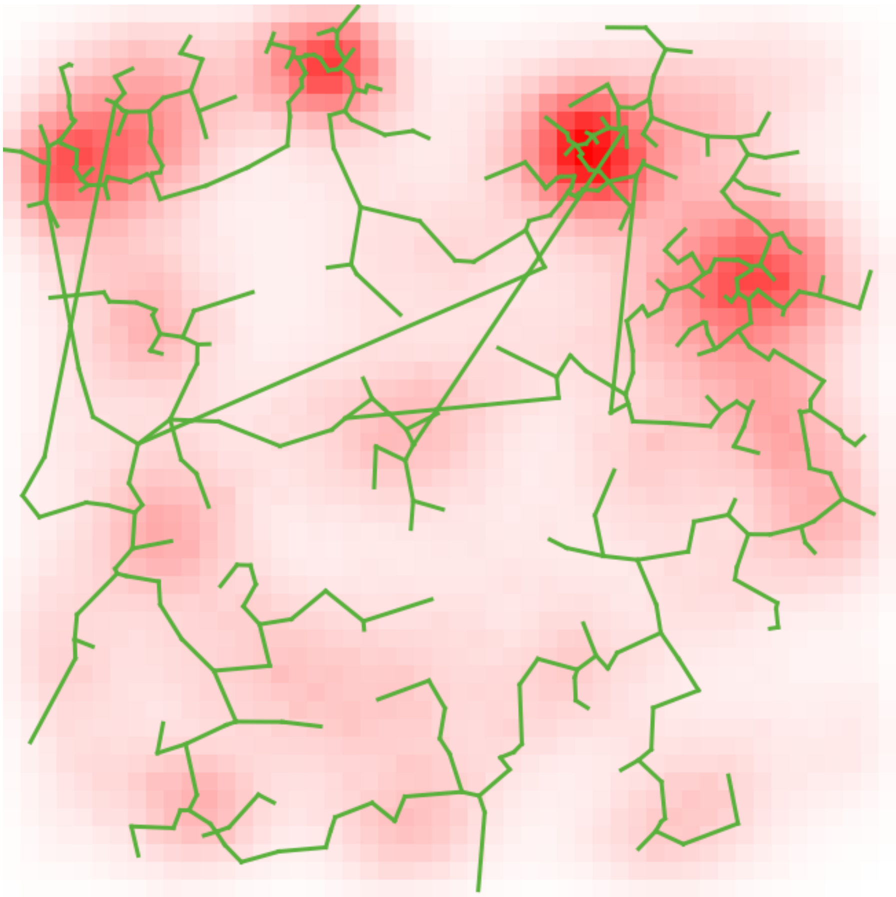
TOWARDS MODELS COUPLING URBAN GROWTH AND TRANSPORTATION NETWORK GROWTH

JUSTE RAIMBAULT



PhD Thesis First Year Preliminary Memoire

March 2016 – version 0.3

[March 8, 2016 at 18:52 – Thesis version 0.3]

Juste Raimbault: *Towards Models Coupling Urban Growth and Transportation Network Growth,* PhD Thesis First Year Preliminary Memoire, © March 2016

ABSTRACT

CONTENTS

LIST OF FIGURES

LIST OF TABLES

INTRODUCTION

It's when you shuffle the anthill that you get a touch of all its complexity.
- Arnaud Banos

"In consequence of a technical issue, traffic is interrupted on the line B of RER, for an unknown duration. More information will be given as soon as available". There is a high probability that someone having lived or spent some time in the metropolitan region of Paris has already heard this frightening announce and endured the difficult consequences the rest of his day. But he might not be aware of the ramifications of causal cascades induced by this not-so-rare event. Territorial Systems, whatever the layers considered in their definitions, will always be extremely complex and interrelations at numerous temporal and spatial scales participate in the emergent behaviors observed at any levels of the system. Martin is a student who daily commutes from Paris to Palaiseau and will miss today a crucial exam, what will have a profound impact on his professional life : implications at a long time scale, small spatial scale and agent granularity. Yuangsi is connecting Orly and Roissy Airports, in his trip from London to Beijing, will miss his plane and his sister's wedding : large spatial scale, short time scale, agent granularity. A collective petition emerges from users, leading to new social organizational patterns and reaction from transportation authority that results in efforts to increase levels of service : mesoscopic temporal and spatial scale, swarm of agents granularity. Looking for causes of the event will also lead to intricate processes at various scales, none of which seems to be a better explication than others : historical railway network in Parisian region shaped further extensions and RER B followed the former *Ligne de Sceaux*, Delouvrier's schema for regional development, and its subsequent partial execution, are elements of explanation of structural weaknesses of Parisian public transportation network [89] ; commuting patterns consequent to territorial organisation induce an overload of particular lines and thus a necessary increase in exploitation incidents. The list could be developed much longer and each approach related to an already mature scientific body of knowledge in different disciplines such as geography, urban economics, transportation. This amusing anecdote is enough to give a touch of the complexity of territorial systems. Our aim here is to dive into this complexity, and in particular to give an original insight into the study of relations between networks and territories. The choice of this reading position will be largely discussed in a further thematic part. Let for now concentrate on the originality of the point of view that we will take.

SCIENTIFIC CONTEXT : COMPLEXITY HAS COME OF AGE

To better introduce our subject, it is necessary to make the reader aware of the particular scientific context we are working in. It is necessary both to understand the general epistemology underlying research questions, and to be aware of the variety of methods and tools used. Contemporaneous science is progressively taking the shift of complexity in many fields. That also implies an epistemological revolution to abandon strict reductionism that failed in most of its synthesis attempts [6]. Arthur recently recalled [9] that a mutation of methods and paradigms was also at stake by the increasing role of computational approaches replacing purely analytical techniques generally self-limited in their modeling and resolution scope. Capturing *emergent properties* in models of complex systems is one of the ways to understand the essence of these new approaches.

These considerations are well known in Social Science (both quantitative and qualitative), in which the complexity of studied agents and systems is the justification of their existence : if humans were particles a whole branch of fields may have never emerged as thermodynamics would have solved most of social issues. [1] They are however less known nor accepted in more "hard" sciences such as physics : Laughlin develops in [118] a view of the discipline at least as at a "frontier of knowledge" then other fields appearing as less mature. Most of knowledge is of classical nature although a majority of structures and systems would be *self-organized*, what means that the single microscopic laws are not enough to determine macroscopic properties unless system evolution is simulated (more precisely this property can be taken as a definition of emergence on which we will come back further, and self-organization is intrinsically emergent). It corresponds to the first nightmare of Laplace's Deamon developed in [67].

As an informal mix of epistemological positions, methods, and fields of applications, *Complexity Science* relies on typical paradigms such as the centrality of emergence and self-organization in most of phenomena of the real world, which make it lie on a frontier of knowledge closer of us than we can think (Laughlin, op.cit.). Such concepts are indeed not new, as they were already enlighten by Anderson [6]. Even cybernetics can be related to complexity by seing it as a bridge between technology and cognitive science [211]. Later, syner-

1 even if it would probably not have been the case as classical physics also failed in their attempts to include irreversibility and evolutions of Complex Adaptive Systems as Prigogine points out in [159]

getics [98] paved the way for a theoretical approach of collective phenomena in physics. Reasons for the recent growth of works claiming a CS approach may be various. The explosion of computing power is surely one because of the central role of numerical simulations [202]. They could also be the related epistemological progresses : apparition of the notion of perspectivism [86], finer reflexions around the notion of model [203][2]. The theoretical and empirical potentialities of such approaches play surely a role in their success[3], as confirmed in various domains of application (see [145] for a general survey), as for example Network Science [12] ; Neuroscience [112]; Social Sciences ; Geography [138][161] ; Finance with the rising importance of econophysics [194] ; Ecology [94]. The Complex Systems Roadmap [37] proposed a double lecture of studies on Complex Systems : an horizontal approach connecting fields of study with transversal questions on theoretical foundations of complexity and empirical common stylized facts, and a vertical conceptions of disciplines, with the aim to construct integrated disciplines and corresponding multi-scale heterogeneous models. Interdisciplinarity is thus central in our scientific background.

INTERDISCIPLINARITY

We must further insist on the role of interdisciplinarity in the positions taken here. This is not a thesis in Geography nor in Complex Adaptive Systems Modeling, but in *Complex Systems Science* that we claim as a proper discipline following Paul Bourgine. It will naturally be seen with defiance by scholars of various concerned disciplines, as recent examples of misunderstandings and conflicts have illustrated [77]. The positioning of Batty proposing *A new Science of Cities* [20] (that he subtly presents as *The* new science of cities) is directed towards an integration of disciplines and methods into a science defined by its object of study, cities. Its theoretical and epistemological weaknesses (no theoretical constructions of studied geographical objects on the one hand, approximative contextualization of complexity) combined with an overall impression of *pot-pourri* of forgotten works (space syntax, land-use models), unfortunately avoid us to use it as we will use geographical theories (e.g. evolutive urban theory) in an appropriated epistemological complexity context. Yet our reading of this work may be the result of a misunderstanding due to different cultural backgrounds.

The scientific evolution of complexity that some see as a revolution [56], or even as *a new kind of science* [213], could indeed face in-

2 In that frame scientific and epistemological progress can not be dissociated and can be seen as coevolving

3 Although the adoption of new scientific practices may be strongly biased by imitation and lack of originality [71], or more ambivalent, by marketing strategies as the fight for funds is becoming a huge obstacle for research [33].

trinsic difficulties due to behaviors and a-priori of researchers as human beings. More precisely, the need of interdisciplinarity that makes the strength of Complexity Science may be one of its greatest weaknesses, since the highly partitioned structure of science organization has sometimes negative impacts on works involving different disciplines. We do not tackle the issue of over-publication, competition, indexes, which is more linked to a question of open science and its ethics, also of high importance but of an other nature. That barrier we are dealing with and we might struggle to triumph of, is the impact of certains *cultural disciplinary differences* and out-coming conflicts on views and approaches. The drama of scientific misunderstandings is that they can indeed annihilate progresses by interpreting as a falsification some work that answers to a totally different question. The example of a recent work on top-income inequalities given in [4], which conclusions are presented as opposed from the one obtained by Piketty [158], follows such a scheme. Whereas Piketty focused on constructing long-time clean databases for income data and showed empirically a recent acceleration of income inequalities, his simple model aiming to link this stylized fact with the accumulation of capital has been criticized as oversimplified. On the other hand, Bergeaud *et al.* prove by a model of innovation economics that *under certain assumptions* income gaps may be beneficial to innovation and consequently a general utility. Thus diverging conclusions about the role of personal capitals in the economy. But diverging *views* or *interpretations* does not mean a scientific incompatibility, and one could imagine try to gather both approaches in an unified framework and model, yielding possibly similar or different interpretations. This integrated approach has chances to contain more information (depending on how coupling is done) and to be a further advance in Science. We shall now briefly develop other examples to give an overview how conflicts between disciplines can be damaging.

PHYSICS REINVENTS GEOGRAPHY. As already mentioned, DUPUY and BENGUIGUI points out in [77] the fact that urban sciences have recently known open conflicts between old tenants of the disciplines and new arrivants, especially physicists. The availability of large datasets of new type of data (social networks, ICT data) have drawn their attention towards the study of objects traditionally studied by human science, as analytical and computational methods of statistical physics became applicable. Although these studies are generally presented as the construction of a scientific approach to cities, implying that existing knowledge was not scientific because of their more qualitative aspect, they have not unveil specifically novel knowledge on urban systems : to give some examples, [16] concludes that Paris has followed a transition during Haussman period and that the evolution of a city is the combination of local transformations and global planning operations, what are facts known for a long time in urban history and urban geography. [50] rediscovers that the gravity model can be improved by adding lags in interactions and theoretically derives the expression of the force of interaction between cities, without any thematic theoretical background. Examples could be multiplied, confirming the current discomfort in communication between physicists and urban geographers. Significant benefices could results from a wise integration of disciplines [148] but the road seems still long.

ECONOMIC GEOGRAPHY OR GEOGRAPHICAL ECONOMICS ? Similar conflict occurred in economics : as [140] describes, the discipline of economic geography, traditionally close from geography, heavily criticized a new stream of thought named *geographical economics*, which purposes is spatialization of mainstream economic techniques. Both do not have the same purposes and aims, and the conflict appears as a total misunderstanding for an external observer.

AGENT-BASED MODELING IN ECONOMY Disciplinary conflicts may also manifest themselves as the reject of novel methods by mainstream currents. Following [80], the operational failure of most classic economic approaches could be compensated by a broader use of agent-based modeling and simulation practices. The lack of analytical framework that is natural in the study of complex adaptive systems seems to be rebutting for most of economists.

FINANCE In Quantitative Finance coexist various stream of research with a very few interactions. Let consider two examples. On the one hand, Statistics are highly advanced in theoretical mathematics, using stochastic calculus and probabilities to obtain very refined estimators of parameters for a given model (see e.g. [13]). On the other hand, Econophysics aims to study empirical stylized facts and infer empirical laws to explain complexity-related phenomena in financial systems [194], such as e.g. cascades leading to market crashes, fractal properties of asset signals, complex structure of correlation networks. Both have their advantages in a particular context and each would benefit from increased interactions between the fields.

These diverse examples illustrate how interdisciplinarity is both crucial and difficult to achieve. We will try to follow that narrow path in our work, borrowing ideas, theories and methods from various disciplines, aiming for the construction of an integrated knowledge. Indeed, coupling heterogeneous approaches at different levels and scales will be a cornerstone of our thesis, skeleton of the underlying philosophy and building brick of the theory we will propose.

COMPLEXITY IN GEOGRAPHY

Coming back to our introducing anecdote, we will focus on our thematic object of study that are territorial systems. More generally, we propose an overview of the role of complexity in geography. Geographers are familiar with complexity for a long time, as the study of spatial interactions is one of its purposes. The variety of fields in geography (geomorphology, physical geography, environmental geography, human geography, health geography to give a few) has certainly been important in the subtlety of the geographical thinking, that considers heterogeneous and multi-scalar processes.

PUMAIN recalls in [162] a subjective history of the emergence of complexity paradigms in geography. Cybernetics yielded system theories as the one developed by Forrester. Later the shift to self-organized criticality and self-organisation concepts in physics conducted to corresponding developments in geography, as [184] witnesses the application of the concepts of synergetics for the dynamics of an urban system. Finally, Complex Systems paradigms as we currently know them appeared from various points of view. For example, the fractal nature of urban shape was introduced in [21] and had numerous application including more recent developments [110]. BATTY also introduced cellular automata in urban modeling and proposed a joint synthesis with agent-based modeling and fractals in [19]. An other incursion of complexity in geography was for the case of urban systems through the evolutive urban theory of PUMAIN. In close relation with modeling from the beginning (the first Simpop model described in [185] enters the theoretical framework of [161]), this theory aims to understand system of cities as systems of co-evolving adaptive agents, interacting in many ways, with particular features emphasized such as the diffusion of innovation. The series of Simpop models [165] focused in testing various assumptions of the theory. For example, different underlying mechanisms were revealed for european city systems and city system of the united states [41]. At other time scales and in other contexts, the SimpopLocal model [189] aimed to investigate the conditions for the emergence of hierarchical urban systems from disparate settlements. A minimal model (in the sense of sufficient and necessary parameter) has been isolated thanks to the use of intensive computation with the model exploration software OpenMole [190], what was a result analytically not derivable for this kind of complex model. The technical progresses of OpenMole [176] were done simultaneously with theoretical and empirical advancements. Epistemological advances were also essential to this framework, as REY develops in [177], and novel concepts such as incremental modeling [60] were found, with powerful concrete applications : [58] implemented it on the soviet city system and isolated dominating socio-economic processes, by systematic testing of thematic assumptions and implementation functions. Directions for the development of such Modeling and Simulation practices in quantitative geography were recently introduced by Banos in [10]. He concludes with nine principles[4], from which we can cite the importance of intensive exploration of computational models and the importance of heterogeneous model coupling, that are among other principles such as reproducibility at the center of the study of complex geographical systems from the point of view described just before. Positioning in the legacy of this line of research, we will conjointly work in the theoretical, empirical, epistemological and modeling domains.

research question

Research question and precise objects are deliberately fuzzy for now, as we postulate that the construction of a problematic can not be dissociated from the production of a corresponding theory. Reciprocally, it makes no sense to ask questions out of the blue, on objects that have been only partially or rapidly defined. Our preliminary question to enter the subject, that we can obtain from concrete cases such as our introducing anecdote or from preliminary literature review, is the following :

Is it possible to produce a definition of territorial systems, and corresponding scales and ontologies, that would yield a natural, consistent and informational view on processes ?

Indeed, a necessary characteristic of territorial systems is their spatio-temporal nature, that is contained in spatio-temporal dynamics. The notion of *process* in the sense of [106] captures furthermore causal relationships in these dynamics, and is thus an interesting approach for an understanding of such systems. *Scale* must be understood here in the operational sense (physical characteristic) and *ontology* as real-world studied objects[5]. Our question may be roughly viewed as a search for theories and models that would unveil some processes involved in complex systems containing at least human settlements, the last requirement being crucial for a convergent problematic construction rather than ending in non-realistic and non-constructive propositions to understand everything between the brain (that can be seen as one building brick of territorial systems as they emerge from hu-

4 I remember René Doursat insisting on the search of the last commandement of Banos

5 this use of ontology here naturally biaises our research towards modeling paradigms as it is close from the notion of ontology used in [128], but we take the position (largely developed further) to understand any scientific construction as *models*, making the frontier between theory and modeling less relevant than in standard views. Any theory has to make choices on described objects, relations and processes, and therefore contains an ontology in that sense.

man social constructions) and the ecosphere that includes territorial systems.

CONTENTS

This provisory Memoire is organized the following way. A first part with four chapters sets the thematic, theoretical and methodological background. The study of geographical systems implies, because of their complexity, a subtle combination of Theoretical constructions and Empirical Analysis, either in an inductive reasoning or in a didactic constitution of knowledge. The first part aims to approach our subject from the theoretical and methodological point of view, and rather as a *necessary foundation* shall be understood as a body of knowledge *coevolving* with Empirical and Modeling Parts. A linear reading is not necessarily the best way to deeply perceive the implications of theory on empirical and modeling experiments and reciprocally. Some methodological developments are necessary but explicit reference will be done when it will be the case. A first chapter starts from the provisory research question given above and frames from a thematic point of view geographical objects and processes to be studied, resulting in precise research questions. The scene is set up for the construction of our theoretical background in a second chapter, that consists in a geographical theory for territorial systems on the one hand and in an epistemological theory of socio-technical systems modeling that frames our approach at a meta-level. We then develop methodological considerations on diverse questions implied by theory and required for modeling. Finally, a chapter of quantitative epistemology finishes to pave the way for modeling directions, unveiling literature gaps precisely linked to our question. A second part develops results obtained from empirical analysis and modeling experiments, along with on-going and planned projects in these fields. It first present empirical analysis aimed at identifying stylized facts. Toy-models of urban growth are then proposed, followed by an example and propositions for more complex models. The third part constructs our research objective for the remaining part of our project and sets a corresponding roadmap. Appendices contain non-digest important parts of our work such as models implementation architecture and details and specific tools developed for a reproducible research workflow.

Part I

THEMATIC, THEORETICAL AND METHODOLOGICAL FOUNDATIONS

1 INTERACTIONS BETWEEN NETWORKS AND TERRITORIES

> *Si la question de la priorité de l'œuf sur la poule ou de la poule sur l'œuf vous embarrasse, c'est que vous supposez que les animaux ont été originairement ce qu'ils sont à présent.*
>
> - Denis Diderot [70]

This analogy is ideal to evoke the questions of causality and processes in territorial systems. When trying to tackle naively our preliminary question, some observers have qualified the identification of causalities in complex systems as "chicken and egg" problems : if one effect appears to cause another and reciprocally, how can one disentangle effective processes ? This vision is often present in reductionist approaches that do not postulate an intrinsic complexity in studied systems. The idea that Diderot suggests is the notion of *co-evolution* that is a central phenomenon in evolutive dynamics of Complex Adaptive Systems as Holland develops in [104]. He links the notion of emergence (that is ignored in a reductionist vision), in particular the emergence of structures at an upper scales from the interactions between agents at a given scale, materialized generally by boundaries, that become crucial in the coevolution of agents at any scales : the emergence of one structure will be simultaneous with one other, each exploiting their interrelations and generated environments conditioned by their boundaries. We shall explore these ideas in the case of territorial systems in the following.

This introductive chapter aims to set up the thematic scene, the geographical context in which further developments will root. It is not supposed to be understood as an exhaustive literature review nor the fundamental theoretical basement of our work (the first will be an object of chapter 4 whereas the second will be earlier tackled in chapter 2), but more as narration aimed to introduce typical objects and views and construct naturally research questions.

1.1 TERRITORIES AND NETWORKS

1.1.1 *Territories and Networks : There and Back Again*

HUMAN TERRITORIES The notion of territory can be taken as a basis to explore the scope of geographical objects we will study. In Ecology, a territory corresponds to a spatial extent occupied by a group of agents or more generally an ecosystem. *Human Territories* are far more complex in the sense of semiotic representations of these that are a central part in the emergence of societies. For RAFFESTIN in [170], the so-called *Human Territoriality* is the "conjonction of a territorial process with an informational process", what means that physical occupation and exploitation of space by human societies is not dissociable from the representations (cognitive and material) of these territorial processes, driving in return its further evolutions. In other words, as soon as social constructions are assumed in the constitution of human settlements, concrete and abstract social structures will play a role in the evolution of the territorial system, through e.g. propagation of information and representations, political processes, conjonction or disjonction between lived and perceived territory. Although this approach does not explicitly give the condition for the emergence of a seminal system of aggregated settlements (i.e. the emergence of cities), it insists on the role of these that become places of power and of creation of wealth through exchange. But the city has no existence without its hinterland and the territorial system can not be summarized by its cities as a system of cities. There is however compatibility on this subsystem between RAFFESTIN approach to territories and PUMAIN's evolutive theory of urban systems [164], in which cities are viewed as an auto-organized complex dynamical systems, and act as mediators of social changes : for example, cycles of innovation occur within cities and propagate between them. Cities are thus competitive agents that co-evolve (in the sense given before). The territorial system can be understood as a spatially organized social structure, including its concrete and abstract artifacts. A imaginary free-of-man spatial extent with potential ressources will not be a territory if not inhabited, imagined, lived, and exploited, even if the same ressources would be part of the corresponding habited territorial system. Indeed, what is considered as a ressource (natural or artificial) will depend on the corresponding society (e.g. of its practices and technological potentialities). A crucial aspect of human settlements that were studied in geography for a long time, and that relate with the previous notion of territory, are *networks*. Let see how we can switch from one to the other and how their definition may be indissociable.

A TERRITORIAL THEORY OF NETWORKS We paraphrase DUPUY in [76] when he proposes elements for "a territorial theory of networks" based on the concrete case of Urban Transportation Networks. This theory sees *real networks* (i.e. concrete networks, including transportation networks) as the materialization of *virtual networks*. More precisely, a territory is characterized by strong spatio-temporal discontinuities induced by the non-uniform distribution of agents and ressources. These discontinuities naturally induce a network of "transactional projects" that can be understood as potential interactions between elements of the territorial system (agents and/or ressources). For example today, people need to access the ressource of employments, economic exchanges operate between specialized production territories. At any time period, potential interactions existed[1]. The potential interaction network is concretized as offer adapts to demand, and results of the combination of economic and geographical constraints with demand patterns, in a non-linear way through agents designed as *operators*. This process is not immediate, leading to strong non-stationarity and path-dependance effects : the extension of an existing network will depend on previous configuration, and depending on involved time scales, the logic and even the nature of operators may have evolved. Raffestin points out in his preface of [150] that a geographical theory articulating space, network and territories had never been consistently formulated. It appears to still be the case today, but the theory developed just before is a good candidate, even if it stays at a conceptual level. The presence of a human territory necessarily imply the presence of abstract interaction networks and concrete networks used for transportation of people and ressources (including communication networks as information is a crucial ressource). Depending on regime in which the considered system is, the respective role of different networks may be radically different. Following Duranton in [78], pre-industrial cities were limited in growth because of limitations of transportation networks. Technological progresses have lead to the end of these limitations and the preponderance of land markets in shaping cities (and thus a role of transportation network as shaping prices through accessibility), and recently to the rising importance of telecommunication networks that induce a "tyranny of proximity" as physical presence is not replaceable by virtual communication. This territorial approach to networks seems natural in geography, since networks are studied conjointly with geographical objects with an underlying theory, in opposition to network science that studies brutally spatial networks with few thematic background [75].

NETWORKS SHAPING TERRITORIES ? However networks are not only a material manifestation of territorial processes, but play their

1 even when nomadism was still the rule, spatially dynamic networks of potential interactions necessarily existed, but should have less chance to materialize into concrete routes.

part in these processes as they evolution may shape territories in return. In the case of *technical networks*, an other designation of real networks given in [150], many examples of such feedbacks can be found : the interconnectivity of transportation networks allows multi-scalar mobility patterns, thus shaping the lived territory. At a smaller scale, changes in accessibility may result in an adaptation of a functional urban space. Here emerges again an intrinsic difficulty : it is far from evident to attribute territorial mutations to some network evolutions and reciprocally materialization of a network to precise territorial dynamics. Coming back to Diderot should help, in the sense that one must not consider network nor territories as independent systems that would have causal relationships but as strongly coupled components of a larger system. The confusion on possible simple causal relationships has fed a scientific debate that is still active. Methodologies to identify so-called *structural effects* of transportation networks were proposed by planners in the seventies [35, 36]. It took some time for a critical positioning on unreasoned and decontextualized use of these methods by planners and politics generally to technocratically justify transportation projects, that was first done by Offner in [149]. Recently the special issue [123] on that debate recalled that on the one hand misconceptions and misuses were still greatly present in operational and planning milieus as [64] confirmed, and on the other hand that a lot of scientific progresses still need to be made to understand relations between networks and territories as Pumain highlights that recent works gave evidence of systematic effects on very long time scales (as e.g. the work of Bretagnolle on railway evolution, that shows a kind of structural effect in the necessity of connectivity to the network for cities to "stay in the game", but that is not fully causal as not sufficient). At a macroscopic level typical patterns of interaction emerge, but microscopic trajectories of the system are essentially chaotic : the understanding of coupled dynamics strongly depends on the scale considered. At a small scale it seems indeed impossible to show systematic behavior, as Offner pointed out. For example, on comparable French mountain territories, [27] shows that reactions to a same context of evolution of the transportation network can lead to very different reactions of territories, some finding a huge benefit in the new connectivity, whereas others become more closed. These potential retroactions of networks on territories does not necessarily act on concrete components : Claval shows in [55] that transportation and communication networks contribute to the collective representation of territories by acting on territorial belonging feeling.

TERRITORIAL SYSTEMS This detour from territories, to networks and back again, allows us to give a preliminary definition of a territorial system that will be the basis of our following theoretical considerations. As we emphasized the role of networks, the definition takes it into account.

Preliminary Definition. *A territorial system is a human territory to which both interaction and real networks can be associated. Real networks are a component of the system, involved in evolution processes, through multiples feedbacks with other components at various spatial and temporal scales.*

This reading of territorial systems is conditional to the existence of networks and may discard some human territories, but it is a deliberate choice that we justify by previous considerations, and that drives our subject towards the study of interactions between networks and territories.

1.1.2 *Transportation Networks*

THE PARTICULARITY OF TRANSPORTATION NETWORKS Already evoked in relation to the question of structural effects of networks, transportation networks play a determining role in the evolution of territories. Although other types of networks are also strongly involved in the evolution of territorial systems (see e.g. the discussions of impacts of communication networks on economic activities), transportation networks shape many other networks (logistics, commercial exchanges, social concrete interactions to give a few) and are prominent in territorial evolution patterns, especially in our recent societies that has become dependent of transportation networks [22]. The development of French High Speed Rail network is a good illustration of the impact of transportation networks on territorial development policies. Presented as a new era of railway transportation, a top-down planning of totally novel lines was introduced as central for developments [220]. The lack of integration of these new networks with existing ones and with local territories is now observed as a structural weakness and negative impacts on some territories have been shown [221]. A review done in [23] confirms that no general conclusions on local effects of High Speed lines connection can be drawn although it keeps a strong place in imaginaries. These are examples of how transportation networks have both direct and indirect impacts on territorial dynamics. Integrated planning, in the sense of a joint planning of transportation infrasctures and urban development, considers the network as a determining component of the territorial system. Parisian *Villes Nouvelles* are such a case, that witnesses of the complexity of such planning actions that generally do not lead to the desired effect [154]. Recent projects as [116] have try to implement similar ideas but we have now not enough temporal scope to judge their success in effectively producing an integrated territory. Transportation networks are anyway at the center of these approaches of urban territories. We will focus in our work on transportation networks for the various reasons given here.

DECONSTRUCTING ACCESSIBILITY The notion of accessibility comes rapidly when considering transportation networks. Based on the possibility to access a place through a transportation network (including transportation speed, difficulty of travel), it is generally described as a potential of spatial interaction[2] [22]. This object is often used as a planning tool or as an explicative variable of agents localisation for example. We must warn here on the potential dangers of its unconditional use. More precisely, it may be a construction that misses a consistent part of territorial dynamics. The mystification of the notion of *mobility* was shown by COMMENGES in [57], which proved than most of debates on modeling mobility and corresponding notions were mostly made-of by transportation administrators of *Corps des Ponts* who roughly imported ideas from the United States without adaptation and reflexion fit to the totally different French context. Accessibility may be such a social construct and have no theoretical root since it is mostly a modeling and planning tool. Recent debates on the planification of *Grand Paris Express* [137], a totally novel metropolitan transportation infrastructure planned to be built in the next twenty years, have revealed the opposition between a vision of accessibility as a right for disadvantaged territories against accessibility as a driver of economic development for already dynamic areas, both being difficultly compatible since corresponding to very different transportation corridors. Such operational issues confirm the complexity of the role of transportation networks in the dynamics of territorial systems, and we shall give in our work elements of response to a definition of accessibility that would integrate intrinsic territorial dynamics.

SCALES AND HIERARCHIES An incontournable aspect of transportation networks that we will need to take into account in further developments is hierarchy. Transportation networks are by essence hierarchical, depending on scales they are embedded in. [133] showed empirical scaling properties for public transportation networks for a consequent number of metropolitan areas across the world, and scaling laws reveal the presence of hierarchy within a system, as for size hierarchy for system of cities expressed by Zipf's law [146] or other urban scaling laws[8, 29]. Transportation network topology has been shown to exhibit such scaling also for the distribution of its local measures such as centrality [183]. Hierarchy seems to play a particular role on interaction processes, as BRETAGNOLLE [39] highlighted an increasing correlation between urban hierarchy and network hierarchy

2 and often generalized as *functional accessibility*, for example employments accessible for actives at a location. Spatial interaction potentials ruling gravity law can also been understood this way.

for French railway network, marker of positive feedbacks between urban rank and network centralities. Different regimes in space and times were identified : for French railway network evolution e.g., a first phase of adaptation of the network to the existing urban configuration was followed by a phase of co-evolution i.e. in the sense that causal relations become difficult to identify. Railway evolution in the United States followed a different pattern, without hierarchical diffusion, shaping locally urban growth. It emphasizes the presence of path-dependance for trajectories of urban systems : the presence in France of a previous city system and network (postal roads) strongly shaped railway development, whereas its absence in the US lead to completely different dynamics. An open question is if generic processes underlie both evolutions, each being different realizations with different initial conditions and different meta-parameters (different *regimes* in the sense of settlement systems transitions introduced in the current ANR Research project TransMonDyn, as a transition can be understood as a change of stationarity for meta-parameters of a general dynamic). In terms of dynamical systems formulation, it is equivalent to ask if dynamics of attractors (long time scale components) obey similar equations as the position and nature of attractors for a stochastic dynamical system give its current regime, in particular if it is in a divergent state (positive local Liapounov exponent) or is converging towards stable mechanisms [184]. To answer this question together with a disentangling of co-evolution processes for that regime, [39] proposes modeling as a constructive element of answer. We will see in next section how modeling can bring knowledge about territorial processes.

INTERACTIONS BETWEEN TRANSPORTATION NETWORKS AND TERRITORY At this state of progress, we have naturally identified a research subject that seems to take a significant place in the complexity of territorial systems, that is the study of interactions between transportation networks and territories. In the frame of our preliminary definition of a territorial system, this question can be reformulated as the study of networked territorial systems with an emphasize on the role of transportation networks in system evolution processes.

1.2 MODELING INTERACTIONS

1.2.1 *Modeling in Quantitative Geography*

Modeling in Theoretical and Quantitative Geography (TQG), and more generally in Social Science, has a long history on which we can not go further than a general context. Cuyala does in [65] an analysis of the spatio-temporal development of French speaking TQG movement and underlines the emergence of the discipline as the combination between quantitative analysis (e.g. spatial analysis or modeling and simulation practices) and theoretical constructions, an integration of both allowing the construction of theories from empirical facts that yield theoretical hypothesis to be tested on empirical data. These approach were born under the influence of the *new geography* in Anglo-saxon countries and Sweden. A broad history of the genesis of models of simulation in geography is done by Rey in [177] with a particular emphasis on the notion of validation of models. The use of computation for simulation of models is anterior to the introduction of paradigms of complexity, coming back to Hägerstrand and Forrester, pioneers of spatial economic models inspired by Cybernetics. With the increase of computational possibilities epistemological transformations have also occurred, with the apparition of explicative models as experimental tools. Rey compares the dynamism of seventies when computation centers were opened to geographers to the democratization of High Performance Computing (transparent grid computing, see [190] for an exemple of the possibilities offered in terms of model validation and calibration, decreasing the computational time from 30 years to one week), that is also accompanied by an evolution of modeling practices [10] and techniques [52]. Modeling (in particular computational models of simulation) is seen by many as a fundamental building brick of knowledge : [128] recalls the combination of empirical, conceptual (theoretical) and modeling domains with constructive feedbacks between each. A model can be an exploration tool to test assumptions, an empirical tool to validate a theory against datasets, an explicative tool to reveal causalities (and thus internal processes of a system), a constructive tool to iteratively build a theory with an iterative construction of an associated model. These are example among others : Varenne proposes in [202] a refined classifications of diverse functions of a model. We will consider modeling as a fundamental instrument of knowledge on processes within complex adaptive systems, as already evoked, and restraining again our question, will focus on *models involving interactions between transportation networks and territories*.

1.2.2 *Modeling Territories and Networks*

Concerning our precise question of interactions between transportation networks and territories, we propose an overview of existing approaches. Following [40], the "*thoughts of specialists in planning aimed to give definitions of city systems, since 1830, are closely linked to the historical transformations of communication networks*". It implies that ontologies and corresponding models addressed by geographers and planners are closely linked to their current historical preoccupations, thus necessarily limited in scope and purpose. In a perspectivist vision of science [86] such boundaries are the essence of the scientific entreprise, and as we will argue in chapter 2 their combination and coupling in the case of models is a source of knowledge.

Land-Use Transportation Interaction Models

A subsequent bunch of literature in modeling interaction between networks and territories can be found in the field of planning, with the so-called *Land-use Transportation Interaction Models*. These works are difficult to be precisely bounded as they may be influenced by various disciplines. For example, from the point of view of Urban Economics, propositions for synthesizing models have existed for a relatively long term [168]. The variety of possible models has lead to operational comparisons [156, 207]. More recently, the respective advantages of static and dynamic modeling was investigated in [114]. Generally these type of models operate at relatively small temporal and spatial scales. [206] reviewed state of the art in empirical and modeling studies on interactions between land-use and transportation. It is positioned in economic, planning and sociological theoretical contexts, and is relatively far from our geographical approach aiming to understand long-time processes. Seventeen models are compared and classified, none of which implements actually network endogenous evolution on the relatively small time scales of simulation. A complementary review done in [47] broadens the scope with inclusion of more general classes of models, such as spatial interaction models (including traffic assignment and four steps models), operational research planning models (optimal localisations), micro-based random utility models, and urban market models. These techniques operate also at small scales and consider at most land-use evolution. [107] covers a similar scope with a further emphasis on cellular automata models of land-use change and agent-based models. These type of models are still largely developed and used today, as for example [69] which is used for Parisian metropolitan region. The short-term range of application and their operational character makes them useful for planning, what is far from our preoccupation to obtain explicative models for geographical processes.

Network Growth

Network growth can be used to design modeling entreprises that aim to endogenously explain growth of transportation networks, generally from a bottom-up point of view, i.e. by exhibiting local rules that would allow to reproduce network growth over long time scales (generally the road network). Economists have proposed such models : [222] reviews transportation economics literature on network growth within an endogenous growth theory [3], recalling the three main features studied by economists on that subject that are road pricing, infrastructure investment and ownership regime, and describes an analytical model combining the three. [215] develops a broad review on network growth modeling extending to other fields : transportation geography early developed empirical-based models but which did concentrate on topology reproduction rather than on mechanisms according to [215] ; statistical models on case studies provide mitigated conclusions on causal relations between offer and demand ; economists have studied infrastructure provision from both microscopic and macroscopic point of views, generally non-spatial ; network science has provided toy-models of network growth based on structural and topological rules rather on mechanism-based rules. An other approach not mentioned that we will develop further is biologically inspired network design. We first give some example of economic-based and geometrical-based network growth modeling attempts. [219] shows through a reinforcement economic model including investment rule based on traffic assignment that local rules are enough to make hierarchy of roads emerge for a fixed land-use. A very similar model in [132] with simpler cost-benefits obtains the same conclusion. Whereas these models based on processes focus on reproducing macroscopic patterns of networks (typically scaling), geometrical optimization models aim to ressemble topologically real networks. [14] proposes a model based on local energy optimization but it stays very abstract and unvalidated. The morphogenesis model given in [62] based on local potential and connectivity rules, even if not calibrated, seems to reproduce more reasonably real street patterns. Very close work is done in [182]. Other tentatives [66, 218] are closer to procedural modeling [122, 205] and therefore not of interest in our purpose as they can difficultly be used as explicative models. Finally, an interesting and original approach to network growth are biological networks. These belong to the field of morphogenetic engineering pioneered by Doursat that aim to design artificial complex system inspired from natural complex systems and in which a control of emerging properties is possible [73]. *Physarum Machines*, that are models of a self-organized mould (slime mould) have been shown to provide efficient bottom-up solution to computationally heavy problems such as routing problems [197] or NP-complete navigation problems such as the Travelling Salesman Problem [223].

It has been shown to produce networks with Pareto-efficient cost-robustness properties [198], relatively close in shape to real networks (under certain conditions, see [2]). This type of models can be of interest for us since auto-reinforcement mechanisms based on flows are analog to mechanisms of link reinforcement in transportation economics.

Hybrid Modeling

Models of simulation implementing a coupled dynamic between urban growth and transportation network growth are relatively rare, and always rather poor from a theoretical and thematic point of view. A generalization of the geometrical local optimization model described before was developed in [15]. As for the road growth model of which it is an extension, no thematic nor theoretical justification of local mechanisms is provided, and the model is furthermore not explored and no geographical knowledge can be drawn from it. [124] adopts a more interesting economic approach, similar to a four step model (gravity-based origin-destination flows generation, stochastic user equilibrium traffic assignment) including travel cost and congestion, coupled with a road investment module simulating toll revenues for constructing agents, and a land-use evolution module updating actives and employments through discrete choice modeling. The experiments showed that co-evolving network and land uses lead to positive feedbacks reinforcing hierarchy, but are far from satisfying for two reasons : first network topology does not really evolve as only capacities and flows change within the network, what means that more complex mechanisms on longer time scales are not taken into account, and secondly the conclusions are very limited as model behavior is not known since sensitivity analysis is done on few one-dimensional spaces : exhaustive mechanisms stay thus unrevealed as only particular cases are described in the sensitivity analysis. From an other point of view, [126] is also presented as a model of co-evolution, but corresponds more to coupled statistical analysis as it relies on a Markov-chain predictive model. [181] gives a model in which coupling between land-use and network growth is done in a weak paradigm, land-use and accessibility having no feedback on network topology evolution. [1] describes a co-evolution model at a very small scale (scale of the building), in which evolution of both network and buildings are ruled by a same agent (influenced differently by network topology and population density) what implies a too strong simplification of underlying processes. Finally, a simple hybrid model explored and applied to a toy planning example in [174], relies on urban activities accessibility mechanisms for settlement growth with a network adapting to urban shape. The rules for network growth are too simple to capture processes we are interested in, but the model produces at a small scale a broad range of urban shapes reproducing typical patterns of human settlements.

Urban Systems Modeling

An approach closer to our current questioning is the one of integrated modeling of system of cities. In the continuity of Simpop models for city systems modeling, Schmitt described in [189] the SimpopNet model which aim was precisely to integrate co-evolution processes in system of cities on long time scales, typically rules for hierarchical network development as a function of cities dynamics coupled with city dynamics depending on network topology. Unfortunately the model was not explored nor further studied, and furthermore stayed at a toy-level. Cottineau proposed transportation network endogenous growth as the last building bricks of her Marius productions but it stayed at a conceptual construction stage. We shall position more in that stream of research in this thesis.

1.2.3 *Sketch of a* Modelography

An ongoing work is the production of a synthesis of this overview, from a modular modeling point of view, combined with a purpose and scale classification. Already mentioned, modular modeling consists in the integration of heterogeneous processes and implementation of processes in order to extract the set of mechanisms giving the best fit to empirical data [60]. We can thus classify models described here according to their building bricks in terms of processes implemented and thus identify possible coupling potentialities. This work is a preliminary step for the analysis in quantitative epistemology developed in chapter 4.

1.3 RESEARCH QUESTION

To close this thematic touring introducing chapter, we can state a general research question that frames our further theoretical constructions and first modeling attempts. It is roughly the same as the problematic given at the end of previous section, but adding the insight of modeling as the approach to understand these complex systems.

General research Question. *To what extent a modeling approach to territorial systems as networked human territories can help disentangling complexly involved processes ?*

This question will be refined by theoretical developments in the next chapter and experiments in the followings.

2 THEORETICAL FRAMEWORK

Your theory is crazy, but not enough to be true.
- Niels Bohr

Theory is a key element of any scientific construction, especially in Human Sciences in which object definition and questioning are more open but also determining for research directions. We develop in this chapter a self-consistent theoretical background. It naturally emerges from thematic considerations of previous chapter, empirical explorations done in chapter 5 and modeling experiments conducted in chapter 6, as a linear structure of knowledge is not appropriate to translate the type of scientific entreprise we are conducting, typically in the spirit of Sanders in [128] for which the simultaneous conjonction of empirical, conceptual and modeling domains is necessary for the emergence of knowledge. This theoretical construction is however presented to be understood independently, and is used as a structuring skeleton for the rest of the thesis.

We propose first to construct the *geographical theory* that will pose the studied objects and their meaning in the real world (their ontology), with their interrelations. This yields precise assumptions that will be sought to be confirmed or proven false in the following. Staying at a thematic level appears however to be not enough to obtain general guidelines on the type of methodologies and the approaches to use. More precisely, even if some theories imply an more natural use of some tools[1], at the subtler level of contextualization in the sense of the approach taken to implement the theory (as models or empirical analysis), the freedom of choice may mislead into unappropriated techniques or questionings (see [172] on the example of incautious use of big data and computation). We develop therefore in a second section a theoretical framework at a meta-level, aiming to give a vision and framing for modeling socio-technical systems.

1 to give a rough example, a theory emphasizing the complexity of relations between agents in a system will conduct generally to use agent-based modeling and simulation tools, whereas a theory based on macroscopic equilibrium will favorise the use of exact mathematical derivations.

2.1 GEOGRAPHICAL THEORETICAL CONTEXT

2.1.1 *Foundation*

Networked Human Territories

Our first pillar has already been constructed before in the thematic exploration of the research subject. We rely on the notion of *Human Territory* elaborated by Raffestin as the basis for a definition of territorial systems. It permits to capture complex human geographical systems in their concrete and abstract characteristics and representation. For example, a metropolitan territorial system can be apprehended simply by the functional extent of daily commuting, or by the perceived or lived space of different populations, the choice depending on the precise question asked. Note that this approach to territory is a position and that other (possibly compatible) entries could be taken [143]. The concrete of this pillar in reinforced by the territorial theory of networks of Dupuy, yielding the notion of networked human territory, as a human territory in which a set of potential transactional networks have been realized, which is in accordance with vision of the territory as networked places [46]. We make therein the assumption that real networks are necessary elements of territorial systems.

Evolutive Urban Theory

The second pillar of our theoretical construction is the Evolutive Urban Theory of Pumain, closely linked to the complexity approach we take. This theory was first introduced in [161] which argues for a dynamical vision of city systems, in which self-organization is key. Cities are interdependent evolutive spatial entities whose interrelations produces the macroscopic behavior at the scale of city system. The city system is also designed as a network of city what emphasizes its view as a complex system. Each city is itself a complex system in the spirit of [28], the multi-scale aspect being essential in this theory, since microscopic agents convey system evolution through complex feedbacks between scales. The positioning within Complex System Sciences was later confirmed [162]. It was shown that this theory provide an interpretation for the origin of pervasive scaling laws, resulting from the diffusion of innovation cycles between cities [167]. The aspect of resilience of system of cities, induced by the adaptive character of these complex systems, implies that cities are drivers and adapters of social change [164]. Finally, path dependance yield non-ergodicity within these systems, making "universal" interpretations of scaling laws developed by physicists incompatible with evolutive urban theory [166]. We will interpret territorial systems following that idea of complex adaptive systems.

Urban Morphogenesis

The idea of morphogenesis was introduced by Turing in [200] when trying to isolate simple chemical rules that could lead to the emergence of the embryo and its form. The morphogenesis of a system consists in self-consistent evolution rules that produce the emergence of its successives states, i.e. the precise definition of self-organization. Progresses towards the understanding of embryo morphogenesis (in particular the isolation of processes producing the differentiation of cells from an unique cell) has been made only recently with the use of Complexity Approaches in integrative biology [68]. In the case of urban systems, the idea of urban morphogenesis, i.e. of self-consistent mechanisms that would produce the urban form, is more used in the field of architecture and urban design [97] (as Alexander generative grammar "Pattern Language" e.g.), in relation with theories of Urban Form [142]. This idea can be pushed into very small scales such as the building [209] but we will use it more at a mesoscopic scale, in terms of land-use changes within an intermediate scale territorial system, in the same ontologies as Urban morphogenesis modeling literature (for example [34] describes a model of urban morphogenesis with qualitative differentiation, whereas [136] give a model of urban growth based on a mono-centric population probability distribution perturbed with correlated noises). The notion of morphogenesis will be important in our theory in link with modularity and scale. Modularity of a complex system consists in its decomposition into relatively independent sub-modules, and modular decomposition of a system can be seen as a way to disentangle non-intrinsic correlations [113] (think of a block diagonalisation of a first order dynamical system). The isolation of a subsystem yields a corresponding characteristic scale. Isolating possible morphogenesis processes imply a controlled isolation (controlled boundary conditions e.g.) of the considered system, corresponding to a modularity level and thus a scale. When self-consistent processes are not enough to explain the evolution of the system (with reasonable action on boundary conditions), a change of scale is necessary, caused by an underlying phase transition in modularity. The example of metropolitan growth is a good example : complexity of interactions within the metropolitan region will grow with size and diversity of functions leading to a change in scale necessary to understand processes. The emergence of an international airport will strongly influence local development, what corresponds to the significant integration within a larger system. The characteristic scales and processes for which these change occur will be precise questions to be investigated through modeling. It is interesting to remark that a territorial subsystem in which morphogenesis has a sense can be seen as an *autopoietic system* in the extended sense of Bourgine in [38], as a network of auto-reproducing processes[2] regulating their boundary conditions, what emphasizes boundaries on which we will last insist.

Co-evolution

Our last pillar is a clarification of the notion of *co-evolution*, on which Holland shed light through an approach of complex adaptive systems by a theory of CAS as signal processing agents operating thanks to their boundaries [104]. In this theory, complex adaptive systems form aggregates at diverse hierarchical levels, that correspond to different level of self-organization, and boundaries are vertically and horizontally intricate in a complex way. That approach introduces the notion of *niche* as a relatively independent subsystem in which ressources circulate (the same way as network communities) : numerous illustrations are given such as economical niches or ecological niches. Agents within a niche are said to be *co-evolving*. Co-evolution thus means strong interdependences (implying circular causal processes) and a certain independence regarding the exterior of the niche. The notion is naturally flexible as it will depend on ontologies, resolution, thresholds etc. considered to define the system. This concept is easily transmissible to the evolutive urban theory and converges with the notion of co-evolution described by Pumain : co-evolving agents in a system of cities consist in a niche with its flows, signals and boundaries and thus co-evolving entities in the sense of Holland. This notion will be important for us in the definition of territorial subsystems and their coupling.

2.1.2 *Synthesis : an theory of co-evolutive networked territorial systems*

We synthesize our pillars as a short self-consistent geographical theory of territorial systems in which networks play a central role in the co-evolution of components of the system. See the foundation subsection for definitions and references. The formulation is intended to be minimalistic.

Definition 1 ***- Territorial System.*** *A territorial system is a set of networked human territories, i.e. human territories in and between which real networks exist.*

At this step complexity and dynamical evolutive characters of territorial systems are implied but not an explicit part of the theory. We will assume to simplify a discrete definition of temporal, spatial and ontological scales under modularity and local stationarity assumptions.

2 which are however not cognitive, making this auto-organized systems fortunately not alive in the sense of autopoietic and cognitive systems

Proposition 1 ***- Discrete scales.*** *Assuming a discrete modular decomposition of a territorial system, the existence of a discrete set* (τ_i, x_i) *of temporal and functional scales for the territorial system is equivalent to the local temporal stationarity of a random dynamical system specification of the system.*

Proof (Sketch of). We underlie that any territorial system can be represented by random variables, what is equivalent to have well defined objects and states and use the Transfer Theorem on events of successive states. If $X = (X_j)$ is the modular decomposition, we have necessarily quasi-independence of components in the sense that $\text{Cov}\left[dX_j, dX_{j'}\right] \simeq 0$ at any time. General stationarity transitions induce modular transitions that are kept or not depending if they correspond to an effective transition within the subsystem, what provide temporal scales as characteristic times of sub-dynamics. Functional scales are the corresponding extent in the state space. ∎

This proposition induce a discrete representation of system dynamics in time. Note that even in the case of no modular representation, the system as a whole will verify the property. This definition of scales allows to explicitly introduce feedback loops and thus emergence and complexity, making our theory compatible with the evolutive urban theory.

Assumption 1 ***- Scales and Subsystems intrication.*** *Complex networks of feedbacks exist both between and inside scales, what impose the existence of weak emergence [24]. Furthermore a horizontal and vertical hierarchical imbrication of boundaries is not the rule.*

Within these complex subsystems intrications we can isolate co-evolving components using morphogenesis. The following proposition is a consequence of the equivalence between the independence of a niche and its morphogenesis. Morphogenesis provides the modular decomposition (local stationarity assumed) needed for the existence of scale, giving minimal vertically (scale) and horizontally (space) independent subsystems.

Proposition 2 ***- Co-evolution of components.*** *Morphogenesis processes of a territorial system are an equivalent formulation of the existence of co-evolutive subsystems.*

Finally we make a key assumption putting real networks at the center of co-evolutive dynamics, introducing their necessity to explain dynamical processes of territorial systems.

Assumption 2 ***- Necessity of Networks.*** *Network evolution can not be explained only by the dynamics of other territorial components and reciprocally, i.e. co-evolving territorial subsystems include real networks. They can thus be at the origin of regime changes (transition between stationarity regimes) or more dramatic bifurcations in dynamics of the whole territorial system.*

On long time scale, an overall co-evolution has been shown for the french railway network by [39]. At smaller scales it is less evident (debate on structural effects) but we postulate that co-evolution effects are present at any scale. Regional examples may illustrate that : Lyon has not the same dynamical relations with Clermont than with Saint-Etienne and network connectivity has necessarily a role in that (among intrinsic interaction dynamics and distance). At a smaller scale, we think that effects are even less observable, but precisely because of the fact that co-evolution is stronger and local bifurcations will occur with stronger amplitude ans greater frequency than in macroscopic systems where attractors are more stable and stationarity scales greater. We will try to identify bifurcation or phase transitions in toy models, hybrid models and empirical analysis, at different scales, on different case studies and with different ontologies.

One difficulty in our construction is the stationarity assumption. Even if it seems a reasonable assumptions on large scales and has already been observed in empirical data [184], we shall verify it in our empirical studies. Indeed, this question is at the center of current research efforts to apply deep learning techniques to geographical systems : Bourgine has recently developed a framework to extract patterns of Complex Adaptive Systems[3]. The issues are then if the stationarity assumption be tackled through augmentation of system states, and if heterogeneous and asynchronous data be used to bootstrap long time-series necessary for a correct estimation of the neural network. These issue are related to the stationarity assumption for the first and to non-ergodicity for the second.

3 Using a representation theorem [111], any discrete stationary process is a *Hidden Markov Model*. Given the definition of a causal state as $\mathbb{P}[\text{future}|A] = \mathbb{P}[\text{future}|B]$, the partition of system states induced by the corresponding equivalence relations allows to derive a *Recurrent Network* that is enough to determine next state of the system, as it is a *deterministic* function of previous state and hidden states [191] : $(x_{t+1}, s_{t+1}) = F[(x_t, s_t)]$. The estimation of Hidden States and of the Recurrent Function thus captures through deep learning entirely dynamical patterns of the system, i.e. full information on its dynamics and internal processes.

2.2 A THEORETICAL FRAMEWORK FOR THE STUDY OF SOCIO-TECHNICAL SYSTEMS

After having set up the thematic theoretical framework, we develop a more general framework in which the previous can enter. At an epistemological level, it is essential to frame generally our directions of research.

2.2.1 *Introduction*

Scientific Context

The structural misunderstandings between Social Sciences and Humanities on one side, and so-called Exact Sciences on the other side, far from being a generality, seems to have however a significant impact on the structure of scientific knowledge [103]. In particular, the place of theory (and indeed the signification of this term itself) in the elaboration of knowledge has a totally different place, partly because of the different *perceived complexities*[4] of studied objects : for example, mathematical constructions and by extent theoretical physics are *simple* in the sense that they are mostly entirely analytically solvable, whereas Social Science subjects such as humans or society (to give a *cliché* exemple) are *complex* in the sense of complex systems[5], thus a stronger need of a constructed theoretical (generally empirically based) framework to identify and define the objects of research that are necessarily more arbitrary in the framing of their boundaries, relations and processes, because of the multitude of possible viewpoints : Pumain suggests indeed in [163] a new approach to complexity deeply rooted in social sciences that "would be measured by the diversity of disciplines needed to elaborate a notion". These differences in backgrounds are naturally desirable in the spectrum of science, but things can get nasty when playing on "common" terrains, typically complex systems problematics as already detailed, as the exemple of geographical urban systems has recently shown [77]. Complex System Science[6] is presented by some as a "new kind of Science" [213], and would at least be a symptom of a shift in scientific practices, from analytical and "exact" approaches to computational and evidence-based approaches [9], but what is sure is that it brings,

4 We used the term *perceived* as most of systems studied by physics might be described as simple whereas they are intrinsically complex and indeed not well understood [118].

5 for which no unified definition exists but of which fields of application range broadly from neuroscience to quantitative finance, including e.g. quantitative sociology, quantitative geography, integrative biology, etc. [145], and for which study various complementary approaches may be applied, such as Dynamical Systems, Agent-based Modeling, Random Matrix Theory

6 that we deliberately call that way although there is a running debate on wether it can be seen as a Science in itself or more as a different way to do Science.

together with new methodologies, new scientific fields in the sense of converging interests of various disciplines on transversal questions or of integrated approaches on a particular field [37].

Objectives

Within that scientific context, the study of what we will call *Socio-technical Systems*, which we define in a rather broad way as hybrid complex systems including social agents or objects that interact with technical artifacts and a natural environment[7], lies precisely between social sciences and hard sciences. The example of Urban Systems is the best example, as already before the arrival of approaches claiming to be "more exact" than soft approaches (typically by physicists, see e.g. the rather disturbing introduction of [131], but also by scientists coming from social sciences such as Batty [20]), many aspects of urban systems were already in the field of exact sciences, such as urban hydrology, urban climatology or technical aspects of transportation systems, whereas the core of their study relied in social sciences such as geography, urbanism, sociology, economy. Therefore a necessary place of theory in their study : following [128], the study of complex systems in social science is an interaction between empirical analysis, theoretical constructions, and modeling.

We propose in this paper to construct a theory, or rather a theoretical framework, that would ease some aspects of the study of such systems. Many theories already exist in all fields related to this kind of problems, and also at higher levels of abstraction concerning methods such as agent-based modeling e.g., but there is to our knowledge no theoretical framework including all of the following aspects that we consider as being crucial (and that can be understood as an informal basis of our theory) :

1. a precise definition and emphasis on the notion of coupling between subsystems, in particular allowing to qualify or quantify a certain degree of coupling : dependence, interdependence, etc. between components.

2. a precise definition of scale, including timescale and scales for other dimensions.

3. as a consequence of the previous points, a precise definition of what is a system.

4. the inclusion of the notion of emergence in order to capture multi-scale aspects of systems.

7 geographical systems in the sense of [72] are the archetype of such systems, but that definition may cover other type of systems such as an extended transportation system, social systems taken with an environmental context, complicated industrial systems taken with users, etc.

5. a central place of ontology in the definition of systems, i.e. of the sense in the real world given to its objects[8].

6. taking into account heterogeneous aspects of the same system, that could be heterogeneous components but also complementary intersecting views.

The rest of this section is organized as follows : we construct the theory in the following part, staying at an abstract level, and propose a first application to the question of co-evolving subsystems. We then discuss positioning regarding existing theories, and possible developments and concrete applications.

2.2.2 *Construction of the theory*

Perspectives and Ontologies

The starting point of the theory construction is a perspectivist epistemological approach on systems introduced by Giere [86]. To sum up, it interprets any scientific approach as a perspective, in which someone pursues some objective and uses what is called *a model* to reach it. The model is nothing more than a scientific medium. Varenne developed [201] model typologies that can be interpreted as a refinement of this theory. Let for now relax this possible precision and use perspectives as proxies of the undefined objects and concepts. Indeed, different views on the same object (being complementary or diverging) have the property to share at least the object in itself, thus the proposition to define objects (and more generally systems) from a set of perspectives on them, that verify some properties that we formalize in the following.

A perspective is defined in our case as a dataflow machine M (that corresponds to the model as medium) in the sense of [90] that gives a convenient way to represent it and to introduce timescales, to which is associated an ontology O in the sense of [128], i.e. a set of elements each corresponds to a *thing* (it can be an object, an agent, a process, etc.) in the real world. We include only two aspect (the model and the objects represented) of Giere's theory, making the assumption that purpose and user of the perspective are indeed contained in the ontology.

Definition 2 *A* perspective on a system *is given by a dataflow machine* $M = (i, o, \mathbb{T})$ *and an associated ontology* O*. We assume that the ontology can be decomposed into atomic elements* $O = (O_j)_j$.

The atomic elements of the ontology can be particular elements such as agents or components of the system, but also processes, interactions, states, or concepts for example. The ontology can be seen

8 *as already explained before, this positioning along with the importance of structure may be related to Ontic Structural Realism [82] in further in further developments.*

as the rigorous description of the content of the perspective. The assumption of a dataflow machine implies that possible inputs and outputs can be quantified, what is not necessarily restrictive to quantitative perspectives, as most of qualitative approaches can be translated into discrete variables as long as the set of possibles is known or assumed.

The system is then defined "reversely", i.e. from a set of perspectives on a system :

Definition 3 *A* system *is a set of* perspectives on a system *:* $S = (M_i, O_i)_{I \in I}$*, where* I *may be finite or not.*

We denote by $\mathcal{O} = (O_{j,i})_{j,i \in I}$ the set of all elements within ontologies.

Note that at this level of construction, there is not necessarily any structural consistence in what we call a system, as given our broad definition could allow for example to consider as a system a perspective on a car together with a perspective on a system of cities what makes reasonably no sense at all. Further definitions and developments will allow to be closer from classical definition of a system (interacting entities, designed artifacts, etc.). The same way, the definition of a subsystem will be given further. The introduced elements of our approach help to tackle so far points three, five and six of the requirements.

PRECISION ON THE RECURSIVE ASPECT OF THE THEORY One direct consequence of these definitions must be detailed : the fact that they can be applied recursively. Indeed, one could imagine taking as perspective a system in our sense, therefore a set of perspectives on a system, and do that at any order. If ones takes a system in any classical sense, then the first order can be understood as an epistemology of the system, i.e. the study of diverse perspectives on a system. A set of perspectives on related systems may in some conditions be a domain or a field, thus a set of perspectives on various related systems the epistemology of a field. These are more analogies to give the idea behind the recursive character of the theory. It is indeed crucial for the meaning and consistence of the theory because of the following arguments :

- The choice of perspectives in which a system consists is necessarily subjective and therefore understood as a perspective, and a perspective on a system if we are able to build a general ontology.
- We will use relations between ontologies in the following, which construction based on emergence is also subjective and seen as perspectives.

Ontological Graph

We propose then to capture the structure of the system by linking ontologies. This approach could eventually be linked to structural realism epistemological positioning [82] as knowledge of the world is partly contained here in structure of models. Therefore, we choose to emphasize the role of emergence as we believe that it may be one practical minimalist way to capture quite well complex systems structure[9]. We follow on that point the approach of Bedau on different type of emergences, in particular his definition of weak emergence given in [24]. Let recall briefly definitions we will use in the following. Bedau starts from defining emerging properties and then extends it to phenomena, entities, etc. The same way, our framework is not restricted to objects or properties and wrapped thus the generalized definitions into emergence between ontologies. We will apply the notion of emergence under the two following forms[10] :

- *Nominal emergence* : one ontology O' is included in an other O but the aspect of O that is said to be nominally emergent regarding O' does not depend on O'.
- *Weak emergence* : one part of an ontology O can be derived by aggregation of elements and interactions between elements of an ontology O'.

As developed before, the presence of emergence, and especially weak emergence, will consist in itself in a perspective. It can be conceptual and postulated as an axiom within a thematic theory, but also experimental if clues of weak emergence are effectively measured between objects. In any case, the relation between ontologies must be encoded within an ontology, which was not necessarily introduced in the initial definition of the system.

We make therefore the following assumption for next developments :

Assumption 3 *A system can be partially structured by extending it with an ontology that contains (not necessarily only) relations between elements of ontologies of its perspectives. We name it the* coupling ontology *and assume its existence in the following. We assume furthermore its atomicity, i.e. if* O *is in relation with* O'*, then any subsets of* O, O' *can not be in relation, what is not restrictive as a decomposition into several independent subsets ensures it if it is not the case.*

9 what of course can not been presented as a provable claim as it depends on system definition, etc.

10 the third form Bedau recalls, *Strong emergence* will not be used, as we need only to capture dependance and autonomy, and weak emergence is more satisfying in terms of complex systems, as it does not assume “irreducible causal powers” to the greater scale objects. Nominal emergence is used to capture inclusion between ontologies.

It allows to exhibit emergence relations not only within a perspective itself but also between elements of different perspectives. We define then pre-order relations between subsets of ontologies :

Proposition 3 *The following binary relationships are pre-orders on $\mathcal{P}(\mathcal{O})$:*

- *Emergence (based on Weak Emergence) :* $O' \preccurlyeq O$ *if and only if* O *weakly emerges from* O'.
- *Inclusion (based on Nominal Emergence) :* $O' \Subset O$ *if and only if* O *nominally emerges from* O'.

Proof With the convention that it can be said that an object emerges from itself, we have reflexivity (if such a convention seems absurd, we can define the relationships as O *emerges from* O' *or* $O = O'$). Transitivity is clearly contained in definitions of emergence.

Note that the inclusion relation is more than an inclusion between sets, as it translates an inclusion "inside" the elements of the ontology.

These relations are the basis for the construction of a graph called the *ontological graph* :

Definition 4 *The ontological graph is constructed by induction the following way :*

1. *A graph with vertices elements of* $\mathcal{P}(\mathcal{O})$ *and edges of two types :* $E_W = \{(O, O')|O' \preccurlyeq O\}$ *and* $E_N = \{(O, O')|O' \Subset O\}$
2. *Nodes are reduced*[11] *by : if* $o \in O, O'$ *and* ($O' \preccurlyeq O$ *or* $O' \Subset O$) *but not* ($O \preccurlyeq O'$ *or* $O \Subset O'$)*, then* $O' \leftarrow O' \setminus o$
3. *Nodes with intersecting sets are merged, keeping edges linking merged nodes. This step ensures non-overlapping nodes.*

Minimal Ontological Tree

The topological structure of the graph, that contains in a way the *structure of the system*, can be reduced into a minimal tree that contains hierarchical structure essential to the theory.

We need first to give consistence to the system :

Definition 5 *A consistent part of the ontological graph is a weakly connected component of the graph. We assume for now to work on a consistent part.*

The notion of consistent system, together with subsystem or nodes timescales that will be defined later, requires to reconstruct perspectives from ontological elements, i.e. the inverse operation of what was done in our deconstruction procedure.

11 the reduction procedure aims to delete redundancy, keeping an entity at the higher level it exists.

Assumption 4 *There exists $\mathcal{O}' \subset \mathcal{P}(\mathcal{O})$ such that for any $O \subset \mathcal{O}'$, there exists a corresponding dataflow machine M such that the corresponding perspective is consistent with initial elements of the system (i.e. machines on ontology overlaps are equivalent). If $\Phi : M \mapsto O$ is the initial mapping, we denote this extended reciprocal construction by $M' = \Phi^{<-1>}(O)$.*

REMARK. This assumption could eventually be changed into a provable proposition, assuming that the coupling ontology is indeed a coupling perspective, which dataflow machine part is consistent with coupled entities. Therein, the decomposition postulate of [90] should allow to identify basic components corresponding to each element of the ontology, and then construct the new perspective by induction. We find however these assumptions too restrictive, as for example various ontological elements may be modeled by an irreducible machine, as a differential equations with aggregated variables. We prefer to be less restrictive and postulate the existence of the reverse mapping on some sub-ontologies, that should be in practice the ones where couplings can be effectively modeled.

Given this assumption, we can define the consistent system as the reciprocal image of the consistent part of the ontological graph. It ensures system connectivity what is a requirement for tree construction.

Proposition 4 *The tree decomposition of the ontological graph in which nodes contains strongly connected components is unique. The corresponding reduced tree, that corresponds to the ontological graph in which strongly connected components have been merged with edges kept, is called the* Minimal Ontological Tree*.*

Proof (sketch of) The unicity is obtained as nodes are fixed as strongly connected components. It is trivially a tree decomposition (with no edges) as in a directed graph, strongly connected components do not intersect, thus the consistence of the decomposition.

Any loop $O \to O' \to \ldots \to O$ in the ontological graph assumes that all its elements are equivalent in the sense of $\preccurlyeq$. This equivalence loops should help to define the notion of strong coupling as an application of the theory (see applications).

The Minimal Ontological Tree (MOT) is a tree in the undirected sense but a forest in the directed sense. Its topology contains a sort of system hierarchy. Consistent subsystems are defined from the set $\mathcal{B}$ of branches of the forest, as $(\Phi^{<-1>}(\mathcal{B}), \mathcal{B})$. The timescale of a node, and by extension of a subsystem, is the union of timescales of corresponding machines. Levels of the tree are defined from root nodes, and the emergence relations between nodes implies a vertical inclusion between timescales.

Scales

Finally, we propose to define scales associated to a system. Following [139], an epistemological continuum of visions on scale is a consequence of differences between disciplines in the way we developed in the introduction. This proposition is indeed compatible with our framework, as the construction of scales for each level of the ontological tree results in a broad variety of scales.

Let (M, O) a subsystem and $\mathbb{T}$ the corresponding timescale. We propose to define the "thematic scale" (for example spatial scale) assuming a representation theorem, i.e. that an aspect (thematic aspect) of the machine can be represented as a dynamic state variable $\vec{X}(t)$. Assuming a scale operator[12] $\|\cdot\|_S$ and that the state variable has a certain level of differentiability, the *thematic scale* if defined as $\|(d^k\vec{X}(t))_k\|_S$

2.2.3 *Application*

The particular case of geographical systems

In [72] Durand-Dastès proposes a definition of geographical structure and system, structure would be the spatial container for systems viewed as complex open interacting systems (elements with attributes, relations between elements and inputs/outputs with external world). For a given system, its definition is a perspective, completed by structure to have a system in our sense. Depending on the way to define relations, it may be easy to extract ontological structure.

Modularity and co-evolving subsystems

For the example of Urban Systems, urban evolutionary theory enters this framework using our previous thematic theory? The decomposition into uncorrelated subsystems yields precisely strongly coupled components as co-evolving components. The correlation between subsystems should be positively correlated with topological distance in the tree. If we define elements of a node before merging as *strongly coupled elements*, in the case of dynamic ontologies, it provides a definition of *co-evolution* and co-evolving subsystems equivalent to the thematic definition.

2.2.4 *Discussion*

LINK WITH EXISTING FRAMEWORKS A link with the Cottineau-Chapron framework for multi-modeling [52] may be done in the case they add the bibliographical layer, which would correspond to the

12 that can be of various nature : extent, probabilistic extent, spectral scales, stationarity scales, etc.

reconstruction of perspectives. [178] proposes the notion of "interdisciplinary coupling" what is close to our notion of coupling perspectives. A correspondance with System of Systems approaches (see e.g. [134] for a recent general framework englobing system modeling and system description) may be also possible as our perspectives are constructed as dataflow machines, but with the significant difference that the notion of emergence is central.

CONTRIBUTIONS TO THE STUDY OF COMPLEX SYSTEMS

- We do not claim to provide a theory of systems (beware of cybernetics, systemics etc. that could not model everything), but more a framework to guide research questions (e.g. in our case the direct outcomes will be quantitative epistemology that comes from system construction as perspectives research ; empirical to construct robust ontologies for perspectives ; targeted thematic to unveil causal relationship/emergence for construction of ontological network ; study of coupling as possible processes containing co-evolution ; study of scales ; etc.). It may be understood as meta-theory which application gives a theory, the thematic theory developed before being a specific implementation to territorial networked systems.
- We Emphasize the notion of socio-technical system, crossing a social complex system approach (ontologies) with a description of technical artifacts (dataflow machines), taking the "best of both worlds".

2.2.5 *Research Directions*

We can draw from the construction of this theoretical framework a set of research directions, that give a general line on how trying to answer to research questions asked after the thematic theory construction.

1. The perspectivist approach implies a broad understanding of existing perspectives on a system, and of possibility of coupling between them ; thus an emphasis on applied epistemology, i.e. **Algorithmic Systematic Review** (exploration of the knowledge space), **Disciplines Mapping**(extraction of its structure) and **Datamining for Content Analysis**(refinement at the atomic level in scientific knowledge) that correspond to the three sections of chapter 4.
2. At a finer level of particularization, the knowledge of perspectives means **Knowledge of stylized facts**, i.e. empirical analysis of cases studies. These are the object of chapter 5.
3. The emphasis on coupled subsystems at different scales implies a deep understanding of coupling mechanisms, thus the need

of methodological and technical developments : **Methods for Statistical Control**, **Methods for Model Exploration**, **Theoretical Study of Coupling**, **Multi-Modeling**, of which some are developed and other proposed in the methodological chapter 3.

4. Furthermore, the possibility of hidden elements within the ontology implies the test for causal relations and intermediate processes at the origin of emergence, thus e.g. the exploration of new paradigms such as role of governance within complex models as done in chapter 7.

5. Finally, the idea behind system structure contained within the ontological forest is a large set of coupled models for a given system : it means that a proper system definition (i.e. thematic problematization and exploration) and construction should yield to a structured family of models : parallel branches can be different implementations of the same process or various processes trying to explain the emerging ontology ; therefore the final objective of a family of models tackling the thematic question.

3 METHODOLOGICAL DEVELOPMENTS

We are now building a rigorous Science of Cities, contrarily to what was done before.
- Marc Barthélémy

Such a shocking phrase was pronounced during the introduction of a *Network* course for students of Complex System Science. Besides the fact that the spirit of CSS is precisely the opposite, i. e. the construction of integrative disciplines (vertical integration that is necessarily founded on the existing body of knowledge of concerned fields) that answer transversal questions (horizontal integration that imply interdisciplinarity) - see e. g. the roadmap for CS [37], it reveals how methodological considerations shape the perceptions of disciplines. From a background in Physics, "rigorous" implies the use of tools and methods judged more rigorous (analytical derivations, large datasets statistics, etc.). But what is rigorous for someone will not be for an other discipline[1], depending on the purpose of each piece of research (perspectivism [86] poses the *model*, that includes methods, as the articulating core of research entreprises). Thus the full role of methodology aside and not beside theory and experiments. We go in this chapter into various methodological developments which may be precisely used later or contribute to the global background.

We first propose a kind of essay insisting on the importance of reproducibility in science. More than a guideline, it is a way to practice science that a necessary condition for its rigor. Any non-reproducible work is not scientific. We then derive technical results on models of urban growth and on the sensitivity of scaling laws, that are both recurrent themes in the modeling of complex urban systems. We then introduce a method in the context of systematic model exploration and model behavior. We finally work on a link between static and dynamic correlations in a geographical system. This chapter is rather heteroclite as sections may correspond to a particular technical need at a point in the thesis, to global methodological directions, or global research directions.

1 a funny but sad anecdote told by a friend comes to mind : defending his PhD in statistics, he was told at the end by economists how they were impressed by the mathematical rigor of his work, whereas a mathematician judged that "he could have done everything on the back of an enveloppe".

3.1 REPRODUCIBILITY

The strength of science comes from the cumulative and collective nature of research, as progresses are made as Newton said "standing on the shoulder of giants", meaning that the scientific enterprise at a given time relies on all the work done before and that advances would not be possible without constructing on it. It includes development of new theories, but also extension, testing or falsifiability of previous ones. In that context

As scientific reproducibility is an essential requirement for any study, its practice seems to be increasing [196] and technical means to achieve it are always more developed (as e.g. ways to make data openly available, or to be transparent on the research process such as `git` [175], or to integrate document creation and data analysis such as `knitr` [216]), at least in the field of numerical modeling and simulation. However, the devil is indeed in the details and obstacles judged at first sight as minor become rapidly a burden for reproducing and using results obtained in some previous researches. We describe two cases studies where models of simulation are apparently highly reproducible but unveil as puzzles on which research-time balance is significantly under zero, in the sense that trying to exploit their results may cost more time than developing from scratch similar models.

3.1.1 *On the Need to Explicit the Model*

A current myth is that providing entire source code and data will be a sufficient condition for reproducibility. It will work if the objective is to produce exactly same plots or statistical analysis, assuming that code provided is the one which was indeed used to produce the given results. It is however not the nature of reproducibility. First, results must be as much implementation-independent as possible for clear robustness purposes. Then, in relation with the precedent point, one of the purposes of reproducibility is the reuse of methods or results as basis or modules for further research (what includes implementation in another language or adaptation of the method), in the sense that reproducibility is not replicability as it must be adaptable [74].

Our first case study fits exactly that scheme, as it was undoubtedly aimed to be shared with and used by the community since it is a model of simulation provided with the Agent-Based simulation platform NetLogo [212]. The model is also available online [66] and is presented as a tool to simulate socio-economic dynamics of low-income residents in a city based on a synthetic urban environment, generated to be close in stylized facts from the real town of Tijuana, Mexico. Beside providing the source code, the model appears to be poorly documented in the literature or in comments and description of the implementation. Comments made thereafter are based on the study of the urban morphogenesis part of the model (setup for the "residential dynamics" component) as it is our global context of study [173]. In the frame of that study, source code was modified and commented, which last version is available on the repository of the project[2].

Rigorous formalization An obvious part of model construction is its rigorous formalization in a formal framework distinct from source code. There is of course no universal language to formulate it [10], and many possibilities are offered by various fields (e.g. UML, DEVS, pure mathematical formulation). No paper nor documentation is provided with the model, apart from the embedded NetLogo documentation since it only thematically describes in natural language the ideas behind each step without developing more and provides information about role of different elements of the interface.

This formulation is a key for it to be understood, reproduced and adapted ; but it also avoids implementation biases such as

- Architecturally dangerous elements : in the model, world context is a torus and agents may "jump" in the euclidian representation, what is not acceptable for a 2D projection of real world. To avoid that, many tricky tests and functions were used, including unadvised practices (e.g. dead of agents based on position to avoid them jumping).

- Lack of internal consistence : the example of the patch variable `land-value` used to represent different geographical quantities at different steps of the model (morphogenesis and residential dynamics), what becomes an internal inconsistence when both steps are coupled when option `city-growth?` is activated.

- Coding errors : in an untyped language such as NetLogo, mixing types may conduct to unexpected runtime errors, what is the case of the patch variable `transport` in the model (although no error occurs in most of run configurations from the interface, what is more dangerous as the developer thinks implementation is secure). Such problems should be avoided if implementation is done from an exact formal description of the model.

Transparent implementation A totally transparent implementation is expected, including ergonomics in architecture and coding, but

Expected model behavior Whatever the definition, a model can not be reduced to its formulation and/or implementation, as expected model behavior or model usage can be viewed as being part of the model itself. In the frame of Giere's perspectivism [86], the

2 at `https://github.com/JusteRaimbault/CityNetwork/tree/master/Models/Reproduction/UrbanSuite`

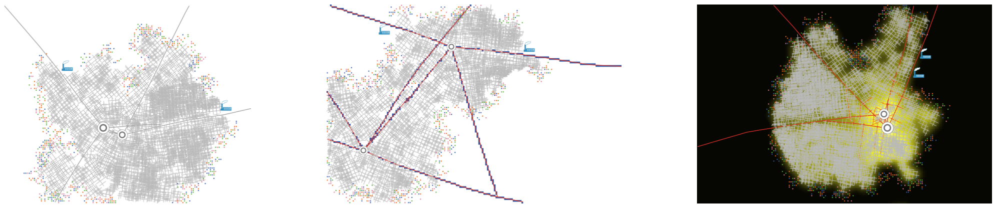


Figure 1: Example of simple improvement in visualization that can help understanding mechanisms implied in the model. *Left :* example of original output ; *Middle :* visualization of main roads (in red) and underlying patches attribution, suggesting possible implementation bias in the use of discretized trace of roads to track their positions ; *Right :* Visualization of land values using a more readable color gradient. This step confirms the hypothesis, through the form of value distribution, that the morphogenesis step is an unnecessary detour to generate a random field for which simple diffusion method should provide similar results, as detailed in the paragraph on implementation.

definition of model includes the purpose of use but also the agent who aims to use it. Therefore a minimal explication of model behavior and exploration of parameter roles is highly advised to decrease chances of misuses or misinterpretations of it. It includes simple runtime charts that are immediate on the NetLogo platform, but also indicators computations to evaluate outputs of the model. It can also be improved visualizations during runtime and model exploration, such as showed in Fig. 1.

3.1.2 *On the Need of Exactitude in Model Implementation*

Possible divergences between model description in a paper and the effectively implemented processes may have grave consequences on the reproducibility of science. The road network growth model given in [14] is one example that we are currently investigating. A strict implementation of model mechanisms provide slightly different results than the one presented in the paper, and as source code is not provided we need to test different hypotheses on possible mechanisms added by the programmer (that seems to be a connexion rule to intersections under a certain distance threshold). Lessons that could be possibly drawn from this examples are

- the necessity of providing source code
- the necessity of providing architecture description along with code (if model description is in a langage too far from architectural specifications) in order to identify possible implementation biaises

- the necessity of performing and detailing explicitly model explorations, that would in that case have helped to identify the implementation bias.

The last point, if first not provided, may ensure a limited risk of scientific falsification as it may be more complicated to fake false exploration results than to effectively explore the model. A joint project currently done is the writing of a false modeling paper in the spirit of [115], in which opposite results to the effective results of a given model are provided, without providing model implementation. A first bunch of test is the acceptance of a clearly non-reproducible paper in diverse journals, possibly with a control on textual elements (using or not "buzz-words" associated to the journal, etc.). Depending on results, a second experiment may be tested with providing open source code for model implementation but still with false results, to verify if reviewers effectively try to reproduce results when they pretend to want the code (in reasonable computational power limits of course, HPC being not currently broadly available in Humanities).

3.1.3 *Perspectives*

Again, reproducibility and transparency is a non-negotiable feature of contemporaneous science, along with Open practices and Open Access. Too much examples (see a very recent one in experimental economics [44]) show in various disciplines the lack of reproducibility of experiments, that is a falsification of previous results or a result in itself. Falsification is a costly practice, and even if necessary [49], could be made more efficient through more transparency and direct reproducibility, increase therein the global workflow of science. We develop in parallel of this thesis various tools aimed to ease reproducibility, for which an overview is given in appendix 10.

3.2 AN UNIFIED FRAMEWORK FOR STOCHASTIC MODELS OF URBAN GROWTH

Urban growth modeling fall in the case of tentatives to find self-consistent rules reproducing dynamics of an urban system, and thus in our logic of system morphogenesis. We examine here methodological issues linked to different frameworks of urban growth.

3.2.1 *Introduction*

Various stochastic models aiming to reproduce population patterns on large temporal and spatial scales (city systems) have been discussed across various fields of the literature, from economics to geography, including models proposed by physicists. We propose here a general framework that allows to include different famous models (in particular Gibrat, Simon and Preferential Attachment model) within an unified vision. It brings first an insight into epistemological debates on the relevance of models. Furthermore, bridges between models lead to the possible transfer of analytical results to some models that are not directly tractable.

Seminal models of urban growth are Simon [193] (later generalized as e.g. [101]) and Gibrat models. Many examples can be given across disciplines. [26] give an equation-based dynamical model, whereas [84] solves a stationary model. [85] reviews urban growth approaches in economics. A model adapted from evolutive urban theory is solved in [81] and improves Gibrat models. The question of empirical scales at which it is consistent to study urban growth was also tackled in the particular case of France [40]. We stay to a certain level of tractability to include models as essence of our approach is links between models but do not make ontologic assumptions.

3.2.2 *Framework*

PRESENTATION What we propose as a framework can be understood as a meta-model in the sense of [60], i.e. an modular general modeling process within each model can be understood as a limit case or as a specific case of another model. More simply it should be a diagram of formal relations between models. The ontological aspect is also tackled by embedding the diagram into an ontological state space (which discretization corresponds to the "bricks" of the incremental construction of [60]). It constructs a sort of model classification or modelography.

We are still at the stage of different derivations of links between models that are presented hereafter.

3.2.3 *Derivations*

Generalization of Preferential Attachment

[217] give a generalization of the classical Preferential Attachment Network Growth model, as a birth and death model with evolving entities. More precisely, network units gain and lose population (equivalent to links connexions) at fixed probabilities, and new unit can be created at a fixed rate.

Link between Gibrat and Preferential Attachment Models

Let consider a strictly positive growth Gibrat model given by $P_i(t) = R_i(t) \cdot P_i(t-1)$ with $R_i(t) > 1$, $\mu_i(t) = \mathbb{E}[R_i(t)]$ and $\sigma_i(t) = \mathbb{E}\left[R_i(t)^2\right]$. On the other hand, we take a simple preferential attachment, with fixed attachment probability $\lambda \in [0, 1]$ and new arrivants number $m > 0$. We derive that Gibrat model can be statistically equivalent to a limit of the preferential attachment model, assuming that the moment-generating function of $R_i(t)$ exists. Classical distributions that could be used in that case, e.g. log-normal distribution, are entirely defined by two first moments, making this assumption reasonable.

Lemma 1 *The limit of a Preferential Attachment model when $\lambda \ll 1$ is a linear-growth Gibrat model, with limit parameters $\mu_i(t) = 1 + \frac{\lambda}{m\cdot(t-1)}$.*

Proof Starting with first moment, we denote $\bar{P}_i(t) = \mathbb{E}[P_i(t)]$. Independence of Gibrat growth rate yields directly $\bar{P}_i(t) = \mathbb{E}[R_i(t)] \cdot \bar{P}_i(t-1)$. Starting for the preferential attachment model, we have $\bar{P}_i(t) = \mathbb{E}[P_i(t)] = \sum_{k=0}^{+\infty} k\, \mathbb{P}[P_i(t) = k]$. But

$$\{P_i(t) = k\} = \bigcup_{\delta=0}^{\infty} \left(\{P_i(t-1) = k-\delta\} \cap \{P_i \leftarrow P_i + 1\}^{\delta}\right)$$

where the second event corresponds to city i being increased δ times between $t-1$ and t (note that events are empty for $\delta \geqslant k$). Thus, being careful on the conditional nature of preferential attachment formulation, stating that $\mathbb{P}[\{P_i \leftarrow P_i + 1\} | P_i(t-1) = p] = \lambda \cdot \frac{p}{P(t-1)}$ (total population $P(t)$ assumed deterministic), we obtain

$$\begin{aligned}\mathbb{P}[\{P_i \leftarrow P_i + 1\}] &= \sum_p \mathbb{P}[\{P_i \leftarrow P_i + 1\} | P_i(t-1) = p] \cdot \mathbb{P}[P_i(t-1) = p] \\ &= \sum_p \lambda \cdot \frac{p}{P(t-1)}\, \mathbb{P}[P_i(t-1) = p] = \lambda \cdot \frac{\bar{P}_i(t-1)}{P(t-1)}\end{aligned}$$

It gives therefore, knowing that $P(t-1) = P_0 + m \cdot (t-1)$ and denoting $q = \lambda \cdot \frac{\bar{P}_i(t-1)}{P_0 + m\cdot(t-1)}$

$$\begin{aligned}
\bar{P}_i(t) &= \sum_{k=0}^{\infty}\sum_{\delta=0}^{\infty} k \cdot \left(\lambda \cdot \frac{\bar{P}_i(t-1)}{P_0 + m\cdot(t-1)}\right)^{\delta} \cdot \mathbb{P}[P_i(t-1) = k-\delta] \\
&= \sum_{\delta'=0}^{\infty}\sum_{k'=0}^{\infty} \left(k'+\delta'\right) \cdot q^{\delta'} \cdot \mathbb{P}\left[P_i(t-1) = k'\right] \\
&= \sum_{\delta'=0}^{\infty} q^{\delta'} \cdot \left(\delta' + \bar{P}_i(t-1)\right) = \frac{q}{(1-q)^2} + \frac{\bar{P}_i(t-1)}{(1-q)} \\
&= \frac{\bar{P}_i(t-1)}{1-q}\left[1 + \frac{1}{\bar{P}_i(t-1)}\frac{q}{(1-q)}\right]
\end{aligned}$$

As it is not expected to have $\bar{P}_i(t) \ll P(t)$ (fat tail distributions), a limit can be taken only through λ. Taking $\lambda \ll 1$ yields, as $0 < \bar{P}_i(t)/P(t) < 1$, that $q = \lambda \cdot \frac{\bar{P}_i(t-1)}{P_0+m\cdot(t-1)} \ll 1$ and thus we can expand in first order of q, what gives $\bar{P}_i(t) = \bar{P}_i(t-1)\cdot\left[1 + \left(1 + \frac{1}{\bar{P}_i(t-1)}\right) q + o(q))\right]$

$$\bar{P}_i(t) \simeq \left[1 + \frac{\lambda}{P_0 + m\cdot(t-1)}\right] \cdot \bar{P}_i(t-1)$$

It means that this limit is equivalent in expectancy to a Gibrat model with $\mu_i(t) = \mu(t) = 1 + \frac{\lambda}{P_0+m\cdot(t-1)}$.

For the second moment, we can do an analog computation. We have still

$$\mathbb{E}\left[P_i(t)^2\right] = \mathbb{E}\left[R_i(t)^2\right] \cdot \mathbb{E}\left[P_i(t-1)^2\right]$$

and

$$\mathbb{E}\left[P_i(t)^2\right] = \sum_{k=0}^{+\infty} k^2\, \mathbb{P}[P_i(t) = k]$$

We obtain the same way

$$\begin{aligned}
\mathbb{E}\left[P_i(t)^2\right] &= \sum_{\delta'=0}^{\infty}\sum_{k'=0}^{\infty} \left(k'+\delta'\right)^2 \cdot q^{\delta'} \cdot \mathbb{P}\left[P_i(t-1) = k'\right] \\
&= \sum_{\delta'=0}^{\infty} q^{\delta'} \cdot \left(\mathbb{E}\left[P_i(t-1)^2\right] + 2\delta'\bar{P}_i(t-1) + \delta'^2\right) \\
&= \frac{\mathbb{E}\left[P_i(t-1)^2\right]}{1-q} + \frac{2q\bar{P}_i(t-1)}{(1-q)^2} + \frac{q(q+1)}{(1-q)^3} \\
&= \frac{\mathbb{E}\left[P_i(t-1)^2\right]}{1-q}\left[1 + \frac{q}{\mathbb{E}[P_i(t-1)^2]}\left(\frac{2\bar{P}_i(t-1)}{1-q} + \frac{(1+q)}{(1-q)^2}\right)\right]
\end{aligned}$$

We have therefore an equivalence between the Gibrat model as a continuous formulation of a Preferential Attachment (or Simon model) in a certain limit. ∎

Link between Simon and Preferential Attachment

A rewriting of Simon model yields a particular case of the generalized preferential attachment, in particular by vanishing death probability.

Link between Favaro-Pumain and Gibrat

[81] generalizes Gibrat models with innovation propagation dynamics, being therefore a generalization of that model. Theoretically, a process-based model equivalent to the Favaro-pumain should then fill the missing case in model classification at the corresponding discretization. Simpop models do not fill that case as they stay at the scale of city systems, as for Marius models [58]. These must also have their counterparts in discrete microscopic formulation.

Link between Bettencourt-West and Pumain

We are considering to study Bettencourt-West model for urban scaling laws [30] as entering the stochastic urban growth framework as stationary component of a random growth model, but investigation are still ongoing.

Other Models

[84] develops an economic model giving a Simon equivalent formulation. They in particular find out that in upper tail, proportional growth process occurs. We find the same result as a consequence of the derivation of the link between Gibrat and Preferential attachment models.

3.3 ANALYTICAL SENSITIVITY OF URBAN SCALING LAWS TO SPATIAL EXTENT

At the center of evolutive urban theory are hierarchy and associated scaling laws. We begin here an methodological investigation on the sensitivity of scaling laws to city definition.

3.3.1 *Introduction*

Scaling laws have been shown to be universal of urban systems at many scales and for many indicators. Recent studies question however the consistence of scaling exponents determination, as their value can vary significantly depending on thresholds used to define urban entities on which quantities are integrated, even crossing the qualitative border of linear scaling, from infra-linear to supra-linear scaling. We use a simple theoretical model of spatial distribution of densities and urban functions to show analytically that such behavior can be derived as a consequence of the type of spatial distribution and the method used. Numerical simulation confirm the theoretical results and reveals that results are reasonably independent of spatial kernel used to distribute density.

Scaling laws for urban systems, starting from the well-known rank-size Zipf's law for city size distribution [84], have been shown to be a recurrent feature of urban systems, at many scales and for many types of indicators. They reside in the empirical constatation that indicators computed on elements of an urban system, that can be cities for system of cities, but also smaller entities at a smaller scale, do fit relatively well a power-law distribution as a function of entity size, i.e. that for entity i with population P_i, we have for an integrated quantity A_i, the relation $A_i \simeq A_0 \cdot \left(\frac{P_i}{P_0}\right)^{\alpha}$. Scaling exponent α can be smaller or greater than 1, leading to infra- or supra-linear effects. Various thematic interpretation of this phenomena have been proposed, typically under the form of processes analysis. The economic literature has produced abundant work on the subject (see [85] for a review), but that are generally weakly spatial, thus of poor interest to our approach that deals precisely with spatial organization. Simple economic rules such as energetic equilibria can lead to simple power-laws [30] but are difficult to fit empirically. A interesting proposition by Pumain is that they are intrinsically due to the evolutionary character of city systems, where complex emergent interaction between cities generate such global distributions [167]. Although a tempting parallel can be done with self-organizing biological systems, Pumain insists on the fact that the ergodicity assumption for such systems is not reasonable in the case of geographical systems and that the analogy cannot be exploited [166]. Other explanations have been proposed at other scales, such as the urban growth model at the mesoscopic scale (city scale) given in [129] that shows that the congestion within transportation networks may be one reason for city shapes and corresponding scaling laws. Note that "classic" urban growth models such as Gibrat's model do provide first order approximation of scaling systems, but that interactions between agents have to be incorporated into the model to obtain better fit on real data, such as the Favaro-Pumain model for innovation cycles propagation proposed in [81], that generalize a Gibrat model and provide better fits on data for French cities.

However, the blind application of scaling exponents computations was recently pointed as misleading in most cases [131], confirmed by empirical works such as [8] that showed the variability of computed exponents to the parameters defining urban areas, such as density thresholds. An ongoing work by Cottineau & *al.* presented at [59], studies empirically for French Cities the influence of 3 parameters playing a role in city definition, that are a density threshold θ to delimitate boundaries of an urban area, a number of commuters threshold θ_c that is the proportion of commuters going to core area over which the unity is considered belonging to the area, and a cut-off parameter P_c under which entities are not taken into account for the linear regression providing the scaling exponent. Remarquable results are that exponents can significantly vary and move from infra-linear to supra-linear when threshold varies. A systematic exploration of parameter space produces phase diagrams of exponents for various quantities. One question raising immediately is how these variation can be explained by the features of spatial distribution of variables. Do they result from intrinsic mechanisms present in the system or can they be explained more simply by the fact that the system is particularly spatialized ? We propose to prove by the tractation of a toy analytical model that even simple distributions can lead to such significant variations in the exponents, along one dimension of parameters (density threshold), directing the response towards the second explanation.

3.3.2 *Formalization*

We formalize the simple theoretical context in which we will derive the sensitivity of scaling to city definition. Let consider a polycentric city system, which spatial density distributions can be reasonably constructed as the superposition of monocentric fast-decreasing spatial kernels, such as an exponential mixture model [5]. Taking a geographical space as $\mathbb{R}^2$, we take for any $\vec{x} \in \mathbb{R}^2$ the density of population as

$$d(\vec{x}) = \sum_{i=1}^{N} d_i(\vec{x}) = \sum_{i=1}^{N} d_i^0 \cdot \exp\left(\frac{-\|\vec{x} - \vec{x}_i\|}{r_i}\right) \qquad (1)$$

where r_i are spread parameters of kernels, d_i^0 densities at origins, $\vec{x}_i$ positions of centers. We furthermore assume the following constraints :

1. To simplify, cities are monocentric, in the sense that for all $i \neq j$, we have $\|\vec{x}_i - \vec{x}_j\| \gg r_i$.
2. It allows to impose structural scaling in the urban system by the simple constraint on city populations P_i. One can compute by integration that $P_i = 2\pi d_i^0 r_i^2$, what gives by injection into the scaling hypothesis $\ln P_i = \ln P_{max} - \alpha \ln i$, the following relation between parameters : $\ln\left[d_i^0 r_i^2\right] = K' - \alpha \ln i$.

To study scaling relations, we consider a random scalar spatial variable $a(\vec{x})$ representing one aspect of the city, that can be everything but has the dimension of a spatial density, such that the indicator $A(D) = \mathbb{E}\left[\iint_D a(\vec{x}) d\vec{x}\right]$ represents the expected quantity of a in area D. We make the assumption that $a \in \{0; 1\}$ ("counting" indicator) and that its law is given by $\mathbb{P}[a(\vec{x}) = 1] = f(d(\vec{x}))$. Following the empirical work done in [59], the integrated indicator on city i as a function of θ is given by

$$A_i(\theta) = A(D(\vec{x}_i, \theta))$$

where $D(\vec{x}_i, \theta)$ is the area centered in $\vec{x}_i$ where $d(\vec{x}) > \theta$. Assumption 1 ensures that the area are roughly disjoint circles. We take furthermore a simple amenity such that it follows a local scaling law in the sense that $f(d) = \lambda \cdot d^\beta$. It seems a reasonable assumption since it was shown that many urban variable follow a fractal behavior at the intra-urban scale [110] and that it implies necessarily a power-law distribution [51]. We make the additional assumption that $r_i = r_0$ does not depend on i, what is reasonable if the urban system is considered from a large scale. This assumption should be relaxed in numerical simulations. The estimated scaling exponent $\alpha(\theta)$ is then the result of the log-regression of $(A_i(\theta))_i$ against $(P_i(\theta))_i$ where $P_i(\theta) = \iint_{D(\vec{x}_i, \theta)} d$.

3.3.3 *Analytical Derivation of Sensitivity*

With above notations, let derive the expression of estimated exponent for quantity a as a function of density threshold parameter θ. The quantity computed for a given city i is, thanks to the monocentric assumption and in a spatial range and a range for θ such that

$\theta \gg \sum_{j\neq i} d_j(\vec{x})$, allowing to approximate $d(\vec{x}) \simeq d_i(\vec{x})$ on $D(\vec{x}_i, \theta)$, is computed by

$$A_i(\theta) = \lambda \cdot \iint_{D(\vec{x}_i,\theta)} d^\beta = 2\pi\lambda {d_i^0}^\beta \int_{r=0}^{r_0 \ln \frac{d_i^0}{\theta}} r \exp\left(-\frac{r\beta}{r_0}\right) dr$$
$$= \frac{2\pi {d_i^0}^\beta r_0^2}{\beta^2}\left[1 + \beta \ln \frac{\theta}{d_i^0}\left(\frac{\theta}{d_i^0}\right)^\beta - \left(\frac{\theta}{d_i^0}\right)^\beta\right]$$

We obtain in a similar way the expression of $P_i(\theta)$

$$P_i(\theta) = 2\pi d_i^0 r_0^2 \left[1 + \ln\left[\frac{\theta}{d_i^0}\right]\frac{\theta}{d_i^0} - \frac{\theta}{d_i^0}\right]$$

The Ordinary-Least-Square estimation, solving the problem $\inf_{\alpha,C} \|(\ln A_i(\theta) - C - \alpha \ln P_i(\theta))_i\|^2$, gives the value $\alpha(\theta) = \frac{\text{Cov}[(\ln A_i(\theta))_i,(\ln P_i(\theta))_i]}{\text{Var}[(\ln P_i(\theta))_i]}$. As we work on city boundaries, threshold is expected to be significantly smaller than center density, i.e. $\theta/d_i^0 \ll 1$. We can develop the expression in the first order of θ/d_i^0 and use the global scaling law for city sizes, what gives $\ln A_i(\theta) \simeq K_A - \alpha \ln i + (\beta - 1)\ln d_i^0 + \beta \ln \frac{\theta}{d_i^0}\left(\frac{\theta}{d_i^0}\right)^\beta$ and $\ln P_i(\theta) = K_P - \alpha \ln i + \ln\left[\frac{\theta}{d_i^0}\right]\frac{\theta}{d_i^0}$. Developing the covariance and variance gives finally an expression of the scaling exponent as a function of θ, where $k_j, k_j{}'$ are constants obtained in the development :

$$\alpha(\theta) = \frac{k_0 + k_1\theta + k_2\theta^\beta + k_3\theta^{\beta+1} + k_4\theta\ln\theta + k_5\theta^\beta\ln\theta + k_6\theta^\beta(\ln\theta)^2 + k_7\theta^{\beta+1}(\ln\theta)^2 + k_8\theta^{\beta+1}\ln\theta}{k_0' + k_1'\ln\theta + k_2'\theta\ln\theta + k_3'\theta^2 + k_4'\theta^2\ln\theta + k_5'\theta^2(\ln\theta)^2} \tag{2}$$

This rational fraction predicts the evolution of the scaling exponent when the threshold varies. We study numerically its behavior in the next section, among other numerical experiments.

3.3.4 *Numerical Simulations*

IMPLEMENTATION We implement empirically the density model given in section 3.3.2. Centers are successively chosen such that in a given region of space only one kernel dominates in the sense that the sum of other contributions are above a given threshold θ_e. In practice, adapting N to world size allows to respect the monocentric condition. Population are distributed in order to follow the scaling law with fixed α and r_i (arbitrary choice) by computing corresponding d_i^0. Technical details of the implementation done in R [169] and using the package `kernlab` for efficient kernel mixture methods [108] are given as comments in source code[3]. We show in figure 2 example of synthetic density distributions on which the numerical study is

3 available at `https://github.com/JusteRaimbault/CityNetwork/tree/master/Models/Scaling`

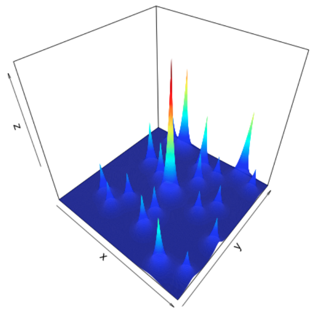


Figure 2: Example of a synthetic density distribution obtained with the exponential mixture, with a grid of size 400×400 and parameters $N = 20$, $r_0 = 10$, $P_{max} = 200$, $\alpha = 0.5$, $\theta_C = 0.01$.

conducted. The validation of theoretical results on these experimental mixtures must still be conducted, along with sensitivity tests to random perturbations, influence of kernel type, and two-parameters phase diagram when adding in the computational model functional density distribution and associated cut-off threshold.

RANDOM PERTURBATIONS The simple model used is quite reducing for maximal densities and radius distribution. We aim to proceed to an empirical study of the influence of noise in the system by fixing d_i^0 and r_i the following way :

- d_i^0 follows a reversed log-normal distribution with maximal value being a realistic maximal density
- Radiuses are computed to respect rank-size law and then perturbed by a white noise.

KERNEL TYPE We shall test the influence of the type of spatial kernel used on results. We can test gaussian kernels and quadratic kernels with parameters within reasonable ranges analog to the exponential kernel.

3.4 STATISTICAL CONTROL ON INITIAL CONDITIONS BY SYNTHETIC DATA GENERATION

3.4.1 *Context*

When evaluating data-driven models, or even more simple partially data-driven models involving simplified parametrization, an unavoidable issue is the lack of control on "underlying system parameters" (what is a ill-defined notion but should be seen in our sense as parameters governing system dynamics). Indeed, a statistics extracted from running the model on enough different datasets can become strongly biased by the presence of confounding in the underlying real data, as it is impossible to know if result is due to processes the model tries to translate or to a hidden structure common to all data.

We formalize briefly a proposition of method that would allow to add controls on meta-parameters, in the sense of parameters driving the represented system at a higher temporal and spatial scale, for a model of simulation. We make the hypothesis that such method is valid under constraints of disjonction for scales and/or ontologies between the model of simulation and the domain of meta-parameters.

3.4.2 *Description*

An advanced knowledge of the behavior of computational models on their parameter space is a necessary condition for deductions of thematic conclusions or their practical application [10]. But the choice of varying parameters is always subjective, as some may be fixed by a real-world parametrization, or other may be interpreted as arbitrarily fixed initial conditions. It raises methodological and epistemological issues for the sensitivity analysis, as the scope of the model may become ill-defined.

Let consider the concrete example of the Schelling Segregation model [187]. One of its crucial features on which the literature has been rather controversial is the influence of the spatial structure of the container on which agents evolve.The thematic aim of the project developed in [61] is to clarify this point through a systematic model exploration. A methodological contribution is the construction of a framework allowing the analysis of the sensitivity of models to *meta-parameters*, i.e. to parameters considered as fixed initial conditions (e.g. the spatial structure for the Schelling model), or to parameters of another model generating an initial configurationyielding thus a *simple coupling* between models (serial coupling). The benefits of such an approach are various but include for example the knowledge of model behavior in an extended frame, the possibility of statistical control when regressing model outputs, a finer exploration of model derivatives than with a naive approach. Some remarks can be made on the approach :

- What knowledge are brought by adding the upstream model, rather than for example in the Schelling case exploring a large set of initial geometries ?

  $\rightarrow$ to obtain a sufficiently large set of initial configuration, one quickly needs a model to generate them ; in that case a quasi-random generation followed by a filtering on morphological constraint will be a morphogenesis model, which parameters are the ones of the generation and the filtering methods. Furthermore, as detailed further, the determination of the derivative of the downstream model is made possible by the coupling and knowledge of the upstream model.

- Statistical noise is added by coupling models

  $\rightarrow$ Repetitions needed for convergence are indeed larger as the final expectance has to be determined by repeating on the first times the second model ; but it is exactly the same as exploring directly many configuration, to obtain statistical robustness in that case one must repeat on similar configurations.

- Complexity is added by coupling models

  $\rightarrow$ In the sense of Varenne [201] , coupling is simple and no complexity is thus added.

3.4.3 *Formal Description*

One has the composition of the derivative along the meta-parameter

$$\partial_\alpha [M_u \circ M_d] = (\partial_\alpha M_u \circ M_d) \cdot \partial_\alpha M_d$$

$\rightarrow$ the sensitivity of the downstream model (Schelling) can be determined by studying the serial coupling and the upstream model ; thematic knowledge : sensitivity to an implicit meta-parameter ; and computational gain : generation of controlled differentiates in the "initial space" is quasi impossible.

The question of stochasticity in simply coupled models causes no additional issue as $\mathbb{E}[X] = \mathbb{E}[\mathbb{E}[X|Y]]$. It naturally multiplies the number of repetition needed for convergence what is the expected behavior.

3.5 LINKING DYNAMIC AND STATIC SPATIO-TEMPORAL CORRELATIONS UNDER SIMPLIFIED ASSUMPTIONS

Space and Time are both crucial for the study of geographical systems when aiming to understand *processes* (by definition dynamical [106]) evolving in a *spatial structure* in the sense of [72]. Space is more than coordinates for elements of the system, but a dimension in itself that drives interactions and thus system properties. Reading geographical systems from the point of view of *spatio-temporal processes* emphasizes the fact that *space actually matters*. Space and time are closely linked in such processes, and depending on the underlying mechanisms, one can expect to extract useful information from one on the other : in certain cases that we will investigate in this part, it is for example possible to learn about dynamics from static information. Spatio-temporal correlations approaches, linked to spatio-temporal dynamics, are present in very broad fields such as dynamical image processing (including video compression) [45, 100, 109], target tracking [25, 204], climate science [63], Earth sciences [135], city systems dynamics [102, 157], among others.

The capture of neighborhood effects in statistical models is a wisely used practice in spatial statistics, as the technique of Geographically Weighted Regression illustrates [43]. A possible interpretation among many definitions of spatial autocorrelation [92] yields that by estimating a plausible characteristic distance for spatial correlations or auto-correlations, one can isolate independent effects between variables from effects due to neighborhood interactions[4]. The study of the spatial covariance structure is a cornerstone of advanced spatial statistics that was early formulated [91]. We propose now to study possible links between spatial and temporal correlations, using spatio-temporal covariance structure to infer information on dynamical processes.

3.5.1 *Notations*

We consider a multivariate spatio-temporal stochastic process denoted by $\vec{Y}[\vec{x}, t]$. At a given point $\vec{x}_0$ in space, we can define temporal covariance structure by

$$\mathbf{C}_t(\vec{x}_0) = \mathrm{Var}\left[\vec{Y}[\vec{x}_0, \cdot]\right]$$

and spatial covariance structure at fixed time by

$$\mathbf{C}_x(t) = \mathrm{Var}\left[\vec{Y}[\cdot, t]\right]$$

It is clear that these quantities will be in practice first ill-defined because of the difficulty in interpreting such a process by a spatio-

4 note that the formal link between models of spatial autocorrelation (see e.g. [93]) is not clear and should be further investigated

temporal random variable, secondly highly non-stationary in space and time. We stay however at a theoretical level to gain structural knowledge, reviewing simple cases in which a formal link can be established.

3.5.2 *Wave Equation*

In the case of propagating waves, there is an immediate link. Let assume that a wave equation if verified by "deterministic" parts of components

$$c^2 \cdot \partial_t^2 \bar{Y}_i = \Delta \bar{Y}_i \tag{3}$$

with $Y_i = \bar{Y}_i + \varepsilon_i$. If errors are uncorrelated and processes are stationary, we have then directly

$$\mathbf{C}_t \left[\partial_t^2 Y_i, \partial_t^2 Y_j\right] = \frac{1}{c^2} \cdot \mathbf{C}_x \left[\Delta Y_i, \Delta Y_j\right] \tag{4}$$

This gives us however few insight on real systems as local diffusion, stationary assumptions and uncorrelated noises are far from being verified in empirical situations.

3.5.3 *Fokker-Planck Equation*

An other interesting approach may when the process verifies a Fokker-Planck equation on probabilities of the state of the system when it is given by its position (diffusion of particles in that case)

$$\partial_t P(x_i, t) = -d \cdot \partial_x P(x_i, t) + \frac{\sigma^2}{2} \partial_x^2 P(x_i, t) \tag{5}$$

With no cross-correlation terms in the Fokker-Planck equation, covariance between processes vanish. We have finally in that case only a relation between averaged spatial and temporal variances that brings no information to our question.

3.5.4 *Master Equation*

In the case of a master equation on probabilities of discrete states of the system

$$\partial_t \vec{P} = \mathbb{W} \vec{P} \tag{6}$$

we have then for state i, $\partial_t P_i = \sum_j W_{ij} P_j$. As this relation is at a fixed time we can average in time to obtain an equation on temporal covariance. It is not clear how to make the link with spatial covariance as these will depend on spatial specification of discrete states. This question is still under investigation.

3.5.5 *Consistent spatio-temporal sampling*

In a more empirical way, we propose to not assume any contraint of process dynamics but to however investigate how the computation of spatial correlations can inform on temporal correlations. We try to formulate easily verifiable assumptions under which this is possible.

We make the following assumptions on the spatio-temporal stochastic processes $Y_i\left[\vec{x}, t\right]$:

1. Local spatial autocorrelation is present and bounded by l_ρ (in other words the processes are continuous in space) : at any $\vec{x}$ and t, $\left|\rho_{\|\Delta\vec{x}\|<l_\rho}\left[Y_i(\vec{x}+\Delta\vec{x}, t), Y_i(\vec{x}, t)\right]\right| > 0$.
2. Processes are locally parametrized : $Y_i = Y_i\left[\alpha_i\right]$, where $\alpha_i(\vec{x})$ varies with l_α, with $l_\alpha \gg l_\rho$.
3. Spatial correlations between processes have a sense at an intermediate scale l such that $l_\alpha \gg l \gg l\rho$.
4. Processes covariance stationarity times scale as $\sqrt{l}$.
5. Local ergodicity is present at scale l and dynamics are locally chaotic.

Assumptions one to three can be tested empirically and allow to compare spatial correlation estimated on spatial samplings at scale l. Assumption four is more delicate as we are precisely constructing this methodology because we have no temporal information on processes. It is however typical of spatial diffusion processes, and population or innovation diffusion should verify this assumption. The last assumption can be tested if feasible space is known, by checking cribbing on image space on the spatial sample. Under these conditions, local spatial sampling is equivalent to temporal sampling and spatial correlation estimators provide estimator of temporal correlations.

4 QUANTITATIVE EPISTEMOLOGY

something on knowledge

-

A corollary of theoretical background proposed in chapter 2 is the need of an understanding of involved disciplines themselves to be able to build integrated heterogeneous models. The potentialities of couplings and integrations are greatly determined by existing approaches and corresponding gaps. This implies an advanced epistemological study in each field, that we propose to tackle in a systematic and quantitative way. This deliberate choice may shadow elaborated epistemological considerations but fits of purpose of preliminary investigations for the construction of models.

4.1 ALGORITHMIC SYSTEMATIC REVIEW

A broad bibliographical study suggests a scarcity of quantitative models of simulation integrating both network and urban growth. This absence may be due to diverging interests of concerned disciplines, resulting in a lack of communication. We propose to proceed to an algorithmic systematic review to give quantitative elements of answer to this question. A formal iterative algorithm to retrieve corpuses of references from initial keywords, based on text-mining, is developed and implemented. We study its convergence properties and do a sensitivity analysis. We then apply it on queries representative of the specific question, for which results tend to confirm the assumption of disciplines compartmentalization.

4.1.1 *In search of models of co-evolution*

Transportation networks and urban land-use are known to be strongly coupled components of urban systems at different scales [42]. One common approach is to consider them as co-evolving, avoiding misleading interpretations such as the myth of structural effect of transportation infrastructures [149]. A question rapidly arising is the existence of models endogeneizing this co-evolution, i.e. taking into account simultaneous urban and network growth. We try to answer it using an algorithmic systematic review. We propose in this section, after a brief state of the art of existing literature, to present such an approach by formalizing the algorithm, which results are then presented and discussed.

4.1.2 *Modeling Interactions between Urban Growth and Network Growth : An Overview*

Land-Use Transportation Interaction Models.

A wide class of models that have been developed essentially for planning purposes, which are the so-called Land-use Transportation Interaction Models, is a first type answering our research question. See recent reviews [47], [107] and [206] to get an idea of the heterogeneity of included approaches, that exist for more than 30 years. Recent models with diverse refinements are still developed today, such as [69] which includes housing market for Paris area. Diverse aspects of the same system can be translated into many models (as e. g.[207]), and traffic, residential and employment dynamics, resulting land-use evolution, influenced also by a static transportation network, are generally taken into account.

Network Growth Approaches

On the contrary, many economic literature has done the opposite of previous models, i.e. trying to reproduce network growth given assumptions on the urban landscape, as reviewed in [222]. In [215], economic empirical studies are positioned within other network growth approaches, such as work by physicists proposing model of geometrical network growth [14]. Analogy with biological networks was also done, reproducing typical robustness properties of transportation networks [198].

Hybrid Approaches

Fewer approaches coupling urban growth and network growth can be found in the literature. [15] couples density evolution with network growth in a toy model. In [174], a simple Cellular Automaton coupled with an evolutive network reproduces stylized facts of human settlements described by Le Corbusier. At a smaller scale, [1] proposes a model of co-evolution between roads and buildings, following geometrical rules. These approaches stay however limited and rare.

4.1.3 *Bibliometric Analysis*

Literature review is a crucial preliminary step for any scientific work and its quality and extent may have a dramatic impact on research quality. Systematic review techniques have been developed, from qualitative review to quantitative meta-analyses allowing to produce new results by combining existing studies [180]. Ignoring some references can even be considered as a scientific mistake in the context of emerging information systems [127]. We aim to take advantage of such techniques to tackle our issue. Indeed, observing the form of the bibliography obtained in previous section raises some hypothesis. It is clear that all components are present for co-evolutive models to exist but different concerns and objectives seem to stop it. As it was shown by [57] for the concept of mobility, for which a "small world of actors" relatively closed invented a notion ad hoc, using models without accurate knowledge of a more general scientific context, we could be in an analog case for the type of models we are interested in. Restricted interactions between scientific fields working on the same objects but with different purposes, backgrounds and at different scales, could be at the origin of the relative absence of co-evolving models. While most of studies in bibliometrics rely on citation networks [144] or co-autorship networks [186], we propose to use a less explored paradigm based on text-mining introduced in [48], that obtain a dynamic mapping of scientific disciplines based on their semantic content. For our question, it has a particular interest, as we want to understand con-

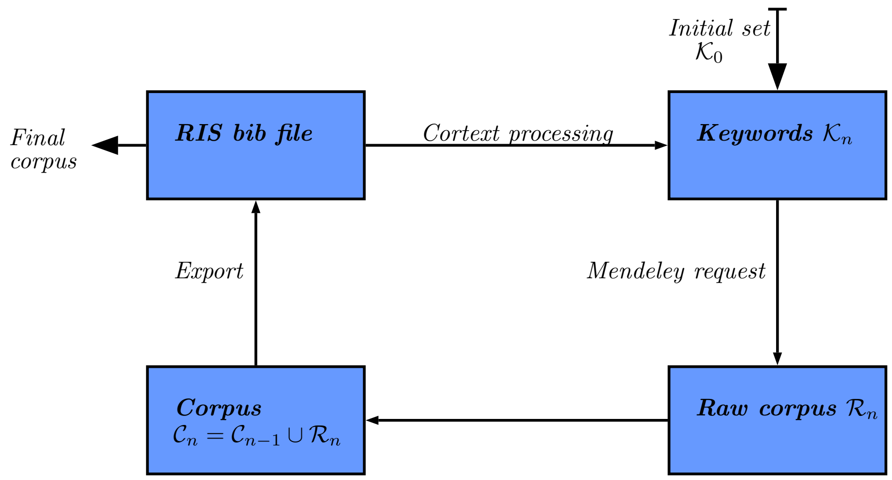


Figure 3: Global workflow of the algorithm, including implementation details : catalog request is done through Mendeley API ; final state of corpuses are RIS files.

tent structure of researches on the subject. We propose to apply an algorithmic method described in the following. The algorithm proceeds by iterations to obtain a stabilized corpus from initial keywords, reconstructing scientific semantic landscape around a particular subject.

Description of the Algorithm

Let A be an alphabet, A^* corresponding words and $T = \cup_{k \in \mathbb{N}} A^{*k}$ texts of finite length on it. A reference is for the algorithm a record with text fields representing title, abstract and keywords. Set of references at iteration n will be denoted $\mathcal{C} \subset T^3$. We assume the existence of a set of keywords $\mathcal{K}_n$, initial keywords being $\mathcal{K}_0$. An iteration goes as follows :

1. A raw intermediate corpus $\mathcal{R}_n$ is obtained through a catalog request providing previous keywords $\mathcal{K}_{n-1}$.
2. Overall corpus is actualized by $\mathcal{C}_n = \mathcal{C}_{n-1} \cup \mathcal{R}_n$.
3. New keywords $\mathcal{K}_n$ are extracted from corpus through Natural Language Processing treatment, given a parameter N_k fixing the number of keywords.

The algorithm stops when corpus size becomes stable or a user-defined maximal number of iterations has been reached. Fig. 1 shows the global workflow.

Results

IMPLEMENTATION Because of the heterogeneity of operations required by the algorithm (references organisation, catalog requests,

Corpuses	1	2	3	4	5
1 (W=3789)	1	0	0.0719	0.0078	0.0724
2 (W=5180)	0	1	0.0338	0	0.0125
3 (W=3757)	0.0719	0.0338	1	0.0100	0.1729
4 (W=3551)	0.0078	0	0.0100	1	0.0333
5 (W=8338)	0.0724	0.0125	0.1729	0.0333	1

Table 1: Symmetric matrix of lexical proximities between final corpuses, defined as the sum of overall final keywords co-occurrences between corpuses, normalized by number of final keywords (100). We obtain very low values, confirming that corpuses are significantly far. Size of final corpuses is given as W.

text processing), it was found a reasonable choice to implement it in Java. Source code is available on the Github repository of the project1. Catalog request, consisting in retrieving a set of references from a set of keywords, is done using the Mendeley software API [141] as it allows an open access to a large database. Keyword extraction is done by Natural Language Processing (NLP) techniques, following the workflow given in [48], calling a Python script that uses [31].

CONVERGENCE AND SENSITIVITY ANALYSIS A formal proof of algorithm convergence is not possible as it will depend on the empirical unknown structure of request results and keywords extraction. We need thus to study empirically its behavior. Good convergence properties but various sensitivities to Nk were found as presented in Fig. 2. We also studied the internal lexical consistence of final corpuses as a function of keywords number. As expected, small number yields more consistent corpuses, but the variability when increasing stays reasonable.

Once the algorithm is partially validated, we apply it to our question. We start from five different initial requests that were manually extracted from the various domains identified in the manual bibliography (that are “city system network”, “land use transport interaction”, “network urban modeling”, “population density transport”, “transportation network urban growth”). We take the weakest assumption on parameter $N_k = 100$, as it should less constrain reached domains. After having constructed corpuses, we study their lexical distances as an indicator to answer our initial question. Large distances would go in the direction of the assumption made in section 2, i.e. that discipline self-centering may be at the origin of the lack of interest for co-evolutive models. We show in Table 1 values of relative lexical proximity, that appear to be significantly low, confirming this assumption.

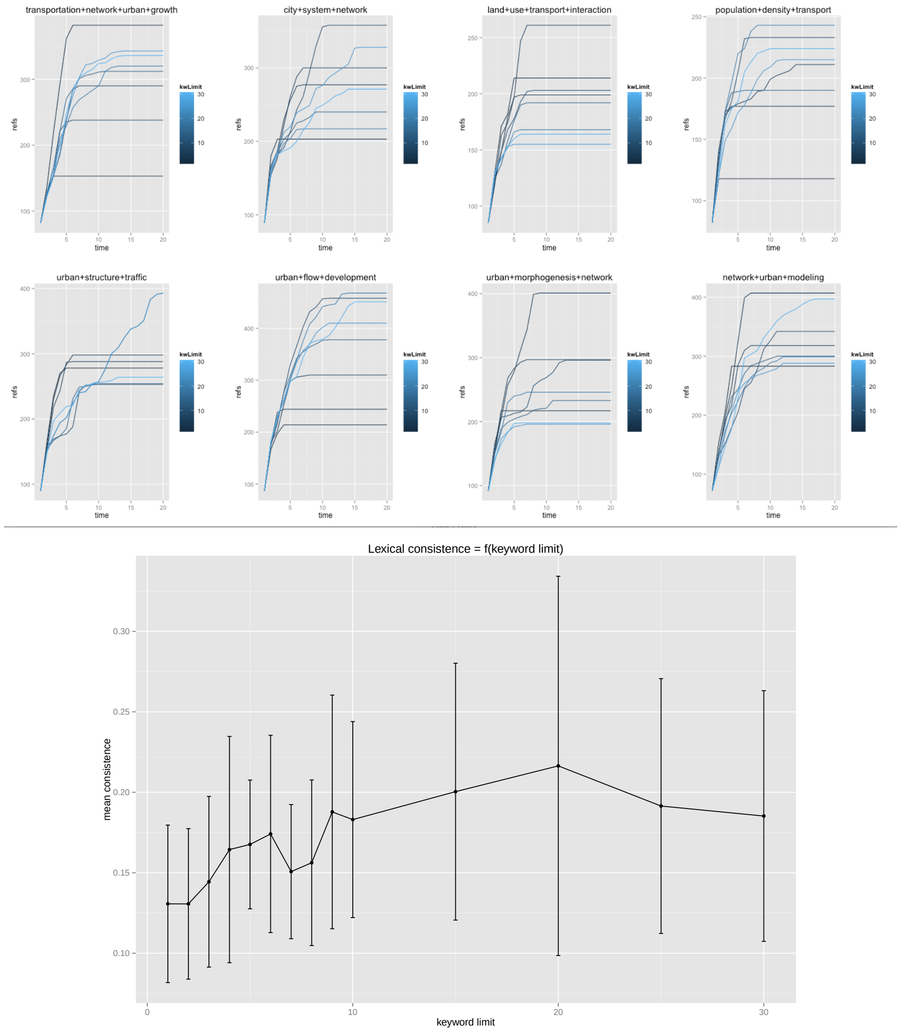


Figure 4: Convergence and sensitivity analysis. Left : Plots of number of references as a function of iteration, for various queries linked to our theme (see further), for various values of N_k (from 2 to 30). We obtain a rapid convergence for most cases, around 10 iterations needed. Final number of references appears to be very sensitive to keyword number depending on queries, what seems logical since encountered landscape should strongly vary depending on terms. Right : Mean lexical consistence and standard error bars for various queries, as a function of keyword number. Lexical consistence is defined though co-occurrences of keywords by, with N final number of keywords, f final step, and $c(i)$ co-occurrences in references, $k = \frac{2}{N(N-1)} \cdot \sum_{i,j\in\mathcal{K}_f} |c(i)-c(j)|$. The stability confirms the consistence of final corpuses.

Further work is planned towards the construction of citation networks through an automatic access to Google Scholar that provides backward citations. The confrontation of inter-cluster coefficients on the citation network for the different corpuses with our lexical consistence results are an essential aspect of a further validation of our results.

The disturbing absence of models simulating the co-evolution of transportation networks and urban land-use, confirmed through a state-of-the-art covering many domain, may be due to the absence of communication between scientific disciplines studying different aspects of that problems. We have proposed an algorithmic method to give elements of answers through text-mining-based corpus extraction. First numerical results seem to confirm the assumption. However, such a quantitative analysis should not be considered alone, but rather come as a back-up for qualitative studies that will be the object of further work, such as the one lead in [57], in which questionnaires with historical actors of modeling provide highly relevant information.

4.2 REFINING BIBLIOMETRICS THROUGH HYPER-NETWORK ANALYSIS

4.2.1 *Context*

As described before, semantic analysis does not contain all the information on disciplinary compartmentation nor on patterns of propagation of scientific knowledge as the ones contained in citation networks for example. Furthermore, data collection in the previous algorithm is subject to convergence towards self-consistent themes because of the proper structure of the method. It may give more insight about scientific social patterns of ontological choices in modeling to study communities in broader networks, that would more correspond to disciplines (or sub-disciplines depending on granularity level).

Previous works in quantitative epistemology using various types of networks have shown interesting potentialities. For the citation network, a good predicting power for citation patterns is for example obtained in [144]. Co-authorship networks can also be used for predictive models [186]. A multilayer network approach was recently proposed in [151], using bipartites networks of papers and scholars, in order to produce measures of interdisciplinarity. Disciplines can be stratified into layers to reveal communities between them and therein collaboration patterns [17].

[210]

4.2.2 *Application to a scientific journal*

Presentation

We briefly describe here an ongoing study that implemented the ideas given above for the particular case of a scientific journal for which bibliographical data is difficult to obtain, that is `cybergeo`, an electronic journal in theoretical and quantitative geography.

[96] [155] [54] [192]

Implementation

The general architecture for data collection is presented in Fig. 5. Citation data is collected from `Google Scholar`, that is the only source for incoming citations [147] in our case as the journal is not referenced in other databases. We are aware of the possible biaises using this single source [32][1], but these critics are more directed towards search results than citation counts.

Text processing is done the same way as in previous section, expect that a particular treatment is done to language detection using *stop-*

1 or see http://iscpif.fr/blog/2016/02/the-strange-arithmetic-of-google-scholars/

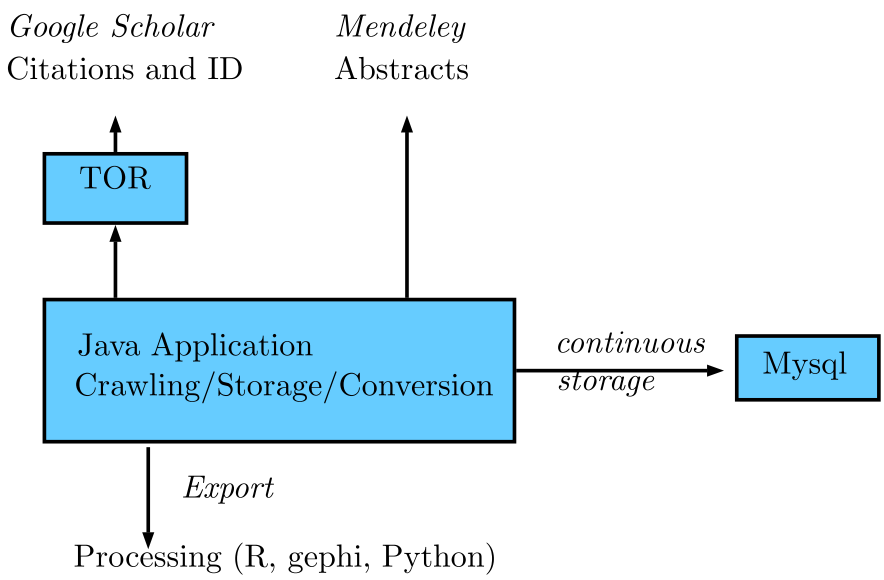


Figure 5: Heterogeneous Bibliographical Data Collection. Architecture of the application for content (semantic data), metadata and citation data collection.

words and a specific tagger `TreeTagger` is used for other languages than english [188].

Results

We show in figures 6 and 7 preliminary results on citation and semantic network.

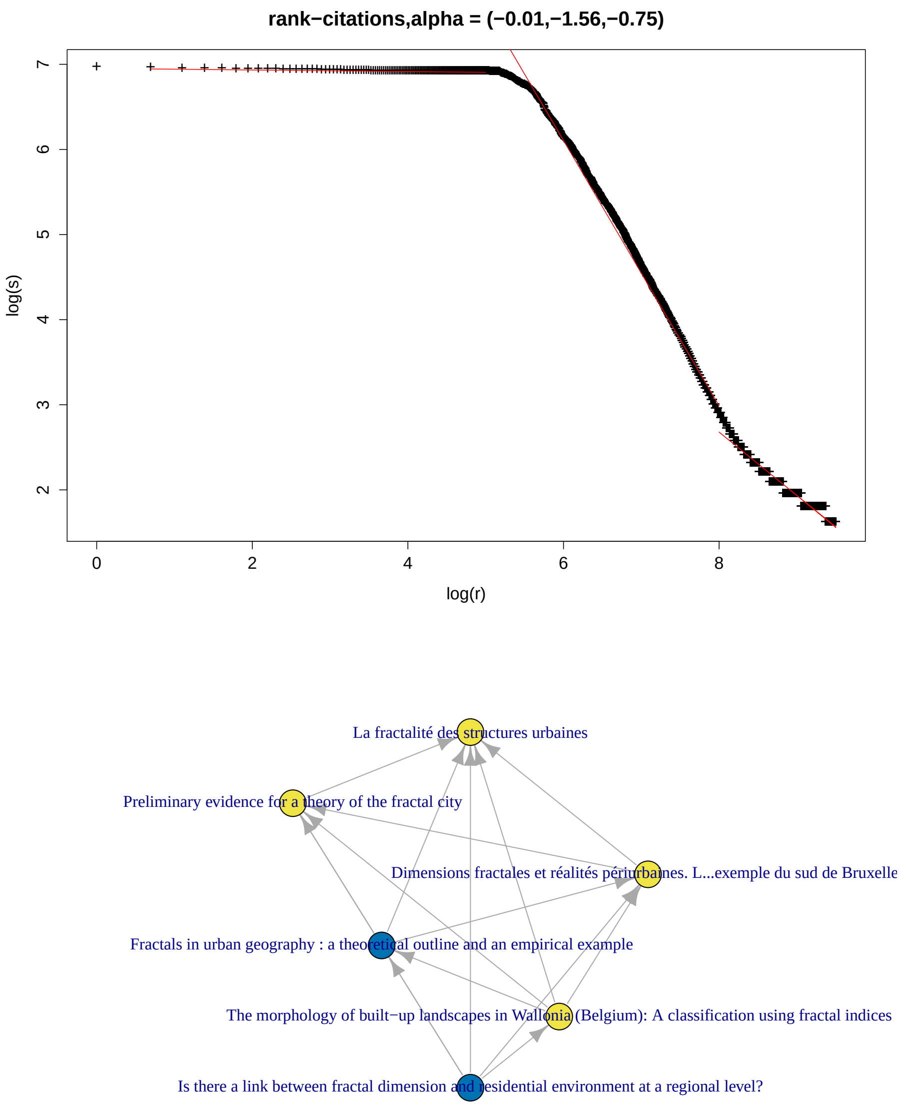


Figure 6: Properties of the citation network

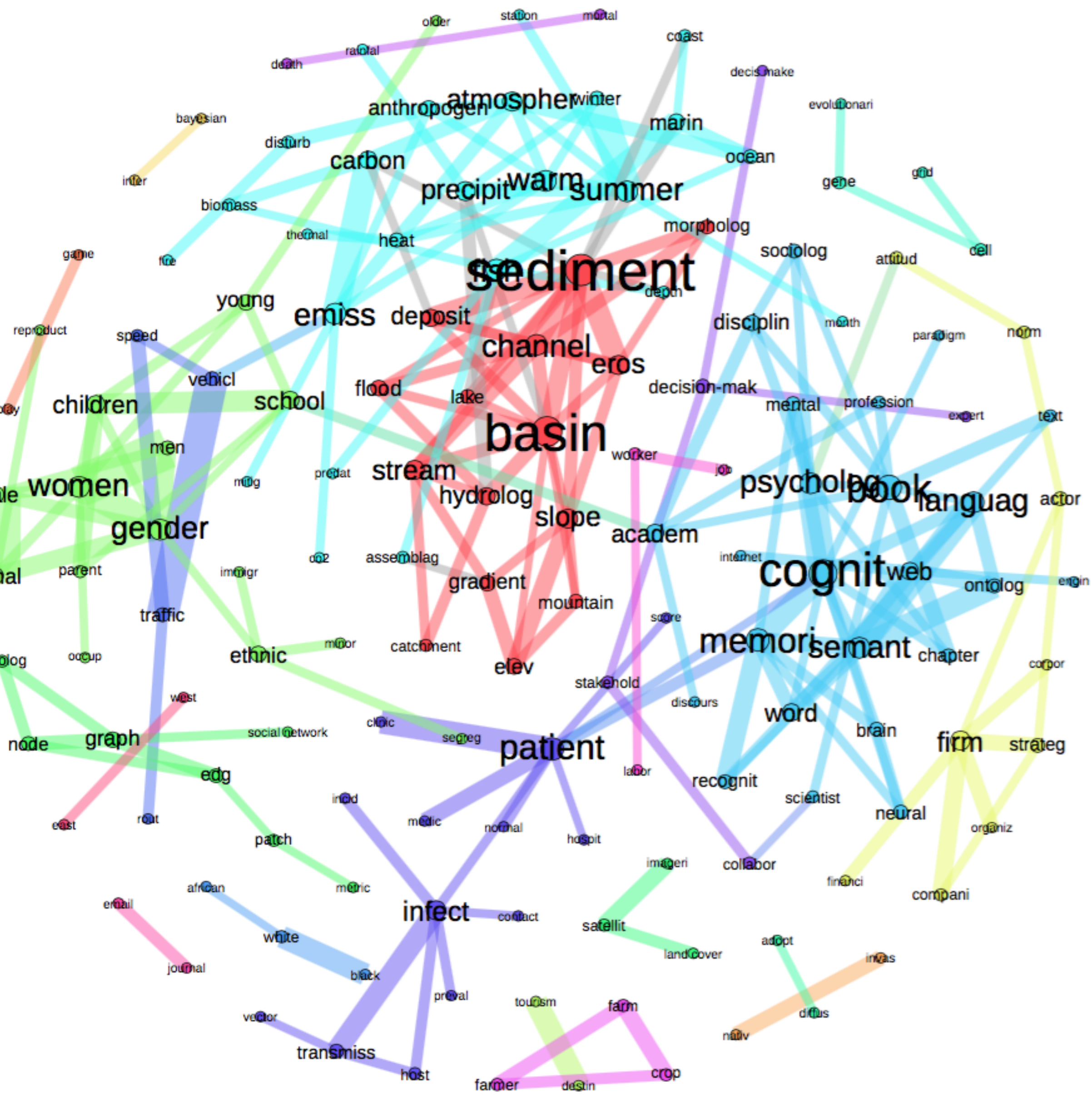


Figure 7: Semantic network of concepts in quantitative geography. Corpus consists of around $2 \cdot 10^5$ abstracts of publications at a topological distance shorter than 2 from the journal `cybergeo` in the citation network. Relevance of keywords were estimated with the bootstrap method,

4.3 TOWARDS MODELING PURPOSE AND CONTEXT AUTOMATIC EXTRACTION

A possible direction to strengthen our quantitative epistemological analysis would be to work on full textes related to the modeling of interaction between networks and territories, with the aim to automatically extract thematics within articles. The idea would be to perform some kind of automatized modelography, with possible features to be extracted that would be ontologies, model architecture or structures, scales, or even typical parameter values. It is not clear to what degree structure of models can be extracted from their description in papers and it surely depends on the discipline considered. For example in a framed field such as transportation planning, using a pre-defined ontology (in the sense of dictionary) and a fuzzy grammar could be efficient to extract information as the discipline is relatively formatted. In theoretical and quantitative geography, beyond the barrier of language, information organisation is surely less subject to unsupervised data-mining because of the more literary nature of the discipline : synonyms and style figures are generally the norm in good level human sciences writing, fuzzing a possible generic structure of knowledge description.

Part II

MODELING AND EMPIRICAL ANALYSIS

This part aims at producing knowledge from the empirical analysis of case studies and from first modeling experiments.

5 EMPIRICAL ANALYSIS : INSIGHTS FROM STYLIZED FACTS

Mais ce n'est pas une question d'âge, de chiffres et de stats
Moi je te parle surtout de rage, de kif et d'espoir
- Youssoupha , *Esperance de Vie*

As this quote suggests, a purely quantitative view of the world makes no sense without qualitative counterbalancing. More precisely, we argue that the *cliché* of an opposition between quantitative and qualitative analysis is an illusion. No distinct boundary exists between both. We propose to call quantitative any process involving computation by a Turing machine, whereas the qualitative will be for us the modeling design process and its interpretations. Therefore both are necessarily closely interlaced in any of our approaches. In particular concerning the construction and the validation or refutation of our theory, empirical analysis on real case studies, implying the extraction and qualification of stylized facts, follows that schema.

We propose in this chapter various empirical analysis on different objects at different scales.

5.1 STATIC CORRELATIONS OF URBAN FORM AND NETWORK SHAPE

Spatio-temporal processes implying diffusion or propagation phenomena generally have a specific structure of correlation. In particular, as derived in section 3.5, a static computation of correlation between different instances of a system may under certain conditions provide information on dynamical correlations implied.

5.1.1 *Morphological Measures of European Population Density*

Context

At the macroscopic scale of system of cities, spatialization of the urban system is reasonably captured by cities position, associated with aggregated city variable to represent entirely the system (see e.g. ontologies of Simpop models [165] or its successor Marius [58]). At the mesoscopic scale at which we aim to capture morphological manifestations of interactions between transportation networks and territories, structure of the territorial system can be specified by more refined indicators for the morphological aspect.

Empirical Analysis

We study systematically morphological indicators for constant size areas covering European Community. The choice of fixed size areas can be questioned regarding definition of a territorial system, that can be otherwise understood as a consistent spatial entity at a given scale and along certain criteria : *Human territories* as defined by Raffestin (op. cit.) or more generally functionally autonomous spaces[1].

Further developments

[195]
`http://www.worldpop.org.uk/`

5.1.2 *Network Measures*

We consider network aggregated indicators as a way to characterize transportation network properties on a given territory, the same way morphological indicators yielded information on urban structure. We propose to compute some simple indicators on same extents as for

1 for example, a tentative of definition of a *Parisian* territory would present many facets. From the subjective territory point of view, intra-muros Parisians consider a strict boundary at *Boulevard Periphérique*, whereas close and even further suburbs will be seen as Parisians from the Province. The functional territory of *Metropolitain* extends slightly further than the administrative boundary. Governance perimeters are currently mutating with the Metropolitan governance project. Complementary perceptions of the territory can thus be multiplied.

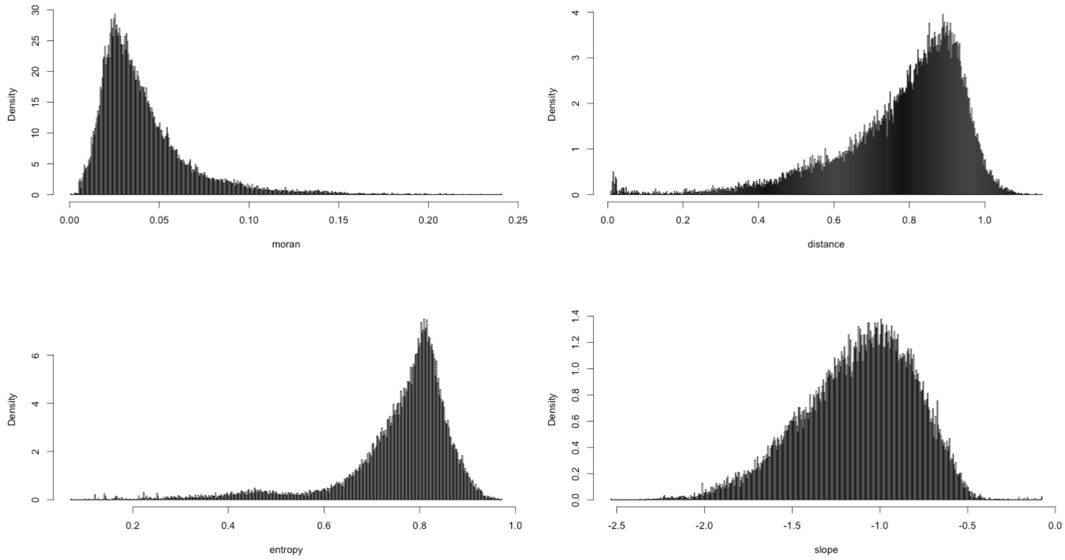


Figure 8: Empirical Distribution of Morphological Indicators

morphology, to be able to explore relations between these static measures.

Network analysises : [130]

Data preprocessing

We work in a first time on road network, which structure is finely conditioned to territorial configuration of population densities. Furthermore, data for present day road network is available through the OpenStreetMap project [152]. Its quality was investigated for different countries such as England [99] and France [88]. It was found to be of a quality equivalent to official surveys for the primary road network.

SIMPLIFICATION ALGORITHM For a given dataset corresponding to a subset of the overall road network, it is necessary to simplify network structure by spatial aggregation as initial data presents very detailed features and thus a very large numbers of nodes ($\simeq 10^10$ for Europe dataset). Such a level of precision is not needed in our study since density data is already aggregated at 500m resolution. It is possible to drastically reduce network size by spatial aggregation of nodes and link replacements. More precisely we use the following procedure :

- a background raster (which resolution r gives the snapping parameter for aggregation) is constructed from a reference raster and the extent of network. This grid gives spatial aggregation units for network nodes.

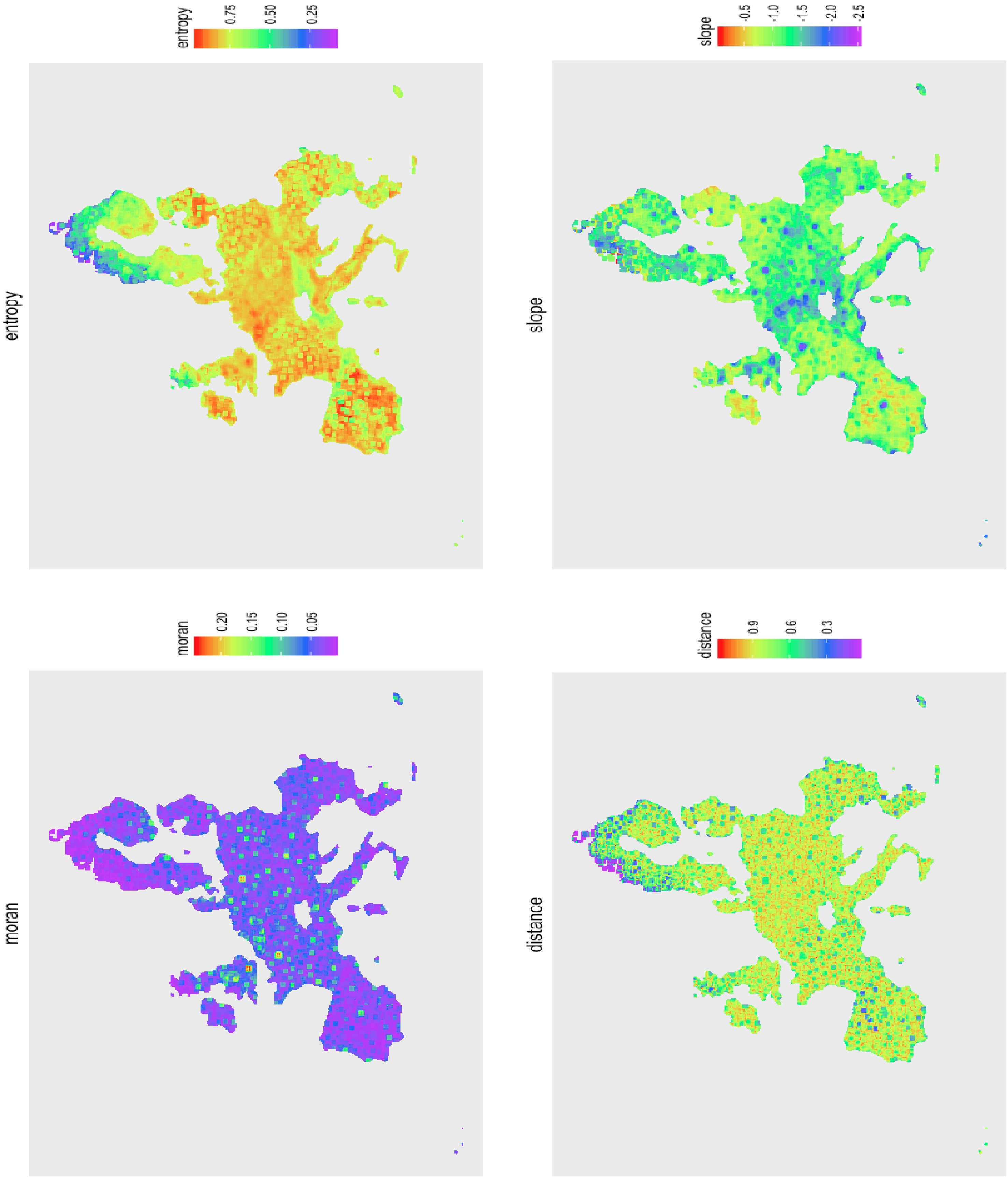


Figure 9: Geographical Distribution of Morphologies

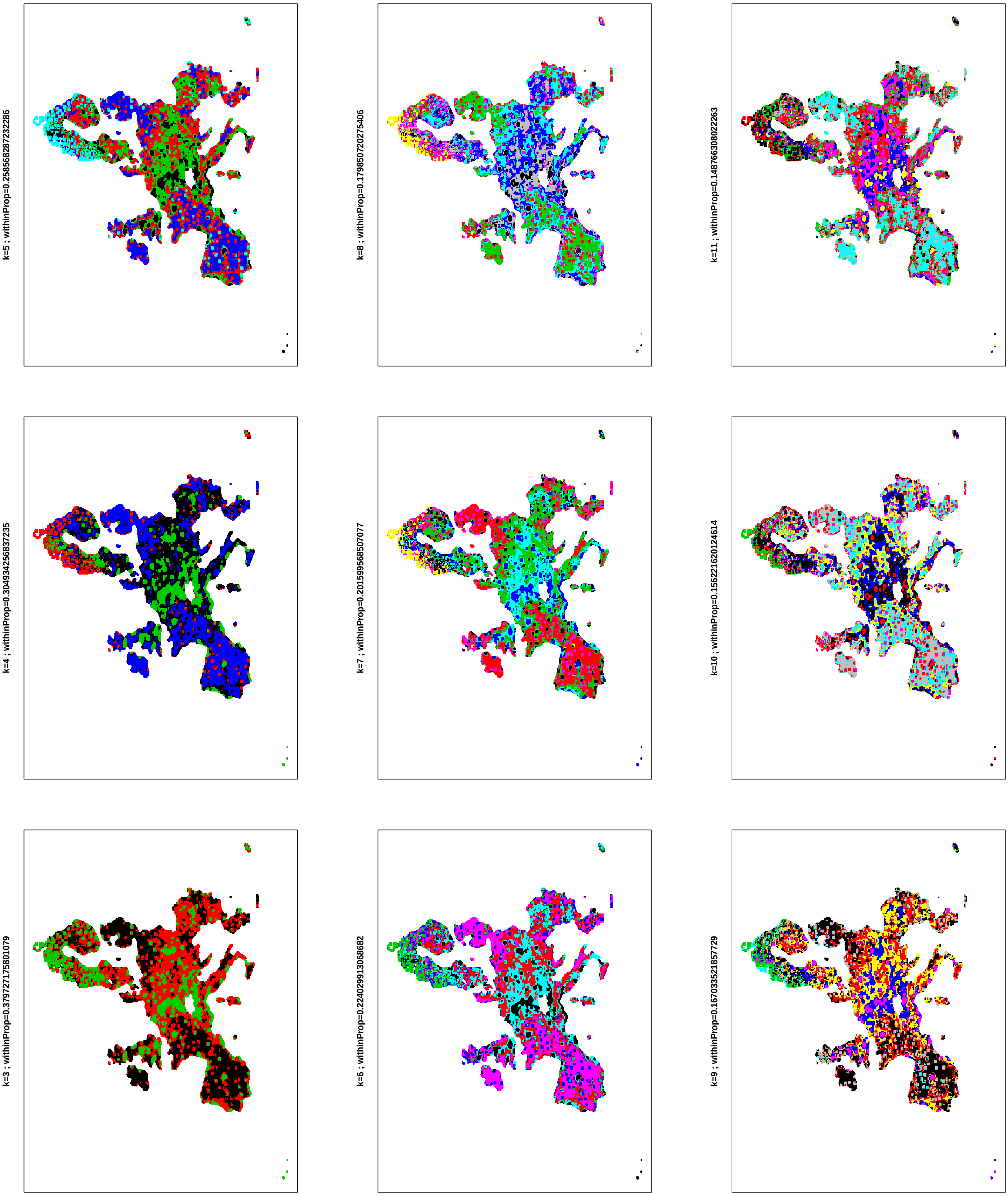


Figure 10: Clustering Analysis of Morphologies

- for each feature of the road dataset, corresponding connected raster cells are stored with corresponding impedance and distance in a sparse adjacency matrix.

IMPLEMENTATION A `PostGIS` database is used to store raw and simplified network, in order to perform efficient spatial requests, compared for example to initial `osm` data formats (`osm` or `pbf`). However

SENSITIVITY TO SIMPLIFICATION PARAMETERS

Indicators

Network macroscopic structure is summarized by the following set of indicators, after the simplifications and reductions done in the previous step. Assuming network given by $N = (V, E)$, nodes spatial positions $\vec{x}(V)$ and edges *effective distances* $d(E)$ taking into account impedances and real distances (to include basically network hierarchy), we have indicators :

- connectivity
- degree distribution
- centrality, taken as normalized mean *betweenness-centrality*
- average path length
- network diameter
- mean network speed

These indicators are used to capture a rough picture of the structure. Refined work at smaller scales (intra-urban road network) and with more elaborated measures that allow to differentiate more precisely local form, was recently done by Lagesse in [117].

Results

5.1.3 *Effective static correlations*

5.1.4 *Insights for interaction processes*

5.2 DISENTANGLING CO-EVOLUTIONS FROM CAUSAL RELATIONS : A CASE STUDY ON *bassin parisien*

[125]

5.2.1 *Context Formalization*

Variables

DESCRIPTION We assume a dynamic transportation network $n(\vec{x}, t)$ within a dynamic territorial landscape $\vec{T}(\vec{x}, t)$, which components are to simplify population $p(\vec{x}, t)$ and employments $e(\vec{x}, t)$. Data is structured the following way :

- Observation of territorial variables are discretized in space and in time, i.e. the spatial field $\vec{T}$ is summarized by $\mathbf{T} = \left(\vec{T}(\vec{x}_i, t_j^{(T)})\right)_{i,j}$ with $1 \leqslant i \leqslant N$ and $1 \leqslant j \leqslant T$. They concretely correspond to census on administrative units (*communes* in our case) at different dates.
- Network has a continuous spatial position but

DEFINITIONS

5.2.2 *On Accessibility*

The notion of accessibility has been central to regional science since its introduction and systematization in planning around 1970.

EXISTENCE OF ACCESSIBILITY ***Is the notion of accessibility crucial for statistical analysis ?***

Weibull has proposed an axiomatic approach to accessibility [208], deriving a canonical decomposition for any *attraction-accessibility* function $A(a, d)$, assuming expected thematic axioms among others technical ones that are :

1. A is invariant regarding the order of the configuration
2. A decrease with distance at fixed attraction and increase with attraction at fixed distance
3. A is invariant when adding null attractions and constant configurations

Then A verifies these *iff* it is of the form

$$A\left[(a_i, d_i)\right] = T\left(\bigoplus_i z(d_i, a_i)\right)$$

where T is increasing with null origin, z is a *distance substitution function* (i.e. verifying axiom 2) and $\oplus$ a *standard composition* associating two attractions at zero distance to th corresponding unique one.

$\rightarrow$ Well suited matrices of autocorrelation should capture accessibility in regressions ; or captured by non-linear regression on $\mathbf{N}$

Accessibility as potential ?

Given any stationary dynamic for $n, \vec{T}$, Helmoltz theorem states that it derives from a potential (can be adapted to non-stationary dynamics with time-varying potential).

CONTINUOUS APPROACH AND ACCESSIBILITY POTENTIAL

5.2.3 *Statistical Tests*

Large set of analysis to be tested (non exhaustive) :

- On data :
  - Multivariate models $\mathcal{L}\left[\mathbf{T},\mathbf{N}\right] \sim \varepsilon$
  - Autocorrelated univariate models $(\mathbf{I} - \Sigma RW)\mathbf{X} \sim \varepsilon$
  - Autocorrelated multivariate models $(\mathcal{L}' - \Sigma RW)\left[\mathbf{T}+\mathbf{N}\right] \sim \varepsilon$
  - Geographically Weighted Regression [43]

    $$\mathcal{L}\left[\mathcal{G}\left(\mathbf{T},\mathbf{N}\right)\right] \sim \varepsilon$$

  - Granger causality tests : [214] use Granger causality to link transit with land-use changes.
- On data returns :
  - Autoregressive multivariate models

    $$\mathcal{L}\left[(\Delta\mathbf{T}(t_{j'}))_{j'\leqslant j}, (\Delta\mathbf{N}(t_{j'}))_{j'\leqslant j}\right] \sim \varepsilon$$

  - Autoregressive autocorrelated multivariate models : idem with spatial autocorrelation term.
  - Synthetic Instrumental Variables : static territory and/or network ?

Bivariate linear models

Autocorrelated univariate models

Autocorrelated multivariate models

Granger causality tests

[214] use Granger causality to link transit with land-use changes.

Autoregressive multivariate models

Autoregressive autocorrelated multivariate models

5.3 early warnings of network breakdowns : socio-economic and real estate trajectories

5.3.1 *Context*

Various aspects of territories are concerned by interactions with networks. In previous empirical studies, no socio-economic attributes of populations inhabiting the territory nor economic values for land and real estate was considered. Both are however crucial elements of territorial dynamics and are extensively studied in fields such as territorial analysis or urban economics : for example, [105] studies households residential choices to understand land-use transportation interactions.

[95]

5.3.2 *Preliminary Results*

5.3.3 *A strategy to investigate early warnings of network breakdowns*

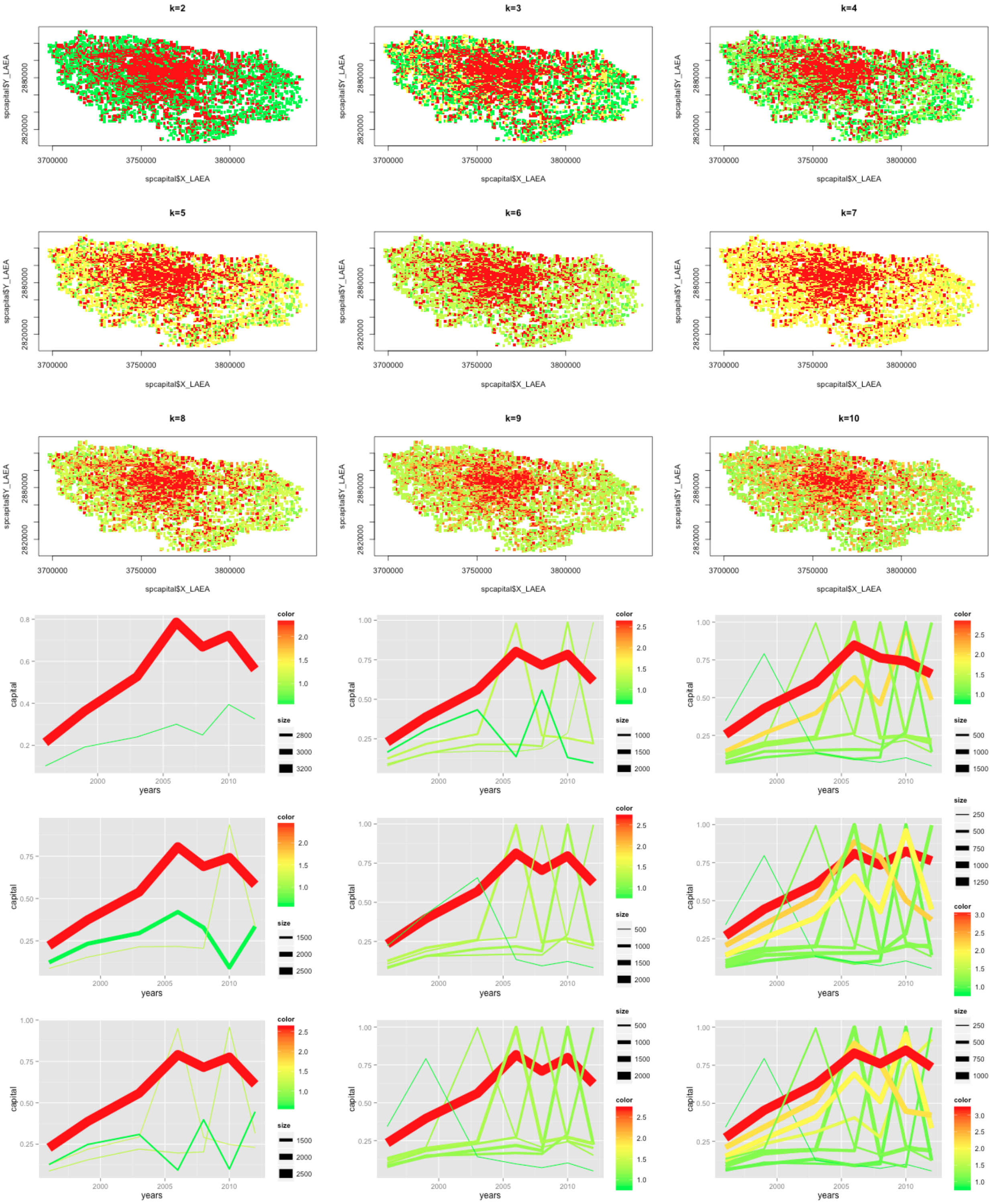


Figure 11: Typology of Real Estate trajectories. Locations were categorized

5.4 south-african historical events as instruments to understand network-territory relations

The method of instruments in statistics is used to identify causal relationships between variables. We already identified the issue of causality in

6 MODELING

6.1 A SIMPLE MODEL OF URBAN GROWTH

We propose a stochastic model of urban growth that generates spatial distributions of population densities, at an intermediate scale between economic models at the macro scale and land-use evolution models focusing on local relations. Integrating simply the two opposite key processes of aggregation ("preferential attachment") and diffusion (urban sprawl), we show that we can capture the whole spectrum of existing urban forms in Europe. An extensive exploration and calibration of the proposed model allows determining the region of parameter space corresponding morphologically to observed european urban systems, providing an validated thematic interpretation to model parameters, and furthermore determining the effective dimension of the urban system at this scale regarding morphological objectives.

6.1.1 *Context*

[7] propose a micro-based model of urban growth, with the purpose to replace non-interpretable physical mechanisms with agent mechanisms, including interactions forces and mobility choices.

6.1.2 *Model Description*

The urban growth model

RATIONALE Our model is an extension of the diffusion-limited aggregation model studied in [18]. Indeed, the tension between antagonist aggregation and sprawl mechanisms may be an important process in urban morphogenesis. [83] opposes centrifugal forces with centripetal forces in the equilibrium view of urban spatial systems, what is easily transferable to non-equilibrium systems in the framework of self-organized complexity : a urban structure is a far-from-equilibrium system that has been driven to this point by this opposite forces.

SETTINGS Precise formulation, description and formalization of the model. ; parameters and their possible interpretation ; def of parameter space.

Indicators

6.1.3 *Results*

Precise description of the implementation (pub openmole exploration etc, importance of intensive computation)

Figure 12: Example of the variety of generated urban shapes

Figure 13: Scatterplots of indicators distribution in an hypercube of the parameter space. We show here the influence of one parameter (diffusion rate). Red points correspond to real data.

Generation of urban patterns

variety of genered forms, examples of extreme shapes.

Model Behavior

convergence - internal model validation convergence properties of indicators ; number of repetitions needed for consistence of results. [histograms and stats of σ

exploration of parameter space Grid, then LHS explorations.

Figure : on 2 first PC of morpho indicators, localization of typical shapes (monocentric city, polycentric city, diffuse rural settlements, aggregated rural settlements) / comparing generated shape with a typical real one

statistical analysis Regression indics = f(params) . TO BE DONE. interpretation ?

Model Calibration

real data

calibration process Specific calib process : PCA that maximize the cumulated distance between generated points and reals points ; then select point cloud that overlaps real points in (PC1,PC2) plan, given a distance threshold.

Figure : Precise calibration of the model. The principal component analysis is conducted to maximize the spread of the differences between real data and model output, i.e. on the set $\{|R_i - M_j|\}$ where R_i is the set of real points, M_j the set of model outputs. We select then the overlapping cloud at threshold θ, by taking models output closer to real point cloud than θ in the (PC1,PC2) plan.

Calibration Results

-> extraction of the exact parameter space covering all real situations. interpretation of its shape (correlations between parameters ?) and its volume in different directions (relative importance of parameters). [possible development : application of Calibration Profile algo to check relative influence of parameters + ad hoc linear algebra on regression of 3.2.3 to do the same]

6.1.4 *Discussion*

Thematic interpretation of growth behavior

interpret positions of typical shapes within param space : confirms thematic interpretation of parameters. depending on results of 3.3.3, necessary and sufficient parameters to explain growth at this scale -> interpretation ?

Integration into a multi-scale growth model

Possible coupling on a gibrat (or favaro-pumain) at europa scale (macro) (with addition of consistence on migration constraints), where meso growth rates which werer exogeneous before, are top-down determined, and bottom-up feedback is done through local aggregation level, that influence attractivity of an area.

-> Accurately calibrated spatialized urban growth model, can reproduce any (european) urban pattern. -> interpretation of parameter influence ; effective independant dimensions of the urban system at this scale.

6.2 CORRELATED GENERATION OF TERRITORIAL CONFIGURATIONS

6.2.1 *Application : geographical data of density and network*

Context

The use of synthetic data in geography is generally directed towards the generation of synthetic populations within agent-based models (mobility, *LUTI* models) [160]. We can make a weak link with some Spatial Analysis techniques. The extrapolation of a continuous spatial field from a discrete spatial sample through a kernel density estimation for example can be understood as the creation of a synthetic dataset (even if it is not generally the initial view, as in Geographically Weighted Regression [43] in which variable size kernels do not interpolate data *stricto sensu* but extrapolate abstract variables representing interaction between explicit variables). In the field of modeling in quantitative geography, *toy-models* or hybrid models require a consistent initial spatial configuration. A set of possible initial configurations becomes a synthetic dataset on which the model is tested. The first Simpop model [185], precursor of a large family of models later parametrized with real data, could enter that frame but was studied on an unique synthetic spatialization. Similarly underlined was the difficulty to generate an initial transportation infrastructure in the case of the SimpopNet model [189] although it was admitted as a cornerstone of knowledge on the behavior of the model. A systematic control of spatial configuration effects on the behavior of simulation models was only recently proposed [61], approach that can be interpreted as a statistical control on spatial data. The aim is to be able to distinguish proper effects due to intrinsic model dynamics from particular effects due to the geographical structure of the case study. Such results are essential for the validation of conclusions obtained with modeling and simulation practices in quantitative geography.

Formalization

We propose in our case to generate territorial systems summarized in a simplified way as a spatial population density $d(\vec{x})$ and a transportation network $n(\vec{x})$. Correlations we aim to control are correlations between urban morphological measures and network measures. The question of interactions between territories and networks is already well-studied [150] but stays highly complex and difficult to quantify [149]. A dynamical modeling of implied processes should shed light on these interactions ([39], p. 162-163). We develop in that frame a *simple* coupling (i.e. without any feedback loop) between a density distribution model and a network morphogenesis model.

DENSITY MODEL We use a model D similar to aggregation-diffusion models [18] to generate a discrete spatial distribution of population density. A generalization of the basic model is proposed in [171], providing a calibration on morphological objectives (entropy, hierarchy, spatial auto-correlation, mean distance) against real values computed on the set of 50km sized grid extracted from european density grid [79]. More precisely, the model proceeds iteratively the following way. An square grid of width N, initially empty, is represented by population $(P_i(t))_{1 \leqslant i \leqslant N^2}$. At each time step, until total population reaches a fixed parameter P_m,

- total population is increased of a fixed number N_G (growth rate), following a preferential attachment such that

$$\mathbb{P}[P_i(t+1) = P_i(t)+1 | P(t+1) = P(t)+1] = \frac{(P_i(t)/P(t))^\alpha}{\sum (P_i(t)/P(t))^\alpha}$$

- a fraction β of population is diffused to four closest neighbors is operated n_d times

The two contradictory processes of urban concentration and urban sprawl are captured by the model, what allows to reproduce with a good precision a large number of existing morphologies.

NETWORK MODEL On the other hand, we are able to generate a planar transportation network by a model N, at a similar scale and given a density distribution. Because of the conditional nature to the density of the generation process, we will first have conditional estimators for network indicators, and secondly natural correlations between network and urban shapes should appear as processes are not independent. The nature and modularity of these correlations as a function of model parameters are still to determine by exploration of the coupled model.

The heuristic network generation procedure is the following :

1. A fixed number N_c of centers that will be first nodes of the network si distributed given density distribution, following a similar law to the aggregation process, i.e. the probability to be distributed in a given patch is $\frac{(P_i/P)^\alpha}{\sum (P_i/P)^\alpha}$. Population is then attributed according to Voronoi areas of centers, such that a center cumulates population of patches within its extent.

2. Centers are connected deterministically by percolation between closest clusters : as soon as network is not connected, two closest connected components in the sense of minimal distance between each vertices are connected by the link realizing this distance. It yields a tree-shaped network.

3. Network is modulated by potential breaking in order to be closer from real network shapes. More precisely, a generalized gravity potential between two centers i and j is defined by

$$V_{ij}(d) = \left[(1-k_h) + k_h \cdot \left(\frac{P_i P_j}{P^2}\right)^{\gamma}\right] \cdot \exp\left(-\frac{d}{r_g(1+d/d_0)}\right)$$

   where d can be euclidian distance $d_{ij} = d(i,j)$ or network distance $d_N(i,j)$, $k_h \in [0,1]$ a weight to modulate role of populations, γ giving shape of the hierarchy across population values, r_g characteristic interaction distance and d_0 distance shape parameter.

4. A fixed number $K \cdot N_L$ of potential new links is taken among couples having greatest euclidian distance potential ($K = 5$ is fixed).

5. Among potential links, N_L are effectively realized, that are the one with smallest rate $\tilde{V}_{ij} = V_{ij}(d_N)/V_{ij}(d_{ij})$. At this stage only the gap between euclidian and network distance is taken into account : $\tilde{V}_{ij}$ does indeed not depend on populations and is increasing with d_N at constant d_{ij}.

6. Planarity of the network is forced by creation of nodes at possible intersections created by new links.

We insist on the fact that the network generation procedure is entirely heuristic and result of thematic assumptions (connected initial network, gravity-based link creation) combined with trial-and-error during first explorations. Other model types could be used as well, such biological self-generated networks [198], local network growth based on geometrical constraints optimization [14], or a more complex percolation model than the initial one that would allow the creation of loops for example. We could thus in the frame of a modular architecture, in which the choice between different implementations of a functional brick can be seen as a meta-parameter [60], choose network generation function adapted to a specific need (as e.g. proximity to real data, constraints on output indicators, variety if generated forms, etc.).

PARAMETER SPACE Parameter space for the coupled model[1] is constituted by density generation parameters $\vec{\alpha}_D = (P_m/N_G, \alpha, \beta, n_d)$ (we study for the sake of simplicity the rate between population

1 Weak coupling allows to limit the total number of parameters as a strong coupling would involve retroaction loops and consequently associated parameters to determine their structure and intensity. In order to diminish it, an integrated model would be preferable to a strong coupling, what is slightly different in the sense where it is not possible in the integrated model to freeze one of the subsystems to obtain a model of the other subsystem that would correspond to the non-coupled model.

and growth rate instead of both varying, i.e. the number of steps needed to generate the distribution) and network generation parameters $\vec{\alpha}_N = (N_C, k_h, \gamma, r_g, d_0)$. We denote $\vec{\alpha} = (\vec{\alpha}_D, \vec{\alpha}_N)$.

INDICATORS Urban form and network structure are quantified by numerical indicators in order to modulate correlations between these. Morphology is defined as a vector $\vec{M} = (r, \bar{d}, \varepsilon, a)$ giving spatial auto-correlation (Moran index), mean distance, entropy and hierarchy (see [121] for a precise definition of these indicators). Network measures $\vec{G} = (\bar{c}, \bar{l}, \bar{s}, \delta)$ are with network denoted (V, E)

- Mean centrality $\bar{c}$ defined as average *betweeness-centrality* (normalized in $[0, 1]$) on all links.
- Mean path length $\bar{l}$ given by $\frac{1}{d_m}\frac{2}{|V|\cdot(|V|-1)}\sum_{i<j} d_N(i,j)$ with d_m normalization distance taken here as world diagonal $d_m = \sqrt{2}N$.
- Mean network speed [11] which corresponds to network performance compared to direct travel, defined as $\bar{s} = \frac{2}{|V|\cdot(|V|-1)}\sum_{i<j}\frac{d_{ij}}{d_N(i,j)}$.
- Network diameter $\delta = \max_{ij} d_N(i,j)$.

COVARIANCE AND CORRELATION We study the cross-correlation matrix $\text{Cov}\left[\vec{M}, \vec{G}\right]$ between morphology and network. We estimate it on a set of n realizations at fixed parameter values $(\vec{M}\left[D(\vec{\alpha})\right], \vec{G}\left[N(\vec{\alpha})\right])_{1\leqslant i\leqslant n}$ with standard unbiased estimator. We estimate correlation with associated Pearson estimator.

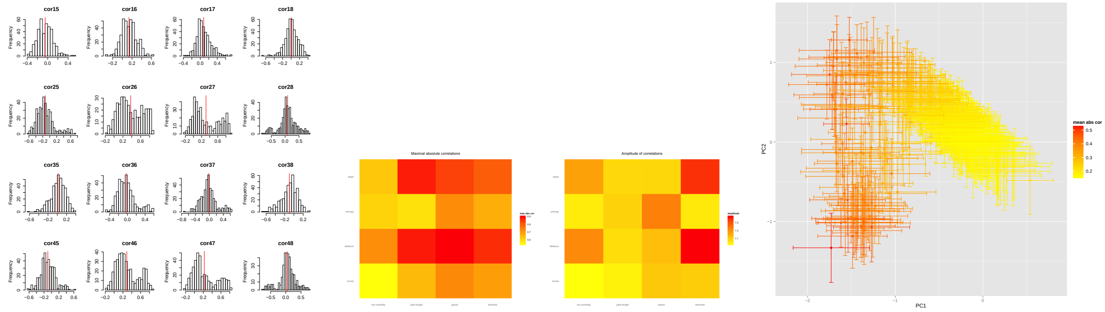
Implementation

Coupling of generative models is done both at formal and operational levels. We interface therefore independent implementations. The OpenMole software [176] for intensive model exploration offers for that the ideal frame thanks to its modular language allowing to construct *workflows* by task composition and interfacing with diverse experience plans and outputs. For operational reasons, density model is implemented in `scala` language as an OpenMole `plugin`, whereas network generation is implemented in agent-oriented language `NetLogo` [212] because of its possibilities for interactive exploration and heuristic model construction. Source code is available for reproducibility on project repository[2].

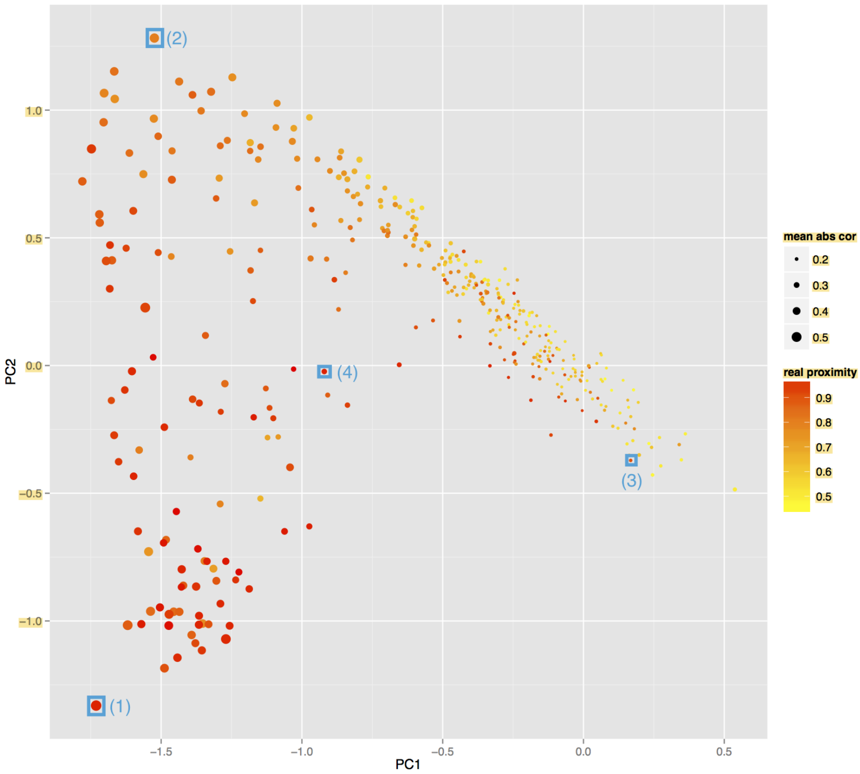
Results

The study of density model alone is developed in [171]. It is in particular calibrated on European density grid data, on 50km width

2 at `https://github.com/JusteRaimbault/CityNetwork/tree/master/Models/Synthetic`



(a) (b) (c)



(d) (e)

Figure 14: **Exploration of feasible space for correlations between urban morphology and network structure** | **(a)** Distribution of crossed-correlations between vectors $\vec{M}$ of morphological indicators (in numbering order Moran index, mean distance, entropy, hierarchy) and $\vec{N}$ of network measures (centrality, mean path length, speed, diameter). **(b)** Heatmaps for amplitude of correlations, defined as $a_{ij} = \max_k \rho_{ij}^{(k)} - \min_k \rho_{ij}^{(k)}$ and maximal absolute correlation, defined as $c_{ij} = \max_k \left|\rho_{ij}^{k}\right|$. **(c)** Projection of correlation matrices in a principal plan obtained by Principal Component Analysis on matrix population (cumulated variances: PC1=38%, PC2=68%). Error bars are initially computed as 95% confidence intervals on each matrix element (by standard Fisher asymptotic method), and upper bounds after transformation are taken in principal plan. Scale color gives mean absolute correlation on full matrices. **(d)** Representation in the principal plan, scale color giving proximity to real data defined as $1 - \min_r \|\vec{M} - \vec{M}_r\|$ where $\vec{M}_r$ is the set of real morphological measures, point size giving mean absolute correlation.

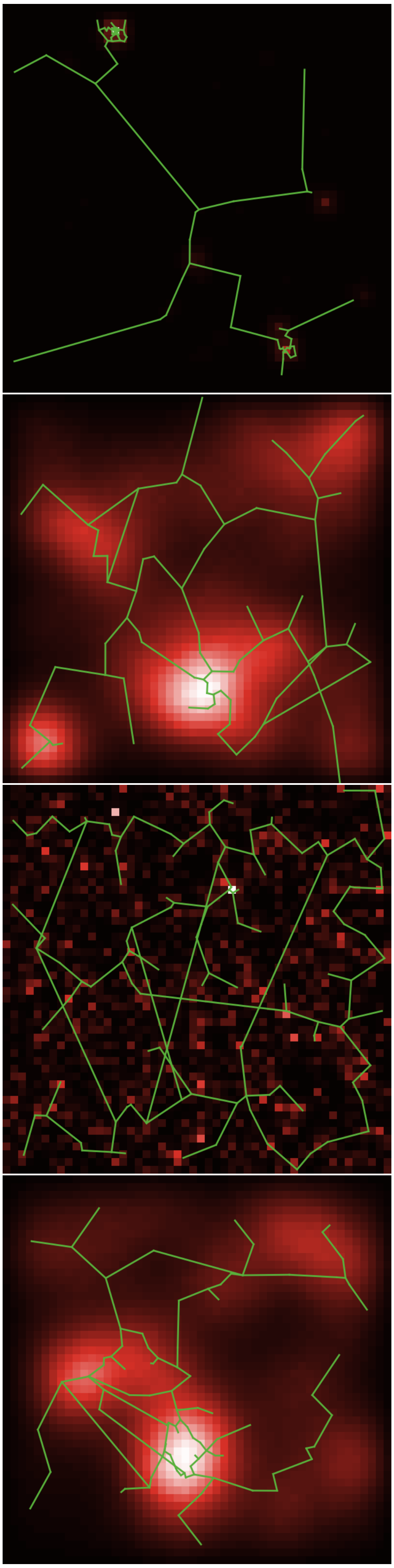


Figure 15: Configurations obtained for parameters giving the four emphasized points in (d), in order from left to right and top to bottom. We recognize polycentric city configurations (2 and 4), diffuse rural settlements (3) and aggregated weak density area (1). See appendice for exhaustive parameter values, indicators and corresponding correlations. For example $\bar{d}$ is highly correlated with $\bar{l}$, $\bar{s}$ ($\simeq$0.8) in (1) but not for (3) although both correspond to rural en-

square areas with 500m resolution for which real indicator values have been computed on whole Europe. Furthermore, a grid exploration of model behavior yields feasible output space in reasonable parameters bounds (roughly $\alpha \in [0.5, 2], N_G \in [500, 3000], P_m \in [10^4, 10^5], \beta \in [0, 0.2], n_d \in \{1, \ldots, 4\}$). The reduction of indicators space to a two dimensional plan through a Principal Component Analysis (variance explained with two components $\simeq 80\%$) allows to isolate a set of output points that covers reasonably precisely real point cloud. It confirms the ability of the model to reproduce morphologically the set of real configurations.

At given density, the conditional exploration of network generation model parameter space suggest a good flexibility on global indicators $\vec{G}$, together with good convergence properties. For a precise study of model behavior, see appendice giving regressions analysis capturing the behavior of coupled model. In order to illustrate synthetic data generation method, the exploration has been oriented towards the study of cross-correlations.

Given the large relative dimension of parameter space, an exhaustive grid exploration is not possible. We use a Latin Hypercube sampling procedure with bounds given above for $\vec{\alpha}_D$ and for $\vec{\alpha}_N$, we take $N_C \in [50, 120], r_g \in [1, 100], d_0 \in [0.1, 10], k_h \in [0, 1], \gamma \in [0.1, 4], N_L \in [4, 20]$. For number of model replications for each parameter point, less than 50 are enough to obtain confidence intervals at 95% on indicators of width less than standard deviations. For correlations a hundred give confidence intervals (obtained with Fisher method) of size around 0.4, we take thus $n = 80$ for experiments. Figure 14 gives details of experiment results. Regarding the subject of correlated synthetic data generation, we can sum up the main lines as following :

- Empirical distributions of correlation coefficients between morphology and network indicators are not simple and some are bimodal (for example $\rho_{46} = \rho[r, \bar{l}]$ between Moran index and mean path length).

- it is possible to modulate up to a relatively high level of correlation for all indicators, maximal absolute correlation varying between 0.6 and 0.9. Amplitude of correlations varies between 0.9 and 1.6, allowing a broad spectrum of values. Point cloud in principal plan has a large extent but is not uniform : it is not possible to modulate at will any coefficient as they stay themselves correlated because of underlying generation processes. A more refined study at higher orders (correlation of correlations) would be necessary to precisely understand degrees of freedom in correlation generation.

- Most correlated points are also the closest to real data, what confirms the intuition and stylized fact of a strong interdependence in reality.
- Concrete examples taken on particular points in the principal plan show that similar density profiles can yield very different correlation profiles.

Possible developments

This case study could be refined by extending correlation control method. A precise knowledge of N behavior (statistical distributions on an exhaustive grid of parameter space) conditional to D would allow to determine $N^{<-1>}|D$ and have more latitude in correlation generation. We could also apply specific exploration algorithms to reach exceptional configurations realizing an expected correlation level, or at least to obtain a better knowledge of the feasible space of correlations [52].

6.2.2 *Discussion*

Scientific positioning

Our overall approach enters a particular epistemological frame. On the one hand the multidisciplinary aspect, and on the other hand the importance of empirical component through computational exploration methods, make this approach typical of Complex Systems science, as it is recalled by the roadmap for Complex Systems having a similar structure [37]. It combines transversal research questions (horizontal integration of disciplines) with the development of heterogeneous multi-scalar approaches which encounter similar issues as the one we proposed to tackle (vertically integrated disciplines). The combination of empirical knowledge obtained from data mining, with knowledge obtained by modeling and simulation is generally central to the conception and exploration of multi-scalar heterogeneous models. Results presented here is an illustration of such an hybrid paradigm.

Direct applications

Starting from the second example which was limited to data generation, we propose examples of direct applications that should give an overview of the range of possibilities.

- Calibration of network generation component at given density, on real data for transportation network (typically road network given the shape of generated networks ; it should be straightforward to use OpenStreetMap open data[3] that have a reasonable quality for Europe, at least for France [88], with however adjustments on generation procedure in order to avoid edge effects due its restrictive frame, for example by generating on an extended surface to keep only a central area on which calibration would be done) should theoretically allow to unveil parameter sets reproducing accurately existing configurations both for urban morphology and network shape. It could be then possible to derive a "theoretical correlation" for these, as an empirical correlation is according to some theories of urban systems not computable as a unique realization of stochastic processes is observed. Because of non-ergodicity of urban systems [166], there are strong chances that involved processes are different across different geographical areas (or from an other point of view that they are in an other state of meta-parameters, i.e. in an other regime) and that their interpretation as different realizations of the same stochastic process makes no sense, the impossibility of covariation estimation following. By attributing a synthetic dataset similar to a given real configuration, we would be able to compute a sort of *intrinsic correlation* proper to this configuration. As territorial configurations emerge from spatio-temporal interdependences between components of territorial systems, this intrinsic correlation emerges the same way, and its knowledge gives information on these interdependences and thus on relations between territories and networks.

- As already mentioned, most of models of simulation need an initial state generated artificially as soon as model parametrization is not done completely on real data. An advanced model sensitivity analysis implies a control on parameters for synthetic dataset generation, seen as model meta-parameters [61]. In the case of a statistical analysis of model outputs it provides a way to operate a second order statistical control.

- We studied in the first example stochastic processes in the sense of random time-series, whereas time did not have a role in the second case. We can suggest a strong coupling between the two model components (or the construction of an integrated model) and to observe indicators and correlations at different time steps during the generation. In a dynamical spatial models we have because of feedbacks necessarily propagation effects and therefore the existence of lagged interdependences in space and time [157]. It would drive our field of study towards a better understanding of dynamical correlations.

3 `https://www.openstreetmap.org`

Generalization

We were limited to the control of first and second moments of generated data, but we could imagine a theoretical generalization allowing the control of moments at any order. However, as shown by the geographical example, the difficulty of generation in a concrete complex case questions the possibility of higher orders control when keeping a consistent structure model and a reasonable number of parameters. The study of non-linear dependence structures as proposed in [53] is in an other perspective an interesting possible development.

6.2.3 *Conclusion*

We proposed an abstract method to generate synthetic datasets in which correlation structure is controlled. Its rapid implementation in two very different fields shows its flexibility and the broad range of possible applications. More generally, it is crucial to favorise such practices of systematic validation of computational models by statistical analysis, in particular for agent-based models for which the question of validation stays an open issue.

6.3 NETWORK GROWTH MODELS : EXPLICATIVE POWER FOR VARIOUS APPROACHES

Considering Network Growth in itself, many heuristics are available to generate a network under some constraints. As already developed,

7 TOWARDS MORE COMPLEX MODELS

7.1 TAKING GOVERNANCE INTO ACCOUNT IN NETWORK PRODUCTION PROCESSES : THE LUTECIA MODEL

7.1.1 *Thematic Context*

We briefly describe a simple game-theory based framework which aims to be integrated as behavioral rules for governing agents in a hybrid model introduced in [119] and formalized then explored in [120]. This model couples land-use dynamics with transportation infrastructure evolution and aims to endogeneize transportation infrastructure development at different levels. The framework proposed extends it by allowing cooperation and fusion between governing entities.

As detailed in [120], a conceptual city system with local administrative boundaries and corresponding governing agents (mayors), and a global governor (state) is the foundation of the model. A land-use evolution (residences and employments localisations) and transportation (gravity flows) are the first step of an iteration. The transportation infrastructure (road network) is then evolved by constructing a new road. First level of decision (global or local) is chosen randomly according to a fixed probability, and in the case of a local decision, the richest mayor will build the new road. The road is then build optimizing the marginal accessibility for the area corresponding to the builder in charge (all world if global, commun if local).

One thematic aspect lacking in the model and that would be interesting to study is the emergence of larger administrative zones, i.e. the emergence of new levels of governance in polycentric metropolitan areas. The reality is of course not as simple, as bottom-up initiatives such as collaboration between neighbor cities are interlaced with top-down decisions such as e.g. the "Métropole du Grand Paris" which is a new administrative structure for Paris Area decided at the state level [87]. It would be however interesting to test conditions for emergence of governance patterns from the bottom-up in a conceptual way by extending the model and adding interactions and fusion between administrative entities.

The extension shall consist in relaxing the assumption of a single road segment built at each time step and attribute one segment to the N richest mayors. That leads to situation where neighbor towns may want to construct both a new road. As they are likely to communicate with each other, we assume that negotiations take place and that they consider eventually to build in common, in which case they merge after (rough simplifying but stylized assumption). Such negotiations may be interpreted as a game in the sense of Game Theory, which as already been widely applied for modeling in social and political sciences for questions dealing with cognitive interacting agents with individual interests [153]. Such a framework as already been used in transportation investment studies, as e.g. in [179] where choices of operators (public and privates) to integrate their system in a global consistent commuter system is explored through the notion of Nash equilibrium.

7.1.2 *Formalization*

The model architecture couples in a complex way a module for land-use evolution with a module for transportation network growth. Sub-modules, detailed in the following, include in particular a governance module that rules processes of network evolution.

Land-use evolution

Transportation Network growth

The workflow for transportation network development is the following :

- At each time step, N new road segments are built. Choice between local and global is still done through uniform drawing with probability ξ. In the case of local building, roads are attributed successively to mayors with probabilities ξ_i, what means that richer areas may get many roads. It stays consistent with the thematic assumption than each road correspond to the allocation of one public market which are done independently (with N becoming greater, this assumption should be relaxed as attribution of subventions to local areas is of course not proportional to wealth, but we assume that it stays true with small N values).
- Areas building a road without neighbors doing it follow the standard procedure to develop the road network.
- Neighbor areas building a road will enter negotiations. We assume in this first simple version of the model that only bilateral negotiations may occur. Therefore, in the case of clusters with more than two areas, pairing is done at random (uniform drawing) between neighbors until all areas are paired.
- Possible strategies for players (negotiating areas, $i = 1, 2$) are : staying alone (A) and collaborating (C). Strategies are chosen simultaneously (non-cooperative game) as detailed after. For

(C,A) and (A,C) couples, the collaborating agent loose its investment and cannot build a road whereas the other continues his buiseness alone. For (A,A) both act as alone, and for (C,C) a common development is done. We denote $Z_i^*(S_1,S_2)$ the optimal infrastructure for area i with $(S_1,S_2) \in \{(A,C),(C,A),(A,A)\}$ which are determined the standard way in each zone separately, and Z_C^* the optimal common infrastructure computed with a 2 segments infrastructure on the union of both areas, which corresponds to the case where both strategies are C. Marginal accessibilities for area i and infrastucture Z is defined as $\Delta X_i(Z) = X_i^Z - X_i$. We introduce the costs of construction which are necessary to build the payoff matrix. They are assumed spatially uniform and noted I for a road segment, whereas a 2 road segment will cost $2 \cdot I - \delta I$ ($\delta I > 0$ cost gain of common technical means, assumed to be equally shared). An interesting generalization would be to divise costs proportionaly to wealth in the case of a collaboration. The payoff matrix of the game is the following, with κ a normalization constant ("price of acessibility") :

1 | 2	C	A
C	$U_i = \kappa \cdot \Delta X_i(Z_C^*) - I + \frac{\delta I}{2}$	$\begin{cases} U_1 = -I + \frac{\delta I}{2} \\ U_2 = \kappa \cdot \Delta X_2(Z_2^*) - I + \frac{\delta I}{2} \end{cases}$
A	$\begin{cases} U_1 = \kappa \cdot \Delta X_1(Z_1^*) - I + \frac{\delta I}{2} \\ U_2 = -I + \frac{\delta I}{2} \end{cases}$	$U_i = \kappa \cdot \Delta X_i(Z_i^*) - I$

We have a typical coordination game for which it is clear that no strategy is dominant for any player. In a probabilistic mixed-strategy case, there always exists a Nash equilibrium that we can easily determine in our case. It is reasonable to make such an assumption since negociations take generally some time during which agents are able to find the way to optimize rationnaly their expected utility. If $\mathbb{P}[S_1 = C] = p_1$ and $\mathbb{P}[S_2 = C] = p_2$, we have

$$
\begin{aligned}
\mathbb{E}[U_1] &= p_1 p_2 U_1(C,C) + p_1 \cdot (1-p_2) U_1(C,A) + p_2 \cdot (1-p_1) U_1(A,C) + (1-p_1)(1-p_2) U_1(A,A) \\
&= p_1 \cdot \left[p_2 \cdot \left(\kappa \cdot \Delta X_1(Z_C^*) - \frac{\delta I}{2} \right) - \kappa \cdot \Delta X_1(Z_1^*) + I \right] + p_2 \cdot \frac{\delta I}{2} + \kappa \cdot \Delta X_1(Z_1^*) - I
\end{aligned}
$$

Optimizing the expected utility along p_1 (the variable on which agent 1 has control) imposes the condition on p_2

$$
\frac{\partial\, \mathbb{E}[U_1]}{\partial p_1} = 0 \iff p_2 = \frac{\Delta X_1(Z_1^*) - \frac{I}{\kappa}}{\Delta X_1(Z_C^*) - \frac{\delta I}{2 \cdot \kappa}}
$$

We obtain the same way

$$p_1 = \frac{\Delta X_2(Z_2^*) - \frac{I}{\kappa}}{\Delta X_2(Z_C^*) - \frac{\delta I}{2\cdot\kappa}}$$

Note that we can directly interpret these expressions, as a player chances to cooperate will decrease with the potential gain of the other player, what is intuitive for a competitive game. It also forces feasibility conditions on I and δI to keep a probability, that are $I \leqslant \kappa \cdot \min(\Delta X_1(Z_1^*), \Delta X_2(Z_2^*))$ (binary positive cost-benefit conditions) and $I - \delta I > \kappa \cdot \max_i(\Delta X_i(Z_i^*) - \Delta X_i(Z_C^*))$. As soon as accessibility difference stay relatively small, both shall be compatible when $\delta I \ll I$, giving corresponding boundaries for I.

- Agents make choice of strategy following uniform drawings with probability computed above. Corresponding infrastructures are built, and in the case of choices (C, C), towns merge in a single one with new corresponding variables (employment, actives, etc.).

REMARK FOR THE IMPLEMENTATION To adapt an existing implementation, one just has to add the negotiation stage if conditions are met, using probabilities given above. The accessibility-dimensioned parameters $\alpha = \frac{I}{\kappa}$ and $\delta\alpha = \frac{\delta I}{\kappa}$ should be more simple to deal with.

7.1.3 *Results*

Implementation

The model was implemented in `NetLogo` [212] because of its exploratory and interactive nature. A particular care was taken for the computation of accessibilities and shortest paths, as a dynamic reevaluation of network distance is necessary for each new potential infrastructure, what become rapidly a computational burden. We use thus a dynamical programming shortest path computation, inspired from [199], using distance matrices updates instead of shortest paths full computation at each step. See details in architectural precisions in Appendix 8.3

Validation

7.1.4 *Perspectives*

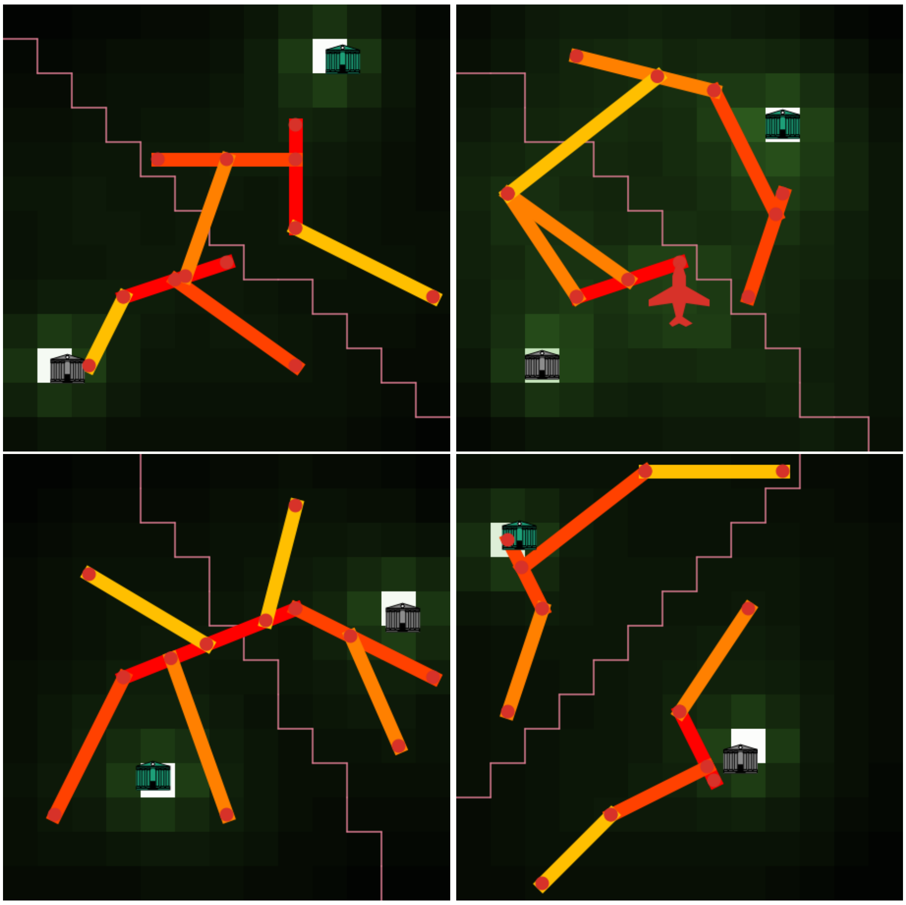


Figure 16: Examples of final configurations

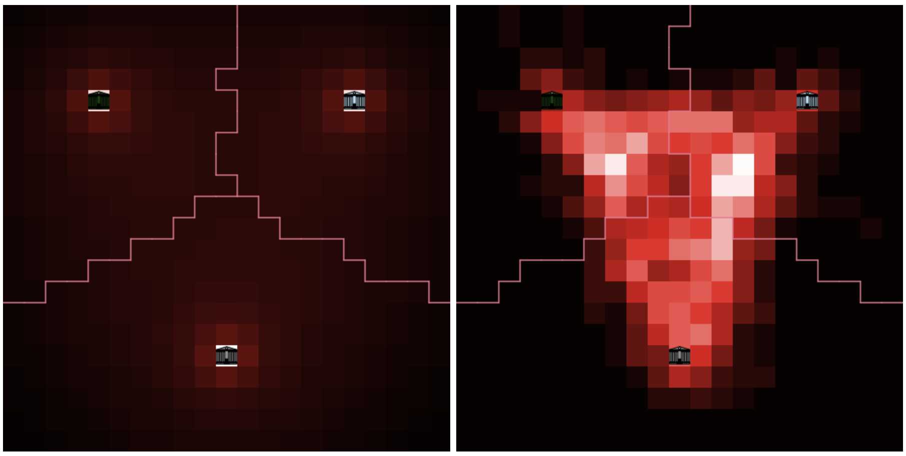


Figure 17

Part III

TOWARDS OPERATIONAL MODELS

Building Bricks

8 A ROADMAP FOR OPERATIONAL FAMILY OF MODELS OF COEVOLUTION

8.1 objectives

8.2 case studies

8.3 roadmap

CONCLUSION

BIBLIOGRAPHY

[1] Merwan Achibet, Stefan Balev, Antoine Dutot, and Damien Olivier. "A Model of Road Network and Buildings Extension Co-evolution." In: *Procedia Computer Science* 32 (2014), pp. 828–833.

[2] Andrew Adamatzky and Jeff Jones. "Road planning with slime mould: if Physarum built motorways it would route M6/M74 through Newcastle." In: *International Journal of Bifurcation and Chaos* 20.10 (2010), pp. 3065–3084.

[3] Philippe Aghion, Peter Howitt, Maxine Brant-Collett, and Cecilia García-Peñalosa. *Endogenous growth theory*. MIT press, 1998.

[4] Philippe Aghion, Ufuk Akcigit, Antonin Bergeaud, Richard Blundell, and David Hémous. *Innovation and Top Income Inequality*. 2015.

[5] Alex Anas, Richard Arnott, and Kenneth A. Small. "Urban Spatial Structure." English. In: *Journal of Economic Literature* 36.3 (1998), pp. 1426–1464. ISSN: 00220515. URL: http://www.jstor.org/stable/2564805.

[6] Philip W Anderson. "More is different." In: *Science* 177.4047 (1972), pp. 393–396.

[7] Claes Andersson, Kristian Lindgren, Steen Rasmussen, and Roger White. "Urban growth simulation from "first principles"." In: *Physical Review E* 66.2 (2002), p. 026204.

[8] E. Arcaute, E. Hatna, P. Ferguson, H. Youn, A. Johansson, and M. Batty. "Constructing cities, deconstructing scaling laws." In: *ArXiv e-prints* (Jan. 2013). arXiv: 1301.1674 [physics.soc-ph].

[9] W. Brian Arthur. *Complexity and the Shift in Modern Science*. Conference on Complex Systems, Tempe, Arizona. 2015.

[10] Arnaud Banos. "Pour des pratiques de modélisation et de simulation libérées en Géographies et SHS." In: *HDR. Université Paris 1* (2013).

[11] Arnaud Banos and Cyrille Genre-Grandpierre. "Towards new metrics for urban road networks: Some preliminary evidence from agent-based simulations." In: *Agent-based models of geographical systems*. Springer, 2012, pp. 627–641.

[12] Albert-Laszlo Barabasi. "Linked: How everything is connected to everything else and what it means." In: *Plume Editors* (2002).

[13] Ole E Barndorff-Nielsen, Peter Reinhard Hansen, Asger Lunde, and Neil Shephard. "Multivariate realised kernels: consistent positive semi-definite estimators of the covariation of equity prices with noise and non-synchronous trading." In: *Journal of Econometrics* 162 (2011), pp. 149–169.

[14] Marc Barthélemy and Alessandro Flammini. "Modeling urban street patterns." In: *Physical review letters* 100.13 (2008), p. 138702.

[15] Marc Barthélemy and Alessandro Flammini. "Co-evolution of density and topology in a simple model of city formation." In: *Networks and spatial economics* 9.3 (2009), pp. 401–425.

[16] Marc Barthelemy, Patricia Bordin, Henri Berestycki, and Maurizio Gribaudi. "Self-organization versus top-down planning in the evolution of a city." In: *Scientific reports* 3 (2013).

[17] F. Battiston, J. Iacovacci, V. Nicosia, G. Bianconi, and V. Latora. "Emergence of multiplex communities in collaboration networks." In: *ArXiv e-prints* (June 2015). arXiv: `1506.01280 [physics.soc-ph]`.

[18] Michael Batty. "Hierarchy in cities and city systems." In: *Hierarchy in natural and social sciences*. Springer, 2006, pp. 143–168.

[19] Michael Batty. *Cities and complexity: understanding cities with cellular automata, agent-based models, and fractals*. The MIT press, 2007.

[20] Michael Batty. *The new science of cities*. Mit Press, 2013.

[21] Michael Batty and Paul A Longley. *Fractal cities: a geometry of form and function*. Academic Press, 1994.

[22] Jean-Jacques Bavoux, Francis Beaucire, Laurent Chapelon, and Pierre Zembri. *Géographie des transports*. Paris, 2005.

[23] Sylvie Bazin, Christophe Beckerich, Corinne Blanquart, Marie Delaplace, and Ligdwine Vandenbossche. "Grande vitesse ferroviaire et développement économique local: une revue de la littérature." In: *Recherche Transports Sécurité* 27.3 (2011), pp. 215–238.

[24] Mark Bedau. "Downward causation and the autonomy of weak emergence." In: *Principia: an international journal of epistemology* 6.1 (2002), pp. 5–50.

[25] Adel Belouchrani, Moeness G Amin, and Karim Abed-Meraim. "Direction finding in correlated noise fields based on joint block-diagonalization of spatio-temporal correlation matrices." In: *Signal Processing Letters, IEEE* 4.9 (1997), pp. 266–268.

[26] Lucien Benguigui and Efrat Blumenfeld-Lieberthal. "A dynamic model for city size distribution beyond Zipf's law." In: *Physica A: Statistical Mechanics and its Applications* 384.2 (2007), pp. 613–627.

[27] Laurence Berne. "Ouverture et fermeture de territoire par les réseaux de transports dans trois espaces montagnards (Bugey, Bauges et Maurienne)." PhD thesis. Université de Savoie, 2008.

[28] Brian JL Berry. "Cities as systems within systems of cities." In: *Papers in Regional Science* 13.1 (1964), pp. 147–163.

[29] L. M. A. Bettencourt and J. Lobo. "Urban Scaling in Europe." In: *ArXiv e-prints* (Oct. 2015). arXiv: `1510.00902 [physics.soc-ph]`.

[30] Luís MA Bettencourt, José Lobo, and Geoffrey B West. "Why are large cities faster? Universal scaling and self-similarity in urban organization and dynamics." In: *The European Physical Journal B-Condensed Matter and Complex Systems* 63.3 (2008), pp. 285–293.

[31] Steven Bird. "NLTK: the natural language toolkit." In: *Proceedings of the COLING/ACL on Interactive presentation sessions*. Association for Computational Linguistics. 2006, pp. 69–72.

[32] John Bohannon. "Scientific publishing. Google Scholar wins raves–but can it be trusted?" In: *Science (New York, NY)* 343.6166 (2014), p. 14.

[33] Johan Bollen, David Crandall, Damion Junk, Ying Ding, and Katy Börner. "From funding agencies to scientific agency." In: *EMBO reports* 15.2 (2014), pp. 131–133.

[34] Olivier Bonin, Jean-Paul Hubert, et al. "Modèle de morphogénèse urbaine: simulation d'espaces qualitativement différenciés dans le cadre du modèle de l'économie urbaine." In: *49è colloque de l'ASRDLF*. 2012.

[35] Alain Bonnafous and François Plassard. "Les méthodologies usuelles de l'étude des effets structurants de l'offre de transport." In: *Revue économique* (1974), pp. 208–232.

[36] Alain Bonnafous, François Plassard, and Didier Soum. "La detection des effets structurants d'autoroute: Application à la Vallée du Rhône." English. In: *Revue économique* 25.2 (1974), pp. 233–256. ISSN: 00352764. URL: `http://www.jstor.org/stable/3500568`.

[37] P. Bourgine, D. Chavalarias, and al. "French Roadmap for complex Systems 2008-2009." In: *ArXiv e-prints* (July 2009). arXiv: `0907.2221 [nlin.AO]`.

[38] Paul Bourgine and John Stewart. "Autopoiesis and cognition." In: *Artificial life* 10.3 (2004), pp. 327–345.

[39] Anne Bretagnolle. "Villes et réseaux de transport : des interactions dans la longue durée, France, Europe, États-Unis." Français. HDR. Université Panthéon-Sorbonne - Paris I, June 2009. URL: http://tel.archives-ouvertes.fr/tel-00459720.

[40] Anne Bretagnolle, Fabien Paulus, and Denise Pumain. "Time and space scales for measuring urban growth." In: *Cybergeo: European Journal of Geography* (2002).

[41] Anne Bretagnolle and Denise Pumain. "Comparer deux types de systèmes de villes par la modélisation multi-agents." In: *Qu'appelle t-on aujourd'hui les sciences de la complexité? Langages, réseaux, marchés, territoires* (2010), pp. 271–299.

[42] Anne Bretagnolle, Denise Pumain, and Céline Vacchiani-Marcuzzo. "The organization of urban systems." In: *Complexity perspectives in innovation and social change*. Springer, 2009, pp. 197–220.

[43] Chris Brunsdon, Stewart Fotheringham, and Martin Charlton. "Geographically weighted regression." In: *Journal of the Royal Statistical Society: Series D (The Statistician)* 47.3 (1998), pp. 431–443.

[44] Colin F Camerer, Anna Dreber, Eskil Forsell, Teck-Hua Ho, Jürgen Huber, Magnus Johannesson, Michael Kirchler, Johan Almenberg, Adam Altmejd, Taizan Chan, et al. "Evaluating replicability of laboratory experiments in economics." In: *Science* (2016), aaf0918.

[45] Junavit Chalidabhongse and CC Jay Kuo. "Fast motion vector estimation using multiresolution-spatio-temporal correlations." In: *Circuits and Systems for Video Technology, IEEE Transactions on* 7.3 (1997), pp. 477–488.

[46] Pierre Champollion. "TERRITORY AND TERRITORIALIZATION: PRESENT STATE OF THE CAENTI THOUGHT." In: *International Conference of Territorial Intelligence*. INTI-International Network of Territorial Intelligence. Alba Iulia, Romania, Sept. 2006, p51–58. URL: https://halshs.archives-ouvertes.fr/halshs-00999026.

[47] Justin S Chang. "Models of the Relationship between Transport and Land-use: A Review." In: *Transport Reviews* 26.3 (2006), pp. 325–350.

[48] David Chavalarias and Jean-Philippe Cointet. "Phylomemetic patterns in science evolution—the rise and fall of scientific fields." In: *Plos One* 8.2 (2013), e54847.

[49] David Chavalarias, Sylvain Charron, Vincent Roger De Gardelle, and Paul Bourgine. "Nobel, Le Jeu De La Decouverte Scientifique." In: (2005).

[50] Yanguang Chen. "Urban gravity model based on cross-correlation function and Fourier analyses of spatio-temporal process." In: *Chaos, Solitons & Fractals* 41.2 (2009), pp. 603–614.

[51] Yanguang Chen. "Characterizing growth and form of fractal cities with allometric scaling exponents." In: *Discrete Dynamics in Nature and Society* 2010 (2010).

[52] Guillaume Chérel, Clémentine Cottineau, and Romain Reuillon. "Beyond Corroboration: Strengthening Model Validation by Looking for Unexpected Patterns." In: *PLoS ONE* 10.9 (Sept. 2015), e0138212. DOI: 10.1371/journal.pone.0138212. URL: http://dx.doi.org/10.1371%2Fjournal.pone.0138212.

[53] Rémy Chicheportiche and Jean-Philippe Bouchaud. "A nested factor model for non-linear dependences in stock returns." In: *arXiv preprint arXiv:1309.3102* (2013).

[54] Jinho Choi and Yong-Sik Hwang. "Patent keyword network analysis for improving technology development efficiency." In: *Technological Forecasting and Social Change* 83 (2014), pp. 170–182.

[55] Paul Claval. "Réseaux territoriaux et enracinement." In: *Cahier/-Groupe Réseaux* 3.7 (1987), pp. 44–60.

[56] David Colander. *The complexity revolution and the future of economics*. Tech. rep. Middlebury College, Department of Economics, 2003.

[57] Hadrien Commenges. "The invention of daily mobility. Performative aspects of the instruments of economics of transportation." Theses. Université Paris-Diderot - Paris VII, Dec. 2013. URL: https://tel.archives-ouvertes.fr/tel-00923682.

[58] Clementine Cottineau. "L'évolution des villes dans l'espace post-soviétique. Observation et modélisations." PhD thesis. Université Paris 1 Panthéon-Sorbonne, 2014.

[59] Clémentine Cottineau. *Urban scaling: What cities are we talking about?* Presentation of ongoing work at Quanturb seminar, April 1st 2015. 2015.

[60] Clémentine Cottineau, Paul Chapron, and Romain Reuillon. "An incremental method for building and evaluating agent-based models of systems of cities." In: (2015).

[61] Clémentine Cottineau, Florent Le Néchet, Marion Le Texier, and Romain Reuillon. "Revisiting some geography classics with spatial simulation." In: *Plurimondi. An International Forum for Research and Debate on Human Settlements*. Vol. 7. 15. 2015.

[62] Thomas Courtat, Catherine Gloaguen, and Stephane Douady. "Mathematics and morphogenesis of cities: A geometrical approach." In: *Physical Review E* 83.3 (2011), p. 036106.

[63] Noel Cressie and Hsin-Cheng Huang. "Classes of nonseparable, spatio-temporal stationary covariance functions." In: *Journal of the American Statistical Association* 94.448 (1999), pp. 1330–1339.

[64] Yves Crozet and François Dumont. "Retour sur les effets économiques du TGV. Les effets structurants sont un mythe (interview)." In: *Ville, Rail et Transports* 525 (2011), pp. 48–51. URL: https://halshs.archives-ouvertes.fr/halshs-01094554.

[65] Sylvain Cuyala. "Analyse spatio-temporelle d'un mouvement scientifique. L'exemple de la géographie théorique et quantitative européenne francophone." PhD thesis. Université Paris 1 Panthéon-Sorbonne, 2014.

[66] FD De Leon, M Felsen, and U Wilensky. "NetLogo Urban Suite-Tijuana Bordertowns model." In: *Center for Connected Learning and Computer-Based Modeling, Northwestern University, Evanston, IL* (2007). URL: http://ccl.northwestern.edu/netlogo/models/UrbanSuite-TijuanaBordertowns.

[67] Guillaume Deffuant, Arnaud Banos, David Chavalarias, Cyrille Bertelle, Nicolas Brodu, Pablo Jensen, Annick Lesne, Jean-Pierre Müller, Édith Perrier, and Franck Varenne. "Visions de la complexité. Le démon de Laplace dans tous ses états." In: *Natures Sciences Sociétés* 23.1 (2015), pp. 42–53.

[68] Julien Delile, René Doursat, and Nadine Peyriéras. "Chapitre 17. Modélisation multi-agent de l'embryogenèse animale." In: *Modélisations, simulations, systèmes complexes* (2016), pp. 581–624.

[69] Jean Delons, Nicolas Coulombel, and Fabien Leurent. "PIRANDELLO an integrated transport and land-use model for the Paris area." Aug. 2008. URL: https://halv3-preprod.archives-ouvertes.fr/hal-00319087.

[70] Denis Diderot. *Entretien entre d'Alembert et Diderot*. Garnier-Flammarion, 1965.

[71] Lynn Dirk. "A Measure of Originality The Elements of Science." In: *Social Studies of Science* 29.5 (1999), pp. 765–776.

[72] O Dollfus and F Durand Dastès. "Some remarks on the notions of 'structure'and 'system' in geography." In: *Geoforum* 6.2 (1975), pp. 83–94.

[73] René Doursat, Hiroki Sayama, and Olivier Michel. *Morphogenetic engineering: toward programmable complex systems*. Springer, 2012.

[74] Chris Drummond. "Replicability is not reproducibility: nor is it good science." In: (2009).

[75] César Ducruet and Laurent Beauguitte. "Spatial science and network science: Review and outcomes of a complex relationship." In: *Networks and Spatial Economics* 14.3-4 (2014), pp. 297–316.

[76] Gabriel Dupuy. "Vers une théorie territoriale des réseaux: une application au transport urbain." In: *Annales de Géographie*. JSTOR. 1987, pp. 658–679.

[77] Gabriel Dupuy and Lucien Gilles Benguigui. "Sciences urbaines: interdisciplinarités passive, naïve, transitive, offensive." In: *Métropoles* 16 (2015).

[78] Gilles Duranton. "Distance, land, and proximity: economic analysis and the evolution of cities." In: *Environment and Planning a* 31.12 (1999), pp. 2169–2188.

[79] EUROSTAT. *Eurostat Geographical Data*. 2014. URL: http://ec.europa.eu/eurostat/web/gisco/geodata/reference-data/administrative-units-statistical-units.

[80] J Doyne Farmer and Duncan Foley. "The economy needs agent-based modelling." In: *Nature* 460.7256 (2009), pp. 685–686.

[81] Jean-Marc Favaro and Denise Pumain. "Gibrat Revisited: An Urban Growth Model Incorporating Spatial Interaction and Innovation Cycles." In: *Geographical Analysis* 43.3 (2011), pp. 261–286.

[82] Roman Frigg and Ioannis Votsis. "Everything you always wanted to know about structural realism but were afraid to ask." In: *European journal for philosophy of science* 1.2 (2011), pp. 227–276.

[83] Masahisa Fujita and Jacques-François Thisse. "Economics of agglomeration." In: *Journal of the Japanese and international economies* 10.4 (1996), pp. 339–378.

[84] Xavier Gabaix. "Zipf's law for cities: an explanation." In: *Quarterly journal of Economics* (1999), pp. 739–767.

[85] Xavier Gabaix and Yannis M. Ioannides. "Chapter 53 The evolution of city size distributions." In: *Cities and Geography*. Ed. by J. Vernon Henderson and Jacques-François Thisse. Vol. 4. Handbook of Regional and Urban Economics. Elsevier, 2004, pp. 2341 –2378. DOI: http://dx.doi.org/10.1016/S1574-0080(04)80010-5. URL: http://www.sciencedirect.com/science/article/pii/S1574008004800105.

[86] Ronald N Giere. *Scientific perspectivism*. University of Chicago Press, 2010.

[87] Frédéric Gilli and Jean-Marc Offner. *Paris, métropole hors les murs: aménager et gouverner un Grand Paris*. Sciences Po, les presses, 2009.

[88] Jean-François Girres and Guillaume Touya. "Quality assessment of the French OpenStreetMap dataset." In: *Transactions in GIS* 14.4 (2010), pp. 435–459.

[89] Jean-François Gleyze. "La vulnérabilité structurelle des réseaux de transport dans un contexte de risques." PhD thesis. Université Paris-Diderot-Paris VII, 2005.

[90] Boris Golden, Marc Aiguier, and Daniel Krob. "Modeling of complex systems ii: A minimalist and unified semantics for heterogeneous integrated systems." In: *Applied Mathematics and Computation* 218.16 (2012), pp. 8039–8055.

[91] Daniel A Griffith. "Towards a theory of spatial statistics." In: *Geographical Analysis* 12.4 (1980), pp. 325–339.

[92] Daniel A Griffith. "What is spatial autocorrelation? Reflections on the past 25 years of spatial statistics." In: *Espace géographique* 21.3 (1992), pp. 265–280.

[93] Daniel A Griffith. *Advanced spatial statistics: special topics in the exploration of quantitative spatial data series*. Vol. 12. Springer Science & Business Media, 2012.

[94] Volker Grimm, Eloy Revilla, Uta Berger, Florian Jeltsch, Wolf M Mooij, Steven F Railsback, Hans-Hermann Thulke, Jacob Weiner, Thorsten Wiegand, and Donald L DeAngelis. "Pattern-oriented modeling of agent-based complex systems: lessons from ecology." In: *science* 310.5750 (2005), pp. 987–991.

[95] Marianne Guérois and Renaud Le Goix. "La dynamique spatio-temporelle des prix immobiliers à différentes échelles: le cas des appartements anciens à Paris (1990-2003)." In: *Cybergeo: European Journal of Geography* (2009).

[96] S. Gurciullo et al. "Complex Politics: A Quantitative Semantic and Topological Analysis of UK House of Commons Debates." In: *ArXiv e-prints* (Oct. 2015). arXiv: `1510.03797 [physics.soc-ph]`.

[97] R. Hachi. *Master Thesis Memoire*. Tech. rep. Université Paris VII, 2013.

[98] Hermann Haken. "Synergetics." In: *Naturwissenschaften* 67.3 (1980), pp. 121–128.

[99] Mordechai Haklay. "How good is volunteered geographical information? A comparative study of OpenStreetMap and Ordnance Survey datasets." In: *Environment and planning B: Planning and design* 37.4 (2010), pp. 682–703.

[100] Michael S Hansen, Christof Baltes, Jeffrey Tsao, Sebastian Kozerke, Klaas P Pruessmann, Peter Boesiger, and Erik M Pedersen. "Accelerated dynamic Fourier velocity encoding by exploiting velocity-spatio-temporal correlations." In: *Magnetic Resonance Materials in Physics, Biology and Medicine* 17.2 (2004), pp. 86–94.

[101] EGP Haran and Daniel R Vining. "A MODIFIED YULE-SIMON MODEL ALLOWING FOR INTERCITY MIGRATION AND ACCOUNTING FOR THE OBSERVED FORM OF THE SIZE DISTRIBUTION OF CITIES*." In: *Journal of Regional Science* 13.3 (1973), pp. 421–437.

[102] A Hernando, R Hernando, A Plastino, and E Zambrano. "Memory-endowed US cities and their demographic interactions." In: *Journal of The Royal Society Interface* 12.102 (2015), p. 20141185.

[103] C. A. Hidalgo. "Disconnected! The parallel streams of network literature in the natural and social sciences." In: *ArXiv e-prints* (Nov. 2015). arXiv: 1511.03981 [physics.soc-ph].

[104] John H Holland. *Signals and boundaries: Building blocks for complex adaptive systems*. Mit Press, 2012.

[105] Marius Homocianu. "Transport-land use interaction modeling - Residential choices of households in urban area of Lyon." Theses. Université Lumière - Lyon II, Jan. 2009. URL: https://tel.archives-ouvertes.fr/tel-00359302.

[106] *Hypergeo*. URL: http://www.hypergeo.eu/spip.php?page=sommaire.

[107] Michael Iacono, David Levinson, and Ahmed El-Geneidy. "Models of transportation and land use change: a guide to the territory." In: *Journal of Planning Literature* 22.4 (2008), pp. 323–340.

[108] Alexandros Karatzoglou, Alex Smola, Kurt Hornik, and Achim Zeileis. "kernlab – An S4 Package for Kernel Methods in R." In: *Journal of Statistical Software* 11.9 (2004), pp. 1–20. URL: http://www.jstatsoft.org/v11/i09/.

[109] Yan Ke, Rahul Sukthankar, and Martial Hebert. "Spatio-temporal shape and flow correlation for action recognition." In: *Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on*. IEEE. 2007, pp. 1–8.

[110] Marie-Laurence Keersmaecker, Pierre Frankhauser, and Isabelle Thomas. "Using fractal dimensions for characterizing intra-urban diversity: The example of Brussels." In: *Geographical analysis* 35.4 (2003), pp. 310–328.

[111] Frank B Knight. "A predictive view of continuous time processes." In: *The annals of Probability* (1975), pp. 573–596.

[112] Christof Koch and Gilles Laurent. "Complexity and the nervous system." In: *Science* 284.5411 (1999), pp. 96–98.

[113] A. Kolchinsky, A. J. Gates, and L. M. Rocha. "Modularity and the Spread of Perturbations in Complex Dynamical Systems." In: *ArXiv e-prints* (Sept. 2015). arXiv: `1509.04386 [physics.soc-ph]`.

[114] Marko Kryvobokov, Jean-Baptiste Chesneau, Alain Bonnafous, Jean Delons, and Vincent Piron. "Comparison of Static and Dynamic Land Use-Transport Interaction Models." In: *Transportation Research Record: Journal of the Transportation Research Board* 2344.1 (2013), pp. 49–58.

[115] *LA POSITION DE LA REVUE SOCIÉTÉS DANS L'ESPACE DISCURSIF DE LA SOCIOLOGIE FRANÇAISE*. URL: `http://zilsel.hypotheses.org/category/canular`.

[116] Alain L'Hostis, Claude Soulas, Gebhard Wulfhorst, et al. "La ville orientée vers le rail." In: *Ville et mobilité* (2012).

[117] C. Lagesse. "Read Cities through their Lines. Methodology to characterize spatial graphs." In: *ArXiv e-prints* (Dec. 2015). arXiv: `1512.01268 [physics.soc-ph]`.

[118] Robert B Laughlin. *A different universe: Reinventing physics from the bottom down*. Basic Books, 2006.

[119] Florent Le Nechet. "Approche multiscalaire des liens entre mobilité quotidienne, morphologie et soutenabilité des métropoles européennes: cas de Paris et de la région Rhin-Ruhr." PhD thesis. Université Paris-Est, 2010.

[120] Florent Le Nechet. "Aménagement urbain et jeux d'échelles : construction à Aménagement urbain et jeux d'échelles : construction à Aménagement urbain et jeux d'échelles : construction à plusieurs niveaux d'un réseau de transport métropolitain." 2012.

[121] Florent Le Néchet. "De la forme urbaine à la structure métropolitaine: une typologie de la configuration interne des densités pour les principales métropoles européennes de l'Audit Urbain." In: *Cybergeo: European Journal of Geography* (2015).

[122] Thomas Lechner, Ben Watson, Pin Ren, Uri Wilensky, Seth Tisue, and Martin Felsen. "Procedural modeling of land use in cities." In: (2004).

[123] *Les effets structurants des infrastructures de transport, L'Espace géographique 2014/1 (Tome 43), p. 51-67*.

[124] David Matthew Levinson, Feng Xie, and Shanjiang Zhu. "The co-evolution of land use and road networks." In: *Transportation and traffic theory* (2007), pp. 839–859.

[125] David Levinson. "Density and dispersion: the co-development of land use and rail in London." In: *Journal of Economic Geography* 8.1 (2008), pp. 55–77.

[126] David Levinson, Wei Chen, et al. "Paving new ground." In: ().

[127] Michael Lissack. "Subliminal influence or plagiarism by negligence ? The Slodderwetenschap of ignoring the internet." In: *Journal of Academic Ethics* (2013).

[128] Pierre Livet, Jean-Pierre Muller, Denis Phan, and Lena Sanders. "Ontology, a Mediator for Agent-Based Modeling in Social Science." In: *Journal of Artificial Societies and Social Simulation* 13.1 (2010), p. 3. ISSN: 1460-7425. URL: http://jasss.soc.surrey.ac.uk/13/1/3.html.

[129] R. Louf and M. Barthelemy. "How congestion shapes cities: from mobility patterns to scaling." In: *ArXiv e-prints* (Jan. 2014). arXiv: 1401.8200 [physics.soc-ph].

[130] Rémi Louf and Marc Barthelemy. "A typology of street patterns." In: *Journal of The Royal Society Interface* 11.101 (2014), p. 20140924.

[131] Rémi Louf and Marc Barthelemy. "Scaling: lost in the smog." In: *arXiv preprint arXiv:1410.4964* (2014).

[132] Rémi Louf, Pablo Jensen, and Marc Barthelemy. "Emergence of hierarchy in cost-driven growth of spatial networks." In: *Proceedings of the National Academy of Sciences* 110.22 (2013), pp. 8824–8829.

[133] Rémi Louf, Camille Roth, and Marc Barthelemy. "Scaling in Transportation Networks." In: *PLoS ONE* 9.7 (July 2014), e102007. DOI: 10.1371/journal.pone.0102007. URL: http://dx.doi.org/10.1371%2Fjournal.pone.0102007.

[134] Dominique Luzeaux. "A formal foundation of systems engineering." In: *Complex Systems Design & Management*. Springer, 2015, pp. 133–148.

[135] Chunsheng Ma. "Spatio-temporal covariance functions generated by mixtures." In: *Mathematical geology* 34.8 (2002), pp. 965–975.

[136] Hernán A Makse, José S Andrade, Michael Batty, Shlomo Havlin, H Eugene Stanley, et al. "Modeling urban growth patterns with correlated percolation." In: *Physical Review E* 58.6 (1998), p. 7054.

[137] David Mangin. *Le Grand Paris, où en est-on ?* Conférence de David Mangin le 14 mars 2014, ENPC et ENSAVT. 2014.

[138] Steven M Manson. "Simplifying complexity: a review of complexity theory." In: *Geoforum* 32.3 (2001), pp. 405–414.

[139] Steven M Manson. "Does scale exist? An epistemological scale continuum for complex human–environment systems." In: *Geoforum* 39.2 (2008), pp. 776–788.

[140] Caterina Marchionni. "Geographical economics versus economic geography: towards a clarification of the dispute." In: *Environment and Planning A* 36.10 (2004), pp. 1737–1753.

[141] Mendeley. *Mendeley Reference Manager*. http://www.mendeley.com/. 2015.

[142] Anne Vernez Moudon. "Urban morphology as an emerging interdisciplinary field." In: *Urban morphology* 1.1 (1997), pp. 3–10.

[143] Alexander B Murphy. "Entente territorial: Sack and Raffestin on territoriality." In: *Environment and Planning D: Society and Space* 30.1 (2012), pp. 159–172.

[144] M. E. J. Newman. "Prediction of highly cited papers." In: *ArXiv e-prints* (Oct. 2013). arXiv: 1310.8220 [physics.soc-ph].

[145] MEJ Newman. "Complex systems: A survey." In: *arXiv preprint arXiv:1112.1440* (2011).

[146] Volker Nitsch. "Zipf zipped." In: *Journal of Urban Economics* 57.1 (2005), pp. 86–100.

[147] Alireza Noruzi. "Google Scholar: The new generation of citation indexes." In: *Libri* 55.4 (2005), pp. 170–180.

[148] David O'Sullivan and Steven M Manson. "Do Physicists Have'Geography Envy'? And What Can Geographers Learn From It?" In: *Annals of the Association of American Geographers* (2015).

[149] Jean-Marc Offner. "Les "effets structurants" du transport: mythe politique, mystification scientifique." In: *Espace géographique* 22.3 (1993), pp. 233–242.

[150] Jean-Marc Offner and Denise Pumain. "Réseaux et territoires-significations croisées." In: (1996).

[151] E. Omodei, M. De Domenico, and A. Arenas. "Evaluating the impact of interdisciplinary research: a multilayer network approach." In: *ArXiv e-prints* (Jan. 2016). arXiv: 1601.06075 [physics.soc-ph].

[152] OpenStreetMap. *OpenStreetMap*. 2012.

[153] Peter C Ordeshook. *Game theory and political theory: An introduction*. Cambridge University Press, 1986.

[154] S. & al. Ostrowetsky. "Les Villes Nouvelles, 30 ans après." In: *Espaces et Sociétés n°119, 4/2004* (2004).

[155] V. Palchykov, V. Gemmetto, A. Boyarsky, and D. Garlaschelli. "Ground truth? Concept-based communities versus the external classification of physics manuscripts." In: *ArXiv e-prints* (Feb. 2016). arXiv: 1602.08451 [cs.DL].

[156] Neil J Paulley and F Vernon Webster. "Overview of an international study to compare models and evaluate land-use and transport policies." In: *Transport Reviews* 11.3 (1991), pp. 197–222.

[157] Bruce Wm Pigozzi. "Interurban linkages through polynomially constrained distributed lags." In: *Geographical Analysis* 12.4 (1980), pp. 340–352.

[158] Thomas Piketty. *Le capital au XXIe siècle*. Seuil, 2013.

[159] Ilya Prigogine and Isabelle Stengers. *The end of certainty*. Simon and Schuster, 1997.

[160] David R Pritchard and Eric J Miller. "Advances in agent population synthesis and application in an integrated land use and transportation model." In: *Transportation Research Board 88th Annual Meeting*. 09-1686. 2009.

[161] Denise Pumain. "Pour une théorie évolutive des villes." In: *Espace géographique* 26.2 (1997), pp. 119–134.

[162] Denise Pumain. "Une approche de la complexité en géographie." In: *Géocarrefour* 78.1 (2003), pp. 25–31.

[163] Denise Pumain. "Cumulativité des connaissances." In: *Revue européenne des sciences sociales. European Journal of Social Sciences* XLIII-131 (2005), pp. 5–12.

[164] Denise Pumain. "Une théorie géographique des villes." In: *Bulletin de la Société géographie de Liège* 55 (2010), pp. 5–15.

[165] Denise Pumain. "Multi-agent system modelling for urban systems: The series of SIMPOP models." In: *Agent-based models of geographical systems*. Springer, 2012, pp. 721–738.

[166] Denise Pumain. "Urban systems dynamics, urban growth and scaling laws: The question of ergodicity." In: *Complexity Theories of Cities Have Come of Age*. Springer, 2012, pp. 91–103.

[167] Denise Pumain, Fabien Paulus, Céline Vacchiani-Marcuzzo, and José Lobo. "An evolutionary theory for interpreting urban scaling laws." In: *Cybergeo: European Journal of Geography* (2006).

[168] Stephen H Putman. "Urban land use and transportation models: A state-of-the-art summary." In: *Transportation Research* 9.2 (1975), pp. 187–202.

[169] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria, 2015. URL: http://www.R-project.org/.

[170] Claude Raffestin. "Repères pour une théorie de la territorialité humaine." In: (1988).

[171] J. Raimbault. "Calibration of a Spatialized Urban Growth Model." In: *Working Paper, draft at https://github.com/JusteRaimbault/CityNetwork/tree/master/Docs/Papers/Den* (2016).

[172] J. Raimbault. "For a Cautious Use of Big Data and Computation." In: *RGS 2016, Session Geocomputation : the next 20 years ; abstract available at https://github.com/JusteRaimbault/CityNetwork/blob/master/Docs/Co* 2016.

[173] J. Raimbault. *Vers des Modèles Couplant Développement Urbain et Croissance des Réseaux de Transports, PhD Project Description.* Tech. rep. Géographie-Cités UMR CNRS 8504/LVMT UMR-T IFSTTAR 9403, October 2014.

[174] J. Raimbault, A. Banos, and R. Doursat. "A hybrid network/-grid model of urban morphogenesis and optimization." In: *Proceedings of the 4th International Conference on Complex Systems and Applications (ICCSA 2014), June 23-26, 2014, Université de Normandie, Le Havre, France; M. A. Aziz-Alaoui, C. Bertelle, X. Z. Liu, D. Olivier, eds.: pp. 51-60.* 2014.

[175] Karthik Ram. "Git can facilitate greater reproducibility and increased transparency in science." In: *Source code for biology and medicine* 8.1 (2013), p. 7.

[176] Romain Reuillon, Mathieu Leclaire, and Sebastien Rey-Coyrehourcq. "OpenMOLE, a workflow engine specifically tailored for the distributed exploration of simulation models." In: *Future Generation Computer Systems* 29.8 (2013), pp. 1981–1990.

[177] Sebastien Rey-Coyrehourcq. "Une plateforme intégrée pour la construction et Une plateforme intégrée pour la construction et l'évaluation de modèles de simulation en géographie." PhD thesis. Université Paris 1 Panthéon-Sorbonne, 2015.

[178] Henri Reymond and Colette Cauvin. "La logique ternaire de Stéphane Lupasco et le raisonnement géocartographique bio-culturel d'Homo geographicus. L'apport de la notion de couplage transdisciplinaire dans l'approche de l'agrégation morphologique des agglomérations urbaines." In: *Cybergeo: European Journal of Geography* (2013).

[179] Athena Roumboutsos and Seraphim Kapros. "A game theory approach to urban public transport integration policy." In: *Transport Policy* 15.4 (2008), pp. 209 –215. ISSN: 0967-070X. DOI: http://dx.doi.org/10.1016/j.tranpol.2008.05.001. URL: http://www.sciencedirect.com/science/article/pii/S0967070X08000280.

[180] Gerta Rucker. "Network meta-analysis, electrical networks and graph theory." In: *Research Synthesis Methods* 3.4 (2012), pp. 312–324.

[181] Yikang Rui and Yifang Ban. "Urban growth modeling with road network expansion and land use development." In: *Advances in Cartography and GIScience. Volume 2*. Springer, 2011, pp. 399–412.

[182] Yikang Rui, Yifang Ban, Jiechen Wang, and Jan Haas. "Exploring the patterns and evolution of self-organized urban street networks through modeling." In: *The European Physical Journal B* 86.3 (2013), pp. 1–8.

[183] Horacio Samaniego and Melanie E Moses. "Cities as organisms: Allometric scaling of urban road networks." In: *Journal of Transport and Land use* 1.1 (2008).

[184] Lena Sanders. *Système de villes et synergétique*. Economica, 1992.

[185] Lena Sanders, Denise Pumain, Hélene Mathian, France Guérin-Pace, and Stephane Bura. "SIMPOP: a multiagent system for the study of urbanism." In: *Environment and Planning B* 24 (1997), pp. 287–306.

[186] E. Sarigöl, R. Pfitzner, I. Scholtes, A. Garas, and F. Schweitzer. "Predicting Scientific Success Based on Coauthorship Networks." In: *ArXiv e-prints* (Feb. 2014). arXiv: 1402.7268 [physics.soc-ph].

[187] Thomas C Schelling. "Dynamic models of segregation." In: *Journal of mathematical sociology* 1.2 (1971), pp. 143–186.

[188] Helmut Schmid. "Probabilistic part-of-speech tagging using decision trees." In: *Proceedings of the international conference on new methods in language processing*. Vol. 12. Citeseer. 1994, pp. 44–49.

[189] Clara Schmitt. "Modélisation de la dynamique des systèmes de peuplement: de SimpopLocal à SimpopNet." PhD thesis. Paris 1, 2014.

[190] Clara Schmitt, Sébastien Rey-Coyrehourcq, Romain Reuillon, and Denise Pumain. "Half a billion simulations: Evolutionary algorithms and distributed computing for calibrating the SimpopLocal geographical model." In: (2014).

[191] Cosma Rohilla Shalizi and James P Crutchfield. "Computational mechanics: Pattern and prediction, structure and simplicity." In: *Journal of statistical physics* 104.3-4 (2001), pp. 817–879.

[192] Naoki Shibata, Yuya Kajikawa, Yoshiyuki Takeda, and Katsumori Matsushima. "Detecting emerging research fronts based on topological measures in citation networks of scientific publications." In: *Technovation* 28.11 (2008), pp. 758–775.

[193] Herbert A. Simon. "On a Class of Skew Distribution Functions." English. In: *Biometrika* 42.3/4 (1955), pp. 425–440. ISSN: 00063444. URL: http://www.jstor.org/stable/2333389.

[194] H Eugene Stanley, Luis A Nunes Amaral, David Canning, Parameswaran Gopikrishnan, Youngki Lee, and Yanhui Liu. "Econophysics: Can physicists contribute to the science of economics?" In: *Physica A: Statistical Mechanics and its Applications* 269.1 (1999), pp. 156–169.

[195] Forrest R. Stevens, Andrea E. Gaughan, Catherine Linard, and Andrew J. Tatem. "Disaggregating Census Data for Population Mapping Using Random Forests with Remotely-Sensed and Ancillary Data." In: *PLoS ONE* 10.2 (Feb. 2015), pp. 1–22. DOI: 10.1371/journal.pone.0107042. URL: http://dx.doi.org/10.1371%2Fjournal.pone.0107042.

[196] Victoria Stodden. "The scientific method in practice: Reproducibility in the computational sciences." In: (2010).

[197] Atsushi Tero, Ryo Kobayashi, and Toshiyuki Nakagaki. "Physarum solver: a biologically inspired method of road-network navigation." In: *Physica A: Statistical Mechanics and its Applications* 363.1 (2006), pp. 115–119.

[198] Atsushi Tero, Seiji Takagi, Tetsu Saigusa, Kentaro Ito, Dan P. Bebber, Mark D. Fricker, Kenji Yumiki, Ryo Kobayashi, and Toshiyuki Nakagaki. "Rules for Biologically Inspired Adaptive Network Design." In: *Science* 327.5964 (2010), pp. 439–442. DOI: 10.1126/science.1177894. eprint: http://www.sciencemag.org/content/327/5964/439.full.pdf. URL: http://www.sciencemag.org/content/327/5964/439.abstract.

[199] Konstantin Tretyakov, Abel Armas-Cervantes, Luciano García-Bañuelos, Jaak Vilo, and Marlon Dumas. "Fast fully dynamic landmark-based estimation of shortest path distances in very large graphs." In: *Proceedings of the 20th ACM international conference on Information and knowledge management*. ACM. 2011, pp. 1785–1794.

[200] Alan Mathison Turing. "The chemical basis of morphogenesis." In: *Philosophical Transactions of the Royal Society of London B: Biological Sciences* 237.641 (1952), pp. 37–72.

[201] Franck Varenne. "Framework for M&S with Agents in Regard to Agent Simulations in Social Sciences." In: *Activity-Based Modeling and Simulation* (2010), pp. 53–84.

[202] Franck Varenne. "Les simulations computationnelles dans les sciences sociales." In: *Nouvelles Perspectives en Sciences Sociales* 5.2 (2010), pp. 17–49.

[203] Franck Varenne, Marc Silberstein, et al. *Modéliser & simuler. Epistémologies et pratiques de la modélisation et de la simulation, tome 1*. 2013.

[204] Mehmet C Vuran, Özgür B Akan, and Ian F Akyildiz. "Spatio-temporal correlation: theory and applications for wireless sensor networks." In: *Computer Networks* 45.3 (2004), pp. 245–259.

[205] Benjamin Watson, M Pascal, Oleg Veryovka, Andy Fuller, Peter Wonka, and Chris Sexton. "Procedural urban modeling in practice." In: *IEEE Computer Graphics and Applications* 3 (2008), pp. 18–26.

[206] Michael Wegener and Franz Fürst. "Land-use transport interaction: state of the art." In: *Available at SSRN 1434678* (2004).

[207] Michael Wegener, Roger L Mackett, and David C Simmonds. "One city, three models: comparison of land-use/transport policy simulation models for Dortmund." In: *Transport Reviews* 11.2 (1991), pp. 107–129.

[208] Jörgen W Weibull. "An axiomatic approach to the measurement of accessibility." In: *Regional Science and Urban Economics* 6.4 (1976), pp. 357–379.

[209] JWR Whitehand, NJ Morton, and CMH Carr. "Urban morphogenesis at the microscale: how houses change." In: *Environment and Planning B: Planning and Design* 26.4 (1999), pp. 503–515.

[210] Jelte M. Wicherts. "Peer Review Quality and Transparency of the Peer-Review Process in Open Access and Subscription Journals." In: *PLoS ONE* 11.1 (Jan. 2016), e0147913. DOI: 10.1371/journal.pone.0147913. URL: http://dx.doi.org/10.1371%2Fjournal.pone.0147913.

[211] Norbert Wiener. *Cybernetics*. Hermann Paris, 1948.

[212] Uri Wilensky. "NetLogo." In: (1999).

[213] Stephen Wolfram. *A new kind of science*. Vol. 5. Wolfram media Champaign, 2002.

[214] Feng Xie and David Levinson. "How streetcars shaped suburbanization: a Granger causality analysis of land use and transit in the Twin Cities." In: *Journal of Economic Geography* (2009), lbp031.

[215] Feng Xie and David Levinson. "Modeling the growth of transportation networks: A comprehensive review." In: *Networks and Spatial Economics* 9.3 (2009), pp. 291–307.

[216] Yihui Xie. "knitr: A general-purpose package for dynamic report generation in R." In: *R package version* 1.7 (2013).

[217] Kazuko Yamasaki, Kaushik Matia, Sergey V Buldyrev, Dongfeng Fu, Fabio Pammolli, Massimo Riccaboni, and H Eugene Stanley. "Preferential attachment and growth dynamics in complex systems." In: *Physical Review E* 74.3 (2006), p. 035103.

[218] Daniel Yamins, Steen Rasmussen, and David Fogel. "Growing urban roads." In: *Networks and Spatial Economics* 3.1 (2003), pp. 69–85.

[219] Bhanu M Yerra and David M Levinson. "The emergence of hierarchy in transportation networks." In: *The Annals of Regional Science* 39.3 (2005), pp. 541–553.

[220] Pierre Zembri. "Les fondements de la remise en cause du Schéma Directeur des liaisons ferroviaires à grande vitesse: des faiblesses avant tout structurelles." In: *Annales de géographie*. JSTOR. 1997, pp. 183–194.

[221] Pierre Zembri. "La contribution de la grande vitesse ferroviaire à l'interrégionalité en France.(High-speed rail and interregionality in France)." In: *Bulletin de l'Association de géographes français* 85.4 (2008), pp. 443–460.

[222] Lei Zhang and David Levinson. "The economics of transportation network growth." In: *Essays on transport economics*. Springer, 2007, pp. 317–339.

[223] Liping Zhu, Masashi Aono, Song-Ju Kim, and Masahiko Hara. "Amoeba-based computing for traveling salesman problem: Long-term correlations between spatially separated individual cells of Physarum polycephalum." In: *Biosystems* 112.1 (2013), pp. 1–10.

Part IV

APPENDIX

9 ARCHITECTURE AND SOURCES FOR ALGORITHMS AND MODELS OF SIMULATION

You must not be afraid of putting
code in your thesis, code is not dirty
- Alexis Drogoul

And yet it is. It makes no sense to put code listings in the core of the text if there is no particular algorithmic detail that requires attention. As soon as implementation biases are avoided, architecture and source for a computational model should be independent from its formal description.

9.1 Algorithmic Systematic Review

Objective

Location

Characteristics

- Language : `Java`
- Size : 7116

Particularities

Architecture

Additional Scripts

9.2 INDIRECT BIBLIOMETRICS

OBJECTIVE

CHARACTERISTICS

- Language : `Python` and `R`
- Size :

PARTICULARITIES

ARCHITECTURE

ADDITIONAL SCRIPTS

9.3 Lutecia model

Objective

Characteristics

- Language : `NetLogo`
- Size : 4791

Particularities

Architecture

Additional scripts

10

TOOLS AND WORKFLOW FOR AN OPEN REPRODUCIBLE RESEARCH

Open for Discovery
- PLoS

10.1 NETLOGO DOCUMENTATION GENERATOR

10.2 GIT AS A REPRODUCIBILITY TOOL

10.3 GIT-DATA

10.4 TOWARDS A GIT-COMPATIBLE FIGURES METADATA HANDLER

11 QUANTITATIVE ANALYSIS OF THESIS REFLEXIVITY