

A Thesis Submitted for the Degree of PhD at the University of Warwick

Permanent WRAP URL:

<http://wrap.warwick.ac.uk/176721>

Copyright and reuse:

This thesis is made available online and is protected by original copyright.

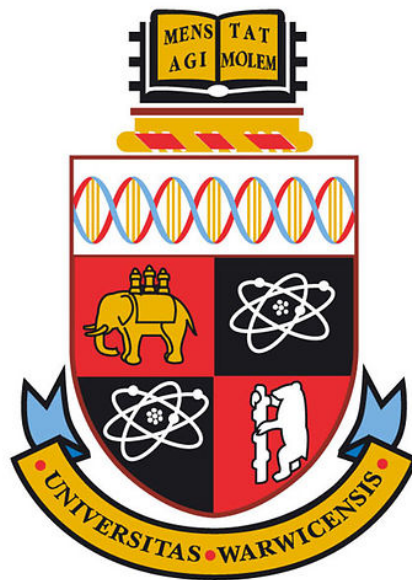
Please scroll down to view the document itself.

Please refer to the repository record for this item for information to help you to cite it.

Our policy information is available from the repository home page.

For more information, please contact the WRAP Team at: wrap@warwick.ac.uk

Structure, Entropy and Evolution of Systems of Cities



by

Matteo Mazzamurro

Submitted to the University of Warwick

in partial fulfilment of the requirements

for admission to the degree of

Doctor of Philosophy

in

Urban Science

Department of Computer Science

November 2022

Contents

List of Figures	v
List of Tables	vii
Acknowledgements	viii
Declaration	ix
Abstract	xi
1 Introduction	1
1.1 Motivation	1
1.2 Research Objectives	2
1.3 Contributions	4
1.4 Thesis outline	6
2 Background: Structure, Entropy and Evolution	8
2.1 What is a city, what is a system?	8
2.1.1 An ideal definition of city	8
2.1.2 An operational definition of city	9
2.1.3 Defining a System	11
2.2 Historical Urban Data	12
2.2.1 Data sources of historical population	12
2.2.2 Estimating human activity and social connectivity, in the present and in the past.	15
2.3 Structure and Evolution of systems of cities	18
2.3.1 In search of universal rules	18
2.3.2 From a static to a dynamic approach	21
2.4 Entropy	26
2.4.1 Definition(s)	26

2.4.2	Entropy in Urban Science	28
2.4.3	Why entropy?	32
2.5	Mathematical Framework	33
2.5.1	Points process statistics	33
2.5.2	Graphs and networks	41
2.5.3	Spatial and temporal networks	49
3	Mathematical Prelude: the Normalisation of First Degree- Based Entropy	55
3.1	Introduction	56
3.1.1	Problem Statement	56
3.1.2	Notation	57
3.1.3	Preliminary Results	60
3.2	Minimum entropy graphs given size	61
3.2.1	Main proof	61
3.2.2	Proof of Lemma 25	63
3.3	Minimum entropy graphs given order and size	67
3.3.1	Graph with small size	67
3.3.2	Main proof	70
3.3.3	Technical lemmas	72
3.3.4	Numerical observations	75
3.3.5	Approximations for applications to large graph	79
3.4	Normalisation	89
4	Cities and their Territory: Population and Spatial Organisa- tion of Systems of Cities	91
4.1	Completion and compatibility of historical population data sets	92
4.1.1	Compatibility of different data sets	92
4.1.2	Data set completion	94
4.2	Population entropy	96
4.2.1	Entropy of urban populations in Europe in 1300-1850 .	96
4.2.2	Entropy of urban populations in England and Wales in 1801-1911	99
4.3	Local entropy	101
4.3.1	Methodology	101
4.3.2	Example	103
4.3.3	A Real-world Study of Local Entropy	104
4.4	Spatial distribution of systems of cities	115

4.4.1	Town locations in England and Wales	116
4.4.2	Homogeneous models for town locations in England and Wales	118
4.4.3	Inhomogeneous models for new town locations in Eng- land and Wales	121
4.4.4	Degree entropy of disc graphs	124
4.4.5	Towards more complete point process models of town locations	129
5	Cities and their Connections: Transportation and Polycentricity of Systems of Cities	131
5.1	Flight network	132
5.2	Transportation networks, network entropy and polycentricity .	133
5.2.1	Data sets	133
5.2.2	Travel time computation	135
5.2.3	The degree entropy of disc graphs based on travel time	137
5.2.4	A human-centred morphological polycentricity: the en- tropy of potential interactions	141
5.2.5	Potential Interactions in England and Wales	142
5.2.6	Human-centred morphological polycentricity of England and Wales	145
5.2.7	Conclusion	151
6	Cities and their People: Human Activities in Systems of Cities	152
6.1	Geo-tagged tweets, dynamic model of their location	153
6.1.1	Data set	153
6.1.2	Spatial Model for aggregated tweets	153
6.1.3	Spatial models in one-hour-long intervals	155
6.1.4	Hourly evolution of the Shannon entropy	159
6.2	Adapting an entropy-based spatial interaction model for con- nections in social networks	161
6.3	Cultural Exchanges	163
6.3.1	Data set	164
6.3.2	Data cleaning	165
6.3.3	Networks of Authors' Bodies of Work	165
6.3.4	Networks of Regions	173

7 Conclusion	179
7.1 Summary	179
7.2 Limitations and future work	180
Bibliography	186

List of Figures

2.1	Central Place Theory	19
2.2	Example of Colex graph	45
2.3	Example of temporal network and its projection	50
3.1	Examples of graphs with maximum first degree-based entropy .	57
3.2	The graph $\mathcal{C}(31, 7)$	59
3.3	Young tableau of a threshold graph	76
3.4	Conjectured graphs with minimum first degree-based entropy .	78
3.5	Extrema of first degree-based entropy for graphs with 24 nodes	80
3.6	Exceptions to the conjectured graphs with minimum entropy .	81
3.7	First degree-based entropy gap for large order graphs	84
4.1	Relations between data sets of historical urban populations . .	92
4.2	Entropy of the population of European cities from 1300 to 1850	97
4.3	Entropy of the population of English and Welsh towns from 1801 to 1911	100
4.4	Neighbourhoods of a cell in raster data	102
4.5	Simple examples of land use in and around a settlement	103
4.6	Local entropy of land use in the British Isles, South Asia, and Italy, in 1700, 1900, and 2015	106
4.7	Local entropy of population density in the British Isles, South Asia, and Italy, in 1700, 1900, and 2015.	107
4.8	Total local entropy in the British Isles	109
4.9	Mean local entropy in the British Isles	110
4.10	Total local entropy in the British Isles, South Asia, and Italy .	112
4.11	Mean local entropy in the British Isles, South Asia, and Italy .	113
4.12	Locations of English and Welsh towns in 1801 and 1881	117
4.13	Test of Complete Spatial Randomness for the locations of En- glish and Welsh towns in 1801 and 1881	119

4.14	Test of a Matérn cluster processes for the locations of English and Welsh towns in 1801 and 1881	120
4.15	Inhomogeneous Matérn cluster model for new towns in England and Wales from 1801 to 1881	122
4.16	Test of an inhomogeneous Matérn cluster processes for the locations of English and Welsh towns in 1801 and 1881	123
4.17	Disc graphs for English and Welsh towns based on geographical distance	125
4.18	First degree-based entropy of disc graphs in England and Wales based on geographical distance	127
5.1	Network of worldwide flights	132
5.2	Railway network in England and Wales in 1851 and 1881	134
5.3	Computation of travel time	136
5.4	Disc graphs for English and Welsh towns based on travel time	138
5.5	First degree-based entropy of disc graphs in England and Wales based on travel time	139
5.6	Largest interaction balances between English and Welsh counties	144
5.7	Largest interaction between English and Welsh towns	145
5.8	Entropy of interactions of English and Welsh towns	146
5.9	Entropy of interactions of English and Welsh towns in 1h radius	147
5.10	Entropy of interactions of English and Welsh towns in 4h radius	149
6.1	Tweets locations in London	154
6.2	Hourly counts of Tweets in London	156
6.3	Models for the spatio-temporal distribution of tweets	157
6.4	Shannon entropy of tweets counts	160
6.5	Troubadours and contrafacta in time	167
6.6	Networks of troubadours with out- and in-degree	169
6.7	Networks of troubadours with out-degree and reachability	171
6.8	Regions in Occitania	173
6.9	Contrafacta networks in Occitan regions	174
6.10	Tensos networks in Occitan regions	176
6.11	Evolution of contrafacta networks in Occitan regions	177
6.12	Entropy of contrafacta networks in Occitan regions	178

List of Tables

3.1	Extremal (n, m) -graphs minimizing the entropy	58
4.1	Compatibility of data sets of historical populations of cities . .	95
5.1	Normalised first degree-based entropy of the flight network . .	133
5.2	Entropy of the interactions for selected cities and regions in England and Wales	150
5.3	Slope of the rank-size relation in selected regions in England and Wales	150
5.4	Shannon entropy of the population distribution in selected re- gions in England and Wales	150
6.1	Robustness of centrality measures in the troubadour networks .	172
6.2	Difference in centrality measures in the troubadour networks .	172

Acknowledgements

I feel extremely privileged because I have so many to thank.

First of all, I thank my supervisors: Weisi Guo, for supporting me during the application process and in the early stages of my PhD, and Oleg Pikhurko, for guiding me to see the end of it. I also thank my collaborators: fellow graph-enthusiast Stijn, whose mathematical prowess tested me and made me relive the pleasure of losing myself in the pursuit of a rigorous proof; and fellow Garganico Stefano, whose friendship has shaped me as a researcher more than I could ever expect.

I warmly thank my PhD cohort, the “Metropopeople”: Aparajita, Chris, Ivana, Jon, Nicole, Natalie, Shuaib, Teddy, and Zander, perfect colleagues and wonderful friends, with whom I have shared part or all of this PhD path. Your presence has been an invaluable support and contributed immensely to my personal growth. In the Computer Science Department, a special thank to Thejaswini and Aditya, whom I regret to have properly met so late. From the University of Warwick at large, I thank all the great people from History, Sociology, and English, whose friendship I cherish and who taught me the importance of interdisciplinary dialogue in research: Aditya, Andrew, Anna(s), Guido, Mantra, Ronan, Saba, Sara, Sue, Shalini, it has been a privilege to have crossed paths with you.

A deep, grateful thanks goes to my family: my parents Luigi and Maria, my sister Michela and my brother Giuseppe, for their continued support, openness, and for teaching me how to responsibly trace my own path in life; and of course to my furry brothers: Smile, Griffin, and Leo, who, as my family knows well, were the main reason why I video-called home every day in the last four years.

An affectionate thank also goes to other furry creatures that entered my life relatively recently: my cats Ludovico and Lucrezia, the former for his unconditional affection, the latter for her unconditional spite.

Finally, it is from the deepest of my heart that I thank my fiancé Somak, for his enduring love, patient spurring, and occasional, unnecessary noise.

Declaration

I, Matteo Mazzamurro, declare that the work presented in this thesis is my own, unless explicitly stated below. The thesis is based exclusively on research conducted while I was a PhD candidate at the University of Warwick and has not been submitted for a degree at another university.

Parts of this thesis are based on the following published papers, mentioned here in inverse chronological order of publication:

- [49] S. Cambie, and **M. Mazzamurro**. Resolution of Yan’s Conjecture on Entropy of Graphs. *MATCH Commun. Math. Comput. Chem.*, 89:389–404, 2023.¹
- [159] S. Milonia, and **M. Mazzamurro**. Temporal Networks of ‘Contrafacta’ in the First Three Troubadour Generations. *Digital Scholarship in the Humanities*, 2022.²
- [151] **M. Mazzamurro**, and W. Guo. Assessing the Impact of Major Historical Events on Urban Landscapes via Local Entropy Measures. *IEEE International Smart Cities Conference*, 2021.
- [152] **M. Mazzamurro**, Y. Wu, and W. Guo. Dynamic Spatial Cluster Process Model of Geo-Tagged Tweets in London. *IEEE International Smart Cities Conference*, 2019.

Other parts of this thesis contain material from the following two articles, currently under review:

- [48] S. Cambie, and **M. Mazzamurro**. Extremal entropy for graphs with given size. Currently under review.

¹The authors are listed alphabetically, as it is customary in mathematical publications. The same remark applies to [48] and [47], listed below.

²Only Matteo Mazzamurro’s contributions to this research were included in this thesis: the mathematical formulation and computational implementation of the temporal network model, and the tests of its robustness. The idea of studying the practice of contrafactum as a spatio-temporal network is Milonia’s.

- [150] **M. Mazzamurro**, and W. Guo. Network-entropy-based morphological polycentricity of systems of cities. Currently under review.

During the PhD, I worked on the following published article and submitted paper, which are not part of this thesis:

- [148] **M. Mazzamurro**. In search of lost space: the changing affordances of physical and virtual teaching spaces during the pandemic. *Journal of PGR Pedagogic Practice*, 2021.

- [47] S. Cambie, Y. Dong, and **M. Mazzamurro**. Extremal values of degree-based entropies of bipartite graphs. Currently under review.

The research was funded by the UK Engineering and Physical Sciences Research Council (EPSRC) grants EP/N509796/1 and EP/R513374/1, and by the EXCELLENT SCIENCE - Marie Skłodowska-Curie Actions grant 778305, as part of the DAWN4IoE project.

Abstract

One of the main questions in Urban Science is whether systems of cities around the world show similarities in their structure and trajectories of development. Shannon entropy has played a crucial role in this line of research, both because it is a versatile measure of uniformity and because of its ability to discriminate significant patterns from only seemingly organised maximum randomness.

In this thesis, we present novel ways to analyse the structure of systems of cities and its evolution using entropy-based measures. We focus on key morphological aspects of a system of cities: the distribution of city sizes, their spatial arrangement, the population density and land use of their surroundings, and their connectivity via transport infrastructure; which we reconnect to human activities via spatial interaction models.

We propose normalisation formulae for the first degree-based graph entropy that facilitate its interpretation as a measure of balance of the degree sequence of a network. We define a local entropy measure for raster data that quantifies the heterogeneity of a variable of interest in the surroundings of each cell. We define a measure of morphological polycentricity for historical systems of cities based on the entropy of the most likely potential interactions between the cities.

We apply our methods to analyse systems of cities in different parts of the world and moments in history. We study the evolution of the entropy of city sizes in the main European powers from 1300 to 1850; the local entropy of land use and population density in Italy, the British Isles, and South Asia from 1700 to modern day; and the spatial organisation and morphological polycentricity of English and Welsh towns in the 19th century, via the entropy of spatial networks informed by the emerging railway system.

Finally, we model the spatial-temporal dynamics of geo-tagged Tweets in London, of the Hungarian social network iWiW, and of the network of literary imitations between medieval Occitan troubadours.

Chapter 1

Introduction

Contemplating these essential landscapes, Kublai reflected on the invisible order that sustains cities, on the rules that decreed how they rise, take shape and prosper, adapting themselves to the seasons, and then how they sadden and fall in ruins.

Invisible Cities, Italo Calvino

1.1 Motivation

Since their emergence, cities have been the main places of exchange and interaction for the human race. More recently, with over 50% of the world population living in urban areas, they have also become its primary abode. With increasing rates of **urbanisation**, the destiny of Man seems to be that of becoming more and more a Homo Urbanus [65]. Yet, the growth of cities has hardly ever been a smooth process: wars, political changes, geographical discoveries, epidemics, and technological innovations have all impacted the pace and form of urbanisation.

In their uneven trajectories, cities have seen their shapes and functions multiply and transform. Emerged as local centres of trade, craftsmanship, culture, and administration, cities have expanded their influence well beyond their immediate surroundings and have become interconnected in sophisticated and complex **systems**, as a resilient solution to the ever-changing constraints of their local ecological and political environment [174]. Constantly evolving **transport infrastructure** and long-distance **social connections** have become vital to their existence and thriving.

With much of their development happening in an unplanned way, it may be tempting to intuitively equate the growth of these systems to an increase in their haphazardness. Yet, as cities grow in number and size, new forms of **structure** have been argued to emerge at the system level, from Christaller’s hierarchical organisation in the spatial distribution of cities [60], to Zipf’s rank-size law of their populations [227]. Whether the observed patterns are genuine and significant or an expected consequence of complete randomness, it remains debatable.

The questions addressed in this thesis come out directly of this ambiguity. We asked ourselves, Can **entropy**, a versatile measure of uncertainty of a distribution, be used to construct intuitive measures of the structure of a system of cities? If reasonably exhaustive entropy-inspired measures exist, do they reveal general trends, or do cities evolve in a constantly renegotiated balance between order and randomness? Can one trace the history of worldwide urbanisation, capturing its salient moments from the variation of these entropy measures in time?

Speculative questions like these, by their own nature, do not offer a definite answer, but rather propose a fascinating problem. Instead of surrendering to the spectre of sterile ineffability, we decided to try and track the trails they opened. This thesis is the result of this effort.

1.2 Research Objectives

This thesis aims to identify meaningful patterns in the physical and social structure of historical systems of cities, and consolidate entropy-based approaches for the analysis of their evolution. It is supported by the EPSRC project “Patterns of City Formation and Development”, that looks at possible ways to study the complexity of systems of cities and its evolution in time, towards the ambitious goal of identifying mathematical principles governing their emergence and development.

This research is grounded on the consideration that the essential aspects of interest of a system of cities include on the one hand its morphological structure, that is the settlements’ locations, populations, local environment, and physical connectivity through transport infrastructure, and on the other hand the human activities that the system sustains, such as the production and exchanges of goods, the social ties, and the transmission of knowledge.

This thesis proposes to tackle the following research objectives:

- RO1 Extend and refine theories on the spatial organisation of system of cities.** Based on empirical observations, it has been suggested that the distribution of settlements in a uniform, mostly rural region either abides to a certain regularity or tends to evolve towards it as a consequence of competition pressure. The spatial organisation of more urbanised and non-geographically uniform regions is much less understood, and real-world examples rather point towards increased clustering. We propose to test the hypothesis that a random cluster point process may describe the locations of cities in an industrialising system (Section 4.4). The underlying assumption is that the spatial organisation of systems of cities is informed by a variety of factors, such as the natural environment and the interaction between the centres, but is not completely deterministic, resulting instead from a realisation of a random process, subject to these constraints.
- RO2 Define meaningful entropy-based measures for the structure of systems of cities.** Shannon entropy, a simple and versatile measure of uniformity of a distribution, has been proposed as an index of concentration or regularity for the spatial distribution of cities, but also for other morphological and functional aspects of interest in the systems. We propose to adapt Shannon entropy to capture the heterogeneity of the land use and population density of the cities' surrounding territory (Section 4.3), the travel time between cities (Section 5.2.3), and the cities' interactions (Section 5.2.4).
- RO3 Identify the correlation, if any, between changes in entropy-inspired measures and major historical events and technological innovations.** Once suitable entropy-based measures of heterogeneity for several aspects of systems of cities have been identified, we plan to assess their significance in analysing the historical evolution of the systems by correlating their variation with historical events and with the emergence of technological innovations that have had a strong impact on urban structure (Sections 4.2, 4.3, and 5.2.3).
- RO4 Improve the interpretability of the first degree-based graph entropy.** Graph-based models are frequently adopted to analyse the interactions between individuals or cities. The first degree-based graph entropy (i.e., the Shannon entropy of the degree sequence of a graph) captures several aspects of potential interest in a system at the same

time: the number of constituent parts, the number of connections, and how equitably these are distributed. A correct interpretation of its value, though, requires to disentangle these aspects. We aim to identify the minima and maxima of this measure given natural constraints on the graph size (Section 3.2), or on both its order and size (Section 3.3). Our goal is that of using this information to define a normalised version of the entropy, that might allow for more easily interpretable comparisons between graphs, and thus between the systems they model.

RO5 Refine measures of morphological polycentricity. Morphological polycentricity is a concept from regional studies that describes the heterogeneity of a system of settlements in terms of their populations and spatial organisation. Existing entropy-based methods do not account for the effects of evolving transport networks on inter-city distances. We propose to define a measure of morphological polycentricity based on evolving travel time and existing entropy-maximising spatial interaction models that could better reflect the potential interactions between cities (Section 5.2.4). Our aim is that of achieving more human-centred measures of polycentricity that are closer to what an inhabitant of the system would perceive, and thus examine the strong connection between the morphological and functional aspects of systems of cities.

RO6 Test these methods with actual historical data on human activity A human-centred perspective calls for the study of individual behaviour, which may deviate from the optimal one predicted by entropy maximisation. We propose to study both modern and ancient data on people’s locations and interactions using the tools adopted and developed in the rest of the thesis: random point processes and their entropy (Section 6.1), and an entropy-based analysis and modelling of their interactions (Section 6.3).

1.3 Contributions

The thesis tackles the above research objectives and makes the following contributions:

C1 We give a mathematical proof of the extremal values of the first degree-based graph entropy for graphs of fixed size m , and for graphs with fixed order n and size m when $n - 1 \leq m \leq 2n - 3$. We provide numerical

results for the case of large order and fixed, arbitrary size. This allows to normalise the first degree-based entropy and compare graphs subject to these natural constraints, thus improving its interpretation (**RO4**). Furthermore, it solves a conjecture by Yan [221].

- C2** We define a local spatial entropy measure for raster data and use it to study the heterogeneity of the land use and population density of urban areas and their surroundings (**RO2**). We apply it to selected systems of cities (Italy, South Asia and Britain from 1700 to modern day), correlating the changes in entropy to historical events and technological innovations (**RO3**).
- C3** We identify an inhomogeneous Matérn cluster point process that describes the locations of new towns emerged in England and Wales between 1801 to 1881, a period of fast urbanisation (**RO1**). The model fits the observed distribution according to major summary characteristics tests, except at the scale of approximately 800 meters, for which it fails to capture the slight repulsion in the pattern.
- C4** We introduce the suitably normalised first degree-based entropy of disc graphs (the spatial graphs where two vertices are connected if and only if their distance is smaller than a certain threshold) as a tool for the analysis of the spatial organisation of systems of cities (**RO2**). Through the example of the English and Welsh systems of towns in the 19th century and its expanding railway system, we show how the method can adopt both geographic distance and travel time. We remark the differences between these two cases and show the correlation between innovation in transport technology and the change in this entropy measure (**RO3**).
- C5** We propose a framework to compute the morphological polycentricity, based on Wilson’s entropy-maximising spatial interaction model, the cities’ populations and intercity travel time. The method expands existing approaches to morphological polycentricity based on spatial interaction models, as it explicitly accounts for changes in travel time as the transport infrastructure evolves, and makes minimal assumptions on the functional form of the most likely intercity interactions (**RO2**, **RO5**). We illustrate the framework with the aforementioned England and Wales system of towns and railways in the 19th century.
- C6** We identify the best fitting homogeneous point processes to describe the

locations of geo-tagged Tweets in an area of central London (**RO6**) at certain times of the day.

C7 We reveal the temporal and spatial properties of the network of knowledge exchange in the Middle Ages between Occitan composers and between Occitan region, based on evidence of literary imitation (**RO6**).

1.4 Thesis outline

This thesis is structured as follows.

In **Chapter 2**, we provide the theoretical background of the thesis. We begin by introducing operational definitions of “city” and “system of cities”, suitable in a variety of geographical and historical contexts (Section 2.1). We then present some of the databases of historical urban data that we use throughout our work (Section 2.2). We review the literature on city population size and spatial organisation of systems of cities, and how these morphological aspects can be combined to estimate intercity flows and other functional interactions using spatial interaction models (Section 2.3). We formally define entropy, discuss its role in urban science, and explain why we adopt it as the main analytical tool in our research (Section 2.4). Finally, we lay out the mathematical framework of the thesis, which includes random point process statistics, graph entropy, and spatial and temporal networks, with particular emphasis on their applications in the analysis of systems of cities (Section 2.5).

In **Chapter 3**, we discuss the normalisation of the first degree-based entropy of graphs, based on its extremal values for graphs subject to size and/or order constraints. More specifically, we determine the minimum entropy of a graph with given size (Section 3.2, based on [48]) and with given order and size for small sizes (Section 3.3, based on [47]), and provide numerical results that could support the solution of the problem when both order and size are fixed and arbitrary.

In **Chapter 4**, we focus on the population distribution and spatial organisation of systems of cities. As preliminary work, we clarify and schematise the relations between different data sets of historical city sizes and population density, and assess their compatibility (Section 4.1). We compute the entropy of city sizes in Late Medieval and Early-Modern Europe, and in 19th century England and Wales, and correlate it with main historical events (Section 4.2). We define local spatial entropy as a measure of heterogeneity of raster data, and apply it to the analysis of the population density and land

use in and around cities in the British Isles, South Asia, and Italy from 1700 to modern day, again correlating the variation of the measure to main historical events (Section 4.3, based on [151]). Finally, we model the location of English and Welsh towns from 1801 to 1881 using point processes, and measure the changing heterogeneity of their spatial distribution at different scales using the normalised first degree-based entropy of the disc graphs (Section 4.4).

In **Chapter 5**, we study the connectivity between cities, based on transport infrastructure. First, we study the heterogeneity of the worldwide flight network using the normalised first degree-based entropy (Section 5.1). Then we focus on the railway network in England and Wales and its evolution from 1851 to 1881 (Section 5.2). We expand our discussion on the heterogeneity of the spatial distribution of towns by using travel time along the network instead of geographical distance (Section 5.2.3). We define a measure of morphological polycentricity by combining travel time and population using an entropy-maximising spatial interaction model (Section 5.2.4, based on [150]).

In **Chapter 6**, we apply our entropy framework to study examples of human activities in cities and systems of cities. We study the spatial and temporal distribution of geo-tagged Tweets in London using point processes and their entropy (Section 6.1, based on [152]). We briefly discuss early results and some difficulties of modelling the impact of physical infrastructure on the evolution of social connectivity in an online social network (Section 6.2). Finally, we study an example of network of cultural exchanges in the Middle Ages, discussing how to deal with temporal uncertainty in the historical data set of literary imitations on which the network is built (Section 6.3, partially based on [159]).

In **Chapter 7**, we close the thesis with a discussion on the limitations of our approach and how to address them in future work.

Chapter 2

Background: Structure, Entropy and Evolution

2.1 What is a city, what is a system?

In a study aimed at tracing the evolution of systems of cities, it is necessary to adopt definitions of “city” and “system of cities” that are sufficiently general to be applicable to potentially very diverse forms of settlements in different places and moments in time, and simple enough to be of practical usage for a quantitative approach based on often scarce historical data.

2.1.1 An ideal definition of city

The problem of finding a sufficiently encompassing definition of city, valid in different places and historical eras is, in fact, a long-standing one. Already at the beginning of the 20th century, Maunier [147] discussed how the notion of city has evolved in time and pointed out the limitations of adopting any universal criterion to define a city, may this be morphological, demographic, juridical, functional, or even a combination of all these. The author proposed to define a city as a relatively dense community having numerous distinct social groups. Relative population density and social complexity are among the few properties of cities that have been found to be consistent since the independently emerged Indus Valley, Egypt, Mesopotamia, and Maya civilisations according to Childe [59]. With the rise of social network studies, social interactions have in some cases taken precedence over other aspects, as cities are studied fundamentally as social networks, embedded in the built environment [32, 105]. But other characteristics, such as the concentration of administrative, reli-

gious, and artistic functions, and the engagement in exchanges with other settlements to import resources not available locally, remain fundamental.

The profundity of these studies lies first of all on their stressing the relativity of the concept of city and its dependence on context, and secondly on highlighting that cities may not exist in isolation, but only as an integral part of a more extensive settlement system that includes their local environment and the other centres with which they are connected, via physical and abstract links, as well as via individual human ties. Over the course of this thesis, we will explore each of these aspects.

2.1.2 An operational definition of city

Ideally, the definition of city adopted in academic studies should include all the above criteria. In practice, though, this is severely limited by data availability. Most official data bases of cities around the world adopt simpler, operational definitions. Some countries, such as Brazil, require a city to be the seat of local administrative power; others, such as Nicaragua, demand it to possess certain kind of urban infrastructure, and others call for the majority of its population to be employed in non-agricultural activities¹. Population thresholds remain by far the most commonly adopted principle to officially define a city. Even these, though, may be widely different across countries: if in both Japan and the Korea 50,000 inhabitants or more are needed for a settlement to qualify as a city, just 200 people are sufficient in Norway or Iceland.

As a matter of convenience, studies of worldwide urbanisation have often adopted a common and arbitrary population threshold to identify cities. [75], for example, used a threshold of 20,000 inhabitants to distinguish a city from a rural settlement, and thus compare global rates of urbanisation. Similarly, [175] used a database of all urban areas in the world having 10,000 inhabitants or more, to study the evolution of the rank-size distribution in various world countries across the years 1950-1990. [41] argued against adopting a universal threshold and proposed to adapt it to each country and moment in time. For example, for European countries in the second half of the 20th century, they argue that 10,000 people is a sensible threshold beyond which one can expect settlements to exhibit the functional and social features that characterise cities. On the other hand, [15] who collected population of European cities from the 9th to 19th century, used 5,000 people as a threshold but argued that, ideally, 2,000 would be a more appropriate threshold for the period under

¹See Table 6 in [204] for details.

consideration.

In recent years, the creation of high resolution spatial data sets has allowed the emergence of alternative, density-based approaches to define cities. [5], for instance, proposed to identify cities via percolation in high-resolution street networks data. In [187, 188], grid-based population density data was used to identify urban areas by merging together neighbouring cells whose density is strictly positive, and selecting those clusters whose total population (given by the sum of the populations of the cells which compose them) is larger than a certain threshold. The articles argued that the proposed raster-based approach captures the presence of new forms of urbanisation, such as ‘conurbations’, ‘metropolitan areas’, and ‘megalopolises’ that may escape the simple rural/urban dichotomy entailed by threshold methods [54], and offers a universally applicable method to construct cities rather than depending on arbitrary administrative definitions. This approach is certainly more satisfactory than a threshold, but available raster data sets of past population density are based on city population databases, which are themselves assembled using thresholds (See Section 2.2.1.2 and Fig. 4.1), which nullifies the effort.

All the above approaches focused exclusively on the morphological aspect of a territory to identify a city, neglecting the social and functional aspects that were integral in all the “ideal” definitions given above. Other studies have proposed definitions of city based on human processes, such as [4], who used commuting patterns to define cities via percolation. The reconstruction of functional and human relations in the past remains, though, a complex and time-consuming process, as will be discussed in Section 2.2.2.2 and Chapter 6. Furthermore, some functional aspects, such as evidence of sufficient social complexity, cannot be obtained directly, without recurring, as in [41] to an educated guess based on city size. This is often the case in historical studies, where one must deduce whether a city engaged in the secondary and tertiary activities that would let it qualify as such exclusively from its morphological aspects, such as evidence of monumentality, or, indeed, its estimated size [106].

In this thesis, while acknowledging the limitations of a population threshold-based approach, we recognise its inevitability. To define cities, we rely on databases that adopt threshold-based definitions, directly or indirectly, as detailed in Section 2.2.1.

2.1.3 Defining a System

Childe’s definition reminds us that an essential property of a city is that of being part of a more extensive system of settlements, without which it would not be possible for it to exist. In the most general sense, a system of cities is a set of cities subject to some constraint (geographical, political, economical, cultural or of other kind) and linked by strong interactions, so that the trajectories of their evolution are strongly interdependent [40].

Even though it may be argued that a handful of global cities are part of a highly integrated system of their own that surpasses geographical boundaries [190], the vast majority of studies on systems of cities assumes the cities to be located in the same region or nation [171]. The boundaries are often taken to be those of contemporary political entities, even in longitudinal studies (see, for example, [173, 15]). This approach is not completely satisfactory given that political, economical, and cultural regions may be subject to major changes in a long time frame, but it is in part justified by the convenience of dealing with fixed, familiar boundaries. In this thesis we will sometimes adhere to this simplification, though we will look explicitly at non-modern geographic and cultural regions, when appropriate².

There are several reasons to adopt spatial proximity as a fundamental criterion for cities to form a system. From a practical perspective, focusing on sets of cities within a well-defined, bounded geographical area helps to ensure that the data about the individual cities is consistent, as historical data sets usually focus on a specific region and period of time and combining different data sets raises issues of compatibility. Furthermore, if the cities under consideration are located in a geographically well-defined area, it becomes possible to apply powerful analytical tools, such as point process and spatial interaction models. If the area is relatively homogeneous, geographically or in terms of available technologies, a condition better approximated at the regional level, the application of such tools is greatly simplified. More conceptually, in spite of the promulgated “Death of Distance” [46] in the contemporary information age, geographical distance has been proved to be a strong influencing factor in many urban phenomena of interest, including the most abstract of intercity

²In Section 4.3 we look at the British Islands (including both Great Britain and Ireland) and South Asia (from the Indus River to the West to the Bay of Bengal to the East) in our study of the evolution of local entropy from 1700 to modern day, because the national boundaries evolved dramatically over the course of this period. Similarly, in Section 6.3, we study literary influences between medieval composers from different areas of the historical cultural region of Occitania, which includes most of modern day Southern France, as well as modern Monaco, and portions of Spain and Northern Italy.

interactions, such as knowledge flows [63].

2.2 Historical Urban Data

The study of the historical evolution of systems of cities is affected by the availability of data, which may be limited in certain regions and time periods. Here, we introduce the population data sets used throughout the thesis and describe how they were assembled combining historical population studies and estimation procedures, before embarking on a more detailed analysis of their relations, compatibility, and completion in Sections 4.1.2 and 4.1.1. We also discuss proxy data as evidence of human activity and social interactions, both in the past and in the present, briefly touching on the opportunities and challenges afforded by modern online social network data.

2.2.1 Data sources of historical population

A great deal of effort has been directed by the scientific community towards assembling data sets suitable for a longitudinal study of the distribution of human population. We distinguish between two categories of these: population estimates for individual cities and raster data sets. The former ones conceptualise cities as discrete entities, often defined via (more or less arbitrary) population thresholds and boundaries. The latter ones attempt to represent more realistically the continuous distribution of human population, including cities as well as sparser settlements, and provide information on the spatial extension of cities. However, this comes at the cost of relying on more speculative and assumption-heavy estimation techniques.

2.2.1.1 City population estimates

Censuses arguably constitute the most reliable sources of historical population estimates for cities. Bennett [27] assembled Census data for 934 English and Welsh towns, covering the period from 1801 to 1911 at 10 years intervals. The data set provides the population of all settlements with at least 2500 inhabitants, a population density of 1 inhabitant per acre or more³, and a spatially contiguous built-up area, so to distinguish towns from rural communities.

To study urban populations in historical periods and regions of the world in which censuses were or are not widespread, other estimates are needed. For

³This is equivalent to approximately 247.1 inhabitants per square kilometer.

the relatively recent period from 1950 to 2018, the United Nations World Urbanisation Prospects [206] provides the population of all urban agglomerations that had 300,000 inhabitants or more in 2018, for a total of 1,860 cities.

For more remote times, it is often regional rather than worldwide studies that offer the most complete perspective. The aforementioned Bairoch [15], for instance, compiled a data set of population estimates of European cities from the year 800 to 1850, including all centres that achieved a population of 5,000 or more at some point during that period. Hanson [107, 106] assembled a comprehensive atlas of cities in the Roman Empire between 100 B.C. and 300 A.D., with their respective area, which can be used to derive population estimates by making reasonable assumptions on the population density.

The Historical Urban Population data set [180] by the National Aeronautics and Space Administration Socioeconomic Data and Application Center (NASA SEDAC) is one of the most complete collections of historical population of cities around the world, covering the period from 3700BC to 2000AD for a total of 1595 cities. It is built upon two major studies of historical urban populations [179]: “Four Thousand Years of Urban Growth: A Historical Census” [55], by historian Tertius Chandler and “World Cities: -3,000 to 2,000” [161], by political scientist George Modelski.

Chandler used an adaptive definition of city, depending on both the year and the geographical location. For the period from 800AD to 1850AD, the author defines a city as a settlement with more than 20,000 inhabitants, except for Asia, where he applied a larger threshold of 40,000 inhabitants. Starting from 1850AD, suburbs are included in the computation of populations and the threshold of 40,000 inhabitants is applied to all locations. From 1962, suburbs outside the municipal area are also counted as part of the urban population.

Modelski focused explicitly on areas underrepresented in Chandler’s work. Three thresholds are applied, depending on the year but independent of the location: in Ancient times (3500 BC to 1000 BC): a city is defined as a settlement with at least 10,000 inhabitants; in the classical period (1000 BC to AD 1000) the threshold is raised to 100,000 inhabitants, and in Modern times (AD 1000 onward) only settlements with 1,000,000 inhabitants or more are included.

In spite of its completeness, [180] presents a few weaknesses. The fact that the data set issued from assembling two different historical studies raises some compatibility issues. First of all, Modelski’s population estimates for the year 2000AD are based on the United Nations’s 2000 Demographic Yearbook [205]

data for ‘cities as proper—governed as one unit’, and thus do not correspond to Chandler’s definition of cities, which include the urban agglomeration. This suggests that population estimates for different cities in recent years in [180] may not be comparable. Furthermore, Chandler’s estimates had been the subject of criticism by Bairoch [15], who claimed that, taking into consideration the land type within city walls, Chandler’s estimate of the populations of European cities should be increased by 15% and those of Latin American cities by up to 50%. Finally, Modelski selectively applied rank-order principles, i.e., Zipf’s law (Section 2.3), to reconstruct the population of individual cities. This implies that any conclusion on general statistical properties of cities in [180] may be inherently biased towards displaying this regularity.

In Section 4.1.2 we discuss how to complete missing records and spatialise some of the data to create data sets suitable for our subsequent analyses in Sections 4.2, 4.4, and 5.2.

2.2.1.2 Raster data

We have seen that conceptualising the cities of a region as well-defined, separate entities poses a number of issues, including the adoption of arbitrary thresholds, and the risk of neglecting sparser settlements forms and the characteristics of the landscape, that have an integral role in the creation and maintenance of the urban system. A different approach is afforded by raster data, that subdivides a region in a grid of small, regular cells and provides for each of these the value of one or more variables of interest associated to its location.

In recent years, several raster data sets have been developed to study population density, in part thanks to the advancement of satellite imagery and computer vision techniques. For instance, [61], [186], and [83] all provide very high resolution population density grids. Unfortunately, the latter two do not provide historical estimates, and advise against using their data for year-on-year comparisons, as changes may reflect advancements in collection and classification techniques rather than actual patterns. The former one only provides historical data from 1970, making it unsuitable for our interest in a long term analysis.

A rare example of raster data set designed for longitudinal studies is HYDE 3.2 [127]. HYDE 3.2 provides worldwide land use and population density estimates from 10,000BCE to 2015CE, with more frequent and accurate data in more recent years. It allows to distinguish between several categories and

subcategories of land use, including the following seven: urban areas, dense settlements, villages, cropland, rangeland, semi-natural, and wild land; cells corresponding to seas and oceans are unclassified. Estimates of the total population counts and density for each cell are also given.

The resolution of the longitude/latitude grid is relatively coarse: $5' \times 5'$ corresponding to a square of side 9.26km on the equator or roughly 6km at the latitude of the United Kingdom. This limits its scope to the study of regions or cities with respect to their surroundings, not allowing for the detection spatial patterns within individual cities, except very large ones.

HYDE 3.2 is assembled via an elaborate procedure, detailed in [127] and [128]. Historical population numbers are taken from [153], [11] and [142], supplemented with the sub-national population numbers of Populstat [130], among others. Spatial patterns are obtained by using population density map patterns for current time periods from [164], and gradually replacing them with weighted maps based on proxies such as distance to water and soil suitability when going back in time. Historical urban densities are taken from [15, 70].

It is important to notice that the data set assumes a functional relation between urban population densities and urban land area. Following the observation that the urban density in Europe and in the United States seems to increase rapidly to very high densities (up to 40,000 inhabitants/km²) and then slowly decrease over time, the data set assumes that the density follows an asymmetric bell-shaped curve and that this applies to all countries of the world. This assumption may not be accurate for individual cities. Despite these limitations, we deem the data set suitable for an analysis that focuses on larger areas and averaging procedures, as the one we carry out in Section 4.3.

2.2.2 Estimating human activity and social connectivity, in the present and in the past.

Resident population data only provides a partial perspective on the way we humans inhabit cities. Other forms of data are necessary to capture the daily fluctuations in the spatial distribution of people as we move around to work, shop, and engage in leisure activities, and, more generally, the relations that we interweave with the city and with one another.

2.2.2.1 The opportunities and challenges of social media data

In a modern, highly digitised world, people leave digital traces of their passage, which are increasingly collected and analysed, often with an eye towards

continuously improving service provision. This data is precious to understand how humans behave, move, and interact.

Twitter data has seen a wide variety of applications to study human behaviour in urban areas, including understanding citizen happiness and well-being [99, 201], crowds [38], urban crime [89], and riots [172]. Tweets, like many other social media data, often contain meta-information, one of which is the location. In Chapter 6.1, we evaluate point process models for a data set of more than 430,000 geo-tagged Tweets posted in a 40km radius disc centred in Trafalgar Square, London, between Sunday 10 June 2012 and Sunday 24 June 2012 [222]. The data was purchased from Twitter and represents the most comprehensive geo-tagged data set of Tweets for that time period and was used, for instance, in [223]. Nonetheless, it should be stressed that the Tweets' spatial distribution may suffer from issues of representativeness, as only 1-2% of Tweets are GPS pin pointed and users from certain socio-economic backgrounds are more involved in the generation of geo-referenced tweets than others [133]. We take this as an unavoidable limitation of the data set.

Citizen's digital data can also be used to study the connectivity between individuals, cities and wider regions. Corporate credit card expense data has been used to assess the economic benefits of the flow of knowledge generated by international business travel [63]. Mobility data from Facebook's Data for Good initiative has been used to track the change in population distribution in a system of settlements as a consequence of disruptive events. For instance, [74] showed that during the first wave of the COVID-19 pandemic, India witnessed a sharp decrease in Facebook users in urban areas (4-11%), paralleled by an increase in their number in rural areas (7%), mostly attributed to urban migrant workers returning to their places of origin, which were often smaller, remote localities. The number of Facebook users was also used to study the change in population density in the UK from March 2020 to September of the same year, evidencing that larger cities (and London in particular) saw a strong decline in users, as opposed to smaller towns, especially along the coast, that witnessed an increase across the period [203]. On the other hand, Facebook's mobility data in Italy showed that people did not move considerably within the country during the first lockdown. Peripheral rural areas saw only marginal increase in the total number of users, whilst most users who appeared to have moved out of large cities, moved to mid-sized towns in their vicinity and urban belts, reflecting a halt in the established commuting pat-

terns rather than a rural exodus [28]. These studies are helpful in suggesting a large displacement of people towards rural and peripheral areas during the lockdown, but are careful in warning that Facebook users may not be representative of the population as a whole, and that not every user shares their location information.

An example of online social network that required its users to provide spatial information about themselves is the Hungarian social network iWiW, shorthand for “International Who is Who”. Launched in 2002, iWiW enjoyed remarkable popularity, with approximately 40% of the Hungarian population joining, but eventually it succumbed to the competition of other social networks, noticeably Facebook, and closed down in 2014 [132]. In iWiW, every user had to confirm their city of residence. In Section 6.2, we show how to adapt an entropy-based model for the evolving connectivity among Hungary’s eighteen regional capitals as well as the national capital Budapest, based on monthly counts of registrations to the platform. By discussing in greater detail the peculiarities of the iWiW, we show the limitations of using online social network data for identifying long term trends in real social connectivity, even in the case of a platform that seems particularly well-suited to study of the spatial patterns of its users and their connections.

2.2.2.2 Finding sources of human interactions in the past

The reconstruction of human interactions in the past is a key object of study in history, sociology, and archaeology.

The sources to reconstruct historical human interactions between different cities and regions are numerous and highly heterogeneous. For instance, [170] estimated the historical migration flows between regions in the United Kingdom in the last three centuries from a combination of census birthplace data, Poor Law certificates, apprenticeship registers, and written and oral personal accounts. Mail exchanges, on the other hand, were used in [96] to estimate the flow of information between British regions, and revealed much greater integration between British cities and London than one would expect from distance and cost alone. For more remote times, other forms of material culture, such as pottery, can be used to reconstruct the trade and cultural connections between settlements and their evolution in time [185, 112].

Making use of these historical data sources to study human interactions implies having to manage some of their inevitable limitations. In most cases, the sources were not collected explicitly for the purpose of studying interactions,

and thus may not contain all the relevant information, and have inconsistent resolution and format. For instance, population movements between two areas may have been recorded only as net demographic effects, underestimating gross migration flows [170]. The records may not carry precise temporal information, which calls for methods to assess robustness of the results [80]. Furthermore, they may be incomplete.

The application of link prediction algorithms has been proposed for the reconstruction of missing connections in data sets of historical interactions, when these are represented as links in mathematical networks (see Section 2.5). In [202], the authors consider a representative sub-graph of a trade network, and evaluate the accuracy of introducing a new link according to different criteria exclusively based on the network structure.

Network analytical tools, more generally, have played a growing role in the study of human connections for a number of essential questions, including their temporal evolution and the relation between connectivity and physical distance, migration and mobility [212]. Networks based on spatial interaction models, for instance, have proven helpful to reconstruct potential human connections from incomplete historical sources by assuming that distance acts as friction, limiting interactions [129] (see Section 2.3.2.4). Yet, a number of other factors influence material exchanges that mean that the patterns observed in the data do not always follow spatially-informed constraint, including conscious bias toward local products, conformism, appreciation of rarity, competition, etc. [185].

In Section 6.3, we reconstruct an evolving network of cultural exchanges in the historical region of Occitania during the 11th to 13th century, based on musical imitation and collaboration between authors, as evidenced by music manuscripts and biographical information. We assess the impact of physical distance on the exchanges, and discuss how to manage the high levels of temporal uncertainty in the data.

2.3 Structure and Evolution of systems of cities

2.3.1 In search of universal rules

Despite having extremely diverse forms and histories, regional and national systems of cities have been argued to present remarkable statistical regularities in their spatial organisation, distribution of city sizes, and trajectories of development. Several theories describing common structural properties of

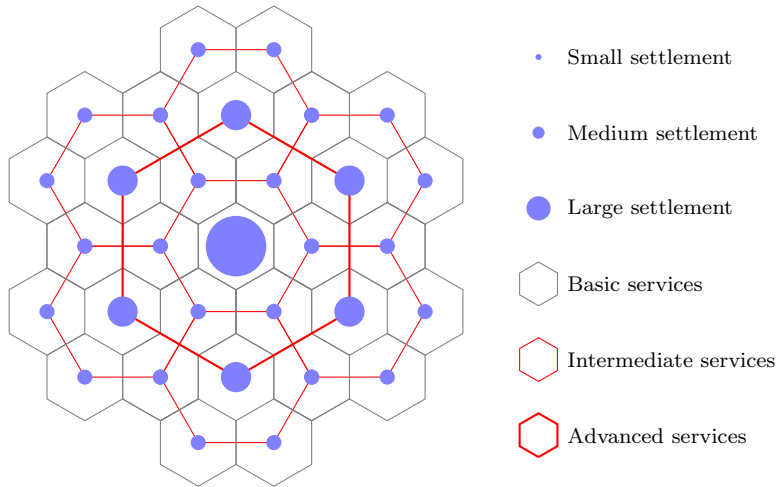


Figure 2.1: Example of hierarchical structure of settlements according to Central Place Theory, based on market principles. Each town is surrounded by nested hexagonal regions, proportional in number and area to the town's size, in which the town acts as provider of services. The regions never overlap at the same level, but they do overlap at different levels, as larger towns offer more specialised services to a larger area. This pattern repeats itself in an infinite uniform surface, so that every point is provided with all types of services.

systems of cities have been proposed in the first half of the 20th century.

Introduced in 1933 by Christaller [60], Central Place Theory asserts that in an ideal infinite, homogeneous, and uniformly densely populated surface, a system of settlements would be arranged in an ordered spatial hierarchical structure dictated by the logic of universal and efficient marketing, transportation, or administration. While having slightly different formulations according to the kind of service provided, the theory predicts in all cases that settlements would be placed along a regular hexagonal lattice (Fig. 2.1). All settlements would provide basic services to a small surrounding hexagonal area, and a few larger centres would provide more specialised services to a larger area, covering several towns. The larger towns would thus be distributed in a spaced-out fashion along the lattice. This structure would be repeated at multiple levels, with fewer and fewer larger cities at each level providing more and more specialised services.

A territory is rarely uniform enough for the conditions of Central Place Theory to hold and indeed the spatial distribution of cities in real-world systems is in general much less regular. When one allows for fluctuations, Okabe and Sadahiro [165] proved that some fundamental properties of Christaller's spatial organisation also hold for a completely random distribution of cities

in a flat surface. For instance, if in this hypothetical random system we look at a city's Voronoi cell (i.e., the points of the surface that are closer to it than to any other city), then the probability that the cell has n sides peaks when $n = 6$ [111] and its average is $\langle n \rangle = 6$ [37, Chapter 5], which means that the hexagonality of service areas arises frequently from the simple fact of operating in a two dimensional surface, without the need of assuming market logic or a rigid hierarchical organisation. Assessing the null hypothesis of complete spatial randomness has long been an essential step in studying the spatial pattern of cities in real world scenarios, as we will discuss more extensively in Section 2.5.1.4. Yet, Central Place Theory remains influential in its fundamental concepts, and indeed contemporary studies have approached the analysis of the spatial organisation of systems of cities via a hierarchical approach based on their size, Voronoi cells, and those of their neighbours [139].

In light of its enduring legacy, another theory on systems of cities that deserves a special mention is Zipf's law [227], which suggests that a city's population size is inversely proportional to a power of its position in its national population ranking. This rank-size relation can be expressed with the formula

$$R_u = \alpha P_u^{-\beta}, \quad (2.1)$$

where P_u and R_u are the population and rank of a city u , and α and β are constants. What is remarkable is not only the apparent ubiquity of this distribution, but also its exceptional stability [174], which persists in spite of the fact that individual cities may see major changes in their ranking [22]. For this reason, Zipf's law has been used to estimate historical populations of cities when data was scarce [161].

A great deal of effort has been directed to finding a mathematical explanation for Zipf's law [17], from the null hypothesis that cities' populations grow with independent, identically distributed rates [92], to more complex, but not necessarily more realistic models of cities populations as random walks constrained to remain positive [87], or driven by random growth and identical intercity migration rates [39]. An equal amount of effort has been dedicated to disproving it statistically [187], evidencing its dependence on the definition of city one adopts [6], and dismissing it as a mere statistical consequence of the fact that the rank R_u of a city is derived from its population P_u , instead of a meaningful observation on the distribution of the city sizes [88]. Alternative rank-size relations have also been proposed [40], as well as less rigid, dynamic models, that incorporate the effects of rare population shocks [209].

2.3.2 From a static to a dynamic approach

Both Central Place Theory and Zipf's law present a static view of systems of cities, that assumes that a system is in equilibrium, and ignores the effects of evolving intercity interactions.

Alternative theories on the spatial distribution of settlements addressed more explicitly this concern. For instance, Hudson [116] proposed that the location of rural settlements is the result of a multi-stage process, whereby new centres are founded in a previously uninhabited territory, which is then densified through short distance dispersal of the centres; this exacerbates local competition pressures, eventually producing the empirically observed evenly spaced patterns.

In other instances, ideas from these laws were adapted to assess the validity of dynamic models of cities' formation and development [21, 33], and to theorise and analyse trends in the evolution of systems of cities.

2.3.2.1 Evolution of Centralisation

The exponent in Zipf's law was used to perform cross-time and cross-regional comparisons of systems of cities and describe trends in their level of centralisation, i.e., the economic and demographic imbalance between one or few major centres and the smaller centres in the system. The cities within a region (or a subset of them) are ranked from the largest to the smallest in terms of population size and a line is fitted to the graph of the logarithm of their population vs the logarithm of their rank. The exponent in Zipf's law is the slope of this line: the steeper the line, the larger the unbalance between the population sizes, and thus the more centralised the region is taken to be.

Pumain and Moriconi-Ebrard [175], for example, remarked that, following a period of increasing centralisation during the 19th and early 20th century described in [15], a clear evolutionary pattern in national systems of cities did not emerge in more recent years from the analysis of the exponents. Hence, the authors advocated for more sophisticated methods that do not only take into account city sizes, but also their territorial contexts and spatial locations to quantify how centralised a system of cities is.

2.3.2.2 Polycentricity

Intuitively antithetical to centralisation, *polycentricity* has gained considerable traction in planning but has remained for a long time an ambiguous concept,

open to multiple interpretations [157]. This vagueness largely derives from the fact that polycentricity has been used to refer to two different aspects of a system of cities: the morphological one, relating to the size and spatial distribution of the cities (see, e.g., [158]), and the functional one, concerning the linkages and relations between them [78, 95].

In this thesis, we are careful about this distinction and say that a system of cities is *morphologically polycentric* when its cities are similarly sized and evenly spaced across its territory [157, 117]. The distance between two cities may be computed as their geographical distance or, more aptly, as the travel time or some other travel cost function [78]. Similarly, we say that a system is *functionally polycentric* when the functional connections between the cities are more balanced [95]. Examples of functional connections include migration flows, commuting patterns, material and knowledge exchanges, among others.

Another source of ambiguity is the fact that polycentricity has been varyingly treated an observable geographical phenomenon or as a policy objective to pursue [95].

Hall and Pain [102] described polycentricity as an emergent and ongoing process, discussing the advent of what they called “Mega-city regions”: polycentric regions clustered around one or more major cities and developed through a process of decentralisation, whereby the cities within the region take advantage of the division of labour and operate synergistically.

When polycentricity is interpreted as a policy objective, the primary aim is often that of analysing the connection between morphological and/or functional polycentricity and economic prosperity. In some cases, the seemingly neutral notion of polycentricity is inflected with the positive connotation of “desirable balance”. The European Spatial Development Perspective [82], for example, maintained that the disproportionate economic and demographic concentration in the core area of the EU (roughly defined as the pentagon with London, Paris, Milan, Munich and Hamburg at its vertices) hinders the full realisation of the economic potential of the peripheral regions, and stressed the importance of multiplying hubs for accelerating growth, reduce regional imbalance and ensure global competitiveness. In other cases, though, the link between polycentricity and economic performance was called into question when regions in the same or different countries were compared. For example, Seymour [192] remarked the different levels of economic success in similarly functionally polycentric regions, such as the Dutch Randstad and German Ruhr, and in the morphologically polycentric North of England compared to

the much more morphologically centralised South East.

2.3.2.3 Measuring Polycentricity

Whether polycentricity is treated as an emerging geographical phenomenon or as a policy objective, a fundamental problem is that of quantifying it. Most approaches to measuring morphological polycentricity consider cities in a region as existing “in isolation”, and focus on one or more aspects of interest, such as their population size, for which several heterogeneity indices can then be used, including the aforementioned exponent in the rank-size relation.

Just as Pumain and Moriconi-Ebrard [175] used it as an index of centralisation for nation-wide systems of cities, the exponent in the rank-size analysis has been widely used as a simple measure of polycentricity at the regional scale [158, 45, 42]. This method is also mentioned in Hall and Pain [102], but the authors stressed therein its crudeness. Major limitations of this measure include its sensitivity to sample size (different slopes may result if a larger or smaller subset of cities in a region is selected), the fact that a line may approximate only very crudely the rank-size relation, and the measure’s aforementioned inability to capture spatial information about the cities’ locations or distance between the cities.

Composite indices are adopted as more specialised and sophisticated measures for polycentricity, their main advantage being that of merging together information about some attributes of interest, such as employment or population, and the spatial location of the cities. Examples include the polycentricity index proposed by ESPON 1.1.1 – Potentials for Polycentric Development in Europe [78], which uses the slope in the rank-size relation for cities’ population and Gross Domestic Product, the Gini coefficient of the sizes of the cities’ service areas, and their multimodal accessibility, and the simpler Urban Centrality in [168], which combines inequity in the distribution of jobs with an index of the spatial dispersion of centres based on their geographical distance. Disadvantages of the polycentricity index are its reliance on the rank-size relation and its complexity, as it involves combining several pieces of information. The Urban Centrality, on the other hand, uses geographical distance to compute heterogeneity in the spatial arrangement, which may not reflect real travel time conditions.

The study of functional polycentricity presupposes instead a network approach [53], in which cities are represented as nodes connected by links that correspond to some functional linkage, such as commuting patterns, trade,

letter exchanges, etc.. In Section 2.5.3, after we formally introduce networks, we discuss network-based measures of polycentricity.

Real world observations suggest that a morphologically polycentric region is not necessarily a functionally polycentric one [192]. Yet, morphological and functional polycentricity are not independent, as the spatial distribution of cities and their sizes impacts the infrastructural linkages between them, which are indispensable for the creation of functional linkages [156, 45].

2.3.2.4 Spatial Interaction Models

Spatial Interaction Models (SIMs) attempt to estimate the intensity of functional connections and the resulting flows of goods and people using exclusively morphological properties of the system. The common assumption of all SIMs is that the intensity T_{uv} of the flows between two locations u and v is a function of some relevant socioeconomic factors P_u and P_v of the locations, such as their population or number of jobs, and of the friction due to their distance d_{uv} , which may mean geographical distance, but also travel time or cost [184].

A plethora of different models exist. The *gravity models* are a family of spatial interaction models whose most general formulation [215] posits that the interaction T_{uv} between two locations u and v is given by

$$T_{uv} = K_u K_v P_u P_v f(d_{uv}),$$

where K_u and K_v are constants that allow to accommodate additional constraints to ensure better fitting of the system, and f is a decreasing function of distance, frequently taken to be $f(d_{uv}) = d_{uv}^{-\phi}$, where ϕ is a parameter regulating the impact of distance. The models derive their name from their affinity with Newton's law of universal gravitation [19], that is retrieved when K_u and K_v don't depend on the locations u and v and $\phi = 2$. It is in this simplified formulation that they were originally introduced to model migration flows in the works of Ravenstein [177, 178], and they still appear occasionally in more modern works [122, 100].

Gravity models, in spite of their popularity, suffer from numerous disadvantages. An immediately apparent one is the fact that they do not allow for self-interactions (since, when $u = v$, the distance term at the denominator would be $d_{uv} = 0$), which may be important in modelling some phenomena such as spending or intercity social ties. Most importantly, they lack a rigorous mathematical derivation from first principles.

The entropy-maximisation model introduced by Wilson [215] overcomes both of these limitations. In Wilson’s model, the interactions are derived from constrained entropy-maximisation, that is mathematically equivalent to finding the most statistically likely or optimal interactions, given basic constraints on total intensity of interactions at each locality. Hence, it corresponds to adopting Occam’s razor principle, imposing minimal assumptions on the functional form of the interactions. This is crucial in the context of historical studies, where information on the interactions is often partial and, in fitting or reconstructing them, one wants to avoid the risk of imposing on them a specific functional form, even one is reasonable in the present, as it may not have been equally valid in the past [183]. The entropy maximisation model finds an explicitly dynamic formulation in [216, 217], where the morphological characteristics of the localities, such as their population, are repeatedly updated so the reduce the unbalance with the predicted, ideal flows. We discuss this model in greater detail in Section 2.4.2.1, after we formally introduce entropy, and adopt it throughout this thesis.

It is worth mentioning that other rigorously constructed SIMs do exist. [214] proposed the Two-Trip model, also based on entropy-maximisation principles, which accounts for the destination’s surroundings when predicting the interactions between two localities. The radiation model [195] is a SIM to predict intercity migrations founded on the principle of intervening opportunities. According to this principle, an individual residing in city u will move to a city v if no better opportunities are offered in u or in all other cities falling in the disc centred at u having as radius the distance d_{uv} . The crucial advantage of the radiation model is that it is parameter-free, which allows to avoid the careful calibration needed for a successful application of both the gravity and the entropy-maximisation model. Nonetheless, the accuracy of the model in its basic formulation has been subject to criticism [169], especially when confronted to more sophisticated versions of the above two models.

However careful the construction, it is important to remark that spatial interaction models such as the entropy maximisation model provide the optimal or most likely functional interactions, but the real-world interactions may not correspond to an optimal state, deterministically derived from morphological constraints. In predicting the interactions between Aegean Islands, [129] utilised a gradient descent model, that retrieves the most likely interactions based on a balance of cost and benefit, but is open to fluctuations and stochastic effects.

Human, political, and technological factors all play an important role in influencing interactions that is not easily captured by spatial interaction models. In their study of the evolution of social connectivity between settlements in South-West U.S. before European contact, Hill et al. [112] highlighted that people’s desire to maintain their connection even after displacement led to the continuation of highly impractical long-distance ties, which would defy optimisation principles. The necessity of guaranteeing high level of integration in a nation state manifests itself in much more intense than expected mail exchanges with London from all parts of Britain than what one would expect from proximity alone [96]. The observation that, at least for Europe, centres of political power tend to grow larger than other cities [15], mean that any SIM for intercity migration would fail to completely explain cities’ growth [40]. Finally, a city’s functional interactions and evolution are strongly linked to its functional specialisation, which depend on the diffusion of randomly emerged innovations [174], a complex phenomenon driven by the presence of suitable local conditions and by the active engagement of individuals, that no SIM could fully predict.

It is with full awareness of these limitations that we adopt SIMs in our analysis.

2.4 Entropy

The concept of entropy has long been central in urban studies, lending itself to multiple applications and interpretations [176]. In this thesis we focus on two historically distinct, but mathematically (nearly) equivalent notions of entropy, emerged from the fields of Information Theory and Statistical Mechanics, due respectively to Shannon, and Boltzmann and Gibbs.

2.4.1 Definition(s)

2.4.1.1 Entropy in Information Theory

Let $X : \Omega \rightarrow \mathbb{R}$ be a discrete random variable, where Ω is the sample space of a probability space with probability function P . Let $\mathbb{X} \subset \mathbb{R}$ be its image and $p : \mathbb{X} \rightarrow [0, 1]$ be its probability mass function, i.e., for all $x \in \mathbb{X}$,

$$p(x) = P(X = x) = P(\{\omega \in \Omega | X(\omega) = x\}).$$

Definition 1 (Shannon entropy [193]). *The Shannon entropy of X is*

$$H(X) = - \sum_{x \in \mathbb{X}} p(x) \log_2 p(x).$$

To keep the notation light, in the following, we will write \log for \log_2 . $H(X)$ can be thought of as a measure of heterogeneity of p [64]. If \mathbb{X} is a finite subset, and $n = |\mathbb{X}|$, then H attains its maximum value $H_{max} = \log(n)$ when all outcomes $x \in \mathbb{X}$ are equally likely, or, in other words, $p(x) = \frac{1}{n}$, for all $x \in \mathbb{X}$. It attains its minimum value $H_{min} = 0$ when one of the outcomes has probability one⁴.

With a slight abuse of notation, we will apply this definition to any finite vector $q = (q_1, \dots, q_n)$, denoting

$$H(q) := - \sum_{j=1}^n p_j \log p_j,$$

where $p_j = \frac{q_j}{\sum_{i=1}^n q_i}$. The interpretation remains similar: $H(q)$ is a measure of the balance of the entries of q , with larger values indicating a more homogeneous distribution and smaller values a more unequal one.

2.4.1.2 Entropy in Statistical Mechanics

In classical physics, to describe a system constituted by several particles, it is enough to describe all their individual positions and momenta. This is referred to as the system's *microstate*. When the number of particles is extremely large (think, for instance, of particles of a gas), this is impractical, and it is usually preferable to describe the system in terms of its macroscopic properties. This is called the system's *macrostate*.

Definition 2 (Boltzmann entropy). *The entropy of a system is*

$$S_B = k_B \ln W,$$

where \ln is the natural logarithm, $k_B = 1.380649 \times 10^{-23} \text{ J/K}$ is the Boltzmann constant, and W is the number of microstates in which the particles may be found given the system's macrostate.

Boltzmann proposed an explicit formula for W . Assuming that the system

⁴For an outcome x such that $p(x) = 0$, we take $0 \log 0 = 0$, in accord with the limit $\lim_{p \rightarrow 0^+} p \log p = 0$.

is formed by N indistinguishable particles, of which N_i are found in a certain state i , the number of microstates is

$$W = \frac{N!}{\prod_i N_i!},$$

where the denominator accounts for the possible permutations of particles in a given state.

Gibbs generalised Boltzmann formula, by allowing microstates to have different probabilities.

Definition 3 (Gibbs entropy). *If the microstate i of a system has probability p_i , then the entropy of the system is $S_G = -k_B \sum p_i \ln p_i$.*

Boltzmann's formulation of the entropy of a system may be retrieved from Gibbs's by assuming that all microstates are equally likely, i.e. $p_i = \frac{1}{W}$. This also corresponds to the maximum value Gibbs entropy can achieve.

Gibbs entropy and Shannon entropy are mathematically equivalent, up to a multiplicative constant. One can intuitively reconnect the Shannon entropy of the discrete random variable X to the number of configurations of individual particles in a microstates by thinking of the event $X = x$ more explicitly as of $\{\omega \in \Omega | X(\omega) = x\}$.

2.4.2 Entropy in Urban Science

The usage of entropy in Urban Science is well-established since its introduction in the pioneering works by Wilson [215] and Batty [19], who followed, respectively, a Boltzmann- and a Shannon-inspired approach, whilst both adopting the more general Shannon's formula.

2.4.2.1 Entropy Maximisation Models

Given a set V of locations, Wilson [215] applied a Boltzmann-inspired interpretation of entropy to find the most likely distribution of spending flows $\{w_{uv}\}_{(u,v) \in E}$ between the locations pairs $E = \{(u, v)\}_{u,v \in V}$. The author interpreted the distribution of flows as a macroscopic description of the system. Such flows could result from assigning to each pair of locations (u, v) , several different sets of individuals residing in u and spending in v . This assignment is interpreted as a microscopic state. Wilson argued that the most likely structure of the spending flows (macrostate of the system) is the one that is associated with the maximum number of possible assignments, where the

assignments are assumed to be equally likely. The maximisation procedure can be subject to additional constraints on the cost and benefit resulting from spending in a given location v by a person residing in u .

More precisely, for each u and v , Wilson considered the “cost” c_{uv} of spending in v for a person in u , which can be some measure of travel cost between locations u and v ; and the benefit b_v derived from spending in v , dependent on characteristics of the location v . He then let $w = \sum_{u,v} w_{uv}$ and assumed the following constraints:

- the total money $\sum_v w_{uv}$ spent by people living in u is known for all u ;
- the total cost $\sum_{uv} c_{uv} w_{uv}$ and the total benefit $\sum_{uv} b_v w_{uv}$ are also known.

Maximising $W(w_{uv}) = \frac{w!}{\prod_{u,v} w_{uv}!}$ under the above constraints yields the most likely configuration of the flows⁵. Using Lagrange multipliers, Wilson proved that the resulting most likely flows are given by

$$w_{uv} = \gamma \cdot p_u \frac{\exp(\alpha b_v - \beta c_{uv})}{\sum_x \exp(\alpha b_x - \beta c_{ux})}. \quad (2.2)$$

Here γ is a proportionality constant, p_u is the population size P_u , and the parameters α and β represent the relative weight of the cost and the benefit, which need to be calibrated to the specific system.

As anticipated in Section 2.3.2.4, Wilson’s model can be thought of as the spatial interaction model with minimal assumptions on the functional form of the interactions, since this is derived exclusively as the most statistically likely or optimal form with respect to cost and benefit considerations. We will use this spatial interaction model in Sections 5.2, 6.2, and 6.3.4.

A dynamic formulation of the model is possible [216], with all the quantities and parameters taken to be dependent on time t . Assuming that the system evolves towards this ideal state, the characteristics of the locations at time $t+1$ are adjusted according to the imbalance between the observed flows and the predicted optimal ones at time t . This dynamic interpretation of the entropy-maximising spatial interaction model was applied to intercity migration by Wilson and Dearden [217], who studied the evolution of population in the U.S. Midwest from 1790 to 1870 by postulating that migration flows would

⁵In fact, maximising $W(w_{uv})$ is equivalent to maximising $\log(W(w_{uv}))$, which, by Stirling’s approximation of the factorial of a logarithm, can be proved to be equivalent to maximising the Shannon entropy of the flow distribution $-\sum_{uv} w_{uv} \log w_{uv}$, which is how Wilson proceeded.

follow the most likely spending flows. The results showed partial alignment with the evolution of the real system.

2.4.2.2 Spatial Entropy

Batty's work, on the other hand, adopted a Shannon-inspired approach. In [19] and [20], the author clarified how to adapt the discrete formula of Shannon entropy to continuous spatial data, such as population density, by subdividing the study area into zones, treating the zone-level populations as a finite distribution, and computing its entropy. The work carefully discusses the effects that the number of zones, zone sizes, and zone aggregation produce on the measure. Batty and al. [23] contains a practical application of these technical observations, in a study of the entropy of the population density in London. By defining and numbering zones concentrically from London's historical core, they showed that if one progressively adds new zones further away from the core, one observes an increase in the entropy, but this is almost entirely due to the increased number of zones. Normalising the entropy by the maximum value it could attain every time a new zone is added reveals a more meaningful pattern, with entropy at first increasing, as the area of study includes both the sparsely populated financial core and the densely populated inner areas London, and then declining, as a large number of vast, less dense suburb and rural areas are also included. This show the importance of both carefully considering the size of the window of observation as well as normalising the entropy to improve its interpretation. The same study also proposed to look at the temporal evolution of the measure, showing that the population density within Greater London became increasingly homogeneous over the course of the 20th century.

Batty's approach is prevalent in the field of spatial statistics, for example, in studies of the expansion and evolution of urban areas, such as [1] and [2].

It is worth remarking that while Batty's work focuses primarily on the issue of zone size and number, further developments have been made on the connection between entropy and scale in the analysis of scale-free spatial patterns, and the association between entropy and fractal dimension [57].

Another aspect which remains under-explored in Batty's work is the effect of the relative location of the zones in which the region of interest is subdivided, e.g., whether two zones are contiguous. Karlstrom and Ceccato [124] addressed this issue, by introducing spatial weights so that the neighbours of a zone also affect its contribution to the entropy. Nonetheless, their measure maintain a

global approach, that is, it focuses on the entropy of the entire area of study (or a subset of it) rather than the entropy of neighbours of individual zones. In Section 4.3, we see how Batty’s approach can be used to study the entropy of a region at the local level, as we focus on urban areas and their immediate surroundings.

2.4.2.3 Entropy of Cities Locations and Population

Shannon entropy has found fruitful applications as a measure of heterogeneity of the spatial distribution of cities in a region and their population.

Medvedkov [155], for instance, proposed to subdivide a region with m towns in n cells of equal areas; count the number m_i of towns in each cell i , for $i = 1, \dots, n$; deduce the distribution of count frequencies $p = \{p_j\}_{j=1, \dots, m}$, where $p_j = |\{i | 1 \leq i \leq n, m_i = j\}|$, and compute its entropy $H(p)$. If all cells contained the same number of towns, say k of them, then the distribution of count frequencies would be maximally concentrated ($p_k = n$, and $p_j = 0$, for all $j \neq k$), and thus the entropy would be minimum. This would indicate a highly regular spatial distribution, while larger values would correspond to more irregular distributions. Note that this definition is based on frequencies, and its interpretation is the opposite of Batty’s, for whom more regular distributions yield larger entropy.

A similar approach to entropy based on frequency has been proposed by Curry [67] in the context of city population sizes, to argue that regularity in the distribution of city sizes may emerge as a consequence of entropy maximisation. Given a system with Z settlements, of which Z_n have population size n , if one assumes that each settlements has the same attractiveness for an individual, the number of ways in which N individuals can be distributed among the settlement sizes is

$$W = \frac{Z!}{\prod_{i=1}^N Z_n!}, \quad (2.3)$$

Maximising W in a Boltmann’s fashion, with a Lagrange multiplier to account for the fact that the average settlement size is known to be N/Z , yields

$$Z_n = \frac{Z^2}{N} e^{-\frac{nN}{Z}}. \quad (2.4)$$

Hence, the probability that the $(i + 1)$ st largest city in the system has a population that is a ratio of the i th city is constant.

The argument was reprised by [29] and clarified and extended in [189]. The latter article reformulated the reasoning in terms of the probability p_n that a randomly selected city has size n , and the probability r_n that a person chosen at random lives in a city of size n , and showed that the entropy maximisation can be used to retrieve Zipf's law for the rank-size distribution⁶.

2.4.3 Why entropy?

The concept of entropy has proved both extremely versatile and meaningful in the study of several aspects of urban systems. In the previous Sections, we have seen some examples, and others will be presented in Sections 2.5.2.6 and 2.5.2.7, as network-specific notions of entropy are introduced.

A great deal of statistical regularities observed by systems of cities can be simply explained as the most likely outcome of a purely random process [30, 92, 165]. Such scenario corresponds in many cases to the one maximising the entropy of a distribution subject to some constraints. It is then natural to use entropy as a measure of regularity (and deviation from it), as it is in the deviations from the most likely scenario that the presence of meaningful structures ought to be sought.

Entropy is fundamentally linked to the crucial if at times nebulous notion of complexity, which, though not explicitly the subject of this thesis, is an established perspective in the study of urban system. Batty [23] argues that Shannon entropy captures several aspects that ought to be included in an intuitive definition of complexity of a system, such as the number of its components, scale, size, and (spatial) distribution.

More philosophically, a fundamental, albeit controversial, theory in urban science for the rationale behind the development of cities from sparser settlements is that of scaling laws (see, e.g., [31]). These suggest that to a linear increase in population corresponds a super-linear growth in economic

⁶[189] argues that the maximisation of W can be interpreted in terms of the energy of the system. The two distributions p_n and r_n are linked by the relation $r_n = \frac{np_n}{\langle n \rangle}$, where $\langle n \rangle = \frac{N}{Z}$ is the average city size. If we identify the population size n as the energy in the set of cities with n people, maximising W as in [67] can be shown to be equivalent to maximising $H(\{p_n\})$ subject to a constraint on the average energy $\sum_n np_n = \langle n \rangle$. This yields the Boltzmann distribution in eq. (2.4). If one, instead, accounts for the pairing of p_n and r_n and assumes r_n to be maximally entropic without constraints (and thus uniform), then a better definition for the energy is $\log n$ and the constraint on its average is $\sum_n p_n \log n = \langle \log(n) \rangle$. This corresponds to the cross entropy of $\{p_n\}$ and the distribution $\{p'_n\}$ obtained if r_n were uniform (see the Appendix in [189] for details). The maximisation, in this case, yields $p_n = \frac{n^{-\gamma}}{Z(\gamma)}$, for constant γ , i.e., the Zipf's distribution is retrieved as the maximally entropic one.

output and a sub-linear growth in infrastructure volume and cost. A possible explanation is that higher density fosters more and more varied connections, facilitating the exchange of knowledge and efficient division of labour. Shannon entropy is a measure of uniformity and concentration of a distribution. Therefore, the entropy of the distribution of population in different cities and in space, of their connections, and of their interactions, are reasonable candidates for a quantity to explore the fundamental mechanisms of urbanisation and the connection between the development of system of cities, major historical events and technological innovation, which define the core subject of this thesis.

2.5 Mathematical Framework

In our analysis of systems of cities, we rely on mathematical techniques from the theory of point processes and network theory, with particular emphasis on spatial and temporal networks, and on network entropy. Here we present the mathematical framework and the notation used throughout the thesis, as well as the fundamental results from the literature on the application of these concepts in Urban Science.

2.5.1 Points process statistics

A *point pattern* is a collection of points in some well-defined region of a Euclidean space. A *point process* is a generalisation of a random variable whose realisations are point patterns.

The theory of point processes is extremely vast and sits at the basis of Stochastic Geometry [198], the branch of Mathematics concerned with the study of random objects in space. In this thesis, we focus exclusively on point process statistics, that is the application of point processes to analyse the locations of objects that can be thought of as points distributed at random in space [119]. Classic examples of these include the location of trees in a forest, cells in biological tissues, and epicentres of earthquakes occurred in a given region in a fixed time interval. We also focus exclusively on processes in regions of the two-dimensional Euclidean space, which we see as a reasonable approximation of small portions of the Earth surface.

2.5.1.1 Point Process Models

Several point process models exist that generate point patterns with different geometric properties. The *Homogeneous Poisson Process (HPP)* is the simplest one. It satisfies two fundamental properties:

1. the number of points in any bounded subregion B of space is distributed as a Poisson random variable with mean $\lambda\nu(B)$, where λ is a fixed proportionality constant, known as the *intensity*, and $\nu(B)$ is the area of the subregion;
2. the number of points in k disjoint bounded subregions are k independent random variables, for any k .

The Homogeneous Poisson Process corresponds to the scenario of *Complete Spatial Randomness (CSR)*, where the intensity does not depend on the location and there is no interaction between the points. Two fundamental properties of the HPP are its *stationarity* and *isotropy*, that is the invariance of the distribution with respect to translation and rotations, respectively.

A more general model is provided by the *Inhomogeneous Poisson Process (IPP)*, that allows for the intensity $\lambda(x)$ to be dependent on the location x . The Inhomogeneous Poisson Process satisfies the latter of the above two properties and a suitably generalised version of the former, where the mean of the Poisson random variable describing the number of points in a bounded region B is substituted by the integral $\int_B \lambda(x)dx$, where the integration is with respect to the usual Lebesgue measure. The Inhomogeneous Poisson Process can be used to model patterns that show visible trends in the area of study and/or depend on exogenous factors that favour or hinder higher density in specific areas (for instance, rain patterns or elevation in the case of the location of trees). On the other hand, in an IPP the points remain mutually independent.

Neyman-Scott processes are a category of point process models that involve the generation of a point pattern in three steps: first a pattern of “parent points” is generated following an HPP with intensity λ_P ; then a random number of “daughter points” are scattered independently with identical distribution around each parent point; finally only the daughter points are kept in the final pattern. If C is the random variable of the number of daughter points around a parent, then the intensity of the process is $\lambda = \lambda_P \bar{C}$, where \bar{C} is the expected value of C . A realisation of a Neyman-Scott process is a *clustered* point pattern.

Different models correspond to different distributions of daughter points in the clusters. The *Matérn cluster process* generates the daughter points in discs of fixed radius R . The number of points C in each cluster follows a Poisson distribution, and the points are independently and uniformly scattered in the disc. The *Modified Thomas process* also generates C points around each parent, where C follows a Poisson distribution, but the points are distributed in a symmetric normal distribution, where the expected distance of a daughter point from its parent is regulated by a parameter σ .

Remark that also an IPP can create clustered conditions: the choice of a Neyman-Scott process model over an IPP can be ascribed to the expected generating mechanism of the pattern, but the two processes are not necessarily distinguishable on the basis of interpoint distance alone. The characteristics of a IPP and a Matérn cluster process can be combined by generating parent and/or daughter points according to an IPP instead of an HPP. We call this the *Inhomogeneous Matérn cluster process*.

A point pattern may also exhibit *regularity*, that is the points may be more distant one from the other than one would expect in the case of complete spatial randomness. Models for regular point patterns include the *Gibbs process*, in which pairs of points interact with a strength dependent on their distance and the points are distributed according to a probability density that depends on the total sum of these interactions⁷. When the interaction is such that no two points are allowed to exist at a distance below a certain threshold, we talk about a *Gibbs hard-core process*. When instead the interaction allows points to be arbitrarily close, albeit with very small probability, the process is said to be *soft-core*.

Finally, note that a point pattern may display several characteristics at once and may be better described by composing or superimposing the realisations of two or more of the above models.

2.5.1.2 Functions for point processes analysis

To analyse a point pattern and identify a suitable model, one may use *summary characteristics*: scalars or functions that capture geometric properties of the

⁷More precisely, if the number of points n in the pattern is fixed, the locations x_1, \dots, x_n of the n points follows the density $f(x_1, \dots, x_n) \propto \exp(-U(x_1, \dots, x_n))$, where $U(x_1, \dots, x_n)$ is the sum of the pairwise interactions and is known as the *total energy* of the system. This formula results from the maximisation of the entropy of f subject to a fixed expected total energy [119]. This opens interesting connections, which are not part of this thesis (see Section 7.2).

pattern.

Clark and Evans’s R index is the ratio of the mean nearest neighbour distance observed in the pattern to that expected in an HPP of the same intensity [62]. For a regular pattern, $R > 1$, whilst for clustered patterns $R < 1$.

The *Empty Space Function* or *Spherical contact distribution function* $F(r)$ is the probability that the disc $B(x, r)$ of radius r centred at an arbitrary location x in the observed region (not necessarily a point in the point pattern) contains at least one point of the pattern.

For a stationary point process, we also define the Nearest neighbour distance function $G(r)$ and Ripley’s K -function $K(r)$.

The *Nearest neighbour distance function* $G(r)$ is the distribution of the distance r from a typical point to its nearest neighbour, where by typical point here we mean a randomly chosen one⁸. The functions G and F differ from each other in that the former is centred on points, while the latter depends on location [119]. Consequently they are distinct, in general, though they do coincide in the case of an HPP, for which

$$F(r) = 1 - \exp(-\lambda\pi r^2) = G(r),$$

for all $r \geq 0$, where λ is the intensity of the process.

Ripley’s K -function is the function $K(r)$ such that $\lambda K(r)$ is the mean number of points in the disc $B(x, r)$, where x is a typical point of the pattern, excluding x itself. In the case of an HPP, $K(r) = \pi r^2$. The advantage of K with respect to F and G is that K captures inter-point distance properties beyond the first contact or nearest neighbour⁹.

The use of F , G and K is preferable to that of the R index, as they allow to study the pattern at different scales. Deviations of F , G and K from their theoretical values in the case of Complete Spatial Randomness can help assess whether the pattern exhibits clustering or repulsion at certain scales,

⁸The assumption of stationarity in the definition of G and K ensures that the notion of *typical point* of a point process, on which these definitions rely, is rigorously defined in the sense of Palm distribution theory. Palm distributions are conditional probabilities that a point process satisfies some property given that the process contains a point at a given location. We will not discuss the details here; instead, we intuitively equate the typical point to its empirical interpretation as “randomly chosen point” [198]. Hence, we will be able to extend these definitions to non-stationary point patterns as well, which can be helpful in fitting or testing point process models (see, for instance, [13] or [119, Example 7.1]).

⁹Indeed, if we generalise $G(r)$ to the k th nearest neighbour distance functions $G_k(r)$, then $\lambda K(r) = \sum_{k=1}^{\infty} G_k(r)$.

keeping in mind that the functions may also be affected by other sources of inhomogeneity.

Alternative approaches to analyse a point pattern include recurring to auxiliary *secondary structures* on point processes, for instance geometric networks, which we discuss in Section 2.5.3.3.

2.5.1.3 Validation of a point process model

Summary characteristics may be used to fit a suitable model to an observed point pattern. Following [119], fitting a point process involves a series of steps:

1. The choice of a possible model, informed by observation of the pattern and a priori knowledge.
2. Testing the CSR hypothesis (if this is not rejected, then no other model than an HPP ought to be sought).
3. Assess simple models, such as Neyman-Scott cluster models or Gibbs Processes, before passing to more elaborate cluster models based on non-Poisson parents.
4. Estimate the model parameters. For this, several methods are possible, including the method of *minimum contrast* that aims to minimise the discrepancy between the observed values of a summary characteristic $\hat{S}(r)$ and the theoretical values of the same function expected from the model for a choice of parameters.
5. Test the agreement of the model with the observed pattern via another summary characteristic $S'(r)$.
6. Perform a formal goodness-of-fit test, such an envelope test: generate a large number of patterns using the model, find the minimum value $S'_{min}(r)$ and maximum value $S'_{max}(r)$ of $S'(r)$ at each r among the generated patterns, and verify that the observed value $\hat{S}'(r)$ satisfies $S'_{min}(r) < \hat{S}'(r) < S'_{max}(r)$ for all r .

The identification of a model is useful to assess hypotheses concerning the nature and generative mechanism of the pattern.

2.5.1.4 Point Process to model city locations

There is a long tradition of studies that use point processes to study the spatial distribution of human settlements, most commonly in small, homogeneous rural areas with regular boundaries.

Dacey [68] studied the distribution of farms in the Tonami Plain in Japan in 1932 and of towns in Iowa with at least 2500 inhabitants, using census data from 1840 to 1960, at 10-year-long intervals. Both areas are characterised by remarkable uniformity in geographical features. The author's method consists in dividing the region under consideration into cells (squares of equal area in the case of the Tonami Plain, and counties of similar size in the case of Iowa) and in counting the number of settlements (farms or towns respectively) in each cell. The resulting distribution of counts is normalised into a probability mass function (pmf), whose first and second moments are matched to those of the following parametric probability mass function:

$$p(x) = (1 - p) \frac{\lambda^x e^{-\lambda}}{x!} + p \frac{x \lambda^{x-1} e^{-\lambda}}{x!},$$

for $x \in \{0, 1, \dots\}$. This allows to estimate of the parameter p of the pmf, which represents the transition from an entirely random distribution of points in the cells (for $p = 0$, one has the pmf of a Poisson distribution with parameter λ) to a more regular pattern (if $p = 1$ every cell has at least one settlement). The distribution of farms in the Tonami Plain in 1932 was found to be more regular than the CSR null hypothesis. For the case of towns in Iowa, the experiment revealed a steady increase in regularity in time, with values of p gradually increasing from $p = 0.008$ in 1840 to $p = 0.830$ in 1960. One may argue, not without reason, that choosing counties as cells in the case of Iowa introduces a bias in the counting, even considering the fact that their areas are very uniform. Yet, the results make perfect sense from a Christallerean perspective [60]: the mostly rural character of the State and the limited size of the towns suggest that they are primarily local bases for agricultural activities, and thus they are bound to be separated one from the other by the expansive rural areas they serve. Furthermore, the increasing regularity is also compatible with the latest stage of the aforementioned Hudson's spatial theory of rural settlements [116], which predicts growing uniformity as competition between rural centres forces inefficient farms to be annexed by more successful nearby ones. Despite the simplicity of its methodology, [68] remains a particularly interesting study for our thesis because of its dealing with the temporal evolution of the process

and its uniformity.

Glass and Tobler [94] made fuller use of methods from the theory of spatial point processes in their study of the distribution of towns in a rectangular portion of the Spanish Plateau south-east of Madrid. The short paper introduces the “radial distribution function” $g(r)$ for $r > 0$ as the probability that a point (by which here we mean a town) is located in a unit square at a distance r away from a typical point, proposing also an estimator. This function is more commonly known as the *pair correlation function* in more recent literature¹⁰. Comparing the radial distribution function $\hat{g}_{obs}(r)$ obtained for the data set to that of an HPP with the same intensity (for which $g(r) \equiv 1$), Glass and Tobler observe that $\hat{g}_{obs}(r)$ is considerably smaller than 1 for small values of the radius r , and then it oscillates about 1. This behaviour is typical of soft-core repulsive processes in which a point inhibits the presence of other points in its vicinity, but the strength of its influence is reduced as the distance from the point increases. The authors suggested that this observation was compatible with a point pattern that is maximally random subject to the constraint of moderately repulsive interactions. It is important to remark that the model tacitly assumes isotropy and stationarity, assumptions that may apply in the flat, homogeneous Spanish Plateau but would fall short in a more general setting.

Isotropy was the object of extensive scrutiny in [109] and [207], both of which used as case study the distribution of towns in the Argentinian Plain with a population of 2000 or more, comparing it for the years 1914, 1947, and 1960. In both articles, the authors connected the towns to their nearest neighbour and then studied the distribution of the off-north angles of the resulting vectors. A significant directionality in the pattern is observed in 1914, reflecting the road infrastructure radiating from Buenos Aires in the north-east. This directionality persists in subsequent years, though not significantly, which the authors ascribed to the addition of interlacing roads to the previous radial pattern and the subsequent development of towns along them. Though the articles agree about directionality, they differ in their conclusions regarding the evolution of the pattern. Haynes and Enders [109] used the nearest-neighbour-based Clark and Evans’s R statistic [62] and a decomposition of Medvedkov’s entropy of towns counts in cells to conclude that the process is degenerating towards greater randomness, contradicting Christaller’s thesis of increasing

¹⁰The pair correlation function $g(r)$ is strongly related to Ripley’s K -function, satisfying the relation $g(r) = \frac{K'(r)}{2\pi r}$ for stationary point processes in two dimensions [119].

uniformity in the time limit. This thesis was rejected by Upton [207], who argued that the R statistics suffers from edge-related effects and proposed to use the spherical contact distribution function F and the nearest neighbour distance distribution function G , instead. Such measures allowed for a finer study of the pattern at different scales and did not show any apparent reversion from regularity towards a more random arrangement. Upton also advocated the usage of double angles in the study of directionality to avoid the complications arising with the frequent reciprocity in the nearest-neighbour relation. Perhaps more interestingly, in the conclusion of the paper, the author called for a directional analysis that takes into account the unequal size of the towns, by weighting links according to distance and size, de facto reconnecting the point process based approach to an analysis based on SIMs (see in Section 2.3.2.4). In addition to warning against the indiscriminate adoption of the assumption of isotropy, the case study shown in [109] and [207] is also a cautionary tale about artificially defining the boundaries of the area under consideration, with little concern for the political entity it belongs to. Indeed, the pattern of cities in the Argentinian Plain seems to be influenced by the growth in the large cities of Buenos Aires, Rosario, and Córdoba, situated in its proximity, but outside its boundaries [109].

A remarkable attempt to address the issue of boundary effects and lack of stationarity is presented in [90]. The paper studies the distribution of population in 1970 in the Chicago area, using one point to represent ten thousand people. The most prominent geographical feature of the region is Lake Michigan, whose shore runs approximately south to north. The lake introduces a remarkable directional bias as a greater than average concentration of population can be observed in a thin strip facing the lake, and just north of the Loop (Chicago's Central Business District). A rectangular observation window was carefully placed in such a way to include the vast majority of the city but exclude any water surface and the denser strip, so to make stationarity plausible. The window was then mapped to a torus, so to avoid boundary effects. An edge-corrected estimator of Ripley's K function was finally used to test the pattern against the hypothesis of complete spatial randomness. Repulsion at small distances was observed, whilst clustering is observed at distances of 7-9 miles, corresponding to the mean journey-to-work distance in Chicago. The adopted methodology ignites a number of necessary reflections. Aggregating the population of a continuously inhabited area in a single point necessarily affects the results of the analysis at small scale. Second-order methods such as

Ripley’s K -function are extremely helpful for the analysis of point patterns because they capture the reciprocal influence of points at different scales, but in order to be meaningful they require the pattern under consideration to be stationary and isotropic. If these assumptions fail, directional trends in the data may incorrectly be identified as clustering. The article addressed the issue by artificially removing areas for which the assumption of stationarity is certainly not valid. But it could not ensure that stationarity was really satisfied.

Second order methods are employed in [101] to study the distribution of rural settlements in Iowa and Sweden. The results in this case show clear repulsion at small scale and clustering at larger scale, leading the author to suggest that a mixed model with large scale Poisson distributed clusters and repulsion among the points in the clusters might be the best at describing the distribution of points. However, such model entails a considerably more complex fitting procedure, that the author does not execute.

2.5.2 Graphs and networks

A *graph* is pair $G = G(V, E)$, where V is a set of objects called *vertices* and E is a set of pairs of vertices, called *edges*. When a graph is used to represent a real-world system of objects and their relations, it is often called a *network*, and its vertices and edges are referred to as *nodes* and *links*, respectively.

2.5.2.1 Graph definitions and notation

We distinguish between *directed* graphs, for which the edges in E are ordered pairs, i.e., $E \subseteq \{(u, v) | u, v \in V\}$, and *undirected* graphs, where the pairs in E are unordered, i.e., $E \subseteq \{\{u, v\} | u, v \in V\}$. An undirected graph such that any two vertices are connected by at most one edge and *self-loops* (i.e., edges connecting a vertex to itself) are not allowed, is said to be *simple*. All undirected graphs in this thesis are assumed to be simple. A *subgraph* is a graph whose vertices and edges are subsets of those of another graph.

If two vertices in an undirected graph are connected by an edge, they are said to be *adjacent*; the vertices are called the *endpoints* of that edge. A *clique* is a subset of vertices in a simple graph such that all pairs of distinct vertices are adjacent. Conversely, a *stable set* is a subset of vertices such that no two distinct vertices are adjacent. The set of all vertices adjacent to a given vertex v is called its *neighbourhood* and denoted by $N(v)$. If a vertex has no adjacent vertex, it is said to be *isolated*. For a graph $G = G(V, E)$ and a subset $S \subseteq V$

of its vertices, the *induced subgraph* $G[S]$ is the subgraph with vertex set S and as edge set all the edges in E whose endpoints are both in S .

The number of vertices $n = |V|$ is called the *order* of G , and the number of edges $m = |E|$ is called the *size* of G . For conciseness, we will call a graph of order n and size m an (n, m) -graph. Note that for a simple graph we have $m \leq \binom{n}{2} = \frac{n(n-1)}{2}$. The order of the largest clique contained in a graph is called its *clique number*.

In a directed network, the number $d^{out}(v)$ of edges issuing from a vertex $v \in \{v_1, \dots, v_n\}$ is called its *out-degree*; the number of edges $d^{in}(v)$ ending in v is called its *in-degree*. In an undirected network these two notions are indistinguishable and we simply call $d(v) = d^{out}(v) = d^{in}(v)$ the *degree* of v .

Definition 4. *The degree sequence of an undirected graph is the sequence $(d_i)_{1 \leq i \leq n}$ of vertex degrees.*

Conventionally, the degree sequence is arranged in non-increasing order, that is $d_i \geq d_j$ if $i \leq j$. Furthermore, a concise notation is adopted that indicates the multiplicity of an entry in the sequence as its exponent. For example, the degree sequence $(4, 1, 1, 1, 1)$ becomes $(4, 1^4)$. Remark that in a simple graph, $\sum_i d_i = 2m$, a result known as the *hand shaking lemma*.

Vertices and edges in a graph may be associated with labels or numerical attributes, that provide additional information about the entities and their relations. For example, the *weight* w_{uv} of the edge (u, v) is a positive value associated to the edge connecting u to v . In general, one takes $w_{uv} = 0$ if no edge exists from u to v . In a directed graph, the sum $s^{out}(v) = \sum_u w_{vu}$ of the weights of the edges issuing from a vertex v is called its *out-strength*; and the sum $s^{in}(v) = \sum_u w_{uv}$ of the weights of the edges ending in v is called its *in-strength*. In an undirected graph, $s(v) = s^{out}(v) = s^{in}(v)$ is simply called the *strength* of v .

A *path* from a node v to a node u in a (directed) graph is a sequence of vertices connected by distinct (directed) edges that starts in v and ends in u , never passing through an already visited node (when the sequence is allowed to include multiple times the same vertex, but never the same edge, one talks, more generally, of a *trail*). If a vertex v is connected via a path to a vertex u , u is said to be *reachable* from v . The set of all vertices reachable from v is called the *forward reachability set* of v . The set of all vertices that reach v is called the *backward reachability set* of v .

When any two vertices in the graphs are connected by at least one path, the graph is said to be *connected*. A *component* of an undirected graph is a

connected subgraph that is not contained in any larger connected subgraph. More than one path may exist between any two vertices in a graph, thus special attention is given to the *shortest paths* connecting them, by which one can mean either the paths containing the smallest number of edges, or, in the context of weighted graphs, the paths whose total sums of edge weights is minimal. The *distance* d_{uv} from a vertex u to a vertex v is the length of any shortest path connecting them, i.e., the number of edges or the sum of edge weights along the path. Remark that d_{vu} need not be equal to d_{uv} in directed graphs, whilst equality always holds in undirected graphs. If no path connects two vertices u and v , then one formally lets $d_{uv} = \infty$ or $\frac{1}{d_{uv}} = 0$.

2.5.2.2 Measures of vertex importance

The degree of a vertex and its strength, if the graph is weighted, are two simple measures of the vertex’s importance. The *size of the forward reachability set* extends the notion of (out-)degree. It can be thought of a measure of the potential spread of information released from v when this can be transmitted to the vertices to which v is connected, and from these to the vertices they are connected to, and so on. Hence, the size of the forward reachability set may be a more appropriate notion of importance of a vertex in the context of information diffusion since, whilst a vertex may have few connections in itself, the connections of these may be numerous and the information it released may thus spread extensively. Similarly, the size of the backward reachability set can be interpreted an indicator of the receptiveness of a vertex.

The literature contains a great deal of sophisticated notions of the importance of single nodes, collectively known as *centrality measures*, each offering a different interpretation of the potential influence of a node in the overall network. The classic [86] contains a clear discussion of the meaning and origin of several centrality measures, while [131] provides a more modern and extensive account. Amongst the most common ones, we mention incidentally the *closeness centrality* and the *betweenness centrality*, that we employed in [159], upon which Section 6.3 is partially based. According to closeness centrality, a vertex is more central when it is close to the other vertices, as measured by shortest paths [25]¹¹. The betweenness centrality of a vertex v is based on the principle that v is more central when it is “strategically located” along the

¹¹The original definition of closeness centrality given in [25] could be applied only to connected graphs. A definition that could be applied to any graph was introduced by [143], but had already been used in the context of real-world networks in [93].

largest number of shortest paths connecting other vertices [24, 85].

2.5.2.3 Graph invariants

Reachability and centrality measures focus on individual, distinguishable vertices in the graph. A different approach consists in considering *graph invariants*, that is properties of the overall structure of the graph, that do not depend on labelling its vertices or edges. The aforementioned size and order of a graph are examples of graph invariants given by a single integer value; the *diameter*, that is the length of longest path in an unweighted graph, is another example. Being connected is an example of boolean invariant, and so is the property of having no *cycles*: non-empty trails connecting a node to itself. Other invariants are given by finite sequences of non-negative integers; the degree sequence is a chief example of these.

Some simple graphs satisfying basic invariants have been named and reserved a special notation for convenience. The following definitions list a few of them, that will be helpful in the rest of the thesis.

Definition 5. *The following are named simple graphs of order n :*

- *A tree T_n is a connected graph with no cycles.*
- *The path P_n is the tree with degree sequence $(2^{n-2}, 1^2)$.*
- *The star S_n is the tree with degree sequence $(n-1, 1^{n-1})$.*
- *A k -regular graph is a graph in which every vertex has degree k . Its degree sequence is (k^n) .*
- *The complete graph K_n is the graph in which all pairs of vertices are connected by an edge. Equivalently, it is the $(n-1)$ -regular graph.*
- *K_n^- is K_n with one edge removed. It is the graph with degree sequence $(n-1^{n-2}, n-2^2)$.*

Definition 6 (Colex graph). *Write the graph size m as $m = \binom{k}{2} + \ell$ where $0 \leq \ell < k$. The colex graph $\mathcal{C}(m)$ is the graph formed by connecting a vertex v to ℓ vertices of a clique K_k .*

An example of colex graph is depicted in Figure 2.2 for $m = 31$, i.e., $k = 8$ and $\ell = 3$.

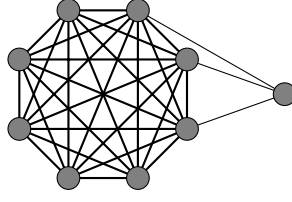


Figure 2.2: The graph $\mathcal{C}(31)$.

Definition 7. A threshold graph $K_a T$ is one whose vertex set can be partitioned into two disjoint sets K_a and T , where K_a is a clique with a vertices, no edge has both ends in T and if $d(u) \geq d(v)$ for $u, v \in T$, then all vertices adjacent to v are also adjacent to u [221].

Remark 8. The colex graph is a threshold graph.

2.5.2.4 Graph entropy measures

We have seen two different ways to analyse a graph: focussing on individual vertices or studying structural graph invariants. A Shannon-entropy based approach can be taken in both cases. In fact, several vertex properties and graph invariants are given by finite sequences for which Shannon entropy is an intuitive measure of uniformity.

In this thesis we focus on the following two notions of Shannon entropy, one at the individual vertex and the other at the global level:

Definition 9 (Entropy of interactions). Let $G = (V, E)$ be a weighted, directed graph (possibly with self loops) and let v be a vertex in G . The entropy of interactions $S(v)$ is the Shannon entropy of the weight w_{vu} of edges (v, u) issuing from v , normalised by the out-strength $s^{out}(v)$:

$$S(v) = - \sum_{u \in V} \frac{w_{vu}}{s^{out}(v)} \log \left(\frac{w_{vu}}{s^{out}(v)} \right). \quad (2.5)$$

An equivalent definition can be formulated for incoming connections, but we will not use it in this thesis.

Definition 10. The first degree-based entropy of a simple graph G with degree sequence $(d_i)_{1 \leq i \leq n}$ and size m equals

$$I(G) = - \sum_{i=1}^n \frac{d_i}{2m} \log \left(\frac{d_i}{2m} \right). \quad (2.6)$$

The adjective “first” in Definition 10 refers to a more general definition of degree-based entropy that rests on degree powers d_i^c , with $c > 0$ an arbitrary positive integer, and reduces to eq. (2.6) if one takes the first power $c = 1$ [91]:

$$I^c(G) = - \sum_{i=1}^n \frac{d_i^c}{\sum_{j=1}^n d_j^c} \log \left(\frac{d_i^c}{\sum_{j=1}^n d_j^c} \right).$$

The entropy of interactions of a vertex v attains its minimum value of $S_{min}(v) = 0$ when only one edge issuing from v has non-zero weight; it attains its maximum $S_{max}(v)$ when all edges issuing from v have equal weight. Hence, normalising by $S_{max}(v)$, one obtains an index between 0 and 1 suitable to compare the entropy of interactions even between graphs with different order.

In absence of constraints on the vertex and the graph (i.e., considering an arbitrary weighted directed graph, possibly with self-loops), the theoretical maximum is $S_{max}(v) = \log(n)$ where n is the order of the graph. Additional information on the graph and on the vertex call for the adoption of more a appropriate normalisation constant. For example, if self-loops are not allowed in the graph, the maximum value $S(v)$ can obtain is $S_{max}(v) = \log(n - 1)$, which is thus a better choice.

2.5.2.5 Normalisation of the first degree-based entropy

The normalisation and interpretation of the first degree-based entropy present a more difficult challenge. When the only constraint on the graph is its order n , the first degree based entropy attains its maximum value $\log(n)$ for all regular graphs. When only the size m is known, the maximum entropy $\log(2m)$ is reached by the graph mK_2 : the graph of order $2m$ formed by m pairs of vertices connected by an edge. When both the order n and the size m are known, Yan [221, Theorem 7] showed that the maximum is obtained by a graph whose nodes have largest and smallest degree differing at most by 1 (see Theorem 11) . This corresponds again to the intuition that the graphs with maximum entropy are those with degree sequence as uniform as possible, given the constraints.

Finding the minimum value is not always equally simple. The problem resides in the fact that not every arbitrary finite sequence of non-negative integers is a valid degree sequence; trivially, it is not possible to have a degree sequence with all but one entry equal to zero that would lead to zero entropy. There are specific conditions to ensure that a finite sequence is a valid degree sequence (see, for example [194]) and they impose major limitations on the

values that the first degree-based entropy can attain.

More generally, determining the extremal values of the entropy of a graph invariant is a non-trivial task. Dehmer [71] proposed a general approach to study the Shannon entropy of probability distributions derived by normalising graph invariants. This approach was motivated by the application of entropy in the study of graph invariants in mathematical Chemistry [162, 36] and was indeed subsequently applied in [73] to compare the properties of real and synthetic chemical structures.

Powers d^c of the vertex degrees d , normalised by their sum, were from the beginning a chief example of such invariants, also in light of their connection to the first Zagreb index [77] in mathematical Chemistry. Cao et al. [52] proved extremal properties for certain classes of graphs (trees, unicyclic, bicyclic, and chemical graphs of given order and size) for the special case of the first degree-based entropy, i.e., when the exponent of the degree powers is $c = 1$. In [51], they extended their work to the case $c > 1$ and provided bounds depending on the smallest and largest degree of the graph. The smallest and largest degree were similarly used in [140] and [141] to provide new bounds, which they proved via Jensen's inequality. Chen et al. [58], instead, focused on the relation between the entropy and different values of the exponent c , providing numerical results for trees, unicyclic, bipartite and triangle-free graph with small number of vertices. Das and Dehmer [69] conjectured that the path graph maximises the degree-based entropy for $c > 0$ among trees, a result eventually proved in [118]. Ghalavand et al. [91] focused again on the first degree-based entropy and used the Strong Mixing Variable method to extend maximality results for trees, unicyclic and bicyclic graphs, following an idea very similar to that used in [52, Lemma 1].

Yan [221] focused on the first degree-based entropy for graphs with given order and size and proved that a connected graph whose degree sequence achieves minimum entropy must be a threshold graph [221, Theorem 4]. The same result holds true when only the graph size is fixed. Yan proceeded to find the minimum entropy for graphs with given order n and size m when $n - 1 \leq m \leq n + 5$ and conjectured that the degree sequence of the graph that minimises the first degree based entropy when $m \geq n + 9$ is

$$(n - 1, m - n + 2, 2^{m-n+1}, 1^{2n-m-3}).$$

Note that the cases $n - 1 \leq m \leq n + 1$ correspond to trees, unicyclic, and bicyclic graphs, solved in [52].

In Chapter 3, we determine the extremal values of the first degree-based entropy for connected graphs with given size, prove an extended version of Yan’s conjecture in the range $n - 1 \leq m \leq 2n - 3$ for graphs and with given order and size, and provide numerical results for the case $m \geq 2n - 2$.

2.5.2.6 Entropy of interactions in Urban Science

As mentioned in Section 2.3.2.3, studies of functional polycentricity start invariably from explicitly defining a network of cities where the nodes are taken to represent cities in a region and the links between them a functional relation of interest. Denoting by V the set of cities in the region, u and v are two cities in V , and w_{uv} , the intensity of the connection from u to v . For example, if the network represents commuting patterns in a region, w_{uv} may denote the number of people commuting from u to v .

Yang et al. [224] and Marin et al. [144] used the entropy of v ’s interactions (Definition 9) as a measure of the node’s functional polycentricity. In this context, $S(v) = 0$ when all connections from v end in a single city, i.e., when the set of connections is maximally centralised, and $S(v) = \log n$ when all links issuing from v have equal weight, i.e., when the relation of v with its neighbours is maximally polycentric.

The construction of SIMs using morphological properties of the cities in a system has allowed to apply the entropy of interaction as a measure of morphological polycentricity of the system. Chapman [56] studied the evolution of the UK and the U.S. system of cities, combining the changing population of cities with their distances via a simple gravity model to obtain two weighted networks. For each city, the author then computed the entropy of interactions. A similar approach has been proposed for Chinese cities by [84], who also included the cities’ Gross Domestic Product in their computations. In [84], the distance between the city is computed using an unrealistic Euclidean metric. In [56], the distance is computed along the railway network, but the method does not account for the evolution of the railway network, neither in terms of its extension nor in terms of its speed. In both cases, the choice of the gravity model to establish potential interactions is arbitrary, and does not allow for self-interactions, as discussed in Section 2.3.2.4. This implies that in an extremely unbalanced system constituted by two cities, one much larger than the other, the measure would counter-intuitively assign maximum possible balance to both the towns. In Section 5.2, we use the entropy of a node’s interactions to measure the evolution of the morphological polycentricity of

the English and Welsh system of cities. We adopt the asymmetric Wilson’s Spatial Interaction Model, based on travel time on an evolving railway network to overcome some of the limitations of the previous studies, and obtain a more intuitive and, we argue, realistic measure of morphological polycentricity.

2.5.2.7 Alternative graph entropy measures

It is worth remarking that the complex networks literature proposes several other network entropy measures derived from the Shannon entropy, of which an extensive survey is presented in [72]. Some of these have been successfully adapted to the urban context. Gudmundsson and Mohajeri [97], for example, studied the Shannon Entropy of the distribution of the lengths and cardinal orientations of streets in 41 cities, tracking their evolution in time as the cities expanded.

The von Neumann entropy of a network introduced by Passerini and Severini [167] has also been interpreted as a measure of centralisation of the connections in a graph by Simmons et al. [196]. Yet, it is computationally expensive, and existing approximations [103, 138, 136], present each its own set of problems [160]. Furthermore, a general interpretation of von Neumann entropy is still an open problem [160], especially in the context of weighted and directed graphs [225], which limits for now its applications to study urban systems.

2.5.3 Spatial and temporal networks

Time and space play an important role in shaping interactions between cities and between individuals. Temporal and spatial networks provide a powerful methodological tool of analysis for these relations.

2.5.3.1 Temporal Networks

In a temporal network, nodes and links are transient and exist only in certain, well defined, intervals of time [114]. Because of this, temporal networks are occasionally also referred to as time-varying [213], or evolving [218] networks.

A temporal network is perhaps easier to imagine as a sequence of static snapshots of the network taken at different moments in time, each capturing a specific instant in the network’s evolution (see Fig. 2.3), a perspective which reconnects them to the more encompassing theory of multilayer networks [34]. For simplicity, it is sometimes convenient to “ignore” the temporal information

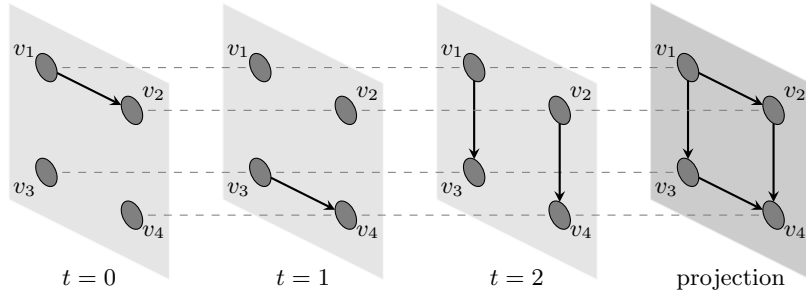


Figure 2.3: Simple temporal network with its projection.

and consider the static network that includes all nodes and links appearing in the temporal network at any point in time. This cumulative view of the network is called the temporal network’s *projection* [114], as it is akin to projecting the multilayer network on a single layer.

Temporal networks have been successfully used in the analysis of communication and the consequent diffusion of information. In [113], they were applied to model email exchanges: each individual is represented as a node and an email is a directed link between two individuals that exists for a single instant in the moment it is read. Similarly, temporal networks are a crucial tool in the study of the spread of diseases. In [182] individuals are again taken as nodes, and if two individuals come in proximity of each other, a link is drawn between them which persists for the duration of their contact; transmission of disease may then happen from an infected individual to a healthy one along such links. Temporal networks are also used as a tool to analyse the emergence of patterns in static networks [154].

2.5.3.2 Node Centrality in Temporal Networks

The definitions of node importance given in Section 2.5.2.2 need some care to be generalised to temporal networks.

One possible approach is that of analysing each of the network’s snapshots, but this ignores the relation between them. Alternatively, one can adopt a cumulative approach, analysing the network’s projection, or several projections including the network’s evolution up to certain time. A couple of remarks are needed.

First of all, it is important to notice that not all paths that appear in the projection network are viable, or time-respecting paths [114], across which transmission of information from a node to another one is possible. To clarify this statement, let us consider the situation presented in Fig. 2.3. In the

projection network, two paths connect v_1 to v_4 : one passing through v_2 and the other one passing through v_3 . The link connecting v_1 to v_2 appears at time $t = 0$, and the link connecting v_2 to v_4 appears at time $t = 2$. Thus, if we study the transmission of information along the network, a piece of information from v_1 would be able to reach v_4 by first being transmitted from v_1 to v_2 at time $t = 0$, and then from v_2 to v_4 at $t = 2$. On the other hand, the link connecting v_1 to v_3 appears at time $t = 2$, while the link connecting v_3 to v_4 appears at $t = 1$. This means that it would not be possible for information from v_1 to reach v_4 via v_3 . Hence the only time-respecting path connecting v_1 to v_4 is the one passing through v_2 .

Second, in a temporal network, the notion of shortest path (among the time-respecting ones) may have two different interpretations, and thus impact some centrality measures. On the one hand, a shortest path connecting an ordered pair of nodes may be, as in a static network, a path that allows to reach the latter from the former in the smallest number of steps; on the other hand, a shortest path may be one that allows to reach the latter from the former in the shortest time [208]. Which of these two interpretations should be preferred, depends on the context.

[163] and [200] define several careful alternative generalisations of centrality measures to temporal networks, and [199] presents how some of them have been applied in real-world networks.

2.5.3.3 Spatial Networks

In a spatial network, the nodes are associated with locations in a space and the structure of the network is at least partially constrained by its spatial properties [16]. The dimension of the space and its metric depend on the system under consideration. The infrastructural and social networks studied in this thesis are constrained on a portion of the Earth's surface, so we adopt a 2-dimensional space, in which the location of a node v is completely identified by its two geographical coordinates (lon_v, lat_v) . The simplest way of computing the distance between two nodes is along great circles. For a network located on a sufficiently small region, the 2-dimensional Euclidean distance might provide a reasonable approximation, but if the region of interest is large enough, such as the network of international flights that we analyse in Section 5.1, the great circle spherical distance or another measure of distance on ellipsoids or geodes is strictly necessary.

Note that in several real-world applications the metric properties of the

space may be considerably more complex. The presence of transport infrastructure and the maximum speed attainable, for instance, affect the ease, cost, and speed of travel, and induce alternative metrics on the space. We explore the role of transport technology on distance in Section 5.2.

The strength of spatial network theory is that it allows to translate metrical properties into graph properties that can be analysed with the above described graph measures. For this reason, spatial graphs are often used as a complementary approach to the analysis of point patterns, alongside the summary statistics that we discussed in Section 2.5.1. Geometric networks are spatial networks built on a point pattern according to some geometric rule [16]. Illian [119] provides a list of geometric networks models that are commonly used in the analysis of point patterns. These include the *maximum threshold graphs* or *disc graphs*, in which nodes are connected if they are closer than a certain distance; nearest-neighbours graphs, in which nodes are connected to a fixed number of their closest neighbours; and Gabriel graphs, where two nodes are connected if and only if no other node falls in the circle having as diameter the segment that connects them.

2.5.3.4 Urban spatial networks

Spatial network theory finds extensive application in the analysis of systems of cities, both because of its great versatility, as the nodes and links in a network allow to effectively represent numerous kinds of geographical entities and relations between them, and because of its well-developed set of models and techniques.

The interaction between systems of cities are often modelled via the deterministic and stochastic spatial interaction models discussed in Section 2.3.2.4. Examples include highway traffic flows between South Korean cities [122] via gravity models; mobility and migration patterns in the United States via the radiation model [195]; maritime interactions between Aegean islands via the minimisation of a cost/benefit Hamiltonian [129]. The choice of the space metric in these models depend on the context. For instance, in [112] the authors used least-cost paths to connect pre-colonial settlements in the Southwestern United States, noticing that the area's topography, rich in valleys and peaks, would make other simpler distance measures inappropriate.

Brughmans and Peeples [44] surveyed a number of spatial network models used to reconstruct interactions between historical settlements, and compared their efficacy in modelling the ancient Roman transport network, including

several of the aforementioned geometric networks¹².

Another context in which spatial networks are widely applied is the study of polycentricity. For morphological polycentricity, the existence of a network is not always explicit. Yet, to account for the spatial distribution of cities, distances between the cities have to be computed, and this defines an implicit network structure where the weight w_{uv} of the link between cities u and v is some function of their distance.

First introduced in Wasserman et al. [210], the aptly named *Polycentricities* are a class of network-based measures that make use of the standard deviation of the nodes' strengths $s(u)$, and of the network density Δ to give an indication of how polycentric the network is.

Polycentricities have been adapted to the urban context by Green [95], who took the nodes in the network to be cities, and the links to be functional relations between them, such as email exchanges, commuting, or migration flows. Green defined the *Ordinary Polycentricity* $P_{F(N)}$ of a directed network N with m links and no loops as

$$P_{F(N)} := \Delta \left(1 - \frac{\sigma_F}{\sigma_{F_{max}}} \right), \quad (2.7)$$

where $\Delta = \frac{m}{n(n-1)}$ is the density of the network; σ_F is the standard deviation of the out-strength of the vertices; and $\sigma_{F_{max}}$ is the standard deviation of the pair $\{0, s_{max}\}$, s_{max} being the maximum vertex strength in the network.

$\sigma_{F_{max}}$ is the maximum value σ_F can attain, so that $P_{F(N)} \in [0, 1]$, with 0 indicating maximal centralisation and 1 denoting complete polycentricity. The subscript F in eq. (2.7) refers to the particular functional relation used to build the network. $P_{F(N)}$ is a purely topological measure, in the sense that it contains information about the network's links and not about the location of its vertices, but it can be extended so that it also accounts for the distance between the network vertices when multiplied by the *Topographical Polycentricity* P_T . This is defined as:

$$P_T := 1 - \frac{\sigma_T}{\sigma_{T_{max}}}, \quad (2.8)$$

where σ_T is the standard deviation of the distances between the cities (possibly subject to some upper bound) and $\sigma_{T_{max}}$, as above, is the standard

¹²Sometimes, geometric network models may appear implicitly in other fields under different names. For instance, the usage of nearest-neighbours graphs in the study of historical settlements in Archaeology is an approach also known as Proximal Point Analysis [43].

deviation of the pair $\{0, d_{max}\}$, where d_{max} is the maximum distance between two cities in the system. As above, $P_T \in [0, 1]$. Thus, one obtains the *Regional Polycentricity*

$$R_{SF} := P_T \cdot P_F, \quad (2.9)$$

whose name hints to the fact that imposing an upper bound on the maximum distance allows to study systems of cities located within a particular distance from each other, and thus within the same region.

Green [95] developed Polycentricities with a functional approach in mind, and, in fact, the computation of Ordinary and Regional polycentricity requires knowledge of additional data on the functional relation between the cities which cannot be derived from the spatial distribution and population sizes of the cities alone (the variables of interest in morphological polycentricity). Nonetheless, there have been attempts to extend Polycentricities to spatial urban networks built using a morphological perspective. For instance, [137, 135, 134] adapted the Ordinary Polycentricity to study the relation between the main centre of a city and its secondary subcentres, substituting the term σ_F in eq. (2.7) with the standard deviation of the products $p_j \cdot d_j$ of the population p_j of a subcentre j and its distance d_j from the main centre. One issue of this adaptation is that it assumes the existence of a single main centre, and this may not be the case within a region. Furthermore, it ignores the relations between smaller centres, which may be of great interest.

Chapter 3

Mathematical Prelude: the Normalisation of First Degree-Based Entropy

Recall that, by Definition 10, the first degree-based entropy of a simple graph G with n vertices, m edges and degree sequence $(d_i)_{1 \leq i \leq n}$ is

$$I(G) = - \sum_{i=1}^n \frac{d_i}{2m} \log \left(\frac{d_i}{2m} \right).$$

The first degree-based entropy is a conceptually and computationally simple measure of uniformity and balance of the connections in a graph. As discussed in Section 2.5, its careful and context-informed normalisation and interpretation pose some challenges, as they require knowing the range of values the measure can take, which is not a trivial problem. Finding the measure's extremal value for networks satisfying natural constraints is crucial to solve it.

In the context of intercity transportation networks, an obvious constraint for the structure of the network is the availability of financial or material resources. Whilst it might be ideal for each city to have a dedicated and direct connection to every other city in the system, this is often not economically viable. Instead, there may be sufficient resources to construct and maintain only a certain number of connections.

Another constraint in transportation networks is the necessity to ensure overall connectivity across the system. In principle, a well-designed transportation network should connect every pair of distinct nodes in the system, albeit possibly only via an indirect path. This is the case, for example, of the

network of flights connecting airports across the globe (Section 5.1), and of a well-developed road and railway network on a landmass.

In this Chapter, we determine the extremal values of the first degree-based entropy given these two natural constraints.

3.1 Introduction

3.1.1 Problem Statement

In the abstract language of graphs, the problem translates into finding the extremal values for graphs with given size, and for connected graphs with given order and size. In both cases, the graphs are assumed to be undirected and simple.

Determining the maximum entropy given these constraints is relatively straightforward. Among all graphs of size m , the (unique) graph attaining the maximum entropy is mK_2 : the graph of order $2m$ formed by m pairs of vertices connected by an edge (Fig. 3.1a). This is because $2m$ is the maximum order of a graph of size m , and thus any graph G , of any order n , satisfies $I(G) \leq \log(n) \leq \log(2m) = I(mK_2)$. Amongst all graphs with given order and size, Yan [221] proved that the maximum entropy is obtained by an “almost regular graph”, that is a graph whose nodes have largest and smallest degree differing at most by 1 (Fig. 3.1b).

Theorem 11 (Theorem 7 in [221]). *The (n, m) -graph $G_{n,m}^{max}$ that maximises the first degree-based entropy has degree sequence*

$$((q+1)^r, q^{n-r}),$$

where $2m = qn + r$, with $q = \lfloor \frac{2m}{n} \rfloor$ and $0 \leq r < n$.

Here we determine the graphs with minimum entropy. More precisely, we prove the following two statements:

Theorem 12. *Among all graphs with m edges, the colex graph $\mathcal{C}(m)$ minimizes the first degree-based entropy.*

The colex graph $\mathcal{C}(m)$ (Definition 6) maximizes the number of triangles (and cliques) among all graphs with size m [126]. Thus it can be considered as the most clustered graph and so a natural candidate for minimising the entropy, which is a measure of balance.

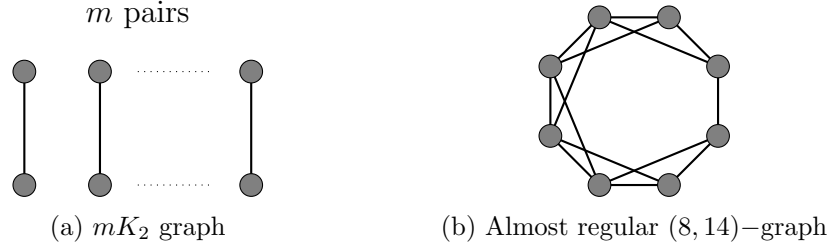


Figure 3.1: Examples of graphs yielding maximum first degree-based entropy when the size m is known, and when both the size and order n are known.

Theorem 13. *Among all (connected) graphs with n nodes and m edges, with $n - 1 \leq m \leq 2n - 3$, every graph minimizing the first degree-based entropy has a degree sequence of the form*

$$\left\{ \begin{array}{ll} (n - 1, m - n + 2, 2^{m-n+1}, 1^{2n-m-3}) & \text{if } n + 6 \leq m \leq 2n - 3 \\ & \text{or } m \in \{n, n + 1, n + 3\}, \\ (n - 1, 4^4, 1^{n-5}) & \text{if } m = n + 5, \\ (n - 1, 4^2, 3^2, 1^{n-5}) \text{ or } (n - 1, 6, 2^5, 1^{n-7}) & \text{if } m = n + 4, \\ (n - 1, 3^3, 1^{n-4}) & \text{if } m = n + 2, \\ (n - 1, 1^{n-1}) & \text{if } m = n - 1. \end{array} \right.$$

These graphs are presented in Table 3.1.

Furthermore, we provide numerical observations for graphs with n nodes and m edges, with $2n - 3 < m \leq \binom{n}{2}$.

3.1.2 Notation

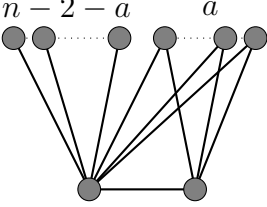
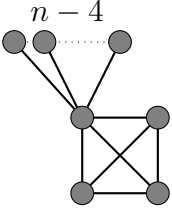
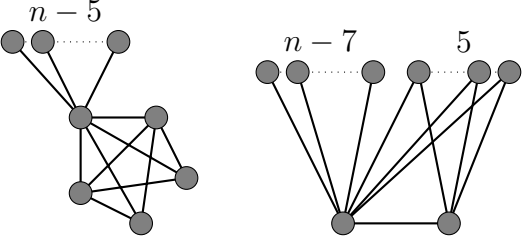
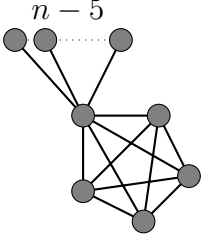
We begin by introducing some notation that will be helpful in the proofs.

Definition 14 (Functions f_c and h_c). *For any constant $c \geq 0$, we define the function $f_c(x) = (x + c) \cdot \log(x + c)$. For a graph G with degree sequence $(d_i)_{1 \leq i \leq n}$, we define $h_c(G) = \sum_i f_c(d_i)$. When $c = 0$, we just write $f(x)$ for $f_0(x)$ and $h(G)$ for $h_0(G) = \sum_i d_i \log(d_i)$.*

With this notation, we can rewrite the definition of first degree-based entropy (Definition 10) as

$$I(G) = \log(2m) - \frac{1}{2m} h(G).$$

Table 3.1: Overview of extremal (n, m) -graphs minimizing the entropy. Compare to [221, Table 1] and [52, Theorems 1,2, and 3].

$m = n - 1 + a$ $0 \leq a \leq n - 2$ $a \notin \{3, 5, 6\}$	
$m = n + 2$	
$m = n + 4$	
$m = n + 5$	

Hence, determining the minimum (resp. the maximum) of $I(G)$ is equivalent to determining the maximum (resp. the minimum) of $h(G)$.

The function h_c allows us to generalise these arguments while imposing additional constraints on the graph. For example, fixing the order n means that all n entries in the degree sequence are greater than or equal to 1, a

condition that $h_1(G)$ embeds directly in its definition. We will see how finding the graph G maximising $h_1(G)$ allows to simplify the solution of Theorem 13. Remark that when $c \in \{0, 1\}$, isolated vertices of G do not contribute to the value of $h_c(G)$, since $0 \cdot \log(0) = 0 = 1 \cdot \log(1)$. When $c \geq 2$, the function $h_c(G)$ depends on the number of vertices as well, since isolated vertices of G contribute $f_c(0) = c \log c > 0$. Thus, we compare graphs with a different order by implicitly extending the order, i.e., adding isolated vertices in such a way that the graphs have the same order¹.

Also note that

$$f(x) = x \log(x) = \int_1^x (\log(t) + \log(e)) dt,$$

a fact² that we use frequently in the proof of Lemma 25 in Section 3.2.2.

In some proofs, we will also make use of the following function.

Definition 15. *The function Δ_c is defined by $\Delta_c(x) = f_c(x) - f_c(x-1) = \log(e) + \int_{x+c-1}^{x+c} \log t dt$.*

Note that Δ_c is a strictly concave, increasing function.

Definition 16 (Generalised colex graph). *Let k and m be integers with $m = \binom{k-1}{2} + a(k-1) + b$ for some integers a, b satisfying $a \geq 0$ and $0 \leq b \leq k-2$. The graph $\mathcal{C}(m, k)$ is formed by a clique K_{k-1} whose vertices are all connected to a stable set of size a and with b vertices of the clique connected to one additional vertex.*

An example of $\mathcal{C}(m, k)$ graph is presented in Fig. 3.2.

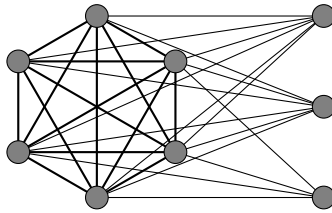


Figure 3.2: The graph $\mathcal{C}(31, 7)$. Here $a = 2$ and $b = 4$.

Remark 17. *When $m < \binom{k+1}{2}$, the graph $\mathcal{C}(m, k)$ is just the colex graph $\mathcal{C}(m)$.*

For instance, the graph $\mathcal{C}(31, 8)$ is just the colex graph $\mathcal{C}(31)$ in Fig. 2.2.

¹If the orders of two graphs to be compared were N and n , with $N \geq n$, one could have defined $h_c^N(G) = \sum_i f_c(d_i) + (N-n)f_c(0)$ and used h_c^N for the comparison, but it would have made the notation excessively heavy.

²The presence of the constant $\log(e) = \log_2(e)$ in the integrand derives from the choice to use the logarithm base 2 in this study, in accord with the main results from the literature.

3.1.3 Preliminary Results

Since we are dealing with degree sequences, some arguments are simplified by using general results on sequences. We start with introducing the notion of majorising sequences. When we speak about a sequence majorising another one, we always assume that the sequences are non-increasing sequences of non-negative reals. If necessary, we add zeros to be sure they are the same length.

Definition 18. *A sequence $(x_i)_{1 \leq i \leq n}$ majorises the sequence $(y_i)_{1 \leq i \leq n}$ if and only if $\sum_{1 \leq i \leq j} x_i \geq \sum_{1 \leq i \leq j} y_i$ for every $1 \leq j \leq n$ and equality holds for $j = n$.*

This definition is used in the following useful inequality, with its immediate subsequent corollary and subsequent lemma.

Theorem 19 (Karamata's inequality, [123]). *Let $(x_i)_{1 \leq i \leq n}$ be a sequence majorising the sequence $(y_i)_{1 \leq i \leq n}$. Then for every convex function g we have $\sum_{1 \leq i \leq n} g(x_i) \geq \sum_{1 \leq i \leq n} g(y_i)$. Furthermore, this inequality is strict if the sequences are not equal and g is a strictly convex function. For concave functions, the same holds with the opposite sign.*

Corollary 20. *Let $c > 0$ and let G and G' be two graphs such that the degree sequence of G' majorises the degree sequence of G . Then $h_c(G') \geq h_c(G)$.*

Proof. By Theorem 19, it is sufficient to note that $f_c(x) = (x + c) \log(x + c)$ is a convex function for any $c > 0$. \square

This corollary has also been observed in [81] for $c = 0$.

Lemma 21. *Let $n, t, \ell > 0$ be fixed positive integers. Write $t = qn + r$, with integers $q \geq 0$ and $0 \leq r < q$. Then, under the condition that all z_i 's in $(z_i)_{i=1}^n$ are non-negative integers and $\sum_{i=1}^n z_i = t$, $\sum_{i=1}^n f(z_i + \ell) - \sum_{i=1}^n f(z_i)$ is maximized when $(z_i)_{i=1}^n = ((q + 1)^r, q^{n-r})$.*

Proof. The function $\delta(z) = f(z + \ell) - f(z)$ is a strictly concave function for every $\ell > 0$, and every (non-increasing) sequence of n integers with sum t majorises the sequence $(z_i)_{i=1}^n = ((q + 1)^r, q^{n-r})$. Hence, the result follows immediately by Karamata's inequality. \square

This allows to give an alternative proof to the following simple result on the entropy of graphs with given order n , already found in [52, Theorem 1], which we will use as a base case in the proof of Theorem 12.

Proposition 22. *Among trees of order n , $h(G)$ is maximised by the star S_n and minimized by the path P_n . The second smallest value is attained by any tree with 3 leaves.*

Proof. The degree sequence of a tree satisfies $\sum_{1 \leq i \leq n} d_i = 2(n-1)$ by the hand shaking lemma. Since a tree has at least 2 vertices of degree 1, and all degrees are at least 1, the degree sequence majorises $(2, 2, \dots, 2, 1, 1)$ and is majorised by $(n-1, 1, 1, \dots, 1)$. The path and the star are the only graphs with these degree sequences. Any tree that is not a path has at least 3 vertices of degree 1. If it has more than 3 vertices of degree 1, its degree sequence majorises $(3, 2, 2, 2, \dots, 2, 1, 1, 1)$, the degree sequence of a tree with 3 leaves. So the conclusion follows from Corollary 20. \square

Note that this statement is also true for the function $h_g(G) = \sum_i g(d_i)$ for any convex function g , in particular this proves [81, Conjecture 1].

We remark here that it will be sufficient to focus on connected graphs.

Proposition 23. *When omitting the isolated vertices, the graph maximizing $h_c(G)$ among all graphs of size m is a connected graph.*

Proof. Identifying two vertices u and v in different components with strictly positive degrees d_u, d_v leads to an increase of the value $h_c(G)$ since f_c is a strictly convex function, i.e., $f_c(d_u + d_v) + f_c(0) > f_c(d_u) + f_c(d_v)$. \square

3.2 Minimum entropy graphs given size

3.2.1 Main proof

Lemma 24. *Let G be a graph maximising $h(G)$ among all graphs of size m . Then G is a threshold graph.*

Proof. Let G be the graph G maximising $h(G)$ among all graphs of size m . By Proposition 23, G is connected. Furthermore, G clearly maximizes $h(G)$ among all graphs with size m and order n as well, where n is the order of G . By [221, Theorem 4], we know that the statement holds for connected graphs with given order and size. Therefore G is a threshold graph. \square

Lemma 25. *Let $k \geq 3$. For every $m \geq \binom{k}{2}$, $h(\mathcal{C}(m, k)) > h(\mathcal{C}(m, k-1))$.*

The proof of the lemma is quite computational and therefore postponed to the subsequent section.

Theorem 26. *Let k and m be integers with $m = \binom{k-1}{2} + a(k-1) + b$ for some integers a, b satisfying $a > 0$ and $0 \leq b \leq k - 2$. Among all threshold graphs G with clique number $\omega(G) = k$ and size m , the graph $\mathcal{C}(m, k)$ is the unique graph maximising h , up to isomorphism.*

Proof. We will prove this by a double induction on k and m .

When $k = 2$, the graph is a tree and the unique extremal graph is a star, as mentioned in Proposition 22. So the base case with $k = 2$ is true for every $m \geq 1$. Now we assume that the statement of the theorem has been proven for all values from 2 to $k - 1$ for every choice of m and prove by induction that this hypothesis is also true for clique number k . When $m = \binom{k}{2}$, the graph $\mathcal{C}(m, k) = K_k$ is the unique one and thus extremal. So now assume it is proven when the size is between $\binom{k}{2}$ and $m - 1$ and let G be a threshold graph with clique number k and size m . Let v be a vertex in the clique whose degree is precisely $k - 1$. Such a vertex exists because G is a threshold graph: if all vertices in the clique had degree k or larger, then there would be a vertex in the stable set connected to all of them, and so G would have a clique of order $k + 1$, but this is impossible since its clique number is k .

Let d_1, d_2, \dots, d_{k-1} be the degrees of the $k - 1$ neighbours of v . In that case, we have

$$h(G) = h(G \setminus v) + f(k - 1) + \sum_{j=1}^{k-1} f(d_j) - \sum_{j=1}^{k-1} f(d_j - 1) \quad (3.1)$$

When $m < \binom{k}{2} + k - 1$, the clique number $\omega(G \setminus v) = k - 1$ and since the result is known for clique number $k - 1$, we have $h(G \setminus v) \leq h(\mathcal{C}(m - (k - 1), k - 1))$. When $m \geq \binom{k}{2} + k - 1$, by the induction hypothesis and Lemma 25 we know that $h(G \setminus v) \leq h(\mathcal{C}(m - (k - 1), k))$. By Lemma 21 the other part of the upper bound in (3.1) is maximized if and only if all d_i differ at most 1. Thus, we conclude that the maximum is precisely $h(\mathcal{C}(m, k))$, and that $\mathcal{C}(m, k)$ is the unique extremal graph. Hence, by induction, the statement is true for every m . Finally, by complete induction, we conclude that the theorem is true for every value of k . \square

We can now proceed with the proof of Theorem 12.

Proof of Theorem 12. Let G be the graph that minimises the first degree-based entropy among all graphs with m edges. By Lemma 24, G is a threshold graph. Let k be its clique number. Note that this implies that $m \geq \binom{k}{2}$. By

Theorem 26, G must be the generalised colex graph $\mathcal{C}(m, k)$. Now, Lemma 25 implies that k is as large as possible, and thus $m < \binom{k+1}{2}$. By Remark 17, we conclude that G is indeed the colex graph $\mathcal{C}(m)$. \square

3.2.2 Proof of Lemma 25

Write

$$m = \binom{k-1}{2} + a(k-1) + b,$$

with $a \geq 0$ and $0 \leq b \leq k-2$, and

$$m = \binom{k-2}{2} + a'(k-2) + b', \quad (3.2)$$

with $a' \geq 2$ and $0 \leq b' \leq k-3$. Then the degree sequence of $C(m, k)$ is

$$((k-1+a)^b, (k-2+a)^{k-1-b}, (k-1)^a, b),$$

and the degree sequence of $C(m, k-1)$ is

$$((k-2+a')^{b'}, (k-3+a')^{k-2-b'}, (k-2)^{a'}, b').$$

Remark that since

$$\begin{aligned} m &= \binom{k-1}{2} + a(k-1) + b \\ &= \binom{k-2}{2} + (k-2) + a(k-2) + a + b, \end{aligned}$$

we have

$$\begin{aligned} a' &= a + 1 + \left\lfloor \frac{a+b}{k-2} \right\rfloor \quad \text{and} \\ b' &\equiv a + b \pmod{k-2}. \end{aligned}$$

We divide the proof of the Lemma in three steps aimed at reducing the general case to a more manageable one:

1. We begin by showing that if Lemma 25 is true for all m for which $b' = 0$ in equation (3.2), then it holds for all m , i.e., for every $0 \leq b' \leq k-3$ as well. We do so by showing that if a counterexample does exist for some m with $1 \leq b' \leq k-3$, then there is a counterexample with size m' that can be represented in such a way that the corresponding b' (in

equation (3.2)) is 0.

2. We then show that if Lemma 25 holds in the case $b' = 0$ and $2 \leq a' \leq k$, then it holds for all values of a' . We do so by proving that if our claim is true for a certain size m , then so it is for $m' = m + (k - 1)(k - 2)$.
3. We conclude by proving that the claim holds for values m that be written as in (3.2) with $2 \leq a' \leq k$ and $b' = 0$.

In the last two steps, we are careful with a reduction to $m = \binom{k}{2} - 1$, since then $\mathcal{C}(m, k) = \mathcal{C}(m, k - 1)$. In the following, we state these steps in the form of claims, which are proven separately³.

Claim 27. *It is sufficient to prove Lemma 25 for the case $b' = 0$ (and every $a' \geq 3$).*

Proof. If $h(\mathcal{C}(m, k)) \leq h(\mathcal{C}(m, k - 1))$ and $k - 2 > b' \geq b$, then $h(\mathcal{C}(m + 1, k)) \leq h(\mathcal{C}(m + 1, k - 1))$ as well, since

$$\begin{aligned}
& h(\mathcal{C}(m + 1, k - 1)) - h(\mathcal{C}(m, k - 1)) \\
&= f(k - 2 + a') - f(k - 3 + a') + f(b' + 1) - f(b') \\
&\geq f(k - 1 + a) - f(k - 2 + a) + f(b + 1) - f(b) \\
&= h(\mathcal{C}(m + 1, k)) - h(\mathcal{C}(m, k)).
\end{aligned}$$

So we can repeat this until $b' + 1 = k - 2$. But then $m + 1 = \binom{k-2}{2} + (a' + 1)(k + 2)$, so it is sufficient to verify a case with $b' = 0$.

If $h(\mathcal{C}(m, k)) \leq h(\mathcal{C}(m, k - 1))$ and $0 < b' < b \leq k - 2$ (note that this implies that $a' \geq 3$), then $h(\mathcal{C}(m - 1, k)) \leq h(\mathcal{C}(m - 1, k - 1))$ as well. For this note that $\frac{b}{b'} \geq \frac{k-2}{k-3}$ and $\frac{k-3+a'}{k-2+a} \leq \frac{k-1}{k-2} < \frac{k-2}{k-3}$, the latter being true since $a' - 1 = a + \left\lfloor \frac{a+b}{k-2} \right\rfloor \leq a + \frac{a+k-2}{k-2}$. This implies that

$$\begin{aligned}
& h(\mathcal{C}(m, k)) - h(\mathcal{C}(m - 1, k)) \\
&= f(k - 1 + a) - f(k - 2 + a) + f(b) - f(b - 1) \\
&\geq f(k - 2 + a') - f(k - 3 + a') + f(b') - f(b' - 1) \tag{3.3} \\
&= h(\mathcal{C}(m, k - 1)) - h(\mathcal{C}(m - 1, k - 1)).
\end{aligned}$$

³Note that one can check the three claims in reverse order as well. In that case, one is increasing the range of values of m for which it is known that Lemma 25 is true.

Inequality (3.3) is true by the following two inequalities

$$\begin{aligned}
& f(b) - f(b-1) - (f(b') - f(b'-1)) \\
&= \int_0^1 \log \left(\frac{b-1+t}{b'-1+t} \right) dt \\
&\geq \int_0^1 \log \left(\frac{b}{b'} \right) dt \\
&= \log \left(\frac{b}{b'} \right) \\
&\geq \log \left(\frac{k-2}{k-3} \right)
\end{aligned}$$

and

$$\begin{aligned}
& f(k-2+a') - f(k-3+a') - (f(k-1+a) - f(k-2+a)) \\
&= \int_0^1 \log \left(\frac{k-3+a'+t}{k-2+a+t} \right) dt \\
&\leq \int_0^1 \log \left(\frac{k-3+a'}{k-2+a} \right) dt \\
&\leq \log \left(\frac{k-2}{k-3} \right).
\end{aligned}$$

◇

Claim 28. *It is sufficient to prove Lemma 25 for the case $2 \leq a' \leq k-1$ and $b' = 0$.*

Proof. It is sufficient to prove that for every $m \geq \binom{k}{2} - 1$ and $m' = m + (k-1)(k-2)$, we have $h(\mathcal{C}(m', k)) - h(\mathcal{C}(m', k-1)) \geq h(\mathcal{C}(m, k)) - h(\mathcal{C}(m, k-1))$ and that this inequality is strict when $m = \binom{k}{2} - 1$.

Write $m = \binom{k-2}{2} + a'(k-2) = \binom{k-1}{2} + a(k-1) + b$ and let $m' = m + (k-1)(k-2) = \binom{k-2}{2} + (a' + k-1)(k-2)$. Then

$$\begin{aligned}
& h(\mathcal{C}(m', k)) - h(\mathcal{C}(m, k)) \\
&= b(f(k-1+a+k-2) - f(k-1+a)) \\
&\quad + (k-1-b)(f(k-2+a+k-2) - f(k-2+a)) + (k-2)f(k-1) \\
&\geq (k-1)(f(k-2+a+k-2) - f(k-2+a)) + (k-2)f(k-1)
\end{aligned}$$

since $f(k-1+a+k-2) - f(k-1+a) > f(k-2+a+k-2) - f(k-2+a)$.

Also,

$$\begin{aligned} & h(\mathcal{C}(m', k-1)) - h(\mathcal{C}(m, k-1)) \\ &= (k-2) (f(k-3+a'+k-1) - f(k-3+a')) + (k-1)f(k-2). \end{aligned}$$

Here $k-3+a' \leq \frac{k-1}{k-2}(k-2+a)$. Hence, we have

$$\begin{aligned} & f(k-2+a+k-2) - f(k-2+a) - (k-2) \\ &= \int_0^{k-2} \log(k-2+a+t) dt \\ &= \frac{k-2}{k-1} \int_0^{k-1} \log\left(k-2+a+\frac{k-2}{k-1}t\right) dt \end{aligned}$$

and

$$\begin{aligned} & f(k-3+a'+k-1) - f(k-3+a') - (k-1) \\ &= \int_0^{k-1} \log(k-3+a'+t) dt \\ &\leq \int_0^{k-1} \log\left(\frac{k-1}{k-2}(k-2+a)+t\right) dt \\ &= \int_0^{k-1} \log\left(k-2+a+\frac{k-2}{k-1}t\right) + \log\left(\frac{k-1}{k-2}\right) dt. \end{aligned}$$

So we conclude

$$h(\mathcal{C}(m', k)) - h(\mathcal{C}(m, k)) \geq h(\mathcal{C}(m', k-1)) - h(\mathcal{C}(m, k-1)),$$

which is equivalent to

$$h(\mathcal{C}(m', k)) - h(\mathcal{C}(m', k-1)) \geq h(\mathcal{C}(m, k)) - h(\mathcal{C}(m, k-1)).$$

Furthermore the difference is strict when $b > 0$, which is the case when $m = \binom{k}{2} - 1$. \diamond

Claim 29. *Lemma 25 is true when $2 \leq a' \leq k$ and $b' = 0$, i.e., when $m = \binom{k-2}{2} + (a+2)(k-2) = \binom{k-1}{2} + a(k-1) + (k-2-a)$ for some $0 \leq a \leq k-2$.*

Proof. For this, we need to prove that

$$\begin{aligned} & (k-2)f(k+a-1) + (a+2)f(k-2) \\ &\leq (k-2-a)f(k+a-1) + (a+1)f(k-2+a) + af(k-1) + f(k-2-a) \end{aligned}$$

and that this is strict when $a > 0$. The latter is equivalent to

$$\begin{aligned} & a[f(k+a-1) - f(k+a-2) - f(k-1) + f(k-2)] \\ & \leq f(k-2+a) + f(k-2-a) - 2f(k-2). \end{aligned} \quad (3.4)$$

We rewrite the right hand side of inequality (3.4) as a telescoping sum:

$$\begin{aligned} & f(k-2+a) + f(k-2-a) - 2f(k-2) \\ & = \sum_{i=1}^a [f(k-2+i) - f(k-3+i) - f(k-2-a+i) + f(k-3-a+i)]. \end{aligned}$$

The inequality now follows from the fact that

$$f(x+1) - f(x) - f(x+1-a) + f(x-a) = \int_0^1 \log\left(\frac{x+t}{x-a+t}\right) dt$$

is a function that is strictly decreasing in x (where the domain is $x \geq a$). \diamond

This concludes the proof of Lemma 25. \square

3.3 Minimum entropy graphs given order and size

By [221, Theorem 4], we know that every graph maximizing $h(G)$ among all (n, m) -graphs is a threshold graph. This implies in particular that it has a universal vertex v , i.e., a vertex adjacent to all other vertices and hence with degree $n-1$. Now, the graph $G \setminus v$, obtained by removing v and all its incident edges from G , is a $(n-1, m-n+1)$ -graph. Considering that $d_G(u) = d_{G \setminus v}(u) + 1$ for every vertex $u \neq v$, we note that it is sufficient to find the $(n-1, m-n+1)$ -graph maximizing $h_1(G)$, where $h_1(G)$ is formed by taking into account that the original degrees are larger by one. We extend this idea towards the setting where there are c universal vertices initially. Then, we compute the extremal graphs maximizing the related function $h_c(G)$ given only the size (and fixed large order essentially). We do so by induction.

We begin by computing some extremal graphs for small size. These are the base cases for the induction.

3.3.1 Graph with small size

We compute the extremal graphs maximizing $h_1(G)$ for $m \leq 10$ and for $h_c(G)$ with $c \geq 2$ for $m \leq 6$.

Lemma 30. For $m \leq 10$, among all graphs with m edges, $h_1(G)$ is maximized by

$$G = \begin{cases} S_{m+1} & \text{if } m \notin \{3, 5, 6\} \\ K_3 & \text{if } m = 3 \\ K_4^- \text{ and } S_6 & \text{if } m = 5 \\ K_4 & \text{if } m = 6. \end{cases}$$

Proof. A computer program can verify this claim⁴. Since $h_1(G)$ only depends on the degree sequence of the graph, for a given $m \leq 10$, it is enough to list all degree sequences of graphs of size m and then compute h_1 for each sequence. To list all degree sequences, it is sufficient to list all integer partitions of $2m$ and then to establish which of these are valid degree sequences using one of several existing criteria (see, e.g., [194]). For example, one can use the function `parts()` from the R-package `partitions` [104] to list all partitions of $2m$ and check which ones are degree sequences using `is_graphical()` from the R-package `igraph` [66]. \square

Lemma 31. For $c \geq 2$ and $m \leq 6$, among all graphs with m edges, $h_c(G)$ is maximized by

$$G = \begin{cases} S_{m+1} & \text{if } m \neq 3 \\ K_3 & \text{if } m = 3. \end{cases}$$

Remark that in this Lemma, one has to take into account isolated vertices when comparing graphs with different order.

Proof. For $m \in \{1, 2\}$ nothing needs to be done, as there is only one connected graph of size m . When $m = 3$, there are precisely 3 connected graphs and we observe that

$$\begin{aligned} h_c(P_4) = 2f_c(1) + 2f_c(2) &< h_c(S_4) = 3f_c(1) + f_c(3) \\ &< h_c(K_3) = 3f_c(2) + f_c(0). \end{aligned}$$

The first inequality follows from strict convexity of the function f_c . The second inequality holds because Δ_c (Definition 15) is strictly concave, and thus $\Delta_c(3) + \Delta_c(1) < 2\Delta_c(2)$.

By Karamata's inequality, it is sufficient to consider the degree sequences of graphs with size m that are not majorized by the degree sequences of other

⁴https://github.com/MatteoMazzamurro/extrema-graph-entropy/blob/main/minimal_entropy_small_size.R

such graphs.

For $m = 4$, these non-majorized degree sequences are

$$v_1^4 = \{4, 1, 1, 1, 1\} \text{ and } v_2^4 = \{3, 2, 2, 1, 0\}.$$

For $m = 5$, they are

$$v_1^5 = \{5, 1, 1, 1, 1, 1\}, v_2^5 = \{4, 2, 2, 1, 1, 0\}, \text{ and } v_3^5 = \{3, 3, 2, 2, 0, 0\}.$$

For $m = 6$, the sequences are

$$v_1^6 = \{6, 1, 1, 1, 1, 1, 1\}, v_2^6 = \{5, 2, 2, 1, 1, 1, 0\},$$

$$v_3^6 = \{4, 3, 2, 2, 1, 0, 0\}, \text{ and } v_4^6 = \{3, 3, 3, 3, 0, 0, 0\}.$$

Now we verify that, $h_c(v_j^m) = \sum_i f_c(d_i)$ is always maximized by the first degree sequence ($j = 1$ for all the above values of m).

For $4 \leq m \leq 6$, we have

$$\begin{aligned} h_c(v_1^m) - h_c(v_2^m) &= \Delta_c(m) + \Delta_c(1) - 2\Delta_c(2) \\ &\geq \int_{c-1}^c \log \left(\frac{(t+1)(t+4)}{(t+2)^2} \right) dt > 0, \end{aligned}$$

the last inequality is true since $(t+1)(t+4) > (t+2)^2$ whenever $t \geq 1$.

For $m \in \{5, 6\}$ we analogously have

$$\begin{aligned} h_c(v_1^m) - h_c(v_3^m) &= \Delta_c(m) + \Delta_c(m-1) + 2\Delta_c(1) - \Delta_c(3) - 3\Delta_c(2) \\ &\geq \int_{c-1}^c \log \left(\frac{(t+5)(t+4)(t+1)^2}{(t+3)(t+2)^3} \right) dt \\ &> 0, \end{aligned}$$

the last inequality being true since $(t+1)(t+4) > (t+2)^2$ for $t > 0$ and $(t+5)(t+1) \geq (t+3)(t+2)$ for $t \geq 1$, and thus $(t+5)(t+4)(t+1)^2 > (t+3)(t+2)^3$ whenever $t \geq 1$.

For the final case, we have

$$\begin{aligned} h_c(v_1^6) - h_c(v_4^6) &= \Delta_c(6) + \Delta_c(5) + \Delta_c(4) + 3\Delta_c(1) - 3\Delta_c(3) - 3\Delta_c(2) \\ &= \int_{c-1}^c \log \left(\frac{(t+6)(t+5)(t+4)(t+1)^3}{(t+3)^3(t+2)^3} \right) dt \\ &> 0, \end{aligned}$$

where the inequality can be verified as follows. When $c = 2$, this can be computed⁵. For $c \geq 3$, this is due to

$$(t+6)(t+5)(t+4)(t+1)^3 > (t+3)^3(t+2)^3$$

for $t \geq 2$. Indeed, on the one hand $(t+6)(t+1)^2 > (t+2)^3$ is equivalent to $2t^2 + t - 2 > 0$, which holds for $t > \frac{-1+\sqrt{17}}{4} \approx 0.78$, and thus for $t \geq 1$; on the other hand $(t+5)(t+4)(t+1) > (t+3)^3$ is equivalent to $t^2 + 2t - 7 > 0$, which holds for $t > -1 + 2\sqrt{2} \approx 1.83$, and thus for $t \geq 2$.

Therefore the extremal degree sequences for $m \leq 6$, $m \neq 3$ correspond to the star S_{m+1} . \square

3.3.2 Main proof

We prove the following theorem that gives the precise characterization of extremal graphs for $h_c(G)$ where $c \geq 1$ is an integer (for $c = 0$, this corresponds to the case when only the size is given, which was treated previously).

Some technical computations are postponed until the following section not to distract from the main argument.

Theorem 32. *Among all graphs with m edges, $h_1(G)$ is maximized by*

$$G = \begin{cases} S_{m+1} & \text{if } m \notin \{3, 5, 6\} \\ K_3 & \text{if } m = 3 \\ K_4^- \text{ and } S_6 & \text{if } m = 5 \\ K_4 & \text{if } m = 6. \end{cases}$$

For any $c \geq 2$, among all graphs with m edges and $n > m$ vertices, $h_c(G)$ is maximized by

$$G = \begin{cases} S_{m+1} & \text{if } m \neq 3 \\ K_3 & \text{if } m = 3. \end{cases} \quad (3.5)$$

Proof. Assume we know the extremal graphs with size at most $m - 1$. By Lemmas 30 and 31, this has been done for $m \leq 6$ and $m \leq 10$ when $c = 1$. So we assume $m \geq 7$ (or $m \geq 11$ if $c = 1$). Let G be an extremal graph with size m for which the minimum (non-zero) degree is equal to b . The latter implies that there are at least $b + 1$ vertices with degree at least b and thus $m \geq \binom{b+1}{2}$.

⁵It is approximately 0.0908

Let v be a vertex with degree b and let d_1, d_2, \dots, d_b be the degrees of the neighbours of v .

If $b = 1$, we have

$$\begin{aligned} h_c(G) &= h_c(G \setminus v) + f_c(1) - f_c(0) + \Delta_c(d_1) \\ &\leq h_c(S_m) + f_c(1) - f_c(0) + \Delta_c(m) \\ &= h_c(S_{m+1}) \end{aligned}$$

and equality occurs if and only if $G = S_{m+1}$.

Now assume $b \geq 2$. Note that $\sum_{i=1}^b d_i \leq m + \binom{b}{2}$ since every edge which is not part of the subgraph $G[N(v)]$ induced by the neighbours of v can be counted at most once.

Since Δ_c is strictly concave, we have

$$\begin{aligned} h_c(G) - h_c(G \setminus v) &= f_c(b) - f_c(0) + \sum_{i=1}^b \Delta_c(d_i) \\ &\leq f_c(b) - f_c(0) + b \cdot \Delta_c\left(\frac{m + \binom{b}{2}}{b}\right) \\ &:= LHS(m, b, c). \end{aligned}$$

On the other hand, we also have

$$\begin{aligned} h_c(S_{m+1}) - h_c(S_{m-b+1}) &= f_c(m) - f_c(m-b) + b\Delta_c(1) \\ &:= RHS(m, b, c). \end{aligned}$$

By the computations performed in the following section, the first is smaller than the second, i.e., $LHS(m, b, c) < RHS(m, b, c)$. Hence

$$h_c(G) < h_c(S_{m+1}) - h_c(S_{m-b+1}) + h_c(G \setminus v).$$

Now, $G \setminus v$ has $m - b$ edges. Here $m - b \geq 4$ (for $c \geq 2$) and $m - b \geq 7$ (for $c = 1$). Due to Lemmas 30 and 31, we have $h_c(G \setminus v) \leq h_c(S_{m-b+1})$. Therefore we conclude that $h_c(G) < h_c(S_{m+1})$.

By complete induction, we have the whole characterization. \square

We are now ready to prove Theorem 13. Indeed, when $n \leq m \leq 2n - 3$, the extremal (n, m) -graph minimizing the entropy is such that, deleting its universal vertex, one obtains the $(n - 1, m - n + 1)$ -graph G maximizing $h_1(G)$,

as described in Theorem 32.

Proof of Theorem 13. By [221, Theorem 4], we know the extremal (n, m) -graph is a threshold graph. This implies in particular that it has a universal vertex v with degree $n - 1$. Now $G' = G \setminus v$ is a $(n - 1, m - n + 1)$ -graph. Taking into account that $d_G(u) = d_{G'}(u) + 1$ for every vertex $u \neq v$, we note that

$$h(G) = f(n - 1) + h_1(G').$$

Now, since $m - n + 1 \leq n - 2$, we note that the extremal structure for G' is determined in Theorem 32 and the conclusion is immediate. \square

3.3.3 Technical lemmas

In this section, we complete the proof of Theorem 32. We show that

$$LHS(m, b, c) = f_c(b) - f_c(0) + b \cdot \left(f_c \left(\frac{m + \binom{b}{2}}{b} \right) - f_c \left(\frac{m + \binom{b}{2}}{b} - 1 \right) \right)$$

and

$$RHS(m, b, c) = f_c(m) - f_c(m - b) + b \cdot (f_c(1) - f_c(0))$$

satisfy $LHS(m, b, c) < RHS(m, b, c)$ for every $b \geq 2$ and $m \geq \binom{b+1}{2}$ whenever $m \geq 7$ and $c \geq 1$, or $m \geq 4$ and $c \geq 2$.

We do this by means of the following lemmas. In Lemma 33, we show that for fixed b and c , it is sufficient to prove the inequality for the smallest m in the range. After that, we prove it in the cases for which $m = \binom{b+1}{2}$ in Lemma 34 and for the remaining cases in Lemma 35.

The proofs are mainly computational and there are alternative computations that lead to the same conclusion⁶.

Lemma 33. *Fix $b \geq 2$ and $c \geq 1$. Then $RHS(m, b, c) - LHS(m, b, c)$ is an increasing function in m .*

Proof. We want to prove that the derivative of this quantity with respect to m is positive. To compute the derivative, taking into account the chain rule

⁶See, for example, <https://arxiv.org/abs/2205.03357> for an alternative proof of Lemma 34.

and $\frac{d}{dx}f_c(x) = \log(x+c) + \log(e)$, we have that

$$\begin{aligned} & \frac{d}{dm} (RHS(m, b, c) - LHS(m, b, c)) \\ &= \log\left(\frac{m+c}{m-b+c}\right) - \log\left(\frac{m + \binom{b}{2} + bc}{m + \binom{b}{2} + bc - b}\right) > 0, \end{aligned}$$

where the inequality follows from the fact that whenever $0 < b < y < z$, we have $\frac{y}{y-b} > \frac{z}{z-b}$. Here it is enough to take $y = m+c$ and $z = m + \binom{b}{2} + bc$. \square

Lemma 34. Fix $b \geq 2$ and $c \geq 1$. Let

$$LL(b, c) = (b+1)f_c(b) - f_c(0) - bf_c(b-1)$$

and

$$RL(b, c) = f_c\left(\binom{b+1}{2}\right) - f_c\left(\binom{b}{2}\right) + b \cdot (f_c(1) - f_c(0)).$$

Then

$$LL(b, c) < RL(b, c)$$

if $c = 1$ and $b \geq 4$, or $c \geq 2$ and $b \geq 3$.

Proof. The cases $1 \leq c \leq 3$ can be verified directly using the formulae: solving numerically the resulting inequalities in the variable b , one finds that the inequality holds as long as $b > 3.24$, $b > 2.53$, and $b > 2.35$ for $c = 1$, $c = 2$, and $c = 3$, respectively⁷.

For $c \geq 4$ and $b \geq 3$, write

$$\begin{aligned} RL(b, c) &= f_c\left(\binom{b}{2} + b\right) - f_c\left(\binom{b}{2}\right) + b\Delta_c(1) \\ &= \sum_{i=1}^b \left[f_c\left(\binom{b}{2} + i\right) - f_c\left(\binom{b}{2} + i - 1\right) \right] + b\Delta_c(1) \\ &= \sum_{i=1}^b \Delta_c\left(\binom{b}{2} + i\right) + b\Delta_c(1), \end{aligned}$$

⁷https://github.com/MatteoMazzamurro/extrema-graph-entropy/blob/main/lemma_4_base_cases.R.

and

$$\begin{aligned}
LL(b, c) &= b(f_c(b) - f_c(b-1)) + f_c(b) - f_c(0) \\
&= b\Delta_c(b) + f_c(b) - f_c(0) \\
&= b\Delta_c(b) + \sum_{i=1}^b [f_c(i) - f_c(i-1)] \\
&= b\Delta_c(b) + \sum_{i=1}^b \Delta_c(i).
\end{aligned}$$

Then

$$\begin{aligned}
RL(b, c) - LL(b, c) &= \sum_{i=1}^b \left[\Delta_c \left(\binom{b}{2} + i \right) - \Delta_c(i) \right] - b(\Delta_c(b) - \Delta_c(1)) \\
&> b \left[\Delta_c \left(\binom{b}{2} + b \right) - \Delta_c(b) \right] - b(\Delta_c(b) - \Delta_c(1)) \quad (3.6) \\
&= b \left[\Delta_c \left(\binom{b+1}{2} \right) + \Delta_c(1) - 2\Delta_c(b) \right]
\end{aligned}$$

where inequality (3.6) follows from $b \geq 2$ and the strict concavity of $\Delta_c(x)$. Then, by Definition 15,

$$\begin{aligned}
&RL(b, c) - LL(b, c) \\
&\geq b \int_{c-1}^c \left[\log \left(t + \binom{b+1}{2} \right) + \log(t+1) - 2\log(t+b) \right] dt
\end{aligned}$$

For the integral to be positive, it is enough that, for $c-1 < t < c$,

$$\left(t + \binom{b+1}{2} \right) (t+1) - (t+b)^2 > 0,$$

which is equivalent to

$$t(b-2)(b-1) > b(b-1). \quad (3.7)$$

Now, since $b \geq 3$, inequality (3.7) holds if and only if $t > \frac{b}{b-2}$. Furthermore, $b \geq 3$ also implies $\frac{b}{b-2} \leq 3$. But $c \geq 4$ so $t > c-1 = 3 \geq \frac{b}{b-2}$. Therefore $RL(b, c) > LL(b, c)$ for $c \geq 4$ and $b \geq 3$ as well. \square

Lemma 35. *It is true that $LHS(7, 3, 1) < RHS(7, 3, 1)$ and $LHS(7, 2, 1) <$*

$RHS(7, 2, 1)$. For any $c \geq 2$, it is true that $LHS(4, 2, c) < RHS(4, 2, c)$.

Proof. By direct computation, we find $RHS(7, 3, 1) - LHS(7, 3, 1) \approx 0.26$ and $RHS(7, 2, 1) - LHS(7, 2, 1) \approx 0.52$, and thus $LHS(7, 3, 1) < RHS(7, 3, 1)$ and $LHS(7, 2, 1) < RHS(7, 2, 1)$. $LHS(4, 2, c) < RHS(4, 2, c)$ is equivalent to

$$2\Delta_c(2.5) + \Delta_c(2) < \Delta_c(4) + \Delta_c(3) + \Delta_c(1).$$

This is true for every $c \geq 2$ since

$$\int_{c-1}^c \log \left(\left(t + \frac{5}{2} \right)^2 (t+2) \right) dt < \int_{c-1}^c \log ((t+4)(t+3)(t+1)) dt,$$

as $(t + \frac{5}{2})^2 (t+2) < (t+4)(t+3)(t+1)$ for every $t \geq 1$. Indeed, expanding this inequality, it simplifies to $t^2 + \frac{11}{4}t - \frac{1}{2} > 0$, which holds for $t > \frac{-11+3\sqrt{17}}{8} \approx 0.17$, and thus for every $t \geq 1$. \square

3.3.4 Numerical observations

In the previous argument, we have seen that, apart from few exceptions for small values of m , the graph $G_{n,m}^1$ with degree sequence

$$(n-1, m-n+2, 2^{m-n+1}, 1^{2n-m-3})$$

minimises the first degree-based entropy. However, $G_{n,m}^1$ is only well defined for $m \leq 2n-3$. Extending the characterisation of (n, m) -graphs minimising the entropy to the case $m > 2n-3$ is not a trivial challenge. A crucial aspect in the proof of Theorem 13 is that, when $m \leq 2n-3$, the minimum entropy graph contains exactly one universal vertex. Deleting such vertex, the problem is reduced to that of maximising h_1 amongst graphs with unconstrained order. When $m \geq 2n-2$, instead, the graph may contain more than one universal vertex, so the order n remains a potentially influential factor even after one universal vertex is deleted.

A computational approach allows to determine the minimum entropy graphs for some values of n and m . As proved by [221, Theorem 4], the (n, m) -graph yielding minimum first degree-based entropy must be a threshold graph $K_a T$ for some value of a .

Now, fix a value of a . We know that any threshold graph $K_a T$ has a universal vertex $v^* \in K_a$. Thus, by construction, any $K_a T$ graph contains the subgraph G , having the same vertex set as $K_a T$, and having as edge set all

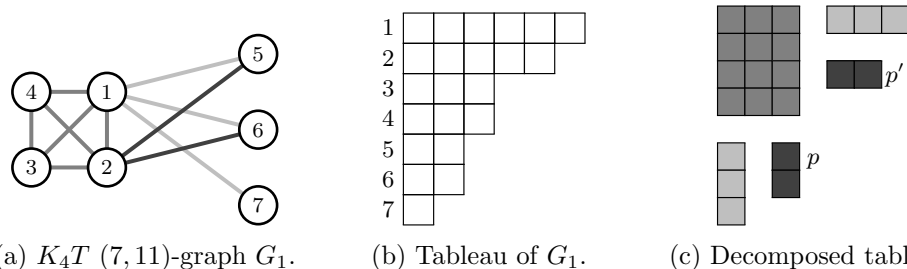


Figure 3.3: K_aT (n, m)-graph G_1 with $a = 4$, $n = 7$, $m = 11$, with corresponding Young tableau. In (b), the numbers on the left correspond to nodes in G_1 . In (c), the shade of gray represents different components of the graph. The lightest boxes correspond to the edges connecting nodes in T to $v_1 = v^*$, the intermediate ones to the edges in K_a , and the darkest ones to the remaining edges between K_a and T . The darker boxes are the partitions p and its conjugate p' , either of which, together with a , n , and m , define the graph uniquely.

edges in the clique K_a and, for each vertex in T , the edge connecting it to v^* .

As the edges in G are in common among all K_aT graphs, listing all K_aT graphs reduces to listing all ways in which the remaining $r = m - \frac{n-1+(a-1)^2+n-a}{2}$ edges can be distributed to connect vertices in K_a and those in T . Now, in a threshold graph, the neighbourhoods of all vertices in T are nested sets of vertices in K_a , and thus, they are completely determined by the degree of the corresponding vertex in T . In other words, it is enough to provide the degree sequence of T and the structure of the graph K_aT is uniquely defined. This degree sequence is in turn completely determined by a partition $p = (p_1, \dots, p_{n-a})$ of r into $n-a$ parts (r being the number of remaining edges and $n-a$ the number of vertices in T). Note that p can be any partition of r into $n-a$ parts as long as all its entries are smaller than $a-1$, that is, the number of vertices in K_a excluding the fully connected node v^* . See Fig. 3.3 for a visual construction.

For a suitable partition $p = (p_1, \dots, p_{n-a})$, with $p_1 \geq p_2 \geq \dots \geq p_{n-a}$, the degree sequence s of the corresponding graph can be retrieved as follows. Write the degree sequence of G as

$$s_G = (n-1) \oplus s_G^K \oplus s_G^T,$$

where $s_G^K = ((a-1)^{a-1})$, and $s_G^T = (1^{n-a})$. Consider the vector

$$w = (w_1, \dots, w_{a-1}),$$

where $w_i = |\{j | p_j \geq i\}|$. Then

$$s = (n - 1) \oplus (s_G^K + w) \oplus (s_G^T + p).$$

Remark that w is nothing other than the conjugate p' of p , filled with zeros at the end so to match the length of s_G^K .

From this construction, it is clear that finding all $K_a T$ graphs for a given a corresponds to finding all partitions p of r into at most $n - a$ parts, such that all parts are less than or equal to $a - 2$, or equivalently⁸, all their conjugates p' , i.e., all the partitions of r into at most $a - 2$ parts, such that all parts are less than or equal to $n - a$.

We thus have an algorithm to compute all degree sequences for $(n, m) - K_a T$ -graphs for fixed values of a , n , and m . Now, although we know from [221] that the graph yielding minimum first degree-based entropy must be of the form $K_a T$ for *some* a , we do not know a priori what value a should take. The possible values of a depend on both n and m . A $K_a T$ -graph with n nodes that is not also a $K_{a+1} T$ graph contains at most $(a - 1) \binom{n - \frac{a}{2}}{2}$ edges. Indeed, the clique K_a contains exactly $\frac{a(a-1)}{2}$ edges, no edges exist among vertices in T and each of the $n - a$ vertices in T is connected to at most $a - 1$ vertices in K_a (if a vertex $t \in T$ were connected to all a vertices in K_a , then $K_a \cup \{t\}$ would form an $(a + 1)$ -clique, and we would have a $K_{a+1} T$ -graph). For m fixed, if we let

$$a_{min} = \min \left\{ a : (a - 1) \binom{n - \frac{a}{2}}{2} \geq m \right\},$$

then $a \geq a_{min}$. Similarly, a $K_a T$ graph has at least $\frac{a(a-1)}{2} + n - a$ edges (the number of edges in G as described above). If we let

$$a_{max} = \max \left\{ a : \frac{a(a-1)}{2} + n - a \leq m \right\},$$

then $a \leq a_{max}$. Then, to find all possible candidates for minimum entropy graph it suffices to find the degree sequences of all $(n, m) - K_a T$ -graphs for $a_{min} \leq a \leq a_{max}$. Solving the quadratic expressions for a_{min} and a_{max} with

⁸Whilst mathematically equivalent, computationally there is a difference, because one still needs to find a large number of partitions and select only the suitable ones. [3, Theorem 1.4] shows that swapping each partition for its conjugate establishes a bijection between the partitions of r in which all parts are less than or equal to $a - 2$ and the partitions of r into at most $a - 2$ parts. Given the bounds on a , this is a substantially smaller set to start from than the set of partitions of r into $n - a$ parts.

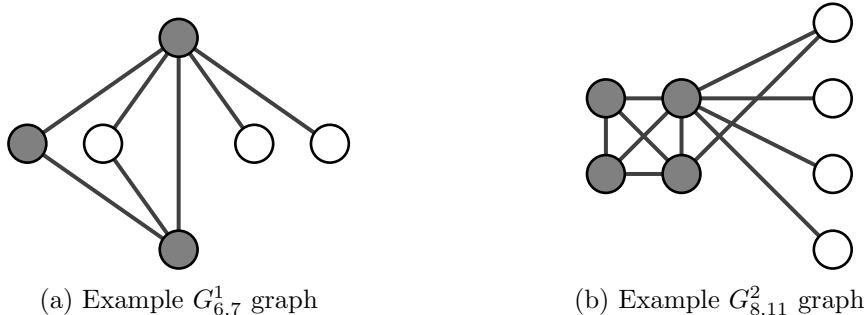


Figure 3.4: Examples of graphs G^1 and G^2 . Both are threshold graphs with clique-numbers $a = 3$ for $G^1_{n,m}$ and a maximal given m for $G^2_{n,m}$. Vertices in the clique are shown in gray.

respect to a we find:

$$a_{min} = \left\lceil n + \frac{1}{2} - \sqrt{n^2 - 2m - n + \frac{1}{4}} \right\rceil,$$

and

$$a_{max} = \left\lfloor \frac{3 + \sqrt{9 + 8(m - n)}}{2} \right\rfloor.$$

With this approach we can explicitly find the degree sequence of the minimum entropy (n, m) -graph for small values of n . Here we show the results for $n \leq 72$.

The above computational approach confirms that the degree sequence of the minimum entropy (n, m) -graph when $m \geq 2n - 2$, does indeed depend on the value of n . When $n \leq 16$, the graph yielding minimum entropy when $m > 2n - 3$ is $G^2_{n,m}$, with degree sequence

$$(n - 1, (k + 1)^j, k^{k-j}, j + 1, 1^{n-k-2}),$$

where k and j are the unique values satisfying

$$m = (n - 1) + \binom{k}{2} + j,$$

$$1 \leq k \leq n - 2,$$

$$1 \leq j \leq k.$$

The graphs $G^1_{n,m}$ and $G^2_{n,m}$ have both a clear construction and interpretation (see Fig. 3.4). $G^1_{n,m}$ is built by considering the smallest possible non-trivial

clique (a 3-clique) and then distributing the remaining edges among the vertices in the only compatible way to have a threshold graph. $G_{n,m}^2$ is built by taking the largest possible clique (a k -clique), connecting all other vertices to a single vertex in the clique, and then distributing the remaining edges by concentrating them as much as possible in one of the vertices outside of the clique (which hence obtains its $j + 1$ degree), so to maximise the unbalance in the degree distribution. We can interpret $G_{n,m}^1$ as the graph whose degree sequence is maximally concentrated in two vertices, but admitting potentially many vertices with degree 1 or 2, whilst we can interpret $G_{n,m}^2$ as the graph where the number of vertices with degree 1 is maximised.

However, when $n > 16$, other (n, m) -graphs attain smaller entropy than $G_{n,m}^2$ for $m \in \{2n - 2, 2n - 1, \dots, 2n - 3 + \ell(n)\}$, where $\ell(n) \geq 1$ is a small non-negative integer that depends on n .

For instance, when $m = 2n - 2$, and $n > 16$, the graph yielding minimum entropy has degree sequence

$$((n - 1)^2, 3^2, 2^{n-4}),$$

that is a graph with two universal vertices and one additional edge. When $m = 2n - 1$, the minimum-entropy graph is $G_{n,m}^2$ for $n \leq 22$, but for $n > 22$, the minimum entropy graph has degree sequence

$$((n - 1)^2, 4, 3^2, 2^{n-5}),$$

that is a graph with two universal vertices and such that, deleting such vertices, one is left with a star S_2 . More generally, the computations suggest that when $m = 2n - 3 + m'$, with $1 \leq m' \leq \ell(n)$, the minimum entropy graph is the graph with two universal vertices and the edges from the graph G of size m' described in equation (3.5) in Theorem 32. When $m > 2n - 3 + \ell(n)$, then $G_{n,m}^2$ is again the minimum entropy graph (Fig. 3.5).

The difficulty in establishing this result via a mathematical proof lies primarily in the seemingly irregular pattern followed by $\ell(n)$ as a function of n (Fig. 3.6), which hinders the formulation of a precise conjecture.

3.3.5 Approximations for applications to large graph

An obvious limitation of a computational approach is that it can only be applied to a limited range of values of n . In the above, for example, we have limited ourselves to the range $n \leq 72$. Many graphs of practical interest are

Extrema of first degree-based entropy for graphs of order $n = 24$

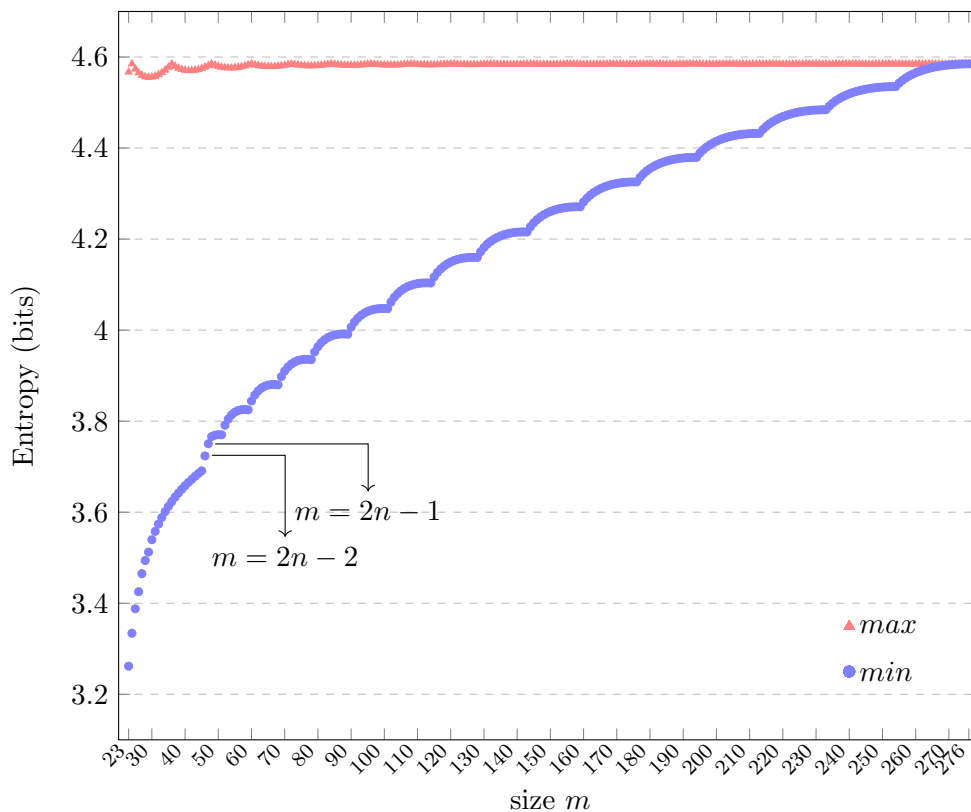


Figure 3.5: Extrema of first degree-based entropy for graphs of order $n = 24$, as function of the size m . Note the three distinct patterns for the minimum values: $I(G_{24,m}^1)$ when $m \in \{n - 1, 2n - 3\}$, with the exceptions described in Theorem 13, $I(G_{24,m}^2)$ when $m \geq 2n$, and an intermediate transition zone of length $\ell(24) = 2$.

considerably larger than that. A complementary approach consists in looking at the behaviour of the entropy for very large n , or, more precisely, in the limit $n \rightarrow \infty$.

We have seen that (in most cases) $G_{n,m}^1$ minimises the first degree-entropy amongst (n, m) -graphs for $n - 1 \leq m \leq 2n - 3$, and that $G_{n,m}^2$ is the most promising candidate for minimising it for $m > 2n - 3 + \ell(n)$, where $\ell(n)$ is a function of n . In this section, we prove that, for large n , the entropy for both these graphs can be approximated by expressions that depend only on the size m and on $\alpha = \lim_{n \rightarrow \infty} \frac{n}{m}$, the limit of ratio of the order and size of the graph. This provides us with two simple expressions for the normalisation of the entropy of large graphs (Propositions 36 and 37).

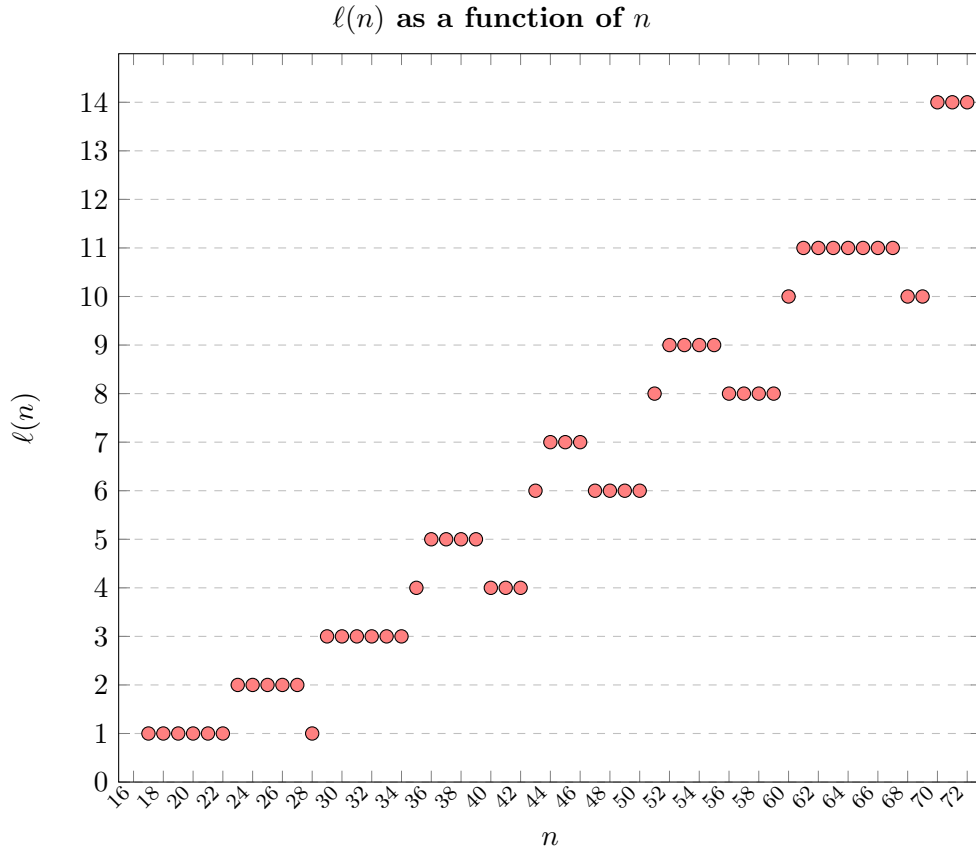


Figure 3.6: Number $\ell(n)$ of values of m for which the (n, m) -graph minimising the first degree-based entropy is neither as described in Theorem 13 nor $G_{n,m}^2$. Remark that $\ell(n) = 0$ for $n \leq 16$, and $\ell(n)$ tends to increase with increasing n , though not monotonically.

Proposition 36. For large n , and any $n + 6 \leq m \leq 2n - 3$,

$$I(G_{n,m}^1) \approx \frac{1}{2} \log m + \frac{1}{2} I(\alpha) + \alpha,$$

where $\alpha = \frac{n}{m}$, and

$$I(\alpha) = -\alpha \log(\alpha) - (1 - \alpha) \log(1 - \alpha)$$

is the binary entropy function.

Proof. If m is large, and since $m \geq n + 6$

$$\begin{aligned}
& \frac{1}{2m}(n-1)\log(n-1) \\
&= \frac{1}{2m}(n-1)\left(\log n + \log\left(1 + \frac{1}{n}\right)\right) \\
&= \frac{1}{2m}(n\log n - \log n) + O\left(\frac{1}{m}\right) \\
&= \frac{1}{2}\alpha\log n + O\left(\frac{\log n}{m}\right), \tag{3.8}
\end{aligned}$$

$$\begin{aligned}
& \frac{1}{2m}(m-n+2)\log(m-n+2) \\
&= \frac{1}{2m}(m-n+2)\left(\log m + \log\left(1 - \frac{n}{m} + \frac{2}{m}\right)\right) \\
&= \frac{1}{2}\left(1 - \frac{n}{m} + \frac{2}{m}\right)\left(\log m + \log\left(1 - \frac{n}{m}\right) + \log\left(1 + \frac{2}{m-n}\right)\right) \\
&= \frac{1}{2}(1-\alpha)\log m + \frac{1}{2}(1-\alpha)\log(1-\alpha) + O\left(\frac{\log m}{m}\right),
\end{aligned}$$

and,

$$\frac{2(m-n+1)}{2m} = 1 - \alpha + \frac{1}{m}.$$

Combining these three results, we obtain

$$\begin{aligned}
I(G_{n,m}^1) &= \log(2m) - \frac{1}{2m}[(n-1)\log(n-1) \\
&\quad + (m-n+2)\log(m-n+2) + 2(m-n+1)] \\
&= \log(m) + 1 - \frac{1}{2}(\alpha\log n + (1-\alpha)\log m + (1-\alpha)\log(1-\alpha)) \\
&\quad - 1 + \alpha + O\left(\frac{\log m + \log n}{m}\right) \\
&= \frac{1}{2}\log m - \frac{1}{2}(\alpha(\log n - \log m) + (1-\alpha)\log(1-\alpha)) \\
&\quad + \alpha + O\left(\frac{\log m + \log n}{m}\right) \\
&= \frac{1}{2}\log m - \frac{1}{2}(\alpha\log \alpha + (1-\alpha)\log(1-\alpha)) \\
&\quad + \alpha + O\left(\frac{\log m + \log n}{m}\right) \\
&= \frac{1}{2}\log m + \frac{1}{2}H(\alpha) + \alpha + O\left(\frac{\log m + \log n}{m}\right),
\end{aligned}$$

with $O\left(\frac{\log m + \log n}{m}\right)$ vanishing for large m . □

Proposition 37. For large n , and $n \leq m \leq \frac{n(n-1)}{2}$,

$$I(G_{n,m}^2) \approx \frac{1}{2} (\log m + I(\alpha) + \alpha + 1),$$

where $\alpha = \frac{n}{m}$.

To prove Proposition 37, we introduce an approximation $f(n, m)$ for $I(G_{n,m}^2)$ that does not depend explicitly on j and k , but only n and m . For a fixed n and variable m , the function $f(n, m)$ is obtained by considering a smooth curve passing through the endpoints of the small arcs described by the graph of $I(G_{n,m}^2)$ (see Fig. 3.7).

Definition 38 (Approximation of $I(G_{n,m}^2)$).

$$f(n, m) = \log(2m) - \frac{1}{2m} ((n-1) \log(n-1) + J_{n,m}), \quad (3.9)$$

where

$$J_{n,m} = \left(\frac{1 + \sqrt{1 + 8(m-n+1)}}{2} \right)^2 \log \left(\frac{1 + \sqrt{1 + 8(m-n+1)}}{2} \right).$$

We now proceed with the proof, that we subdivide in Lemmas for convenience.

Lemma 39. Write $m = n - 1 + \frac{k(k-1)}{2} + j$, $k \geq 1$, $1 \leq j \leq k$.

39.1 If $k > 1$ then $\left| \frac{4(2j-k)+1}{4k^2} \right| < 1$.

39.2 If $m \geq n + 6$, then $\left| \frac{2j-k+2}{k^2} \right| < 1$.

Proof. 1. On the one hand, since $j \geq 1$,

$$\frac{4(2j-k)+1}{4k^2} \geq \frac{9-4k}{4k^2},$$

and

$$\frac{4(2j-k)+1}{4k^2} > -1 \iff 4(k(k-1)) > -9,$$

which is true for all k . On the other hand, since $j \leq k$,

$$\frac{4(2j-k)+1}{4k^2} \leq \frac{4k+1}{4k^2},$$

First degree-based entropy for $G_{n,m}^1$ and $G_{n,m}^2$ for $n = 100$ nodes

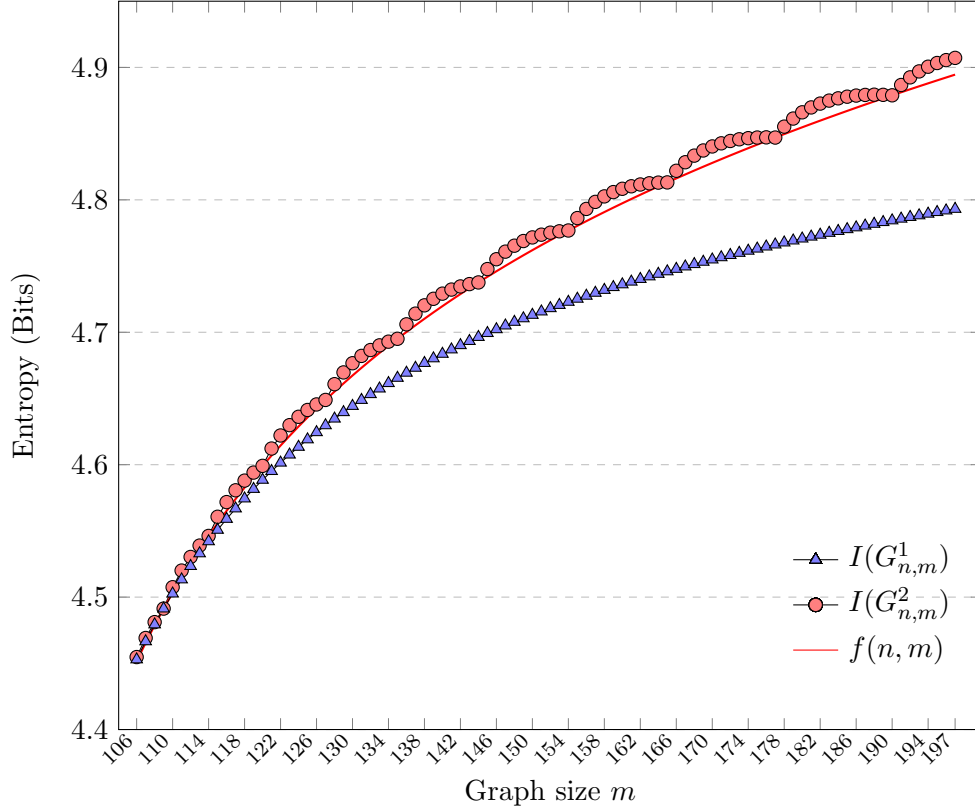


Figure 3.7: Unnormalized first degree-based entropy $I(G_{n,m}^1)$ and $I(G_{n,m}^2)$ for $n = 100$ nodes as $n + 6 \leq m \leq 2n - 3$. For a fixed n , the gap $I(G_{n,m}^2) - I(G_{n,m}^1)$ tends to grow larger as m increases. Remark that $l_2 = \lim_{n \rightarrow \infty} (I(G_{n,m}^2) - I(G_{n,m}^1)) = \frac{1-\alpha}{2}$, where $\alpha = \lim_{n \rightarrow \infty} \frac{n}{m}$ is the limit of ratio of the order and size of the graph. Remark that for $G_{n,m}^1$ to be defined, we need $n + 6 \leq m \leq 2n - 3$, thus $\frac{1}{2} \leq \alpha \leq 1$, and therefore $0 \leq l_2 \leq \frac{1}{4}$.

and

$$\frac{4k+1}{4k^2} < 1 \iff 4k(k-1) > 1,$$

which hold for all $k > 1$. Therefore if $k > 1$ then $\left| \frac{4(2j-k)+1}{4k^2} \right| < 1$.

2. $m \geq n + 6$ and $j \leq k$, so $\frac{k(k+1)}{2} \geq 7$. Hence $k \geq 3$. Now, since $j \geq 1$,

$$\frac{2j-k+2}{k^2} \geq \frac{4-k}{k^2},$$

and

$$\frac{4-k}{k^2} > -1 \iff k(k-1) > -4,$$

which holds for all k . On the other hand, since $j \leq k$,

$$\frac{2j - k + 2}{k^2} \leq \frac{k + 2}{k^2}.$$

Now, $\frac{k+2}{k^2} < 1$ as long as $k > 2$, and we've just seen that $k \geq 3$. Therefore if $m \geq n + 6$, then $\left| \frac{2j-k+2}{k^2} \right| < 1$. □

Lemma 40. $\lim_{n \rightarrow \infty} (I(G_{n,m}^2) - f(n, m)) = 0, \forall m \in \mathbb{N} : n \leq m \leq \frac{n(n-1)}{2}$.

Proof. Write $m = (n - 1) + \frac{k(k-1)}{2} + j$ with k and j as above. Then the difference becomes

$$I(G_{n,m}^2) - f(n, m) = \frac{I_{k,j} - J_{k,j}}{2(n-1) + k(k-1) + 2j}, \quad (3.10)$$

where

$$I_{k,j} = j(k+1) \log(k+1) + k(k-j) \log k + (j+1) \log(j+1),$$

and

$$J_{k,j} = \left(\frac{1 + \sqrt{1 + 4(k(k-1) + 2j)}}{2} \right)^2 \log \left(\frac{1 + \sqrt{1 + 4(k(k-1) + 2j)}}{2} \right).$$

If $k = 1$, then $j = 1$ and $I_{k,j} = 4 = J_{k,j}$, so $I(G_{n,m}^2) - f(n, m) = 0$ for any value of n .

Assume $k > 1$. Then by Lemma 39.1, $\left| \frac{4(2j-k)+1}{4k^2} \right| < 1$. We can thus rewrite the square root term using a binomial series expansion as

$$\begin{aligned} \sqrt{1 + 4(k(k-1) + 2j)} &= 2k \sqrt{1 + \frac{4(2j-k)+1}{4k^2}} \\ &= 2k \left(1 + \frac{1}{2} \frac{4(2j-k)+1}{4k^2} + O\left(\frac{1}{k^2}\right) \right) \\ &= 2k \left(1 + O\left(\frac{1}{k}\right) \right) \end{aligned} \quad (3.11)$$

$$= 2k + O(1), \quad (3.12)$$

where eq. (3.11) follows from the fact that $j \leq k$.

Now, when $|x| > 1$,

$$\begin{aligned}\log(x+1) &= \log\left(x\left(1+\frac{1}{x}\right)\right) \\ &= \log x + \log\left(1+\frac{1}{x}\right) \\ &= \log x + \frac{1}{\ln(2)}\left(\frac{1}{x} + O\left(\frac{1}{x^2}\right)\right)\end{aligned}\tag{3.13}$$

$$= \log x + O\left(\frac{1}{x}\right),\tag{3.14}$$

where in eq. (3.13) we've used a Taylor expansion of the natural logarithm.

By taking $x = \sqrt{1 + 4(k(k-1) + 2j)}$ in eq. (3.14), we can then write

$$\begin{aligned}\log\left(\frac{1 + \sqrt{1 + 4(k(k-1) + 2j)}}{2}\right) \\ &= \log\left(1 + \sqrt{1 + 4(k(k-1) + 2j)}\right) - 1 \\ &= \log\left(\sqrt{1 + 4(k(k-1) + 2j)}\right) + O\left(\frac{1}{\sqrt{1 + 4(k(k-1) + 2j)}}\right) - 1 \\ &= \frac{1}{2}\log(1 + 4(k(k-1) + 2j)) + O\left(\frac{1}{k}\right) - 1 \\ &= \frac{1}{2}\log\left(4k^2\left(1 + \frac{4(2j-k)+1}{4k^2}\right)\right) + O\left(\frac{1}{k}\right) - 1 \\ &= \log k + 1 + \frac{1}{2}\log\left(1 + \frac{4(2j-k)+1}{4k^2}\right) + O\left(\frac{1}{k}\right) - 1 \\ &= \log k + O\left(\frac{4(2j-k)+1}{4k^2}\right) + O\left(\frac{1}{k}\right)\end{aligned}\tag{3.15}$$

$$= \log k + O\left(\frac{1}{k}\right),\tag{3.16}$$

where eq. (3.15) follows from a Taylor expansion (note that $k > 1$, and so $\left|\frac{4(2j-k)+1}{4k^2}\right| < 1$).

By eqs. (3.12) and (3.16), we have thus

$$\begin{aligned}J_{k,j} &= \left(\frac{1 + 2k + O(1)}{2}\right)^2 \left(\log(k) + O\left(\frac{1}{k}\right)\right) \\ &= k^2 \log k + O(k \log k).\end{aligned}\tag{3.17}$$

Now, applying eq. (3.14) to $I_{j,k}$, and recalling that $1 \leq j \leq k$, we have

$$\begin{aligned}
I_{j,k} &= j(k+1) \left(\log k + O\left(\frac{1}{k}\right) \right) \\
&\quad + k(k-j) \log k + (j+1) \log(j+1) \\
&= jk \log k + j \log k + O(k) + k^2 \log k \\
&\quad - kj \log k + (j+1) \log(j+1) \\
&= k^2 \log k + O(k \log k).
\end{aligned} \tag{3.18}$$

And thus by equations 3.17 and 3.18,

$$\begin{aligned}
J_{j,k} - I_{j,k} &= k^2 \log k + O(k \log k) - (k^2 \log k + O(k \log k)) \\
&= O(k \log k).
\end{aligned} \tag{3.19}$$

The sum in the denominator of eq. (3.10), contains n and k^2 terms which dominates $O(k \log k)$, as long as an appropriately large n is chosen. Therefore we can make the difference $I(G_{n,m}^2) - f(n, m)$ arbitrarily small for large n , that is

$$\lim_{n \rightarrow \infty} (I(G_{n,m}^2) - f(n, m)) = 0.$$

□

Lemma 41. For large n , and $n+6 \leq m \leq \frac{n(n-1)}{2}$,

$$f(n, m) \approx \frac{1}{2} (\log m + I(\alpha) + \alpha + 1),$$

where $\alpha = \frac{n}{m}$

Proof.

$$f(n, m) = \log(2m) - \frac{1}{2m} ((n-1) \log(n-1) + J_{n,m}),$$

where

$$J_{n,m} = \left(\frac{1 + \sqrt{1 + 8(m-n+1)}}{2} \right)^2 \log \left(\frac{1 + \sqrt{1 + 8(m-n+1)}}{2} \right).$$

Let $x = m - n + 1$, then

$$J_x = \left(\frac{1 + \sqrt{1 + 8x}}{2} \right)^2 \log \left(\frac{1 + \sqrt{1 + 8x}}{2} \right).$$

Now,

$$\begin{aligned}
J_x &= \frac{1}{4}(1 + 2\sqrt{1+8x} + 1 + 8x) \left(\log(\sqrt{1+8x}) + \log\left(1 + \frac{1}{\sqrt{1+8x}}\right) - 1 \right) \\
&= \frac{1}{2}(1 + \sqrt{1+8x} + 4x) \left(\frac{1}{2} \log(8x) + \frac{1}{2} \log\left(1 + \frac{1}{8x}\right) \right. \\
&\quad \left. + \log\left(1 + \frac{1}{\sqrt{1+8x}}\right) - 1 \right) \\
&= \frac{1}{2}(1 + \sqrt{1+8x} + 4x) \left(\frac{1}{2} \log x + O\left(\frac{1}{x}\right) + O\left(\frac{1}{\sqrt{x}}\right) + \frac{1}{2} \right) \tag{3.20}
\end{aligned}$$

$$= x \log x + x + O(\sqrt{x} \log x), \tag{3.21}$$

where for eq. (3.20), we've used the fact that $m \geq n + 6$, so $x \geq 7$, and thus we can use a Taylor expansion for \log .

By eq. (3.21), we have

$$\begin{aligned}
\frac{J_{n,m}}{2m} &= \frac{(m-n+1)}{2m} \log(m-n+1) \\
&\quad + \frac{m-n+1}{2m} + O\left(\frac{\log(m-n+1)}{m}\right) \\
&= \frac{1}{2} \left(1 - \alpha + \frac{1}{m}\right) \left(\log m + \log(1 - \alpha) + O\left(\frac{1}{m}\right)\right) \\
&\quad + \frac{1 - \alpha}{2} + O\left(\frac{\log m}{m}\right) \\
&= \frac{1 - \alpha}{2} (\log m + \log(1 - \alpha) + 1) + O\left(\frac{\log m}{m}\right). \tag{3.22}
\end{aligned}$$

Thus by equations 3.22 and 3.8, we have

$$\begin{aligned}
I_{n,m} &= \log m + 1 - \frac{1}{2} \alpha \log n \\
&\quad - \frac{1 - \alpha}{2} (\log m + \log(1 - \alpha) + 1) + O\left(\frac{\log m}{m}\right) \\
&= \frac{1}{2} \log m - \frac{\alpha}{2} (\log n - \log m) \\
&\quad - \frac{1 - \alpha}{2} \log(1 - \alpha) + \frac{\alpha + 1}{2} + O\left(\frac{\log m}{m}\right) \\
&= \frac{1}{2} (\log m - \alpha \log \alpha - (1 - \alpha) \log(1 - \alpha) + \alpha + 1) + O\left(\frac{\log m}{m}\right) \\
&= \frac{1}{2} (\log m + H(\alpha) + \alpha + 1) + O\left(\frac{\log m}{m}\right), \tag{3.23}
\end{aligned}$$

with the term $O\left(\frac{\log m}{m}\right)$ small for large values of m . □

Proof of Proposition 37. The result follows from Lemmas 40 and 41. □

3.4 Normalisation

Normalising the first degree-based entropy $I(G)$ on the basis of one's knowledge of basic graph properties of G provides a more informative index $\tilde{I}(G)$ of the balance of the graph connections than normalising by the extrema obtained by the measure amongst all graph.

When we are interested in comparing G with other connected graphs with order n , a possible approach is that of situating its entropy $I(G)$ in a range going from the entropy of the most centralised tree $I(S_{n-1}) = 1 + \frac{1}{2} \log(n-1)$ (as proved in Proposition 22), to the entropy of the complete graph $I(K_n) = \log(n)$ (or any other regular graph with n nodes). This is achieved by normalising $I(G)$ as

$$\tilde{I}(G) = I_n(G) = \frac{I(G) - I(S_{n-1})}{I(K_n) - I(S_{n-1})}.$$

By the results of Theorem 12, when one compares G with other graphs of size m , one should normalise $I(G)$ as

$$\tilde{I}(G) = I_m(G) = \frac{I(G) - I(\mathcal{C}(m))}{I(mK_2) - I(\mathcal{C}(m))}.$$

Here $I(mK_2) = \log(2m)$ and

$$I(\mathcal{C}(m)) = \log(2m) - \frac{1}{2m} (\ell f(k) + (k - \ell)f(k - 1) + f(\ell)),$$

where $m = \binom{k}{2} + \ell$ with $0 \leq \ell < k$. Then, $I_m(G)$ falls within the range $[0, 1]$ with the minimum (resp. maximum) obtained precisely when the graph's degree sequence is maximally concentrated (resp. balanced).

By Theorems 11 and 13 and by the above numerical observations, a more appropriate normalisation for an (n, m) -graph G , for which both the order n and the size m are known, is

$$\tilde{I}(G) = I_{n,m}(G) = \frac{I(G) - I(G_{n,m}^i)}{I(G_{n,m}^{max}) - I(G_{n,m}^i)},$$

where $i = 1$ when $n + 6 \leq m \leq 2n - 3$ and $i = 2$ when $m > 2n - 3 + \ell(n)$.

Here

$$I(G_{n,m}^{max}) = \log(2m) - \frac{1}{2m}(rf(q+1) + (n-r)f(q)),$$

where $2m = qn + r$, with $q = \lfloor \frac{2m}{n} \rfloor$ and $0 \leq r < n$,

$$I(G_{n,m}^1) = \log(2m) - \frac{1}{2m}(f(n-1) + f(m-n+2) + m-n+1),$$

and

$$I(G_{n,m}^2) = \log(2m) - \frac{1}{2m}(f(n-1) + jf(k+1) + (k-j)f(k) + f(j+1)),$$

with j and k satisfying $m = n - 1 + \binom{k}{2} + j$, with $1 \leq k \leq n - 2$, and $1 \leq j \leq k$. Then, the index $I_{n,m}(G) \in [0, 1]$ with $I_{n,m}(G) = 0$ when the degree sequence of G is maximally unbalanced given the constraints on its order and size, and when $I_{n,m}(G) = 1$ when the degree sequence is as balanced as possible.

In Chapters 4 and 5, we use the normalised indices I_n , I_m , and $I_{n,m}$ to study the balance of the distribution of connections in various urban networks: the disc graphs connecting towns in a region when their geographical distance is smaller than a certain threshold (Section 4.4.4); the worldwide network of airports connected via direct flight routes (Section 5.1); and the disc graphs based on the evolving travel time between towns as the railway system expands and modernises (Section 5.2.3). Whilst the indices are easy to calculate for a given network, the proofs and computations presented in this chapter to verify their validity show the intricacy of the problem even in these simple cases. In practice, much more complex spatial, geographic, economic, political, cultural, and historical constraints are in place on the structure of real urban networks.

Chapter 4

Cities and their Territory: Population and Spatial Organisation of Systems of Cities

In this chapter, we look at fundamental aspects of the morphology of a system of cities: the cities' population sizes, the land use and population density of their surrounding territory, and their spatial organisation in terms of locations and reciprocal distances.

We open the chapter with a discussion on the compatibility of some of the historical population data sets introduced in Section 2.2, and the rationale behind our choice to focus on three of them. We show a simple application of Shannon entropy to study the balance of city sizes in several European powers from 1300 to 1850 in Section 4.2. In Section 4.3, we introduce local entropy: a spatial entropy measure for raster data, and apply it to study the heterogeneity in and around urban areas in the British Isles, Italy, and South Asia. In both sections, we discuss the correlation between variations in the measures and major historical events. Finally, in Section 4.4, we focus on the spatial distribution of towns in England and Wales in the 19th century, which we study using point processes and the first degree-based entropy of disc graphs, which connect any two towns when their geographical distance is less than a fixed threshold, normalising the entropy as seen in Section 3.4.

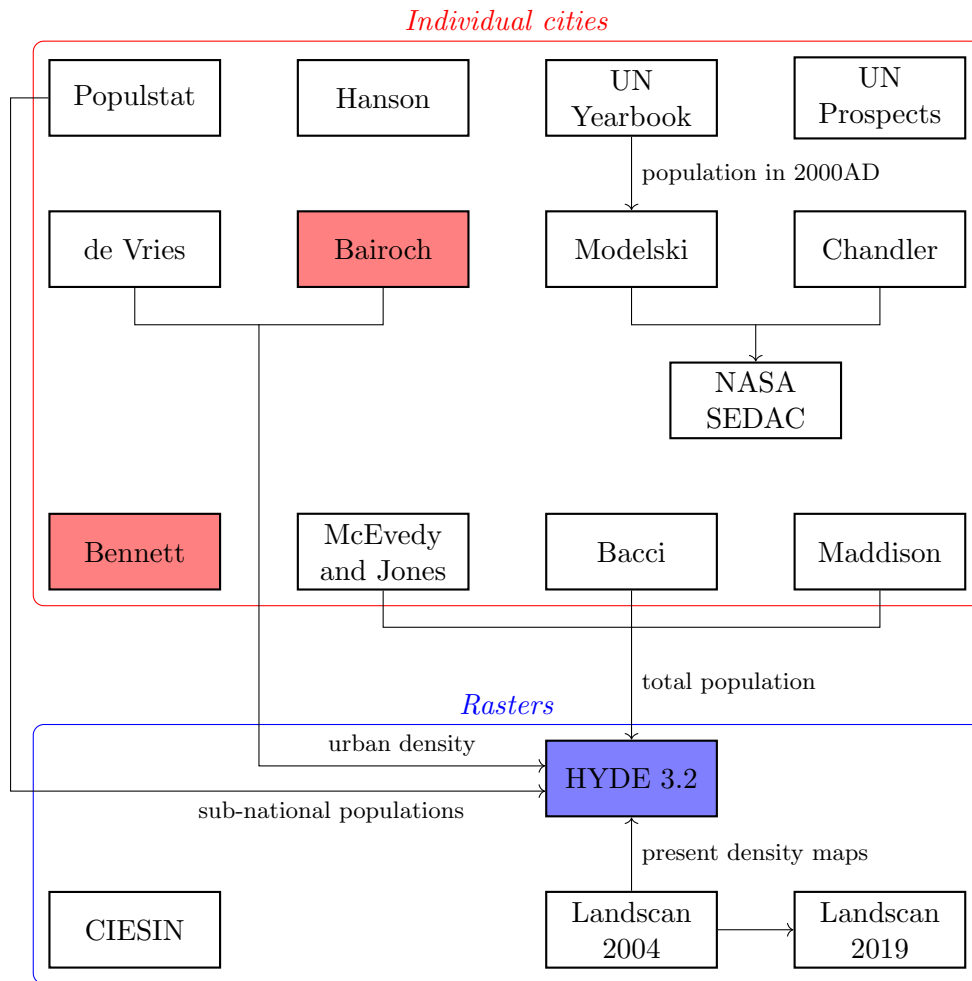


Figure 4.1: Population data sets and studies and their relations with one another. Shaded data sets are used in the thesis.

4.1 Completion and compatibility of historical population data sets

4.1.1 Compatibility of different data sets

The data sets described in Section 2.2.1 adopt different criteria to compute historical estimates of the population, and often focus on different regions or time periods. Yet, as we have seen, they are deeply interconnected with one another, being often derived from common historical studies. A diagram of their relations is given in Fig. 4.1. This raises the question of whether the data sets are compatible, or, in other words, whether analyses carried out with different data sets would lead to consistent results.

The NASA SEDAC [180] and UN World Urbanisation Prospects [206] are formally unrelated data sets, but have the same worldwide geographic scope, cover overlapping periods of time, and contain a similar number of cities. Thus, we check here their compatibility.

We create a more complete data set, merging together the two data sets. Discrepancies in the population data of the same city are to be expected, given that they use different criteria to define urban populations, as detailed in Section 2.2.1. To avoid major clashes, the population data from year 1000BC to 1950AD was taken from the NASA data set, whilst data from 1955AD onward was taken from the UN one. Cases of different spellings for the same cities and countries were manually accounted for and appropriately matched. The resulting merged data set contained 2738 cities, but only 717 had records in both original data sets. The remaining 2021 cities appear in only one of the two data sets, and thus the data on their population refers to either the 1000BC-1950AD period or the 1955AD-2015AD period, but not to both.

We compare the thus assembled data set with the population data in HYDE 3.2 [127]. Both the NASA and UN data sets provide population data for each individual city u , and a single pair of coordinates (lon_u, lat_u) , corresponding to the city centre. On the other hand, HYDE 3.2 is a raster data set of historical population counts and densities in small cells covering the globe, estimated by combining present-day high-resolution population maps, historical sources for the population of cities and wider regions, and estimates of historical urban densities and water and soil accessibility, through the procedure outlined in Section 2.2.1.2 and detailed in [127] and [128]. For each city u , the coordinates (lon_u, lat_u) fall in exactly one cell (i_u, j_u) in HYDE 3.2.

We matched the city's population $p_{u,y}$ in a given year y in the merged data set to the population of the corresponding cell $p_{(i_u, j_u), y}$ in HYDE 3.2. The population of the cell vastly underestimated the population of the city ($p_{(i_u, j_u), y} \ll p_{u,y}$), especially in recent years and for larger cities. This was to be expected since a cell has an area of at most $86km^2$ and many modern urban areas are larger than that.

To contrast this effect, we mapped each city not just to one cell but to a set of cells. For each cell (i, j) in the raster data set, we considered the neighbourhoods $C_{(i,j)}^n$ formed by the cell itself and its closest, $n - 1 = 4, 8$ or 20 closest neighbours, as measured by great-circle distance between the centres of the cells. The resulting neighbourhoods correspond to discs around the centre of the cell (i, j) , having radii 8km, 12km and 20km at the equator,

respectively. The neighbourhood $C_{(i,j)}^n$ acts as an observation window of size n around the cell (i, j) , as depicted on the top of Table 4.1.

For each year y , we found the most appropriate value \hat{n}_y of window size n by minimising the mean “error”

$$\varepsilon_{n,y} = \frac{1}{N} \sum_{i=1}^N \frac{p_{u,y}}{p_{C_{(i,j)}^n},y} - 1,$$

across the whole data set, i.e.,

$$\hat{n}_y = \underset{n}{\operatorname{argmin}} \varepsilon_{n,y}.$$

We report the values of $\varepsilon_{n,y}$ for all years and window sizes in Table 4.1, as an indicator of the compatibility of the data sets.

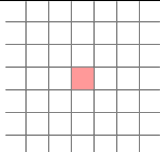
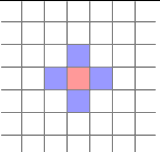
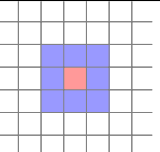
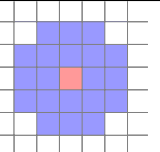
Note that the selection of a window size had to be performed a priori on the whole data set and not at the individual city level, because of the missing records. Hence, the mapping from a city to a set of cells tends to underestimate the population of larger centres and overestimate that of smaller centres. On average the error of the best-matching city to cell set size is 17%. We consider this an excessive error, and therefore we proceed to use Bairoch [15] and Bennett’s [27] data sets in Section 4.2 and HYDE 3.2 in Section 4.3, as the former data set is incorporated directly in the latter, avoiding incorporating data from other data sets.

4.1.2 Data set completion

Historical data sets, even the most accurate ones such as census data, often have missing records. Whether the lack of a data point should be addressed as an issue, accepted as a limitation, or interpreted as a significant piece of information in and of itself, it depends on the specific case. On the one hand, if a town’s population is not recorded at some point in time, but it is known shortly before and after, one can reasonably assume that the town was there all along, and estimate the missing record(s) via interpolation. On the other hand, if a town’s population is not recorded for a long period of time, the settlement may have faced a disruptive event that made it impossible to count its population, or even led it to be momentarily abandoned and then rebuilt on the same location.

In [27], complete population records are not available for every town in England and Wales between 1801 and 1911, but the time interval between

Table 4.1: Average error $\varepsilon_{n,y}$ when comparing the population of cities from the data sets NASA SEDAC [180] (until 1950AD) and UN World Urbanisation Prospects [206] (1955AD onwards) to the population of the corresponding cells in HYDE 3.2 [127], depending on the size of the observation window.

				
year	cell (i, j)	$C_{(i,j)}^5$	$C_{(i,j)}^9$	$C_{(i,j)}^{21}$
1000BC	-0.034	3.071	5.863	13.453
0	-0.64	0.533	1.605	4.562
500AD	-0.692	0.296	1.188	3.63
700AD	-0.738	0.106	0.876	2.969
1000AD	-0.813	-0.209	0.343	1.839
1100AD	-0.784	-0.082	0.562	2.313
1200AD	-0.783	-0.077	0.57	2.329
1300AD	-0.804	-0.175	0.404	1.979
1400AD	-0.837	-0.315	0.164	1.456
1500AD	-0.832	-0.295	0.197	1.523
1600AD	-0.844	-0.348	0.108	1.332
1700AD	-0.837	-0.335	0.128	1.362
1750AD	-0.744	-0.058	0.537	2.066
1800AD	-0.619	0.267	0.986	2.762
1850AD	-0.5	0.495	1.252	3.031
1900AD	-0.647	-0.034	0.398	1.34
1950AD	-0.685	-0.19	0.135	0.796
1960AD	-0.655	-0.118	0.229	0.924
1970AD	-0.672	-0.17	0.15	0.783
1980AD	-0.68	-0.2	0.101	0.69
1990AD	-0.688	-0.229	0.055	0.603
2000AD	-0.707	-0.288	-0.031	0.461
2005AD	-0.722	-0.329	-0.089	0.365
2010AD	-0.729	-0.351	-0.122	0.31
2015AD	-0.74	-0.381	-0.165	0.24

successive records is only 10 years long, and the data is more frequently missing in small towns with little population changes. Hence, if a town's population is not recorded for one or more years, we chose to complete the data via a simple linear interpolation. For some towns, records only begin from a certain year onward, and it is not possible to establish a priori whether the town was a newly established one, or the data was not recorded for other reasons. For these cases, we preferred not to intervene on the original data set. Thus, we could limit the effects of missing data, whilst retaining information on the increase in number of settlements in the system over time.

In [15], the data sets presents similar gaps in the time series of city populations in major European powers from 1300 to 1851, but the time interval between two records, if an intermediate one is missing, is between 100 years and 200 years long. We considered this too long a time to assume a smooth transition between the available records, and thus decided against applying interpolation on this data set.

On a different note, neither [27] nor [15] provided the geographical locations of the settlements. We retrieved these via Geonames¹, associating each town to the coordinates of a modern settlement with the same name. Whilst sanity checks were performed by mapping the data, there remains the possibility of rare incorrect assignment of geographical coordinates in [15] due to the presence of more than one locality with the same name in the same country, or change of city names.

4.2 Population entropy

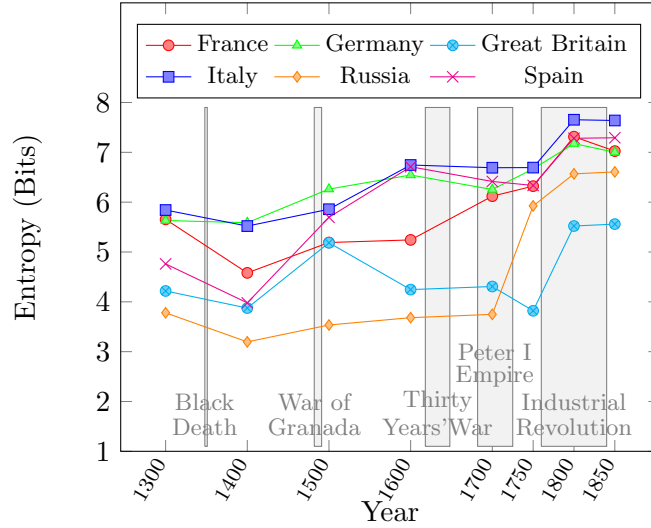
A first, simple application of entropy to the study of the evolution of systems of cities is the analysis of the temporal evolution of the Shannon entropy of the proportion of urban population residing in each city.

4.2.1 Entropy of urban populations in Europe in 1300-1850

Figure 4.2 shows the evolution of this measure in several European systems of cities, from 1300 to 1850. The data is taken from Bairoch [15], whose data set was described in Section 2.2.1. Beyond merely compiling data, [15] also highlighted the urbanisation trends in major European countries and how they were impacted by major historical events. The Shannon entropy of the urban

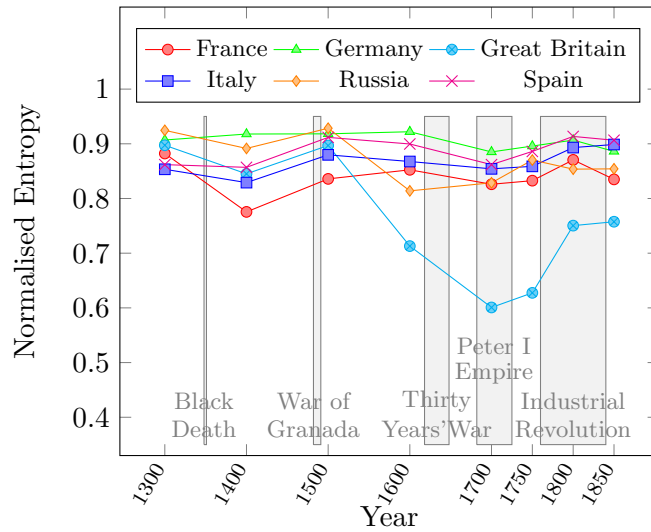
¹<https://www.geonames.org/>.

Entropy of urban population in Europe (1300-1850)



(a) Unnormalised Shannon entropy of the population of cities in major European countries from 1300 to 1850.

Entropy of urban population in Europe (1300-1850), normalised



(b) Shannon entropy of the population of cities in major European countries, normalised for each country C and year y by its maximum possible value $\log(N_{C,y})$, where $N_{C,y}$ is the number of cities in C in year y .

Figure 4.2: Shannon entropy of the population of cities in major European countries from 1300 to 1850, based on data from Bairoch [15]. The country borders are matched to modern ones. Some major events in European history are indicated along the time axis by shaded blocks.

population, in spite of its simplicity, is able to capture some of these structural changes.

The general though uneven growth in the unnormalised Shannon entropy shown in Fig. 4.2a reflects primarily the increase in the number of cities as the population of Europe slowly climbs for most of the period under consideration, but also the improvements in records. The most noticeable exception to this growth is observed in 1400, after the Black Death struck Europe, killing at least a third of the continent's population between 1348 and 1351. The epidemic had a remarkable impact on the urban system. As the total population decreased, the urbanisation rate increased in the vast majority of the countries [15], fuelled by rural to urban migration [121]. The uneven spatial pattern of mortality-driven population decline and subsequent migration-led population recovery produced major shifts in the population rankings of cities [121] and the population concentrated into fewer centres, as the cities engaged in a "survival of the fittest" competition [50], in which only few managed to grow or even avoid decline. These factors are reflected in the decrease in unnormalised entropy of every country from 1300 to 1400, visible in Fig. 4.2a, as well in normalised entropy (Fig. 4.2b), that follows a similar decline in all countries but Germany.

The final stages of the Reconquista, with the War of Granada, had a major impact on the Spanish urban system as a very urbanised Islamic territory passed under the domain of a more rural Christian one [15]. The sharp increase in both the unnormalised and normalised Shannon entropy reflects the decline in the population of the main cities in Andalusia, such as Granada, Cordoba and Seville, which were at the time among the largest in Europe, and thus the levelling-down to a more equalised urban system.

The Thirty Years' War (1618-1648) affected especially Germany, whose cities saw a decline of 60% in their populations. This correlates to the decline in the country's Shannon entropy in the 17th century.

The Reign of Peter I (also known as Peter the Great) signed an enormous expansion of the Russian Tsardom and the creation of the Russian Empire. Several cities were founded, including the capital Saint Petersburg, explaining the spectacular growth in entropy shown in Fig. 4.2a.

The main effects of the Industrial Revolution (1760-1840) would not be felt in Europe until much later than the observed period, but Great Britain is major exception. Not only does the overall entropy reach its peak, signalling the expansion of the system of cities, but also the normalised entropy shows

a remarkable increase after two centuries of decline. This reflects the fact that, for the first time since the 1500, the epicentre of urban growth starts to move away from London towards other smaller cities, creating a more uniform system. In fact, industrialisation emerged outside of London and the other main British cities of the time, in a small number of minor centres that saw their population grow unprecedentedly [14]. In spite of its increase in the 19th century, the distribution of population in British cities remained significantly more unbalanced than in other European countries from the 1600s onward, as reflected by the low values of both the normalised and unnormalised entropy, primarily for the dominating position of London².

The Shannon entropy of the population distribution in a system of cities is thus able to capture the change in uniformity of the system, with greater values of entropy associated to more numerous cities, and growth in the relative share of the population in mid-to-small sized cities. Yet, this simple measure is not able to capture other crucial trends in the urbanisation, such as the progressive shift of the epicentre of European urbanisation from Southern to Western and Northern Europe from the 16th century onward. To capture trends like this, more explicitly spatial measures are needed.

4.2.2 Entropy of urban populations in England and Wales in 1801-1911

We perform the same analysis for the population of English and Welsh towns from 1801 to 1911. We use Census data [27] completed via linear interpolation, as detailed in Section 4.1.2³.

Fig. 4.3 shows the evolution of the Shannon entropy, in its non-normalised and normalised form. The growth in the former is almost exclusively due to the increase in the number of settlements in the system. The latter follows a generally decreasing trajectory until 1891, indicating a progressive increase in concentration of the population in fewer centres. Both these observations reflect the period of fast urbanisation and significant internal migration that

²As late as 1520, most major towns in England had failed to recover from the population losses caused by the Black Death, as evidenced by a stagnating land market and large prevalence of derelict, empty buildings [108]. London was a rare exception. Whilst towns with more than 10,000 inhabitants grew faster than the population as a whole starting from the 1550s, London's manifold advantages of being a court, capital city, and major port [115] cemented its dominating position in the 16th and 17th century, with the primate city housing 11.5% of the national population in 1700, up from 5% in 1600, that is 19 times the population of Norwich, the second largest centre [115].

³The results obtained with the original data set are almost identical, but we chose to present those of the completed data set to ensure consistency with the analysis in Section 5.2.

Entropy of the population in English and Welsh towns (1801-1911)

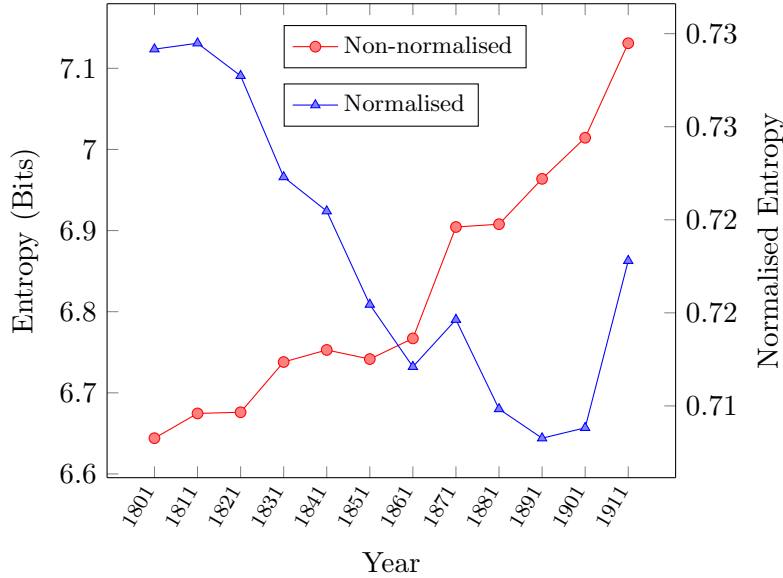


Figure 4.3: Shannon entropy of the population of English and Welsh towns from 1801 to 1911, in its non-normalised form (scale on the left) and normalised form (scale on the right). The normalisation is performed by dividing the entropy in each year y by its theoretical maximum value $\log(N_y)$, where N_y is the number of towns in the year y . The data is taken from [27].

characterised England and Wales for most of the 19th century [211, 170]. The years after 1880 witnessed a change in the trajectory of urbanisation, as new industrial sectors emerged, old ones declined, and the bulk of the growth shifted from larger centres to suburban towns [170]. Again, this is captured by the measures, with a sharp increase in both the un-normalised entropy and, more significantly, the normalised entropy at the turn of the 20th century.

It is important to remark, though, that the absolute variation in the normalised entropy is relatively small⁴, and the measure is significantly smaller than what is observed in other European countries in the previous centuries (See Fig. 4.2b). This indicates that the English and Welsh system of cities remained an extremely unbalanced and centralised one throughout the 19th century, as already remarked above for earlier periods.

⁴The measure attains a minimum of 0.713 in 1891 and a maximum of 0.734 in 1811.

4.3 Local entropy

Until now, we have studied cities and the evolution of the Shannon entropy of their population by considering them as discrete, individual entities (Section 4.2). But, as discussed in Section 2.1, cities may not exist in isolation.

In this Section, we conceptualise cities as entities emerging from a continuum, that remain inextricable from their territory, and shift our attention on the relation between a city and its immediate geographical surroundings. We focus on two fundamental dimensions of urbanisation and their effects on a territory: increased land consumption and growing population density. In Section 4.3.1, we define local entropy, that expands recent efforts to extend Batty’s spatial entropy to create measures accounting not only for size and number of spatial units but also for spatial proximity, as discussed in Section 2.4.2.

We retain our interest in a long-term historical study, and, through the case studies of the British Isles, Italy, and South Asia, reconnect the temporal evolution to some of the main socioeconomic, political, and technological changes and epidemic events these regions went through during the last three centuries.

4.3.1 Methodology

For this analysis, we use the HYDE 3.2 raster data set [127], introduced in Section 2.2.1.2. Recall that HYDE 3.2 provides estimates of historical population densities in a grid of $5' \times 5'$ cells covering the whole world surface, and classifies each cell according to its prevailing land use into one of the seven categories: urban area, dense settlement, village, cropland, rangeland, semi-natural, and wild land, with the exception of oceans, that are unclassified. We assess the level of heterogeneity and disorder in land use and population density in and around human settlements via a local entropy measure, adapted to the raster form of the above described data set.

Consider a region R , and fix some window size $n \in \{5, 9, 21\}$, as formally defined Section 4.1.1 and depicted in Fig. 4.4. Let (i, j) be any cell in R whose land use classifies it as a human settlement (urban, dense settlement, or village). Suppose n' of the cells in its neighbourhoods $C_{(i,j)}^n$ fall within R and on land. Let p_k be the fraction of these n' cells in $C_{(i,j)}^n$ with land use k , where k ranges along the aforementioned categories as well as cropland, rangeland, semi-natural, and wild lands (so that $\sum_k p_k = 1$). We define the

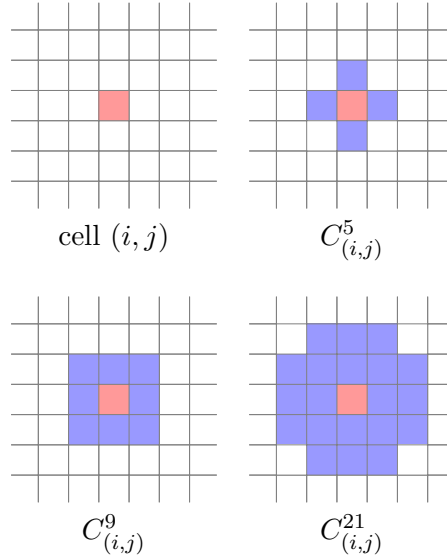


Figure 4.4: Neighbourhoods $C_{(i,j)}^n$ of a cell (i, j) for $n = 5, 9$ and 21 .

land use local entropy $E_{(i,j)}^n$ of the cell (i, j) with respect to its $n - 1$ closest neighbours as the Shannon entropy

$$E_{(i,j)}^n = - \sum_k p_k \log p_k.$$

The minimum value of the entropy ($E_{(i,j)}^n = 0$) is achieved when all cells in the neighbourhood $C_{(i,j)}^n$ have the same land use as (i, j) , i.e., when we have a homogeneous landscape around (i, j) . The more heterogeneous the landscape is, the larger the value of the entropy.

Note that the maximum possible value of $E_{(i,j)}^n$ depends on n' , rather than n directly. For example, for any value of n , if the neighbourhood $C_{(i,j)}^n$ contains only $n' = 5$ cells falling on land and within the borders of the region R , the most heterogeneous case is the one in which the 5 cells have any 5 distinct land uses, i.e., $p_k = \frac{1}{5}$ for any 5 of the k 's and $p_k = 0$ for the remaining two, giving a maximum value of $\log 5$. Normalising $E_{(i,j)}^n$ by the appropriate maximum values allows to avoid border effects and compare the local entropy at (i, j) for different window sizes. The normalised values fall between 0 (complete homogeneity) and 1 (maximum heterogeneity).

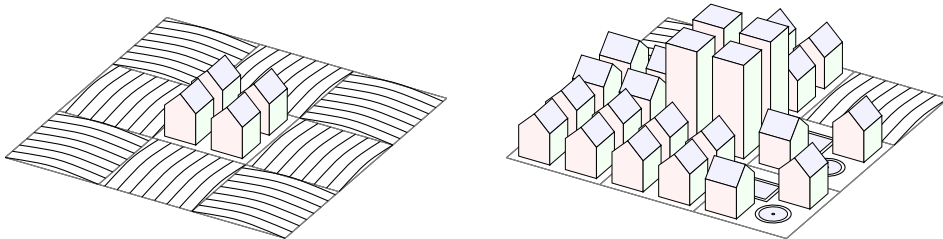
We use the normalised values to define a new raster of land use heterogeneity around human settlements (see, for example, Figs. 4.6 and 4.7 in Section 4.3.3). Summing over all cells (i, j) , we obtain E_{land}^n , a measure of the total local entropy of human landscapes in R . Similarly, averaging over the cells,

we obtain \overline{E}_{land}^n , a measure of the mean local entropy of urban landscapes in R . By computing E_{land}^n and \overline{E}_{land}^n at different moments in time, it is finally possible to study the evolution of the local heterogeneity of land use in the region R .

In a completely analogous way, we classify the density raster data into different categories, and study the evolution of the local entropy of the population density in time. For simplicity and symmetry, here we choose seven categories: zero density (uninhabited lands), densities between 1 and 50 inh/km² (sparsely populated), 51 to 100 (moderate-low density), 101 to 200 (moderate-high density), 201 to 500 (high density), 501 to 1,000 (very high density), and 1,001 or more inh/km² (extremely high density). We proceed as before, calculating the local entropy of the density in the neighbourhoods of each cell classified as human settlement, and normalising the values accordingly. Summing over the cells we obtain $E_{density}^n$, a measure of the total local entropy of population density in R . Averaging, we obtain $\overline{E}_{density}^n$, a measure of the mean local entropy of population density in R .

Remark that the averaging procedure allows the local entropy to be applied at different scales. We will see an analogous approach in studying cities connectivity via the entropy of interactions (Definition 9 in Section 2.5.2), and its application to polycentricity of systems of cities (Section 5.2).

4.3.2 Example



(a) A simple region: a moderate density settlement surrounded by fields. (b) A more complex region: a mixture of high, moderate and low density areas, as well as agricultural land.

Figure 4.5: Examples of a simple and a more complex region.

As a simple example of the computation of local entropy, we consider the two 9-cell regions in Fig. 4.5. As reference systems, we take the top cell in each region to be $(1, 1)$. We choose $n = 5$ as the window size.

In the region in Fig. 4.5a, we have agricultural land with a moderately

dense settlement at its centre. In this case, only the central cell $(2, 2)$ contains a human settlement, so we only deal with $C_{(2,2)}^5$. Here $n' = 5$ and $p_{dense} = \frac{1}{5}$, $p_{cropland} = \frac{4}{5}$, with $p_k = 0$ for every other land use k . Thus

$$E_{(2,2)}^5 = \log 5 - \frac{8}{5}.$$

Normalising by the maximum value $\log 5$, we obtain

$$E_{land}^5 = \bar{E}_{land}^5 = \frac{\log 5 - 8/5}{\log 5} \approx 0.31.$$

In Fig. 4.5b, we have a more complex region with one urban cell at the centre $(2, 2)$, four cells with dense settlement around it $\{(1, 2), (2, 1), (3, 1), (3, 2)\}$, two low density villages $\{(2, 3), (3, 3)\}$ and two cells of cropland $\{(1, 3), (3, 1)\}$. In this case, we have to compute $C_{(i,j)}^n$ for all cells except the two allocated to cropland. As an example of cell on the border, consider $(1, 2)$. $C_{(1,2)}^5$ contains $n' = 4$ cells, with $p_{urban} = \frac{1}{4}$, $p_{dense} = \frac{1}{4}$, $p_{cropland} = \frac{1}{2}$, and $p_k = 0$ for all other land use k . Hence

$$E_{(1,2)}^5 = \frac{3}{2},$$

and the normalising constant is $\log 4 = 2$. After computing and normalising appropriately each cell entropy value, we sum or average over the 7 cells with human settlements, and obtain

$$E_{land}^5 \approx 4.2, \text{ and } \bar{E}_{land}^5 \approx 0.6,$$

reflecting the fact that this region is considerably more heterogeneous than the one in Fig. 4.5a, both overall, and on average around each settled cell.

4.3.3 A Real-world Study of Local Entropy

We apply local entropy to study the change in heterogeneity of land use and population density in the British Isles, South Asia, and Italy.

These regions are all characterised by relatively high population density by global standards, but have radically different urban systems. Britain's industrial revolution propelled it from a mostly rural, sparsely populated area to the first country to industrialise and achieve high rates of urbanisation. It has several large cities, but its urban system is dominated by the primate city of London. Parts of South Asia, on the other hand, have long been amongst the most densely populated in the world, but the region has seen urbanisation rates

rise only in more recent years. The region has large number of metropolises, none of which contains a large proportion of the total population and can be said to dominate above the others. Italy represents an intermediate situation: it kept relatively high urbanisation rates throughout its history, even before its industrialisation, which truly began only after the country's unification in the late nineteenth century. It has a relatively heterogeneous system of cities, with three major, similar sized urban areas: Milan, Rome and Naples, respectively in the North, Centre, and South of the peninsula.

We consider the years from 1700 to 2015, a period in which the regions were shaken by important socioeconomic and technological changes (e.g., the industrialisation) and political events (e.g., the colonisation and decolonisation of India, the unification of Italy, the World Wars), which had lasting effects on their urban structure.

4.3.3.1 Raster of local entropy

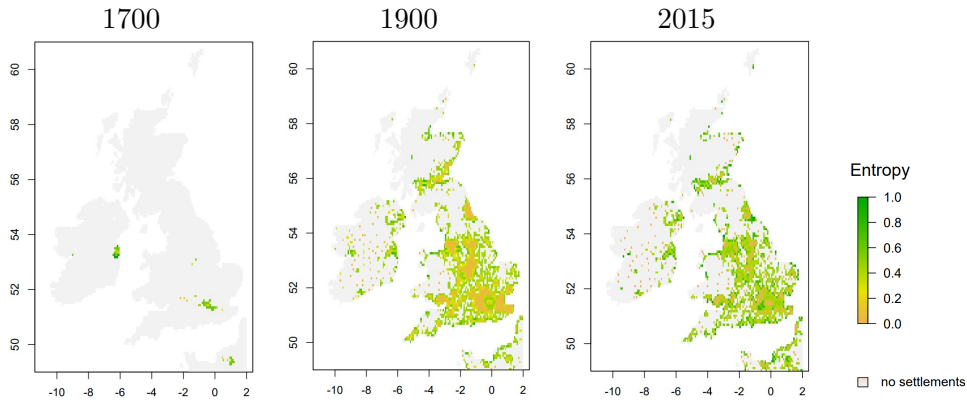
Figure 4.6 shows the regions' rasters of local entropy of land use and in 1700, 1900 and 2015.

In the British Isles, after a dramatic expansion from 1700 to 1900, the number of cells occupied by human settlements remains roughly unchanged to 2015. Comparing the 1900 to 2015, some of the cells classified as villages turn into dense settlements and dense settlements are incorporated into the growing urban areas around major metropolises, resulting in a shift in their color from orange (low local entropy) to green (high local entropy), reflecting the growing heterogeneity in land use of the areas around them. This process is exemplified by Greater London, where the urban core expands, resulting in lower entropy values in the centre, surrounded by higher local entropy areas at the fringes of the city's Green Belt.

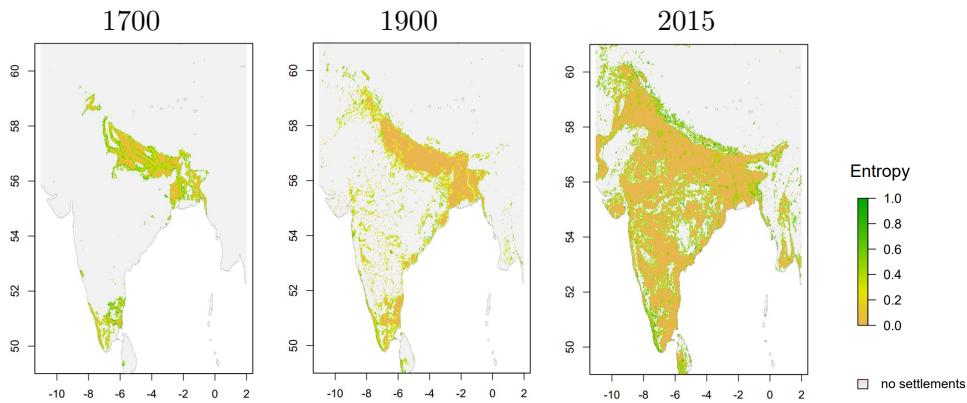
In Italy, a similar expansion is observed from 1700 to 1900, but the increase in settled areas continues to 2015. The majority of new cells occupied by settlements appear in the vicinity of already established urban areas, especially in the Po Valley, or lie along the coast. These newly settled cells are characterised by large local entropy, reflecting their peripheral nature.

In India, the Valley of the Ganges and the South were already hosting several settlements by 1700. The period to 1900 sees the expansion of settlements especially along the coast, in part as a consequence of the growth of port cities during the British colonial rule. A great expansion of the settled areas in the interior follows. Most of the cells are classified as villages surrounded by other

Local entropy of land use in the British Isles



Local entropy of land use in South Asia



Local entropy of land use in Italy

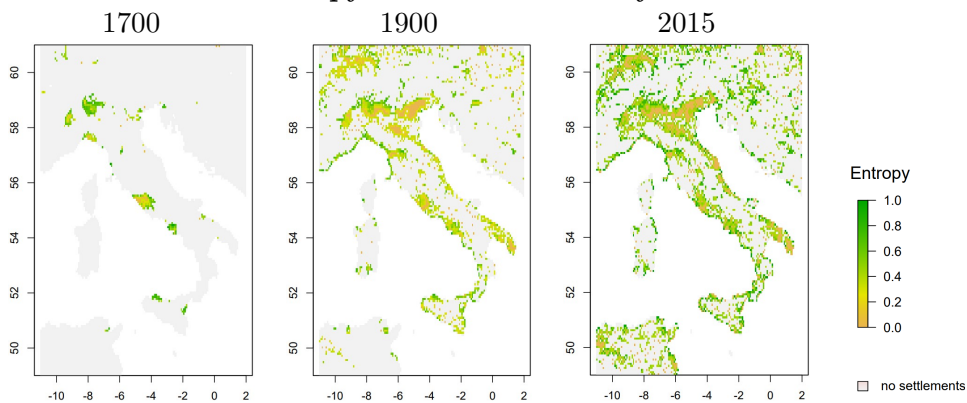
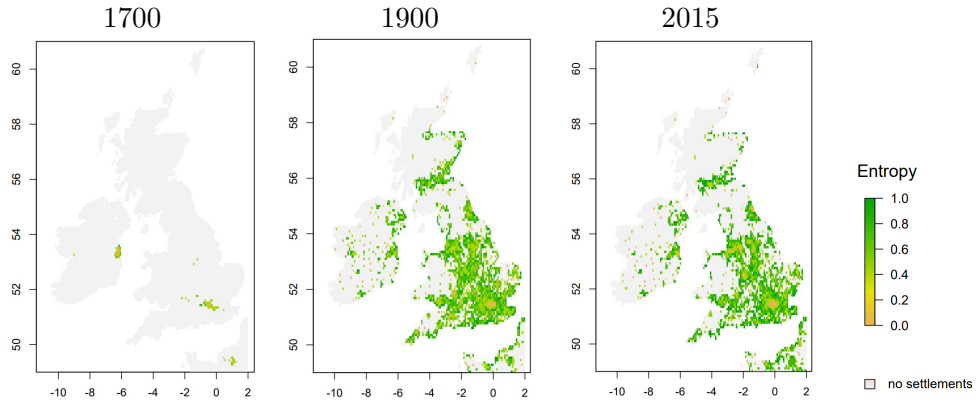
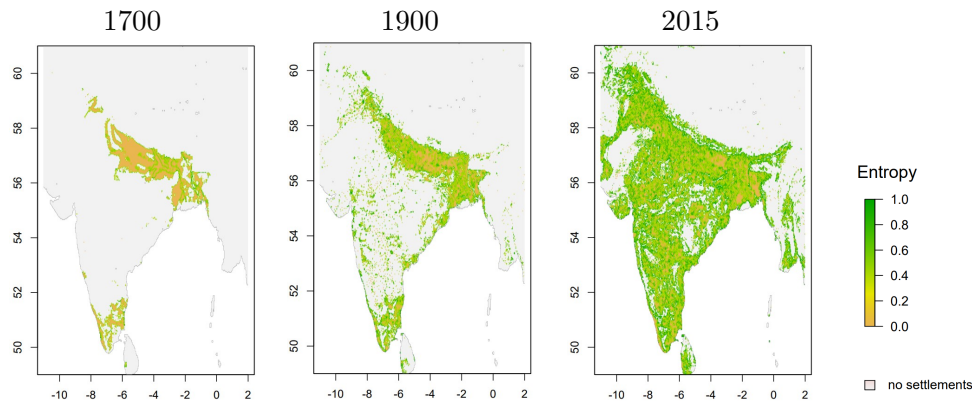


Figure 4.6: Local entropy of land use for each cell classified in HYDE as a human settlement (urban, dense settlement, or village) in the British Isles, South Asia, and Italy, in 1700, 1900, and 2015. The window size is $n = 5$.

Local entropy of population density in the British Isles



Local entropy of population density in South Asia



Local entropy of population density in Italy

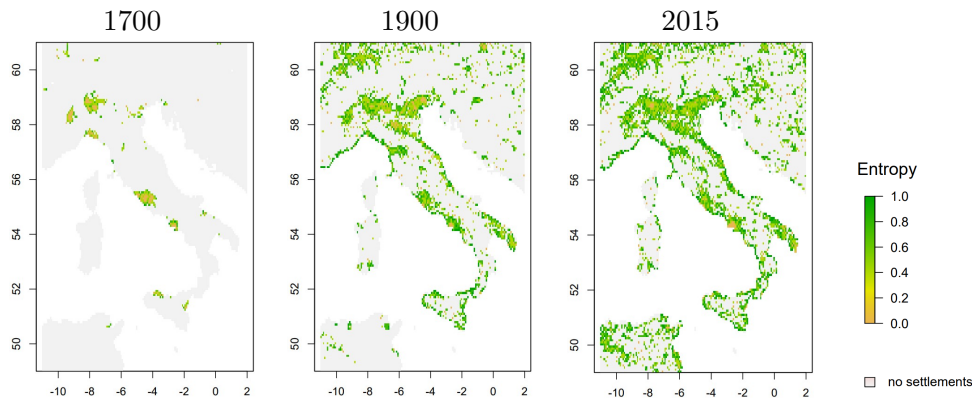


Figure 4.7: Local entropy of population density for each cell classified in HYDE as a human settlement (urban, dense settlement, or village) in the British Isles, South Asia, and Italy, in 1700, 1900, and 2015. The window size is $n = 5$.

villages or cropland, and this is reflected by the predominance of low entropy (orange) cells. Nonetheless, higher values are observed in either more remote areas of the interior and along the Himalayas, where settlements border wilder landscapes, and around established large urban centres.

Figure 4.7 shows the regions' rasters of local entropy of population density in the same period.

In the British Isles, the local entropy of cells corresponding to large cities tends to decrease from 1900 to 2015, as their urban areas expand and the population becomes less concentrated in the urban core and more uniformly spread in its surroundings. Urban fringes, on the other hand, retain large entropy values. Remark this phenomenon, for example, in the highly densely populated regions in North-Western England, that display increasingly distinguishable low entropy cores.

A similar phenomenon is noticeable in some areas in Italy as well, for example around Naples, in the South-West, and Milan, in the North-West, but is less apparent elsewhere.

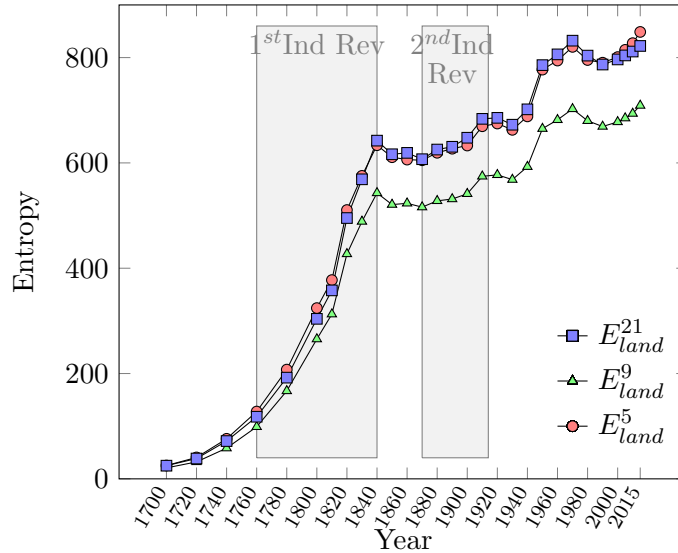
In South Asia, from 1700 to 1900, the relatively uniformly populated areas in the North, see the population become more concentrated, resulting in a more heterogeneous landscape of large and small entropy values. The progressive concentration of population proceeds to 2015, with generally higher entropy values and some areas of low entropy, resulting from consistently high densities, visible in Mumbai, on the Arabian Sea coast; in Uttar Pradesh and Bihar, in North India; Bangladesh; and along the Kerala coast, in the South West.

4.3.3.2 Historical evolution of local entropy in the British Isles

In Fig. 4.8 we show the evolution of the total local entropy of land use E_{land}^n and population density $E_{density}^n$ in the British Isles from 1700 to 2015 for various choices of window size $n = 5, 9, \text{ and } 21$. The value of n is found not to affect the overall trends.

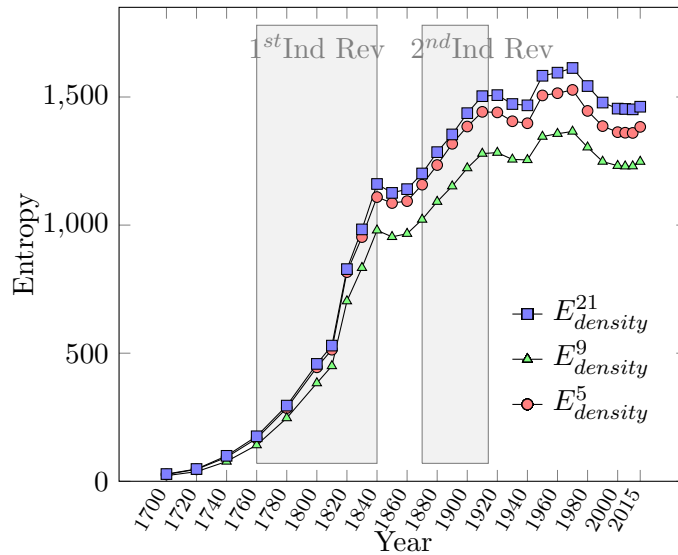
The evolution of E_{land}^n (Fig. 4.8a) and $E_{density}^n$ (Fig. 4.8b) reflects quite clearly the two industrial revolutions that Great Britain experienced between 1760-1840 and from the 1870s to the beginning of World War I in 1914. Both periods show a sustained growth in total local entropy, and were followed by more unstable decades of plateauing or slowly decreasing entropy values. Recent years are characterised by different behaviours in the two measures: whilst the entropy of land use continues to grow, the entropy of population density enters a phase of relative decline.

Total local entropy of land use in the British Isles



(a) Total local entropy E_{land}^n of land use.

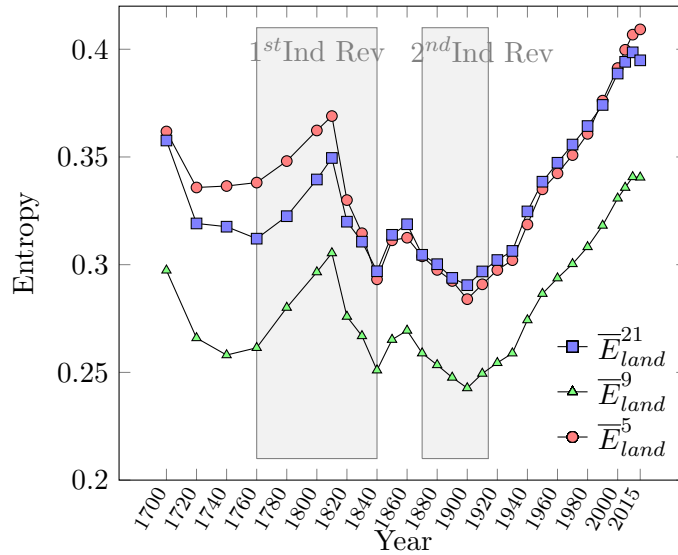
Total local entropy of population density in the British Isles



(b) Total local entropy $E_{density}^n$ of population density.

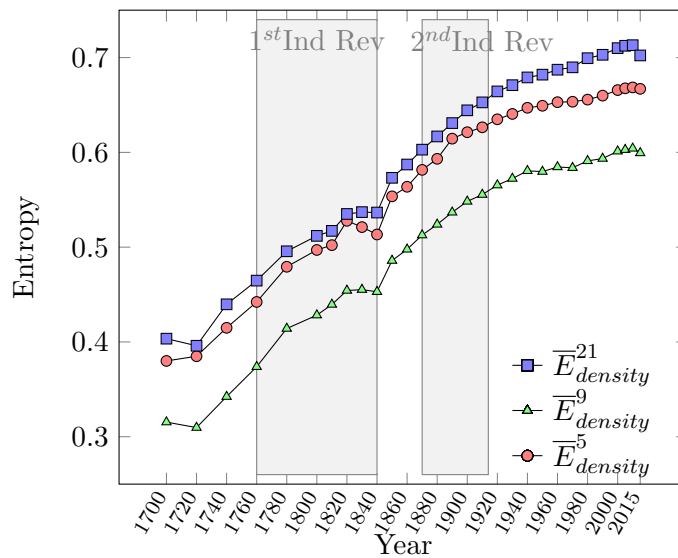
Figure 4.8: Total local entropy E_{land}^n of land use and $E_{density}^n$ of population density in the British Isles, for $n = 5, 9$, and 21 nearest neighbours. Whilst differences in absolute values exist, the general trends are not impacted by the window size, and are mostly consistent between the measures.

Mean local entropy of land use in the British Isles



(a) Mean local entropy \overline{E}_{land}^n of land use.

Mean local entropy of population density in the British Isles



(b) Mean local entropy $\overline{E}_{density}^n$ of population density.

Figure 4.9: Mean local entropy \overline{E}_{land}^n of land use and $\overline{E}_{density}^n$ of population density in the British Isles, for $n = 5, 9$, and 21 nearest neighbours. The general trends are not greatly affected by the window size but reveal important differences between the measures, with the entropy of population density steadily increasing while that of land use sees a more irregular evolution.

To distinguish the purely local effects from the mere increase in the number of cells occupied by settlements, we look at the mean entropy \overline{E}_{land}^n and $\overline{E}_{density}^n$ in Fig. 4.9. As above, the value of n does not affect the main trends. These are not constant during the two industrial revolutions and may represent more nuanced consequences of industrialisation.

The first industrial revolution sees at first a rapid increase in \overline{E}_{land}^n , followed by a sharp decrease starting from 1800 (Fig. 4.9a). This could reflect the fact that after a period of emergence of large number of new settlements, urban areas become better established and development happens mostly around them, creating more homogeneous urban and dense landscapes in their surroundings. We will see in Section 4.4 that this is indeed a time when industrial and urban cluster emerged in England and Wales.

The second industrial revolution, on the contrary, shows at first a decline in \overline{E}_{land}^n , followed by a modest recovery toward the dawn of the First World War after an all time minimum in 1900. This suggests that more complex phenomena were already in place concerning land use in the British Isles. As we will discuss in greater details in Section 5.2, this is a period of strong expansion of the railway, and the decline in the measure may denote growing concentration of population in compact cities before the rise of urban sprawling.

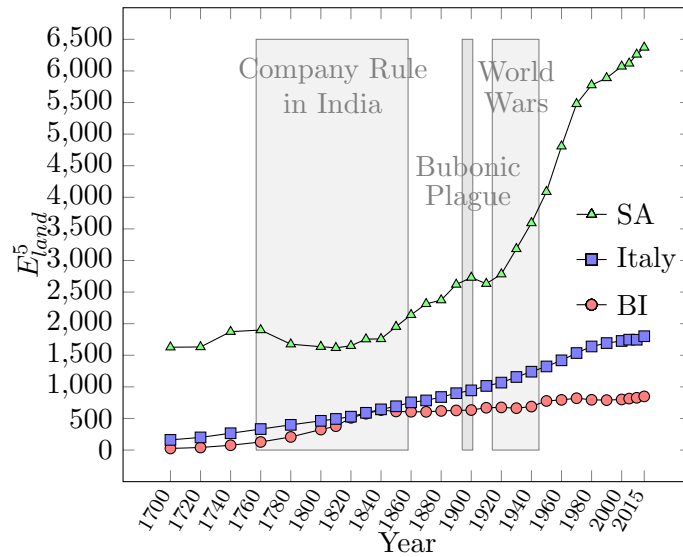
The evolution of $\overline{E}_{density}^n$ during the first industrial revolution is of fast growth which plateaus towards the end of the period, followed by an increase in subsequent decades until present day (Fig. 4.9b). A possible interpretation is that the technologies of the first industrial revolution allow only up to a certain maximum urban density, and, thus, stratification and heterogeneity of densities. It's only as new technologies emerge that maximum population density can again increase.

4.3.3.3 Historical evolution of local entropy - South Asia and Italy

In Figs. 4.10 and 4.11, we perform a similar analysis for the total and mean entropy of land use and population density, in the British Isles, South Asia, and Italy. The choice of $n = 5$ corresponds to the value minimising the average error in Table 4.1 in the period 1700AD to 2015AD. Nonetheless, the choice of n was found not to have a major impact on the overall trends.

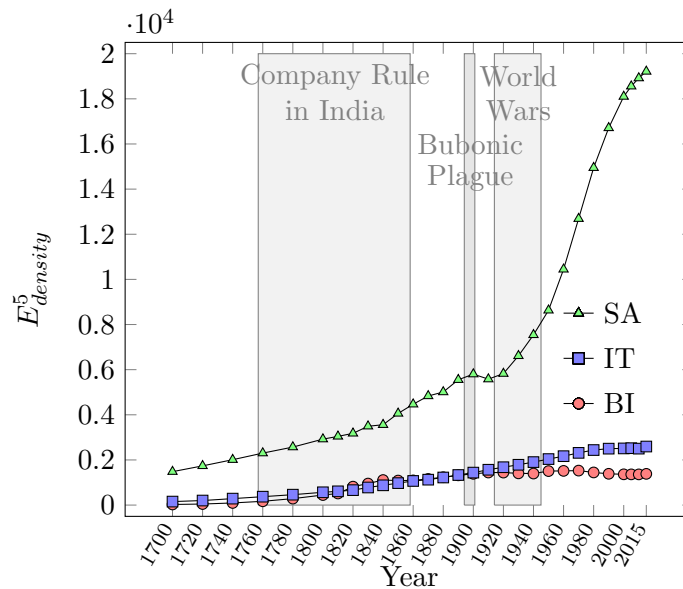
The analysis of the total local entropy in Italy reveals a steady growth in both E_{land}^5 (Fig. 4.10a) and $E_{density}^5$ (Fig. 4.10b), that slows down in the 1980s and reignites in 2015. This trend correlates with the country's demographic stagnation of the late 20th century, followed by a brief period of demographic

Total entropy of land uses in the British Isles (BI), South Asia (SA), and Italy (IT)



(a) Total local entropy E_{land}^5 of land use.

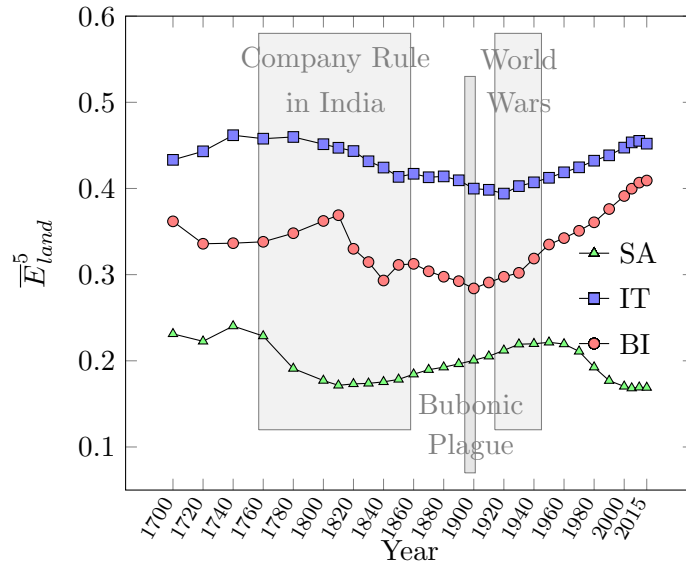
Total entropy of density in the British Isles (BI), South Asia (SA), and Italy (IT)



(b) Total local entropy $E_{density}^5$ of population density.

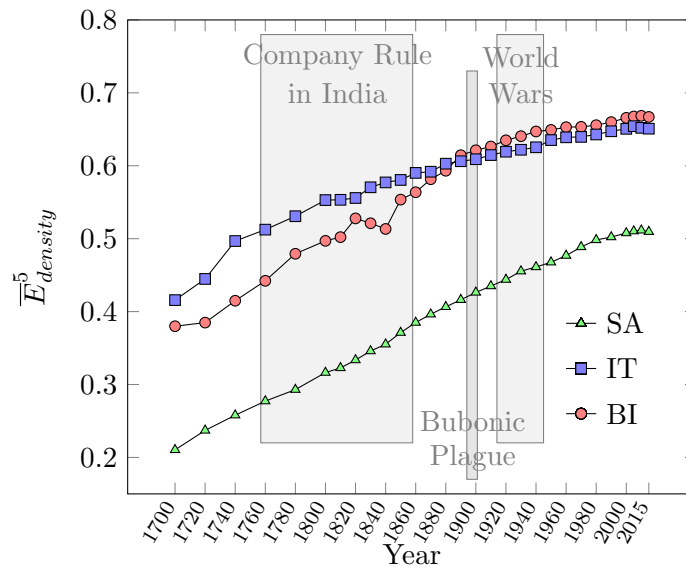
Figure 4.10: Comparison of local entropy \bar{E}_{land}^5 of land use and $\bar{E}_{density}^5$ of population density in the British Isles, South Asia, and Italy.

Mean entropy of land uses in the British Isles (BI), South Asia (SA), and Italy (IT)



(a) Mean local entropy \overline{E}_{land}^5 of land use.

Mean entropy of density in the British Isles (BI), South Asia (SA), and Italy (IT)



(b) Mean local entropy $\overline{E}_{density}^5$ of population density.

Figure 4.11: Comparison of local entropy \overline{E}_{land}^5 of land use and $\overline{E}_{density}^5$ of population density in the British Isles, South Asia, and Italy.

expansion in the new millennium.

The case of South Asia is more revealing. The change in E_{land}^5 in India correlates with some of the major historical events that affected the area (Fig. 4.10b). As the rule of the British East India Company on India is established in 1757, the entropy of land use enters a period of decline. The measure returns to grow in the early 19th century but it dips at the end of the century, at a time when India is struck by a major epidemic of bubonic plague. Twenty years later, the measure begins a long phase of fast expansion. The bubonic plague also correlates with a dip in $E_{density}^5$. The correlation of this latter measure with the beginning of the Company rule is absent, possibly masked by the general demographic growth.

The change in \bar{E}_{land}^5 in South Asia follows very closely the evolution of the political situation of the area as well (Fig. 4.11a). The Company rule corresponds to a period of fast decline and stagnation in \bar{E}_{land}^5 . The measure returns to grow in the early 19th century and continues to do so until the end of WWII, at a time when India regain its independence. On the other hand, the evolution of \bar{E}_{land}^5 for Italy does not seem to be deeply affected by major events in the 19th century, including the unification of the country in 1861, proceeding instead a steady decline begun in the mid 18th century. It is only in the 1920s that the measure returns to growth. This is a period which sees Italy under Fascist rule, with fast demographic expansion of Rome as capital, and the foundation of a few new urban centres in swampy areas reclaimed for cultivation, but the growth in the measure may also reflect more general industrialisation and urbanisation trends seen elsewhere in Europe, including the British Isles. The evolution of $\bar{E}_{density}^5$ is monotonically increasing and relatively smooth for both South Asia and Italy (Fig. 4.11b).

4.3.3.4 Historical evolution of local entropy - Observations

On a comparative note, the larger values of both $E_{density}^5$ and E_{land}^5 in South Asia compared to Italy and Britain reflect the large size of its system of settlements; the smaller values of both $\bar{E}_{density}^5$ and \bar{E}_{land}^5 reflect most likely its much lower urbanisation rate.

A long term historical vision help to shed some light on the inherently unpredictable nature of urbanisation. Looking at the past confirms that correlation, let alone causation, between major events and changes in urban form are often far from obvious. Political events were strongly correlated with the evolution of local entropy in the case of South Asia, but much less so in Italy.

Cities, in virtue of their density and enhanced social contact, are impractical places to live in without technological intervention [120]. It is thus unsurprising that it is the diffusion of technological innovation that seems to correlate more strongly with the evolution of the land and population entropy of urban systems, in good accord with the aforementioned theory of innovation waves [174].

4.4 Spatial distribution of systems of cities

Together with their population and land use of their surrounding territory, another fundamental morphological aspect of a system of cities is the locations of the cities. As mentioned in Section 2.5.1, the location of a city may depend on a large number of factors, including characteristics of the natural environment (such as natural resources or access to rivers and seas), but also on the location of other cities and on its interactions with them. Given the infinite amount of possible configurations a systems of cities could take in a given territory, any attempt to completely explain and unequivocally determine the location of cities using this or any list of factors alone would probably prove futile. Rather, the factors should be more aptly thought of as a large set of constraints that make some layouts more likely than others, while still allowing for many possible concrete realisations. When adopting this perspective, point processes become a natural tool to analyse the location of cities. Inhomogeneous processes may reflect the constraints of the local environment and cluster point processes or Gibbs processes may capture the reciprocal influence of cities locations in terms of clustering, attraction and repulsion.

In this Section, we model the spatial distribution of towns in England and Wales in 1801 and 1881 using point processes. We base our analysis on Bennett's data set [27], completed via linear interpolation and spatialised via Geonames, as detailed in Section 4.1.2. For each year, the data set includes only the settlements with a population of 2500 inhabitants or more, a population density of at least 1 inh/acre (approximately 250 inh/km²), and whose built-up area is spatial contiguous. These criteria are satisfied by 530 towns in 1801, and by 811 towns in 1881. In Section 4.4.1, we discuss the main spatial properties of the system and highlight the differences between the proto-industrial period in 1801 and the Second industrial revolution in 1881. We discuss homogeneous models for the patterns and their limitations in Section 4.4.2, and propose an inhomogeneous model for new cities emerged

between 1801 to 1881 in Section 4.4.3. We then analyse the patterns using the secondary structure of disc graphs, applying the first degree-based entropy. Remark that we will not explicitly use spatial entropy as a method of analysis. Whilst a theoretical literature on the entropy of point process models is in development (see for instance [10]), applied methods rely on subdividing the observed area into a grid and computing the Shannon entropy of the frequency of cell counts [155]. Spatial entropy methods, such as the one developed in Section 4.3, would also be applicable. This study would be very similar to the one on the local entropy of the UK described in the previous section, but neglecting the cities populations and focusing only on the locations, thus we preferred not to pursue it⁵. We conclude this Section with a few remarks and limitation in Section 4.4.5.

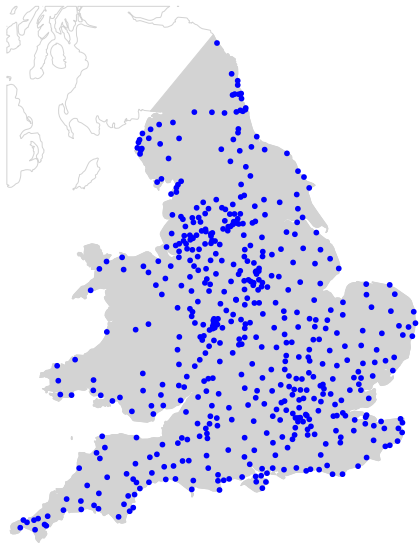
4.4.1 Town locations in England and Wales

Fig. 4.12 shows the locations of the towns in 1801 and 1881. The distribution in 1801 is far from uniform (Fig. 4.12a). Central Wales and some areas of the extreme North of England host fewer cities, as one may expect given their mountainous topography. Small clusters of cities characterise other areas of England, most noticeably around London in the South East, in the Midlands and in the North-West. The concentration of cities along the coast is higher than average, with the majority of Welsh and North-Eastern cities lying on the coast, as well as a large number of cities dotting the shores of the English Channel in the South. In the rest of the territory, cities look more regularly spaced than completely random (see, for instance, the East), suggesting the presence of some repulsion effects in their location, not uncommon in agrarian systems of settlements [116].

In 1881, the spatial pattern (Fig. 4.12b) maintains many of the earlier characteristics, including clustering in Northern England, Midlands, around London and along the coast and regularity outside of clusters, but the number of towns increases by more than half, from 530 to 811. Areas that had already higher than average density of centres in 1801, such as the North, the Midlands, and the surroundings of London seem to have reinforced their primacy. This can be in part ascribed to suburbanisation: especially from 1851, English cities that were already quite large grew further, their municipal area rapidly filled, and their population overflowed to the neighbouring territory, creating

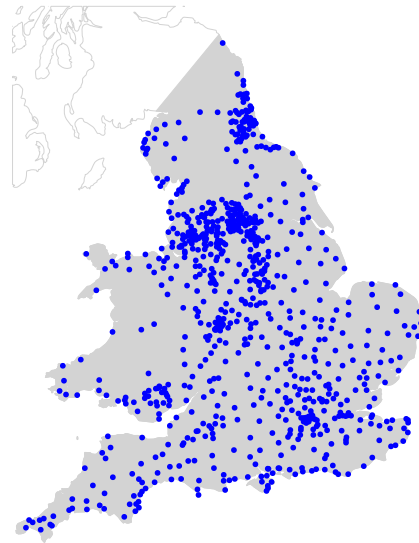
⁵For an example of an application of spatial entropy methods to the analysis of point processes in a different context, see Section 6.1.

Locations of English and Welsh towns in 1801



(a) Locations of English and Welsh towns in 1801. The distribution appears inhomogeneous: cities are more regularly distributed than random in the East of England and in Wales; aggregated in clusters in the North, the Midlands, and around London; and in higher number along the coast.

Locations of English and Welsh towns in 1881



(b) Locations of English and Welsh towns in 1881. The distribution is inhomogeneous: clustering is apparent in North East, in the North, Midlands, South Wales, and around London. The pattern is more regular elsewhere, with above average concentration along the coast.

Figure 4.12: Locations of English and Welsh towns in 1801 and 1881. New centres primarily appear around established population centres in the North of England and around London. Other areas where new centres sprout are South Wales and North East.

new towns or transforming existing rural settlements into urban ones [211]. Exceptions to this are new clusters of cities emerged around the estuary of River Tyne, in the extreme North East of England, and in Southern Wales, that emerged as a prominent coal and iron mining region in the 19th century [211]. More generally, if we look exclusively at new cities, we notice that they tend to form clusters, with most new singletons being new port cities along the Southern and Western coast (see also Fig. 4.15a in Section 4.4.3).

4.4.2 Homogeneous models for town locations in England and Wales

The above description of the spatial patterns already suggests that homogeneous processes may not be a suitable models for the distribution of cities in neither 1801 nor 1881.

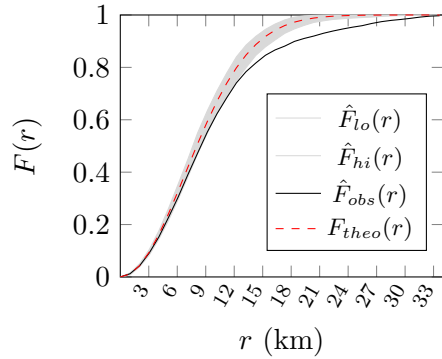
The null hypothesis of Complete Spatial Randomness is formally tested in Fig. 4.13, where we show, for 1801 and 1881, the observed values \hat{F}_{obs} and \hat{G}_{obs} of the Empty Space Function F and the Nearest Neighbour Distance Function G , and compare them to their theoretical values F_{theo} and G_{theo} for Homogeneous Poisson Processes with the same number of points. For each test function and year, the plots also show confidence bands, obtained by performing 100 simulations of the proposed model, and selecting the smallest and largest values ($\hat{F}_{lo}(r)$ and $\hat{F}_{hi}(r)$, and $\hat{G}_{lo}(r)$ and $\hat{G}_{hi}(r)$ respectively) at each radius r .

In all cases, the observed values $\hat{F}_{obs}(r)$ and $\hat{G}_{obs}(r)$ fell outside of the confidence bands for many values of r , confidently rejecting the null hypothesis. A more detailed interpretation of the values of $\hat{F}_{obs}(r)$ and $\hat{G}_{obs}(r)$ is not possible at this stage, as we haven't yet established homogeneity, and deviations of the values from $F_{theo}(r)$ and $G_{theo}(r)$ due to interactions between the points would not be distinguishable from those originated from other sources of inhomogeneity.

The apparent presence of clusters invites us to formally test the hypothesis of homogeneous cluster processes, i.e., cluster processes in which the parent process is a homogeneous Poisson process. Fig. 4.14 performs the same tests with the F and G functions comparing the values observed for the locations of cities in 1801 and 1881 to those of the best fitting Matérn cluster processes according to the method of minimum contrast, applied to Ripley's K -function (See Section 2.5)⁶.

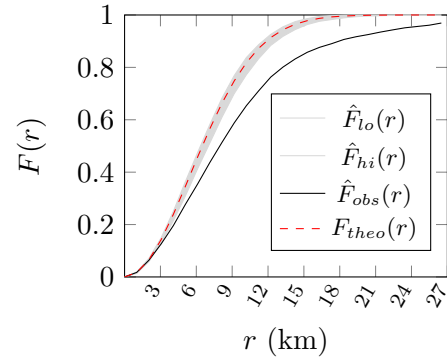
⁶In 1801, the best fitting homogeneous Matérn cluster process has an intensity of 0.000337

CSR vs towns in 1801, F -test



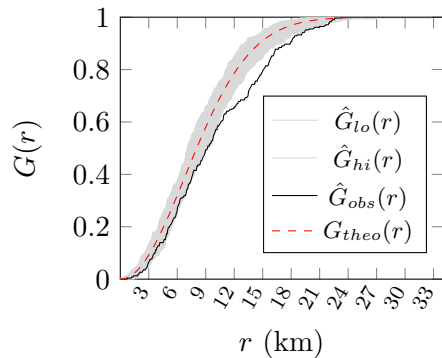
(a) The Empty Space Function F of the location of English and Welsh towns in 1801 lies below the lower limit of the confidence band when $r \geq 10\text{km}$, rejecting the Complete Spatial Randomness assumption.

CSR vs towns in 1881, F -test



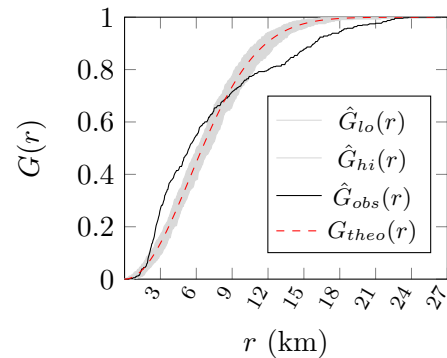
(b) The Empty Space Function F of the location of English and Welsh towns in 1881 lies below the lower limit of the confidence band when $r \geq 4\text{km}$, which is incompatible with the Complete Spatial Randomness assumption.

CSR vs towns in 1801, G -test



(c) The Nearest Neighbour Distance Function $G(r)$ of the locations of towns in 1801 lies below the lower limits of the confidence band when $r \approx 8\text{km}$, and when $10\text{km} \leq r \leq 22\text{km}$, rejecting the hypothesis of Complete Spatial Randomness.

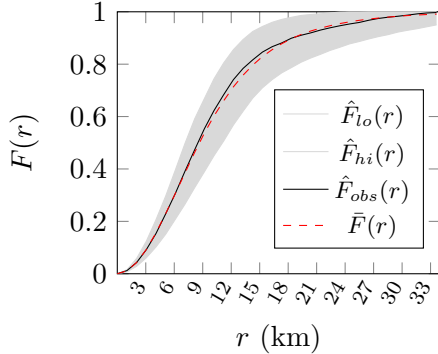
CSR vs towns in 1881, G -test



(d) The Nearest Neighbour Distance Function $G(r)$ of town locations in 1881 exceeds the upper limit of the confidence band when $2.2\text{km} \leq r \leq 7.2\text{km}$, and lies below the lower limit when $10.5\text{km} \leq r \leq 23\text{km}$, clearly violating the Complete Spatial Random assumption.

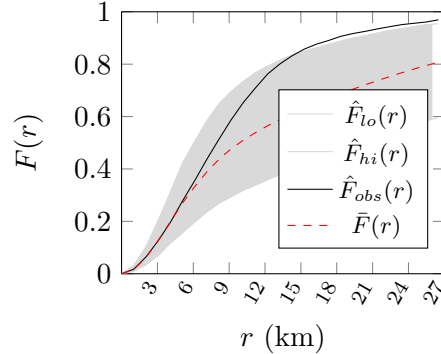
Figure 4.13: Test Complete Spatial Randomness of the distribution of towns in England and Wales in 1801 and 1881 using the Empty Space Function F and the Nearest Neighbour Distance Function G . In all cases, the test show that the patterns are incompatible with the assumption of Complete Spatial Randomness.

Matérn vs towns in 1801,
F-test



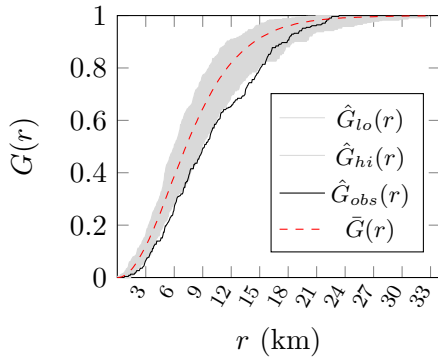
(a) The Empty Space Function F of the location of English and Welsh towns in 1801 falls within the confidence band of the best fitting Matérn cluster process, but this may be due to inhomogeneity.

Matérn vs towns in 1881,
F-test



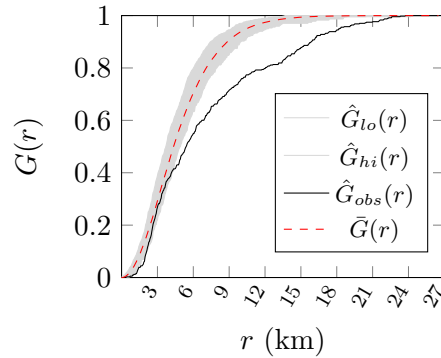
(b) The Empty Space Function F of the location of English and Welsh towns in 1881 lies above the confidence band of the best fitting Matérn cluster process when $r \geq 15$ km.

Matérn vs towns in 1801,
G-test



(c) The Nearest Neighbour Distance Function $G(r)$ of the locations of towns in 1801 lies below the lower limits of the confidence band of the best fitting Matérn cluster process when $1.8\text{km} \leq r \leq 7\text{km}$, and again when $8.2\text{km} \leq r \leq 15.6\text{km}$.

Matérn vs towns in 1881,
G-test



(d) The Nearest Neighbour Distance Function $G(r)$ of the locations of towns in 1881 lies below the lower limits of the confidence band of the best fitting Matérn cluster process when $0.7\text{km} \leq r \leq 2\text{km}$, and again when $4.3\text{km} \leq r \leq 22\text{km}$.

Figure 4.14: Testing the hypothesis that a Matérn cluster processes built on a Homogeneous Poisson Process may describe the spatial distribution of towns in England and Wales in 1801 and 1881. The best fitting Matérn Cluster processes are identified via the minimum contrast method applied to Ripley's K function. For both 1801 and 1881, at least one of the Empty Space Function F or the Nearest Neighbour Distance Function G falls outside the confidence band of the model, thus rejecting the hypothesis.

In Fig. 4.14a, we see that the observed values of \hat{F}_{obs} in 1801 are compatible with those of the best fitting model and its realisations. Yet, the hypothesis is rejected by the G function (Fig. 4.14c), for which the observed values $\hat{G}_{obs}(r)$ fall below the lower limit of the confidence band when $1.8\text{km} \leq r \leq 7\text{km}$, and again when $8.2\text{km} \leq r \leq 15.6\text{km}$, indicating that the model is excessively clustered compared to the real distribution. Figs. 4.14b and 4.14d show that the Matérn cluster process for 1881 is not suitable: with both the F and G functions pointing towards excessive clustering of the model.

Similar experiments also reject other homogeneous cluster models, such as the Thomas process.

4.4.3 Inhomogeneous models for new town locations in England and Wales

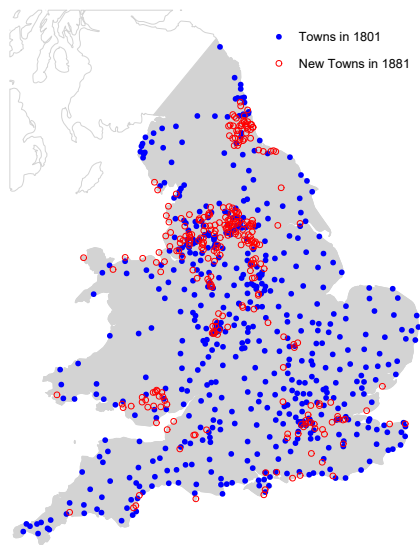
The tests in the previous section confirm what the maps in Fig. 4.12 already suggested: a suitable point process to describe the locations of English and Welsh cities is likely to be an inhomogeneous one.

The inhomogeneous intensity $\lambda(lon, lat)$ of cities in England and Wales may depend on a potentially very large number of covariate spatial densities. Instead of embarking on the likely unsuccessful identification of all of them, here we take a different approach. We consider the locations of English and Welsh towns in 1801 as our base knowledge of the system, encoding a lot of what can be known about the local natural constraints at play. We use this local knowledge to build a model for the new towns that emerged from 1801 to 1881.

In Section 4.4.1, we have already highlighted some properties of the locations of new cities displayed in Fig. 4.15a. Amongst these the most prominent, confirmed by historical studies [211], is that many new cities formed through a process of suburbanisation, creating clusters around already existing large centres, which tended to be located in generally more densely populated areas. This observation suggests that a suitable model for the formation of new cities might be a cluster process, in which both the number of clusters and the number of points in the cluster is positively influenced by the existing density of centres.

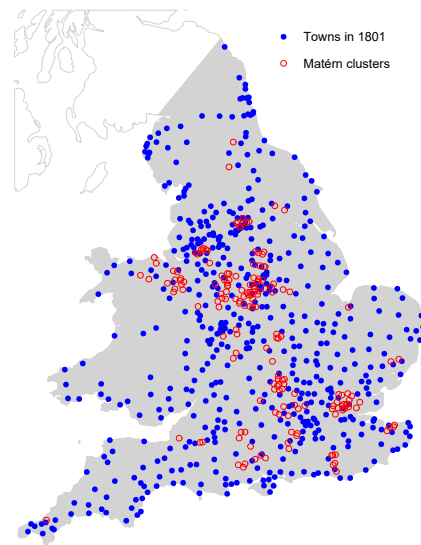
points/km², i.e., a parent point every 2967.359km² on average, and a cluster radius of 50.228km, with an average of 10.14 points in each cluster. In 1881, the best fitting model has a lower parent density of 0.000156 points/km², corresponding to a parent point every 6410km², and smaller cluster radius of 35.776km, but much higher density of daughter points within each cluster: 38.505 on average.

New towns in England and Wales from 1801 to 1881



(a) Locations of new English and Welsh towns emerged between 1801 and 1881. The new towns are clustered primarily around already denser areas in the North, Midlands, and South East. New clusters are visible in the North East and South Wales, and sporadic towns emerge elsewhere.

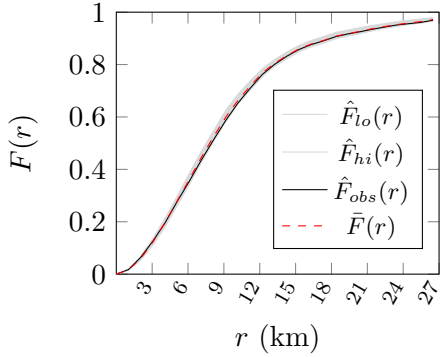
Example of inhomogeneous Matérn clusters



(b) Locations of English and Welsh towns in 1801 and the clusters of a realisation of the inhomogeneous Matérn cluster model. Most of the large clusters of simulated new towns are located in areas with higher density in 1801 by design. A few isolated centres emerge.

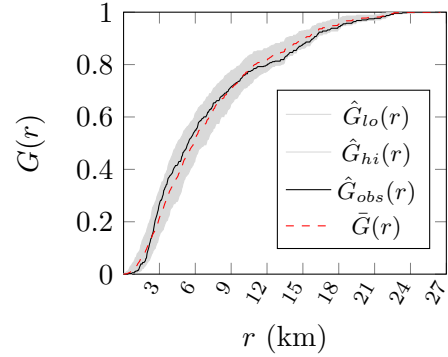
Figure 4.15: Comparison of the real distribution of new towns emerged in England and Wales from 1801 to 1881, and a realisation of the inhomogeneous Matérn cluster model, superimposed to the distribution of towns in 1801. The model captures the new clusters in dense areas, but fails to account for sporadic large clusters in areas with low density in 1801, such as South Wales and North East, and, more generally, of new centres along the coast.

Inhomogeneous Matérn, new towns, F -test



(a) The Empty Space Function F of the location of new English and Welsh towns between 1801 and 1881 falls within the confidence bands of the selected inhomogeneous Matérn cluster model.

Inhomogeneous Matérn, new towns, G -test



(b) The Nearest Neighbour Distance Function G of the location of new English and Welsh towns between 1801 and 1881 falls within the confidence bands of the selected inhomogeneous Matérn cluster model, except when $0.8\text{km} \leq r \leq 0.9\text{km}$.

Figure 4.16: Testing the hypothesis that an inhomogeneous Matérn cluster process may describe the new towns emerged in England and Wales from 1801 to 1881. The selected model assumes that the predicted intensity $\hat{\kappa}$ of the parents points and of the points in each cluster reinforces the density $\kappa_{1801}(\text{lon}, \text{lat})$ in 1801. The summary functions reject the hypothesis, but point towards an improvement in the goodness of fit of the selected model, which is incompatible only for very small values of r .

In Fig. 4.16, we test the hypothesis that the locations of new towns in 1881 may be described via an inhomogeneous Matérn cluster process, for which the intensity of the parent point reinforces the existing density in 1801, and the average cluster size is larger in areas that were denser in 1801⁷. The model is compatible with both test functions F , G for most values of r . The only exception is when $0.8\text{km} \leq r \leq 0.9\text{km}$, for which the model appears excessively clustered according to $G(r)$ as opposed to the slight repulsion evidenced in the real pattern. Given its small scale, the repulsion observed in the real data may be due to the surface area of the town, which physically impedes other towns

⁷More precisely, we take the cluster radius to be 12km, or roughly equivalent to the radius of a large metropolitan area. We let $\lambda(\text{lon}, \text{lat})$ be the intensity of the process in 1801, computed via kernel smoothing. Then we take the intensity $\kappa(\text{lon}, \text{lat})$ of the parent points to be $\kappa(\text{lon}, \text{lat}) \propto \lambda(\text{lon}, \text{lat})^3$ and the intensity $\mu(\text{lon}, \text{lat})$ of daughter points in the disc around a parent point to be $\mu(\text{lon}, \text{lat}) \propto \lambda(\text{lon}, \text{lat})^4$. These values are somewhat arbitrary and found by trial and error, in absence of a more suitable and efficient procedure to fit inhomogeneous point processes to an existing pattern.

to be so close to its centre.

A realisation of the inhomogeneous Matérn cluster model is presented in Fig. 4.15b. The model captures reasonably well the new clusters around London and in the North of England, as well as the sporadic smaller cluster of cities emerged outside of the areas of highest density in 1801. The model, though, does not capture the new urbanisation phenomena described in Section 4.4.1, such as the emergence of ports and clusters in areas which in 1801 hosted relatively few towns, such as the River Tyne estuary and South Wales. This is to be expected, given that the model was designed to reinforce the existing structure of the urban system.

4.4.4 Degree entropy of disc graphs

We conclude our study of the spatial distribution of English and Welsh towns by analysing the evolving inhomogeneity of the pattern via an auxiliary structure: disc graphs.

For each year $y \in \{1801, 1811, \dots, 1881\}$ and non-negative distance threshold r , we define the network $G_{y,r}$ by taking as nodes all towns that were part of the system in year y , and connecting a pair of towns by a link if their geographic distance is at most r . We study the first degree-based entropy of the networks $\{G_{y,r}\}_{y,r}$ as r is gradually increased from $r = 0\text{km}$, for which one obviously obtains the empty graph, to $r = 672\text{km}$, the maximum intercity distance, for which the complete graph is retrieved.

In Fig. 4.17, we show the networks $G_{y,r}$ for thresholds $r = 13\text{km}$ and $r = 30\text{km}$ in the years $y = 1801$ and 1881 .

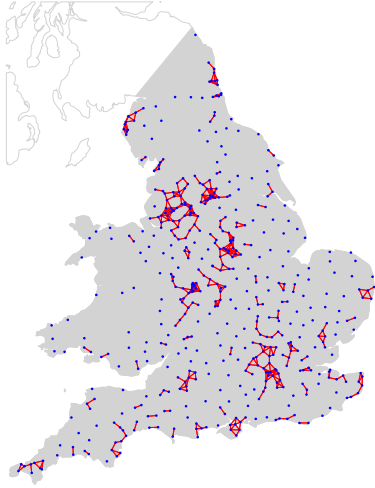
For $r = 13\text{km}$, a large number of cities remain disconnected and the networks are divided into several, mostly small components, including a large number of dyads. Some of the components in $G_{1801,13\text{km}}$ merge into larger ones in $G_{1881,13\text{km}}$. That is the case for the towns around the Pennines and the Tyne estuary in the North of England, Southern Wales, and near Birmingham and London.

For $r = 30\text{km}$, in both 1801 and 1881, most towns are part of a single network component, with the exception of an isolated dyad in central Wales⁸, and a few isolated peripheral nodes⁹. The networks are characterised in large

⁸The neighbouring but otherwise remote towns of Newton and Llanidloes.

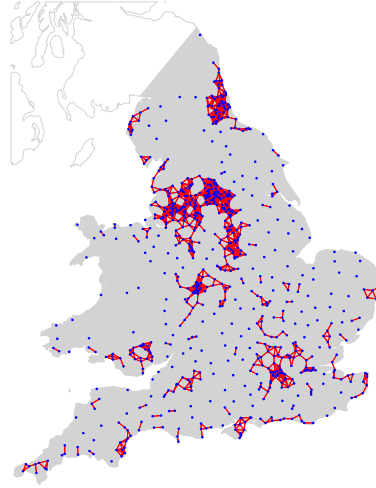
⁹Berwick in Northumberland; Minehead in Somerset, South-West England; Aberystwyth on the coast of central Wales; and, in 1881, Holyhead, a former fishing village in Northern Wales that in the early 19th century was redeveloped into a port town with links to nearby Ireland, and railway connection to Chester and Shrewsbury in England.

Disc graph in 1801, $r = 13\text{km}$



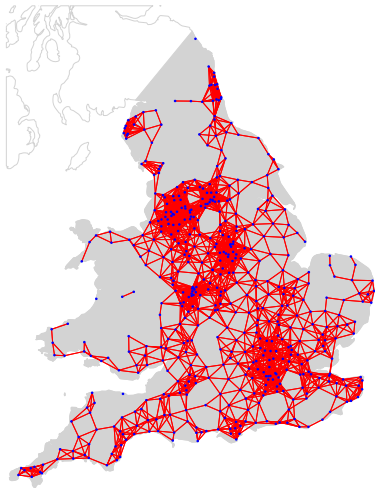
(a) The disc graph in 1801 for a small threshold of $r = 13\text{km}$ is extremely sparse, with only a few, small clusters and several isolated towns.

Disc graph in 1881, $r = 13\text{km}$



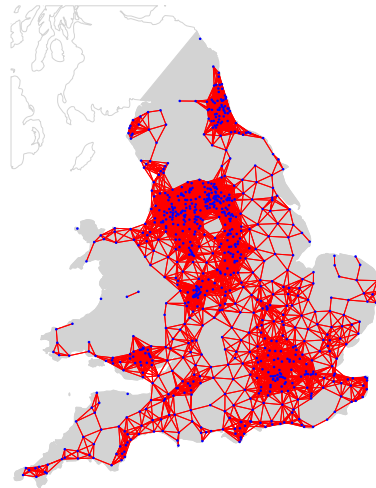
(b) The disc graph in 1881 for a small threshold of $r = 13\text{km}$ shows larger clusters, but they generally contain several nodes with small degree.

Disc graph in 1801, $r = 30\text{km}$



(c) For a moderate threshold $r = 30\text{km}$ almost all towns in 1801 are connected. The graph resembles a lattice with a few denser areas.

Disc graph in 1881, $r = 30\text{km}$



(d) When $r = 30\text{km}$, the disc graph in 1881 connects almost all towns, with more numerous, denser and larger clusters compared to 1801.

Figure 4.17: Disc graph of the pattern of English and Welsh towns in 1801 and 1881, for the thresholds $r = 13\text{km}$ and $r = 30\text{km}$.

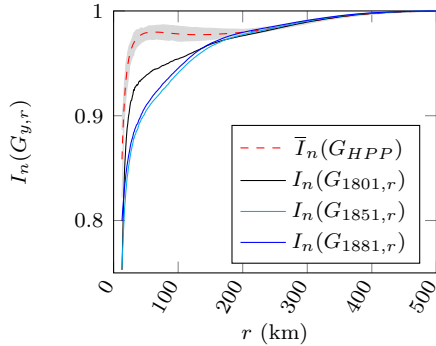
part by a lattice structure, which is almost the same in 1881 as in 1801, implying a large number of relatively low degree nodes. The denser clusters of towns in 1881, though, create a small number of nodes with very high degree (up to 85), and thus a clearly less balanced degree sequence. This is captured by the appropriately normalised first degree-based entropy, that is smaller in 1881 than in 1801 for this threshold.

Fig. 4.18 shows the degree entropy of the disc graphs $G_{y,r}$ when $y = 1801, 1851, \text{ and } 1881$, and $13\text{km} \leq r \leq 500\text{km}$. The entropy is normalised according to the graph order n or both the graph order n and size m , as detailed in Section 3.4. The entropy of the graphs are not shown for $r > 500\text{km}$ because for very large values of the distance threshold, the network is almost complete. On the one hand, this means that the order-normalised degree entropy $I_n(G_{y,r})$ is almost 1 in all years. On the other hand, the difference between the minimum and maximum unnormalised entropy of graphs with the same order and size as $G_{y,r}$ is very small. Normalising by this difference makes the order- and size-normalised entropy $I_{n,m}(G_{y,r})$ unstable, in a non-meaningful way. Remark, however, that a normalisation is still advisable in most cases, because, when the entropy is not normalised, the effect of the increase in the number of towns on the measure is so large that any other structural property is masked.

In Fig. 4.18a, the entropy $I_n(G_{y,r})$ shows a monotonic increase in all years as a function of r , reflecting primarily the growth in the network sizes. For very small values of r , $I_n(G_{1801,r})$ is smaller than both $I_n(G_{1851,r})$ and $I_n(G_{1881,r})$, but the relation is reversed when $14\text{km} \leq r \leq 175\text{km}$, and $16\text{km} \leq r \leq 138\text{km}$, respectively. The large drop in entropy from 1801 to 1851 for these values of r is explained by the increased clustering of the pattern of towns, especially around Manchester, Sheffield and Leeds, and, to a lesser extent, on the river Tyne. By 1881, the entropy has grown modestly compared to 1851. This reflects the fact that the new towns appeared in the period from 1851 to 1881 are more widespread, with a reinforcement of the main Northern cluster and the river Tyne area, but also the emergence of new clusters in industrial valleys of South Wales, and a large number of towns in the South East of England, on the Southeastern and Northwestern coast, and along the newly constructed railway corridors from London to the North¹⁰. Note, however, that in all years the entropy is smaller than $\bar{I}_n(G_{HPP})$, the first degree-based entropy expected

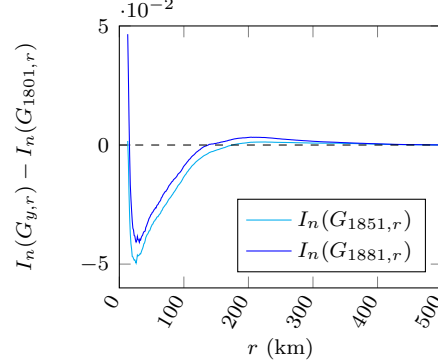
¹⁰We'll discuss the railways and their impact on travel distance in Section 5.2.

**Degree entropy of disc graphs
(normalised by n)**



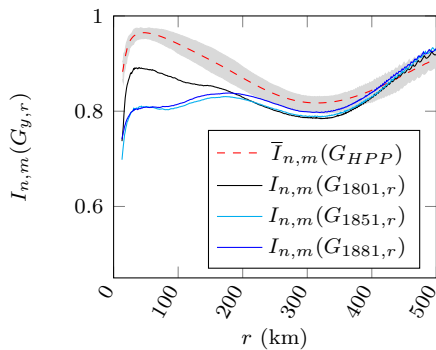
(a) Entropy of the disc graphs in 1801, 1851, and 1881 for changing disc radius threshold r , normalised by n . In all years, the entropy grows monotonically as a function of r . This is mostly due to the increase in the graphs' sizes as r increases.

**Degree entropy difference
(normalised by n)**



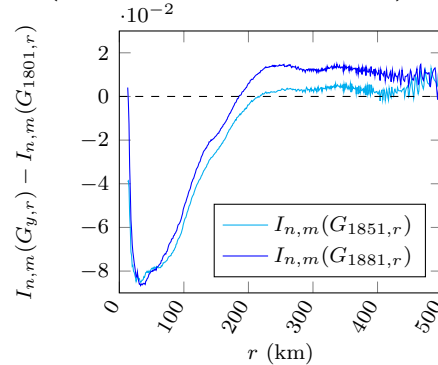
(b) Difference between the entropy of the disc graphs in 1851 and 1801, and 1881 and 1801, for changing disc radius threshold r , normalised by n . The thus normalised entropy reveals clustering for $r < 175\text{km}$, but is not very sensitive for medium and large r .

**Degree entropy of disc graphs
(normalised by n and m)**



(c) Entropy of the disc graphs in 1801, 1851, and 1881 for changing r , normalised by n and m . For small r , the degree sequence became more unbalanced by 1851 than in 1801, and then changed very little. In 1801, the system was more balanced for small r than medium values, unlike in 1851 and 1881.

**Degree entropy difference
(normalised by n and m)**



(d) Entropy difference of the disc graphs in 1851 and 1801, and 1881 and 1801, for changing r , normalised by n and m . The measure is sensitive at small and medium scale, picking up the change in the spatial distribution, but becomes noisy at large scales.

Figure 4.18: Normalised degree entropy of the disc graphs in 1801 and 1881, for the disc radius threshold $13 \text{ km} \leq r \leq 500\text{km}$. The normalisation by the order n is affected by the dependency between r and the size m . This is not the case for the normalisation by both n and m , but this measure is too sensitive and noisy for large r , when the disc graph approaches the complete network.

for a disc graph built on a Homogeneous Poisson Process¹¹, indicating greater clustering than complete spatial randomness at all points in time. For larger r , $G_{1851,r}$ and $G_{1881,r}$ have larger entropy than $G_{1801,r}$ but the measures are almost indistinguishable. This is more clearly visible in Fig. 4.18b, where the difference $I_n(G_{y,r}) - I_n(G_{1801,r})$ is shown for $y = 1851$ and 1881 . The results are partially explained by the fact that when the entropy is normalised by the order n alone, graphs with large sizes achieve consistently high entropy, which masks the difference between them. In other words, $I_n(G)$ is not a sensitive measure when G has large size.

In Fig. 4.18c, the entropy $I_{n,m}(G_{y,r})$, normalised by both n and m , shows a more varied behaviour as a function of r . In all years, the entropy increases for small r , which is expected as many pairs of cities become connected and thus several nodes see their degree increase from 0 to 1, creating a more balanced degree sequence¹². The effects are more marked in 1801 than in 1851 and 1881, because in later years the patterns are more clustered, and thus an increase in r even to modest values also yield an increase in the number of relatively high degree nodes, counterbalancing the uniformity of the degree sequence. In 1801, the entropy starts declining as a function of r as soon as $r > 40\text{km}$, as all nodes become connected and the density in the relatively sparser clusters increases. In 1851 and 1881, we observe a general increase in the entropy as r increases to 170km , ascribable to the fact that the clustering is mostly happening at the local and regional scale. At a finer look, though, the entropy sees a slight decrease for $43\text{km} \leq r \leq 84\text{km}$ in 1851 and plateaus in 1881. The interval's extrema are consistent with the radius of the main Northern England cluster and the distance between the barycentres of the main Northern England cluster and the Tyne estuary cluster. As r reaches the intermediate value of 170km , $I_{n,m}(G_{y,r})$ begins to decline in all years, and returns to grow markedly for values of $r > 320\text{km}$ in all years, as the network grows towards the complete graph¹³. This is likely a consequence of the geometric shape of England and Wales rather than any other property of the pattern, as revealed by a comparison with the normalised entropy $\bar{I}_{n,m}(G_{HPP})$ Homogeneous Poisson Process in the same territory. The slightly higher values in 1881 for large r may be due to the larger number of clusters, which yield a marginally more balanced degree distribution, but the structure of the system

¹¹The intensity of this process is irrelevant, since we normalise by the number of towns n .

¹²We have observed, in fact, that the first degree entropy is particularly sensitive to increase when a previously isolated node increases its degree to 1.

¹³Remark that $r = 320\text{km}$ is approximately the radius of the networks.

is similar in term of balance at this scale in all years. Again, the described features are clarified by the difference $I_{n,m}(G_{y,r}) - I_{n,m}(G_{1801,r})$ in Fig. 4.18d. This figure also reveals that the normalised entropy becomes noisy for large r , thus suggesting that this measure is too sensitive to analyse graphs that are almost complete.

4.4.5 Towards more complete point process models of town locations

We have used a point process approach to describe the location of towns in England and Wales in 1801 and 1881, with a particular focus on new cities emerged during the intervening period. The experiments show the necessity to use rather complex inhomogeneous models to capture both the variations in the natural environment and the interactions between the cities. Whilst an inhomogeneous Matérn cluster process is able to successfully capture some of the salient properties of the new towns and generate patterns whose first and second order properties are generally compatible to those observed in reality, several challenges emerge.

The first is that of creating a model that can at the same time reproduce the observed reinforcement of the existing patterns while allowing for the emergence of new significant structures. This will necessarily require the introduction of supplementary information in the form of relevant spatial covariates. These may be of multiple kinds, and their identification is not trivial. For instance, the emergence of a cluster of cities in Southern Wales over the course of the 19th century was strongly tied to the mining of iron. The presence of such mineral may not have been such a crucial factor for economic development before the consolidation of the industrial revolution, in times of lower demand and ineffective extraction technologies, and thus would have not been seen as a sufficiently attractive factor to establish such cluster in 1801.

The second is the careful consideration of border issues, and the influence of cities beyond the observed areas. The expansion of ship-building industries and overseas commerce, for instance, were at the basis of the emergence of port cities, which would have been impossible to predict given the distribution of cities in England and Wales alone.

The third is the necessity to account for the surface area of cities, which imposes a minimum distance between them that may be incorrectly picked up by point process methods as a sign of regularity at small scales. For this, more complex objects and methods, such as random closed sets [198], may be more

suitable.

A final one, which we have not discussed so far, is the inhomogeneity of distance when one substitutes more realistic metrics such as travel time to the Euclidean (or great circle) metric implied in a point process approach. In recent years, new mathematical and computational methods have been developed to study random lines [125] and point processes on linear networks [220, 12]. These methods are a very promising way to better account for the impact of geographical discontinuities such as rivers and mountains, and of transport infrastructure on distances and interactions between towns. In the next Chapter, we focus in greater detail on transport infrastructure, its role in connecting cities, and the quantification of the heterogeneity of such connections.

Chapter 5

Cities and their Connections: Transportation and Polycentricity of Systems of Cities

After having explored the population sizes, surrounding territory, and spatial arrangement of systems of cities, we now turn our attention to the transport networks that connect them. A transportation link, such as a scheduled bus or flight, may be considered as an example of a functional connection between two cities, enabling the flow of people, goods, and ideas. On the other hand, transport infrastructures, such as roads and railways, are also an integral part of the morphological arrangement of cities, as they modify the spatial relations between them by shrinking the travel time and generating non-uniform metrics on the space. In this chapter, we adopt the latter perspective.

In Section 5.1, we compute the first degree-based entropy of the worldwide network of flights, and see how the different normalisation methods described in Section 3.4 yield different conclusions on the level of centralisation of the system. In Section 5.2, we look at the evolution of the railway system in England and Wales between 1851 and 1881. We use the first degree-based entropy to analyse the evolving spatial organisation of the pattern of cities, applying it to disc graphs built using minimum theoretical travel time instead of geographical distance. This extends our work in Section 4.4.4. We also study the evolution of the entropy of interactions of a theoretical network of optimal flows retrieved via an entropy-maximisation model. We show how accounting

for the presence of transport infrastructure can yield a more realistic and human-centred view of morphological polycentricity.

5.1 Flight network

We consider the worldwide network of flights as recorded in the Openflights database [166] (Fig. 5.1). For each pair of cities connected by a direct flight, the database provides information on which airlines operate on the route. We neglect this information, and merge in one link all routes with the same endpoints, irrespective of the number of airlines by which they are served. This ensures that the network is simple.

Worldwide flight network



Figure 5.1: Network of worldwide flights. Longer distance flights are depicted in darker shades for aesthetic reasons, but distance plays no role in this analysis

The network contains 3334 airports, connected by 19203 distinct routes. The network is divided into 7 components, but it has a clear giant component containing 3308 nodes. The other six components are small local systems of minor airstrips, located in remote areas of the United States and its territories, in New Caledonia, and in Namibia. Here we focus on the main component, that we call G .

G contains $n = 3308$ airports, connected by $m = 19178$ distinct routes. A quick computation shows that the unnormalised entropy $I(G) \approx 10.14$.

In Table 5.1, we compare the three normalisation of $I(G)$ presented in Section 3.4. $I_n(G) \approx 0.68$, which gives the network a relatively high entropy with respect to the minimum and maximum values other connected graphs of order n can achieve. When compared to graphs with given size m , the normalised entropy $I_m(G) \approx 0.33$, which means that the system is quite centralised given

Table 5.1: Normalised first degree-based entropy of the flight network.

	Constraint		
	order n	size m	n and m
I_{min}	6.85	7.62	7.99
I_{max}	11.69	15.23	11.69
$\tilde{I}(G)$	0.68	0.33	0.58

the “budget” of connections. The most realistic situation, where we compare G to graphs with its same order n and size m , yields $I_{n,m}(G) \approx 0.58$, showing that the flight network sits in between centralised and balanced systems.

5.2 Transportation networks, network entropy and polycentricity

Consider a region R , with a set V of cities. To each city $u \in V$, associate a pair of coordinates $(lon_u, lat_u) \in R$, for example, the coordinates of its city hall or other landmark. The travel cost c_{uv} to move one unit of population from u to another city v may be affected by several factors: geographical distance, of course, but also the presence of transport infrastructure in R , and the state of its technology. In other words, the cost of travel metric in the region R is generally not uniform nor constant in time.

Here we take the cost of travel $c_{uv} = t_{uv}$, where t_{uv} is the minimum theoretical time to travel from u to v or vice versa¹. Via the example of the system of English and Welsh towns in the period from 1851 to 1881, and its rapidly evolving railway network, we show the effects of the transportation network on the spatial organisation of the system of cities, and their potential interactions.

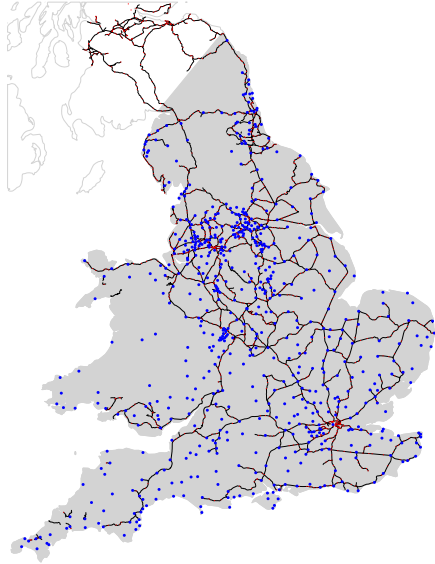
5.2.1 Data sets

For the historical population sizes of English and Welsh towns, we use again the UK Data Service database [27], completed via linear interpolation, as discussed in Section 4.1.2.

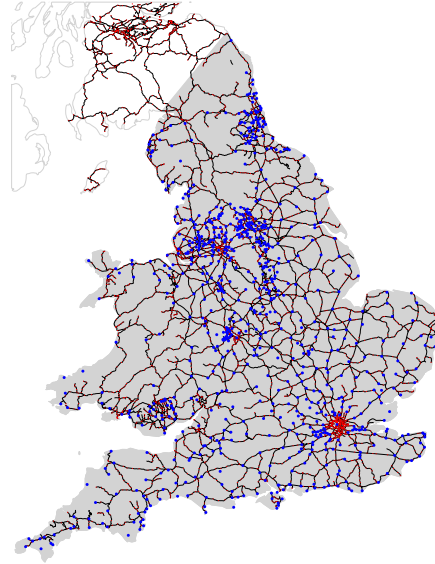
To estimate travel time between localities at different times, we assume that people use the fastest land transport available. We use data on historic

¹This metric need not be symmetric in general, but in our computation of travel we do not account for this asymmetry. See Section 7.2 for a discussion on limitations and future improvements of this method.

Railway network in 1851



Railway network in 1881



(a) Railway network in 1851, with stations and towns locations. There are 656 towns. The network is sparse and connects primarily the main urban clusters, with large areas having no lines nor stations.

(b) Railway network in 1881, with stations and towns locations. The number of towns has increased to 811. The network is considerably denser, with more stations, and most towns are directly connected to it.

Figure 5.2: Railway network in 1851 and 1881, with stations (in red) and towns locations (in blue). This is a period of extremely fast expansion of both the network and the urban system. The data on the historical railway network is taken from [145, 110, 146, 191, 35], while the data on historical towns is taken from [27], spatialised through Geonames.

railway lines [191, 146], and railway stations locations [110, 146] in England, Wales and Scotland. The data sets do not explicitly contain travel time nor speed along the lines.

The “Bradshaw’s Monthly Railway and Steam Navigation Guides” provide a comprehensive collection of train timetables for Britain, covering the period from October 1839 to May 1961 at monthly intervals. Some of the guides are available online as pdf scans of the original printed versions². The remarkable length of each guide (from 145 pages for a guide in 1850 to more than 500 pages for the October 1882 issue) made it impractical to manually transform the timetables into machine-readable data and the idiosyncratic style of train

²See, for instance, the online archive <https://timetableworld.com/>.

timetables defied conventional image to text converters. Whilst recognising that, ideally, train timetables should be used to accurately find the travel time between any pair of localities, for the scope of this work we decided to estimate it based on average train speeds and distance along the railway network.

We derive our estimate for average speed from the online atlas that accompanies the railway data [35]. The atlas mentions that in 1830, the year that saw the beginning of the railway era with the opening of the Liverpool to Manchester Railway, the average train speed was 12 miles per hour (mph), equivalent to 19.3 kilometres per hour (km/h). In 1900, the average speed had increased to 40 mph (64.4 km/h). Assuming a linear increase in the transport speed, we can estimate the average speed in 1851 and 1881 to be 20.4 mph (32.8 km/h) and 32.4 mph (52.1 km/h), respectively. This makes railway transport about 2.55 and 4.05 times faster in 1851 and 1881 respectively when compared to the 8 mph (12.9 km/h) average speed reached by the stagecoach, which was the fastest form of land transport available to passengers before the advent of the railways.

5.2.2 Travel time computation

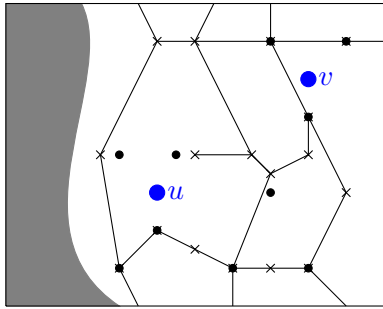
For each year and each pair of towns u and v , we computed the travel time as follows. Refer to Fig. 5.3 for a visual depiction of this method, applied to a fictional example region (Fig. 5.3a).

First of all, we found the geographic distance d_{uv} separating u and v (Fig. 5.3b). We divided the distance by the speed of 12.9 km/h, and obtained the theoretical minimum time needed to reach v from u along the geodesic via stagecoach³. This value was taken as a baseline.

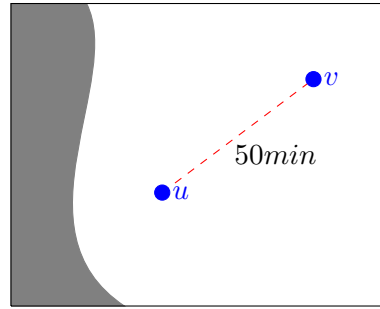
For each town, we found the three closest stations in the stations data sets, and recorded the geodesic distances that separate them from the town (Fig. 5.3c). The assumption here is that one may accept to travel out of a town using a station other than the closest one, if this allows one to access a shorter railway connection to one's destination.

In the railway lines data sets, each line is defined as a set of segments. Railway lines that allow for interchanges have a point in common, but the endpoints of the segments do not necessarily coincide with the coordinates of a station in the station data sets. For each station, we associated the closest segment endpoint in the rail line data sets (Fig. 5.3d). We assumed that the lines that include that point among their segment endpoints are accessible

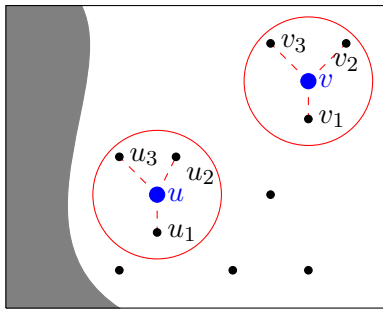
³For simplicity, this ignores the topography.



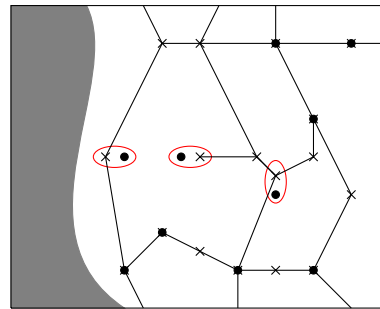
(a) Example region.



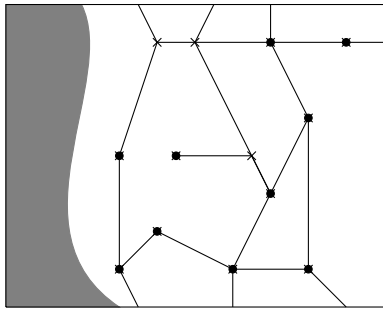
(b) Step 1: Geodesic travel time.



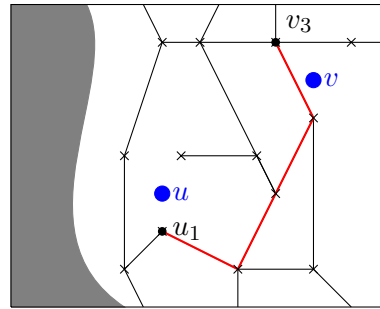
(c) Step 2: Find closest 3 stations for each city by geodesic travel time.



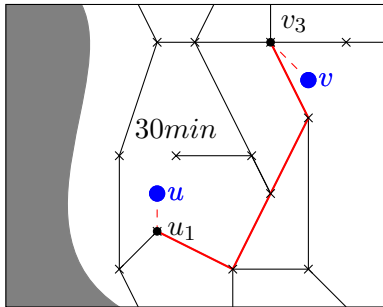
(d) Step 3: Identify each station with its closest endpoint (if not already).



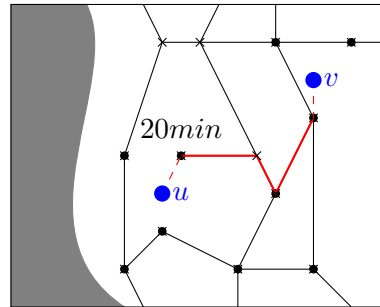
(e) Step 4: Delete unnecessary breakpoints, keeping record of the length.



(f) Step 5: Find shortest paths between each pair $(u_i, v_j)_{1 \leq i, j \leq 3}$. Here u_1 to v_3 .



(g) Step 6: For each shortest path, compute travel time using train speed and including city-station path.



(h) Step 7: Identify fastest path overall (remarking that it could be the geodesic from Step 1).

Figure 5.3: Travel time computation. Here blue dots are cities, black dots are stations, and crosses are railway segments endpoints.

from the station.

We simplified the network by merging together two segments if one of their endpoints coincided and no other line intersected them at that point (Fig. 5.3e). We associated to the new segment a length attribute equal to the sum of the lengths of two segments that were merged to form it, so to keep track of the real length while simplifying the network.

Using the `igraph` package in R [66], we computed the shortest railway path connecting each pair of stations $\{u_i, v_j\}_{1 \leq i, j \leq 3}$ amongst the three closest stations to u and v , and recorded its length (Fig. 5.3f). Dividing by the maximum railway speed for the year under consideration, we obtained the travel time between these stations.

For the pair of towns (u, v) , we computed the travel time along nine paths, each corresponding to a combination of travelling from u to one of the three stations closest to it along the geodesic, from this station to one of the closest three stations to v , and from there to v along the geodesic (Fig. 5.3g).

We finally defined the travel time $t_{u,v}$ between cities u and v as the minimum between the stagecoach travel time along the geodesic and the fastest among these nine paths (Fig. 5.3h).

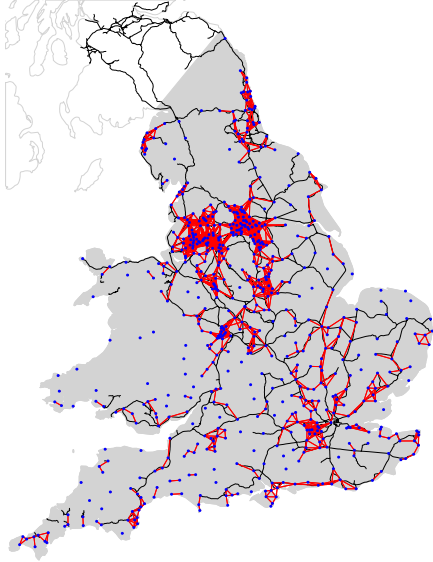
5.2.3 The degree entropy of disc graphs based on travel time

We extend our study of the spatial organisation of the English and Welsh towns, which we begun in Section 4.4.4. We construct disc graphs $G_{y,t}$, for a year $y \in \{1851, 1881\}$ and a travel time threshold t , this time connecting two towns u and v if their minimum theoretical travel time $t_{u,v} \leq t$.

Fig. 5.4 shows the dramatic change in the theoretical connectivity within 1h between 1851 and 1881, both as a consequence of a vastly expanded railway network and of faster average train speed along the lines. For this threshold, the network in 1851 (Fig. 5.4a) is sparse, with many isolated towns and small components. The largest components are found in the Midlands and Pennine area, on the Tyne estuary, and in the South East. Many links follow the railway lines. By 1881 (Fig. 5.4b), all but the most remote towns in the extreme North of England, coastal Wales, and Devon are connected, and the network is visibly denser. Whilst part of the same giant component, the Pennine/Midlands, Tyne estuary, South Wales and London clusters remain visually distinct for this choice of threshold.

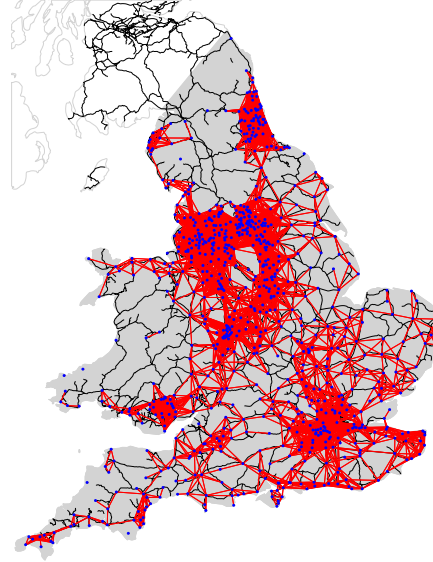
The analysis of the first degree-based entropy in Fig. 5.5 reveals complex effects of the presence and development of railways on the level of morpholog-

Disc graph in 1851, $t = 1\text{h}$



(a) The disc graph with the minimum theoretical travel time threshold $t = 1\text{h}$ in 1851. Most links (in red) are in the major urban cluster or follow the sparse railway lines (in black). Several towns remain disconnected.

Disc graph in 1881 $t = 1\text{h}$



(b) The disc graph with travel time threshold $t = 1\text{h}$ in 1881. The higher speed of trains and the more capillary railway network greatly expand the link set. Almost every town is part of the same giant component.

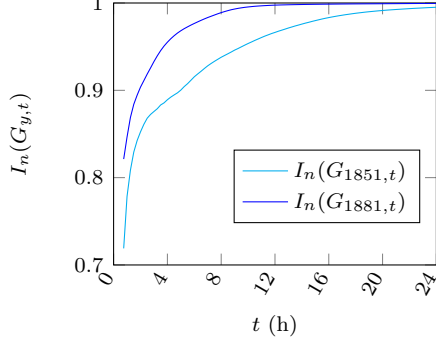
Figure 5.4: Disc graphs of English and Welsh towns in 1851 and 1881, for a travel time threshold of $t = 1\text{h}$, accounting for the development of railways.

ical polycentricity of the system.

On the one hand, an expanded and faster railway system makes the network in 1881 generally more balanced with respect to the order-normalised entropy than in 1851, for all given travel time threshold t (Figs. 5.5a and 5.5b). This is primarily, but not exclusively, a consequence of the vastly expanded link set. A local maximum of the difference $I_n(G_{1881,t}) - I_n(G_{1851,t})$ is observed when $t = 4\text{h}45\text{min}$. In 1881 this travel time was enough to reach London and the Tyne estuary from the main Pennines/Midlands cluster, which means that the clusters become more balanced and integrated. On the other hand, normalised values of the entropy tend to be smaller than the ones observed in Section 4.4.4 when only geographic distance was employed.

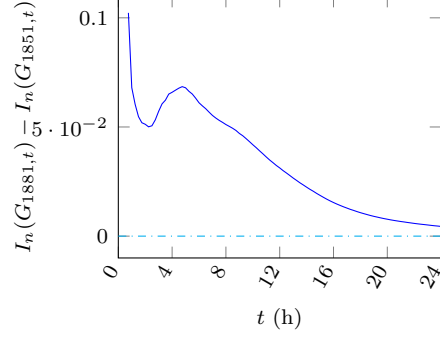
Fig. 5.5c shows the entropy in 1851 and 1881, normalised by n and m . $I_{n,m}(G_{1851,t})$ reaches its minimum for small t , due to the fact that only few

**Degree entropy of disc graphs
(normalised by n)**



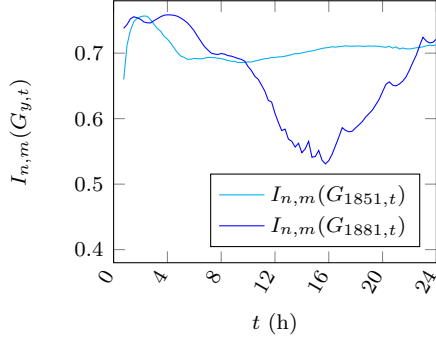
(a) Entropy of the disc graphs in 1851 and 1881 for changing travel time threshold t , normalised by the order n . In both years, the entropy grows monotonically, as the graph size increases.

**Degree entropy difference
(normalised by n)**



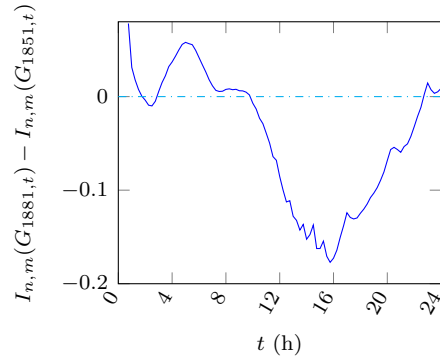
(b) Difference between the entropy of the disc graphs in 1881 and 1851, for different travel time t , normalised by n . $I_n(G_{1881,t}) > I_n(G_{1851,t})$ at all times, with a local maximum at $t = 4\text{h}45\text{min}$.

**Degree entropy of disc graphs
(normalised by n and m)**



(c) Entropy of the disc graphs in 1851 and 1881 for changing travel time t , normalised by n and m . The oscillations are connected to the spatial distribution of towns, but the presence of the railway accentuates the imbalance for moderate to large t . For $t > 10\text{h}$, $I_{n,m}(G_{1881,t})$ may not be indicative of the centralisation of the system due to the excessive sensitivity of the measure for large size graph.

**Degree entropy difference
(normalised by n and m)**



(d) Difference between the entropy of the disc graphs in 1881 and 1851, for different travel time threshold t , normalised by n and m . The system is more balanced in 1881 for small travel time, but less so for $t \approx 2\text{h}15\text{min}$, when $I_{n,m}(G_{1851,t})$ reaches its peak. When $4 \leq t \leq 10\text{h}$, most towns in the core become increasingly connected in 1881, making the entropy more balanced than in 1851.

Figure 5.5: Normalised degree entropy of the disc graphs in 1851 and 1881, for travel time $45\text{min} \leq t \leq 24\text{h}$. The oscillations reveal the complex interaction of spatial organisation and the presence of railways, and their impact on morphological polycentricity.

towns are connected to each other. It reaches its maximum when $t = 2\text{h}15\text{min}$. Multiplied by the maximum speed of the stagecoach, which would still have been the fastest mean of transport available in many towns, this corresponds approximately to the distance threshold of $r = 30\text{km}$ which in Fig. 4.17 saw most towns become connected in a lattice. Notice, however, that the maximum value of the entropy $I_{n,m}(G_{1851,2\text{h}15\text{min}}) \approx 0.758$ remains smaller than $I_{n,m}(G_{1851,30\text{km}}) \approx 0.800$, signalling that the presence of railways in its early days has favoured the connectivity around some towns only, and thus increased the centralisation of the system. The coverage of the railway network was rather unequal: all of the major clusters of towns in 1851 were served by railways, but this was not the case for many other towns. The entropy $I_{n,m}(G_{1851,t})$ declines for $t > 2\text{h}15\text{min}$, and remains relatively stable.

The entropy in 1881 sees larger fluctuations. A local maximum and the global maximum are reached for $t \approx 1\text{h}30\text{min}$ and $t \approx 4\text{h}$, respectively. These oscillations are compatible with the pattern of slight peaks, plateaus and declines remarked in Fig. 4.18c for a distance threshold $r < 170\text{km}$. When $4 \leq t \leq 7\text{h}$, the entropy declines, and plateaus when $7\text{h} \leq t \leq 10\text{h}$. When multiplied by the average train speed in 1881 (52.1 km/h), the plateau corresponds approximately to the increase observed in $I_{n,m}(G_{1881,r})$ for $350\text{km} \leq r \leq 500\text{km}$. The fact that in this case the entropy does not increase but only slows down its decline may be ascribed to the fact that the few towns not yet reached by the railway network in 1881 were all located in the more peripheral areas of the country. This meant that they are comparatively worse connected in the travel-time-based networks than in the geographic-distance-based ones, when compared to the towns in the well-served and geographically central core of the region. When $t > 10\text{h}$, we observe a steep decline. These values are not visible in Fig. 4.18c, as the distance threshold only stops at 500km. The decline may be ascribed in part to the aforementioned phenomenon of sharpened isolation of geographically peripheral towns, disconnected from the railway network, but the large oscillation of the measure and its steep, uneven increase for $t > 16\text{h}$ suggest that it may be quite simply the result of the excessive sensitivity of $I_{n,m}$ for near-complete graphs.

Apart from its peak $t = 2\text{h}15\text{min}$, $I_{n,m}(G_{1851,t})$ remains smaller than $I_{n,m}(G_{1881,t})$ for all values of t for which the latter value is a valid indicator of polycentricity (Fig. 5.5d).

5.2.4 A human-centred morphological polycentricity: the entropy of potential interactions

We apply Wilson’s spatial interaction model [215] discussed in Section 2.4.2, to obtain the most likely flows $\{w_{uv}\}_{u,v}$ between cities u and v , based on their travel time distance t_{uv} and populations p_u, p_v , in 1851, 1861, and 1881.

Following [216] and [215], as the population is the only morphological attribute known for each city, we assume the benefit b_v of interacting with city v to be $b_v = \log p_v$. With these assumptions, eq. (2.2) for the most likely interactions simplifies to

$$w_{uv} = \gamma \cdot p_u \frac{\exp(\alpha \log p_v - \beta t_{uv})}{\sum_x \exp(\alpha \log p_x - \beta t_{ux})}. \quad (5.1)$$

Note that we have substituted the travel time for the more generic travel cost. Since we are interested in the entropy of interactions (Definition 9) and thus in w_{uv} normalised by their sums, we can safely take the proportionality constant $\gamma = 1$. Recall that α regulates the benefit that u receives from interacting with a city v given its population p_v : if $\alpha = 1$, the interaction with v is directly proportional to p_v , if $\alpha > 1$, then the attractiveness of v depends superlinearly on its population. On the other hand, β is a weight for the disadvantage of interacting, for instance, the negative impact of geographical distance and poor infrastructure on travel time or cost: a larger β signals higher sensitivity to these aspect, while $\beta = 0$ represents the extreme case of distance, travel time or cost having no influence at all on the interactions.

Thus, we obtain a family of networks, one for each choice of α and β , whose connections represent the most likely interactions between the pairs of cities, given some assumption on the balance of cost/benefit of interacting, their population and the travel time between them.

For each city u , we can now compute the entropy of interactions $S(u)$ and obtain a measure of heterogeneity of u ’s most likely connections. As these depend exclusively on its population, its location with respect to other cities in V , and the transport infrastructure and technology in R , $S(u)$ can be interpreted as a dynamic, human-centred measure of morphological polycentricity.

Note that this measure can be adapted to different scales and purposes. In the Topographical and Regional Polycentricity (eqs. (2.8) and (2.9)), one could impose an upper threshold on distances d_{uv} . The same can be done here with the travel time t_{uv} , and thus one can study the polycentricity of each town within a certain travel time distance t , such as commuting distance

(say, $t \leq 1\text{h}$) or day-trip distance (say, $t \leq 4\text{h}$).

On the other hand, one can average $S(u)$ among all cities in R and obtain an overall measure of polycentricity of the region:

$$S_R = \frac{1}{|V|} \sum_{u \in V} S(u).$$

Similarly, one can average $S(u)$ across subregions of R for a comparative approach. For instance, for subregions R_1 and R_2 of R , with set of cities V_1 and V_2 , one can write $S_{R_1} = \frac{1}{|V_1|} \sum_{u \in V_1} S(u)$, and $S_{R_2} = \frac{1}{|V_2|} \sum_{u \in V_2} S(u)$, and compare the two values.

5.2.5 Potential Interactions in England and Wales

As mentioned in Section 2.4.2.1, spatial interaction models based on entropy maximisation have been used as dynamical models of the evolution of the system [216, 217]. The validation of a particular model and choice of parameters often passes through the model's ability to describe the system's evolution. In our case, assuming that the migration flows follow the most likely interactions identified by the model, one can use actual migration flows and population evolution to calibrate the model.

Adopting this approach for a historical study such as our study of mid-19th century English and Welsh towns poses some challenges linked to the limited availability of migration data.

The censuses between 1851 and 1881 did not record explicitly migration flows between British towns. They did record, though, the number of residents in each county who were born in a different one. On the basis of this data and on additional qualitative information provided in the Census Notes, Smith created a map [197, Fig. 3] representing the major inter-county migration flows in 1861. This map highlights two aspects in particular. The first is that most migrations in the Midlands and the North were short distance ones. Lancashire, whilst attracting most migrants from neighbouring counties, especially the West Riding of Yorkshire, was perhaps an exception, as it attracted a large number of individuals from as far as Devon. The second is that London was by far the main centre of attraction for the whole of England and Wales. Its influence was especially strong in the Greater South East, but could be felt as far as Northumberland. In particular, counties near London saw very little reciprocal migration. The places where London's hegemony was rivalled were the Midlands and the North, in proximity of attractive counties. The family

historians Pooley and Turnbull [170] confirmed and expanded these findings remarking the stability of these migration patterns across the period.

We used Smith’s map to validate our model and choice of parameters. We aggregated the towns by county and, for any two counties R_1 and R_2 , we considered the total weight

$$w_{R_1 R_2} = \sum_{u \in R_1, v \in R_2} (w_{uv} - w_{vu}), \quad (5.2)$$

representing the expected interaction balance between R_1 and R_2 . We compared the largest weights $w_{R_1 R_2}$ in the year 1861 to the map of major migrations flows [197, Fig. 3], computing the w_{uv} for different choices of the parameters α and β in eq. (5.1).

The effect of changing α in the reasonable range of scaling effects $1 \leq \alpha \leq 1.2$ was negligible. Hence we fixed $\alpha = 1$. No choice of β led to completely satisfactory results. A small value such as $\beta = 0.5$ lessened the effects of distance, and predicted spurious connections from the northernmost counties to Lancashire and overestimated the influence on London on the Midlands, but correctly predicted some long-distance migration flows to the capital. Increasing values of β led to a sharp decrease in predicted flows to London. A good compromise was found with $\beta = 1$, which reproduced most of the map’s patterns, although underestimating London’s influence on the North-East and Devon.

The values $\alpha = \beta = 1$ in eq. (5.1) led to networks generally consistent with historical studies on the evolution of the British urban system in the 19th century [211] and migration flows [197, 170]. Yet, this remains to some extent, an arbitrary choice, and a full exploration of the method should account for a range of values⁴.

The resulting networks are shown in Fig. 5.7. Following the example of [144], here the towns are distinguished in three categories to highlight the areas of greatest interest: the Greater South East (corresponding in our case to the historical counties of Buckinghamshire, Middlesex, Sussex, Hampshire, Kent, Oxfordshire, Berkshire, Surrey, Bedfordshire, Cambridgeshire, Essex, Hertfordshire, Norfolk, Suffolk and London), the North (Cumberland, Northumberland, Westmorland, Durham, Lancashire, Cheshire and Yorkshire) and the rest of England and Wales. The first thing to remark is that p_u , p_v , and t_{uv} have changed substantially from 1851 to 1881, as a result of the combined

⁴See Section 7.2 for more details on the limitations of this parameter fitting.

**Largest interaction balances between
English and Welsh counties in 1861 ($\alpha = \beta = 1$)**

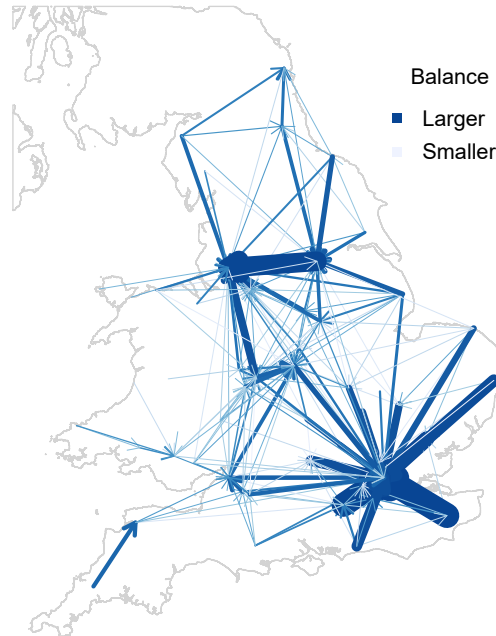


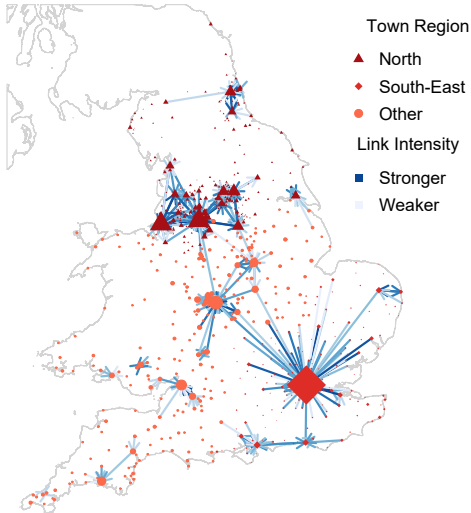
Figure 5.6: The largest interaction balances between English and Welsh counties in 1861 for parameters $\alpha = 1$ and $\beta = 1$. The interactions largely correspond to the migration flows in 1861 shown in [197, Fig. 3], but underestimate long-distance migration to London from peripheral counties.

effects of population growth, railway expansion and faster railway speeds. All these factors led to an overall increase in the strength of the interactions.

In 1851, a number of clusters of strong connections are apparent: a very centralised star pattern around London in the South-East, a strongly interconnected set of cities in the Pennine area, smaller star-shaped patterns in the Midlands that reconnect the previous two larger clusters, and a few small isolated clusters in Yorkshire and in the South West of the country. The situation in 1881 sees two major changes: the expansion of the London-centred star-shaped pattern to the rest of the South-East, a sign of the consolidation of London's preponderance on the region, and a greater level of integration of the clusters in the North and in the Midlands.

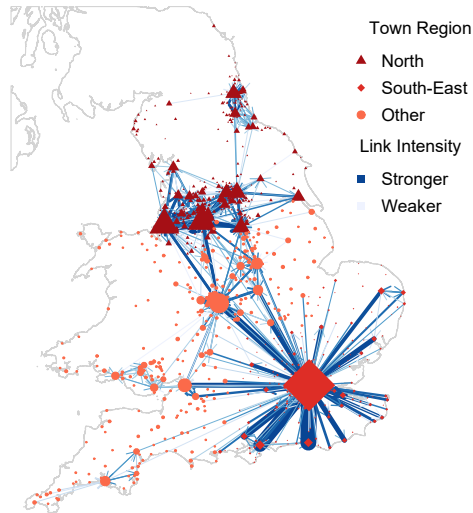
Neglecting directionality, these networks are not too unlike the undirected networks in Figs. 5.4a and 5.4b. This is to be expected, since travel time is one of the main factors in the model. Note, however, that the connections are directed towards or issuing from the main centres in the clusters due to the effect of their population, instead of connecting nearby centres, as is

Largest interactions between English and Welsh towns, 1851



(a) Largest potential interactions in 1851. Much of the flow in the South East is directed towards London, whilst the interactions in the other clusters are more polycentric.

Largest interactions between English and Welsh towns, 1881



(b) Largest potential interactions in 1881. The dominance of London in the South East is reinforced, and the urban clusters in the North become more integrated.

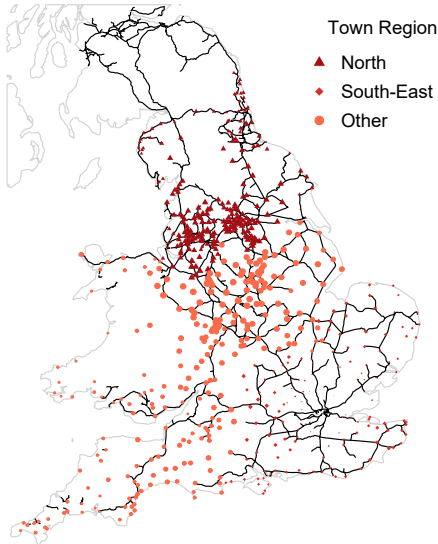
Figure 5.7: Largest interaction between English and Welsh towns in 1851 and 1881. The size of the symbol representing a town is proportional to its population. The thickness of the link is proportional to the absolute magnitude of the interaction. The shade represents the relative weight of the interactions when compared to other represented in the same figure. For clarity, only the most intense connections are shown, but the network is complete: all pairs of towns are connected in the model, albeit some of them only very weakly.

particularly apparent in the case of London and the South East.

5.2.6 Human-centred morphological polycentricity of England and Wales

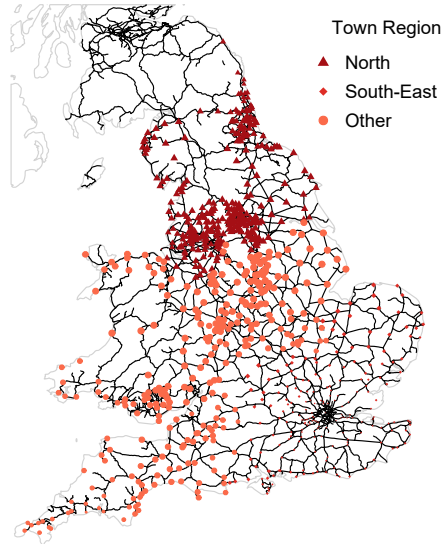
We computed the entropy of interactions $S(u)$ for each town u as a measure of its morphological polycentricity. In Figs. 5.8a and 5.8b we show, respectively, $S(u)$ in 1851 in 1881 for each city, superimposing it to the maps of railway lines. London and cities in the South East have substantially smaller entropy when compared to cities in the North, a result which agrees with what one would expect from observing the networks in Fig. 5.7. Indeed, most flows from cities in the South East are directed towards London, and the greatest part of London's flows is directed towards itself.

Entropy of interactions for English and Welsh towns, 1851



(a) Entropy of potential interactions in 1851. Higher entropy values are found in the Midlands and Pennines area, while geographically peripheral areas, and the entire South East have very small entropy.

Entropy of interactions for English and Welsh towns, 1881



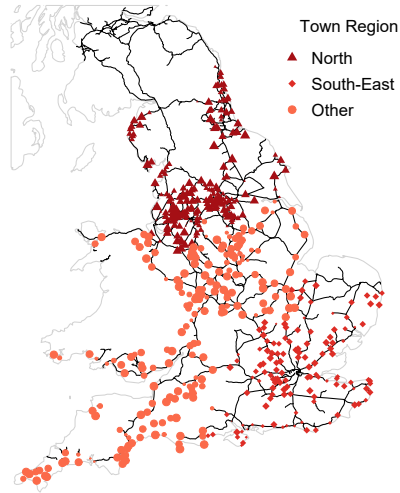
(b) Entropy of potential interactions in 1881. Higher entropy values are found in most of England and Wales except for London and the South East. Railways have mitigated the effects of geographical isolation.

Figure 5.8: Entropy of the potential interactions for each English and Welsh town in 1851 and 1881. The size of a symbol is proportional to the entropy of the town, normalised so to fall in the range $[0,1]$. The shape of the symbol depends on the region the town belongs to. The black lines represent railway lines. Remark the vast expansion of railway during this period.

Note that in 1851, towns at the geographic extremes of England and Wales (such as South-West Wales, Northumberland, and Cornwall) tend to have small entropy since the number of cities they can reach is limited by their peripheral locations and poor infrastructure reach. In the case of the Northumberland, this is due, at least in part, to artificial boundary effects. Indeed, the population data refers to England and Wales alone, and including Scottish cities in the analysis might yield changes in the results. In 1881 the situation is mitigated by better overall connectivity and faster transport speeds along the railway lines.

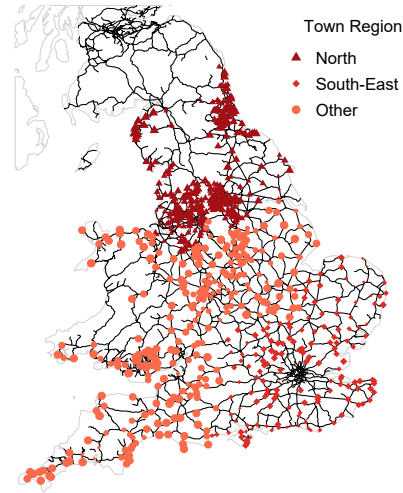
When a maximum travel time threshold t is imposed in the computation of the entropy, the effects of being located at the geographical extremes or at centre of the area are alleviated.

Normalised entropy of interactions in 1h radius, 1851



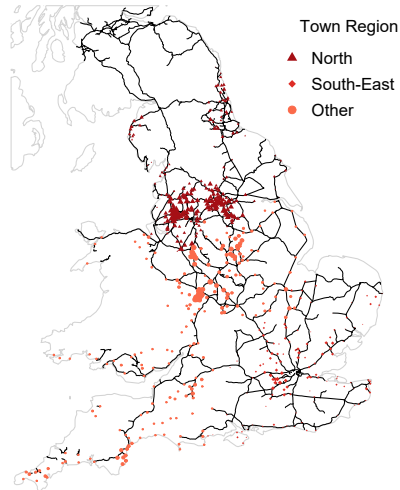
(a) Entropy of interactions within a 1h travel time radius in 1851, normalised. The entropy of towns in the South East is the smallest, but it is larger than if no travel time threshold is imposed.

Normalised entropy of interactions in 1h radius, 1881



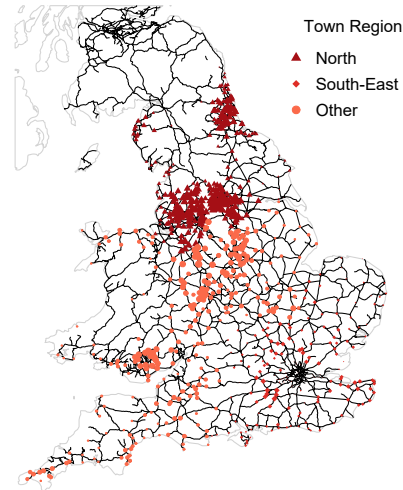
(b) Entropy of interactions within a 1h travel time radius in 1881, normalised. The entropy in the peripheral regions and in the South-East is larger than if no travel time threshold is imposed.

Entropy of interactions in 1h radius, 1851



(c) Entropy of interactions within a 1h travel time radius in 1851, not normalised. The values are generally small.

Entropy of interactions in 1h radius, 1881



(d) Entropy of interactions within a 1h travel time radius in 1881, not normalised. The values are larger.

Figure 5.9: Entropy of interactions for each English and Welsh town in 1851 and 1881, within one-hour travel distance from the town. The size of a symbol is proportional to the entropy of the town. The shape of the symbol depends on the region the town belongs to. The black lines correspond to railway lines.

Adopting a small threshold $t = 1\text{h}$ (Fig. 5.9), the normalised entropy indicates that connections in the South are substantially more polycentric than when no threshold is imposed, although they still remain below the values reached in the North (see Fig. 5.9a and 5.9b). This is because London remains out of reach from most areas of the South East when such a short travel time is allowed. When the entropy is not normalised (Fig. 5.9c and 5.9d), we can appreciate the overall increase in railway connectivity from 1851 to 1881, and the consequent increase in entropy in peripheral areas, particularly apparent in Northumberland and Southern Wales.

When a larger threshold $t = 4\text{h}$ is imposed (Fig. 5.10), the results of the normalised entropy (Figs. 5.10a and 5.10b) are already nearly identical to the unconstrained case. When the entropy is not normalised (Fig. 5.10c and 5.10d), the effects of the expansions of the railways are again apparent in the increase of the entropy. In both the normalised and not-normalised case, the expected pattern of extreme centralisation emerges in all of the “home counties”, whose name alludes to the fact that their residents could travel to London and come back within one day.

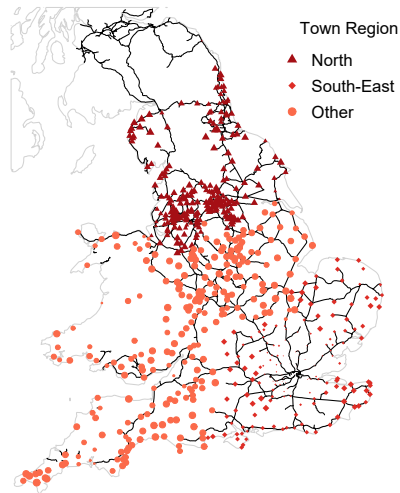
In Table 5.2, we show the average entropy of interactions S_R in the North and South-East of England, in selected cities, and in the whole system, normalised so to fall in the range $[0,1]$ and be comparable. All areas except the South East become increasingly polycentric according to the average entropy measure as a result of better transport connections that allow more heterogeneous interactions. The connections coming out of London, though still extremely centralised, also become remarkably more polycentric, as a result of increased and faster connections to the neighbouring areas of the South East. This also explains the reduction of the entropy in the South East from 1851 to 1881, after a temporary increase in 1861.

Tables 5.3 and 5.4 allow to compare these results with more classic measures of polycentricity that do not employ a network approach.

The slope in the rank-size relations (i.e., the exponent in the Zipf’s law) in Table 5.3, suggests a tendency for the population distribution to become more polycentric in the North and progressively less so in the South⁵. It should be pointed out, though, that the measure is very sensitive to the number of cities used to estimate it: the small value of the slope in the Greater South East suggests that the region is more polycentric than the North, because the best-fitting line in the rank-size plot ignores London as an outlier. When

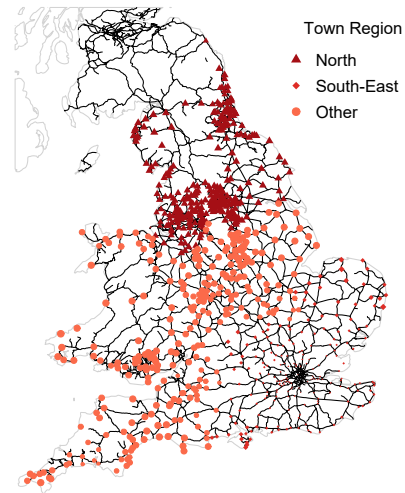
⁵Recall that the slope is steeper when the population distribution is more concentrated.

Normalised entropy of interactions in 4h radius, 1851



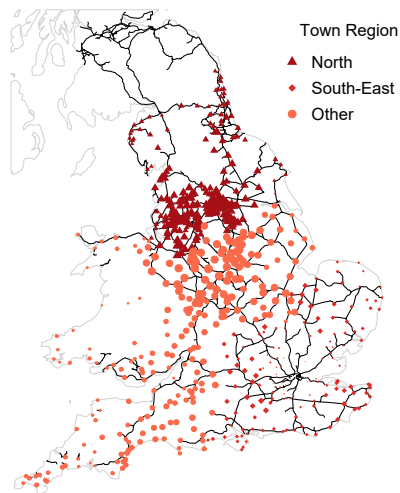
(a) Entropy of interactions within a 4h travel time radius in 1851, normalised. Values in the South East are much smaller than elsewhere.

Normalised entropy of interactions in 4h radius, 1881



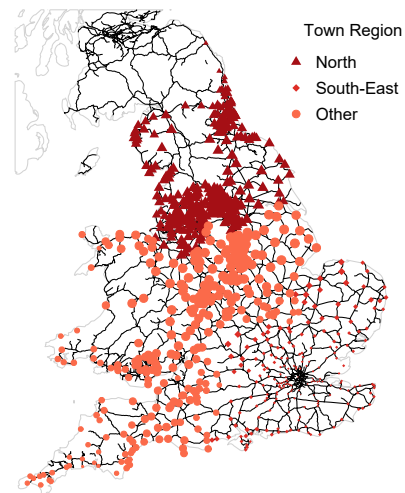
(b) Entropy of interactions within a 4h travel time radius in 1881, normalised. Values in the South East are much smaller than elsewhere.

Entropy of interactions in 4h radius, 1851



(c) Entropy of interactions within a 4h travel time radius in 1851, not normalised. Peripheral areas and the South east have the largest entropy.

Entropy of interactions in 4h radius, 1881



(d) Entropy of interactions within a 4h travel time radius in 1881, not normalised. Most towns have visibly larger entropy than in 1851.

Figure 5.10: Entropy of interaction for each English and Welsh town in 1851 and 1881, within four-hour travel time from the town. The size of a symbol is proportional to the entropy of the town. The shape of the symbol depends on the region the town belongs to. The black lines correspond to railway lines.

only a few towns for this region are used, including London, the results do point more realistically towards great level of centralisation, but information regarding smaller cities is inevitably lost.

The Shannon entropy of the population in Table 5.4 tells a more varied story, with a tendency towards centralisation between 1801 and 1881, that slows down and inverts its trend by the early 20th century.

Table 5.2: Entropy of the interactions for selected cities and average entropy in different regions in England and Wales.

		Entropy of interactions		
Year		1851	1861	1881
Region				
North		0.499	0.558	0.645
Greater South East		0.279	0.314	0.26
London		0.035	0.061	0.134
Manchester		0.430	0.490	0.595
Birmingham		0.409	0.438	0.580
England and Wales		0.435	0.485	0.555

Table 5.3: Slope of the rank-size relation in different regions of England and Wales between 1851 and 1881 (computed with all cities).

		Slope in rank-size relation				
Year		1801	1851	1861	1881	1911
Region						
North		1.171	1.20	1.169	1.149	1.078
Greater South East		0.904	0.984	0.992	1.007	1.072
England and Wales		0.94	1.02	1.025	1.038	1.078

Table 5.4: Shannon entropy of the population distribution for different regions in England and Wales between 1851 and 1881.

		Shannon entropy of town population				
Year		1801	1851	1861	1881	1911
Region						
North		0.841	0.781	0.764	0.771	0.779
Greater South East		0.457	0.430	0.426	0.418	0.434
England and Wales		0.734	0.720	0.717	0.715	0.723

5.2.7 Conclusion

The years between 1850 and 1880 were characterised by rapid urbanisation and industrialisation in Britain, primarily in larger centres [211, 170]. They were also decades of major expansion and modernisation of the transport infrastructure. The results of our analysis of morphological polycentricity suggest that, while this period witnessed an increased concentration of the population into fewer urban centres, cities became better integrated into a more heterogeneous network of connections, supported by an expanded and improved transport infrastructure. The years after 1880 brought on economic restructuring as older industries declined and newer sectors emerged and the growth shifted from large to smaller centres, especially suburban towns [170]. The economic and social factors driving these processes are complex, but certainly the conditions for this decentralisation process were made possible by the capillary expansion of the transport system in previous decades and by the strengthening of connections between larger and smaller centres.

Looking at the greater heterogeneity in population sizes after 1881 suggests that the population growth of cities might eventually have followed the dynamics of their increasingly polycentric most-likely interactions. A network-entropy-based approach to measuring morphological polycentricity could thus give a more accurate description of the state of a region compared to other non-network based approaches, judging by the system's subsequent evolution.

Chapter 6

Cities and their People: Human Activities in Systems of Cities

The previous chapters of this thesis studied systems of cities concentrating on their morphology: their location, population, land use, and the physical infrastructure connecting them. This chapter adopts a different perspective to study systems of cities, shifting the focus to the human activities that take place therein: the daily dynamics of the population distribution, the social connectivity, and the diffusion of ideas and innovations.

On the one hand, this chapter explores the role of entropy in studying cities when these are interpreted as “social networks embedded in the built environment” [32]. On the other hand, it wants to reconnect the morphological perspective adopted so far to the functional one, where in this case by “function” of a city or system of cities we mean the human connectivity they foster.

These two perspective are highly interdependent: the morphology of a system constitutes the backbone on which human activities take place, enabling and constraining them at the same time. The functional aspect, in turn, is amongst the factors that influence the shape of cities and their connectivity via infrastructure.

We begin by developing a model for the distribution of Tweets in London in Section 6.1 based on point processes, and study its short-term spatio-temporal dynamics. In Section 6.2, we discuss the complexity of modelling long-term dynamics with unstable and evolving social network data, using the case study

of connections between Hungarian cities as measured in the online social network iWiW. Finally, in Section 6.3, we apply a temporal network approach to model cultural exchanges through imitation and collaboration among medieval lyricists, discussing the different challenge of dealing with high level of uncertainty in the data.

6.1 Geo-tagged tweets, dynamic model of their location

6.1.1 Data set

Our data set comprises the locations and time stamps of 432,067 geo-tagged tweets posted in a 40km radius disc centred in Trafalgar Square, London, between Sunday 10th June 2012 and Sunday 24th June 2012 [222]. Neither the tweets' content nor any information about the users is provided.

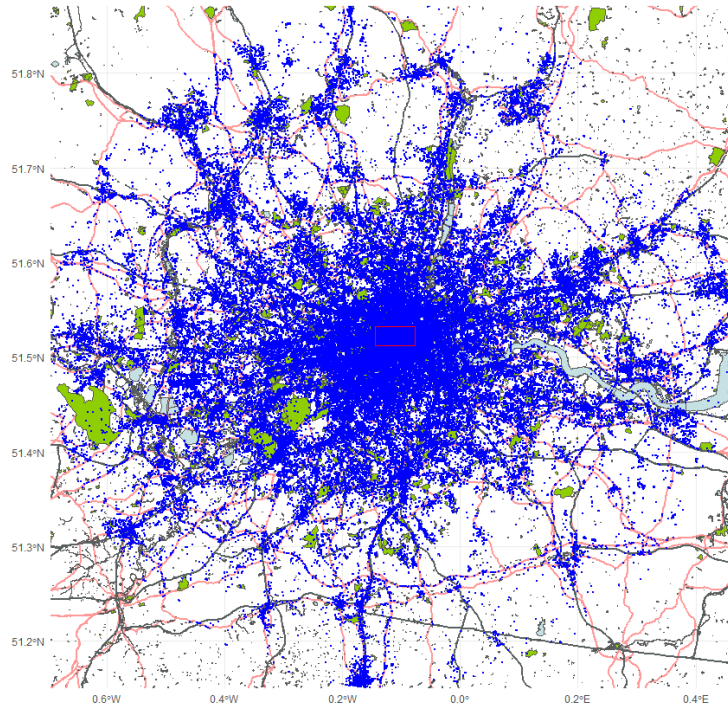
A first inspection of the data revealed large heterogeneity in the spatio-temporal distribution of the tweets. The hour of the day naturally played an important role (Fig. 6.2a), with an above average number of tweets posted between 7am and 22pm, and a marked peak in the evening. A number of exceptionally high peaks of activity are apparent in the data, likely due to specific events¹, although it was not possible to confirm any association in absence of the tweets' content. An unusually low activity at 4pm on Thursday 21st is ascribable to a temporary Twitter outage.

6.1.2 Spatial Model for aggregated tweets

Aggregating the tweets throughout the period (Fig. 6.1) reveals a number of complex features of the point pattern. The density of tweets is remarkably higher in the city centre than in the outskirts, reflecting the higher densities of residential, working, and tourist population in the central areas of the capital. More generally, the impact of the city's topography is apparent on the tweets distribution: the Thames, major parks and major roads such as the M25 are all visible from a first inspection of the data. The spatial distribution of tweets in the disc 6.1a is certainly not stationary nor isotropic. This impedes most

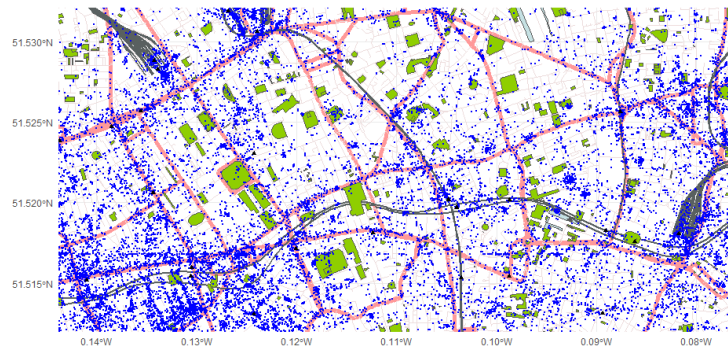
¹The largest hourly counts are observed between 8pm and 11pm on Sunday 24th June, 8pm to 9pm on Friday 15th, 7pm to 9pm on Tuesday 19th, and 4pm to 6pm on Monday 11th, with tweets distributed across the region of study. At each of these times, a match of the 2012 UEFA European Football Championship was being aired. The highest peak corresponds temporally to the penalty shoot-out of the quarter-final match England vs Italy (eventually won by the latter).

Aggregated Tweets in and around London



(a) Locations of all the Tweets in the data set. The boundaries of the rectangular region of study are provided for spatial orientation. The process is inhomogeneous and non-stationary and carries the evident impact of London's topography: parks, main roads and railways, the Green Belt, the Thames.

Aggregated Tweets in rectangular region



(b) Locations of all the Tweets in the rectangular region. Some Tweets are clustered and other seem to follow the main transportation lines. The region is purposely selected to lie in the city centre and avoid major parks and the river.

Figure 6.1: Locations of Tweets in the data set, aggregated across the entire study period. The background map (from OpenStreetMap) shows some of the main topographical features of the city.

traditional modelling attempts at this temporal resolution and observation window size.

An attempt to model the distribution of the aggregated tweets using two separate point processes (one for the city centre and one for the outskirts) and to model the process in the outskirts as a cluster process with a fixed number of clusters is presented in [152]. In the article, we clustered the tweets using a k -means algorithm and tested the hypothesis that in the outskirts the cluster centres were distributed according to a Homogeneous Poisson Process (the null hypothesis for the distribution of the centres, but also the base for describing the overall distribution as a cluster process such as a Neyman-Scott process). For some values of k and definitions of outskirts, the pattern was compatible with the Complete Spatial Randomness hypothesis according to the Empty Space Function; nonetheless, it consistently failed more sophisticated tests involving second order methods such as Ripley's K function.

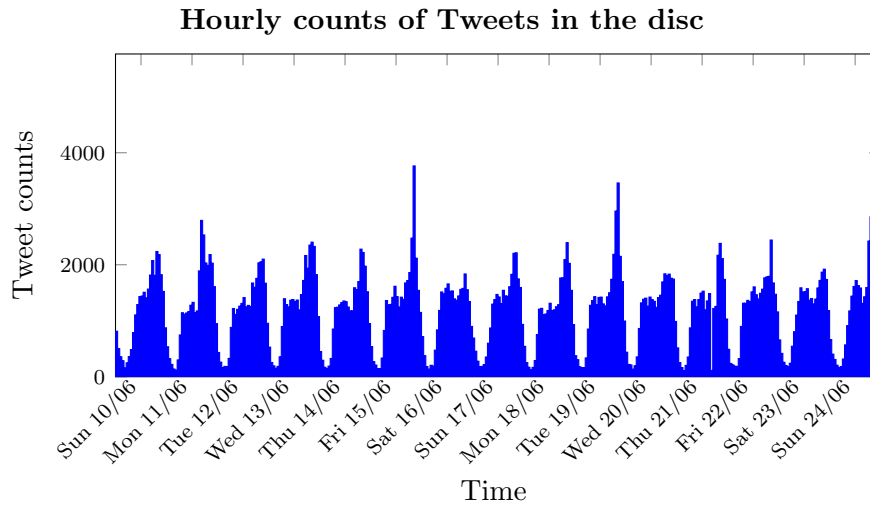
To reduce the issues that plagued this approach whilst keeping as much as possible its generality, we focused on a smaller but still sufficiently varied area²: a rectangle comprised between the Tube stations Blackfriars to the South, and Angel to North, and Great Portland Street to the West and Aldgate to the East. This area was carefully selected so to avoid the Thames, major parks and ensure a high density of points throughout, but is not uniform in land use, being both commercial and residential, nor in population density. The spatial distribution of the aggregated tweets in this area (Fig. 6.1b) is mostly organised in clusters (which suggest a manageable underlying Neyman-Scott process), but is still affected by the same lack of stationarity. Main roads, indeed, are still visible and mostly oriented in an irregular grid.

The modelling of the aggregated distribution would require a considerably more complex and ad hoc approach, accounting for the street network and may not be possible without additional information on the users or the content of the tweets. This defeats the purpose of seeking simple processes that might describe, at least in certain circumstances, the spatial distribution of tweets.

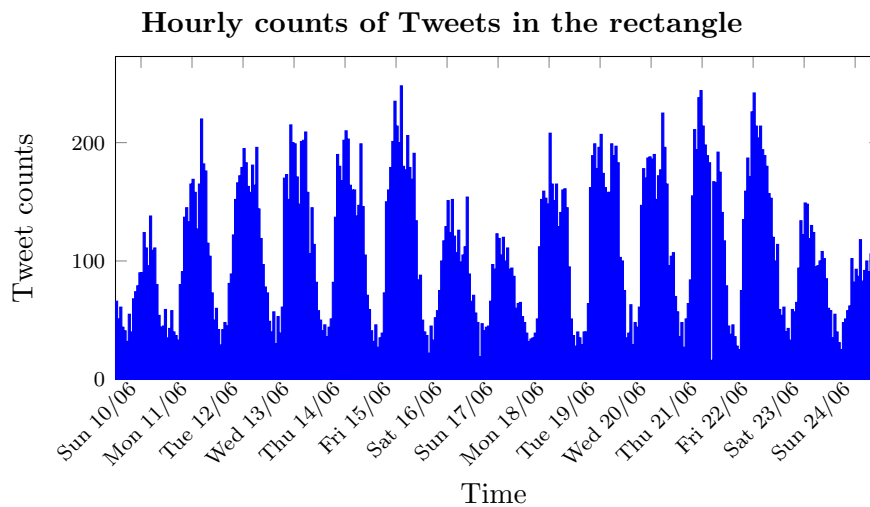
6.1.3 Spatial models in one-hour-long intervals

Once established that in non-trivial areas the aggregated distribution of tweets cannot be captured via a simple process, this raises the question of whether at the same spatial-scale but at a finer temporal resolution, a simple process

²Much smaller areas, uniform and artificially selected areas might exhibit simpler patterns, easier to model but have the risk of lacking of representativeness.



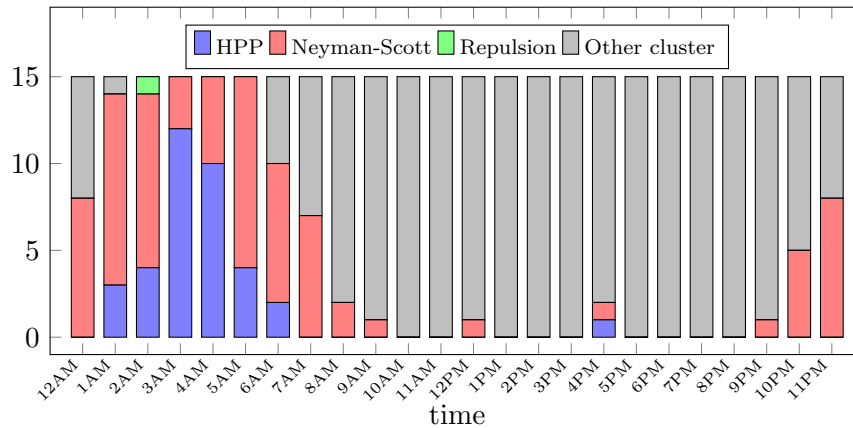
(a) Hourly counts of Tweets in the disc. Occasional peaks emerge, presumably linked to events.



(b) Hourly counts of Tweets in the rectangular area of study. Weekly variations are apparent in this case, as more Tweets are posted on weekdays than on weekends.

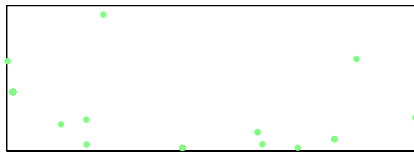
Figure 6.2: Hourly counts of Tweets in the whole data set and in the rectangular area of study. In both cases, daily oscillations are apparent: the highest level of activity is in the evening and lowest is at night.

Valid models by hour



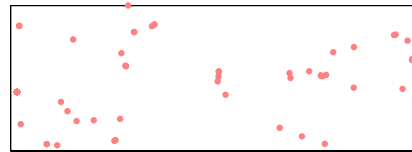
(a) Number of instances during the 15 day period under scrutiny when one of the proposed models (Homogeneous Poisson Process, Neyman-Scott cluster processes, or other repulsive process) is valid, by hour of the day. The proposed models work better at night, when the density of points is lower. Alternative models are needed to model the distribution of Tweets in the daytime. Many Neyman-Scott processes are actually capturing repulsion in the patterns rather than clustering.

10 June 2am-3am



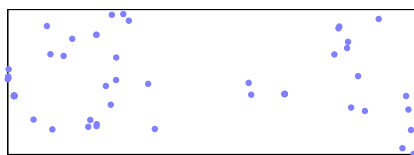
(b) Tweets in the rectangular area between 2am and 3am on Sunday 10th June. Distance to nearest neighbour evidences possible repulsiveness between the points.

12 June 10pm-11pm



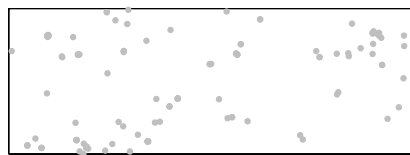
(c) Tweets in the rectangular area between 10pm and 11pm on Tuesday 12th June. The patterns is compatible with a Matérn cluster process with cluster radius of 22m and average size of 0.9.

22 June 6am-7am



(d) Tweets in the rectangular area between 6am and 7am on Friday 22nd June. The pattern is compatible with a Homogeneous Poisson Process.

24 June 9pm-10pm



(e) Tweets in the rectangular area between 9pm and 10pm on Sunday 24th June. The pattern is not compatible with a Homogeneous Poisson Process, Neyman-Scott process, and it does not shows repulsion.

Figure 6.3: Models for the spatio-temporal distribution of tweets in the rectangular study region.

may still be sufficient to describe the process. To this purpose, we aggregate the tweets in 360 one-hour-long intervals.

The hourly counts (Fig. 6.2b) show a periodic pattern: as well as the daily oscillation also observed in the whole data set, we also notice a remarkable weekly variation, with higher counts on weekdays and smaller counts on weekends.

In Fig. 6.3, we describe some salient aspects of the spatio-temporal variation of the hourly-aggregated tweets. An inspection of the data sets reveals the presence of repeated points in aggregated hour-long intervals. These may be due, for example, to a user who tweets from the exact same location multiple times during one hour. Most point process models assume that almost surely there will not be two points with the same location. In absence of information on the users, in this instance, we delete repeated entries from the data sets (see Section 7.2 for a discussion of the limitations of this approach).

For each hour-long interval, we compute a simple Clark and Evans Aggregation Index [62], as a crude measure of whether clustering or regularity should be expected. The index shows an overwhelming tendency towards clustering, and only in a limited number of cases (26 cases out of 360, and all at night or early morning), the index points towards regularity (see, for instance, Fig. 6.3b).

For each hour, we test the hypothesis that the pattern may be described via a Matérn or Thomas cluster process³. Following the fitting procedure proposed by [119] and detailed in Section 2.5.1.3, we find the most suitable parameters for the cluster models. We use a mincon approach, that optimises the observed parameters against the theoretical Ripley’s K -function of the model. Subsequently, we perform an envelope test, where the process with the estimated parameters is generated 99 times, and for each the spherical contact distribution function F and the nearest neighbour distance G are computed. The observed values \hat{F} and \hat{G} for the data are compared to the minimum and maximum values obtained by the simulations. If \hat{F} or \hat{G} fall outside the range defined by these extremal values, the model is considered invalid.

The Matérn and Thomas processes provide suitable models for a large number of patterns (105 and 108 respectively), again mostly at night and in the early morning. However, in a large number of cases, both \hat{F} and \hat{G} are also

³We do not test exclusively clustered patterns because even regular ones may still be modelled using a Neyman Scott process with an average cluster size less than one. In fact, this turned out to be the case for almost all of the regular patterns. The case shown in Fig. 6.3b being the only exception.

compatible with the Complete Spatial Randomness hypothesis, which ought to be preferred standing by Occam’s razor. This still leaves respectively 70 and 74 cases in which the Matérn and Thomas processes provide acceptable, simple models⁴. In Fig. 6.3c, we see an example of process compatible with a Matérn cluster process.

In a tenth of the cases (36 out of 360), the null hypothesis of a Homogeneous Poisson Process remains the best option (see for instance Fig. 6.3d). In the majority of cases though (221 out of 360), and almost the totality of peak-hours activity, none of the models considered is valid: the pattern shows no evidence of regularity and is not compatible with the Complete Spatial Randomness hypothesis, but its clustering is not describable using a Neyman-Scott process (Fig. 6.3e).

6.1.4 Hourly evolution of the Shannon entropy

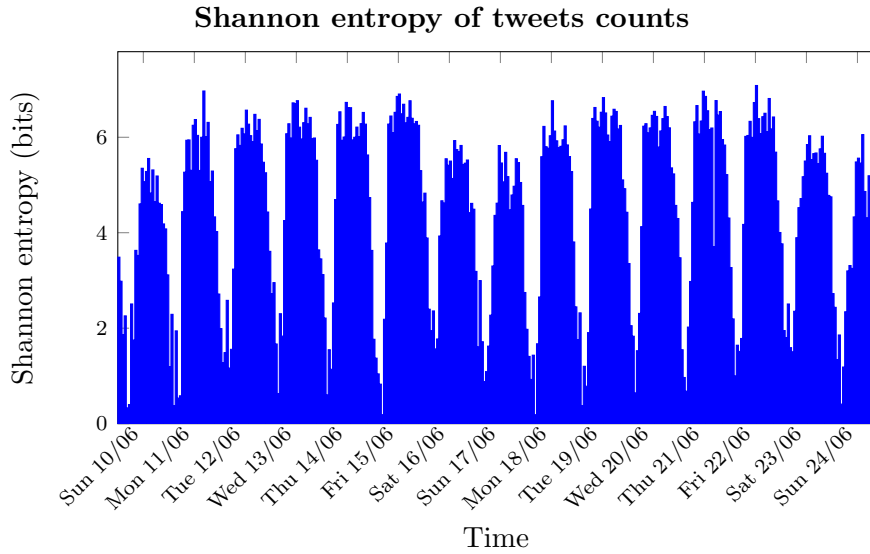
Finally, we study the heterogeneity of the spatial patterns using Shannon entropy. We subdivide the region of study into a grid of square cells. For each hour, we count the tweets falling in each cell. Normalising by the number of tweets posted in that hour, we obtain a distribution of tweets, of which we compute the Shannon entropy, as a measure of uniformity (See Section 2.4).

Fig. 6.4 shows the Shannon entropy for each hour when each cell is a square with side 25m (taken as an approximation for the size of a housing or office block). The results when the cell side is 50m or 100m are nearly identical and thus are omitted.

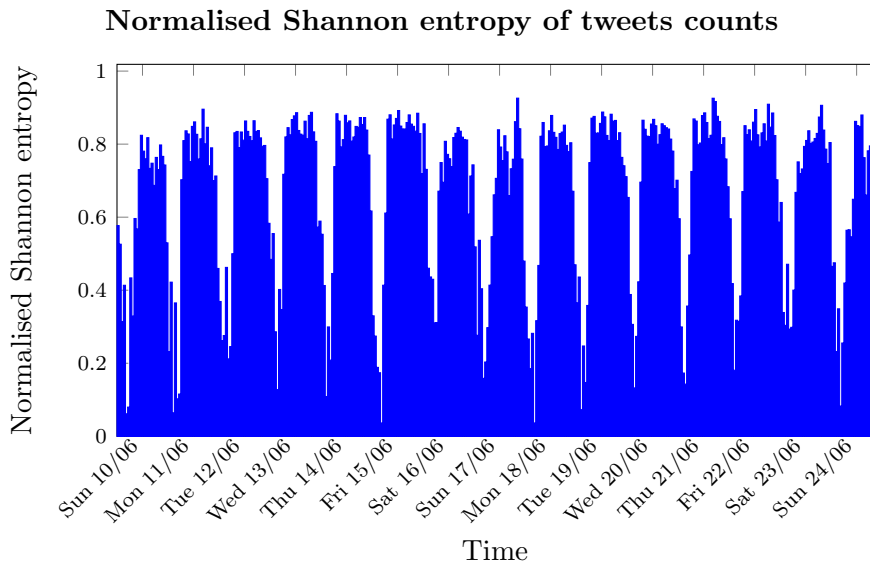
The pattern in Fig. 6.4a largely retraces the hourly counts already shown in Fig. 6.2b, with marked daily and weekly oscillations. This is in part to be expected given that the Shannon entropy depends also on the number of occupied cells, and thus between two equally spatially heterogeneous patterns, one with more tweets is likely to have larger entropy.

To discard the effect of the hourly counts, we normalise the entropy by its maximum theoretical value given the number of tweets. Given that the number of cells exceeded the tweet counts at each hour, this simply corresponded to $\log n_h$, where n_h is the number of tweets posted in hour h . The normalised

⁴Remark, however, that the cluster sizes are generally small and all but a handful of these patterns (13 and 16 respectively) show values of \hat{F} and \hat{G} that both refute the Complete Spatial Randomness hypothesis. Note, furthermore, that these tests alone do not allow to determine whether alternative cluster or inhomogeneous processes would provide better models, as measured, for instance, by minimising some distance metric between the values of \hat{F} and \hat{G} observed for the patterns and the average values \bar{F} and \bar{G} obtained for the models.



(a) Shannon entropy of tweets counts in the rectangular area in central London. The values are heavily influenced by the hourly counts.



(b) Normalised Shannon entropy of tweets counts in the rectangular area in central London. The pattern is more concentrated at night and more uniform during the day. No major differences are visible between different days.

Figure 6.4: Shannon entropy of tweets counts in the rectangular area. in both its unnormalised and normalised form. The area is subdivided into a grid of square cells with side 25m. Results are almost identical when larger cells of 50m and 100m side are used. The robustness of Shannon entropy to the grid resolution was also remarked in [152] for the process in the overall region.

entropy in Fig. 6.4b still shows remarkable daily oscillations: the tweets tend to be more concentrated during the night and more uniformly distributed during the day. This applies to all days, with no remarkable oscillations between different days of the week.

6.2 Adapting an entropy-based spatial interaction model for connections in social networks

We have seen in Section 6.1 that geotagged data from a social media platform captures spatio-temporal fluctuations in its users' behaviour and special events, and allows to recognise salient aspects of urban morphology such as transport infrastructure, to which human activities are intimately tied. A more complex question is whether this kind of data can also reflect long-term changes in urban structure and the human activities enabled by it; for example, whether it allows to evaluate the effects of the construction a new transportation link on social connectivity.

A necessary step for adopting such a longitudinal approach is the identification of a baseline behaviour of the social network against which comparing anomalies and ascertain the presence of a trend. A key issue with using modern social media platforms data to this scope is that they themselves tend to evolve in time, as they increase or wane in popularity or change their features in a way that affect their users' engagement, making it more difficult to extricate actual social trends from purely platform-related ones.

In this section, we discuss a practical case of this problem. It stems out of a collaborative project aimed at identifying the effects of new highways and bridges on inter-city social connectivity in Hungary⁵. We use the Hungarian social network iWiW, shorthand for "International Who is Who", introduced in Section 2.2.2.1.

iWiW had a number of peculiarities that make it particularly suited for a longitudinal study of social connectivity and how it is impacted by new physical infrastructure. First of all, its popularity in the country: 40% of the population joined the platform. Registrations to the platform were only possible by invitation from a user, and each user had only a limited number of opportunities to send an invitation. More precisely, each new user was given the chance to send exactly one invitation, but occasionally the invitation

⁵The project was led by Eszter Bokányi, and I contributed with the present adaptation of an entropy-maximisation model. The data and hence the results, however, are proprietary.

opportunities would be replenished for all users. This factor suggests that an invitation was likely the sign of a strong social bound in real life. Furthermore, every user had to provide information on their city of residence, which means that some spatial information was known for each user. Finally, in the years the platform was active, Hungary saw the inauguration of several new highways and bridges that reduced travel times between several cities.

Our data set consisted of monthly counts of registrations in the Hungary’s eighteen regional capitals as well as the national capital Budapest. For each pair of cities u and v , we knew how many invitations were sent from u to v in each month. To see whether the reduction of travel time between two cities following the construction of new infrastructure yielded an increase in invitations being sent between them, we needed a baseline. This required, first of all, filtering out the effects of the internal dynamics of the social network, such as the change in frequency of the invitations at time t in the overall network. We could then focus only on the relative prevalence of links between u to v with respect to pairs of cities that did not see a reduction of their travel time.

To define a baseline, we adopted an adapted version of Wilson’s retail model, described by eq. (2.2). Invitations were interpreted as a special “currency” that a user in a location u could spend in a location v . The basic assumption of the model is that the probability that a user from u has a social tie with a person in v and thus spends their invitation in v at time t is larger when the travel time $c_{uv}(t)$ from u to v at time t is smaller.

Nonetheless, a few differences with Wilson’s model have to be remarked. The number of invitations sent from u at time t does not depend directly on the number of people in u , but rather on the total number of unsent invitations $C_u(t)$ still available to active users in u . The benefit b_v of sending an invitation to someone in v does not depend on the population of v (as assumed, for instance, in eq. (5.1)), but rather on the number $n_v(t)$ of people who live in v and have not joined the platform yet.

With these assumptions, we obtain that the expected number $w_{uv}(t)$ of invitations sent from u to v is

$$w_{uv} \propto C_u(t) \frac{n_v(t)^\alpha \cdot e^{-\beta c_{uv}(t)}}{\sum_x (n_x(t)^\alpha \cdot e^{-\beta c_{ux}(t)})},$$

for some parameters α and β .

In practice, more complex temporal dynamics are at play in this social

network that cannot be captured by a simple spatial interaction model. As remarked in [132], the number of invitations sent in iWiW followed typical patterns of the diffusion of innovation: the first users tended to be concentrated in Budapest, followed by university cities, and only later the density of connections became sufficiently homogeneous to allow a comparison.

Furthermore, new features were added to the social network during the late stages of its life cycle, such as the possibility to trade invitations with other users, that further complicate the modelling and weaken the assumption that an invitation represents a strong social tie. In conclusion, the rapid evolution of online social network makes the identification of long term trends such as the impact of infrastructure construction on social ties a challenging task.

6.3 Cultural Exchanges

In Section 6.2, we have modelled social connectivity between cities from ties between individuals in an online platform, and discussed the issue of reconstructing the current state of a social network via rapidly evolving data. Even if a fully encompassing view of the present social connections were available, though, it would not fully encapsulate the complexity of meaningful human relations. People, in fact, have the ability to transcend the limits of their immediate social circle, time and environment, and leave traces of themselves through their works and creations. A different point of view on human processes is centred on cultural exchanges, or, in other words, how one's ideas and innovations may spread and influence others across space and time. Whilst the diffusion of ideas is by its own nature rather nebulous, in many cases, it is possible to have explicit, unambiguous and spatio-temporally located signs of this passage, e.g. in the form of patents and scholarly citations, both of which naturally lend themselves to be modelled via networks [18, 226] (see Section 2.5.3.1). In other cases, the transmission of an idea remains rather implicit, and establishing its existence as a link, its direction, and its spatio-temporal properties requires a much more attentive analysis of the items that bear its traces.

In this section, we reconstruct an evolving network of cultural exchanges between poets and regions in the historical territory of Occitania in South-Western Europe, and discuss how to manage high levels of uncertainty in the data. We look at metrical and musical imitations in Occitan lyrics of

the 12th and 13th century⁶, more specifically at the practice of *contrafaction*, whereby an author reused a preexisting melody for composing a new text. The existence of a link between a model and a *contrafactum* is established through a musicological and philological analysis of the songs⁷. Nonetheless, the direction of the relationship between two songs is often uncertain, as is the temporal information on when the songs were composed.

6.3.1 Data set

The data on Occitan lyrics is taken from the Connecting Medieval Music database⁸, which is based on the Bibliografia Elettronica dei Trovatori (BEdT) [7], a relational database of Provençal troubadour texts, and the Dizionario biografico dei trovatori [98], complemented by critical editions and secondary literature.

For each text, the database specifies its author; composition year or year range, when the exact year is not known; whether it is a *contrafactum*, a model for one or more *contrafacta*, or, more generally, it has metrical analogies with other songs. For each author, the database specifies his area of origin, and the period in which he was active as a year range. Based on this information, the BEdT subdivides the authors into six roughly chronological generations.

We selected the authors of the first three generations, who were active between the early 12th and the early 13th century. We selected all songs by these authors which bore metrical similarities with songs by authors of any generation, to study the lasting influence and impact of texts composed by authors of the first three generations on the whole corpus of Troubadour songs. Thus, we obtained a data set with 312 authors and 761 songs, with 483 cases of *contrafaction* and 383 instances of more general metrical similarities.

It is important to remark that a database could never be complete since most Occitan songbooks were only composed between the end of the 13th and

⁶This work is taken from [159], that I cowrote with Dr Stefano Milonia. The article dealt with the identification of influential authors amongst Occitan troubadours and is part of a larger project on reconstructing the complex network of relations between authors, patrons, works, and political events in European Medieval Music.

⁷The structural correspondence between the number of syllables and melodic units, the cross-examination of metrical schemes, the reuse of rhyme scheme, sounds, and words, and the analysis of the compositions' themes all contribute to the identification of a link, even as lyrics in the manuscript sources are often not provided with musical notation. For a more detailed discussion on how links between songs are established, see the Introduction in [159] and the references within.

⁸The data set was compiled under the supervision of Stefano Milonia and is accessible at <https://medmus.warwick.ac.uk/contrafacta-page4?view=map>.

the beginning of the 14th century [9], and a non-quantifiable amount of works are lost [159].

6.3.2 Data cleaning

Beyond the incompleteness of the data set, a more technical problem concerns the uncertainty in the temporal information. As discussed in Section 2.5.3.1, the construction of temporal networks requires that each link be provided with precise temporal information about its onset and duration.

In the database, temporal indications are often vague and sometimes absent. To manage this uncertainty, it was necessary to adopt reasonable conventions, summarised in the following steps:

1. If the composition date of a song was absent, we used the activity period of its author. Note that also this may be an approximate time span.
2. If the temporal indications were vague, we used arbitrary but sensible and consistent numeric translation: “Beginning of the 13th century” became 1201-1210, “End of the 13th century” became 1291-1300, “Between the 12th and the 13th century” became 1191-1220, and “Middle of the 13th century” became 1226-1275.
3. If there were multiple hypotheses about the composition date, we used the most inclusive range (e.g., for “after 1194, and before 1198, around 1196” we used the range 1194-1198).
4. To ensure temporal consistency, if a contrafactum’s composition year was given as a year range, we made sure that the lower end of the range was larger than or equal to the composition year or lower end of the year range of its model.

Applying these conventions allowed to provide each song in the data set with a time interval in which it was composed (possibly a single year), and each author with an approximate activity period, in a way that made the temporal data self-consistent.

6.3.3 Networks of Authors’ Bodies of Work

6.3.3.1 Temporal Network Construction

We use the data to construct a temporal network in which each node represents the body of works of an author, and each link represents the borrowing of the

metric structure of a song. A node comes into existence in the first year of the author’s reported activity period, and remains in the network thenceforth, reflecting the fact that an author’s work keeps existing even after they stop composing. We add a directed link from an author’s oeuvre to another’s for each contrafactum the latter wrote based on a model by the former. When two songs by different authors show metrical similarities but the direction of the influence is unclear, we add a bidirectional link between the nodes.

Ideally, each link should be associated with the specific instant t when the imitation was composed. As the temporal information about the composition of a song is often provided only as a year range, say $[t_1, t_2]$, one needs to make a choice on how to fix the composition time t in this interval. Any choice, deterministic or random, is in some way arbitrary, and in Section 6.3.3.5 we discuss the effects of selecting t at the lower end of the interval, its upper end, its mid-point, or uniformly at random in the interval. Here we present the network obtained by fixing t as the lower end of the interval, that is $t = t_1$.

6.3.3.2 Remarks on the construction

A few remarks are needed in order to appreciate the properties and limitations of the above network model and its computational implementation.

Some of the songs in the database are *tensos* (dialogic compositions in which different poets respond to each other in alternating stanzas), and thus they have two or more authors. According to the network construction, if an author writes a contrafactum of a tenso, there will be links connecting each author of the tenso to the author of the contrafactum. Similarly, if the contrafactum of a song is a tenso, then several links will connect the author of the model to the authors of the contrafactum. Hence, there is no one-to-one correspondence between a link in the network and a contrafactum. The presence of *tensos* means that the act of writing a contrafactum is not necessarily a binary relation between two authors. Using simplicial complexes [8] instead of networks may thus provide a better model for the connections between the authors, but this goes beyond the scope of this analysis.

For the construction of the temporal network, we used the R package *tsna* [26]. The package is one of the most advanced options available for the construction and analysis of temporal networks, but does not currently allow for a temporal network to be a multigraph at any time t . Hence, it does not distinguish whether an author has composed one or more contrafacta of models by the same author in a given year. This potentially affects some network

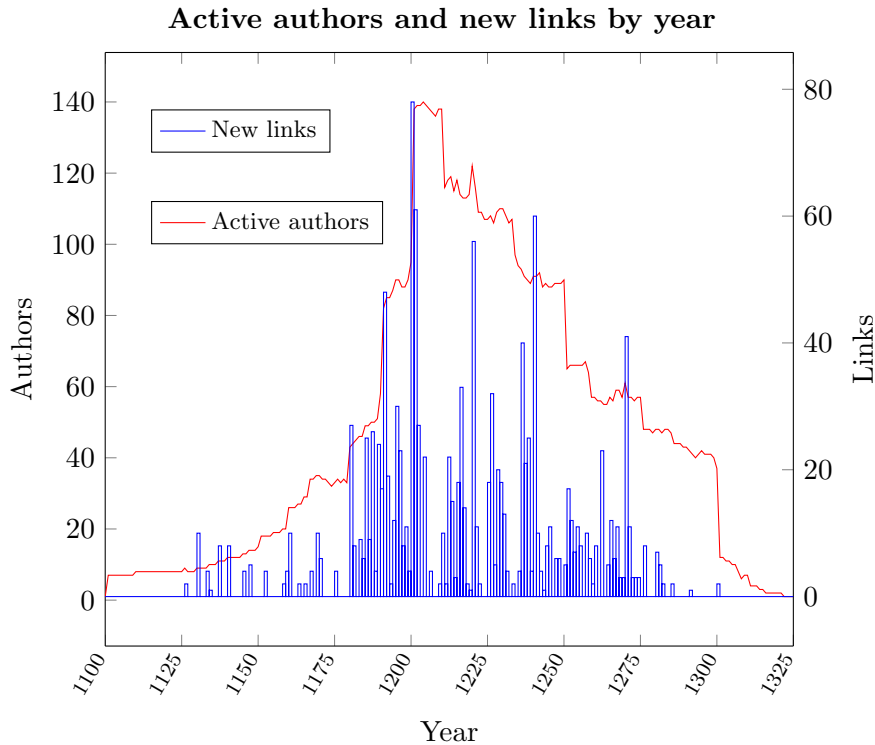


Figure 6.5: Number of authors active (left) and contrafacta or songs with metrical analogies composed (right) in each year.

measures that rely on link counts, such as degree and betweenness centrality. However, the occurrence of duplicate links is relatively rare (less than one fifteenth of the total links), so the overall effects are limited. Closeness centrality and the size of the reachability set, on the other hand, are unaffected by the presence of duplicate links.

All this considered, we deem the network model valid for our scopes. The final temporal network used in this study contains 312 nodes (one for each author's body of work), and 1011 directed dynamic links representing similarities between them, of which 580 signifying established evidence of contrafaction.

6.3.3.3 Temporal properties of networks of authors

Fig. 6.5 shows the number of authors who were active on a given year (continuous line plot, scale on the right hand side of the y-axis) and the number of new links in the temporal network (bar plot, scale on the left hand side of the y-axis). The temporal trajectories of these quantities are strongly related, as one would expected. The decrease in the number of authors who were active after 1220 is partly due to the fact that this is the chronological limit of au-

thors of the first three generation. Nonetheless we observe that compositions of the earlier generations kept being re-purposed until the last years of the troubadour era (1280-90).

6.3.3.4 Temporal influence measures

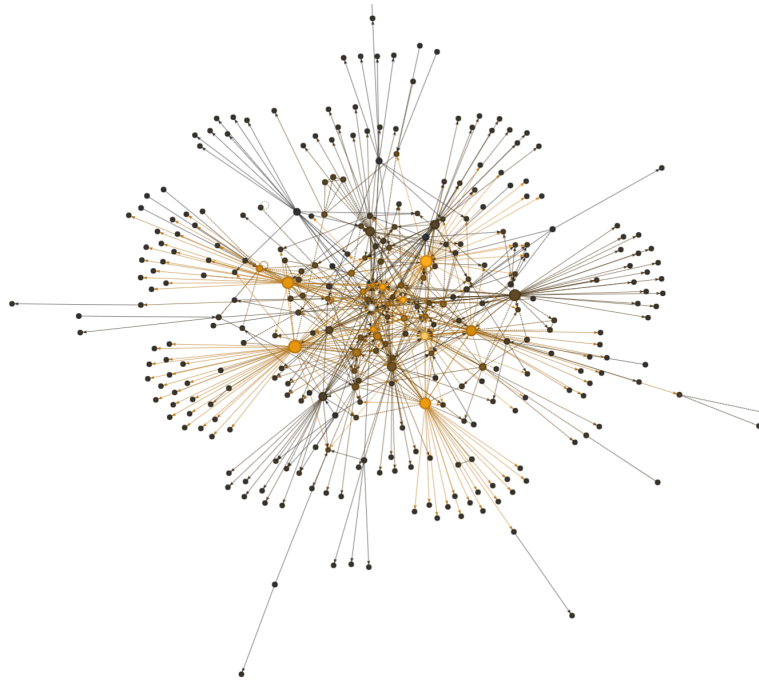
We use the above constructed temporal network to study the potential influence of a troubadour’s body of work on those of other troubadours. The underlying assumption here is that an author who has imitated songs written by others might have internalised some of their stylistic features, and thus his subsequent compositions might be signed by the their influence. Thus, a style propagates beyond the mere individual imitation, and is reflected in an author’s body of work, and in the work of those who imitate his works, and so on.

We display the network and the associated information in an interactive widget at <https://medmus.warwick.ac.uk/networks>⁹. The user can search the name of an author to highlight the corresponding node, together with all its adjacent nodes or all nodes in its reachability set, thus allowing an intuitive visualisation of the authors’ potential influence. Beyond focusing on a single author in the network, one can choose to display different measures of influence of all authors using the size and the colour of the corresponding nodes, thus obtaining a more holistic view on the network. The measures can be chosen among out-degree, in-degree, size of the reachability set, closeness centrality, and betweenness centrality, so to more easily identify authors of great influence or receptivity according to the various criteria of vertex importance described in Sections 2.5.2 and 2.5.3.2. The user can also choose to display all potential connections, including metrical analogies, or established contrafacta only. Some examples of the visuals produced with the above-described widget for different centrality measures are shown in Figs. 6.6 and 6.7.

In Fig. 6.6, the size of each node displays its out-degree (the larger the node, the larger the out-degree), while its brightness represents its in-degree (the lighter the node, the larger its in-degree). This provides a simple visualisation of an author’s influence in the overall network in terms of how many songs his works have directly inspired (out-degree) and how much he borrowed from others (in-degree). The lightest node at the centre of Fig. 6.6a represents Peire Cardenal as the most receptive imitator of authors from the first to the third

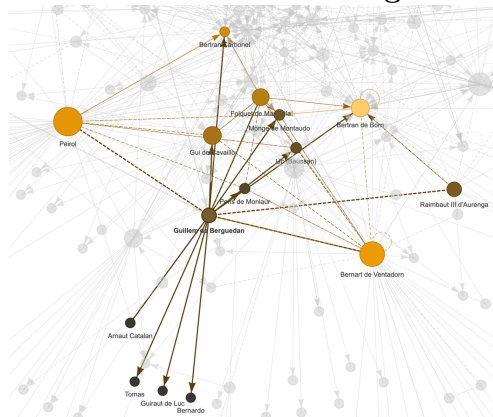
⁹The R code to generate, visualise, and analyse the networks can be found at <https://github.com/MatteoMazzamurro/contrafacta-networks/releases/tag/v1.1.0> [149].

Network of Authors (out- and in-degree)



(a) Main component of the network of authors' bodies of work.

Detail of Guillem de Berguedan



(b) Guillem de Berguedan's connections

Figure 6.6: Network of authors' oeuvres and their connections, based on the practice of contrafaction (solid lines) and on general metrical analogies (dashed lines). The size of a node is proportional to the number of contrafacta the corresponding author has inspired (plus metrical analogies), while its colour represents the number of contrafacta that the author's body of work has produced (plus metrical analogies): lighter shades denote larger values while darker shades denote smaller values. In the zoomed-in view, connections of one of the authors, Guillem de Berguedan, have been highlighted.

generation. Fig. 6.6b shows what happens when one selects an author in the widget, and zooms in to read the details of their connections. In this case, we chose the author Guillem de Berguedan, and highlighted his adjacent nodes for greater clarity. We notice that Guillem de Berguedan and Raimbaut III d’Aurenga appear to have similar sizes, and thus a similar degree of influence in terms of number of imitations of their works written by others, while Peirol and Bernart de Ventadorn are the most influential in terms of contrafacta inspired by their works and songs that can be connected to them by virtue of metrical analogies. Furthermore, Bertran de Born is qualified by a lighter colour, indicating that he has produced a greater number of imitations: he is indeed the most notorious composer of sirventes in the troubadour tradition and the main author of contrafacta among this selected group.

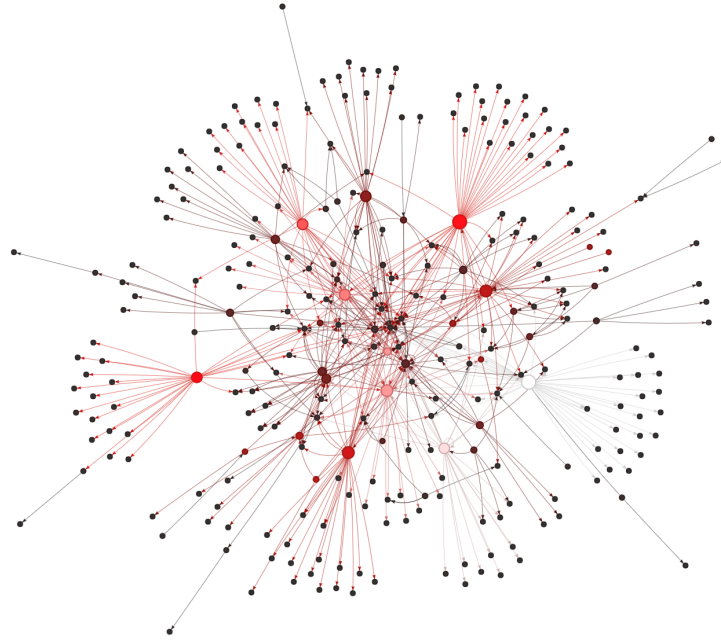
In Fig. 6.7, the size and the colour of nodes represent respectively their out-degree and size of the time-respecting reachability set. Only established links by contrafacta are shown. In Fig. 6.7a, we see the overall structure of the main component of the network, with a large numbers of imitators whose work was not imitated, a small group of authors whose works were widely imitated, and an even smaller group of authors who were both very receptive and influential. The lighter shade of Folquet de Marseilla in Fig. 6.7b reveals that he had a potentially larger impact on the overall network, as his work has indirectly reached a large number of authors, even if he is similar to Bertran de Born in terms of the number of contrafacta inspired by their works. Note also that Folquet has the largest reachability set in this group, as shown by its light shade, in spite of the fact that Raimbaut de Vaqueiras’ songs have proved very popular for direct imitations, as shown by its large size. Given that both authors’ activity periods are similar, this difference is due to the fact that authors who have imitated Raimbaut’s works further down the line were probably relatively minor ones compared to those who imitated Folquet’s works.

6.3.3.5 Robustness of the network measures

As mentioned in Section 6.3.3.1, the construction of the temporal network of authors requires to make an arbitrary assumption on the year t in which a contrafactum was composed within the year range $[t_1, t_2]$ provided in the data set. Potentially, this choice might impact the structure of the temporal network and its node centrality measures.

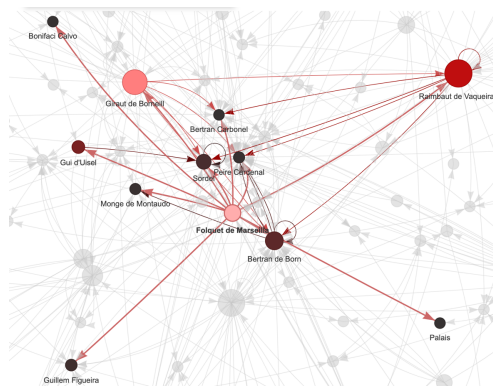
To evaluate this impact, we start by considering the temporal networks N_1 ,

Network of Authors (out-degree and reachability set)



(a) Main component of the network of authors' bodies of work.

Detail of Folquet de Marseilla



(b) Folquet de Marseilla's connections

Figure 6.7: Network of authors' bodies of work connected via the compositions of contrafacta. The size of a node is proportional to the number of contrafacta the corresponding author has inspired. The brightness of the colour of a node represents the size of its time-respecting reachability set, that is the number of authors he may have inspired with his work, directly or indirectly. In the zoomed-in view, the connections of one of the authors, Folquet de Marseilla, have been highlighted.

N_2 , and $N_{1,2}$, obtained by choosing $t = t_1$, $t = t_2$, and $t = \frac{t_1+t_2}{2}$, respectively. Note that the network N_1 is the one considered in the above analysis.

Table 6.1 shows the statistical properties of the in-degree and forward reachability for N_1 and N_2 : their maximum value, mean, and standard deviations. The values for $N_{1,2}$ are similar, and thus were omitted. Table 6.2 shows the difference between the measures when comparing node by node the networks N_1 and N_2 , in terms of the largest difference, the mean of the differences, and their standard deviation. These measures are shown for both the case in which only links corresponding to contrafacta are considered and that in which links representing metrical analogies are also included.

Choosing the year of composition t at the beginning or at the end of the proposed range has negligible effect on the in-degree; the impact on the forward reachability is larger on certain nodes, which is to be expected given the cumulative nature of this metric, but still contained overall.

Table 6.1: Statistics of centrality measures for N_1 and N_2

Measure	Analogies?	N_1 mean	N_2 mean	N_1 sd	N_2 sd	N_1 max	N_2 max
in-degree	no	1.737	1.731	3.133	3.03	36	34
in-degree	yes	3.058	3.026	5.08	4.917	45	42
reachability	no	5.676	5.978	17.156	18.167	131	118
reachability	yes	54.984	55.462	89.281	89.484	271	247

Table 6.2: Statistics of the differences in centrality measures for N_1 and N_2

Measure	Analogies?	mean $N_1 - N_2$	std deviation $N_1 - N_2$	max $N_1 - N_2$
in-degree	no	0.006	0.212	2
in-degree	yes	0.032	0.350	3
reachability	no	-0.302	3.200	30
reachability	yes	-0.478	35.297	196

To further test robustness, we also employ a non-deterministic approach. For each contrafactum or metrical analogy, we assign the creation date t by sampling a year uniformly at random within its range $[t_1, t_2]$. We repeat this operation one hundred times, each time obtaining a creation time for each link. We thus construct one hundred potentially different temporal networks. For each network, and for each node, we compute the reachability. We finally compute the coefficient of variation of the reachability of each node, that is the ratio of the standard deviation of the reachability for this node across the one hundred networks, and its mean.

When only contrafacta links are considered, the maximum coefficient of variation is 2.135, showing a large variability of reachability for this node, and thus a strong impact of the date selection procedure on the node. Yet, the

mean coefficient of variation is 0.047, indicating that the impact of the date selection on reachability is very limited on average. When metrical analogies are also included, the largest coefficient of variation is 1.486, whilst the mean is 0.10, again indicating that the impact of the date selection procedure on single nodes may be large, but it is small on average.

6.3.4 Networks of Regions

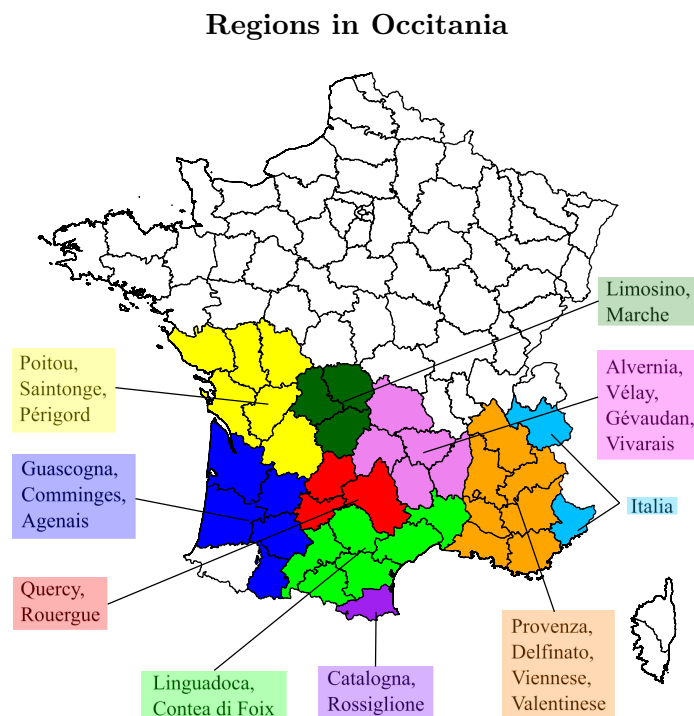
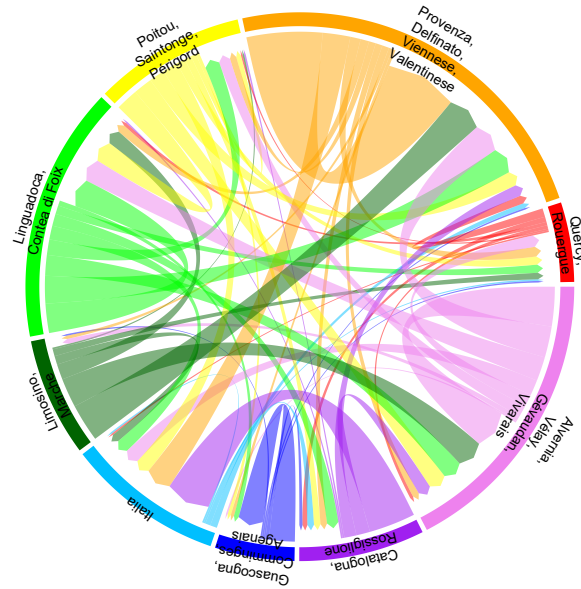


Figure 6.8: Approximate borders of Occitan regions in the database, superimposed to modern departments of France. Some regions (noticeably Italia and Catalogna-Rossiglione) extend beyond the limits of modern France.

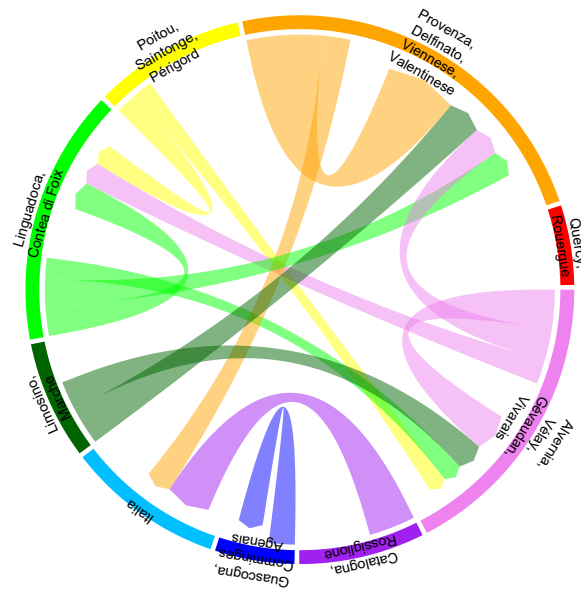
We aggregate the interactions between authors on the basis of the authors' area of origin to create a network of cultural exchanges between nine regions in Occitania (see Fig. 6.8): Alvernia, Velay, Gévaudan, Vivarais (AVGV); Catalogna, Rossiglione (CR); Guascogna, Comminges, Agenais (GCA); Italia (IT); Limosino, Marche (LM); Linguadoca, Contea di Foix (LCF); Poitou, Saintonge, Périgord (PSP); Provenza, Delfinato, Viennese, Valentinese (PDVV); and Quercy, Rouergue (QR).

The circle plot in Fig. 6.9a shows the intensity of the interactions between all pairs of regions. Each sector on the exterior annulus corresponds to a

Cultural exchanges in Occitania based on contrafacta



(a) All interactions by contrafacta.



(b) Main interactions by contrafacta only.

Figure 6.9: Directed networks of Occitan regions based on the production of contrafacta: arrows points from the regions where the authors of models were from to regions where the authors of their contrafacta were from. The thickness of the arrows represent the number of contrafacta, and is taken as an indication of literary influence.

region. The sector’s angle represents the proportion of all connections in the data set either starting or ending in that region, i.e., the proportion of contrafacta or models in the data set that were composed by authors from that region. It is clear that in some regions, such as PDVV, and AVGV, the practice of contrafaction was more prolific than in others, such as QR and GCA. Inside the annulus, the thickness of the arrows between different sectors corresponds to how many contrafacta authors from the region at the head of the arrow have composed based on models by authors from the region at its tail. For greater clarity, Fig. 6.9b highlights only the main connections. Some of the most intense connectivity is between authors within the same region, for instance in PDVV and AVGV, GCA, or LCF; or neighbouring regions, such as PDVV with AVGV and IT, AVGV with LM with LCF. In other cases, though, intense exchanges are found also between non-neighbouring regions, such as PSP and LCF or AVGV, or CR and IT.

To analyse the spatial pattern of literary influences weaved by the practice of contrafaction, we fitted the entropy-maximisation model from eq. (2.2) to this system of interactions. We used the distance between the centroids of the regions as the “cost” of interactions and the region’s total lyric production as its “population” and “benefit” of interacting with it. The best choice of the parameter β in terms of mean square error of the prediction turns out to be just $\beta = 0$, indicating that distance does not play a strong role on the interactions. In other words, the cultural interactions between these regions are dominated by factors other than their physical distance.

Perhaps not surprisingly, this is not the case when we analyse the production of tenzos (Fig. 6.10). The authors who worked closely together in the composition of these collaborative lyrics, indeed, were for the greatest part originally from the same region or neighbouring regions (such as LCF with PDVV and QR, or IT with PDVV, see Fig. 6.10b).

In Fig. 6.11, we show in greater detail the evolution of the main interactions in time, cumulatively (Figs. 6.11a, 6.11c, 6.11e), and in selected ten-year-long intervals (Figs. 6.11b, 6.11d, 6.11f).

From Fig. 6.11a, we can see that a large number of contrafacta were originally produced in GCA based on models by authors from the same region. In 1180-1190, LCD joined GCA as the most productive region (Fig. 6.11b), but LCD’s interaction were much more heterogeneous, as its models were imitated in neighbouring regions (PDVV, CR) and non-neighbouring regions (PSP) alike. By 1210 (Fig. 6.11c), the heart of the troubadour period, the

Cultural exchanges in Occitania based on tensos

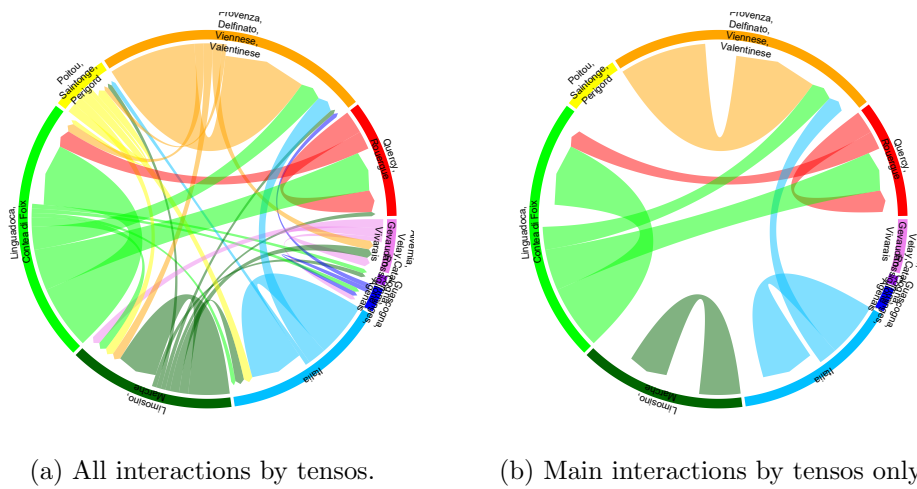
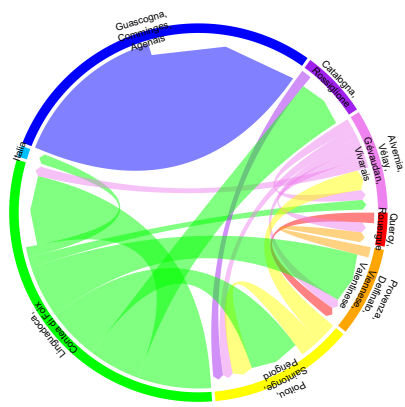


Figure 6.10: The network of Occitan regions based on the production of tensos by authors from each region. The arrows pointing from a region to itself are thicker than in the case of interactions by contrafacta.

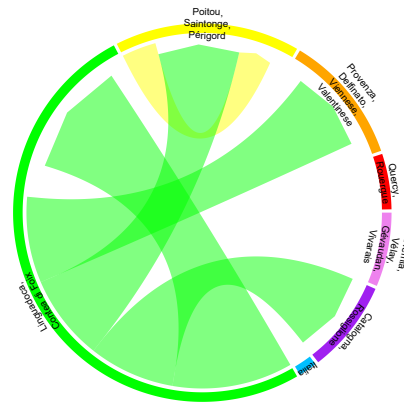
general structure of exchanges was already very similar to its final configuration. AVGV and PDVV establish themselves as the main areas of production of contrafacta, with the former being, just as LM and PSP, also an important source of models (Fig. 6.11d). In 1220-1230, (Fig. 6.11f), authors from IT produce a large number of contrafacta, based on models from neighbouring PDVV, and to a lesser extent, AVGV; slowly increasing its share in the total production (Fig. 6.11e), that would further increase by the end of the period.

We complement the analysis of the cumulative networks by shifting our attention from the total production to its heterogeneity via a study of the temporal evolution of the entropy of interactions in Fig. 6.12. Overall (black line with diamond), the interactions between the regions tended to become more heterogeneous until the central period (1200-1210), plateauing for a couple of decades before experiencing again a small increase. When we look at individual regions and where their models proved influential, we see that songs by authors from AVGV and LCF were already imitated in a large number of regions by the 1190's and these two regions remained amongst the top three in terms of heterogeneity ever since. The years leading to 1210 saw several regions increase their heterogeneity: with PDVV reaching its peak at the turn of the century and GCA, LM and GCA all witnessing an increase in their interaction entropy values before stagnating. After 1230, a last wave of changes in the heterogeneity is observed as QR and IT see their songs being imitated

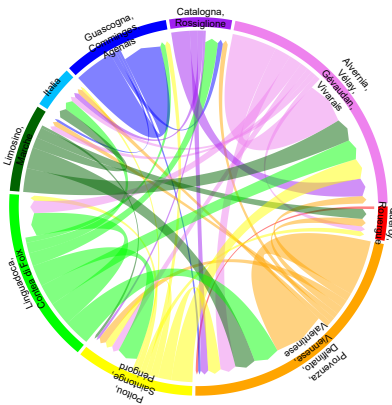
Evolution of cultural exchanges in Occitania based on contrafaction



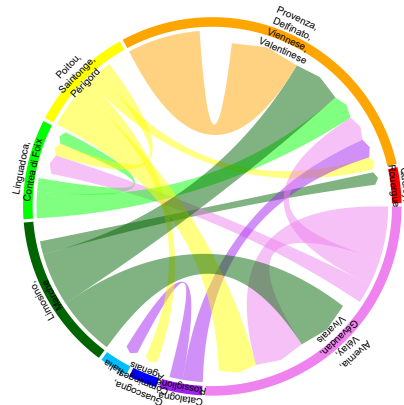
(a) Cumulative contrafacta until 1190.



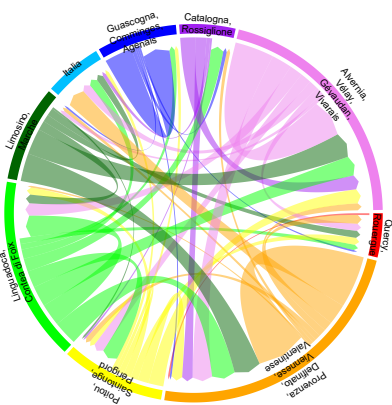
(b) Contrafacta 1180-1190.



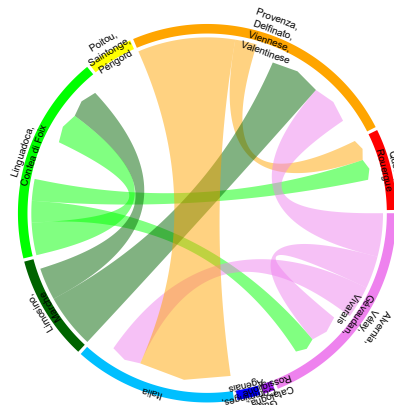
(c) Cumulative contrafacta until 1210.



(d) Contrafacta 1200-1210.



(e) Cumulative contrafacta until 1230.



(f) Contrafacta 1220-1230.

Figure 6.11: Temporal evolution of the network of Occitan regions based on the production of contrafacta by authors from each region.

Interaction entropy of Occitan regions

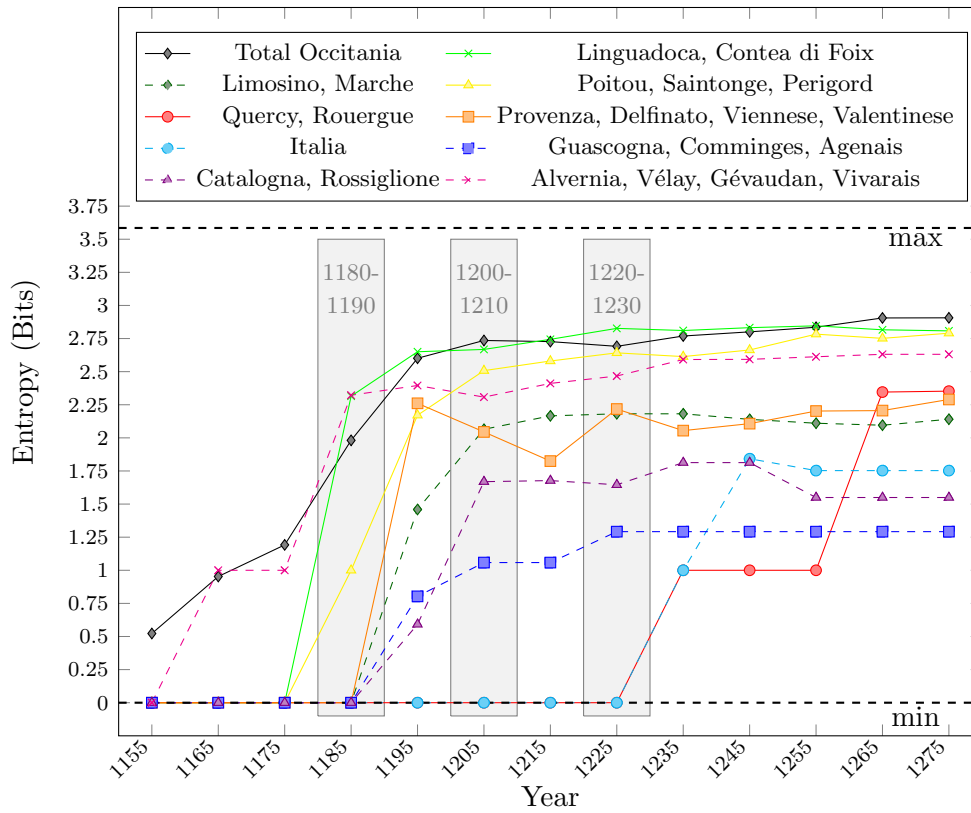


Figure 6.12: Temporal evolution in the interaction entropy of the network of Occitan regions. Recall that the weight of a link connecting a region to another one is equal to the number of contrafacta that authors from the latter have composed based on models by authors from the former.

by more than one region for the first time, which leads to positive entropy. Remark that neither of these regions become a major source of models, but a small number of songs by authors from QR are imitated throughout eastern Occitania, making the region's influence one of the most heterogeneous.

Chapter 7

Conclusion

7.1 Summary

In this thesis, we studied the structure and evolution of historical systems of cities using entropy-based approaches.

We analysed fundamental morphological aspects of the systems: the distribution of city sizes, the spatial organisation of their locations, and the population density and land use of their surrounding territories (Chapter 4). We discussed the impact of changes in the transport infrastructure on intercity distances and potential interactions, and how to capture this via updated measures of morphological polycentricity (Chapter 5). We looked at dynamics of the population distribution, social connectivity, and diffusion of ideas and innovations, discussing the opportunities and limitations of new and old forms of data in representing them (Chapter 6).

In doing so, we expanded and refined existing entropy measures and analytical tools. We proposed normalisation formulae for the first degree-based entropy of graphs, based on its extremal values for graphs subject to size and/or order constraints, thus improving its interpretation as a measure of uniformity of the degree sequence of a graph (Chapter 3). We defined local entropy: a spatial entropy measure for raster data that indicates, for each cell, how heterogeneous a quantity of interest is in its surroundings (Chapter 4). We introduced the normalised first degree-based entropy of the disc graphs as a tool to analyse the clustering of point processes (Chapters 4 and 5).

We illustrated our methods through examples of systems of cities in different places and historical periods. We discussed to what extent the Shannon entropy of city sizes in the main European powers from 1300 to 1850 was correlated to the main political events of the time. We performed a similar study

of the local entropy of the land use and population density in Italy, the British Isles and South Asia from 1700 to modern day, highlighting that the strongest structural changes happened in periods when new technological innovations were being adopted, whilst the correlation with political event was feebler. We analysed the spatial arrangement of English and Welsh towns in the 19th century, modelling its increasing clustering, and showing that railways made the system less balanced overall with respect to travel time, whilst allowing, in general, greater heterogeneity of connectivity at the city level.

We studied the spatial-temporal dynamics of social media activity using geo-tagged Tweets in London, discussing the limitations in adopting homogeneous point process models, and describing the heterogeneity of the distribution of Tweets via the entropy of cell counts. We studied the properties of transport, social, and cultural links between different localities via the worldwide network of flights, the Hungarian social network iWiW, and the literary imitations between medieval Occitan troubadours, respectively. We discussed the difficulties of adopting an entropy maximisation model for the intercity social connectivity in iWiW, despite the similarity of the problem with that of computing spending flows between localities. We showed that distance does not seem to have a major impact on the networks of cultural exchanges via the practice of *contrafactum* in Medieval Occitania, whilst it is more influential on the interactions via *tensos*.

We conclude the thesis highlighting some limitations of our work and discussing future research directions.

7.2 Limitations and future work

The research expounded in this thesis presents a number of limitations that open it to future expansion and improvement.

Minimum of the first degree-based graph entropy for graphs of arbitrary order and size. In Section 3.3, we determined the graphs attaining minimum entropy amongst all graphs with n nodes and m edges, when $n - 1 \leq m \leq 2n - 3$, and provided numerical observations, and limits for the case of large n and m , but we did not solve the problem for arbitrary n and m . We believe that some of the ideas introduced in the thesis, such as the function $h_c(G)$ in Definition 14, may be helpful to deal with the general case. Nonetheless, the problem remains difficult, as it depends on determining a

closed formula for the number $\ell(n)$ of non-standard minimal-entropy graphs (see Section 3.3.4).

Population data sets. Our comparative study of the entropy of historical city sizes is hindered by the high level of uncertainty in historical population estimates and by the scarce compatibility of different data sets (Section 4.1.1). These can be considered, to some extent, inevitable limitations. Indeed, though some data sets (including the employed [127]) have attempted to harmonise data from several sources, their data can only be considered reliable on average and not at the individual city level.

Another limitation is the greater focus on Europe than on other regions, primarily due to our greater familiarity with European history and languages, which simplified the consultation of sources. Other areas around the world have long (in fact longer) traditions of collecting relatively accurate population data through censuses, for instance China [79]. Data sets of city-level data do exist, such as an atlas of walled cities and urban extent [219], not dissimilar to [107]. Their analysis and comparison would be extremely enriching. This would require building more expert local knowledge and intuition.

Local entropy. Our definition of local entropy for raster in Section 4.3, employs neighbourhoods $C_{i,j}^n$, which act as observation windows around a cell. For now, these observation windows are defined according to a pseudo-Euclidean metric that ignores the presence of public infrastructure, and thus may not correspond to the actual accessibility of a city’s surrounding territory from the perspective of an inhabitant. Refined definitions using isochrones would provide a more intuitive description and reflect the change in the meaning of “neighbouring territory” when transport technology is taken into account, in line with our discussion in Section 5.2.

Entropy and energy considerations in point process modelling of settlement locations. In Section 4.4.5, we already discussed the limitations of modelling the location of new English and Welsh cities via an inhomogeneous Matérn cluster process, and provided possible solutions to some of its issues. In particular, we mentioned that, although satisfactory for most cases, the model fails to account for the repulsion between towns at small distance. A composite model in which the points in the clusters are distributed according to a Gibbs hard-core process might help solve the issue, at the expense of a

more complex fitting procedure.

One interesting aspect of employing an Gibbs hard-core process to model the interactions between towns is that this model is intrinsically connected to entropy maximisation of the total energy of the process (i.e., the sum of the interactions), as hinted at in Section 2.5.1.1. A further exploration of the concept of energy in patterns of town locations could open a fruitful research direction.

Normalisation of graph entropies for transportation networks. In real-world transportation networks, a settlement's location and the maximum range and speed of transport technology have a strong influence on its potential connections and thus on its entropy. This is of special interest when studying phenomena, such as commuting, that are heavily influenced by travel time.

A more appropriate normalisation constant for the entropy of interactions within a certain travel time t in Section 5.2 could be $\log(n(v, t))$, where $n(v, t)$ is the number of nodes attainable from v , given its location and t . Similarly, limitations in the flying range of airplanes mean that direct links between some extremely far airports are not technically possible or at least commercially viable with current technology, and this may impact an airport's connectivity (Section 5.1). This calls for an even more careful normalisation of the first degree-based graph entropy, with conditions on the maximum degree in the graph, or, even more ambitiously, given an arbitrary majorising sequence for the degree sequence.

Multi-modality and real travel conditions in the computation of travel time. In our computation of travel time between 19th century English and Welsh towns Section 5.2, we considered geographical distance divided by the stagecoach speed as a baseline and complemented it with distance along the railway network divided by an estimate of the average railway speed in the year of study. This approach has a number of limitations. First, it tacitly assumes the presence of direct stagecoach roads between towns and ignores their conditions (for instance, their steepness). Secondly, it ignores variations in train travel speed, train frequency, cost, line capacity, line transfer times, and time needed by a typical person to reach a station along the actual street network of the towns. Lastly, it ignores the presence of alternative means of transport, such as ferries or waterway, and thus the effects of multi-modality.

Our study aimed to provide a proof of concept, and any improvement in

these directions would contribute to strengthen it. The aforementioned Bradshaw's Monthly Railway and Steam Navigation Guides provide detailed train timetables for Britain from 1839 to 1961; transforming them into machine-readable data would greatly benefit our study, as well as any other study on historical transport networks and intercity interactions. The Cambridge Group for the History of Population and Social Structure, who created and published the railway data sets used in this thesis, also created multi-modal transport networks for Britain at various points in time. Whilst not all of this data is readily accessible, enquiries or collaborations with the group could allow further expansion of our work.

The limits of Spatial Interaction Models with historical data and parameter choice for most likely interaction. The qualitative nature of the fitting in Section 5.2.5 is certainly a limitation of difficult solution, in absence of complete, quantitative data on migration. Historical data on migration flows may be available at the county level at best, which makes it impossible to verify finer-grained connections in our network.

The fact that London occupies such a special position in the British system of cities also poses a challenge in the definition of a network via a Spatial Interaction Model with minimal assumptions. The migration flows, indeed, seem to follow the rule: if there is an attractive location nearby, migrate there, otherwise migrate to London. A realistic model would need to assign London a special premium, thus imposing some functional form on the predicted interactions. Substituting population size with data on economic activity, as predicted in Wilson's original model, might help with this challenge, but, as far as we are aware, such data is not generally available at such a capillary resolution.

Another limitation worth mentioning is the complex interplay between local migration and international emigration, or, more generally, interactions within a system and the interactions of a system with the outside. By drawing a boundary around a set of cities, we isolate them from the rest of the world, which is an unsatisfactory but necessary assumption. In the case of England and Wales, for example, we inevitably neglect the intense emigration that these nations were facing in the 19th century and their impact on inter-county or inter-town migration.

Inhomogeneous models for Tweets. Modelling the locations of geo-tagged

Tweets in London, we only tested homogeneous models in an attempt to keep the models as general and assumption free as possible. This approach led to a satisfactory modelling only in a minority of cases and in a limited area of the city. If one is interested in more accurate fitting, at the expense of simplicity and generality, inhomogeneous models are required. Based on our analysis, these ought to include spatial covariates such as population density, account for the street networks, and be applied to a data set that links Tweets to individual users, so to be able to satisfactorily address the issue of repeated points. Future models should also exploit the well-developed theory of spatio-temporal point processes [76].

Network of contrafacta. Our current work on cultural exchanges between Occitan regions looked at the first three generations of troubadours and at the practice of contrafactum (Section 6.3). In the future, we plan to extend it to other troubadour generations and incorporate weighted edges proportional to objective measures of similarity of the musical imitations to better capture the degree of relatedness between troubadours' bodies of works.

Potential interactions beyond spatial interaction models. The multiplication of virtual interactions opportunities in the modern world, though not quite spelling the once popular notion of “death of distance”, make the assumptions of the spatial interaction models less solid, as the ratio between cost and benefit of a physical interaction now is less clear and need to be more carefully calibrated. Furthermore, the entropy maximisation principle rests on the assumption that individuals movements at the basis of cities interactions can be modelled as those of particles, identical and deprived of agency. More realistic models of interactions should attempt to include considerations on individuals' skills, socio-economic conditions, as well as the complexity of intentionality and decision making, to achieve a truly human-centred perspective on intercity interactions.

Alternative notions of entropy. In this thesis, we have studied the structure and evolution of systems of cities through the lenses entropy, to which we have given primarily a statistical-mechanical and information-theoretical interpretation. We would like to conclude the thesis by mentioning that alternative notions of entropy do exist and have been applied to the study of cities. A particularly interesting one comes from ecology and sees cities as dissipative

systems that maintain their structure only by drawing in energy and resources, increasing the randomness and entropy in their host environment [181]. As it has been highlighted in [176], these studies are rather speculative, not technical, and based on the questionable assumption that the world is an isolated system. A rigorous exploration of the role that systems of cities may play in regulating the internal structure of individual cities and managing their entropy resonates with ours “philosophical” remarks from Section 2.4.3 and would be a fascinating line of research to pursue.

Bibliography

- [1] L. Altieri, D. Cocchi, and G. Roli. A new approach to spatial entropy measures. *Environmental and ecological statistics*, 25(1):95–110, 2018.
- [2] L. Altieri, D. Cocchi, and G. Roli. Measuring heterogeneity in urban expansion via spatial entropy. *Environmetrics*, 30(2):e2548, 2019.
- [3] G. E. Andrews. *The theory of partitions*. Cambridge university press, 1998.
- [4] E. Arcaute, E. Hatna, P. Ferguson, H. Youn, A. Johansson, and M. Batty. Constructing cities, deconstructing scaling laws. *Journal of the royal society interface*, 12(102):20140745, 2015.
- [5] E. Arcaute, C. Molinero, E. Hatna, R. Murcio, C. Vargas-Ruiz, A. P. Masucci, and M. Batty. Cities and regions in Britain through hierarchical percolation. *Royal Society open science*, 3(4):150691, 2016.
- [6] S. Arshad, S. Hu, and B. N. Ashraf. Zipf’s law and city size distribution: A survey of the literature and future research agenda. *Physica A: Statistical mechanics and its applications*, 492:75–92, 2018.
- [7] S. Asperti and L. de Nigro. Bibliografia elettronica dei trovatori- v. 2.5-2012. http://www.bedt.it/BEdT_04_25/, 2012. [Accessed: 09/02/2021].
- [8] R. H. Atkin. *Combinatorial connectivities in social systems: An application of simplicial complex structures to the study of large organizations*. Springer, 1977.
- [9] Avalle and Leonardi. *I manoscritti della letteratura in lingua d’oc*. Einaudi, Turin, 1993.

- [10] F. Baccelli and J. O. Woo. On the entropy and mutual information of point processes. In *2016 IEEE International Symposium on Information Theory (ISIT)*, pages 695–699. IEEE, 2016.
- [11] M. L. Bacci. *A concise history of world population*. John Wiley & Sons, 2017.
- [12] A. Baddeley, G. Nair, S. Rakshit, G. McSwiggan, and T. M. Davies. Analysing point patterns on networks—a review. *Spatial Statistics*, 42:100435, 2021.
- [13] A. Baddeley and R. Turner. spatstat: an R package for analyzing spatial point patterns. *Journal of statistical software*, 12:1–42, 2005.
- [14] P. Bairoch. Urbanisation and economic development in the western world: some provisional conclusions of an empirical study. In H. Schmal, editor, *Patterns Of European Urbanisation Since 1500*, pages 63–75. Routledge, 1981.
- [15] P. Bairoch, J. Batou, and C. Pierre. *Population des villes européennes de 800 à 1850: banque de données et analyse sommaire des résultats (la)*. Librairie Droz, 1988.
- [16] M. Barthélemy. Spatial networks. *Physics reports*, 499(1-3):1–101, 2011.
- [17] M. Barthélemy. *The structure and dynamics of cities*. Cambridge University Press, 2016.
- [18] V. Batagelj, P. Doreian, A. Ferligoj, and N. Kejžar. Citation pattern in temporal United States patent data. In *Understanding large temporal networks and spatial networks: Exploration, pattern searching, visualization and network evolution (Vol. 2)*. John Wiley & Sons, 2014.
- [19] M. Batty. Spatial entropy. *Geographical analysis*, 6(1):1–31, 1974.
- [20] M. Batty. Entropy in spatial aggregation. *Geographical Analysis*, 8(1):1–21, 1976.
- [21] M. Batty. Polynucleated urban landscapes. *Urban studies*, 38(4):635–655, 2001.
- [22] M. Batty. Rank clocks. *Nature*, 444(7119):592–596, 2006.

- [23] M. Batty, R. Morphet, P. Masucci, and K. Stanilov. Entropy, complexity, and spatial information. *Journal of geographical systems*, 16(4):363–385, 2014.
- [24] A. Bavelas. A mathematical model for group structures. *Human organization*, 7(3):16–30, 1948.
- [25] M. A. Beauchamp. An improved index of centrality. *Behavioral science*, 10(2):161–163, 1965.
- [26] S. Bender-deMoll and M. Morris. tsna: Tools for temporal social network analysis. *R package version 0.2. 0*, URL <https://CRAN.R-project.org/package=tsna>, 2016.
- [27] R. J. Bennett. Urban population database, 1801-1911 [data collection]. uk data service. sn. 7154, 2012.
- [28] P. Beria and V. Lunkar. Presence and mobility of the population during the first wave of covid-19 outbreak and lockdown in Italy. *Sustainable Cities and Society*, 65:102616, 2021.
- [29] B. J. Berry. Cities as systems within systems of cities. *Papers in regional science*, 13(1):147–163, 1964.
- [30] B. J. Berry and P. J. Schwind. Information and entropy in migrant flows. *Geographical Analysis*, 1(1):5–14, 1969.
- [31] L. M. Bettencourt. The origins of scaling in cities. *Science*, 340(6139):1438–1441, 2013.
- [32] L. M. Bettencourt. Impact of changing technology on the evolution of complex informational networks. *Proceedings of the IEEE*, 102(12):1878–1891, 2014.
- [33] M. Bida and C. Rozenblat. Modelling hierarchy and specialization of a system of cities from an evolutionary perspective on firms’ interactions. In *Theories and Models of Urbanization*, pages 279–302. Springer, 2020.
- [34] S. Boccaletti, G. Bianconi, R. Criado, C. I. Del Genio, J. Gómez-Gardenes, M. Romance, I. Sendina-Nadal, Z. Wang, and M. Zanin. The structure and dynamics of multilayer networks. *Physics reports*, 544(1):1–122, 2014.

- [35] D. Bogart, L. Shaw-Taylor, and X. You. The development of the railway network in Britain 1825-1911. *Transport, Urbanization and Economic Development in England and Wales C*, pages 1670–1911, 2018.
- [36] D. Bonchev. *Information Theoretic Indices for Characterization of Chemical Structures*. Research Studies Press, Chichester, 1983.
- [37] B. Boots, K. Sugihara, S. N. Chiu, and A. Okabe. *Spatial tessellations: concepts and applications of Voronoi diagrams*. John Wiley & Sons, 2009.
- [38] F. Botta, H. S. Moat, and T. Preis. Quantifying crowd size with mobile phone and Twitter data. *Royal Society Open Science*, 2, 2015.
- [39] J.-P. Bouchaud and M. Mézard. Wealth condensation in a simple model of economy. *Physica A: Statistical Mechanics and its Applications*, 282(3-4):536–545, 2000.
- [40] A. Bretagnolle and D. Pumain. Simulating urban networks through multiscalar space-time dynamics: Europe and the United States, 17th-20th centuries. *Urban Studies*, 47(13):2819–2839, 2010.
- [41] A. Bretagnolle, D. Pumain, and C. Vacchiani-Marcuzzo. The organization of urban systems. In *Complexity perspectives in innovation and social change*, pages 197–220. Springer, 2009.
- [42] M. Brezzi and P. Veneri. Assessing polycentric urban systems in the OECD: Country, regional and metropolitan perspectives. *European Planning Studies*, 23(6):1128–1145, 2015.
- [43] C. Broodbank. *An island archaeology of the early Cyclades*. Cambridge University Press, 2002.
- [44] T. Brughmans and M. A. Peeples. Spatial networks. In *Archaeological Spatial Analysis*, pages 273–295. Routledge, 2020.
- [45] M. Burger and E. Meijers. Form follows function? Linking morphological and functional polycentricity. *Urban studies*, 49(5):1127–1149, 2012.
- [46] F. Cairncross. *The Death of Distance*. Harvard Business School Press, Cambridge, 1997.

- [47] S. Cambie, Y. Dong, and M. Mazzamurro. Extremal values of degree-based entropies of bipartite graphs. *arXiv preprint arXiv:2206.00922*, 2022.
- [48] S. Cambie and M. Mazzamurro. Minimum entropy of graphs with given size. *arXiv preprint arXiv:2204.08251*, Apr. 2022.
- [49] S. Cambie and M. Mazzamurro. Resolution of Yan’s conjecture on entropy of graphs. *MATCH Commun. Math. Comput. Chem.*, 89:389–404, 2023.
- [50] B. M. Campbell. *The great transition*. Cambridge University Press, 2016.
- [51] S. Cao and M. Dehmer. Degree-based entropies of networks revisited. *Applied Mathematics and Computation*, 261:141–147, 2015.
- [52] S. Cao, M. Dehmer, and Y. Shi. Extremality of degree-based graph entropies. *Information Sciences*, 278:22–33, 2014.
- [53] R. Cardoso and E. Meijers. Metropolisation: The winding road toward the citification of the region. *Urban Geography*, 42(1):1–20, 2021.
- [54] A. Champion, G. Hugo, and A. Lattes. Towards a new conceptualization of settlement for demography: Beyond the urban/rural dichotomy. *Population and Development Review*, 29:277–97, 2003.
- [55] T. Chandler. *Four thousand years of urban growth: An historical census*. Mellen, 1987.
- [56] G. Chapman. The spatial organization of the population of the United States and England and Wales. *Economic Geography*, 49(4):325–343, 1973.
- [57] Y. Chen and L. Huang. Spatial measures of urban systems: From entropy to fractal dimension. *Entropy*, 20(12):991, 2018.
- [58] Z. Chen, M. Dehmer, and Y. Shi. Bounds for degree-based network entropies. *Applied Mathematics and Computation*, 265:983–993, 2015.
- [59] V. G. Childe. The urban revolution. *The town planning review*, 21(1):3–17, 1950.

- [60] W. Christaller. *Central places in southern Germany*, volume 10. Prentice-Hall, 1966.
- [61] CIESIN - Center for International Earth Science Information Network - Columbia University. Global population count grid time series estimates. Accessed 02-May-2021.
- [62] P. J. Clark and F. C. Evans. Distance to nearest neighbor as a measure of spatial relationships in populations. *Ecology*, 35(4):445–453, 1954.
- [63] M. Coscia, F. M. Neffke, and R. Hausmann. Knowledge diffusion in the network of international business travel. *Nature Human Behaviour*, 4(10):1011–1020, 2020.
- [64] T. M. Cover. *Elements of information theory*. John Wiley & Sons, 1999.
- [65] P. Crane and A. Kinzig. Nature in the metropolis. *Science*, 308(5726):1225–1226, 2005.
- [66] G. Csardi, T. Nepusz, et al. The igraph software package for complex network research. *InterJournal, complex systems*, 1695(5):1–9, 2006.
- [67] L. Curry. Explorations in settlement theory. The random spatial economy. Part I. *Annals, Association of American Geographers*, 1963.
- [68] M. F. Dacey. Modified Poisson probability law for point pattern more regular than random. *Annals of the Association of American Geographers*, 54(4):559–565, 1964.
- [69] K. C. Das and M. Dehmer. A conjecture regarding the extremal values of graph entropy based on degree powers. *Entropy*, 18(5):183, 2016.
- [70] J. De Vries. *European Urbanization, 1500-1800*, volume 4. Routledge, 2006.
- [71] M. Dehmer. Information processing in complex networks: Graph entropy and information functionals. *Applied Mathematics and Computation*, 201(1-2):82–94, 2008.
- [72] M. Dehmer and A. Mowshowitz. A history of graph entropy measures. *Information Sciences*, 181(1):57–78, 2011.

- [73] M. Dehmer, K. Varmuza, S. Borgert, and F. Emmert-Streib. On entropy-based molecular descriptors: Statistical analysis of real and synthetic chemical structures. *Journal of chemical information and modeling*, 49(7):1655–1663, 2009.
- [74] E. Denis, O. Telle, S. Benkimoun, P. Mukhopadhyay, and S. Nath. Mapping the lockdown effects in India: how geographers can contribute to tackle covid-19 diffusion. *The Conversation*, 2020.
- [75] Department of Economic and Social Affairs, Population Studies. Growth of the world’s urban and rural population, 1920-2000. 1969.
- [76] P. J. Diggle. Spatio-temporal point processes: methods and applications. *Monographs on Statistics and Applied Probability*, 107:1, 2006.
- [77] T. Došlic, B. Furtula, A. Graovac, I. Gutman, S. Moradi, and Z. Yarahmadi. On vertex-degree-based molecular structure descriptors. *MATCH Commun. Math. Comput. Chem.*, 66(2):613–626, 2011.
- [78] S. Dühr. Potentials for polycentric development in Europe: The ESPON 1.1.1 project report. *Planning, Practice & Research*, 20(2):235–239, 2005.
- [79] J. D. Durand. The population statistics of China, AD 2–1953. *Population Studies*, 13(3):209–256, 1960.
- [80] M. Düring. How reliable are centrality measures for data collected from fragmentary and heterogeneous historical sources? A case study. In *The Connected Past*. Oxford University Press, 2016.
- [81] M. Eliasi. On extremal properties of general graph entropies. *MATCH Commun. Math. Comput. Chem.*, 79(3):645–657, 2018.
- [82] European Commission. ESDP- European Spatial Development Perspective: Towards a balanced and sustainable development of the territory of the European Union. 1999.
- [83] Facebook Connectivity Lab and Center for International Earth Science Information Network – CIESIN – Columbia University. High resolution settlement layer (hrsl), 2016. Accessed 07-May-2021.
- [84] Y. Fan, R. Guo, Z. He, M. Li, B. He, H. Yang, and N. Wen. Spatio-temporal pattern of the urban system network in the Huaihe River Basin based on entropy theory. *Entropy*, 21(1):20, 2018.

- [85] L. C. Freeman. A set of measures of centrality based on betweenness. *Sociometry*, pages 35–41, 1977.
- [86] L. C. Freeman. Centrality in social networks conceptual clarification. *Social networks*, 1(3):215–239, 1978.
- [87] X. Gabaix. Zipf’s law for cities: an explanation. *The Quarterly journal of economics*, 114(3):739–767, 1999.
- [88] L. Gan, D. Li, and S. Song. Is the Zipf law spurious in explaining city-size distributions? *Economics Letters*, 92(2):256–262, 2006.
- [89] M. Gerber. Predicting crime using Twitter and kernel density estimation. *Decision Support Systems*, 61, 2014.
- [90] A. Getis. Second-order analysis of point patterns: the case of Chicago as a multi-center urban region. In *Perspectives on Spatial Data Analysis*, pages 83–92. Springer, 2010.
- [91] A. Ghalavand, M. Eliasi, and A. Ashrafi. First degree-based entropy of graphs. *Journal of Applied Mathematics and Computing*, 59(1-2):37–46, 2019.
- [92] R. Gibrat. Les inégalités économiques. 1931.
- [93] J. Gil-Mendieta and S. Schmidt. The political network in Mexico. *Social Networks*, 18(4):355–381, 1996.
- [94] L. Glass and W. R. Tobler. General: Uniform distribution of objects in a homogeneous field: Cities on a plain. *Nature*, 233(5314):67–68, 1971.
- [95] N. Green. Functional polycentricity: A formal definition in terms of social network analysis. *Urban studies*, 44(11):2077–2103, 2007.
- [96] D. Gregory. The friction of distance? Information circulation and the mails in early nineteenth-century England. *Journal of Historical Geography*, 13(2):130–154, 1987.
- [97] A. Gudmundsson and N. Mohajeri. Entropy and order in urban street networks. *Scientific reports*, 3(1):1–8, 2013.
- [98] S. Guida and G. Larghi. *Dizionario biografico dei trovatori*. Mucchi editore Modena, 2014.

- [99] W. Guo, N. Gupta, G. Pogrebna, and S. Jarvis. Understanding happiness in cities using Twitter: Jobs, children, and transport. In *2016 IEEE International Smart Cities Conference (ISC2)*, pages 1–7, Sep. 2016.
- [100] W. Guo, X. Lu, G. M. Donate, and S. Johnson. The spatial ecology of war and peace. *arXiv preprint arXiv:1604.01693*, 2016.
- [101] R. Haining. Describing and modeling rural settlement maps. *Annals of the Association of American Geographers*, 72(2):211–223, 1982.
- [102] P. G. Hall and K. Pain. *The polycentric metropolis: Learning from mega-city regions in Europe*. Routledge, 2006.
- [103] L. Han, F. Escolano, E. R. Hancock, and R. C. Wilson. Graph characterizations from von Neumann entropy. *Pattern Recognition Letters*, 33(15):1958–1967, 2012.
- [104] R. K. S. Hankin. Additive integer partitions in R. *Journal of Statistical Software, Code Snippets*, 16, May 2006.
- [105] J. Hanson and T. Brughmans. Settlement scale and economic networks in the Roman Empire. In T. Brughmans and A. Wilson, editors, *Simulating Roman Economies: Theories, Methods, and Computational Models*. Oxford University Press Oxford, 2021.
- [106] J. W. Hanson. *An urban geography of the Roman world, 100 BC to AD 300*. Archaeopress, 2016.
- [107] J. W. Hanson. Cities database. <http://oxrep.classics.ox.ac.uk/databases/cities/>, 2017.
- [108] J. Hatcher et al. *Plague, population and the English economy 1348-1530*. Published for the Economic History Society by Macmillan & Co., 1977.
- [109] K. E. Haynes and W. T. Enders. Distance, direction, and entropy in the evolution of a settlement pattern. *Economic Geography*, 51(4):357–365, 1975.
- [110] J. Henneberg, M. Satchell, X. You, L. Shaw-Taylor, E. Wrigley, and M. Cobb. 1881 England, Wales and Scotland Railway Stations [data collection]. 10.5255/ukda-sn-852996, 2021.
- [111] H. Hilhorsta. Statistical properties of planar voronoi tessellations. *The European Physical Journal B*, 64(3):437–441, 2008.

- [112] J. B. Hill, M. A. Peeples, D. L. Huntley, and H. J. Carmack. Spatializing social network analysis in the late precontact US Southwest. *Advances in Archaeological Practice*, 3(1):63–77, 2015.
- [113] P. Holme. Network reachability of real-world contact sequences. *Physical Review E*, 71(4):046119, 2005.
- [114] P. Holme and J. Saramäki. A map of approaches to temporal networks. In *Temporal Network Theory*, pages 1–24. Springer, 2019.
- [115] R. A. Houston. *The population history of Britain and Ireland 1500-1750*, volume 18. Cambridge University Press, 1995.
- [116] J. C. Hudson. A location theory for rural settlement. *Annals of the Association of American Geographers*, 59(2):365–381, 1969.
- [117] U. L. p. IGEAT (Institut de Gestion de l’Environnement et d’Aménagement du Territoire. ESPON 1.4.3: Study on Urban Functions, Final report, 2007.
- [118] A. Ilić. On the extremal values of general degree-based graph entropies. *Information Sciences*, 370:424–427, 2016.
- [119] J. Illian, A. Penttinen, H. Stoyan, and D. Stoyan. *Statistical analysis and modelling of spatial point patterns*. John Wiley & Sons, 2008.
- [120] J. Jacobs. *The economy of cities*. Vintage, 2016.
- [121] R. Jedwab, N. D. Johnson, and M. Koyama. Pandemics, places, and populations: Evidence from the Black Death. 2019.
- [122] W.-S. Jung, F. Wang, and H. E. Stanley. Gravity model in the korean highway. *EPL (Europhysics Letters)*, 81(4):48005, 2008.
- [123] J. Karamata. Sur une inégalité relative aux fonctions convexes. *Publ. Math. Univ. Belgrade*, 1:145–148, 1932.
- [124] A. Karlström and V. Ceccato. A new information theoretical measure of global and local spatial association. 2000. draft paper submitted to the European Regional Science Association, https://static.sys.kth.se/abe/safeplaces/publikationer/S_statistics.pdf (Accessed 17 November 2022).

- [125] W. S. Kendall. Lines and networks. *Markov Processes and Related Fields*, 20(1):81–106, 2014.
- [126] R. Kirsch and A. J. Radcliffe. Many triangles with few edges. *Electron. J. Combin.*, 26(2):Paper 2.36, 23, 2019.
- [127] K. Klein Goldewijk, A. Beusen, J. Doelman, and E. Stehfest. Anthropogenic land use estimates for the Holocene–HYDE 3.2. *Earth System Science Data*, 9(2):927–953, 2017. Accessed 02-May-2021.
- [128] K. Klein Goldewijk, A. Beusen, and P. Janssen. Long-term dynamic modeling of global population and built-up area in a spatially explicit way: HYDE 3.1. *The Holocene*, 20(4):565–573, 2010.
- [129] C. Knappett, T. Evans, and R. Rivers. Modelling maritime interaction in the Aegean Bronze Age. *Antiquity*, 82(318):1009–1024, 2008.
- [130] J. Lahmeyer. Populstat database. growth of the population per country in a historical perspective, including their administrative divisions and principal towns, 2004.
- [131] V. Latora and M. Marchiori. A measure of centrality based on network efficiency. *New Journal of Physics*, 9(6):188, 2007.
- [132] B. Lengyel, E. Bokányi, R. Di Clemente, J. Kertész, and M. C. González. The role of geography in the complex diffusion of innovations. *Scientific reports*, 10(1):1–11, 2020.
- [133] L. Li, M. F. Goodchild, and B. Xu. Spatial, temporal, and socioeconomic patterns in the use of Twitter and Flickr. *Cartography and Geographic Information Science*, 40(2):61–77, 2013.
- [134] Y. Li. Towards concentration and decentralization: The evolution of urban spatial structure of Chinese cities, 2001–2016. *Computers, Environment and Urban Systems*, 80:101425, 2020.
- [135] Y. Li and X. Liu. How did urban polycentricity and dispersion affect economic productivity? A case study of 306 Chinese cities. *Landscape and Urban Planning*, 173:51–59, 2018.
- [136] H. Lin and B. Zhou. On the von Neumann entropy of a graph. *Discrete Applied Mathematics*, 247:448–455, 2018.

- [137] X. Liu and M. Wang. How polycentric is urban China and why? A case study of 318 cities. *Landscape and urban planning*, 151:10–20, 2016.
- [138] J. Lockhart, G. Minello, L. Rossi, S. Severini, and A. Torsello. Edge centrality via the Holevo quantity. In *Joint IAPR International Workshops on Statistical Techniques in Pattern Recognition (SPR) and Structural and Syntactic Pattern Recognition (SSPR)*, pages 143–152. Springer, 2016.
- [139] T. Louail and M. Barthelemy. A dominance tree approach to systems of cities. *Computers, Environment and Urban Systems*, 97:101856, 2022.
- [140] G. Lu, B. Li, and L. Wang. Some new properties for degree-based graph entropies. *Entropy*, 17(12):8217–8227, 2015.
- [141] G. Lu, B. Li, and L. Wang. New upper bound and lower bound for degree-based network entropy. *Symmetry*, 8(2):8, 2016.
- [142] A. Maddison. *A millennial perspective*. OECD, 2001.
- [143] M. Marchiori and V. Latora. Harmony in the small-world. *Physica A: Statistical Mechanics and its Applications*, 285(3-4):539–546, 2000.
- [144] V. Marin, C. Molinero, and E. Arcaute. Uncovering structural diversity in commuting networks: global and local entropy. *Scientific Reports*, 12(1):1–13, 2022.
- [145] J. Marti-Henneberg, M. Satchell, X. You, L. Shaw-Taylor, and E. Wrigley. 1851 England, Wales and Scotland Railway Stations [data collection]. 10.5255/ukda-sn-852994, 2021.
- [146] J. Marti-Henneberg, M. Satchell, X. You, L. Shaw-Taylor, and E. Wrigley. 1881 England, Wales and Scotland Rail Lines [data collection]. 10.5255/ukda-sn-852993, 2021.
- [147] R. Maunier. The definition of the city. *American journal of sociology*, 15(4):536–548, 1910.
- [148] M. Mazzamurro. In search of lost space: The changing affordances of physical and virtual teaching spaces during the pandemic. *Journal of PGR Pedagogic Practice*, 1:15–21, 2021.
- [149] M. Mazzamurro. Contrafacta networks first three generations, v1.01.0, zenodo. <https://doi.org/10.5281/zenodo.6363437>, 2022.

- [150] M. Mazzamurro and W. Guo. Network-entropy-based morphological polycentricity of systems of cities, 2020. [Under Review].
- [151] M. Mazzamurro and W. Guo. Assessing the impact of major historical events on urban landscapes via local entropy measures. In *2021 IEEE International Smart Cities Conference (ISC2)*, pages 1–7. IEEE, 2021.
- [152] M. Mazzamurro, Y. Wu, and W. Guo. Dynamic spatial cluster process model of geo-tagged tweets in London. In *2019 IEEE International Smart Cities Conference (ISC2)*, pages 678–684. IEEE, 2019.
- [153] C. McEvedy and R. Jones. World atlas of population history. *Hammondsworth: Penguin Books Ltd*, 1978.
- [154] M. Medo, G. Cimini, and S. Gualdi. Temporal effects in the growth of networks. *Physical review letters*, 107(23):238701, 2011.
- [155] Y. V. Medvedkov. The concept of entropy in settlement pattern analysis. In *Papers of the Regional Science Association*, volume 18, pages 165–168. Springer, 1967.
- [156] E. Meijers. Polycentric urban regions and the quest for synergy: is a network of cities more than the sum of the parts? *Urban studies*, 42(4):765–781, 2005.
- [157] E. Meijers. Measuring polycentricity and its promises. *European planning studies*, 16(9):1313–1323, 2008.
- [158] E. J. Meijers and M. J. Burger. Spatial structure and productivity in US metropolitan areas. *Environment and planning A*, 42(6):1383–1402, 2010.
- [159] S. Milonia and M. Mazzamurro. Temporal networks of ‘contrafacta’ in the first three troubadour generations. *Digital Scholarship in the Humanities*, 2022.
- [160] G. Minello, L. Rossi, and A. Torsello. On the von Neumann entropy of graphs. *Journal of Complex Networks*, 7(4):491–514, 2019.
- [161] G. Modelski. *World cities:-3000 to 2000*. Faros 2000, 2003.
- [162] A. Mowshowitz. Entropy and the complexity of graphs: I. An index of the relative complexity of a graph. *The bulletin of mathematical biophysics*, 30(1):175–204, 1968.

- [163] V. Nicosia, J. Tang, C. Mascolo, M. Musolesi, G. Russo, and V. Latora. Graph metrics for temporal networks. In *Temporal networks*, pages 15–40. Springer, 2013.
- [164] Oak Ridge National Laboratory. Landscan 2004 global population database. <https://landscan.ornl.gov/>, 2005.
- [165] A. Okabe and Y. Sadahiro. An illusion of spatial hierarchy: spatial hierarchy in a random configuration. *Environment and Planning A*, 28(9):1533–1552, 1996.
- [166] Openflights. Routes data. <https://openflights.org/data.html> (Accessed 22 August 2022).
- [167] F. Passerini and S. Severini. The von Neumann entropy of networks. *arXiv preprint arXiv:0812.2597*, 2008.
- [168] M. L. Pereira RHM, Nadalin V and A. PH. Urban centrality: a simple index. *Geographical analysis*, 45(1):77–89, 2013.
- [169] D. Piovani, E. Arcaute, G. Uchoa, A. Wilson, and M. Batty. Measuring accessibility using gravity and radiation models. *Royal Society open science*, 5(9):171668, 2018.
- [170] C. Pooley and J. Turnbull. *Migration and mobility in Britain since the eighteenth century*. Routledge, 2005.
- [171] A. Pred. *City-systems in advanced economies: past growth, present processes and future development options*. Routledge, 1977.
- [172] R. Procter, J. Crump, S. Karstedt, A. Voss, and M. Cantijoch. Reading the riots: what were the police doing on Twitter? *Policing and Society*, 23(4):413–436, 2013.
- [173] D. Pumain. Pour une théorie évolutive des villes. *L’Espace géographique*, pages 119–134, 1997.
- [174] D. Pumain. An evolutionary theory of urban systems. In *International and transnational perspectives on urban systems*, pages 3–18. Springer, 2018.
- [175] D. Pumain and F. Moriconi-Ebrard. City size distributions and metropolisation. *Geojournal*, 43(4):307–314, 1997.

- [176] B. Purvis, Y. Mao, and D. Robinson. Entropy and its application to urban systems. *Entropy*, 21(1):56, 2019.
- [177] E. Ravenstein. The laws of migration 1. *Journal of the Statistical Society of London*, 48:167–235, 1885.
- [178] E. Ravenstein. The laws of migration 2. *Journal of the Statistical Society of London*, 52:241–305, 1889.
- [179] M. Reba, F. Reitsma, and K. C. Seto. Spatializing 6,000 years of global urbanization from 3700 BC to AD 2000. *Scientific data*, 3(1):1–16, 2016.
- [180] M. Reba, F. Reitsma, and K. C. Seto. Historical urban population: 3700 bc - ad 2000. <https://doi.org/10.7927/H4ZG6QBX>, 2018. Accessed 02-May-2021.
- [181] W. E. Rees. Cities as dissipative structures: Global change and the vulnerability of urban civilization. In *Sustainability science*, pages 247–273. Springer, 2012.
- [182] C. S. Riolo, J. S. Koopman, and S. E. Chick. Methods and measures for the description of epidemiologic contact networks. *Journal of Urban Health*, 78(3):446–457, 2001.
- [183] R. J. Rivers and T. S. Evans. How do we avoid imposing the present on the past when modelling spatial interactions? *Documenta Praehistorica*, 47:462–475, 2020.
- [184] J.-P. Rodrigue. *The geography of transport systems*. Routledge, 2020.
- [185] I. Romanowska, T. Brughmans, A. Lichtenberger, and R. Raja. Urban networks seen through ceramics: formal modelling approaches to pottery distribution in Jerash. In *Urban network evolutions. Towards a high definition archaeology*, pages 131–137. Aarhus University Press Aarhus, 2018.
- [186] A. N. Rose, J. J. McKee, K. M. Sims, E. A. Bright, A. E. Reith, and M. L. Urban. Landscan 2019. Accessed 02-May-2021.
- [187] H. D. Rozenfeld, D. Rybski, J. S. Andrade, M. Batty, H. E. Stanley, and H. A. Makse. Laws of population growth. *Proceedings of the National Academy of Sciences*, 105(48):18702–18707, 2008.

- [188] H. D. Rozenfeld, D. Rybski, X. Gabaix, and H. A. Makse. The area and population of cities: New insights from a different perspective on cities. *American Economic Review*, 101(5):2205–25, 2011.
- [189] A. Sanchirico, G. Andrulli, and M. Fiorentino. City size distribution analyses based on the concept of entropy competition. *Geographical Analysis*, 2022.
- [190] S. Sassen. The global city. *New York*, 1991.
- [191] M. Satchell, E. Wrigley, L. Shaw-Taylor, X. You, and J. Henneberg. 1851 England, Wales and Scotland Rail Lines [data collection]. 10.5255/ukda-sn-852991, 2018.
- [192] T. Seymour. Urban polycentricity in northern England: economic catalyst or chimera. Technical report, DPU working paper, 2017.
- [193] C. E. Shannon. A mathematical theory of communication. *The Bell system technical journal*, 27(3):379–423, 1948.
- [194] G. Sierksma and H. Hoogeveen. Seven criteria for integer sequences being graphic. *Journal of Graph theory*, 15(2):223–231, 1991.
- [195] F. Simini, M. C. González, A. Maritan, and A.-L. Barabási. A universal model for mobility and migration patterns. *Nature*, 484(7392):96–100, 2012.
- [196] D. Simmons, J. Coon, and A. Datta. The quantum theil index: characterizing graph centralization using von Neumann entropy. *arXiv preprint arXiv:1707.07906*, 2017.
- [197] C. T. Smith. The movement of population in England and Wales in 1851 and 1861. *The Geographical Journal*, 117(2):200–210, 1951.
- [198] D. Stoyan, W. S. Kendall, S. N. Chiu, and J. Mecke. *Stochastic geometry and its applications*. John Wiley & Sons, 2013.
- [199] J. Tang, I. Leontiadis, S. Scellato, V. Nicosia, C. Mascolo, M. Musolesi, and V. Latora. Applications of temporal graph metrics to real-world networks. In *Temporal Networks*, pages 135–159. Springer, 2013.
- [200] D. Taylor, M. A. Porter, and P. J. Mucha. Supracentrality analysis of temporal networks with directed interlayer coupling. In *Temporal network theory*, pages 325–344. Springer, 2019.

- [201] A. Tsakalidis, M. Liakata, T. Damoulas, B. Jelinek, W. Guo, and A. Cristea. Combining heterogeneous user generated data to sense well-being. In *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics*, pages 3007–3018, Dec. 2016.
- [202] C. Tsirogiannis and C. Tsirogiannis. Uncovering the hidden routes: Algorithms for identifying paths and missing links in trade networks. *The Connected Past: Challenges to Network Studies in Archaeology and History*, pages 103–20, 2016.
- [203] UK: Office of National Statistics. Coastal towns in England and Wales: October 2020, 2020.
- [204] United Nations. Demographic Yearbook. <https://unstats.un.org/unsd/demographic/products/dyb/dyb2007/notestab06.pdf>, 2007. Accessed 13-Nov-2022.
- [205] United Nations, Department of Economic and Social Affairs. 2000 Demographic Yearbook, 2002.
- [206] United Nations, Department of Economic and Social Affairs, Population Division. World urbanization prospects: The 2018 revision, online edition. <https://population.un.org/wup/Download/>, 2018. Accessed 02-May-2021.
- [207] G. J. Upton. Distance and directional analyses of settlement patterns. *Economic Geography*, 62(2):167–179, 1986.
- [208] D. Vega and M. Magnani. Metrics for temporal text networks. In *Temporal network theory*, pages 147–160. Springer, 2019.
- [209] V. Verbavatz and M. Barthelemy. The growth equation of cities. *Nature*, 587(7834):397–401, 2020.
- [210] S. Wasserman, K. Faust, et al. *Social network analysis: Methods and applications*. Cambridge university press, 1994.
- [211] A. F. Weber. *The growth of cities in the nineteenth century: A study in statistics*, volume 29. Macmillan, 1899.
- [212] B. Wellman and C. Wetherell. Social network analysis of historical communities: Some questions from the present for the past. *The History of the Family*, 1(1):97–121, 1996.

- [213] J. Whitbeck, M. Dias de Amorim, V. Conan, and J. L. Guillaume. Temporal reachability graphs. In *Proceedings of the 18th annual international conference on Mobile computing and networking*, pages 377–388, 2012.
- [214] J. Wilkinson, T. Emms, and T. S. Evans. Dynamical analysis of spatial interaction models. *arXiv preprint arXiv:1911.10123*, 2019.
- [215] A. Wilson. *Entropy in Urban and Regional Modelling*. Pion Ltd, 1970.
- [216] A. Wilson. Boltzmann, Lotka and Volterra and spatial structural evolution: an integrated methodology for some dynamical systems. *Journal of The Royal Society Interface*, 5(25):865–871, 2008.
- [217] A. Wilson and J. Dearden. Tracking the evolution of the populations of a system of cities. In *Population Dynamics and Projection Methods*, pages 209–222. Springer, 2011.
- [218] B. B. Xuan, A. Ferreira, and A. Jarry. Computing shortest, fastest, and foremost journeys in dynamic networks. *International Journal of Foundations of Computer Science*, 14(02):267–285, 2003.
- [219] Q. Xue, X. Jin, Y. Cheng, X. Yang, and Y. Zhou. The dataset of walled cities and urban extent in late imperial China in the 15th–19th centuries. *Earth System Science Data*, 13(11):5071–5085, 2021.
- [220] I. Yamada and J.-C. Thill. Local indicators of network-constrained clusters in spatial point patterns. *Geographical analysis*, 39(3):268–292, 2007.
- [221] J. Yan. Topological structure of extremal graphs on the first degree-based graph entropies. *MATCH Commun. Math. Comput. Chem.*, 85(2):275–284, 2021.
- [222] B. Yang, W. Guo, B. Chen, G. Yang, and J. Zhang. Data from: Estimating mobile traffic demand using twitter, 2016.
- [223] B. Yang, W. Guo, B. Chen, G. Yang, and J. Zhang. Estimating mobile traffic demand using Twitter. *IEEE Wireless Communications Letters*, 5(4):380–383, Aug 2016.
- [224] H. Yang, M. Dijst, P. Witte, H. van Ginkel, and J. Wang. Comparing passenger flow and time schedule data to analyse high-speed railways and urban networks in China. *Urban Studies*, 56(6):1267–1287, 2019.

- [225] C. Ye, R. C. Wilson, C. H. Comin, L. d. F Costa, and E. R. Hancock. Entropy and heterogeneity measures for directed graphs. In *International Workshop on Similarity-Based Pattern Recognition*, pages 219–234. Springer, 2013.
- [226] D. Zhao and A. Strotmann. Analysis and visualization of citation networks. *Synthesis lectures on information concepts, retrieval, and services*, 7(1):1–207, 2015.
- [227] G. K. Zipf. *Human behaviour and the principle of least-effort*. Addison-Wesley Press, 1949.