Mathematical Modelling of Demographic and Migratory Dynamics

By

CHARLOTTE R. JAMES

Department of Engineering Mathematics
UNIVERSITY OF BRISTOL

A dissertation submitted to the University of Bristol in accordance with the requirements of the degree of DOCTOR OF PHILOSOPHY in the Faculty of Engineering.

JUNE 2018

Word count: 43630

Abstract

In the following we develop analytical and numerical methods to study the emergence of patterns in complex systems, with a particular focus on human mobility, migrations and population dynamics. The purpose is to provide a quantitative description of the urbanisation process by defining a general and flexible modelling framework able to reproduce the universal patterns of population distribution observed empirically in different countries, as well as to estimate the demographic evolution of cities.

In 2012 the world's population exceeded 7 billion, and since 2008 the number of individuals living in urban areas has surpassed that of rural areas. This is the result of an overall increase of life expectancy in many countries that has caused an unprecedented growth of the world's total population during recent decades, combined with a net migration flow from rural villages to urban agglomerations. While it is clear that the rate of natural increase and migration flows are the driving forces shaping the spatial distribution of population, a general consensus on the mechanisms that characterise the urbanisation process is still lacking. In order to address this problem we focus on three areas of research: the size and spatial distribution of cities, growth processes at the level of individual citizens and occupational migration. Our aims are threefold: to determine which model of migration best describes the empirical relationship between the number of urban agglomerations and the population of a region, testing Heaps' Law for cities; to propose a theoretical framework to explain the emergence of the power law distribution of city sizes, Zipf's Law, from a microscopic birth-death process without fine tuning; to measure the relevance of social connections in determining migration decisions, analysing relocations of scientists.

We initially present a deterministic model of population dynamics incorporating a logistic population growth with both gravity and intervening opportunities models of migration. In this framework, we analytically assess the spatial distribution of cities that each model of migration produces finding that if individuals relocate according to an intervening opportunities model, the number of cities in a region increases linearly with the region's population. To empirically assess this result we analyse two distinct data sets, presenting two fundamental relationships that are quantitatively supported by the empirical evidence: 1) the number of cities in a country is proportional to the country's total population, irrespective of the country's area, and 2) the average distance between cities scales as the inverse of the square root of the country's population density. Using these relationships, we verify that a null model of urbanisation, where cities whose populations follow Zipf's law are randomly distributed in space, produces correct estimates of the expected number of cities in regions of various sizes worldwide. This result supports the hypothesis that spatial correlations in the distribution of cities are absent and suggests that, within the deterministic framework considered, an intervening opportunities model may be most appropriate for describing migratory dynamics during the urbanisation process.

We next address how the populations of non-interacting cities grow. With regards to city

growth, a general model has emerged that accounts for how the population of cities as a whole changes in time; Gibrat's law, or proportionate random growth. Despite its wide acceptance, a general consensus on how the underlying stochastic processes within a city, namely births and deaths, interact to give rise to this growth mechanism is lacking. To address this, we show that proportionate random growth can emerge from a general class of birth-death processes characterised by two mechanisms: correlations (i.e. dependence between individuals) and environmental variability. Our microscopic processes also demonstrate the relationship between Zipf's law for the distribution of city sizes and Taylor's law for the scaling of the fluctuations in population increments. The model may be applied to other systems with an explicit time dependence where the distribution of group sizes follows Zipf's law.

Migrations play a primary role in determining a city's demographic and economic growth. Consequentially, understanding the factors behind individual relocation decisions can improve predictive models of population projection. Here we focus on occupational migration, specifically the migration of scientists. In the academic community there is a widely accepted belief that movement between institutions is beneficial to, possibly even essential for, a successful career and many individuals relocate at some point in their career. Despite its common occurrence, it remains unclear how a scientist looking to relocate selects their next institution and at which point in time they decide to make this move. Using a comprehensive dataset on scientific publications, we reconstruct career trajectories of scientists in order to determine the driving forces behind the decision to change institutions. We apply methods originating from machine learning, including decision tree classifiers and logistic regression, in order to determine which factors are most influential in a scientist's decision to relocate. Using this insight, we introduce the quality-social-gravity model; a modified version of a traditional gravity model to estimate mobility flows. We demonstrate that the quality-social-gravity model places the true destination of a scientist in the 10 most probable destinations for over 23% of cases compared to 7% of the traditional gravity model, hence improving the model's predictive power. The insight gained from this work provides us with a deeper understanding of the factors that influence the migration decisions of scientists alongside a general modelling approach to describe migration dynamics.

Human migration and demographic growth are examples of complex phenomena showcasing the typical features of complex systems, namely the presence of heterogeneous distributions, long-range interactions, complicated individual (microscopic) dynamics and emergent collective (macroscopic) behaviour. The modelling techniques developed to describe the emergence of these general patterns in the context of urban dynamics and human migration can also find application in other complex systems, such as those in ecology and biology, where these patterns are also present.

Dedication and acknowledgements

Firstly I owe a huge thank you to my supervisor Filippo Simini for his all of his support, guidance and patience and for everything he has taught me. Without his expertise this Ph.D. would not have been possible. Thank you also to my collaborators Luca Pappalardo and Alina Sirbu for making me feel welcome during my time in Pisa and teaching me the basics of machine learning, and to Sandro Azaele for the time and effort he put in to the work presented in Chapter 3. Thank you to the Engineering Maths Department for the education and opportunities I have received and to EPSRC for the funding.

During the last few years I have made many friends who have helped make the Ph.D experience an enjoyable one. While there are too many names to mention, a big thank you goes to everyone in the Buncaer, old and new, for creating a social and supportive environment. A special thank you goes to Oscar for his advice, ideas and encouragement and for keeping me positive over the last few months.

Finally, thank you to my family for always being there and believing in me; this thesis is dedicated to you.

AUTHOR'S DECLARATION

I declare that the work in this dissertation was carried out in accordance with the requirements of the University's Regulations and Code of Practice for Research Degree Programmes and that it has not been submitted for any other academic award. Except where indicated by specific reference in the text, the work is the candidate's own work. Work done in collaboration with, or with the assistance of, others, is indicated as such. Any views expressed in the dissertation are those of the author.

SIGNED: ..... DATE: .....

TABLE OF CONTENTS

List of Tables

List of Figures

Figure **Page**

CHAPTER 1

INTRODUCTION

Increasing urbanisation rates, generally defined as the increase of the proportion of people living in urban areas or the proportion of buildings belonging to urban agglomerations [91], is a trend that has happened in waves throughout human history, with a dramatic acceleration in the last 300 years [113]. In 2015, 56% of China's population lived in cities, a figure that has more than doubled compared to the 26% of 1990. The Ministry of Housing and Urban-Rural Development estimates that by 2025 300M Chinese now living in rural areas will move into cities. State spending is planned on new houses, roads, hospitals, schools, which could cost up to 600 billion USD a year. A great rate of urbanisation is also expected in Sub-Saharan African countries. As a result, by 2030 it is estimated that the world's population will have increased by over 1 billion people, most of whom will dwell in the rapidly growing cities of Asia and Africa [86]. Such large migration events may have both positive and negative effects. Positive effects can include access to better health care and education systems, and more job opportunities. Negative effects include depletion of resources, traffic, pollution, crime, rising inequalities and segregation in cities. As a consequence, a quantitative understanding of the mechanisms that drive urbanisation is important to help governments and decision makers plan their investments in order to achieve sustainable urban development and growth. These decisions will have a huge impact on the lives of millions of people, the economy and the environment.

In the following we consider two mechanisms that drive urbanisation: the migration of individuals from rural areas to towns or cities and overall population growth, which increases the proportion of individuals living in cities as opposed to rural areas. Specifically we focus on three main aspects. In Chapter 2 we investigate how population growth and migration interact to give rise to the size and spatial distributions of cities seen globally. In particular we empirically validate Heaps' law for cities and determine whether deterministic models of migration flows are able to reproduce it. In Chapter 3 we introduce a stochastic model of population growth at

the level of individual inhabitants of cities, ie. births and deaths. We assess the model's ability to reproduce the empirical size distribution of cities, establishing a connection between Taylor's law and Zipf's law. Finally, in Chapter 4 we consider the factors that drive individuals to migrate between cities, focussing on relocations due to job transitions and specifically the movement of scientists. We introduce an adapted gravity model and demonstrate that it is more accurate in predicting the next location of a scientist compared to a traditional gravity model.

1.1 Cities as Complex Systems

Complex systems have properties that arise from the interactions and relationships of many underlying components. Such properties are considered emergent and include self organisation, heterogeneity, hierarchal structure, long-range interactions and decentralised control, to name a few [24]. Examples of systems that exhibit complex behaviour can be found in many disciplines including (but not limited to) Physics, Economics, Computer Science and Biology. As such, numerical and analytical frameworks developed for the study of complex systems may be applied across many varied domains.

Cities themselves are considered complex adaptive systems; complex systems in which the individual and collective behaviour of the underlying components displays self-organisation and adapts to changes in the environment. The study of cities as complex systems has revealed many interesting properties. Stemming from the fact that most cities develop gradually over time, cities have been shown to have a hierarchal structure [18]. At each point in development the structure of the city is organised in such a way that there is a high level of connectivity between different parts, for example roads connecting one side of the city to another. However, when viewed at a later time, the city appears highly complex, with underlying patterns at both the macro- and micro- scale such as power law distributions for the length of street segments [89]. It is this hierarchal development that leads to the hierarchal structure of cities and also their self-similarity; cities appear the same when viewed at different scales. As a result of this self-similarity, research has looked into the fractality of cities [19, 21]; fractals, by definition, are self-similar objects. The fractal dimension of an object is the ratio of the change in detail to the change in scale. If fractal dimension is fixed and independent of the scale at which an object is viewed, that object is generated by a single set of processes that operate at every scale. With regards to cities, the evolution of the perimeter of a city over a period of urbanisation has been shown to have a fractal dimension that increases with scale and declines over time [22]. Thus the growth of a city is not controlled by a single set of processes and urbanisation can cause the fractality of a city's perimeter to decrease.

Self-similarity in urban systems suggests that the processes driving urbanisation and city formation are similar across all cities. There is evidence that many urban quantities scale with the size, or population, of cities. These quantities include urban supply networks [67], average wage,

number of patents and the total amount of housing [25]. Explanations for these scalings have suggested that they are emergent features of cities caused by the underlying interactions of the individuals residing in them [23, 111]. The scaling exponents, η, for different urban quantities fall into three different categories: sub-linear ($\eta < 1$); linear ($\eta \sim 1$); super-linear ($\eta > 1$). Each category can be associated with similar urban indicators. An exponent $\eta \sim 1$ is associated with quantities relating to individual needs such as total housing and total employment. A sub-linear exponent describes quantities termed 'economies of scale'; quantities associated with the infrastructure of the city such as petrol sales and total length of roads. Finally, a super-linear exponent, $\eta > 1$ is linked to quantities attributed to the social nature of a city, such as innovation, wealth, crimes and patents [25]. When studying cities as complex systems, particularly when using data to investigate properties such as the scaling of urban quantities, the definition of a city becomes important. The scaling laws discussed so far were obtained using extensive datasets from sources such as the US Census Bureau and Eurostat Urban Audit, among others. City data obtained from sources such as these are linked to definitions of urban areas that are not necessarily consistent with one another; there is no global consensus on how cities should be defined [128]. In order to overcome the inconsistencies in administrative definitions of cities, which are based on historical and political considerations, clustering algorithms [107] have been used [4, 32, 33]. This allows for the study of complex features of cities, such as the scaling of urban quantities, using a consistent definition of a city based solely on urban morphology. Repeating the scaling analysis using a consistent definition of cites has demonstrated that, apart from some small deviations, almost all urban quantities scale linearly with city population [4, 32, 33]. However, there is also variation in the scaling exponent when different definitions of cities are considered [33]. A metropolitan area can be defined as a large central city combined with smaller satellite settlements from which people commute to the central district. If a clustering algorithm is combined with commuting data to define metropolitan areas the scaling of certain quantities can change from the sub- to super-linear regime, and vice versa [4]. Alongside this, the scaling exponents for certain quantities are highly dependent on the size of the cities being considered: a city with a population of $10,000$ differs from on with $100,000$ inhabitants. In [4] it is demonstrated that the scaling of the number of patents with city population varies with the minimum city population: for a minimum population of $10,000$ the scaling is super-linear whereas for a minimum population of $50,000$ it is sub-linear. It has been suggested that this result, namely that the scaling of urban quantities depends on the definition of a city, provides an opportunity for understanding the relation between the features of cities and their size at the micro-, meso- and macro-scales [33]. However, the sensitivity of the scaling exponents to the definition of a city suggests that a theory of cities cannot rest on scaling alone [4].

A further issue with scaling laws arises from the approach most often used to fit the parameters to the data [69]. The most common method for fitting exponent η is to use least-squares on the log-transform of the data, ie. $\ln y = \eta \ln x + c$, where y represents the quantity of interest, such

as average wage or number of patents, x represents the population of the city and c is a constant corresponding to an intercept in log-log space. While this approach is simple to implement, it contains a number of assumptions and limitations that are often ignored. For example, assumptions about the fluctuations of y given x are rarely verified in real data, and if the variable of interest y can have a value of 0, which is the case for patents, due to the log-transform these datapoints must be removed prior to fitting.

In the following we focus on the urbanisation process. With regards to city formation and growth, some of the earliest research stems from human geography and urban economics [29, 47, 66]. Christaller's Theory of Central Places suggests that cities form as an urban hierarchy: settlements are regularly spaced on a hexagonal lattice with larger cities being spaced further apart and surrounded by smaller satellite cities [29]. The 'Edge City Model' of urban economics [66] describes how self-organisation of businesses into compact areas within a city can evolve from a perturbation to an initially homogeneous business distribution due to economic competition. From a complex systems perspective, a city is formed by the interactions of many underlying components. A simple yet effective way of understanding how local interactions of these components can produce a global order, or city, is through cellular automata [19]. Cellular automata allow for mathematical analysis due to their analytical simplicity while producing properties, such as self organisation, that are compatible with cities.

For the purpose of our work, we wish to focus on how the complicated individual (microscopic) dynamics within a city give rise to the emergent collective (macroscopic) behaviour. Here, the underlying components are the individual inhabitants of the cities and the urban systems, or systems of cities, in which they reside. We consider population growth and human migration: processes of individual and collective behaviour associated with cities. These processes contribute to emergent phenomena such as the size and spatial distributions of cities and their scaling with the population of urban systems.

1.2 Zipf's Law and the growth of cities

A requirement of any general model of population growth in cities it that is must be able to reproduce the empirical patterns observed in data. One of the most documented empirical findings in human geography is Zipf's law for the distribution of city sizes. Formally, a random variable follows Zipf's law if its probability density function is a power law with exponent -2. For the case of cities, Zipf's law states that the probability to find a city with a given population, x, is inversely proportional to the square of the population:

$$P(x) \sim x^{-\beta}, \tag{1.1}$$

where $\beta \simeq 2$ and $P(x)$ is a probability density function (PDF). In other words and equivalently, the size (population) x of the r^{th} largest city (ie. a city with rank r) is given by

$$x = c\, r^{-\frac{1}{\beta-1}}, \tag{1.2}$$

where c is the population of the largest city [81]. Power laws such as Zipf's law provide a good description of many empirical distributions such as word frequency, earthquake magnitudes, personal income, scientific citations and astronomical masses to name a few [6, 17, 74, 95, 146]. Zipf's law itself has been found in a large number of diverse areas, including systems where observations correspond to groups of individuals and the variable of interest is the group size, such as the number of employees in firms [6] or the distribution of family names [95]. With regards to cities, the exponent $\beta \simeq 2$ has been shown to apply to the distribution of city sizes both globally and historically with surprisingly small deviations [49]. Zipf's law also applies to the population distribution of larger regions, for example countries in Europe [53] as well as counties in the United States (Chapter 3).

While there is a vast amount of evidence validating Zipf's law for cities, it is important to note that there are situations where the distribution of city sizes can not be fully described by a power law. For example, countries with "primate cities" by definition do not have a city size distribution that follows Zipf's law: primate cities are at least twice as large as the next largest city [61] and therefore Equation 1.2 does not hold. Examples of countries with primate cities include France, where the population of Paris is over 6 times larger than Lyon, and the United Kingdom, where London's population is more than 10 times greater than that of Birmingham. In the economics literature, the existence of primate cities has been attributed to, for example, imbalances in development and economic influence [61]. From a complex systems perspective, outliers, such as primate cities, have been found to co-exist with power laws in a large variety of systems. In [122] these outliers are coined "dragon-kings" which emerge due to mechanisms of self-organisation and cities such as Paris and London are extreme events that are not expected if the distribution of all other cities follows a power law. In [37] it is demonstrated that primate cities are also present in Simon's model (section 1.2.2), an elementary rich-get-richer model, and are caused by a dominant first-mover advantage.

Even for regions without primate cities, there exists debate around whether a power law is truly the best fit for the distribution of city sizes: some have claimed that the lognormal distribution is a better fit than Zipf's law [38]. The argument that cities are log-normally distributed stems from the fact that the lognormal distribution arises asymptotically from proportionate growth processes (section 1.2.1) and empirical evidence has repeatedly shown that the growth of cities follows this process: larger cities grow at the same rate as smaller cities. In log-log space the PDF of a lognormal distribution is a parabola whereas for Zipf's law it is a straight line. As the parabola of the lognormal distribution has two free parameters, as opposed to the one free parameter of Zipf's law, given that the data is often noisy the lognormal may appear to be a better fit: it has an extra degree of freedom. Alongside this if a parabola has a small curvature, locally it can appear to be a straight line. However, there is no clear evidence to reject Zipf's law (see section 3.3) and Zipf's law provides a simpler explanation for the distribution of city sizes. For this reason we assume that the stationary distribution of city sizes, above a minimum population,

follows Zipf's law with exponent $\beta \sim 2$.

The existence of a global distribution of city sizes places a constraint on models of city growth. A number of general mechanisms exist to account for the emergence of Zipf's law in various systems [10, 95, 140]. In [10] it is suggested that when many diverse systems share a common characteristic, such as a power-law stationary distribution, it must be due to a global feature that is shared between all of the systems rather than system-specific details. Here this global feature is presented as the division into groups; people are divided into cities, employees are divided between firms, people are divided into families based on their surname. They demonstrate that information theory can be used to find the group-size distribution as follows. Consider a system that consists of a total of M individuals and N cities. The number of cities with population x is given by $N(x)$: a city of size x has x inhabitants. If no further information is known about the system of cities or the individuals, a best guess would be that each person has equal probability to occupy one of the x spaces in one of the N cities: the chance of finding a specific person at a specific location is $1/M$. In this framework, where there is an equal probability of finding each of the M individuals in any of the N cities, the stationary distribution of city sizes is given by $P(x) \sim e^{-ax}x^{-1}$. If the same framework is adapted such that there exists additional knowledge relating to the location preferences of individuals, there is no longer an equal probability of finding each of the M individuals in any of the N cities. The consequence of this knowledge is that there is an additional constraint on the model which results in $P(x) \sim x^{-\beta}$ where β depends on how much additional knowledge there is. In the case of $\beta = 2$, Zipf's law for cities is obtained.

A series of recent work [2, 112] has demonstrated that Zipf's law arises naturally when there are underlying, unobserved variables; latent variables. Systems with latent variables that control observations can lead to Zipf's law without fine tuning by mixing together narrow distributions with very different means. However, this work focuses on relating Zipf's law to latent variables in static systems, without an explicit time dependence.

Several models have been proposed to explain how Zipf's law can emerge as the stationary distribution of dynamical processes for the sizes of groups of individuals. These models can be divided into two classes based on the scale considered: mesoscopic models at the scale of the groups (e.g., cities) and microscopic models at the scale of the individuals (e.g., city inhabitants).

1.2.1 Mesoscopic City Growth: Gibrat's Law

Mesoscopic models are stochastic processes describing the evolution of a group's population as a whole, usually in terms of growth rates. Examples of mesoscopic models are random multiplicative processes (RMP) [104], also called Gibrat's law [52] or proportionate random growth in economics literature [48, 121], and those based on the interplay between intermittency and diffusion [83].

Proportionate random growth In a RMP the population x of a city varies stochastically in time. The change in size in a time step dt is determined by the growth rate at time t, g_t, and the

new size at time $t+dt$ is given by

$$x_{t+1} = (1+g_t)x_t, \tag{1.3}$$

where the growth rate, g_t, is a random variable. From Equation 1.3 it is clear that growth rates are independent of city size; x_{t+1}/x_t is independent of x_t. The homogeneity in growth rates is known as Gibrat's Law. In order for this to apply, growth rates g_t must be identically distributed and independent of city size. Using the diffusive approximation for large populations, $x_t \gg 1$, Equation 1.3 may be written in continuous time as a stochastic differential equation (SDE):

$$dx_t = \bar{g}\, x_t dt + \sigma_g x_t\, dW_t. \tag{1.4}$$

Here $\bar{g}$ and σ_g are the mean and standard deviation of the growth rate respectively, dW_t is a Brownian motion and $dx_t = x_{t+dt} - x_t$. If the growth rates are extracted from a distribution with small and negative mean, $\bar{g}$, and finite variance, σ_g^2, introducing a reflecting boundary at small populations to avoid extinction, ie. by imposing $x_t \geq x_0$, it can be shown that a stationary distribution of city sizes exists. Asymptotically (ie. at $t \gg 1$) the distribution follows Zipf's Law with exponent $\beta = 2 - 2\bar{g}/\sigma_g^2 \approx 2$ if the growth rate's mean is much smaller than the variance [48]. An alternative way of stating this result, from Equation 1.4, is that in a random multiplicative process both the mean population growth and its standard deviation are proportional to the city size:

$$\langle dx_t | x_t = x \rangle \propto x; \tag{1.5}$$

$$\langle dx_t^2 | x_t = x \rangle - \langle dx_t | x_t = x \rangle^2 \propto x^2. \tag{1.6}$$

These relationships define proportionate random growth: the mean change in city size (population) and the standard deviation of the population change in time dt is proportional to the population of the city at time t.

Intermittency Another mesoscopic model of urban development and growth, introduced in [141] and analysed in [83], which is an alternative to Gibrat's law, is based on the interplay between intermittency and diffusion. Here, a reaction-diffusion process is proposed as a mechanism for cities to form from an initially homogeneous distribution of individuals. Denoting the population at location i at time t as $x(i,t)$ with $x(i,0) = 1$ for all i, the population evolution in a time step is split into two sub-processes. The first sub-process corresponds to the reaction, based upon the Zeldovich model for intermittency [142]. At this stage, the population $x(i,t)$ changes according to the following:

$$x(i,t') = \begin{cases} (1-q)p^{-1}x(i,t), & \text{with probability } p, \\ q(1-p)^{-1}x(i,t), & \text{with probability } 1-p. \end{cases} \tag{1.7}$$

where t' is an intermediate time between t and $t+1$, p is a probability ($0 < p < 1$) and q is a parameter ($0 \leq q \leq 1$). This reaction conserves the average population of the cities and hence has a well defined mean. However, the variance and higher moments of the population diverge and consequentially strong inhomogeneities in the distribution of population, ie cities, emerge. The second sub-process corresponds to diffusion; $x(i,t+1) = (1-\alpha)x(i,t')$. Here, α represents the fraction of the population at location i that moves to a different location. In [141], this diffusing fraction of city dwellers is re-distributed uniformly between neighbouring locations of i. In analogy with Gibrat's law, in the first sub-process, cities are growing according to a RMP and in the second sub-process diffusion acts as a mechanism for implementing a soft lower boundary condition. In the framework described, the stationary distribution of city sizes, $P(x)$ can be described as $P(x) \sim x^{-\beta}$ where β is a function of the model's parameters and is given by: $\beta = 1 + \ln p / \ln[p/(1-\alpha)]$. Thus, for $\alpha \ll 1$, the stationary distribution of city sizes can be described by Zipf's law.

1.2.2 Microscopic City Growth

Microscopic models provide a more fundamental description of population dynamics. Focussing on the growth of cities, microscopic models are stochastic processes describing the events experienced by an individual, namely births, deaths and migrations, that ultimately determine the change in the size of a population. There are many examples of microscopic models that attempt to provide a general description of how power law stationary distributions can arise from underlying processes.

Yule's and Simon's model is based on the rich-get-richer mechanism [117, 139]. In the context of cities, Simon's model suggests that a system of cities can evolve in two ways; an individual is born within an existing city, increasing that city's population by 1, or an individual is born and relocates to an uninhabited area creating a new city with population 1. If the probability in time dt that an individual is born to a city inhabitant and stays in the city is s and the probability that in time dt an individual is born and relocates to a new (uninhabited) city is g then the distribution of city sizes at any time is given by $P(x) = \frac{g}{s} x^{-\beta}$, with $\beta = 1 + g/s$. Thus, providing that the rates s and g are approximately equal the distribution of city sizes will follow Zipf's law. It is important to note the distinction between s and g: s is the probability that *any* individual within a city reproduces in time dt, g is the probability that *one* individual is born and relocates from a single city. This in fact corresponds to proportionate random growth as the expected amount an existing city of size n will grow in time dt is $s \cdot x \cdot dt$. In contrast, the expected number of new cities to be created in time dt, assuming i cities already exist, is $g \cdot i \cdot dt$.

A similar microscopic model to that of Yule and Simon is introduced in [84]. The difference here is that rather than an existing city growing at a rate s, transition rates for the growth, $w_g(x)$, and decrease, $w_d(x)$ of a city of size x are used. Alternatively, in a time dt each city loses a citizen with probability $w_d(x)dt$ or gains a citizen with probability $w_g(x)dt$. Akin to Yule and Simon,

there is also a probability pdt that in time dt a new city is created with a population of 1. In this case the stationary distribution of city sizes depends on the transition rates. In particular, if $w_g(x)$ and $w_d(x)$ are linear functions of x, which corresponds to the case of there being no interactions among individuals, the stationary distribution of city sizes does not follow Zipf's law. However, if a pairwise interaction among individuals is considered, corresponding to transition rates $\propto x^2$, then the stationary distribution of city sizes can be described by Zipf's law.

Other microscopic models have focussed on cluster growth and aggregation to account for features of cities such as the fractality of their perimeter [108] and the area distribution of an urban system [82]. In [108], a simple lattice-based model, based on the gravity model, is used to populate sites on an empty lattice. The central site is initially populated and at each successive step further sites are populated with a probability proportional to the distance from already occupied sites. Performing cluster analysis on the results it is found that the size distribution of cluster areas follows Zipf's law apart from the largest cluster. Alongside this, growth rates of clusters are independent of cluster area; proportionate random growth is an emergent feature of this simple model. A further class of microscopic models focuses on how migration shapes the size distribution of cities. In [72] it is suggested that city growth is driven by migration; individuals have a preference to move to large urban areas. In this model, the power law exponent $\beta = 2$ is accounted for if city population is growing at a faster rate than the overall population, a result confirmed by historical data on the populations of US cities.

Frasco et al [45] model city growth using a social network which incorporates the notion that individuals are more likely to move to areas where they are well socially connected. In this model, N individuals are placed into a square area with side length L. The first individual (node) is placed randomly. Each remaining individual is successively placed according to a network growth and redirection model. Each time a new individual y is added to the network, one of the existing individuals, z, is chosen at random and y is connected to z with probability $(1-r)$; otherwise, with probability r, y is connected to a neighbour of z, z', selected at random. The spatial position of a new individual depends on how they join the network. If they connect directly to z, they are placed at a point (r,θ) from z where θ is selected randomly and uniformly from the range $[0,2\pi)$ and r is drawn from a power law distribution. If an individual instead connects by redirection to a neighbour of z, z', they are placed at (b,θ) where θ is again selected randomly and b represents the closest an individual can be placed to the one it connects to. A clustering can be performed on the spatial distribution of individuals in order to obtain cities. The distribution of city sizes is given by $P(x) \sim x^{-\beta}$, where $\beta = 1+(-\frac{r}{2}-\frac{1}{2}\sqrt{r(2-r)})^{-1}$. As such, as r approaches 1, the distribution of city sizes can be described by Zipf's law.

An important point to note about the microscopic models discussed is that they assume population can only grow. In Yule and Simon's model, the lattice based model in [108] and Frasco et al's social network [45], in each step or time interval existing cities get bigger, new cities are created, or both. In the variation of Yule and Simon's model presented in [84], while existing

cities can decrease in size, because the changes in city size are caused by migration rather than births and deaths, the population of the system of cities cannot decrease. Furthermore, here Zipf's law with exponent $\beta = -2$ can only be obtained when the transition rates scale as x^2: this implies that the average population change in dt, $\langle dx \rangle$ also scales as x^2 which does not agree with empirical evidence. Any model of urbanisation that aims to explain the distribution of city sizes but does not allow for a decline in population is incomplete; empirically urban populations must both increase and decrease.

1.2.3 Comparison

Mesoscopic models, such as proportionate random growth, are able to explain the emergence of Zipf's law without the need to fine-tune their parameters to specific values [48]. However, while there exists a satisfactory explanation of the presence of Zipf's Law in the distribution of large city sizes, there is no consensus on a model that describes the underlying stochastic processes at the level of the individuals, namely births, deaths and migrations: mesoscopic models are coarse-grained descriptions of population dynamics and lack an explicit link to the underlying microscopic processes. Microscopic models, in contrast, are able to produce the power law exponent $\beta = 2$ for the distribution of city sizes, however they are only able to do this for specific values of their parameters. Whilst these models link the distribution of city sizes to the underlying individual stochastic dynamics, they require fine-tuning.

It has been noted that when Zipf's Law fits the stationary distribution of a system, it is often only present in the tail of the distribution [95]. In particular, Zipf's Law for cities only describes the size distribution of large cities: it does not fit the distribution as a whole. In order to account for the full distribution of city sizes a microscopic understanding of population dynamics, describing how cities form and grow without the need for fine tuning, is required.

The rule of proportionate random growth is also unable to answer fundamental questions about the urbanisation process [19, 20, 23, 25, 46, 77, 80]. Urbanisation can happen in two ways: diffusion (or sprawl) and aggregation. Diffusion corresponds to existing cities growing and increasing in size because of either net migration from rural areas or a greater rate of natural increase in urban areas. Aggregation corresponds to new villages and towns being created in rural areas that were previously considered non-urbanised. In order to properly characterise urbanisation patterns we should consider both aspects: the distribution of city sizes, describing the size and growth of existing cities, and the overall number of cities, describing the abundance and formation of new urban areas. While proportionate random growth can account for the distribution of city sizes, it provides no explanation for the number of cities in a region or their spatial distribution.

1.3 Human Mobility

Up to now, the literature discussed has focussed on the growth of cities and the evolution of their size distribution. These models, with the exception of [84], are mostly concerned with how the population of cites changes due to births and deaths, or rather natural increase. While natural increase also accounts for external (or international) migration, a key component that is missing is an explicit dependence of the models on the movement of individuals between existing cities. Human mobility models can be used to estimate and predict migration flows between cities. For this reason they are an important tool when considering how the size and spatial distribution of cities evolves during urbanisation. Alongside this, human mobility models can be used to characterise specific types of migration, providing a deeper understanding of why individuals migrate between different locations.

In recent years, the introduction of widely available, high resolution geotagged data sources has resulted a rapid increase in the number of papers published using the key phrases 'human migration' and 'human mobility' [14]. Human mobility models can be broken down into two general classes reproducing either individual mobility patterns or population flows. Individual mobility patterns have a high degree of stochasticity due to the uncertainty associated with free will and the actions of an individual. Consequentially, the most basic models of individual mobility are based on stochastic processes, the most simple of which is a random walk. Population level models aim to estimate the average number of people moving between two locations per unit time. These mobility flows often depend on variables relevant to the locations such as distance between them and their populations.

1.3.1 Individual Level

As aforementioned, when modelling the movement of individual people the most basic approach is to assume that movement is random and can be described by a random walk; a path formed by successive discrete random steps. In the simplest version, spatial displacements Δx_i are taken at regular times t_i. Considering a 1D example, if the initial position of an individual at time $t = 0$ is x_0, then after N steps the individual will be located at $x(t_N)$:

$$x(t_N) = x_0 + \sum_{i=1}^{N} \Delta x_i. \tag{1.8}$$

Here, each displacement Δx_i is an independent random variable taken from a probability distribution $f(\Delta x)$, called the jump length distribution, which can be used to determine the probability density function (PDF), $P(x,t)$ for the individual at be at position x at time t.

Of particular interest when considering individual mobility patterns is how the square root of the mean squared displacement (RMSD), $R(t)$ scales with time. The mean square displacement corresponds to the second moment of $P(x,t)$,

$$MSD(t) = \langle x(t)^2 \rangle = \int_{-\infty}^{\infty} x^2 P(x,t) dx, \tag{1.9}$$

and therefore $R(t)$ is given by $R(t) = \langle x(t)^2 \rangle^{1/2}$, characterising the growth of an individual's displacement from their origin with time. The scaling of $R(t)$ with time can be used to categorise the motion of an individual and is therefore important when it comes to finding the correct model of individual mobility to describe data. If $f(\Delta x)$ has a finite variance the limit of the random walk is Brownian motion. For Brownian motion, $R(t) \sim t^{1/2}$: the distance of an individual from their origin is proportional to the square root of the time elapsed. As Brownian motion corresponds to diffusion, random walks with displacement growing at a slower rate than $t^{1/2}$ are said to be sub-diffusive whereas if the displacement grows at a rate faster than $t^{1/2}$ the random walk is super-diffusive.

From Equation 1.9 it is clear that the behaviour of $R(t)$ with time is determined solely by $P(x,t)$ and hence $f(\Delta x)$: the probability distribution that the random walker's steps are extracted from. An example of this is the Levy Flight: a random walk in which an individual regularly moves short distances but occasionally moves further. In this case, $f(\Delta x)$ is long tailed to account for the small probability of Δx being large:

$$f(\Delta x) \sim \frac{1}{\Delta x^{1+\beta}} \tag{1.10}$$

with $0 < \beta < 2$. For a Levy flight, the root mean square displacement scales super-diffusively: $R(t) \sim t^{1/\beta}$.

While random walks provide a good starting point for the modelling of individual mobility, in order to accurately reproduce individual patterns of movement more complex extensions of the simple random walk are required. One extension is the continuous time random walk (CTRW) where displacements do not occur at regular time steps but rather happen at intervals extracted from a second probability distribution $\phi(\Delta t)$ known as the wait-time distribution. Both the jump-length distribution $f(\Delta x)$ and the wait-time distribution $\phi(\Delta t)$ are independent. From a CTRW four types of random walks are obtained, depending on whether none, either of both distributions have heavy tails.

For the cases where $\phi(\Delta t)$ has a finite variance, the CTRW corresponds to either Brownian motion or a Levy Flight depending on the form of $f(\Delta x)$, as discussed above. Conversely, if the distribution of jump-lengths has a finite variance but the wait time distribution is described by a power law, $\phi(\Delta t) \sim \Delta t^{-(1+\alpha)}$, $(0 < \alpha < 2)$, the random walk is described as fractional Brownian motion. In this case, the behaviour of $R(t)$ depends on the exponent α; if α is less than 1, the process is sub-diffusive whereas if it is greater than 1 it is super-diffusive. Finally, if both the distribution of jump lengths and the distribution of wait times are described by power-law distributions with exponents β and α respectively then $R(t) \sim t^{\alpha/\beta}$; the nature of the diffusive behaviour is fully specified by α and β. Analysis of various data sources (GPS, CDRs, Dollar Bills) has found that both jump length distributions and the distribution of wait times display power-law behaviour [26, 55, 119, 144]. The parameter ranges measured from these data sources for exponents α and β are given by $0.42 \leq \alpha \leq 0.8$ [119, 144] and $0.31 \leq \beta \leq 0.75$ [55, 144].

One aspect of human behaviour that is missing from the above models is our tendency to return to previously visited locations. To account for this, Song et al [119] introduced two extensions to the CTRW model: exploration and preferential return. Here, exploration is the probability for an individual move to a previously unvisited location. In contrast, preferential return is the probability to return to a location that the individual has explored before. In [119] it is shown that on incorporating these two features into a CTRW and fitting the new parameters to data, results correspond to an individual's motion being dominated by their most visited location which is more in line with expected human behaviour. It also captures the ultra-slow diffusion of $R(t)$ and other statistics of human mobility such as the visitation frequency and radius of gyration. Further adaptations of random walks have included factors such as recency [15], where individuals return to previously visited places based on their total number of visits and the time since their last visit, and social information [132] to account for the fact that individuals who interact socially don't always move independently.

1.3.2 Population Level

Information about the movement of individuals can be aggregated to describe the flow of people moving between locations. These flows are often organised in an origin destination matrix which contains information on the flow of people between all possible origin and destination pairs. As it is not possible to disaggregate flows, in order to model how the flow of individuals between an origin and destination might change in the future the dependences between mobility flows and attributes of each location need to be determined. Spatial interaction models predict flows of individuals based on a small number of key local attributes of the origin and destination. In essence, these models are designed to estimate the number of trips, T_{ij} between locations i and j from the spatial distribution and socio-economic characteristics of the populations of i and j. In the most basic form, a model aims to infer flows using the product of variables specific to each location, such as the population of i and j, and a variable that relates the two locations, such as the distance between i and j.

Two main spatial interaction models for human population dynamics have emerged. These differ in the choice of variables considered and the functional form in which these variables are used. Gravity models, akin to Newton's Law of Gravity, assume that the number of trips, or flow, between two locations decreases with the distance, travel time or travel cost between the locations. Intervening opportunities models assume that the flow between two locations can be determined by the number of intervening opportunities, or other potential destinations, that lie between the locations. Together these models provide a mathematical framework for modelling human mobility flows [16, 62] along with other types of spatial flows [41, 63, 65].

1.3.2.1 Gravity Models

Gravity models, first proposed by George Zipf (of Zipf's law) in 1946 [145] as a method of determining mobility flows, suggest that the migratory flow T_{ij} between locations i and j can be approximated by

$$T_{ij} \propto \frac{P_i P_j}{r_{ij}^2}, \tag{1.11}$$

where P_i and P_j are the populations of i and j and r_{ij} is the (geographic) distance between them.

The model assumes that the number of individuals leaving i is proportional to the population of i, the number of individuals arriving at j, or its attractiveness, is proportional to the population of j, and that there is some distance related cost in moving from i to j. Using these ideas, gravity models can be formulated in the more general form:

$$T_{ij} = K m_i m_j f(r_{ij}). \tag{1.12}$$

The masses m_i, m_j, relate to the number of trips leaving i or the number attracted by j and $f(r_{ij})$ is a deterrence function; a decreasing function of r_{ij}. Here, the masses are often some function of the population of i and j but not necessarily linear; it is common in literature to use the population raised to some power which is different for each location. The deterrence function $f(r_{ij})$ is often a power law or exponential function, however other more complex functions can be considered. The use of a power law or exponential deterrence function allows for the parameters of the model to be easily fit; for these cases the gravity model belongs to the family of Generalised Linear Models [93].

Gravity models have been popular in many fields including transport planning [34, 40], geographical studies [136] and spatial economics [64, 99] due to their ability to estimate trip flows, and consequently traffic demand, between two different locations based on local properties of the locations. These models can also be used in situations where it is essential to have knowledge of mobility flows, such as in epidemic modelling [12, 75]. Despite their widespread use, gravity models are a simplification of travel flows that may not accurately describe observations made from data [70, 114]. Furthermore, they require estimations of a number of free parameters making them sensitive to incomplete data or fluctuations in the data [62, 114].

One way of overcoming some of these limitations is by constraining the gravity model. For example, if the number of individuals leaving location i is known to be O_i, then the task of estimating the flow of individuals between location i and j is simplified; the number of unknown parameters in Equation 1.12 has been reduced as m_i is replaced by O_i. This type of model is called a singly-constrained gravity model and takes the form

$$T_{ij} = K_i O_i m_j f(r_{ij}) = O_i \frac{m_j f(r_{ij})}{\sum_k m_k f(r_{ik})}. \tag{1.13}$$

where the constant K_i now depends on the origin and its distance from all other locations. The idea of constraining a gravity model can be taken further; a doubly-constrained gravity model also

fixes the number of individuals arriving at location j as $D_j = \sum_i T_{ij}$. While the use of constrained models overcomes some of the limitations of unconstrained gravity models, these models are only suitable if in-going and out-going flows are known quantities.

A derivation and justification of Equation 1.13 comes from Alan Wilson [135]. He argues that if no information about an origin destination matrix is known, then the most probable configuration of a set of flows $\{T_{ij}\}$ is one that maximises the number of configurations of trips associated with it. Considering $\{T_{ij}\}$ as a state, we know from information theory that the less information we have about this state the higher its entropy. In the case of an origin-destination matrix, if we have no information then for each origin i and destination j we know only the total number of individuals that leave i, O_i, and the total number of individuals that arrive at j, D_j: we do not know the destinations of the individuals leaving i nor the origins of the individuals arriving at j.

If we have missing or no information about the flows between locations, the set of flows $\{T_{ij}\}$ can be approximated by maximising the entropy of the number of distinct arrangements of individuals, $\Omega(\{T_{ij}\})$, that give rise to the set of flows $\{T_{ij}\}$ [14]. Here, $\Omega(\{T_{ij}\})$ represents the number of ways in which T_{11} individuals can be selected from the total number of travellers $T = \sum_{ij} T_{ij}$; T_{12} from the remaining $T - T_{11}$, etc. This maximisation can be performed using Lagrange multipliers, subject to the constraints: $\sum_j T_{ij} = O_i$; $\sum_i T_{ij} = D_j$; $\sum_{ij} T_{ij} C_{ij} = C$ where C_{ij} is the cost of travel from location i to location j. In the limit of a large number of trips, the configuration that maximises Ω is given by

$$T_{ij} = K_i O_i L_j D_j e^{-\beta C_{ij}}. \tag{1.14}$$

Here, the Lagrange multiplier β controls the effect of the cost on flows and can be calibrated using the data. The values of K_i and L_j are set to fulfil the constraints on O_i and D_j respectively, namely $\sum_j T_{ij} = O_i$ and $\sum_i T_{ij} = D_j$.

1.3.2.2 Intervening Opportunities Model

The intervening opportunities model of human mobility was formally introduced by Stouffer in 1940 [125]. The intervening opportunities model suggests that the key factor in determining the mobility flow between two locations is not the distance between them but rather the cumulative number of opportunities between them. Opportunities can be defined based upon the social phenomena under investigation, for example the number of job opportunities if commuting flows are being estimated. The probability of making a trip is related to the accessibility of opportunities for satisfying the purpose of the trip.

Formally, if destinations j are ranked by their travel cost form the origin then the flow T_{ij} of individuals from location i to the location with rank j is given by:

$$T_{ij} = O_i \frac{e^{-LV_{ij-1}} - e^{-LV_{ij}}}{1 - e^{-LV_{in}}}. \tag{1.15}$$

where O_i corresponds to the number of individuals leaving the origin i, V_{ij} is the cumulative number of opportunities up to the location with rank j and n is the total number of possible destinations in the region considered. The number of opportunities at locations j is often taken as the population, m_j, or the number of arrivals to j, $D_j = \sum_i T_{ij}$. Parameter L can be considered the probability of relocating to any destination; this is a free parameter of the model which can be adjusted when fitting the model to data.

1.3.2.3 Radiation Model

The radiation model [114] can be considered an extension of the intervening opportunities model. In this case, for every individual each possible destination is given a fitness z_j, extracted from a distribution $p(z)$, which represents the opportunities for the individual at that destination. Each individual is also assigned a threshold fitness, z_t, which represents the minimum number of opportunities that individual requires from a possible destination. All destinations j are ranked according to their distance from origin i and the destination is selected such that it is the closest destination with enough opportunities, $z_j \geq z_t$, to satisfy the individual.

In this framework, the average number of travellers from location i to location j, T_{ij} is given by

$$T_{ij} = O_i \frac{1}{1 - \frac{m_i}{M}} \frac{m_i m_j}{(m_i + s_{ij})(m_i + m_j + s_{ij})}, \tag{1.16}$$

where O_i represents the number of individuals leaving location i which is sampled from a distribution of probabilities that a trip originating in i ends in j. Parameters m_i and m_j correspond to the number of opportunities at the origin and destination respectively and s_{ij} represents the total number of opportunities in a circle of radius r_{ij} centred on the origin. The normalising factor $1/(1 - m_i/M)$ ensures that all trips starting in the region under consideration end there. Here M represents the total number of opportunities; $M = \sum_i m_i$.

As with the intervening opportunities model, in the framework of the radiation model opportunities can be approximated by the population of each location or by the inflow of individuals arriving at a location D_j [70, 71, 85]. The radiation model has the advantage over the gravity and intervening opportunities models in that there are no free parameters that need to be calibrated with data. However, this may also be considered a weakness of the model as it is not robust to changes in the spatial scale being considered [70, 71, 76, 85]. Extensions have been proposed to overcome this limitation [116, 138] which have been shown to improve the model's performance. In the extended version of the model, a free parameter α is introduced to control the strength of the effect of the opportunities between the origin and destination on the location choice.

1.3.2.4 Model comparison

Naturally, since the gravity, intervening opportunities and radiation models all provide a framework for estimating mobility flows, the models have been compared on several occa-

sions [68, 85, 102, 137, 143]. Generally both gravity and intervening opportunities models perform comparatively well, however the intervening opportunities model has not seen such widespread use. This is possibly due to a lack of research into its implementation and calibration and also due to the practical advantage of using geographical distance between two locations to estimate mobility flows as opposed to inferring opportunities from data. Due to the recency of its introduction, less research exists on the radiation model. However, that which does suggests results are competitive with pre-existing spatial models of population dynamics [85].

Despite the fact that both the gravity and intervening opportunities models are able to estimate migration flows with comparable accuracy when fitted to empirical data, currently there is no objective quantitative criterion for selecting one modelling approach over the other in order to infer which between geographic distance or intervening opportunities is the variable that best describes domestic migration flows.

Spatial models of population dynamics lend themselves to many diverse applications. In [62], it is shown that the traffic flow between any two cities in Korea is well described by an unconstrained gravity model with a power law deterrence function $f(r_{ij}) = r_{ij}^{-2}$. Research into the spread of invasive species has looked into the movement of cargo ships between ports [63] due to their potential to carry such species. They find that a doubly-constrained gravity model with a truncated power law deterrence function best describes global ship movements. An unconstrained gravity model has also been shown to be a good fit for commuting flows worldwide [11] however in this case an exponential form of the deterrence function, $f(r_{ij}) = e^{r_{ij}/R}$, is the most appropriate to describe the data. The radiation model has also been applied to commuting flows. In fact, the formulation of the radiation model is based upon the process of how an individual selects a job from multiple job offers [114]. The model is evaluated against commuting flows in the United States where it is demonstrated that it out-performs the gravity model at predicting commuting patterns at both the intra-state and national spatial scales.

1.3.3 Occupational Migration and the Science of Science

Commuting flows can be considered a type of occupational migration: movement occurring due to an individual's specific job or the sector that they work in. An alternative aspect of occupational migration is permanent migration due to job transitions. Job transitions are nowadays seen to be essential for a successful career across many professional industries. Factors that this may be attributed to include; higher living costs increasing the pressure to gain a position with a higher salary, personal and professional goals and the highly competitive nature of the job market. The normality of job transitions is no less evident in academia; in the academic community there is a widely accepted belief that movement between institutions is beneficial to, possibly even essential for, a successful career. Consequentially, changing institution is a key part of academic life and a key career decision of any scientist, playing an important role in scientific productivity and the generation of scientific knowledge. For scientists, movement between institutions (henceforth

referred to as scientific migration) allows them to build their scientific profile; a strong academic environment can allow a scientist to be more effective in their own research and contribute to the success of their institution.

Various aspects of scientific profiles have been analysed in recent years, with publication data being central in the process. One line of research focusses on collaboration networks, which has brought important insights and results in the field of complex network theory, and its applications on analysing real-world networks [31, 94, 100]. In such networks, each scientist is a node and each link represents a collaboration. Early work on scientific collaboration networks [94] found that they exhibit small world structure; only ~ 6 steps are needed to get from one randomly chosen scientist to another. Furthermore, two scientists have at minimum a 30% probability of collaboration if they have a collaborator in common, demonstrating a strong clustering effect. It has also been shown that a rich club phenomenon exists in scientific collaboration networks [31]. In this case, high degree nodes are very well connected to each other and nodes with high centrality form tightly interconnected communities. These results suggest that sociability plays a large role in scientific collaborations; the more scientists you collaborate with, the larger the probability of forming new collaborations in the future. More recently Scellato et al [109] found that scientists who have lived and worked in a different country have a larger collaboration network and a higher number of international collaborations. This demonstrates a link between scientific migration and collaboration networks, highlighting the importance of scientific migration with regards to scientific productivity and education.

A large proportion of existing work on scientific migration focuses on large-scale surveys of country-level movements. These surveys are designed to reveal long-term cultural and economic priorities [5, 88, 96]. Appelt et al. [3] looked at OECD data on changes in affiliation of authors between 1996 and 2011, using a gravity-based empirical framework to investigate the factors that influence the international mobility of scientists. They found that, whilst geographic distance and scientific proximity negatively correlate with the mobility of scientists between two countries, researchers are more likely to move to a country that has no visa restrictions based on their nationality, speaks the same language and is in a similar or better economic position. Research into the mobility of elite scientists has investigated the professional and personal determinants of the decision to relocate to a new institution [9]. Elite scientists are defined as scientific researchers that are highly funded or cited, have a high number of patents, are a member of a national academy of scientists or are award winners. Findings show that elite scientists are less likely to move if they have recently received funding or have children of high-school age, suggesting they find it costly to disrupt the social networks of their children. However highly productive scientists, ie. scientists with more publications, are more likely to move, particularly if their local environment is of lower quality compared to a distant environment. A recent survey of over 2000 Nature readers [96] mostly from the US and Europe, found that improved quality of life, increased funding opportunities and a better salary were all incentives for scientists to move

and work abroad. In contrast, yet in agreement with [3], they also find that unfavourable socio-economic conditions such as visa restrictions and an unstable political situation act as barriers to movement. The GlobSci survey [44], involving around 17,000 researchers in four scientific fields across 16 countries has generated further results that align with those aforementioned: better career prospects and stronger environments are the most important factors in a scientist's decision to move [96].

Another line of research has looked at the evaluation of the productivity and performance of scientists. Scientific productivity can be considered relevant for career progression in academia and is therefore an important part of a scientific profile and scientific migration. Multiple performance indices have been proposed over the years including the number of papers published (in a year) [118], and the time it takes for a paper to collect the majority of its citations [133]. Research into understanding the impact of a scientist's relocation on their scientific performance also features prominently in this area. Analysing the relocations and the scientific performance of scientists, it has been found that while moves from elite to lower-rank institutions lead to a moderate decrease in scientific performance, moves to elite institutions do not necessarily result in subsequent performance gain [36] . In [118] empirical evidence demonstrates that scientific impact is randomly distributed within the sequence of papers published by an individual during their scientific career. This implies that temporal changes in impact can be explained by temporal changes in productivity or luck. An analysis of the migration traces of scientists extracted from Web of Science had revealed that, regardless the nation of origin, scientists who relocate are more highly cited than their non-moving counterparts [126]. Further outcomes from the GlobSci survey [43, 110] show that on average the impact of foreign born scientists is higher than that of natives who have no international mobility experience. Findings suggest that scientific migration between countries correlates with a performance boost; scientific migration enhances scientific performance. These results all support the belief that migration between institutions is beneficial to a scientific career.

In the context of studying occupational migration, the availability of massive datasets of individuals' career paths has led to models that aim to predict what an individual's next job will be. In [97] a system is built to recommend new jobs to people who are seeking a job, using all their past job transitions as well as their employment data. They train a machine learning model to show that job transitions can be accurately predicted, significantly improving over a baseline that always predicts the most frequent institution in the data. Recently a system to predict next career moves based on profile context matching and career path mining from a real-world LinkedIn dataset has been proposed [73]. The system is demonstrated to accurately predict future career moves, revealing interesting insights into occupational migration at the micro-level.

Despite this wealth of research, our understanding of the factors influencing a scientist's decision to relocate, such as their scientific profile or the quality of their scientific environment, is limited. Furthermore, while the role of social relationships is known to influence human activity

in several contexts [28, 133], it is not clear what the contribution of a scientist's collaboration network is on their decision to relocate.

1.4 Thesis outline

In the following, we address the knowledge gaps highlighted in the previous sections. In Chapter 2 we introduce a deterministic model of population dynamics combining population growth and migration. Using both gravity and intervening opportunities models of migration, we analytically assess the spatial distribution of population that each model produces. We use data on the size and spatial distribution of cities globally to assess whether it is possible to determine which spatial model of migration correctly describes the migration of individuals within a fixed region. We conjecture 'Heaps' Law of Cities': a scaling relation between the total population of a region and the number of cities within it. This is validated using the same data. Finally we assess the extent to which the size and spatial distribution depends on the definition of a city using a gridded population dataset. We use a clustering algorithm to obtain cities from the gridded population and investigate the presence of Zipf's law for population and area of cities and Heaps' law for the number of cities. In Chapter 3 we present a microscopic model of births and deaths that is able to reproduce Zipf's Law for the distribution of large cities without fine tuning. We also characterise the scaling between the expected population growth and its variance, relating it to a relationship originating from ecology: Taylor's law. We validate our model using empirical data on the populations of cities and counties in the United States. In Chapter 4 we use machine learning techniques to determine which factors are most influential in a scientist's decision to migrate. Using these results we present an adaptation of a singly constrained gravity model which incorporates additional factors in the decision to relocate. We demonstrate that this model is more successful than a singly constrained gravity model at predicting where a scientist will move to once they have made the decision to relocate.

Our work provides further insight into important aspects of the urbanisation process. Our results from Chapter 2 demonstrate that the number of cities in a region can be fully described by the total population of the region and is independent of the region's population density or area. We also show that the spatial distribution of cities suggests cities are randomly distributed in space; there are no significant spatial correlations. Our novel deterministic model, while producing interesting analytical results, needs adapting into a stochastic form in order to provide a realistic description of the interplay between migration and population growth. In Chapter 3 we demonstrate that in order to produce a size distribution of cities that is in line with empirical distributions, the assumption that growth rates are independent and identically distributed, must be relaxed. While aggregately (ie. at the level of a city) this assumption is valid, we show that, at the level of the individual inhabitants, introducing correlations and environmental variability results in a distribution of city sizes that, above a threshold population, can be described by

Zipf's law: our model provides a microscopic explanation of Gibrat's law. Finally in Chapter 4 we demonstrate that when a scientist moves between institutions it is personal factors that are most important in determining when they move and where they move to. The methodology implemented in this chapter, using machine learning methods to inform a modified model of migration, marks a possible advancement for future models of migration and a general model of urbanisation.

CHAPTER 2

THE SIZE AND SPATIAL DISTRIBUTION OF CITIES

2.1 Introduction

Compared to the great efforts made to characterise the distribution of city sizes both empirically and theoretically, much less work has been done to answer the other fundamental question about the urbanisation process: What determines the number of cities in a country? In this chapter we investigate the relationship between the population of a region and the number and spatial distribution of cities within it. Initially we consider the analytical relationship between population and number of cities, introducing a deterministic model of population dynamics in section 2.2 in order to understand how cities form due to the interplay between natural increase and migrations. To account for migrations, in section 2.3 we consider a gravity model and in section 2.4, an intervening opportunities model. For each migration model we linearise the deterministic equation about a steady-state population density and study the evolution of a small perturbation to the steady state. We obtain analytical relationships between the population of a region and the number of cities expected for each of the gravity and intervening opportunities models. Analytically, we find that if individuals migrate according to a gravity model, the number of cities is independent of the population of a region whereas if they migrate according to an intervening opportunities model, the number of cities increases linearly with population. We assess the extent to which the analytical results agree with numerical simulations of our deterministic model in section 2.6. Following on from this, in section 2.7 we investigate the empirical relationship between the total population of a region, the number of cities within it and their spatial distribution using a dataset on the population of all cities globally. Here we introduce the concept of Heaps' law for cities: an empirical relationship between the number of cities in a region and the region's population. The validity of Heaps' law demonstrates that cities are randomly distributed in space; that is, there is no evidence of a relationship between the size

of urban areas and their spatial distribution. In 2.8 we demonstrate that these results are also valid for urban clusters in the United States. The empirical evidence for Heaps' law for cities, where the number of cities increases with a region's population, suggests that an intervening opportunities model may offer a more realistic description of migration flows: Heaps' law is equivalent to the analytical result for this model obtained in 2.4. In section 2.9 we discuss the merits and limitations of our deterministic approach to modelling population dynamics.

The work presented in this chapter has been carried out in collaboration with F. Simini and can, in part, be found in [115]. Specifically, F. Simini derived the analytical results presented in sections 2.7.1 and 2.8 and performed the data analysis presented in section 2.8.

2.2 A deterministic model of population dynamics

Our first step in investigating how the number of cities within a region is affected by the region's population is to theoretically understand how cities are formed. Our approach is to use a deterministic model, similar to those used in economic geography [47, 66], which describes the evolution of the population density.

A general theory aiming at describing human demographic dynamics, i.e. the spatio-temporal evolution of the population distribution in a region, must consider all factors that contribute to the population change in any location. These contributions are reduced to the following two fundamental categories: births, deaths and external (international) migrations on one end, and internal (national) migrations or relocations on the other. The first contribution modifies the total population of the region, while the second is responsible for the redistribution of individuals within the region's area.

If relocations from one point to another are to be considered, alongside population growth and international migrations, a deterministic equation to describe the change in population density, $\rho(x,t)$, at location x is given by:

$$\frac{\partial \rho(x,t)}{\partial t} = g(\rho(x,t)) - T^{out}(x,t,\rho) + T^{in}(x,t,\rho). \tag{2.1}$$

T^{out} and T^{in} are the number of individuals leaving and relocating to point x from other parts of the region respectively; these terms represent internal migrations. The form of these functions is determined by the mathematical model used to describe the flow of individuals who relocate from one point to another. Here we consider a gravity model alongside intervening opportunities and radiation models. The probability that an individual will relocate to a given location is dependent on both the distance to and opportunities at the destination (gravity models) or at all intermediate locations (intervening opportunities and radiation models). Here, an opportunity refers to any property of a location that may be of interest to the relocating individual. We distinguish between two classes of opportunities; those that depend on the population, such as the availability of goods and services, and those that are independent of population, such as the presence of naturally occurring resources.

The function $g(\rho(x,t))$ represents the contribution to population change caused by births, deaths, and external (international) migrations. It can be zero, $g(\rho(x,t)) = 0$, if the system has constant population, linear $g(\rho(x,t)) = g \cdot \rho(x,t)$ for an exponential population growth, or Logistic $g(\rho(x,t)) = g \cdot \rho(x,t)(1 - \rho(x,t)/\rho_0)$, characterising population dynamics with a uniform stationary state, $\rho(x,t) = \rho_0$ in the absence of internal migrations. Here g is the population growth rate and ρ_0 is often called the carrying capacity.

In the following we consider a logistic form of growth and study the time evolution of a continuous population distribution when subjected to small perturbations about a uniform stationary state $\rho(x,t) = \rho_0$. While this approach neglects the inherent randomness of stochastic models required to reproduce phenomena such as Zipf's law, it allows for a fully deterministic analysis of the migration models and their parameters, particularly the parameter constraints under which cities will form. Furthermore we note that additional contributions, such as a diffusion term responsible for the local relocation of individuals from highly populated areas to less populated suburbs, can easily be added, however these will not be considered here.

2.3 Gravity Model

In the context of migratory dynamics, the gravity model suggests that the probability per unit time that an individual moves between two locations, say x_i and x_j, is proportional to the product of the population densities at x_i and x_j, each to some power, and a function of the distance, r_{ij}, between x_i and x_j, ie:

$$P_{x_i}(x_j) = \frac{\rho(x_i)^\alpha \rho(x_j)^\beta f(r_{ij})}{\int_D dx_j \rho(x_i)^\alpha \rho(x_j)^\beta f(r_{ij})}. \tag{2.2}$$

The denominator ensures that the probability, $P_{x_i}(x_j)$, is correctly normalised over the domain D. Exponents α and β are usually positive, and often $\beta = 1$. $f(r)$ may be any continuous function, usually an exponential, e^{-rR}, or power law, $r^{-\gamma}$.

In Equation 2.2 it is assumed that individuals will relocate to another position only if there is a non-zero population at the destination. In reality, the probability for an individual to relocate is also dependent upon the availability of natural resources at the new location or more generally, any opportunity which is not related to the population, such as the presence of fertile soil or minerals. If these resources are combined into a single term, w, which has the same units as the population opportunities, then we can re-write Equation 2.2 as

$$P_{x_i}(x_j) = \frac{[\rho(x_j) + w(x_j)] f(r_{ij})}{\int_D dx_j [\rho(x_j) + w(x_j)] f(r_{ij})}. \tag{2.3}$$

where β has been replaced with 1, $\rho(x_i)^\alpha$ has cancelled, and $\rho(x_j) + w(x_j)$ denotes the total number of opportunities at location x_j. For simplicity we assume that w is stationary in time; this may be justified by the fact that the rate of recovery of such resources is greater than the rate of consumption. Setting the value of β to equal 1 means that the probability of migration to x_j

is linearly proportional to the population density at that location. For $\beta > 1$, the effect of large cities is amplified: the probability per unit time of an individual relocating to a given city grows at a faster rate than the population of the city itself resulting in a small number of very large cities. In contrast, for $\beta < 1$, the probability per unit time of an individual relocating to a given city increases at a slower rate than the population of the city, resulting in many smaller cities forming.

In order to model the change in population density at location x_i, ie $\rho(x_i)$, in time, the average number of people both leaving from and moving to x_i must be included. The number of people leaving i per unit time is given by:

$$T^{out}(x_i) = T(\rho(x_i), w(x_i))\rho(x_i) \int_D dx_j P_{x_i}(x_j) = T(\rho(x_i), w(x_i))\rho(x_i). \tag{2.4}$$

In the above, $T(\rho(x_i), w(x_i))$ is the average migration rate, i.e. the fraction of people in x_i that will relocate in a time unit. To a first approximation this is assumed to be constant and independent of location; $T(\rho(x_i), w(x_i)) = T$. The average number of people relocating to x_i per unit time is given by:

$$\begin{aligned} T^{in}(x_i) &= \int_D dx_j T(\rho(x_j), w(x_j))\rho(x_j) P_{x_j}(x_i) \\ &= [\rho(x_i) + w(x_i)] \int_D dx_j \frac{T(\rho(x_j), w(x_j))\rho(x_j) f(r_{ij})}{\int_D dx_i [\rho(x_i) + w(x_i)] f(r_{ij})}. \end{aligned} \tag{2.5}$$

Inserting these expressions into Equation 2.1 allows for the change in population density with time, as described by a gravity model, to be studied.

2.4 Intervening Opportunities & Radiation Models

The intervening opportunities model suggests that the probability of migration is more strongly influenced by the opportunities at the new destination as opposed to distance or population density. According to this model, the probability per unit time of observing an individual at location x_i moving to location x_j is given by:

$$P_i(x_j) = [\rho(x_j) + w(x_j)] f\left(\int_{B_{x_i}(r_{ij})} dz [\rho(z) + w(z)] \right), \tag{2.6}$$

where $B_{x_i}(r_{ij})$ is a ball of radius $|x_j - x_i|$ centred on x_i and f is a continuous function. The normalisation of Equation 2.6 over a domain D with total population N may be imposed as a condition on f:

$$\begin{aligned}\int_D dx_j P_{x_i}(x_j) &= \int_0^\infty dr r \int_0^{2\pi} d\theta[\rho(r,\theta)+w(r,\theta)] f\left(\int_0^r dz z \int_0^{2\pi} d\theta[\rho(z,\theta)+w(z,\theta)]\right) \\ &= \int_0^\infty dr \frac{da(r)}{dr} f(a(r)) \\ &= \int_0^N da f(a) \\ &= F(0)-F(N)=1.\end{aligned} \tag{2.7}$$

Here, $a(r) = \int_0^r dz z \int_0^{2\pi} d\theta[\rho(z,\theta)+w(z,\theta)]$, and $F(a) \equiv \int_a^\infty du f(u)$ is a decreasing function such that $F(0) = 1$ and $F(N) = 0$. The deterrence function, $F(a)$, may be either exponential, e^{-aR}, as is the case for the intervening opportunities model, or power law, $\frac{1}{1+a}$, which corresponds to the radiation model.

For these models, the total number of travellers leaving x_i per unit time is the same as for the gravity model, described by Equation 2.4. The total number of people moving to x_i per unit time is given by:

$$\begin{aligned}T^{in}(x_i) &= \int_D dx_j T(\rho(x_j), w(x_j))\rho(x_j) P_{x_j}(x_i) \\ &= [\rho(x_i)+w(x_i)] \int_D dx_j T(\rho(x_j), w(x_j))\rho(x_j)] f\left(\int_{B_{x_i}(r_{ij})} dz[\rho(z)+w(z)]\right).\end{aligned} \tag{2.8}$$

These terms combine to give the general dynamic equation for the time evolution of the population density as per Equation 2.1.

2.5 Analytical results

The uniform distribution $\rho(x) = \rho_0$ is a stationary state of Equation 2.1 because the growth term is equal to zero and T^{in} is equal to T^{out} for all x, hence the time derivative on the left-hand side is zero. In order to analytically assess the differences between the gravity and intervening opportunities models of migration we linearise Equation 2.1 about the stationary state and study the time evolution of a small perturbation $\tilde{\rho}(x,t) = \sum_k h_k(t) e^{ikx}$; the perturbation is a the sum over normal modes of wavenumber k and amplitude $h_k(t)$ (see Appendix A). For each model of migration we consider an exponential deterrence function, $f(r) = e^{-rR}$, and a power-law deterrence function, $f(r) = (1+r)^{-\gamma}$ with $\gamma > 0$.

On performing this analysis we find that if a population is described by a logistic growth with individuals relocating according to a gravity model (see A.1), cities will develop if the following two conditions are met simultaneously:

1. The carrying capacity is much larger than the density of resources: $\rho_0 \gg w$.

2. The population is sufficiently mobile: $T > 4g$.

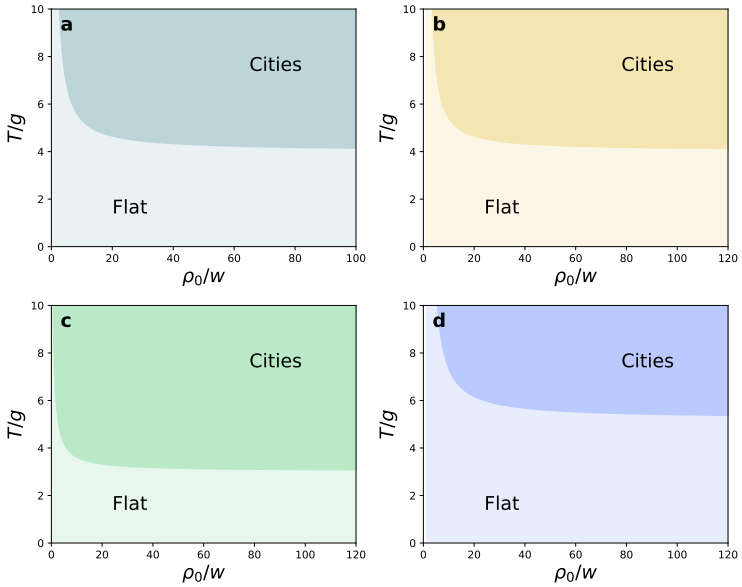

FIGURE 2.1. **a** Gravity Model - exponential f(r): Parameter space (ρ_0/w, T/g). Above the critical curve the formation of cities is possible whilst below it, any perturbations to the steady state distribution will decay back to ρ_0 and the landscape will be flat. **b** Gravity Model - power law f(r): Parameter space (ρ_0/w, T/g). By comparison with **a** we demonstrate that for the gravity model the particular deterrence function has no effect on the condition for cities to form. **c** Intervening Opportunities Model: Parameter space (ρ_0/w, T/g). From comparison with **a** and **b** it can be seen that for an intervening opportunities model cities may form when the ratio between the migration rate and growth rate is lower than that required for gravity models. **d** Radiation Model: Parameter space (ρ_0/w, T/g). By comparison with **c** it can be seen that if people relocate according to a radiation model, the ratio between the migration rate and growth rate must be higher than that for an intervening opportunities model in order for cities to form.

When the population density is sufficiently high, the average distance between cities depends only on the deterrence function f: it is independent of the growth rate, migration rate and carrying capacity. These results are shown in Figures 2.1**a** and 2.1**b** where it is seen that the critical curves for the gravity model with both exponential and power-law deterrence functions are the same. As the carrying capacity increases, $\rho_0 \to \infty$, the wavelength of the mode with the highest instability, k_m, which is proportional to the average number of cities per unit length, saturates to a constant value: the number of cities within a region is independent of the region's population. The first condition, namely that the carrying capacity must be much larger than the density of resources in order for cities to form, suggests that resources relating to the population, such as the availability of goods and services, are far more important for the urbanisation process than resources, such as fertile soil, that are independent of the population.

If instead the population relocates according to an intervening opportunities or radiation

model (see A.2), an initial perturbation to the steady state population distribution will result in the development of urban settlements if the following two conditions are met simultaneously:

1. The carrying capacity is much larger than the density of resources: $\rho_0 > c_0 w$, with $c_0 = 1/3$ and 1 for the intervening opportunities and radiation models respectively.

2. The population is sufficiently mobile: $T > c_1 g$ with $c_1 = 3$ and 5 for the intervening opportunities and radiation models respectively.

Here, the intervening opportunities model corresponds to the exponential deterrence function, $F(r) = e^{-rR}$, and the radiation model corresponds to the power-law deterrence function, $F(r) = (1+r)^{-\gamma}$ with $\gamma > 0$. These results are shown in Figures 2.1**c** and 2.1**d** where it is seen that the critical curves for the intervening opportunities and radiation models are not the same; in contrast to gravity models, for intervening opportunities type models these curves do depend on the deterrence function used. In this case as the carrying capacity ρ_0 increases, k_m increases linearly with ρ_0: the number of cities within a region is linearly dependent on the region's population. These results apply for both the 1-dimensional and 2-dimensional versions of our deterministic equation.

It is important to note the distinct contrast between the gravity model and an intervening opportunities model; the variation of k_m with increasing density ρ_0. The value of k_m may be related to the number of cities, C, that will grow from a small perturbation through the equation

$$C = L k_m / 2\pi. \tag{2.9}$$

For the gravity model, as $\rho_0 \to \infty$, k_m saturates to a constant value, whereas for the intervening opportunities model, k_m increases proportionally to ρ_0. This result is summarised in Figure 2.2. From this difference is it possible to determine which model of migration best describes patterns seen in data.

2.6 Numerical Simulations

We simulate both the gravity and intervening opportunities models using numerical methods in order to assess the accuracy of the analytical results derived above. Specifically, as we are interested in the spatial distribution of cities, we simulate our deterministic model, Equation 2.1, for both the gravity and intervening opportunities models of migration and assess how the number of cities varies with an increasing carrying capacity ρ_0. Numerical simulations demonstrating the validity of the conditions for city formation can be found in [115].

2.6.1 Gravity Model

To simulate the gravity model in 1-dimension, we evolve the dynamic equation A.1 in time for an exponential deterrence function, starting with an initial distribution at $t = 0$ of small

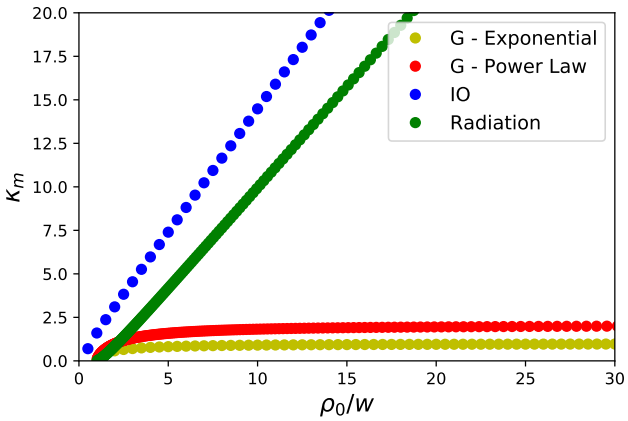

FIGURE 2.2. The wavelength of the mode with the highest instability, k_m, which is proportional to the number of cities per unit length, as a function of the ratio ρ_0/w. For gravity models k_m tends to a constant in the limit of high carrying capacity, $\rho_0 \gg w$, whereas in the same limit k_m grows proportionally to ρ_0 for intervening opportunities models.

random perturbations about the steady state solution ρ_0. The integrals are evaluated using the trapezoidal rule. By varying the model parameters, the validity of the relationship between the wavenumber and the number of cities, Equation 2.9, is determined. Unless otherwise stated, we rescale variables by setting growth rate and resources, g and w respectively, to equal unity: this allows for the model to be specified with 3 variables rather than 5 and is equivalent to rescaling the units of time and population. We consider the case where resources, w, are uniformly distributed in space.

We simulate the model with fixed parameters $T = 10$ and $R = 1.5$ and increasing carrying capacity $10 \leq \rho_0 \leq 90$. We use an average migration rate of $T = 10$ based on the condition that in order for cities to form, $T > 4g$: for the case of $g = 0$ (Figure 2.3**a**), cities will form for any value of T, however for $g = 1$ (Figure 2.3**b**), cities will form for any $T > 4$ however the time for the population distribution to reach equilibrium increases as $T \to 4$. The value of R was selected to align with empirical values found for European countries [124]. From the simulations we find that increasing the carrying capacity increases the size of cities however the number of cities remains constant, determined by Equation 2.9. The results of this simulation are displayed in Figure 2.3**a,b**.

2.6.2 Intervening Opportunities Model

In order to numerically simulate the intervening opportunities model, we make an adjustment to the expression for the probability of observing an individual at x_i moving to x_j, originally formulated in Equation 2.6. Rather than using the probability density function $f(a)$, we use the

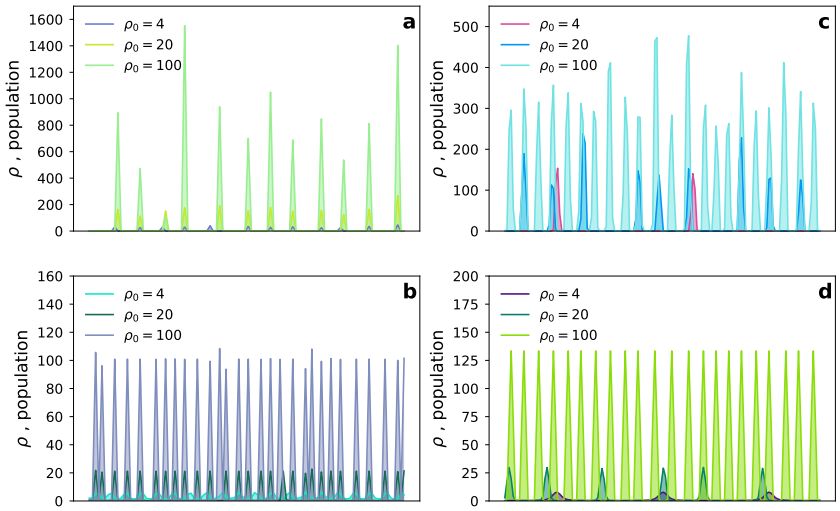

FIGURE 2.3. The final distribution obtained from **a** Gravity model, with parameters T=10, R=1.5, g=0 **b** Gravity model, with parameters T=10, R=1.5, g=1 **c** Intervening opportunities model, with parameters T=10, R=0.1, g=0 **d** Intervening opportunities model, with parameters T=10, R=0.1, g=1. ρ_0 is varied between 4 and 100 to demonstrate the relationship between the number of peaks and the steady state population density for each model.

cumulative distribution function, $F(a)$. On doing this, we write the dynamic equation, analogous to Equation A.14, as:

(2.10)
$$\begin{aligned}\frac{\partial\rho(x_i)}{\partial t} &= g\rho(x_i)(1-\frac{\rho(x_i)}{\rho_0}) - T\rho(x_i) + T\sum_j P_{ij} \\ &= g\rho(x_i)(1-\frac{\rho(x_i)}{\rho_0}) - T\rho(x_i) + T.\sum_j \left[\frac{F(a_{ij}) - F(a_{ij}+\rho(x_j)+w(x_j))}{F(\rho(x_i)+w(x_i)) - F(N)}\right].\frac{\rho(x_j)+w(x_j)}{\sum_{x_k\in R_{ij}}\rho(x_k)+w(x_k)}\end{aligned}$$

where R_{ij} is the set of locations on the ring of radius r_{ij}, centered on x_i, therefore the summation:

(2.11)
$$\sum_{x_k\in R_{ij}} \rho(x_k)+w(x_k)$$

is a sum over all locations x_k at the same distance as x_j from the origin x_i. We define D_{ij} such that it represents the set of locations within the disc of radius r_{ij} centered in x_i and the opportunities between but not including locations x_i and x_j may be written:

(2.12)
$$a_{ij} = \sum_{x_k\in D_{ij}} \rho(x_k)+w(x_k).$$

Using this formulation, the dynamical equation corresponding to a population relocating according to an intervening opportunities model with an exponential deterrence function is simulated.

Simulations are carried out for fixed T and R whilst varying ρ_0, the results of which are shown in Figure 2.3**c-d**. From this we find that the number of cities increases with ρ_0 for an intervening opportunities model as expected from Equations A.19 and A.14.

2.6.3 Model Comparison

From both the analytical results and those obtained from simulations, we note that the main contrast between the gravity and intervening opportunities models is their behaviour with increasing ρ_0, specifically how the number of cities changes due to a change in the steady state population density. From the analysis of the gravity model, we find that the number of cities does not vary with ρ_0; this is shown in Equation A.7 and supported by Figure 2.3**a-b**. In comparison to this, for the intervening opportunities model, we find that increasing ρ_0 also results in an increase in the number of cities. For this model, the expression for k_m suggests a linear relationship between ρ_0 and the number of cities, when all other parameters are fixed, and this is supported for low values of ρ_0 by Figure 2.3**c-d**.

The dynamic equation, 2.1, used for each model only differs in the migration term; both models follow a logistic growth. As a result, by setting this growth to equal zero the differences between gravity driven migration and migration driven by an intervening opportunities model may be effectively analysed. Setting the growth to zero to study the number of peaks is also justified by the fact that the number of peaks is proportional to k_m; for both models this is independent of the value of g.

2.6.3.1 1D

Figures 2.3**a** and 2.3**c** demonstrate the difference between the two models noted above when the growth term is set to zero. Here we simulate both models with an identical initial distribution and equal parameters except for R, the value of which is specific to the model being used. It is seen that as ρ_0 is increased the final distribution from the gravity model displays peaks of increasing height whereas that for the intervening opportunities model displays an increasing number of peaks alongside a less significant increase in height.

In Figures 2.3**b** and 2.3**d** we demonstrate the effect of a small growth on the final distributions for the gravity and Intervening Opportunity models respectively. We observe that the effect of a non-zero growth rate on the final distribution of cities is that all cities evolve to approximately the same size; ρ_0. This is a result of the logistic form of growth, $g \cdot (1 - \rho(x,t)/\rho_0)$; as $\rho(x,t)$ approaches ρ_0, the growth rate tends to zero, all cities have reached the same size and have equal opportunities, therefore there is also no further migration.

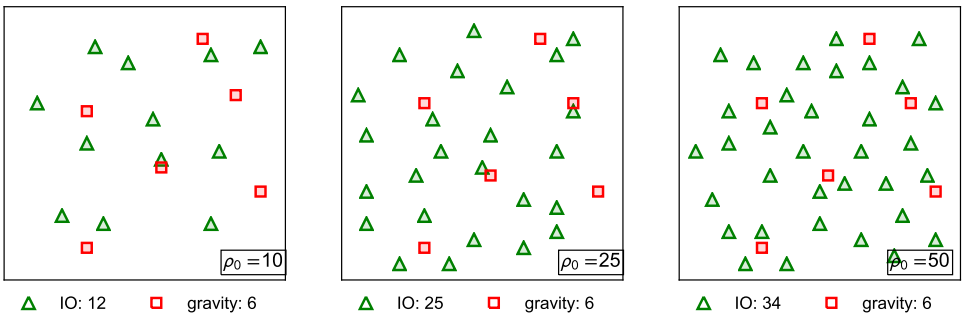

FIGURE 2.4. The final stationary distribution of cities when individuals relocate according to a gravity model (red) and an intervening opportunities model (green) for increasing values of $\rho_0 = 10, 25, 50$ with fixed parameters $L = 15$, $T = 30$, $R = 0.5$ (gravity model) and $R = 0.2$ (intervening opportunities model). It is observed that the final number of cities remains constant with increasing ρ_0 if individuals relocate according to a gravity model whereas the number of cities increases with ρ_0 if individuals relocate according to an intervening opportunities model.

2.6.3.2 2D

In Figure 2.4 we display the final stationary distributions of cities obtained from simulations of both the gravity model (red squares) and intervening opportunities model (green triangles) for initial distributions randomised about increasing values of ρ_0 over the range [10,50]. From this we observe that, in correspondence with the 1-dimensional simulations, the number of cities for the gravity model remains fixed and constant with increasing ρ_0 whereas for the intervening opportunities model the number of cities increases with ρ_0. As the initial distribution of population is randomised about ρ_0 this result demonstrates that if individuals relocate according to a gravity model the number of cities is unaffected by the total population of the region. In contrast, if it is an intervening opportunities model by which people relocate then the number of cities within a region will increase with the total population of that region.

2.6.3.3 Numerical results : summary

Numerical simulations have confirmed the analytical results summarised in section 2.5. Specifically, we have demonstrated that if ρ_0, the carrying capacity, is increased, the number of cities remains constant if individuals relocate according to a gravity model whereas it increases if individuals relocate according to an intervening opportunities model.

2.6.4 Summary

In the previous sections we assume that population grows logistically and migration can be described by either a gravity or intervening opportunities model. Our deterministic model suggests that if the number of cities in a country increases with the country's population,

individuals relocate according to an intervening opportunities model. In contrast, if the number of cities is independent of the population of a region, a gravity model is a better description of the demographic dynamics. However, so far our results are only analytical; in the following sections we empirically investigate the relationships between the number of cities in a region and some of the region's properties, such as the region's total population and built-up area. In particular, we consider how the total population (or the total built-up area) of a region affects the number of cities.

2.7 Heaps' Law for Cities

The relationship between the total population of a region and the number of cities is analogous to Heaps' Law in linguistics [51, 58], which describes the empirical scaling relationship between the number of distinct words, W, in a document and the total number of words in the document (or text length), N: $W \sim N^{\gamma}$, where $\gamma \leq 1$ is the Heaps exponent.

In fact, the analogy of Heaps' Law in the context of cities would state that the number of cities C in a total population of N individuals would scale as

$$C \sim N^{\gamma}. \tag{2.13}$$

Previous research has shown that Zipf's law and Heaps' law often appear together [81], suggesting that the presence of Zipf's law implies Heaps' law.

However, this relationship does not necessarily hold for spatially extended systems, such as cities, because evidence for Zipf's law at the country (global) scale does not necessarily imply the presence of Zipf's law and Heaps' law at the regional (local) scale. In fact, even if Zipf's law for the distribution of city sizes holds globally at the level of countries, it might not hold locally at smaller spatial scales if correlations in the spatial distribution of urban clusters are present. This would be the case, for example, if urban clusters were aggregated by size, so that it is more common to find clusters of similar sizes close to each other compared to the case in which clusters are randomly distributed among the regions, irrespective of their size. The overall (global) distribution of cluster sizes would not change and still be a power-law, but the size distributions in the regions would not follow Zipf's law anymore and as a consequence Heaps' law would not hold. Indeed, this is what happens in ecological systems, where macro-ecological statistical patterns of species distribution and abundance display a strong dependence on the spatial scale considered [8]. One of the most relevant statistics used to characterise the degree of biodiversity of ecosystems is the species-area relationship (SAR), which measures the number of different species expected to be found in areas of increasing size. Since the density of individuals per unit area is constant, the SAR is the equivalent of Heaps' law for ecosystems, as it measures the relationship between a region's total population and the expected number of different groups of individuals in the region, where here groups correspond to species instead of cities. Empirical measurements of the SAR show a different functional behaviour as the region's area increases,

and this is due to the fact that the shape of the distribution of species sizes, called the relative species abundance, depends on the spatial scale considered.

There is another reason to investigate the relationship between Zipf's and Heaps' laws for cities. Zipf's law for the distribution of city sizes usually holds only for the tail of the distribution, or the distribution of large cities. However the fact that in a region the distribution of city sizes has a power-law tail does not give any information regarding the relationship between the number of cities in the region and its total population. In other words, when Zipf's law holds only for large cities, there is no guarantee that Heaps' law holds as well. To understand this, consider a region in which city sizes follow Zipf's law. If the population of each city is doubled and hence the total population of the region is also doubled, yet no new cities are created, Zipf's law will still be present, albeit with a larger scale parameter (i.e. the minimum city size is doubled). However, Heaps' law will not hold in this case, or it will have exponent $\gamma = 0$, because the total population, N, is doubled, but the number of cities, C, has not changed. This result would in fact correspond to our deterministic equation with individuals relocating according the the gravity model.

In the following, we use a dataset on the population and location of cities globally to assess if Heaps' law holds for all countries in all continents (except Australia and Antartica), and to test the predicted relationship between Heaps' and Zipf's exponents. Cities can be defined in many different ways and various relevant properties of urban agglomerations, including the scaling relationships between population size and urban indicators such as area of roads and number of patents, depend on the method used to define cities [4, 23]. In particular, the relationship between the number of cities in a region and the region's total population, i.e. Heaps' law, can also depend on the definition of city considered. To understand how Heaps' law depends on the definition of city, we use a second dataset of the spatial distribution of population in the United States that allows us to consider various definitions of urban clusters and provide additional support to our results.

2.7.1 Analytical Results

When considering the distribution of cities in space, the most simple assumed (or null) model is that cities are in fact distributed randomly and that there are no spatial correlations between them. If we start by assuming that Zipf's law, ie. the global distribution of city sizes, holds for all regions then we can populate space by drawing from this distribution. For a total (target) region population N, the sizes of the cities populating the region are drawn from the distribution of city sizes. This process stops when the sum of the city sizes exceeds N and the number of cities C is given by the number of drawings from the distribution. Repeating this process many times and averaging over the results from all realisations we obtain an estimate for the number of cities expected in a region of total population N: $\langle C|N\rangle$.

For this null model, it is possible to obtain an analytical expression for $\langle C|N\rangle$, or Heaps' law. The probability to find a city with population x in a region with total population N and C cities is

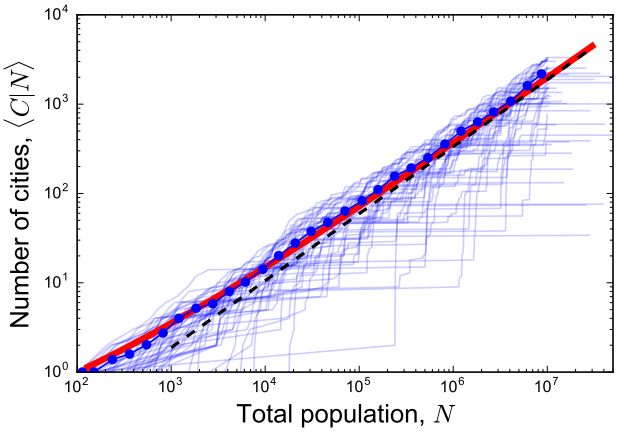

FIGURE 2.5. Relationship between the expected number of cities and the total population for the null model, which assumes a Zipf law with exponent $\beta = 1.75$ and minimum city size $X = 10^2$ for the distribution of city sizes and no spatial correlations between cities. The thin blue lines are 100 realisations of the model and the blue circles denote the average number of cities, $\langle C|N \rangle$, for a fixed value of the total population, N. The solid red line is the theoretical prediction of Equation 2.17 and the dashed black line is Heaps' law, $\langle C|N \rangle \sim N^{\beta-1}$, which holds in the limit of very large population $N \gg X$

$p(x|C,N)$. The average population of these C cities is given by N/C, therefore

$$N = C \cdot \langle x|C,N \rangle \tag{2.14}$$

where $\langle x|C,N \rangle$ corresponds to the conditional expectation of x given C and N. If we multiply this by $p(C|N)$ on both sides and integrate with respect to C we obtain:

$$N = \int dC p(C|N) C \int dx x p(x|C,N). \tag{2.15}$$

If the probability to find a city with population x in a group of C cities is assumed to be independent of the number of cities, ie. $p(x|C,N) \approx p(x|N)$ then we can say that the expected number of cities C in a region of total population N is $\langle C|N \rangle \approx N/\langle x|N \rangle$. We have already assumed that city sizes follow Zipf's law, with exponent $\beta < 2$, we also know that the maximum city size cannot be greater than the total population of the region. Using this, we can therefore write:

$$p(x|N) = \frac{\beta - 1}{X^{1-\beta} - N^{1-\beta}} x^{-\beta} \mathscr{H}(N - x), \tag{2.16}$$

where $\mathscr{H}(\cdot)$ is the Heaviside step function and X is a minimum city size. From this and Equation 2.15 we obtain an equation relating the average number of cities to the total population:

$$\langle C|N \rangle \approx \frac{2-\beta}{\beta - 1} \cdot N \frac{(X^{1-\beta} - N^{1-\beta})}{(X^{2-\beta} - N^{2-\beta})}. \tag{2.17}$$

Continent	β	σ_β	γ	σ_γ
Africa	1.77	0.05	0.78	0.09
Asia	1.94	0.04	0.85	0.05
America	1.96	0.04	0.97	0.07
Europe	2.02	0.05	1.01	0.06

TABLE 2.1. Exponents β of Zipf's law and γ of Heaps' law for Africa, Asia, America and Europe. Values σ_β and σ_γ correspond to the standard deviation in the exponents of Zipf's and Heaps' laws respectively. We observe that for Europe and America the expected relationship $\beta = \gamma + 1$ is satisfied within 1 standard deviation. Exponents β and γ were fit to the data using non-linear least squares.

Noting that $\langle C|N\rangle$ is a function of the ratio N/X, when the population is large, $N \gg X$ we obtain Heaps' law: $\langle C|N\rangle \sim (N/X)^{\beta-1}$. Figure 2.5 shows 100 realisations of the null model and the theoretical prediction given by Equation 2.17 for $\beta = 1.75$.

2.7.2 Empirical Results

Initially, we wish to assess whether Heaps' law holds for the four most populated continents: Asia, Africa, America and Europe. In order to do so, we analyse the *Geonames* data [130], grouping countries by their associated continent. This dataset contains the population of all cities with 1,000 or more inhabitants globally. Information relating to the location of each city, such as the latitude and longitude, country, state and county is also included in the dataset. Data on the area and population of all countries was obtained from Worldbank [127].

Firstly we verify the existence of Zipf's law for each continent. For all countries in each continent we find that the city size distribution, for each continent, follows Zipf's law above a minimum city population of $\sim 10^5$, as shown in Figure 2.6**a** where the darker parts of the distribution correspond to city sizes above this threshold. Fitting a power law to the tails of each city size distribution, we obtain a measure of the Zipf exponent β. Exponents β of the PDFs are displayed in Table 2.1. We find that, while Europe and America both satisfy Zipf's law, ie. their exponents are given by $\beta = 2$ within errors, Asia and Africa have exponents smaller than 2. Our power-law fitting is implemented using non-linear least squares. The population at which the distribution of city sizes follows Zipf's law (Figure 2.6**a**), is 10^5 for consistency across all continents. The validity of selecting this population was confirmed by minimising the Kolmogorov-Smirnov distance between the data and the fitted power-law, as suggested in [30]. This method gives values for the minimum population in the range $[47000, 55000]$. The values of Zipf's exponent β for the continents using these values are consistent with those obtained using a minimum city population of 10^5 for all continents.

Following this we assess whether Heaps' law holds for all countries in our four continents. To do this, for each country within a continent we calculate the number of cities with population $x > X$, where X is the minimum city population. We then fit a power law to a scatter plot of

Continent	Population	Density	Area
Africa	0.84	0.04	0.45
Asia	0.95	-0.09	0.45
America	0.99	-0.23	0.82
Europe	0.95	0.10	0.67

TABLE 2.2. The correlation coefficients between the number of cities and the population density, area and total population for countries in Africa, Asia, America and Europe.

the total population of a country vs the number of cities for which $x > X$, thus obtaining Heaps' exponent γ. From Table 2.1 we observe that Heaps' law, $C \sim N^{\beta-1}$ is satisfied, within errors, by all continents. This is further demonstrated in Figure 2.6**d**.

Examples of the scatter plots used to fit Heaps' law for each continent are given in Figure 2.6**b,c,e,f**. Here, the total population of a country (x-axis) is plotted against the number of cities with population $x > 10^5$. The marker size for each country corresponds to the logarithm of the area and the colour corresponds to the country's population density. From these plots it is clear that the number of cities does not scale with a country's area or density; in order to determine the number of cities in a region it is the total population that is the informative feature. The correlation coefficients for cities-density, cities-area and cities-population are displayed in Table 2.2. For figures displaying the number of cities vs country density and area see A.4.

2.7.3 Heaps' Law for States

As aforementioned, the presence of Zipf's law and Heaps' law at the continental scale does not imply its presence at a smaller scale where regions can be more homogeneous. To assess the presence of Heaps' law as a smaller scale we look as states within the United States. Figure 2.7**a** provides clear evidence that the number of cities grows proportionally to the state population. Furthermore, it is clear that there is at most only a small relationship between the number of cities and the area or population density of a state: for this data, the correlations coefficients for cities-population, cities-area and cities-density are given by 0.95, 0.04 and -0.08 respectively.

Combining our results with Zipf's law we obtain an expression for $C(N,X)$, the number of cities above a minimum city population X given a total population N: $C(N,X) = C(N)P(>X) \sim N/X$. This is equivalent to Heaps' law, with $\gamma = 1$. In Figure 2.7**b** we plot the number of cities with more than X inhabitants as a function of the ratio N/X for values of X ranging from $5,000$ to $5,000,000$ inhabitants. As all points collapse on the same line, this confirms that the Equation $C(N,X) \sim N/X$ holds over several orders of magnitude of N and X. We find further evidence for the validity of Heaps' law throughout time in the state of Iowa in the United States. Historical data shows that between 1850 and 2000 the number of incorporated places (i.e. self-governing cities, towns or villages) grew at the same rate as the state population. This result is displayed in Figure 2.7**c**.

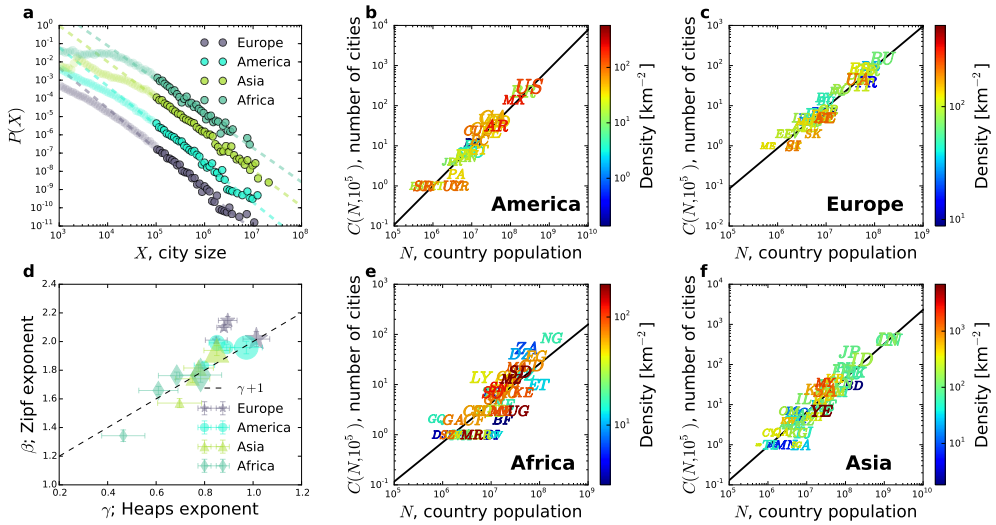

FIGURE 2.6. **a**, Zipf's Law: PDF (y-axis) of city sizes X (x-axis) for all cities in Europe, America, Asia and Africa. The darker regions correspond to cities with population $X > 10^5$, above which the distributions are a power law with exponent β given in Table 2.1. The dashed lines correspond to $y = x^{-\beta}$. Distributions have been shifted in the y-axis for clarity. **b-c** and **e-f**, Heaps' law for America, Europe, Africa and Asia. The following information is displayed for each country: population (x-axis), number of cities with more than 100k inhabitants (y-axis), 2-letter country code (marker), logarithm of the area (marker size) and population density (color). The black line is a power law fit of the scaling relationship between the number of cities and the total population; Heaps exponents γ are reported in Table 2.1. **d**, The exponent of the Zipf PDF, β (y-axis) and the corresponding exponent γ of Heaps' law for Europe, America, Asia and Africa. Marker size corresponds to the minimum city population used in determining the values of γ and β: values used are $10^3, 5\times10^3, 10^4$ and 10^5, where 10^3 is represented by the smallest marker and 10^5 by the largest. The black dashed line corresponds to the relationship between the exponents, $\beta = \gamma + 1$.

2.7.4 Spatial distribution of cities

If cities are randomly and uniformly distributed in space, the average number of cities in a region with uniform population density [1] is proportional to the region's area, or equivalently the density of cities scales as $\chi \sim C/A$. Combining this result with Heaps' law, $C \sim N/X$, and observing that the average distance to the closest city, $\langle d_c \rangle$, scales as the inverse square root of the density of cities, we obtain the following relation:

$$\langle d_c \rangle \sim 1/\sqrt{\chi} \sim \sqrt{A/C} \sim \sqrt{X(A/N)} \sim \sqrt{X/\rho}\,. \tag{2.18}$$

[1] If measured on a length scale larger than the average distance between cities

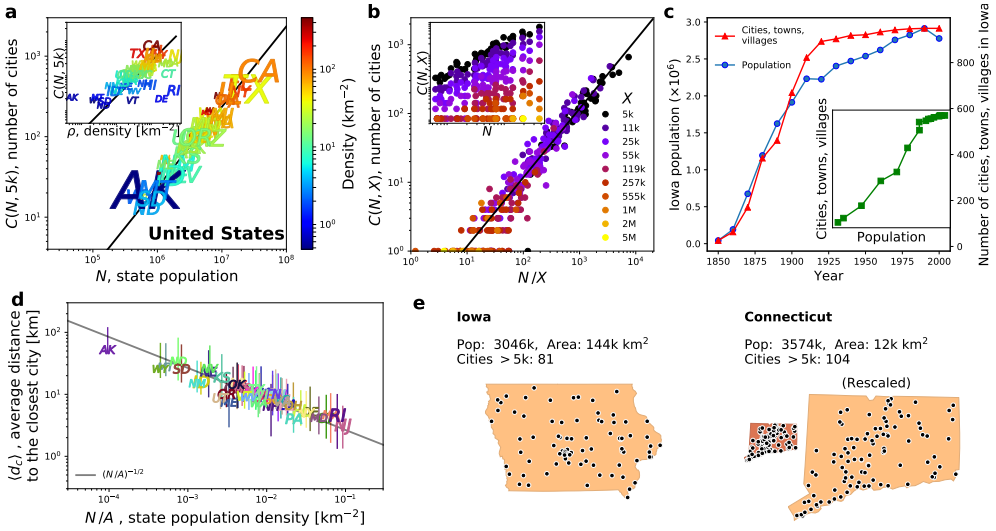

FIGURE 2.7. **a**, The number of cities with more than 5,000 inhabitants in the United States is proportional to the state's population, $\mathrm{corr}(C,N) = 0.95$. The correlations with area (0.04) and population density (−0.08, see inset) are negligible, as illustrated by the following pairs of states with similar area or density and very different number of cities: Alaska ("AK": $A = 1.5\text{M km}^2$, $C(5k) = 22$) vs Texas ("TX": $A = 0.7\text{M km}^2$, $C(5k) = 392$), and Rhode Island ("RI": $\rho = 393\ \text{km}^{-2}$, $C(5k) = 35$) vs New Jersey ("NJ": $\rho = 467\ \text{km}^{-2}$, $C(5k) = 316$). **b**, Combining Heaps' law with Zipf's law it is possible to estimate the number of cities with more than X inhabitants in a country with population N as $C(N,X) \sim N/X$. As a consequence, the scattered cloud of points resulting when plotting $C(N,X)$ against N for various X's in the range $5 \cdot 10^3 - 5 \cdot 10^6$ (inset) collapses on a straight line when $C(N,X)$ is plotted against the ratio N/X. **c**, Historical records of the number of incorporated places (C, red triangles) and the state population (N, blue circles) in Iowa from 1850 to 2000 [123]. The similar growth rates of C and N entail the validity of the first law of urbanisation $C \sim N$ during the 150-year period (inset). **d**, The average distance to the closest city in the United States scales as the inverse of the square root of the state's population density (here all cities with more than 5,000 inhabitants are considered). The asymmetric error bars denote the standard deviations above and below the average. **e**, Illustration of the relationships between total population, number of cities, and their average distance in Iowa and Connecticut. In agreement with Heaps' law, $C \sim N^{\beta-1}$, Iowa and Connecticut have similar populations and a similar number of cities with more than 5,000 inhabitants, despite Connecticut having one-twelfth the area of Iowa. In agreement with cities in Connecticut are closer than cities in Iowa because of the higher population density in Connecticut. By rescaling distances such that Connecticut's area becomes equal to Iowa's area, the two states would have the same population density and consequently the same average distance between cities.

We assess the validity of this assumption using the *Geonames* dataset. Figure 2.7**d** shows the average distance to the closest city with more than $X = 5000$ inhabitants for states in the United States, as a function of the states' population density. We observe that the average distance to the closest city is inversely proportional to the square root of the population density, thus confirming the relationship of Equation 2.18.

This finding supports some of the conclusions of the Central Place Theory of human geography, whilst disproving others [29]. On the one hand, it is true that for regions with a given population density the larger the cities are, the fewer in number they will be, and the greater the distance, i.e. increasing X in Equation 2.18 results in a greater average distance $\langle d_c \rangle$. On the other hand, the average distance between cities of a given size X is not the same for all the states, but depends on the state's population density; cities of a given size are closer in densely populated states than in sparsely populated ones, i.e. for a fixed city size X and state area A the distance between cities decreases as the inverse square root of the state population, N. This result is demonstrated in Figure 2.7**e** for the states of Iowa and Connecticut in the United States. Here we observe that while both states have a similar total population and number of cities with over $5,000$ inhabitants, Connecticut has a much smaller area. As a result, the average distance to the closest city in this state is much less than in Iowa. However, if the area of Connecticut is rescaled such that it is equal to that of Iowa, the average distance to the closest city becomes comparable to Iowa's.

2.8 Heaps' law for urban clusters

To understand how Heaps' law depends on the definition of a city, we analyse data from the Global Rural-Urban Mapping Project [27] (GRUMPv1) consisting of estimates of the population of the United States for the year 2000 at a resolution of 30 arc-seconds ($\sim$ 1km). In particular, we wish to determine the relationship between the area and population of urban clusters and whether the area of urban clusters, a, is also valid as the relevant size variable for Zipf's and Heaps' laws. Indeed, the PDF of the distribution of urban areas is also known to follow Zipf's law with exponent $\beta \simeq 2$ [82], hence our null model predicts that the number of urban clusters is given by Equation 2.17, where N now denotes the total urbanised area and $\beta - 1$ the exponent of the counter cumulative distribution function (CCDF) of city areas.

In the GRUMP data the spatial distribution of population is represented as a matrix, whose elements denote the estimated number of individuals resident within each of the grid cells. We apply a city clustering algorithm [107] (CCA) to the GRUMP data and define cities as spatial clusters of neighbouring grid cells with population over a given threshold, X, which also corresponds to the minimum cluster population. Initially we consider how the number, size and areas of cities in a fixed region depends on the minimum population X. We vary the parameter X over the interval [10-600] persons per km^2, clustering adjacent cells with population above

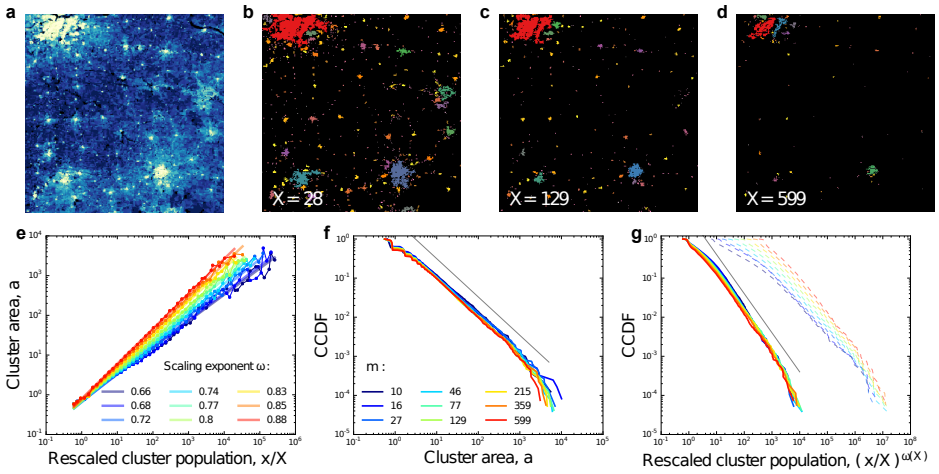

FIGURE 2.8. **a**, Portion of the GRUMP dataset representing the distribution of population in a region of the Midwest US. The color denotes the logarithm of the population: light yellow for high population, dark blue for low population. **b-d**, Urban clusters in the region depicted in panel a obtained applying the City Clustering Algorithm for different values of the minimum population parameter: $X = 28$ in b, 129 in c, 599 in d. **e**, Scaling relationship between area and population of clusters. The different colors denote different values of the CCA parameter X (see the legend of panel f for the X values). The points indicate the average area of clusters with a given rescaled population. Data are fit to the power law in Equation 2.19 and the legend reports the values of the scaling exponent ω for the various X. **f**, Counter Cumulative Distribution Functions of cluster areas. The different colors denote different values of the CCA parameter X (see the legend for the X values). The grey line is a power law with exponent -1 as a guide for the eye. **g**, Counter Cumulative Distribution Functions of cluster populations, x (dashed curves), and rescaled populations, $(x/X)^{\omega(X)}$ (solid curves). The grey line is a power law with exponent -1 as a guide for the eye.

the threshold X. As a reference, the official definition of urban area adopted by the US census considers values of X between 193 and 386 people per square kilometer [103]. In the range of X considered, the numbers and sizes of clusters obtained with the CCA are very different. Panels b-d of Figure 2.8 show the clusters within a square region in the Midwest (Figure 2.8**a**) for $X = 28, 129$, and 599. Both the number and area of clusters decrease as X increases and some large clusters split into multiple smaller clusters. This suggests that the exponents of both Zipf's law and Heaps' law may depend on the definition of a city being used and our minimum population X.

The area and population of urban clusters are strongly correlated variables. Measuring the scaling relationship between the area, a, and population, x, of the clusters provides a means of characterising the expansion of urban areas. We use the gridded population data to measure

urban sprawl for different definitions of city, i.e. different values of the CCA parameter X (see Figure 2.8**e**). We observe that the scaling relationship between a and x has the following dependence on the minimum population parameter X:

$$a \sim (x/X)^{\omega(X)} \tag{2.19}$$

Note that the area of a cluster scales with the ratio x/X. This can be interpreted as the maximum area that a cluster of population x can have, given X. The scaling exponent ω depends on X. In particular, $\omega(X)$ is an increasing function of X, which grows from 0.66 to 0.88. The sub-linear scaling between a cluster's area and population, $\omega(X) < 1$ for all X, implies an increase in the population density of large clusters: the population density scales as $x/a = x^{1-\omega}$, which is a growing function of x when $\omega < 1$. This is a result of the economies of scale in the use of urban space. In large clusters space is organised more efficiently than in small clusters, so that each square kilometre of land can host a larger number of individuals, hence increasing the cluster's population density [23]. Urban sprawl happens when the exponent ω has a large value, indicating a reduced efficiency in the utilisation of space as the size of clusters grows. The fact that ω increases with X means that the estimated urban sprawl is bigger when clusters are defined using a large X and smaller when X is small.

The scaling relationship between area and population of clusters, Equation 2.19, implies that the Zipf exponents of the distributions of cluster areas and populations, β_a and β_x respectively, are not independent, but related by the equation

$$\beta_x - 1 = (\beta_a - 1) \cdot \omega(X). \tag{2.20}$$

The empirical distributions of cluster areas for different values of the CCA parameter X, shown in Figure 2.8**f**, indicate that the Zipf exponent for the areas is $\beta_a - 1 \simeq 1$, independent of X. The distributions of cluster populations, instead, have exponents that depend on X. If the populations are rescaled by X and elevated to the power of $\omega(X)$, the curves for different X collapse on the same power law with exponent $\beta_x - 1 \simeq 1$, verifying the relationship $\beta_x - 1 = (\beta_a - 1) \cdot \omega(X)$ (see Figure 2.8**g**).

2.8.1 Local distributions of areas and populations of urban clusters

To further understand how the number, areas and populations of clusters depend on the CCA parameter X, we perform a systematic analysis of the GRUMP data, considering regions at a smaller spatial scale. We divide the US into non-overlapping square regions of size $L = 128$ km, and obtain urban clusters in each region by applying the CCA for all values of X between 10 and 600. We use geographical areas of $128km^2$ to ensure that each region contains enough clusters to compute the distribution of cluster sizes and areas. In order to avoid finite-size effects, we only consider regions with low urbanisation, having a percentage of built-up area smaller than 5% (this condition is satisfied by 49% of the regions for m = 10 and up to 93% for m = 599). We

group together regions with similar total population, N_x, and built-up area, N_a, and compute the distributions of cluster sizes (i.e. populations and areas) separately for each each group (see Figure 2.9**a,d**). If the assumptions of our null model hold, the Counter Cumulative Density Functions (CCDFs) of cluster areas and populations should be truncated power laws and have the forms

$$P(>a|N_a) \sim a^{1-\beta_a} f_a(a/N_a) \tag{2.21}$$

$$P(>x|N_x) \sim (x/X)^{1-\beta_x(X)} f_x(x/N_x), \tag{2.22}$$

where f_a and f_x are scaling functions that rapidly go to zero when their argument is larger than 1, to account for finite-size effects. We find that the assumptions of the null model hold: the local distributions of city sizes are power laws with the same exponent as the global distribution at the country scale, with cutoffs that account for the finite sizes of the regions. The scaling collapses shown in Figure 2.9**b,e** provide a validation to the predicted functional forms of the CCDFs, Equations 2.21 and 2.22.

2.8.2 Heap's law at the local scale

Finally, we wish to verify that the average number of clusters is related to the total size of the region as predicted by the null model (Equation 2.17). This means that cities are randomly distributed among the regions, even at small spatial scales. For each group of regions with similar total population, N_x, and built-up area, N_a, we compute the average number of clusters for all values of the CCA parameter X, $\langle C|N_a,X\rangle$ and $\langle C|N_x,X\rangle$, and we check if these empirical values are compatible with the estimates of the null model. To this end we perform 100 realisations of the null model, sampling cities with replacement from the list of all the empirical values of cluster areas, and we repeat the same procedure for the cluster populations. For each value of N_a and N_x, we compute the confidence intervals for the null model, defined as the 10th and 90th percentiles of the number of clusters in the 100 realisations. We observe that the empirical estimates of the average number of clusters lie within the null model's confidence intervals (see Figure 2.9**c,f**), confirming that empirical data is compatible with a random distribution of clusters within the regions.

2.9 Conclusion

In this chapter we presented a deterministic model of population dynamics. This model incorporated a logistic growth of the total population and either a gravity model or intervening opportunities model to describe the movement of individuals between cities. We demonstrated analytically that if individuals relocate according to a gravity model, the number of cities in a region is independent of the region's population. In contrast, if individuals relocate according to

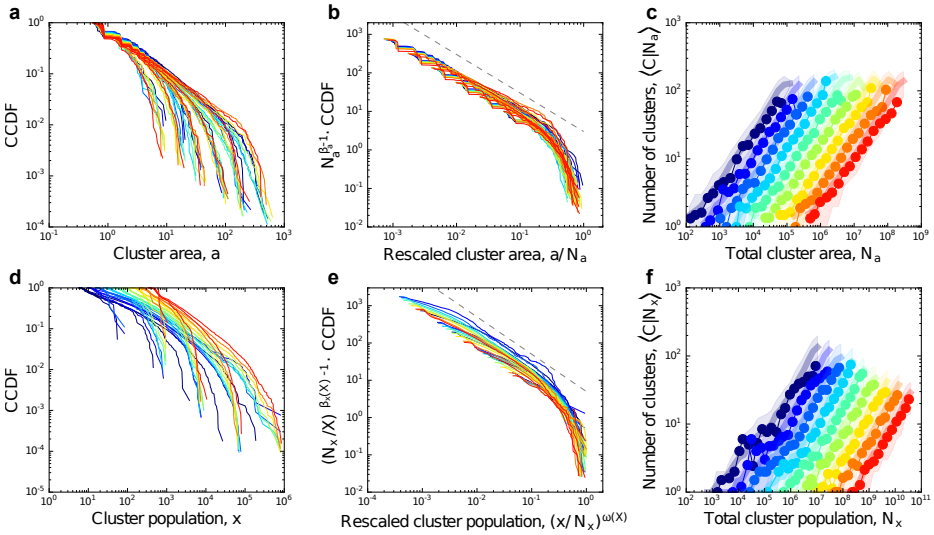

FIGURE 2.9. **a**, Counter Cumulative Distribution Functions of areas of the clusters in all US regions of 128×128 km^2 having urbanised area up to 5%. Regions are grouped in six groups according to their total urbanised area, N_a, and the CCDFs of each group are computed separately for the different values of the CCA parameter X (see the legend in Figure 2.8**f** for the X values). **b**, The CCDFs of panel a collapse on the same curve when the axes are properly rescaled. The dashed grey line is a power law with exponent -1 as a guide for the eye. **c**, Average number of clusters as a function of the total urbanised area, N_a, for the 128×128 km^2 US regions (circles). The lower and upper values of the dashed areas denote the 10th and 90th percentile of 100 realisations of the null model. For clarity, curves have been shifted by X^2 along the x-axis. **d**, Counter Cumulative Distribution Functions of populations of the clusters in all US regions of 128×128 km^2 having urbanised area up to 5%. Regions are grouped in six groups according to their total population, N_x, and the CCDFs of each group are computed separately for the different values of the CCA parameter X. **e**, The CCDFs of panel d collapse on the same curve when the axes are properly rescaled. The dashed grey line is a power law with exponent -1 as a guide for the eye. **f**, Average number of clusters as a function of the total population, N_x, for the 128×128 km^2 US regions (circles). The lower and upper values of the dashed areas denote the 10th and 90th percentile of 100 realisations of the null model. For clarity, curves have been shifted by X^2 along the x-axis.

an intervening opportunities model, the number of cities in a region is proportional to the total population of that region.

To formalise these results, we introduced the concept of Heaps' Law of cities to describe how the number of cites within a region depends on the total population of the region. Using the Geonames dataset we found that in the context of cities, Heaps' Law has an exponent $\gamma = 1$; the number of cities within a region is proportional to the total population of that region. Following on from Heaps' Law, we also investigated how the spatial distribution of cities depends on the population of a region. We demonstrated analytically, using Heaps' Law, that if cities are randomly distributed in space, the average distance to the closest city must scale as the square root of the inverse of the population density, Equation 2.18. Using the Geonames dataset we validated this result for states within the United States. This result suggests that cities are randomly distributed in space; spatial interactions between cities are not strong enough to produce spatial correlations. Finally we assessed whether the presence of Heaps' law depends on the definition of a city. Using the GRUMP dataset we applied a city clustering algorithm to obtain urban clusters with a minimum population X. We found that Zipf's law for urban areas is independent of parameter X and has exponent $\beta_a - 1 \sim 1$. In contrast, Zipf's law for urban populations does depend on X and the exponent $\beta_x - 1$ is given by $(\beta_a - 1) \cdot \omega(X)$. Rescaling cluster populations by X and raising to the power $\omega(X)$, Zipf's law for cities is obtained with exponent $\beta \sim 2$. Following on from this we confirmed that Heap's law successfully describes the distribution of the number of urban clusters as a function of the total population, N_x, and built-up area, N_a, of a region. Further evidence for both Heaps' Law for cities and the spatial distribution of cities can be found in [115].

Our empirical investigation suggests that intervening opportunities rather than geographic distance may be the better variable at describing migration flows; Heaps' exponent $\gamma = 1$ and the number of cities increases with the total population of a region. However, our deterministic model is unable to reproduce Zipf's Law for city sizes at it neglects the inherent randomness of stochastic models required to reproduce this phenomena. As a consequence, our modelling framework fails to properly describe key statistical patterns of demographic dynamics; a general framework must reproduce both Zipf's Law and Heaps' Law in order to be compatible with patterns observed in data. We have designed and assessed a stochastic adaptation of our model [115] however further work is required. A description of the stochastic model can be found in Appendix A.3. In particular, the linear stability analysis, Appendix A, performed on both models of migration is only valid for small perturbations about the stationary distribution. With regards to the simulations, this means that the conditions for the number of cities (Equation 2.9) are only valid during the initial stages of the simulation; at later times the effects of the non-linearities present in Equation 2.1 become more pronounced. From the linear stability analysis it was concluded that the growth and migration rates affect whether or not cities may form from the initial perturbation (Equations A.8 and A.20) however the number of cities that form depend on the parameters ρ_0, R and w

only. Numerical simulations indicate that on longer time scales, the number of cities that are present also depends on the growth rate g and the migration parameter T [115]. In particular the number of cities is constant for a given ratio T/g, however if the ratio is varied, by changing either g or T, the number of cities also changes.

In order to find empirical confirmation for the model's predictions, we would need a complete series of historical data on both growth and migration rates. While it is possible to estimate the growth rate of each state within the US since 1790 using census data, we were not able to determine the migration rates of the states prior to 1940, the first year in which the US census included questions about people's mobility [101]. Due to the lack of historical data on migration rates prior to 1940 we are unable to estimate the ratio T/g of each state, hence precluding the possibility to test the presence of a correlation between the ratio T/g and the number of cities. The limitations of our deterministic model of population dynamics suggest that a description of urbanisation based on stochastic processes might be more appropriate in order to capture both Heaps' law and Zipf's law for cities.

CHAPTER 3

STOCHASTIC FLUCTUATIONS IN CITIES

3.1 Introduction

Driven by increasing global population and urbanisation, recent years have seen a wealth of research into how cities evolve. With regards to city growth, a general model has emerged that accounts for how the population of cities as a whole changes in time; Gibrat's law, or proportionate random growth. Despite its wide acceptance, a general consensus on how the underlying stochastic processes within a city, namely births, deaths and migrations, interact to give rise to this growth mechanism is lacking. In particular, and as aforementioned, existing microscopic models of population dynamics that are able to account for Zipf's Law for the distribution of city sizes are only able to do so for specific values of their parameters.

In this chapter we present a microscopic model of births and deaths that does not require fine-tuning of parameters and is able to reproduce both Zipf's Law for the size distribution of cities and the fluctuation scaling in population sizes associated with Gibrat's Law (Equation 1.5). The assumptions of the central limit theorem, that demographic processes have uniform and constant rates and that individuals are independent and identical, are almost never satisfied in natural systems; empirically, people are essentially different from one another and they interact with each other. To relax these assumptions, we consider processes with environmental variability, section 3.2.1, and correlations between individuals, section 3.2.2.

Environmental variability is easily justified by the development of a country; as access to healthcare improves, birth rates and death rates will fall. Correlations in birth and death rates can occur due to political and economic effects: post World War II there was a large increase in birth rates attributed to improving economies and veterans returning home [105].

In section 3.2 we show that if the assumptions of the central limit theorem are relaxed,

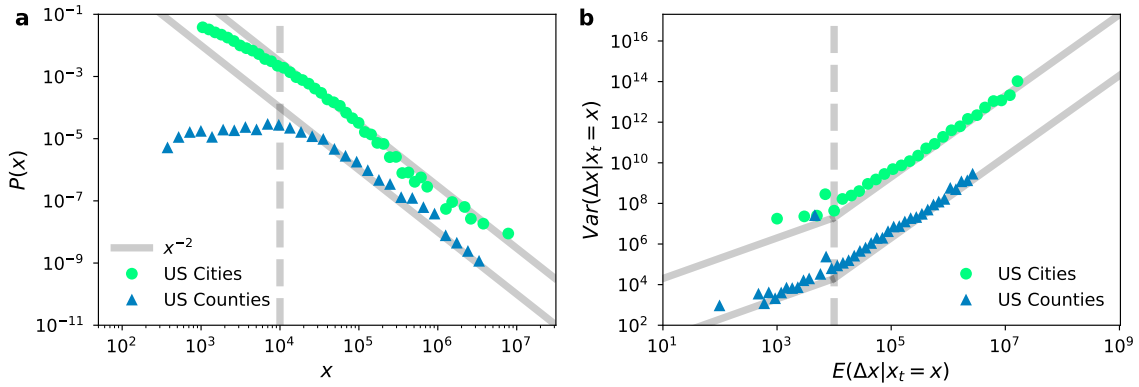

FIGURE 3.1. **a** Zipf's law, data. Probability density function $P(x)$ (y-axis) vs size x (x-axis) for cities (circles) and counties (triangles) within the United States. The solid lines are guides for the eye corresponding to $P(x) \propto x^{-2}$. Data on the population of cities is obtained from the *Geonames* dataset [130]. County level data is obtained from the US Census Bureau [129]. The distribution for US cities has been shifted for clarity. **b** Taylor's law, data. The variance of population in year $t+1$ conditioned to the population in year t (y-axis) vs the average population in year $t+1$ conditioned to the population in year t (x-axis) for cities (circles) and counties (triangles) in the United States during the period 1970 to 2010. The vertical dashed line denotes the cross-over city size x_c at which Taylor's exponent defined in Equation 3.3 transitions from $\alpha = 1/2$ to $\alpha = 1$ (solid lines); as shown in panel a), x_c corresponds to the cross-over city size at which the distributions start following Zipf's law. The distribution for US cities has been shifted for clarity.

that is birth and death rates are not either independent or identically distributed but rather realistic features like temporal variability and dependence between individuals are considered, the resulting processes are able to reproduce Zipf's law with exponent $\beta = 2$ without fine tuning. These processes are characterised by an anomalous scaling of the temporal fluctuations of the population increments, commonly known as Taylor's law in ecology [35, 39, 54];

$$\sigma_x^2 \propto \langle x \rangle^{2\alpha}. \tag{3.1}$$

The exponent α often takes a value of either 1/2 or 1 with the latter corresponding to a random multiplicative process (Figure 3.1**b**, Equation 1.5). Taylor's law, specifically temporal fluctuation scaling, appears in many diverse areas. For example: fluctuations in the weekly traffic of internet routers have exponent $\alpha = 0.75$; the size of stock market transactions have exponent $\alpha = 0.69$; the yearly reproductive variability of trees has an exponent $\alpha = 1$ [39].

We demonstrate that a microscopic process of individual population dynamics where the variance of the population increments scales as the square of the population and hence has Taylor's exponent ~ 1 has a stationary distribution that follows Zipf's law. This connection is

empirically investigated in section 3.3 using data on the population of the largest cities, $x \geq 3 \times 10^5$, globally. We demonstrate that countries which satisfy Zipf's law, $P(x) \sim x^{-2}$, also satisfy Taylor's law with exponent $\alpha \sim 1$. The work presented in this chapter has been carried out in collaboration with S. Azaele, A. Maritan and F. Simini. Specifically S. Azaele performed the initial calculations and F. Simini implemented the method of generating correlated variables that was used for Figure 3.3**c-d**.

3.2 The Microscopic Model

We present our microscopic model as one in which individuals belong to different cities.To describe how the population in each city evolves in time due to births and deaths we use a modified version of the Galton-Watson branching stochastic process. The original Galton-Watson process [57] is a discrete-time stochastic process describing the evolution of a population of x_t individuals at time t according to the equation

$$x_{t+1} = \sum_{i=1}^{x_t} z_i, \tag{3.2}$$

where z_i are independent and identically distributed random variables over the integers, with finite mean and variance. The probability density function for these variables, z_i, is given by $P(z|\lambda)$, where λ corresponds to the parameters. At each time step, the population x_t can change in one of three ways. If $z_i = 0$, individual i dies, if $z_i = 1$, they do nothing and if $z_i = 2,3,4...$ then they have $1,2,3...$ children.

In the original Galton-Watson process, extinction will occur with probability 1 if the average number of offspring per individual is less than or equal to one, $\mathbb{E}(z) \leq 1$. To avoid extinction we include the boundary condition $x_t \geq 1$ for all t, which accounts for immigration. Modelling immigration using a boundary condition is a solution adopted in several models of population dynamics [8]. We verified that the specific implementation of the boundary (as a hard reflecting boundary or as a constant influx of individuals) and its value does not affect our results. If $\mathbb{E}(z) > 1$, then the population will experience an exponential growth and there will be no stationary state. For $\mathbb{E}(z) \leq 1$, the stationary distribution does not follow Zipf's law; it is a power law with exponent -1.

The conditional mean and variance of the population increments, defined as $\mathbb{E}(\Delta x|x_t = x)$ and $Var(\Delta x|x_t = x)$ respectively, where $\Delta x = x_{t+1} - x_t$, can be used to measure the scaling of the fluctuations in the population size. Given that the random variables z_i are independent and identically distributed with finite mean and variance, the conditional mean and variance of population increments both scale as x: $\mathbb{E}(\Delta x|x) \sim x$ and $Var(\Delta x|x) \sim x$.

This scaling behaviour can be summarised considering the relationship between these two quantities, i.e. Taylor's law:

$$Var(\Delta x|x) \propto \mathbb{E}(\Delta x|x)^{2\alpha}. \tag{3.3}$$

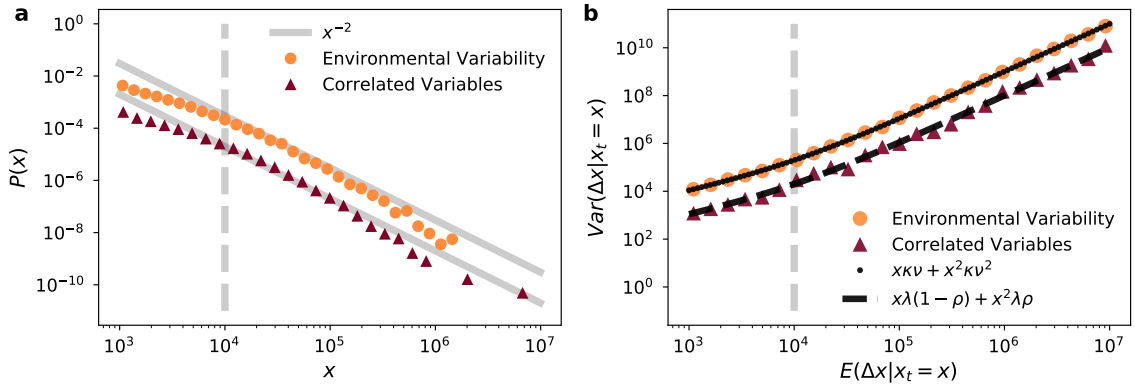

FIGURE 3.2. **a** Zipf's law, models. Stationary distributions of city sizes $P(x)$ for modified branching processes with environmental variability (circles) and correlated individuals (triangles). See the main text for details on the numerical simulations and the parameter values. The distributions have been shifted along the y-axis for clarity. The vertical line denotes the cross-over city size x_c at which the distributions start following Zipf's law: $x_c = 10^4$ for both models. **b** Taylor's law, models. The variance of population change in time interval $[t, t+1]$ conditioned to the population at time step t (y-axis) vs the average population change in time interval $[t, t+1]$ conditioned to the population at time step t (x-axis) for branching processes with environmental variability (circles) and correlated variables (triangles). Simulations and parameter values are the same as in a). We observe a transition of Taylor's exponent from $\alpha = 1/2$ for $x < x_c$ to $\alpha = 1$ for $x > x_c$. The black lines correspond to the analytical results of Equation 3.5 and Equation 3.12. Curves have been shifted along the y-axis for clarity.

Taylor's law usually refers to the scaling of the (unconditioned) variance versus the mean of a random variable [35, 39, 54]. Here we extend it to the case of the scaling of the variance of the random variable increments versus the mean of the random variable, both conditioned to the value of random variable itself before the increment. As aforementioned, the exponent α often takes a value of either 1/2, as in the original Galton-Watson process, or 1, as in a random multiplicative processes. As random multiplicative processes are known to produce Zipf's law, this suggests Zipf's law will be present with exponent $\gamma = 1$ when Taylor's exponent is $\alpha = 1$. On the other hand, when Taylor's exponent is $\alpha = 1/2$ we hypothesise that Zipf's law will not be present. This suggests a connection between Taylor's law and Zipf's law, which we characterise in the following sections with analytical arguments and numerical simulations. Formalising this intuition, we propose two variations of the Galton-Watson process, namely processes with environmental variability and processes with correlated individuals, and show that for large populations they have Zipf exponent $\beta = 2$ and Taylor exponent $\alpha = 1$.

3.2.1 Environmental variability

We first present the case of processes with environmental variability. In order to introduce environmental variability we consider a modified Galton-Watson process where the parameters λ of the distribution of the individuals are not constant values but random variables drawn from a distribution $G(\lambda)$ at each time step. To be specific, we assume that the z_i are Poisson random variables with distribution $P(z|\lambda) = \text{Poiss}(\lambda) = e^{-\lambda}\lambda^z/z!$. We select a Poisson distribution due to its analytical amenability, however any discrete distribution with infinite support that satisfies $z_i \geq 0$ may be used. For further discussion see section 3.4.

In order to determine the fluctuations in the population increments, we compute the variance of the population change, $\Delta x = x_{t+1} - x_t$, conditioned to the population at time t by applying the law of total variance [134];

$$Var(\Delta x|x_t = x) = \mathbb{E}_\lambda[Var_{\Delta x|\lambda}(\Delta x)] + Var_\lambda(\mathbb{E}_{\Delta x|\lambda}[\Delta x]). \tag{3.4}$$

Here $\mathbb{E}_{\Delta x|\lambda}$ and $Var_{\Delta x|\lambda}$ denote the mean and variance with respect to $P(z|\lambda)$, and $\mathbb{E}_\lambda$ and Var_λ denote the mean and variance with respect to $G(\lambda)$. For a Poisson distribution, where both the mean and variance are given by parameter λ, we have that $Var_{\Delta x|\lambda}(\Delta x) = \mathbb{E}_{\Delta x|\lambda}[\Delta x] = x\lambda$. Substituting this into Equation 3.4 we obtain

$$Var(\Delta x|x_t = x) = x\,\mathbb{E}_\lambda(\lambda) + x^2\,Var_\lambda(\lambda). \tag{3.5}$$

Note that the fluctuations in the size of a city's population are proportional to x and follow Equation 3.3 with exponent $\alpha = 1/2$ for small populations, $x \ll x_c \equiv \mathbb{E}_\lambda(\lambda)/Var_\lambda(\lambda)$, whereas they scale as x^2 for large populations $x \gg x_c$. The crossover population x_c marks the transition between these two scaling regimes.

Empirical evidence of the presence of this crossover can be found analysing the fluctuations of the populations of cities and counties in the United States, where data shows a transition between Taylor exponents $\alpha = 1/2$ and $\alpha = 1$ around $x_c \sim 10^4$ (Figure 3.1**b**). It is important to note that the cross-over population we use, x_c, is only approximately 10^4 and is taken from Figure 3.1**b**: we observe that at this population, the fluctuation scaling is smoothly transitioning from a power law with exponent $\alpha = 1/2$ to a power law with exponent $\alpha = 1$. Within our framework, a cross-over population of $x_c \sim 10^4$ result means that on average the ratio between the variance and the mean of the growth rates is very small, around $x_c^{-1} = Var_\lambda(\lambda)/\mathbb{E}_\lambda(\lambda) \sim 10^{-4}$. For example, if the distribution of λ is a Gamma distribution, $G(\lambda) = \text{Gamma}(\lambda|\kappa, \nu)$, then Equation 3.5 becomes $Var(\Delta x|x_t = x) = x\kappa\nu + x^2\kappa\nu^2$ and $x_c = \nu^{-1}$. We verify the validity of this prediction with numerical simulations of Equation 3.2 shown in Figure 3.2**b**, where the z_i are independent and identical Poisson random variables with distribution Poiss(λ) and a new parameter λ is drawn at each time step from a Gamma distribution with fixed parameters $\kappa = 10^4$ and $\nu = 10^{-4}$.

Next we demonstrate that the stationary distribution of city sizes follows Zipf's law for large populations, as shown empirically in Figure 3.1**a**, numerically in Figure 3.2**a** and summarised

Distribution	γ	D	p
US Counties	0.86	0.04	<0.00
US Cities	0.96	0.07	0.01
Environmental variability	0.91	0.02	<0.00
Correlated individuals	1.00	0.01	<0.00

TABLE 3.1. Exponent γ for the distribution of city sizes obtained using a Maximum likelihood estimate, Kolmogorov-Smirnov statistic D and p-value for the distributions in Figures 3.1**a** and 3.2**a**. All numbers are rounded to two decimal places.

in Table 3.1. When the total population is sufficiently large, we expect that Equation 3.2 can be approximated as

$$\Delta x \equiv x_{t+1} - x_t \approx \hat{\mu}_x + \hat{\sigma}_x \xi(t), \tag{3.6}$$

where $\Delta x(t)$ is a continuous random variable, $\hat{\mu}_x = \mathbb{E}(\Delta x | x_t = x)$, $\hat{\sigma}_x = \sqrt{Var(\Delta x | x_t = x)}$, and $\xi(t)$ is a zero-mean Gaussian white noise with autocorrelation $\langle \xi(t)\xi(t') \rangle = 2\delta(t - t')$. Figures 3.3**a-b** demonstrate that numerical simulations support the validity of the ansatz of Equation 3.6.

Using the law of total expectation [134] and Equation 3.5, we obtain

$$\hat{\mu}_x = x(\mathbb{E}_\lambda(\lambda) - 1) \equiv x\hat{\mu} \tag{3.7}$$

$$\hat{\sigma}_x = \sqrt{x\mathbb{E}_\lambda(\lambda) + x^2 Var_\lambda(\lambda)} \equiv \sqrt{x\hat{\sigma}_1^2 + x^2\hat{\sigma}_2^2}. \tag{3.8}$$

Using the formal substitutions $t + 1 \to t + \mathrm{d}t$, $\hat{\mu} \to \mu \mathrm{d}t$, $\hat{\sigma_1} \to \sigma_1 \mathrm{d}t$ and $\hat{\sigma_2} \to \sigma_2 \mathrm{d}t$, expanding in Taylor series to first order in $\mathrm{d}t$, we obtain the following stochastic differential equation:

$$\dot{x}(t) = \mu x(t) + \sqrt{\sigma_1^2 x(t) + \sigma_2^2 x(t)^2}\,\xi(t), \tag{3.9}$$

with a reflecting boundary at $x = 1$. Solving the Focker-Planck equation for this stochastic differential equation gives us the stationary distribution of city sizes;

$$P(x) = \mathcal{N} \frac{(\sigma_1^2 + \sigma_2^2 x)^{\frac{\mu}{\sigma_2^2}}}{(\sigma_1^2 x + \sigma_2^2 x^2)}, \tag{3.10}$$

where $\mathcal{N}$ is a normalisation constant depending on parameters μ, σ_1 and σ_2. For large populations, $x \gg x_c = \sigma_1^2/\sigma_2^2$, Equation 3.9 becomes a random multiplicative process with growth rate of mean μ and variance σ_2^2. In this limit the stationary distribution of Equation 3.10 can be approximated as $P(x) \sim x^{-2+\frac{\mu}{\sigma_2^2}}$. From this we observe that, providing $|\mu| \ll \sigma_2^2$, the distribution of city sizes follows Zipf's Law with exponent $\beta = 2$. If the average city population is large this condition is always satisfied.

From Equation 3.10 we also observe that for small cities, with population $x \ll x_c$, the stationary distribution does not follow Zipf's Law; in this limit it can be approximated as $P(x) \sim x^{-1+\frac{\mu}{\sigma^2}}$. Thus, our microscopic model is able to account for both Zipf's Law in the tail of the distribution of city sizes and its absence in the distribution of smaller cities.

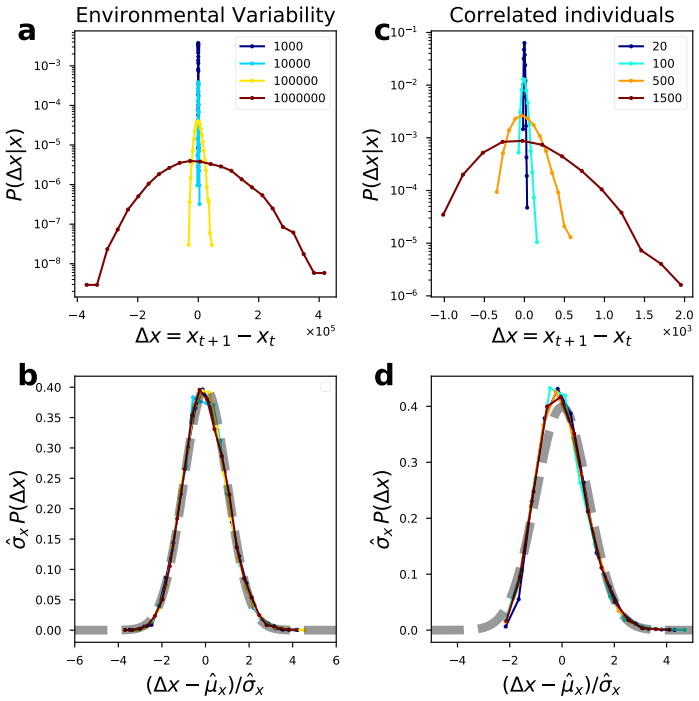

FIGURE 3.3. Numerical simulations support the ansatz of Equation 3.6 for processes with environmental variability (panels a,b), and with correlated individuals (panels c,d). **a** Distribution of the fluctuations of the population increments in consecutive time steps, $\Delta x = x_{t+1} - x_t$, for a process with environmental variability where $P(z|\lambda) = \text{Poiss}(\lambda)$ and $G(\lambda) = \text{Gamma}(\lambda|\kappa, \nu)$ with $\kappa = 100, \nu = 0.01$. Different curves correspond to different values of populations $x_t = x$. **b** The curves in panel **a** collapse on the same distribution, a standard normal distribution (dashed line), when the population increments are shifted by removing the mean and rescaled by the square root of the variance derived in Equation 3.5. **c** Distribution of the fluctuations of the population increments in consecutive time steps for a process with correlated individuals, where the n Poisson random variables have covariance matrix $Cov(z_i, z_j) = \delta_{ij}\lambda + (1 - \delta_{ij})\rho\lambda$ with $\lambda = 0.999, \rho = 0.1$. **d** The curves in panel **c** collapse on the standard normal distribution (dashed line) when the population increments are shifted by removing the mean and rescaled by the square root of the variance derived in Equation 3.12.

3.2.2 Correlated individuals

The second case we present is the class of processes where individuals are correlated. To introduce correlations between individuals we again consider a modified Galton-Watson process, where the joint probability $P(z_1,\ldots,z_x|\lambda)$ does not factorise into the product of the individual probabilities $P(z_i|\lambda)$. To be specific, we assume that the z random variables have a Poisson distribution with fixed parameter λ, which is the same for all individuals, and correlation matrix $\rho_{ij} = Cov(z_i,z_j)/\lambda$, where $Cov(z_i,z_j) \equiv \mathbb{E}[(z_i-\lambda)(z_j-\lambda)]$. When individuals are correlated the off-diagonal terms of the covariance matrix are non-zero. For simplicity we consider the simplest case where correlations between individuals are all equal;

$$Cov(z_i,z_j) = \delta_{ij}\lambda + (1-\delta_{ij})\rho\lambda. \tag{3.11}$$

The fluctuations of the population increments for the process with correlated individuals have the same scaling form of the fluctuations for the process with environmental variability. The conditional variance of $\Delta x = x_{t+1} - x_t$ for a given population $x_t = x$ is the sum of all the elements of the covariance matrix:

$$Var(\Delta x|x_t = x) = \sum_{i,j} Cov(z_i,z_j) = x\lambda(1-\rho) + x^2\rho\lambda \tag{3.12}$$

Here the cross-over population between the regimes with Taylor exponents $\alpha = 1/2$ and $\alpha = 1$ is $x_c = 1/\rho - 1$. As seen before, for cities and counties in the United States the transition between these two scaling regimes occurs at $x_c \sim 10^4$ (Figure 3.1**b**). Within the framework of a process with correlated individuals, this cross-over population corresponds to a very small value of correlation between people in a city or county, $\rho \sim 1/x_c \sim 10^{-4}$. The physical meaning of this correlation can be interpreted considering that the inverse of the correlation matrix corresponds to the interaction between individuals [7, 131]. Since the off-diagonal elements of the inverse of the covariance matrix vanish in the small ρ limit we get that only a small amount of interaction is needed for the fluctuation scaling in large cities to follow Taylor's law with exponent $\alpha = 1$, yet in the absence of correlations this will not be present. With regards to our modelling framework, as the cross-over population is inversely proportional to the amount of correlation, $\rho \sim 1/x_c$, any increase or decrease in the strength of the correlation will alter the exponent of the stationary distribution of city sizes. In the case of strong correlation, as ρ approaches 1, the cross-over population x_c also approaches 1: for $x \gg x_c = 1$ the distribution of city sizes can be described by Zipf's law with exponent $\beta = 2$ and the fluctuation scaling for all cities will follow Taylor's law with exponent $\alpha = 1$. In contrast, as the correlation approaches 0, such that individuals are non-interacting, $x_c \to \infty$: the stationary distribution of city sizes, up to population x_c, is a power-law with exponent $\beta = 1$ and fluctuations follow Taylor's law with $\alpha = 1/2$.

Using the ansatz of Equation 3.6 we follow the same steps taken for the case of environmental variability to demonstrate that the stationary distribution of city sizes is a power law. The validity of the ansatz for correlated individuals is once again supported by numerical simulations, Figures

3.3**c-d**. Using the law of total expectation and Equation 3.12 we obtain the mean and variance of a population change for correlated individuals:

$$\hat{\mu}_x = x(\lambda - 1) \equiv x\hat{\mu} \tag{3.13}$$

$$\hat{\sigma}_x = \sqrt{x\lambda(1-\rho) + x^2\rho\lambda} \equiv \sqrt{x\hat{\sigma}_1^2 + x^2\hat{\sigma}_2^2} \tag{3.14}$$

Again using the substitutions $t+1 \to t+\mathrm{d}t$, $\hat{\mu} \to \mu\mathrm{d}t$, $\hat{\sigma_1} \to \sigma_1\mathrm{d}t$ and $\hat{\sigma_2} \to \sigma_2\mathrm{d}t$ we obtain Equation 3.9, where the variables μ, σ_1 and σ_2 now correspond to those in Equation 3.14. As such, the stationary distribution of Equation 3.10 is also valid for the case where individuals are correlated. In this case, the stationary distribution of city sizes for large populations, $x > x_c$, is a power-law given by $P(x) \sim x^{-2+\frac{\lambda-1}{\rho\lambda}}$, that follows Zipf's law when $|\lambda - 1|/\lambda \ll \rho$. Furthermore, for small populations $x < x_c$, the size distribution of cities will not follow Zipf's law thus correlations between individuals are also able to account for both Zipf's Law in the tail of the distribution and its absence in the distribution of smaller cities. Numerical simulations support presence of Zipf's law in the distribution tail for cities with population $x > x_c$ and its absence in the distribution of smaller cities (Figure 3.2**a**).

In the numerical simulations, we used two methods to generate correlated Poisson variables. In the first method, used in the simulations of Figure 3.2, we consider the modified Galton-Watson process $x_{t+1} = \sum_{i=1}^{x_t}(z_i + k)$ where the z_i are independent and identical random variables with Poisson distribution Poiss$(\lambda(1-\rho))$ and we introduce the random variable k, independent of the z_i, with Poisson distribution Poiss$(\rho\lambda)$. One can show that for this process the conditional variance of Δx is identical to Equation 3.12. In Figure 3.2 we use $\lambda = 1$ and $\rho = 10^{-4}$. In the second method, used in the simulations of Figure 3.3**c-d**, we use a Gaussian copula model to link the Poisson marginals [13]. We verified with numerical simulations that for appropriate values of the multivariate Gaussian's parameters, this method generates Poisson variables with the desired values of λ and covariance matrix $Cov(z_i, z_j) = C_{ij}$; the conditional variance of Δx is again identical to Equation 3.12. Both methods give results compatible with the theoretical predictions.

3.3 Empirical evidence of Zipf's and Taylor's Laws

In order to further demonstrate the connection between Zipf's and Taylor's Laws we analyse a comprehensive dataset on the population, x, of cities with more than 300,000 inhabitants globally [92]. This dataset consists of population estimates for the largest cities from 1950 to 2030 where cities are defined as urban agglomerations according to each country's definition of urban. To be consistent with our previous data analysis, Figure 3.1, we use only the data corresponding to 2010.

Figure 3.4**a** shows Zipf's Law for the countries with over 50 large cities ($x > 3 \times 10^5$). From this we observe that alongside the United States, the distribution of city sizes in India, China,

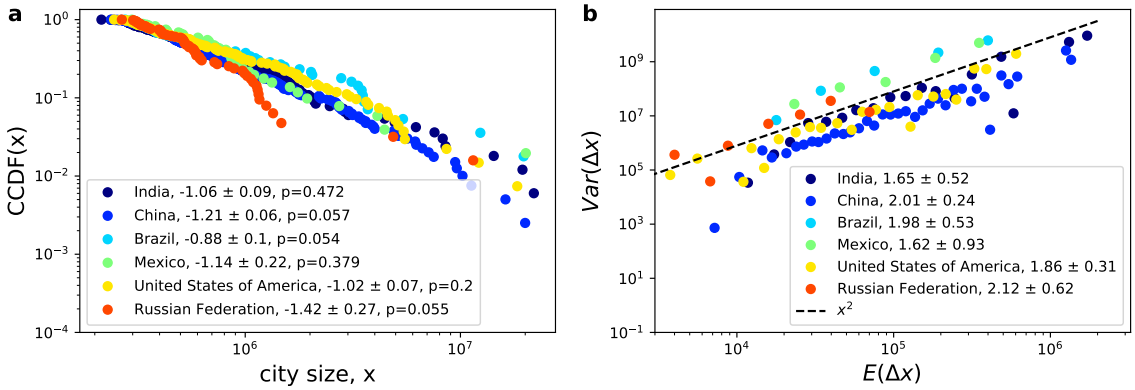

FIGURE 3.4. **a** Zipf's law, data. The CCDF (y-axis) of city sizes x (x-axis) for Brazil, China, India, Mexico, USA and Russia. Exponents quoted are equivalent to $\beta-1$. All countries follow Zipf's law with $\beta = 2$ within a 2σ confidence interval. The quoted p-values provide further support at the 95% confidence level that the distribution of city sizes follow Zipf's law with $\beta = 2$. **b** Taylor's law, data. The variance in the population change, $Var(\Delta x)$ (y-axis) of mean city sizes $E(\Delta x)$ (x-axis) for Brazil, China, India, Mexico, USA and Russia. All countries follow Taylor's law with $\alpha =$ within a 1σ confidence intreval.

Brazil, Mexico and the Russian Federation follow Zipf's Law with exponent $\beta - 1 = 1$, within a 2σ confidence interval. Exponents are determined using the counter-cumulative distribution function (CCDF) of the populations of each city. As we are only considering large cities, we are dealing with the tail of the distribution of city sizes which is often noisy. By using the CCDF rather than the PDF, any noise associated with binning the data is removed and the value of the fitted exponent is generally more accurate [95]. The errors in the exponent were calculated using bootstrapping [90] by performing $N = 1000$ trials where for each trial we sample with replacement from the list of true city sizes until our sample contains C cities. Here C corresponds to the number of cities with $x > 3 \times 10^5$ for each country. By doing this we are able to fit Zipf's exponent to the distribution for each trial and hence we obtain a distribution of N values of the exponent.

Taking this further, we compute the p-value for each country's exponent. This is obtained by performing $N = 1000$ trials of sampling C city sizes from a power law distribution with pdf exponent -2: the Zipf distribution. Fitting the exponent to the distribution of city sizes for each trial we obtain a p-value: the fraction of exponents that are more extreme than the exponent of the true city size distribution. From Figure 3.4 we observe that for all countries the distribution of city sizes follows Zipf's Law at the 95% confidence level.

In Figure 3.4**b** we demonstrate the validity of Taylor's Law for all countries that have over 50 cities with population $x > 3 \times 10^5$. We observe that all countries follow Taylor's Law with

exponent $\alpha = 1$ within a 1σ confidence interval. The errors in the values of α fitted to the data were obtained in the same way as the errors on the Zipf's exponents.

3.4 Conclusion

In this chapter we have presented two general classes of microscopic birth-death processes that can be shown to be equivalent to proportionate random growth and can reproduce Zipf's law for city sizes without fine tuning. For the processes considered the statistical properties of city sizes observed empirically can be obtained by relaxing the assumptions that individuals are independent and homogeneous. To summarise our results, we have shown that environmental variability and correlations between individuals are able to produce Zipf's law and Taylor's law. The exponent of Taylor's law shifts from $\alpha = 1/2$ for small populations to $\alpha = 1$ for large populations. Also, when population is large Zipf's law emerges naturally without fine tuning whenever the growth rate's mean is much larger than the variance. This reveals a general connection between Zipf's law and Taylor's law in microscopic stochastic processes of population dynamics under realistic assumptions. Our processes are applicable to dynamical systems with an explicit time dependence where the stationary distribution of group sizes can be described by Zipf's law. In this respect, our derivation of Zipf's law differs from static models without an explicit time dependence and from models where groups can only grow [2, 117, 140].

The models outlined are the most basic form possible; using a Poisson distribution reduces each model to just two parameters, however the use of other probability distributions is possible. Using an alternative distribution results in increasing the number of parameters from 2 to 4, thus allowing for more flexibility when fitting the model to data. The parameter values used in the simulations of Figure 3.2 are chosen to fit the models to city data, capturing both the Zipf exponent of the tail and the large value of n_c. Because of the specific choice of a Poisson distribution for $P(z|\lambda)$, this requires to use a set of parameter values that are very close to criticality (i.e. $\lambda = 1$). However, it is important to emphasise that this is not an intrinsic limitation of the models, and it is possible to find various distributions $P(z|\lambda)$ and $G(\lambda)$ that satisfy the conditions $|\mu| \ll \sigma_2^2 \ll \sigma_1^2$ and produce non-critical systems with stationary distributions close to Zipf's law for any value of n_c. In general, for generic distributions $P(z|\lambda)$ and $G(\lambda)$, we have:

$$\begin{cases} \sigma_1^2 &= \mathbb{E}_\lambda \left(Var_{\Delta x|\lambda}(\Delta x) \right) \\ \sigma_2^2 &= Var_\lambda \left(\mathbb{E}_{\Delta x|\lambda}(\Delta x) \right) \\ \mu &= \mathbb{E}_\lambda \left(\mathbb{E}_{\Delta x|\lambda}(\Delta x) \right) - 1 \end{cases}$$

so it is possible to find many combinations of functions and parameter values such that $|\mu| \ll \sigma_2^2 \ll \sigma_1^2$, corresponding to Zipf's law and large n_c. For example, using a Negative Binomial distribution for $P(z|\lambda)$, whose mean and variance can be independently set, and choosing an appropriate distribution over its parameters, $G(\lambda)$.

Within the proposed framework, migrations between cities are not explicitly modelled. A comprehensive microscopic model of human population dynamics should include a more thorough treatment of migration. In particular, we wish to have more accurate descriptions of the mechanisms behind individuals' choices to relocate to a new city and to understand what factors influence their choice of destination. These topics will be addressed in the next chapter.

CHAPTER 4

WHY, WHEN , WHERE: UNCOVERING THE DECISION MAKING BEHIND SCIENTIFIC MIGRATION

4.1 Introduction

The increasing availability of rich data sources related to research inputs and outputs, in particular publication datasets, has resulted in a new field emerging; the science of science (SciSci) [42]. The purpose of this field is to quantitatively understand how scientific creativity and discovery evolve by determining the driving forces behind successful science. One aspect of this emerging field is scientific mobility: the movement of scientists between institutions. As aforementioned in 1.3.3, so far the study of scientific mobility has either been based upon data obtained from large scale surveys with no modelling or predictive component or it has looked at how movement between institutions effects a scientist's performance and productivity.

In this chapter we investigate a new perspective; how a scientist's profile influences their decision to move. In order to do this we consider two unanswered questions: Why does a scientist decide to move institutions? When they have made this decision, where do they choose to move to? To answer these questions we present a novel model of scientific migration that combines data mining with a customised model of human mobility. This work can have important implications, helping research institutions and governments understand scientific mobility and implement policies to attract the best researchers, improving the quality of their research. Alongside this they can facilitate the construction of services that recommend new jobs to scientists based on their profile, or help search committees seek successful candidates for research jobs. Scientific migration is a specific type of occupational migration. While the answers to these questions cannot be generalised to the general population, the development of a modelling framework which uses data mining and machine learning techniques to inform a novel model of human mobility presents

an advancement in the modelling of human migration. Therefore, the framework used to study scientific migration can be adapted to other types of migration for which there are large-scale datasets.

We approach this problem by answering each question successively. Initially, in section 4.3, we define a career trajectory and in section 4.4, the profile of a scientist. In section 4.5 we present a model designed to predict whether or not a scientist will move in the next year based on their scientific profile. From this model we identify the most important factors that lead a scientist to make the decision to move. Finally, for those researchers who are predicted to move, in section 4.6 we predict which institution they will choose. We do this by introducing a novel version of a singly-constrained gravity model; an adaptation of the traditional gravity model of human mobility to include the factors identified at the first step.

The work presented in this chapter has been carried out in collaboration with L. Pappalardo, A. Sirbu and F. Simini. Specifically L. Pappalardo pre-processed the data, A. Sirbu obtained data on the citation network and F. Simini wrote the code to fit the gravity models presented in section 4.6 to the data.

4.2 Dataset

Our dataset consists of all articles published in the American Physical Society (APS) journals from 1950 to 2009. For each article, the date of publication, the author names and the corresponding affiliations are stored. In addition to this, location information (latitude and longitude) is available for every affiliation that appears in the dataset, as is citation information for every paper; we have a list of publications each paper cites and a list of publications in which each paper is cited. In total, the dataset consists of $\sim 1,000,000$ papers, over $200,000$ scientists and over $3,500$ institutions.

For our aims, we consider only scientists in the dataset with 2 or more distinct affiliations; a change in affiliation represents a move (see section 4.3). This constraint reduces the dataset to over $60,000$ scientists, $3,500$ institutions and $360,000$ articles. Some of the measures obtainable from this reduced dataset, are displayed in Figure 4.1. These measures, along with others, are used to determine the profile of each scientist.

From Figure 4.1 we observe that the number of papers published in APS journals has been increasing at a constant rate since ~ 1980 (panel **a**). Comparing panels **c** and **d** it is interesting to observe that the number of new authors publishing in these journals has seen a sharp decline from 2005 onwards (**c**) yet in contrast, many authors who had previously published in APS journals published again during this period (**d**). This suggests that the number of new scientists is declining, while experienced scientists are becoming more productive. While this observation is of interest, and warrants further investigation, we will not explore it here.

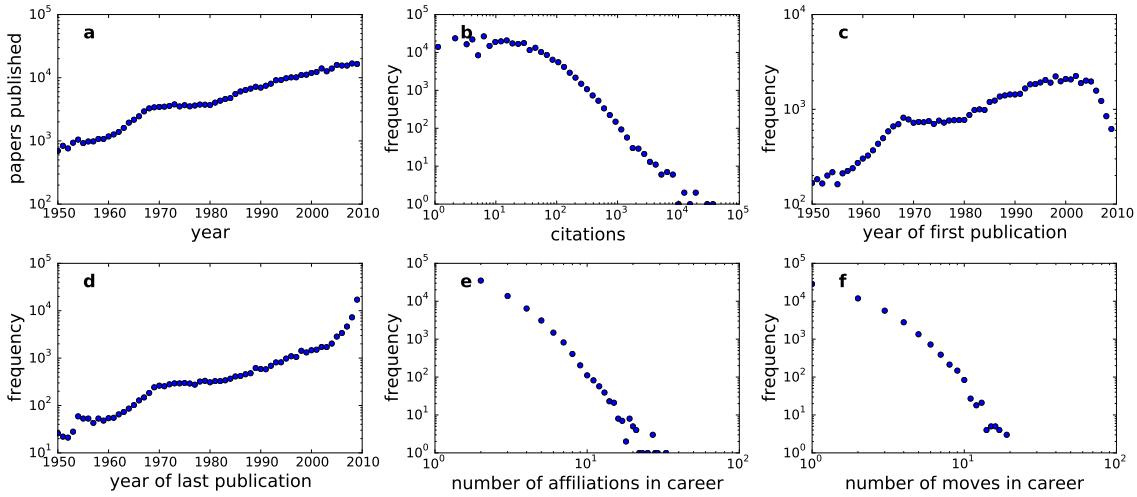

FIGURE 4.1. Features of the dataset: **a** the number of papers published in each year from 1950 to 2009; **b** histogram of the number of citations received by each paper; **c** the number of scientists publishing for the first time from 1950 to 2009; **d** the number of scientists publishing for the last time between 1950 and 2009; **e** histogram of the number of institutions scientists are affiliated with; **f** histogram of the number of moves a scientist makes during their career.

4.3 Career Trajectory

We define a career trajectory as a time-ordered list of institutions that a scientist has worked at. Formally, if a scientist s has worked at n institutions then their career trajectory, $T(s)$, is a list of n tuples;

$$T(s) = \langle (t_1, i_1), \ldots, (t_n, i_n) \rangle,$$

where t_j for $1 \leq j \leq n$ corresponds to the year that the scientist moved to their j^{th} institution and i_j is the corresponding institution that the scientist moved to. Two consecutive affiliations in a career trajectory correspond to a move, i.e., that the scientist moved from one institution to another. A move is therefore formally defined as a pair of consecutive tuples in $T(s)$. For example,

$$T(s) = \langle (1968, \text{Evanston}), (1973, \text{Hamburg}) \rangle$$

is a career trajectory indicating that scientist s moved from Evanston, Illinois to Hamburg, Germany in the year 1973.

Figure 4.2 shows an illustrative example of a 40-year long career trajectory: the scientist s is initially at Stanford University and moves to four other institutions during their career, each migration being detected by the changing of the main affiliation in s's publications.

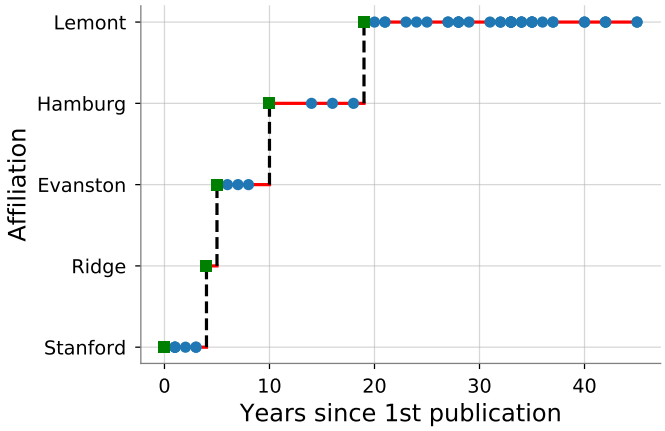

FIGURE 4.2. The career trajectory of a scientist. Blue circles represent publications. Black dashed lines represent moves from an institution to another. Green squares represent the publication at the new institution corresponding to the move.

4.3.1 Computation of career trajectories

For each scientist s in our dataset, we construct their career trajectory $T(s)$ as follows. First, we sort all their publications by time, from the least recent to the most recent publication. We then link to each publication their affiliation. Note that a scientist may specify more than one affiliation in a publication. We disambiguate these cases using the first affiliation reported by the scientist, as suggested in the literature [36]. From this time-ordered list of tuples, we remove all tuples (t_j, i_j) for which $i_{j-1} = i_j$, ie. tuples where the consecutive affiliations are the same. The resulting list is a career trajectory for scientist s where each pair of tuples represents a move the scientist has made between institutions.

4.4 Scientific Profile

We define the scientific profile of scientist s as a multidimensional feature vector:

$$\boldsymbol{p}^{(h,t)}(s) = [\underbrace{c_1^{(h,t)}, \ldots, c_n^{(h,t)}}_{\text{career}}, \underbrace{e_1^{(h,t)}, \ldots, e_m^{(h,t)}}_{\text{environment}}, \underbrace{r_1^{(h,t)}, \ldots, r_w^{(h,t)}}_{\text{relationships}}],$$

where each element of $\boldsymbol{p}^{(h,t)}(s)$ is a feature describing a specific aspect of scientist s's activity during a time window $(t-h, t)$ of h years preceding time t.

We consider a scientific profile to be composed of three main elements: *(i)* their *scientific career*, which corresponds to features that are personal to the scientist such as their experience, publications and citations; *(ii)* their *scientific environment*, which relates a scientist to their colleagues at their current institution, comparing their respective scientific output; and *(iii)* their *scientific relationships*, indicating the working relationships a scientist has established

with collaborators at external institutions during the h years. Table 4.1 describes the features composing the scientific profile.

4.4.0.1 Scientific career

The scientific career includes features describing individual characteristics of scientist s in the past h years. As proxies of scientific production, we consider the number of publications the scientist has produced and the citations they have accrued during the considered period. Moreover, we estimate a scientist's experience as the number of years since their first publication and define a scientist's scientific mobility using boolean values which represent whether they have or have not changed institutions in the last h years.

4.4.0.2 Scientific environment

A scientist's production shapes, and it is shaped by, the scientific environment in which they operate. For this reason we estimate the level of production of a scientist's environment as the number of citations and the number of publications of their colleagues during the h years. A colleague is a scientist working at the same institution as scientist s at time t. We also consider the differential of citations, i.e., the mean difference between s's citations and their colleagues' citations. This allows us to compare the productivity of a scientist to the productivity of their peers.

4.4.0.3 Scientific relationships

Scientific collaboration is a proven mechanism for promoting excellence in scientific research, as higher collaboration rates are linked to higher citation rates [1, 3, 126]. For this reason we take into account a scientist s's collaborations by estimating the size of their collaboration circle using three features: the number of institutions they have collaborated with during the h years, the total number of distinct collaborators and their tendency to collaborate with scientists at external institutions (that we call xenophilia), computed as the ratio of external collaborators to the total number of collaborators in the h years.

4.4.1 Computation of scientific profiles

Given time t and history window h, from our dataset we compute the career, environment and relationship features for the scientist's activity in the h years preceding t in the following way. The number of publications by a scientist s is given by the total number of papers in the dataset for which s is an author and the publication date falls within the period $(t-h,t)$. We compute the number of citations of s as the sum of citations to all papers in the dataset for which s is an author and for which the citing paper is published in the period $(t-h,t)$. The experience of scientist s is computed as the difference $t-t_1$, where t_1 is the time of s's first publication in the

category	variable	description
(i) career	publications	number of papers published in the last h years
	citations	number of citations received in the last h years
	experience	years since the first publication
	mobility	whether she changed institutions in the last h years
(ii) environment	colleagues' citations	mean number of citations received by colleagues in the last h years
	colleagues' publications	mean number of papers published by colleagues in the last h years
	differential of citations	mean difference between citations and colleagues' citations in the past h years
(iii) relationships	institutions	number of institutions she has collaborated with in the past h years
	collaborators	number of scientists she collaborated with in the last h years
	xenophilia	ratio of external collaborators to total collaborators in the last h years

TABLE 4.1. The features describing scientific profile with the corresponding description. They are grouped in three macro-categories: (i) career; (ii) environment; and (iii) relationships.

dataset. Finally, the mobility of s relates to how recently they have changed institution: if s has moved within the period $(t-h,t)$, the mobility feature has value 1, if they have not moved the feature has value 0, if at time t they are at their first institution (i.e., the only institution they have been affiliated with so far in the dataset) they are assigned a value of -1.

To compute the environmental features we define the *colleagues* (or peers) of scientist s as all the scientists who publish a paper during the period $(t-h,t)$ that are affiliated with s's institution at time t. For each colleague we compute their publications and citations as described above. The colleagues' citations and publications features are then given by the mean number of citations and the mean number of publications for all colleagues, respectively. The differential of citations is determined by taking the difference between the scientist's citations and each peer's citations individually, and then taking the mean of these differences.

To compute the relationships features we define a *collaborator* as a scientist who is a co-author of s for at least one paper published in the period $(t-h,t)$. The collaborators feature is then the number of distinct collaborators in the list of co-authors and the institutions feature is the number of distinct affiliations. We compute xenophilia as the ratio between the number of collaborators at external institutions divided by the total number of collaborators.

4.5 Why does a scientist decide to move institutions?

In order to understand why a scientist chooses to move institutions we construct a predictive model. This model aims to predict if a scientist will migrate in the next year given the features in their scientific profile. We consider two types of machine learning models to classify scientists s. A classification model is one in which, given an observation (or feature vector) determines the probability of that observation belonging to one of two or more classes. When there are only two classes, this is termed binary classification. For our first model we have a binary classification task; given the feature vector $\boldsymbol{p}^{(h,t)}(s)$, does scientist s move (class 1) or not (class 0). We will consider two binary classifiers, a logistic regression and a decision tree, and compare the accuracy of each one to a baseline classifier. For technical descriptions of the classifiers and evaluation measures, see Appendix B.1.

For each scientist s in our dataset we compute their feature vector $\boldsymbol{p}^{(h,t)}(s)$ that describes their scientific profile between years $t-h$ and t. Since the features in $\boldsymbol{p}^{(h,t)}(s)$ have different ranges of values and distributions, we standardise them by computing their quantiles with respect to each feature of the other scientists. The transformed features are all uniformly distributed. By doing this, each element in $\boldsymbol{p}^{(h,t)}(s)$ lies between 0 and 1 and represents how the features of s compare to all other scientists in our dataset. Standardising the features in this way allows us to compare the values of their coefficients once the model has been trained; a feature with a large coefficient suggests it is an important factor in a scientist's decision to relocate.

After standardising the features we assign each scientist s to a label; $m(s) = 0$ or $m(s) = 1$. Here 1 indicates that scientist s will migrate the next year, and 0 that s will not migrate. From the scientists' feature vectors we construct a dataset of examples $\boldsymbol{p}^{(h,t)} = \{\boldsymbol{p}^{(h,t)}(s) | s \in D_{\text{APS}}\}$ each associated with the corresponding label in $m = \{m(s) | s \in D_{\text{APS}}\}$. Here, D_{APS} denotes our dataset. For examples where $m(s) = 1$, we use the move events of all 60,000 scientists: all values of t for which the affiliation of s in year $t+1$ is different from the affiliation of s in year t. For examples where $m(s) = 0$, for each scientist s we generate a random value t between the scientist's first and last publications in the dataset until $t+1$ is not a year in which s moved. We repeat this process 3 times to ensure that our classes, $m(s) = 0$ and $m(s) = 1$ are approximately balanced. In total, we have 290,000 points in $\boldsymbol{p}^{(h,t)}$ to train and test our classifiers on: 140,000 examples for which $m(s) = 1$ and 150,000 examples for which $m(s) = 0$.

4.5.1 Results

To investigate to what extent we can predict whether a scientist will migrate next year given their history, for each value $h = 1, \ldots, 10$ we train two predictors L_h and T_h on the dataset $\boldsymbol{p}^{(h,t)}$ and the associated labels m, where L_h is a logit and T_h is a decision tree. We evaluate the predictors with 10-fold cross-validation and investigate the goodness of the predictions – in terms of accuracy, recall, precision, F1-score and AUC. Table 4.2 shows the values of the evaluation

measures for the best tree and the best logit. We find that $h_{\text{best}} = 1$ for T_h and $h_{\text{best}} = 10$ for L_h.

model	h_{best}	**ACC**	**recall**	**prec**	**F1**	**AUC**
tree	1	**0.65**	**0.84**	**0.61**	**0.71**	**0.65**
logit	10	0.62	0.64	0.62	0.63	0.62
dummy	-	0.50	0.50	0.50	0.50	0.50

TABLE 4.2. Predictive performance of the best tree (T_1) and the best logit (L_{10}), compared with a baseline classifier (dummy). The models are evaluated with a 10-fold cross-validation, and the goodness of predictions are measures in terms of accuracy (ACC), recall, precision (prec), F1-score (F1) and Area Under the Curve (AUC).

Both predictors perform better than a baseline classifier which generates predictions based upon the class distribution of the training set.

Figure 4.3 shows how the AUC score changes with the size of the history window for T_h and L_h, $h = 1, \ldots, 10$. We observe that: (i) the decision tree performs slightly better than the logit; (ii) the performance of the logit improves with increasing h; (iii) the decision tree performs best with a small history window and performance declines with increasing h suggesting that a greater history window adds no predictive power to the model.

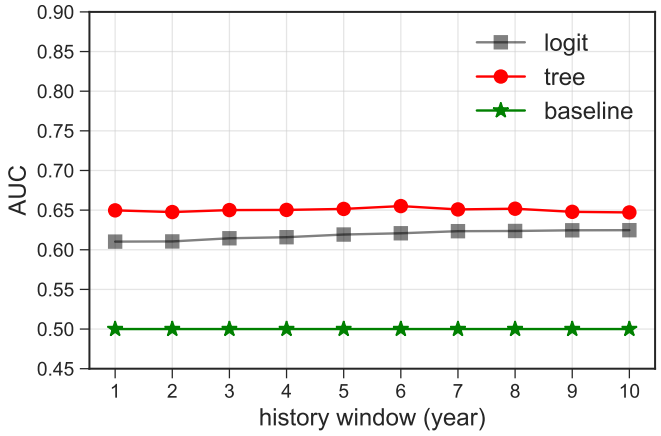

FIGURE 4.3. AUC score of predictors T_h and L_h, for $h = 1, \ldots, 10$. The baseline classifier has $AUC = 0.5$.

Figure 4.4**a** shows the feature importances resulting from T_1, indicating that a scientist's experience is the strongest predictor of the probability to migrate. Figure 4.4**b** shows the standardised features of a scientist correctly predicted to move (red bars) and a scientist correctly predicted to stay (grey bars) using model T_1. We observe that the scientist who moves scores highly in experience and mobility compared to the scientist who does not move. In contrast, the

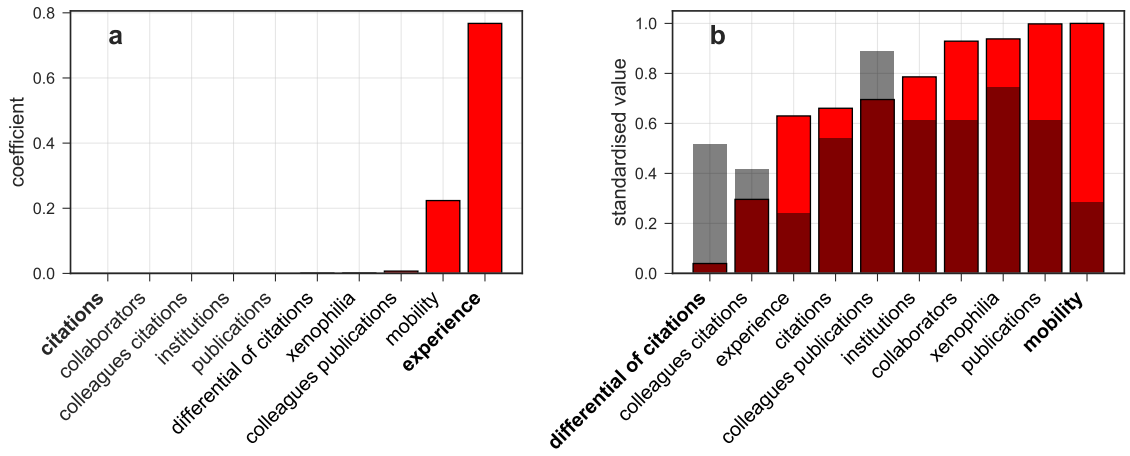

FIGURE 4.4. **a** Coefficients resulting from model T_1. The experience of a scientist is the strongest predictor of future migration. **b** Feature values for a scientist predicted to move by model T_1 (red) and a scientist predicted to stay (grey).

scientist who does not move scores highly against other scientists for features, such as differential of citations and colleague's citations, that are of little importance according to Figure 4.4**a**.

4.6 Where do scientists move to?

The second stage of our model aims at predicting the destination institution that a scientist will move to. The task of estimating the probability that a scientist will relocate to a given institution can be interpreted as a classification problem with as many classes as institutions. The state-of-the-art model to estimate mobility and migration flows is the gravity model [40, 62, 98], which is a multinomial logistic regression based on distance and size (population) of locations. Specifically, the features of a traditional singly-constrained gravity model are the size of the potential destination, estimated here as the logarithm of the number of scientists affiliated to the destination institution, and the logarithm of the distance between the scientist's current institution and the potential destination. We compare the performance of the traditional gravity model with a quality-social-gravity model, an extended model that includes additional indicators relating to the quality of the potential destination and social interactions a scientist has had with researchers at the potential destination. The new social and quality features are: *(i)* the fraction of collaborators at the potential destination; *(ii)* the average number of papers published by the scientists at the potential destination; *(iii)* the average number of citations received by the scientists at the potential destination. We compute these features using $h = 5$ and we standardise them by computing their quantiles, as done for the move prediction. The probability that a

scientist s will relocate to institution i is then estimated as

$$P(s \to i) = \frac{e^{z(\mathbf{x}_{s,i})}}{\sum_j e^{z(\mathbf{x}_{s,j})}} \tag{4.1}$$

where the score $z(\mathbf{x}_{s,i}) = \boldsymbol{\beta} \cdot \mathbf{x}_{s,i}$ is a linear combination of the features defined above. The parameters $\boldsymbol{\beta}$ are estimated by maximising the model's likelihood using stochastic gradient ascent. For technical details see Appendix B.2.

The models are assessed using the cumulative distribution function (CDF) of the ranks r of possible destinations and the Mean Percentile Ranking (MPR) [73]. A rank corresponds to the probability of scientist s relocating to a given destination, with $r = 1$ representing the most likely destination. The MPR is defined as

$$MPR = \frac{1}{|S|} \sum_{s \in S} \frac{r_s}{|I|}, \tag{4.2}$$

where r_s is the rank of scientist s's actual destination, S represents all scientists the model is tested on and I corresponds to the number of possible destinations. This is therefore a measure of a model's ability to correctly predict a scientist's destination. A low value of MPR is more desirable as means that the model can assign the true destination a high rank.

We train the models using a subset of 200 of the scientists in our dataset; we find that there is no performance improvement using a larger number of scientists and training the models on 200 reduces the computational cost. In order to reduce the computational cost of the optimisation, which can be quite high when the total number of destination locations is over 10^3, we consider an approximation of the likelihood computed considering a subset of 100 potential destinations, as proposed in [73]. A random subset of potential destinations is extracted for each move of each scientist in our training set. This subset of potential destinations always includes the true destination, while the other locations are randomly selected with a probability that increases linearly with their sizes and slowly decreases with the distances from the origin location. This ensures that the most relevant potential destinations, i.e., the larger and closer to the origin, are included in the likelihood's estimate. Numerical tests show that the optimal size of the subset of potential destinations is 100 locations, i.e., considering more than 100 potential destinations does not significantly improve the model's performance.

The model's performance is evaluated considering all the remaining scientists in the dataset. For each scientist s in our testing set we compute the probabilities, $P(s,i)$, to relocate to any institution in the dataset according to the model. We then sort all institutions in decreasing order of $P(s,i)$ and define the rank of each institution so that the model's top prediction has the largest $P(s,i)$ and rank 1. We then consider the rank of the scientist's actual destination, r_s, and we use it to compute two statistics: (1) the cumulative distribution function (CDF) of the ranks r of all the moves of all scientists in our testing set (Figure 4.5); and (2) the Mean Percentile Ranking (MPR).

4.6.1 Results

Results show that including information on the quality of the institutions improves the model's performance. In particular, the quality-social-gravity model that includes quality and social information has CDF(r=10) = 0.23, i.e., for 23% of the scientists in our testing set the real destination is among the top 10 model's predictions. This is in contrast to the original gravity model without quality and social information, which has CDF(r=10) = 0.07, i.e., for only 7% of the scientists the real destination is among the top 10 model's predictions (see Figure 4.5). This result, that a quality-social-gravity model which incorporates social and quality information is superior, is further supported by the MPR: $MPR_{\text{quality-social}} = 0.069$ while $MPR_{\text{gravity}} = 0.074$, corresponding to an error reduction of 7%. These results are summarised in Table 4.3.

model	**MPR**	**CDF (r=10)**	$\bar{r}$
gravity	0.074	7%	275
quality-social-gravity	**0.069**	**23%**	**257**

TABLE 4.3. Performance comparison between the original gravity model and the quality-social-gravity model

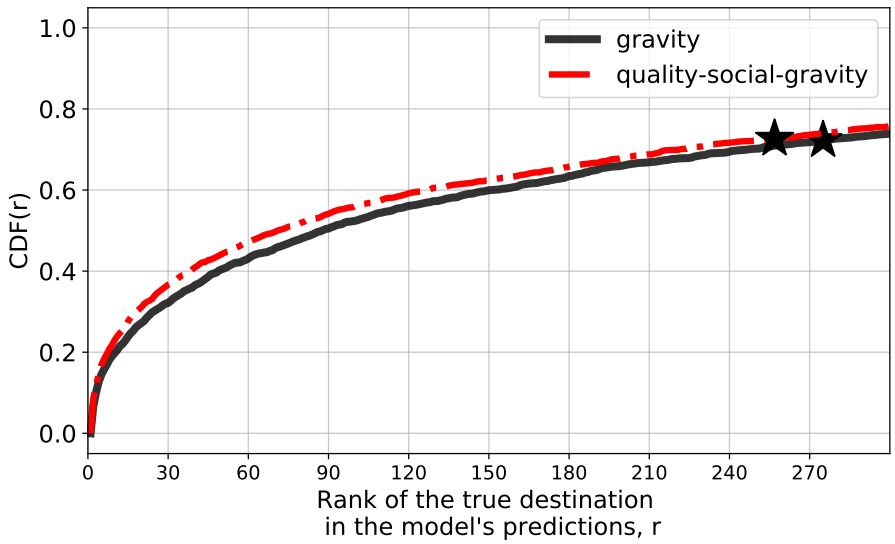

FIGURE 4.5. CDF of the ranks of scientists' true destinations according to the original gravity model (black) and the quality-social-gravity model (red). Black stars represent the mean rank: we find that $\bar{r} = 275$ for the original gravity model and $\bar{r} = 257$ for the quality-social-gravity model.

4.7 Conclusion

In this chapter we proposed a solution to the problem of predicting the next institution a scientist will move to. In the first part, we used data mining to predict whether or not a scientist will migrate in the next year on the basis of the quality of their career, environment and collaborations. The use of data mining and machine learning methods revealed which features where most influential in determining whether a scientist will move or not. Using these features we moved to the second, and more challenging, task of predicting where a scientist will move to. We showed that our quality-social-gravity model, obtained by injecting information about the quality of institutions and the scientist's collaboration network into the traditional gravity model, achieves a prediction error which is up to 7% less than a traditional gravity model.

Our work is placed at the point of intersection of previous research, outlined in section 1.3.3. We explore the characteristics of scientific performance which most affect the decision to relocate, focusing on aspects that have not been previously investigated, such as a scientist's propensity to collaborate with external institutions and their relationship with their scientific environment. Moreover, we propose an algorithm to predict the next career move which is tailored for scientists, hence considering science-specific features and the distance between scientific institutions. From a methodological point of view, our work provides a novel solution to the next (scientific) career move problem, as we combine data mining and predictive models with global generative migration models.

CHAPTER 5

DISCUSSION

In the preceding chapters we have presented new advances in addressing knowledge gaps in three distinct areas relating to the mathematical modelling of human mobility and population dynamics, with a focus on urbanisation.

Initially we aimed to determine which model of migration best describes the spatial distribution of cities. In order to do this we introduced a deterministic model of population dynamics and demonstrated analytically that if migration is described by a gravity model, the spatial pattern of the population distribution is independent of the total population. In contrast, if migration is described by an intervening opportunities model of migration the number of cites, and hence their spatial distribution, varies with the total population of the region. These results were confirmed using numerical simulations. Following on from this we conjectured Heaps' Law for cities; an analytical relationship between the number of cities in a region and the region's population, $C \sim N^{\gamma}$. Here, $\gamma = 0$ represents the analytical result from a gravity model of migration and $\gamma = 1$ an intervening opportunities model. To validate the existence of Heaps' Law we used data on the population of all cities with over 1000 inhabitants. We found that the number of cities in a country or region scales linearly with the total population of the region; $\gamma \sim 1$. Hence, according to our deterministic model the spatial distribution of population is best described by an intervening opportunities model of migration. We also investigated whether Zipf's law and Heaps' law depend on the definition of a city and whether they exist for the areas of cities, as well as the populations. To do this we analysed a gridded dataset for the United States, which reports the number of people living in each square kilometre. Using a clustering algorithm to obtain cities we found that Zipf's law is present for the size distributions of both cluster populations and cluster areas. For populations the exponent of Zipf's law depends on parameter X of the clustering algorithm which defines the minimum population of a cell within a cluster. We also validated Heaps' law

for urban clusters; the number of clusters within a given area is linearly proportional to the population of that area.

The deterministic equation combined with both gravity and intervening opportunities models of migration is a novel attempt at describing the evolution of the population density. This allowed us to highlight the differences between the two models in terms of the number of cities expected in a region of increasing total population. However, the logistic function is not a realistic description of population growth. In order to mathematically model city formation, a stochastic approach is more appropriate as this can naturally reproduce Zipf's law for the distribution of city sizes.

While the deterministic model may have limitations, the analytical results obtained from its analysis led us to empirically test the validity of Heaps' Law for Cities. Heaps' Law provides a quantitative description of the number of cities expected from a given total population. We empirically verified that a null model of urbanisation where cities are randomly distributed in space produces correct estimates of the expected number of cities in regions of various sizes worldwide. While this suggests that cities are not spatially correlated, it does not mean that cities are non-interacting and independent of each other. On the contrary, it is apparent that urban systems are strongly interacting [59]: migrations, for example, create a dependency in the dynamics of the population in various cities, with some cities increasing in size because others are decreasing. However, our analysis demonstrates that such interactions do not produce urbanisation patterns characterised by significant spatial correlations. These results provide a new insight into the urbanisation process. They suggest that as a region's population increases, existing cities will not only grow in size but new cities will also form, as demonstrated for Iowa in the United States. The absence of spatial correlations suggests that the growth or formation of a city does not depend on its location in relation to existing cities.

In the analysis of gridded population data we only consider regions with low urbanisation, having a percentage of urbanised area smaller than 5% (this condition is satisfied by 49% of the regions for $X = 10$ and up to 93% for $X = 599$). We do this because in highly urbanised areas deviations from Zipf's law and Heaps' law are inevitable. In fact, in regions with large population density, urban clusters start to merge and, as a result, when population keeps increasing the number of clusters decreases instead of increasing. Alongside this, the distribution of cluster sizes loses its characteristic power law tail because of the emergence of one giant cluster. The characterisation of urban patterns in the regime of large population density requires the development of a different theoretical framework, which is a task left for future work. The theoretical result Zipf's and Heaps' laws in finite-size systems, Equation 2.17, is completely general and applicable to all systems characterised by Zipf's law for the distribution of group sizes, including word counts, the size of biological genera, the number of firm employees and views/popularity of Youtube videos.

Secondly, we introduced a novel microscopic model of population growth that does not require fine tuning of its parameters. Our aim was to understand how Zipf's Law for cities appears as

a mesoscopic distribution of city sizes from the underlying microscopic dynamics at the level of the individual inhabitants. We considered a simple birth-death process and demonstrated that by relaxing the assumptions of the central limit theorem, namely that individual birth and death events are independent and identically distributed, the stationary distribution of city sizes follows Zipf's Law. Analytical results were validated using numerical simulations and data on the populations of cities and counties in the United States. We used the scaling between the mean and variance of population increments to conjecture a link between Taylor's Law and Zipf's Law: when Zipf's Law describes the stationary distribution of city sizes, the population increments follow Taylor's Law with exponent $\alpha = 1$. This result was also validated using both numerical simulations and data from the United States and other countries globally.

Our models differ from other microscopic models of population dynamics in several ways. Previous models are only able to obtain a distribution of city sizes that can be described by Zipf's law for certain values of their parameters [45, 117, 139]. In contrast, our models can reproduce Zipf's law for a range of parameters; the only constraint is that $|\mu|/\sigma_2^2 \ll 1$ which can be satisfied by many (μ, σ_2) pairs. Alongside this, other models that do reproduce Zipf's law do so for the whole distribution of city sizes. This is in spite of the fact that, historically, is has been noted that Zipf's Law is only present in the tail of the distribution of city sizes [95] and does not fit the distribution as a whole. Our models are able to account for the absence of Zipf's law in the distribution of smaller cities and relate the population from which Zipf's law is present to the transition between Taylor's exponents 1/2 and 1.

Considering the probability density function in Equation 3.10, it is important to highlight that the values of parameters μ and σ_2 allow for deviations from the exponent of Zipf's law, $\beta = 2$. While an exponent of 2 is traditionally associated with the distribution of city sizes, this is not stable when preforming a cross country comparison; exponents both higher and lower than this value are also common [106, 120]. On the one hand, our model may account for the variation in Zipf's exponent globally, suggesting that if overall population growth is negative, exponent $\beta > 2$ whereas if population growth is positive, $\beta < 2$. On the other hand, in section 3.3 we demonstrated that although the exponents $\beta - 1$ fitted to the data deviate from 1, the p-values suggest that the distributions of city sizes are still compatible with Zipf's law and an exponent $\beta = 2$. Further exploration of this relationship is a task for the future however it suggests that our methods may be a starting point for explaining the distribution of city sizes in all countries.

The microscopic models presented are not only applicable to cities; they can be applied to any other dynamical system where group sizes follow Zipf's law. In fact, if a stationary distribution of group sizes is known to approximately follow Zipf's law from a minimum size, our models can be exactly fit to the data. Thus, the models are both flexible and accurate in describing the evolution of groups whose stationary distribution follows Zipf's law.

Finally, we focussed on the migration of scientists with the aim of understanding why a scientist makes the decision to move and how they select their next institution. In order to

determine the factors that are most important in a scientist's decision to relocate we used decision tree and logistic regression classifiers. These classifiers have built in feature selection and revealed that a scientist's experience along with their mobility are the most significant in determining whether they will move or not. Using these results we constructed a novel quality-social-gravity model; a singly constrained gravity model with a deterrence function that incorporates quality and social factors along with the distance. We demonstrated that the incorporation of these additional features improves the model's accuracy at predicting which institution a scientist is most likely to move to when compared to a traditional singly-constrained gravity model.

Predicting when a scientist will move and where they will move to is inherently a challenging task, however our results provide us with a deeper understanding of the factors influencing a scientist's decision to relocate. Our first model, predicting whether or not a scientist will move in the next year, performed better than the baseline classifier on all measures. In the future, the predictive power of our model could be further improved by considering a longer future window: rather than predicting whether a scientist will move next year, predicting whether they will move in, for example, the next 2, 5 or 10 years.

While the inclusion of additional quality and social features improved the performance of the gravity model, the difference in performance between the traditional gravity model and quality-social-gravity model was minimal, with an error reduction of only 7%. This suggests that quality and social features alone do not provide enough information to accurately predict where a scientist will next move. The APS dataset used in our work has been widely used by other researchers [31, 36, 118]. Of particular interest is the work by F. Gargiulo and T. Carletti [50] which is similar to ours; the data is used to reconstruct researchers' career trajectories and their evolution is studied in terms of some selected features. However it also differs from ours: features selected are focussed on cultural-geographical aspects and there is no predictive model, rather they embed trajectories in a spatial network and use methods from network theory for analysis. They find that distance is not a good predictor of a scientist's next institution but rather cultural similarity such as shared language and history are important factors. In the future it would be interesting to incorporate features such as these into our quality-social-gravity model: additional features relating to cultural aspects could further improve the predictive power of the model.

Our work also paves the road to many other research lines. For example it would be interesting to generalise our approach to other classes of high-skilled individuals, such as senior managers [60] or soccer players [79]. As the relocation of those individuals strongly affects the success of both the origin and destination companies, predicting their relocation decisions can have wide economic consequences in the companies' revenues and future profits. If the inclusion of additional or alternative features improves the accuracy of our adapted gravity model, we plan to exploit the proposed framework for the creation of a human migration model for the general population. In this context, we can use mobile phone data and social media data to describe

individual relocations and social relationships, respectively, and construct an adapted gravity model for human migration. A general model of human migration would mark a significant advancement in the modelling of urbanisation. Such a model could allow governments and city planners to estimate how the population of a city will grow due to migration. Combined with a model of population growth within cities, such as the one presented in Chapter 3, sustainable urban planning and growth in countries undergoing rapid urbanisation would become more possible.

APPENDIX A

MATHEMATICAL ANALYSIS OF THE DETERMINISTIC MODEL AND A STOCHASTIC MODEL OF POPULATION DYNAMICS

A.1 Gravity Model

A.1.1 Pattern Formation and Growth in 1D

Considering a 1-dimensional line of infinite length, we write the dynamic Equation, 2.1, for the gravity model as:

$$\begin{aligned}\frac{\partial\rho(x,t)}{\partial t} =& g(1-\frac{\rho(x,t)}{\rho_0})\rho(x,t) - T\rho(x,t) \\ &+ T(\rho(x,t)+w)\Big(\int_0^\infty \frac{f(r)\rho(x-r,t)}{\int_0^\infty f(z)(\rho(-r+x-z,t)+\rho(-r+x+z,t)+2w)dz} \\ &+ \frac{f(r)\rho(r+x,t)}{\int_0^\infty f(z)(\rho(r+x-z,t)+\rho(r+x+z,t)+2w)dz}dr\Big).\end{aligned} \tag{A.1}$$

The uniform distribution $\rho(x)=\rho_0$ is a stationary state of Equation 2.1 because the growth term is equal to zero and T^{in} is equal to T^{out} for all x, hence the time derivative on the left-hand side is zero. In order to determine the stability of the uniform steady state we linearise Equation

A.1 about ρ_0 and study the time evolution of a small perturbation $\tilde{\rho}(x,t)$;

$$\begin{aligned}\frac{\partial\tilde{\rho}(x,t)}{\partial t} =&(-g-T)\tilde{\rho}(x,t)+T\int_0^\infty \frac{2\rho_0 f(r)\tilde{\rho}(x,t)}{\int_0^\infty f(z)(2\rho_0+2w)\,dz}\,dr\\ &+T(u+w)\int_0^\infty\Bigg(-\frac{\rho_0 f(r)\int_0^\infty f(z)(\tilde{\rho}(-r+x-z,t)+\tilde{\rho}(-r+x+z,t))\,dz}{\left(\int_0^\infty f(z)(2\rho_0+2w)\,dz\right)^2}\\ &-\frac{u f(r)\int_0^\infty f(z)(\tilde{\rho}(r+x-z,t)+\tilde{\rho}(r+x+z,t))\,dz}{\left(\int_0^\infty f(z)(2\rho_0+2w)\,dz\right)^2}\\ &+\frac{f(r)\tilde{\rho}(x-r,t)}{\int_0^\infty f(z)(2\rho_0+2w)\,dz}+\frac{f(r)\tilde{\rho}(r+x,t)}{\int_0^\infty f(z)(2\rho_0+2w)\,dz}\Bigg)\,dr.\end{aligned} \tag{A.2}$$

Here, $\tilde{\rho}(x,t)$ represents a small fluctuation in the stationary density, ie $\rho(x,t)=\rho_0+\tilde{\rho}(x,t)$, and can be decomposed as a sum of normal modes of wave number $k\in\mathrm{R}$ and amplitude $h_k(t)$.

Substituting $\tilde{\rho}(x,t)=h_k(t)e^{ikx}$ into Equation A.2 we obtain an equation for the time evolution of the amplitude;

$$\begin{aligned}\frac{\partial h_k(t)}{\partial t} =&(-g-T)h_k(t)+\frac{T\rho_0 h_k(t)\left(\int_0^\infty f(r)\,dr\right)}{\int_0^\infty f(z)(\rho_0+w)\,dz}\\ &+T(\rho_0+w)h_k(t)\left(\frac{\int_0^\infty f(r)\cos(kr)\,dr}{\int_0^\infty f(z)(\rho_0+w)\,dz}-\frac{\rho_0\left(\int_0^\infty f(r)\cos(kr)\,dr\right)^2}{\left(\int_0^\infty f(z)(\rho_0+w)\,dz\right)^2}\right),\end{aligned} \tag{A.3}$$

which may be written as

$$\frac{\partial h_k(t)}{\partial t}=h_k(t)\Lambda_k(\rho_0,g,T,w,f). \tag{A.4}$$

Here Λ_k is the growth rate of the instability of the normal mode with wavenumber k; in order for cities to develop and patterns to emerge Λ must be greater than zero for some k. We consider the cases in which the deterrence function, $f(r)$ takes an exponential or power law form.

A.1.1.1 Exponential Deterrence Function

If $f(r)$ takes an exponential form (ie. $f(r)=e^{-rR}$) then the growth function Λ is given by:

$$\Lambda_k(\rho_0,g,T,w,R)=-g-T\frac{\rho_0}{\rho_0+w}+T\frac{w+(k/R)^2(\rho_0+w)}{(1+(k/R)^2)^2(\rho_0+w)}. \tag{A.5}$$

This expression simplifies using rescaled variables to give:

$$\lambda_\kappa(\mu_0,\tau)=-1-\tau\frac{\mu_0}{\mu_0+1}+\tau\frac{1+\kappa^2[\mu_0+1]}{[1+\kappa^2]^2[\mu_0+1]}. \tag{A.6}$$

Here, $\kappa=k/R$, $\tau=T/g$, $\mu_0=\rho_0/w$ and $\lambda=\Lambda/g$. The curve defined by Equation A.6 is displayed in Figures A.1a and A.1b for different values of μ_0 and τ.

As aforementioned, in order for cities to form, λ_κ must be greater than zero at its maximum κ. The maximum is found by differentiating the expression for λ_κ with respect to κ. On doing this, we find that $\lambda_\kappa(\mu_0,\tau)$ has a maximum in κ_m=0 of $\mu_0\le 1$ and in

$$\kappa_m=\sqrt{(\mu_0-1)/(\mu_0+1)} \tag{A.7}$$

otherwise. This corresponds to a wavenumber $k_m = R\sqrt{(\rho_0 - w)/(\rho_0 + w)}$ which, providing $\Lambda_{k_m} > 0$, corresponds to the average number of cities per unit length. For the case that $\rho_0 \gg w$ it can be seen that k_m only depends on R, therefore the density, or number, of cities is fully determined by the characteristic length of travel, R^{-1}, and is independent of all other variables such as the growth and migration rates.

The condition on the migration parameter T in order for λ_κ to be greater than zero is found by substituting the expression for κ_m back into Equation A.6 and solving for τ. On doing this, the conditions:

$$T_c \geq 4g\rho_0 \frac{(\rho_0 + w)}{(\rho_0 - w)^2},\ \tau_c \geq 4\mu_0 \frac{(\mu_0 + 1)}{(\mu_0 - 1)^2} \tag{A.8}$$

are obtained. From these relations we can conclude that if $\rho_0 \gg w$ then cities will emerge if $\tau > 4$, or $T > 4g$. It is important to note that Equation A.8 is independent of R; the characteristic travel distance has no effect on whether or not cities will emerge.

A.1.1.2 Power Law Deterrence Function

The power law form for the deterrence function is given by $f(r) = (1+r)^{-\gamma}$ with $\gamma > 0$. The analysis of Equation A.2 is more complicated for this case, however the maximum of $\lambda_k(\mu_0, \tau, \gamma)$ will occur at a κ_m that is only dependent on μ_0 and the function f with its parameter γ. This is a general result for every function f due to the fact that κ_m is defined as $\lambda'_{k_m}(\mu_0, \tau, \gamma) = 0$,

$$\begin{aligned} \lambda'_k &= \frac{d}{dk}\left\{ -1 - \tau\frac{1}{\mu_0+1} + \tau\left[\frac{\int_0^\infty f(r)cos(kr)dr}{\int_0^\infty f(z)dz} - \frac{\mu_0}{\mu_0+1}\left(\frac{\int_0^\infty f(r)cos(kr)dr}{\int_0^\infty f(z)dz} \right)^2 \right] \right\} \\ &= \tau\left[\frac{d}{dk}\left(\frac{\int_0^\infty f(r)cos(kr)dr}{\int_0^\infty f(z)dz} \right) - \frac{\mu_0}{\mu_0+1}\frac{d}{dk}\left(\frac{\int_0^\infty f(r)cos(kr)dr}{\int_0^\infty f(z)dz} \right)^2 \right], \end{aligned} \tag{A.9}$$

a function of μ_0 and f only. For the case of $\mu_0 \gg 1$ (that is $\rho_0 \gg w$), it is only the function f, defining the spatial range of the migration process, that the characteristic distance between cities is dependent upon.

The function $\lambda'_{k_m} = 0$ may be solved numerically for different values of γ. On doing this we find that the critical curve collapses on the same universal function, independent of γ. Moreover, it turns out that the critical curve of Equation A.8 determining the condition for the formation of cities is valid for any continuous function f, including a power law deterrence function and therefore holds universally. We may prove this by defining $A(k) \equiv \left(\int_0^\infty f(r)cos(kr)dr\right)/\left(\int_0^\infty f(z)dz\right)$. From Equation A.9 we have

$$\begin{aligned} 0 &= A'(k_m) - \frac{2\mu_0}{\mu_0+1}A(k_m)A'(k_m) \\ A(k_m) &= \frac{\mu_0+1}{2\mu_0}. \end{aligned} \tag{A.10}$$

Inserting this into $\lambda_{k_m}(\mu_0, \tau_c) = 0$, the critical curve defined in Equation A.8 is re-obtained.

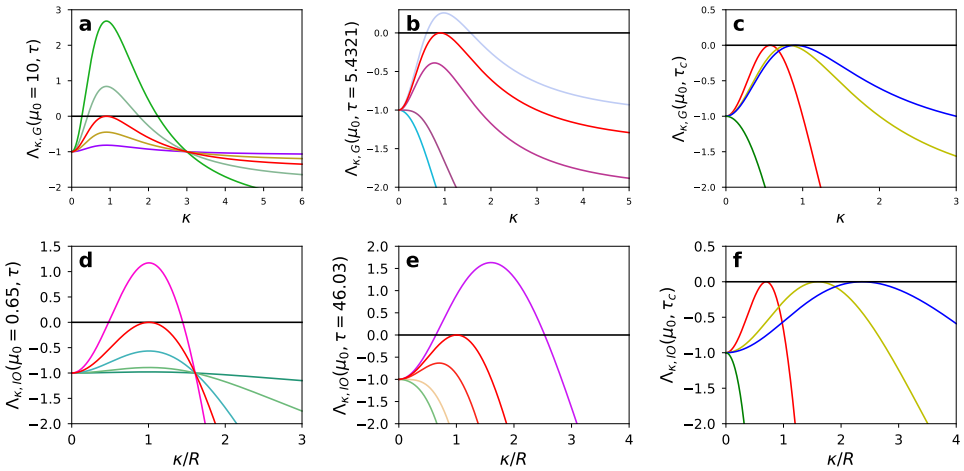

FIGURE A.1. **a** The function $\lambda_\kappa(\mu_0,\tau)$ from Equation A.6 with fixed μ_0 =10 and varying τ: 1,3,10,20. The red line corresponds to τ_c = 5.4321, as described by Equation A.8. **b** The function $\lambda_\kappa(\mu_0,\tau)$ from Equation A.6 with fixed τ =5.4321 and varying μ_0: 0.5, 1, 4, 40. The red line corresponds to μ_0=10 for which τ = 5.4321 is critical, as described by Equation A.8. **c** The function $\lambda_\kappa(\mu_0,\tau)$ from Equation A.6 at the critical value τ_c defined in Equation A.8 for different μ_0: 0.5(green), 2(red), 5(yellow), 10(blue). If $\mu_0 < 1$ the maximum of λ_κ is in $\kappa_m = 0$. As μ_0 grows, the maximum approaches the limiting value of $\kappa_m = 1$. **d** The function $\lambda_{\tilde{\kappa}}(\mu_0,\tau)$ from Equation A.18 with fixed μ_0 =0.65 and varying τ: 1,5,20,100. The red line corresponds to τ_c = 46.03, as described by Equation A.20. **e** The function $\lambda_{\tilde{\kappa}}(\mu_0,\tau)$ from Equation A.18 with fixed τ =46.03 and varying μ_0: 1/4, 1/3, 1/2, 1. The red line corresponds to μ_0 = 0.65, for which τ = 46.03 is critical (Equation A.20). **f** The function $\lambda_{\tilde{\kappa}}(\mu_0,\tau)$ from Equation A.18 at the critical value τ_c defined in Equation A.20 for different μ_0: 0.25(green), 0.5(red), 1(yellow), 1.5(blue). If $\mu_0 < 1/3$ the maximum of λ_κ is in $\kappa_m = 0$. As μ_0 grows, the maximum κ_m increases indefinitely.

A.1.2 Pattern formation and Growth in 2D

In 2 dimensions, Equation 2.1 describing the evolution of the density of population for a gravity model of human migration becomes a function of $\rho(x,y,t)$, the population density in 2 dimensions.

(A.11)

$$\frac{\partial\rho(x,y,t)}{\partial t} = g\rho(x,y,t)\left(1-\frac{\rho(x,y,t)}{\rho_0}\right) - T\rho(x,y,t) + T(\rho(x,y,t)+w).$$
$$\left(\int_0^\infty\int_0^{2\pi}\frac{f(r)\rho(r\cos(\theta)+x, r\sin(\theta)+y, t)}{\int_0^\infty\int_0^{2\pi} f(z)(\rho(r\cos(\theta)+x+z\cos(\phi), r\sin(\theta)+y+z\sin(\phi), t)+w)d\phi dz}d\theta dr\right).$$

Linearising this equation about the stationary solution, ρ_0, results in an expression for the evolution of a small perturbation in the population density $h_{k,l}$:

$$\begin{aligned}\frac{\partial h_{k,l}(t)}{\partial t} =& h_{k,l}(t)\Big\{T(\rho_0+w)\int_0^\infty\int_0^{2\pi}\Big(\frac{f(r)e^{i(kr\cos(\theta)+lr\sin(\theta))}}{\int_0^\infty 2\pi f(z)(\rho_0+w)dz}\\ &-\frac{\rho_0 f(r)\int_0^\infty 2\pi f(z)J_0\left(\sqrt{k^2+l^2}z\right)e^{i(kr\cos(\theta)+lr\sin(\theta))}dz}{\left(\int_0^\infty 2\pi f(z)(\rho_0+w)dz\right)^2}\Big)d\theta dr\\ &+T\rho_0\int_0^\infty\frac{f(r)}{\int_0^\infty f(z)(\rho_0+w)dz}dr-g-T\Big\}.\end{aligned} \tag{A.12}$$

The amplitude $h_{k,l}$ corresponds to a 2 D plane wave with wave-vector components k,l.

A.1.2.1 Exponential Deterrence Function

For the 2D case, the exponential form of $f(r)$ is given by $f(r)=e^{-Rr}$, where $r=\sqrt{x^2+y^2}$, the classic Euclidean distance. Substituting this into the expression for the evolution of $h_{k,l}$, Equation A.12, we obtain an expression for the growth function $\Lambda_{k,l}$:

$$\Lambda_{k,l}=T\left(\frac{\rho_0\left(k^2+l^2\right)}{(\rho_0+w)\left(k^2+l^2+R^2\right)}+\frac{R}{\sqrt{k^2+l^2+R^2}}-1\right)-g. \tag{A.13}$$

The function $\Lambda_{k,l}$ depends only on the magnitude of the wavevector, $p=\sqrt{k^2+l^2}$. Substituting p^2 for k^2+l^2 in Equation A.13 results in an identical expression as that for Λ_k, Equation A.5. Hence the critical value of the ratio between migration and growth above which cities are able to form is the same for both the 1D and 2D gravity models, given by Equation A.8.

A.1.2.2 Power Law Deterrence Function

The power law form of the deterrence function in 2D is given by $f(r)=(1+r)^{-\gamma}$, as for the 1D case, but with r equal to the Euclidean distance. Substituting this into the equation for the time evolution of a perturbation, Equation A.12, we find that an analytical expression for Λ is not obtainable. Solving the resulting equation for the formation of cities numerically, in analogy with the 2D model with an exponential $f(r)$, the critical curve of Equation A.8 is obtained. Summarising, in both 1 and 2 dimensions, if population dynamics are described by a logistic growth with individuals relocating according to a gravity model cities will form if $T>4g$ and $\rho_0 \gg w$ and this is independent of the deterrence function considered.

A.2 Intervening Opportunities & Radiation Models

A.2.1 Pattern formation and Growth in 1D

The full 1D dynamic equation for the intervening opportunities and radiation models is given by

$$\begin{aligned}\frac{\partial\rho(x,t)}{\partial t} =& g\rho(x,t)[1-\frac{\rho(x,t)}{\rho_0}]-T\rho(x,t)+T[\rho(x,t)+w].\\ &\int_0^\infty\left[\rho(x-r,t)f\left(\int_{x-2r}^{x}[\rho(z,t)+w]dz\right)+\rho(x+r,t)f\left(\int_x^{x+2r}[\rho(z,t)+w]dz\right)\right]dr.\end{aligned} \tag{A.14}$$

Using the same method as used for the gravity model, we write this equation for small fluctuations in $\rho(x,t)$, denoted $\tilde{\rho}(x,t)$, by linearising about the uniform stationary state ρ_0.

$$\begin{aligned}\frac{\partial\tilde{\rho}(x,t)}{\partial t} =& -(g+T)\tilde{\rho}(x,t)+2T\rho_0\tilde{\rho}(x,t)\int_0^\infty f(2r[\rho_0+w])dr+\\ &T[\rho_0+w]\int_0^\infty\Big(f(2r[\rho_0+w])[\tilde{\rho}(x-r,t)+\tilde{\rho}(x+r,t)]+\\ &\rho_0 f'(2r[\rho_0+w])\int_{x-2r}^{x+2r}\tilde{\rho}(z,t)dz\Big)dr.\end{aligned} \tag{A.15}$$

$$\begin{aligned}\Lambda_k =& -(g+T)+\frac{T\rho_0}{\rho_0+w}+\\ &2T[\rho_0+w]\left[\int_0^\infty cos(kr)f(2r[\rho_0+w])dr+\rho_0\int_0^\infty\frac{sin(2kr)f'(2r[\rho_0+w])}{k}dr\right].\end{aligned} \tag{A.16}$$

Introducing the rescaled variables: $\tau = T/g$, $\mu_0 = \rho_0/w$, $\kappa = k/w$ and $\lambda = \Lambda/g$, the above expression becomes:

$$\begin{aligned}\lambda_\kappa(\mu_0,\tau,f) =& -(1+\tau)+\frac{\mu_0}{\mu_0+1}+2\tau[\mu_0+1].\\ &\left[\int_0^\infty cos(\kappa z)f(2z[\mu_0+1])dz+\mu_0\int_0^\infty\frac{sin(2\kappa z)f'(2z[\mu_0+1])}{\kappa}dz\right].\end{aligned} \tag{A.17}$$

A.2.1.1 Exponential f(a) - Intervening Opportunities Model

For the intervening opportunities model, the deterrence function takes the form of an exponential; $f(a) = Re^{-Ra}$. With this, Equation A.17 is evaluated exactly to give

$$\lambda_{\tilde{\kappa}}(\mu_0,\tau) = -(1+\tau)+\tau\frac{\mu_0}{\mu_0+1}-\tau(\mu_0+1)\left[\frac{\mu_0}{\tilde{\kappa}^2+(\mu_0+1)^2}-\frac{4(\mu_0+1)}{\tilde{\kappa}^2+4(\mu_0+1)^2}\right], \tag{A.18}$$

where $\tilde{\kappa} = \kappa/R = k/(wR)$. The curve described by Equation A.18 is shown in Figures A.1d and A.1e for different values of the parameters.

The function $\lambda_{\tilde{\kappa}}(\mu_0,\tau)$ has a maximum in $\tilde{\kappa}_m = 0$ if $\mu_0 \leq 1/3$ or in

$$\tilde{\kappa}_m = \sqrt{\frac{-4\mu_0(\mu_0+2)+6\sqrt{\mu_0(\mu_0+1)^5}-4}{3\mu_0+4}} \tag{A.19}$$

otherwise. The parameter $\tilde{\kappa}_m$ corresponds to a wavenumber $k_m = Rw\tilde{\kappa}_m$. If $\lambda_{k_m} > 0$, cities are able to form and k_m is proportional to the number of cities per unit length. From the expression for k_m we find that, in contrast to the gravity model, as $\mu_0 \to \infty$ (ie $\rho_0 \gg w$), $k_m \to \infty$. This implies that for a fixed R, or travel distance, the density of cities will be greater in more populated regions.

The condition for $\lambda_{\tilde{\kappa}_m} > 0$ is found by inserting the expression for $\tilde{\kappa}_m$ into Equation A.17 and solving for τ, to obtain an expression for τ_c; if $\tau > \tau_c$, cities will emerge. τ_c is given by:

$$\tau_c = \frac{3(5\mu_0^3 + 11\mu_0^2 + 7\mu_0 + 4\sqrt{\mu_0(\mu_0+1)^5} + 1)}{(\mu_0+1)(3\mu_0-1)^2}. \tag{A.20}$$

It should be noted that this equation is independent of R. Furthermore, if $\mu_0 \gg 1$ ($\rho_0 \gg w$), the condition for growth reduces to $\tau > 3$, or $T > 3g$. Here we may draw a parallel with the gravity model; in both cases cities can only emerge if the migration rate is sufficiently higher than the growth rate.

A.2.1.2 Power Law f(a) - Radiation Model

For the radiation model, $f(a) = 1/(1+a)^2$ and, as with a power law deterrence function in the gravity model, the evaluation of Equation A.17 is more complicated. By solving $\lambda_{\tilde{\kappa}}(\mu_0, \tau, f) = 0$ numerically, we find that $\tilde{\kappa}_m = 0$ if $\mu_0 \leq 1$, and $\tilde{\kappa}_m > 0$ otherwise. From this we also obtain critical values of the parameters that will allow the emergence of cities. The results are shown in Figures 2.1c and 2.1d where it is seen that the critical curves for the intervening opportunities and radiation models are not the same; in contrast to gravity models, for intervening opportunities type models these curves do depend on the deterrence function used.

A.2.2 Pattern formation and Growth in 2D

The intervening opportunities model in 2 dimensions is

$$\begin{aligned} &\frac{\partial\rho(x,y,t)}{\partial t} = g\rho(x,y,t)\left(1 - \frac{\rho(x,y,t)}{u}\right) - T\rho(x,y,t) + T(\rho(x,y,t)+w). \\ &\left(\int_0^\infty \int_0^{2\pi} \rho(x + r\cos(\theta), y + r\sin(\theta), t). \right. \\ &\left. f\left[\int_0^r \int_0^{2\pi} (\rho(x + r\cos(\theta) + z\cos(\phi), y + r\sin(\theta) + z\sin(\phi), t) + w) d\phi dz\right] d\theta dr\right). \end{aligned} \tag{A.21}$$

This is analogous to the equation in 1 dimension, Equation A.14, however now we consider the opportunities within a circle of radius r, with r being the Euclidean distance between the origin and destination for a 2 dimensional population density $\rho(x,y,t)$.

Performing linear stability analysis on this equation, with a perturbation to the stationary distribution of the form $e^{i(kx+ly)}$, we find that the conditions for growth of cities are unchanged from the 1-dimensional case if the 1 dimensional wavenumber k is replaced by the magnitude of

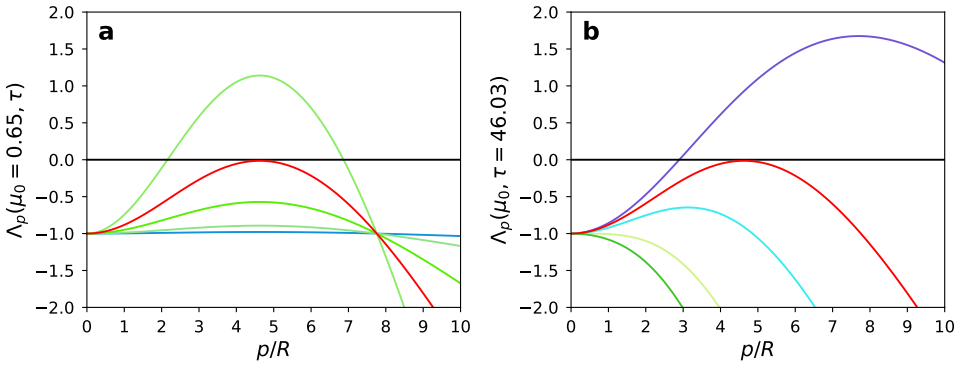

FIGURE A.2. **a** The function $\lambda_{\tilde{p}}(\mu_0, \tau)$ for fixed μ_0 = 0.65 and varied τ = 1,5,20,100. The red line corresponds to τ_c = 46.03, the same as the critical value for the 1D case. **b** The function $\lambda_{\tilde{p}}(\mu_0, \tau)$ for fixed τ = 46.03 and varied μ_0 = 1/4, 1/3,1/2,1. The red line corresponds μ_0 = 0.65, for which τ = 46.03 is critical.

the 2 dimensional wave-vector p. This result is demonstrated in Figures A.2a and A.2b where all parameters are equal to those used to generate Figures A.1d and A.1e respectively, for the 1D case. Here, $\tilde{p} = p/R$.

A.3 A stochastic model of population dynamics

We present a stochastic model of population dynamics which combines the models of migration, presented in Chapter 2, with Gibrat's proportionate random growth. Such an approach is expected to reproduce Heaps' Law for cities, Equation 2.13, and Zipf's law for city sizes.

For this approach, we model the population as individuals rather than a continuous density. Space is discretised into cells of size $(L_x/N_x)\times(L_y/N_y)$; L_x and L_y are the sizes of the 2-dimensional region being considered and N_x, N_y are the number of cells in each dimension. For simplicity, we will start by only considering square regions with $L_x = L_y = L$, and $N_x = N_y = N$, thus the total number of cells is N^2 and each has an area of $(L/N)^2 \equiv dl^2$. Each cell may be specified by a pair of coordinates (x, y) with $0 \leq x < N$ and $0 \leq y < N$. The population of cell (x, y) at time t is given by $n(x, y, t)$.

The population in any cell may change in two ways; natural increase and migrations. Here, natural increase refers to the difference between the number of births and the number of deaths within a cell therefore it may be both positive or negative. Migrations refers to any person relocating between cells; both inward migrations and outward migrations must be accounted for.

The mechanism that dictates the natural increase in the population of a cell is proportionate random growth, the growth mechanism known to produce Zipf's Law. If we consider the change in the population of cell (x, y) due to births and deaths between time t and $t + dt$, calling this change $\delta_g(x, y, t)$, then by the rule of proportionate random growth we have $\delta_g(x, y, t) = \xi(t)n(x, y, t)$. Here,

$\xi(t)$ is a normally distributed random variable with mean μ and variance σ^2. In order for the resulting distribution of city sizes to be a power-law [48], μ must be slightly negative, σ must be small (of order ≈ 0.1) and there must be a reflecting boundary condition preventing the population of a cell from going below a minimum value, $n_{min} = n_0$.

We model relocations between cells using either a gravity or an intervening opportunities model. To implement this, we define a probability of migration, T, which corresponds to the probability that an individual in cell (x,y) will relocate to any other cell between time t and $t+dt$. The total number of outgoing migrants from cell (x,y) in time dt, $m(x,y,t)$ is extracted from a binomial distribution with migration probability $T = 0.4$. We then relocate these outgoing migrants according to the gravity or intervening opportunities model as described in the following sections.

A.3.1 Gravity Model

If individuals relocate according to a gravity model, the probability of a migrant relocating from (x,y) to (i,j) is given by

$$P(x,y \to i,j) = \frac{F(r(x,y \to i,j))[n(i,j,t)+w(i,j,t)]}{\sum_{i,j} F(r(x,y \to i,j))[n(i,j,t)+w(i,j,t)]} \tag{A.22}$$

where $r(x,y \to i,j)$ is the euclidean distance between locations (x,y) and (i,j). The deterrence function F is assumed to be exponential, e^{-rR}.

A.3.2 Intervening Opportunities Model

According to the intervening opportunities model, the probability of a migrant relocating from (x,y) to (i,j) is given by

$$P(x,y \to i,j) = \left[\frac{F(A_{x,y \to i,j}) + F(A_{x,y \to i,j} + A_{i,j})}{F(n(x,y,t)+w(x,y)) - F(N)}\right] . \frac{n(i,j,t)+w(i,j)}{A_{i,j}} \tag{A.23}$$

where $A_{i,j}$ corresponds to the sum of the population and opportunities in all cells at an equal distance from (x,y) as (i,j) and F is the deterrence function that we assume to be exponential, e^{-AR}. $A_{x,y \to i,j}$ is the sum of the population and resources of all intervening cells; all cells that lie at a distance from (x,y) that is less than the distance to (i,j), including the population and resources of cell (x,y) itself.

Using Equations A.22 or A.23, we compute the probability of relocation to all cells according to the model in question. The number of migrants moving between any two locations are then extracted from a multinomial distribution. The number of incoming migrants to cell (x,y) in between time t and $t+dt$, $\delta_{in}(x,y,t+dt)$, is given by the sum of the outgoing migrants from all other cells that have destination (x,y). We denote this sum $\Phi(x,y,t)$.

Using this framework, the population dynamics can be simulated. In a time dt, the population in each cell will change due to natural increase, outgoing migrants and incoming migrants. The

population of a cell (x, y) after dt is therefore given by:

$$\begin{aligned} n(x,y,t+dt) &= n(x,y,t) + \delta_g(x,y,d) - \delta_{out}(x,y,d) + \delta_{in}(x,y,d) \\ &= n(x,y,t) + \xi(t)n(x,y,t) - m(x,y,t) + \Phi(x,y,t). \end{aligned} \tag{A.24}$$

Simulations start at time $t = 0$ from an initial condition where the population of each cell is drawn from a uniform distribution within the range $n_0 \leq n \leq 1.2 \cdot n_0$, where n_0 is the minimum population of a cell.

After each time step, dt, the state of the system is assessed by counting the total number of cities and their populations. Cities correspond to clusters of population and may therefore be defined as adjacent populated geographical spaces (cells) [107]. We define a cluster as a set of adjacent cells for which the Manhattan distance of a single cell from the closest cell that is a member of the cluster has a maximum of 1. This satisfies the adjacency condition and is equivalent to a condition on the cell edges; any cell that is a member of a cluster must share at least one edge with another cell within the same cluster. Alongside this, the population of every cell within the cluster must be greater than a minimum value X to ensure that all cells are populated.

A.4 Heaps' law for Continents

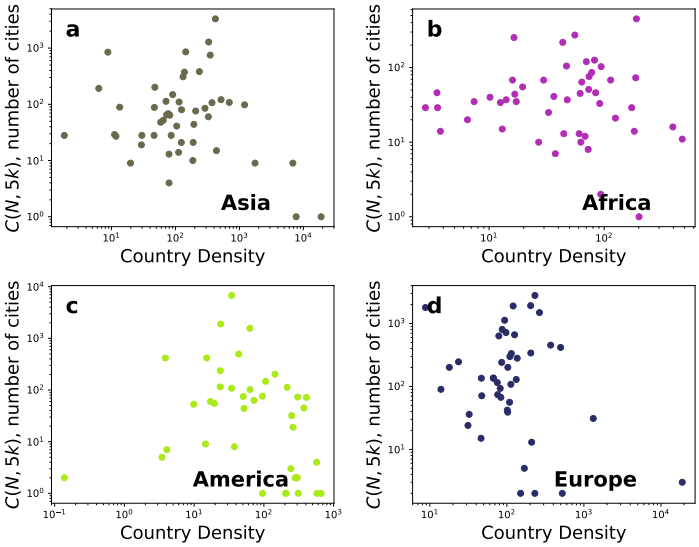

FIGURE A.3. Number of cities (y-axis) vs population density (x-axis) for countries in Europe, America, Asia and Africa. For all plots we observe that the cities-density correlation is very low (for values see table 2.2).

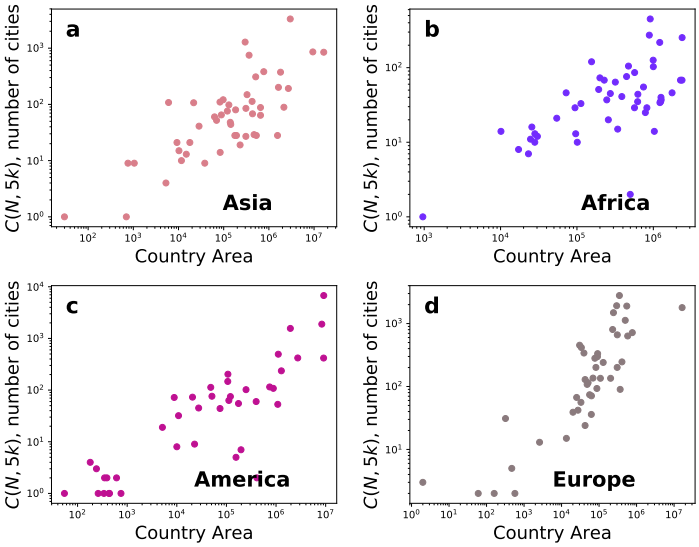

FIGURE A.4. Number of cities (y-axis) vs area (x-axis) for countries in Europe, America, Asia and Africa. For all plots we observe that the cities-area correlation is positive. The strongest correlation exists for countries in America however this is not as strong as the cities-total population correlation (see table 2.2).

APPENDIX B

SCIENTIFIC MIGRATION: TECHNICAL ASPECTS

B.1 Technical aspects: prediction 1

B.1.1 Logistic Regression

Logistic regression is a linear model used for classification. In this framework the probability for scientist s to move is given by the logistic function;

$$F(\boldsymbol{x}(s)) = \frac{1}{1 + e^{-\boldsymbol{x}(s)}}$$

where $\boldsymbol{x}(s) = \beta_0 + \sum_i \beta_i \boldsymbol{p}_i^{(h,t)}(s)$. If $F(\boldsymbol{x}(s)) > 0.5$ the model predicts that the scientist s will move, otherwise it predicts that they will stay. The coefficients, β_i, of each feature in $\vec{p}^{(h,t)}$ are determined using maximum likelihood estimation [87].

B.1.2 Decision Tree

A Decision tree is a classifier that predicts the class of a target using simple decision rules learned from the features of that target. The root of the tree contains all samples in the dataset. From the root there are branches and subsequent nodes. Starting with the root node, each node corresponds to a division of the dataset according to a rule. For example, if one of the features of the dataset is age, anyone less than 40 takes the left branch to a new leaf node and anyone greater than or equal to 40 takes the right branch. The feature value on which the node is split is selected such that it maximises the information (or minimises the entropy) of the leaf nodes. This process will terminate when some condition, such as a minimum number of samples in each leaf, has been met [78].

B.1.3 Evaluation measures

In order to evaluate the above models, we use compare them in terms of accuracy, recall, precision, F1-score and AUC, to a baseline classifier. We define P_t as the number of true positives, or scientists correctly predicted to move, and P_f as the number of false positives, or scientists incorrectly predicted to move. Analogously, we define the number of true negatives, scientists correctly predicted to stay, as N_t and N_f as the number of false negatives; scientists incorrectly predicted to stay. Using these definitions, accuracy refers to the number of true results ($T_p + T_f$) as a fraction of the total number of points in the dataset. Recall is given by $P_t/(P_t + N_f)$; it represents the number of scientists correctly predicted to move as a fraction of the total number of movers. Precision is the number of true positives as a fraction of the total number of positives returned: $P_t/(P_t + P_f)$. The F1-score is the harmonic mean of precision and recall. Finally, the AUC is the area under the graph when the recall is plotted against the fall - out, or false positive rate. Fall-out is given by $P_f/(P_t + N_f)$, defined such that the sum of recall and fall-out is equal to 1.

B.1.4 Baseline classifier

For comparison and completeness, we compare the results of both the logistic regression and decision tree classifiers to a baseline classifier. As our dataset is balanced a stratified classifier seems most informative. This classifier generates predictions according to the datasets class distribution, therefore we expect the baseline classifier to score ~ 0.5 for all evaluation measures.

B.2 Technical aspects: prediction 2

B.2.1 Stochastic gradient ascent

Stochastic gradient ascent (SDA) is an optimisation method [56]. For a given function $L(\omega)$ with unknown parameter ω, SDA finds the value of ω that maximises the likelihood of L. It approximates the gradient of L, $dL/d\omega$ by considering a single sample of a training set at each step. At each step, the value of ω is updated according to

$$\omega \leftarrow \omega + \eta \frac{dL}{d\omega}$$

.

where η corresponds to a step size or learning rate and controls the speed at which the algorithm converges on the optimal value of ω.

For our task of predicting where a scientist will relocate to, L is the model's log-likelihood and ω is a vector of the features of the gravity model. The optimisation is of the coefficients for each feature. We use mini-batch gradient ascent, where the gradient of L is approximated considering

a small subset of samples from the training set at each step in order to minimise computational time.

BIBLIOGRAPHY

[1] G. ABRAMO, A. CIRIACO D'ANGELO, AND G. MURGIA, *The relationship among research productivity, research collaboration, and their determinants*, Journal of Informetrics, 11 (2017), pp. 1016–1030.

[2] L. AITCHISON, N. CORRADI, AND P. E. LATHAM, *Zipf's law arises naturally when there are underlying, unobserved variables*, PLoS Computational Biology, 12 (2016), p. e1005110.

[3] S. APPELT, B. VAN BEUZEKOM, F. GALINDO-RUEDA, AND R. DE PINHO, *Which factors influence the international mobility of research scientists?*, OECD STI Working Papers, (2015).

[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 (2015), p. 20140745.

[5] L. AURIOL, *Careers of doctorate holders*, OECD STI Working Papers, 4 (2010).

[6] R. L. AXTELL, *Zipf distribution of us firm sizes*, Science, 293 (2001), pp. 1818–1820.

[7] S. AZAELE, R. MUNEEPEERAKUL, A. RINALDO, AND I. RODRIGUEZ-ITURBE, *Inferring plant ecosystem organization from species occurrences*, Journal of Theoretical Biology, 262 (2010), pp. 323–329.

[8] S. AZAELE, S. SUWEIS, J. GRILLI, I. VOLKOV, J. R. BANAVAR, AND A. MARITAN, *Statistical mechanics of ecological systems: Neutral theory and beyond*, Reviews of Modern Physics, 88 (2016), p. 035003.

[9] P. AZOULAY, I. GANGULI, AND J. G. ZIVIN, *The mobility of elite life scientists: Professional and personal determinants*, Research Policy, 46 (2017), pp. 573 – 590.

[10] S. K. BAEK, S. BERNHARDSSON, AND P. MINNHAGEN, *Zipf's law unzipped*, New Journal of Physics, 13 (2011), p. 043004.

[11] D. BALCAN, V. COLIZZA, B. GONCALVES, H. HU, J. J. RAMASCO, AND A. VESPIGNANI, *Multiscale mobility networks and the spatial spreading of infectious diseases*, Proceedings National Academy of Science USA, 106 (2009), pp. 21484–21489.

[12] D. BALCAN, V. COLIZZA, B. GONÇALVES, H. HU, J. J. RAMASCO, AND A. VESPIGNANI, *Modeling the spatial spread of infectious diseases: The global epidemic and mobility computational model*, Journal of Computational Science, 1 (2010), p. 132.

[13] A. BARBIERO AND P. A. FERRARI, *Simulation of correlated poisson variables*, Applied Stochastic Models in Business and Industry, 31 (2015), pp. 669–680.

[14] H. BARBOSA, M. BARTHELEMY, G. GHOSHAL, C. R. JAMES, M. LENORMAND, T. LOUAIL, R. MENEZES, J. J. RAMASCO, F. SIMINI, AND M. TOMASINI, *Human mobility: Models and applications*, Physics Reports, (2018).

[15] H. BARBOSA, F. B. DE LIMA-NETO, A. EVSUKOFF, AND R. MENEZES, *The effect of recency to human mobility*, EPJ Data Science, 4 (2015), pp. 1–14.

[16] M. BARTHELEMY, *Spatial networks*, Physics Reports, 499 (2011), pp. 1–101.

[17] S. BASU AND C. JONES, *On the power-law tail in the mass function of protostellar condensations and stars*, Monthly Notices of the Royal Astronomical Society, 347 (2004), pp. L47–L51.

[18] M. BATTY, *Hierarchy in cities and city systems*, in Hierarchy in Natural and Social Sciences, Springer, 2006, pp. 143–168.

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

[20] ——, *The size, scale, and shape of cities*, Science, 319 (2008), pp. 769–771.

[21] ——, *Cities as complex systems: Scaling, interaction, networks, dynamics and urban morphologies.*, (2009).

[22] M. BATTY AND P. A. LONGLEY, *Urban shapes as fractals*, Area, (1987), pp. 215–221.

[23] L. M. BETTENCOURT, *The origins of scaling in cities*, Science, 340 (2013), pp. 1438–1441.

[24] L. M. BETTENCOURT, *Cities as complex systems*, Modeling Complex Systems for Public Policies, (2015), pp. 217–238.

[25] L. M. BETTENCOURT, J. LOBO, D. HELBING, C. KÜHNERT, AND G. B. WEST, *Growth, innovation, scaling, and the pace of life in cities*, Proceedings of the National Academy of Sciences, 104 (2007), pp. 7301–7306.

[26] D. BROCKMANN, L. HUFNAGEL, AND T. GEISEL, *The scaling laws of human travel*, Nature, 439 (2006), pp. 462–465.

[27] CENTER FOR INTERNATIONAL EARTH SCIENCE INFORMATION NETWORK COLUMBIA UNIVERSITY, *Global rural urban mapping project (GRUMP) alpha: Gridded population of the world, version 2, with urban reallocation (GPW-UR).*, 2007.

[28] E. CHO, S. A. MYERS, AND J. LESKOVEC, *Friendship and mobility: User movement in location-based social networks*, in Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '11, New York, NY, USA, 2011, ACM, pp. 1082–1090.

[29] W. CHRISTALLER, *Central places in southern Germany*, Prentice-Hall, 1966.

[30] A. CLAUSET, C. R. SHALIZI, AND M. E. NEWMAN, *Power-law distributions in empirical data*, SIAM Review, 51 (2009), pp. 661–703.

[31] V. COLIZZA, A. FLAMMINI, M. A. SERRANO, AND A. VESPIGNANI, *Detecting rich-club ordering in complex networks*, Nature Physics, 2 (2006), pp. 110–115.

[32] C. COTTINEAU, O. FINANCE, E. HATNA, E. ARCAUTE, AND M. BATTY, *Defining urban clusters to detect agglomeration economies*, Environment and Planning B: Urban Analytics and City Science, (2018), p. 2399808318755146.

[33] C. COTTINEAU, E. HATNA, E. ARCAUTE, AND M. BATTY, *Diverse cities or the systematic paradox of urban scaling laws*, Computers, Environment and Urban Systems, 63 (2017), pp. 80–94.

[34] J. DE DIOS ORTÚZAR AND L. WILLUMSEN, *Modeling Transport*, John Wiley and Sons Ltd, New York, 2011.

[35] M. A. DE MENEZES AND A.-L. BARABÁSI, *Fluctuations in network dynamics*, Physical Review Letters, 92 (2004), p. 028701.

[36] P. DEVILLE, D. WANG, R. SINATRA, C. SONG, V. D. BLONDEL, AND A.-L. BARABÁSI, *Career on the move: Geography, stratification, and scientific impact*, Scientific Reports, 4 (2014), pp. 4770 EP –.

[37] P. S. DODDS, D. R. DEWHURST, F. F. HAZLEHURST, C. M. VAN OORT, L. MITCHELL, A. J. REAGAN, J. R. WILLIAMS, AND C. M. DANFORTH, *Simon's fundamental rich-get-richer model entails a dominant first-mover advantage*, Physical Review E, 95 (2017), p. 052301.

[38] J. EECKHOUT, *Gibrat's law for (all) cities*, The American Economic Review, 94 (2004), pp. 1429–1451.

[39] Z. EISLER, I. BARTOS, AND J. KERTÉSZ, *Fluctuation scaling in complex systems: Taylor's law and beyond*, Advances in Physics, 57 (2008), pp. 89–142.

[40] S. ERLANDER AND N. F. STEWART, *The Gravity model in transportation analysis: theory and extensions*, Topics in Transportation, VSP, Utrecht, The Netherlands, 1990.

[41] P. EXPERT, T. S. EVANS, V. D. BLONDEL, AND R. LAMBIOTTE, *Uncovering space-independent communities in spatial networks*, Proceedings of the National Academy of Sciences, 108 (2011), p. 7663.

[42] S. FORTUNATO, C. T. BERGSTROM, K. BÖRNER, J. A. EVANS, D. HELBING, S. MILOJEVIĆ, A. M. PETERSEN, F. RADICCHI, R. SINATRA, B. UZZI, A. VESPIGNANI, L. WALTMAN, D. WANG, AND A.-L. BARABÁSI, *Science of science*, Science, 359 (2018).

[43] C. FRANZONI, G. SCELLATO, AND P. STEPHAN, *The mover's advantage: The superior performance of migrant scientists*, Economics Letters, 122 (2014), pp. 89 – 93.

[44] ——, *Chapter 2 - international mobility of research scientists: Lessons from globsci*, in Global Mobility of Research Scientists, A. Geuna, ed., Academic Press, San Diego, 2015, pp. 35 – 65.

[45] G. F. FRASCO, J. SUN, H. D. ROZENFELD, AND D. BEN-AVRAHAM, *Spatially distributed social complex networks*, Physical Review X, 4 (2014), p. 011008.

[46] M. FUENTES, M. KUPERMAN, AND V. KENKRE, *Nonlocal interaction effects on pattern formation in population dynamics*, Physical Review Letters, 91 (2003), p. 158104.

[47] M. FUJITA, P. R. KRUGMAN, A. J. VENABLES, AND M. FUJITA, *The spatial economy: cities, regions and international trade*, vol. 213, Wiley Online Library, 1999.

[48] X. GABAIX, *Zipf's law for cities: an explanation*, Quarterly Journal of Economics, (1999), pp. 739–767.

[49] X. GABAIX AND Y. M. IOANNIDES, *The evolution of city size distributions*, Handbook of Regional and Urban Economics, 4 (2004), pp. 2341–2378.

[50] F. GARGIULO AND T. CARLETTI, *Driving forces of researchers mobility*, Scientific Reports, 4 (2014), p. 4860.

[51] M. GERLACH AND E. G. ALTMANN, *Scaling laws and fluctuations in the statistics of word frequencies*, New Journal of Physics, 16 (2014), p. 113010.

[52] R. GIBRAT, *Les inégalités économiques*, Recueil Sirey, 1931.

[53] K. GIESEN AND J. SÜDEKUM, *Zipf's law for cities in the regions and the country*, Journal of Economic Geography, (2010), p. lbq019.

[54] A. GIOMETTO, M. FORMENTIN, A. RINALDO, J. E. COHEN, AND A. MARITAN, *Sample and population exponents of generalized Taylor's law*, Proceedings of the National Academy of Sciences, 112 (2015), pp. 7755–7760.

[55] M. C. GONZALEZ, C. A. HIDALGO, AND A.-L. BARABASI, *Understanding individual human mobility patterns*, Nature, 453 (2008), pp. 779–782.

[56] P. HARRINGTON, *Machine learning in action*, vol. 5, Manning Greenwich, CT, 2012.

[57] T. E. HARRIS, *The theory of branching processes*, Courier Dover Publications, 2002.

[58] H. S. HEAPS, *Information retrieval: Computational and theoretical aspects*, Academic Press, Inc., 1978.

[59] A. HERNANDO, R. HERNANDO, AND A. PLASTINO, *Space-time correlations in urban sprawl*, Journal of the Royal Society Interface, 11 (2013), p. 20130930.

[60] R. IREDALE, *High-Skilled Migration*, Springer Netherlands, Dordrecht, 2016, pp. 1–10.

[61] M. JEFFERSON, *Why geography? the law of the primate city*, Geographical Review, 79 (1989), pp. 226–232.

[62] W. S. JUNG, F. WANG, AND H. E. STANLEY, *Gravity model in the Korean highway*, EuroPhysics Letters, 81 (2008), p. 48005.

[63] P. KALUZA, A. KÖLZSCH, M. T. GASTNER, AND B. BLASIUS, *The complex network of global cargo ship movements*, Journal of The Royal Society Interface, 7 (2010), pp. 1093–1103.

[64] D. KAREMERA, V. I. OGULEDO, AND B. DAVIS, *A gravity model analysis of international migration to North America*, Applied Economics, 32 (2000), pp. 1745–1755.

[65] G. KRINGS, F. CALABRESE, C. RATTI, AND V. D. BLONDEL, *Urban gravity: a model for inter-city telecommunication flows*, Journal of Statistical Mechanics: Theory and Experiment, 2009 (2009), p. L07003.

[66] P. KRUGMAN, *The Self Organizing Economy*, Blackwell Publishers, 1996.

[67] C. KÜHNERT, D. HELBING, AND G. B. WEST, *Scaling laws in urban supply networks*, Physica A: Statistical Mechanics and its Applications, 363 (2006), pp. 96–103.

[68] H. C. LAWSON AND J. A. DEARINGER, *A comparison of four work trip distribution models*, Proceedings of American Society of Civil Engineering, 93 (1967), pp. 1–25.

[69] J. C. LEITÃO, J. M. MIOTTO, M. GERLACH, AND E. G. ALTMANN, *Is this scaling nonlinear?*, Royal Society Open Science, 3 (2016), p. 150649.

[70] M. LENORMAND, A. BASSOLAS, AND J. J. RAMASCO, *Systematic comparison of trip distribution laws and models*, Journal of Transport Geography, 51 (2016), pp. 158–169.

[71] M. LENORMAND, S. HUET, F. GARGIULO, AND G. DEFFUANT, *A Universal Model of Commuting Networks*, PLoS ONE, 7 (2012), p. e45985.

[72] F. LEYVRAZ AND S. REDNER, *Scaling theory for migration-driven aggregate growth*, Physical Review Letters, 88 (2002), p. 068301.

[73] L. LI, H. JING, H. TONG, J. YANG, Q. HE, AND B.-C. CHEN, *Nemo: Next career move prediction with contextual embedding*, in Proceedings of the 26th International Conference on World Wide Web Companion.

[74] W. LI, *Random texts exhibit zipf's-law-like word frequency distribution*, IEEE Transactions on Information Theory, 38 (1992), pp. 1842–1845.

[75] X. LI, H. TIAN, D. LAI, AND Z. ZHANG, *Validation of the gravity model in predicting the global spread of influenza*, International Journal of Environmental Research and Public Health, 8 (2011), pp. 3134–3143.

[76] X. LIANG, J. ZHAO, L. DONG, AND K. XU, *Unraveling the origin of exponential law in intra-urban human mobility*, Scientific Reports, 3 (2013).

[77] C. LINARD, A. J. TATEM, AND M. GILBERT, *Modelling spatial patterns of urban growth in Africa*, Applied Geography, 44 (2013), pp. 23–32.

[78] R. LIOR ET AL., *Data mining with decision trees: theory and applications*, vol. 81, World Scientific, 2014.

[79] X. F. LIU, Y.-L. LIU, X.-H. LU, Q.-X. WANG, AND T.-X. WANG, *The anatomy of the global football player transfer network: Club functionalities versus network properties*, PLOS ONE, 11 (2016), pp. 1–14.

[80] R. LOUF AND M. BARTHELEMY, *Modeling the polycentric transition of cities*, Physical Review Letters, 111 (2013), p. 198702.

[81] L. LÜ, Z.-K. ZHANG, AND T. ZHOU, *Zipf's law leads to heaps' law: Analyzing their relation in finite-size systems*, PloS ONE, 5 (2010), p. e14139.

[82] H. A. MAKSE, J. S. ANDRADE, M. BATTY, S. HAVLIN, H. E. STANLEY, ET AL., *Modeling urban growth patterns with correlated percolation*, Physical Review E, 58 (1998), p. 7054.

[83] S. C. MANRUBIA AND D. H. ZANETTE, *Intermittency model for urban development*, Physical Review E, 58 (1998), p. 295.

[84] M. MARSILI AND Y.-C. ZHANG, *Interacting individuals leading to zipf's law*, Physical Review Letters, 80 (1998), p. 2741.

[85] A. P. MASUCCI, J. SERRAS, A. JOHANSSON, AND M. BATTY, *Gravity versus radiation models: On the importance of scale and heterogeneity in commuting flows*, Physical Review E, 88 (2013), p. 022812.

[86] G. MCNICOLL, *United Nations, Department of Economic and Social Affairs: World economic and social survey 2004: International migration*, Population and Development Review, 31 (2005), pp. 183–185.

[87] S. MENARD, *Applied logistic regression analysis*, vol. 106, Sage, 2002.

[88] H. F. MOED, M. AISATI, AND A. PLUME, *Studying scientific migration in scopus*, Scientometrics, 94 (2013), pp. 929–942.

[89] N. MOHAJERI, P. LONGLEY, AND M. BATTY, *City shape and the fractality of street patterns*, Quaestiones Geographicae, 31 (2012), pp. 29–37.

[90] C. Z. MOONEY, R. D. DUVAL, AND R. DUVALL, *Bootstrapping: A nonparametric approach to statistical inference*, no. 94-95, Sage, 1993.

[91] U. NATIONS, *Principles and recommendations for population and housing censuses*, 1997.

[92] ——, *Global city population estimates*, 2014.

[93] J. A. NELDER AND R. J. BAKER, *Generalized linear models*, Wiley Online Library, 1972.

[94] M. E. NEWMAN, *The structure of scientific collaboration networks*, Proceedings of the National Academy of Sciences, 98 (2001), pp. 404–409.

[95] ——, *Power laws, pareto distributions and zipf's law*, Contemporary Physics, 46 (2005), pp. 323–351.

[96] R. V. NOORDEN, *Global mobility: Science on the move*, Nature, 490 (2012), pp. 326–329.

[97] I. PAPARRIZOS, B. B. CAMBAZOGLU, AND A. GIONIS, *Machine learned job recommendation*, in Proceedings of the Fifth ACM Conference on Recommender Systems, RecSys '11, New York, NY, USA, 2011, ACM, pp. 325–328.

[98] L. PAPPALARDO, S. RINZIVILLO, AND F. SIMINI, *Human mobility modelling: Exploration and preferential return meet the gravity model*, Procedia Computer Science, 83 (2016), pp. 934 – 939.
The 7th International Conference on Ambient Systems, Networks and Technologies (ANT 2016) / The 6th International Conference on Sustainable Energy Information Technology (SEIT-2016) / Affiliated Workshops.

[99] R. PATUELLI, A. REGGIANI, S. P. GORMAN, P. NIJKAMP, AND F.-J. BADE, *Network analysis of commuting flows: A comparative static approach to German data*, Networks and Spatial Economics, 7 (2007), pp. 315–331.

[100] N. PERRA, B. GONÇALVES, R. PASTOR-SATORRAS, AND A. VESPIGNANI, *Activity driven modeling of time varying networks*, Scientific Reports, 2 (2012).

[101] C. POTTER, *New questions in the 1940 census*, Prologue-Quarterly of the National Archives and Records Administration, 42 (2010), pp. 46–52.

[102] C. E. PYERS, *Evaluation of intervening opportunities trip distribution models*, Highway Research Record, 114 (1966), pp. 71–88.

[103] M. RATCLIFFE, C. BURD, K. HOLDER, AND A. FIELDS, *Defining rural at the US Census Bureau*, American Community Survey and Geography Brief, (2016).

[104] S. REDNER, *Random multiplicative processes: An elementary tutorial*, American Journal of Physics, 58 (1990), pp. 267–273.

[105] D. S. REHER, *Baby booms, busts, and population ageing in the developed world*, Population Studies, 69 (2015), pp. S57–S68.

[106] K. T. ROSEN AND M. RESNICK, *The size distribution of cities: an examination of the pareto law and primacy*, Journal of Urban Economics, 8 (1980), pp. 165–186.

[107] 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 (2008), pp. 18702–18707.

[108] D. RYBSKI, A. G. C. ROS, AND J. P. KROPP, *Distance-weighted city growth*, Physical Review E, 87 (2013), p. 042114.

[109] G. SCELLATO, C. FRANZONI, AND P. STEPHAN, *Migrant scientists and international networks*, Research Policy, 44 (2015), pp. 108 – 120.

[110] G. SCELLATO, C. FRANZONI, AND P. STEPHAN, *A mobility boost for research*, Science, 356 (2017), pp. 694–694.

[111] M. SCHLÄPFER, L. M. BETTENCOURT, S. GRAUWIN, M. RASCHKE, R. CLAXTON, Z. SMOREDA, G. B. WEST, AND C. RATTI, *The scaling of human interactions with city size*, Journal of the Royal Society Interface, 11 (2014), p. 20130789.

[112] D. J. SCHWAB, I. NEMENMAN, AND P. MEHTA, *Zipf's law and criticality in multivariate data without fine-tuning*, Physical Review Letters, 113 (2014), p. 068102.

[113] K. C. SETO, B. GÜNERALP, AND L. R. HUTYRA, *Global forecasts of urban expansion to 2030 and direct impacts on biodiversity and carbon pools*, Proceedings of the National Academy of Sciences, 109 (2012), pp. 16083–16088.

[114] F. SIMINI, M. C. GONZÁLEZ, A. MARITAN, AND A.-L. BARABÁSI, *A universal model for mobility and migration patterns*, Nature, 484 (2012), pp. 96–100.

[115] F. SIMINI AND C. JAMES, *Discovering the laws of urbanisation*, ArXiv e-prints, (2015).

[116] F. SIMINI, A. MARITAN, AND Z. NÉDA, *Human mobility in a continuum approach*, PLoS ONE, 8 (2013), p. e60069.

[117] H. A. SIMON, *On a class of skew distribution functions*, Biometrika, (1955), pp. 425–440.

[118] R. SINATRA, D. WANG, P. DEVILLE, C. SONG, AND A.-L. BARABÁSI, *Quantifying the evolution of individual scientific impact*, Science, 354 (2016).

[119] C. SONG, T. KOREN, P. WANG, AND A.-L. BARABÁSI, *Modelling the scaling properties of human mobility*, Nature Physics, 6 (2010), pp. 818–823.

[120] K. T. SOO, *Zipf's law for cities: a cross-country investigation*, Regional Science and Urban Economics, 35 (2005), pp. 239–263.

[121] D. SORNETTE, *Multiplicative processes and power laws*, Physical Review E, 57 (1998), p. 4811.

[122] D. SORNETTE, *Dragon-kings, black swans and the prediction of crises*, ArXiv preprint arXiv:0907.4290, (2009).

[123] STATE DATA CENTER, *State Library of Iowa*, 2015.

[124] J. STILLWELL, K. DARAS, M. BELL, AND N. LOMAX, *The image studio: A tool for internal migration analysis and modelling*, Applied Spatial Analysis and Policy, 7 (2014), pp. 5–23.

[125] S. A. STOUFFER, *Intervening opportunities: A theory relating mobility and distance*, American Sociological Review, 5 (1940), pp. 845–867.

[126] C. R. SUGIMOTO, *Scientists have most impact when they're free to move*, Nature, 550 (2017), pp. 29–31.

[127] THE WORLD BANK, *World development indicators (2010)*.

[128] UNITED NATIONS, DEPARTMENT OF ECONOMIC AND SOCIAL AFFAIRS, POPULATION DIVISION, *The world's cities in 2016*, 2016.

[129] US CENSUS BUREAU, *County population totals*, 1970-2010.

[130] B. VATANT AND M. WICK, *Geonames ontology*, 2012.

[131] I. VOLKOV, J. R. BANAVAR, S. P. HUBBELL, AND A. MARITAN, *Inferring species interactions in tropical forests*, Proceedings of the National Academy of Sciences, 106 (2009), pp. 13854–13859.

[132] D. WANG, D. PEDRESCHI, C. SONG, F. GIANNOTTI, AND A.-L. BARABASI, *Human mobility, social ties, and link prediction*, in Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining, ACM, 2011, pp. 1100–1108.

[133] D. WANG, C. SONG, AND A.-L. BARABÁSI, *Quantifying long-term scientific impact*, Science, 342 (2013), pp. 127–132.

[134] N. A. WEISS, *A course in probability*, Addison-Wesley, 2006.

[135] A. G. WILSON, *Entropy in urban and regional modelling*, London: Pion, 1970.

[136] ——, *Urban and regional models in geography and planning*, Wiley, New York, 1970.

[137] D. K. WITHEFORD, *Comparison of trip distribution by opportunity model and gravity model*, Pittsburgh Area Transportation Study, 1961.

[138] Y. YANG, C. HERRERA, N. EAGLE, AND M. C. GONZÁLEZ, *Limits of predictability in commuting flows in the absence of data for calibration*, Scientific Reports, 4 (2014).

[139] G. U. YULE, *A mathematical theory of evolution, based on the conclusions of Dr. JC Willis, FRS*, Philosophical Transactions of the Royal Society of London.Series B, Containing Papers of a Biological Character, (1925), pp. 21–87.

[140] D. ZANETTE AND M. MONTEMURRO, *Dynamics of text generation with realistic zipf's distribution*, Journal of Quantitative Linguistics, 12 (2005), pp. 29–40.

[141] D. H. ZANETTE AND S. C. MANRUBIA, *Role of intermittency in urban development: a model of large-scale city formation*, Physical Review Letters, 79 (1997), p. 523.

[142] Y. B. ZEL'DOVICH, S. MOLCHANOV, AND RUZMAĬ, *Intermittency in random media*.

[143] F. ZHAO, L.-F. CHOW, M.-T. LI, A. GAN, AND S. D. L., *Refinement of FSUTMS trip distribution methodology*, tech. rep., Technical Memorandum 3, Florida International University, 2001.

[144] M. ZHAO, L. MASON, AND W. WANG, *Empirical study on human mobility for mobile wireless networks*, in Military Communications Conference, 2008. MILCOM 2008. IEEE, IEEE, 2008, pp. 1–7.

[145] G. K. ZIPF, *The P1 P2/D hypothesis: On the intercity movement of persons*, American Sociological Review, 11 (1946), pp. 677–686.

[146] ——, *Human behavior and the principle of least effort: An introduction to human ecology*, Ravenio Books, 2016.