathbb{R}$:

$$u_R(R,G+1) - u_R(R,G) = u_G(R+1,G) - u_G(R,G) = b \tag{A.22}$$

To summarize, the existence of the potential function imposes the existence of three

real constants (a, b, d) such that the utility functions take the bilinear form:

$$\begin{aligned} u_R(R,G) &= u_R(0,0) + aR + bG \\ u_G(R,G) &= u_G(0,0) + bR + dG \end{aligned} \tag{A.23}$$

Reciprocal proof

Suppose that the utility of red and green agents are given by Eqs. A.23. Our goal is here to demonstrate that these utility function being given, there exist a potential function of the system.

Suppose that an agent moves from a neighborhood composed of R red and G green agents to a neighborhood composed of R' red and G' green agents. If the moving agent is red, then the variation of his utility can be written as

$$\Delta u = u_R(R', G') - u_R(R, G) = a(R' - R) + b(G' - G)$$

while the variation of the collective utility is:

$$\begin{aligned} \Delta U &= a(R' - R) + b(G' - G) \quad \text{which is the variation of the moving agent utility} \\ &\quad + G'b - Gb \quad \text{which is the sum of the variation of the other green agents} \\ &\quad + R'a - Ra \quad \text{which is the sum of the variation of the other red agents} \\ &= 2a(R' - R) + 2b(G' - G) \;=\; 2\Delta u \end{aligned}$$

Similarly, if the moving agent is green, the variation of his utility can be written as

$$\Delta u = u_G(R', G') - u_G(R, G) = b(R' - R) + d(G' - G)$$

while the variation of the collective utility is:

$$\begin{aligned} \Delta U &= b(R' - R) + d(G' - G) \quad \text{which is the variation of the moving agent utility} \\ &\quad + G'd - Gd \quad \text{which is the sum of the variation of the other green agents} \\ &\quad + R'b - Rb \quad \text{which is the sum of the variation of the other red agents} \\ &= 2b(R' - R) + 2d(G' - G) \;=\; 2\Delta u \end{aligned}$$

Hence, whatever the starting configuration and whatever the move, the variation in utility of the moving agent is always half the variation of the collective utility, a global

function depending on the configuration of the city. Hence, if the utility functions of the agents are given by Eqs. A.23, the function $\mathcal{F} : x \to U(x)/2$ is a potential function of the system $\square$

Appendix B

Annex to Chapter 3 - Collective *vs* Individual Dynamics

B.1 Phase separation

Focusing on the large H case, the problem gets back to finding the set $\{\rho_q\}$ which maximize the potential $F(x) = H\sum_q f(\rho_q)$ with the constraint $\sum_q \rho_q$ fixed. We are interested to know whether the stationary state is statistically homogeneous or inhomogeneous. Following standard physics textbooks methods, the homogeneous state at density ρ_0 is unstable against a phase separation if there exists two densities ρ_1 and ρ_2 such that

$$\gamma f(\rho_1) + (1-\gamma) f(\rho_2) > f(\rho_0). \tag{B.1}$$

The parameter γ $(0 < \gamma < 1)$ corresponds to the fraction of blocks that would have a density ρ_1 in the diphasic state, while a fraction $1-\gamma$ would have a density ρ_2. This condition simply means that the value of the sum $\sum_q f(\rho_q)$ is higher for the diphasic state than for the homogeneous state, so that the diphasic state has a much larger probability to occur. Geometrically, the inequality (B.1) corresponds to requiring that $f(\rho)$ is a non-concave function of ρ. The values of ρ_1 and ρ_2 are obtained by maximizing $\gamma f(\rho_1') + (1-\gamma) f(\rho_2')$ over all possible values of ρ_1' and ρ_2', with γ determined by the mass conservation $\gamma\rho_1' + (1-\gamma)\rho_2' = \rho_0$.

Further, the equilibrium coexistence points to a given temperature can be determined by a double tangent method where the equilibrium densities of the individual phase fall on the same tangent line of $f(\rho)$. The first derivatives of f are equivalent at these two

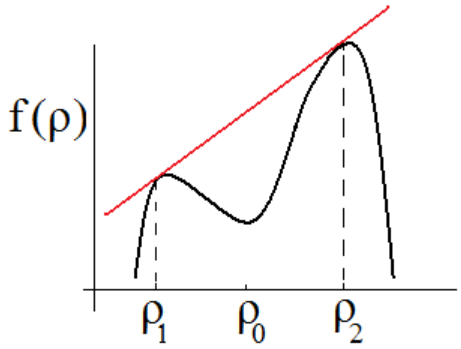

Figure B.1: **Phase separation**. The system of density ρ_0 splits into two phases of densities ρ_1 and ρ_2 if it increases its potential. The standard double tangent construction determines the densities of the two phases at equilibrium.

densities and also equal to the slope connecting these two points, *ie*,

$$\frac{f(\rho_2) - f(\rho_1)}{\rho_2 - \rho_1} = f'(\rho_1) \tag{B.2}$$

$$\frac{f(\rho_2) - f(\rho_1)}{\rho_2 - \rho_1} = f'(\rho_2) \tag{B.3}$$

For the computation of functions depending on the global state of the city such as the normalized collective utility $U^*(x) = U(x)/\sum_q n_q$, two cases have to be distinguished

- The case when there is one phase of density ρ_0. In this case, the utility of each agent is equal to the normalized collective utility :

$$U^*(x) = u(\rho_0) \tag{B.4}$$

- The case when there are two phases of densities ρ_1 and ρ_2. In this case, the normalized collective utility can be written as

$$U^*(x) = \gamma \frac{\rho_1}{\rho_0} u(\rho_1) + (1-\gamma)\frac{\rho_2}{\rho_0} u(\rho_2) \tag{B.5}$$

  with the conservation of the number of agents providing the value of the fraction $\gamma = (\rho_2 - \rho_0)/(\rho_2 - \rho_1)$ of blocks of density ρ_1.

B.2 Asymmetrically peaked utility function

The specific form of the utility function is an input of the model, and it can be postulated on a phenomenological basis, or rely on a theory of the interactions among agents. To illustrate the influence of the parameter α, we choose to work with the asymmetrically peaked utility function defined as:

$$\begin{aligned} u(\rho) &= 2\rho && \text{if } \rho \leq \frac{1}{2} \\ u(\rho) &= m + 2(1-m)(1-\rho) && \text{if } \rho > \frac{1}{2} \end{aligned}$$

where $m < 1$ is a real parameter.

It is straightforward to verify that the function $f(\rho)$ reads for $\rho \leq 1/2$

$$f(\rho) = -T\big(\rho \ln \rho + (1-\rho)\ln(1-\rho)\big) + (1+\alpha)\rho^2 \tag{B.6}$$

and similarly, for $\rho > 1/2$

$$f(\rho) = -T\big(\rho \ln \rho + (1-\rho)\ln(1-\rho)\big) - (1+\alpha)(1-m)\rho^2 + (2-m)\rho - (1-\alpha)(2-m)/4 \tag{B.7}$$

B.2.1 Limiting case T goes to 0

Let us first consider the limiting case $T \to 0$. From the above expression of $f(\rho)$, it turns out that $f(\rho)$ is convex for $0 < \rho < 1/2$ and concave for $1/2 < \rho < 1$, as $1 - m > 0$. Thanks to Fig. B.2, it is pretty clear that there exists $\rho_2(\alpha, m) > 1/2$ such that a phase of mean density ρ_0 is stable if $\rho_0 \geq \rho_2(\alpha, m)$. In the opposite case, a phase separation occurs, and the densities ρ_1 and ρ_2 can be computed as previously explained.

However, in the limit $T = 0$, the line joining ρ_1 and ρ_2 does not correspond to a double tangent. Due to the concavity of f on $[0, 1/2]$, one has $\rho_1 = 0$, but $f'(0) = 0$ while the slope of the line is negative. To determine ρ_2, we first assume that $1/2 < \rho_2 < 1$, so that the line joining $\rho_1 = 0$ to ρ_2 is a tangent to f at ρ_2, which is expressed as:

$$f'(\rho_2) = \frac{1}{\rho_2}\big(f(\rho_2) - f(0)\big), \tag{B.8}$$

yielding

$$\rho_2 = \frac{1}{2}\sqrt{\frac{1-\alpha}{1+\alpha}\,\frac{2-m}{1-m}}. \tag{B.9}$$

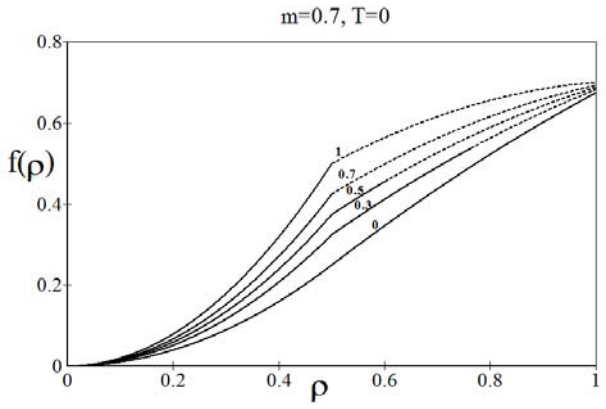

Figure B.2: Graphic representation of the f function for $m = 0.7$, $T = 0$ and $\alpha = 0, 0.3, 0.5, 0.7, 1$. The dash lines represent the part of the curves merging with their concave hulls. The solid line hence corresponds to the range of mean densities ρ_0 for which there is phase separation.

From Eq. (B.9), we find that ρ_2 is in the range $1/2 < \rho_2 < 1$ if (and only if) the following condition is satisfied:

$$\frac{3m-2}{6-5m} = \alpha_t(m) < \alpha < \alpha_c(m) = \frac{1}{3-2m}. \tag{B.10}$$

Hence for $\alpha \geq \alpha_c(m)$, ρ_2 sticks to the value $\rho_2 = 1/2$. Similarly, for $\alpha \leq \alpha_t(m)$, one has $\rho_2 = 1$. These results are illustrated on Fig B.3. The dependency of the outcome with the mean density of agents is quite simple. For $\rho_0 < \rho_2(\alpha, m)$, two kinds of blocks coexist in the stationary states: empty blocks and blocks of density ρ_2. The collective utility can then be written as

$$U^*(x) = u(\rho_2) = 2 - m - \sqrt{\frac{1-\alpha}{1+\alpha}(2-m)(1-m)} \tag{B.11}$$

This expression clearly increases with α, as expected.

In the opposite case for which $\rho_0 \geq \rho_2(\alpha, m)$, the density of the blocks in the stationary states is homogeneous and the collective utility is then

$$U^*(x) = u(\rho_0) = 2 - m - 2(1-m)\rho_0 \tag{B.12}$$

Notice furthermore that the independence of collective utility with the tax parameter α observed on Fig 3 of the article for $\alpha > \alpha_c$ and $\alpha < \alpha_t$ correspond to domains for

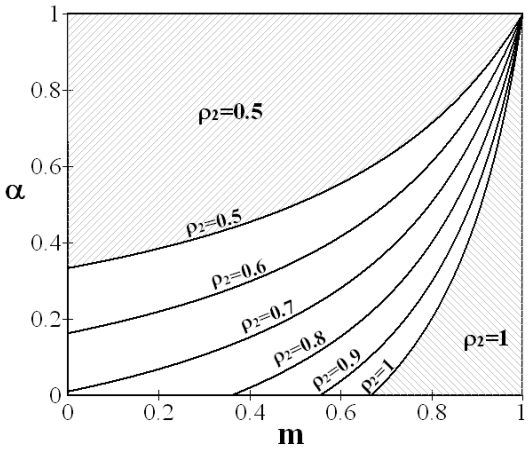

Figure B.3: **Values of** $\rho_2(\alpha, m)$ **at** $T = 0$. If $\rho_0 < \rho_2(\alpha, m)$, the system of density ρ_0 splits into two phases of densities $\rho_1 = 0$ and $\rho_2 = \rho_2(\alpha, m)$ to increase the value of the potential $F(x)$. Otherwise, the equilibrium corresponds to the homogeneous phase of density ρ_0.

which the density ρ_2 has reached a saturation value (respectively 1 or 1/2).

The phase diagrams presented on Fig. B.4 give a more precise idea of the influence of the parameter ρ_0 over the different phases in the stationary states.

B.2.2 Finite temperatures

Finally, we turn to the analysis of the model for finite values of T and show that the behavior of the model remains qualitatively similar to that obtained previously in the $T \to 0$ limit. The high T case is the simplest to analyze. For $2T/(1+\alpha) \geq \max_{[0,1]} \big(4\rho(1-\rho)\big) = 1$, f is concave on the two intervals $[0, 1/2[$ and $]1/2, 1]$ where it is regular. One can moreover verify that at the singular point $\rho = 1/2$, $f'(1/2^+) > f'(1/2^-)$, which ensures that f is concave on the whole interval $[0, 1]$. Hence for $T/(1+\alpha) > 1/2$, there is a single phase of density ρ_0.

In the opposite case $0 < T/(1+\alpha) < 1/2$, the analysis is somewhat similar to the zero T limit. The function f is convex on the interval

$$\frac{1}{2}\left(1 - \sqrt{1 - \frac{2T}{1+\alpha}}\right) < \rho < \frac{1}{2} \tag{B.13}$$

and concave on the complementary interval. As $f(\rho)$ has an infinite slope in $\rho = 0$ and $\rho = 1$, the densities ρ_1 and ρ_2 satisfy $0 < \rho_1 < 1/2$ and $1/2 \leq \rho_2 < 1$. Assuming that

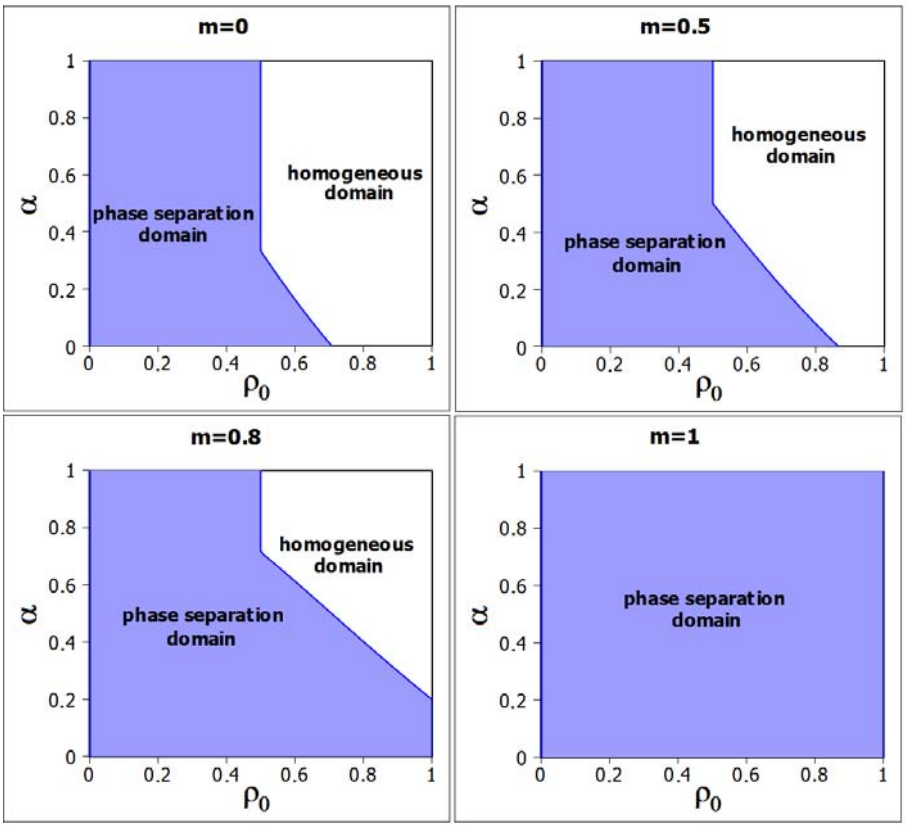

Figure B.4: **Phase diagrams** at $T = 0$ for different values of m. For $\rho_0 > 1/2$, phase separation is always a disadvantage in terms of collective utility. The homogeneous phase, which maximizes the collective utility is stable from a certain value of α. For $\rho_0 \leq 1/2$, the collective utility is maximal for a separation into two phases of densities $\rho_1 = 0$ and $\rho_2 = 1/2$, a separation obtained in the stationary states when $\alpha > 1/(3-2m)$. For lower values of the tax, phase separation is a disadvantage.

$\rho_2 = 1/2$, the density ρ_1 is given by the implicit expression

$$\frac{(1-2\rho_1)^2}{\ln\left(4\rho_1(1-\rho_1)\right)} = -\frac{2T}{1+\alpha}. \tag{B.14}$$

Then the assumption $\rho_2 = 1/2$ is consistent as long as $f'(1/2^+) \geq (f(1/2)-f(\rho_1))/(1/2-\rho_1)$, which can be rewritten as

$$\varphi(\rho_1) \geq \frac{1-\alpha(3-2m)}{1+\alpha} \tag{B.15}$$

where the function φ is defined by

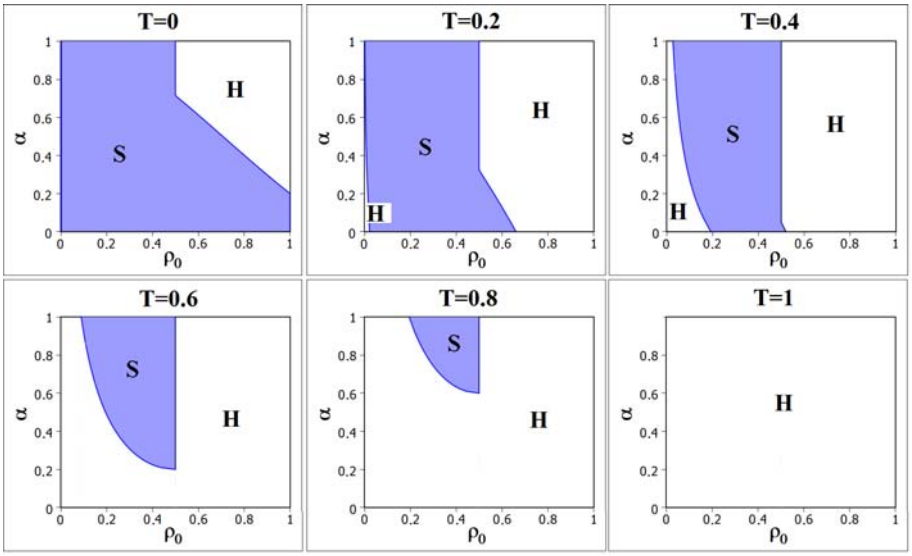

Figure B.5: **Phase diagrams** at $m = 0.8$ for different values of T. Increasing the "temperature" T tends to favour homogeneous states. For $T \to 0$, the phase diagram is affected only for only extremal values of ρ_0, as can be expected from the entropic term $Ts(\rho) = -T\rho\ln\rho - T(1-\rho)\ln(1-\rho)$. As T is increased, all the diagram is affected by the entropic term. Compared to the $T = 0$ case, the main change at low T is the apparition of a second homogeneous phase for $\rho_0 < 1/2$. But whereas for $\rho_0 > 1/2$ homogeneity corresponds to the best interest of the agents, for $\rho_0 < 1/2$, collective utility is not maximized in an homogeneous city. This homogeneous domain is here purely induced by noise. Note that an increase in α tends to reduce this domain while it tends to increase the homogeneous domain for $\rho_0 > 1/2$.

$$\varphi(\rho) = 4\rho - 1 + (1-2\rho)^2 \frac{\ln\rho - \ln(1-\rho)}{\ln(4\rho(1-\rho))}. \tag{B.16}$$

Note that the inequality (B.15) is automatically verified if $\alpha \geq 1/(3-2m)$, as the function φ is positive. If the inequality (B.15) is not satisfied, then $\rho_2 > 1/2$, and the values of ρ_1 and ρ_2 are solutions of two coupled non-linear equations, that can be solved numerically.

The phase diagrams presented on Fig B.5 give a idea of the influence of the "temperature" T over the stationary states of the system.

B.3 Model with two types of agents

B.3.1 Bases of the model

Notations

In this section, describing a city inhabited by two types of agents (that we refer to as red and green agents), we will note:

Q the number of blocks the city is divided in, each block being composed of H cells;

x a configuration of the city, corresponding to the knowledge of the state (empty, red or green) of each cell;

$n_{qr}(x)$ and $n_{qg}(x)$ the numbers of red and green agents living in the block q;

$u(n_{qr}/H)$ (resp $u(n_{qg}/H)$) the utility of a red (resp green) agent living in block q, with $u(0) = 0$ by convention;

$N_0 = \sum_q n_q \leq QH$ the total number of agents;

$N_R = \sum_q n_{qr}$ the total number of red agents (idem for the green ones);

$U(x) = \sum_q (n_{qr}u(n_{qr}/H) + n_{qg}u(n_{qg}/H))$ the total utility in configuration x;

$L(x) = \sum_q \left(\sum_{m=0}^{n_{qr}} u(m/H) + \sum_{m=0}^{n_{qg}} u(m/H) \right)$ the value of the "linking function" in configuration x.

$0 \leq \alpha \leq 1$ the tax parameter.

Dynamic rule

At each iteration, one picks at random an agent and a vacant cell. The agent moves in this empty cell with a probability

$$Pr\{move\} = \frac{1}{1 + e^{-(\Delta u + \alpha(\Delta U - \Delta u))/T}} = \frac{1}{1 + e^{-((1-\alpha)\Delta u + \alpha \Delta U)/T}} \tag{B.17}$$

where Δu is the variation of utility which the chosen agent can achieve by moving, ΔU is the variation of global utility which would result from this same move and $0 \leq \alpha \leq 1$ is an "altruism" parameter (for $\alpha = 0$ the move only depends on the egoistic interest of the agent, for $\alpha = 1$ it only depends on the collective interest).

A potential function

Let us define two states x and y as *immediately communicating states* (ICS) if we can switch from state x to state y by moving one single agent. Whatever the form of the

utility function u, one has for every move $\Delta u = \Delta L$. The transition probability from a configuration x to a configuration y in one iteration can thus be written:

$$\begin{aligned} P_{xy} &= \gamma_{xy} \frac{1}{1 + e^{-((1-\alpha)(L(y)-L(x))+\alpha(U(y)-U(x)))/T}} \\ &= \gamma_{xy} \frac{e^{((1-\alpha)L(y)+\alpha U(y))/T}}{e^{((1-\alpha)L(x)+\alpha U(x))/T} + e^{((1-\alpha)L(y)+\alpha U(y))/T}} \end{aligned}$$

where γ_{xy} takes into account the probability to pick the right agent and the right vacant cell that allow to pass from x to y:

$$\gamma_{xy} = \frac{1}{N_0(QH - N_0)} \quad \text{if } x \text{ and } y \text{ are ICS,} \tag{B.18}$$

$$\gamma_{xy} = 0 \quad \text{if } x \text{ and } y \text{ are not ICS.} \tag{B.19}$$

Since the function

$$\Pi(x) = \frac{e^{(1-\alpha)L(x)+\alpha U(x)}}{\sum_z e^{(1-\alpha)L(z)+\alpha U(z)}} \tag{B.20}$$

is the unique normalized function that verifies for all x and y the detailed balance:

$$\Pi(x) P_{xy} = \Pi(y) P_{yx} \tag{B.21}$$

one can identify Π as the stationary distribution function.

There is $\frac{H!}{n_R! n_G! (H-n_R-n_G)!}$ ways of ordering n_R undifferentiated red agents and n_G undifferentiated green agents in H cells. Indeed, there is $\frac{H!}{(n_R+n_G)!(H-n_R-n_G)!}$ ways of placing the vacant cells and $\frac{(n_R+n_G)!}{n_R! n_G!}$ ways of placing the agents' colors. So one can compute the stationary distribute function for the coarse-grained states $\{\rho_q\}$:

$$\Pi(\{n_q\}) = \frac{1}{Z} \prod_q \frac{H!}{n_R! n_G! (H - n_R - n_G)!} e^{\left((1-\alpha)L(x)+\alpha U(x)\right)/T} \tag{B.22}$$

$$= \frac{1}{Z} e^{H/T \sum_q f(n_q, T, H)} \tag{B.23}$$

where

$$\begin{aligned} f(n_R, n_G, T, H) &= -\frac{T}{H}\ln\Big(\frac{n_R! n_G! (H-n_R-n_G)!}{H!}\Big) \\ &+ \alpha\frac{n_R}{H}u_R\Big(\frac{n_R}{H}\Big) + \alpha\frac{n_G}{H}u_G\Big(\frac{n_G}{H}\Big) \\ &+ (1-\alpha)\frac{1}{H}\sum_{m=0}^{n_R} u_R\Big(\frac{m}{H}\Big) + (1-\alpha)\frac{1}{H}\sum_{m=0}^{n_G} u_G\Big(\frac{m}{H}\Big) \end{aligned}$$

The configurations that maximize the potential $F(x) = \sum_q f(n_R, n_G, T)$ are the more probable to come up. In the limit $H/T \to \infty$, these configurations are even the only ones that will appear in the stationary states (since $\Pi(x)/\Pi(y) = e^{H/T(F(x)-F(y))} \to 0$ for $F(x) - F(y) < 0$ and $H/T \to \infty$).

Continuous limit

In the limit $H \to \infty$, by keeping constant the mean density $\rho_0 = N_0/H$ and the density of each block $\rho_q = n_q/H$ (ρ_q hence becoming a continuous variable), one has thanks to Stirling's formula:

$$\ln\Big(\frac{n_R! n_G! (H-n_R-n_G)!}{H!}\Big) \simeq H\Big(\rho_{qR}\ln\rho_{qR} + \rho_{qG}\ln\rho_{qG} + (1-\rho_{qR}-\rho_{qG})\ln(1-\rho_{qR}-\rho_{qG})\Big)$$

and the stationary distribution can be written as:

$$\Pi(\{\rho_q\}) = \frac{1}{Z}\prod_q e^{H/T f(\rho_{qR}, \rho_{qG}, T)} \tag{B.24}$$

where the "block-potential" is

$$\begin{aligned} f(\rho_R, \rho_G, T) &= -T\rho_R\ln\rho_R - T\rho_G\ln\rho_G - T(1-\rho_R-\rho_G)\ln(1-\rho_R-\rho_G) \\ &+ \alpha\rho_R u_R(\rho_R) + \alpha\rho_G u_G(\rho_G) \\ &+ (1-\alpha)\int_0^{\rho_R} u_R(\rho')d\rho' + (1-\alpha)\int_0^{\rho_G} u_G(\rho')d\rho' \end{aligned}$$

The problem hence gets back to find the set $\{\rho_{qR}, \rho_{qG}\}$ which maximize the potential $F = \sum_q f(\rho_{qR}, \rho_{qG}, T)$ with the constraints $\sum_q \rho_{qR} = Q\rho_{0R}$ and $\sum_q \rho_{qG} = Q\rho_{0G}$.

If one compares this result with the result of the one population model, it can be seen that for a zero temperature the 'two populations model' is obtained by summing

two 'one population models', one for each color. But at a non-zero temperature the $-T(1-\rho_R-\rho_G)\ln(1-\rho_R-\rho_G)$ term links both populations.

B.3.2 Homogeneous-inhomogeneous transitions

The homogeneous phase may be unstable with respect to phase separation.
Let us split the system into two phases of densities $\rho_1 = (\rho_{1R}, \rho_{1G})$ and $\rho_2 = (\rho_{2R}, \rho_{2G})$. The constraint that the overall densities of particles/agents are $\rho_0 = (\rho_{0R}, \rho_{0G})$ is expressed by the lever rule:

$$\begin{cases} Q_1 + Q_2 & = Q \\ Q_1\rho_1 + Q_2\rho_2 & = Q\rho_0 \end{cases}$$

where Q_1 and Q_2 are respectively the number of blocks of density ρ_1 and ρ_2. The homogeneous phase is stable against phase separation if for all ρ_1 and ρ_2

$$Q_1 f(\rho_1) + Q_2 f(\rho_2) < Q f(\rho_0) \tag{B.25}$$

Geometrically, this inequality corresponds to requiring that $f(\rho)$ is a concave function. When the concavity requirement is violated, phase separation will occur for certain values of ρ_0. The equilibrium densities ρ_1 and ρ_2 are such that the line that joins the points (ρ_1, $f(\rho_1)$) and (ρ_2,$f(\rho_2)$) is part of the concave hull of the function.
In the 2 populations model there is a possibility that the system is split into 3 phases of densities $\rho_1 = (\rho_{1R}, \rho_{1G})$, $\rho_2 = (\rho_{2R}, \rho_{2G})$ and $\rho_3 = (\rho_{3R}, \rho_{3G})$. The constraint that the overall densities of particles/agents are $\rho_0 = (\rho_{0R}, \rho_{0G})$ is now:

$$\begin{cases} Q_1 + Q_2 + Q_3 & = Q \\ Q_1\rho_{1R} + Q_2\rho_{2R} + Q_3\rho_{3R} & = Q\rho_{0R} \\ Q_1\rho_{1G} + Q_2\rho_{2G} + Q_3\rho_{3G} & = Q\rho_{0G} \end{cases}$$

where Q_1, Q_2 and Q_3 are respectively the number of blocks of density ρ_1, ρ_2 and ρ_3.
And the equilibrium densities ρ_1, ρ_2 and ρ_3 are now such that the plane that joins the points (ρ_1, $f(\rho_1)$), (ρ_2,$f(\rho_2)$) and (ρ_3,$f(\rho_3)$) is part of the concave hull of the function. For some values of the parameters there may even be 4 points of the same plane belonging to the f function and its concave hull. In this case there will be a continuum of possible values of Q_1, Q_2, Q_3 and Q_4 verifying the global density constraints.

B.3.3 With a peaked utility function

Expression of the f function

Let us consider for both populations the asymmetrically peaked utility function defined for $m < 1$ as:

$$\begin{aligned} u(\rho) &= 2\rho \quad \text{if } \rho \leq 0.5 \\ u(\rho) &= m + 2(1-m)(1-\rho) \quad \text{if } \rho > 0.5 \end{aligned}$$

For $\rho_R \leq 0.5$ and $\rho_G \leq 0.5$, the f function is:

$$\begin{aligned} f(r,g) &= -T\big(r\ln r - g\ln g - (1-r-g)\ln(1-r-g)\big) + (1+\alpha)(r^2+g^2) \\ \frac{\partial f}{\partial r}(r,g) &= -T\big(\ln r - \ln(1-r-g)\big) + 2(1+\alpha)r \\ \frac{\partial^2 f}{\partial r^2}(r,g) &= -T/r - T/(1-r) + 2(1+\alpha) \end{aligned}$$

(Partial derivatives relative to g are obtained by replacing $g \leftrightarrow r$)

For $\rho_R > 0.5$ and $\rho_G \leq 0.5$:

$$\begin{aligned} f(r,g) &= -T\big(r\ln r + g\ln g + (1-r-g)\ln(1-r-g)\big) - (1+\alpha)(1-m)r^2 + (2-m)r \\ &- (1-\alpha)(2-m)/4 + (1+\alpha)g^2 \\ \frac{\partial f}{\partial r}(r,g) &= -T\big(\ln r - \ln(1-r-g)\big) - 2(1+\alpha)(1-m)r - (2-m) \\ \frac{\partial^2 f}{\partial r^2}(r,g) &= -T/r - T/(1-r-g) - 2(1+\alpha)(1-m) \\ \frac{\partial f}{\partial g}(r,g) &= -T\big(\ln g - \ln(1-r-g)\big) + 2(1+\alpha)r \\ \frac{\partial^2 f}{\partial r^2}(r,g) &= -T/g - T/(1-r-g) + 2(1+\alpha) \end{aligned}$$

The situation $\rho_R \leq 0.5$ and $\rho_G > 0.5$ can be obtained by replacing $g \leftrightarrow r$ in the previous paragraph.

For $T = 0$

f is concave in ρ_R and ρ_G for ρ_R and $\rho_G \leq 0.5$. For $\rho_R > 0.5$ and $\rho_G \leq 0.5$, f is concave in ρ_R and convex in ρ_G (and conversely for $\rho_R \leq 0.5$ and $\rho_G > 0.5$, f is concave in ρ_G

and convex in ρ_R).

The concave hull of the function has a different form for different values of the parameters α and m:

- for $\alpha \geq \frac{1}{3-2m}$ the points $\left(0, \frac{1}{2}, f(0, \frac{1}{2})\right)$ and $\left(\frac{1}{2}, 0, f(\frac{1}{2}, 0)\right)$ belong to the concave hull whereas for $\alpha \leq \frac{1}{3-2m}$ they are replaced by the points $\left(0, \rho_2, f(0, \rho_2)\right)$ and $\left(\rho_2, 0, f(\rho_2, 0)\right)$ with $\rho_2(\alpha, m) = \frac{1}{2}\sqrt{\frac{1-\alpha}{1+\alpha}\,\frac{2-m}{1-m}}$ (see the resolution of the one population model).
- for $\alpha \geq \frac{m}{4-3m}$ the point $\left(\frac{1}{2}, \frac{1}{2}, f(\frac{1}{2}, \frac{1}{2})\right)$ belongs to the concave hull whereas for $\alpha < \frac{m}{4-3m}$ it does not.

So there are three possible situations, shown on figure B.6:

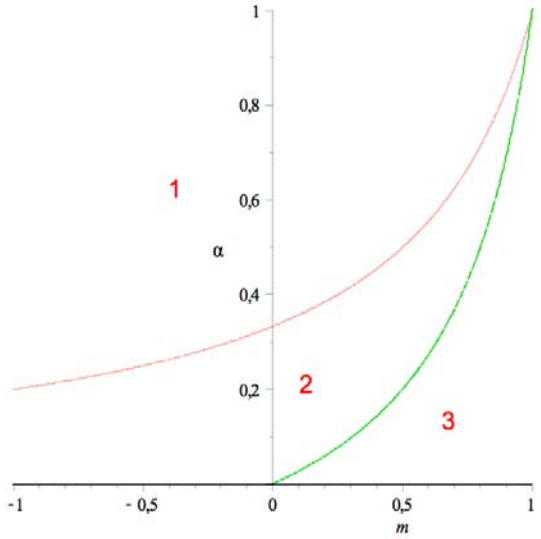

Figure B.6: The domains of different concave hulls for different values of m and α

- $\alpha \geq \frac{1}{3-2m}$ (which will be case 1)
- $\frac{m}{4-3m} \leq \alpha \leq \frac{1}{3-2m}$ (case 2)
- $\alpha < \frac{m}{4-3m}$ (case 3)

Case 1

The number and composition of the phases depend on the global densities $\rho_0 = (\rho_{0R}, \rho_{0G})$ (see figure B.7).

In part A of figure B.7, the system separates into 3 or 4 phases of densities $(0,0)$, $(0,\frac{1}{2})$,

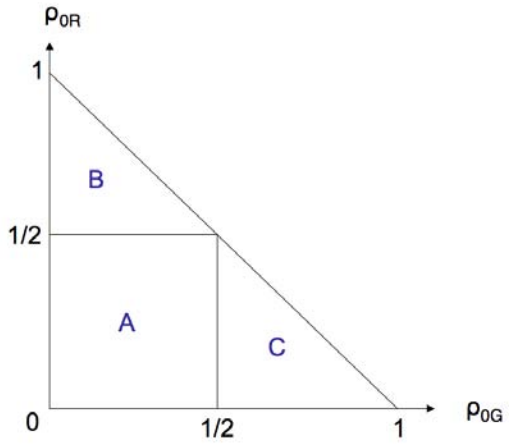

Figure B.7: Domains of different phases for different global densities in case 1

$(\frac{1}{2},0)$ and $(\frac{1}{2},\frac{1}{2})$ in respective quantities Q_1, Q_2, Q_3 and Q_4, which must verify

$$\begin{cases} Q_2+Q_4 & = 2Q\rho_{0G} \\ Q_3+Q_4 & = 2Q\rho_{0R} \\ Q_1+Q_2+Q_3+Q_4 & = Q \end{cases}$$

The system can build either 3 or 4 phases because red and green agents do not "see" each other: their utility is maximal when half of the block is filled with agents of their color, the other half being either empty or filled with agents of the other color.
In part B, the system separates into 2 phases of densities $(\frac{\rho_{0R}-\rho_{0G}}{1-2\rho_{0G}},0)$ and $(\frac{1}{2},\frac{1}{2})$ with respective weights $Q_1 = Q(1-2\rho_{0G})$ and $Q_2 = 2Q\rho_{0G}$.
And symmetrically in part C the system separates into 2 phases of densities $(0,\frac{\rho_{0G}-\rho_{0R}}{1-2\rho_{0R}})$ and $(\frac{1}{2},\frac{1}{2})$ with respective weights $Q_1 = Q(1-2\rho_{0R})$ and $Q_2 = 2Q\rho_{0R}$.

Case 2

The number and composition of the phases depend again on the global densities as shown on figure B.8.
In part A of figure B.8, the system separates into 3 phases of densities $(0,0)$, $(0,\rho_2(\alpha,m))$ and $(\rho_2(\alpha,m),0)$ in respective quantities $Q_1 = Q(1-\frac{\rho_{0R}+\rho_{0G}}{\rho_2})$, $Q_2 = Q\frac{\rho_{0G}}{\rho_2}$ and $Q_3 = Q\frac{\rho_{0R}}{\rho_2}$.

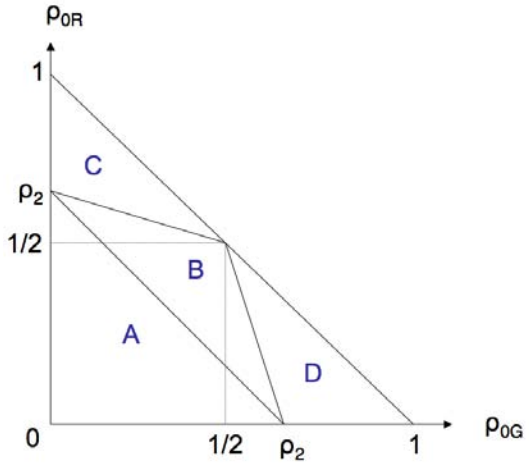

Figure B.8: Domains of different phases for different global densities in case 2

In part B the system is split into 3 phases of densities $(0, \rho_2)$, $(\rho_2, 0)$ and $(\frac{1}{2}, \frac{1}{2})$ in respective quantities $Q_1 = Q\frac{\rho_2(2\rho_{0G}-1)+\rho_{0R}-\rho_{0G}}{2\rho_2(\rho_2-1)}$, $Q_2 = Q\frac{\rho_2(2\rho_{0R}-1)+\rho_{0G}-\rho_{0R}}{2\rho_2(\rho_2-1)}$ and $Q_3 = Q\frac{\rho_2-\rho_{0R}-\rho_{0G}}{\rho_2-1}$.
In part C the phase decomposition is the same as in part B of case 1 and in part D it is the same as in part C of case 1.

Case 3

The different domains of phase decomposition are shown on figure B.9.
In part A of figure B.9 the system separates into 3 phases like in part A of case B.
In part B there are 2 phases of densities $(\rho_{0R} + \rho_{0G}, 0)$ and $(0, \rho_{0R} + \rho_{0G})$ in respective quantities $Q_1 = Q\frac{\rho_{0R}}{\rho_{0R}+\rho_{0G}}$ and $Q_2 = Q\frac{\rho_{0G}}{\rho_{0R}+\rho_{0G}}$.

Interpretation

At zero temperature, the two populations model is indeed very similar to the one population model: both populations of agents see each other only as occupied cells (the utility of a red agent does not depend on the number of green agents in the block). The only difference with two superimposed one population models lies in the fact that blocks with density $(\frac{1}{2}, \frac{1}{2})$ are full of agents with maximal utilities and cannot take in more agents, so that for certain values of the global densities, in cases 1 and 2, there

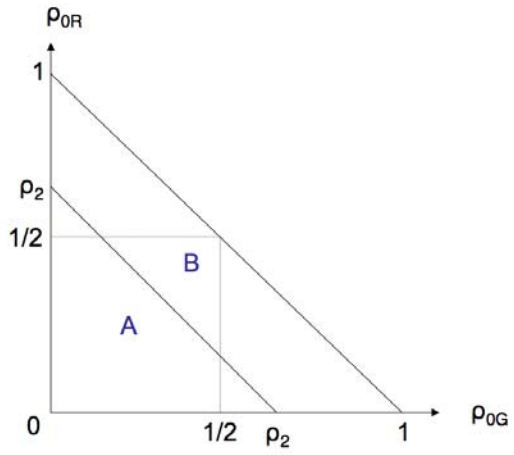

Figure B.9: Domains of different phases for different global densities in case 3

is a phase of half-red, half-green blocks and another phase containing the excess of the more numerous type of agents with a higher density (and thus an inferior utility).

For $T > 0$

At high temperatures, the entropic term of the f function is the leading term, and as it is a concave one, the function is concave everywhere : for any density $\rho_0 = (\rho_{0R}, \rho_{0G})$, the system stays in an homogeneous phase because of the strong noise.
For intermediary values of the temperature, the system has a behavior between a noise driven one and the one it has at zero temperature, but we will not further study it here.

Appendix C

Annex to Chapter 5 - Complex Systems Science, Dreams and Reality

C.1 "Self-Organization" subfields

Community	size	%(SO)	Community	size	%(SO)
GrowthSO	1192	34.64	NanoSO	1995	18.85
SOM	3495	33.88	CondPolymers	471	18.05
SOC	4447	33.48	SurfSO	1511	15.55
NanoFabr	457	31.29	QDots	921	14.12
CytoskSO	651	29.19	CA	846	8.98
PattformSO	691	26.19	PattForm	1403	8.34
TDA	628	25.48	NN	2902	6.48
MolecularSO	2684	24.25	NeuralSynchr	1451	5.86
SwarmSO	608	23.52	Econophys	738	5.42

Table C.1: Communities in which more than 5% of the articles use a keyword containing "self" and "organ*". Examples of such keywords are: self-organisation, self organization, self organizing, self organizing maps, self organized systems, self organized molecules,...

C.2 Most networking references

Abramowitz, M., & Stegun, IA. 1968. *Handbook of Mathematical Functions.* Dover Publications.

Alberts, Bruce, Bray, Dennis, Lewis, Julian, Raff, Martin, Roberts, Keith, & Watson, James D. 1994. *Molecular Biology of the Cell.*

Chomczynski, P., & N., Sacchi. 1987. Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Analytical Biochemistry*, **162**, 156–159.

Cohen, J. 1988. *Statistical power analysis for the behavioral sciences.* Lawrence Erlbaum.

Feller, W. 1958. *An introduction to probability theory and its applications.* Wiley.

Hebb, D.O. 1949. *The organization of behavior: A neuropsychological approach.* Wiley.

Hopfield, J.J. 1982. Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences of the United States of America*, **79**(8), 2554.

Kauffman, S.A. 1993. *The origins of order: Self organization and selection in evolution.* Oxford University Press, USA.

Lakowicz, JR. 1999. Principles of Fluorescence spectroscopy. *Kluwer Academic/Plenum Publisher.*

Marquardt, D.W. 1963. An algorithm for least-squares estimation of nonlinear parameters. *Journal of the Society for Industrial and Applied Mathematics*, **11**(2), 431–441.

Metropolis, N., Rosenbluth, AW, Rosenbluth, MN, Teller, AH, & Teller, E. 1953. Equations of state calculated by fast computing machines. *Journal of Chemical Physics*, **21**, 1087–1092.

Nicolis, G., & Prigogine, I. 1977. *Self-organization in nonequilibrium systems.* Wiley New York.

Press, W.H., Flannery, B.P., Teukolsky, S.A., Vetterling, W.T., *et al.* 1992. *Numerical recipes.* Cambridge university press.

Shannon, C.E. 1948. A mathematical theory of communication. *The Bell System Technical Journal*, **27**, 379–423, 623–656.

Stanley, H.E. 1971. *Introduction to phase transitions and critical phenomena.* Oxford University Press.

Turing, A.M. 1952. The chemical basis of morphogenesis. *Philosophical Transactions*

of the Royal Society of London. Series B, Biological Sciences, **237**(641), 37–72.

Venter, J.C., Adams, M.D., Myers, E.W., Li, P.W., Mural, R.J., Sutton, G.G., Smith, H.O., Yandell, M., Evans, C.A., Holt, R.A., *et al.* 2001. The sequence of the human genome. *Science*, **291**(5507), 1304.

Watson, JD, & Crick, FHC. 1953. Molecular structure of nucleic acids: a structure for deoxyribose nucleic acid. *Nature*, **171**(4356), 709–758.

Whitesides, George M., & Grzybowski, Bartosz. 2002. Self-Assembly at All Scales. *Science*, **295**(5564), 2418–2421.

Witten, T.A., & Sander, L.M. 1981. Diffusion-limited aggregation, a Kinetic Critical Phenomenon. *Physical Review Letters*, **47**(19), 1400–1403.

C.3 Communities "ID cards"

We present in this section the "ID Cards" of the six most discussed communities of the subfield community network, *ie* lists of the most frequent keywords, journals of publication, articles refered to and journal refered to. These ID Cards allow to grasp in a fairly good way the scientific content of a given community.

Table C.2: **Community "ID Card".** The community CN contains $N = 3684$ articles.
Its average internal link weight is $< \omega_{in} > \simeq 1/22$.

Keyword	prop	σ
complex networks	0.53	260.804
dynamics	0.17	48.214
small-world networks	0.13	129.264
model	0.11	23
internet	0.1	96.383
networks	0.08	41.746
evolution	0.08	30.27
scale-free networks	0.07	98.245
systems	0.07	13.515
organization	0.07	44.6
synchronization	0.06	36.731
topology	0.05	66.625
metabolic networks	0.05	78.053
web	0.04	70.114
stability	0.04	14.258
graphs	0.03	55.364
random graphs	0.03	66.888
small-world	0.03	67.309
complex network	0.03	62.893
community structure	0.03	37.757
saccharomyces-cerevisiae	0.02	9.179
social networks	0.02	56.672
emergence	0.02	30.031
percolation	0.02	38.158
robustness	0.02	24.817

Refs	Times used
Albert R, 2002, REV MOD PHYS (74), 47	2006
Barabasi AL, 1999, SCIENCE (286), 509	1659
Watts DJ, 1998, NATURE (393), 440	1556
Newman MEJ, 2003, SIAM REV (45), 167	1288
Strogatz SH, 2001, NATURE (410), 268	877
Dorogovtsev SN, 2002, ADV PHYS (51), 1079	671
Albert R, 2000, NATURE (406), 378	635
Jeong H, 2000, NATURE (407), 651	486
Albert R, 1999, NATURE (401), 130	445
Boccaletti S, 2006, PHYS REP (424), 175	420
Pastorsatorras R, 2001, PHYS REV LETT (86), 3200	390
Amaral LAN, 2000, P NATL ACAD SCI USA (97), 11149	388
Jeong H, 2001, NATURE (411), 41	388
Erdos P, 1959, PUBL MATH-DEBRECEN (6), 290	356
Erdos P, 1960, PUBL MATH I HUNG (5), 17	352
Milo R, 2002, SCIENCE (298), 824	321
Newman MEJ, 2002, PHYS REV LETT (89), 8701	317
Ravasz E, 2002, SCIENCE (297), 1551	305
Faloutsos M, 1999, COMP COMM R (29), 251	298
Cohen R, 2000, PHYS REV LETT (85), 4626	296
Barabasi AL, 1999, PHYSICA A (272), 173	294
Girvan M, 2002, P NATL ACAD SCI USA (99), 7821	282
Dorogovtsev SN, 2003, EVOLUTION NETWORKS B	280
Newman MEJ, 2001, P NATL ACAD SCI USA (98), 404	262
Callaway DS, 2000, PHYS REV LETT (85), 5468	258

Journal	prop	σ
PHYS REV E	0.331	119.35
PHYSICA A	0.217	91.76
LECT NOTE COMPUT SCI	0.082	15.05
PHYS REV LETT	0.067	37.27
EUR PHYS J B	0.065	64.21
PROC NAT ACAD SCI USA	0.041	16.63
INT J MOD PHYS C	0.035	39.6
CHIN PHYS LETT	0.031	40.65
CHAOS	0.029	28.95
IEEE INT SYMP CIRC SYST PROC	0.024	17.26
NEW J PHYS	0.024	39.3
EUROPHYS LETT	0.024	27.81
PHYS LETT A	0.024	21.77
INT J BIFURCATION CHAOS	0.023	17.29
J STAT MECH-THEORY EXP	0.023	38.98

Refs (Journals)	Times used
PHYS REV LETT	9246
NATURE	9101
PHYS REV E 2	8864
SCIENCE	5482
P NATL ACAD SCI USA	4450
PHYSICA A	3563
REV MOD PHYS	2240
PHYS REV E	2047
EUR PHYS J B	1707
SIAM REV	1401
EUROPHYS LETT	1136
J THEOR BIOL	906
PHYS REV E 1	903
ADV PHYS	775
NUCLEIC ACIDS RES	729

Table C.3: **Community "ID Card".** The community SOC contains $N = 4447$ articles.
Its average internal link weight is $< \omega_{in} > \simeq 1/199$

Keyword	prop	σ
self-organized criticality	0.276	180.616
model	0.114	26.379
dynamics	0.103	26.429
evolution	0.075	29.45
systems	0.060	8.954
models	0.039	14.474
fluctuations	0.037	36.555
turbulence	0.035	21.968
self-organization	0.034	4.983
behavior	0.034	8.762
avalanches	0.034	64.80
earthquakes	0.033	58.09
noise	0.031	24.895
transport	0.029	15.886
criticality	0.026	47.168
1/f noise	0.025	48.331
deformation	0.024	28.72
complexity	0.024	3.959
growth	0.021	0.808
patterns	0.021	6.377
flow	0.020	9.583
field	0.020	18.164
simulation	0.019	3.255
system	0.019	0.716
phase-transitions	0.018	18.573

Refs	Times used
Bak P, 1987, PHYS REV LETT (59), 381	1101
Bak P, 1988, PHYS REV A (38),, 364	712
Bak P, 1996, NATURE WORKS SCI SEL	245
Bak P, 1993, PHYS REV LETT (71), 4083	240
Jensen HJ, 1998, SELF ORG CRITICALITY	207
Olami Z, 1992, PHYS REV LETT (68), 1244	202
Paczuski M, 1996, PHYS REV E A (53), 414	146
Bak P, 1989, J GEOPHYS RES-SOLID (94), 15635	139
Burridge R, 1967, B SEISMOL SOC AM (57), 341	122
Dhar D, 1990, PHYS REV LETT (64), 1613	118
Frette V, 1996, NATURE (379), 49	118
Kadanoff LP, 1989, PHYS REV A (39), 6524	117
Bak P, 1996, NATURE WORKS	115
Manna SS, 1991, J PHYS A	112
Turcotte DL, 1999, REP PROG PHYS (62), 1377	102
Vespignani A, 1998, PHYS REV E (57), 6345	100
Drossel B, 1992, PHYS REV LETT (69), 1629	99
Boffetta G, 1999, PHYS REV LETT (83), 4662	96
Lu ET, 1991, APJ	90
Malamud BD, 1998, SCIENCE (281), 1840	90
Dhar D, 1999, PHYSICA A (263), 4	89
Barabasi AL, 1995, FRACTAL CONCEPTS SUR	88
Chang T, 1999, PHYS PLASMAS (6), 4137	86
Hwa T, 1992, PHYS REV A (45), 7002	86
Newman DE, 1996, PHYS PLASMAS 2 (3), 1858	83

Journal	prop	σ
PHYS REV E	0.078	38.48
PHYSICA A	0.042	22.7
PHYS REV LETT	0.033	25.88
PHYS PLASMAS	0.024	44.37
PHYS REV B	0.018	13.53
J GEOPHYS RES-SOLID EARTH	0.016	38.8
AIP CONF PROC	0.013	9.14
TECTONOPHYSICS	0.013	39.82
PLASMA PHYS CONTROL FUSION	0.011	33.05
GEOPHYS RES LETT	0.011	22.95

Refs (Journals)	Times used
PHYS REV LETT	12181
NATURE	3426
J GEOPHYS RES	2874
PHYS REV B	2815
J GEOPHYS RES-SOL EA	2672
GEOPHYS RES LETT	2605
PHYS REV E	2446
PHYS PLASMAS	2426
SCIENCE	2166
PHYSICA A	2121

Table C.4: **Community "ID Card".** The community ComputSystBio contains $N = 1799$ articles. Its average internal link weight is $< \omega_{in} > \simeq 1/324$

Keyword	prop	σ
gene-expression	0.105	29.811
escherichia-coli	0.09	31.696
expression	0.087	13.642
systems	0.066	7.592
model	0.066	6.579
activation	0.063	15.884
protein	0.061	14.106
complex	0.057	7.46
phosphorylation	0.056	23.345
signal-transduction	0.055	33.656
networks	0.052	17.788
systems biology	0.049	62.627
cells	0.049	11.644
transcription	0.046	18.39
dynamics	0.043	3.358
rhythms	0.043	65.654
simulation	0.04	10.691
binding	0.039	9.445
drosophila	0.035	20.722
robustness	0.033	25.4
mechanism	0.032	9.972
kinetics	0.031	13.075
saccharomyces-cerevisiae	0.031	7.647
receptor	0.03	11.805
in-vivo	0.029	6.412

Refs	Times used
Gillespie DT, 1977, J PHYS CHEM-US (81), 2340	159
Dunlap JC, 1999, CELL (96), 271	109
Gardner TS, 2000, NATURE (403), 339	96
Arkin A, 1998, GENETICS (149), 1633	93
Elowitz MB, 2000, NATURE (403), 335	91
Gillespie DT, 1976, J COMPUT PHYS (22), 403	90
Mcadams HH, 1997, P NATL ACAD SCI USA (94), 814	85
Elowitz MB, 2002, SCIENCE (297), 1183	84
Bhalla US, 1999, SCIENCE (283), 381	66
Barkai N, 1997, NATURE (387), 913	65
Kitano H, 2002, SCIENCE (295), 1662	63
Young MW, 2001, NAT REV GENET (2), 702	62
Gibson MA, 2000, J PHYS CHEM A (104), 1876	60
Kume K, 1999, CELL (98), 193	59
Marshall CJ, 1995, CELL (80), 179	59
Gekakis N, 1998, SCIENCE (280), 1564	56
Tyson JJ, 2003, CURR OPIN CELL BIOL (15), 221	56
Shearman LP, 2000, SCIENCE (288), 1013	55
Glossop NRJ, 1999, SCIENCE (286), 766	54
Huang CYF, 1996, P NATL ACAD SCI USA (93), 10078	54
Darlington TK, 1998, SCIENCE (280), 1599	54
Lee K, 2000, SCIENCE (289), 107	53
Reppert SM, 2002, NATURE (418), 935	52
Schoeberl B, 2002, NAT BIOTECHNOL (20), 370	52
Vondassow G, 2000, NATURE (406), 188	52

Journal	prop	σ
PROC NAT ACAD SCI USA	0.036	13.86
BIOPHYS J	0.028	28.81
J BIOL CHEM	0.025	6.27
J THEOR BIOL	0.025	22.29
NATURE	0.012	8.16
LECT NOTE COMPUT SCI	0.012	-3.58
PLOS COMPUT BIOL	0.011	18.17
GENE DEVELOP	0.01	9.2
BIOINFORMATICS	0.01	10.42
CURR BIOL	0.009	8.44

Refs (Journals)	Times used
J BIOL CHEM	4986
P NATL ACAD SCI USA	4928
SCIENCE	4663
NATURE	4556
CELL	3051
EMBO J	1485
J THEOR BIOL	1415
BIOPHYS J	1371
MOL CELL BIOL	1209
NEURON	996

Table C.5: **Community "ID Card".** The community Transcriptomics Data Analysis (TDA) contains $N = 628$ articles. Its average internal link weight is $< \omega_{in} > \simeq 1/43$

Keyword	prop	σ
patterns	0.273	65.409
self-organizing maps	0.216	99.058
gene-expression	0.159	28.356
identification	0.116	17.682
microarray	0.108	79.859
classification	0.100	28.106
cancer	0.095	28.214
gene expression	0.082	36.526
saccharomyces-cerevisiae	0.068	13.212
cell-cycle	0.062	21.492
prediction	0.060	16.642
expression	0.060	4.145
gene-expression data	0.058	95.821
oligonucleotide arrays	0.058	85.248
discovery	0.054	47.311
microarray data	0.052	68.108
cluster-analysis	0.051	67.97
clustering	0.051	29.864
DNA microarray	0.049	66.328
microarrays	0.046	48.629
cDNA microarray	0.044	60.323
hybridation	0.044	31.714
cDNA microarrays	0.043	76.89
gene-expression patterns	0.039	73.138
DNA microarrays	0.039	59.58

Refs	Times used
Eisen MB, 1998, P NATL ACAD SCI USA (95), 14863	314
Tamayo P, 1999, P NATL ACAD SCI USA (96), 2907	273
Golub TR, 1999, SCIENCE (286), 531	142
Toronen P, 1999, FEBS LETT (451), 142	123
Spellman PT, 1998, MOL BIOL CELL (9), 3273	96
Schena M, 1995, SCIENCE (270), 467	89
Alizadeh AA, 2000, NATURE (403), 503	88
Tavazoie S, 1999, NAT GENET (22), 281	78
Alon U, 1999, P NATL ACAD SCI USA (96), 6745	77
Brown MPS, 2000, P NATL ACAD SCI USA (97), 262	77
Derisi JL, 1997, SCIENCE (278), 680	68
Cho RJ, 1998, MOL CELL (2), 65	62
Lockhart DJ, 1996, NAT BIOTECHNOL (14), 1675	53
Iyer VR, 1999, SCIENCE (283), 83	51
Wen XL, 1998, P NATL ACAD SCI USA (95), 334	46
Herrero J, 2001, BIOINFORMATICS (17), 126	45
Tusher VG, 2001, P NATL ACAD SCI USA (98), 5116	44
Hastie T, 2001, ELEMENTS STAT LEARNI	42
Alter O, 2000, P NATL ACAD SCI USA (97), 10101	39
Khan J, 2001, NAT MED (7), 673	39
Brown PO, 1999, NAT GENET S (21), 33	38
Bendor A, 1999, J COMPUT BIOL (6), 281	37
Chu S, 1998, SCIENCE (282), 699	36
Lipshutz RJ, 1999, NAT GENET S (21), 20	36
Perou CM, 2000, NATURE (406), 747	34

Journal	prop	σ
BIOINFORMATICS	0.057	39.49
BMC BIOINFORMATICS	0.044	34.68
LECT NOTE COMPUT SCI	0.039	2.12
PHYSIOL GENOMICS	0.023	34.87
PROC NAT ACAD SCI USA	0.02	3.75
IEEE IJCNN	0.019	5.98
GENOME RES	0.011	13.64
J BIOMED INFORM	0.011	23.57
J COMPUT BIOLOGY	0.009	15.55
P SOC PHOTO-OPT INSTRUM ENG	0.009	-2.07

Refs (Journals)	Times used
P NATL ACAD SCI USA	2002
SCIENCE	1057
BIOINFORMATICS	844
NATURE	601
NAT GENET	543
NUCLEIC ACIDS RES	486
J BIOL CHEM	392
CANCER RES	345
GENOME RES	286
CELL	268

Table C.6: **Community "ID Card".** The community Transcriptomics contains $N = 2043$ articles. Its average internal link weight is $< \omega_{in} > \simeq 1/439$

Keyword	prop	σ
saccharomyces-cerevisiae	0.17	67.43
gene-expression	0.12	37.784
messenger-RNA	0.11	48.577
identification	0.11	32.298
expression	0.1	20.177
complex	0.09	17.985
yeast	0.08	48.156
gene	0.07	26.323
protein	0.07	21.22
caenorhabditis-elegans	0.05	57.728

Refs	Times used
Lander ES, 2001, NATURE (409), 860	108
Ho Y, 2002, NATURE (415), 180	97
Bartel DP, 2004, CELL (116), 281	95
Gavin AC, 2002, NATURE (415), 141	94
Fire A, 1998, NATURE (391), 806	90
Gygi SP, 1999, NAT BIOTECHNOL (17), 994	79
Venter JC, 2001, SCIENCE (291), 1304	72
Uetz P, 2000, NATURE (403), 623	71
Hilleren P, 2001, NATURE (413), 538	67
Lacava J, 2005, CELL (121), 713	63

Journal	prop	σ
J BIOL CHEM	0.039	12.93
MOL CELL BIOL	0.027	19.89
NUCL ACID RES	0.025	26.13
PROC NAT ACAD SCI USA	0.023	8.22
RNA	0.02	41.06
MOL CELL	0.018	24.08
GENE DEVELOP	0.017	18.06
BIOCHEM BIOPHYS RES COMMUN	0.015	13.98
CELL	0.014	15.77
EMBO J	0.013	12.26

Refs (Journals)	Times used
J BIOL CHEM	5265
CELL	5119
P NATL ACAD SCI USA	5056
NATURE	5043
MOL CELL BIOL	4497
SCIENCE	4243
EMBO J	3835
GENE DEV	3411
NUCLEIC ACIDS RES	2634
MOL CELL	2052

Table C.7: **Community "ID Card".** The community Self-Organized Maps (SOM) contains $N = 3495$ articles. Its average internal link weight is $< \omega_{in} > \simeq 1/169$

Keyword	prop	σ
self-organizing map	0.1	111.867
neural networks	0.08	38.377
self-organizing maps	0.08	85.996
classification	0.06	39.352
model	0.03	-0.268
neural network	0.03	25.875
clustering	0.03	42.989
algorithm	0.03	14.097
prediction	0.02	15.602
maps	0.02	30.932

Refs	Times used
Kohonen T, 1982, BIOL CYBERN (43), 59	530
Kohonen T, 1995, SELF ORG MAPS	506
Kohonen T, 1990, P IEEE (78), 1464	401
Haykin S, 1994, NEURAL NETWORKS COMP	275
Kohonen T, 1997, SELF ORG MAPS	269
Kohonen T, 2001, SELF ORG MAPS	263
Vesanto J, 2000, IEEE T NEURAL NETWOR (11), 586	160
Kohonen T, 1989, SELF ORG ASS MEMORY	150
Fritzke B, 1994, NEURAL NETWORKS (7), 1441	134
Sammon JW, 1969, IEEE T COMPUT (18), 401	131

Journal	prop	σ
LECT NOTE COMPUT SCI	0.135	40.23
IEEE IJCNN	0.053	46.01
P SOC PHOTO-OPT INSTRUM ENG	0.039	7.33
LECT NOTE ARTIF INTELL	0.036	20.76
IEEE TRANS NEURAL NETWORKS	0.023	30.39
NEURAL NETWORKS	0.023	31.27
NEUROCOMPUTING	0.02	22.97
EXPERT SYST APPL	0.012	22.67
IEEE SYS MAN CYBERN	0.008	9.6
PATT RECOG	0.007	15.54

Refs (Journals)	Times used
IEEE T NEURAL NETWOR	2202
NEURAL NETWORKS	1612
BIOL CYBERN	1158
SELF ORG MAPS	1133
NEURAL COMPUT	1074
IEEE T PATTERN ANAL	823
P IEEE	757
PATTERN RECOGN	630
SCIENCE	592
NEUROCOMPUTING	532

Appendix D

Complex Systems Science - An Historical Perspective

In this appendix devoted to the bibliometric study of Complex Systems Science (CSS), we explore the historic evolution of Complex Systems Science. We present some preliminary maps of CSS corresponding to different time periods and we build on the properties of bibliometric coupling to detect the chronological affiliation of communities.

D.1 Complex Systems History

D.1.1 Database

The "Complex Systems" database gathered from WoS contains in total 215 000 records of articles published from 1950 to 2009. The results presented in the last chapter were only based on data corresponding to articles published after 2000, *ie* in the most recent period. In this section, we present results based on the whole database.

Fig. D.1 displays the number of articles within our database by year of publication. Its overall increase with time reflects the general increase of scientific production. A large jump in the number of published articles can be observed around 1990. While it might be tempting to interprete this jump as a historical event (a sudden interest for the Complex Systems science or a massive increase in the scientific production due to the wide circulation of computers), it is in fact merely a bias introduced by the WoS's extraction tools. Indeed, when a query is put in WoS, the Topic Keywords of the query are looked for in the titles, article's keywords and, in case of publication posterior to 1990, abstracts. This means that our query looking for specific Topic Keywords

(as defined in Table 5.5) reaches many more articles published after 1990, those which contain these Topic Keywords exclusively in their abstracts.

To study the historical evolution of the science of complex systems, we divided the records in several 10-years slices according to their publication year. For each decade, we build a network where the nodes correspond to the articles published in the corresponding period of time and where a link between two nodes indicates that the papers corresponding to these nodes share at least one reference. The weight of the link is taken as the BC cosine distance defined in Eq. 5.1.

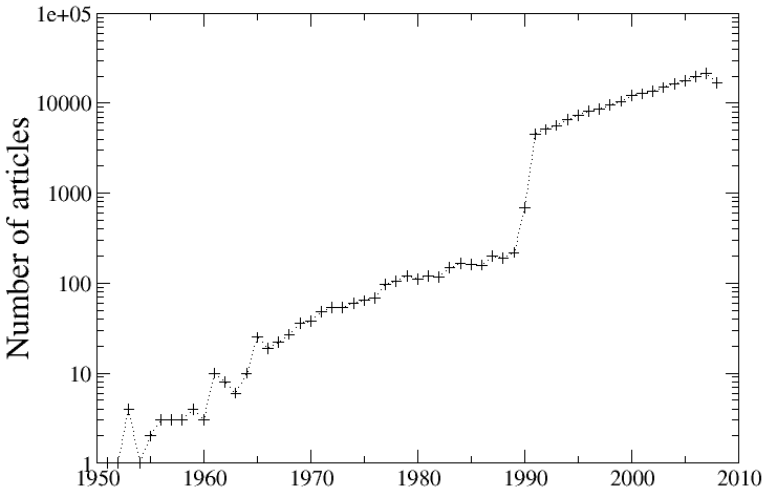

Figure D.1: **Number of papers within the *Complex Systems* database by year of publication.**

The general characteristics of these networks are displayed in Table D.1. For the first decades (roughly until 1985), the number of articles is rather low. The networks can be easily built and submitted to a manual analysis. For the last decades, the numbers of articles becomes very important, the mean density $2M/N(N-1)$ decreases with time while the mean distance $< \omega_{ij} >^{-1}$ increases, a sign that the networks are on average more heterogeneous in recent periods. On the other side, the mean degree M/N of an article increases with the time period, a sign that the networks contain more homogeneous subparts.

Decade	N_a	N	$2M/N(N-1)$	M/N	$< \omega_{ij} >^{-1}$
1960-1969	166	41	0.0902	1.80	55
1965-1974	383	138	0.0315	2.16	266
1970-1979	710	284	0.0298	4.23	490
1975-1984	1118	519	0.0273	7.09	727
1980-1989	1590	895	0.012968	5.80	874
1985-1994	23501	20286	0.004882	49.62	4484
1990-1999	66787	62040	0.003211	99.61	7283
1995-2004	114268	109458	0.002114	115.68	11299
2000-2009	146030	141098	0.001969	138.94	10848

Table D.1: **General characteristics of the articles' BC Networks.** For each decade, N_a is the numbers of articles in our database, N is the number of article in the network (*ie* sharing at least one reference with at least an other articles of our basis), M is the number of BC links (hence $2M/N(N-1)$ is the density of number of links in the graph and M/N is the mean number of links in which an article takes part, its degree). The average $< \omega_{ij} >$ is done on the $N(N-1)/2$ pairs of articles. It can be taken as a measure of a "bibliographic coupling" network inner-coherence, while $< \omega_{ij} >^{-1}$ can be seen as a characteristic "bibliographic distance" of each network. Indeed, if the links' weights were homogeneously distributed, two randomly chosen articles would share one reference over $< \omega_{ij} >^{-1}$.

D.1.2 Chronologically successive maps

Methods

Starting from the last decades (1985-1995, 1990-1999, 1995-2004 and 2000-2009), we first obtain a network by applying the Louvain algorithm (Blondel *et al*, 2008). This step gives a general perspective on the structuration and evolution of the composition of our database in terms of large scientific fields. We typically obtain a dozen large field communities in each decade (see Table D.2). In order to gain a better understanding on the composition of these fields and on the nature of their links, we apply the same two-steps procedure introduced last chapter. We group the papers in two hierarchical community networks for each decade, a rough one (field structuration) and a specific one (subfields). Applying the Louvain algorithm once to the entire network yields the field communities. The subfield communities are obtained by

- applying the Louvain algorithm to each field community
- checking if the obtain substructure of each field has is meaningful (internal modularity $Q^i \geq 0.4$) and whether each subfield check Eq. D.1 in the global subfield network. If these two conditions are not respected for a given field, it is kept as a whole in the subfield network.

Decade	N	Q_f	$C_f\,(N_f)$	Q_{sf}	$C_{sf}\,(N_{sf})$	Thr
1985-1994	20286	0.7421	12 (19814)	0.6822	43 (17262)	100
1990-1999	62040	0.75	12 (61460)	0.7034	67 (59739)	100
1995-2004	109458	0.7592	14 (108739)	0.7586	82 (106753)	100
2000-2009	141098	0.7128	11 (139749)	0.6375	79 (135834)	300

Table D.2: **Quantitative characterization of the obtained community networks.** For each decade, we report the number of articles within the bibliographic network, the modularity Q_f, the number of field communities C_f containing more than Thr articles (and total number of articles N_f within) of the field community network obtained by the first run of the Louvain algorithm. The second run of Louvain algorithm performed independently on each field lead to subfields community networks whose modularity Q_{sf} and number of communities C_{sf} (of size superior than Thr) are also reported. Note that the finer partition of the subfield network comes with a lowering of the modularity.

Decade	Community	N	$< \omega >^{-1}$	Q^i
	Self-Organized Criticality (SOC)	925	39.75	0.165
	Mitosis	352	53.27	-
	Hematology (Hemato)	149	181.28	-
1985	**Statistical physics (StatPhys)**	302	249.04	-
	Molecular Biology (MolBio)	1859	284.55	-
-	**Neural Networks**	2526	569.66	0.407
	Cellular Biology (CellBio)	3050	752.38	0.476
1994	**Control**	825	896.94	0.771
	Dynamical Systems	3395	963.51	0.508
	Fluid Mechanics & Geoscience (FluidMech Geosc)	971	1502.93	0.738
	Material Sciences	2328	1755.14	0.76
	Neurosciences	3132	3150.53	0.758
	Self-Organized Criticality(SOC)	2650	61.83	0.169
	Surface Self-Organization (SurfSO)	975	75.67	0.207
	Statistical Physics (StatPhys)	295	126.94	-
1990	**CellBio**	1671	172.45	-
	Epigenomics	10494	896.8	0.475
-	**Fluid Mechanics & Geosciences (FluidMech Geosc)**	2184	1235.54	0.632
	Dynamical Systems	9105	1307.73	0.499
1999	**Neural Networks**	11118	1336.69	0.692
	Immunology (Immuno)	4343	2278.94	-
	Material Sciences	7472	2534.48	0.679
	Neurosciences	6563	3033.74	0.692
	Ecology	4403	3952.66	0.663
	Circadian Clock (CircClock)	161	19.23	-
	Complex Networks (CN)	875	24.2	0.093
	Self-Organized Criticality (SOC)	4824	137.08	0.219
	Quantum Chaos (Qchaos)	931	292.8	-
1995	**Surface Self - Organization (SurfSO)**	3441	247.48	0.397
	Digital Communication (DigitCom)	3252	719.89	0.573
-	**NeuralNetworks**	13834	964.24	0.512
	Dynamical Systems	11837	1311.07	0.542
2004	**Molecular and Cellular Biology (MolBio CellBio)**	20524	2051.06	0.571
	Immunology & Genetic Deseases (Immuno GenDis)	6301	2397.72	-
	Material Sciences	13551	2674.78	0.665
	Neurosciences	8966	2775.75	0.651
	Fluid Mechanics & Geosciences (FluidMech Geosc)	5688	3265.16	0.723
	Ecology - Management - Computational Model (EMC)	14279	7437.35	0.704
	Complex Networks (CN)	3684	21.87	0.087
	Self-Organized Criticality (SOC)	4447	199.3	0.202
	Surface Self-Organization (SurfSO)	3809	493.33	0.517
2000	**Digital Communication (DigitCom)**	6094	988.63	0.579
	Dynamical Systems (DynSystems)	13115	1661.96	0.571
-	**Neural Networks (NN)**	21913	1759.12	0.532
	Fluid Mechanics (FluidMech)	5534	2689.56	0.685
2009	**Material Sciences**	19531	3148.04	0.656
	Neurosciences	12567	3231.99	0.708
	Cellular and Molecular Biology (Biology)	32107	4852.62	0.615
	Ecology - Management - Computational Model (EMC)	16948	5560.05	0.708

Table D.3: **Fields' size N, inner coherence $< \omega >^{-1}$ and internal modularity Q^i.** Fields whose subfield structure was found to be unsignificative (roughly $Q^i < 0.2$) are enlightened in colors. In all cases, the biological fields were gathered as a single field just before the computation of the subfield communities. Hence the lack of some values of internal modularity and the fact that these communities share the same color on Fig D.3. In the same way, the *Quantum Chaos* and *Statistical Physics* field were merged with the *Dynamic Systems* fields before the second partitions.

This double recursive use of a modularity optimization was also used by (Fortunato & Barthélemy, 2007), on smaller networks (less than 1000 nodes) and with another community detection algorithm. They noted that *by restricting modularity optimization to a [single community during the second application of Louvain algorithm], we neglect all links between the original communities and we have no guarantee that we accurately detect its substructure and that this is a safe way to proceed. Thus, we have to check whether all substructures we detected are real [communities]*, *ie* if a subcommunity of a given up community is also a community of the entire network. To do so, they propose a simple criterion:

$$\text{a set } I \text{ of articles is a suitable community if:} \quad q_I \geq 0 \tag{D.1}$$

where q_I is the module of a community as defined in Eq. 5.2. We found that this criterion was well respected for all the subfield communities in all the networks we built.

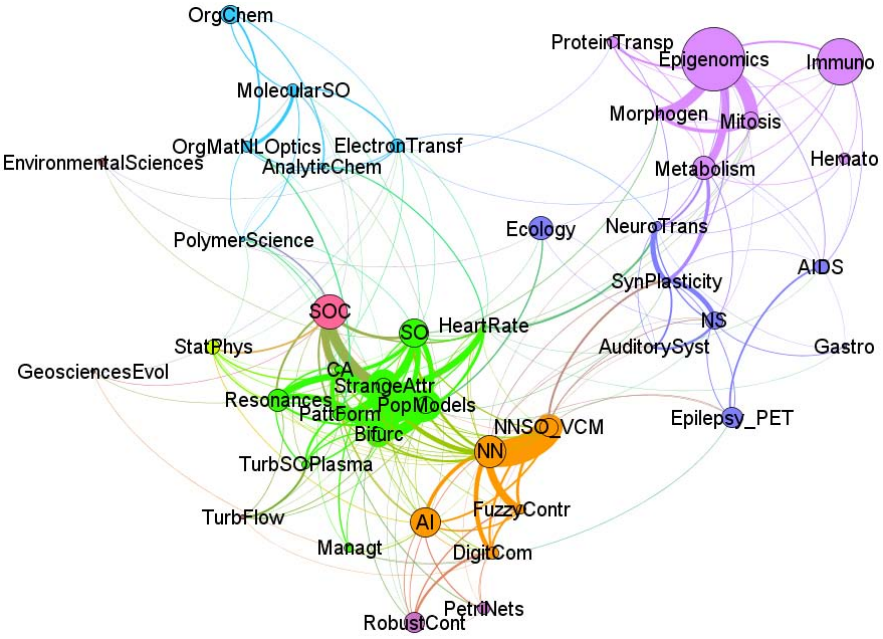

Figure D.2: **1985-1994's subfield community network.** The surface of a community I is proportional to its number of articles N_I and the width of the link between two communities I and J is proportional to the mean bibliographic coupling $\omega_{IJ} = \sum_{i \in I,\, j \in J} \omega_{ij} / N_I\, N_J$. For the sake of clarity, communities with less than 100 articles and links with a mean weight $\omega < 2.10^{-5}$ are not displayed. Labels are based on a frequency analysis.

Results

Fig. D.3 displays the largest field community networks obtained for the four successive and overlapping decades. We first note that these different networks show a remarkable continuity. At the center of these networks, we find theoretical domains: self-organized criticality, dynamical systems, complex networks and neural networks. These fields are linked to more experimental fields (materials science, biology or neurosciences) lying at the periphery of the networks. Among all these communities, two in particular stand out. First, the *complex networks* (CN) community, which appears only from the 1995-2004 decade since its main founding references (namely Albert & Barabasi, 2002; Barabasi & Albert, 1999; Watts & Strogatz, 1998; Newman, 2003, see appendix B) were all published after 1995. This community has the strongest coherence : on average, the mean bibliographic coupling weight between two articles of this community is $< \omega > \sim 1/20$ (see Table D.3). This value means that two papers taken randomly from this community very often share a reference. This coherence remained stable while the number of articles increased from around 900 in the 1995-2004 decade to around 3700 the 2000's (see Table D.3). The *self-organized criticality* (SOC) community presents similar characteristics: its main founding references are well-defined (namely Bak *et al*, 1987, 1988, see appendix B) and its coherence is among the strongest. However, the scientific production of this community has begun to decrease (around 4800 articles in the 1990-1999 decade *vs* around 4400 in the 1995-decade. This decrease is significant when compared to the general increase of scientific production. In the meantime, the average coherence decreased from $< \omega > \sim 1/40$ in the 1985-1994 decade to $< \omega > \sim 1/200$ in the 2000-2009 one (see Table D.3). Compared to *complex networks*, *self-organized criticality* seems to be declining as a scientific community.

A point worth noting is that while most communities are close to "standard" disciplines such as those of the Web of Science subject index, CN and SOC communities do not. This shows the importance of using bibliographic coupling to define communities instead of taking as starting points the Web of Science "subject categories".

We now come to the level of subfield communities. The subfield networks displayed on Fig D.2 to D.5 show a rich structure that appears to be quite stable and similar to the structure of the 2000-2009 decade presented in the previous chapter. We do not intend to give here a full analysis of the inner composition and the inter-relations of the subfields displayed on those maps. However, a quick analysis of the database allow us to check two points. First, the most networking references (see Chapter 5) are still mainly methodological references, mathematical handbooks or data analysis tools, although references to Self Organization are also present. Second, trading zones are

always present. For example, in the 1995-2004 decade, the two communities *Circadian Clock* (CircClock) and *SelfOrganized Maps and Biological Prediction* (SOMBioPred) are detected as trading zones. They connect the biological community respectively to *CN* and *Dynamical Systems* for the former and to *Neural Networks* for the latter. It is hence tempting to identify them to the 1995-2004 decade's precursor of the *Computatinal Systems Biology* and *Transcriptomics Data Analysis* of the 2000-2009 decade we studied in Chapter 5. We present in next section a tentative method for detecting such relations of filiation between communities identified on different time periods.

D.1.3 Communities's filiation detection

We present in this section a method for detecting and visualizing the chronological affiliation of communities from one time slice to another. A first way to characterize the relation between a community A of of given time slice to a community B of a successive time slice is to compute the relative overlap

$$o(A,\, B) = \frac{|A \cap B|}{|A \cup B|}$$

where $|A \cap B|$ is the number of common articles between both communities (remember that there is a 5-years overlap between successive time slices) and $|A \cup B|$ the total number of distinct articles within A and B. Communities from successive time slices can thus be matched one by one in descending order of their relative node overlap. Hence for example, the communities *Complex Network* of the 2000-2009's decade and *Complex Network* of the 1995-2004's decade are matched together with an overlap of 0.17, *Self Oganized Criticality* and *Self Organized Criticality* are matched with an overlap of 0.22, *SOMBioPred* and *TDA* with an overlap of 0.44 and *Circadian Clock* and *ComputSystBio* with an overlap of 0.05.

This last value seems rather low - even by taking into account the exponential increase of the overall scientific production which is expected to bound the overlap value, and raise the question of the validity of this method. Indeed, if the *Circadian Clock* and *ComputSystBio* share some articles, how can we be sure that the remaining articles are really related to a common scientific subject?

We have used a second method, based on the computation of the mean 'Bibliographic Coupling' (BC) weight between articles of two communities belonging to two different time slices. We hence characterize the relation between communities of different time slices by their shared references in the same way that we characterize the relation between communities of the same time slice. The advantage of measuring BC

weight instead of the overlap is the fact that it allows to estimate a scientific proximity even when the overlap is null.

To illustrate the informations bring by this method, we propose on Fig D.6 four "filiation diagrams". Starting from a 2000-2009 community C (at the top, in red), these diagrams show the 1995-2004 communities C' (in orange) whose references are most similar to those of community C (we use a BC weight threshold of $\omega_{CC'} > 10^{-4}$). The same procedure is repeated on the $1995-2004$ community most similar to community C, allowing to see the 1990-1999 communities C'' (in yellow) most similar to it and so on. For example, the *Complex Network* community of the 2000-2009 decade builds on its 1995-2004 counterpart, but also on SOC, on some Dynamic systems' communities (*Chaos* and *Neural Synchronization*) some Neural Networks' communities and *Pattern Formation and Self Organization.* The 1995-2004 *Complex Network* community is itself based mainly on communities related to the idea of Self Organization. The Self-Organized Criticality community offers a clear affiliation: SOC is mainly related to SOC, which is mainly related to SOC... The secondary ascendants show communities known to be related to SOC (*Heart Rate*, *Geophysics*, *NeuralSynch*).

The case of the *ComputSystBio* is also interesting from a historical perspective: models of Circadian Clock and of the enzymatic kinetics implied in Photosynthesis are known to be at the foundation of modern Systems Biology.

Of course all these observations need to be deepened by a close historical analysis of the precise shared references between all these communities, the institutions and personal links at the origin of these evolutions.

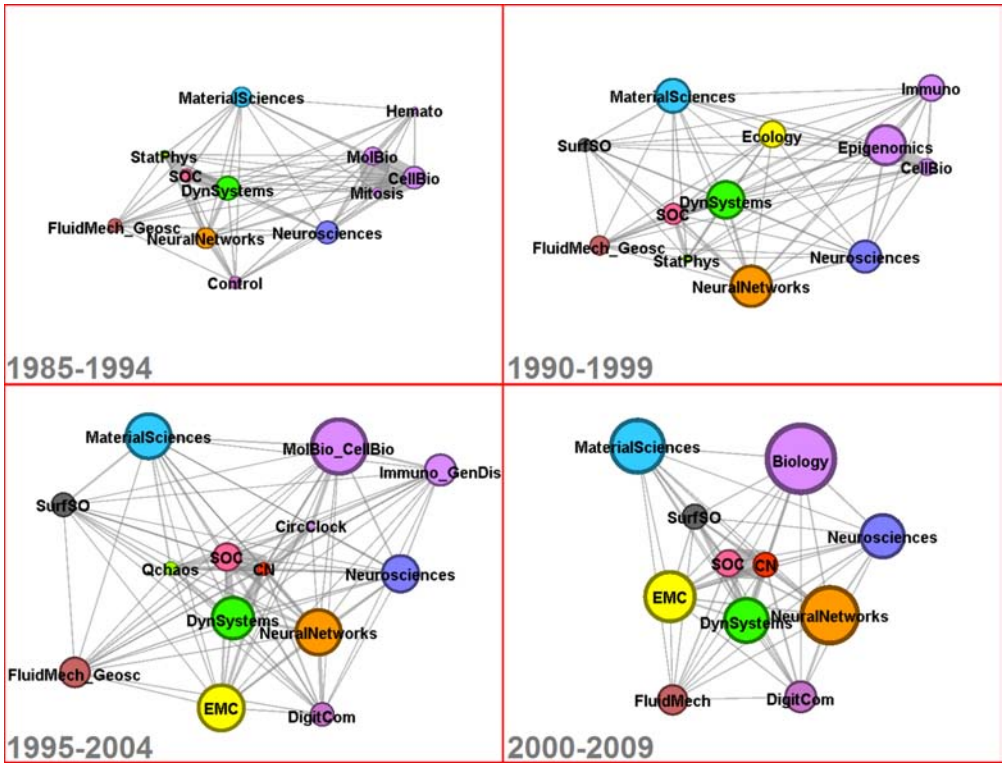

Figure D.3: **Community structure obtained for the four studied decades with a first run of the modularity maximization (Blondel *et al*, 2008).** The surface of a community I is proportional to its number of articles N_I and the width of the link between two communities I and J is proportional to the mean bibliographic coupling $\omega_{IJ} = \sum_{i\in I,\, j\in J} \omega_{ij}/N_I\, N_J$. The layout of the graph is obtained thanks to a spring-based algorithm implemented in the Gephi visualization software (Bastian *et al* , 2009). For the sake of clarity, communities with less than 100 articles (300 for the 2000-2009 decade) are not displayed. Labels are based on a frequency analysis. Refer to Table D.3 for accronyms significations.

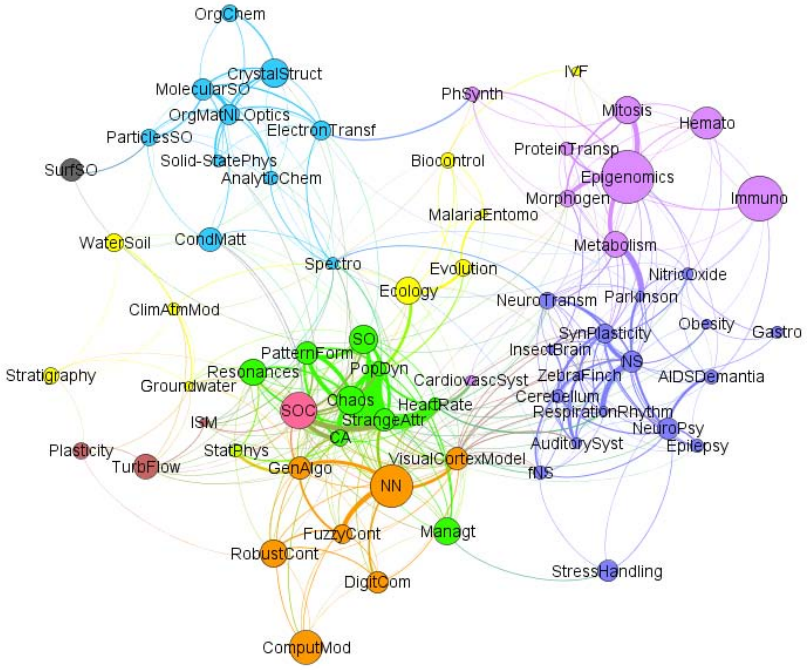

Figure D.4: **1990-1999's subfield community network.** The surface of a community I is proportional to its number of articles N_I and the width of the link between two communities I and J is proportional to the mean bibliographic coupling $\omega_{IJ} = \sum_{i \in I,\, j \in J} \omega_{ij} / N_I\, N_J$. For the sake of clarity, communities with less than 100 articles and links with a mean weight $\omega < 2.10^{-5}$ are not displayed. Labels are based on a frequency analysis.

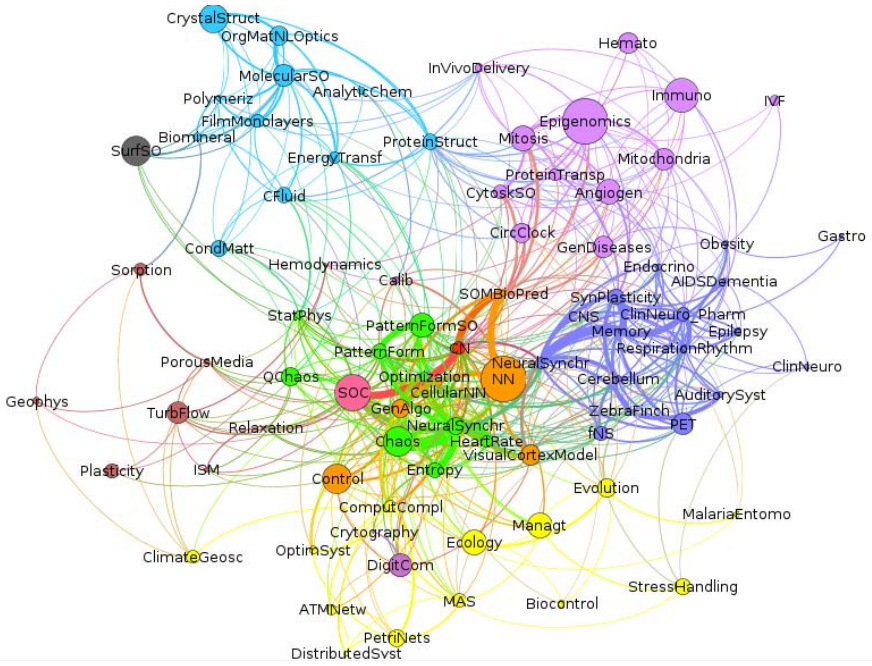

Figure D.5: **1995-2004's subfield community network.** The surface of a community I is proportional to its number of articles N_I and the width of the link between two communities I and J is proportional to the mean bibliographic coupling $\omega_{IJ} = \sum_{i \in I,\, j \in J} \omega_{ij} / N_I\, N_J$. For the sake of clarity, communities with less than 100 articles and links with a mean weight $\omega < 2.10^{-5}$ are not displayed. Labels are based on a frequency analysis.

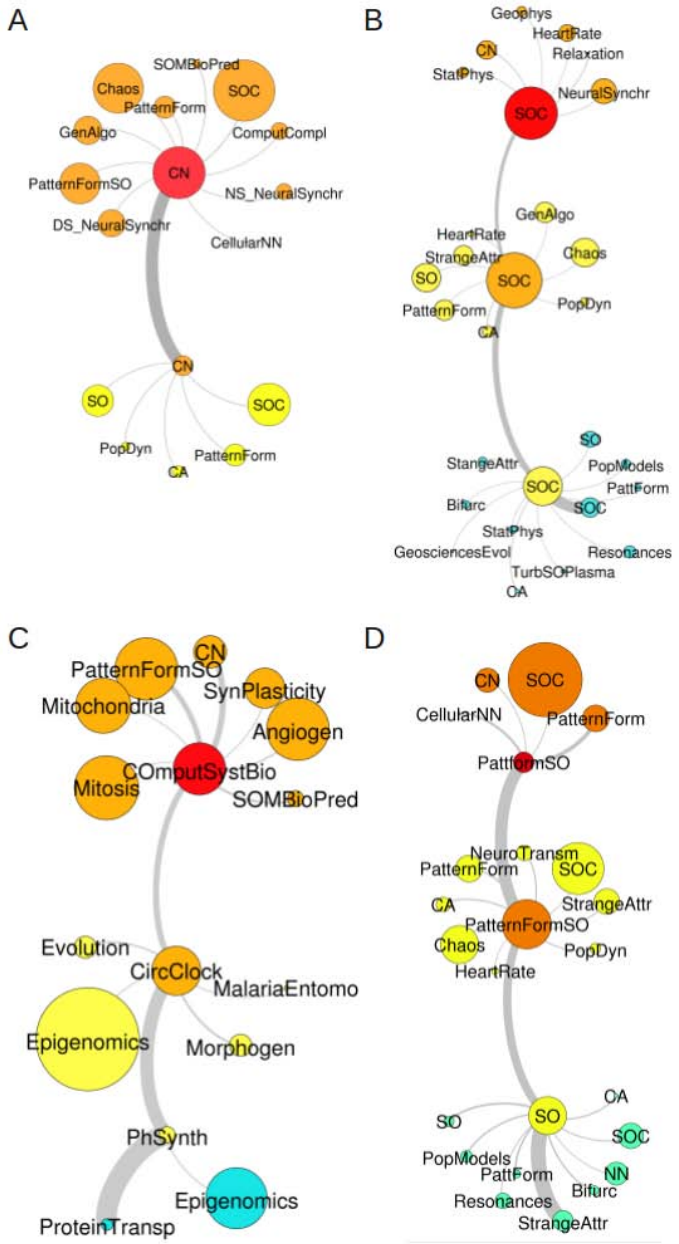

Figure D.6: **Community filiation diagrams.** **A.** *Complex Networks* **B.** *Self-Organized Criticality* **C.** *Computational Systems Biology* **D.** *Pattern Formation & Self-Organisation* The surface of a community I is proportional to its number of articles N_I and the width of the link between two communities I and J is proportional to the mean bibliographic coupling $\omega_{IJ} = \sum_{i \in I,\, j \in J} \omega_{ij} / N_I\, N_J$. Colors corresponds to different chronological periods (from top to bottom: red, orange, yellow and blue for respectfully the 2000-2009, 1995-2004, 1990-1999 and 1985-1994 decades). The procedure leading to these diagrams is explained in the main text. The relative position of the communities are here meaningless.

Abstract

This thesis explores the problems raised by the aggregation of entities into a global, collective level, an old problem encountered in many fields of science. We work on three projects related to the aggregation problem in social systems, using tools derived from statistical physics, and more generally quantitative tools. Great care is taken to tackle the questions at the heart of these projects in a mutually beneficial way for both social and natural sciences.

The first project focus on a paradigmatic model of the emergence of puzzling macroscopic behavior from simple individual rules, Schelling's segregation model. This model simulates the evolution of the spatial repartition of two types of agents living in a virtual city. It is widely known for this paradoxical effect: if the agents have a mild preference for one's neighbors to be of the same kind, their move lead to segregative pattern at the global scale, even if total segregation does not maximize the collective utility. We first use simulations to show that introducing small amount of coordination in the agents' moving decision can significantly reduce segregation. We then propose an analytical resolution of Schelling's model for a wide range of utility functions. Using evolutionary game theory, we provide existence conditions for a potential function which characterizes the global configuration of the city and is maximized in the stationary states. We use this potential function to derive several analytical results. Switching on a physicist point of view, we generalize our potential function in a simplified version of the model which interpolate between cooperative and individual dynamics.

The second project is based on the exploration of huge databases on scientific literature (mostly Web of Science) to investigate the existence and evolution of paradigms or scientific institutions. We mostly use the old but quite unused bibliographic coupling (BC) approach, based on a normalized number of shared references, to measure relations between articles. Using standard techniques to group similar articles, we can define 'natural' communities characterized by their references. Thanks to a large database (141 098 records) of relevant articles, we used this approach to empirically study the 'complex systems' field. We show that the overall coherence of the field does not arise from a universal theory but rather from computational techniques and fruitful adaptations of the idea of self organization to specific systems. We also investigate the idea of 'trading zones', small communities creating an interface between disciplines around specific tools or concepts. We also apply our approach to develop a set of routines allowing to draw different maps of the research carried out in a scientific institution, specifically co-occurrence (of authors, keywords, institutions) maps and BC communities maps. We

use the example of the ENS de Lyon to discuss why these maps may become a valuable tool for institutions' directors.

Finally, the third project deals with the emergence of 'institutions' or 'structures' in social systems. Our collaboration with a team of sociologists has lead us to question the assumption of a clear dichotomy between two 'levels', namely individuals and society. Building on the social theory developed by Gabriel Tarde at the end of the 19th century, we explore different possibility to visualize (and conceptualize) the evolution of social phenomena without making a distinction between two levels. Bibliometric data are used as an example. We also propose an attempt to formalize Tarde's theory in the scope of an algorithmic model. The point is to show how one can obtain 'wholes' through the simplification of complex individuals. While our prototype model fulfil several of Tarde's precept, it raises many more questions. Finally, we focus on a single question raised by our colleagues sociologists: the existence of lasting structure from non lasting entities. We build on the physicist's approach developed in opinion models. While most papers focus on stationary properties, we choose to build a model to investigate the dynamical properties of social structures which are always changing. The key ingredients of our model are the introduction of noise in the agents' interactions, a turnover in the population of agents and a generation effect, the agents taking into account their opinion and age difference in their interaction. The outcomes of our model display a rich phenomenology of group dynamics. Our point is not to produce realistic representations of reality, but much more to help the sociologists enrich their conceptualizations of social phenomena.

Keywords: social systems, collective phenomena, segregation, coordination, complex systems, maps of science, bibliographic coupling, Tarde, opinion model.