



# Walk on City Maps - Mathematical and Physical phenomenology of the City, a Geometrical approach

Thomas Courtat

► **To cite this version:**

Thomas Courtat. Walk on City Maps - Mathematical and Physical phenomenology of the City, a Geometrical approach. Modeling and Simulation. Université Paris-Diderot - Paris VII, 2012. English. <tel-00714310>

**HAL Id: tel-00714310**

**<https://tel.archives-ouvertes.fr/tel-00714310>**

Submitted on 3 Jul 2012

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

THÈSE

Présentée pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ PARIS-DIDEROT

Discipline : Physique

Spécialité : Interdisciplinarité

École Doctorale Frontières Du Vivant (ED 474)

Laboratoire Matières et Systèmes Complexes et Orange Labs



# Promenade dans les cartes de villes - Phénoménologie mathématique et physique de la ville - une approche géométrique

présentée par  
Thomas COURTAT

Thèse effectuée sous la direction de  
Stéphane DOUADY et Catherine GLOAGUEN

Soutenue le 31 Janvier 2012 devant le Jury composé de :

Marc BARTHÉLÉMY	Chercheur, CEA	Examineur
Gilles COPPIN	Professeur, Telecom Bretagne	Rapporteur
Vincent DANOS	Professeur, University of Edinburgh	Rapporteur
Laurent DECREUSEFOND	Professeur, Telecom Paris Tech	Examineur
Stéphane DOUADY	Directeur de recherche, CNRS	Directeur
Jean-Pierre FREY	Professeur, Université Paris Est Créteil	Examineur
Pablo JENSEN	Professeur, ENS de Lyon	Examineur
Catherine GLOAGUEN	Ingénieur de recherche, Orange Labs	Directrice



# Contents

<b>I</b>	<b>Introduction</b>	<b>15</b>
<b>1</b>	<b>Cities and Science</b>	<b>19</b>
1.1	Social Physics and Urban Thermodynamics . . . . .	21
1.2	Morphology analysis . . . . .	23
1.3	From Graph to Complex Network Theory . . . . .	27
1.4	A representation of the city as a collection of objects arranged in space . . .	33
1.5	Growth models . . . . .	37
1.6	Synthesis . . . . .	40
<b>2</b>	<b>Mixing town-planning with Physics</b>	<b>41</b>
2.1	Science and Modelling . . . . .	43
2.2	Shape emergence . . . . .	44
2.3	What is a city ? . . . . .	45
<b>3</b>	<b>Approach of this thesis</b>	<b>53</b>
3.1	Quantitative principles shaping the city . . . . .	53
3.2	Problems . . . . .	54
3.3	Summary . . . . .	54
3.4	Style of the thesis . . . . .	57
3.5	Main notations . . . . .	59
<b>II</b>	<b>Mathematical and Physical Modeling of the City</b>	<b>61</b>
<b>4</b>	<b>Mathematical and Computational Representation of Cities</b>	<b>63</b>
4.1	Data . . . . .	65
4.2	City Graphs . . . . .	67
4.3	Algorithms to Recover Street Hypergraph structures . . . . .	72
4.4	Tuning and Performances . . . . .	73
4.5	Computational implementation . . . . .	77
4.6	Conclusion . . . . .	82
<b>5</b>	<b>Phenomena</b>	<b>83</b>

5.1	Return on phenomena described in the literature . . . . .	85
5.2	Street scaling . . . . .	86
5.3	Small World effect . . . . .	89
5.4	Distance transformation . . . . .	93
<b>6</b>	<b>Map Description and analysis</b>	<b>97</b>
6.1	Topological and Geometrical descriptors . . . . .	99
6.2	Map analysis and Centralities . . . . .	102
6.3	Discussion . . . . .	106
<b>7</b>	<b>Analyzing maps and planning with the Simplest Centrality</b>	<b>111</b>
7.1	Introduction . . . . .	113
7.2	Avignon: testing the robustness of simplest centrality toward windowing . .	113
7.3	Villers-Sur-Mer: testing town-planning scenarios . . . . .	117
7.4	Conclusion: comparison with Space Syntax and "dual" (street) graph . . . .	117
<b>8</b>	<b>Models</b>	<b>121</b>
8.1	A continuous models that came last but we present first . . . . .	123
8.2	Simple network models . . . . .	127
8.3	Morphogenesis of the City . . . . .	130
8.4	Object Morphogenesis of the City . . . . .	138
8.5	Conclusion . . . . .	141
<b>III</b>	<b>Stochastic Geometry</b>	<b>145</b>
<b>9</b>	<b>Sketching a city with Stochastic Geometry</b>	<b>147</b>
9.1	Principles . . . . .	149
9.2	Low scale models . . . . .	155
9.3	High scale model: Gabriel Graph . . . . .	158
9.4	Identification . . . . .	161
9.5	Algorithm . . . . .	162
9.6	Conclusion . . . . .	163
<b>10</b>	<b>Map Segmentation</b>	<b>165</b>
10.1	Spectral Clustering . . . . .	167
10.2	City Segmentation Algorithm . . . . .	173
10.3	Results . . . . .	175
<b>IV</b>	<b>Conclusion</b>	<b>183</b>
<b>11</b>	<b>Conclusion</b>	<b>185</b>
<b>V</b>	<b>Appendices</b>	<b>189</b>
<b>A</b>	<b>Geometry</b>	<b>191</b>

---

A.1	Some basis on Euclidian planar geometry . . . . .	191
A.2	Polygons . . . . .	192
A.3	Tessellations . . . . .	192
A.4	Fractal . . . . .	194
<b>B</b>	<b>Complex Networks'theory</b>	<b>195</b>
B.1	What is a complex network ? . . . . .	195
B.2	Analysis of complex networks . . . . .	195
B.3	Canonical complex networks . . . . .	197
B.4	Spatial embedding . . . . .	199
<b>C</b>	<b>Measure and probability theory for stochastic geometry</b>	<b>201</b>
C.1	Measure and integration theory . . . . .	201
C.2	Real random variables . . . . .	204
C.3	Stochastic geometry . . . . .	209



# Abstract

We are interested in the phenomenology of cities by restricting them to the geometry of their street network. This study aims at being synthetic, functional and interdisciplinary. It follows the large work that has been performed from the early XXth century by town-planners, social scientists, geographers, statisticians, physicists.

We try to demonstrate that the street - as a coherent alignment of street segments - can be considered as an essential elementary structure of the city. How much information is encoded in the street network? To what extent does it constraint the city use? How are the current urban layout and its evolution determined jointly by traffic axis and structuring elements?

We present a mathematical and computational framework allowing to consider the map of a city as a geometric continuum associated to the topology of a planar graph. To this graph we juxtapose a hypergraph structure using the street geometry to obtain easily the notion of axis and a multi-scale representation of the city.

In spite of an apparent shape diversity, we show that the street network of a city is subjected to general laws that leave hallmarks on a city map. We propose several morphogenesis models of the city, implementing the idea that the city's growth follows a structured extension / division of space logic and able to reproduce hallmarks observed on actual maps.

The understanding of regulation mechanism of the city allows us to propose functional algorithms whose computational efficiencies are very interesting. We present an algorithm recovering streets from a collection of street segments; the notion of simplest centrality whose calculus on a map allows a hierarchical analysis of it, revealing for instance main traffic axis and ill-deserved area; a fast approximate algorithm to find the shortest path between two random points; and a Spectral Clustering based algorithm to produce morphological segmentations of maps. We also work on the identification of random tessellation models to be substituted to an actual road network and to solve large optimization problems using statistical equivalence.





# Résumé

Nous nous intéressons à la phénoménologie des villes en nous limitant à la géométrie induite par le squelette de leur réseau de rues. C'est une étude à volonté synthétique, fonctionnelle et interdisciplinaire qui vient s'ajouter aux travaux qui ont été menés à grande cadence depuis le début du XXème siècle par des urbanistes, sociologues, géographes, statisticiens, physiciens.

Nous cherchons à montrer que la rue, en tant qu'alignement cohérent de segments de rues peut être considérée comme structure élémentaire de la ville. Quelle quantité d'information est donnée par la géométrie du réseau routier ? Dans quelle mesure contraint-il nos échanges ? Comment le paysage urbain actuel est-il déterminé par son évolution le long d'axes de circulation et d'éléments structurants ?

Nous présentons un cadre mathématique permettant de considérer la carte d'une ville comme un continuum géométrique défini par la topologie d'un graphe planaire. Nous superposons à ce graphe une structure d'hypergraphe pour manipuler aisément la notion d'axes ainsi qu'une représentation multi-échelles de la ville.

En dépit d'une grande diversité apparente de formes, nous montrons que le réseau de rues d'une ville se soumet à un certain nombre de lois générales qui laissent des traces sur le plan de la ville. Nous proposons des modèles de croissance et de morphogénèse de la ville, implémentant l'idée que l'évolution de la ville suit une logique d'extension / division structurée de l'espace et reproduisant les signatures observées sur les plans de villes réelles.

La compréhension des mécanismes régulateurs de la ville nous permet de proposer des algorithmes fonctionnels dont le temps de calcul est très intéressant. Ainsi nous présentons un algorithme reconstituant les rues à partir de segments de rues ; la notion de centralité simple dont le calcul sur une carte permet une analyse hiérarchique de celle-ci, met en valeur les axes de trafic principaux et en évidence les zones mal desservies ; un algorithme permettant d'approximer rapidement le plus court chemin entre deux points aléatoires ; un algorithme prenant appui sur le Spectral Clustering pour produire des segmentations morphologiques de cartes et retravaillons l'identification de modèles de mosaïques aléatoires pour les substituer à un réseau urbain particulier dans la résolution par équivalents statistiques de grands problèmes d'optimisation.



## Part I

# Introduction



This introductory part is constituted of three chapters that aim at leading naturally the reader to the stakes of this manuscript. The first chapter is a state of the art reviewing considerations on city modelling from the XVIII-th century. It depicts, synthesizes and critics the main theories ordered according to thematics and chronology. The second chapter is more personal and presents definitions of modelling, shape and city. The observation of real cities and their comparison to other transportation networks lead to isolate main principles of city systems in the third chapter that presents the problem of the thesis and a quick outlook of the contents of the manuscript.



# Chapter 1

## Cities and Science

### Contents

---

1.1	Social Physics and Urban Thermodynamics . . . . .	<b>21</b>
1.1.1	Around Zipf Law . . . . .	21
1.1.2	Quantitative Geography, Spatiality . . . . .	21
1.1.3	Dynamics, Multi Agent modeling . . . . .	23
1.2	Morphology analysis . . . . .	<b>23</b>
1.2.1	Typologies . . . . .	23
1.2.2	The Fractal City . . . . .	24
1.2.3	Density profile . . . . .	24
1.2.4	Stochastic Geometry . . . . .	26
1.3	From Graph to Complex Network Theory . . . . .	<b>27</b>
1.3.1	Koenigsberg bridges, Operational research . . . . .	27
1.3.2	Geographic Information System (GIS) . . . . .	28
1.3.3	Topological Statistics . . . . .	28
1.3.4	Centralities . . . . .	30
1.3.5	Random walks . . . . .	32
1.3.6	Other Complex Networks . . . . .	33
1.4	A representation of the city as a collection of objects arranged in space . . . . .	<b>33</b>
1.4.1	Human perception of Space . . . . .	33
1.4.2	Space Syntax . . . . .	33
1.4.3	Axial representations of the City . . . . .	36
1.4.4	The "dual" analysis . . . . .	36
1.5	Growth models . . . . .	<b>37</b>
1.5.1	Cellular automata . . . . .	37
1.5.2	DLA . . . . .	37
1.5.3	L-Systems . . . . .	37
1.5.4	Complex Network Growth . . . . .	37
1.6	Synthesis . . . . .	<b>40</b>

---



**Cities And Sciences : Synthesis**

Cities have become a subject of scientific investigation since the early XXth century for economists, geographers, town-planners, social scientists, architects and recently physicists. The first motivation was economical: to understand the variation of costs in space and to minimize them in practical tasks. The resulting literature is large, making it difficult to synthesize in a few pages. Two levels of observation complement each other: the macro and the micro. The macro gets interested in the statistical distributions of measures on cities (size, population, production...) and their correlations. The micro models the behaviour of agents or local infrastructures and their interactions to predict displacement flux or emergence of phenomena. These approaches can also be divided in spatial and non spatial approaches. To end with, the spatial approach considers the global shape of a city or the one of its road network.

**Contributions of this chapter**

1. A synthetic state of the art on urban modelling.

## 1.1 Social Physics and Urban Thermodynamics

### 1.1.1 Around Zipf Law

The first famous work on quantitative social science was [4] extended by Zipf [138] who exhibited that the distribution of cities' populations follows a power law distribution:

$$\mathbb{P}(\text{Population} > N) \propto 1/N^\alpha \quad \text{with } \alpha \gtrsim 1 \quad (1.1)$$

The validity of this law has been empirically proved in [120] for 73 countries with some variations in  $\alpha$ , one of the difficulty being the lack of an universal definition of the notion of city. This phenomenon is explained in [3] by a model of preferential attachment or multiplicative stochastic process (see App B). Wealth is created in a system of cities and the probability that a city captures wealth is proportional to its size (total amount of wealth). In [19], the authors show that all quantitative indicators of a city's size (population, creation, innovation, electrical consumption...) are linked with power laws: if  $Y(t)$  is a urban indicator and  $N(t)$  the population at time  $t$ ,

$$Y(t) = Y_0 \cdot N(t)^\beta \quad (1.2)$$

For the majority of indicators,  $\beta > 1$  which means parameters have to grow faster than the population. This is antithetic of stable development where all internal metabolism has to scale sub-linearly to be viable.

### 1.1.2 Quantitative Geography, Spatiality

Zipf's law and its developments provide macroscopic insights on the city's resources. A problem of practical importance in economics is to know how these resources are organized spatially, to determine costs related to mercantile exchanges and production.

The trends we are going to depict fit together under the generic term "regional science". Their paternity can be attributed to Adam Smith: most of them are a spatial translation of his *Homo Economicus* principle, considering agents are acting rationally to maximize their profit.

**Von Thünen model, marginal productivity theory** The Von Thünen model [131] of agricultural land was one of the first model which encapsulates the most important ingredients for a monocentric, isolated city where farmers transport their production directly to the city center and behave rationally to maximize their profits. The rent cost decreases with the distance to the center but the cost associated with good transportation increases with this distance. Following this argument, Von Thünen concluded that for each product there is a certain distance from the city where its production would be worthwhile. The center of the city is the marketplace and for a product there is a particular distance of profitability: beyond the production is not profitable and before there exists another product more profitable to deal with. The consequence is the organization of the monocentric city into regular crowns devoted to particular types of production.

**Location, central place theory** Location theory is the science that aims at explaining in a general way the spatial distribution of resources in an economical system. It directly steps on the work of Von Thünen. It has particularly been developed in the 30's by German geographers and economists. The most well known model is Christaller's Central Place Theory [39].

This work is presented as a rigorous deductive theory. It starts by making assumptions: space is homogeneous, isotropic, producers and consumers are rational agents who try to maximize their profit, reduce their cost and transportation. Activities gathers into "places" represented by points with their influence area in the plane. A more central place has a bigger influence area. The resulting "optimal" distribution of central places is a two layer hierarchy of cities regularly distributed on a perfectly hexagonal lattice. One can reproach to Christaller ( apart from is adhesion to Hitler's Reich), some fantasy on his hypothesis, choice of parameters (that are more symbolic than rational) and epistemological approach [100]. The problem with Christaller is that he modelled some ideal in a contestable fashion rather than the reality and tried with the Nazi regime to impose its model to the reality. We can also refer to August Lösch [80] that has a more rigorous and soft reasoning and who reaches similar configurations of central places.

**Regional science** Encompassing location theory, Regional Science has started by 40's and thrived until 80's. For instance Isard's [71] is somewhat the cornerstone of the movement. It gathers tries in explaining and modelling any spatio-economical phenomena: transportation, migration, land use and urban development, inter-industry relations, environmental and ecological forecast, resource management, urban and regional policy analysis, geographical information systems.

An interesting part of Regional Science is **gravitational models** [121, 47] which aims at describing human movements between various locations. This theory refuses to consider that the cost of a path is proportional to its length and postulates similarly to Newton's gravity law that [121]:

$$F_{ab} = \frac{P_a P_b}{d(a,b)^2} \quad (1.3)$$

where  $a$  and  $b$  are two cities of population  $P_a$  and  $P_b$  separated by a distance  $d(a,b)$ .  $F_{ab}$  is the "demographic force" between  $a$  and  $b$ : a measure of the interactions between people of  $a$  and  $b$  such as migration, travels, commercial exchanges... Note that  $F_{ab} = F_{ba}$ . This means that the "attraction" of  $a$  is weighted by the population of  $b$  to obtain the number of interactions from  $b$  to  $a$ .

This leads to Reilly's law of retail gravitation: the influence zone of each city is not defined by the perpendicular bisector between the cities but by the distance (from  $b$ ):

$$D_{b||a} = \frac{d(a,b)}{1 + \sqrt{\frac{P_b}{P_a}}} \quad (1.4)$$

More generally one could postulate a law of the form:

$$F_{ab} = C.P_a P_b . f(d(a,b)) \quad (1.5)$$

where  $f$  is a decreasing function and  $C$  some constant. The choice of  $f$  and  $C$  depends on the application. For instance [9] synthesizes that one obtains  $f(d) = 1/d$  for railway expresses,  $f(d) = 1/d^2$  for Korean highways and telecommunication flows,  $f(d) = \exp(-d/\kappa)$  for cargo ships.

For theoretical convenience, one could replace  $P_a$  by  $F_a = \sum_b F_{ab}$  in Eq.1.5 as a measure of the importance of the city. It is then possible to show by entropy minimization [136] that for a discrete and large system of cities, the "less informative gravitational law" for trips is

$$F_{ab} = A_a B_b F_a F_b e^{-\beta d(a,b)} \quad (1.6)$$

where  $A_a$ ,  $B_b$  and  $\beta$  are theoretical formal parameters to be tuned for a particular system.

More recently, Jensen [72] has shown that some commercial activities have interest to gather in a same point of a city and other can be uniformly distributed in the city.

### 1.1.3 Dynamics, Multi Agent modeling

As pointed out in [2] by M. Batty, the previous works consider the city as an inert matter, a system at equilibrium. In the 90's, computational improvement and the emergence of Multi Agent framework enabled the possibility to consider the city as a dynamical system, far from equilibrium and constituted of numerous little agents whose interactions dynamically change the shape and utilization of space.

A multi agent model (MAM) is a system compounded of an environment (for instance a planar field with a superimposed grid or a predefined map) on which evolve sequentially a large number of "microscopic" agents communicating and interacting. Several types of agents are defined by an interface of characteristics and behaviours toward environment. Each agent evolves according to the variations of the environment and the population modifying dynamically the land use.

Generally speaking if multi agent modelling tends to reproduce realistic trends, it needs a large calculus power and does not naturally propose a synthetic point of view on a phenomenon. The "simplified" version of MAM are Cellular automata, which work with one or two kinds of agents on a square grid with a very limited number of possible actions. There are many applications of MAM to city dynamics at various scales (micro or macro simulations, urban or regional levels). The book [2] presents along its chapters typical applications: asystematic mobility, pedestrian and vehicle traffic simulation, journey to work flows, land use changes, economic in a poli-nucleated urban area. There exist several software packages to handle common problems with urban multi agent modelling: MaGIA, The Time Machine, Cage, GioCoMo, EuroSim, SimPop...

## 1.2 Morphology analysis

### 1.2.1 Typologies

A first look at the internal structure of the city comes with **human geography** at the beginning of the 20th century. Geographers proposed a typology of cities [32], showing how

to recognize visually some characteristic patterns or "Figures de style", and relate them to the typography, or some basic structure (linear or radial). Fig.1.1 is an extract of [32] which shows examples of typical typologies of French cities.

The architect Stephen Marshall also presents in [86] a zoology of street patterns with an original and interesting reflection on the functionality and the perception of street arrangement completing the work of [81].

### 1.2.2 The Fractal City

In [52, 11, 123] the authors observe that a metropolitan organizes fractally (Fig.1.2). This is a feature of the non stable condition of urban growth: the fractal dimension, empirically calculated from the center of the area increases through time and tends to 2. Theoretical consideration on fractals and fractal dimensions can be found in App.A.

The city is a non compact pattern, there is a continuity between highly dense zones: city cores and their country side. Added to that some green areas remain to air the city core and create scattering. Together with a hierarchical comprehension of the city they are the two key ideas of the fractal analysis of cities. In fact this theory restricts itself to auto affine shapes such as Fourier's dust or Sierpinski's carpet and compares the cities features to these reference models.

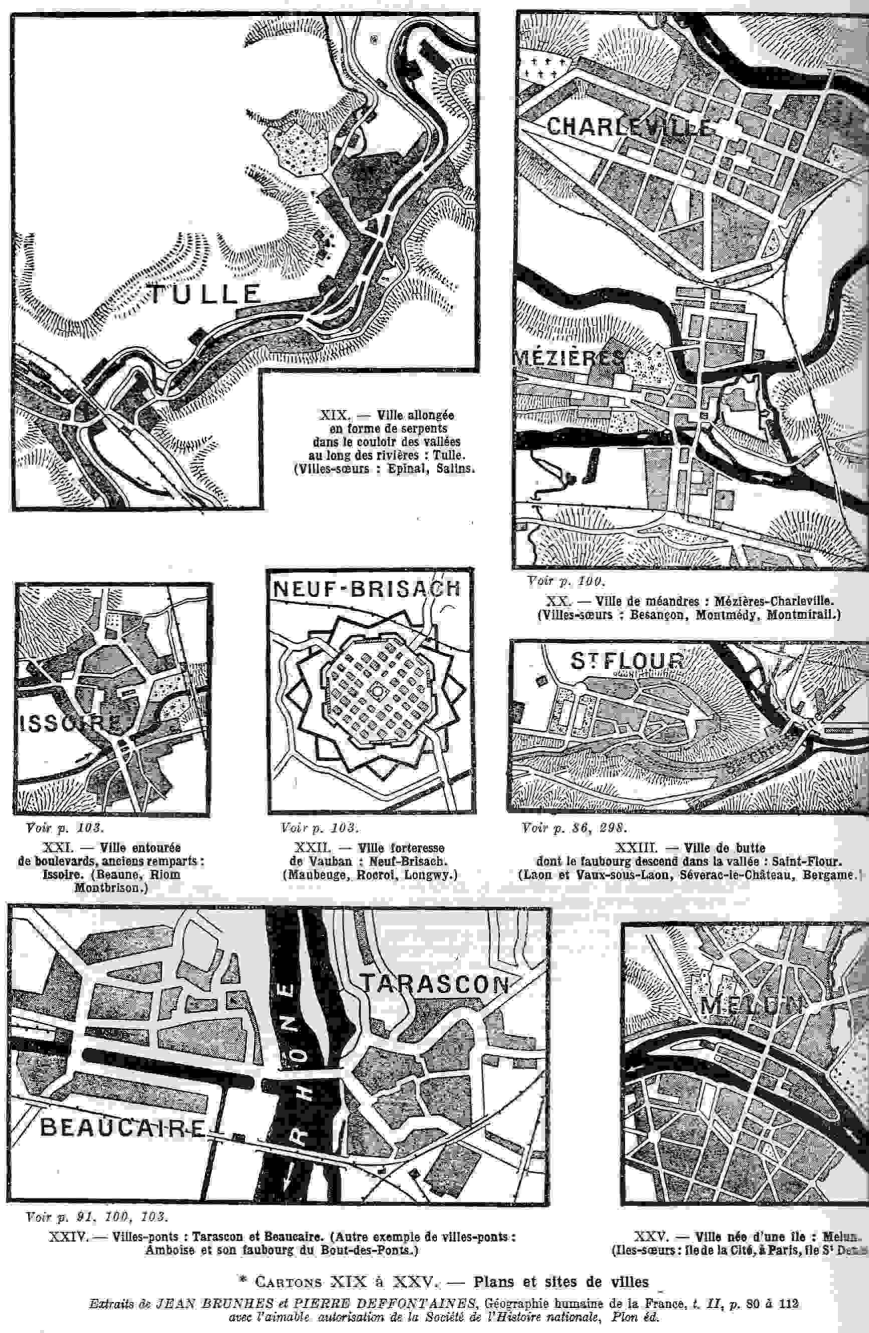
This theory raises some problems.

From an epistemological point of view, the "Fractal Theory of Cities" cannot pretend to be a scientific theory: they make an observation or hypothesis according to which the city shapes as a fractal (auto-affine would be a more coherent expression with the presentation of their work). But they get no deductive results from this hypothesis, they assume it is a fractal and so compute its fractal dimension. In science, the corpus of deductions has to be stronger than the hypothesis made at the beginning of the theory.

From a purely technical point of view, the dimension they measure highly depends on arbitrary methods. They have to choose if the city is lineal or circular. When circular, they have to decide of a single center point for the city and then they measure the fractality of the trace of the built up area radially from this point. The notion of "built up area" is also arbitrary, it results essentially from the thresholding of the density map. Added to that the box counting method they use to determine the dimension is known to be very sensitive and the practical curves they plot for instance in "Fractal Geometry for measuring and modelling Urban Pattern" ([2]) is very noisy.

### 1.2.3 Density profile

The first to get interested in density profiles of cities is Clark in the 50's with [40]. In spite of very modest techniques to compute this density for several cities, he shows for a dozen of cities around the world that the population density is a decreasing exponential of the distance to the center of the city. To obtain this he needs to define the notion of city center and to integrate population density along rings of constant radius. Remark this work was not easy and maybe not accurate with the tools of the 50's, indeed, he fits density profile



**Figure 1.1:** Different typologies of cities, street networks and general shapes illustrated by French town examples from [32]. In most cases, cities are structured by the local topography: a valley, meanders...

with only a five or six points. The article is very clear, methodical and pleasant to read but it would be worthwhile to redo the calculus with modern computers. He also shows that density gradient decreases globally with time. This observed phenomenon is taken as a postulate in [85].

More recent studies [18] lead geographers as Denise Pumain to consider density profiles can



**Figure 1.2:** From [123]. Observation of the fractal spatial layout of Besançon in France and the first steps of the Sierpinski Carpet used by the authors to explain fractal via the notion of auto-affinity.

be of different shapes: exponential, power law or Gaussian. Agreements are the decrease of density gradient with time and the existence of a trough in the density next to the very center of the city.

#### 1.2.4 Stochastic Geometry

To solve optimization problems from Telecommunication engineering, [57] suggests to replace the actual road network of a city with a stochastic model [122] "statistically equivalent". They get it down in [59] and obtain good results for Access Network Cost assessment in [61]. Two main ingredients are to be chosen: the intensity of the road network and its morphology. The intensity is a real positive parameter and the morphology is a referent tessellation model from a collection of classical and well mathematically studied tessellation: Poisson Line Tessellation, Poisson Voronoï Tessellation or Poisson Delaunay Tessellation. This supposes that a city is geometrically stationary: the observation of mean intensive parameters on a window does not depend on the choice of this window.

In fact this hypothesis is solely locally legitimated and the rigorous use of this method calls

for a first step of segmentation of the city into homogeneous area (see Ch.9 for models, Ch.10 for segmentation and App.C for a review on stochastic geometry).

## 1.3 From Graph to Complex Network Theory

When working on the network of streets in a city, the city is generally seen as a graph. Possibly vertices (or nodes) of this graph are intersections between streets and edges are portions of street that bind these vertices. This section reviews the joint history of city modelling and graphs, the practical interest of this modelling, its popularization with Geographical Information Systems and the systematic statistical study that follows.

### 1.3.1 Koenigsberg bridges, Operational research

Graph theory and city modelling were "invented" in the same time by Leonhard Euler in 1735. The question was to know whether it was possible to walk through each bridge of the city of Koenigsberg once and only once.

He observed that the organization of the bridges around the local geography can be - for this problem - reduced to a purely topological object: a graph. Furthermore this graph exhibits the property of planarity whose constraint is strong enough to provide a negative answer to the question. This illustrates how space constraints our transportation and way of life.

We will be more exhaustive in Ch.4 but let us give here some definitions on graphs to understand the state of the art. A graph  $G$  is a couple  $G = (V, E)$  where  $V$  is a finite set called vertex set and  $E$  a part of  $V \times V$  called edge set. If a couple  $(v_1, v_2)$  is in  $E$ , this means that vertices  $v_1$  and  $v_2$  are "bounded" or "in relation". For instance in a city graph, two intersections are in relation if there exists an actual piece of road between them (and this road must not contain any other intersection). If  $E$  is symmetric ( $(v_1, v_2) \in E \Rightarrow (v_2, v_1) \in E$ ), the graph is undirected and the two opposite edges are considered as a single edge. The degree  $k_i$  of a vertex  $i$  is the number of edges that pass through  $i$  (half this quantity in an undirected graph context).

When it is possible to represent  $V$  by point on the plane,  $E$  by a collection of curves binding the vertices in relation without having edge-crossings outside of  $V$ , the graph is planar. Street networks are planar (see Fig.1.3, right). Planarity is a huge constraint on the features of the graph which will induce difficulties to compare actual street network with relevant reference planar graph models. When the graph is planar, a new set of objects can be defined: the cells. They are portions of the plane delimited by elementary cycles of edges. In the following section, the number of vertices will be written  $m$ , the number of edges  $n$  and the number of cells  $c$ . These three quantities are linked by the Euler's formula:

$$m - n + c = 1 \tag{1.7}$$

Such a topological point of view is rich in operational research to solve tricky optimization problems such as the shortest path problem, the salesman problem or the minimum spanning tree problem. This section mainly gets interested in the analysis of the topology of a graph



representing a city.

The graph theory [16, 64] leads to the complex network theory [23, 98, 1] when graphs are large and feature some global, non trivial phenomena like small-world behaviour (the path length between two nodes is very small compared to the size of the network) or scale-free degree distribution (a hierarchical organization that may lead to a small-world behaviour), see App.B for more theoretical considerations.



**Figure 1.3:** The map of Königsberg (left), a diagram to emphasize the structure of the river (middle) and the abstraction of the topology of this structure as a graph (right). From Wikipédia.

### 1.3.2 Geographic Information System (GIS)

A GIS is a software allowing to handle rich geographic data: capture, modify, apply queries to them and producing thematic maps. It deals with vector or raster data. Vector data are compounded of spatial objects (points, polylines, polygons) to which are associated additional information (thickness of the street, street name, population...). The objects can be a representation of a road network (polylines), regions etc... Rasters are simply images, a grid of color pixels to produce a map on a particular topic. Web sites like Google Maps present a raster rendering of underlying vector data.

### 1.3.3 Topological Statistics

The advent of GIS has allowed an exhaustive study of cities, in particular the research of statistical correlations between the different topological features that may present a map. In what follow, we consider a map as a (large) undirected graph  $G = (V, E)$  with  $m$  the size of  $V$ ,  $n$  the size of  $E$ ,  $k_i$  the degree of the vertex  $i$  and  $\langle k \rangle = \sum k_i/n$  the average vertex degree.

**Vertices and Edges topology analysis** [35] and [33] get interested in the statistics of topology in worldwide cities. They work on square extractions of maps (Fig.1.4).

They describe at first basic indicators such as the number of vertices, edges, mean vertex degree, total length of edges, mean edge length. They show that the mean degree is very low, typically between 2 and 3 which forbids the network to behave as a scale-free network. The given explanation is that the planarity strongly constraints the topology. This is partly



**Figure 1.4:** Square extraction of the footprint of Verona. Element of the data base used in [33] to study statistical properties of street networks.

true even if the planarity only constraints the mean degree to be smaller than 6 the bound being reached for instance in a Delaunay triangulation (see App.B and App.A). [35] remarks that the degree is hardly ever larger than 5 but [33] describes the whole degree distribution by an exponential fitting, which is exaggerated toward the support of the distribution.

[33] shows that the assortativity coefficient  $\Gamma$  which describes if vertices connect to vertices of the same degree or not is rather negative: there is no connection between vertices of high degree.

The meshdness coefficient:

$$M = \frac{m - n + 1}{2n - 5} \quad (1.8)$$

measures in [33] if the map is highly connected ("constructed") or not. Indeed  $m - n + 1$  is the actual number of cells in the network and  $2n - 5$  is the maximal number of cells reachable for a network with the same number of vertices. But [9] shows that for large network  $M$  writes as a function of the mean degree  $\langle k \rangle$ .  $M$  typically ranges between 0.01 and 0.3.  $M$  is a measure of the cost of the network. Other measures of the cost can be defined, we will write them  $M$  generically and the efficiency Eff can be defined for instance by:

$$\text{Eff} = \frac{1}{n(n-1)} \sum_{\substack{i,j \in V \\ j \neq i}} \frac{d_{ij}^{\text{eucl}}}{d_{ij}} \quad \text{or} \quad \frac{1}{n(n-1)} \sum_{\substack{i,j \in V \\ j \neq i}} \frac{1}{d_{ij}} \quad (1.9)$$

where  $d_{ij}$  and  $d_{ij}^{\text{eucl}}$  are respectively the shortest path and the Euclidian distances between vertices  $i$  and  $j$ . Indistinctly, the efficiency is high if it is possible to go fast from a vertex to another. This is the case if the graph is highly connected: if  $M$  is high. But having an high  $M$  asks in practice for work and money. In an actual city,  $M$  and Eff are two antagonistic parameters: they want to minimize  $M$  and to maximize Eff. The resulting city is a compromise between these two parameters.

In these articles, results are normalized by constructing two extreme graphs from the same set of vertices as the actual city: a "cheap" tree with the Minimum Spanning Tree and an "expensive" highly connected graph with the Greedy Triangulation. A measure  $\mu$  is normalized to  $\mu^*$

$$\mu^* = \frac{\mu - \mu^{\text{MST}}}{\mu^{\text{GT}} - \mu^{\text{MST}}} \quad (1.10)$$

It was a great and new idea to use two structurally different planar graphs as reference models or "O-models", the planarity being a strong constraint, the classical reference random graph (App.B) is not a relevant point of comparison for planar graphs.

$\text{Eff}^*$  is an increasing function of normalized  $M^*$  with a saturation:  $\text{Eff}^* \sim_{M^* \rightarrow 1} 0.8$ . The grid like cities obtain a better efficiency than medieval cities for a relative cost between 0.3 and 0.4. Cycles are also under study which shows that a city has very few triangles: a street network is not transitive. [34] explores also a set of positive correlations between the city robustness (number of nodes to remove so that the largest component size is 50% of the original network) and  $M$ . The result is quite general: the network is as robust as it is costly. Note that in the limit case of the tree, the network is more robust to random attacks but much less to selective ones.

**Vertices and edges Topology general statistics** [8, 10, 9] present several empirical laws binding  $L$  the total length of streets,  $n$  the number of vertices and  $m$  the number of edges. First  $m \propto n$  which means the average degree of cities is quite constant. It is observed that the total length of streets in a city follows a power-law distribution:  $\mathbb{P}(L) \propto L^{-\gamma}$  with  $\gamma \simeq 3.36 > 3$  which means that both mean and variance exist and that there is a typical edge length. From the two previous laws, it predicts that  $L \propto \sqrt{n}$ .

**Cells features** [79] studies the "dual" topology: the connectivity of cells in the network of German cities. Since the dual graph is also planar the mean number of cell neighbours or cell degree cannot exceed 6. It exhibits indeed a predominance of degree 4, with much more variance than vertex degree (there are cell degrees of 19). As for the geometry of cells, it shows that the probability of an area  $A$  is  $\mathbb{P}(A) \propto A^{-\gamma}$  with  $\gamma \simeq 1.9$  for  $10^4 < A < 10^7$  m<sup>2</sup>. The distribution of form factors:  $\Phi = 4A/(\pi D^2)$  ( $D$  is the diameter of the incircle) indicates a broad intra and intercity diversity (Fig.1.5).

### 1.3.4 Centralities

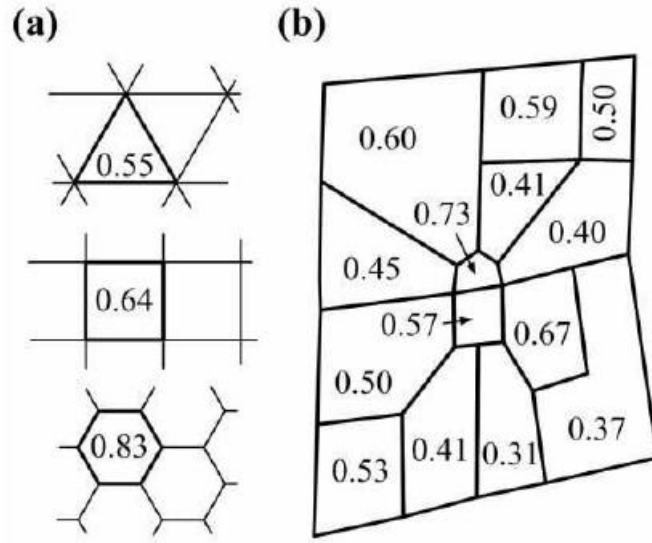
First introduced in social science [12], a centrality is a measure defined on the set of a graph's vertices or on its edges to quantify to what extent an element is important, "central" for the graph. A central point is a point "easy to reach", a point "one has to cross to go from a point to an other". For instance the root of a spanning tree should have an high centrality whereas its terminations should not. Here are classical centralities defined for the vertices  $i \in V$ . Notice that the same notion can be defined for edges  $\in E$

The **degree centrality** measures a very local centrality:

$$C_i^D = \frac{k_i}{n-1} \quad (1.11)$$

The **closeness centrality** measures to what extent a vertex is near to the others with respect to the street network:

$$C_i^C = \frac{n-1}{\sum_{\substack{j \in V \\ j \neq i}} d_{ij}} \quad (1.12)$$



**Figure 1.5:** Form factor  $\Phi$  for three reference regular lattices (a) and for the cells of a street network (b). In real street networks, the form factor exhibits a wide diversity. From [79].

The **betweenness centrality** measures the proportion of shortest paths that pass through the vertex  $i$ :

$$C_i^B = \frac{1}{(n-1)(n-2)} \sum_{\substack{j,k \in V \\ j \neq k \neq i}} \frac{\sigma_{jk}(i)}{\sigma_{jk}} \quad (1.13)$$

with  $\sigma_{jk}$  is the number of shortest paths between  $j$  and  $k$  and  $\sigma_{jk}(i)$  is the number of these paths that cross  $i$ .

The **straightness centrality** measures if the network is "star-shaped" from a vertex:

$$C_i^S = \frac{1}{n-1} \sum_{\substack{j \in V \\ j \neq i}} \frac{d_{ij}^{eucl}}{d_{ij}} \quad (1.14)$$

A variant of straightness is the **efficiency centrality**:

$$C_i^E = \frac{\sum_{j \in V, j \neq i} \frac{1}{d_{ij}}}{\sum_{j \in V, j \neq i} \frac{1}{d_{ij}^{eucl}}} \quad (1.15)$$

The **information centrality** measures the non robustness of the network toward the removal of a vertex:

$$C_i^I = \frac{C^S(G) - C^S(G - \{i\})}{C^S(G)} \quad (1.16)$$

where  $C^S(G)$  is the average straightness centrality of the graph  $G$ :  $C^S(G) = \langle C_i^S \rangle_{i \in V}$ . Two kinds of studies are (jointly) applied to centralities in road networks: the spatial distribution and the global distribution of centrality in the city [42, 117, 106]. The spatial distribution is represented with the city graph and colours on vertices coding for different values of centrality, which provides a direct reading of the map. If the closeness presents

above all a radial decrease from the center of the image, the other centralities capture routes and area, emphasizing main streets or ill deserved zones. [108] shows correlations between centralities and densities of retail and services.

$C^B$  is single scaled, well approximated by a Gaussian curve ([79] finds a power law, interpreting that only a few nodes maintain the transportation and might be subject to congestions),  $C^C$  and  $C^S$  are seen as mainly linear through the city and the authors claim that  $C^I$  has distinct behaviour if it is a planned city (exponential) or if its is a self-organized one (power law with exponent between 2 and 3).

Centralities have been defined firstly for abstract graphs. Only the straightness and efficiency centralities take in account the spatiality of a street network (not really the planarity but the spatiality) by introducing Euclidian and Shortest Path distances in the measure. In Ch.4 we will define a mathematical framework allowing to handle easily with the geometrical and spatial aspects of a city and apply this framework in Ch.6 to redefine centralities coherently for street networks.

### 1.3.5 Random walks

A random walk on a graph is a stochastic (or random) process on that graph. A number  $N$  of agents evolve sequentially and independently on that graph with respect of a transition kernel providing probabilities to move from a vertex to one of its neighbours. Generally, after a large number of steps, agents are distributed on the graph with respect of a stationary distribution (see [14] for the theory and more precise considerations). This distribution is only determined by the transition kernel and the topology of the graph.

Monograph [21] is an exhaustive study of random walks on topological graphs representing maps. They use classical discrete Markovian processes on graphs (purely random walk, or walk biased by a centrality) to characterize and compare various urban structures (Venice, Rothenburg, Amsterdam, Bielefeld, Manhattan). As indicators, authors study:

- The mixing rate of a simple random walk.
- The relative entropy and the entropy rates of the simple random walk and a random walk biased by centrality.
- The first passage time and the hitting time of streets.
- Community structures and map segmentation by diffusion processes.
- Various thermodynamic quantities such as the city entropy.

The book remains quite general on random walks and its dedication to cities is more a pretext than a real concern. Indeed it is not realistic to imagine agents jumping from a street intersection to an other, it would be interesting to study random walk in a continuous framework, for instance the one we will introduce in Ch.4. Added to that it would be interesting to find back a centrality as a stationary distribution rather than impose it a priori in the random process.

### 1.3.6 Other Complex Networks

The conclusion of the works presented in this section is that the city does not behave like a classical complex network: it does not exhibit classical small-world or scale-free features. It presents rather trivial behaviours, partially due to their spatial embedding.

The city should be compared to other spatial networks. [9] reviews general and particular features of spatial complex network, including cities. [34] is the similar study as in [33] for ant networks of galleries where we see the same qualitative relationships between cost, efficiency and robustness. In these articles they built reference models as two extremal kinds of planar graphs: the minimum spanning tree and the greedy triangulation. Nonetheless they construct these graphs with the same vertex set as the actual city. One could imagine that the modification of the edges induces also a modification on the presence or position of some vertices. [24, 103] show from the geometry of cells and that a city is a hierarchical network to be compared to crack patterns.

## 1.4 A representation of the city as a collection of objects arranged in space

### 1.4.1 Human perception of Space

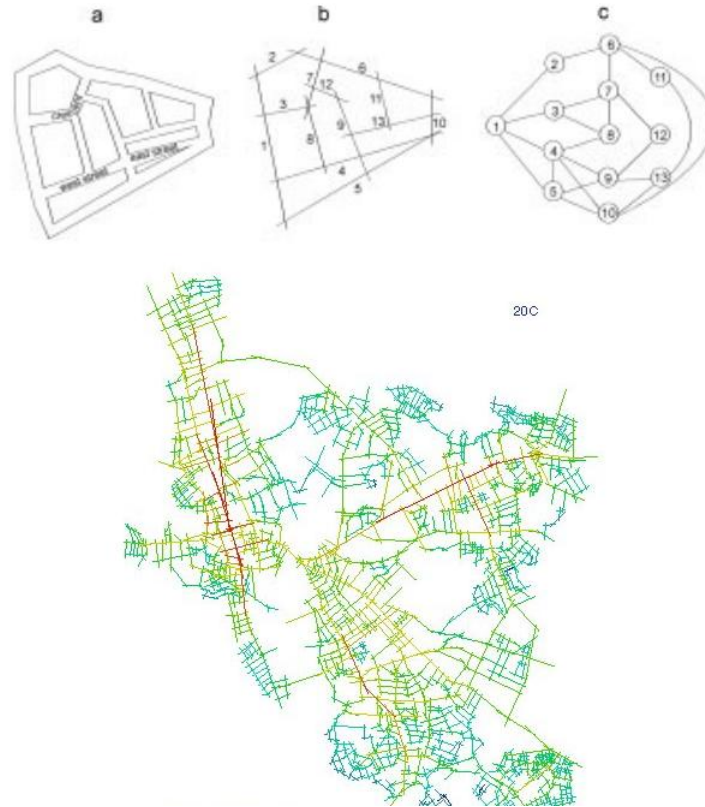
In [81] the architect Kevin Lynch reports through several and various examples his study of perception of space by people. He concludes with the notion of **mental map** that constitutes our perception of space through five elements: paths (streets allowing to travel from a point to another), edges (that provide spatial delimitations to streets, like walls, buildings...), districts (coherent local zones in the city), nodes (central and strategic points), landmarks (identifiable places and objects that serve as reference points). In short urban space can be divided into objects relevant to our perception and our way-finding: on the one hand axes, on the other hand blocks or "islets". The one being the "dual" of the other.

This is the foundation of the whole "Space Syntax" theory and application. These foundations have been re-explored by the main authors of Space Syntax for instance in pedestrian way-finding [68, 67]. These two references are representative but not exhaustive.

### 1.4.2 Space Syntax

Space Syntax [73, 66, 67, 2] is a framework dedicated to the analysis of cities. From a city layout, it defines a "visibility graph" called the "axial map" (Fig1.6 Up) and then measures the topological integration of the axes. This integration is supposed to be representative of flow of people in this axis and thus of its importance. Integration is plotted into a color map which provides a visual tool to diagnose the valuation of elements in a map (Fig1.6 Down).

**Visibility Graph** Axial lines are used to represent directions of visibility; they represent the longest visibility lines in two-dimensional urban space [73, 66, 67]. Axial lines are



**Figure 1.6:** Space Syntax framework. On Top: a virtual map with street width (a), the visibility lines of this map (b) and the associated connectivity graph (c). Down: the color plot of integration for the city of Oxford. From [73] and Bin Jiang's web site.

calculated from a map of a city representing streets as thick lines (Fig1.6,a). The drawing of these lines is called the visibility graph (Fig1.6,b). Each axis has an ID (identification number) and two axis that meet are adjacent. This principle induces a graph structure (connectivity graph): axis are vertices and their intersections are edges (Fig1.6,c).

**Integration** From the connectivity map, [66] defines various observation parameters. From one node  $i$  in the connectivity graph,  $s \in \mathbb{N}$ ,  $N_s(i)$  the number of nodes reachable with  $s$  steps and  $q$  the maximal number of steps. Then

$$l_i = \sum_{s=1}^q s \cdot N_s \quad (1.17)$$

is called the connectivity of node  $i$  if  $q = 1$ , the global depth  $l_i$  if  $q = +\infty$  and the local depth if  $1 < q < +\infty$  (typically  $q = 3$  to compute the local depth). The integration of an axis is calculated by its "Relative Asymmetry" index:

$$RA(i) = 2 \frac{l_i / (n - 1) - 1}{n - 2} \quad (1.18)$$

which is the mean global depth of node  $i$  normalized by the mean global depth of the complete graph  $\mathbb{K}_n$  [21] shows that the distribution of  $RA$  can adopt various shapes according

to the city.

**Implementation** Space Syntax is implemented as a GIS module. It is widely used in architecture and town-planning to quantify the navigability of a network and study socio cultural factors as crime, traffic congestion, hospitals, shops or museums placing...

**Criticisms** If there is a large literature on the use of space syntax, there is a little on its mathematical and computational foundations. [111] is an article pointing out some inconsistencies of Space Syntax:

- The decision making of a pedestrian does depend on the metric, not only on the topology.
- The axial map discards 3D information or land use.
- There is not uniqueness of the axial map for a given "urban texture".
- The visibility graph is very sensible to small deformations of the map. It provides the example of a grid and the same grid with a little variation in vertices position: axial analysis passes through long line to little segments.
- Curved street are very different from straight lines in the space syntax framework: they are not represented by a single axe.
- Edges effect and segregation paradox: consider two grids, the main axes are at the middle of these grids. Then bind these grids by a segment, the main axe is this new joint and middle axes have then a poor integration.

To this list of "inconsistencies" follows an exchange between the Space Syntax creator Hillier: [69] and Carlo Ratti: [111, 110].

From our point of view, inconsistencies listed by Carlo Ratti are more or less legitimate. Indeed space syntax suffers of some inconsistencies on its definition and mathematical formulation. The visibility map is not robustly defined with respect of variations on the representation of the city (small deformations or also sampling of the map), the resulting integration map depends arbitrary on the way the map is stored in memory. We will provide a mathematical framework to compensate rigorously for these difficulties in Ch.4. Ratti reproaches to space syntax not to take in account 3D information. We will neither consider these information in this manuscript. We imagine 3D information are strongly correlated to spatial information and we will try to see how much information is contained in the planar geometry of a map, how the particular arrangement of a city can determine its local and global evolution. To come to the segregation paradox, it is possible that a small new street segment becomes an important transportation hub after the whole city has reached an equilibrium with the new structure in Ch.6 we will present a comparable measure and emphasize its robustness toward local and global scales. What is more disappointing is that former main axes that keep an important local function have a poor integration. It would be interesting to have a measure of integration that has both a global and a local interpretation. To end with, space syntax uses the argument of "human perception" to legitimate its



integration relevance. In Ch.6 we will avoid to use such argument and show mathematically the relevance of our measure.

### 1.4.3 Axial representations of the City

As said above, the visibility map is not robustly defined with respect of small variations of a map. It is very sensitive to local curvature and to the sampling of the map. Various methods have been proposed to overcome this inconsistency. The notion of axis is replaced in [74] by the notion of named-street: two axes are the same if they have the same name in the data base. In [49] two axes are melt if their angle is lesser or equal to an arbitrary threshold ( $45^\circ$ ). [105] presents the Intersection Continuity Principle: two axes are melt at an intersection if they make the largest convex angle between all angles at the intersection.

### 1.4.4 The "dual" analysis

The graph induced by axes (whatever is the method used to recover this notion) is called in the literature the "dual graph" and the plain graph representing the map the "primary graph". The study of this graph is referred in [105] as the "dual approach" even if the meaning of dual is completely different from the common mathematical sense as used previously (the cells connections is the dual of the cell's edges). [114] speaks of an "information network".

The "dual" street graph (where each street is now a vertex and an edge is drawn between streets that intersect) is no more planar, similarly as for the visibility graph in Fig.1.6,c. It behaves much more like a classical scale-free or small-world network than the primary graph. If [105, 21] suggest that the connectivity of streets follows a power law distribution  $1/k^\gamma$  with  $2 < \gamma < 3$ . [74] finds for various cities (Munich, San Francisco, Gäval) polynomial fitting in the log-log scale plot, [75] shows it is heavy tailed but not necessarily with a power law. [74] also studies the  $k$ -clustering coefficient for several parts of Sättra:

$$C^{(k)}(i) = \frac{2l_i^{(k)}}{m_i^{(k)}(m_i^{(k)} - 1)} \quad (1.19)$$

where  $l_i^{(k)}$  is the number of edges at a topological distance smaller than  $k$  and  $m_i^{(k)}$  the number of vertices with that property. The clustering coefficient is much higher than the one of a random graph plus there is for cities under study a very low maximal topological distance between streets (7). It concludes the scale-free behaviour is not obvious but that the dual graph is a small-world network. For [75] the connectivity distribution shows that the city is a hierarchical structure and they propose a fractal model for the placement of roads whose fractal dimension is related to the exponent in the connectivity distribution. [105] computes classical centralities on the dual graph.

In conclusion, there exist two approach to study street network: the "primal" approach and the "dual" approach. The "primal" only exhibits trivial phenomena, the planar embedding of a map being a strong constraint. The "dual" approach even if not always robustly defined by authors seems to be more promising.

## 1.5 Growth models

### 1.5.1 Cellular automata

A cellular automaton is somehow the simplest ancestor of Multi Agent models. A square grid is given and each site can take values in a finite alphabet. Local rules are defined to make the grid evolve. They have become well known from John Horton Conway's Game of Life in the 70's. One of the most famous cellular automata is Eden's model [46] which is to mimic the growth of a colony of bacteria. The Eden model has been modified to reproduce trends of the urban growth [15] A huge work has been realized on the modelling of city by cellular automata. It deals with the global evolution of the city morphology [26] or the differentiation of space [2].

### 1.5.2 DLA

A Diffusion-Limited Aggregation is a process where particles displace randomly in space, following a Brownian motion and end up to cluster when they meet other particles. They are a particular case of cellular automata when the size of the pixel tends to 0. DLA models for instance the growth of Saffman-Taylor instability. They generally produce fractal shapes. That is why they are proposed in [11] to model the growth of a city. Nonetheless the result is a single tree-shaped cluster. In [87, 85] they use more sophisticated correlated percolation models in presence of a density gradient to produce several clusters still fractal on the boundary but with compact center Fig.1.7.

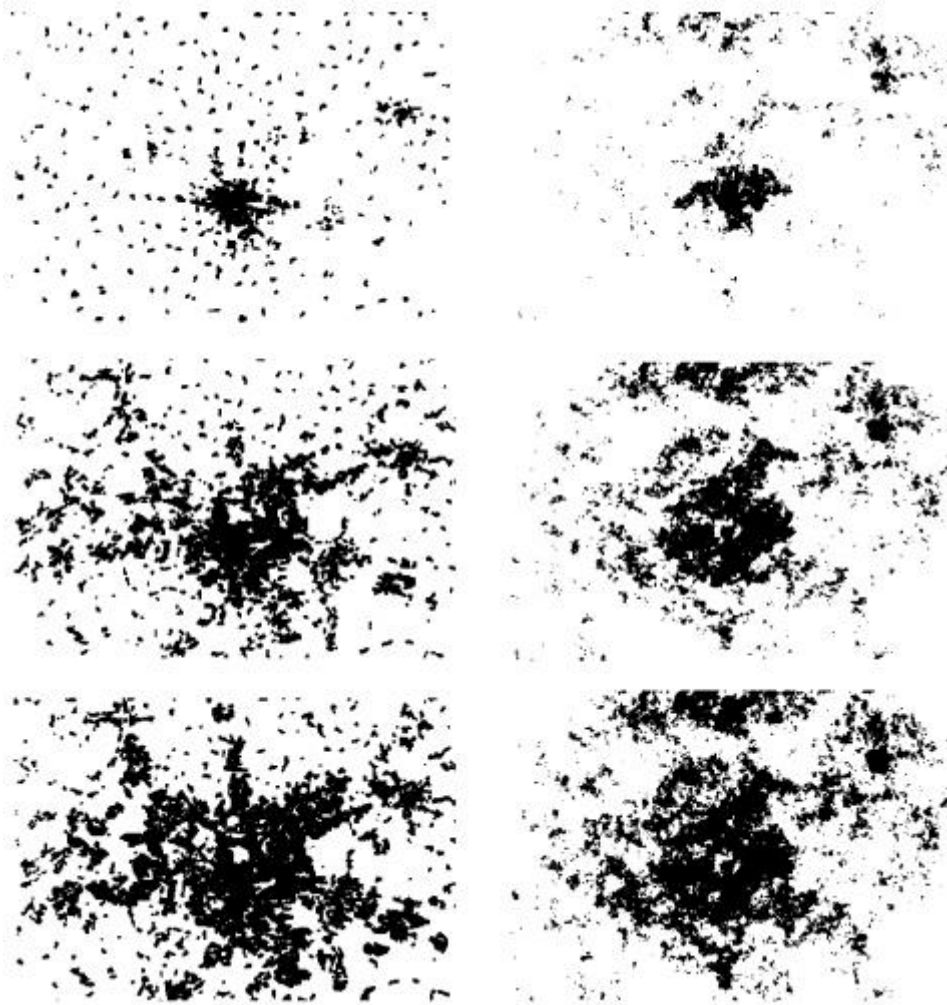
### 1.5.3 L-Systems

[101, 104] extend L-Systems (or Lindenmayer Systems [109]) to create real-like maps for graphics and video games. A L-system is compounded of an alphabet  $A$ , a set of words  $A^*$ , a set of postulates or rules  $P$  defined on  $A$  and an initial state  $w_0$ . The system makes  $w_0$  evolve recursively to  $w_0$  with the evolution rules, some stochasticity and by taking care of preserving the global coherence of the system.

From a population density map and a geographical constraints (elevation map, boundaries see Fig.1.8, Left), the program proposes at each step to make the city grow of a segment. To this it chooses the best road pattern (raster, radial, branching) and coherency is checked (for instance that a new segment connects in an existing edge) and if necessary corrected. A resulting map is shown in Fig.1.8, Right.

### 1.5.4 Complex Network Growth

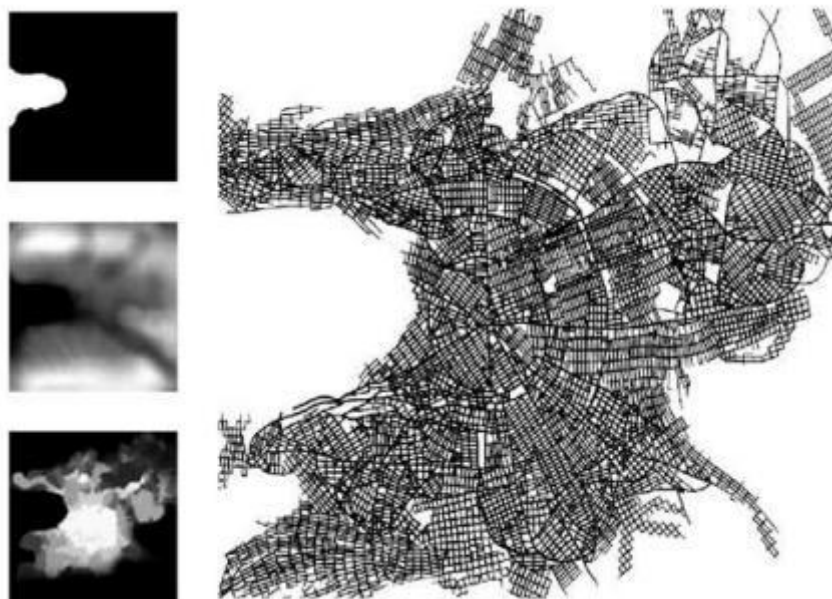
The first attempt to model the growth of a city from the network point of view is [8]. The presented model seeks out to mimic the observed cell area distribution observed in [79]:  $\mathbb{P}(A) \propto A^{-\gamma}$  with  $\gamma \simeq 1.9$ . The model is sequential: the city is a graph  $(G_n = (V_n, E_n))_{n \in \mathbb{N}}$  initialized with  $G_0 = ((0,0), \emptyset)$ . At each step a new *center*  $V_{n+1}$  is added with respect



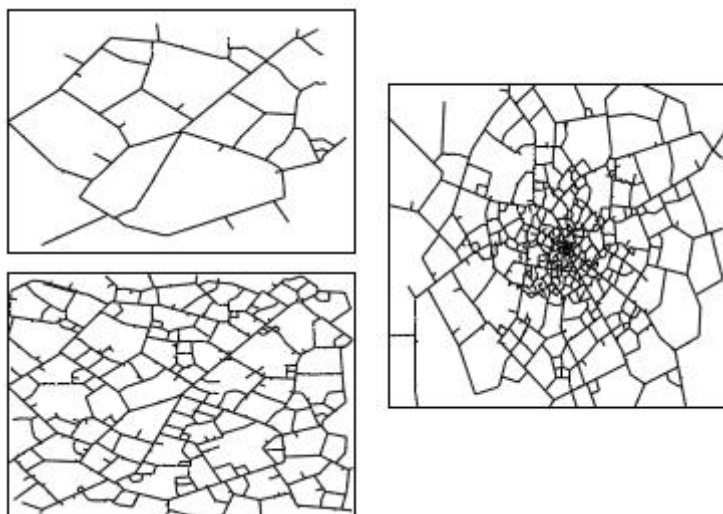
**Figure 1.7:** From [87]. Left column: infrastructure distribution in real cities ; right column: simulated correlated percolations to mimic these distributions.

to a radial isotropic probability density  $P(r) \propto \exp(-|r|/|r_c|)$  with  $r_c = 0.1\text{km}$ . Then it connects to  $G_n$  that become  $G_{n+1}$  by adding to  $E_n$  all segments whose end points are  $V_{n+1}$  and points in the relative neighbourhood (see Sec.A.3.3) of  $V_{n+1}$  in  $V_n$ .

In [10] the same authors provide refinements of this model by coupling the evolution of the road network and the population density. They subdivide a bound window into square sectors and assume that existing network on each square defines a probability of settlement. This probability on square  $i$  depends 1) on the rent price  $C_R(i) = A\rho(i)$  where  $A$  is a constant and  $\rho(\cdot)$  the network density 2) on the accessibility  $C_T(i) = B(g_m - \bar{g}(i))$  where  $B$  and  $g_m$  are constant and  $\bar{g}(i)$  is the mean betweenness of nodes in square  $i$ . The income of a new center in a sector  $i$  is then of the form  $K(i) = Y - C - R(i) - C_T(i)$ ,  $Y$  being a constant which is increasing with the mean centrality and decreasing with the cost of the location. A center is chosen in square  $i$  with a probability proportional to  $e^{\beta_0 K(i)}$  which rewrites  $\exp(\beta(\lambda\bar{g}(i) - \rho(i)))$ ,  $\beta$  and  $\lambda$  being two tunable parameters. If  $\lambda$  is small, new centres appear where the density is low (cheapest places) which produces an homogeneous spatial network. If  $\lambda$  is higher, typically  $\lambda = 8.0$  for  $\beta = 1$  clusters and a particular central



**Figure 1.8:** From [104]. A map (on the right) generated by a L-System (CityEngine) from maps of water, height and population (on the left).



**Figure 1.9:** From [8]. On the left evolution of a spatial network with uniform distribution of centres on a window ; on the right, the same with an exponential distribution of centres.

core appear. The radial density variation from the core is well fitted by an exponential, the same exponential as it was enforced in the first article [8].

A "dual" approach can be found in [20]: they do not consider an intrinsic evolution of the network but an evolution by a succession of division of the cells an parcels which induces evolution of the road network and ensures a power law on the distribution of the areas.

The DLA models are used to recreate fractal cities. The only purpose was to propose a model that has a fractal result. But DLA are indeed generic models known to produce fractals.

And the articles never give a city-based interpretation of their results. City Engine with L-systems produces impressive maps and can easily be extended to produce 3D landscapes. Their results are visually satisfactory but they are not explicative. The network growth model is the basis of the work we will present in Ch.8. From local rules emerge global features observed in real cities. When biasing by centrality the choice of new centres, the local choice is made with a global integrated criterion. In our model we will replace the topological point of view (centres are added and create new street intersection) with a geometrical point of view: the growth of the city will not be determined by the position of former centres but by the whole continuum of infrastructures along the streets. Our model will also be auto-sufficient: we will not impose a priori density distribution for new centres but make this distribution emerge from simple local interactions.

## 1.6 Synthesis

In the first chapter of the collaborative book [2], "Fifty Years of Urban Modeling: Macro-Statics to Micro-Dynamics", Michael Batty describes the advances of urban modelling through XXth century. Michael Batty, architect-planner and geographer is a major contributor of modern urban modelling. He has participated to fractal theory, DLA models, Space Syntax, Urban perception and traffic... He presents the evolution of urban modelling with tree parallel "time lines": Models, Cities and Planning. These time lines present independently the evolution of modelling, actual cities and planning method from the industrial "monocentric" city to the modern global city passing through the poly-nucleated city. He also proposes a "deconstruction" of this time line which shows the intricate links between these three lines.

We propose here to synthesize modern urban modelling through a list of complementary points of view. Two levels of observation complement each other: **the macro and the micro**. The macro gets interested in the statistical distribution of measures of cities (size, population, production...) and their correlations. The micro models the behaviours of agents or local infrastructures and their interaction to predict displacement flux or emergence of phenomena.

These approaches can also be divided in **spatial and non spatial** approaches.

The spatial approach considers **the global shape of a city or this of its road network**.

# Mixing town-planning with Physics

## Contents

---

2.1	Science and Modelling . . . . .	<b>43</b>
2.2	Shape emergence . . . . .	<b>44</b>
2.3	What is a city ? . . . . .	<b>45</b>
2.3.1	Origins . . . . .	45
2.3.2	Town planning . . . . .	46
2.3.3	Elements of geometry . . . . .	46
2.3.4	Functions . . . . .	47
2.3.5	Diversity . . . . .	48
2.3.6	Construction . . . . .	48
2.3.7	Comparison with other networks . . . . .	49

---

—**Mixing town-planning with Physics: Synthesis**—

The purpose of this chapter is to explain simply to a town-planner what a **model** is and to a physicist what a **city** is. We also discuss the notion of **shape** and the recent theory of **morphogenesis** that shows that a lot of shapes in nature are determined by endogenous phenomena: the dynamics of an evolving object imposes strong constraints on its shape. To what extent can we develop a morphogenesis of the city?

We will simply define a city as a set of **distributed infrastructures spatially evolving** and whose global wealth is sufficient to maintain the pace of life.

A city structures around the **duality of its blocks and of street network**. Former blocks are divided (as in crack patterns) into smaller blocks with a strong constraint: the cutting segments must arrange into long lines forming streets.

We will distinguish between **organic cities and planned ones**. The first is the result of local non concerted interactions trying to optimize at a certain moment the efficiency of a particular new point. In the second the city is in one piece: all the infrastructures are placed in the same time, creating globally supposedly optimal patterns.

The geometry of the city is adapted to the **local geography** that acts as an outside constraint shell in its development.

—**Contributions of this chapter**—

1. A qualitative study of cities with comparison to veins and crack networks.

## 2.1 Science and Modelling

" What is a model ?

- You know it is somewhat like the ships on your shelves, it is reduced boats scientific people construct.

- Then scientists play all day long ? They build toy models to have fun ?

- To have fun ? No they do not !

Ah, indeed they do, you are right. Let me tell you a story little girl. Science is necessary to our survival but that is not why we made it. Science, do not believe what the dictionary says is above all a vice. "A man who thinks is a corrupted animal". Science is the set of all the things we know. Well, we have made a huge step :)

You have to admit that you and your awareness are surrounded by a world that has an intrinsic reality, that is to say that does not depend on you and your observation of it. A very well-known man whose name was Plato thought and this is widely obvious or admitted by the whole scientific community that you and in the continuity all the other persons do not have a direct access to truth. He says that all we can and agree to see of the truth are its projections on the wall of a dark cavern. That is the same as if you saw a rabbit only by its shadow.

Science implicitly postulates the existence of a truth that only makes its presence known to us through reality that is another time filtered by our perception. Science tends to understand truth by going past our senses to limit to reality that does not depend - by definition or maybe by hypothesis - on the person that considered it.

Here we are, Science is the set of all what people have said and will say about reality as an approximation of truth. Truth would be an idea we cannot reach directly. In fact I would prefer to see science as an approach, a way to comprehend world that wants to be independent of perception. To this a lot of work has been done to define coherence rules, reasoning and methods.

All what has occurred, that have been reality at one time is now true as a story. That is the point of history. It is a science also. But what we are speaking about is what will always be true.

Physics is the part of science that is concerned by this. If I want to be coherent I would have to claim biology is a part of physics. In fact traditionally people split physic in several parts. Biology is the part of physic that looks at living systems. For a long time physic and biology have been totally separated. They have started to melt partially again.

Now that we know what physic is we can speak of models.

A scientific model you were right is the same as a reduced ship model. A model is a way to handle with reality. As we reduce ships to put them on the shelves, we reduce reality to put it on our minds. When reducing we simplify.

In fact, if I have understood what I was telling, the first and most important model that we have had is language. This model says that reality is constituted of elementary elements we can replace by symbols and describe their interactions. That is the power of models. Language even if it is not the necessary condition of thinking is a highly efficient catalyst of it. We have extracted from the world a relatively small number of words that interact in sentences. And this allows us thinking beyond our perception.

So, what we ask to a model is to organize our perception of reality by simplifying it or let



us say by adapting it to the structure of our mind. "

## 2.2 Shape emergence

I think we have to define separately ideal shapes and physical shapes.

An ideal shape is a characteristic of a set on a particular metrical space. "Shape" is an equivalence relationship: two sets are of the same shape if there exist a rotation, a translation and a dilatation transforming one into the other. Shape is all the information of a set without consideration of size, position and orientation.

In physics - the science of real objects - this definition seems to be too rigid. It is generally connoted to the outlines of the object, its exterior lines. It may depend on the perception we have of the object, the scale of observation. Let us consider a rabbit (rather than a piece of wax). This rabbit is articulated, its stoutness is variable with time but we would require this rabbit keeps its "rabbit shape" as an intrinsic attribute. All the same in the set of all rabbits, there is a variation in features, they can be more or less hairy, they can have lost their tail, the relative size of ears can change.

The "shape" of a rabbit is distinguished from its posture. The posture would be the mathematical description of an instantaneous picture of the rabbit. The shape of a rabbit is an ideal, an object is said to be a rabbit if it presents a tolerable variation to that ideal. We will call that ideal "the typical rabbit". It may not be a rabbit. It is a variable object that generates realisations.

A problem in physics is the presence of matter in the shape. Does matter import ? I mean differentiation of elements that constitutes the shape ? Has the interior of a rabbit, stomach, lungs etc. to enter to the definition of "rabbit shape". I think for real object there is no absolute accessible definition of shape. Shape is a model that depends highly of the approximation induced by the observation scale. And if we want to be coherent with this idea we have to accept that the skin compactness, the hair color etc. are elements of shape at a certain scale.

And can we speak of shape for a non compact object ? Fractal theory presents objects whose boundaries are not clearly defined or at least highly change with the observation scale. The notion of shape is thus dependent of the nature of the object and the observation scale. The stake is to succeed in describing a shape.

There are simple shapes, that can be easily written and constructed in mathematics such as circles, polygons, conics. And the recursive notion of compound shapes: a shape is a simple shape or a shape to which we add another shape with a particular size and relative position. If shape is translation invariant, the relative position of elements in a compounded shape are essential.

Shapes can be described by some properties. Notions can be topological properties: properties that do are stable with respect of a smooth distortion of space. Or they can be geometrical with metrics and angles.

A functional point of view had been adopted to explain emergence of shapes in nature:

physical object exist and have their particular shape since this shape is somehow optimal to last. Morphogenesis theory [25, 45] shows through numerous cases that shapes of surrounding objects can be explained not by their function but by their causes. In the growth of an object, the physical constraints that excerpt make the shape of the object evolve in a particular way. Shape is then explained by endogenous principles: the formation process of an object determine its shape and the possible variation of the actual object to an ideal shape.

## 2.3 What is a city ?

This section is to define the city, its components and present qualitatively the mechanisms of its growth. It is a mixing of the references [113, 38, 92, 22, 127, 26], discussions with social-scientists, town-planners and architects (I thank Clément-Noël Douady, Jean-Pierre Frey, Philippe Bonin) and my advisers.

### 2.3.1 Origins

To give a definition vague enough to handle the numerous cases the notion have to cop with let us define a city as "the spatial infrastructures in place to organize a community of people". We will be more specific latter.

The very first city then dates back to Neolithic area from -6000 a.c when men became geographically stable and invented agriculture. In the occidental antiquity cities started to look like our current cities with streets and compact organization of houses. Cities were closed spaces surrounded by walls, the streets were straight organized around the *Cardo* (north / south axis) and the *Decumanus* (est / west axis) and their intersection was the Forum, a public meeting place mainly dedicated to politics. The city has then a symbolic meaning: it is the spatial projection of a divine order. From Prehistory to Antiquity cities have been enlarged to gather and exchange in the same secure place necessary goods.

In the middle age the city loses its symbolism, with irregular maps and intricate network of small streets. Three centers are generally distinguished: the market, the castle and the cathedral or church. The city is a spontaneously organized system around these three places of exchange, symbol and power.

Renaissance rehabilitates symbolism and aesthetic, it breaks up with middle age by extending the city, making boulevards of former walls, straightening streets, introducing geometrical shapes, radial and orthogonal layout.

The concept of city had to adapt to the industrialization and the appearance of new transportation means. The willing at this time is coherent with renaissance: the city's pace of life has to be improved, humanized. Transport is more efficient, straight lines are added in extension of the city and in addition of its initial network.

### 2.3.2 Town planning

The idea of town-planning could gather all work and thoughts on the construction and optimization of the city. Nonetheless, the term itself and the work of town-planner has only been institutionalized in the beginning of 20th century. It has a scientific claim but it is above all a practice. Practice that calls for interdisciplinary competencies: architects, social scientists, engineers...It is "the practise that aims at creating through time an ordered layout of space creating harmony, well-being and economy". Town planners are then decision-makers acting rationally at a local or regional scale.

Town-planning intervene to **(1)**: Create a new city, **(2)**: Re-structure an existing city or district **(3)**: Extend a city.

There are two main modern theories of town-planning:

**Progressive**: represented for instance by Cerda, Haussman, Le Corbusier. Progressive theory has a real scientific claim, the city has to be organized rationally, to ensure efficient circulation, well-fare and hygiene. The streets are thus long and large lines opening up view on spectacular monuments. Functions are separated, there is a hierarchy in places, and the main city core dilutes into country side.

**Culturalist**: represented for instance by Marcel Poète. This theory gives priority to the expression of humanism in the city. The city is above all a meeting place. Space has to be varied, intimist, surprising. Present and past coexist. City is directly compared to a living organism.

### 2.3.3 Elements of geometry

Here are a few definitions related to city geometry.

A **town** is a political / historical entity of adjacent infrastructures. Towns are gathered into **cities or agglomerations or urban unities**. A city is a spatially coherent reunion of towns. In graph theory we would say cities are "connected components" or clusters of infrastructures. Cities have a physical meaning contrary to towns whose limits are arbitrary fixed.

A city generally includes a **main city core** and suburbs. Not to be confused with suburbs, **faubourgs** are quarters that first have grown at the boundary of a city center that generally ends up to absorb them.

A **Metropolis**: is an artificial gathering of agglomerations created politically to encourage their exchanges and their joint development.

A **block**: is a small compact gathering of building or un-built zone surrounded by streets. For mathematical convenience we will consider block are the same as cells in graph theory: a small portion of space delimited by streets and that does not contain any loop.

Blocks have to be distinguished from **parcels** that are land-law unities and have owner. They can be re-divided or melt together. The document revising the state of parcels is the **land registry** (a detailed map of parcels with supplementary information).

**The street** is a transport unity. In occidental culture it is defined by its name. Geometrically it is the continuity of coherently aligned pieces of roads. Contrary to parcels, they are

public area.

A city does not pop up from nothing. There is always a geographical reason, a river, a power resource that makes the place appealing. A community gathers next to it to exploit it efficiently. They create infrastructures, these infrastructures are added to the initial wealth of the geography and attract new people. Little by little a city grows by adding new elements, keeping most of the previous ones. During this evolution, buildings in blocks are often destroyed and reconstructed to increase the local density but an element is stable: the empty space of the street.

We want here to define the important notion of **structuring elements**. Qualitatively when building a city, people will try to have an efficient access to local resources and thus to construct streets orthogonal to these resources. Former streets become structuring elements. Recursively, the impact of structuring elements are reinforced through time. For instance in the map of Venice (Fig.2.1) we clearly observe that the intricate structure of the channels act as a shell that shapes actively the growing network.



**Figure 2.1:** The city of Venice. The river structure is an essential element in the Venice's transportation network. The intricate shape of the channels sculpts the shape of the street network that tends to be locally perpendicular to it. Extracted from Google Maps by Andreas Perna.

#### 2.3.4 Functions

**A city is a gathering of people in a fuzzy area, to create goods that at least permits the perenity of the community.**

Thus a city attempts to lodge people, ensure their survival, well-being, circulation and their communication; to optimize the economical production.

### 2.3.5 Diversity

A wide diversity of city systems is observable as well on their overall shape as on their street network structure.

For a town, the boundaries are more or less intricate.

Most of the times cities or agglomerations do not have clear boundaries, which has lead people to define it as "fractaly shaped". We will discuss this description. Let us say there is no clear universal definition to limit cities in space, there is a continuity between city and countryside, the growth of a city being the progressive conquest of the countryside area. Cities tend to sprawl along long pre-existing connecting roads.

One can consider **mono-centric or poly-centric cities**. A monocentric city is the "traditional" vision of a city we have: it is compounded of a main city core highly dense and density decreases in the suburbs. Polycentrism corresponds to the idea of metropolis. Metropolis are artificial entities, an interesting question would be to wonder if from a monocentric system new centres can naturally appear for instance when the center is saturated. Another question is to wonder if polycentrism is a stable descriptor of a system that evolves. Won't centres end up to melt into a single core?

The road network of cities also exhibits a wide diversity both between cities and between different parts of the city: diversity expresses at the inter-city scale but also at the inner-city scale. Network can be regular (square for instance) lattice, radial, tree-like, labyrinthine difficult to describe simply or "organic". General street networks are a mixing of these patterns [86].

### 2.3.6 Construction

Some street patterns seem intuitively "optimal" in regard to a particular geography. The square grid of American cities for instance is optimal in the sense that it reduces the mean distance between two points of the city and allows a regular compact distribution of buildings. The radial scheme is optimal when there is a very central resource: the goal of people living in the city is not to go uniformly from a point to another but to reach this center.

We also picture that more generally an "optimal" city could be an homotopy of a square grid or radial pattern. The homotopy being a smooth distortion of the network to adapt it to the geography.

But this supposes that the city is entirely built in a short span of time. We will speak of **planned cities**. And most of the time it is not the case. Local structures have to evolve and the city has to extend spatially as much as it growths in terms of economy or population. Square grids are not "stable" structures.

If a city is progressively made, we will speak of **organic cities**. In this case an original frame is built around structuring elements and various agents (individual persons, industries...) come to settle in the city, they have to choose a place and to connect to the network. For instance in Oulan-Battor (Fig.2.2) the city is constituted by many persons that progressively arrive and build their houses to be close to other inhabitants. The houses are aligned more or less to leave a path allowing transportation. Another typical case is the city of Xian (Fig.2.3). It was originally a planned square city around the imperial palace. But square



**Figure 2.2:** The city of Ulaan-Battor (Mongolia). The growth of this city consists in the addition of houses. The growth starts from an pre-existing road, the density increases with the proximity to this road. Other roads are not explicitly created: when adding a new house, urban actors make sure that it is aligned with neighbours along main resulting streets and secondary streets also appear. From Google Maps.

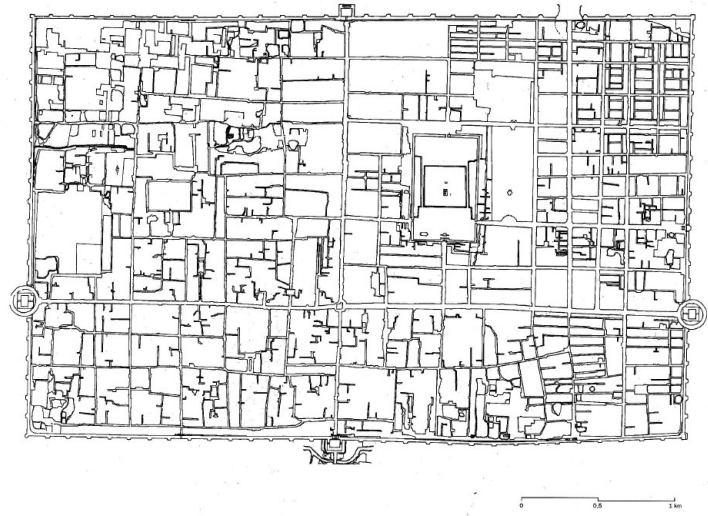
blocks were large enough which allowed an organic evolution of the city. Various populations constructed street segments in the initial squares. As we can see in Fig.2.3 these populations did not have the same wealth or the same politics and a morphological gradient results in the 1948 map.

### 2.3.7 Comparison with other networks

To understand what a city is we have compared it to other planar networks whose shapes look similar: leaf venation network and crack pattern on the surface of a pottery. This comparison is not original, it has already been done in [24] from the iterative division of space perspective and in [103] from the hierarchical representation of transport network perspective.

As the city, the boundaries of the leaf and its inside network evolve jointly. The leaf exhibits a highly hierarchical network, from a long (and large) stalk to a little segment of vein. The network includes loops at a medium scale. But these loops are not observable at general scale (Fig.2.4, Up) and the small scale structure is tree like (Fig.2.4, Down). The vein network in a leaf is above all hierarchical, tree like, the loops intervene at a medium scale to make redundancies in the network and make it more robust. In a street network the notion of loop or cell is essential, it is a structure element which plays a role in the evolution of the system.

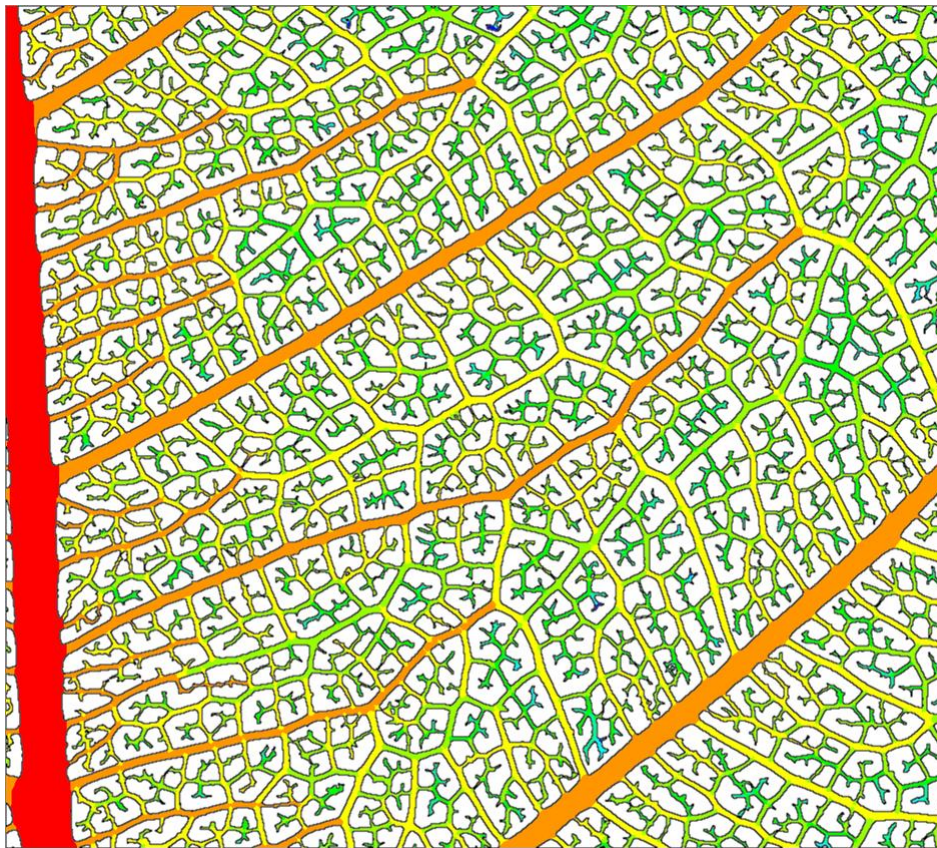
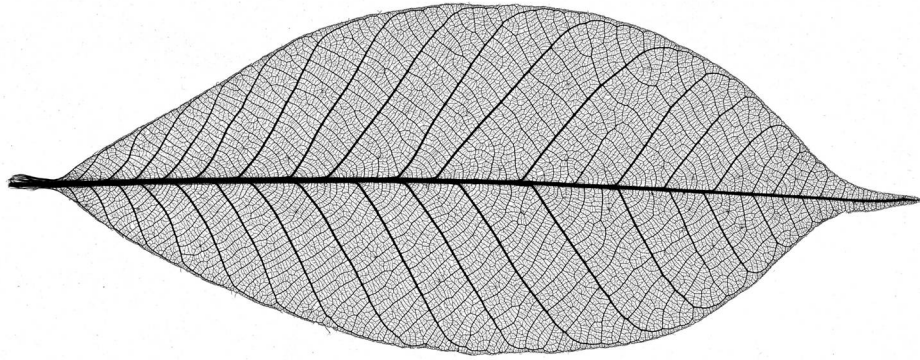
Crack patterns are obtained when a bounded surface of clay for instance dries. The initial structure cracks (Fig.2.5) into a line to reduce global tension. There are two resulting cells.



**Figure 2.3:** The Chinese town of Xi'an in 1949, with various subdivision patterns inside a regular grid. The square grid was created as an original frame around the imperial palace: the city was originally planned. But through time organic patterns have re-divided the initial structure. We can also notice a "morphological gradient" of these patterns: the network is more structured at East and more tree-like at South-West.

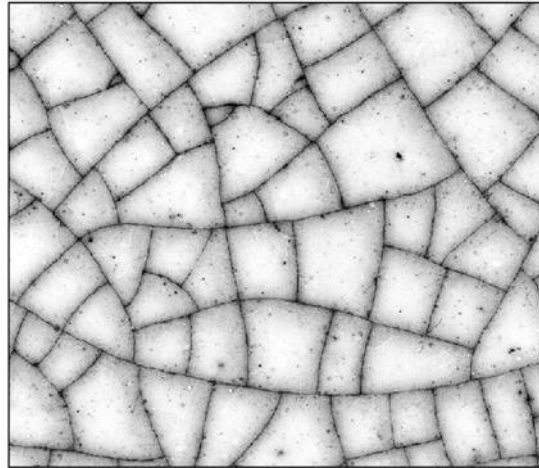
These cells redivide independently. The new division curves are perpendicular to the vase boundaries and to the first line. And so on until cells reach a minimal size (determined by the thickness of the vase). The result looks like a city map but there are structural and dynamical differences.

**(1):** For the vase, boundaries are still, in a city they move. **(2):** The divisions of two adjacent cells in the vase are independent, in a city, blocks may be cut coherently to create long resulting axes. **(3):** The perpendicularity constraint is stronger for crack patterns: a new separator has to be perpendicular to the two older separators it meets. Consequently the new separator cannot be a line, it is a curve. In a city the constraint seems to be looser: the new separator have to be a line and thus can be perpendicular only to one of the former streets it meets.



**Figure 2.4:** Observation of a leaf's venation network at two scales. From Andreas Perna.





**Figure 2.5:** Picture of a Crack pattern at the surface of a terra-cotta vase. From Steffen Bohn.

## Approach of this thesis

### 3.1 Quantitative principles shaping the city

The two previous chapters of this introduction have put in light important principles arising in urban context:

1. Classically monocentric and polycentric cities are distinguished. In fact a system of cities responds to a hierarchical principle illustrated by Zipf's law: the distribution of cities' size follows a power law. Small cities orbit around big cities and again smaller cities orbit around the small cities... At the scale of a city, we find more or less important district or streets also hierarchically distributed. Monocentric and polycentric cities are ideal cases, first or second order spatial approximation of this hierarchy principle. Both cases can be considered under the same idea of cities with distributed resources.
2. The city may be planned that is to say constructed of one piece or organic that is to say constructed little by little by independent agents. These two cases correspond to two different optimization procedures and produce two very different resulting patterns. In the planned case, the global structure is optimal to adapt to the local geography. In the organic case, agents or settlers add sequentially infrastructures to the current city's layout. Each infrastructure addition is locally optimal at a given time for the settler that makes it.
3. The city organizes into coherent geometrical structures called streets. These streets are subjected to the hierarchy principle: there are long and not so long streets with a more or less central importance. New streets are generally perpendicular to previous streets or geographical elements of importance (such as rivers) we call generically structuring elements. This rule is determinant in the construction of the network: it induces anisotropy, clusters or gradient of density along main axis. Main axis may be pre-existing or are the hallmark of the extension of the city into the countryside. Another geometrical constraint arises when considering the duality between blocks and streets. Former large blocks are divided into smaller blocks with the constraint to make aligned streets.
4. The "primal graph" of a city map exhibits quite trivial behaviours common to most planar networks. Nonetheless one of the most evident particularity of city maps is their

organization in long axes or streets. The "dual graph" constituted of these streets as vertices and their intersections as edges may display more noticeable features.

## 3.2 Problems

**In this thesis we will question the phenomenology of the city from its network structure. We want to show how the street network structure jointly constraints the further development of a city and the use of space by its inhabitants.**

We will not use complementary information such as population, street width, building height... We will consider the skeleton of a map, see it as a planar graph and go into its description, analysis and study the repercussion it has on our use of space (traffic, engineering...).

## 3.3 Summary

The principal scientific notions our work will rely on are Euclidian Geometry, Graph and Complex Network Theory, Measure Theory and Stochastic Geometry. They are presented in three independent appendices: App.A, App.B, App.C.

In Ch.4 we provide mathematical structures to represent city maps. We want to consider the geometry of a map without limiting to its topology as done traditionally in the complex network framework. A city is represented by a geometrical graph approximated (with an arbitrary accuracy) by a straight graph  $G$ . The straight graph is a geometrical object whose topology is sufficient to code the geometry which is pleasant for computational concerns. To this graph is associated its geometrical projection that allows to define functions and integrate them all along the streets segments of the city. For instance a function  $f$  can be defined on each point of the city (not only vertices), it could represent the population on each point and it will be possible to "integrate" this population all along the street system with the uniform measure  $\mu_G$ :

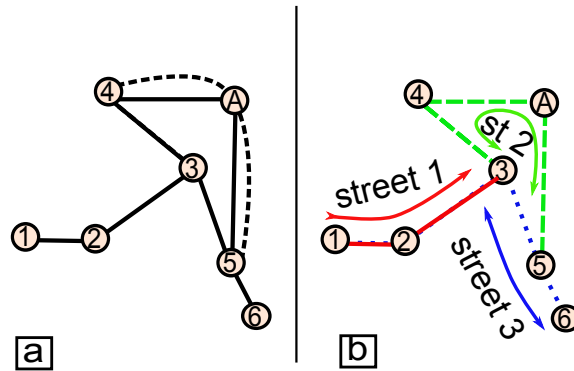
$$\text{Total population} = \int_{\text{all along the street system}} f \, d(\text{Uniform measure on the city}) \quad (3.1)$$

$$= \int_G f(g) d\mu_G(g) \quad (3.2)$$

To the city graph is superimposed an hypergraph structure that provides a multi-scale representation of the graph by recovering the notion of street from street segments. A city graph is then a graph  $G$  with its vertices  $V$  and its edges  $E$ . To this graph is associated its "projection" on  $\mathbb{R}^2$ :  $\pi_G$  and a uniform measure  $\mu_G$ . We automatically provide  $G$  with an additional hypergraph structure  $H$

$$G = ((V, E), H, \mu_G(\cdot)) \quad (3.3)$$

This mathematical framework comes with its computational implementation. We address the choice of optimal structures to code city graphs and present classical and new algorithms



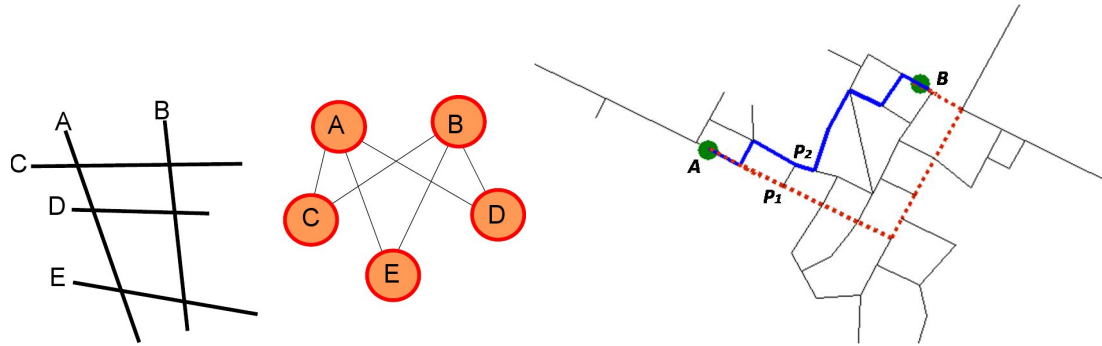
**Figure 3.1:** On the left a straight graph  $G$  that is the approximation of a geometrical graph: the vertex  $A$  is an additional vertex to sample the curvature of the dotted edge. On the right, we have provided this straight graph with an hypergraph structure  $H$  constituted of three hyperedges called streets that are disjoint sets of edges of  $G$ .

to efficiently solve classical problems on city maps as for instance the shortest path problem in weighted or unweighted graphs (classical) or the street hypergraph recovering with two original algorithms and their implementation in linear time. The new Cab Driver's algorithm for fast approximation of the shortest path problem in cities presented in Ch.6 is an illustration of the efficiency of hypergraphs in urban modelling.

In Ch.5 we will observe and test two phenomena occurring in cities' hypergraph. The length of streets follows a mixture of log-normal laws: there are long and small streets in a city, the resulting distribution is heavy tailed but admits an expectancy. The street hypergraph can be represented as a graph. The shortest path distance in this graph corresponds to the number of times one has to turn to go from a street to an other. And this graph exhibits a Small-World property. We propose a clear definition of this property but the main idea is that the number of times one has to turn to go from the street of a random point to the street of an other random point scales logarithmically with a measure of the size of the city (we take the total length of streets as a reference measure but they are all linked by power relations according to [19]).

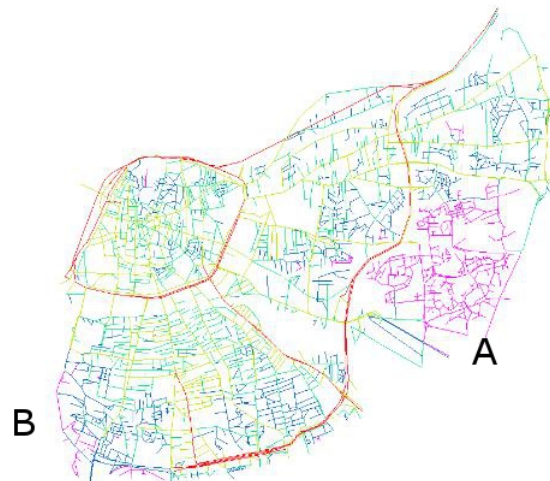
$$\mathbb{E}(\text{distance between two random points}) = O(\log(\text{Size of the city})) \quad (3.4)$$

In Ch.6 we present a sequence of measures to characterize and compare cities: the size, the topology, the efficiency, the anisotropy, the compactness. These measures could be used to automatically classify cities for instance to cluster them according to the typical cities or "figures de style" presented in [32], see Fig.1.1. But the inner-city diversity is generally too broad and we will see in Ch.10 these measures are interesting to cluster a particular city in morphologically homogeneous zones. Ch.6 ends with the redefinition of centralities in our mathematical framework. From the definition of simplest distance (number of turns to go from one point to an other), we define the simplest centrality and compare it to the closeness and the straightness centrality. This new centrality is faster to compute and propose a hierarchical interpretation of the city. We will justify the relevance of this centrality by using the phenomena observed in Ch.5. The analysis of simplest centrality naturally leads to a brand new algorithm to solve the shortest path problem: the Cab Driver's Algorithm.



**Figure 3.2:** From left to right: a city is compounded of 5 streets forming an hypergraph. To this hypergraph is associated a graph whose vertices are streets and edges are drawn between two streets if they intersect in the map. The shortest path distance between two streets - vertices in the "street hypergraph graph" called simply hypergraph corresponds to the simplest distance between two points in the streets. The last picture shows between points A and B the shortest path in blue and one of the simplest path in red.

In Ch.7 we use the simplest centrality to analyse the map of two French towns: Avignon and Villers-Sur-Mer. Avignon is studied through three scales: the historical center, the city center and the extended town. Simplest centrality allows to recover main roads in the city at its various scales of functioning and to diagnose ill-deserved area. With Villers-Sur-Mer, we test the interest of simplest centrality to assess town-planning scenarios to homogenize the transportation in the city.



**Figure 3.3:** Visualization of the simplest centrality of Avignon. It enhances main boulevards and radial axis (in red) and allows to diagnosis ill-deserved zones in pink.

In Ch.8 we propose several models to explain the growth of the city. The first model is continuous: the city is a scalar field of wealth. An integro-differential equation is proposed for the spatio-temporal evolution of the city. The main idea is that each infinitesimal area of the city creates at time  $t$  additional wealth proportional to the wealth of the area. This

new wealth is then dispatched through the city:

$$\frac{\partial f(\vec{x}, t)}{\partial t} = \lambda \cdot \text{Cste} \int_{\mathbb{R}^2} f(\vec{y}, t) K(\|\vec{x} - \vec{y}\|) d\vec{y} \quad (3.5)$$

with  $\text{Cste} \int \int K(\|\vec{x} - \vec{y}\|) d\vec{x} d\vec{y} = 1$  and  $\lambda \geq 0$ . This equation can be applied to a system of several cities and is coherent with the Zipf's law. We study the equation numerically for a monocentric city and show in this case the wealth profile is an inverse power law, the global growth of the city is exponential, the local growth is a power-law followed by an exponential and the border of the city drifts away exponentially.

The next models are network based. We define a simple division of space process and show it produces a Small-World network. To reproduce also the distribution of cells'area observed in [79] we propose a model of division / extension of space.

Finally, we introduce models for the "Morphogenesis of the organic City". It implements more realistically the division / extension of space principle. The idea is that the current layout of a city induces a potential field  $P$  quantifying the interest of a place  $x$  for new settlers. The potential is a sum of elementary potentials  $V$  induced in  $x$  by the collection of objects (streets, rivers...) forming the city.

$$P_{G \rightarrow x} = \sum_{h \text{ object in } G} V_{h \rightarrow x} \quad (3.6)$$

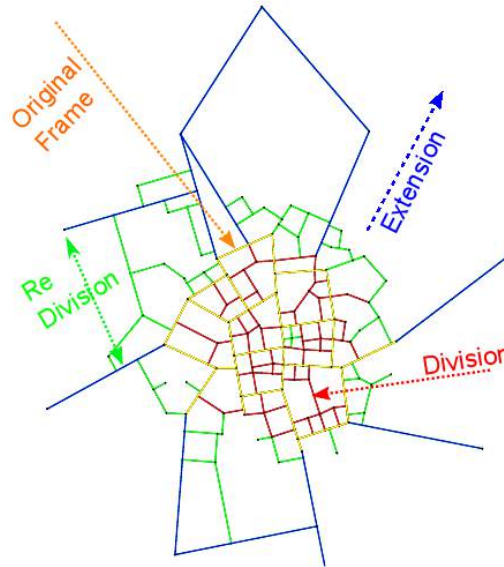
Sequentially new settlers arrive in the city, chose a place and link this place to the existing infrastructure. This model reproduces the log-normal scaling of streets and the Small-World behaviour of the street hypergraph. A few parameters to describe the politic of inhabitants of the city allow to produce a wide range of city patterns.

Last two chapters (Ch.9 and Ch.10) place in the continuity of [57] in the stochastic geometry framework. The purpose is to solve large optimization problems from telecommunication engineering by statistical equivalence. To an actual map is substituted a stationary random tessellation. We choose low scale models (Poisson Line Tessellations and Crack Tessellations with or without anisotropy distribution) and an high scale model (Poisson Gabriel Graph) that are relevant in a urban context. We propose an algorithm to simulate the typical vertex of a Poisson Gabriel Graph and use it to assess the mean topology of this random graph. To low scale models we add a new segment process to take dead-ends into account. We propose a robust algorithm to identify a map with a good equivalent random graph. The problem is a city is generally not stationary. We solve it with a step of morphological segmentation of the map. To this we introduce a new spectral clustering ([7]) based algorithm. The whole procedure of segmentation - identification can be performed in a linear time.

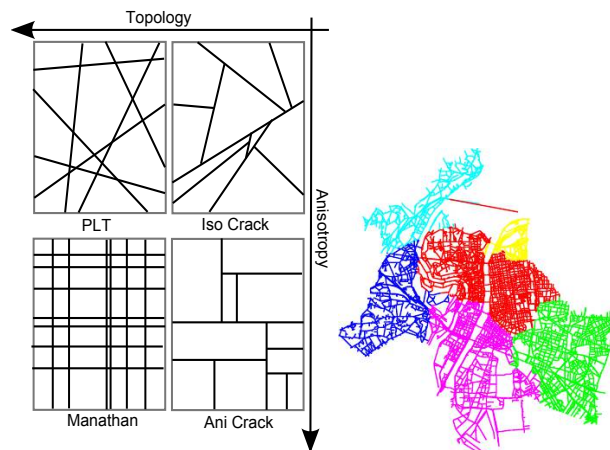
### 3.4 Style of the thesis

Our approach seeks to be interdisciplinary, functional and synthetic.

**Interdisciplinary:** We place in a physical science approach to work on a subject that belongs to social science. To complete our project we will have to handle relatively new mathematical notions and implement them into computer code since our data are numeric. We will try to use a vocabulary and give interpretations of our results into terms that are



**Figure 3.4:** Sketch of the division - extension of space idea. From an original frame (in yellow) the city produces division patterns when it locally densifies (in red). Sometimes the city produces a few lines of extension at its outskirts (in blue). And then the global result is re-divided (in green). Various implementations of this principles will allow use to recover main observed phenomena in the city: small-world in the street hypergraph, log-normal scaling of street lengths, power-law distribution of blocks' area.



**Figure 3.5:** On the right, the result of a morphological segmentation of the city of Lyon. To each coloured zone is associated one of the stationary random tessellation displayed on the left, according to the measured topology and anisotropy of the zone. Optimization problems can be analytically solved for each model and thus substituting a real planar graph to one of this tessellation allows a fast solution of optimization problems by statistical equivalence rather than using heavy back-tracking methods.

relevant for physicists, mathematicians, engineers, computer scientists and town-planners.

**Functional:** The results of this work are implemented into automatic procedures that can

be used in town planning, traffic forecast or telecommunication engineering. The algorithms are optimized.

**Synthetic:** our presentation of the subject will not be wordy. We present our hypothesis, compare them to reality and try to explain them by mathematical models as minimal as possible. In particular the growth models we present in Ch.8 do not take into account all the ingredients that would be necessary to mimic the complexity of a human community. We only want to exhibit a minimal set of principles that permits to explain some phenomena occurring on cities and reproduce quite a diversity of shapes. This framework is made to be flexible. In a second step interested persons can easily make it evolve in a more elaborated structure to observe more subtle behaviours our predict trends.

### 3.5 Main notations

Generally a graph will be called  $G$ , its vertices  $V$ , its edges  $E$  and its cells  $C$ . An hypergraph will be  $H$  and we will use also  $H$  for the graph associated to the hypergraph. To designate vectors we use the classical  $\vec{x}$ ,  $x_k$  for its components,  ${}^t\vec{x}$  for its transpose and we will designate matrix with bold letters. So if  $G$  is a graph, its adjacency matrix is  $\mathbf{G}$ . Its shortest path matrix will be  $\mathbf{G}^*$ . If a matrix stores information related to a collection of object  $(v_i)$  the components of the matrix  $\mathbf{G}$  can be written  $\mathbf{G}_{ij}$  or  $\mathbf{G}(v_i, v_j)$ . The cardinal of a set  $S$  is written  $\#S$  or shorten into  $S$  for instance when studying computational complexity.

We use classical geometrical notations:  $\langle \cdot, \cdot \rangle$  is a scalar product, when this product depends on matrix  $\mathbf{A}$  we write  $\langle \cdot | \mathbf{A} | \cdot \rangle$ . The  $n$ -norm is  $\|\cdot\|_n$  ( $\|{}^t(x_1, \dots, x_m)\|_n^n = \sum |x_i|^n$ ) and for the Euclidian norm ( $n = 2$ ) we simplify with  $\|\cdot\|$ . The same notation  $\|\cdot\|$  is sometimes also used to designate the length of an orientable path, for instance if  $h$  is a (curved) street,  $\|h\|$  is its length. The shortest path distance on a graph is written  $d_G^{sp}(\cdot, \cdot) = d^{sp} = \mathbf{G}(\cdot, \cdot)$ . When we compare shortest and Euclidian distance we will prefer to use  $d^e$ . The simplest distance on a geometrical graph equipped with an hypergraph is  $d^{sim}(\cdot, \cdot)$ . To compute this distance we have to compute the shortest path distance on the hypergraph  $H$  written coherently  $\mathbf{H}^*$ . By abuse of notation  $\mathbf{H}^*(x, y)$  can be the simplest path distance between two points of the original graph.

If  $x$  is a point in  $(x_i)$  a collection of points then  $\text{Vor}(x|(x_i))$  is the Voronoï (see App.A) cell of  $x$  relatively to  $(x_i)$  for the Euclidian distance.

$\mu_d$  is the Borelian measure on  $\mathbb{R}^d$ . We use erf to designate the error function, log for the natural logarithm,  $\mathcal{N}(m, \sigma^2)$  both to designate a normal law of mean  $m$  and variance  $\sigma^2$  and its density. Random variable are in capital letters and  $X \sim f$  means the variable  $X$  follows the law  $f$ . Probability, expectancy and variance are respectively  $\mathbb{P}$ ,  $\mathbb{E}$ ,  $\mathbb{V}$ .

We use the notation of Landau for functions especially to express complexities:  $f = O(g)$  means  $f/g \rightarrow \text{Constant} \neq 0$ .





## Part II

# Mathematical and Physical Modeling of the City



# Mathematical and Computational Representation of Cities

## Contents

---

4.1	Data . . . . .	<b>65</b>
4.1.1	Geographical Information Systems . . . . .	65
4.1.2	MIF files . . . . .	65
4.1.3	Import routine . . . . .	65
4.1.4	Data base . . . . .	66
4.2	City Graphs . . . . .	<b>67</b>
4.2.1	Graphs and planar graphs . . . . .	67
4.2.2	Geometrical and straight graphs . . . . .	69
4.2.3	Hypergraphs . . . . .	70
4.2.4	Measure . . . . .	71
4.2.5	Terminology and notations . . . . .	71
4.3	Algorithms to Recover Street Hypergraph structures . . . . .	<b>72</b>
4.3.1	Angular tolerance (AT) . . . . .	72
4.3.2	The minimal reciprocal alignment (MRA) . . . . .	73
4.4	Tuning and Performances . . . . .	<b>73</b>
4.4.1	Criteria . . . . .	73
4.4.2	Analysis . . . . .	74
4.5	Computational implementation . . . . .	<b>77</b>
4.5.1	Structures . . . . .	77
4.5.2	Algorithms . . . . .	78
4.6	Conclusion . . . . .	<b>82</b>

---

**—Mathematical and Computational Representation of Cities: Synthesis —**

Studied under the Complex Network approach, cities' maps have been naturally considered as **graphs** whose edges are portion of streets and vertices their intersections. This purely topological representation loses all the geometric information of the map. For instance the city's geometry particularity is to gather its segments into coherent sets called **streets**. We define the notion of **geometric graph**, it is always possible to sample as a **straight graph** whose topological structure is sufficient to recover geometry. These objects are naturally equipped with an **integral operator** allowing measuring function all along their geometry. An additional **hypergraph** structure reports the multi-scale street structure of the system. We discuss two methods to recover automatically the notion of street from the map's geometry without information such as the streets' name.

**Computational representation** of city graphs is discussed as a corpus of **optimized algorithms** to solve recurrent problems.

**—Contributions of this chapter —**

1. A city graph is seen as a continuous object with a Borelian measure.
2. The notion of "dual graph" is replaced by the structure of hypergraph
3. Study of two algorithms to recover street hypergraph structure. After confrontation with real data, the best algorithm to manage this task appears to be parameter free. It approximately minimizes the global torsion at street intersections.
4. Corpus of optimized classical algorithms on city graphs.

## 4.1 Data

### 4.1.1 Geographical Information Systems

We have evoked GIS in 1.3.2. In practice, Vector data are collections of lines, polylines (streets), polygons or pierced polygons (regions) to which are associated alphanumerical attributes such as street name, street type, length, width, population... Data are georeferenced into projection coordinates. In France the classical coordinates in use are the various variations of Lambert system. Commonly used SIG are MapInfo, ArcView or also the open source Open Jump.

### 4.1.2 MIF files

We will not use a full SIG file(".shp") but only the map of streets it contains. From a .shp it is possible to extract two files with an open format. The ".MIF" file is an extraction of the geometrical information of a map. It is a standard "MapInfo" Interchange file, generally associated with a ".MID" file that contains the additional textual information. A street map is encoded into a list of polylines ("Plines"), each representing a portion of road.

A typical extraction of the structure of a .MIF file is:

```
Data
Pline 2
433634.57305836806 2381798.4086203203
433631.8246595768 2381740.657403816
Pline 2
433859.8172549281 2384204.937368185
433998.1683535401 2384085.947034154
Pline 4
433773.95605450065 2381542.5069964
433794.1700535319 2381538.780212023
433809.68250936584 2381535.9222575817
433818.186687505 2381533.576159634
```

The data is compounded of three polylines, each polyline is characterized by its number of vertices and a list of line, each line containing the  $(x, y)$  coordinates of a vertex separated by a blank.

### 4.1.3 Import routine

From a ".MIF" file we want to create a structure "GRAPH" that contains a field "Vertices" which is a  $v \times 2$  numerical array gathering the coordinates of all street intersections and a field "Edge" which is a  $e \times 2$  array, the line  $i$  contains the references of the two extremities of the  $i$ -th edge in the array "Vertices". This data structure to represent graphs will be discussed in 4.5.2. To limit the size of the resulting structure, each extremity of poly-lines are conserved but intermediate points are transformed into vertices in a proportion  $s$  defined

by the operator.

When we will recover the notion of street from the plain collection of street segments of the map it will be very important to be sure we have conserved actual angles between street segments at every intersection. To that purpose, we in fact always keep the extremities of a polyline but also their direct neighbour, the other intermediary points are sampled.

Data in "MIF" file are noisy: there is redundancy in edges and some of them are detached to the global structure, we have to erase double and detached segments. To this we first import the whole graph and then compute its connected components (with the algorithm described in 4.5.2) and only keep the largest one.

---

**Algorithm 1** Map Import

---

- 1: Open the file "City.MIF"
  - 2: Scan it
  - 3: For each Pline, transform the extremities and their direct neighbour as vertices and do the same for a proportion  $s$  of intermediate points
  - 4: Compute the connected components of the resulting graph
  - 5: Keep the largest connected component and erase the others
  - 6: Erase edges that appear several times
  - 7: Translate the map such as the point  $(0, 0)$  is the center of gravity of the map
- 

#### 4.1.4 Data base

We have constituted from the Environmental Systems Research Institute's maps a data base of 109 French cities. We selected major towns except from Paris (that was too large at the beginning of this work to allow reasonably short calculus) plus a few smaller ones from Parisian suburbs, country and seaside. To each city is associated a complete version and an underscanned one. Typically the resulting graphs contain 5000 to 10000 vertices.

Aix-en-Provence	Beauvais	Chaumont
Ajaccio	Belfort	Cherbourg-Octeville
Albi	Besançon	Clermont-Ferrand
Alençon	Béziers	Colmar
Amiens	Blois	Courbevoie
Angers	Bordeaux	Dax
Angoulême	Boulogne-Billancourt	Digne-les-Bains
Annecy	Bourges	Dijon
Arras	Brest	Douai
Auch	Caen	Epinal
Aurillac	Carcassonne	Evreux
Auxerre	Châlons-en-Champagne	Foix
Avignon	Chambéry	Gap
Bar-le-Duc	Charleville-Mézières	Grenoble
Bastia	Chartres	Guéret
Bayonne	Châteauroux	Issy-les-Moulineaux

La Rochelle	Mulhouse	Rouen
La Roche sur Yon	Nancy	Royan
Laon	Nanterre	Saint-Brieuc
Laval	Nantes	Saint-Etienne
Le Creusot	Narbonne	Saint-Germain-en-Laye
Le Havre	Neuilly-sur-Seine	Saint-Lô
Le Mans	Nevers	Saumur
Le Vésinet	Nice	Strasbourg
Levallois-Perret	Nîmes	Tarbes
Lille	Niort	Toulon
Limoges	Orange	Toulouse
Lons-le-Saunier	Orléans	Tours
Lorient	Pau	Troyes
Lyon	Périgueux	Tulle
Mâcon	Perpignan	Valence
Marseille	Poitiers	Vannes
Mende	Privas	Vesoul
Metz	Quimper	Vichy
Montauban	Rennes	Villers-sur-Mer
Mont-de-Marsan	Rochefort	
Moulins	Rodez	

## 4.2 City Graphs

Classically a city map is represented by a graph. This is a purely topological representation. In this section we present the notion of graph and adapt it to be relevant for cities. If graphs are interesting since they are easily implemented in a computer and since there is a large corpus of efficient algorithms related to these objects, they do not a priori permit to handle the geometry of a map. For instance we have seen above that the processing of a .MIF file produces "artificial" vertices (of degree two) to sample the curvature of some streets. These vertices have no physical meaning and should not be counted as other vertices. Or we can consider that an edge is a continuous set of vertices of degree two, which is both consistent topologically and geometrically. We will show here that in a general way maps are geometrical graphs: to each edge is associated a particular curve in the plane. But geometrical graphs are not practical computationally. We will show it is always possible to sample a geometrical graph to a "close" straight graph for which the geometrical information is encoded in the topology. To a geometrical or straight graph is canonically associated an abstract measure which allow to integrate information "all along the geometry of the city", replacing classical topological and biased by sampling vertices  $\sum_{\text{vertices}}$  by  $\int_{\text{all points in the city}}$

### 4.2.1 Graphs and planar graphs

A graph  $G$  is a couple  $(V, E)$  where  $V$  is a finite set  $\{v_1, \dots, v_n\}$  called vertices and  $E$  a part of  $V \times V$  called edges. If  $E$  is symmetric  $((a, b) \in E \Rightarrow (b, a) \in E)$  the graph is said to be



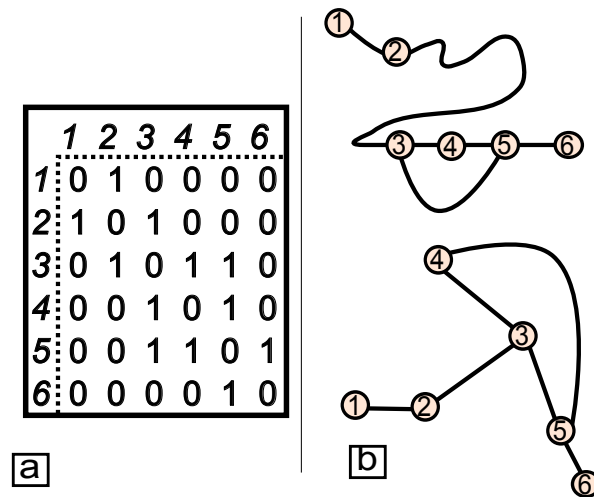
undirected. For an undirected graph, the degree function  $d_G$  is defined on  $V$  and counts the number of edges that pass through a vertex:  $d_G(v) = \#\{x \in V, (v, x) \in E\}$ .

To a graph one can associate its adjacency matrix (Fig.4.1 Left)  $\mathbf{G}$  of size  $\#V \times \#V$  whose entry  $\mathbf{G}(v_i, v_j)$  is 1 if  $v_i$  and  $v_j$  are adjacent i.e. if there is an edge between  $v_i$  and  $v_j$ , 0 otherwise. We write  $v_i \sim v_j$  when  $v_i$  and  $v_j$  are adjacent. Similarly if two edges  $e_1$  and  $e_2$  share a vertex we write  $e_1 \sim e_2$ .

To an edge  $e$  linking  $v_1$  and  $v_2$  can be associated a weight  $w(e) = w(v_1, v_2)$  for instance to model a variable resistance in the branches of an electrical system. When no weight is associated to the edges of the graph, this one is said unweighted.

A path is a sequence of vertices  $(v_1, \dots, v_m)$  such as  $v_i \sim v_{i+1}$ . If  $v_1 = v_m$  the path is called a loop or circuit.

The length of the path  $(v_1, \dots, v_m)$  is  $m - 1$  if the graph is unweighted,  $\sum w(v_i, v_{i+1})$  if it is weighted. Among all possible path between  $v_1$  and  $v_m$ , there exist at least one whose length is minimal. It is call the (a) shortest path and its (their) length is the shortest path distance between  $v_1$  and  $v_m$ .



**Figure 4.1:** The representation of a graph by its symmetric adjacency matrix (a). This graph is planar: it admits at least two geometrical graphs as drawings (b).

The graph is said to be planar (Fig.4.1 Right) if it is possible to draw  $G$  on the plane, i.e. to materialize,  $V$  by a set of points in  $\mathbb{R}^2$  and  $E$  by a set of curves that do not intersect outside of  $V$ . One can show it is possible to draw a planar graph by imposing to the edges to be straight segments.

In a drawing appears the notion of cell. A cell is a geometrical object. It is a region of the plane delimited by a minimal loop i.e. a loop whose associated region does not contain any other loop. The set of cells is written  $C$  and the Euler's relationship holds: If  $G = (V, E)$  is a planar graph and  $C$  the set of its cells then

$$\#V - \#E + \#C = 1 \tag{4.1}$$

### 4.2.2 Geometrical and straight graphs

Among all possible planar representations of a graph, we want to define a notion of graph that can contain the real information of a city map.

A geometrical graph can be seen as a particular drawing of a planar graph.

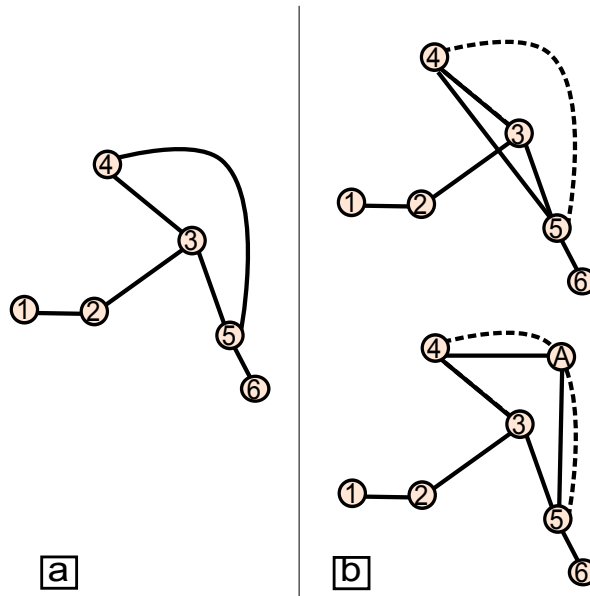
Let the available space  $\mathbb{A}$  be a connected and compact subset of  $\mathbb{R}^2$ ,  $V$  a finite subset of  $\mathbb{A}$  and  $E$  a set of almost everywhere derivable paths included in  $\mathbb{A}$  from one element of  $V$  to another that do not intersect outside of  $V$ . Then  $G = (V, E)$  is an element of the space of geometrical graphs  $\mathcal{G}_g(\mathbb{A})$ . If  $E$  is restricted to straight segments ( $G \in \mathcal{G}_s(\mathbb{A})$ ),  $G$  is a straight graph. To a geometrical graph  $G$ , one associates the "geometrical projection of  $G$  in the plane"  $\pi_G$ , the subset of  $\mathbb{A}$  defined by :

$$\pi_G = \{x \in \mathbb{A}, \exists e \in E, x \in e\} \quad (4.2)$$

$\pi_G$  is compact so we can provide  $\mathcal{G}_g(\mathbb{A})$  with an Hausdorff distance :

$$d_H(g_1 || g_2) = \max_{x \in \pi_{g_1}} \min_{y \in \pi_{g_2}} \|x - y\| \quad (4.3)$$

A drawing  $G' = (V', E')$  is a rectification (Fig.4.2 Right-Bottom) of the geometrical graph  $G = (V, E)$  if  $V \subset V' \subset \pi_G$  and if each element of  $E'$  is a segment.



**Figure 4.2:** A geometrical graph (a) and two of its rectifications (b). The upper one is not straight because two edges intersect outside of the vertices set.

$G'$  is not necessarily a planar straight graph since edges can possibly intersect outside of vertices (Fig.4.2 Right-Up). The idea is that one should be able to add to a geometrical graph as many vertices of degree two as he wishes and still consider the same mathematical object.

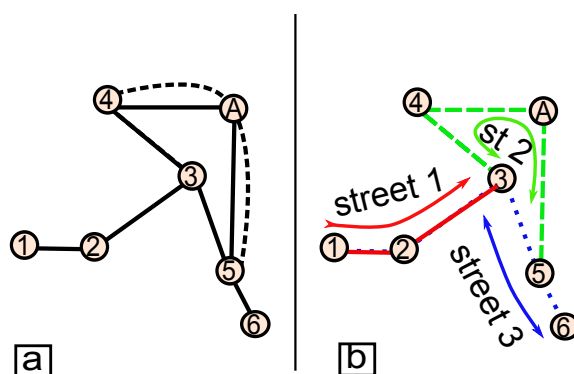
**Every geometrical graph admits a planar rectification**  
**Every geometrical graph is the limit of a sequence of straight graphs**

A geometrical graph is always tacitly weighted: the weight of an edge  $e$  between  $v_1$  and  $v_2$  is the geometrical length of  $e$ :  $\|e\|$  which equals  $\|v_1 - v_2\|$  when the graph is straight.

### 4.2.3 Hypergraphs

A hypergraph is a couple  $V$  of vertices (as for graphs) and  $E$  of hyperedges. An hyperedge is a subset (if the hypergraph is undirected, a  $n$ -uple of vertices otherwise) of  $V$ . In short, a hypergraph is a graph whose edges can contain more than two vertices. We will use hypergraphs to represent multi-scale structures in a city map i.e. scales that are larger than the segments between two vertices.

Let  $G = (V, E)$  a graph, if  $R$  is an equivalence relationship on  $E$  then  $\hat{G} = (V, E/R)$  is an hypergraph. For instance we will seek out to reconstruct the notion of street by relations of the kind "these edges belong to the same street" whose result constitutes the **street hypergraph**.



**Figure 4.3:** A straight graph (a) and its hypergraph structure (b) deduced from  $R_{\pi/20}$ . Viewed as a city's map, this graph contains 7 streets segments but 3 streets.

In Sec.4.3 we will try to recover a street hypergraph structure from a graph. The relation "be in the same street" is hard to decide for two street segments far from each other. To solve that problem, the idea is to define a first local relationship  $\hat{R}$ : "these edges share a vertex and are aligned" and propagate it to a global relationship  $R$ : "these edges are in the same street". We will use the following property: If  $\hat{R}$  is a reflexive relationship on  $E^2$  then the relationship  $R$  on  $E$  defined by:

$$e_1 R e_2 \text{ iif } \exists \alpha_1 = e_1, \alpha_2, \dots, \alpha_n = e_2 \in E \mid \alpha_1 \hat{R} \alpha_2, \alpha_2 \hat{R} \alpha_3, \dots, \alpha_{n-1} \hat{R} \alpha_n \quad (4.4)$$

is an equivalence relationship (transitive closure). This permits to define equivalence classes from local relationships ((Fig.4.3 Right).

Another simple example of hypergraph is the **trunk hypergraph** defined by the local relationship "these edges meet at a vertex of degree 2". The resulting hyperedges or trunks are pieces of the street network contained between two "real" vertices i.e vertices that do not come from sampling.

#### 4.2.4 Measure

To represent a geometrical graph as a continuum, we want to define the notion of infinitesimal piece of street. As the Lebesgue's measure  $\mu(\cdot)$  on  $\mathbb{R}$ , we want to create a measure  $\mu_G(\cdot)$  on a graph  $G$ .

As a compact part of  $\mathbb{A}$ ,  $\pi_G$  is a polish space (complete and separable) on which one can define a Borelian measure  $\mu_G$  (see App.C or [78, 89] for more precisions on measure theory). For instance  $\mu_G(G)$  is the total length of edges in  $G$  or if  $A$  is a subgraph or an extraction of  $G$  (the intersection between  $\pi_G$  and a compact set  $W$ ),  $\mu_G(A)$  is the total length of edges in  $A$ . That measure permits to integrate functions along the edges of a geometrical graph. If  $h(g)$  is the height of buildings at the position  $g$  in the graph  $G$  that represents a city then

$$\int_G h(g) d\mu_G(g) \quad (4.5)$$

is the total height of buildings in the graph.

If  $f$  is a measurable function on  $G$ , we seek out to estimate  $\int_G f(g) d\mu_G(g)$ . An easy to implement algorithm is a Monte-Carlo method. The measure  $P_G(\cdot) = \mu_G(\cdot)/\mu_G(G)$  is a "uniform" probability measure on  $G$  with its Borelian  $\sigma$ -algebra. A random variable  $X$  following  $P_G$  provides a random point on the graph  $G$ . In addition,

$$\mathbb{E}(f(X)) = \frac{1}{\mu_G(G)} \int_G f(g) d\mu_G(g) \quad (4.6)$$

and if  $X_1, \dots, X_n$  are  $n$  independent random variables that follow  $P_G$ ,

$$\frac{\mu_G(G)}{n} \sum f(X_i) \rightarrow \int_G f(g) d\mu_G \quad (4.7)$$

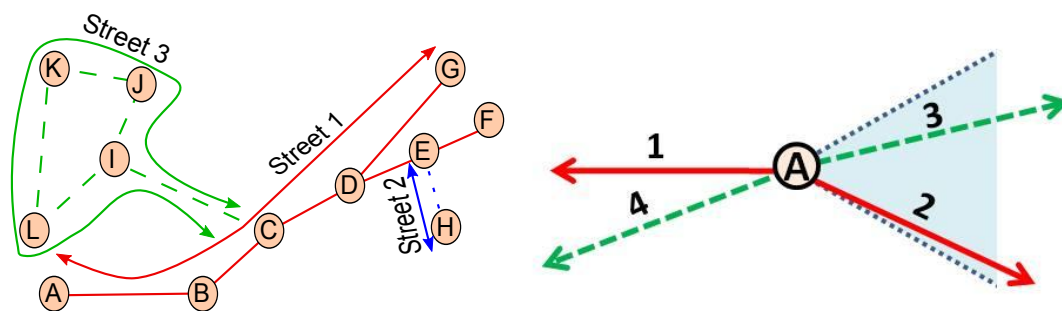
which allows to approximate any integral on  $G$ . If  $H$  is an hypergraph then  $h \in H$  can be seen as a subgraph of  $G$  and we have trivially:

$$\int_G f(g) d\mu_G(g) = \sum_{h \in H} \int_h f(g) d\mu_G(g) \quad (4.8)$$

#### 4.2.5 Terminology and notations

A city graph  $G$  is a straight graph that represents the map of a city or a city that could exist. It writes  $G = ((V, E), T, H, \mu_G)$  where  $V$  is the set of vertices,  $E$  the set of edges,  $T$  the hypergraph of trunks and  $H$  an additional hypergraph structure that represents streets and will be defined later,  $\mu_G$  is the Borelian measure associated to  $\pi_G$  that will be identified with  $G$ . Elements in  $E$  are called **street segments**, in  $H$  **streets** and in  $T$  **trunks**.

We partition  $V$  into  $V = V_1 \cup V_2 \cup V_+$  where  $V_1$  are vertices of degree 1 called **dead ends**,  $V_2$  vertices of degree 2 called **junctions** that will be seen as sample artefacts and  $V_+$  vertices of higher degree called **intersections** that make sense in the topology of the city. Two vertices are adjacent if they are the end points of a same edge. Two edges, trunks or streets are adjacent if they share a same vertex. In all cases for two **adjacent elements**  $a$  and  $b$  of



**Figure 4.4:** LEFT: Street 1 is branched (vertex  $D$ ) and the Street 3 contains a loop. RIGHT: At intersection  $A$ , segment 1 could be associated to 3 and 2. The closest angle to  $\pi$  is made by 3 but 3 and 4 correspond to an angle reciprocally minimal. They are associated and 1 is then associated with 2. The same reasoning leads to the same associations whatever the first segment considered.

these set we write  $a \sim b$ . If  $v$  is a vertex,  $E(v)$ ,  $H(v)$  and  $T(v)$  are respectively the sets of edges, streets and trunks that pass through  $v$ . Similarly we write  $V(e)$  the end points of the edge  $e$ ,  $V(t)$ ,  $V(h)$ ... The **degree function** of a city  $G$  is defined for  $V$ ,  $E$ ,  $H$ ,  $T$ , written  $d_G(x)$  and counts the number of elements in the same set that are directly adjacent to  $x$ .

### 4.3 Algorithms to Recover Street Hypergraph structures

Let  $C = (V, E)$  be a city graph. We have proposed to represent mathematically coherent structures in the city by hypergraphs. The trunk hypergraph is evident to define. But when coming to street hypergraph, we have to deal with the idea of alignment. Intuitively, a street is a collection of coherent street segments. We present here two algorithms, depending on an angular parameter  $\alpha$  that defines the notion of alignment, and check them against our data base to asses their relevance and determine the values for  $\alpha$  that make sense in our urban context.

#### 4.3.1 Angular tolerance (AT)

We use the local reflexive relationship  $\hat{R}_\alpha$  depending on the angular parameter  $\alpha$ :

$$e_1 \hat{R}_\alpha e_2 \quad \text{iif} \quad \exists v, v_1, v_2 \in V, e_1 = [vv_1], e_2 = [vv_2], \\ | (d(v) = 2) \vee (|\angle(\overrightarrow{vv_1}, \overrightarrow{vv_2}) - \pi| \leq \alpha) \quad (4.9)$$

this relation considers that two adjacent street segments are part of the same street if they meet at a junction or if they meet at an intersection but remain almost aligned (Fig. 4.4 left). This algorithm strongly risks producing "branched streets" (red solid line in Fig. 4.4 left, Fig. 4.5).

### 4.3.2 The minimal reciprocal alignment (MRA)

To avoid the above cited problem, let us define  $\hat{S}_\alpha$ . We position at particular vertex  $v$  and consider the set of the edges passing through it  $E(v) = \{e_1 = [v_1v], \dots, e_n = [v_nv]\}$ . We iteratively define  $\hat{S}_\alpha$  with the variable  $s$ :

1. the initial "remaining edges" is set for  $s = 0$ :  $E_0 = E(v)$ ,
2. we consider all pairs of edges  $(e_i, e_j) \quad 1 \leq i < j \leq n, e_i S_\alpha e_j$  iff

$$\begin{aligned} & |(\angle(\overrightarrow{vv_j}, \overrightarrow{vv_j}) - \pi) \leq \alpha) \\ & \text{and } \forall e_k = [vv_k] \in E_s \neq e_i, e_j, \quad |(\angle(\overrightarrow{vv_k}, \overrightarrow{vv_i}) - \pi) < |(\angle(\overrightarrow{vv_i}, \overrightarrow{vv_j}) - \pi| \\ & \text{and } |(\angle(\overrightarrow{vv_k}, \overrightarrow{vv_j}) - \pi) < |(\angle(\overrightarrow{vv_i}, \overrightarrow{vv_j}) - \pi| \quad (4.10) \end{aligned}$$

Two edges are associated if they are the most aligned in  $E_s$ .

3.  $E_{s+1}$  is  $E_s$  without the edges associated in the  $s$  step.
4. We go on till  $(E_s)$  stabilizes.

The reflexivity on the minimal condition induces the reflexivity of  $\hat{S}_\theta$ .

For instance in Fig.4.4 right:  $E_0 = \{1, 2, 3, 4\}$ , 3 and 4 are associated,  $E_1 = \{1, 2\}$ , 1 and 2 are associated and the algorithm ends.

Roughly speaking this algorithm tries to minimize the global "torsion" in the set of streets. Compared to "AT", this algorithm will not produce streets with branches (excepted in very special cases).

## 4.4 Tuning and Performances

We have specified *AT* and *MRA* with a single angular parameter  $\alpha$ . In practice we want the algorithm to recover the actual streets of a city.

It is hard to access to these information with our data. But in a particular city, their number can be extracted although not reliable: there are as many streets as there are different **street names** in the data base. . We just try to reach the true number of streets. (AT) and to a lesser extent (MRA) risk producing branched rather straight streets. We define **the branching coefficient** to describe this tendency and seek out to minimize it.

In this section we assess the performances of the algorithms and deduce an optimal tuning for  $\alpha$  from a corpus of  $N = 109$  major French towns:  $(C_1, \dots, C_N)$ .

### 4.4.1 Criteria

#### Number of street recovering

We assume we know for  $N$  cities their actual number of streets:  $T_1, \dots, T_N$ . Let  $\alpha \longrightarrow f_k(\alpha)$  the function that associates to an angle  $\alpha$  the number of streets one of our algorithm assess

for the city  $k$ . If the algorithm is relevant, the quadratic error

$$\Delta^2(\alpha) = \frac{1}{N} \sum_{k=1}^N \left( \frac{f_k(\alpha) - T_k}{T_k} \right)^2 \quad (4.11)$$

is small. However  $(T_k)$  is not accurate because some street segments have a blank "NAME" field. The data base underestimates the number of streets. To get around this problem, we assume the error in the data base is proportional to the proposed number of streets:  $\tilde{T}_k = (1 + \lambda)T_k \quad \forall k$ . The criterion rewrites in function of  $\alpha$  and  $\lambda$ :

$$\Delta^2(\alpha, \lambda) = \frac{1}{N} \sum_{k=1}^N \left( \frac{f_k(\alpha) - (1 + \lambda)T_k}{(1 + \lambda)T_k} \right)^2 \quad (4.12)$$

A quick study of the data base permits to assess that  $0.1 < \lambda < 0.7$ . This can seem surprisingly high but these blank fields are above all related to small streets, side-streets, short paths...  $\frac{\partial \Delta^2}{\partial \lambda}(\alpha, \lambda) = 0$  leads to a functional relationship between  $\alpha$  and  $\lambda = \lambda_\alpha$ :

$$\lambda(\alpha) = \frac{\sum f_i(\alpha)^2 / T_i^2}{\sum f_i(\alpha) / T_i} - 1 \quad (4.13)$$

and the criterion rewrites only in function of  $\alpha$ :

$$\Gamma^2(\alpha) = \frac{1}{N} \sum_{k=1}^N \left( 1 - \frac{f_k(\alpha)}{T_k \cdot \frac{\sum f_i(\alpha)^2 / T_i^2}{\sum f_i(\alpha) / T_i}} \right)^2 \quad (4.14)$$

### Branching coefficient

Let  $H$  a hypergraph structure computed from  $C$  and  $h \in H$  a street, seen as an extracted subgraph of  $C$ . The number of branches in  $h$  is defined by:

$$\xi(h) = \sum_{v \in h} \max(d_h(i) - 2, 0) \quad (4.15)$$

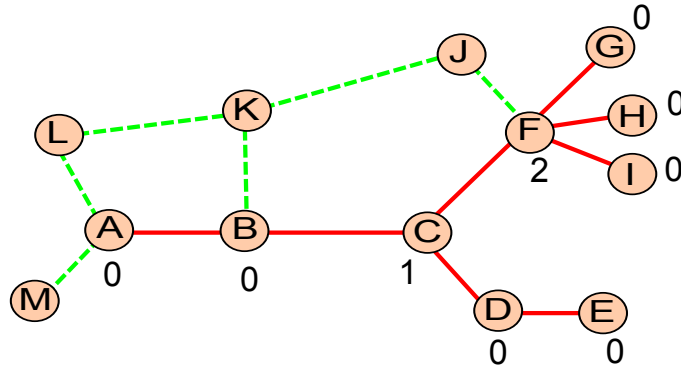
To measure the branched aspect of  $H$  we define its branching coefficient from the number of branches of its streets (see Fig.4.5):

$$\Xi(H) = \frac{\sum_{h \in H} \xi(h)}{\sum_{k > 2} (k - 2) \cdot d^\circ(k)} \quad (4.16)$$

If none of the streets is branched,  $\Xi = 0$  and if  $H$  is compounded of a single non straight street,  $H$  is maximally branched wit  $\Xi = 1$ .

### 4.4.2 Analysis

For the 109 cities of the data base and several values of the angle  $\alpha$  we have calculated the mean criteria  $\Gamma$  and  $\Xi$ . The algorithm is relevant and its angle is good if  $\Gamma$  and  $\Xi$  are minimal. This permits for each algorithm to find the best angle and assess the relevance of the algorithms.



**Figure 4.5:** The red solid graph is a street. Its number of branches is  $1 + 2 = 3$ .

### AT

The function  $\Gamma(\alpha)$  reaches its absolute minimum in 0. But this value is eliminated since it drives to an aberrant value of  $\lambda$ . The second best value of the criterion (21%) is reached around  $\pi/5$  (Fig. 4.6 top-left). This corresponds to a  $\lambda = 0.5$  (Fig. 4.6 top-right) which is coherent with the order of magnitude we expressed when looking at the data base. The absolute minimum in the branching coefficient is in average  $\Xi = 0.15$  which is slightly high but stays reasonable (Fig. 4.6 bottom-right). Fig. 4.6 bottom-left shows the criterion for  $\lambda$  constant equal to 0.5. With the corrected number of streets the criterion is convex and  $\pi/5$  appears as a rather good and stable minimum.

### MRA

For this algorithm, the function  $\Gamma(\alpha)$  is almost constant equal to 0.2 (Fig. 4.7 top-left).  $\lambda_\alpha$  is exponentially decreasing with an asymptotic value of 0.56 (Fig. 4.7 top-right).

But  $\forall \lambda \in [0, 1]$   $C(\alpha, \lambda)$  has an asymptotic minima (when  $\alpha \rightarrow \pi/2$ , Fig. 4.7 bottom-left). The choice of  $\lambda$  is hence not clear but for every reasonable value of  $\lambda$ , the criteria is optimized for  $\alpha \rightarrow \pi/2$ .

Conversely  $\lambda = 0.56$  is stable since almost every angle provides a value of  $\lambda$  close to 0.56:  $0.56 \simeq \lambda_\alpha \quad \forall \alpha \in [\pi/5, \pi/2]$

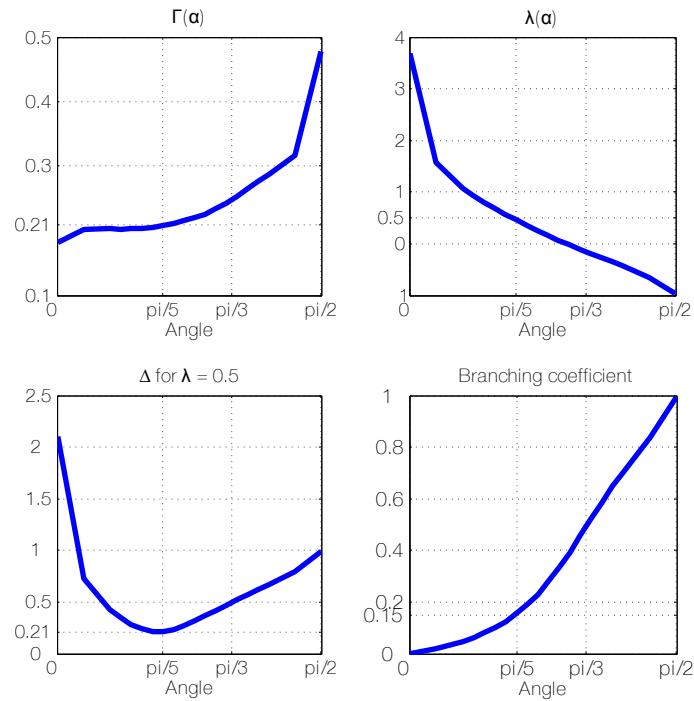
Moreover this is the optimal value we found for  $\lambda$  with AT which is comforting.

$\Xi_{MRA} < 0.01 = \Xi_{AT}/10$  (Fig. 4.7 bottom-right) which is very satisfactory: (MRA) almost does not produce branches.

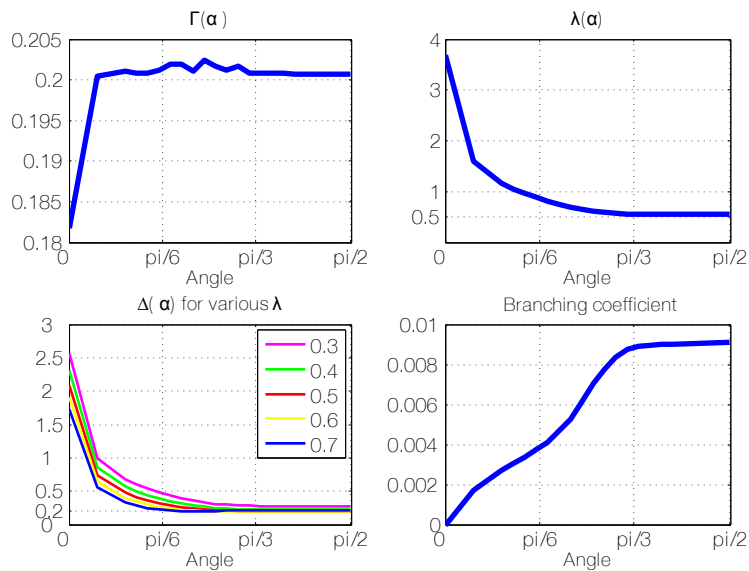
In fact the found optimal value  $\alpha = \pi/2$  means that the best tuning of the algorithm is "angle free". Either the vertex under consideration is a junction or there is at least an angle smaller than  $\pi/2$ . Consequently the condition on  $\alpha$  is relaxed from  $S_\alpha$  to  $S = S_{\geq \pi/2}$ . The global minimum is the same for the two algorithms: 0.21 but the branching coefficient is much smaller for MRA. Branches in streets are anecdotal when using MRA (one can appear in a very particular case even if instinctively we would say branches are an impossible configuration).

We will in practice use the MRA in its maximal version that does not depend on the angle.





**Figure 4.6:** Tuning and performances of the Angular Tolerance algorithm (AT). Top-left:  $\Gamma(\alpha)$ , Top-Right:  $\lambda(\alpha)$ , Bottom-Left:  $\Delta(\alpha)$  for  $\lambda = 0.5$  and Bottom-right: the increasing of the mean branching coefficient  $\Xi$  with  $\alpha$ .



**Figure 4.7:** Tuning and performances of the Minimal Reciprocal Angle algorithm (MRA). Top-left:  $\Gamma(\alpha)$ , Top-Right:  $\lambda(\alpha)$ , Bottom-Left:  $\Delta(\alpha, \lambda)$  for  $\lambda = 0.3$  to  $0.7$  (its optimal value) and Bottom-right: the increasing of the mean branching coefficient with  $\alpha$  asymptotically inferior to 1%

From $\downarrow$ to $\rightarrow$	Space	Adjacency matrix	Incidence list	Adjacency list	Mixed
Adjacency matrix	$O(V^2)$	-	$O(V^2)$	$O(V^2)$	$O(V^2)$
Incidence list	$O(V + E)$	$O(E)$	-	$O(E)$	$O(E)$
Adjacency list	$O(V + V \cdot \bar{d}_G)$	$O(E)$	$O(V \cdot E^2)$	-	$O(V \cdot E^2)$
Mixed	$O(V + V \cdot \bar{d}_G + E)$	$O(V)$	$O(1)$	$O(1)$	-

**Table 4.1:** Storage size of each structure and the complexity of the algorithm converting a structure into another.

## 4.5 Computational implementation

In this section we present the way one can code a city graph on a computer and discuss the efficiency of various structures from the complexity of essential algorithms on graphs. It appears that the incidence list is the best compromise between storage size and computational complexity of common algorithms. We review classical algorithms (connected component, shortest path distances) and propose an efficient algorithm for the new problem of deriving an hypergraph from a graph. A crucial remark to legitimate the hypergraph structure in 6.3.2 will be that the point to point shortest path problem is solved in  $\#V^2$  or  $\#V \log \#V$  for a weighted graph and in  $\#V$  for an unweighted graph.

### 4.5.1 Structures

A straight graph is totally defined by the positions of its vertices and its **adjacency matrix**. Nonetheless the adjacency matrix is sparse and thus is not efficient in terms of storage space. We can use an **incidence list** representation to overcome this difficulty. Vertices are stored into an array or list with two columns and as many lines as there are vertices and edges into another one with two columns, each one containing a reference to the extremities.

Another structure which is redundant but efficient is the **adjacency list**. To each vertex  $v$  is associated the list of vertices adjacent to it  $V(v)$ .

The last structure we use is **mixed**: it contains a vertices array, an edge array and for each vertex  $v$  the list  $E(v)$ . A hypergraph can either be represented by an additional array containing for each edge a label or a list of edges for each hyperedge. The incidence list is in general the best compromise between storage size and efficiency in algorithms. The table Tab.4.1 sums up the size of each structure and the complexity necessary to pass from one structure to another. The following paragraph lists the common algorithm we have to deal with in our framework. In most of the cases incidence list is an efficient structure and when it is not the case, the cost to convert that structure to a more efficient one is negligible.

In the following we write for an hypergraph  $G = ((V, E), H)$ ,  $V = \#V$ ,  $E = \#E$ ,  $\bar{d}_G$  the mean degree in  $V$  and  $H = \#H$ . As an order of magnitude,  $500 \leq V \leq 20000$ ,  $\bar{d}_G \simeq 2.5$  and thus  $E$  and  $V$  are almost the same,  $H \simeq V/10$ .

### 4.5.2 Algorithms

Here are the most important algorithms we will have to handle.. We present them in pseudo-code and calculate their complexity with the adapted structure.

**Find the vertices adjacent to a vertex and Find the edges passing through a vertex** These trivial problems are solved in  $O(E)$  with an incidence list by looking at the extremities of each edge. An adjacency list permits to do the task in a constant time but the structure conversion is  $O(E)$ .

**Get  $n$  random points on the graph** To compute integral numerically along the street system with  $\mu_G$ , it is necessary to sample randomly the map in  $n$  independent points. This can be done with incidence list in  $O(nE)$ :

---

**Algorithm 2** Random points on a geometrical graph

---

```

1: INPUT Graph  $G = (V, E)$ , Number of Points  $n$ 
2: vector CumLength size  $E$   $\text{CumLength}(i) = \sum_{j \leq i} \|e_j\|$ 
3:  $L =$  total street length
4: for  $i=1$  to  $E$  do
5:    $u =$  Uniform Random Number in  $[0L]$ 
6:    $e = \text{argmin}(\text{CumLength}(e_i) \geq u)$ 
7:   Vertices  $v_1, v_2 =$  extremities of  $e$ 
8:    $\lambda =$  Uniform Random Number in  $[0, 1]$ 
9:   OUTPUT( $i$ ) =  $\lambda.v_1 + (1 - \lambda)v_2$ 
10: end for

```

---

An edge is chosen at random with a probability proportional to its length. Then a number  $\lambda$  between 0 and 1 is uniformly picked and the resulting random point is the barycentre of the edge's extremities weighted by  $\lambda$  and  $1 - \lambda$ .

**Compute the connected components of a graph** When importing a map, it is necessary to erase some disturbing detached edges. To this we compute the connected components of the whole graph and only keep its largest class of equivalence. This is done linearly in time with a mixed structure or in  $O(V.E)$  with an incidence list.

The idea is to take a vertex at random, to give it a label 1 and to spread the same label to its neighbours and then to the neighbours of its neighbours... The successive neighbours are stored in a FIFO, when it is empty the algorithm increments the label and re-starts from the first vertex non-labeled it finds.

**Compute the shortest path distance between all pair of vertices** We want in a first time find the shortest path distance along the street system for any couple of vertices. A classical and very simple to understand algorithm is the Floyd Warshall's algorithm: The time complexity is  $O(V^3)$  and the storage capacity is  $O(V^2)$ . There exist some modifications

**Algorithm 3** Connected components of a graph

---

```

1: INPUT Graph  $G = (V, E)$ 
2: Component = Vector size  $V$  initialized to 0
3: Visited = Vector size  $V$  initialized to false
4: FIFO Current =  $[V(1)]$ 
5: Visited( $i$ ) = true
6: currentLabel = 1
7: while (not(isvoid(Current))) do
8:   Vertex  $v = \text{dequeue}(\text{Current})$ 
9:   Component( $v$ ) = currentLabel
10:   $N = \{w \in V(v), \text{not}(\text{Visited}(w))\}$ 
11:  Visited(all in  $N$ ) = true
12:  Current = queue(Current , N)
13:  if isVoid(Current) then
14:    Current =  $[\text{argmin} \{i, \text{Component}(i) = 0\}]$ 
15:    currentComp ++
16:  end if
17: end while

```

---

**Algorithm 4** Floyd Warshall's algorithm for all shortest paths in a positively weighted graph

---

```

1: INPUT Adjacency Matrix A size  $V \times V$ 
2: Matrix Distances size  $V \times V$ 
3: Each element in Distances =  $\infty$ 
4: If  $A(i,j)=1$  then  $D(i,j)=1$ 
5: Each diagonal element in Distances = 0
6: for  $k=1$  to  $n$  do
7:   for  $i=1$  to  $n$  do
8:     for  $j=1$  to  $n$  do
9:       Distances( $i, j$ ) =  $\min(\text{Distances}(i, j), \text{Distances}(i, k) + \text{Distances}(k, j))$ ;
10:    end for
11:   end for
12: end for

```

---

of this algorithm providing a time complexity of  $O(V^2 \log V)$  but they are much more difficult to implement.

Given two vertices, their single shortest path can be found in  $O(V \log V)$  by the Dijkstra's algorithm using appropriated structures.

For two random points  $x$  and  $y$  on the graph we can adapt the previous algorithms by remarking that:

$$d_C^{sp}(x, y) = \min_{e_x \in E(x), e_y \in E(y)} \{d_C^{sp}(e_x, e_y) + \|e_x - x\| + \|e_y - y\|\} \quad (4.17)$$

For an unweighted graph, this can be done much more efficiently: in  $O(V)$  for a particular couple of points and in  $O(V^2)$  for every pair. We will see it is a primordial element that

---

**Algorithm 5** Dijkstra's algorithm for the point to point shortest path in a positively weighted graph

---

```

1: INPUT Graph  $G = (V, E)$ , Vertex start, Vertex end
2: Vector currentDistances of size  $\#V$  initialized to  $\infty$ 
3: currentDistances(start) = 0
4: for all  $v \in V(\text{start})$  do
5:   currentDistances( $v$ ) =  $d(\text{start}, v)$ 
6: end for
7: while  $\exists x \in V \mid \text{currentDistance}(x) = \infty$  do
8:    $v_1 = \text{argmin}(\text{currentDistances})$ 
9:   for all  $v_2 \in V(v_1)$  do
10:    currentDistances( $v_2$ ) =  $\min(\text{currentDistances}(v_2), \text{currentDistances}(v_1) + d(v_1, v_2))$ 
11:   end for
12: end while
13: OUTPUT currentDistances(end)

```

---

makes the hypergraph structure appealing in the next section.

The idea is again to explore progressively the neighbourhoods of a vertex with a FIFO (or a Stack).

---

**Algorithm 6** Point to Point shortest path in unweighted graphs

---

```

1: INPUT Graph  $G = (V, E)$ , Vertex  $v_1$ , Vertex  $v_2$ 
2: Vector Distances of size  $V$  each element initialized to  $\infty$ 
3: FIFO Current = [ $v_1$ ]
4: FIFO Next = []
5: CurrentDistance = 0
6: while Distances( $v_2$ ) =  $\infty$  do
7:   for all  $v \in \text{Current}$  do
8:     Distances( $v$ ) = CurrentDistance
9:     Next = Next  $\cup \{w \in V(v) \mid \text{Distances}(w) = \infty\}$ 
10:   end for
11:   Current = Next
12:   Next = []
13:   CurrentDistance ++
14: end while

```

---

If the graph under consideration is a small-world (the mean shortest path length scales as the logarithm of the size of the graph, we will be more rigorous on this definition in 5.3 and App.B), we can show the algorithm solves the point to point problem in  $\log V$  is in average, the point to all other points in  $V$ .

We note  $n_i^v$  the number of vertices at a distance  $i$  of a the vertex  $v$ . To find the distance to  $v$  of a vertex at a distance  $i$ , it is necessary to explore  $\sum_{j=0}^i n_j^v$  vertices. We write  $M(v)$  the maximal distance to vertex  $v$ . The diameter  $D(G)$  of the graph is the maximum of  $M(v)$ ,  $\alpha(G)$  is the mean shortest path distance between two random vertices. If the graph

is small-world,  $D(G)$  and  $\alpha(G)$  are  $O(\log \#V)$  (see 5.3 for the relations between average, radius and diameter). The total time to find independently all the shortest paths is

$$\sum_{i=0}^{M(v)} \sum_{j=0}^i n_j^v = \sum_{i=0}^{M(v)} (M(v) - i)n_i^v \quad (4.18)$$

$$= M(v)\#V - \mathbb{E}(d^{sp}(v, \cdot)) \quad (4.19)$$

Thus the expectancy of complexity for the shortest path distance between a random vertex and an other is

$$\frac{1}{(\#V)^2} \sum_{v \in V} (M(v)\#V - \mathbb{E}(d^{sp}(v, \cdot))) \leq D(G) - \frac{\alpha(G)}{(\#V)^2} \quad (4.20)$$

$$= O(\log \#V) \quad (4.21)$$

If this algorithm does not work with general weighted graphs it is because of loops and branches: it is impossible to explore progressively the increasing neighbourhoods, a vertex can be in a remote neighbourhood but close because of weights in the path linking it to the source vertex.

Nonetheless the algorithm can be easily adapted for weighted graphs that present no branches: "line graphs". This may seem a too restricted class of graph to be interesting but we will see in Ch.6 it is a very important preamble for our Cab Driver's Algorithm that approximates shortest path problem in a logarithmic time.

---

**Algorithm 7** Point to Point shortest path in weighted line graph

---

```

1: INPUT Graph  $G = (V, E)$ , Vertex  $v_1$ , Vertex  $v_2$ 
2: Vector Distances of size  $V$  each element initialized to  $\infty$ 
3: Distances( $v_1$ ) = 0
4: FIFO Current = [( $v_1, \emptyset$ )]  $\triangleright$  FIFO of tuples vertex - vertex, the second vertex is the
   previous vertex in the path from the source  $v_1$  to the current vertex.
5: FIFO Next = []
6: while Distances( $v_2$ ) =  $\infty$  do
7:   for all ( $v, u$ )  $\in$  Current do
8:     Distances( $v$ ) = Distances( $v$ ) +  $d(u, v)$ 
9:     Next = Next  $\cup$  {( $w \in V(v)$ )Distances( $w$ ) =  $\infty, v$ }
10:  end for
11:  Current = Next
12:  Next = []
13: end while

```

---

**Compute an hypergraph** Both algorithms (AT) and (MRA) can be implemented within the same skeleton by encapsulating two functions "Relation" with a Boolean output, taking as parameters a vertex  $v$ , the set  $E(v)$  and two distinct elements of it. The algorithm divides in two steps: (1) determine local relations between segments (2) transform this relation into equivalence classes by using Eq. 4.4. In the following code we mix up objects and their

---

**Algorithm 8** Hypergraph Computation

---

```

1: Input Graph  $G = (V, E)$ 
2: Vector  $H$  of size  $E$ 
3: Array of Sets Connection size  $E$ 
4: for all  $v \in V$  do                                     ▷ Step 1: local relationship
5:   Neighbours =  $E(v)$ 
6:   for  $1 \leq j < k \leq \#$ Neighbours do
7:      $e_1 =$  Neighbours( $j$ )
8:      $e_2 =$  Neighbours( $k$ )
9:     if Relation(  $v, e_1, e_2, \text{Neighbours}$ ) then
10:       Connection( $e_1, \text{next index available}$ ) =  $e_2$ 
11:       Connection( $e_2, \text{next index available}$ ) =  $e_1$ 
12:     end if
13:   end for
14: end for
15: CurrentStreetMark = 1                                     ▷ Step 2: global propagation of the relationship
16: for all  $e \in E$  do
17:   if  $H(e) = 0$  then
18:     stack = [ $e$ ]
19:     while notEmpty(stack) do
20:       current = pop(stack)
21:        $H(\text{current}) = \text{CurrentMark}$ 
22:       push(stack, { $f | f \in \text{Connection}(\text{current}), H(f) = 0$ })
23:     end while
24:     CurrentMark ++
25:   end if
26: end for

```

---

index in an array. With plain graph structure, the complexity is  $O(V \times E)$  (Step 1) and  $O(E)$  (Step 2) thus globally in  $O(V^2)$ . With an adjacency list (calculated in  $O(E)$ ) Step 1 becomes  $O(V)$  and the whole algorithm is  $O(V)$ .

## 4.6 Conclusion

In this chapter we have presented a proper mathematical framework and its optimized implementation to work on city maps. We will use this formalism in the next chapters. It is theoretically relevant, allows physical interpretation of measures such as centralities. One of the success of this chapter will be presented in Sec.6.3.2 with the Cab Driver's Algorithm that will use all the knowledge and formalism of the first chapter of this thesis to approximate the shortest path problem in a map in a logarithmic time, using an hypergraph structure as an intermediary.

# Chapter 5

## Phenomena

### Contents

---

5.1	Return on phenomena described in the literature . . . . .	<b>85</b>
5.2	Street scaling . . . . .	<b>86</b>
5.2.1	Empirical street length fitting . . . . .	86
5.2.2	Interpretation . . . . .	87
5.2.3	Recall on Kolmogorov-Smirnov test . . . . .	88
5.3	Small World effect . . . . .	<b>89</b>
5.3.1	Definitions . . . . .	89
5.3.2	Observations . . . . .	90
5.3.3	Discussion . . . . .	92
5.4	Distance transformation . . . . .	<b>93</b>

---



**Phenomena: Synthesis**

Both by their global shapes and the structure of their street network, cities exhibit a large diversity. This **diversity** is sometimes interpreted as the expression of the human freedom and creativity. Nonetheless, we can measure some general features on cities' map and interpret them as hallmarks of the functional constraints that build cities. In the next chapter we will attempt to show that in fact these hallmarks can be explained by some **emerging phenomena**.

It has been shown in the literature that the area of cells in a city follows a power law distribution. We show here that **the length of street is log-normally distributed**, that **the street graph** (or hypergraph) of a city exhibits a **small-world behaviour** and that **Euclidian and Shortest path distances** in a city dominate each other.

**Contributions of this chapter**

1. Street length scales as a log-normal function, at least for the 109 French towns under study.
2. Street hypergraph exhibits a small-world behaviour, rigorous definition of this notion in the context.
3. Relation between Euclidian and shortest path distances in cities.

## 5.1 Return on phenomena described in the literature

The literature presented in Ch.1 has pointed out some phenomena occurring in the urban context. The most well known is the hierarchical repartition of resources in a system of cities induced by the Zipf's law [138, 120]. This is a multi - city scale phenomenon in a polycentric system. If the empirical fitting of data by a power law may be statistically contestable, the main idea is relevant: the distribution of a "measure" of city size is scale invariant, with a very few big entities and a lot of small ones. This can be explained by a preferential attachment model: the bigger a city is the more it tends to attract new resources created in the whole city system.

At a single city scale, [40] shows that the radial distribution of resources follows an exponential law. This result can be challenged considering the measurement tools that were available in the 50's when was written the article. More recently [30] proposed to model this distribution by an exponential, a power law or a Gaussian. In short, resources tend to decay with the distance to the gravity center. At the same time, a large body of literature [11, 51, 52, 53, 123] points out the fractal organization of a city. This is coherent with the hierarchical distribution of sizes in a city systems. In fact it is the geometrical formulation of the hierarchy principle on the different parts of the city itself. But a city is compound of a few scales only. We will only remember from this theory that a city presents several scales on its street system and the boundary of a city is not a simple shape and is difficult to define objectively: there is continuity between the city and its countryside.

The network has been studied from a functional point of view in [74, 114, 33, 105, 106]. The main conclusions are that on the plain graph of a city's map there are no very noticeable phenomena. The topology is very constrained by the planarity of the graph, the associativity is trivial and the robustness of the network is quite variable according to its structure but remains as classical as other planar networks [34]. Emerging phenomena seem to appear not on the map or "primal graph" but on the "dual" (street) graph, that is to say the graph that represents the relationship between streets or axes. An interesting result relative to the distribution of cells in a street network has been found in [79]. The distribution of cell area approximately follows a power law with exponent 2. We do not have easily access to cells statistics with our data base. We consider this result as true and will try to mimic it in the models we present in Ch.8. A model based on addition and connections of centers is built in [8]. This model also reproduces other topological statistics the author presents in the beginning of its articles.

Here we will observe two phenomena occurring on the street hypergraph: the street length repartition follows a log normal law and the dual graph present somewhat of a small-world behaviour. We will prefer to use the expression "hypergraph" to designate the graph obtained by considering streets as vertices and their intersections as edges.

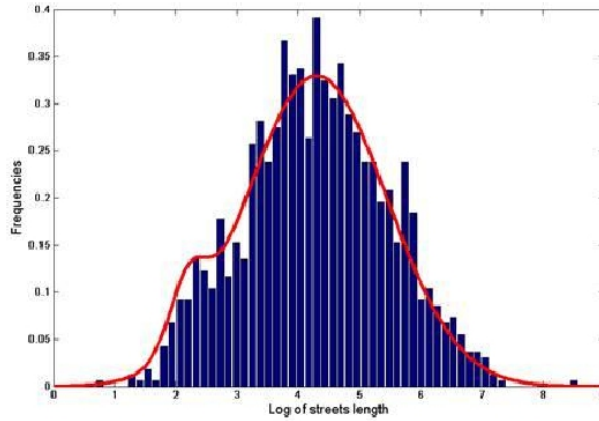
## 5.2 Street scaling

### 5.2.1 Empirical street length fitting

In a city, there are long streets assuring an efficient transportation system and small streets let us say "densely" distributed to provide habitation space. We thus expect that the distribution of street lengths  $L$  exhibits a wide range of values or scales logarithmically. Fig.5.1 plots the distribution of the logarithm of street lengths (derived from (MRA) but the result is robust with other methods) in the French city of Amiens. The global shape of this histogram suggests two maxima and two (different) normal tails. We assume that  $\log L$  follows a mixture of two Gaussians (or similarly that  $L$  follows a mixture of log-normal laws):

$$\log L \sim p_- \mathcal{N}(m_-, \sigma_-) + (1 - p_-) \mathcal{N}(m_+, \sigma_+) \quad (5.1)$$

with  $m_- < m_+$ . The identification of this model has been performed with an Expectation Maximization algorithm. Other cases show that the multi log normal distribution is robust



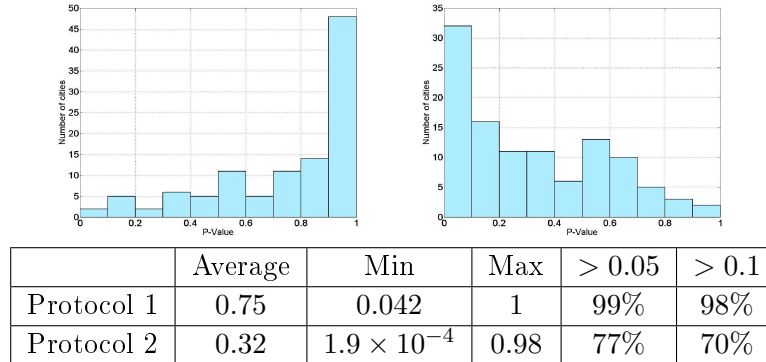
**Figure 5.1:** The distribution of the logarithm of street lengths in Amiens (France). The red curve is the fitting of this distribution by a mixture of two Gaussians.

even if it is possible to observe one or two maxima. For our whole data base of French towns we calculated a bi-normal fitting of  $L$  and calculated the p-value of this fitting from a Kolmogorov - Smirnov test: " $L$  follows a mixture of two log normal laws" against " $L$  does not follow a mixture of two log normal laws". We have chosen this test rather than a Chi-2 for its robustness to distribution supports which is an important requirement when testing heavy distribution tails. Theoretical considerations and a discussion on the choice of this test can be found in 5.2.3.

In Protocol 1 we have for each city calculated the best parameters with an EM algorithm and calculated the p-value of " $L$  follows a mixture of two log normal laws with these estimated parameters" against " $L$  does not follow a mixture of two log normal laws with these estimated parameters", as often done in the literature. Nonetheless the statistics of the test is changed if parameters are estimated with the same data as used for the test.

We built a second protocol: since Kolmogorov -Smirnov is relevant from 100 samples and our cities typically contain 500 to 1000 streets, we randomly divide each length distribution

in two parts, used one to estimate parameters and the other to perform the test. The estimation and the test are done with less data and are then less accurate. Results for both methods are summed-up in Fig. 5.2. The hypothesis is as relevant as the p-value is close to



**Figure 5.2:** Main characteristics of the p-value distribution for the test "the distribution follows a mixture of log-normal" in 109 French Cities. Protocol 1 estimates parameters and performs a Kolmogorov -Smirnov test with the same data. Protocol 2 used the (randomly chosen) half of the streets to assess parameters and the other half to perform the test.

1. Traditionally one considers that the hypothesis cannot be rejected if  $p\text{-value} > 0.1$ .

The first protocol shows a really impressive p-value of 0.75.

Let's focus on the second method. It is theoretically valid but needs randomization. From a realization to another the p-value of a particular city may highly change but the average p-value remains between 0.3 and 0.4. In 77% of cases the hypothesis is not rejected and in average the p-value is 0.32 which is quite high: if with a p-value of 0.1 one cannot assert the hypothesis is false, it begins to be true when the p-value is above 0.3.

### 5.2.2 Interpretation

Log-normal laws are common in nature [77]. They appear in concentration of elements, latency periods of disease, rainfall, permeability in plant physiology... They are characteristic of multiplicative processes. We then could think that a city shapes itself by dividing in smaller blocks former blocks. This would lead to consider the city is the result of a division process. [124] recalls that for isotropic planar tessellations stable under iteration the length of the typical "I segment" (a street) is long-tailed but the result is not a log-normal. It is necessary to add another phenomenon to get the log-normal distribution.

This could be the consequence of the extension of the city: people have a typical transportation length:  $\lambda$ . They accept to settle in a place where they have access to a constant volume of resources at a distance smaller than  $\lambda$ . Then when they cannot further divide blocks they place at the exterior of the city into larger blocks.

To come to bimodality: this one does not appear clearly on each city. A social science explanation is the succession of several transportation modes along time or that several populations built the city with two different policies (inhabitants and industries for instance). We do not have the material to discuss deeply this observation but we would rather incline

toward the intrinsic hallmarks of the duality between division and extension of space when a city grows.

### 5.2.3 Recall on Kolmogorov-Smirnov test

The Kolmogorov-Smirnov test is a non parametric statistical test. Let  $f$  be a continuous density of probability and  $(x_1, \dots, x_n)$   $n$  samples. We want to measure whether these samples could have been generated by real random variables of density  $f$ . We test  $H_0$ : " $(x_1, \dots, x_n)$  are independent samples from random variables following  $f$ " versus  $H_1 = \bar{H}_0$ .

Let  $F(x) = \int^x f(t)dt$  the "theoretical" distribution function associated to  $f$  and  $F_n(x) = \#\{x_i \leq x, 1 \leq i \leq n\}$  the "empirical" distribution function of the samples. One creates the statistics

$$D_n = \sqrt{n} \sup_x |F(x) - F_n(x)| \quad (5.2)$$

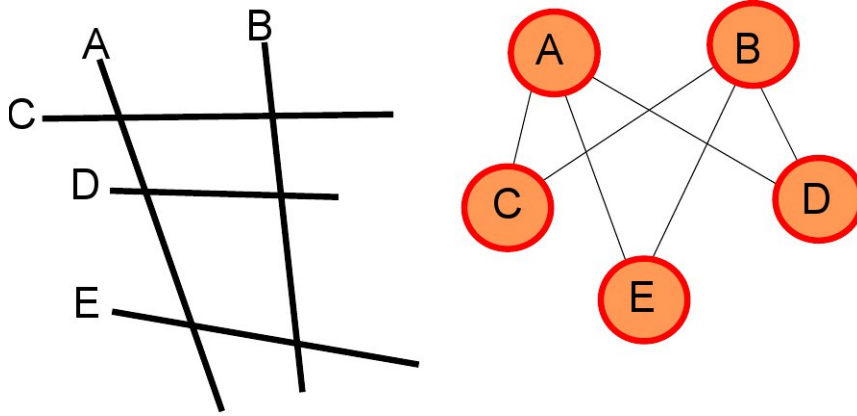
Since  $\{x_i\}$  is finite, that supremum can be calculated as a minimum. Under  $H_1$ ,  $D_n \rightarrow \infty$  but under  $H_0$ ,  $D_n \rightarrow K = \sup_{[0,1]} |B(t)|$  (in distribution) where  $B$  is the Brownian bridge. The distribution function of  $K$  is known:

$$\mathbb{P}(K \leq x) = 1 - \sum_{i=1}^{\infty} (-1)^{i-1} e^{-2ix^2} \quad (5.3)$$

Thus for the measure  $D_n$  it is easy to compute the p-value of the test :  $\mathbb{P}(K \geq D_n)$ . In our case (a Gaussian mixture), the theoretical repartition function  $F$  writes simply in function of the function erf. We have preferred a KS-test rather than a  $\chi^2$  test since:

1. The test is independent of the choice of bin size in an histogram
2. The test is robust toward the support of distribution. Indeed  $\chi^2$  assumes there is a reference compact interval on which is calculated the difference between theoretical frequencies and measured ones. This interval is generally chosen with min and max observations as boundary and this tacitly introduces a lot of information into the test.

Nonetheless, contrary to  $\chi^2$  test, there is no general variation of KS test that permits to test adequacy when parameters have been estimated from the data used to perform the test. We could have used boot strap methods to overcome this difficulty. But the implementation was not realistic for the five free parameters. Since we had numerous data for each city, we decided to randomly cut these data into two sets: one to assess parameters and one to perform the test.



**Figure 5.3:** On the left, a city graph with its streets A,B,C,D,E forming the hypergraph  $H$ . On the right the representation of  $H$  as a graph: each street label is now a vertex and an edge is drawn between two of them if the street intersect in the geometrical projection of the city graph.

## 5.3 Small World effect

### 5.3.1 Definitions

Let  $G$  a city with  $H$  an hypergraph structure over it. Let us call topological distance of  $H$  the function  $d_H^{topo} : H \times H \rightarrow \mathbb{N}$  that satisfies:

$$d_H^{topo}(h_1, h_2) = \begin{cases} 0 & \text{if } h_1 = h_2 \\ \min_{h \in H, h \cap h_2 \neq \emptyset} d_G^{topo}(h, h_1) + 1 & \text{otherwise} \end{cases} \quad (5.4)$$

The topological distance counts the number of times one needs to turn to go from a street to another one.  $H$  can be naturally seen as another graph we still write  $H$  with an adjacency matrix  $\mathbf{H}$ . The hyperedges become vertices and an edge between two vertices is drawn only if the two associated hyperedges meet in  $G$  (as geometrical subgraphs).  $d^{topo}(h_1, h_2)$  is then simply the shortest path distance between  $h_1$  and  $h_2$  in the unweighted graph  $H$  with the shortest path matrix  $\mathbf{H}^*$  (Fig.5.3) .

**The topological average distance of a street  $h_0$  is :**

$$\frac{1}{\#H} \sum_{h \in H} d^{topo}(h, h_0) \quad (5.5)$$

A central street would be a street that minimizes the previous sum.

One drawback of this wholly topological definition is that, since topological distances are integers, several streets can be defined simultaneously as central streets. Common sense would then be to take all of these streets as simultaneous central streets, and to calculate the distance of any street as the minimum of the distance to any of these streets to define the notion of distance of a street to the center of the city.

Another way is to weight with the street lengths:

$$\bar{d}_C^{topo}(h_0) = \frac{1}{\mu_C(C)} \sum_{h \in H} \|h\| \cdot d^{topo}(h, h_0) \quad (5.6)$$

( $\|h\|$  is the length of the street  $h$ ). This definition has an interesting interpretation: if  $x$  is a point on  $h_0$  and if  $Y$  is a uniformly chosen point on  $C$  (with respect to  $\mu_C(\cdot)$ ), then the expectancy of the number of turns to do to go from  $x$  to  $Y$  is  $\bar{d}_C^{topo}(h_0)$ . Based on this observation, we define for the whole hypergraph  $H$  an **average topological distance coefficient for a city**:  $\alpha$  that is the mean number of turns to do to go from a random point to another in the city. It represents the "inverse of the efficiency" of the city.

$$\alpha(H) = \frac{1}{\mu_C(C)^2} \sum_{h_1, h_2 \in H} \|h_1\| \|h_2\| \cdot d^{topo}(h_1, h_2) \quad (5.7)$$

If  $L$  is the vector containing the length of streets,  $\alpha(H)$  writes as a bilinear form:

$$\alpha(C) = \frac{1}{\mu_C(C)^2} {}^t L H^* L \quad (5.8)$$

A street that minimizes  $\bar{d}_C^{topo}(\cdot)$  is unique if the network is not too regular (like a square lattice). It is called the **center of the city**  $h_C$ . The topological distance of streets to the center of the city constitutes the **topological map** of the city and can be represented with color scales as for the city of Amiens and its central district in Fig.5.4.

The **topological radius** of the city is then defined by

$$r_C = \max d_C^{topo}(h, h_C) \quad (5.9)$$

and its **diameter** by :

$$\text{diam}_C = \max_{h_1, h_2 \in H} d_C^{topo}(h_1, h_2) \quad (5.10)$$

In an evident way,

$$r_C \leq \alpha_C \leq \text{diam}_C \leq 2r_C \quad (5.11)$$

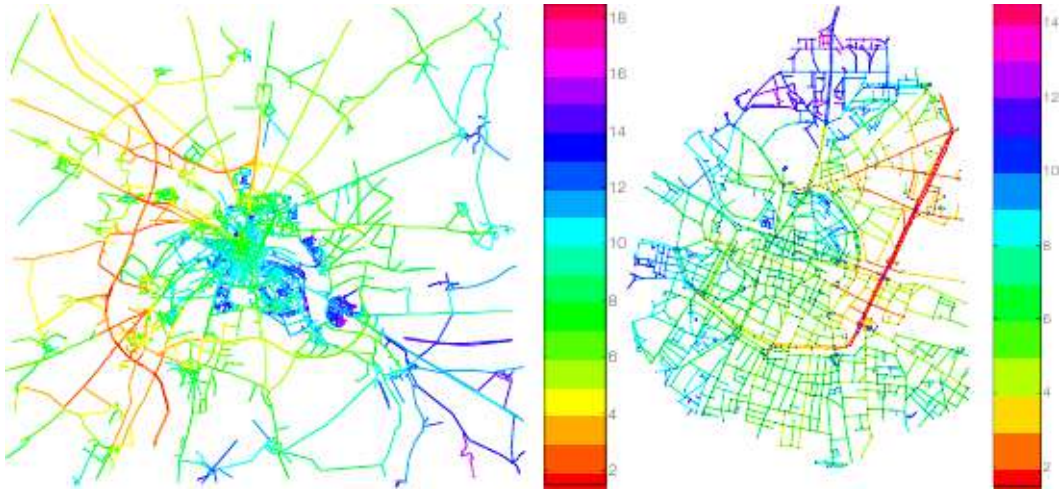
In Amiens, the topological center appears to be a part of its highway-belt. In Fig.5.4 we have plotted the distance of each street to the topological center. This map gives a hierarchical vision of the space. There is no radial component of the increase of the topological distance: a scale of long streets serves the whole city, allowing the variation of the topological distance to be mainly local.

Added to that the topological radius of the city grows very slowly with the size of the city (14 in the center of Amiens, 18 in the whole city that is eight times bigger).

### 5.3.2 Observations

More generally we would not be surprised to observe a Small World behaviour in  $H$ . A graph exhibits a Small World behaviour (see [1, 98, 23] or App.B for more details) when the average path length between two random nodes is very small compared to the number of nodes in the network. The problem is to define properly the notion of "small". In a dynamical context, when the number of nodes  $N_t$  and the average path length  $P_t$  are functions of time, "small" is when  $P_t \leq O(\log N_t)$ .

We consider a set of cities of total street length (size)  $L_i$  and of average topological distance  $\alpha_i$ . Fig.5.5 represent the relation between  $\log L_i$  and  $\alpha_i$  in our data basis. This scatter plot



**Figure 5.4:** The "topological maps" of Amiens (left) and of its central district (right). In each map, the red street is the topological center and the color of each street refers to its distance to that center. The maximum distance to the center is the radius of the map (18 in the whole city and 16 in the center). It is striking that this radius increases slowly with the size of the considered street system. We can even infer that the construction of surrounding highway belt (the found center) is made precisely to keep the topological radius of the city small. The XIXth century boulevard is a aside structuring element of the city and thus has become a historical center for the core of Amiens.

is fitted by a line whose slope of 1.0086 is very close to 1. For the sake of simplicity we have fixed the slope at exactly 1 and redone a median fit to find a robust gap at the origin. Then for a city of size  $L$ , the **expected average topological path distance** is (defined as):

$$\alpha_{th} = \log \frac{L}{L_0}, \quad L_0 = 0.905 \cdot 10^3 \text{m} \quad (5.12)$$

One would say the points in Fig.5.5 are quite dispersed around the fitted line. Indeed but the important thing is that the order of magnitude of  $\alpha$  is the same than  $\log L$ . And the gap  $\Delta$  between  $\alpha_{th}(L)$  and the observed  $\alpha_{obs}$  is an information on the efficiency of a particular city relatively to an average city independently of its size Fig.5.6.

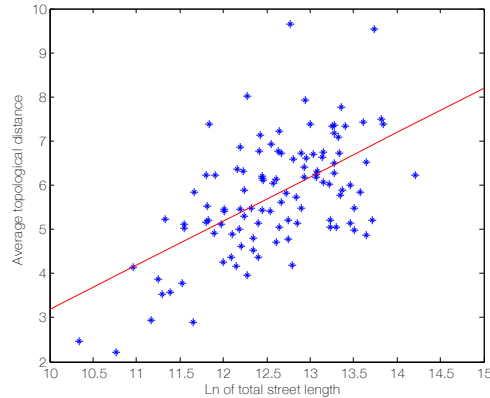
$$\Delta = \alpha_{obs} - \log \frac{L_{obs}}{L_0} \quad (5.13)$$

We have chosen to define  $\Delta$  from the theoretical efficiency function of  $L_{obs}$  rather than from the line of Fig.5.5 since this way  $\Delta$  appears more like an additive noise.

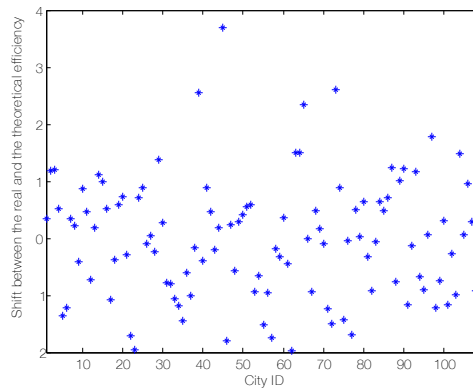
If  $\Delta < 0$  as for Boulogne-Billancourt and Lyon, the city is more efficient, if  $\Delta > 0$  as for Digne-les-Bains, Gap, Nantes and Mende the city is less efficient.

In Fig.5.7, we represented for each city the ratio between its diameter and its radius. This ratio lies theoretically between 1 and 2. If the ratio is 2 as in Amiens, Annecy, Belfort, Bourges, Colmar, Evreux, Le Mans, Levallois-Perret, Tarbes, Troyes, Vannes and Villers-sur-Mer, the city is called "centralized": the topological distance to go from a point to another is the same than to go from the first point to the center and then to switch to the





**Figure 5.5:** For the 109 French towns of our data basis, in blue, the mean topological  $\alpha$  distance between two points in a city in function of the logarithm of total street length (as a measure of the size) of the city. The equation of the red line is  $y = x - 6.8079$ .



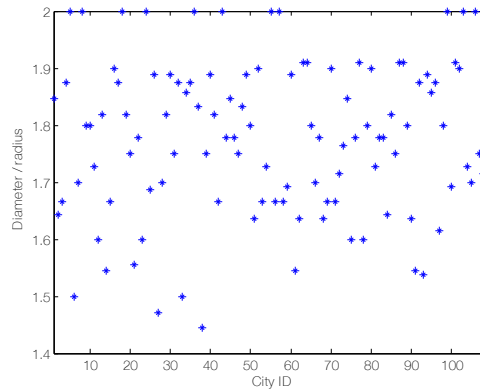
**Figure 5.6:** The shift  $\Delta$  between the measured mean efficiency of a city and the expectancy according to the total street length from Eq.5.12. Toward this criterion, cities with a negative  $\Delta$  are more efficient and these with a positive  $\Delta$  are less efficient. The x-axis corresponds to the name of each city in our data base ordered by alphabetic order (ID). We have chosen this x-axis since  $\Delta$  is uncorrelated to the length of streets.

second point (even if the "centre" is a surrounding boulevard or highway). The city center is a necessary passage point.

Conversely if the ratio is smaller as in Angers, Carcassonne, Chaumont or Dax, the city is "with distributed resources": it is not necessary to pass through the center to displace, there exist efficient streets everywhere in the city.

### 5.3.3 Discussion

In Eq.5.12 defining the theoretical efficiency of a city,  $L_0$  can be considered as a typical street length, which could depend on the culture of the country and the age of the city system under consideration. In France, this typical length has the order of magnitude of the kilometre. But what is amazing is that the factor in front of the logarithm is almost exactly 1 which means that the phenomenon expresses in the Neperian basis.



**Figure 5.7:** The topological radius vs the topological diameter of the cities in our data basis. By definition  $\text{radius} \leq \text{diameter} \leq 2 \cdot \text{radius}$ . If  $\text{diameter} \simeq 2 \cdot \text{radius}$ , the city is "centralized": to move from one point to another, a person needs to pass through the center of the city; contrary to that is  $\text{diameter} < 2 \cdot \text{radius}$ , there often exist an efficient path from a point to another that does not pass through the topological center of the city: we will say the city has "distributed resources".

The x-axis corresponds to the ID of each city in our data base.

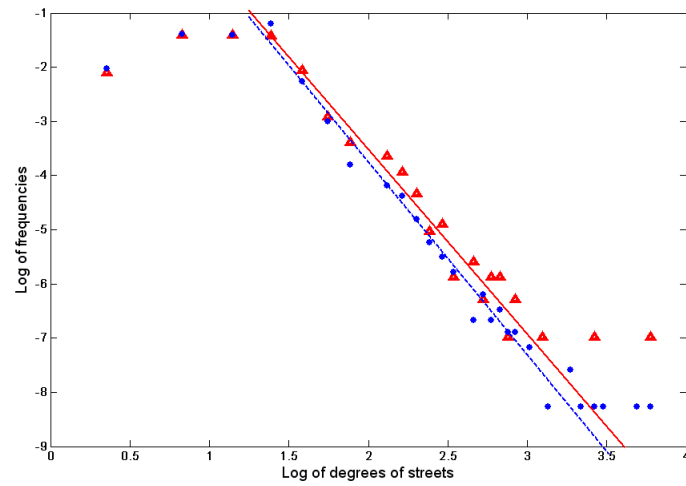
We are tempted to explain this small world behaviour by an underlying scale-free behaviour on the degree of streets. Nonetheless even if in Amiens and its center (Fig.5.8) the degree function looks like a line, the range is not wide enough to claim a scale-free behaviour. Then we will content ourselves to note a hierarchical distribution of streets that is strongly correlated with the street length and leads to a small-world behaviour all the same. There are long streets creating a upper scale that desert the city area, to which connect smaller streets. We also note that the street graph is not transitive: streets do not form triangles which has for consequence that the robustness of the street system is very sensitive to strategic attacks.

## 5.4 Distance transformation

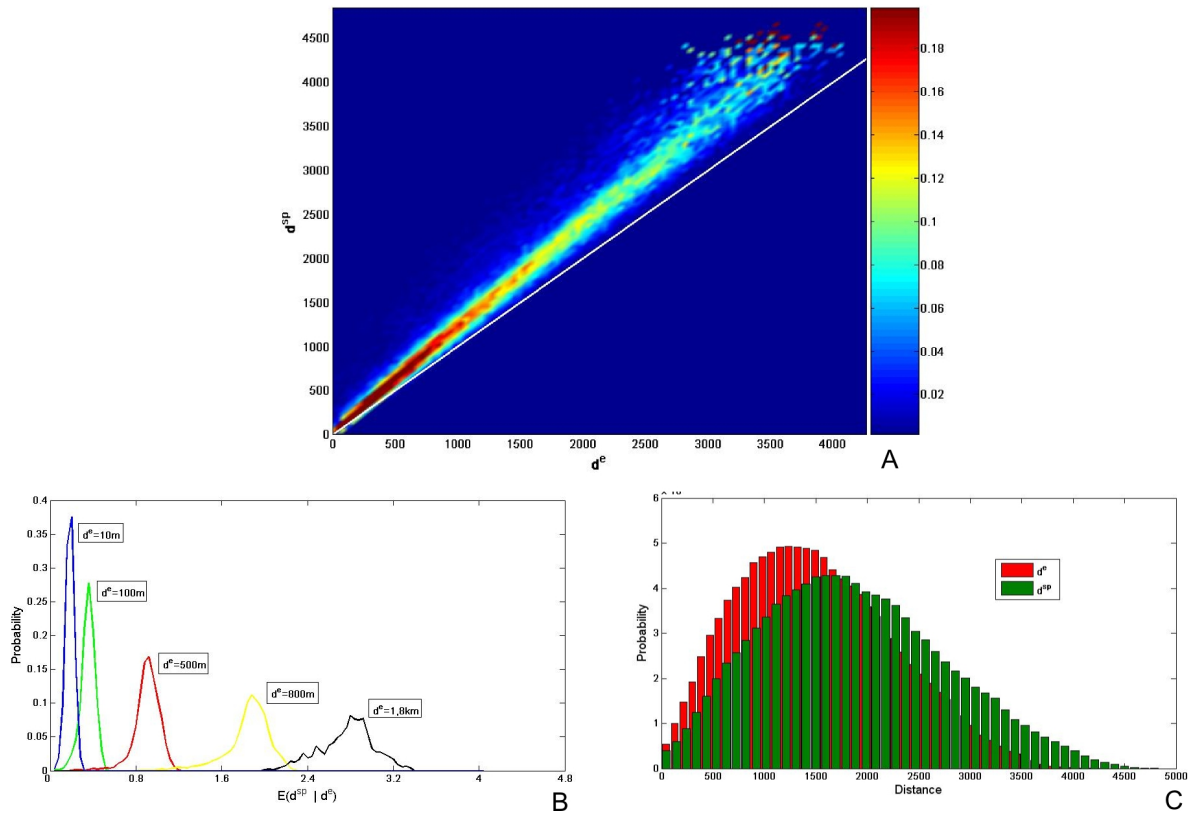
Several distances can be defined given any pair of points  $x, y \in C$ . The most natural are the Euclidian or fly-bird distance:  $d^e(x, y) = ||x - y||$  and the shortest-path distance with respect to the streets network  $d_C^{sp}(x, y)$ . To observe the distribution of  $d^e$ ,  $d_C^{sp}$  and their correlation, we randomly sampled Amiens with respect to the probability measure  $\mu_C(\cdot)/\mu_C(C)$  over 5000 couples of points  $S = (X_i, Y_i)$ . It is noticeable that distributions of  $d^e$  and  $d_C^{sp}$  have the same global shape (Fig.5.9-c) and that conditionally to  $d^e$ ,  $d_C^{sp}$  is localized around a value close to a constant plus  $d^e$  (Fig.5.9-b) as if  $d_C^{sp}$  could write  $\alpha \cdot d^e + \beta$  (Fig.5.9-a).

The conditional distributions look like normal distribution and the global distribution of distances could be fitted by beta laws converging to Weibull laws when the observation window is large enough.

Note that a priori the proportionality coefficient depends on the global shape of the city. In a city dense enough, relatively homogeneous, with no evident ill deserved zones, Euclidian distance between two points is a good order of magnitude of the shortest path distance.



**Figure 5.8:** Log-Log plot of the street degree distributions in Amiens (red triangles) and its center (blue points). The scatter plots can be fitted by lines but the variation of street degree is not sufficient to speak of scale-free distributions.



**Figure 5.9:** The distribution of  $d^{sp}$  conditionally to  $d^e$ . The colormap shows  $\mathbb{P}(d^{sp}|d^e)$ . Probabilities have been truncated at 0.2 get a better color dynamics. The white line is the first bisector. The second graph shows the probability distribution of  $d^{sp}$  for different values of  $d^e$ . Curves have pretty much the same global shape with a variance increasing with  $d^e$ . The last plot C shows the repartition of Euclidian and Shortest Path distances through the city, the two resulting histograms have the same shape, Shortest Path distance is Euclidian distance translated slightly and dilated.



# Chapter 6

## Map Description and analysis

### Contents

---

6.1	Topological and Geometrical descriptors . . . . .	<b>99</b>
6.1.1	Size of a city . . . . .	99
6.1.2	Topology . . . . .	99
6.1.3	Anisotropy . . . . .	100
6.1.4	Density, Compactness . . . . .	102
6.1.5	Network efficiency . . . . .	102
6.2	Map analysis and Centralities . . . . .	<b>102</b>
6.2.1	Classical centralities . . . . .	103
6.2.2	Simplest distance and simplest centrality . . . . .	103
6.2.3	Display . . . . .	104
6.2.4	Comparison of centralities . . . . .	105
6.2.5	Iterated centrality . . . . .	106
6.3	Discussion . . . . .	<b>106</b>
6.3.1	Justifications for simplest centrality . . . . .	106
6.3.2	A new algorithm to find shortest path: The Cab Driver's Algorithm . . . . .	107
6.3.3	Conclusion . . . . .	108

---

---

**Map Description and analysis : Synthesis**

---

We present here a few parameters allowing describing and comparing the shape of maps: **size, topology, anisotropy, compactness, efficiency**. We redefine in our mathematical framework the notion of **centrality**. We study two classical centralities and compare them to a new one we propose: **the simplest centrality**.

It is based on the average topological distance between streets. We compare its capacity to interpret a map, detect ill deserved zones and main axis and its time complexity with the classical centralities.

We **justify** a posteriori and a priori its consistency and relevance. We propose to interpret this centrality as the approximation of a stationary flow of displacements.

We propose a new algorithm to solve the point to point shortest path problem: the Cab Driver's Algorithm that approximates the solution in a logarithm time.

---

**Contributions of this chapter**

---

1. Definition of a few description parameters: size, anisotropy, topology, compactness, efficiency.
2. Redefinition of centrality in the city graph framework.
3. Definition of the simplest centrality and comparison with classical centralities. Simplest centrality presents a more coherent hierarchical view on the city, putting in light main axis, ill deserved zones. It is boundary effects free. Furthermore it is much faster to run.
4. A priori mathematical justification for simplest centrality.
5. Logarithmic algorithm for point to point shortest path on city maps: the Cab Driver's Algorithm.

## 6.1 Topological and Geometrical descriptors

Bettencourt et Al. have statistically shown in [19] that every macroscopic measure of the "size" of the city writes as a power function of another measure of the size. Here we only get interested to the measures that only depend on the extracted city graph of a map  $G = ((V, E), H)$ .

### 6.1.1 Size of a city

In complex network theory, the measure of the size of a graph is naturally its number of vertices. The number of vertices is similarly taken in [33, 73, 74, 35] as the measure of the size of the city when looking only at its map.

But this measure can be largely biased by the artificial number of degree 2 vertices as said in Ch.4 and observed in Fig.6.1. One possibility is to remove these nodes. The number of vertices does not take into account the scale of the city: for instance picture a city only constituted by a a very long street compared to a city made of four little streets forming a square. From the number of vertices point of view, the second city is the biggest but in fact, the first one gathers more resources. The total length of streets in the graph  $\mu_G(G)$  measures the intrinsic size of the city.

### 6.1.2 Topology

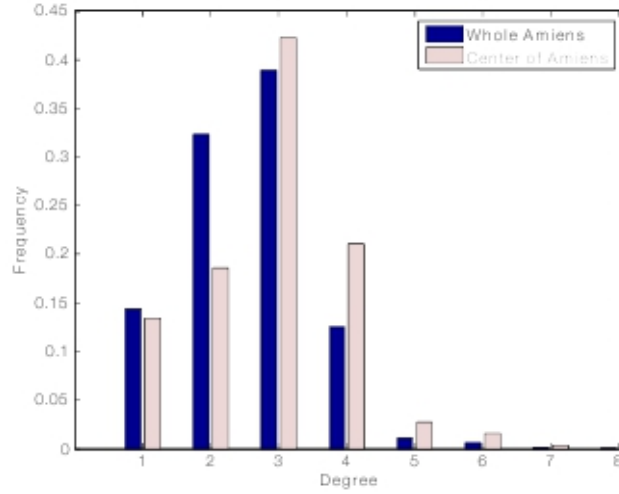
Let  $N(k) = \#\{v \in V, N(v) = k\}$  be the number of vertices of degree  $k$  in  $G$  and  $\bar{N}(k) = N(k)/\sum N(i)$  be their mean number. As said above, the set  $V_2$  (junctions) should not be taken into account since it only represents sampling artefacts to preserve the shape of streets. In [33] the histogram of  $N$  is studied by means of an exponential tail of distribution. Nonetheless, this distribution Fig.6.1 is very peaked in 3 or in 4. It is thus sufficient to describe the histogram  $N$  by **the organic ratio**:

$$r_N = \frac{N(1) + N(3)}{\sum_{j \neq 2} N(j)} \quad (6.1)$$

This allows to discriminate quickly whether the city had been planned or not. Indeed a planned city is filled with a homotopy of a rectangular grid (only  $\bar{N}(4) \neq 0$  leading to  $r_N \simeq 0$  in the limit case). This grid is clearly useful to settle buildings but also sticks to human perception of space since we have the intuition of left - right / front - behind. Higher order connections are also characteristic of planned pattern creating radial places on purpose. In unplanned or organic cities i.e. created by the interaction between non concerted settlements, there is little probability for streets segments to be coherent and so the largest number is  $N(3)$  with also some dead-ends leading to  $r_N \simeq 1$ . This parameter we have introduced is thus a very simple measure of the degree of planning or organicity of the city.

Following [35] we characterize the topology of a city by its "meshedness coefficient". It's easy to count  $\#V$  and  $\#E$  then the number of cells,  $\#C$  is deduced from Euler's formula.





**Figure 6.1:** The histograms of degrees' distribution in Amiens (whole city and center). We observe that both distributions are peaked in 3 even in the more regular city centre, where the number of connections 4 is still half. The large number of 2 comes from straightening, especially in the suburb curved streets and should not be taken into account.

Given  $V$ , the maximum number  $2\#V - 5$  of faces is obtained by the Greedy Triangulation algorithm [35]. So the quantity  $M = (\#E - \#V + 1)/(2\#V - 5)$  equals to 0 if the city is a tree and is close to 1 if it is a highly connected graph. For real cities [74]  $M$  typically ranges between 0.08 and 0.35.

To be coherent with our preceding remark we should not take junctions into account. It has no incidence on the numerator:  $\#E - \#V$  is constant but we have to change  $2\#V - 5$  into  $2\#V.(1 - \bar{N}(2)) - 5$

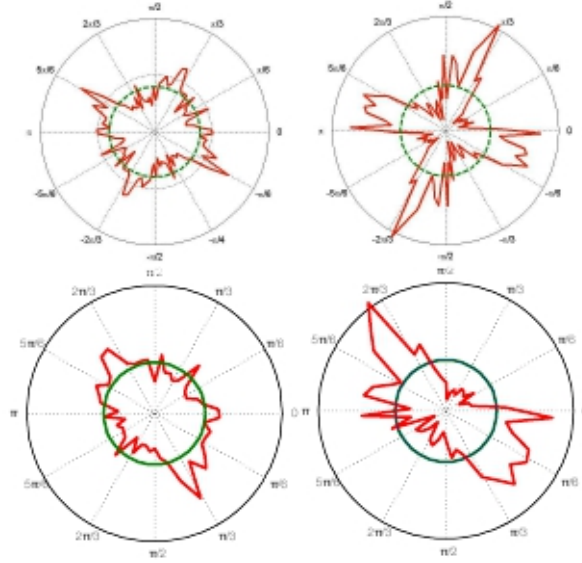
$$M_3 = \frac{\#E - \#V + 1}{2\#V.(1 - \bar{N}(2)) - 5} \quad (6.2)$$

$M_3$  is quite small because of the general lack of triangles in the topology of a city contrary to a greedy triangulation network. As a trapezoid contains two triangles, we rescale it in  $M_4 = 2M_3$  whose maximum is hit when the considered city contains the maximal number of trapezoids. We thus modify the meshedness coefficient of [35] to take into account the lack of meaning of junctions and the lack of triangles in actual cities.

Amiens appears as an "average" organic city, with a meshedness coefficient  $M_4$  of 0.41 (0.54 when restricted to the center) and  $r_N = 0.79$  (0.68 in the center).

### 6.1.3 Anisotropy

In order to grant an efficient access to physical resources, the street system locally tends to be perpendicular locally to **structuring elements** as for example rivers or older streets (as explained in Ch.2). Then a city is not "isotropic" but generally presents two main directions. Although everybody agrees that a city is anisotropic, we have found any definition of anisotropy in the literature or a way to measure it. The measure we have introduced in Ch.4 allows to define rigorously this instinctive notion.



**Figure 6.2:** Angular distribution of Amiens (top left : whole city - top right : center). We notice on both distributions an approximate four-fold symmetry. The doubled angular distributions (bottom) look like ellipsoids which allows to define the anisotropy coefficient as the ratio of the eigenvalues of the inertia matrix:  $Ani = 0.42$  in the whole city and  $Ani = 0.71$  in the center.

Let  $\vec{u}_0 \in \mathbb{R}^2$  be an arbitrary vector, taken as an angular origin. For angle  $\alpha \in [0, \pi]$  we want a measure of the "total length" of streets in  $G$  that are oriented in directions  $[0, \alpha]$ :

$$\Psi^*(\alpha) = \mu_G(g \in G, \angle(g, \vec{u}_0) \in [0, \alpha]) \quad (6.3)$$

where  $\angle(., .)$  is the angle measure between two vectors in  $[0, \pi]$ . In the special case of straight graphs, the impact of each street-segment is proportional to its length. From this, we define the angular density by

$$\Psi(\alpha) = \frac{d\Psi^*(\alpha)}{d\alpha} \quad (6.4)$$

The drawing of  $\Psi$  reveals the anisotropy of the city graph (Fig.6.2). For an isotropic city, the angular density  $\Psi_I$  would be a continuous and uniform density  $\Psi_I(\alpha) = \frac{1}{\pi}$ .

We notice a fuzzy four-fold symmetry as the result of local street perpendicularity. It would be useful to sum up the four-fold symmetry of this angular density as a single normalized indicator. We looked for a bound distance measure between  $\Psi$  and  $\Psi_I$ .

Since the angle is defined modulo  $\pi$ , we can "fold"  $\Psi$  :  $\Psi^{(2)}(\theta) = \Psi(\theta)e^{i2\theta} - \int \Psi(u)e^{i2u} du$ . The observed distribution  $\Psi$  is discrete because of the limited number of streets segments, measures like  $\int |f - \frac{1}{\pi}|^n$  highly depends on the size of the bean chosen to estimate the integral.

The inertia matrix

$$\begin{pmatrix} \int \Re(\Psi^{(2)})^2 & - \int \Re(\Psi^{(2)})\Im(\Psi^{(2)}) \\ - \int \Re(\Psi^{(2)})\Im(\Psi^{(2)}) & \int \Im(\Psi^{(2)})^2 \end{pmatrix}$$

which is symmetric and positive (from Cauchy-Schwartz's inequality) with two eigenvalues  $\lambda_1 > \lambda_2$  this allows us to define

$$Ani = 1 - \frac{\lambda_2}{\lambda_1} \quad (6.5)$$

as an anisotropy coefficient.

For Amiens, its anisotropy coefficients varies from 0.42 in the whole city to 0.71 in the center. This corresponds to the visual impression of Fig.6.2. Indeed the historical grid-like Roman center of the city has diluted with distance and time to a more isotropic radial distribution of resources.

#### 6.1.4 Density, Compactness

We propose here an easy to calculate indicator that expresses in the same time the compactness of the city's shape and of its street system.

Let  $A$  be the area of the convex hull (App.A) of a city graph  $G$ .  $\mu_G(G)$  is the total length of its street system.

Then we imagine a city with a square convex hull of the same area that is to say with a side of length  $\sqrt{A}$  and a regular square lattice filling this hull. The constraint for the mesh-size is to obtain the same total length of streets. The area of mesh divided by the area of the hull is a number between 0 and 1 and one minus this quantity is a measurement of the idea of compactness in the city. Thus we define:

$$\text{Comp} = 1 - \frac{4A}{(\mu_G(G) - 2\sqrt{A})^2} \quad (6.6)$$

Comp is generally located between 0.9 and 1 (0.936 for Amiens): cities we study are very compact.

#### 6.1.5 Network efficiency

If we pick two points at random on a map  $G$ , the expectancy of the shortest path distance between these two points is the integral of  $d^{sp}(\cdot, \cdot)$  over all couple of points with respect of the uniform measure  $\mu_G \times \mu_G$  divided by  $\mu_G(G)^2$ . The expectancy is obviously comprised between 0 and  $\mu_G$ .

$$\int \int d^{sp}(x, y) d\mu_G(x) d\mu_G(y) = \mu_G(G)^3 \alpha_G \quad (6.7)$$

$\alpha_G \in [0, 1]$ . Scaling arguments shows that  $\alpha_G$  only depends on the global shape of the city graph. It can tend to 0 for a highly dense square network and to 1 for an infinite binary tree.

Thus the average path length in a city is directly proportional to the total length of its streets:  $\mu_C(C)$ . The ratio between the average path length and the total street length is an indicator of the efficiency of the street network.

## 6.2 Map analysis and Centralities

The parameters previously defined permit a global characterization of maps, we here address the problem of comparing the quality of different points in a map. We want to define a measure on the geometry of a map that quantifies if a point is well located or not. This extends the classical topological notion of centrality.

### 6.2.1 Classical centralities

There are two natural distances on a geometrical graph: the Euclidian  $d^e$  and the shortest path one  $d^{sp}$ . In the complex network framework [1, 23] a centrality is a measure defined on  $V$  that quantifies whether a node has a central location or not. Among already defined centralities (closeness, betweenness, straightness, information, see Ch.1), we focus here on those that directly derive from distances.

**The closeness centrality**  $C_i^C$  of a node  $i$  measures if the node is close to the others in average.

$$C_i^C = \frac{\#V - 1}{\sum_{\substack{j \in V \\ j \neq i}} d^{sp}(i, j)} \quad (6.8)$$

**The straightness centrality**  $C_i^S$  of a node  $i$  measures if the node is rather in a straight line or not and if that node efficiently transforms Euclidian distances into shortest path distance

$$C_i^S = \frac{1}{\#V - 1} \sum_{\substack{j \in V \\ j \neq i}} \frac{d^e(i, j)}{d^{sp}(i, j)} \quad (6.9)$$

For a continuous and geometrical description of the city, avoiding topological bias, these two centralities can be redefined by means of  $\mu_G$  for each point  $x$  of the city graph:

$$C^C(x) = \frac{\mu_G(G)}{\int_G d^{sp}(x, g) d\mu_G(g)} \quad (6.10)$$

$$C^S(x) = \frac{1}{\mu_G(G)} \int_G \frac{d^e(x, g)}{d^{sp}(x, g)} d\mu_G(g) \quad (6.11)$$

$C^S(x)$  is well defined since  $d^e(x, g) \underset{x \rightarrow g}{\sim} d^{sp}(x, g)$ .

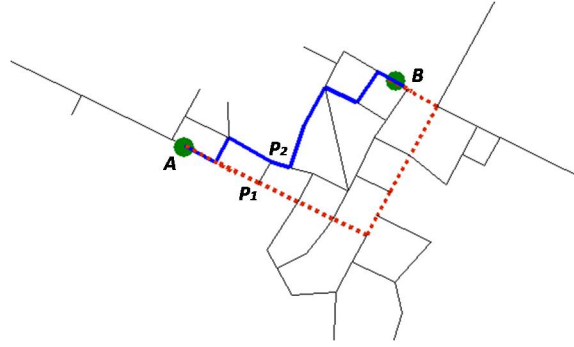
### 6.2.2 Simplest distance and simplest centrality

A plausible behaviour for a human being to go from place  $A$  to place  $B$  would be to adopt **the simplest path** instead of the shortest one. To model this choice, we define the information distance  $d_{C,p}^i(A, B)$  from a point to another along a path  $p$  in the city. And we define the information distance between those two points as :  $\min_{p \text{ path from } A \text{ to } B} d_{C,p}^i(A, B)$ . Let's compute this distance from the toy map Fig.6.3. To direct somebody from  $A$  to  $B$  along the dotted red path ( $P_1$ ) instructions are: (1) From  $A$  take the 1st street to the right with respect to the house, (2) go straight (keep on the same street) while meeting **6** street-crossings (an intersection of degree  $n$  counts for  $n - 2$  street-crossing), (3) on the 7th intersection, take the second street from the right, (4) go straight - **4** street-crossings, (5) on the 5th intersection take the third one, (6) go straight - **1** streets-crossing, (VII) walk for 35 meters.

This is encoded into the "path information vector" where the odd components are the number of the street to take at the current intersection (the first street being the right one), the even components are the number of intersections to cross straight before turning, and the last component is the distance to go in the last street segment:  $\vec{I}(A \rightarrow B || P_1) = [1, \mathbf{6}, 2, \mathbf{4}, 3, \mathbf{1}|35]$  and for  $P_2$  the blue dotted path that is the shortest path between  $A$  and

$B$ ,  $\vec{I}(A \rightarrow B || P_2) = [1, \mathbf{0}, 2, \mathbf{0}, 1, \mathbf{1}, 2, \mathbf{0}, 2, \mathbf{0}, 2, \mathbf{0}, 1, \mathbf{0} | 15]$ . The information length of a path is its number of components: 7 and 13 respectively for  $P_1$  and  $P_2$ . We define the information distance between two points as the minimal information distance among all paths that go from  $A$  to  $B$ . In Fig. 6.3 the red dotted path was the simplest so the information distance from  $A$  to  $B$  is 7. By convention, **the simplest distance** between  $A$  and  $B$   $d^{sim}(A, B)$  is 0 if  $A = B$ , 1 if  $A$  and  $B$  are in the same street and the information distance minus 2 otherwise (the first and the last component in the information vector being always necessary). The simplest path distance is then a well defined distance for all points in a geometrical graph. This distance corresponds generally to the topological distance of the streets the points are in plus one.

There may exist several simplest paths between two points. To each simplest path we associated its metric simplest distance: the total length of the path. The simplest path that minimizes the metric simplest distance is called the shortest simplest path.



**Figure 6.3:** Two points  $A$  and  $B$  located on a city graph. The blue solid path is the shortest path between the two points and the red dotted one is a simplest path but not the shortest simplest path.

As for the closeness, the averaging of the simplest distance defines a centrality, (**the simplest centrality**):

$$C^{sim}(x) = \frac{\mu_G(G)}{\int_G d^{sim}(x, g) d\mu_G} \quad (6.12)$$

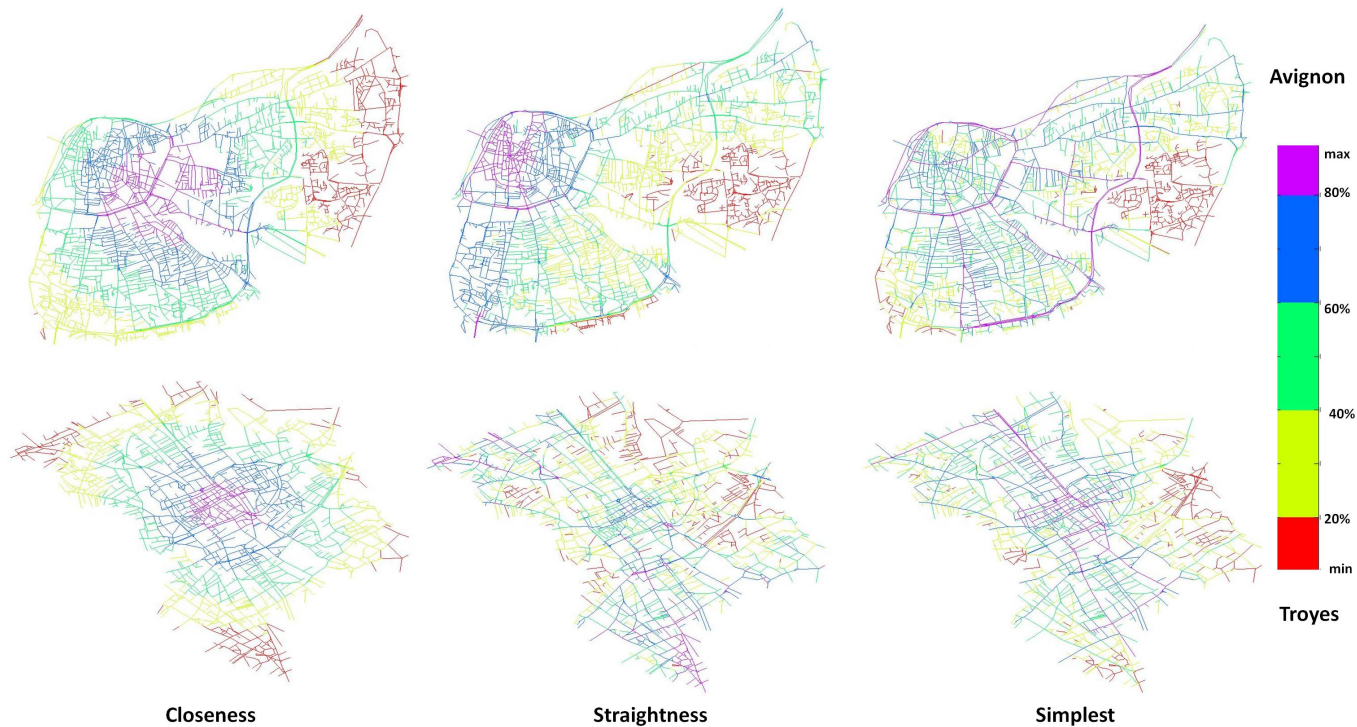
Since  $d^{sim}$  is constant on a street  $h$ , the simplest centrality rewrites if  $x$  is on the street  $h_0$ :

$$C^{sim}(x \in h_0) = \frac{\mu_G(G)}{\sum_{h \in H} \|h\| \cdot d^{top}(h, h_0)} \quad (6.13)$$

where  $d^{topo}$  is the shortest path distance on the hypergraph presented in Ch.5.

### 6.2.3 Display

We represent the street network in  $\mathbb{R}^2$  and associate to each point a color that cods for its centrality. To this we map the centrality using a color-map with  $C$  colours. The correspondence can be **linear** (Fig.7.3) or **adaptive** (Fig.6.4). For the adaptive one, we used a histogram equalization in order to get a more contrasted plot. That way a color represents a proportion  $1/C$  of the city.



**Figure 6.4:** The different centralities studied in this article for two french cities: Avignon (the extended city center) and Troyes. The color mapping is obtained with 5 colours and a histogram equalization so that in each picture a color represent a proportion  $1/C$  of the city.

#### 6.2.4 Comparison of centralities

We see in Fig.6.4 that each centrality produces a different interpretation of the map. A radial one for the closeness, a local one for the straightness and a hierarchical one for the simplest centrality.

For the closeness, we see essentially a radial decrease. For Troyes the maximal centrality corresponds to the city center, for Avignon the maximum is hit in the center of the image rather than in the center of the city (on the top left). This shows closeness depends highly on the chosen borders of the map (boundary effect).

That would be the result obtained by replacing the shortest path distance by an Euclidian distance and calculating the centrality on a homogeneous grid. This confirms the observation made in Sec.5.4: in "normal" cities, the layout is rather dense and homogeneous leading to  $d^e \simeq d^{sp}$ .

The straightness appears to be boundary effect free. In the case of Avignon, it finds correctly the center but for Troyes, but the overall impression is rather disorganized, this centrality does not provide a global coherent vision of the city.

On the contrary, the simplest centrality we proposed provides a hierarchical view of the city. Both in Avignon and in Troyes, it reveals the main historical or modern axis.

On each map of Avignon, a zone is particularly striking (right) for it produces a discontinuity in the overall variation of the centralities. The centralities are able to put in light ill-deserved zones. Troyes is quite homogeneous so a centrality does not reveal much but Avignon is a

composite city and the centrality analysis is in accord with reality (this will be discussed more deeply in the next chapter). The main axis are the old city walls, a high way and radial axis. The ill-deserved zone detected on the right is a residential area, voluntary isolated and the one on the bottom is a poor area known by town planners.

### 6.2.5 Iterated centrality

One of the problems of simplest centrality is its sensibility to roundabouts that break crossing street into two segments. To get around this difficulty, one can iterate the computation of centralities, using the previous centrality result instead of the street length. To this we have to put aside the continuous representation of the city and stick to the topological representation of the hypergraph street network.

Imagine at step  $n$  a centrality vector  $\vec{C}^{(n)}$  gathers centralities of streets. We recall that  $\mathbf{H}^*$  is the matrix whose component  $\mathbf{H}_{ij}^*$  is the topological distance between streets  $i$  and  $j$ . Then if we want to replace in Eq.6.12 the length of streets by their current centrality  $\vec{C}^{(n)}$ , it should look like:

$$\frac{1}{C_i^{(n+1)}} = \sum_j C_j^{(n)} \mathbf{H}_{ij}^* \quad (6.14)$$

The result is the inverse of a centrality because a centrality should decrease with increasing mean topological distance. Taking the direct and the inverse centrality in the same formula is not practical so we rather introduce the inverse of the topological distance. To avoid divergences, we add 1 to each topological distance. These two constraints lead to consider the matrix  $\tilde{\mathbf{H}}^*$  whose components are  $\tilde{\mathbf{H}}_{ij}^* = 1/(\mathbf{H}_{ij}^* + 1)$  and whose lines are normalized to 1. Then  $\tilde{\mathbf{H}}^*$  is a stochastic matrix, positive and irreducible and the Perron-Frobenius theorem shows that

$$\vec{C}^{(n)} = (\tilde{\mathbf{H}}^*)^n \vec{C}^{(0)} \rightarrow \vec{C}^{(\infty)} \quad \text{with} \quad \tilde{\mathbf{H}}^* \vec{C}^{(\infty)} = \vec{C}^{(\infty)} \quad (6.15)$$

where  $C_\infty$  is the positive eigenvector of  $\tilde{\mathbf{H}}^*$  associated to the eigenvalue 1 and whose sum of coefficients is 1. Iterating this way we obtain a stable centrality  $C_\infty$  which is more robust to structures like roundabouts that parasite our analyse.

## 6.3 Discussion

### 6.3.1 Justifications for simplest centrality

To justify simplest centrality we could use an a posteriori argument: it gives a great **interpretation of maps** that both puts in light obvious main axes and reveals some hidden ill deserved areas.

We can also remark that this centrality is **pleasant** and easy to interpret since it is constant on coherent lines. In fact we have seen that contrary to other centrality it is **boundary effect free and coherent**. It is adapted to the hierarchical and looped structure of cities' maps.

An interesting point is that simplest centrality is **very fast to compute**: if the initial

graph size is  $V$  then getting the hypergraph can be done in  $O(V)$  and all the topological distances on the hypergraph are unweighted distances on a small-world graph that can be computed in  $O(H^2)$  where  $H$  is the number of streets (approximately  $V/10$  in practical cases). Whereas closeness, straightness or betweenness are computed (with more memory demands) in  $O(V^2 \log V)$  with simple structures. With typical map, simplest centrality calculus is 1000 times faster.

We technically remark here that both for simplest and shortest distances, if the time to compute all the path distances from a particular point is  $F(V)$  then the known algorithms to find all the point to point distances is  $VF(V)$ . Thus it is not more efficient to calculate all the distances and then to calculate all the centralities. The storage of distances matrix can lead to memory overflows then we will prefer to compute distances for a single point, its centrality, just store this centrality and redo it for the other points.

A good centrality tells about the distance to travel from a point to the rest of the map. We have seen that using a shortest path distance (closeness) leads to too much spurious effects. However using the simplest distance is a good approximation of the shortest path distance and it is also independent of boundary effects.

The total distance from a point to another along the simplest path is greater than the one along the shortest path. But this distance is a good upper bound of shortest distance since streets are straight and organized to connect hierarchically: **most of the time the (metric) simplest distance is also the shortest!** The problem is to convert our integer simplest distance to a real metric distance. The metric simplest distance is the total distance through the simplest path. From a point  $A$  to a point  $B$  the simplest path pass through street intersections  $a_0 \dots a_n$  with  $n$  the topological simplest distance. It can be written:

$$d(A \rightarrow B) = \sum d(a_i \rightarrow a_{i+1}) \simeq \sum \|a_i - a_{i+1}\| \simeq d^{sim}(A, B).l_0 \quad (6.16)$$

where  $l_0$  is a typical distance. The last approximation is justified by the log normal scaling of street lengths and thus the existence of the expectancy of distances  $\|a_i - a_{i+1}\|$ :  $l_0$  and of its variance. The simplest distance is the number of straight pieces of path we have to go from  $A$  to  $B$  and also the number of terms in the sum. Up to a scaling parameter  $l_0$ , simplest distance is a good approximation of shortest distances.

The log-normal scaling also shows there are very long streets that are main axes. The existence of these main axes crossing the whole map explain the robustness of simplest centrality toward boundary effects contrary to closeness centrality. **In conclusion simplest centrality is both an approximation and a better conditioning of closeness or shortest path centrality.** This surprising property comes from the particularities (heavy tail and finite expectancy) of street length distribution.

### 6.3.2 A new algorithm to find shortest path: The Cab Driver's Algorithm

From the previous remarks, we can build a very simple and brand new algorithm to compute shortest path of two points on the street network.

Shortest path distances in a planar graph are calculated as shortest path distances in a weighted graph.



An origin and a destination point are given. The graph is transformed into a street hypergraph. The simplest path between the streets of these two vertices is found (the hypergraph structure is smallest than the graph structure and above all the hypergraph is unweighted which allows to use the shortest path algorithm for unweighted graphs that is faster, see Ch.4, Algo.6). We write  $a_i$  the successive intersections of streets in the simplest path. The only thing to assess is the various shortest path distances between  $a_{i-1}$  and  $a_i$ . Since these two points are on a same street, and a street is rather straight, the shortest path distance can be approximated by their Euclidian distance:  $\|a_{i-1} - a_i\|$ . But it is possible to be more precise taking into account the curvature of the portion of street joining  $a_{i-1}$  and  $a_i$  while keeping the same complexity order. The portion of street between  $a_{i-1}$  and  $a_i$  is a weighted graph but not branched and it is in fact very easy to adapt Algo.6 for distances in unweighted graphs to a weighted graph that has no branches.

---

**Algorithm 9** Fast Algorithm to find shortest paths on a city map

---

- 1: INPUT Graph  $G = (V, E)$ , Vertices  $v, w$
  - 2: Compute Hypergraph  $H$
  - 3: Compute Intersection Matrix of  $H$ :  $\tilde{H}$
  - 4:  $v \in h_v, w \in h_w$
  - 5: Compute Path, simplest path between  $h_v$  and  $h_w$ :  $a_0, \dots, a_n$
  - 6: Compute distance between  $v$  and  $w = \sum \|a_i - a_{i+1}\|$  (with Euclidian distances or shortest path distances in a weighted line graph)
- 

Once the hypergraph computed, this algorithm is simply a shortest path algorithm on an unweighted graph: the path between two vertices can be found in  $O(H)$  at worst and since the graph is a small-world, it can be done in  $O(\log H)$  in average. The hypergraph is computed in  $O(V)$ . This is to be compared to the complexity of the best known algorithm: Dijkstra's which runs in  $O(V \log V)$  with the suitable data structures.

Using a slight modification of Alg.6 it is possible to find all the simplest paths from a street to another with the same time complexity and then to choose the shortest simplest path.

### 6.3.3 Conclusion

We have defined new scalar parameters such organic coefficient, anisotropy, compactness to describe a map or a part of a map... We will reuse them with benefits in Ch.10 to cluster large maps. We could propose an automatic classification of parts of cities using these parameters and PCA. Indeed a whole city is composed of parts of different characteristics. Thus parameters have not a strong meaning globally and it would more interesting to seek out to classify extractions of various cities.

We have introduced the simplest distance which permits to define a new centrality (coherently with our continuous framework), the simplest centrality which is very significant providing a global, hierarchical view of a map. The use of this centrality in practical town-planning cases is discussed in the next chapter.

The hallmarks of the city tested in Ch.5 allow to show that the simplest path distance is a good approximation of the shortest path distance and thus to propose and prove the validity

of a brand new algorithm: the Cab Driver Algorithm. It computes in a logarithmic time the shortest path between two points in the map by approximating it by the simplest path.



# Analyzing maps and planning with the Simplest Centrality

## Contents

---

7.1	Introduction . . . . .	<b>113</b>
7.2	Avignon: testing the robustness of simplest centrality toward windowing	<b>113</b>
7.3	Villers-Sur-Mer: testing town-planning scenarios . . . . .	<b>117</b>
7.4	Conclusion: comparison with Space Syntax and "dual" (street) graph . .	<b>117</b>

---

**—Analysing maps and planning with the Simplest Centrality : Synthesis —**

In this chapter we apply and test the simplest centrality in two real town-planning problems.

At first we study at various scales the city of Avignon and show this centrality is meaningful at various scales, allows to find principal transportation axis and diagnose ill-deserved area. The use of an adaptive color scale to display centrality is necessary.

Then we test several scenarios to open up a new district to the whole city in Villers-Sur-Mer.

## 7.1 Introduction

With a team of social scientists and town planners, we have submitted to the urban office of a few cities our analysis of their city. The question was to know if the simplest centrality rendering makes sense to urban actors (recovering important streets and ill deserved zones), if added to that it permits to put in light some dysfunctions in the network that are not trivial and if it can be used to simulate the impact of urban modifications, help to test scenarios to relieve congestion or to open up isolated area. The first study we present is a multi scale analysis of Avignon ; in the second, we observe an ill deserved zone in Villers-Sur-Mer and compare three simple scenarios to homogenize the city. Town planners have received our work with enthusiasm.

## 7.2 Avignon: testing the robustness of simplest centrality toward windowing

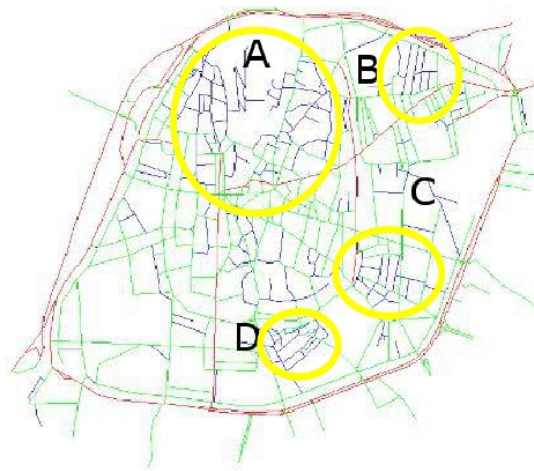
Avignon is a large city in the south of France, born into the hollow of the river Le Rhône with the Durance river. It dates back to the prehistory but the current city is built on the basis of the medieval city. The medieval walls have been progressively engulfed, leaving tracks as a belt surrounding the medieval center of the city. We have studied the city on three different scales to assess of the robustness of the simplest centrality Fig.7.1: the historic city center delimited by a remain of a city wall (1), a qualitatively dense area (2) and the whole city area (3).



**Figure 7.1:** The city of Avignon and its neighbourhood. We have studied it on three scales: the historic city center (1), the dense area (2) and the whole city (3)

The intra-muros city center is rather homogeneous with distributed resources. We have

plotted it with three colours (Fig.7.2) that allow to distinguish between main axis, and side streets, leaving a middle category for roads in between. The found main axis correspond to the boulevard cycling around the medieval wall. It also recovers the curved street east to center which is an important medieval access road from the nearby big city. It finds also the North to South new XIXth century road. These two roads are the main two axes for public transport. Other small variations in the order of the streets turned out to be very significant. It first finds the historical mentions around the Papal Palace (A in Fig.7.2). The other three lower (blue) color zones (B, C and D in Fig.7.2) correspond to burrows with small streets not very well deserved. The town-planners informed us that these burrows where under scrutiny for their bad state and their rehabilitation is now planned.



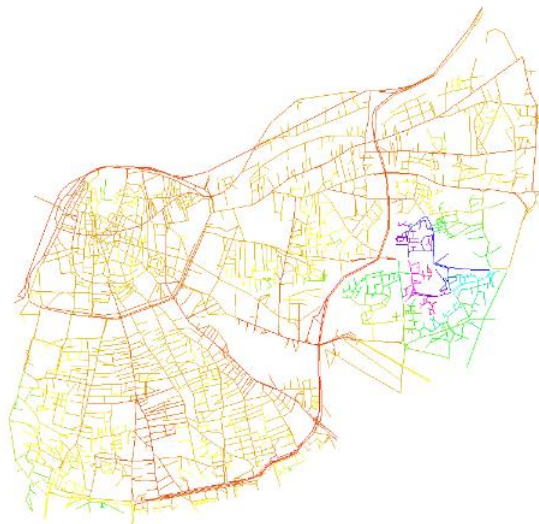
**Figure 7.2:** The centrality map of the center of Avignon with three colours which represent high central street (red), medium (green) and low (blue). It is to remark that the notion of "low" centrality is relative to this particular map, the center of Avignon is nearly homogeneous. Apart from the historical Papal Palace and the bishops mansions (noted A), we have three more homogeneous blue places (B, C and D) which are precisely the burrows where rehabilitation is due. It also underlined an essential middle age curved road, from right to center, that is not considered by people even if it is still essential for transportation.

The result for the dense area Fig.7.3 and Fig.7.4 is more subtle. Fig.7.3 represents the centrality map with a linear scale. Almost all the map is of the same orange color, just an area in the east differentiates by a drop of centrality. This could be disappointing for the defence of simplest centrality. But we plotted the same centrality with an adaptive scale of colours in Fig.7.4 (see 6.2.3) and the new created dynamics reveals structures. We first have to remark that the wall remains an essential structure at this scale even if intra-muros main streets seem to vanish at this resolution of observation. In fact a characteristic of the adaptive color scale is that if one zooms on the city center on the second map and rescales the colours, the main local roads will re-appear (to a certain extent).

At this scale, we find again obvious main axis: the city boulevard and the modern surrounding highway running east and south at first and secondarily radial streets. Two zones in pink are emphasized as are ill deserved zones. The east one (Fig.7.4-A) with extreme

values was at the origin of the bad conditioning of the linear color scale. The other part (Fig.7.4-B) surrounded by the modern highway does not have the same efficient structure as its surroundings. Interestingly these two zones are ill-deserved for two opposite reasons. The west zone (Fig.7.4-B) is a poor district with security problems. The east zone (Fig.7.4-A) on the contrary is a residential district that badly connects to the network on purpose, to keep the streets quiet. It is almost a tree, well-known for the inhabitants for the labyrinthine impression it makes.

It is one of the surprising lessons of this case study, that ill-deserved zones are not necessary poor but could also be middle-class neighbourhoods researched for their tranquillity. The other zones around A and B outside the city center has also a different structure that corresponds to different histories: the north-east one is more recent and gained on industrial lands, the south one is older and gained in the XIXth century on agricultural lands. In this last part we can even distinguish the part closer to the city, older with a better centrality than the part near to the modern highway corresponding to lower class buildings.

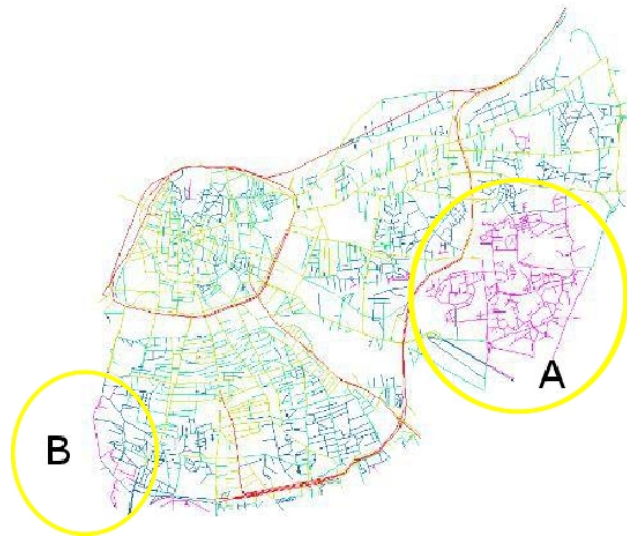


**Figure 7.3:** The simplest centrality of dense area of Avignon with a linear color scale.

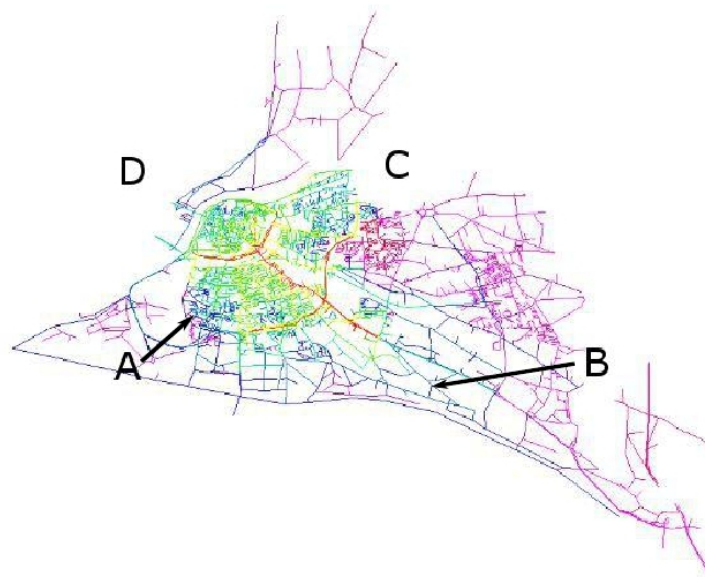
At last, the whole city area centrality Fig.7.5 reflects the notion of urban continuity. The centrality decreases with the distance to the center but this decrease is smooth thanks to the hierarchy of important streets. A historical radial street linking the former wall and the main boulevard to an outside town appears to have a more important role at the global scale (Fig.7.5) than at the dense area scale (Fig.7.3). We thus see it as a high level hub around which the industries were first built. It shows its importance to link different parts of the city and the city to the outside but has not really a local function.

The comparison of the centrality of Avignon at three different scales shows that roughly the centrality is stable and does not depend on the map borders: we always find the same main boulevards and access roads. At the second order however we find differences which shows how the main elements function differently at different scales.





**Figure 7.4:** The simplest centrality of dense area of Avignon with an adaptative color scale. One place comes out, noted A. This very ill-deserved place is well known as a labyrinth where only local people venture. It is in fact a well-considered middle class villa settling, local people enjoying the total absence of through traffic. On B, the place is not so well deserved, at the border and surrounded by highway, and this is the poor area with security problems.



**Figure 7.5:** The simplest centrality of the whole city of Avignon with an adaptative color scale. Town planners still despair on the choice of A as the pace for the new TGV train station, on a ill-deserved place, (blocked by the two rivers), while B would have been traditionally irrigated, C a place with high development potential. The actual booming place is in D (proposed for the station), also on the main circulation axes, but away from already constructed area which raised administrative and political problems.

### 7.3 Villers-Sur-Mer: testing town-planning scenarios

For an other test, we chose the city of Villers-Sur-Mer.

Villers-Sur-Mer is a small seaside city in the North of France.

Since we have seen in Ch.5 that Villers is a centralized city (it is necessary to pass through the center to go from a point to an other), we have plotted the topological distance of streets to the center rather than the centrality as in Avignon's case. Fig.7.6 (up-left) is the result for the actual city.

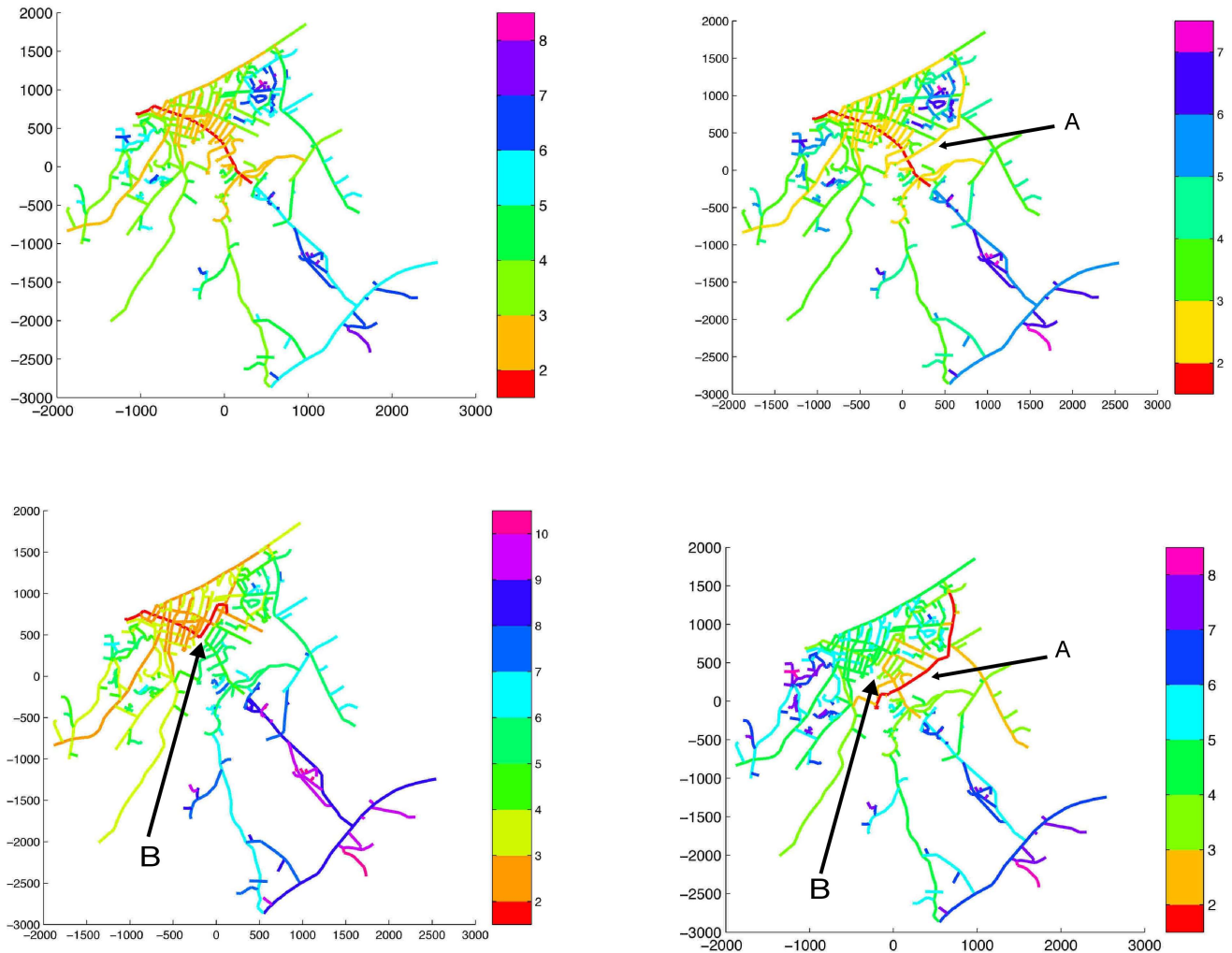
What surprised us at first was that the streets are not perpendicular to the sea. This was contradictory with our idea of structuring elements. But historically the city has not been built to be a seaside resort but as a village inside the ground, connected to a small port protected by the cliffs. The city expanded around the west rocks *Les Vaches noires*, people came to paint and the path to these rocks was the primordial axis. It is actually what the analyse found as the city center. This axis defines the direction of streets, the secondary ones being simply perpendicular to it. It is the local structuring element, kind of backbone of the city with its aside ribs.

Then the city developed from west to east and became lately a seaside resort, which makes streets perpendicular to the sea. The recent area on the east is relatively badly deserved, its topological distance to the center going from 5 to 8 whereas the average for the city is about 3. It contains a commercial center *Villers 2000* that it would be interesting to irrigate. We thus proposed three scenarios to modify slightly the city to reduce the topological distance of this area or at least make it enter into the average topological distance. Scenario 1, Fig.7.6 (up-right): build a belt around the area. Scenario 2, Fig.7.6 (bottom-left): cut the main road into to pieces to have people make a detour in their displacements and then produce new attractions. Scenario 3, Fig.7.6 (bottom-right): combine the two previous scenarios.

The addition of a boulevard is practically inefficient: it decreases the global radius of 1 but the commercial area stays on the margins. Interestingly, cutting the main road makes obviously the radius and the mean distance increase but the commercial area becomes an average well deserved zone: we want here to interpret centrality as a stationary flow of a random walk. People cannot use the main street efficiently, by cutting it we make them use auxiliary paths and thus they have to visit the commercial area that becomes artificially a popular place. The combination of the two scenarios provides the same radius as the original city but the distribution of centrality is more centred around the average. The boulevard belt becomes the city center, the old center is split into two average streets and the commercial area is as well deserved as the seaside or the access to rocks.

### 7.4 Conclusion: comparison with Space Syntax and "dual" (street) graph

At first sight our approach of topological distance and simplest centrality is very similar to the Space Syntax framework [66]. They analyse the city into continuity lines, which are axis on which an agent has a total visibility. Then they define local and global indices to measure the quality of the integration of each line. They justify the relevance of their analysis by



**Figure 7.6:** The distance to the center of streets in Villers-Sur-Mer (up-left). To open up the new east commercial region we have tested the impact of three scenarios. The addition of a boulevard or belt around this area (up-right), the cut of the main road to have people by-pass this hole through the commercial region (down-left) and the combination of the two previous scenarios (bottom-right).

showing their measure is strongly correlated with the pedestrian perception of space and choice of orientation [68]. As said in Ch.1, the main inconsistency of space syntax is very sensitive to a small variation of the road network. This inconsistency have been overcome by replacing the notion of line by that of streets (with different definitions) for instance in [73]. In a nutshell space syntax and its by-product divide space into objects and measure their integration by counting a topological distance between them. They justify themselves by a human perception argument.

Contrary to space syntax, we present a coherent and synthetic framework. Structures and measure are clearly and well defined. We have shown that hypergraph can be computed only from the topological structure of the map and this linearly in time and independently of any arbitrary parameter. We make the link between local measure and hypergraph measure

with the simplest distance which approach the simplest distance. We show that one can pass from a continuous graph to a simply topological structure the hypergraph by replacing the simplest distance by a weighted topological distance. This is a good approximation highly justified by the optimization of computational time. We do use a mathematical argument to prove the relevance of our measure rather than a human perception one. To plot centrality we used an adaptive color scale which is the same than to consider only the ranking of streets. Thus our centrality has both a local and a global meaning, a zoom on a map and a rescaling of colours defining a centrality map.

We must admit nonetheless that [105] reaches almost the same results as us. They work on the notion of "dual graph" i.e. the graph associated to street network that can be obtained by streets' name or some other algorithm. Then they compute classical centralities on this graph. We have a robust method to identify streets and our simplest centrality is somehow the closeness centrality on the "dual graph". Indeed, our weighting by lengths permits to have a physical interpretation of the centrality but the result is almost the same. This can be justified by the existence of  $C_\infty$  (Ch.6, iterated centrality) and the exponential convergence of Eq.9.2. The simplest centrality is the first iterated centrality,  $C^{(1)}$  when  $C^{(0)}$  is the length proportion of streets while the closeness on the "dual graph" is  $C^{(1)}$  obtained when  $C^{(0)} = \frac{1}{\#H} (1, \dots, 1)$ . Starting from close starting points and converging toward the same stationary point, it thus reasonable that the first iterated are close.

Maybe the visual representation of the "dual centrality" in [105] is not convincing and so they did not seek to extend their investigation. The reason might just be they used a linear scale of colours.



# Models

## Contents

---

8.1	A continuous models that came last but we present first . . . . .	<b>123</b>
8.1.1	Principles and formulation . . . . .	123
8.1.2	Solutions . . . . .	124
8.1.3	Simulations . . . . .	124
8.1.4	Saturation effect . . . . .	126
8.2	Simple network models . . . . .	<b>127</b>
8.2.1	Square lattice . . . . .	127
8.2.2	Division of space . . . . .	127
8.2.3	Division - extension of space . . . . .	129
8.3	Morphogenesis of the City . . . . .	<b>130</b>
8.3.1	Potential field . . . . .	132
8.3.2	Connection . . . . .	133
8.3.3	Parameters . . . . .	134
8.3.4	A few simulations . . . . .	135
8.4	Object Morphogenesis of the City . . . . .	<b>138</b>
8.5	Conclusion . . . . .	<b>141</b>

---

### —Models: Synthesis—

In the previous chapters, we have quantitatively studied cities, tried to describe them and to show that in spite of a great apparent variability of shapes, cities' maps exhibit common hallmarks. In this chapter we present models that reproduce and thus can explain these signatures. The main idea of these models is that in the **organic case**, the city's layout is determined by a process of **extension and division of space**.

The first model we present consider the city as **a continuum of wealth concentrations** that interact and evolve. This model implies that the local growth of a city is globally **exponential** and that it is the same asymptotically locally. This model predicts an exponential movement of the "border" of a city and **a power law distribution of wealth** through space.

The other models are **dynamic planar graphs**. The first is simply a division process which reproduces the "small world" effect on street graph. The second adds an extension component that keeps the small world effect and explains the statistics of blocks' sizes and the logarithmic scaling of street lengths.

The last two models are more intricate and do not allow explicit calculus. It implements more realistically the idea of extension division of space with a few parameters that allow to reproduce a huge diversity of networks while keeping the main features. We named it "**Morphogenesis of the city**".

### —Contributions of this chapter—

1. An evolution model for the density of spatial resources.
2. Numerical simulation of this model in the monocentric case.
3. Two very simple models of division and extension / division of space that allow calculus to exhibit features: small-world effect and distribution of blocks' area.
4. The morphogenesis model is a flexible framework that reproduce main trends and diversity. It can easily used as a basis to incorporate refinements: diversity of agents, anisotropy of space...

## 8.1 A continuous models that came last but we present first

### 8.1.1 Principles and formulation

We built the model bellow to answer the question "at what speed does a city's boundary move?" Indeed in the division / extension model we present further in this chapter, we had to assume an exponential growth of the borders to reproduce blocks' area distribution. This assumption seemed surprising. We did not find a clear answer in the literature or in discussions with experts of the city. We thus tried to build a justification.

This model is a spatial translation of the Zipf's principle. As a scale free function the distribution of city sizes leads to think that in a closed city system the growth of a particular city has a multiplicative logic (see Ch.1) or App.B). We extend this principle to every piece of the city. A particular place produces wealth proportionally to its current wealth. The total wealth of the city is then growing as an exponential. The question is to model how the created wealth of a particular place dispatches in space.

We define a field  $f(\vec{x}, t)$  which represents the density of "wealth" in the city at place  $\vec{x}$  at time  $t$ . We consider every point in space has an influence on the whole space: a point  $\vec{y}$  of wealth  $f(\vec{y}, t)$  at  $t$  creates a "wealth quantity" proportional to  $f(\vec{y}, t)$  and which is dispatched on each point  $\vec{x}$  of space following a decreasing function (or "kernel") of  $||\vec{x} - \vec{y}||$  written  $K$ . Thus the wealth at each point  $\vec{x}$  evolves at time  $t$  with a time derivative:

$$\frac{\partial f(\vec{x}, t)}{\partial t} = \lambda \cdot \text{Cste} \int_{\mathbb{R}^2} f(\vec{y}, t) K(||\vec{x} - \vec{y}||) d\vec{y} \quad (8.1)$$

with  $\text{Cste} \int \int K(||\vec{x} - \vec{y}||) d\vec{x} d\vec{y} = 1$  and  $\lambda \geq 0$ .

Let  $I(t)$  be the total wealth of the city at time  $t$ :

$$I(t) = \int_{\mathbb{R}^2} f \quad (8.2)$$

The spatial integration of both sides in Eq.8.1, the permutation integral - derivative on the left and use of Fubini's theorem combined with the normalization of  $K$  on the right permit to show that  $I$  is an exponential of time:

$$dI = \lambda I dt \implies I(t) = I_0 \exp \lambda t \quad (8.3)$$

$\lambda$  is the growth factor of the city typically of 1 percent per year. The exponential growth of the whole city is coherent with Zipf's law. The question our model tries to answer is how is this growth dispatched through space?

An infinite number of kernels can be chosen, the conditions being their positivity and integrability. In what follow, we study two particular kernels. The Newtonian Kernel:

$$\frac{\partial f(\vec{x}, t)}{\partial t} = \lambda \cdot \text{Cste} \int_{\mathbb{R}^2} \frac{f(\vec{y}, t)}{1 + \gamma \cdot ||\vec{x} - \vec{y}||^n} d\vec{y} \quad (8.4)$$

where  $n > 2$  to ensure integrability and Cste is a normalization constant such as the integral of the kernel is 1:

$$\text{Cste} = \frac{2\pi^2}{n \sqrt[n]{\gamma} \sin(2\pi/n)} \quad (8.5)$$



Newtonian Kernels describe a long range influence. Conversely the Exponential Kernel  $\exp(-\gamma||\vec{y}||)$  normalized by the constant:  $2\pi/\gamma^2$  will be our test model for small range influences.

### 8.1.2 Solutions

We want to solve Eq.8.1 in the case of a "**monocentric and radial city**", for the two kinds of toy kernels: Newtonian and Exponential. If  $f(., 0)$  is radial and if there is no perturbations in the evolution of the system then  $f(\vec{x}, t)$  is radial and can write:  $\tilde{f}(r, t)$ . Eq.8.1 writes as a convolution:

$$\frac{\partial f(\vec{x}, t)}{\partial t} = \lambda f * K \quad (8.6)$$

The equation "could" be easily solved by using Fourier Transform:

$$\tilde{f}(r, t) = \int \bar{r} J_0(r\bar{r}) e^{(2\pi t \int \bar{r} J_0(\bar{r}\bar{r}) K(\bar{r}) d\bar{r})} d\bar{r} \quad (8.7)$$

$$J_n(x) = \frac{1}{\pi} \int_0^\pi \cos(n\tau - x \sin \tau) d\tau \quad (\text{Bessel functions}) \quad (8.8)$$

Nonetheless Eq.8.7 is not easy to simplify and it is not convenient to use in numerical simulations. Indeed  $J_0$  oscillates a lot and can be negative. Parametric integrals involved depend on scales factors ( $\bar{r}$  and  $\bar{r}$ ) which makes the approximation toward infinity unstable. This is the general problem of Fourier transform that are always oscillating and reach a constant (for instance zero) value only by the mean of infinite destructive interferences.

Nonetheless, Eq.8.7 permits to show that the solution at time  $t$  is continuous with respect of conditions at time 0: the initial city can be represented by a Dirac. We then opt for a numerical simulation of the evolution and to this use the following positive and robust equation:

$$\frac{\partial \tilde{f}(r_0, t)}{\partial t} = \int r \tilde{f}(r, t) \left[ \int_0^{2\pi} \bar{K}(r, r_0, \theta) d\theta \right] dr \quad (8.9)$$

with

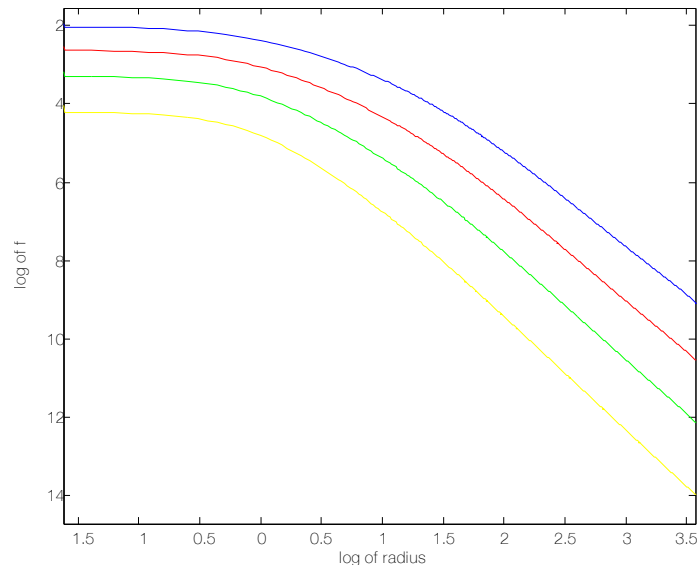
$$\bar{K}(r, r_0, \theta) = K \left( \sqrt{r^2 + r_0^2 - 2rr_0 \cdot \cos \theta} \right) \quad (8.10)$$

### 8.1.3 Simulations

We present here the observed result of numerical simulations with Eq.8.9 for a Newtonian Kernel.

The profile distribution at time  $t$  is an inverse power function whose asymptotic exponent  $-\gamma(t)$  is not constant but increases with time: the profile tends to be smoother. Fig.8.1 shows the evolution of the profile through time.

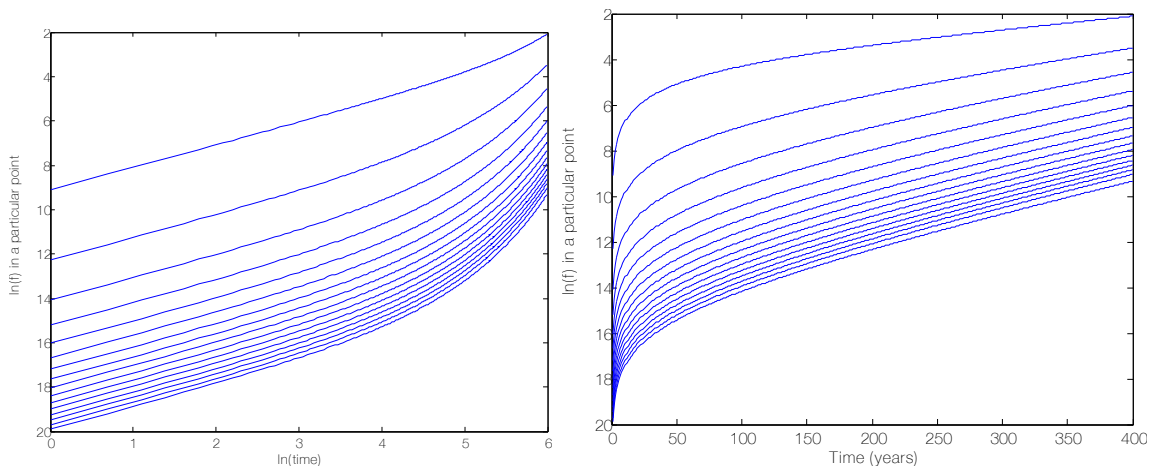
This result is not in agreement with [40] which observes in the 50's (and thus with limited measures tools) that the density profile of several cities around the world is exponential. It would be worth to re-do the measurements with modern tools. The fittings in [40] are based on few points, the density at a distance  $r$  being the average of densities in a crown centered



**Figure 8.1:** Log-log plot of the density profile of a city with 1 percent of annual growth after 100, 200, 300 and 400 years of evolution. The profile is power-law shaped: when far from the center,  $f(r, t) \sim_{\infty} A(t)/r^{\gamma(t)}$  with  $\gamma$  decreasing through time. This resulting curve has the same shape than the initial kernel.

to the city core (maybe difficult to define in a practical context). We found nonetheless a coherent result: the density gradient vanishes with time.

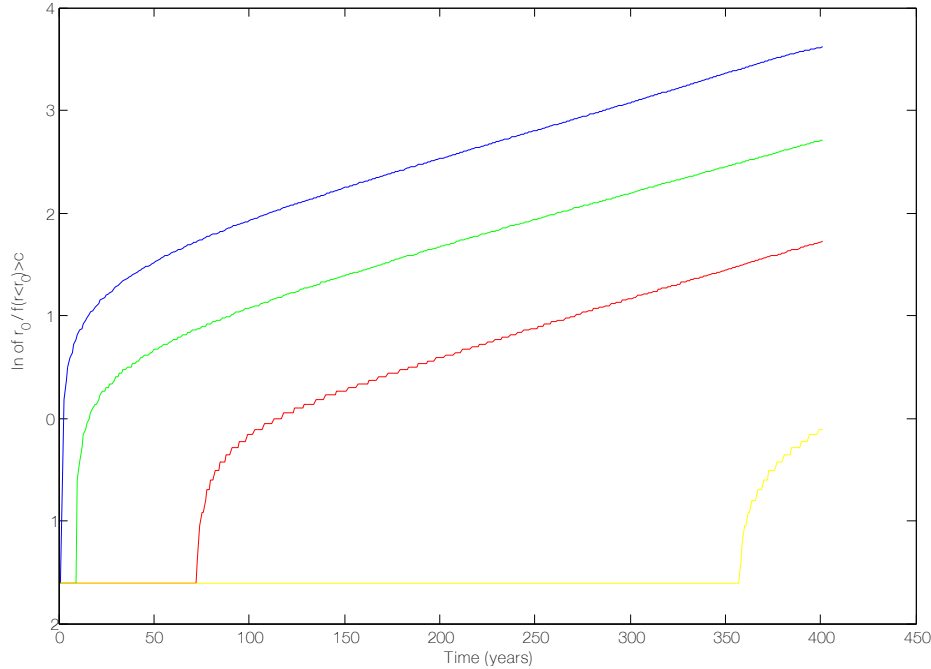
Let us consider a fixed point in space and look at the evolution of wealth through time in this point. Fig.8.2 shows that there are two regimes of growth: at first the local wealth is a power function (Fig.8.2, left), then it transform to an exponential function (Fig.8.2, right). Ultimately the local growth in the Newtonian city is exponential.



**Figure 8.2:** Evolution of the density observed in a particular fixed set of points through time. Points are regularly spaced. The local growth is at first a power function (line in the log-log plot on the left) and then reaches an exponential scheme (line in the linear-log plot on the right).

At last we want to assess the celerity of city's border move. The first difficulty is to define the notion of border for a city. We fix an arbitrary threshold  $c > 0$  and look at the evolution

of  $\tilde{f}^{-1}(c, \cdot)$ . The results for various  $c$  are reported in Fig.8.3. In fact whatever the arbitrary choice of  $c$ , the evolution of the "arbitrary" radius of the city ends up to be exponential. In consequence, this model implies that the "wave front" of a city has an exponential celerity.



**Figure 8.3:** Distance  $r_0$  at which the density profile reaches different arbitrary  $c$  values in a linear-log plot. For all values of threshold  $c$ , the distance ends up to be an exponential function of time: the border of the city growth exponentially whatever is the arbitrary definition of border.

If the kernel  $K$  is not an inverse power function but an **exponential**, the local growth remains a power law followed by an exponential (Fig.8.2), the border of the city still drifts away exponentially (Fig.8.3) but the general profile is no more a power-law, it is an exponential which is in agreement with [40].

In conclusion it seems that whatever the profile, the border of the city always drifts away exponentially and the local growth starts as a power-law to ends up as an exponential. On the contrary, the density profile seems to reflect the shape of the kernel.

### 8.1.4 Saturation effect

In this model, the density tends to  $\infty$  locally. To avoid this we modify the evolution equation to get a "saturation effect". The total growth of the city has to remain proportional to the total wealth of the city. But if in a particular point an upper bound is reached the growth in this point is set to 0 and the wealth that should have been created in this point is re-distributed in the rest of the city. Thus:

$$\frac{\partial f(\vec{x}, t)}{\partial t} = \frac{\Lambda(f(\vec{x}, t))}{\int_{\mathbb{R}^2} \Lambda(f(\vec{y}, t)) f(\vec{y}, t) d\vec{y}} \int_{\mathbb{R}^2} f(\vec{y}, t) K(\|\vec{x} - \vec{y}\|) d\vec{y} \quad (8.11)$$

with

$$\Lambda(z) = \begin{cases} \lambda & \text{if } z \leq c \\ 0 & \text{otherwise} \end{cases} \quad (8.12)$$

keeps the exponential evolution of the whole system, wealth is distributed spatially as in Eq.8.1 and resources created in saturated regions are re-distributed in the rest of the city with respect of the distance. This equation creates naturally a saturated "city center" whose boundary evolves exponentially. The main properties of Eq.8.1 are maintained: the density profile is an inverse power function (whose exponent increase faster) if the kernel is a power-function, an exponential if the kernel is exponential ; the local growth is a power function then an exponential ; the integral of  $f$  is an exponential.

## 8.2 Simple network models

From now on we will get particularly interested in the network growth of the city, and present a few simple models that reproduce street network features. We will show that the previous result of exponential spatial growth of the city (according to our continuous model) is coherent with a network model of division / extension of space that reproduces block area distribution.

### 8.2.1 Square lattice

Symbol of the planned city, the square lattice is a "very small world": its mean topological distance between streets tends to a constant (1.5) when the network densifies. This network exhibits a peaked cell area distribution and all the same for the street length distribution.

### 8.2.2 Division of space

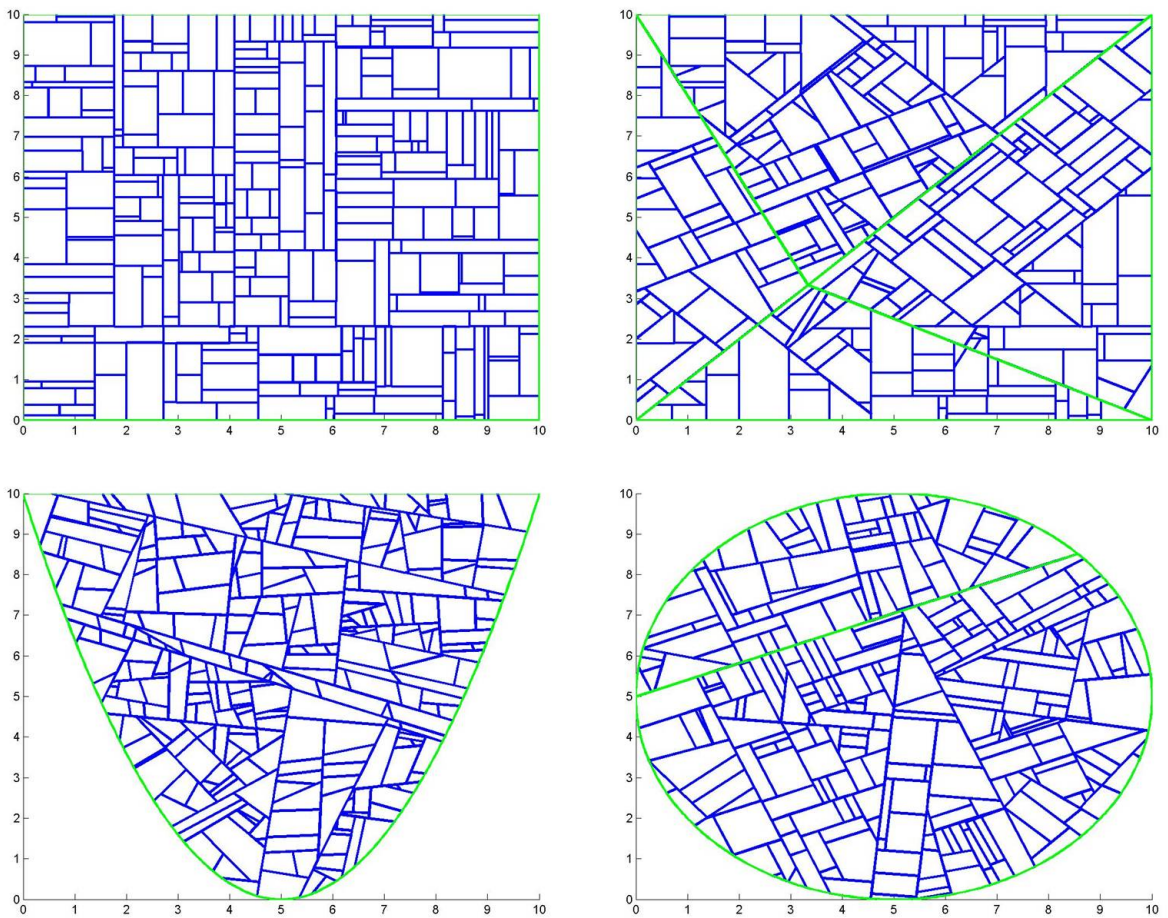
We now propose coherently with the observation in Fig.8.2 of a locally exponential growth of infrastructures to study a space division model and look if it recovers the main "street" properties presented in Ch.5: cell area distribution, street length scaling and small world behaviour in the street hypergraph.

Several methods to define space division processes are known in mathematics ([95, 41], Ch.9). Basically, they place in a bounded window, and divide sequentially this window. They divide it a first time, then they apply a measure (perimeter, area...) to each resulting cell and divide by a random line at the next step a cell with a probability proportional to its measure.

To adapt this to the city problem and for the sake of simplicity, we chose a window and a number of iterations. At each iteration a cell is chosen at random with a probability proportional to its area. This is the same as to pick at each iteration a uniformly distributed random point in the window and divide the cell it lays on. To adapt to local geometry, the cutting segment is not taken with respect of an angular distribution but is perpendicular to the edge of the initial cell to which the point is the closest.

This choice in the orientation of new segments implements the influence of previous (structuring) elements. The very first structuring elements (Ch.2) are drawn in green in Fig.8.4 where different configurations are imposed at the beginning.

To carry out rough calculus we simplify this model by making it "almost deterministic": the city evolves sequentially, the initial window is a square of area  $W$ , at each step each cell is divided in two by a segment almost at its center but not exactly just to ensure inter-cell segments are not coherent. Horizontal and vertical segments are alternated. We consider the resulting structure after  $n$  iterations.



**Figure 8.4:** Division of space easily adapts to structuring elements. For each picture, we have defined a window with primordial curves (in green). Then we have simulated a division process on this window. A cell is divided with a probability proportional to its area and a new line is always orthogonal to one of the sides of the cell it appears on.

The number of cells is  $K_n = 2^n$ . If  $L$  is the length of the side of the initial square, the total length of streets after  $n$  iterations is  $L_n = 2L(1 + 2\sqrt{K_n})$ . Since cuttings are made almost in the center of cells, the cell area distribution is still peaked around  $W/K_n$

We can find by recursive arguments that the maximal topological distance from the border

of the window is:

$$r_n = \left\lfloor \frac{\log_2 K_n - 1}{3} \right\rfloor + 1 \quad (8.13)$$

Since  $L_n$  is a power function of  $K_n$ ,  $r_n$  is a logarithmic function of  $L_n$ . And the diameter of the network  $d_n$  is such as  $d_n \leq 2r_n$  (see sec.5.3). Consequently the average topological distance in the system  $\alpha_n$  is between two logarithmic functions of the size of the city, we consider it is a logarithmic function of the size of the city.

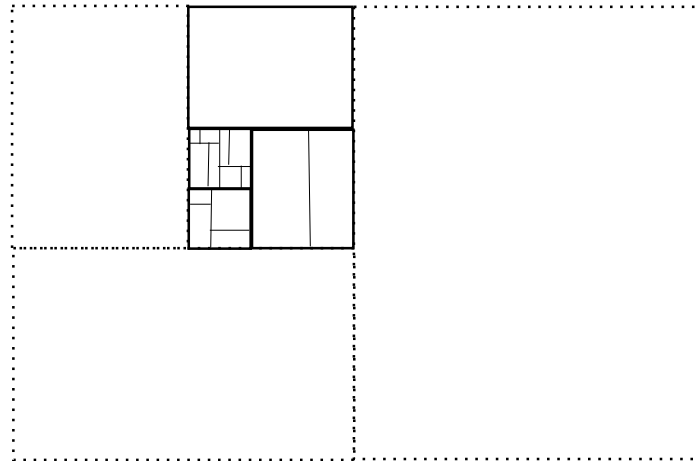
### 8.2.3 Division - extension of space

The division model is not sufficient to explain the distribution of cells area ( $\propto 1/A^2$ , [79]). To that purpose we want to test the additional effect of the growth of the city.

In a first time we implemented this idea: the initial city is a square of area  $W$ , at each step every cell in the city is divided in two parts and a new square of area  $W$  is added to the system at the same time (for instance successive squares are aligned from left to right). The distribution of cell area after  $n$  step was then  $\propto 1/A...$

This is close to the target but an hypothesis had to be changed. We were led to consider that successive squares have size increasing as an exponential. It astonishingly turned out to be a necessary hypothesis to get the good result. As said above, it was not possible to get data to measure the effective growth of cities and we did not find literature on this point. That is why we built the previous continuous model that corroborate this hypothesis.

Thus we want to build a model that implement division of space with exponential growth of surface.



**Figure 8.5:** Simple model for the division / extension of space growth of a city network. 1) A square is built 2) this square is divided in two piece, a new square of the same size is added at its bottom. Three cells result. 3) Each cell is divided in two with alternated horizontal and vertical segments; a new square is added on the right of the two first squares etc. At each step existing cells are divided and a new square is added following a logarithmic spiral.

For instance we consider the city is at first a square of side length  $L$ . We juxtapose to this square another square of side  $L$  and successively at step  $n$  we add a square of side  $u_{n-1} + u_{n-2}$  as in Fig.8.5. The sequence of squares creates a particular case of logarithmic

spiral. Roughly, side of square  $n$  is of side  $\phi^n.L$  with  $\phi$  the golden number.

This process only shows it is possible to pave space with squares of geometrically growing size. At each step every existing cell is also divided in two roughly similar cells. Then at step  $n$  there are  $n$  squares, the first is divided into  $2^n$  cells of area  $L^2/2^n\dots$  and the  $m$ -th with  $m \leq n$  is divided into  $2^{n-m}$  cells of area  $\phi^{2m}L^2/2^{n-m}$ . There are in whole  $2^{n+1} - 1 \simeq 2^{n+1}$  cells.

The area of the cells of  $m$ -th square is

$$A_m^{(n)} = \frac{L^2 \cdot (\phi^2)^m}{2^{n-m}} \quad (8.14)$$

and the frequency of this area is:

$$f_m^{(n)} \simeq \frac{2^{n-m}}{2^{n+1}} = 2 \cdot 2^{-m} \quad (8.15)$$

If we write  $\theta = \log_2(\phi^2)$ , mixing the two previous equations leads to

$$A_m^{(n)} \simeq \underbrace{\left(2^n \cdot L^2 / 2^{(1-\theta)}\right)}_{C^{(n)}(\mathcal{M})} \left(f_m^{(n)}\right)^{(1-\theta)} \quad (8.16)$$

In other words, after a given number of steps  $n$ , there are  $n$  different area values and there is a power relation between each area and its appearance frequency. With the growth factor  $\phi \simeq 1.62$  we have chosen we have numerically  $f \propto A^{-2.57}$ . This is close to the empirical 2. In fact the choice of  $\phi$  was arbitrary to obtain easily a construction with coherent squares, small deformations of these squares permit to have different growth factors and the factor  $2^{3/4} \simeq 1.68$  quite close to  $\phi$  leads to  $f \propto A^{-2}$

The overall structure created by squares with their increasing side length allows to keep a global small-world effect.

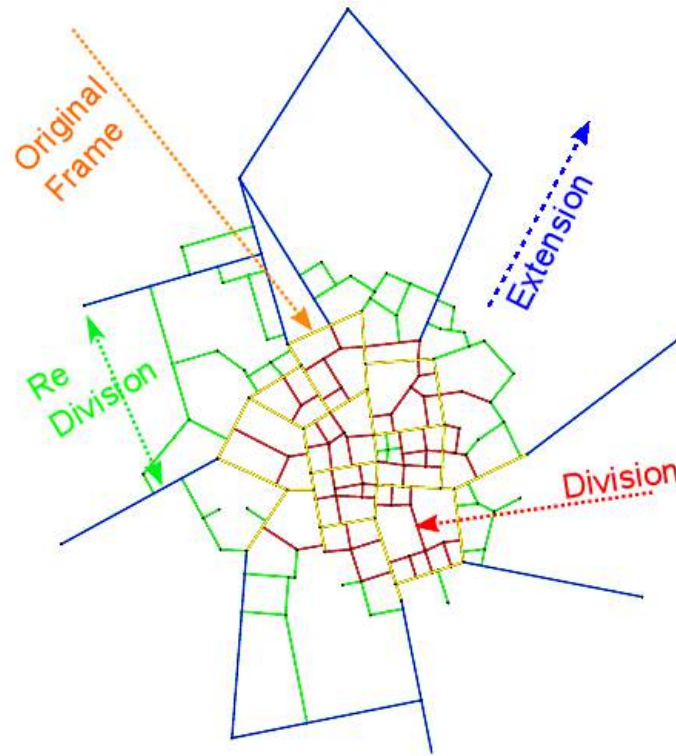
### 8.3 Morphogenesis of the City

This section presents a model of the growth and development of a city. The city is reduced to its streets and we build a dynamical model allowing to add street segments one after another. We want to measure to what extent the condition of the city - of being constituted by segment structures - constraints its development. The main purpose is to translate the idea of division and extension of space in a **general and flexible framework** (Fig.8.6). The morphogenesis is only implemented for **the organic city**: the city is the result of local non-concerted interactions, time can be seen as discrete, at each interaction a new settlement (generic term to designate a commercial infrastructure, a private individual...) is produced to maximize the profit of the new settler. Then the settler sees the space around the city as a **potential field**, each particular location having an economical or geographical quantification of its attractiveness. When he has chosen his place, he connects to the network. The resulting new street segments are considered to be filled fast by other settlers.

Our model is first built on three assumptions, two principles (installation and connection) and a few parameters. The whole gives a coherent and consistent vision of the problem.

We aim at building a model that can reproduce several limit cases of urban growth but also point out continuity between them. We will then introduce parameters to build a continuum between these limit cases.

The principles and parameters we use are meaningful, expressed in an interface language that allows the mathematical and physical communities to exchange with town planners, architects and social scientists. As a dynamical system and a geometrical graph, we will



**Figure 8.6:** Sketch of the division - extension of space idea. From an original frame (in yellow) the city produces division patterns when it locally densifies (in red). Sometimes the city produces a few lines of extension at its outskirts (in blue). And then the global result is re-divided (in green).

see a city as a function  $C : \mathbb{R}_+ \rightarrow \mathcal{G}_C \subset \mathcal{C}_d$  with  $C(t) = \{V(t), E(t)\}$ . Then we make the following postulates on the evolution of  $C$  :

**P1** A city is the result of a sequence of operations occurring at increasing times  $(t_i)_{i \in \mathbb{N}}$  such that :

$$C(t) = C(t_i) \quad \forall t \in [t_i, t_{i+1}[$$

**P2** Infrastructures are conserved:

$$C(t_1) \subseteq C(t_2) \quad \text{if } t_1 \leq t_2$$

**P3** There exist two functions  $P_t$  (price) and  $V_t$  (potentiality) such that the city is a com-



promise between them:

$$C(t + \Delta_t) = \underset{\substack{c \subset C(t) \\ P_t(c - C(t)) \leq 0}}{\operatorname{argmin}} V_t(C(t), c)$$

Functions  $V$  and  $P$  are not obvious to define. They should be in a "microscopic" point of view aggregation of economical parameters. We can avoid developing them if we observe a city's growth is determined by "macroscopic" insights:

**Its construction** The capacity to add new elements to the map and to build new streets.

**Its organization** From a random settlement to a highly structured one.

**Its sprawling** A city has to make a compromise between its inner development and outer growth.

The city  $C(t)$  induces a spatial potential field describing the attractiveness of any point of the available space. A new settlement (either an individual settler or a facility) has its own policy (8.3.3) of choice with respect to this potential. After having chosen its location, it connects to the existing street system by creating one or several new roads. This model explicitly decouples the problems of positioning and of connecting.

### 8.3.1 Potential field

For each point  $x$  in the available space the potential  $P_{C \rightarrow x}$  quantifies to what extent  $x$  is a good choice to locate a new center. This potential should mimic the following ideas:

- A large scale behaviour such that the global attraction of a part of the city should be proportional to the global mass of infrastructures in place and have a long range (power-law) influence with some distance  $d$ :  $P_{C \rightarrow x} \propto -\frac{\int d\mu_C}{d^\gamma(x, C)}$
- A very short scale behaviour that forbids a new center to be located on existing infrastructures:  $P_{C \rightarrow x} = +\infty$
- A medium scale deduced from the two previous ones, that should display some local minimum, potentially with a particular geometry favouring a particular geometrical organization.

Thus among several possible fields we choose Fig. 8.7:

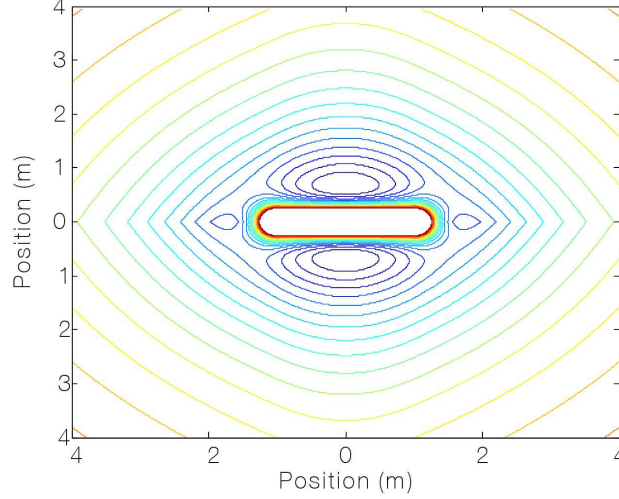
$$P_{C \rightarrow x} = \left( \frac{\alpha}{d_{\min}(x, C)} - \frac{\beta}{\sqrt{d_{\min}(x, C)}} \right) \int \frac{d\mu_C}{\sqrt{d_\perp(c, x)}} \quad (8.17)$$

$d_{\min}(x, C) = \min_{c \in C} (x, c)$  is used in Eq.8.17 so that the rejecting zone is hard : there is a tube around the city streets where new settlements are impossible. The radius of this tube is  $\lambda_0 = (\alpha/\beta)^2$  (Fig.8.7).

$d_\perp(x, c)$  is the  $\|\cdot\|_1$  norm in the local basis formed by the unitary tangent and the normal to  $C$  in  $c$ . The choice of such a distance is purely technical, to simplify explicit integral calculus compared to Euclidian distance and to produce peaks (Fig.8.7) at the extremities

of street segments. Towards  $\infty$ ,  $P_{C \rightarrow x} \sim \frac{\beta}{d(x,C)}$  and between those two extreme positions, interferences between streets segments produce local minima. To choose parameters  $\alpha$  and  $\beta$ , one sets  $\lambda_0$  the hard rejection radius and  $\beta$  the long-range influence. The choice of  $\beta$  influences the local geometry of the city but won't be discussed here.

How a new settlement is chosen from this potential field will be discussed in sec.8.3.3, the idea will be to find in a more or less hard way a local or global minimum in the field.



**Figure 8.7:** The level lines of the potential field for a city reduced to a single segment of length 1 with  $\lambda_0 = 1$  and  $\beta = 10$ .

### 8.3.2 Connection

Once a settlement is added in a location  $x$ , it links to the existing network  $C$ . Not all connections are eligible.

From a point  $x$  we define the **visible set of points**:

$$V_{x|C} = \{x \in C, [c, x] \cap C = \{c\}\} \quad (8.18)$$

This is the set of points in the city graph one can connect to without having to cross other streets.

The **optimal set of points** from  $x$  of a part  $E$  of  $C$  is the set of points that locally minimize the connection distance from  $x$  to the existing network:

$$\dot{E}_x = \{e \in E \mid \exists \epsilon \mid \forall e' \in E \cap c \oplus B_\epsilon, d(x, e) \leq d(x, e')\} \quad (8.19)$$

New connections are made between  $x$  and points in **the optimal visible set**  $V_{x|C}$ . This one being a finite set included in  $C \cup (x \perp C)$  where  $x \perp C$  is **the set of orthogonal projections** of  $x$  on the city in the visible set.

To avoid connections too close from each other, we introduce **the relative neighbourhood** ([10], App.A), a subsampling of the Delaunay neighbourhood. In a general way, if  $P$  is a

point and  $E$  a points set. Then  $s \in E$  is said to be in the relative neighbourhood of  $P$  ( $s \in RN[E||P]$ ) if and only if

$$\forall u \in S, d(P, s) \leq \max\{d(u, P), d(u, s)\} \quad (8.20)$$

ie there is no point both closer to  $s$  and to  $P$ . All candidates to become new connections are segments from  $x$  to its relative neighbourhood in the optimal visible set  $RN(V_{x|C}||x)$ . Now that we have found good candidates we will determine in sec.8.3.3 which of them will become actual new street segments.

### 8.3.3 Parameters

One of the aim of the morphogenesis model is to reproduce different morphologies of cities. These morphologies are seen as variation in the political decision made to build the city. We present in the following sections parameters that permit to differentiate politics and to define the behaviour of new settlers toward the potential field of the city and the way they connect to former streets.

**Organization** The global field induces minima. These minima represent points where it is the most interesting to settle.

The question is to find a parameter  $P_e$  that describes whether the city is organized or not. The idea is that when a city is organized it sticks strictly to optimal settlement places and when it is purely unorganized, new settlements are added at random without any influence of the potential field.

Then a new settlement is selected by a Monte Carlo method with a number  $n$  of iterations and the new point is be chosen as  $X = \text{Argmin}_n P(X_i)$ . For random cities  $n$  is close to 1 and for organized one it is much higher.

Let  $W$  be the area of a part of the plan that contains the current city, let  $X_1, \dots, X_n$  be  $n$  points on this part, uniformly and independently chosen. And let  $X = \text{argmin} P(X_i)$  and  $\text{Argmin}(P)$  the set of local minima of the potential field  $P$ . Then let

$$P_e = \mathbb{P}(|X - \text{Argmin} P| \leq e) \quad (8.21)$$

represents the probability that the Monte Carlo method throws a point in a radius  $e$  of a local minimum.

We want to give  $P_e$  as an input parameter and traduce it into an iterations number. Let  $N$  be the number of local minima. If  $e$  is quite small, comparing the surface of the window  $W$  and of the zone in a distance smaller than  $e$  of local minima we obtain:

$$P_e \approx 1 - \left( \frac{W - N \cdot \pi \cdot e^2}{W} \right)^n \quad (8.22)$$

$$n \simeq \frac{\log(1 - P_e)}{e^2} \cdot \frac{W}{N \cdot \pi} \quad (8.23)$$

$N$  is estimated roughly by noticing that a local minimum is often due to the interaction between two close streets segments:  $N \simeq 3 \cdot \#V$  since 3 is roughly the average connectivity

number of intersections.

$n$  diverges when  $P_e = 1$ . To model **very organized populations** ( $P_e = 1$ ), we take at first  $P_e = 1 - 10^{-10}$  (for instance) and then apply an additional steepest descent algorithm from this point to get a point that is exactly in a local (almost all the time global) minimum of the potential field. Such a case is simulated for instance in Fig.8.14.

**Connection and construction** There are typically about four or five possible new street segments in  $[x, RN(V_+||x)]$  for a new center  $x$ . If the city shapes as a slum it would be tree like so we link the number of streets segments indeed added with the construction  $\omega \in [0, 1]$  of the city. We sort segments in  $[x, RN(V_+||x)]$  by increasing length :  $(s_1, \dots, s_n)$ .  $s_1$  is drawn with probability 1. We pick  $n'$  at random with a Binomial law:  $n' \sim \mathcal{B}(\omega, n - 1) + 1$  and segments  $s_2, \dots, s_{n'}$  are also added. If  $\omega = 1$  every admissible segment is added and if  $\omega = 0$  only the shortest one.

**Sprawling** When constructing with a rejection radius  $\lambda_0$ , the city gets a typical mesh width. If at a particular urban operation a potential field with a rejection radius of  $K\lambda_0$  with  $K > 1$  is considered then the city's inner meshes will appear as filled up with the rejection zone of this potential field and new points of interest will position outside of the city.

With this observation we will consider that in a proportion  $f_{ext}$  centres are added with respect to a potential of rejecting radius  $K_{ext}\lambda_0$ . This creates foils at the outskirts of the city and thus an extension of the city that represents for instance an industrial zone which needs a large surface that might be cut later into offices or housing when the city continues its expansion.

**Refinements** To enhance the realism of this model, some empirical parameters are added. **The length of a new street segment** is bounded to  $l_{max} = k_{l_{max}}\lambda_0$ . This can avoid too long costly connections (possibly  $l_{max} = \infty$ ).

**The windows  $W$**  out of where new settlements are chosen can be either definitely set Fig.8.8 or dynamically change with the overall city Fig.8.9 and Fig.8.11.

Since a Monte-Carlo method is used to pick new centres, there is a very little probability that a new settlement is added in line with an existing street. If geometrically this does not have many consequences, it may strongly change the local topology. That is why, if an orthogonal projection is in a radius  $c\lambda_0$  with typically  $c \simeq 0.3$  of an intersection that is visible from the center then this **orthogonal projection is removed** and replaced with a connection with the nearby vertex. This rises the vertex degree and allows longer streets.

### 8.3.4 A few simulations

To summarize an individual simulation of the city growth we need to provide our algorithm with the parameters:

1. The number of settlements:  $N$

$\Omega \setminus P_e$	0	0.5	0.8	0.99999
1	0.37	0.43	0.46	0.48
0.6	0.26	0.31	0.33	0.36
0.3	0.14	0.16	0.18	0.20
0	0	0	0	0

**Table 8.1:** Variation of the meshedness coefficient  $M_4 = 2(\#E - \#V + 1)/(2\#V(1 - \bar{N}(2)) - 5)$  for the 16 simulated cities. Each result is the averaging of 30 simulations. The variation for each case is almost constant equal to 0.04.  $M_4$  is a increasing function of both  $P_e$  and  $\omega$ .

2. The organization probability  $P_e$
3. The radius of the rejecting tube:  $\lambda_0$
4. The long scale influence:  $\beta$
5. The construction:  $\omega$
6. The sprawling factor  $K_{ext}$  and the sprawling probability  $f_{ext}$

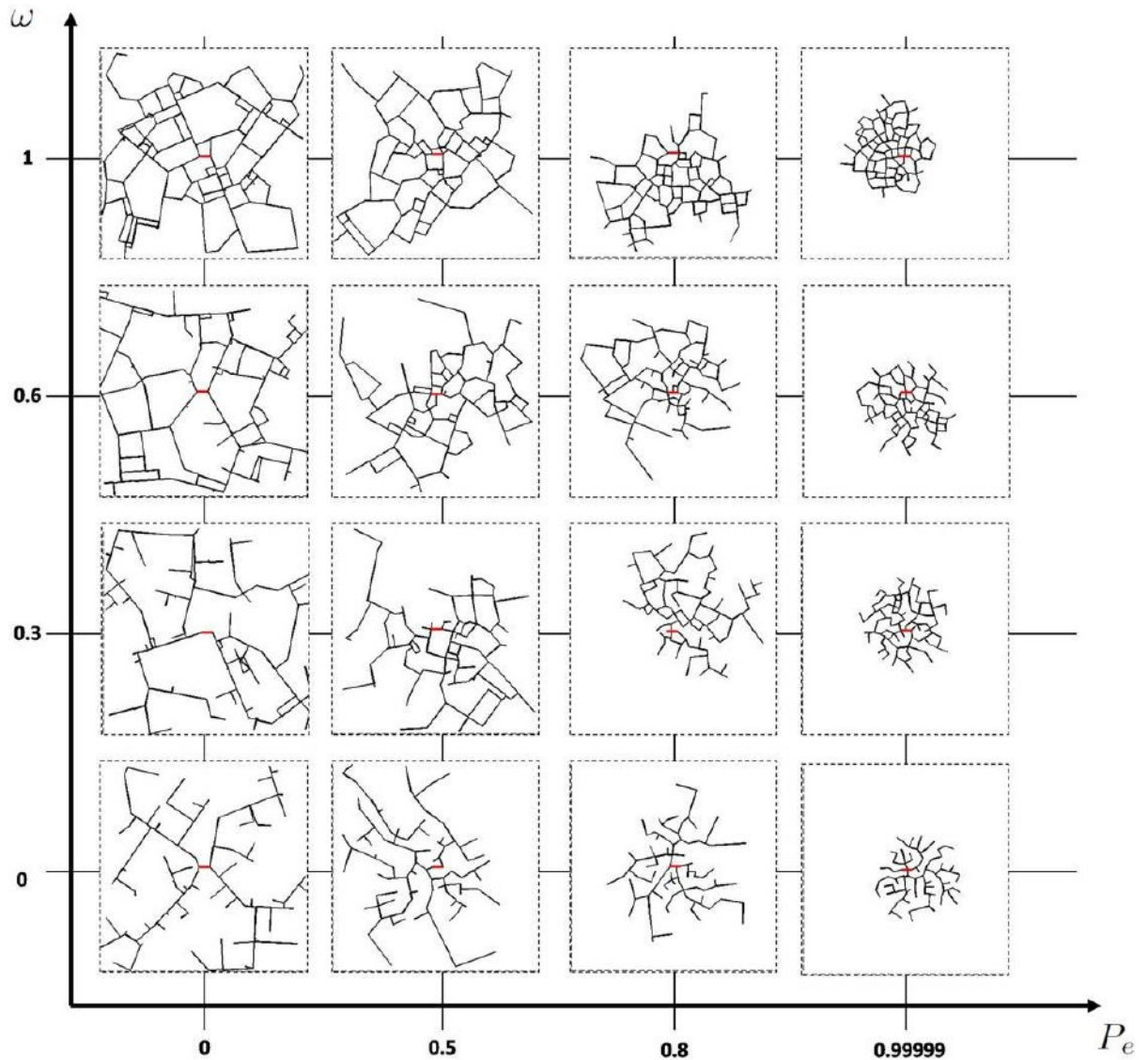
Only four parameters will actually shape the simulated city ( $P_e, \beta, \omega, f_{ext}$ ) the others being scaling parameters and that the influence of  $\beta$  won't be discussed here.

**Simple patterns** Fig.8.8 shows the result of 16 simulations. The organization probability  $P_e$  and the construction  $\omega$  are varying jointly when the same number of operations  $N = 80$ , the same rejecting radius  $\lambda_0 = 10m$ , the same available space (a square with an area of  $1.6 \text{ km}^2$ ), the same initial city (a segment of length 20 m at the center of the available space) and the same extension probability of  $O$  are used. The first result is that this model is able to reproduce very different type of growth with very few "physical" parameters.

We observe on this matrix representation that the meshedness  $M_4$  (Tab.8.1) is an increasing function of both  $P_e$  and  $\omega$  Tab.8.1. This result has been obtained by averaging the meshedness coefficient of 30 simulations for each couple  $(P_e, \omega)$ . Added to that the standard deviation of  $M_4$  for each couple is of 4 percent so that this coefficient is characteristic of the conditions of simulation. Contrary to that, the anisotropy coefficient  $Ani$  is almost the same in each case (between 0.31 and 0.46) with a large standard deviation of 20 percent. This  $Ani$  is quite large in the absolute: for the first iterations some directions have to be arbitrary chosen, which creates favoured directions. But it is the same order of weight as for the most isotropic French towns. Of course the organic ratio  $r_N$  is in every case close to 1.

When  $\omega \simeq 0$ , the resulting simulations are to be compared to the Saffman-Taylor instability. It seems when  $\omega \simeq 0$  that only a bounded number of ramifications are possible from the initial segment (4 on this figure) as if first created branches shielded the initial center from newer ones. When  $\omega > 0$ , the resulting cities are to be compared to crack patterns: their dynamics follows a logic of division / subdivision of space.

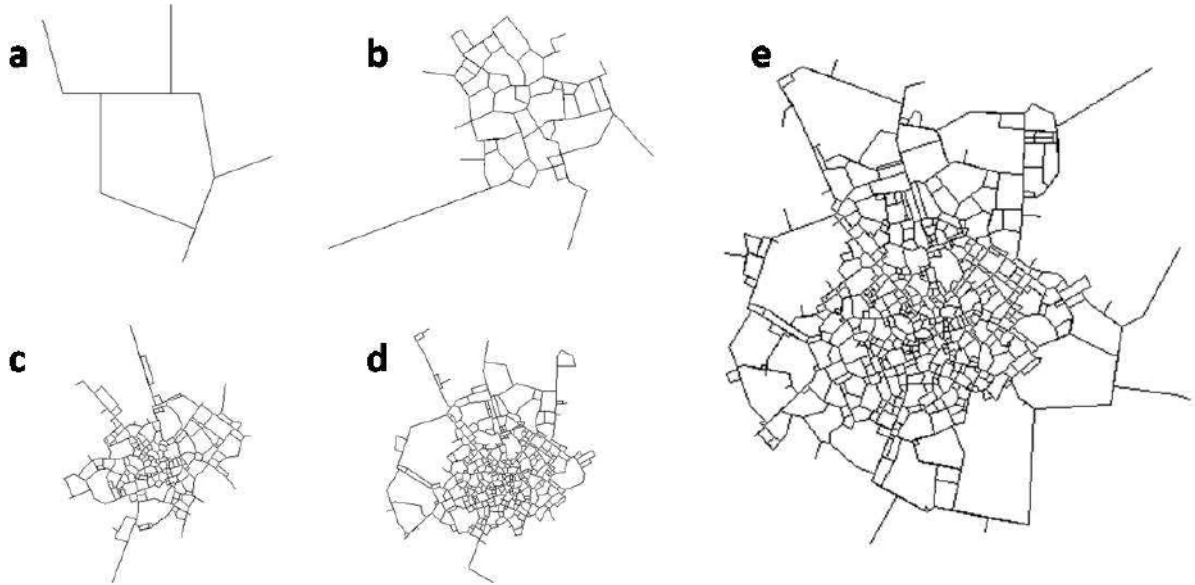
**A city built with constant parameters** Fig.8.9 presents a city evolving with  $\lambda_0 = 10m$ ,  $K = 10$ ,  $K_f = 0.1$ ,  $P_e = 0.8$ ,  $\omega = 0.7$ . The regular need of larger surface for activities



**Figure 8.8:** Simulations of the morphogenesis model with for various constant organization  $P_e$  and the construction  $\omega$ . On each thumbnail, the rejecting radius is  $\lambda_0 = 10m$ , there is no sprawling:  $k = 0$ , the number of settlements is  $N = 80$  and the available space is bound in a square with sides of  $400m$ . The red and bold segment represents an initial street-segment and a scale of 20 meters.

such as industries, big institutions, etc. is well reproduced here. During the history, as the development of the city center progresses, it eventually absorbs the peripheral larger surfaces, to split them into smaller surfaces, with thus new larger places appearing at the new periphery. This reproduces and explains the situation of economical zones always outside at the periphery of towns. It explains as well the successive subdivision of space that leaves so many traces, first in the log normal distribution of streets length but also in the hierarchical distributions of streets Fig.8.10. For this simulation, the ratio  $r_N$  is equal to 0.93 so that the term "organic" fits. The meshedness coefficient  $M_4 = 0.48$  is quite close to Amien's (between 0.41 and 0.54) as the anisotropy (0.69 to be compared to 0.71 in the

center of Amiens).



**Figure 8.9:** (*a, b, c, d*) : four steps in the development of the city (*e*) with 600 urban operations. For this simulation,  $\lambda_0 = 10m$ ,  $K = 10$ ,  $K_f = 0.1$ ,  $P_e = 0.8$ ,  $\omega = 0.7$  and the windows is adapted to the size of the current city. The main phenomenon at work is the dynamics between the inner development and the extension of the city that creates two hierarchical scales.

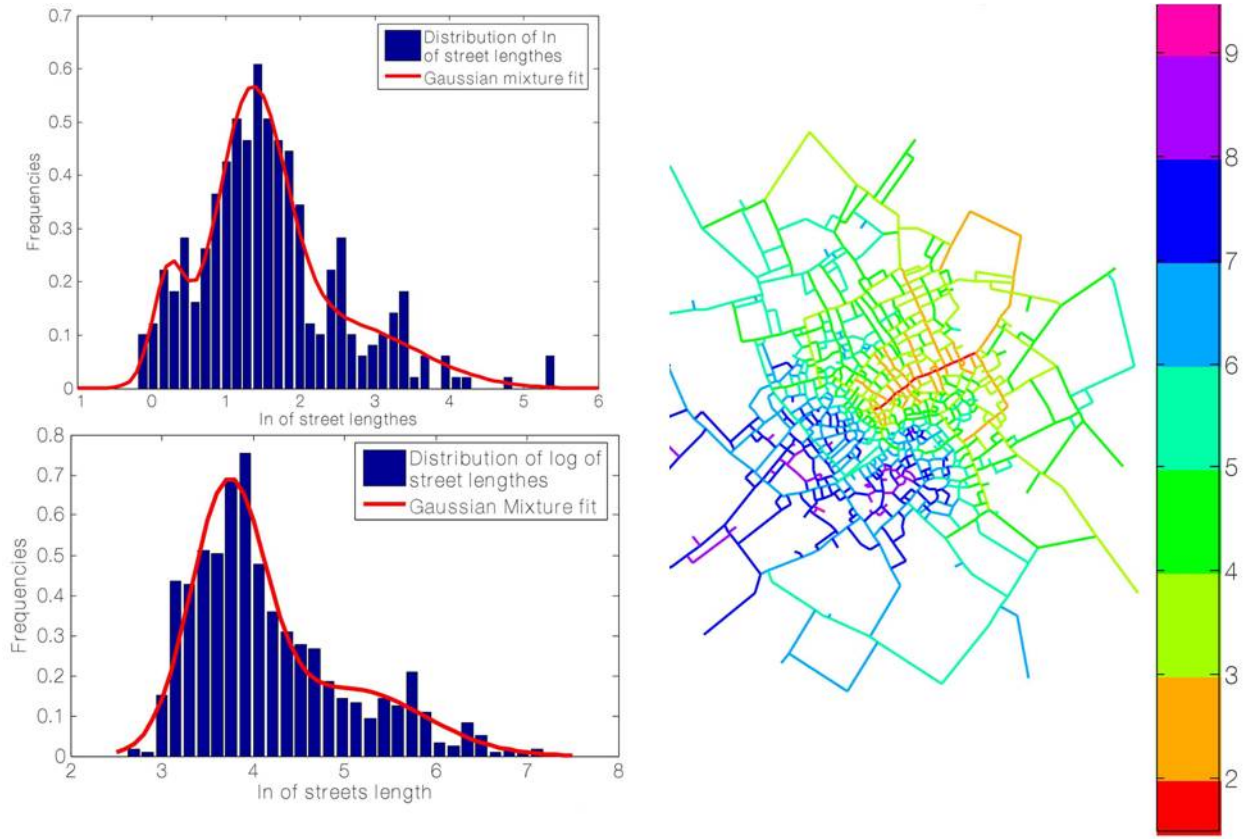
Fig. 8.10 shows that the morphogenesis reproduces the small topological radius and the log-scaling of street lengths observed in real cities (Ch.5). This shows the morphogenesis model reproduces non trivial emergent behaviours in the city map.

**A city built with varying parameters** Constant parameters are not realistic to model a real city, this one being shaped by its history which is from the morphogenesis point of view a variation of input parameters.

We represent the history of a city by a piecewise constant function  $t \rightarrow (P_e, \omega, l_{\max}, f_{ext}, K, \beta, \lambda_0)$ . For instance, the city of Fig.8.11 has been obtained by simulating at first a city with a low construction and no sprawling ( $\omega = 0.2$  and  $f_{ext} = 0$ ) and then changed to a sprawling and constructed city ( $\omega = 0.8$  and  $f_{ext} = 0.15$ ). The simulation starts with two perpendicular streets with a length 20 times larger than  $\lambda = 10m$ . These pre-existing streets are structuring elements as could be a river or long distance crossing roads. This kind of variation in parameters recalls Casbah in Morocco where the historical center of the city is a souk.

## 8.4 Object Morphogenesis of the City

A problem with the model presented above is that simulated cities are too "circular". We want to emphasize in the simulation engine the importance previous line streets becoming structuring elements and favouring particular developments like fingers along primitive long



**Figure 8.10:** The distribution of the logarithm of streets length for two synthetic cities (top) and the topological street distance to the center (bottom) for the resulting city of simulation Fig. ???. As for real cities (Amiens), the street topology presents a bounded hierarchical representation of the city and the streets length is well-fitted by a mixture of log-normal random variables.

streets. To this we make a paradigm shift. The previous potential was the result of the sum of infinitesimal potentials. This caused us technical problems to propose a distance that produces peaks at the extremities of a street.

Here we change of paradigm: the potential is not the sum of infinitesimal potentials but the sum of potentials induced by objects. Then a river or a street will have an influence all around it but the weight of this influence will not depend on the size of the object. A longer street will have more influence not in a particular place but this influence expresses in a larger zone.

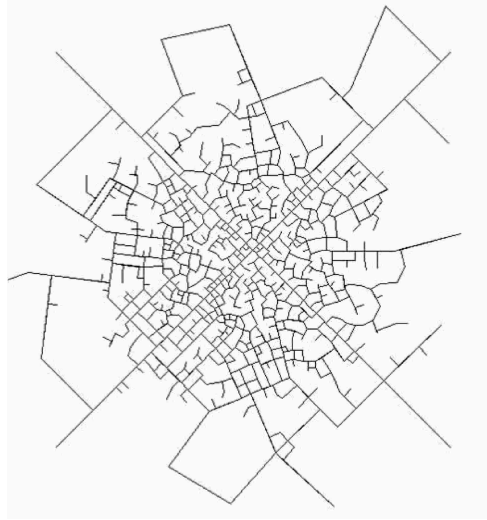
This idea will not be studied theoretically, we just present a simple principle and observe visually its repercussions on the city's geometry.

We define the potential function  $V$  (largely inspired from Lennard-Jones potential) by:

$$V_{\alpha,\beta,n}(r) = \alpha \cdot \left( \left( \frac{\beta}{r} \right)^{2n} - \left( \frac{\beta}{r} \right)^n \right) \quad (8.24)$$

The mathematical parameters  $\alpha, \beta, n$  of this function can be tuned via physical parameters as the **rejecting radius**:  $r_0 = \beta$ , the **optimal radius**  $r_m^n = 2 \cdot \beta^n$  and the **minimal potential**  $V_m = -\alpha \cdot \left( \frac{2^n - 1}{4^n} \right)$ . In  $[0, r_0]$  the potential is positive (as an approximation of





**Figure 8.11:** A city developing with varying parameters from two perpendicular long structuring elements. The historic center of the city has been built with a low construction parameter and no sprawling ( $\omega = 0.2$  and  $f_{ext} = 0$ ) to recall the tree aspect of a central souk in the Casbah. Then parameters are changed to  $\omega = 0.8$  and  $f_{ext} = 0.15$  which products an industrial crown. The resulting town is very circular, even with primitive long structuring elements.

$+\infty$ , the idea is to forbid settlements at a distance smaller than  $r_0$ ).  $r_m$  is the optimal distance to settle and  $V_m$  describes the relative attractiveness of that optimal radius. The shape of this potential will allow a displacement of the local minimal potential from the optimal radius to make a global compromise when several infrastructures interfere.

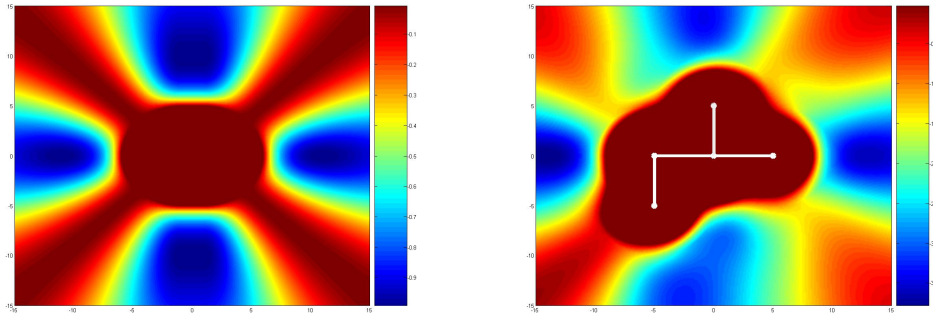
Two exponents are involved in Eq.8.24. To be more general they should have been  $m$  and  $n$  with no a priori link. We have chosen to take  $m = 2n$  since it allows simple calculus and we only try to fit 3 physical parameters ( $r_0$ ,  $r_m$  and  $V_m$ ) thus 3 parameters in the formula are enough.

To induce the geometry we want around a street, we add to the potential an angular effect. An object  $h$  produces an **unitary potential**  $P_h(x)$  which depends not only on the distance of  $x$  to  $h$  but also on the angle the segment limited by  $x$  and its projection on  $h$  makes with  $h$ :

$$P_h(x) = V(d(h, x)) \cdot \cos(2 \cdot |\angle(h, x)|) \quad (8.25)$$

The angle in the cosine function is double to ensure the spatial continuity of the unitary potential. It has the positive consequence to penalize points that make an angle of  $\pi/4$  with a street segment and to favour those which are in the alignment (Fig.8.12, Left). The global potential of the network is simply the sum of unitary potentials (Fig.8.12, Right).

We can keep the same philosophy for tuning simulation by introducing as in sec.8.3 an organization, a construction, an extension factor, an extension probability and some refinements. We want to test with this model the influence of axis in the current layout of a city. We thus define a simulation routine: a few random lines are made on the plane. Then we apply the object morphogenesis model to fill these primitive structuring elements with



**Figure 8.12:** Potential field for a single segment (left) and a collection of connected segments (right).

inhabitants. The results are various as shown in Fig.8.13. We observe through these four simulations that the global shape of the cities are no more circular, new street segments spouse and emphasize the shape of long lines. The curvature of some streets in the first (Top-Left) simulation has been for us astonishing, for the coherence they reveal. An other striking characteristic is to present large fluctuation of the density with clusters. Do the size of these clusters follow Zipf's law?

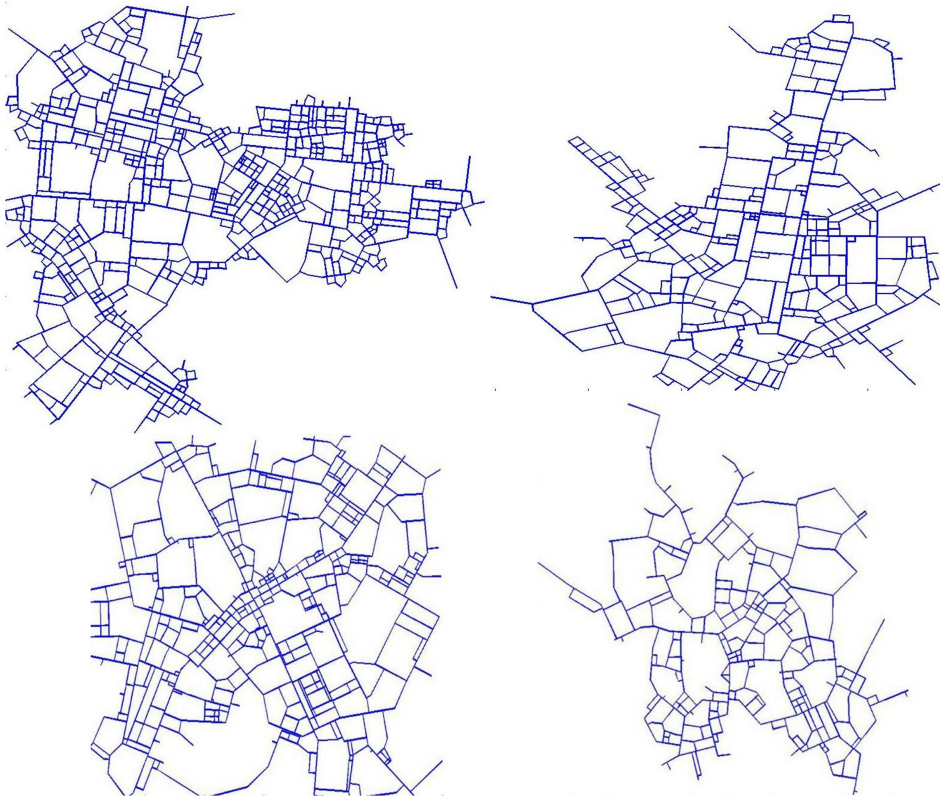
Fig.8.14 shows several steps in the construction of a city first constituted of an "very organized" population ( $P_e = 1$ ) to which follows a laxer one. A strong dense square-shaped center appears without more constraints.

## 8.5 Conclusion

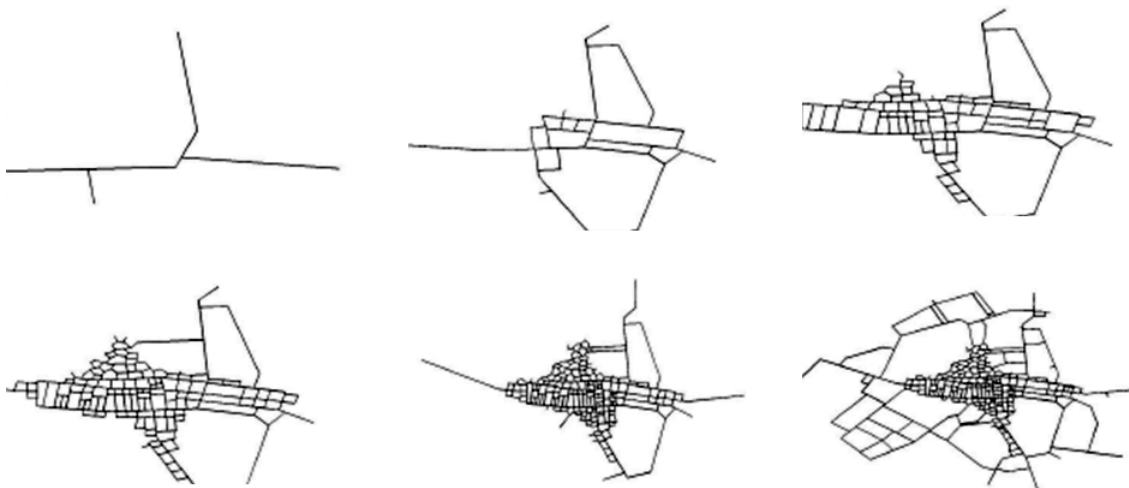
In this chapter we have presented a collection of nested models to test the influence of constitutive elements of the city on its development (spatiality, streets).

The first continuous model is above all the mathematical translation of a principle in accordance with Zipf's law. We postulate that there is a general scale-free principle in city growth. This principle is admitted and observed when considering systems of cities. We claim that it is also in action at the scale of a particular city. We only obtain numerical results but the generic mathematical problem of the evolution of a density function according to an integro-differential equation is of simple and flexible formulation. The fits of numerical results are very good and robust, it is clear that they could be reached theoretically. We have studied it in the mono-centric, radial case. But it should be studied the same way by adding anisotropy in space, several initial center, noise... This will show how cities' proximity influence the shape of each other.

The division and division/extension models are simple sketch of a city evolution that allow explicit calculus. The two morphogenesis models can only be simulated. They reproduce both city maps' hallmarks and a quite large diversity of layout. The simple morphogenesis makes too regular systems when they are large but the object morphogenesis puts the notion of axis at the center of its principles which makes emerge spontaneous intricate shapes and



**Figure 8.13:** A few realizations of the object morphogenesis. For each of these four cities, random long lines are at first drawn, creating primordial structuring elements and then the city grows with constant parameters, a high construction and a high organization. The variety of global shapes is explained by the randomness of first lines, the



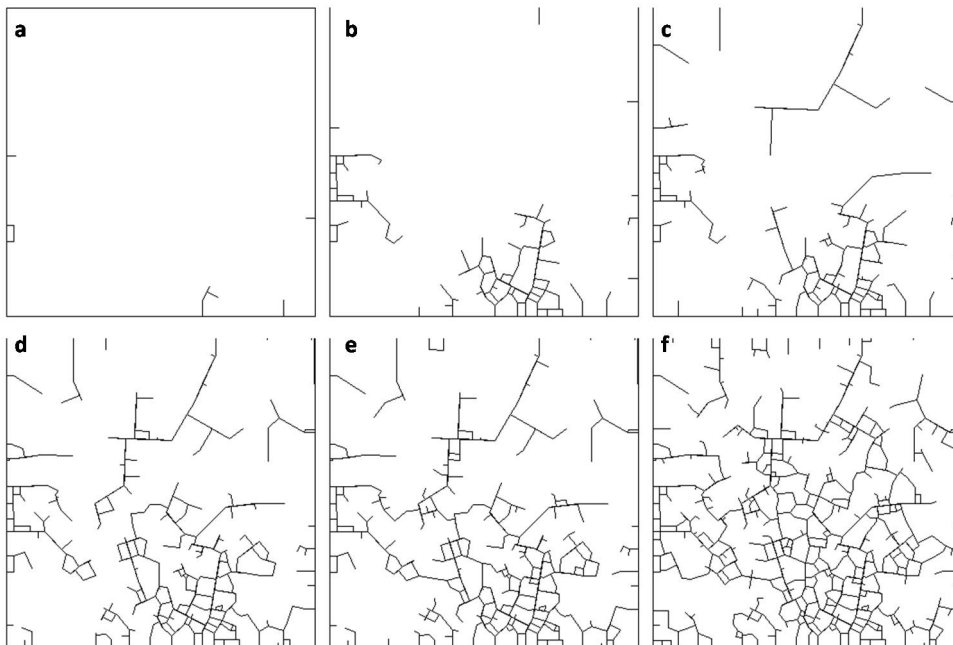
**Figure 8.14:** History of a city starting from random lines as structuring elements, knowing a period of high organization producing a square like dense center and then a softer period.

local clustering in the distribution of infrastructure.

We want again to emphasize the distinction between global optimization of a system created

in one piece and sequential local optimization of a system naturally evolving and produces "organic" shapes.

Our models are coherent. For instance in the extension/division model, new squares are added in places where they optimize their distance to existing infrastructures. The age and thus the development of former squares make naturally the new square spin as a spiral. The large scale evolution of the city is "snail-like". The morphogenesis model is able to spontaneously produce clusters that distribute with different sizes: Fig.8.15 displays the evolution of a city growing on the interior border of a square structuring element. The connection rule forbids two distant segments to connect. Several clusters appear on the different sides of the structuring square but they end up to melt and are no more observable in the final city. This simulation can be seen as the assimilation of faubourgs by a main city core.



**Figure 8.15:** Simulation of the morphogenesis in an area surrounded by a square structuring element. Connections of a too large distance are not allowed. Several centres emerge (b, c and d), their size is sparsely distributed (maybe a Zipf's law?). From (e) different clusters start melting, the distribution of clusters' size is no more observable.

In fact in the chronology of this thesis preparation models have come in the inverse order than we have presented them here. At first we wanted to work on a segment space: find the segment that at one time optimize the growth of the city. But there were continuity problems induced by the connectivity constraint. We then had to uncouple the choice of a position and the connection to the network. We wanted not to have to impose a priori probability distribution for the choice of centres (as done in correlated percolation [87, 20], L-Systems [104, 37] or topological network growth [8, 10]). The probability density had to be determined by the actual geometry of the city and change with time. We first based on centrality in the existing network but we remarked that Euclidian distance is for an outer point a good approximation of shortest path distance. From the structure of our data into

segments the most natural thing was to propose a potential induced by street segments. But we had a metaphysical problem: two aligned segments of length  $l$  and a single segment of length  $2l$  did not produce the same potential field... We then proposed to see the potential as the integral of contributions of infinitesimal roads. We studied in depth this model which reproduced the city features (small-world and street scaling).

But this model was too intricate and we were not able to make calculus on it. Notably studying the distribution of cells area asked for a lot of tedious computing work. We then went back to the very first sketch we had made at the beginning of our work when we were particularly interested to structuring elements: division of space. But division of space was not sufficient to obtain a power law distribution of cell area. We had to add the notion of extension in this model. We first tried to add sequentially squares and divide previous square. But the distribution was not the good one. We had to make the hypothesis that new squares grow geometrically. This hypothesis has first shocked us. We looked for literature on this subject and asked to town-planners "at what speed does a city's boundaries drift away?". They did not know. It was impossible with our data to measure these information. And we proposed the continuous model. The idea was "city will evolve as in Newtonian systems, every thing will interact with every thing but instead of making other object move, interactions will make objects grow". And finally this model is in accordance with Zipf's law, and the boundary of a city (whatever its arbitrary definition is) indeed drifts away exponentially.

At last our work on simplest distance and simplest centrality made us understand that the cutting of the city network was essential and that in fact our first idea for the morphogenesis model was better. We re-worked on it and the contradictions we had found in a first time erased.

## Part III

# Stochastic Geometry



# Sketching a city with Stochastic Geometry

## Contents

---

9.1	Principles . . . . .	<b>149</b>
9.1.1	Basics and intuition in Stochastic Geometry . . . . .	149
9.1.2	Telecommunication assessment . . . . .	153
9.1.3	Identification procedure . . . . .	154
9.1.4	Problems with the method . . . . .	155
9.2	Low scale models . . . . .	<b>155</b>
9.2.1	Poisson Line Tessellations . . . . .	156
9.2.2	Crack STIT . . . . .	157
9.2.3	Dead-ends . . . . .	157
9.3	High scale model: Gabriel Graph . . . . .	<b>158</b>
9.3.1	Construction . . . . .	158
9.3.2	Properties . . . . .	158
9.3.3	Typical node . . . . .	159
9.3.4	Mean Formulae . . . . .	160
9.4	Identification . . . . .	<b>161</b>
9.4.1	Rewriting of intensity vectors as feature vectors . . . . .	161
9.4.2	Identification . . . . .	162
9.5	Algorithm . . . . .	<b>162</b>
9.6	Conclusion . . . . .	<b>163</b>

---



---

**Sketching a city with Stochastic Geometry: Synthesis**


---

The morphogenesis model has the physical interest to exhibit the principal mechanisms that shape a city. But this model is not "operational". In this chapter we are going to present more simple models from **stochastic geometry** that sketch a city map. Certainly less realistic, they allow solving fast optimization problems occurring in urban engineering whose solutions are strongly constrained by the road network.

We place in the continuity of [57] in stochastic geometry. A map extraction being under consideration we seek to identify a **random tessellation** to substitute statistically to the actual street pattern.

Our contribution is modest, the purpose is to reconsider the work that has already been done, choose better models by the light of the previous chapters' results, simplify and "robustificate" the identification procedure.

We propose to consider a map as a superposition of two scales. **The upper scale** is identified to a **Poisson Gabriel Graph** (we study here the statistics of this graph), and the **lower** to a Crack STIT or a Poisson Line Tessellation according to the actual morphology of the map. The lower tessellation is completed by a **Poissonian Dead-end Process** we define and statistically study. We provide an identification method robust and fast to run based on **simplest centrality filtering** and model selection with **intrinsic parameters**.

---

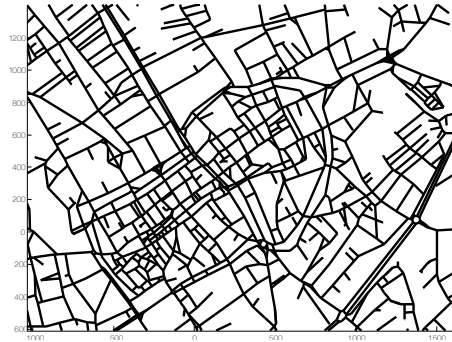
**Contributions of this chapter**


---

1. Short review on stochastic geometry and its utilization to statistical optimization.
2. Study of the mean features of the Poisson Gabriel Graph by the simulation of its typical vertex.
3. Definition and study of a dead-end process via the calculus of the mean distance of the typical point of a stationary Poisson Process and a random stationary tessellation.
4. Rewriting of identification criteria and specification of a fast and robust identification algorithm.

## 9.1 Principles

We aim at replacing a map extraction such as in Fig.9.1 by a random graph statistically equivalent to make the resolution of telecommunication problem easier. We first present main principles of stochastic geometry [122] and their applications to telecommunication network forecasting [61, 62, 60, 59, 57]. A more theoretical review on stochastic geometry can be found in App.C and the basic notions related to geometry such as Voronı and Delaunay tessellations are presented in App.A.



**Figure 9.1:** An extraction of the city of Troyes. The purpose of this section is to model this city graph by a stationary model of random geometry

### 9.1.1 Basics and intuition in Stochastic Geometry

**Poisson Point Process** A spatial Poisson Point Process of intensity  $\Lambda$  a measure on  $\mathbb{R}^d$  is an infinite countable random set of (distinct) points of  $\mathbb{R}^d$ :  $\Psi = \{X_1, X_2, \dots\}$  which respects:

1.  $(\#\Psi \cap A_1) \dots (\#\Psi \cap A_n)$  are independent discrete variables when  $A_1, \dots, A_n$  are disjoint bounded Borel sets.
2.  $\mathbb{P}(\#\Psi \cap A) = e^{-\Lambda(A)} \frac{(\Lambda(A))^k}{k!}$  where  $A$  is a bounded Borel set

It is important to admit the existence of such processes and picture them as the representation of total randomness.

A point process is said stationary and isotropic if when  $\mathcal{T}$  is a translation and  $\mathcal{R}$  is a rotation then  $\Psi$ ,  $\mathcal{T}\Psi$  and  $\mathcal{R}\Psi$  are equal in distribution: the translated or rotated process behaves the same way. If a process is stationary and if  $W$  is a compact window then

$$\exists \lambda > 0 \text{ the "intensity" } \mathbb{E}(\#\Psi \cap W) = \mu_2(W) \cdot \lambda \quad (9.1)$$

A Poisson Point Process whose intensity writes  $\Lambda(\cdot) = \lambda \cdot \mu_2(\cdot)$  is an homogeneous Poisson Point Process. It is both stationary and isotropic.

**Simple Random Tessellations** A tessellation is an infinite sequence  $(C_1, C_2, \dots)$  of compact polygons such as  $\cup C_i = \mathbb{R}^2$  and if  $i \neq j$ ,  $\text{Int}(C_i) \cap \text{Int}(C_j) = \emptyset$ . It is somehow an infinite planar graph with no vertices of degree 1. The term "random tessellation" has to

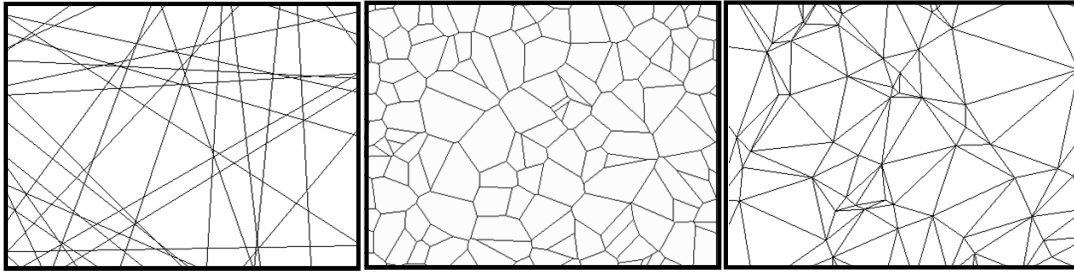
be defined properly but the practical idea is to define some construction rules of tessellation from a random set of points. The classical models taken into account in city modelling are (Fig.9.2):

**Poisson Line Tessellation:** a Poisson Point process is dropped on the real line. To each point is associated a line with a random orientation.

**Poisson Voronoï Tessellation:** A planar Poisson Process is dropped and the Voronoï diagram of that set is drawn

**Poisson Delaunay Tessellation:** A planar Poisson Process and its Delaunay diagram is drawn

These tessellations are "stationary" and "isotropic".



**Figure 9.2:** Simulation in a bound window of a Poisson Line Tessellation, a Poisson Voronoï Tessellation and a Poisson Delaunay Tessellation.

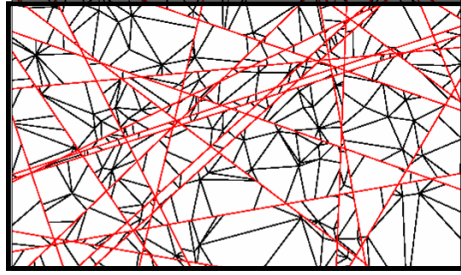
A tessellation process  $\Xi = (C_1, \dots)$  induces other random object processes: the point process of vertices, the segment process of edges, the point process of edges' mid point, the point process of cells' centroid. If the tessellation is "stationary" then all the above listed process also are. And one can consider the associated intensities  $\lambda_0$  of vertices,  $\lambda_1$  of edges' mid points,  $\lambda_2$  of cells' centroid and  $L_A$  which is the expectancy of the total length of the tessellation's edges in a window of area 1. For convenience we will write  $L_A = \lambda_3$  and consign these four intensities in a vector  $\vec{\lambda}$  called the intensity vector.

The statistics of the resulting topology of each model only depends on the single parameter  $\gamma$  (Tab.9.1) which defines the stationary underlying Poisson Point Process.

	PLT	PVT	PDT
$\lambda_0$	$\frac{\gamma^2}{\pi}$	$2 \cdot \gamma$	$\gamma$
$\lambda_4$	$\frac{2 \cdot \gamma^2}{\pi}$	$3 \cdot \gamma$	$3 \cdot \gamma$
$\lambda_2$	$\frac{\gamma^2}{\pi}$	$\gamma$	$2 \cdot \gamma$
$\lambda_3$	$\gamma$	$2 \cdot \sqrt{\gamma}$	$\frac{32}{3 \cdot \pi} \cdot \sqrt{\gamma}$

**Table 9.1:** Intensities of stationary tessellation features for classical models generated from an homogeneous Poisson Point Process of intensity  $\gamma$ .

**Iterated tessellations** Let  $\Xi^{(1)} = (C_1^{(1)}, C_2^{(1)}, \dots)$  a random tessellation and  $\Xi^{(2)}$  another, with  $\Xi^{(2),1}, \Xi^{(2),2}, \dots$  an infinite sequence of independent random tessellations that are equal to



**Figure 9.3:** A sample of the iteration of a PLT (in red) followed by a PDT (in black).

$\Xi^{(2)}$  in distribution. Then the new random tessellation written  $\Xi = \Xi^{(1)}/\Xi^{(2)}$  is constructed by simulating  $\Xi^{(1)}$  and filling each cell  $C_k$  by a truncated tessellation  $\Xi^{(2),k} \cap C_k$ . If  $\Xi^{(1)}$  and  $\Xi^{(2)}$  are stationary  $\Xi^{(1)}/\Xi^{(2)}$  also is. Its intensity vector is written  $\vec{\lambda}^{1/2}$ . If at least one of the tessellation is isotropic,  $\vec{\lambda}^{1/2}$  can be simply calculated from  $\vec{\lambda}^{(1)}$  the intensity vector of  $\Xi^{(1)}$  and  $\vec{\lambda}^{(2)}$  the intensity vector of  $\Xi^{(2)}$ :

$$\begin{aligned}
 \lambda_0^{1/2} &= \lambda_0^{(1)} + \lambda_0^{(2)} + \frac{4}{\pi} \cdot \lambda_3^{(1)} \cdot \lambda_3^{(2)} \\
 \lambda_1^{1/2} &= \lambda_1^{(1)} + \lambda_1^{(2)} + \frac{6}{\pi} \cdot \lambda_3^{(1)} \cdot \lambda_3^{(2)} \\
 \lambda_2^{1/2} &= \lambda_2^{(1)} + \lambda_2^{(2)} + \frac{2}{\pi} \cdot \lambda_3^{(1)} \cdot \lambda_3^{(2)} \\
 \lambda_3^{1/2} &= \lambda_3^{(1)} + \lambda_3^{(2)}
 \end{aligned} \tag{9.2}$$

**Cox Processes** A Poisson Cox Process is a point process which is Poissonian with respect of a random variable: the intensity of the process is itself random.

For instance let  $\Xi$  be a random tessellation and  $\Psi$  a random Poisson point process on the edges of this tessellation. With respect of a realization of  $\Xi$ ,  $\Psi$  is a linear Poisson Point Process and the variable integrating all realization of  $\Xi$ :  $\Psi$  is said to be a Poisson Cox Process.

**Typicality** Intuitively, if you have a collection of homogeneous objects  $O = (O_i)$ , the "typical" object of this collection is an object that have all the properties of  $O$ . For instance if  $O$  is a family of rabbits, the "ideal" rabbit has two ears and one tail. But the "typical" rabbit has let's say 1.99 ears on average and 0.98 tails... In fact the typical rabbit has a distribution of number of tails and a distribution of number of ears and a joint distribution of these two features. Consider it is a rabbits family from Chernobyl, each rabbit has a chance  $p$  to have been irradiated. If he has been irradiated is a a certain probability to lose on ear, the other and/or the tail. The typical rabbit is then a "random" rabbit which reproduces in distribution all the characteristics of the rabbits' family. In average we said the typical rabbit has 1.99 ears and 0.98 tails, he is irradiated with probability  $p$  and if so he correlates the lake of ears and tail...

Typical object is a process on the set of objects that has "in average" "all" the properties of the set and their correlations. It can be seen as a "well sampled" object. If you "pick" a rabbit at random you get a realization of the typical rabbit, if you pick a ear at random

you do not get a realization of the typical rabbit... you get the rabbit to whose belongs the typical ear!!

To formalize this idea we have to clearly define the context, the notion of object, picking, average and property. We will restrict now to geometry and stationary processes that is to say process that respects the Lebesgues measure. If  $O = (o_i)$  is a stationary process of object on a measurable space  $\Omega$  in  $\mathbb{R}^2$  then there exist a typical process  $o^*$  such that for all bounded window  $W$  and for all  $f$  measurable functional on  $\Omega$  that is translation invariant,

$$\mathbb{E} \left( \frac{\sum_{o_i \in O \cap W} f(o_i)}{\#O \cap W} \right) = \mathbb{E}(f(o^*)) \quad (9.3)$$

this definition is not convenient because of the fraction in the expectancy of the first term. The above formula is then replaced by:

$$\frac{1}{\lambda \cdot \mu_2(W)} \mathbb{E} \left( \sum_{o_i \in O \cap W} f(o_i) \right) = \mathbb{E}(f(o^*)), \quad \lambda = \mathbb{E} \left( \frac{O \cap W}{\mu_2(W)} \right), \forall W \text{ bounded (intensity)} \quad (9.4)$$

Similarly, if you pick an object uniformly at random in a random realization of  $O \cap W$ :  $\omega$  then

$$\mathbb{E}(f(\omega)) \rightarrow_{W \rightarrow \mathbb{R}^2} \mathbb{E}(f(o^*)) \quad (9.5)$$

The distribution of  $o^*$  is the **Palm distribution** of  $O$ .

You take a tessellation process and extract it in a large window. If you number each cell, pick a number and look at the associated cell you obtain a random variable: the typical cell of the tessellation. If you pick a point at random in  $W$  and look at the cell it belongs to you do not get the same random variable.

In a city graph context, let us take the following example. You consider all streets of a city, take uniformly at random one of them and look at its length. Then the random measure resulting of this procedure is "the length of the typical street", we have seen that it follows a log-normal law. Contrary to that if you pick a point at random on the street network and look at the length of the street it belongs to, you get a different random variable, shifted to higher length values. In fact, if  $f$  is the distribution of length of the typical street,  $g(k) = \frac{1}{\mathbb{E}f} \cdot lf(l)$  is the distribution of the length of the street of a random point.

When in addition we suppose the process is Poissonian, it is possible to simulate the process around the typical point without having to simulate the whole process. In fact if  $X$  is a stationary Poisson Point Process with intensity  $\lambda$ ,

$$\mathbb{E} \left( \sum_{x \in X} f(x, X) \right) = \lambda \mathbb{E}(f(0, X + \{0\})) \quad (9.6)$$

Thus the Palm distribution of a stationary point process  $X$  is  $\mathbb{P}_0(X \in \cdot) = \mathbb{P}(X \cup 0 \in \cdot)$ . More general if  $X$  is not stationary but has a diffuse measure  $\Lambda(\cdot)$ :

$$\mathbb{E} \left( \sum_{x \in X} f(x, X) \right) = \int \mathbb{E}(f(x, X + \{x\})) d\Lambda(x) \quad (9.7)$$

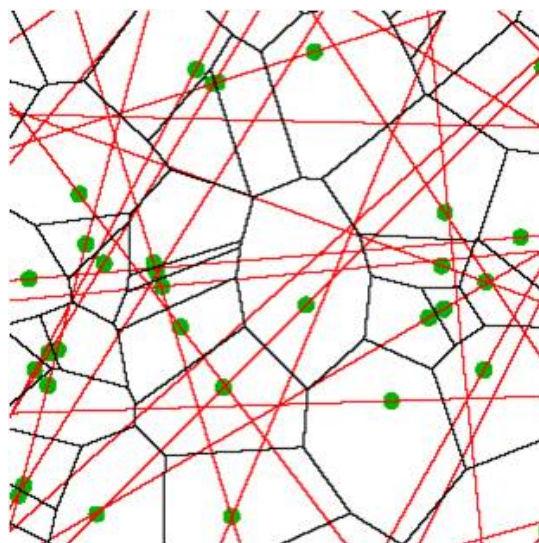
This is known as the **Mecke's or Slivnak's formula** and expresses the idea that you do not disrupt a Poisson Point Process by adding a point to it. In fact it is a characteristic

property of Poisson Processes [5]. As a consequence you can measure the landscape created by the process and observed in a point of this process by placing in 0. The main technical consequence is that to assess a functional on a typical cell of a tessellation, you can place its center in 0 and construct iteratively the closest neighbours. Since tessellations are generally defined by local interaction rules you do not have to simulate all the neighbours but all until you know further points will not interact with the cell.

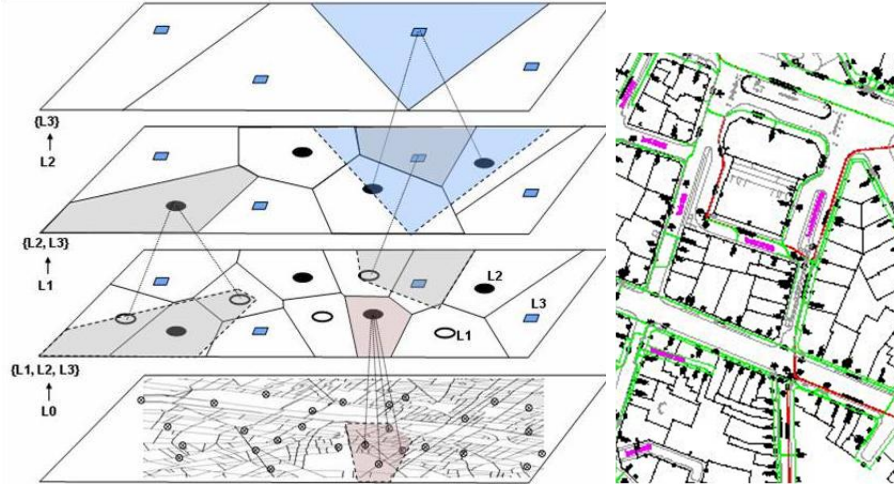
### 9.1.2 Telecommunication assessment

Generally speaking, a telecommunication network is constituted of several layers ( $L_1, L_2, \dots$ ) of routers hierarchically ordered. Then each router in  $L_k$  has to be connected to an upper router in  $L_{k+1}$ . The cost of the whole network is an increasing function of the number of routers and the total length of cable used to link the whole system. Hence a router of  $L_k$  is linked to the closest router of  $L_{k+1}$ . Moreover in a urban context, routers and cable can solely be laid on the road system. So minimizing the cost of a network is equivalent to optimize a combinatorial problem on a city graph with a lot of variables. Engineers have to prospect for a network that performs a good compromise between cost (total cable length) and Quality Of Service (number of routers). They have to test a lot of scenarios to make their choice. Calculate an optimal network for each scenario is much too time greedy. [57] proposed to solve the problem statistically:

1. Consider the road network is in fact the result of a Poissonian Random Tessellation
2. Consider that each router layer is a Poisson point process dropped on the previous tessellation (Cox Process Fig.9.4)
3. Assess statistically the distribution of cable length



**Figure 9.4:** The road network is replaced by a PLT (in red). The routers of the higher layer  $L_1$  are replaced by a Poisson Point Process on the PLT (in green) which is a Poisson Cox Process. In black, the Euclidian serving zone of each router. From Catherine Gloaguen's.



**Figure 9.5:** A telecommunication network is constituted of several layers of routers here  $L_0, L_1, L_2, L_3$ . Routers are put on the road network and are connected with respect of that network. From Catherine Gloaguen's.

We consider at first the problem with a road network  $\Xi$ , a low router layer  $\Psi_0$  and a high router layer  $\Psi_1$ ,  $\Xi$  being the support of each layer. The problem is to know for each element of  $\Psi_0$  its shortest path distance to the closest router of  $\Psi_1$ . Statistically we only seek out for the distribution of these distances. This problem is the simplest and in fact provides the fundamental element for the solution of general case. We want to assess the distribution  $d_{\Psi_0 \rightarrow \Psi_1(\cdot)}$  of length of a typical point of  $\Psi_0$  to the process  $\Psi_1$ . This distribution does not depend on  $\lambda_0$ . As an approximation we consider  $x$  connects to  $y \in \Psi_1$  iif  $x \in \text{Vor}(y|\Psi_1)$ , the Voronoï cell being defined in the Euclidian meaning. Scaling arguments show that  $d_{\Psi_0 \rightarrow \Psi_1(\cdot)}$  only depend on the shape of the road tessellation  $\Xi$  and a scale parameter  $\kappa$ :

$$\kappa = \frac{\text{Mean total road network length per unit area}}{\text{Linear intensity of } \Xi \text{ (per unit length)}} = \frac{\lambda_3}{\lambda_l} \quad (9.8)$$

we the write  $d_{\Psi_0 \rightarrow \Psi_1(\cdot)} = d_{\kappa, \Xi(\cdot)}$ . It is possible to simulate exactly (and with no runtime or memory problems) [62] a typical node of  $\Psi_1$ , its serving zone (Voronoi cell) and the portion of  $\Xi$  that belongs to this zone.  $d_{\kappa, \Xi(\cdot)}$  is also the distribution of distances of all points of  $\Xi \cap C^*$  to the center of  $C^*$ . The general features of the simulated curves and the exact calculus in the cases  $\kappa \rightarrow 0$  and  $\kappa \rightarrow \infty$  leads to fit the general density by the positive function  $f$  ([62]):

$$f(x|\kappa, \alpha = \alpha(\kappa), \beta = \beta(\kappa)) = C_0(\kappa) (x + C_1(\kappa))^{\beta-1} e^{(-\alpha(x+C_1(\kappa)))^\beta} \quad (9.9)$$

with  $C_0$  and  $C_1$  known constants that depend on  $\kappa$  (and  $\alpha, \beta$ ) and the shape or "type" of  $\Xi$ .

### 9.1.3 Identification procedure

The routine to identify an urban extraction to a stochastic tessellation presented in [57] is:

1. Pre-processing of the map

- (a) Choice of a convex window
  - (b) Erasing of dead-ends
  - (c) Erasing of squares and roundabouts
  - (d) Rectification and underscanning of points fitting streets' curves.
2. Calculus of  $\vec{\lambda}^{(C)}C$  the real topological vector of the studied city  $C$ .
  3. Calculus for each model  $i \in \{ \text{PLT, PVT, PDT} \}$  or their iteration of

$$\gamma_i = \underset{\gamma}{\operatorname{argmin}} d(\vec{\lambda}^{(i)}(\gamma), \vec{\lambda}_C) \quad \text{with} \quad d(\vec{a}, \vec{b}) = \sqrt{\sum_{j=0}^N \left( \frac{\vec{a}_j - \vec{b}_j}{\vec{b}_j} \right)^2} \quad (9.10)$$

4. Choice of the best model  $I$  from  $I = \underset{i}{\operatorname{argmin}} \gamma_i$
5. Validation of this choice by a sequence of Monte-Carlo simulations

#### 9.1.4 Problems with the method

The above described procedure presents some problems:

1. The algorithm calls for a very time consuming preprocessing
2. Tessellations models have been chosen for their (relative) simplicity but are they relevant to model cities?
3. There is a symmetry in the formula Eq.9.2: it is impossible to make a distinction between  $H/K$  and  $K/H$
4. 2 parameters are assessed from 3 independent measures which is not robust
5. Parameter assessments are strongly dependant of map's quality, sampling and the choice of window

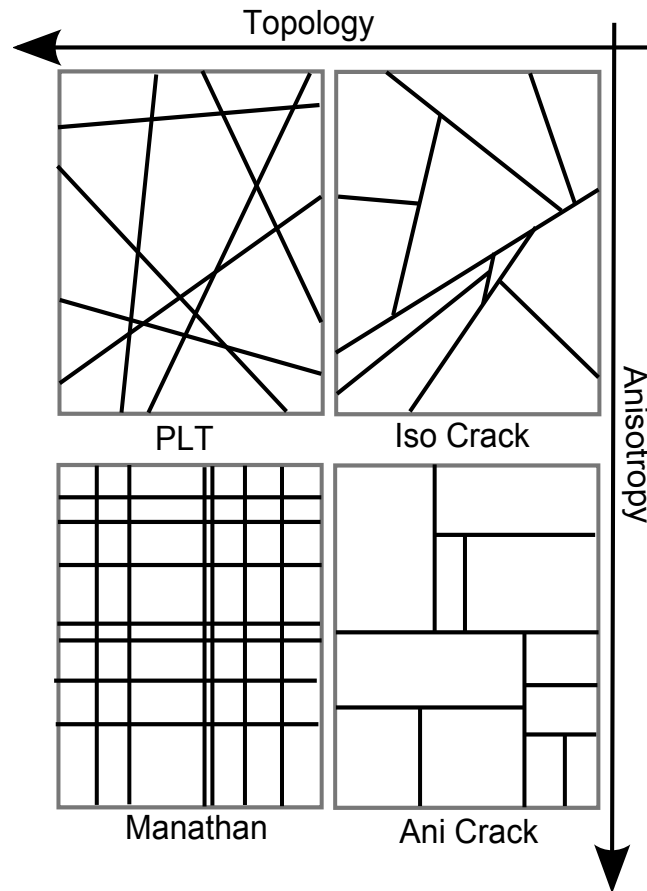
In the following, we will choose stochastic tessellations that are suitable to model cities. We propose to see the city as a superimposition of two scales. At each scale a few models emerge naturally. We present a fast identification method based on intrinsic criteria rather than on average criteria and break the symmetry of the iteration formula (Eq.9.2) by introducing a simplest centrality filtering.

## 9.2 Low scale models

If a city is **planned** it is essentially locally constituted of long streets intersecting in vertices of degree 4. In the **organic** case, we have seen that the map is essentially the result of a space division process with vertices of degree 3.

Two well documented tessellation models that simply reproduce these features are the Poisson Line Tessellation and the Crack STIT tessellation. Besides in both model an isotropy





**Figure 9.6:** Matricidal representation of the four models we consider to model the lowest scale of a street network. Each depends only on one parameter tuning its intensity. Models are identified by crossing a topological and an anisotropy criterion.

function can naturally be introduced which is very interesting to consider the influence of structuring elements. In the following we present the mathematical formulation of these models and their main properties. To end with, we propose a geometrical construction that allows adding simply "Poissonian" **dead ends** to the tessellations.

### 9.2.1 Poisson Line Tessellations

Let  $X = \{X_i\}_{i \in \mathbb{N}}$  be a homogeneous Poisson Point Process of intensity  $\lambda$  on the real line,  $(\theta_i)$  a sequence of independent random variable following  $\phi$  a probability measure on  $[0, 2\pi[$ . To a couple  $(X_k, \theta_k)$  one associates the line  $H_k$  perpendicular to the circle of radius  $[0, X_k]$  and passing through the point  $\mathfrak{R}_{\theta_k}(X_k)$ . The set  $\{H_k\}$  is said to be a Poisson Line Process. If  $\phi$  is uniform the result is isotropic: and Poisson Line Tessellation and if  $\phi$  is constituted by two atoms separated by  $\pi/2$  the result is maximally anisotropic, it is a Manathan tessellation (Fig.9.6).

In both cases, the resulting topology is lonely compounded of degree 4 vertices.

### 9.2.2 Crack STIT

There are several possibilities to define a tessellation resulting from division of space. And [41] proposes a generic framework. Every construction is iterative and the final tessellation is always the limit of a division process on a bound window that is dilated progressively. A measure on compact bodies is chosen: area, perimeter... At a given step, each cell of the current tessellation is divided by a random line (with respect of an isotropy function  $\phi$ ) with a probability proportional to its measure. The new resulting tessellation is dilated and the process restarts.

The Crack STIT tessellation is obtained by choosing the perimeter as the selection measure. It is more classically obtained by another construction process: [90, 95, 95, 112] and has some particularities that make it appealing. The resulting topology is compound of degree 3 vertices.

The tessellation is stable under iteration (in distribution). The interior of typical cell of a Crack STIT is equal (in distribution) to the one of a PLT with the same isotropy function and linear intensity.

### 9.2.3 Dead-ends

Tessellations models do not allow to represent a very common phenomenon on city maps: dead-ends (around 10 percent of vertices). In real cities and especially in country zones, dead-ends organize into trees. Nonetheless we will approximate these trees by a single segment which is accurate in urban zones.

In a stochastic framework, we consider a stationary tessellation  $\Xi$  with linear intensity  $\lambda_3$  and a Poisson Point Process  $X$  of intensity  $\gamma$ . We consider the Point Process connects to the tessellation by making a segment between each point and its projection on the tessellation. The tessellation with the addition of the Dead-End process is a stationary random graph. We are interested in its vector of intensities  $\vec{\zeta} = {}^t(\zeta_0, \zeta_1, \zeta_2, \zeta_3)$ . The three first components are obvious and some scaling considerations with the use of the Mecke formula shows that the mean length of such a segment is

$$\zeta_3 = \frac{\gamma}{\lambda_3} \alpha_{\Xi} \quad (9.11)$$

where  $\alpha_{\Xi}$  is a constant parameter depending only on the shape of the tessellation  $\Xi$ .  $\alpha$  is the mean distance of the origin to the same stationary dilated tessellation with edge intensity 1. In general  $\alpha$  is an integral non calculable, here are nonetheless particular results for the models under consideration:

**PLT** The PLT of intensity 1 is generated by a Poisson Point Process  $P$  of intensity 1 on the real-line and by picking for each of this point a direction at random. Consequently the distance of 0 to the tessellation is a random variable  $Z$  whose distribution is:

$$1 - \mathbb{P}([-x, x] \cap P = \emptyset) \quad (9.12)$$

The density of  $Z$  is  $2e^{-2x}$  and  $\alpha_{PLT} = 2$  whatever the anisotropy distribution:  $\alpha_{Man} = 2$ .

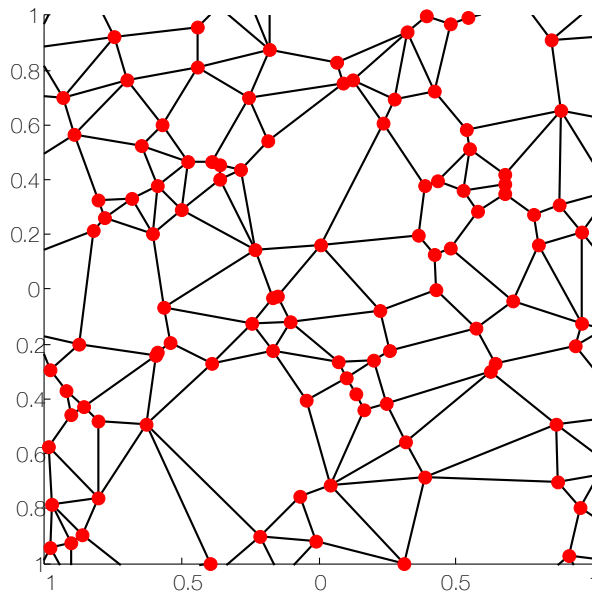
**Crack STIT:** The interior of the typical cell of a Crack STIT is equal in distribution to the PLT's one. Consequently their 0-cells are also the same and  $\alpha_{Crack} = \alpha_{PLT} = 2$  whatever the anisotropy distribution.

### 9.3 High scale model: Gabriel Graph

To the scale previously described superposes most of the time an upper scale. This scale can be an original frame such as a former rampart or an additional non coherent structure to improve traffic such as Haussman's streets in Paris. These streets provide an efficient access to strategically points in the city: it is somehow a relational geometrical graph. Which leads to think of a Poisson Delaunay Tessellation to model it. But the mean degree of PDT is 6 which is too much. We propose thus to study a sub-sampling of the Delaunay graph: the Gabriel graph.

#### 9.3.1 Construction

Let  $S = \{x_1, \dots, x_k, \dots\}$  a discrete set of points. To construct the Gabriel Graph of  $S$  we consider all element of  $S$  is a vertex. For each pair of vertices  $(x_i, x_j)$  an edge is drawn if and only if the circle of diameter  $[x_i, x_j]$  contains no points of  $S$  excepted from  $x_i$  and  $x_j$ .



**Figure 9.7:** Realization in a bound window of a Poisson Gabriel Graph.

#### 9.3.2 Properties

The Gabriel Graph is planar, it is a sub graph of the Delaunay graph. If we replace the deterministic point set  $S$  by a Poisson Point Process and construct the associated Gabriel Graph, we obtain a Poisson Gabriel Graph. This graph is connected with probability 1

since the Minimum spanning tree is connected with probability 1 [97] and is a subgraph of Gabriel. We will see that this graph is not necessarily a tessellation.

### 9.3.3 Typical node

To assess mean formulae of a PGT we simulate the edges emanating from the typical point of the tessellation.

**Radial Simulation of a Poisson Point Process** We fix a point  $o = (0, 0)$  and simulate radially from it a Poisson Point Process  $\Psi$  of intensity  $\gamma$ . We consider the random points  $P_1, P_2, \dots, P_n$  being the  $n$ -th closest point of  $\Psi$  to  $o$ . We write  $R_n = \|P_n\|$  the distance of  $P_n$  to the origin and  $\Theta_n$  the angle between  $oP_n$  and a reference vector, let us say  ${}^t(0, 1)$ .  $\Theta_n$  are independent variables uniformly distributed on  $[0, 2\pi[$ .  $R_{n+1} = R_n + X_{n+1}$  where  $X_n$  are positive random variables, independent and with the same distribution than a variable  $X$ :

$$\mathbb{P}(X > r) = e^{-\lambda\pi r^2}, \quad f_X(r) = 2\pi\lambda r e^{-\lambda\pi r^2} \quad (9.13)$$

$X$  can be simulated from a variable  $Y$  uniformly distributed with the variable change:

$$X = \sqrt{-\frac{\ln(1-Y)}{2\pi\lambda}}, \quad Y \sim \mathcal{U}([0, 1]) \quad (9.14)$$

**Simulation of the edges emanating from the typical node**  $P_1$  is necessarily connected to  $o$ . Then we draw  $P_2$  will he be connected ?  $P_2$  is connected if and only if

$$\|P_2\|^2 < \left\| P_1 - \frac{P_2}{2} \right\|^2 \quad (9.15)$$

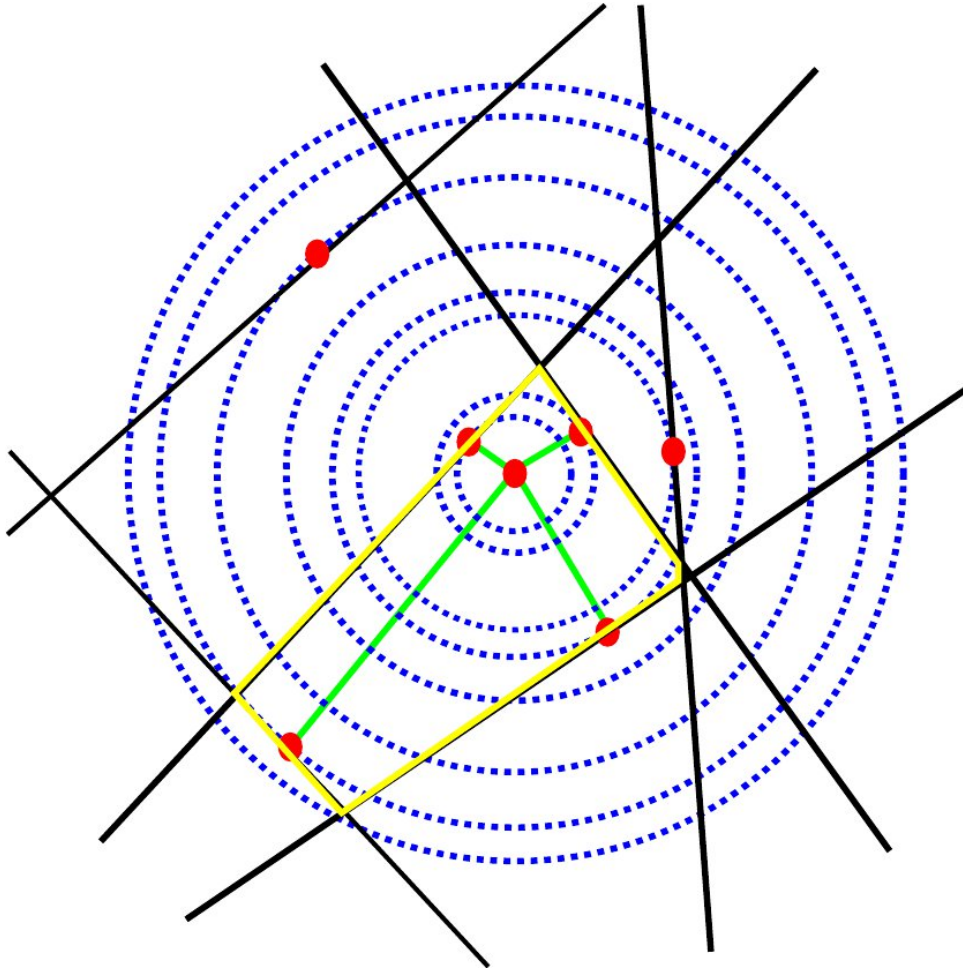
a short calculus with an appropriate basis rotation leads to consider the line  $H_1$  that passes through  $P_1$  and is tangent to the circle of center  $o$  and radius  $R_1$  and the half plane delimited by  $H_1$  that contains  $o$ :  $\Delta_1^o$ .  $P_2$  is connected iff  $P_2 \in \Delta_1^o$ . More generally to each point  $P_i$  we associate  $H_i$  and  $\Delta_i^o$ . Remark that only  $P_1, \dots, P_n$  can prevent  $P_{n+1}$  from connecting to  $o$ . Thus  $P_{n+1}$  connects iff

$$P_{n+1} \in \bigcap_{i=1}^n \Delta_i^o = \text{Poly}_n$$

Note that  $P_n \in \Delta_k^o$  iff  $(\langle P_n | P_k \rangle) < \|P_k\|$

$\bigcap \Delta_i^o$  is a polytope that contains  $o$  and is decreasing with  $n$ . In the beginning the polytope is infinite, then it ends up to close with probability 1. We write  $M_n = \max_{x \in \bigcap \Delta_i^o} \|x\|$ . From these remarks and notations results an algorithm simulating the edges emanating from the typical node of a PGT:

It ends with probability 1 and a large number (10000) of independent simulations permits to estimate the mean degree of a typical node  $\bar{n}_{02} = 3.043$  and the mean length of edges emanating from the typical node:  $\bar{l}_0 = 3.3013/\sqrt{\gamma}$ .



**Figure 9.8:** Radial simulation of the edges emanating from the typical node of a Poisson Gabriel Graph.

Remark that we do not get a tessellation but a graph since nodes can have degree 1 with a probability of about 0.04: PGG. In this very exceptional case, we transform a PGG to a PGT: if  $x$  is a node of degree 1 in PGG, we can for instance add to the graph its Delaunay edges; the graph stays planar and there is no more nodes of degree 1.

### 9.3.4 Mean Formulae

From  $\bar{n}_{02} \simeq 3.043$  and  $\bar{l}_0 \simeq \frac{3.3013}{\sqrt{\gamma}}$  classical formulae permits to get the value of components of the intensity vector of a PGG:

$$\begin{aligned}\lambda_0 &= \gamma \\ \bar{n}_{02} &= 2 + \frac{2\lambda_2}{\lambda_0} \implies \lambda_2 \simeq 0.5215\gamma \\ \lambda_1 &= \lambda_0 + \lambda_2 \implies \lambda_1 = 1.5215\gamma \\ \bar{l}_0 &= \frac{2L_A}{\lambda_0} \implies \lambda_3 = L_A = 1.6507\sqrt{\gamma}\end{aligned}$$

---

**Algorithm 10** Simulation of edges emanating from the typical vertex of a Gabriel Graph

---

```

1: n=0
2: Set  $M_n = \infty$ , set  $\text{Poly}_n = \mathbb{R}^2$ , set  $\text{Edges} = \emptyset$ 
3:  $R_n = 0$ 
4: while  $R_n < M_n$  do
5:   n=n+1
6:   Generate  $P_n$ 
7:   if
8:     then  $\text{Edges} = \text{Edges} \cup [oP_n]$ 
9:   end if
10:  Compute
11:   $\text{Poly}_n = \text{Poly}_{n-1} \cap \Delta_n^o$ 
12:   $M_n = \max_{x \in \text{Poly}_n} \|x\|$ 
13: end while
14: OUTPUT:  $\#\text{Edges}$ ,  $\text{length}(\text{Edges})$ 

```

---

## 9.4 Identification

We propose to see the city as an iterated tessellation. The upper tessellation is a Gabriel graph, the lower either a Crack or a PLT. At first we compute a simplest centrality and threshold it to an arbitrary quantile to select the upper scale. The parameter of the PGT is then assessed. We propose to select low models with a matrix, by classifying them according to the Anisotropy and the Topology of the street network. Once the model selected, we will assess its intensity with an additional dead end process.

### 9.4.1 Rewriting of intensity vectors as feature vectors

We define the modified intensity vector:

$$\vec{\Lambda} = {}^t(\lambda_0, \lambda_1, \lambda_2, \lambda_3^2) \quad (9.16)$$

If we write

$$L = \lambda_3^2 \quad (9.17)$$

then there exists for each stationary tessellation of type  $i$  a **feature vector**  $\vec{v}_i$  such as

$$\vec{\Lambda}_i = L \cdot \vec{v}_i \quad (9.18)$$

For the models under consideration Tab.9.2 gathers vectors  $\vec{v}_i$ : exactly when possible and numerically. We can also calculate  $\vec{\Lambda}$  for the iteration of two models (at least one have to be isotropic) and deduce from it the feature vector of the iteration:

$$\vec{\Lambda}^{(1/2)} = L_1 \vec{v}_1 + L_2 \vec{v}_2 + \sqrt{L_1 L_2} \vec{I} \quad (9.19)$$

where

$$\vec{I} = {}^t\left(\frac{4}{\pi}, \frac{6}{\pi}, \frac{2}{\pi}, 0\right) \quad (9.20)$$

is an interference vector that does not depend on the particular tessellation models used.

Comp.	PVT	PDT	PLT	Manathan	Crack	IsoCrack	Gabriel
0	1/2	$9\pi^2/1024$	$1/\pi$	1	$2/\pi$	1	?
	0.5	0.0867	0.3183	1	0.6366	1	0.3670
1	3/4	$27\pi^2/1024$	$2/\pi$	2	$3/\pi$	3/2	?
	0.75	0.2602	0.6366	2	0.9549	1.5	0.1914
2	1/4	$9\pi^2/512$	$1/\pi$	1	$1/\pi$	1/2	?
	0.25	0.1735	0.3183	1	0.3183	0.5	0.5584
3	1	1	1	1	1	1	1

**Table 9.2:** Feature vectors for the random graph models under consideration (exact value when known and numerical value). The normalization of the last component allows to compare the topology of models.

### 9.4.2 Identification

Let  $u$  be the real feature vector of a map extraction. The classical identification of simple models can then be done through the criterion:

$$C_i = \|L.\vec{v}_i - \vec{u}\|^2 \quad (9.21)$$

The minimum is reached for

$$L = \frac{\langle \vec{v}_i, \vec{u} \rangle}{\|\vec{v}_i\|^2} \quad \text{and is} \quad C_i^{min} = \|\vec{u}\|^2 - \frac{\langle \vec{v}_i, \vec{u} \rangle^2}{\|\vec{v}_i\|^2} = \|\vec{u}\|^2 \cdot \sin^2(\vec{u}, \vec{v}_i) \quad (9.22)$$

To consider relative distances rather than absolute ones, we are led to optimize criteria of the form:

$$C_i = \langle L.\vec{v}_i - \vec{u} | \mathbf{A} | L.\vec{v}_i - \vec{u} \rangle \quad \mathbf{A} = \text{diag}(\vec{u})^{-1} \quad (9.23)$$

Similarly the minimum is reached for

$$L = \frac{\langle \vec{v}_i | \mathbf{A} | \vec{u} \rangle}{\langle \vec{v}_i | \mathbf{A} | \vec{v}_i \rangle} \quad \text{and is} \quad C_i^{min} = \langle \vec{u} | \mathbf{A} | \vec{u} \rangle - \frac{\langle \vec{v}_i | \mathbf{A} | \vec{u} \rangle^2}{\langle \vec{v}_i | \mathbf{A} | \vec{v}_i \rangle} = \langle \vec{u} | \mathbf{A} | \vec{u} \rangle \cdot \sin^2(\vec{u}, \mathbf{A}.\vec{v}_i) \quad (9.24)$$

## 9.5 Algorithm

We now have models and basic methods to choose the best one in a practical context. Rather than using directly the formulae in [57] that average information quite brutally and present symmetry problems, we first separate each scale of the city and identify the best model via intrinsic parameters before assessing their intensity. The algorithm we present here is not the single method respecting the general philosophy.

In a first time we compute simplest centrality on the map to threshold the most central streets as the upper scale of the system. This subgraph is identified to a PGG with linear intensity  $L_{Up}$ . The rest of the network (lower scale) is identified to a Crack STIT or to a PLT (Fig.9.6) from its morphology (intrinsic parameters). The choice is firstly made on the mean degree of the lower scale. To this we calculate the organic ratio  $R$  which essentially

depends on the number of degree 3 vertices  $N(3)$  which is biased because of the interference between upper and lower scales. There are  $\frac{4}{\pi}\sqrt{L_{Up}}(\sqrt{L} - \sqrt{L_{Up}})$  interference vertices of degree 3.

When the model class is chosen, we decide if the lower scale is rather isotropic or anisotropic with the calculus of the anisotropy coefficient.

The last thing is to assess the intensities of the lower scale tessellation and of the dead-ends process. This can be robustly with various methods in the same time or dead-ends first and tessellation after.

---

**Algorithm 11** Choice of similar iterated random tessellation

---

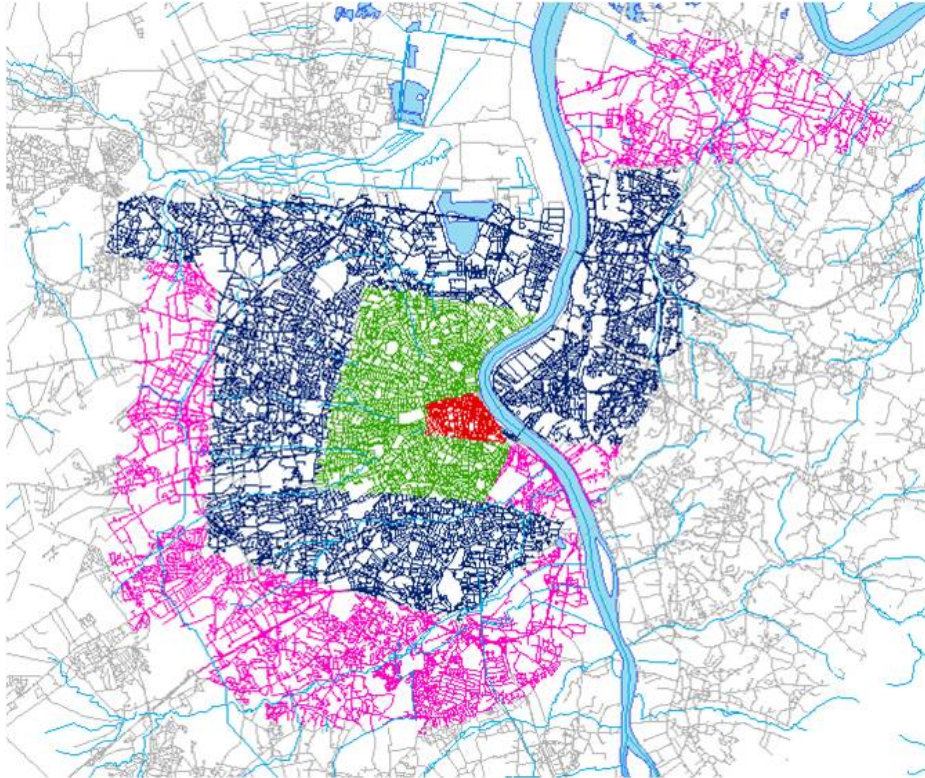
- 1: Extract the map,  $L$  is the total street length
  - 2: Compute the simplest centrality on the road network
  - 3: Extract the sub graph  $Up$  constituted of roads whose centrality  $> c$  (arbitrary around 10 percent for instance).
  - 4: Compute  $L_{Up}$  the square of the linear intensity of the upper scale.
  - 5: Compute  $R = \tilde{N}(3)/(\tilde{N}(3) + N(4))$  where  $\tilde{N}(3) = N(3) - N(1) - \frac{4}{\pi}\sqrt{L_{Up}}(\sqrt{L} - \sqrt{L_{Up}})$  the organic ratio of the lower scale (the degree 3 interference vertices are subtracted).
  - 6: **if**  $R \leq 0.5$  **then**
  - 7: Model class is CRACK
  - 8: **else** Model class is LINE
  - 9: **end if**
  - 10: Compute the anisotropy  $A$
  - 11: **if**  $A \leq 0.5$  **then**
  - 12: model is isotropic
  - 13: **else** Model class is anisotropic.
  - 14: **end if**
  - 15: Compute  $N(1)$  the number of dead ends.
  - 16: Estimate parameters of the tessellation.
- 

## 9.6 Conclusion

For dead ends, we can either consider that it projects to the iteration of high and low scale either that it only projects to the low scale. In the first case we must approximate  $\alpha_{\text{iteration}}$  by  $\alpha_{\text{Low scale}}$ , in the second case we must admit that because of a very few occurrences the graph is not planar. Also, the subgraph resulting from the truncation of highly central streets may not be connected which induces a bias. An other fundamental assumption in this chapter is zones under consideration are stationary. It is obviously not the case when looking at whole cities or agglomerations. The stable condition of the network is modified by an intensity modulation, cultural differences inside the city and structuring elements. But one could consider that locally the city is stationary. The next step is then to find a way to cut a map into a small number of zones where the street network can indeed be considered as stationary. This can be done "by hand" as in Fig.9.9 but the result is subjective, not



trustable, not accurate and painful to get. The next chapter presents an algorithm that realizes in a satisfactory fashion this task.



**Figure 9.9:** Hand segmentation of the city of Bordeaux. From Catherine Gloaguen's.

# Chapter 10

## Map Segmentation

### Contents

---

10.1	Spectral Clustering . . . . .	<b>167</b>
10.1.1	K-Means . . . . .	167
10.1.2	Principle of Spectral-Clustering . . . . .	167
10.1.3	Formal approach . . . . .	170
10.1.4	Algorithm . . . . .	172
10.2	City Segmentation Algorithm . . . . .	<b>173</b>
10.2.1	Elementary region graphs . . . . .	173
10.2.2	Similarity function . . . . .	174
10.2.3	Algorithm . . . . .	174
10.2.4	Tunning . . . . .	174
10.3	Results . . . . .	<b>175</b>
10.3.1	Toy cases . . . . .	175
10.3.2	Real cities . . . . .	177

---

### Map Segmentation: Synthesis

The stochastic geometry models only apply to regions where the map can be considered as "stationary". Nonetheless a whole map is in any way stationary or homogeneous.

The purpose of this chapter is to build and assess the performances of algorithm to segment a map in a given number of zones one can consider as "morphologically homogeneous". We deal with *morphological* segmentation and not *functional* segmentation. The flexibility of the recent **Spectral Clustering** has caught our attention: it uses graph formalism, allows incorporating varied segmentation criteria, generates non linear boundaries between zones contrary to the well known **K-means** algorithm.

Roughly, the principle of Spectral Clustering is to operate a change of representation space on the data, to apply then K-means which makes linear boundaries and re project these boundaries on the first space, the result presenting varied shapes adapted to data. At first we will present Spectral Clustering: heuristically its principle in terms of matrix perturbation, more rigorously but also more obscurely we will see it minimizes some criteria. In a second part we will present our refinements to Spectral Clustering for **city segmentation**. To end with, we will study the tuning and the relevance of our algorithm with toy cases and real cities.

### Contributions of this chapter

1. A review and an understanding on spectral clustering .
2. A spectral clustering based algorithm to morphologically cluster city maps.
3. The tuning of this algorithm with representative toy cases and real cities; its implementation in linear time.
4. Its application to equivalent stationary tessellation identification.

## 10.1 Spectral Clustering

Before presenting Spectral Clustering, we have to say a few words on the more classical K-means algorithm. K-Means is a clustering algorithm which classifies data with linear separators. Spectral Clustering is an evolution of K-Means producing non linear separators and using K-Means as an element. If not recalled in the text, our presentation of K-means and spectral clustering is based on [7, 130, 94, 99, 65].

### 10.1.1 K-Means

Let  $(x_1, \dots, x_n)$  be  $n$  samples of  $d$ -dimensional data in a metric space. The purpose of K-means is to partition these data into  $K$  classes or clusters such as samples in a same class are "close" and samples of different classes are "distant".

For instance if  $(x_1, \dots, x_n)$  is sampled from a Gaussian mixture, classes should contain samples from the same Gaussian generator.

More formally, we aim at finding the partition  $S$ :

$$S = (S_1, \dots, S_K) = \underset{s=(s_1, \dots, s_K)}{\operatorname{argmin}} \sum_{i=1}^K \sum_{x_j \in S_i} \|x_j - \mu_i\|^2, \quad \mu_i = \frac{1}{\#S_i} \sum_{x \in S_i} x \quad (10.1)$$

The problem as formulated here is NP-hard. The K-means algorithm is a fast to run heuristic to solve that problem. Expectation-Minimization like, it divides into 2 steps and is iterative. To start with,  $K$  vectors  $(\mu_i^{(0)})_i$  are chosen at random (in a set adapted to data, for instance their convex hull). A sample is affected to the  $i$ -th cluster if its belongs to the Voronoï cell of  $\mu_i^{(0)}$  relatively to the sequence  $(\mu_j)$ :  $\operatorname{Vor}(\mu_i^{(n)} | \mu_1^{(n)}, \dots, \mu_K^{(n)})$  (Assignment). Each  $\mu_i^{(n)}$  is updated to the mean of all sample in its Voronoï cell (Update). And so on till the clusters stabilize.

---

#### Algorithm 12 K-means

---

- 1:  $n=1$
  - 2:  $\mu_1^{(n)}, \dots, \mu_K^{(n)}$  at random
  - 3: **repeat**
  - 4:    $S_i^{(n)} = \{x_j, x_j \in \operatorname{Vor}(\mu_i^{(n)} | \mu_1^{(n)}, \dots, \mu_K^{(n)})\}$  ▷ Assignment Step
  - 5:    $\forall i, \mu_i^{(n+1)} = \frac{1}{\#S_i^{(n)}} \sum_{x \in S_i^{(n)}} x$  ▷ Update Step
  - 6: **until** clusters stabilize
- 

### 10.1.2 Principle of Spectral-Clustering

The K-Means algorithm is based on Voronoï diagrams and thus cannot produce non linear separators. The basic idea of Spectral Clustering is to carry out a change of representation space and to use a linear separation in this new space. The resulting segmentation in the initial space is not linear and can adopt more intricate shapes.

To this Spectral Clustering starts by transforming data into a weighted graph and operates its changes via the eigenspaces of the Laplacian matrix.

**Similarity graph** Let  $(x_1, \dots, x_n)$  be the  $d$ -dimensional samples to cluster. Let  $w(.,.)$  be a positive function intended to measure to what extent data  $x_i$  and  $x_j$  are "close" or "similar". The similarity graph of  $(x_1, \dots, x_n)$  is the graph having these samples as vertices and that make an edge between  $x_i$  and  $x_j$  of weight  $w(x_i, x_j)$  ( $x_i$  and  $x_j$  are not connected if  $w(x_i, x_j) = 0$ ).

**Laplacian Matrix** If  $G = (V, E)$  is a graph, with a weighted adjacency matrix  $\mathbf{W}$  and if  $\mathbf{D}$  is the  $\#V \times \#V$  diagonal matrix whose entry  $d_{ii}$  is the degree of vertex  $i$ , the Laplacian matrix  $\mathbf{L}$  of  $G$  is defined as

$$\mathbf{L} = \mathbf{D} - \mathbf{W} \quad (10.2)$$

**Theorem 10.1.1**

The Laplacian  $\mathbf{L}$  checks:

1. If  $f \in \mathbb{R}^n$ ,

$${}^t f \mathbf{L} f = \frac{1}{2} \sum_{i,j} w_{ij} (f_i - f_j)^2$$

2.  $\mathbf{L}$  is symmetric and positive semi-definite, all its eigenvalues are positive

3. 0 is eigenvalue of  $L$  with  ${}^t(1, \dots, 1)$  as eigenvector

4. The multiplicity of the eigenvalue 0 is the number of connected components in the graph

**Ideal case** If the data is compounded of  $k$  obvious clusters i.e  $k$  connected components then modulo a basis change, the Laplacian is a square block diagonal matrix:

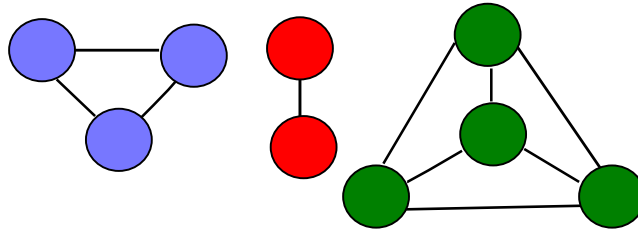
$$\mathbf{L} = \begin{pmatrix} \mathbf{L}_1 & & & \\ & \mathbf{L}_2 & & \\ & & \ddots & \\ & & & \mathbf{L}_k \end{pmatrix} \quad (10.3)$$

$\mathbf{L}_i$  runs from index  $l_i$  to  $l_{i+1}$ . 0 is eigenvalue of multiplicity  $k$ . The associated eigenspace is spanned by vectors

$$\vec{\mathbf{1}}_k = {}^t(0, \dots, 0, \underbrace{1, \dots, 1}_{\text{indices } l_k \text{ to } l_{k+1}}, 0, \dots, 0) / (l_{k+1} - l_k) \quad (10.4)$$

which form an orthonormal basis. Let  $\mathbf{V}$  the  $n \times k$  matrix that contains the orthonormal eigenvectors spanning the eigenspace of 0.  $\mathbf{V}$  cannot be defined in an univocal way but since  $(\vec{\mathbf{1}}_k)_k$  is an orthogonal basis, there exist an orthogonal matrix  $\mathcal{R}$  such that:

$$\mathbf{V} = \mathcal{R} \begin{pmatrix} \vec{\mathbf{1}}_1 & & & \\ & \vec{\mathbf{1}}_2 & & \\ & & \ddots & \\ & & & \vec{\mathbf{1}}_k \end{pmatrix} \quad (10.5)$$



**Figure 10.1:** A graph constituted of 3 connected components

let us write  $(u_i)_{i=1:n}$  the lines of  $\mathbf{V}$ . Then two points  $x_i$  and  $x_j$  are in the same cluster if and only if  $u_i = u_j$ . Thus clustering  $(x_i)$  is the same than clustering  $(u_i)$ .

This equivalence remains if the data is not ideal: the Laplacian  $\mathbf{L}$  is a little "perturbed" but if  $x_i$  and  $x_j$  are strongly connected,  $u_i$  and  $u_j$  keep being close as shown by the example below.

**Example** We show here the different matrix  $\mathbf{W}$ ,  $\mathbf{L}$  and  $\mathbf{V}$  of the graph in Fig.10.1:

$$\mathbf{W} = \begin{pmatrix} 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \end{pmatrix} \tag{10.6}$$

$$\mathbf{L} = \begin{pmatrix} 2 & -1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ -1 & 2 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ -1 & -1 & 2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 3 & -1 & -1 & -1 \\ 0 & 0 & 0 & 0 & 0 & -1 & 3 & -1 & -1 \\ 0 & 0 & 0 & 0 & 0 & -1 & -1 & 3 & -1 \\ 0 & 0 & 0 & 0 & 0 & -1 & -1 & -1 & 3 \end{pmatrix} \tag{10.7}$$

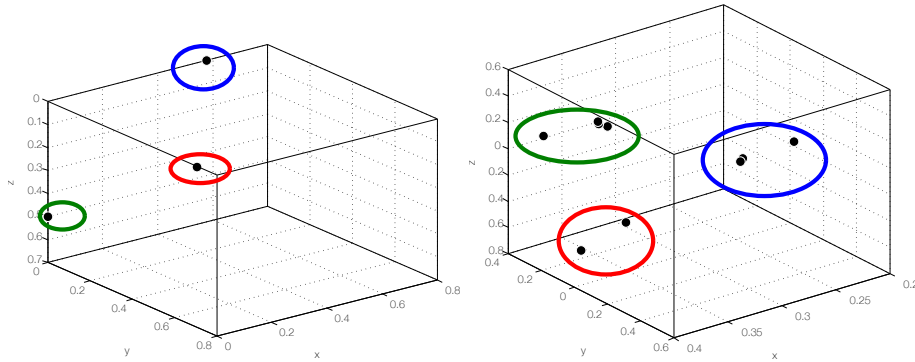
The eigenvalues are 0, 0, 0, 2, 3, 3, 4, 4, 4 with eigenvectors:

$$\mathbf{V} = \begin{pmatrix} 0.5774 & 0 & 0 & 0 & 0.2673 & 0.7715 & 0 & 0 & 0 \\ 0.5774 & 0 & 0 & 0 & -0.8018 & -0.1543 & 0 & 0 & 0 \\ 0.5774 & 0 & 0 & 0 & 0.5345 & -0.6172 & 0 & 0 & 0 \\ 0 & -0.7071 & 0 & -0.7071 & 0 & 0 & 0 & 0 & 0 \\ 0 & -0.7071 & 0 & 0.7071 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -0.5000 & 0 & 0 & 0 & -0.1296 & 0.7344 & -0.4402 \\ 0 & 0 & -0.5000 & 0 & 0 & 0 & -0.6130 & -0.5745 & -0.2101 \\ 0 & 0 & -0.5000 & 0 & 0 & 0 & 0.7785 & -0.3226 & -0.1996 \\ 0 & 0 & -0.5000 & 0 & 0 & 0 & -0.0359 & 0.1627 & 0.8499 \end{pmatrix} \tag{10.8}$$

Let us now pretrubate  $\mathbf{W}$  by introducing loose links between each pair of vertices:

$$\mathbf{W} = \begin{pmatrix} 1.0000 & 1.1914 & 1.1760 & 0.1959 & 0.0505 & 0.1475 & 0.0024 & 0.0398 & 0.1323 \\ 1.1045 & 1.0000 & 1.0938 & 0.1977 & 0.0847 & 0.0668 & 0.0452 & 0.1521 & 0.1281 \\ 1.0346 & 1.0130 & 1.0000 & 0.0760 & 0.1362 & 0.1136 & 0.0118 & 0.0101 & 0.0610 \\ 0.0543 & 0.1166 & 0.1567 & 1.0000 & 1.0030 & 0.1942 & 0.1578 & 0.0997 & 0.1287 \\ 0.1751 & 0.1031 & 0.0922 & 1.1536 & 1.0000 & 0.1920 & 0.0824 & 0.0536 & 0.1867 \\ 0.0273 & 0.0866 & 0.1588 & 0.1980 & 0.1453 & 1.0000 & 1.0425 & 1.1258 & 1.0414 \\ 0.1788 & 0.1160 & 0.1206 & 0.0877 & 0.1489 & 1.1678 & 1.0000 & 1.1260 & 1.1150 \\ 0.0597 & 0.1060 & 0.0831 & 0.0428 & 0.0880 & 1.0268 & 1.0741 & 1.0000 & 1.0088 \\ 0.0569 & 0.0418 & 0.1749 & 0.0640 & 0.1367 & 1.1214 & 1.0903 & 1.0054 & 1.0000 \end{pmatrix} \tag{10.9}$$

$$\mathbf{L} = \begin{pmatrix} 2.6913 & -1.1914 & -1.1760 & -0.1959 & -0.0505 & -0.1475 & -0.0024 & -0.0398 & -0.1323 \\ -1.1045 & 2.7743 & -1.0938 & -0.1977 & -0.0847 & -0.0668 & -0.0452 & -0.1521 & -0.1281 \\ -1.0346 & -1.0130 & 3.0561 & -0.0760 & -0.1362 & -0.1136 & -0.0118 & -0.0101 & -0.0610 \\ -0.0543 & -0.1166 & -0.1567 & 2.0157 & -1.0030 & -0.1942 & -0.1578 & -0.0997 & -0.1287 \\ -0.1751 & -0.1031 & -0.0922 & -1.1536 & 1.7932 & -0.1920 & -0.0824 & -0.0536 & -0.1867 \\ -0.0273 & -0.0866 & -0.1588 & -0.1980 & -0.1453 & 4.0300 & -1.0425 & -1.1258 & -1.0414 \\ -0.1788 & -0.1160 & -0.1206 & -0.0877 & -0.1489 & -1.1678 & 3.5064 & -1.1260 & -1.1150 \\ -0.0597 & -0.1060 & -0.0831 & -0.0428 & -0.0880 & -1.0268 & -1.0741 & 3.6124 & -1.0088 \\ -0.0569 & -0.0418 & -0.1749 & -0.0640 & -0.1367 & -1.1214 & -1.0903 & -1.0054 & 3.8020 \end{pmatrix} \tag{10.10}$$



**Figure 10.2:** The lines of the matrix constituted of the tree first eigenvectors ( $V$ ) of the similarity matrix ( $W$ ) as columns. On the LEFT: the ideal case where green, red and blue subgraphs are not connected at all. On the RIGHT: the perturbed case where these subgraphs are loosely connected. In both cases the clusters resulting from K-means algorithm are surrounded by the color corresponding to the subgraph of Fig.10.1.

Eigenvalues: 0, 0.879, 1.0646, 2.9919, 3.8798, 4.006, 4.666, 4.7556, 5.0386

$$\mathbf{v} = \begin{pmatrix} -0.3240 & -0.5080 & 0.2970 & 0.0239 & 0.6785 & 0.4444 & 0.0251 & -0.0626 & 0.0041 \\ -0.3147 & -0.4603 & 0.2688 & 0.0039 & -0.7260 & 0.3240 & -0.0480 & 0.0160 & 0.0479 \\ -0.2604 & -0.4185 & 0.2335 & -0.1042 & 0.0469 & -0.8205 & 0.0414 & -0.0012 & -0.0258 \\ -0.3330 & 0.1253 & -0.5494 & 0.7146 & 0.0477 & 0.0506 & 0.0156 & 0.0103 & -0.0098 \\ -0.3748 & 0.1245 & -0.6628 & -0.6901 & -0.0495 & -0.0375 & 0.0079 & -0.0382 & -0.0019 \\ -0.3277 & 0.2699 & 0.0757 & 0.0387 & -0.0297 & -0.0374 & -0.8379 & 0.1607 & -0.1809 \\ -0.3827 & 0.2993 & 0.1212 & 0.0074 & 0.0542 & 0.0961 & 0.2003 & -0.5639 & 0.6975 \\ -0.3339 & 0.2870 & 0.1152 & 0.0101 & -0.0397 & 0.0595 & 0.1035 & -0.2782 & -0.6758 \\ -0.3340 & 0.2809 & 0.1008 & -0.0044 & 0.0175 & -0.0792 & 0.4920 & 0.7571 & 0.1448 \end{pmatrix} \quad (10.11)$$

### 10.1.3 Formal approach

**Graph cut** The graph cut problem for a weighted graph  $G$  (weight  $w$ ) consists in dividing  $G$  into two or more disjoint subgraphs that minimize a criterion. This criterion expresses the idea that vertices in a same subgraph should be highly connected and vertices in different subgraphs should be loosely connected. For a graph cut into two components, the simplest criterion to work on is the minimization of the  $\text{Cut}_2$  of a partition:

$$\text{Cut}_2(A, B) = \sum_{i \in A, j \in B} w(x_i, x_j) \quad (10.12)$$

This problem can be solved in  $O(\#V^2\#E)$  with the Edmonds Karp algorithm. Nonetheless, the result is somewhat unbalanced, the solution of the problem being mainly the isolation of a few vertices. Another criterion based on the  $\text{RatioCut}_2$  minimization gets around this problem by introducing a regularization factor:

$$\text{RatioCut}_2(A, B) = \text{Cut}_2(A, B) \left( \frac{1}{\#A} + \frac{1}{\#B} \right) \quad (10.13)$$

the  $\text{RatioCut}_2$  being put up if a subgraph is small. Finding the solution of the  $\text{RatioCut}_2$  problem is  $NP$ -hard. The purpose of the spectral clustering algorithm is to set up an heuristic that approximates its solution.

The two criteria can be generalized to cuts into  $k$  subgraphs:

$$\text{Cut}(A_1, \dots, A_k) = \frac{1}{2} \sum_i \text{Cut}_2(A_i, \bar{A}_i) \quad (10.14)$$

$$\text{RatioCut}(A_1, \dots, A_k) = \frac{1}{2} \sum_i \frac{\text{Cut}_2(A_i, \bar{A}_i)}{\#A_i} \quad (10.15)$$

**Resolution with two clusters** Given a subset  $A$  of  $G$ , we define the vector  $f = {}^t(f_1, \dots, f_n)$  by:

$$f_i = \begin{cases} \sqrt{\#A/\#\bar{A}} & \text{if } x_i \in A \\ -\sqrt{\#\bar{A}/\#A} & \text{if } x_i \in \bar{A} \end{cases} \quad (10.16)$$

then

$$\text{RatioCut}(A, \bar{A}) = \frac{1}{\#A} {}^t f \mathbf{L} f \quad (10.17)$$

furthermore,

$$\|f\| = \sqrt{n} \quad (10.18)$$

and

$$f \perp {}^t(1, \dots, 1) \quad (10.19)$$

The problem of minimization of RatioCut can be equivalently rewritten as

**Find  $f$  as defined in 10.16 that minimizes 10.17 which respects 10.18 and 10.19**

This problem is still *NP*-hard. But we are on the verge to use the theorem:

**Theorem 10.1.2 (Rayleigh-Ritz (1))**

If  $A$  is a positive definite symmetric matrix with eigenvalues  $\lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_n$  associated to normalized eigenvectors  $V_1, \dots, V_n$  then

$$\min_{x, \|x\|=\sqrt{C}} {}^t x A x = C \cdot \lambda_1$$

and that minimum is reached for  $x = \sqrt{C} \cdot V_1$ .

$$\min_{\substack{x, \|x\|=\sqrt{C}, \\ x \perp V_1, \dots, x \perp V_k}} {}^t x A x = C \cdot \lambda_{k+1}$$

and the minimum is reached for  $x = \sqrt{C} \cdot V_{k+1}$

If we relax the 10.16 condition (on discrete values of  $f$ ), the problem rewrites in the hypothesis of the theorem and thus its solution is the second eigenvector  $V_2$  of the Laplacian matrix.

The only thing left to do is to transform the solution  $V_1$ . The idea is to cluster its components into two groups, for instance by using a K-means algorithm.



**General resolution with  $k$  clusters** Similarly we rewrite the RatioCut criterion with the Laplacian matrix. Given a partition  $A_1, \dots, A_k$  we define the vectors  $h_j = {}^t(h_{1,j}, \dots, h_{n,j})$  by:

$$h_{i,j} = \begin{cases} \sqrt{1/\#A_j} & \text{if } x_i \in A_j \\ 0 & \text{otherwise} \end{cases} \quad (10.20)$$

and  $H$  is the  $n \times k$  matrix containing the vectors  $(h_j)$  as columns.  $H$  is orthonormal :

$${}^t\mathbf{H} \mathbf{H} = \mathbf{I} \quad (10.21)$$

and a short calculus leads to express the RatioCut as:

$$\text{Tr}({}^t\mathbf{H}\mathbf{L}\mathbf{H}) = \text{RatioCut}(A_1, \dots, A_k) \quad (10.22)$$

**Minimizing the ratiocut is equivalent to minimizing 10.22 with  $H$  as defined in 10.20 and the constraint 10.21.**

If we relax the condition on the discrete form of  $H$ : 10.20, the problem can be easily solved by invoking a second form of the Rayleigh-Ritz theorem:

**Theorem 10.1.3 (Rayleigh-Ritz (2))**

If  $\mathbf{A}$  is a positive definite symmetric matrix with eigenvalues  $\lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_n$  associated to normalized eigenvectors  $V_1, \dots, V_n$  then

$$\min_{\mathbf{H} \in \mathbb{R}^{n \times k}, {}^t\mathbf{H}\mathbf{H} = \mathbf{I}} {}^t\mathbf{H}\mathbf{A}\mathbf{H} = \lambda_1 + \dots + \lambda_k$$

and that minimum is reached for  $\mathbf{H} = [V_1, \dots, V_k]$

The only thing left to do is to cluster the lines of  $V$  by using K-Means (for instance).

#### 10.1.4 Algorithm

Here is the simple code to implement Spectral Clustering in the general case once the similarity function  $w(.,.)$  has been chosen.

---

**Algorithm 13** Spectral Clustering

---

- 1: INPUT Data  $(x_1, \dots, x_n)$ , Similarity Function  $w(.,.)$ , Number of Clusters  $k$
  - 2: Compute the similarity graph  $G$  of  $(x_1, \dots, x_n)$
  - 3: Compute  $\mathbf{L}$  the Laplacian of  $G$
  - 4: Compute the  $k$  first eigenvectors of  $\mathbf{L}$ :  $v_1, \dots, v_k$
  - 5: Let  $\mathbf{V} \in \mathbb{R}^n \times \mathbb{R}^k$  the matrix containing  $v_1, \dots, v_k$  as columns
  - 6: Let  $u_1, \dots, u_n \in \mathbb{R}^k$  be the lines of  $\mathbf{V}$
  - 7: Apply k-means on  $(u_i)$ :  $(u_i)$  are clustered into  $C_1, \dots, C_k$
  - 8: Clusters  $A_1, \dots, A_k$  are defined by  $A_i = \{j, u_j \in C_i\}$
-

## 10.2 City Segmentation Algorithm

In the following section we present a Spectral Clustering based algorithm to segment a city graph  $G = (V, E)$  into a given number of clusters. We seek out to divide the map into regions where it can be considered as "morphologically homogeneous" this expression needing to be defined.

To this we start by "**pixellating**" the city graph into **elementary regions**. The topology of this pixellization induces a graph whose vertices are elementary regions and edges are drawn between two regions if they are adjacent. We weight this graph with a positive function quantifying the morphological likeness between two regions. This function is the linear aggregation of likeness criteria based for instance on the density, topology or anisotropy of regions.

To end with we run spectral clustering on the resulting weighted graph.

The stake of our work is then **(1): To define a relevant pixelling (2): To define a relevant criterion to measure similarity.**

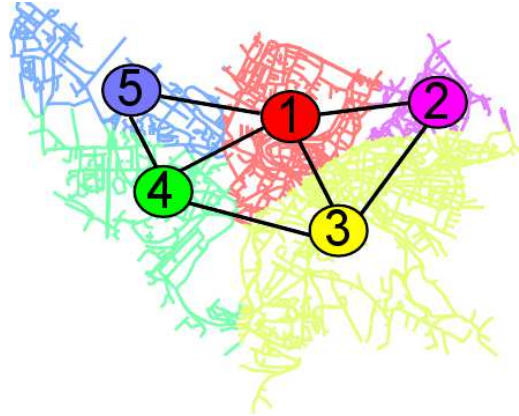
### 10.2.1 Elementary region graphs

Let  $N \in \mathbb{N}$  and  $x_1, \dots, x_n$  be  $N$  random points on  $G$ . To each point  $x_i$  we associate the subgraph or region  $V_i$  that is the Voronoï cell of  $x_i$  relatively to  $x_1, \dots, x_n$  calculated on  $G$  that is to say the set of edges whose mid-point is closer to  $x_i$  than to any other  $x_j$ .

There are several interests to use a Voronoï pixellization. (1): The pixellization is much more isotropic than a square grid and this avoids directional artefacts. (2): The pixellization adapts to the structure of the city. Since centres are chosen uniformly at random on the map the mean length of streets inside a pixel is constant: each pixel contains the same amount of information which would not be the case with a Euclidian Voronoï with centres picked at random on a window (convex hull) containing the map. (3): This Voronoï is easily computed in  $O(n \# E)$ ,  $E$  the number of edges in the map. This is much more simple than an Euclidian Voronoï in  $O(n \log n)$  with an intricate implementation followed by the extraction of the map in each cell (which can be fast but needs the creation of a new heavy structures).

Let  $W$  be a positive function defined for each couple  $(V_i, V_j)$ . The idea will be to define  $W$  to take high value if two regions are morphologically the same and a value close to 0 otherwise.

From regions  $(V_i)_i$  we define various region graphs. **The full region graph** has for vertices the  $V_i$  and is fully connected, each edge  $(V_i, V_j)$  is weighted by  $W(V_i, V_j)$ . **The adjacent region graph** has  $(V_i)$  as vertices but the edge  $(V_i, V_j)$  is drawn only if the two regions are adjacent i.e if they share at least one edge. The edge is then weighted by  $W$ . We also introduce **the mixed region graph** as the complete graph with a positive offset on edges between two adjacent elementary regions. Fig.10.3 illustrates the adjacent region graph for a little number of regions in Angoulême.



**Figure 10.3:** The city of Angoulême divided into 5 elementary regions. Superimposed to it, the associated adjacent region graph.

### 10.2.2 Similarity function

To each elementary region  $V_i$  we extract a feature vector  $\vec{F}_i = (F_i)_k$  where each component is a real number

For two real numbers, we define  $K_\sigma(.,.)$  as the non normalized Gaussian kernel of standard deviation  $\sigma$ :

$$K_\sigma(x, y) = \exp\left(-\frac{(x - y)^2}{2\sigma^2}\right) \quad (10.23)$$

Two regions are compared by mean of a distance  $D$ : the similarity function that aggregates linearly the difference between the features.  $D$  applies to feature vectors and the weight on the region graph  $W$  is canonically associated.

$$W(V_i, V_j) = D(F_i, F_j) = \sum_{k \text{ a feature}} \alpha_k \cdot K_{\sigma_k}((F_i)_k, (F_j)_k) \quad (10.24)$$

The features taken into account here are: **the density, the anisotropy, the ratio between intersections of degree  $\leq 3$  and intersections of degree  $\geq 4$  and the proportion of dead ends.**

$$\vec{F}_i = (\lambda_i, N_i, DE_i, Ani_i, ) \quad (10.25)$$

### 10.2.3 Algorithm

We sum up in the following the algorithm implementing the idea presented above to segment a city map.

### 10.2.4 Tuning

**Scale parameters in the similarity function** The **density** varies quite smoothly on a map. Thus the standard deviation of the intensity on different Voronoi zones is a good assessment for the scale parameter of the intensity likeness function. It could be auto tuned but for real French cities this parameter is typically  $0.0055\text{m}^{-1}$ .

---

**Algorithm 14** Map Segmentation Algorithm

---

- 1: INPUT Graph  $C$ , Similarity Distance  $D$ , Number of Regions  $N$ , Number of Clusters  $k$
  - 2: Pick  $N$  points at random on  $C$
  - 3: Compute the Voronoï cell of these points  $V_1, \dots, V_N$ : the elementary regions
  - 4: Compute a region graph  $G$
  - 5: For each cell calculate the feature vector  $\vec{F}_i = (\lambda_i, N_i, DE_i, Ani_i,)$
  - 6: Weight each edge in  $G$ ,  $e_{ij}$  by  $D(F_i, F_j)$
  - 7: Run Spectral Clustering on  $G$  for  $k$  clusters
- 

Contrary to that, standard deviation for other parameters highly depends on the city under consideration. A city can be homogeneous and thus present an almost null variance: the standard deviation of these features is not a good fashion to tune the scale parameter. It has to be arbitrary chosen to set a relevant variation on features. The parameter 0.02 is quite satisfactory for **the other parameters**.

**Number of regions** There is a compromise to do with the number of regions. If this number is too small, the resulting segmentation will strongly depend on the initial Voronoï graph and will be rough. Conversely, if the number is too large, parameter estimation on each zone will be unstable.

Regions centres are drawn uniformly on the street network rather than on the available space. This choice permits to have an almost constant amount of information among regions.

Heuristically, we found that one region for 10 km of streets is a good compromise. It corresponds to 60 to 80 regions for an average large town.

## 10.3 Results

In this section we asses the quality of the City Segmentation Algorithm, at first in toy cases, secondly in real maps (Lyon Fig.10.7 and Lille Fig.10.8).

In real cases, it is not easy to judge qualitatively of the goodness of segmentation (which shows the interest of our algorithm: to carry out a task an operator cannot objectively do). Thus for each simulation we introduce **the reordered similarity matrix**. It is simply the result of a permutation applied on rows and columns of the similarity matrix in such a way that elementary regions in the same cluster follow each other. If the algorithm is relevant, this matrix should be compounded of diagonal blocks with high values and low values elsewhere. It is displayed with a color scale on a square grid  $n \times n$ .

### 10.3.1 Toy cases

In a first time, to validate the method and chose the type of graph (adjacency, complete, mixed) we created toy cases. The first case is a Gabriel Graph whose underlying vertex process follows spatial Gaussian law. The topology and the anisotropy of this graph is thus

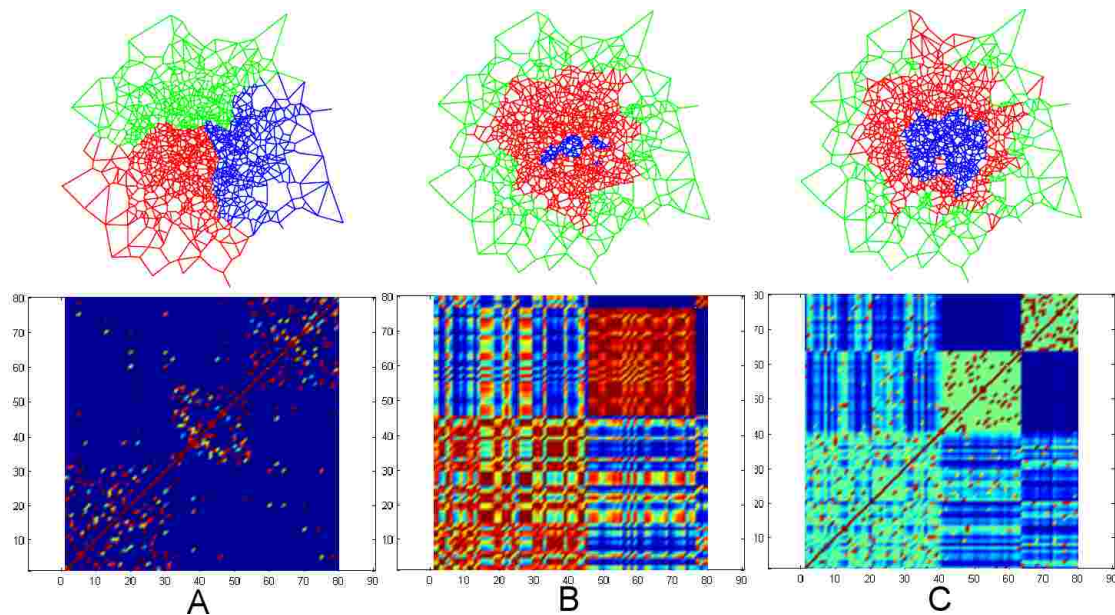
stationary and the intensity is radially modulated.

The second case is compounded of 4 morphologically different parts (Tab.10.1):

<b>Topology:</b> 4 <b>Density:</b> high, <b>Anisotropy:</b> no	<b>Topology:</b> 3.3 <b>Density:</b> low <b>Anisotropy:</b> yes
<b>Topology:</b> 4 <b>Density:</b> high <b>Anisotropy:</b> yes	<b>Topology:</b> 4 <b>Density:</b> low <b>Anisotropy:</b> yes

**Table 10.1:** A toy case constituted of 4 morphologically different square zones.

**Intensity segmentation** Let us start with a single parameter: the density. We created a spatially modulated Gabriel Graph and applied the City Segmentation Algorithm with the three kinds of elementary region graphs under consideration: the adjacency graph, the complete graph and the mixed graph. The density is the only feature used to define the similarity function. The scale parameter is set to the standard deviation of the intensity on elementary regions and we are looking for let us say 3 clusters. The expected result is three crowns of approximately equal areas Fig.10.4. The result with a adjacency graph is not radial, with the complete graph, the central cluster is small than the others. The best result is obtained by the mixed region graph.

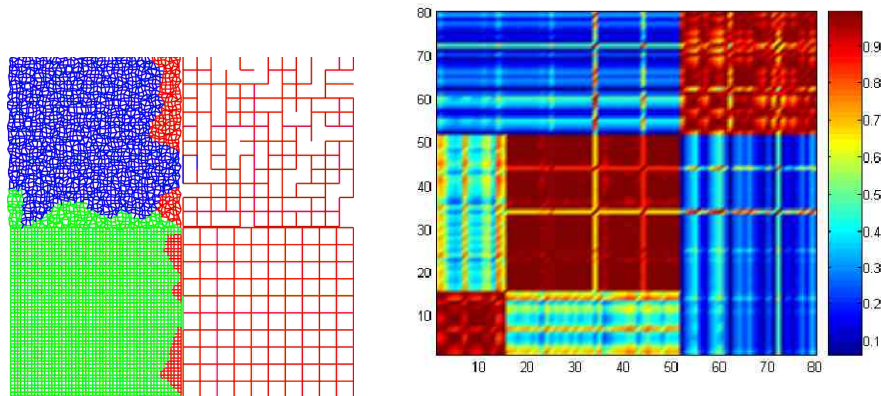


**Figure 10.4:** Segmentation of a Gabriel Graph modulated in intensity into 3 zones. A: the result with an adjacency graph, B: the result with a complete graph, C: the result with a mixed graph

**General segmentation** We use the toy case of Tab.10.1 in different scenarios.

In Fig.10.5 we want to cluster the toy case into three regions discriminated only on the density and the anisotropy. The result is satisfactory whatever the region graph. We just

have to notice expected little smudges at the boundary between different zones, they are inevitably caused by the non zero size of the elementary regions.

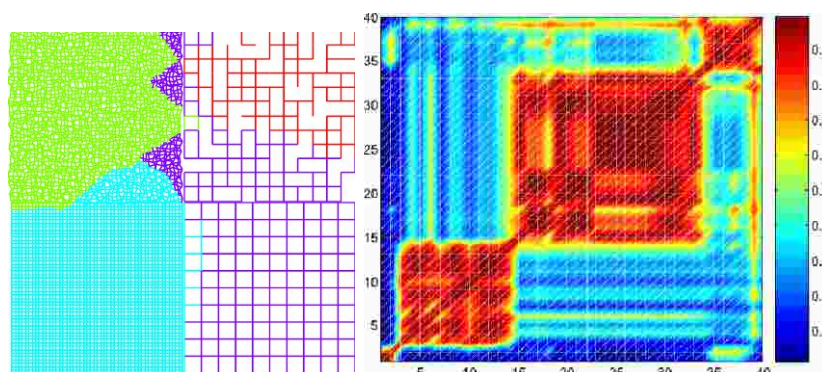


**Figure 10.5:** Segmentation of the toy case into 3 clusters with two criteria: density and anisotropy. The likeness function is calculated on the full graph of 80 Voronoi zones. On the left the toy geometrical graph and its 3 found clusters into 3 different colours. On the right the likeness matrix reordered according to the clusters.

In Fig.10.5 we want to cluster the toy case its for artificial regions. The displayed result is obtained with complete region graph. We notice more smudges and in fact here their causes are multiple. The region size at first but also a edge-effect between the two zones on the right: at the border the topology is three which introduces a bias in the feature vector. This problem can be solved by using rather an adjacent region graph.

### 10.3.2 Real cities

Eventually, in real city cases Fig.10.7 and Fig.10.8, adjacent graph appears to be the most efficient option. It provides the most spatially coherent results. Furthermore it is faster to compute. The overall result is difficult to judge visually, the notion of homogeneity being



**Figure 10.6:** Segmentation of a toy case into 4 clusters with three criteria: density, topology and anisotropy. The likeness function is calculated on the full graph of 80 Voronoi zones. On the left the toy geometrical graph and its 4 found clusters into 4 different colours. On the right the similarity matrix reordered according to the clusters.

not really accessible to our senses. But the reordered similarity matrix exhibit diagonal block which shows the algorithm has succeeded.

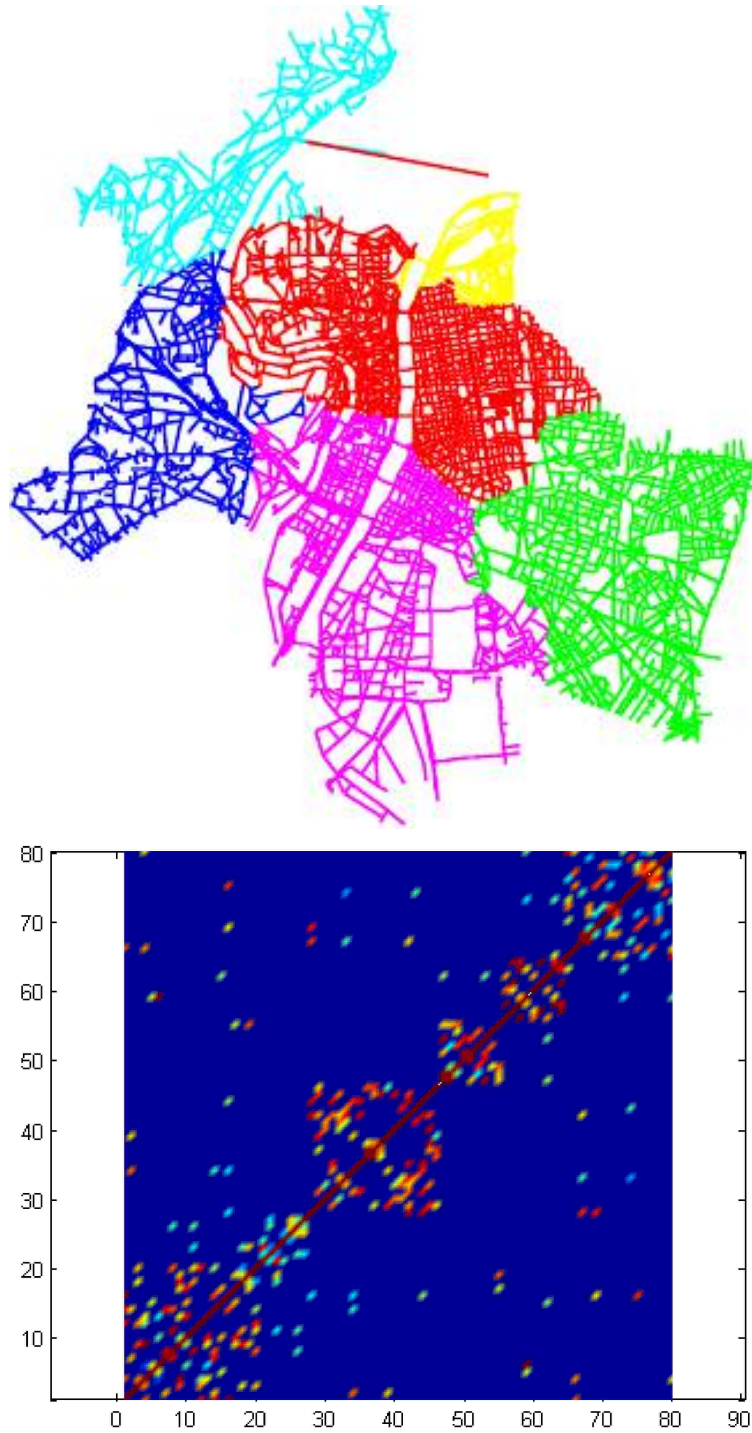
The result segmented map is robust with respect of the random Voronoï diagram that is used if it has been chosen with an appropriated number of centres. One of the main difficulty is to chose a relevant number of clusters. For the moment we do it "by hand", running the algorithm with a variable number of clusters and assessing their results visually via the segmented map and its reordered similarity matrix.

Once the segmentation is obtained, we want to identify each homogeneous zone to a stationary random tessellation (Ch.9).

At first we compute the hypergraph (Ch.4) and then the simplest centrality (Ch.6) of the map. We consider the subgraph constituted of the 5 percent more central streets. They form the upper scale of the city which is identified to a Poisson Gabriel Graph whose intensity is calculated with the ratio between the total upper scale length and the city's convex hull area.

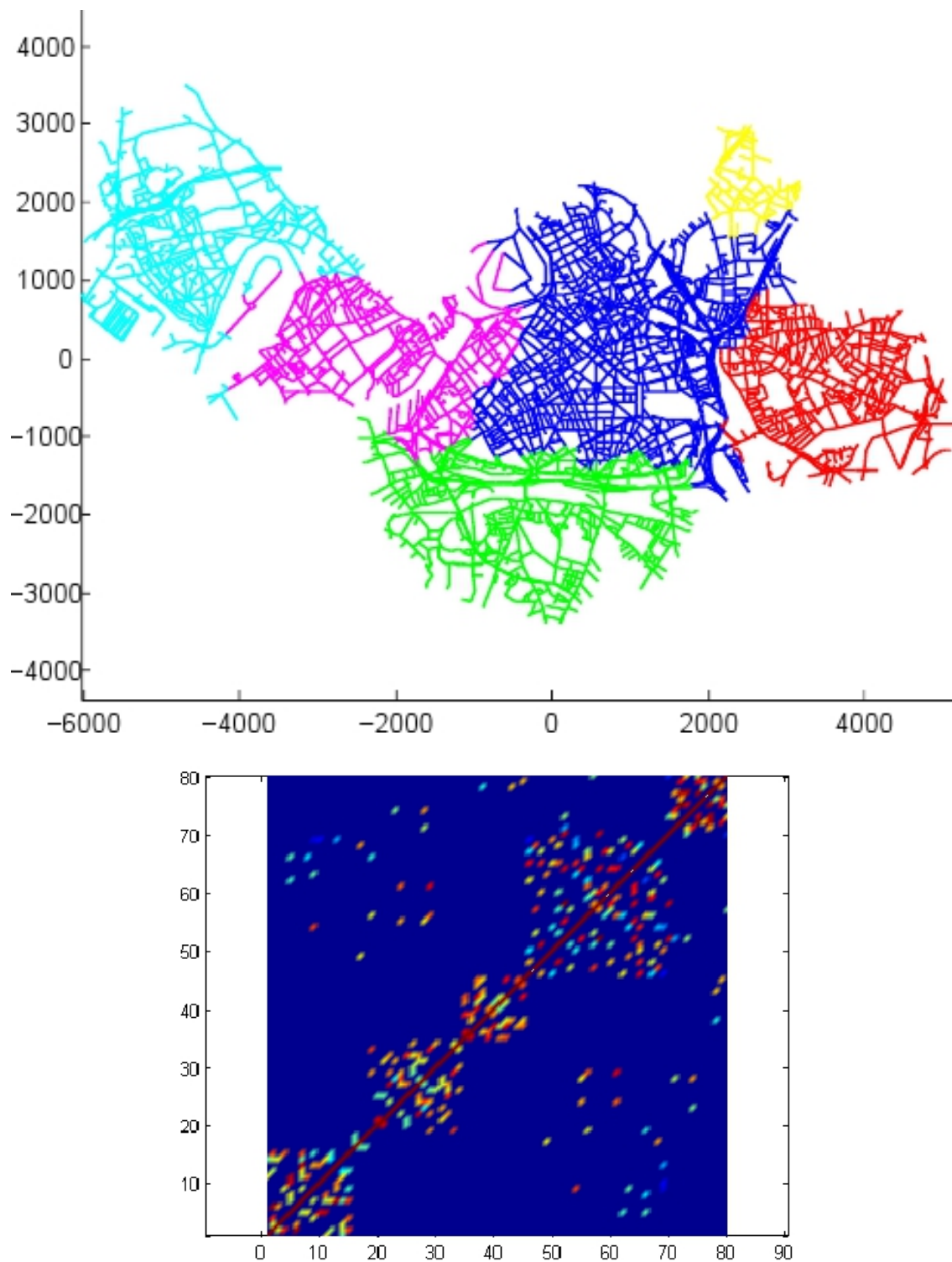
Then each zone is identified to a Poisson Line Tessellation or a Crack Tessellation (isotropic or anisotropic) model with an additional dead ends process. Once the model is chosen on each zone, its intensity is determined with Sec.9.4.2 method, the mean parameters being the ration between the features of subgraph on each zone and the area of its convex hull.

Fig.10.9 shows another representation of the segmentation of Lille. In bold blue we have emphasized the upper scale graph, each zone is delimited by its convex hull polygon drawn in bold red.

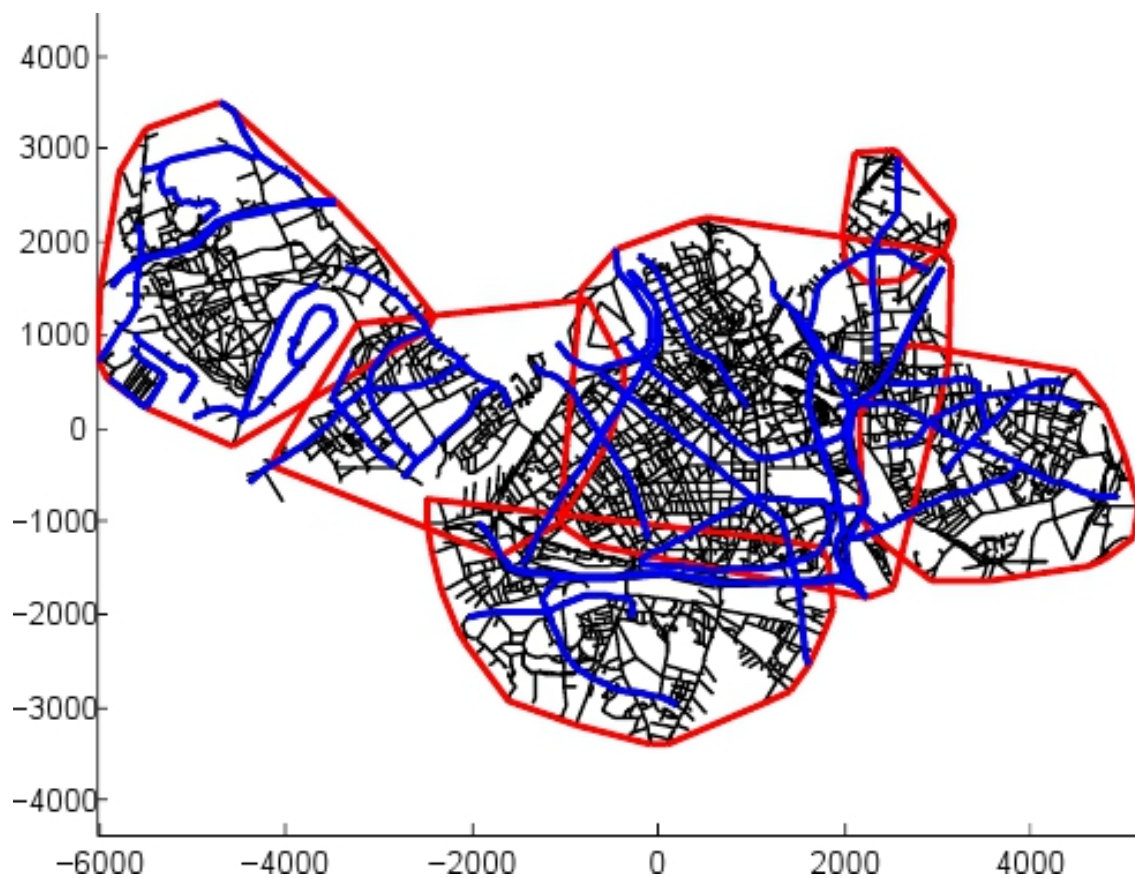


**Figure 10.7:** Segmentation into 6 zones of the city of Lyon (Left) and the reordered similarity matrix (Right).





**Figure 10.8:** Segmentation into 6 zones of the city of Lille (Left) and the reordered similarity matrix (Right).



**Figure 10.9:** Post processing of Fig.10.8. For each found cluster we compute the convex hull (in red). We compute for the wall city the simplest centrality and filter the 5 percent more central streets (in blue). The blue network is identified to a Poisson Gabriel Graph and each zone is identified to a Poisson Line Tessellation or a Crack Tessellation (with or without anisotropy) completed by a dead-end process.



## Part IV

# Conclusion



## Conclusion

In this thesis we have adopted a "hard science" approach to study a central object from social sciences: the City.

We have defined the city as a system of people and infrastructures spatially distributed to produce wealth that at least allows the continued existence of the community. Cities organize next to structuring elements or resources such as rivers or former cities to have a semi-optimal use of these local resources. The resulting structure is a more or less intricate network of streets that allows communication, transport, exchange of goods and people. We wanted to understand how this network structure organizes and how it jointly constraints the further development of a city and the use of space by people.

Under complex network study, the **map of a city** is represented as a topological graph. Not to loose the geometrical information on the shape of streets, we propose to see a map not as a plain graph but as a **geometric graph**, a continuum we provide with a Borelian measure and its associated integral operator. It is possible to sample a geometric graph into a **straight graph** arbitrary close to the original. The topological structure of the straight graph is sufficient to recover its geometry.

The graph of the city seems to display only trivial features. The topology indeed is strongly constrained by the planarity of the map. The distribution of block or **cell area**  $A$  is proportional to  $1/A^2$ .

The particularity of city graphs is that their edges are coherently arranged into separated lines called **streets**. We represent these streets by an additional **hypergraph structure**. And all happens in this hypergraph or street graph.

**The length of streets** is distributed as a mixture of log-normal laws, which implies a wide diversity of street lengths but also that the distribution admits an average and a variance. The street hypergraph exhibits a **small-world behaviour** that is to say the number of time one has to change of street to go from a random point in the city to another is small compared to the size of the city.

The dual of the city graph is the set of cells induced by the street system. When the city evolves, cells are divided to produce streets or axes that are to be coherent long alignments. The relation between division and alignment is not trivial and thus constraints the devel-

opment of cities.

Contrary to classical crack patterns, block divisions in cities are correlated; highly in the planned case, loosely in the organic case. There is certainly a continuum between organic and planned regimes.

**In the planned case**, the city is built of one piece, optimizing an overall structure. There exist globally optimal networks depending on the purpose of the city: the square grid if the resources are distributed and the radial scheme if the city is organized around a center of dominating importance. These patterns can adapt to the local constraints. We picture a simple homotopy adapted to the geography transforms a planned city next to an intricate relief to a planned city in a flat ground.

**In the organic case** the city is constituted by the progressive addition of local patterns that are optimal for the settlement of new infrastructures. When the place available inside the city is not sufficient (saturation or need of large infrastructure), the city extends making long streets and large cells at the outskirts that will be re-divided latter in smaller pieces. We have tried to understand organic cities from simple generic principles. **City is above all a running process of division and extension of space.** This process is the macro-description of the sum of local micro-interactions between independent agents, infrastructures and authority's regulation.

We have presented a coherent set of **continuous and network growth models**. In the **continuous** model, each place on a city produces a quantity of wealth proportional to its current wealth. This wealth is dispatched on all the space. It is coherent with **Zipf's law**: we consider a city as a distribution of infinitesimal centres in interaction. Some results of this model are independent of technical hypothesis we have to make: the local growth of wealth is exponential, the border of the city drifts exponentially and the wealth gradient becomes smoother with time. A **morphogenesis model** complement these results by considering the street network as an essential element of the urban growth. The current layout of the city induces a potential field on the whole space, an arriving agent chooses an available place to settle with respect of a politic toward the potential field. When he has chosen the place, he connects to the existing network. This endogenous model via a few parameters recover a large number of morphologies (as much in the network structure as in the overall shape) and reproduces hallmarks on the street network.

This understanding of cities' phenomenology has allowed us to propose a **corpus of efficient algorithms** dedicated to map related problems.

**The streets in the city can be recovered** fast without other information than the topology of the map and based on this recovering we propose an algorithm providing a good approximation of the **point to point shortest path problem** in a logarithmic time.

**The simplest centrality** and its display with an adaptive color scale presents a robust **automatic method to analyse a map**, emphasizing main axes and ill-deserved zones. We have shown its interest in town planning and analysis with two concrete cases.

Based on previous works on **stochastic geometry**, we have proposed relevant models and fast methods to identify a map to a "statistically equivalent" random geometric graph. Our contribution is the selection of low scale models, the addition of a dead-end process, the proposal of a **Poisson Gabriel Graph** as a high scale model, the simulation of its typical vertex and a simple fast and robust procedure to make the identification. Since these models are stationary, a **morphological segmentation** step is performed by a Spectral Clustering

---

based algorithm.

One of the first social scientist we presented our morphogenesis model to confronted us with the intrinsic freedom of humans. How do we consider in our model the choice for a family to settle in a place because of the beauty of the landscape, the color of the sun at dusk? I would like to provide three arguments to answer this remark. **(1):** This level of detail does not intervene in our model. We aim at sketching generically the growth of a city. Our model does not include crucial information such as elevation, sun, climate. But this model has been thought to be flexible and for instance one can easily add relief to the framework by a plain curvature of the plane. In addition models do not pretend to reproduce the totality of situations, they are synthetic tools to exhibit minimal set of hypothesis that explain a trend. **(2):** In fact we can consider the human freedom to choose a place for non apparent reasons is included tacitly in our model. Indeed in the potential generated by a city there are symmetries or points with the same attractiveness. A strong hypothesis we have made is that settlements are made sequentially, at discrete times, one after another. The new settler has then to choose between locations of similar potential. Once the choice is made, at random, a symmetry break appears on the map of the potential. And the follow of the city history reinforce this arbitrary choice: potential is higher next to places where pioneers have settled. **(3):** To come to the human freedom, freedom would be the ability to make decisions that outwardly seem to be random choices. Human enjoy thinking they escape from the determinism of physical objects and animals. But are not town-planners, engineers, independent pioneer agents of Nature as elements of an overall project? Is not the sum of our acts a path to the trap that Nature has set to us? Simulations from the morphogenesis model have been shown to town-planners and social scientists. Some of them have mixed up these artificial maps to real maps which we consider as a success. All the same they appreciated the map view generated by simplest centrality. In the same time this centrality put in light some obviousness of the city network and presents element that were not obvious but that make sense.

Originally, the subject of this PhD was "Darwinian network synthesis". The stepping-stone was the previous works performed on stochastic geometry. The project was limited to map segmentation with evolutionary algorithms. A set of baroque influences and the difficulty to start the work made us think there was a prior need to generate a synthetic knowledge on cities. That is what we have done, specifying tools to handle and to understand cities. In the end we have presented a coherent mathematical framework dedicated to city map and a few observations on the city's phenomenology. These naturally conduct to the definition of useful adapted algorithms. For instance we have presented a very simple procedure to solve the point to point problem on a city graph logarithmically. When this work had been



done, stochastic geometry modelling and maps segmentation became easier to comprehend. The literature on cities was in the same time very large and let me dare to say a little poor. Added to that it is to point out the cultural gap between different disciplines that make social science articles hard to access for physicists and conversely. Sometimes I was close to reinvent the wheel. For instance for a long time we did not know our simplest centrality was very close to Space Syntax or the "Dual Graph" theory. We were very disappointed when updating the state of the art to discover that what we thought was our main contribution was in fact already known and our work was to be discarded. But in fact working on the subject as if it was brand new bared fruits. The success is not only in results but in approach that creates benefits. The state of mind our approach put us into permitted to read critically the literature, to understand obscurity, errors and approximations in it and to add our personal contribution in a coherent fashion. For instance contrary to Space Syntax, we have insisted on the computational optimality of simplest centrality, its robustness and we do not try to justify it by "human perception" arguments: it naturally and simply comes mathematically as a boundary effect free version of the shortest path centrality.

A lot of outcomes and questions rise now.

Among them are obviously how to make the different growth models more robust, and study them theoretically. Morphogenesis should be implemented in an efficient framework allowing interactions with an operator. The geography around the city could be placed or adapted to real situations. Morphogenesis could be run on actual cities or situations to assess if it could be a relevant tool for town-planning.

We have started to study the distribution of path lengths in a city for random points. It seems that beta distributions are natural and relevant to model distances. When the size of a city increases with some technical hypothesis, these distributions are similar to Weibull distributions. And the thing is that Weibull distributions are stable with respect of minimum. A direct consequence is that the local efficiency of a point and its global efficiency are quantified with a single parameter. Integrating the whole points in every situation provides a single parameter description of the pace of the city.

Another study axis would be the link between centrality, random walk and traffic forecasting. We have the intuition that for each normalized centrality there exists a Markovian process whose stationary distribution is the centrality. Our idea for simplest centrality is to say that looked from the sky a pedestrian seems to adopt a biased random walk: when at an intersection he chooses its new direction with a probability proportional to the angle it makes with its current direction. The stationary distribution of this walk integrated on each street could provide a result close to simplest centrality.

Part V

Appendices



# Appendix **A**

## Geometry

### A.1 Some basis on Euclidian planar geometry

Topology defines the notion of continuity of applications between sets and gets interested in the properties of a set that do not change with respect of a continuous application. It can be done more abstractly but a convenient way to work on topology is to deal with the notion of distance in metrical spaces or metrical vector spaces. Geometry seeks out to characterize subsets of a set locally (differential geometry) or globally (integral geometry) with the notion of measure and especially of Lesbesgues's measures or volumes. Since geometry is interested in intrinsic features of subsets, an interesting distance is Euclidian distance that derives from a scalar product and which is in a vector space context independent of the coordinate system. The typical frame of geometry is thus Euclidian geometry: the metrical study of subsets (curves, manifolds, volumes) on  $\mathbb{R}^d$  equipped with a scalar product with  $d = 2$  or  $d = 3$ .  $\mathbb{R}^d$  can be curved by tensors or replaced by manifolds (for instance hyperbolic geometry).

#### A.1.1 Balls and spheres

In dimension  $\mathbb{R}^d$ , the ball of radius  $r$  is  $B_d(r) = \{x, \|x\| \leq r\}$  and the sphere is  $S_{d-1} = \{x, \|x\| = r\}$ . They are respectively a  $d$  and  $d - 1$  dimensional object whose Lesbesgue's measures are:

$$\mu_d(B_d(r)) = \frac{\pi^{n/2} R^n}{\Gamma(n/2 + 1)} \quad \mu_{d-1}(S_{d-1}(r)) = \frac{2\pi^{n/2} R^{n-1}}{\Gamma(n/2)} \quad (\text{A.1})$$

#### A.1.2 Minkowsky's sum

If  $X$  is a set and  $Y$  another set, the Minkowski's sum of  $X$  and  $Y$

$$X \oplus Y = \bigcup_{x \in X} (Y + x) = \{x + y, x \in X, y \in Y\} \quad (\text{A.2})$$

### A.1.3 Connectivity and convexity

Connectivity is a topological property to describe a set that is of a single component. Formally a set is connected if it cannot be written as the union of two disjoint open sets. A set  $A$  is convex if for  $x, y \in A, 0 \leq \lambda \leq 1 \quad (\lambda x + (1 - \lambda)y) \in A$

### A.1.4 Convex hull

Let  $S$  be a set of points in  $\mathbb{R}^d$  then the convex hull of  $S$ :  $\text{ConvHull}(S)$  is the smallest convex set of  $\mathbb{R}^d$  that contains  $S$ .

$$\text{ConvHull}(S) = \left\{ \sum_{i=0}^k \alpha_i x_i, \alpha_i \geq 0, \sum \alpha_i = 1, x_i \in S, k \in \mathbb{N} \right\} \quad (\text{A.3})$$

If  $S$  is a finite collection of points then  $\text{ConvHull}(S)$  is a polygon and it can be computed for  $d = 2$  in  $O(\#S \log \#S)$  with a divide and conquer algorithm [43].

## A.2 Polygons

In geometry, a **simple polygon** is a closed polygonal chain of line segments in the plane which do not have points in common other than the common vertices of pairs of consecutive segments.

A convex polygon is a polygon delimiting a convex zone of  $\mathbb{R}^d$ .

Let  $\Pi$  be a simple polygon with ordered vertices  $P_0, P_1, \dots, P_{n-1}, P_n = P_0$  of coordinates  $((x_i, y_i))_i$ .

The area of  $\Pi$  is easily calculated from:

$$\mathcal{A} = \left| \frac{1}{2} \sum_{k=0}^{n-1} (x_k y_{k+1} - x_{k+1} y_k) \right| \quad (\text{A.4})$$

And if  $x$  is a point,  $d$  a half-line of random direction and passing through  $x$ . Then

$$x \in \text{Int}(P) \iff \#P \cap d \in (2\mathbb{Z} + 1) \quad (\text{A.5})$$

## A.3 Tessellations

### A.3.1 Basic definitions

A tessellation on a set (let's say locally compact)  $A \subset \mathbb{R}^d$  is a collection at least countable of sets (also locally compact)  $(C_i)$  such as:

- $\text{int}(C_i) \neq \emptyset \forall i$
- $\text{int}(C_i) \cap \text{int}(C_j) = \emptyset \quad \text{if } i \neq j$

- $\cup C_i = A$

A practical tessellation case is when  $A$  is a convex locally compact subset and every  $C_i$  a finite convex polygon. Then one can derive from the tessellation a set of edges ( $E_i$ ) and a set of vertices ( $V_i$ ). Then the tessellation is equivalent to a planar graph (possibly infinite). Thus we will mix the notion of tessellation and planar graph with no dead ends.

### Dual tessellation

Let  $(C_1, \dots)$  be a convex tessellation. To each  $C_i$  is associated a centroid  $c_i \in C_i$  (for instance the barycentre of the cell). We create a new graph by linking two centroids  $c_i$  and  $c_j$  if and only if  $C_i$  and  $C_j$  share an edge. The resulting graph is planar with no dead ends: it is the dual tessellation ("the" dual in a topological sense but "a" dual in the geometrical sense).

### A.3.2 Voronoï tessellations

We give the main definitions and results relatively to Voronoï graphs in  $\mathbb{R}^d$ , the basic idea can be generalized to subsets of  $\mathbb{R}^d$  but we have to beware of irregularities emerging from side effects. If  $x$  is a point of a discrete point set (germ set)  $X = (X_i)$ , we associate to  $x$  its Voronoï cell with respect to  $X$ :

$$\text{Vor}(x|X = (x_1, x_2, \dots)) = \{y \in \mathbb{R}^d, \|x - y\| \leq \|x_i - y\|, \forall x_i \neq x\} \quad (\text{A.6})$$

The set of closed cells  $\{\text{Vor}(x_i|X)\}$  is a tessellation called the Voronoï Tessellation of  $X$ . This tessellation has some remarkable properties: each cell is a polytope, each cell is convex; if  $X$  is not too regular (i.e the realization of Poisson Point Process for instance), each vertex is of degree 3. Some variations have been investigated [137] such as the power diagram. To each  $x_i$  is associated a positive real number  $p_i$  (its power) and the power function of  $(x_i, p_i)$  is defined for a point  $y$  in the space as  $\sigma_i(y) = \|x_i - y\|^2 - p_i^2$ . The cell associated to  $x_i$  is the set of points  $y$  such as  $\sigma_i(y) \leq \sigma_j(y) \forall j \neq i$ . We can also mention anisotropic Voronoï diagrams, Möbius or Apollonius diagrams. We can also consider a set of geometrical objects rather than a set of points as germs.

### A.3.3 Delaunay tessellations

When taking the germ set as centroid set, the dual of the Voronoï Tessellation is the **Delaunay tessellation**. Each cell of the Delaunay tessellation is a triangle. A direct construction of Delaunay is possible. If  $X$  is a discrete point set, and  $x_i, x_j, x_k$  three different points of  $X$ , the triangle  $x_i x_j x_k$  is drawn if and only if there is no other points of  $X$  inside the circumcircle of  $x_i x_j x_k$ . **Relative neighbourhood** and **Gabriel graph** are planar subgraph of the Delaunay triangulation that are not in general tessellations. They all belong to the class of neighbourhood graphs. An edge  $x_i x_j$  is drawn in the relative neighbourhood graph if no point of  $X$  is both closer to  $x_i$  or  $x_j$ : if the intersection of circles of centres  $x_i$  and  $x_j$  and of radius  $\|x_i - x_j\|$  is empty. An edge is drawn in the Gabriel graph if there is no other points of  $X$  in the circle of diameter  $x_i x_j$ .

## A.4 Fractal

Fractals have been a trendy object to model cities. They are very convoluted objects, they have been introduced in the 70's by Benoît Mandelbrot as a class of non trivial shapes that contravene human intuition of shape. Indeed, in dimension  $d$ , a simple object  $O$  such as a ball, a polytope etc... are such that

$$\mu_{d-1}(aO) = a^{d-1}\mu_{d-1}(O) \quad (\text{A.7})$$

fractal objects are objects that do not respect this intuition:

$$\mu_{d-1}(aF) = a^\alpha \mu_{d-1}(F) \quad (\text{A.8})$$

with  $d - 1 < \alpha \leq 2$  is the "fractal dimension" of  $F$ . A consequence is that  $F$  has no characteristic scale: it is the same whatever the observation scale is. In fact the above equation makes no sense: the measure of  $F$  is infinite, one cannot rectify it and define properly a measure. But the idea is that given a reference length, a measure of the fractal dilated by  $a$  diverges as  $a^{\alpha-d-1}$ . To be more formal, we have to introduce the Hausdorff's dimension. If  $F \subset \mathbb{R}^d$ ,  $r, s \geq 0$ :

$$H_r^s(F) = \inf_{\text{diam}(A_i) < r} \left\{ \sum \text{diam}(A_i)^s \mid \bigcup A_i \supseteq F \right\} \quad (\text{A.9})$$

As a function of  $r$ ,  $H_r^s(F)$  is decreasing. We can thus consider

$$H^s(F) = \lim_{r \rightarrow 0} H_r^s(F) \quad (\text{A.10})$$

If there exist  $s_0$  such as  $H^{s_0}$  is finite then  $H^{s < s_0} = \infty$  and  $H^{s > s_0} = 0$ . Thus there exist a number  $\dim_H$  intrinsic to  $F$  called the fractal dimension of  $F$  such as:

$$\dim_H = \inf\{s, H^s(F) = 0\} = \sup\{s, H^s(F) = \infty\} \quad (\text{A.11})$$

This definition is mathematically rigorous and allows theoretical calculus. Nonetheless it is not utilisable to measure the dimension of real curves. Thus in practice we use the Minkowski's dimension also known as box counting method.

$$\dim_M^+ = \limsup_{r \rightarrow 0} \frac{\log \mu_d(F \oplus B_r)}{\log r} \quad \dim_M^- = \liminf_{r \rightarrow 0} \frac{\log \mu_d(F \oplus B_r)}{\log r} \quad (\text{A.12})$$

A typical and well known case of fractal is auto affine shapes such as Sierpinski carpet, Von Kock's curve, Peano's curve. These shapes are defined as the limit of a recursive construction process. A generator shape is given, and on that shape are chosen  $m$  similar seeds, the shape is initialized to that generator. Then each seed is replaced by a rotation translation of the generator shaped reduced by a factor  $h$ . This process is iterated, the shape of generation  $n + 1$  is obtained by replacing each seed in the shape of generation  $n$  by a reduction of a factor  $nh$  of the generator shape. The fractal dimension  $\alpha$  of a shape resulting of such a constriction is simply:

$$\alpha = \frac{\log m}{\log h} \quad (\text{A.13})$$

# Complex Networks' theory

This additional chapter synthetizes main notions and results from Complex Network's theory that has been an emerging field from the 60's. It is based on the reports [1, 23, 98] for general complex networks and [9] for the particular question of planarity.

## B.1 What is a complex network ?

In physical, biological or social science network is a conceptual representation of homogeneous entities (vertices) and their interactions (edges). Vertices can be for instance people in the scientific community, Facebook profiles, web sites, softwares, neurons... Edges do not have to exist physically. They only express a link or relation between entities (reflexive most of the time). To follow up the vertex example above, edges can be co-authorship, "be friend with", "point out to", "have a package dependency", "share an axon". The vertices linked to a given vertex constitute its neighbourhood: they directly communicate with it. Network is a purely topological representation. A generic problem on a network is to assess how information spread from neighbourhood to neighbourhood, how a signal is regularized from an input vertex to an output vertex with possibly feedbacks. Network can be natural and science is to describe and understand them. They can also be engineered.

Typically in physical science, a complex network is a **large** graph resulting from a **dynamical** process from which emerges some non trivial **global features**.

If  $G = (V, E)$  is a network then  $n = \#V$  is the number of vertices and  $m = \#E$  is the number of edges. If two vertices  $i$  and  $j$  are adjacent we write  $i \sim j$ .

## B.2 Analysis of complex networks

### B.2.1 Size

The size of the network is generally defined as  $n$  the number of its vertices.



### B.2.2 Degree distribution

To each vertex  $i$  of the network, we associate its degree  $k_i$  that is to say the number of edges that are incident to it. Then we consider the degree distribution among all vertices  $P(k)$ .  $P(k)$  may be peaked (spatial networks, random graphs) or display a long tail (scale-free networks). The features of  $P(k)$  are summarized by: the average degree,  $\langle k \rangle = \sum k.P(k) = \frac{m}{n}$ , the fluctuation  $\langle k^2 \rangle = \sum k^2.P(k)$  and their ratio  $\kappa = \langle k^2 \rangle / \langle k \rangle$  plays a role in dynamical processes on a complex network.

### B.2.3 Clustering

The clustering coefficient is a measure that tends to quantify if the network is transitive. For instance in social networks it is a quantitative answer to the question "are the friends of someone also friends".

$$C = 3 \cdot \frac{\#\{(x, y, z) \in V^3, (xy) \in E, (xz) \in E, (yz) \in E\}}{\#\{(x, y, z) \in V^3, (xy) \in E, (yz) \in E\}} \quad (\text{B.1})$$

### B.2.4 Degree correlation

The question is to know if high degree vertices connect to each other (**assortative**) or rather connect to low degree vertices (**disassortative**) and conversely. To this is introduced the average degree in the neighbourhood of vertex  $i$ :  $K_i$  and then the average degree of neighbours of nodes of degree  $k$ :  $K(k)$ . This function is globally increasing if the network is assortative, decreasing if its disassortative and constant if not correlated. A numerical measure is presented in [33].

### B.2.5 Robustness

We want to measure to what extent a network is robust to failures or attacks, that is to say if information transport in the network can effectively adapt if some vertices of the network are removed. To this we consider we remove at random a proportion  $f$  of vertices and after that operation we compute the relative size of the largest connected component of the network:  $S(f)$ . To be smooth  $S$  has to be calculated in average for several random removal. The curve  $S(f)$  is decreasing. The robustness of the network is defined as  $f_c = S^{-1}(S_c)$  where  $S_c$  is arbitrary chosen.

Two kinds of robustnesses are can be distinguished: the **resilience** when vertices are chosen uniformly and **robustness** properly speaking when that choice is biased with the degree of vertices or a measure of their centrality (B.2.7).

### B.2.6 Distances

The shortest path distance or geodesic distance from a vertex  $i$  to a vertex  $j$  is written  $d_{ij}$ . It can be computed in polynomial time. The diameter of the network is  $\max_{i,j} d_{ij}$ . But this

measure is very sensitive to noise. We seek out to define an average measure of the total amount of distances in a network. For instance  $\langle d_{ij} \rangle_{ij}$  but if the network is split into several connected component this mean is  $\infty$ . A solution is to consider the harmonic mean of geodesic distance,  $l$  defined by:

$$l^{-1} = \frac{2}{n(n-1)} \sum_{i>j} d_{ij}^{-1} \quad (\text{B.2})$$

### B.2.7 Centrality

A centrality is a measure calculated on each vertex of a network. It is to quantify to what extent a vertex is important: a hub necessary to cross, a place easy to reach. In the literature one can find very classical centralities such as: The closeness centrality of a node  $i$ :

$$C_i^C = \frac{N-1}{\sum_{\substack{j \in V \\ j \neq i}} d_{ij}^{short}} \quad (\text{B.3})$$

The betweenness centrality of a node  $i$ :

$$C_i^B = \frac{1}{(N-1)(N-2)} \sum_{\substack{j,k \in V \\ j \neq k \neq i}} \frac{n_{jk}(i)}{n_{jk}} \quad (\text{B.4})$$

where  $n_{jk}$  is the number of geodesic from  $j$  to  $k$  and  $n_{jk}(i)$  the number of geodesics from  $j$  to  $k$  that pass through  $i$ . When the network is embedded in an Euclidian space, one can define the Euclidian distance between to vertices and the straightness centrality of a node  $i$ :

$$C_i^S = \frac{1}{N-1} \sum_{\substack{j \in V \\ j \neq i}} \frac{d_{ij}^{eucl}}{d_{ij}} \quad (\text{B.5})$$

We can also consider that the stationary distribution of a Markov chain on the network is a centrality.

## B.3 Canonical complex networks

### B.3.1 Random graphs

Random graphs are used as a 0 model to test relevance of observations in real networks. Consider  $n$  points and a number  $0 \leq p \leq 1$ . To form the random graph  $G_{n,p}$ , we consider for instance that a possible edge in the graph is indeed drawn with probability  $p$ . The degree distribution of a resulting graph is binomial, it can be approximated by a Poisson Law when  $n$  is large, with mean  $pn$ . What is striking is that for a given property exists most of the time a critical probability  $p_c(n)$  such as the property is almost surely true for higher probabilities and almost surely false for lower.

For instance there is a percolation probability  $p_c = 1/n$  such as

- If  $p < p_c$ , the graph has no component of size greater than  $O(\log n)$ .
- If  $p = p_c$ , the graph's largest component size is  $O(n^{2/3})$ .
- If  $p > p_c$ , the graph has a single giant component of size  $O(n)$  and no other component has a size greater than  $O(\log n)$ .

If the graph is efficiently connected ( $p > p_c$ ) the diameter of the graph (after exclusion of small connected component) is  $O(\log n)$ . The clustering coefficient is simply  $p$ , that is to say that for a fixed mean degree  $\langle k \rangle$  the clustering coefficient tends to 0 when  $n \rightarrow \infty$ . There exist several modifications of the random graph definition that are more or less convenient to work theoretically or computationally with.

### B.3.2 Scale-free networks

A scale-free network is a network whose degree distribution asymptotically exhibits a power law:  $P(k) \sim k^{-\gamma}$ .

It means that the network has a wide variety of degrees (the distribution is long-tailed) and that one can observe this diversity at any scale (power laws are the only functions such that  $f(ax) = g(a)f(x)$ ,  $\forall a, x$ ).

Some authors even impose the condition  $1 < \gamma \leq 3$  so that variance does not exist and the diversity is wide.

### B.3.3 Small-world networks

A small-world network is a network whose diameter or average geodesic length is "very small" compared to the size of the network. The notion of very small can only be defined if there exist several realizations of the network with different sizes or in an evolutionary context: if the network  $G = G(n)$  grows then it is a small world if

$$l(G(n)) = O(\log n) \tag{B.6}$$

Some authors also impose to small world networks to have an high clustering coefficient.

### B.3.4 Simulation

**Barabási and Albert** In the continuity of Price's model and the preferential attachment principle: "a new vertex will have tendency to connect to vertices with high degree" ("the rich gets richer"), Barabási and Albert propose a growth model for network that exhibit a scale free behaviour. They consider that a network evolves: at each time a new vertex  $j$  is added and connects to  $m$  existing vertices. But the choice of these vertices is biased by their degree:  $\mathbb{P}(j \rightarrow i) \propto k_i$ . Then they show that after a large number of iterations,

$$\mathbb{P}(k_i = k) = \frac{2m(m+1)}{(k+2)(k+1)k} \sim 2m(m+1)k^{-3} \tag{B.7}$$

Some modifications of the model allow to get exponent different from 3. Preferential attachment is thus an appealing model to explain the emergence of scale free networks in nature.

**Small World network** Simulating a small world is much more easy and thus it is harder to find a quasi universal model to explain them.

For instance a random graph can present small-world behaviour, a fully connected graph or a scale free network. It is even possible to design sequentially a small world network with a predetermined objective mean path function. The Watts and Strogatz model permits to construct a small world network with an high clustering coefficient.

An interesting problem is to design a graph with a particular diameter and a maximal degree. In fact it has been shown that the number of nodes in such a graph is bounded by the "Moore Bound".

## B.4 Spatial embedding

A case of practical importance is the case of spatial networks. Loosely speaking a spatial network is a network whose vertices are embedded in an Euclidian space. The idea is that a spatial connection between two points has a cost that is an increasing function of their distance. In a general way, a spatial network is equivalent to a weighted network, with a weight 0 if two vertices are not connected and a weight increasing with their distance otherwise.

Planar networks are a particular class of spatial networks: their vertices are embedded in an Euclidian space and their edges are geometrical paths across the space that bind two vertices and that do not intersect outside of vertices.

Planar networks will display particular features due to the important physical constraints that are applied on them:

- Vertices are distributed on space and an edge creation has a real cost. Connections are then mainly local.
- Degree of vertices are bounded by the Euler's equality: there are at most  $3n - 6$  edges and thus the mean degree cannot exceed  $2(3n - 6)/n \sim 6$ .



# Measure and probability theory for stochastic geometry

The aim of this document is to present the basis in measure and probability theory to tackle stochastic geometry. If there is a lot of literature on the two first, stochastic geometry is more discrete and there is no global textbook available with simple explanations. We aim at providing the reader with the major structure of the theory, the main proofs the whole being illustrated with simple example.

This text is based on Mazet's *Intégration* [89] and *Measure theory* by V. Liskevich [78] both available on the Internet, with also the reference books *Stochastic geometry and its application* by Stoyan [122], *Markov point processes and their applications* by M. N. M. van Lieshout [129], *New Perspectives in Stochastic Geometry* by Wilfrid S. Kendall [76] the textbooks *Stochastic Geometry and Wireless Networks* by F. Baccelli [5, 6] and the lecture notes *A short course on stochastic geometry* by G. Last available on the Internet. Particular tanks to Pr. Pierre Calka for his time and wise advises.

## C.1 Measure and integration theory

### C.1.1 Measure space

**$\sigma$ -algebra and measurable spaces** Let  $\Omega$  a set and  $\mathcal{A} \subset \mathcal{P}(\Omega)$  if added to that

1.  $\Omega \in \mathcal{A}$
2. If  $A \in \mathcal{A}$  then  $A^c \in \mathcal{A}$
3. If  $A_1, A_2, \dots \in \mathcal{A}$  then  $\bigcup A_i \in \mathcal{A}$

then  $\mathcal{A}$  is a  $\sigma$ -algebra on  $\Omega$  and  $[\Omega, \mathcal{A}]$  is a measure space.

**Measures** On a measurable space, one can define a measure  $\mu$  that is a function  $\mathcal{A} \rightarrow \bar{\mathbb{R}}^+$  that holds:

1.  $\mu(\emptyset) = 0$
2.  $\mu(\bigcup A_i) = \sum \mu(A_i)$  if  $A_1, A_2, \dots \in \mathcal{A}$  and  $i \neq j \Rightarrow A_i \cap A_j = \emptyset$

With such a measure,  $[\Omega, \mathcal{A}, \mu]$  is called a measured space.

A measure whose image is reduced to  $[0, 1]$  with  $\mu(\Omega) = 1$  is called a **probability measure**.

**Complete measure space** A measure space  $[\Omega, \mathcal{A}, \mu]$  is complete if for all  $N \in \mathcal{A}$ ,  $\mu(N) = 0$ , every subset of  $N$  is measurable (then its measure is 0).

It is always possible to extend a measure space  $[\Omega, \mathcal{A}, \mu]$  to a complete measure space  $[\Omega, \mathcal{A}_0, \mu_0]$ . To this:

1. Define the upper measure of  $\mu$  on  $\mathcal{P}(\Omega)$ :

$$\mu^*(A) = \inf_{A \subseteq \cup E_i, E_i \in \mathcal{A}} \sum \mu(E_i)$$

2. Remark that if  $A \in \mathcal{A}$ ,  $\mu^*(A) = \mu(A)$ .

3. Define the Carathèdory measurable sets: a set  $A$  is called measurable if for any  $E \in \mathcal{P}(\Omega)$ ,

$$\mu^*(E) = \mu^*(E \cap A) + \mu^*(E \cap A^c)$$

4. Invoke the Carathèdory theorem: the set of measurable sets is a  $\sigma$ -algebra and the restriction of  $\mu^*$  to this  $\sigma$ -algebra is a complete measure.

**Measurable functions** If  $[\Omega_1, \mathcal{A}_1]$  and  $[\Omega_2, \mathcal{A}_2]$  are two measurable spaces, a function  $f : \Omega_1 \rightarrow \Omega_2$  is measurable iff  $\forall A_2 \in \mathcal{A}_2, f^{-1}(A_2) \in \mathcal{A}_1$ .

A measurable function allows to transform measures: if  $\mu_1$  is a measure on  $[\Omega_1, \mathcal{A}_1]$  then  $\mu_2(\cdot) = \mu_1(f^{-1}(\cdot))$  is a measure on  $[\Omega_2, \mathcal{A}_2]$

### C.1.2 Classical $\sigma$ -algebra on topological spaces

**$\sigma$ -algebra generated by a part** If  $S$  is a subset of  $\mathcal{P}(\Omega)$ , the  $\sigma$ -algebra generated by  $S$  is the smallest  $\sigma$ -algebra that contains  $S$ . This  $\sigma$ -algebra always exists and is written  $\sigma(S)$ .

**$\sigma$ -algebra generated by a function** If  $f$  is a function from  $[\Omega_1, \mathcal{A}_1]$  to  $\Omega_2$  then  $(f(A), A \in \mathcal{A})$  is a  $\sigma$ -algebra on  $\Omega_2$ .  $(f(A), A \in \mathcal{A})$  is the smallest  $\sigma$ -algebra that makes the mapping  $f$  measurable.

Similarly, If  $f$  is a function from  $\Omega_1$  to  $[\Omega_2, \mathcal{A}_2]$  then  $(f^{-1}(A), A \in \mathcal{A})$  is a  $\sigma$ -algebra on  $\Omega_1$ .

**Product  $\sigma$ -algebra** If  $[\Omega_1, \mathcal{A}_1]$  and  $[\Omega_2, \mathcal{A}_2]$  are two measurable spaces,  $\mathcal{A}_1 \times \mathcal{A}_2 = \mathcal{A}_1 \times \mathcal{A}_2$ ,  $A_1 \in \mathcal{A}_1, A_2 \in \mathcal{A}_2$  is a  $\sigma$ -algebra for  $\Omega_1 \times \Omega_2$ .

**Borelian** If  $X$  is a topological space, the Borelian  $\sigma$ -algebra of  $X$  is the  $\sigma$ -algebra generated by the set of the opens of  $X$ . The elements of that  $\sigma$ -algebra are called Borelians of  $X$ .

For instance think of  $\mathbb{R}$  and the  $\sigma$ -algebra generated by  $\{]a, b[, a, b \in \mathbb{R}\}$  written  $\mathcal{B}_1$ .

We define recursively the Borelian  $\sigma$ -algebra of  $\mathbb{R}^d$ :  $\mathcal{B}_d$

### C.1.3 Classical measures

**Mass measures** If  $x$  is a particular point in  $\mathbb{R}^d$  then:  $\delta_x : A \mapsto 1$  if  $x \in A$ , 0 otherwise is a measure on  $[\mathbb{R}^d, \mathcal{B}]$  called the Dirac measure.

All the same, if  $X = (x_1, x_2, \dots) \in \mathbb{R}^d$ ,  $\delta_X : A \mapsto \#(A \cap X)$  is a measure on  $[\mathbb{R}^d, \mathcal{B}]$  called the counting measure.

**Borel measure** The Borelian set of  $\mathbb{R}^d$  is generated by open PAVE. So  $]a_1, b_1[ \times \dots \times ]a_d, b_d[ \mapsto (b_1 - a_1) \dots (b_d - a_d)$  can be extended into a measure on  $[\mathbb{R}^d, \mathcal{B}_d]$ . It is written  $\mu_d$  and called the Borelian measure on  $\mathbb{R}^d$ .

**Lebesgue measure** It is the completion of the Borel measure to the  $\sigma$ -algebra generated by  $\{A \cup N\}$  where  $A \in \mathbb{B}$  and  $N$  is a negligible part  $\mu_d$  (Lebesgue  $\sigma$ -algebra).

It is the only invariant by translation measure.

**Product of measures** If  $[\Omega_1, \mathcal{A}_1, \mu_1]$   $[\Omega_2, \mathcal{A}_2, \mu_2]$  are two measure space then  $(\mu_1 \otimes \mu_2)(A_1 \times A_2) = \mu_1(A_1) \times \mu_2(A_2)$  is a measure on  $[\Omega_1 \times \Omega_2, \mathcal{A}_1 \times \mathcal{A}_2]$ .

### C.1.4 Integral

In the following sections, we work on a measure space  $[\Omega, \mathcal{A}, m]$ .

Every positive measurable function writes  $f = \sum_{i \in \mathbb{N}} a_i \cdot \mathbf{1}_{A_i}$  it is then natural to define  $\int f d\mu = \sum_{i \in \mathbb{N}} a_i \cdot \mu(A_i)$ .

In the case of  $\mathbb{R}^d$ , this integral called the Lebesgue's integral extend the notion of Riemann's integral. Thus classical calculus methods can be applied.

#### Convergence theorems

**Beppo-Lévi** If  $(f_n)$  is an increasing sequence of strictly positive measurable functions and if  $f_n \rightarrow f$  simply, then

$$\int f_n dm \rightarrow \int f dm$$

**Monotonous convergence** If  $(f_n)$  is an increasing sequence of measurable functions and if  $f_n \rightarrow f$  simply, then if  $\int f_n dm$  is upper bounded then  $f$  is measurable and  $\int f_n dm \rightarrow \int f dm$ .

**Dominated convergence** If  $(f_n)$  is an increasing sequence measurable functions that tends simply to  $f$  and if there exist an integrable function  $g$  such that  $|f_n| < g$  then  $\int f dm$  exists and  $\int f_n dm \rightarrow \int f dm$



### C.1.5 Radon-Nikodym theorem

If  $\mu_1$  and  $\mu_2$  are two measures on  $[\Omega, \mathcal{A}]$  and if there exist a function  $f$  such that  $\forall A \in \mathcal{A}$ ,  $\mu_2(A) = \int f(x)d\mu_1(x)$  then  $\mu_2$  is said to be absolutely continuous with respect to  $\mu_1$  ( $\mu_2 \ll \mu_1$ ) and  $f$  is a density function written  $f = \frac{d\mu_2}{d\mu_1}$ . The Radon-Nikodym theorem states that  $\mu_2 \ll \mu_1$  if and only if  $\mu_1(A) = 0 \Rightarrow \mu_2(A) = 0$ .

## C.2 Real random variables

### C.2.1 Definition

**Random variable** If  $[\Omega, \mathcal{A}, \mathbb{P}]$  is a measure space with  $\mathbb{P}$  a probability measure (called respectively sample space, event space and of course probability), then a measurable function  $X : [\Omega, \mathcal{A}] \rightarrow [\mathbb{R}^d, \mathcal{B}]$  is called a random variable.

We will write  $\mathbb{F}$  the image measure of  $\mathbb{P}$  by  $X$ .

$[\Omega, \mathcal{A}]$  is never explicitly defined, in fact one postulates their existence to model the observation we get. We know them only on their capacity to make some result of interesting random variables measurable.

**Probability distribution, density** The function  $F : x \mapsto \mathbb{P}(X(\cdot) \in ]-\infty, x])$  is the probability distribution of  $X$ . It contains the same information as  $\mathbb{F}$ .

If  $F$  is derivable almost everywhere,  $f = F'$  is called the density of  $X$ .  $f$  is then the Radon-derivative of  $\mathbb{F}$  by  $\mu_d$ .

### C.2.2 Expectancy

If  $X$  is a random variable its expectancy is defined by:

$$\mathbb{E}(X) = \int_{\Omega} X(\omega)d\mathbb{P}(\omega) = \int_{\mathbb{R}} x.d(\mathbb{F}(x))$$

If the random variable is real and admits a density,

$$\mathbb{E}(X) = \int_{\mathbb{R}} x.d(\mathbb{F}(x)) = \int_{\mathbb{R}} x.f(x).dx$$

**Markov's inequality:**

$$\mathbb{P}(|X| > a) \leq \frac{\mathbb{E}(|X|)}{a}$$

### C.2.3 Moments

The  $n$ -th order moment of a random variable  $X$  is (if it exists):

$$\mathbb{E}((X - \mathbb{E}(X))^n)$$

The second-order moment, called the variance is of particular interest since it describe the variability of a random variable and appears in the Central Limit theorem. It is written  $\mathbb{V}(X)$ .

Moreover, the set of random variables that admit a variance ( $\mathcal{L}^2$ ) is a vector space with an Euclidian structure induced by the scalar product

$$\langle X, Y \rangle = \mathbb{E}(XY)$$

which make  $\mathcal{L}^2$  a pleasant space to work in.

**Chebyshev's inequality:** The existence of an expectancy and a variance induces a constraint on the dispersion of the values taken by a random variable  $X$  expressed by the Chebyshev's inequality:

$$\mathbb{P}(|X - \mathbb{E}(X)| > a) \leq \frac{\mathbb{V}(|X|)}{a^2}$$

#### C.2.4 Conditional distribution

Let  $(X, Y)$  be a random variable on  $\mathbb{R}^d \times \mathbb{R}^c$  with probability  $\mathbb{P}_{X,Y}$ .

$\mathbb{P}_X(A) = \mathbb{P}_{X,Y}(A, \Omega)$  is a probability function called the marginal probability of  $X$ . One defines similarly  $\mathbb{P}_Y$ . For all  $A$ ,  $\mathbb{P}_{X,Y}(A, \cdot)$  is absolutely continuous with respect to  $\mathbb{P}_Y(\cdot)$  so there exist  $\mathbb{P}(X = \cdot | Y = y)$  a probability such that:

$$\mathbb{P}_{X,Y}(A, B) = \int_B \mathbb{P}(X \in A | Y = y) d\mathbb{P}_Y(y)$$

this probability is called the conditional probability.

If  $\mathbb{P}_{X,Y}$  is a discrete probability with distribution  $p(X = n_x, Y = n_y)$  then  $\mathbb{P}(\cdot | Y = n_y)$  is a discrete probability with distribution

$$p(X = n_x | Y = n_y) = \frac{p_{X,Y}(X = n_x, Y = n_y)}{p_Y(n_y)}$$

All the same if  $\mathbb{P}_{X,Y}$  admits a density  $f_{X,Y}$  then  $\mathbb{P}_X$  admits a density  $f_X$  and  $\mathbb{P}(\cdot | Y = y)$ ,  $f(\cdot | Y = y)$  with

$$f(X = x | Y = y) = \frac{f_{X,Y}(X = x, Y = y)}{f_Y(Y = y)}$$

**Conditional expectancy** There is an only random variable  $\sigma(Y)$ -measurable  $Z$  such that for every random variable  $U$   $\sigma(Y)$ -measurable,  $\mathbb{E}(XU) = \mathbb{E}(ZX)$ . It is called the conditional expectancy of  $X$  with respect of  $Y$  and written  $\mathbb{E}(X|Y)$ .

If  $\mathbb{E}(X|Y = y) = e(y)$  then  $\mathbb{E}(X|Y) = e(Y)$  and if  $X$  and  $Y$  are independent then  $\mathbb{E}(X|Y) = \mathbb{E}(X)$ .

The conditional expectancy can be seen as the best guess one can make about  $X$  when knowing the result  $Y$ .

In  $\mathcal{L}^2$ ,  $\mathbb{E}(X|Y)$  is the closest random variable to  $X$  that is  $\sigma(Y)$ -measurable.

### C.2.5 Independence

Two random variables  $X$  and  $Y$  are said to be independent if:

$$\mathbb{P}(X \in A_1 | Y \in A_2) = \mathbb{P}(X \in A_1)$$

it is the same than  $\mathbb{P}(Y \in A_2 | X \in A_1) = \mathbb{P}(Y \in A_2)$  or than

$$\mathbb{P}(X \in A_1, Y \in A_2) = \mathbb{P}(X \in A_1)\mathbb{P}(Y \in A_2)$$

and conditions of the same shape are true for distributions or density when possible to write them.

Intuitively, it means that the realizations of  $X$  and  $Y$  do not influence on each other.

### C.2.6 Classical laws

We describe here the most common discrete distributions and continuous densities used in modeling of "random" phenomena.

We say "random" in quotes since randomness is not an intrinsic attribute of phenomena, it is a convenient point of view a scientific takes to reduce complexity.

These law can be used as parametric families to fit data or their use can be justified by theoretical considerations.

#### Discrete laws

##### Uniform

$$\mathbb{P}(X = k) = \frac{1}{b-a+1} \mathbb{1}_{[a,b]} \quad \mathbb{E}(X) = \frac{a+b}{2}$$

$$\mathbb{V}(X) = \frac{(b-a+1)^2-1}{12}$$

##### Geometric

$$\mathbb{P}(X = k) = (1-p)^{k-1}p \quad \mathbb{E}(X) = \frac{1}{p}$$

$$\mathbb{V}(X) = \frac{1-p}{p^2}$$

The geometric distribution is the solution of the problem "I repeat infinitely an experiment that as a probability  $p$  to succeed, when is the first time I may succeed?".

##### Binomial

$$\mathbb{P}(X = k) = \binom{n}{k} p^k (1-p)^{n-k} \quad \mathbb{E}(X) = np$$

$$\mathbb{V}(X) = np(1-p)$$

This distribution is the solution of the problem "I repeat  $n$  times an experiment with a probability  $p$  of success, how many times may I may I succeed?".

##### Poisson

$$\mathbb{P}(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}$$

If the delay between two events follows an exponential law of parameter  $\lambda$  then the number of events in a given interval of length  $l$  follow a Poisson distribution with parameter  $l\lambda$ .

**Continuous laws****Uniform**

$$f(x) = \frac{1}{b-a} \mathbf{1}_{[a,b]} \quad \mathbb{E}(X) = \frac{a+b}{2}$$

$$\mathbb{V}(X) = \frac{(b-a)^2}{12}$$

**Exponential**

$$f(x) = \lambda e^{-\lambda x} \mathbf{1}_{[0,\infty[} \quad \mathbb{E}(X) = \frac{1}{\lambda}$$

$$\mathbb{V}(X) = \frac{1}{\lambda^2}$$

A random variable  $X$  exponentially distributed as the "no memory" property:  $\mathbb{P}(X > t + s | X > t) = \mathbb{P}(X > s)$ .

The exponential law is in practice used to model life times of various objects (bulbs, bacteria) or waiting periods in queue theory due to that property.

**Normal**

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \quad \mathbb{E}(X) = \mu$$

$$\mathbb{V}(X) = \sigma^2$$

Due to the Central Limit Theorem, the Normal law is one of the most common. It appears on physics when a macroscopic phenomenon is the result of a large number of additive independent microscopic phenomena.

The information theory [] that is not studied here tells that when knowing only the average tendency of a phenomenon and its dispersion, the normal law is the only model that do not introduce additional information.

**Log-Normal**

$$f(x) = \frac{1}{x\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(\ln x - \mu)^2}{2\sigma^2}\right) \quad \mathbb{E}(X) = e^{\mu + \sigma^2/2}$$

$$\mathbb{V}(X) = (e^{\sigma^2} - 1)e^{2\mu + \sigma^2}$$

If  $X \sim \mathcal{N}(\mu, \sigma^2)$  then  $\ln X$  follows a log-normal law whose density is:

As the normal law appears naturally to model additive phenomena, the log-normal is to be used for multiplicative phenomena.

**Gamma**

$$f(x) = \frac{1}{\Gamma(k)\theta^k} x^{k-1} e^{-x/\theta} \quad \mathbb{E}(X) = k\theta$$

$$\mathbb{V}(X) = k\theta^2$$

The function gamma is the extension of the factorial to  $\mathbb{R}$ :  $\Gamma(x) = \int_0^{+\infty} t^{x-1} e^{-t} dt$  and  $\Gamma(n \in \mathbb{N}^*) = (n-1)!$ . The Gamma law is stable under addition and the exponential law is a particular case of Gamma law. Thus the use of Gamma law is in general necessary to carry out calculus when basic phenomena are exponentially distributed.

**Weibull**

$$f(x) = \frac{k}{\lambda} \left(\frac{x}{\lambda}\right)^{k-1} e^{-(x/\lambda)^k} \quad \mathbb{E}(X) = \lambda\Gamma\left(1 + \frac{1}{k}\right)$$

$$\mathbb{V}(X) = \lambda^2\Gamma\left(1 + \frac{2}{k}\right) - \mathbb{E}(X)^2$$

With  $k = 2$  the Weibull law is the Rayleigh law.

Considering that when  $k = 1$  the Weibull law is an exponential law, we can use Weibull to model durations in a more general case, with possibly a "memory".

## C.3 Stochastic geometry

### C.3.1 General Point processes

**A great topological space to work on**

Let  $(\chi, d)$  be a metric space.

It is said to be **complete** if every Cauchy sequence converges in  $\chi$  or in an equivalent way if every sequence of decreasing non empty closed subsets has a non-empty intersection.

It is said to be **separable** if there exist a sequence of points that is dense in the overall.

It is said to be **compact** if from each of its cover by opens can be extract a finite subcover.

If added to that  $\chi$  is a finite dimensional vector space, a compact is closed bounded subset of  $\chi$ . A compact is complete and separable.

The framework to work on stochastic geometry is a complete separable metric space on which it is always possible to define a Borelian  $\sigma$ -algebra:  $[\chi, \mathcal{B}]$ .

For instance one can take  $\mathbb{R}^d, \mathbb{N}$ , a compact of  $\mathbb{R}^2$  (the case in most of the applications), a geometrical graph in the sense of [102],  $[0, 1] \bmod 1$ .

**Definition**

A point process on  $[\chi, \mathcal{B}]$  is a mapping  $\Phi$  from a probability space  $[\Omega, \mathcal{A}, \mathbb{P}]$  to  $\mathfrak{N}$ , the set of locally finite and simple sets of points in  $\chi$  such that for  $B \in \mathcal{B}$  bounded,  $\Phi(B)$  is a random variable counting the number of points in  $\Phi$  that lay on  $B$ .

$\mathcal{N}$  is the smallest  $\sigma$ -algebra that makes the mappings  $\Phi_B(\omega) = (\Phi(\omega))(B)$  measurable when  $B$  is a bounded Borel set.

An element  $N$  of  $\mathfrak{N}$  can be seen both as a countable set of random points:  $N = \{X_1, \dots, \}$  or a random counting measure:  $N = \sum \delta_{X_i}, \int f(x)dN(x) = \sum f(X_i)$ .

Remark that since  $\Phi$  is simple,  $\mathbb{P}(\Phi(\{x\}) > 1) = 0$ .

**Distributions, void probability**

The distribution of a point process is the function

$$\mathbb{P}(\cdot \in \mathcal{N})$$

The **finite dimensional distribution** is the function

$$(B_i), (n_i) \rightarrow \mathbb{P}(\Phi(B_1) = n_1, \dots, \Phi(B_k) = n_k)$$

To end with, the **void probability** is the function

$$B \rightarrow \mathbb{P}(\Phi(B) = 0)$$

**The void probabilities determine the finite dimensional distribution that determines the distribution.** For the first implication it is sufficient to remark that the sets  $\{\}$  forms a semi-ring generating  $\mathcal{N}$  (since it is defined as the smallest  $\sigma$ -algebra such mappings  $\omega \rightarrow \Phi(\omega)(A) \in B$  are measurable) and thus their probabilities (the finite dimensional distributions) fully define the distribution of  $\Phi$ .

To come to the second part, we will show that  $\mathbb{P}(\Phi(A_1) \leq n_1, \dots, \Phi(A_k) \leq n_k)$  can be written with  $\nu : A \rightarrow \nu(A) = \mathbb{P}(\Phi(B) = 0)$ .

To start with we only consider a bounded Borelian  $A$ . Since  $\chi$  is separable, we can define on it a sequence of nested partition  $(\Xi^{(n)})_n$  with  $\Xi^{(n)} = \{\Xi_k^{(n)}\}_{1 \leq k \leq n}$  and  $\mu(\Xi_k^{(n)}) \rightarrow 0$  when  $n \rightarrow \infty$ .

We write  $A_k^{(n)} = A \cap \Xi_k^{(n)}$  and  $H(A_k^{(n)}) = \mathbb{1}_{\Phi(A_k^{(n)}) \geq 1}$ .  $H_n(A) = \sum H(A_k^{(n)})$  counts the number of sets in the  $n$ -th partition and in  $A$  that contain points of  $\Phi$ .  $H_n$  is increasing and since  $\Phi$  is simple, from a particular rank  $H_n(A) = \Phi(A)$ .

$\mathbb{P}(H_n(A) = N) = \sum \mathbb{P}(H(A_1^{(n)}) = i_1, \dots, H(A_{n_k}^{(n)}) = i_{n_k})$  with  $i_j \in \{0, 1\}$  and  $\sum i_j = N$ .

When passing to the limit, we get an expression of  $\mathbb{P}(\Phi(A) = n)$  as a function of  $\nu$ .

**Intensity** The mean value of points for a bound Borelian  $B$  writes:  $\Lambda(B) = \mathbb{E}(\Phi(B))$  it is called the intensity measure.

$$\mathbb{E}\left(\int f(x)d\Phi(x)\right) = \int f(x)d\Lambda(x)$$

which is more or less the principle of the Monte-Carlo method: if  $\Phi$  is a 1-dimensional stationary Poisson Point Process of intensity  $\lambda$  and  $f(x) = x^2 \mathbb{1}_{[-1,1]}(x)$  then

$$\mathbb{E}\left(\int f(x)d\Phi(x)\right) = \lambda \int_{-1}^1 x^2 dx = \frac{2}{3}\lambda$$

Proof : First consider functions of the form  $f(x) = \mathbb{1}_A(x)$  then  $\mathbb{E}(\int f(x)d\Phi(x)) = \mathbb{E}(f(A)) = \Lambda(A) = \int f(x)d\Lambda(x)$  by linearity of the expectancy and the integral, this result is also true for finite sums of indicator function and when passing to the limit, true for every measurable function  $f$ .

### Stationarity, isotropy

A point process  $\Phi$  is stationary if its distribution is invariant under translation:  $\mathbb{P}(\Theta_x \Phi) = \mathbb{P}(\Phi)$ ,  $\forall x \in \mathbf{R}^d$ .

It is invariant under rotation or isotropic if for every rotation matrix  $\mathbb{R}$ ,  $\mathbb{P}(\mathbf{R}(\Phi)) = \mathbb{P}(\Phi)$ .

A stationary point process is such that  $\Lambda(B) = \lambda \cdot \mu(B)$  and  $\lambda$  is then the intensity of the process.

**Marked point processes**

A marked point process is a point process when to each point is associated a random mark in a measurable space  $[\mathfrak{M}, \mathcal{M}]$ . In fact it can be defined as a simple point process  $\Psi$  in  $[\chi \times \mathbb{M}, \mathcal{B} \otimes \mathcal{M}]$  where  $\Psi(\cdot \times \mathfrak{M})$  has to be a point process also.

Its intensity is then simply defined on a  $\sigma$ -algebra  $\mathcal{B} \otimes \mathcal{M}$  by  $\Lambda(B \times L) = \mathbb{E}(\Phi(B \times L))$ . And we can also consider the intensity measure of the point process without marks  $\Lambda_p(\cdot) = \Lambda(\cdot \times \mathfrak{M})$ . Then we observe that  $\Lambda(\cdot \times L)$  is absolutely continuous with respect of  $\Lambda_p$  whatever  $L$  and thus there exist a measure  $M_x$ :

$$\Lambda(B \times L) = \int_B M_x(L) \Lambda_p(d\mu(x))$$

$M_x(\cdot)$  can be seen as the probability distribution of marks knowing there is a point in  $x$ . We have the Campbell formula for marked point processes:

$$\mathbb{E}\left(\sum_{(x,y) \in \Psi} f(x,y)\right) = \int \int f(x,y) dM_x(y) d\Lambda_p(x)$$

If  $\Psi$  is a marked point process, the shifted process  $\Theta_x \Psi$ ,  $x \in \mathbb{R}^d$  is defined by  $\Theta_x \Psi(B \times L) = \Psi((B+x) \times L)$  if  $\Psi$  and every  $\Theta_x \Psi$  are equal in distribution, the marked point process is stationary.

If  $\Psi$  is stationary, the mark distribution  $M_x$  do not depend on  $x$  and thus writes  $M$  and  $\Lambda_p(x) = \lambda \mu(x)$ . The Campbell formula then simplifies to

$$\mathbb{E}\left(\sum_{(x,y) \in \Psi} f(x,y)\right) = \lambda \int \int f(x,y) dM(y) d\mu(x)$$

**Palm distribution**

Let  $\Phi$  be a point process of intensity  $\Lambda$ .

Then one defines the Campbell distribution of this process as the measure on  $\mathbb{B} \times \mathbb{N}$

$$\mathcal{C}(B, Y) = \mathbb{E}(\Phi(B) \mathbf{1}(Y))$$

that is to say the mean number of points in  $B$  when  $\Phi$  has property  $Y$ . It generalizes the notion of intensity.  $\mathcal{C}(\cdot \times Y)$  is absolutely continuous with respect to  $\Lambda$  so we can write for any property  $Y$ :

$$\mathcal{C}(B \times Y) = \int_B P_x(Y) d\Lambda(x)$$

which is interpreted as the sum of the probabilities that  $\Phi$  respects  $Y$  when  $\Phi$  has a point in  $x$  times the probability that  $\Phi$  has a point in an infinitesimal ball centred in  $x$ .

$$\mathbf{E}\left(\sum_{x \in \Phi} h(x, \Phi)\right) = \int h(x, \phi) d\mathcal{C}(x, \phi) = \int \int h(x, \phi) d\mathbb{P}_x(\phi) d\Lambda(x)$$

If  $\Phi$  is stationary then  $P_x(Y) = P_0(\mathcal{T}_{-x}Y)$  which is of practical interest: one can evaluate

$P_0(Y)$  on a single realization  $\phi$  of  $\Phi$  by averaging: We have just defined the notion of "typical point" of a point process. The typical point of a process is another process giving a point-centred- and its neighbourhood. Its distribution provides for a property  $A$  the probability that the process has property  $A$  when it has a centred point.

If we have a particular realization of the point process, we can estimate the probability of the typical point by taking a sequence of bounded windows whose size tends to infinity. When taking a point at random in these windows (which is possible since it is bounded and the process is locally finite), the distribution of the process centred on that point is the Palm distribution.

For a point process, this concept is trivial but it will show its interest for objects processes: for instance we will speak of tessellation processes. We will not mix up the "typical" cell and the 0-cell. They are both polygon random variables. To define the "typical" one imagine you take a large but bounded window, number each cell it contains, pick a number at random and then look at the cell with that number. Then you get a "typical" cell with its "typical" distribution. To get the 0-cell you simulate a large number of process and look at the cell containing the origin on each simulation. Then you get the 0-cell. Generally speaking, the 0-cell is larger than the typical cell.

### C.3.2 Poisson Point processes

#### Definition and existence

If  $\Lambda(\cdot)$  is a given locally finite measure, there exist a point process  $\Phi$  called the Poisson-process of intensity measure  $\Lambda$  that checks:

1.  $\Phi(A_1), \dots, \Phi(A_m)$  are independent random variables when  $A_1, \dots, A_m$  are measurable and pairwise disjoint
2.  $\mathbb{P}(\Phi(A) = n) = e^{-\Lambda(A)} \frac{\Lambda(A)^n}{n!}$  when  $(A)$  is measurable

If  $\Lambda(d\mu) = \lambda d\mu$  we will talk of an homogeneous Poisson process of intensity  $\lambda$ .

#### Simulation

Let  $A$  be a bounded Borel set and  $\Phi$  a Poisson point process of intensity  $\Lambda$ . Conditional on  $\{\Phi(A) = n\}$ ,  $\Phi|_A = (X_1, \dots, X_n)$  where  $(X_i)$  are independent  $\Lambda$ -uniformly distributed random variables in  $A$ .

Then to simulate a Poisson Point process in  $A$ , one has to pick a number of points  $N$  as a Poisson random variable and then dispatch  $n$  points in  $A$ , at random with respect to  $\Lambda$ .

Similarly, Poisson processes are the only point process that provides a uniform distribution of points on each bounded Borelian, which make them be the "most random point processes".



### Characterization

**The Laplace functional characterization** If  $\Lambda$  is a locally finite measure on  $\mathbb{R}^d$ , then a point process  $\Phi$  is a Poisson point process with intensity  $\Lambda$  iff

$$\mathbb{E}(\exp(-\int f(x)d\Phi(x))) = \exp(-\int (1 - e^{-f(x)})d\Lambda(x))$$

**Mecke's formula** If  $\Lambda$  is a locally finite measure on  $\mathbb{R}^d$ , then a point process  $\Phi$  is a Poisson point process with intensity  $\Lambda$  iff

$$\mathbb{E}(\int f(x, \Phi)d\Phi(x)) = \mathbb{E}(\int f(x, \Phi + \delta_x)d\Lambda(x))$$

### The Palm distribution for stationary point processes

$$\mathbb{P}_{\Phi}^0(\cdot) = \mathbb{P}(\Phi + \delta_0 \in \cdot)$$

### C.3.3 Finite point processes

#### Definition by distributions

They are a useful for applications since generally speaking a statistician only accesses to a limited region of observation, "a window"  $W$ .

To define a finite point process on  $W$ , one can define:

1. A probability  $p(n)$  giving the number of points in  $W$
2. A family of symmetric probabilities  $j_n : W^n \rightarrow [0, 1]$  giving the position of each point conditionally to the number of points.

In fact we do not define directly a point process from this definition but a random variable on  $W \cup W^2 \cup \dots$ . Nonetheless since  $j_n$  is chosen to be symmetric, it is possible to "forget" the order of points and to get a point process.

#### Definition by density

Let  $\Phi_\lambda$  be an homogeneous Poisson Point Process with intensity  $\lambda$  and probability distribution  $\pi_\lambda$ . Let  $W$  be a window and  $\mathfrak{N}^f$  the set of finite point processes on  $W$ .  $\mathfrak{N}^f$  is  $\mathcal{N}$ -measurable. A positive measurable function  $p$  such that

$$\int_{\mathfrak{N}^f} p(\vec{x})d\pi(\vec{x}) = 1$$

is a density function.

Point processes defined this way can be simulated by a Metropolis algorithm.

For instance  $p(\vec{x}) = Cste\beta^{n(\vec{x})}\gamma^{s(\vec{x})}$  with  $\beta, \gamma > 0$ ,  $Cste$  a normalization constant (incalculable),  $n(\vec{x}) =$  number of points in  $\vec{x}$  and  $s(\vec{x}) =$  number of pairs of points with a distance  $>$

$\epsilon$  is called a Strauss process. Its behaviour depends on  $0 < \gamma < 1$  (points are "repulsed" by each others) or  $\gamma > 1$  (points are "attracted" by each others).

### C.3.4 Tessellation process

If a tessellation process is stationary, its associated nodes, edges mid-points, cell centroids are stationary point processes. We can also consider the mean geometrical parameters of the typical cell :

$\lambda_0$	Mean number of nodes per unit area
$\lambda_1$	Mean number of side mid points per unit area
$\lambda_2$	Mean number of cell centroids per unit area
$\lambda_3$	Mean length of side per unit area
$\bar{l}$	Length of the typical side
$\bar{U}$	Perimeter of the typical cell
$\bar{A}$	Area of the typical cell
$n_{\bar{0}2}$	Number of sides from a typical node the same as number of cells touching the typical node
$n_{\bar{2}0}$	Number of node on the boundary of the typical cell

These various parameters are linked by the following relationships:

$$\lambda_1 = \lambda_3 + \lambda_2$$

$$\lambda_2 \bar{U} = 2\lambda_1 \bar{l}$$

$$\lambda_2 \bar{A} = 1$$

$$n_{\bar{0}2} = 2 + 2\frac{\lambda_2}{\lambda_0}$$

$$n_{\bar{2}0} = 2 + 2\frac{\lambda_0}{\lambda_2}$$



# Bibliography

- [1] Réka Albert and Albert-Lazlo Barabasi. Statistical mechanics of complex networks. *Reviews of modern physics*, 74:47–97, 2002.
- [2] Sergio Albeverio, Denise Andrey, Paolo Giordano, and Alberto Vancheri. *The Dynamics of Complex Urban Systems*. 2008.
- [3] C. Andersson, A. Hellervik, and K. Lindgren. A spatial network explanation for a hierarchy of urban power laws. *Physica A*, 345:227–244, 2005.
- [4] F. Auerbach. Das gesetz der bevölkerungskonzentration. *Petermanns Geographische Mitteilungen*, 59:74–76, 1913.
- [5] Francois Baccelli and Bartłomiej Błaszczyszyn. *Stochastic Geometry and Wireless Networks, Part I: Theory*. Now Publishers Inc, 2009.
- [6] Francois Baccelli and Bartłomiej Błaszczyszyn. *Stochastic Geometry and Wireless Networks, Part II: Applications*. Now Publishers Inc, 2009.
- [7] F. Bach and M. I. Jordan. Learning spectral clustering. *Advances in Neural Information Processing Systems*, 16, 2004.
- [8] Barthélemy and Flammini. Modeling urban street patterns. *Phys Rev Lett*, 100:138702, 2008.
- [9] Marc Barthélemy. Spatial networks. *arXiv:1010.0302v2*, 2010.
- [10] Marc Barthélemy and Alessandro Flammini. Co-evolution of density and topology in a simple model of city formation. *Networks and spatial economics*, 9:401–425, 2009.
- [11] M. Batty and P. Longley. *Fractal Cities*. Academic, San Diego, 1994.
- [12] A. Bavelas. A mathematical model for group structures. *Human Organization*, 7:16–32, 1948.
- [13] M. Beil, H. Braxmeier, F. Fleischer, V. Schmidt, and P. Walther. Qualitative analysis of keratin filament networks in scanning electron microscopy images of cancer cells. *Journal of Microscopy*, 220:84–95, 2005.

- [14] Michel Benaïm and Nicole El Karaoui. *Promenades aléatoires*. Editions de l'Ecole Polytechnique, 2007.
- [15] L. Benguigui. A new aggregation model. application to town growth. *Physica A*, 219:13–26, 1995.
- [16] Claude Berge. *Graphes et Hypergraphes*. Dunod, 1973.
- [17] Marc Bernot, Vicent Caselles, and Jean-Michel Morel. *Branched transportation networks*. Springer Complexity, 2007.
- [18] A. Bertaud. The spatial organization of cities. *A. Bertaud's web site*, 2005.
- [19] Luis M. A. Bettencourt, José Lobo, Dirk Helbing, Christian Kühnert, and Geoffrey B. West. Growth, innovation, scaling, and the pace of life in cities. *PNAS*, 104:7301–7306, 2007.
- [20] Agnieszka Bitner, Robert Holyst, and Marcin Fialkowski. From complex structures to complex processes: Percolation theory applied to the formation of a city. *Physical Review E*, 80:037102, 2009.
- [21] Philippe Blanchard and Dimitri Volchenkov. *Mathematical analysis of urban spatial networks*. Springer Complexity, 2009.
- [22] Paul Blanquart. *Une histoire de la Ville Pour repenser la société*. La découverte, 1997.
- [23] S. Boccaletti, V. Latorab, Y. Morenod, M. Chavezf, and D.-U. Hwanga. Complex networks: Structure and dynamics. *Physics Reports*, 424:175–308, 2006.
- [24] S. Bohn, S. Douady, and Y. Couder. Four sided domains in hierarcical space dividing patterns. *Phys Rev Lett*, 94:054503, 2005.
- [25] Paul Bourguine and Annick Lesne. *Morphogenèse, L'origine des formes*. Echelles, 2006.
- [26] Hervé Le Bras. *Essai de géométrie sociale*. Editions Odile Jacob, 2000.
- [27] A. Bretagnolle and D. Pumain. Comparer deux types de systèmes de villes par la modélisation multi-agents. *Qu'appelle t-on aujourd'hui les sciences de la complexité? Langages, réseaux, marchés, territoires*, G. Weisbuch et A. Zwirn (Ed.), *On Line*: <http://halshs.archives-ouvertes.fr/docs/00/53/03/23/PDF/Comparer-Texte-Hal.pdf>, pages 271–299, 2010.
- [28] Anne Bretagnolle, Eric Daudet, and Denise Pumain. From theory to modelling : urban systems as complex systems. *CyberGeo: European Journal of Geography*, 335:1–17, 2006.
- [29] Anne Bretagnolle, Fabien Paulus, and Denise Pumain. Time and space scales for measuring urban growth. *CyberGeo: European Journal of Geography*, 219, 2002.

- [30] Anne Bretagnolle, Denise Pumain, and Céline Vacchiani-Marcuzzo. Les formes des systèmes de villes dans le monde. *Données urbaines, Mattei M.-F., Pumain D. (Ed.); On Line on: <http://halshs.archives-ouvertes.fr/docs/00/15/00/65/PDF/DU5dpabcvm.pdf>, pages 301 – 314, 2007.*
- [31] Anne Bretagnolle, Denise Pumain, and Céline Vacchiani-Marcuzzo. The organisation of urban systems. *Complexity perspective in innovation and social change, D. Lane, D. Pumain, S. Van der Leeuw, G. West (Ed.). Online on: <http://halshs.archives-ouvertes.fr/docs/00/45/97/13/PDF/Organization-Bretagnolle-Hal.pdf>, pages 197–220, 2009.*
- [32] Jean Brunhes and Pierre Deffontaine. *Géographie humaine de la France*. Plon, 1947.
- [33] J. Buhl, J. Gautrais, N. Reeves, R.V. Solé, S. Valverde, P. Kuntz, and G. Theraulaz. Topological patterns in street networks of self-organized urban settlements. *Eur. Phys. J. B*, 49:513–522, 2006.
- [34] J. Buhl, J. Gautrais, R.V. Solé, P. Kuntz, S. Valverde, J.L. Deneubourg, and G. Theraulaz. Efficiency and robustness in ant networks of galleries. *Eur. Phys. J. B*, 42:123–129, 2004.
- [35] Alessio Cardillo, Salvatore Scellato, Vito Latora, and Sergio Porta. Structural properties of planar graphs of urban street patterns. *Physical Review E*, 73:066107, 2006.
- [36] Rui Carvalho and Alan Penn. Scaling and universality in the micro-structure of urban space. *Physica A*, 332:539–547, 2004.
- [37] Guoning Chen, Gregory Esch, Peter Wonka, Pascal Mü, and Eugene Zhang. Interactive procedural street modeling. *ACM Transactions on graphics*, 27:103, 2008.
- [38] Françoise Choay. *L’urbanisme, utopies et réalités*. Seuil, 1965.
- [39] W. Christaller. *Centrale Places in Southern Germany*. Prentice-Hall (translated from the 1933 German edition), 1966.
- [40] C. Clark. Urban population densities. *Journal of the Royal Statistical Society. Series A*, 114:490–496, 1951.
- [41] Richard Cowan. New classes of random tessellations arising from iterative division of cells. *Advances in Applied Probability*, 42:26–47, 2010.
- [42] Paolo Crucitti, Vito Latora, and Sergio Porta. Centrality measures in spatial networks of urban streets. *Physical Review E*, 73:036125, 2006.
- [43] Mark de Berg, M. van Krefeld, M. Overmars, and O. Schwarzkopf. *Computational Geometry: Algorithms and Applications*. Springer Complexity, 2000.
- [44] Jean-François Delmas. *Cours de statistique et analyse de données*. Ecole nationale des Ponts et Chaussées <http://perso-math.univ-mlv.fr/users/martinez.miguel/Cours/CoursStatDelmas.pdf>, 2007.

- [45] Stéphane Douady, Audrey Manning, and Bernard Hennion. In the beginning there was shape. *Origins of Life*, pages 47–60, 2009.
- [46] Murray Eden. A two-dimensional growth process. *Proc. Fourth Berkeley Symp. on Math. Statist. and Prob.*, 4:223–239, 1961.
- [47] Erlander and Stewart. *The Gravity Model in Transportation Analysis – Theory and Extensions*. 1990.
- [48] Lucas Figueiredo and Luiz Amorim. Decoding the urban grid: or why cities are neither trees nor perfect grids. In *6th International Space Syntax Symposium*, 12-15 June 2007.
- [49] Lucas Figueiredo and Luiz Amorim. Continuity line in the axial system. In *5th International Space Syntax Symposium*, 2005.
- [50] F. Fleischer, C. Gloaguen, H. Schmidt, V. Schmidt, and F. Schweiggert. Simulation algorithm of typical modulated poisson-voronoi cells and application to telecommunication network modelling. *Japan Journal of Industrial and Applied Mathematics*, 25:305–330, 2008.
- [51] Pierre Frankhauser. Morphologie des "villes émergentes" en europe à travers les analyses fractales. Technical report, ThêMA UMR 6049 CNRS, 2000.
- [52] Pierre Frankhauser. La ville fractale et la fractalité des villes. *La ville émergente, résultats de recherche*, 2:147–161, 2002.
- [53] Pierre Frankhauser. Lanalyse fractale pour décrire la structure spatiale des villes. *Images de Franche-Comté*, 26:6–9, 2002.
- [54] Thomas Fricker. Statistical analysis and segmentation of inhomogeneous random graphs. Master’s thesis, Univesitat Ulm, 2009.
- [55] Sebastian Frommherz. Exhaustive representation of the real roadway system data for the construction of a data base. Master’s thesis, Universitat Ulm, 2007.
- [56] Laetitia Gauvin, Jean-Pierre Nadal, and Jean Vannimenus. Schelling segregation in an open city: a kinetically constrained blume-emery-griffiths spin-1 system. *Phys Rev E*, 81, 2010.
- [57] C. Gloaguen, P. Coupe, R. Maier, and V. Schmidt. Stochastic modelling of urban access networks. In *Proc. 10th Internat. Telecommun. Network Strategy Planning Symp. (Munich, June 2002)*, 2002.
- [58] C. Gloaguen, F. Fleischer, H. Schmidt, and V. Schmidt. Simulation of typical cox-voronoi cells, with a special regard to implementation tests. *Mathematical Methods of Operations Research*, 62:357–373, 2005.
- [59] C. Gloaguen, F. Fleischer, H. Schmidt, and V. Schmidt. Fitting of stochastic telecommunication network models via distance measures and monte-carlo tests. *Telecommun. Syst.*, 31:4, 2006.

- [60] C. Gloaguen, F. Fleischer, H. Schmidt, and V. Schmidt. Analysis of shortest paths and subscriber line lengths in telecommunication access networks. *Networks and Spatial Economics*, 10:15–47, 2010.
- [61] C. Gloaguen, F. Voss, and V. Schmidt. Parametric distributions of connection lengths for the efficient analysis of fixed access network. *Annals of Telecommunications*, 66:103–118, 2011.
- [62] Catherine Gloaguen, Florian Voss, and Volker Schmidt. Parametric distance distributions for fixed access network analysis and planning. In *ITC21*, 2009.
- [63] Sébastien Grauwina, Eric Bertin, Rémi Lemoy, and Pablo Jensen. Competition between collective and individual dynamics. *PNAS*, 106:20622–20626, 2009.
- [64] Jonathan L. Gross and Jay Yellen. *Handbook of graph theory*. CRC Press, 2003.
- [65] T. Hastie, R. Tibshirani, and J. Friedman. *Elements of statistical learning*. Springer Complexity, 2009.
- [66] B. Hillier and J. Hanson. *The Social Logic of Space*. Cambridge University Press, 1984.
- [67] Bill Hillier. A theory of space as object: or, how spatial laws mediate the social construction of urban space. *Urban Des. Int.*, 7:153–179, 2002.
- [68] Bill Hillier, A. Penn, J. Hanson, T. Grajewski, and J. Xu. Natural movement: or, configuration and attraction in urban pedestrian movement. *Environment and Planning B : Planning and design*, 20:29–66, 1993.
- [69] Bill Hillier and Alan Penn. Rejoinder to carlo ratti. *Environment and Planning B : Planning and design*, 31:501–511, 2004.
- [70] Y. Ioannides and Henry Overman. Zipfs law for cities: An empirical examination. Technical report, Centre for Economic Performance, LSE, 2000.
- [71] W. Isard. *Location and Space-Economy: A General Theory Relating to Industrial Location, Market Areas, Land Use, and Urban Structure*. The MIT Press, 1972.
- [72] Pablo Jensen. A network-based prediction of retail stores commercial categories and optimal locations. *Phys Rev E*, 74, 2006.
- [73] Bin Jiang and Christophe Claramunt. Integration of space syntax into gis: New perspectives for urban morphology. *Transactions in GIS*, 6:295–309, 2002.
- [74] Bin Jiang and Christophe Claramunt. Topological analysis of urban street networks. *Environment and Planning B : Planning and design*, 31:151–162, 2004.
- [75] Vamsi Kalapala, Vishal Sanwalani, Aaron Clauset, and Cristopher Moore. Scale invariance in road networks. *Physical Review E*, 73:026130, 2006.
- [76] Wilfrid S. Kendall and Ilya Molchanov. *New Perspectives in Stochastic Geometry*. 2010.



- [77] E. Limpert, W. Stahel, and M. Abbt. Log-normal distributions across the science: Keys and clues. *BioScience*, 51:341, 2001.
- [78] V. Liskevich. *Measure theory*. <http://www-maths.swan.ac.uk/staff/vl/c98.pdf>, 1998.
- [79] Stefan Lämmner, Björn Ghlsen, and Dirk Helbing. Scaling laws in the spatial structure of urban road networks. *Physica A*, 363:89–95, 2006.
- [80] August Lösch. Population cycles as a cause of business cycles. *The Quarterly Journal of Economics*, 1937.
- [81] Kevin Lynch. *The Image of the City*. The MIT Press, 1960.
- [82] R. Maier, J. Mayer, and V. Schmidt. Distributional properties of the typical cell of stationary iterated tessellations. *Mathematical Methods of Operations Research*, 59:287–302, 2004.
- [83] R. Maier and V. Schmidt. Stationary iterated tessellations. *Advances in Applied Probability*, 35:337–353, 2003.
- [84] Roland Maier. *Iterated Random TEssellations with Application to Spatial Modelling of Telecommunication Networks*. PhD thesis, Universität Ulm, 2003.
- [85] Hernán A. Makse and Shlomo Havlin and H. Eugene Stanley. Modelling urban growth patterns. *Nature*, 377:608 – 612, 2002.
- [86] Stephen Marshall. *Streets and patterns*. Spon, 2005.
- [87] Hernan A. Maske, José S. Andrade Jr, Michael Batty, Shlomo Havlin, and H. Eugene Stanley. Modeling urban growth patterns with correlated percolation. *Physical Review E*, 58:7054–7062, 1998.
- [88] J. Mayer, V. Schmidt, and F. Schweigger. A unified simulation framework for spatial stochastic models. *Simulation Modelling Practice and Theory*, 12:307–326, 2004.
- [89] Pierre Mazet. *Intégration*. <http://www.math.jussieu.fr/mazet/integration/A1poly.pdf>, 2003.
- [90] J. Mecke, W. Nagel, and V. Weiss. Length distributions of edges in planar stationary and isotropic stit tessellations. *Journal of Contemporary Mathematical Analysis*, 42:28–43, 2007.
- [91] K. R. Mecke and D. Stoyan. *Statistical Physics and Spatial Statistics*. Springer Complexity, 2000.
- [92] Pierre Merlin. *L’urbanisme*. PUF, 2010.
- [93] Michel Métivier and Jacques Neveu. *Cours de probabilités*. Ecole Polytechnique, 1979.
- [94] B. Nadler and M. Galun. Fundamental limitations of spectral clustering methods. In *Neural Information Processing Systems Conference, Vancouver, NIPS-06*, 2006.

- [95] W. Nagel and V. Weiss. Crack stit tessellations : Characterization of stationary random tessellations stable with respect to iteration. *Advances in applied probability*, 37:859–883, 2005.
- [96] W. Nagel and V. Weiss. Mean values for homogeneous stit tessellations in 3d. *Image Anal Stereol*, 27:29–37, 2008.
- [97] D. Neuhaeuser, C. Hirsch, C. Gloaguen, and V. Schmidt. On the distribution of typical shortest-path lengths in connected random geometric graphs. *Submitted to Queueing Systems, Preprint: <http://www.mathematik.uni-ulm.de/stochastik/personal/schmidt/publications/ConnectedGraphs.pdf>*, 2012.
- [98] M. E. J. Newman. The structure and function of complex networks. *SIAM REVIEW*, 45:167–256, 2003.
- [99] Andrew Y. Ng, Michael Jordan, and Yair Weiss. On spectral clustering: Analysis and an algorithm,. *Neural Information Processing Systems*, 14, 2002.
- [100] G. Nicolas. Le prétendu "modèle cristalléen". [http://cyberato.pu-pm.univ-fcomte.fr/sites/default/files/cyberato/nicolas-georges/forum/disputatoire/christaller-walter-modele/nicolas-georges\\_pretendu - modele - christallerien250.pdf](http://cyberato.pu-pm.univ-fcomte.fr/sites/default/files/cyberato/nicolas-georges/forum/disputatoire/christaller-walter-modele/nicolas-georges_pretendu - modele - christallerien250.pdf).
- [101] Y. I. H. Parish and P. Müller. Procedural modeling of cities. *Proceedings of ACM SIGGRAPH*, pages 301–308, 2001.
- [102] Mathew Penrose. *Random Geometric Graphs*. Oxford University Press, 2003.
- [103] Andrea Perna, Pascale Kuntz, and Stéphane Douady. Large scale coherence of spatial network-like patterns identified using local geometry. *arXiv:1101.1133v1*, 2010.
- [104] Julien Perret. *Modélisation d'environnements urbains virtuels*. PhD thesis, Université de Rennes 1, 2006.
- [105] Sergio Porta, Paolo Crucitti, and Vito Latora. The network analysis of urban streets: A dual approach. *Physica A*, 369:853, 2006.
- [106] Sergio Porta, Paolo Crucitti, and Vito Latora. The network analysis of urban streets: A primal approach. *Environment and Planning B: Planning and Design*, 33:705–725, 2006.
- [107] Sergio Porta, Paolo Crucitti, and Vito Latora. Multiple centrality assessment in parma: a network analysis of paths and open space. *Urban Des. Int.*, 13:41–50, 2008.
- [108] Sergio Porta, Emanuel Strano, and Valentino Iacoviello. Street centrality and densities of retail and services in bologna, italy. *Environment and Planning B : Planning and design*, 36:450–465, 2009.
- [109] Przemyslaw Prusinkiewicz and Aristid Lindenmayer. *The Algorithmic Beauty of Plants*. Springer-Verlag, 1996.
- [110] Carlo Ratti. Rejoinder to hillier and penn. *Environment and Planning B : Planning and design*, 31, 2004.

- [111] Carlo Ratti. Urban texture and space syntax: some inconsistencies. *Environment and Planning B: Planning and Design*, 31, 2004.
- [112] C. Redenbach and C. Thäle. On the arrangement of cells in planar stit and poisson line tessellations. *Preprint: arXiv:1105.0770*.
- [113] Marcel Roncayolo. *La ville et ses territoires*. Folio Essais, 1990.
- [114] M. Rosvall, A. Trusina, P. Minnhagen, and K. Sneppen. Networks and cities: An information perspective. *Phys Rev Lett*, 94:028701, 2005.
- [115] A. Runions, A. M. Fuhrer, B. Lane, P. Federl, A.-G. Rolland-Lagan, and P. Prusinkiewicz. Modeling and visualization of leaf venation patterns. *ACM Transactions on Graphics*, 24:702–711, 2005.
- [116] Lena Sanders, Jean-Marc Favaro, H el ene Mathian, Denise Pumain, and Beno t Glisse. Intelligence artificielle et agents collectifs : le mod ele eurosim. *Cybergeo : Revue europ eenne de g eographie.*, 392, 2007.
- [117] Salvatore Scellato, Alessio Cardillo, Vito Latora, and Sergio Porta. The backbone of a city. *European Physical Journal B*, 50:221–225, 2006.
- [118] Hendrik Schmidt. Telecommunication access network modelling. Master’s thesis, Universitat Ulm, 2002.
- [119] Viktoriya Semeshenko, Mirta B. Gordon, and Jean-Pierre Nadal. Collective states in social systems with interacting learning agents. *Physica A*, 387:4903–4916, 2008.
- [120] K. W. Soo. Zipf’s law for cities: a cross-country investigation. *Regional Science and Urban Economics*, 35:239–263, 2005.
- [121] J. Q. Stewart. Demographic gravitation: Evidence and applications. *Sociometry*, 11:31–58, 1948.
- [122] D. Stoyan, W.S. Kendall, and J. Mecke. *Stochastic Geometry and its applications*. J. Wiley & Sons, 1995.
- [123] C ecile Tannier and Denise Pumain. Fractals in urban geography : a theoretical outline and an empirical example. *CyberGeo: European Journal of Geography*, 307, 2005.
- [124] C. Thale. Moments of the length of line segments in homogeneous planar stit tessellations. *Image Anal Stereol*, 28:69–76, 2009.
- [125] Ralf Thiedmann. Analysis of the stochastic subscriber line model in paris. Master’s thesis, Universitat Ulm, 2006.
- [126] Bruno Augusto Nassif Traven olo and Luciano da Fontoura Costa. Outward accessibility in urban street networks: Characterization and improvements. *arXiv:0802.3665*, 2008.
- [127] Jean-Fran ois Tribillon. *L’urbanisme*. La d ecouverte, 1990.

- 
- [128] G. Troll and P. beim Graben. Zipfs law is not a consequence of the central limit theorem. *Phys Rev E*, 57:1347–1355, 1998.
- [129] M. N. M. van Lieshout. *Markov point processes and their applications*. Imperial College Press, 2000.
- [130] Ulrike von Luxburg. A tutorial on spectral clustering. *Statistics and Computing*, 17, 2007.
- [131] Johann Heinrich von Thünen. *Recherches sur l'influence que le prix des grains, la richesse du sol et les impôts exercent sur les systèmes de culture*. Guillaumet et Cie, 1851.
- [132] F. Voss, C. Gloaguen, F. Fleischer, and V. Schmidt. Distributional properties of euclidean distances in wireless networks involving road systems. *Distributional properties of Euclidean distances in wireless networks involving road systems*, 27:1047–1055, 2009.
- [133] F. Voss, C. Gloaguen, F. Fleischer, and V. Schmidt. Densities of shortest path lengths in spatial stochastic networks. *Stochastic Models*, 27, 2011.
- [134] F. Voss, C. Gloaguen, and V. Schmidt. Capacity distributions in spatial stochastic models for telecommunication networks. *Image Analysis and Stereology*, 28:155–163., 2009.
- [135] F. Voss, C. Gloaguen, and V. Schmidt. Scaling limits for shortest path lengths along the edges of stationary tessellations. *Advances in Applied Probability*, 42:936–952, 2010.
- [136] A.G. Wilson. A statistical theory of spatial distribution models. *Transportation Research*, 1:253269, 1967.
- [137] Camille Wormser. *Generalized Voronoi Diagrams and Applications*. PhD thesis, University of Nice-Sophia Antipolis, 2008.
- [138] Georg K. Zipf. *Human Behavior and the Principle of Least Effort*. Addison-Wesley, New York, 1949.





