Università degli studi di Padova
Facoltà di Scienze MM.FF.NN.
Dipartimento di Fisica "G. Galilei"

SCUOLA DI DOTTORATO DI RICERCA IN FISICA
CICLO XXII

# EMERGENCE OF SCALING IN ECOLOGICAL COMMUNITIES FROM TROPICAL FORESTS TO HUMAN MOBILITY

**Direttore della Scuola** Ch.mo Prof. Attilio Stella

**Supervisore** Ch.mo Prof. Amos Maritan

**Dottorando** Dott. Filippo Simini

## Abstract

This work is mainly focused on the study of the interrelationships between transportation networks and the structure of the ecological communities in which the transport takes place.

Transportation is a common need in many and diverse natural systems composed by individuals that interact with each other by competing for the same resources or by cooperation.

For the cases we investigated we found that the features of the transportation system are closely related to the structure of the community. For example in tropical forests we showed that the optimization of the transportation of water and resources to the leaves poses strict restrictions on the distribution of tree sizes. Also, in the context of human mobility we found that the commuting fluxes of people within a country are driven by the spatial distribution of population.

In particular we discovered that in both tropical forests and human settlements the adaptation/evolution of the transportation network responds to simple optimality principles of efficiency, distinctive of each system.

In forests, the transportation network is constituted by trees, whose number, size and shape characterize the efficiency in the transportation of water. The relationship between size and abundance as well as the scaling relationships between diameter, height, crown extension can be determined by assuming that each tree maximizes its metabolic rate for a given mass, and that the entire forest fully utilizes all available resources.

The transportation network describing human mobility is defined as a weighted network where the nodes are the locations (e.g. municipalities or counties) and a link of weight $w$ between two nodes indicates a commuting flux of $w$ people

among the two locations. Again, we found that one is able to reproduce the observed mobility patterns with remarkable accuracy assuming that each individual choses her/his destination trying to balance between the attractiveness versus the distance of the locations (i.e. choosing the closest location with higher attractiveness).

Scaling plays a major role in our analysis.

In both systems we were able to identify scaling relationships between the important variables, and thanks to finite-size scaling we succeeded in linking together seemingly unrelated quantities. This provided new insights and helped in discovering universal relationships.

## Sommario

Il principale obiettivo di questo lavoro è lo studio delle relazioni tra le reti di trasporto e la struttura delle comunità ecologiche in cui il trasporto ha luogo.
Il trasporto è un'esigenza comune in molti dei sistemi naturali costituiti da individui che interagiscono tra loro attraverso la competizione per le stesse risorse o tramite meccanismi di cooperazione.
Nei casi che abbiamo trattato abbiamo riscontrato che le caratteristiche del sistema di trasporto sono strettamente legate alla struttura della comunità. Per esempio, abbiamo mostrato come nelle foreste tropicali l'ottimizzazione del trasporto dell'acqua e delle sostanze nutritive alle foglie impone dei rigidi vincoli alla distribuzione delle dimensioni degli alberi.
Inolte, nel contesto della 'mobilità umana' abbiamo dimostrato che i flussi dei pendolari all'interno di uno stato sono determinati dalla distribuzione spaziale della popolazione.
In particolare abbiamo scoperto che sia nelle foreste tropicali che negli insediamenti umani l'adattamento e l'evoluzione della rete di trasporto sono determinati da semplici principi di ottimizzazione dell'efficienza, caratteristici del particolare sistema.
Nelle foreste la rete di trasporto è costituita dagli alberi, il cui numero, dimensioni e forma determinano l'efficienza nel trasporto dell'acqua. Il legame tra le dimensioni e la numerosità degli alberi e le relazioni di scala tra diametro, altezza, raggio della chioma, possono essere determinate assumendo che ogni albero massimizza il metabolismo a parità di massa, e che l'intera foresta utilizza tutte le risorse disponibili.
D'altro canto, la rete di trasporto con cui si descrive la mobilità umana è un grafo pesato, in cui i nodi sono i luoghi (le municipalità o le contee) e un legame con peso $w$ tra due nodi indica un flusso di $w$ persone tra i due luoghi. Ancora

una volta abbiamo trovato che è possibile riprodurre con sorprendente precisione i flussi osservati assumendo che ogni individuo scelga la sua destinazione cercando di bilanciare l'attrattività con la distanza dal luogo in cui si trova (scegliendo cioè la più vicina località che abbia una attrattiva maggiore del luogo in cui si trova). Lo scaling ha un ruolo fondamentale nella nostra analisi. In entrambi i contesti siamo stati in grado di identificare le relazioni di scala tra le principali variabili e, grazie alle tecniche delle leggi di scala finite, siamo riusciti a trovare un legame tra quantità apparentemente indipendenti. Questo ci ha portato a una comprensione più profonda dei sistemi analizzati e ci ha permesso di scoprire l'esistenza di relazioni di universalità.

# Contents

# Chapter 1

# Introduction

This work is mainly focused on the study of the interrelationships between transportation networks and the structure of the ecological communities in which the transport takes place.

Transportation is a common need in many and diverse natural systems composed by individuals that interact with each other by competing for the same resources or by cooperation.

For the cases we investigated we found that the features of the transportation system are closely related to the structure of the community. For example in tropical forests we showed that the optimization of the transportation of water and resources to the leaves poses strict restrictions on the distribution of tree sizes. Also, in the context of human mobility we found that the commuting fluxes of people within a country are driven by the spatial distribution of population.

In particular we discovered that in both tropical forests and human settlements the adaptation/evolution of the transportation network responds to simple optimality principles of efficiency, distinctive of each system.

In forests, the transportation network is constituted by trees, whose number, size and shape characterize the efficiency in the transportation of water. The relationship between size and abundance as well as the scaling relationships between diameter, height, crown extension can be determined by assuming that each tree

maximizes its metabolic rate for a given mass, and that the entire forest fully utilizes all available resources.
The transportation network describing human mobility is defined as a weighted network where the nodes are the locations (e.g. municipalities or counties) and a link of weight $w$ between two nodes indicates a commuting flux of $w$ people between the two locations. Again, we found that one is able to reproduce the observed mobility patterns with remarkable accuracy assuming that each individual choses her/his destination trying to balance between the attractiveness versus the distance of the locations (i.e. choosing the closest location with higher attractiveness).

It is thus useful to briefly recall some noteworthy results on the theory of transportation networks, systems in which the "matter" produced (or absorbed) by a given region –spanned e.g. by a regular lattice or a network– is collected at (distributed from) one or more sinks (sources).
There are many examples of both natural and artificial systems that can be described in this way: the network of tributaries in fluvial basins (collection of rain water), circulatory system in animals (delivery of oxigen to the cells), xilems and roots in vascular plants (delivery of water to the leaves), car traffic in a urban area.
Several studies have pointed out that the properties and configurations of transportation networks are usually selected with respect to their efficiency (*1*).
For example the network of tributaries in fluvial basins emerges as a local optimal configuration that minimizes the cost of downhill transportation of rain water (*2*). More specifically, the total power dissipated in the whole basin can be expressed as a function of the water currents in every channel $I_b$, $C\left(\{I_b\}\right) \propto \sum_{b\in \text{channels}} |I_b|^\gamma$, where $\gamma \simeq 1/2$. Scaling exponents characterizing the shape of the basin and the winding of tributaries can be calculated analytically for the intensities of water

currents $I_b$ that correspond to the *global* minimum of $C$ (*3*). The configurations of *local* minima of the cost function $C$, found through numerical simulations, have shape and winding exponents consistent with the ones measured in real basins. More generally it has been proved that every cost function $C$ with the above form and with $0 < \gamma \leq 1$ has local minima that are directed spanning trees, where currents tend to merge without forming loops (*3*).

Recently the case $\gamma > 1$ has also been considered (*4*). In this case the cost function $C$ is convex and thus only one global minimum exists. The optimal currents configuration exhibits a probability distribution function characterized by a scaling behavior given by $P(> I|L) \sim I^{1-\tau}$ for $L \lesssim I \lesssim L^2$ (here $L$ is the system size) with $\tau = (2d-1)/(d-1)$, and contrarily to the case $\gamma < 1$ currents tend to split forming loops. The scaling exponent of the current distribution proves robust with respect to: (*i*) the choice of the transportation cost, as far as it is convex and has finite first derivatives with respect to the currents; (*ii*) the distribution of injected currents; (*iii*) position–dependent (convex) cost functions. The analytical results show that the exponent of the asymptotic power–law behavior of the current probability distribution function varies continuously from $\tau = 3$ in two dimensions to $\tau = 2$ at infinite dimensions with no evidence of an upper critical dimension.

Other variational principles have been proposed in order to select the network configurations which are optimal for navigation. Indeed it has been investigated which configuration permits to minimize the average distance between any two nodes for individuals who navigate the network not knowing its global structure (*5*). The network is built adding long-range connections to a two-dimensional lattice: a pair of sites $i$, $j$ is randomly chosen to receive a long range connection with probability proportional to $r_{ij}^{-\alpha}$ ($r_{ij}$ is the euclidean distance between $i$ and $j$). It has been proved that the network configuration with exponent $\alpha = 2$ minimizes the average length of the travel between any two nodes when the individual choses

his/her next step having only local information, i.e. moving to the neighbor that has the small Manhattan distance from the target.
Furthermore, when a global constraint on the total cost of the long range connections is introduced (i.e. $\sum r_{ij} = \Lambda$), the network configuration with $\alpha = 3$ has proved to minimize the average distance for navigators with both local *and* global information (*6*). Indeed, this result seems to be supported by experimental data about the US airport network (*7*), where the probability of a flight connection within US decays as a power-law with the distance $r$ between airports, $r^{-3}$.

These examples reveal that in order to understand the causes of the structure of a transportation network it is important to find which quantity is optimized in the process.
Thus let us take a closer look at the complex systems we will analyze in this work, in order to comprehend which of their features can be safely neglected and which instead play a key role in defining the variational principles that determine the network structures.

Forests cover $\sim 42$ million km$^2$ in tropical, temperate, and boreal lands, $\sim 30\%$ of the land surface, and provide ecological, economic, social, and aesthetic services to natural systems and humankind. They influence world's climate through physical, chemical and biological processes that affect planetary energetics, the hydrologic cycle, and atmospheric composition (*8*).
In particular, forests play a key role in carbon dioxide cycle, being able to adsorb and emit large quantities of $CO_2$ in the atmosphere. Indeed forests store $\sim 45\%$ of terrestrial carbon, uptaking about $\sim 30\%$ of all anthropogenic $CO_2$ emissions from fossil fuel burning and land-use change.
It is thus possible to develop strategies to mitigate carbon emissions through forestry activities like reforestation, increase the carbon density of existing forests,

and reduce emissions from deforestation (*9*).

However, climate mitigation through forestry carries the risk that carbon stores may return to the atmosphere by disturbances such as fires and insects outbreaks. In fact while, on average, about $98.0$ to $99.7\%$ of forest land is in a carbon-sequestering stage, and the remaining $0.3$ to $2\%$ is emitting carbon, if we integrate over long periods and large areas, uptake and emissions from these areas nearly balance each other (*10*). The reason is that net carbon uptake is slow, in essence representing tree growth, while carbon emissions are generated from natural breakdown (tree death, gap formation), disturbance (wind break, fire), wave mortality (pest outbreak), or timber harvest – all of which are rapid processes.

Modern technology permits to determine the carbon balance of forests with unprecedented precision using local $CO_2$ flux measurements. However, such measurements have limited potential to contribute to a quantification of a region's, a nation's, or a subcontinent's carbon budget.

A correct estimate of carbon stores and releases requires the understanding of mass allocation, metabolic fluxes and resources use in forests. Our work aims to define the correct framework to analyze these processes in tropical forests.

In order to reduce the complexity of the numerous interactions among trees and identify the fundamental aspects that permit an analytical treatment, previous theories for the study of biodiversity in tropical forests established a convenient and significative approximation.

When in the last decades an increasing number of data on species richness in tropical forests has permitted a quantitative analysis of biodiversity, a novel theory proposed by S. Hubbell (*11*) claimed to be able to account for the diversity and the relative abundance of species under the hypothesis of *neutrality*. Neutrality is defined as "per capita ecological equivalence among all individuals of every species in a given trophically defined community. This definition is not the same as 'nothing going on' because it permits complex ecological interactions among individuals so long as all individuals obey the same interaction rules". The neutrality assumption implies, for example, that all individuals have the same birth and death rates irrespective of the species they belong to, and under this assumption the theory is able to predict the relative abundance of species and the species-area relationships (the rate at which species diversity increases with area). Further studies (*12*) have shown how under the assumptions of neutrality and non-interaction among species, it is possible to calculate the species turnover distribution (the probability $P_{STD}(\lambda, t)$ that the ratio of the populations of a species separated by a time interval $t$ is equal to $\lambda$) and the extinction times.

The sound success of neutral theory's predictions suggests that the neutrality hypothesis might be a good starting point also for the study of size distribution in tropical forests. Of course, when studying resources partitioning and the relationship between size and abundance, the interactions between individuals plays a fundamental role and can not be neglected, but a substantive simplification derives from the neutrality assumption. A further greater simplification comes from a stronger assumption we will use: the kind and number of species in a tropical forest is non influent to the distribution of tree sizes. This allows to focus our analysis on physical and geometrical considerations, like for example the mechanisms of space occupation, rather than ecological interactions related to biodiversity.

Despite the radicalism of these approximations, the results we obtain prove to agree with the available experimental data.

The discussion of our theory on size distribution in tropical forests is presented in chapter 2.

The other complex process we investigated is human mobility.

The recent outbreaks of new pandemic influenza, like SARS in 2003 and A H1N1 in 2009, have posed to politics and researchers the problem of how to predict and control the spreading of viruses (*13*). Sophisticated mathematical and numerical tools have been proposed to simulate the dynamic of the infections starting from the index case (*14, 15*). A central issue for these models is the definition of a human dynamic, i.e. the law of motion for human beings who are the carriers of the epidemic. At the present time the rules of human mobility are elusive, and all models utilize the commuting flux data obtained from census surveys or the air transportation database. The air transportation database is very accurate and detailed, and it can be a valid instrument to model the spreading of infectious diseases at the country level. On the other hand, if one is interested in the diffusion within a country the effects of land transportation must be taken into account (*16*).

The census databases on commuting fluxes usually provide a survey on home-to-work trips at counties/municipalities level. Unfortunately these data sets are available only for a limited number of countries, and they are carried out rather infrequently (once every 5-10 years). Thus for most of the developing countries those data are absent or considerably outdated.

The classical approach to obtain empirical laws from census data is the gravity model (*17*). As suggested by its name, the gravity model is in analogy with Newton's law of gravity: the number of people traveling between two locations (the attractive 'demographic' force) is proportional to the populations (masses) of the two locations, and decay with the distance. The rate of decay with the distance has to be empirically measured, and varies by context. Thus many generalizations of this law have been proposed, also with the addition of other variables (time, costs) and parameters to better fit experimental data. However any attempt to find a universal law of human mobility using the fitting procedure of the gravity model has not succeeded.

In the last years a different approach originated when the diffusion of personal mobile devices capable to frequently monitor the user's position (like GPS navigators or cell phones) has made available a huge amount of direct measurements. The possibility to "follow" each single individual during his/her daily route permits to have a complete information about the mobility process and to identify some important features of human trajectories. For example it has been observed (*18*) that the distribution of trips length is a power law, like in Lévy flights, but differently from a stochastic process with independent random steps, human trajectories show a high degree of temporal and spatial regularity: each individual has a characteristic length scale and a significant probability to return to few highly frequented locations.

However the definition of a stochastic process capable to generate patterns with the same temporal and spatial correlations observed in human dynamics is still an open problem.

On the basis of the conclusions of the above mentioned studies, our approach to human mobility is to look at the commuting fluxes at a coarse grained scale, the municipality/county level, which is suitable for epidemic prevention and easier to model than the individual trajectories. In particular we propose a model to reproduce the observed data assuming that the spatial distribution of population is the only necessary information, together with an optimality principle to determine the most probable trip's destination for each individual.

Chapter 3 is devoted to the description of our framework to model human mobility.

# Chapter 2

# Forests

## Abstract

Ecological communities exhibit pervasive patterns and inter-relationships between size, abundance, and the availability of resources. We use scaling ideas to develop a unified, model-independent framework for understanding the distribution of tree sizes, their energy use and spatial distribution in tropical forests. We demonstrate that the scaling of the tree crown at the individual level drives the forest structure when resources are fully used. Our predictions match perfectly with the scaling behaviour of an exactly solvable self-similar model of a forest and are in good accord with empirical data.

This chapter is organized as follows.

We start in section 1 with an introduction on size-frequency distribution in tropical forests and a review of the most relevant literature on the subject.

Section 2 is devoted to the description of the experimental data-set used in our analysis.

In section 3 we formulate the key concept of our work, i.e. that scaling laws may arise imposing space filling constraints, and we present two examples to support it.

We then proceed in section 4 to state the fundamental hypotheses for the optimal

transportation process, and we derive from them all the ecological quantities of interest.

Finally, we discuss the relevance of our findings and the possible application of our theory to non-tropical forests, in section 5.

## 2.1 Introduction

Understanding the interrelationships between patterns of size, abundance and resource availability in tree-dominated communities has proved to be a daunting challenge (*19, 20, 21, 22, 1, 23, 24, 25, 26, 11, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39*). Despite the differences in climate and biodiversity, power law patterns are ubiquitous in forest communities and this suggests the possibility that a general underlying principle can account for them.
Two main theories have been proposed to explain the observed similarities in size-frequency distribution among tropical forests: metabolic ecology and demographic equilibrium theory.

Metabolic ecology (*24, 38, 39, 23, 26*) builds on a model of ideal tree that satisfies various allometric relationships between the different features of the plant (diameter, height, mass). The idea is to represent the tree as a collection of identical tubes of equal length; each tube supplies a single leaf (which is assumed to be an invariant unit independent of tree's size) with water and nutrients. The tree is assumed to be an exact self-similar object which consists of the doubling of branches at each step (i.e. each branch generates two smaller branches). The tubes run into the branches and their number is preserved. At every step tubes are allowed to shrink in radius, and branches also in length, both following power laws relations. The values of the exponents of these relations are determined by minimizing the energy dissipated in fluid flow and assuming an optimal scaling relationship between length and radius of branches to resist buckling. From the model is then possible to derive many scaling relationships between the parts of the tree, like for example diameter ($r$) vs height ($h$), $h \sim r^{2/3}$; metabolic rate ($B$ proportional to the total leaf area) vs mass ($M$), $B \sim M^{3/4}$; and $B \sim r^2$. Experimental data are in accord with these predictions, but the question arises

whether this particular model of tree and the assumptions made in the calculations are really necessary or rather there can be a more general and model-independent derivation for those scaling laws.
A recent couple of papers (*38, 39*) aims to extend the metabolic theory from individual to stand-level in order to explain the structure and dynamics of forests. To this end it is further assumed that the forest is space-filling and in resource and demographic steady state.
Indeed the general idea is insightful, but the practical application and the derived results are quite unsatisfactory. For example, it is stated that space-filling implies the energy equivalence principle, i.e. that the total number of leaves of all individuals within any size class equally fills the same amount of area (and thus utilizes the same amount of energy which is proportional to leaf area). Actually it is possible to build a forest that is space-filling without assuming the energy equivalence principle. Besides, the energy equivalence principle cited above is ambiguous because it is not specified which size variable (diameter, height,...) is used to define the size classes. Indeed this erroneous statement is used to derive the power law's exponent of the distribution of diameters, and this exponent is different if the size classes are defined with respect to diameter or height or total leaf area. This ambiguity affects many of the consequent predictions, such as all the size frequency distributions, the scaling of nearest neighbor distance, the limiting of resources supply.

On the other hand, demographic equilibrium theory proposes a different approach, arguing that tree size distribution is the demographic consequence of size-dependent variation in growth and mortality (*33, 34, 35*).
In old-growth forests where disturbances are absent and demographic equilibrium has been reached, it is straightforward to show that the tree diameter distribution, $p(r)$, is related to growth rate as a function of size, $g(r)$, and mortality rate as a

function of size, $m(r)$, by

$$p(r) \propto \frac{1}{g(r)} \exp\left[-\int_{r_0}^{r} dx\, \frac{m(x)}{g(x)}\right] \tag{2.1}$$

The diameter distribution can assume different forms, depending on which functions $g(r)$ and $m(r)$ are considered. The most common cases for $p(r)$ are the inverse exponential, the power law, the Weibull, and the quasi-Weibull distributions.

Recent comparative studies (*33, 37*) have pointed out that the demographic equilibrium theory provides a better fit of the diameter distributions of tropical forests than the prediction of metabolic ecology. However the above mentioned comparison is not fair for a couple of reasons. First, the prediction of metabolic ecology on size-distribution is free of parameters ($p(r) \sim r^{-2}$), while the best candidates of demographic equilibrium theory, the Weibull and quasi-Weibull distributions, have respectively 2 and 3 fitting parameters. It is clear that it is easier to obtain a better fit if you can adjust some parameters than if you hazard a precise prediction [1]. Moreover the best fit values of the parameters of demographic equilibrium almost never agree with the corresponding experimental estimates of growth and mortality rates, within errors.

Second, the most remarkable observation about size-frequency distributions in topical forests is their surprising similarity, suggesting the existence of a universal explanation for it. Now, while the best fit of each single distribution is good, the values of the parameters obtained by the demographic equilibrium theory are quite different among the distributions. This could be a sign of a potential data over-fit, and it is possible that the effort to provide a good fit for each single distribution has compromised the attempt to find the simple general laws common to all forests.

[1] "With four parameters I can fit an elephant, and with five I can make him wiggle his trunk." John von Neumann

Having said that, the main criticism drew to metabolic ecology is that it does not explain the cut-off observed in diameter distribution at large diameters. In fact there are less large trees than what metabolic ecology predicts, and the theoretical result $p(r) \sim r^{-2}$ is verified only up to a characteristic size.
We will show that the finite-size scaling approach provides an extension and improvement to the results obtained with simple scaling, allowing to understand the causes of the lacking of large trees and, most importantly, demonstrating the existence of a link to bridge metabolic ecology with demographic equilibrium theory.

## 2.2 Data analysis

The data we used in our analysis come from the Barro Colorado Island (BCI) forest census (*40*), one of the largest data-sets available on tropical forests.
The BCI forest is a rectangle of 50 ha (500 x 1000 meters) on one of the islands of the Panama Canal, located at coordinates 9.149398 -79.855486.
Six complete censuses have been carried out in the years 1981-83, 1985, 1990, 1995, 2000, 2005. All trees with diameter at breast height (dbh, measured at $\simeq 1.3$ m from the ground) greater than 10 mm were tagged, measured, mapped and identified to species.
In particular we utilized the most recent survey (2005) which consists of 368,122 total trees counted, corresponding to an average density of 0.76 trees per square meter. In our analysis we need to know at least two informations for each tree: the diameter and the position. 368,036 trees ($99.9\%$) are provided only with coordinates, and 208,387 ($57\%$) have both position and diameter. In Figure 2.1 the spatial locations of trees in the two categories is displayed.

The BCI rainforest possesses a high degree of biodiversity, counting almost 300 different species. In this work we will focus on the distribution of tree sizes, thus

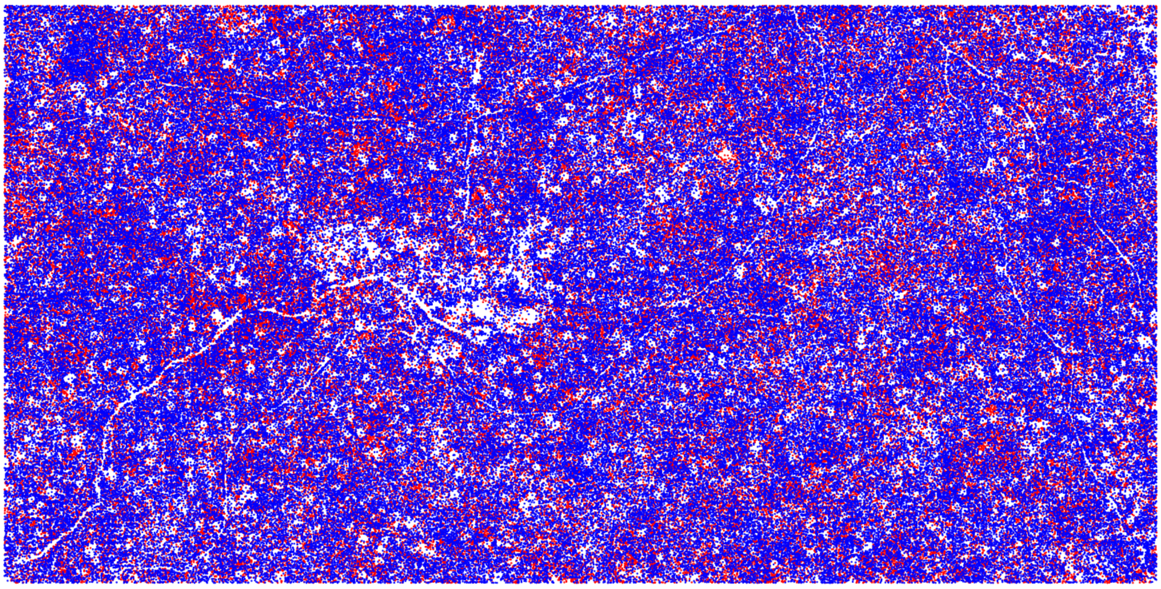


Figure 2.1: Map of the 50 ha BCI forest. The plot represents the 500m x 1000m rectangle with the 368,036 trees censused. The blue dots correspond to trees for which both position and diameter is known, while the red dots correspond to trees for which diameter information is missing. Blue and red dots appear to be uniformly distributed, ensuring that neglecting the trees with missing information does not affect the spatial analysis performed over the whole forest considering only the blue trees (for example when calculating the distribution of distances between trees).

we will not analyze the BCI forest from the point of view of Relative Species Abundance and beta-diversity. We only provide a brief overview about how the species are different with respect to the size of their individuals.

On Table 2.1 there is a list of the 20 most numerous species with their average diameter and the variance. Looking at the table it is hard to detect a clear connection between the characteristic size and the number of individuals belonging to a specie. Indeed among the 20 most numerous species the average diameter is found to be both smaller and larger than the forest average ($\simeq 5$ cm), thus the number of individuals belonging to a specie is not related to the average size of that specie (i.e. species with small characteristic size are not more populated than species with large size, or vice versa). In addition, we can consider the second moment of the diameter distribution (the variance) to check if it has any influence

| Specie code | Number of trees | Average diameter [cm] | Variance [cm] |
|---|---|---|---|
| HYBAPR | 29846 | 2.2 | 1.0 |
| FARAOC | 26038 | 4.5 | 3.2 |
| TRI2TU | 11344 | 5.5 | 7.9 |
| DES2PA | 11327 | 2.6 | 1.5 |
| ALSEBL | 7754 | 5.6 | 9.2 |
| MOURMY | 6540 | 2.0 | 0.9 |
| GAR2IN | 4602 | 3.1 | 2.6 |
| HIRTTR | 4566 | 5.7 | 5.7 |
| TET2PA | 4493 | 4.6 | 8.2 |
| PSYCHO | 3119 | 1.8 | 0.7 |
| SWARS2 | 2926 | 3.8 | 3.0 |
| PROTPA | 2853 | 3.2 | 2.2 |
| PROTTE | 2829 | 5.6 | 7.2 |
| SWARS1 | 2784 | 3.8 | 4.4 |
| CAPPFR | 2749 | 2.7 | 1.2 |
| SOROAF | 2539 | 3.7 | 2.0 |
| RINOSY | 2277 | 2.2 | 1.0 |
| TACHVE | 2234 | 3.8 | 6.3 |
| DRYPST | 2180 | 5.7 | 6.6 |
| QUARAS | 2137 | 13.8 | 18.3 |

Table 2.1: The twenty most populated species of BCI forest.

on the "success" of a specie. In fact if the variance is small (compared to the mean) then the distribution of diameter is highly peaked around the mean and the specie is specialized in exploiting the resources at that characteristic scale, while if the variance is large then the distribution is broad and the specie competes for resources at all scales. Again the table shows that the most numerous species have both picked and broad distributions, meaning that not even the broadness of the distribution of sizes within a specie seems to be related to its abundance. A complete picture of the independence of average diameter, variance and the ratio of them from the number of trees of the specie is given in Figure 2.2.

Anyway it is also true that species are not just sets of trees of random sizes. This can be easily seen by comparing the distribution of average diameters of real species versus the distribution of average diameters of "random" species with the same number of trees of real ones but with diameters randomly reassigned. Figure 2.3 shows that the distribution for random species is highly peaked around the average diameter $\simeq 5$ cm, while the distribution for real species is fat-tailed. This means that the distribution of size (and thus the use of resources) among species is not a trivial issue and deserves to be investigated further.

However the goal of the present work is to propose an explanation to the distribution of sizes of trees in the entire forest. Thus the question we must answer is whether the subdivision of trees into species is relevant for the distribution of

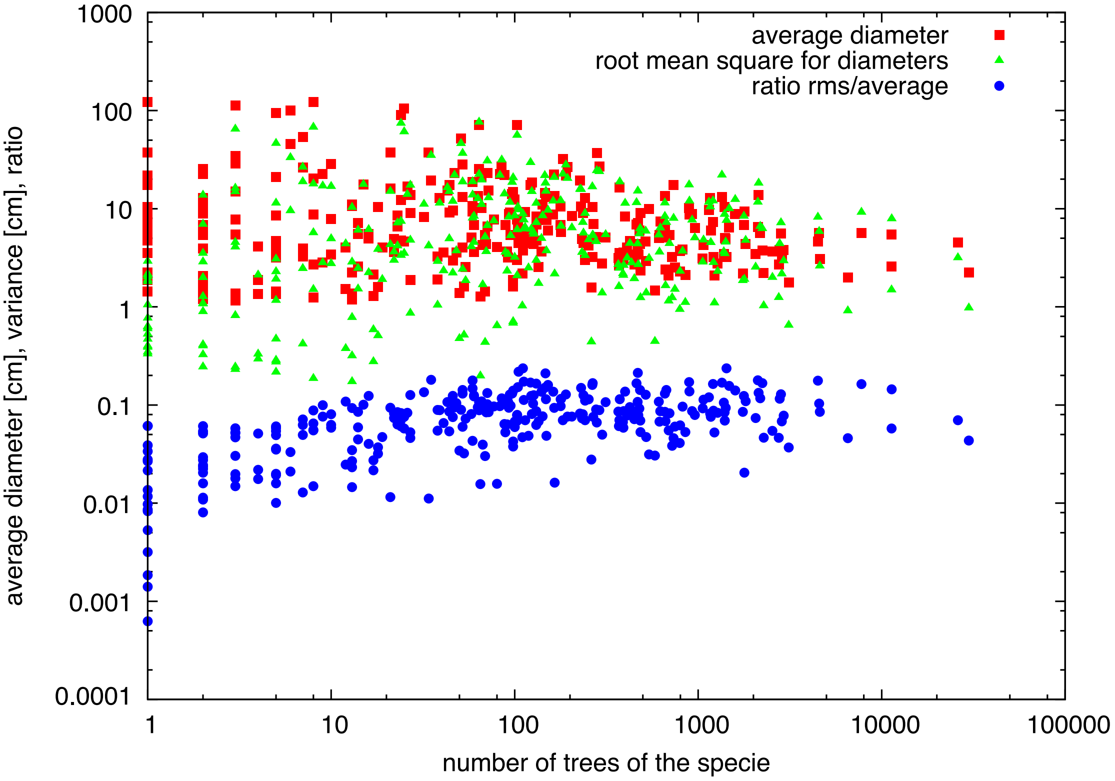


Figure 2.2: The average diameter of species (red squares), the root mean square (rms) (green triangles) and the ratio rms/average diameter (blue dots) versus the number of trees of the specie. The three plots are almost constant meaning that the size distribution within a specie is independent of the population of that specie.

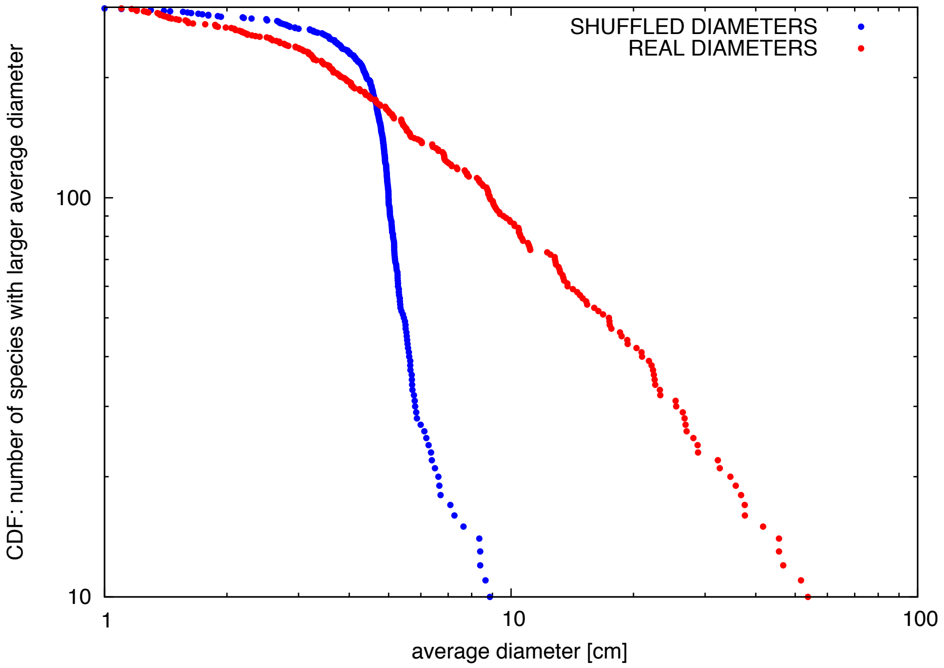


Figure 2.3: The cumulative distribution of average diameter of species. The red dots correspond to real species of BCI forest, while blue dots are the randomized species obtained by randomly reassigning the trees to each specie.

sizes in the forest, i.e. is it important to know which kind and how many species are present to determine the size distribution?

Diameter data collected in rainforests of different countries and with different species surprisingly show remarkable similarities in the size distribution. Consider in fact the six forests of Table 2.2; all of them have a diameter distribution close to a power law with exponent $\simeq -2$, while the number of species ranges from less than 200 for Sinharaja (Sri Lanka) to over 1000 of Lambir (Malaysia).

These considerations suggest that number and kind of species does not significantly influence the size distribution of trees. This is the reason why we will not consider the specie of each tree a meaningful variable in the rest of this work.

| Forest Name | Number of species (*41*) | Power law exponent (*33*) |
|---|---|---|
| BCI, Panama | 298 | -1.97 |
| Yasuni, Ecuador | 821 | -1.86 |
| Pasoh, Malaysia | 678 | -1.93 |
| Korup, Cameroon | 308 | -1.96 |
| Lambir, Malaysia | 1004 | -1.95 |
| Sinharaja, Sri Lanka | 167 | -2.05 |

Table 2.2: Comparison between species richness and the power law exponent of diameter distribution for six tropical forests.

Hence let us analyze the diameter distribution of all trees ignoring the species. In figure 2.4 the cumulative distribution of diameters for the BCI forest is displayed. It looks like a power law distribution for small trees with a cut-off for large values of diameters. The reason of this cut-off of the power law behavior at large diameters has not been fully explained. Of course there must be a physiological limit to the size of a tree because of the stability problems and the difficulties in transporting water to the top of the tree. Indeed, the tallest trees (e.g. sequoias) manage to deliver water and nutrients to leaves at over 100 meters above the ground. This tree height is by an order of magnitude greater than the physical limit that a column of water can reach by suction. Moreover the tree is lacking of any moving part which could help in the water circulation (like for example the heart in animals). Despite the efforts of researchers in order to explain this phenomenon, the mechanisms that limit the height of plants have not been completely understood at the moment (*42, 43*).

Anyways the biggest sizes of trees in tropical forests are usually well below those limit values: the largest diameters are less than 3 meters and the highest heights are about 50 meters. Thus the cut-off in size distributions is related to other factors than physical or physiological limits.

As showed in Figure 2.5, the cut-off is not due to the limited size of the sample.

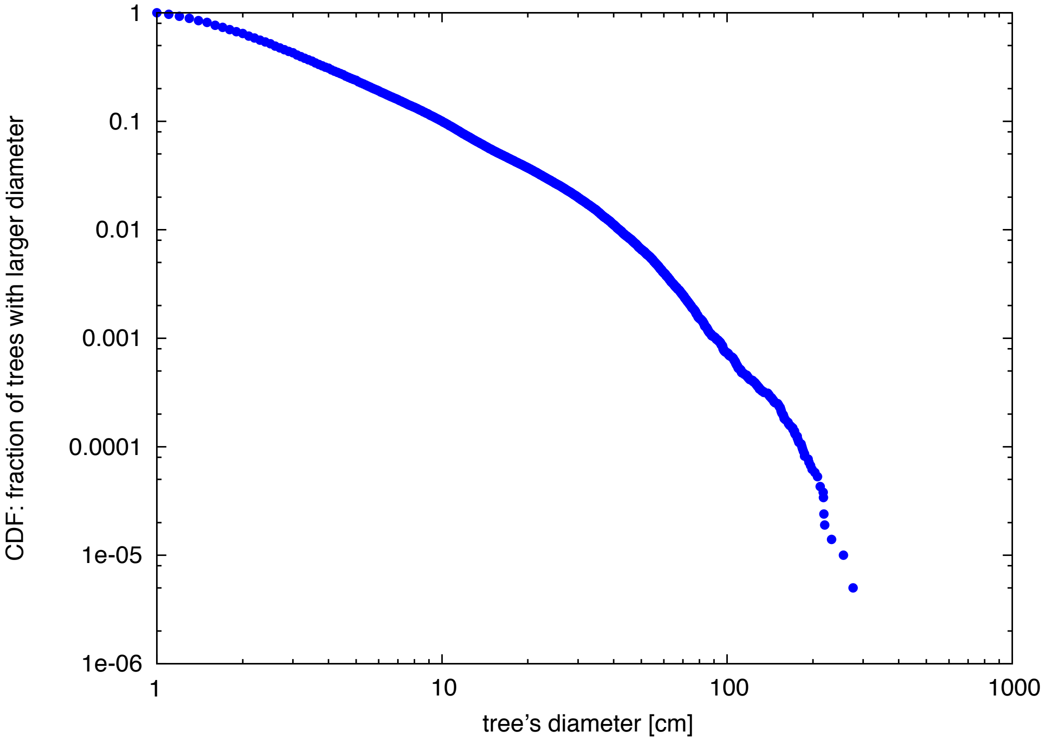


Figure 2.4: The cumulative distribution of diameters.

Indeed one might think that considering larger portions of forest the cut-off of size distribution should increase because the probability to find bigger trees is higher. But the distribution of sizes of the entire BCI forest (50 ha) is almost indiscernible from a sample of radius 50 m ($\simeq 0.8$ ha, $1.5\%$ of the total area). This consideration suggests that the cut-off is related to a local variable, since it is not extensive with the forest area. As proposed in (*38*) the cut-off arises from resource limitation, and the size of the largest individuals is related to the resources available per unit area. While this intuition is correct, the explanation provided for it is not consistent. We will propose and justify a new scaling relationship that links the density of resources to the height of the tallest trees, proving that the height of a forest is proportional to the resources available per unit area.

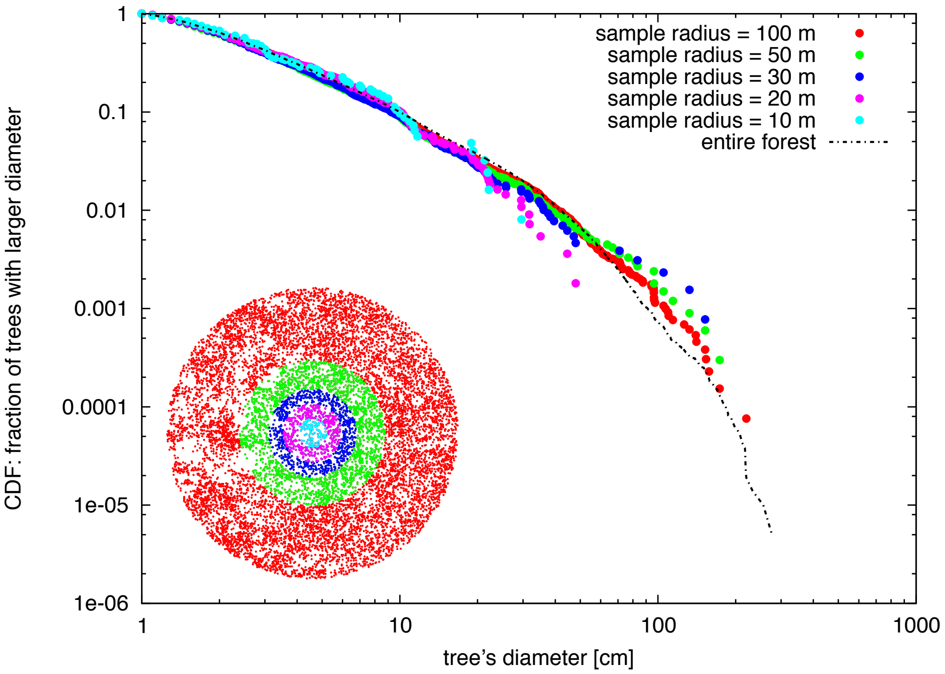


Figure 2.5: Comparison of cumulative distributions of diameters by increasing the plot size.

## 2.3 Space filling generates scaling

The goal we have in this work is to find a general explanation for the widespread fat-tailed distribution of tree size in tropical forests.

The size-frequency distribution of trees in indigenous forests is the result of ecological community's organization in resources partitioning. In an environment where the amount of nutrients is limited, individuals of different sizes tend to occupy the available space in order to exploit all environment's resources.

In this section we show how power law distributions may arise from space filling constraints, and we propose two simple models in which this happens. This mechanism which yields fat-tailed distributions, is related to the size distribution in tropical forests because the space filling condition can be translated in ecology to the condition of maximum use of available resources (and thus of space). Indeed, our approach is to consider the forest as a transportation network, that collects water and nutrients from the underground and distributes them to leaves in order to perform photosynthesis. The trees are the pipes of the network, and their size and geometry are adjusted in order to optimize the transport of water and fully utilize the resources. Further details about these variational principles in ecology can be found in the next section.

Let us now discuss how the space filling condition can generate scaling.
As an intuitive and simple example of space occupation's mechanism, consider a square of size $L$ with periodic boundary conditions; in this two-dimensional ecological system the density of nutrients is uniformly distributed on the square. We define the individuals that populate this ecosystem as disks of different sizes starting from a minimum diameter $r_m \ll L$ up to a maximum diameter $r_M$; each individual utilizes an amount of resources proportional to its area $\sim r^2$. We begin to populate the empty ecosystem adding individuals in the following way: we randomly choose a point on the square, if this point is already occupied or if there is another individual at a distance less than $r_m$ we discard it and repeat the choice. Otherwise we assign a size to the new individual randomly chosen between $r_m$ and the double of the smallest distance from the boundary of any neighbor. The sizes of circles are extracted from a uniform distribution, so that every size has equal prior probability to be chosen. We repeat this algorithm until we reach a stationary state (i.e. it is impossible to add a new individual) and then we look at the distribution of sizes $P(r)$ (see Figure 2.6). We discover that it is a power

law: $P(r) \sim r^{-\gamma}$ with $\gamma \simeq 2.75$. In this simple case a power law distribution

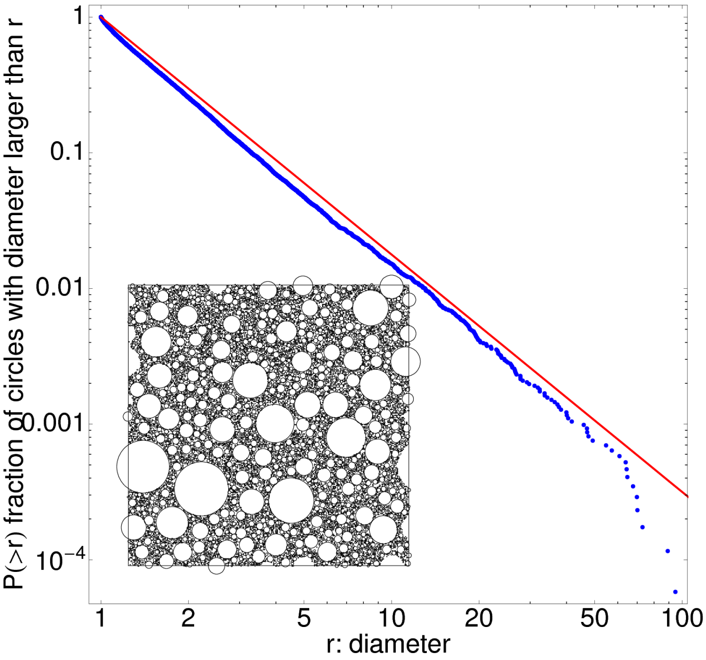


Figure 2.6: Cumulative probability distribution of diameters: $P(\text{diameter} > r)$ is the probability to find an individual with diameter greater than $r$ ($P(> r) \equiv \sum_{i=r}^{r_M} N(i) / \sum_{i=r_m}^{r_M} N(i)$). In this example $L = 500$, $r_m = 1$ and $r_M = 100$. The red line is the power law $P(> r) = r^{-1.75}$.

arises quite naturally when we randomly fill the space available with individuals of different sizes. This power law means that the system is scale free, i.e. there is not a characteristic diameter and the system viewed at different length scales is self similar. This result is both due to statistical reasons and geometrical constraints: it is the most probable way to randomly pack circles of different radii in a square.

This model of disks packing is a generalization of the Random Sequential Adsorption (RSA) problem proposed by Rényi (*44*) and known as the car parking problem. The original version of the model is in one dimension: on an infinite line, segments of unit length are dropped randomly and sequentially and they are adsorbed only if they do not overlap previously adsorbed segments. For this problem, the analytical expressions for the time-dependence of the density of segments ($45$) and the probability of each configuration at the stationary state ($46$) have been calculated.

Many generalizations have been proposed, where asymmetric objects of different sizes are adsorbed on multi-dimensional supports. In particular, numerical simulations of RSA of disks of different sizes on a square have shown that, starting from a uniform distribution of sizes, the final distribution of diameters of the adsorbed disks follows a power law whose exponent depends on the width of the starting uniform distribution ($47$). When the ratio between the width of the uniform distribution ($r_M - r_m$) and the smallest size ($r_m$) is sufficiently large, our simulations show that the diameter distributions of the adsorbed disks are power laws with exponent $-2.75$ (see Figure 2.7). However exact results for the general cases are not known.

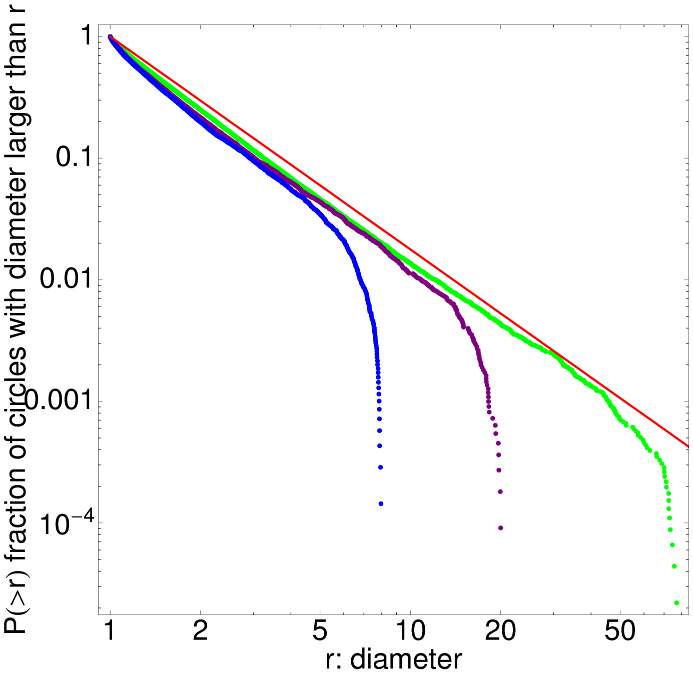


Figure 2.7: Cumulative probability distributions of diameters, $P(\text{diameter} > r)$, of the RSA for three values for the maximum diameter $r_M = 8$ (blue dots), 20 (purple dots) and 80 (green dots). The red line is the power law $P(> r) = r^{-1.75}$.

The key feature of our description is that we do not consider each individual isolated from the others. Indeed the size of each organism is determined considering the presence of the neighbours and that all individuals contribute to enable a complete development of the ecosystem as a whole.

We show now another example based on the above considerations: a simple model of growing networks in three dimensions that allows to obtain an analytical expression for the size-frequency distibution. Again in this case, it is a power law.

Let us consider a square lattice of size $L$ with periodic boundary conditions. At each lattice's site there is a small tree (a seed) that will grow in height and mass through the following steps: we add a $L \times L$ lattice just above the old one. We randomly choose a site from the old level and one of its nearest neigbors of the upper new level (provided that no other old site has chosen it before) and link them. We repeat this process until all sites of the new level are linked to a site of the old level; then we add another level and go on until we reach the desired maximum height $h_c$. This procedure represents a prototype of trees' growth by ramification that incorporates two key aspects: the space-filling character (i.e. all available space is occupied) and the competition for resources (i.e. every tree is competing with its neighbours for space).

Once the growing process is completed we can define the metabolic rate $B$ of each tree considering that it is poportional to the number of sites forming that tree, i.e. its volume (assuming that each site/leaf uses the same energy per unit time).

It is possible to calculate analytically the distribution of trees with a given metabolic rate (this model is analogous to the diffusion-aggregation models with injection that are solved in every dimension (*48*)): $P(B) = B^{-3/2} f(B/{h_c}^2)$ where $f$ is a cutoff function. The $-3/2$ exponent is due to the directness of the network (the growth is from the bottom to the top) and is rather robust with respect to changes in the model. It is also proved that height $h$ and transverse extension $r_{cro}$ are related by $r_{cro} \sim h^{1/2}$, the same scaling relationship between length and transverse extension of a random walk. Thus for the metabolic rate/volume we have $B \propto h \cdot r_{cro}^2 \sim h^2$. Using this scaling relationship we can obtain the probability distribution for heights $P(h) \sim P(B(h))\frac{dB(h)}{dh} \sim h^{-3} f(h^2/{h_c}^2)h \sim h^{-2} f(h/h_c)$. Figure 2.8 displays the comparison between the analytical results and numerical simulations.

The forest generated by the diffusion-aggregation model with injection described above exhibits a power law distribution of tree sizes, but unfortunately the exponents do not agree with experimental measures. Indeed, assuming that $B \sim r^2$ (in accordance with empirical measures (*26*)) we obtain $h \sim r$, in contrast with the established scaling relationship $h \sim r^{2/3}$. In order to retrieve the correct scaling exponents we need to modify some ingredient of the previous model. It might be reasonable, for example, that the growth of branches is not driven by diffusion but responds to another principle. A general framework that improves

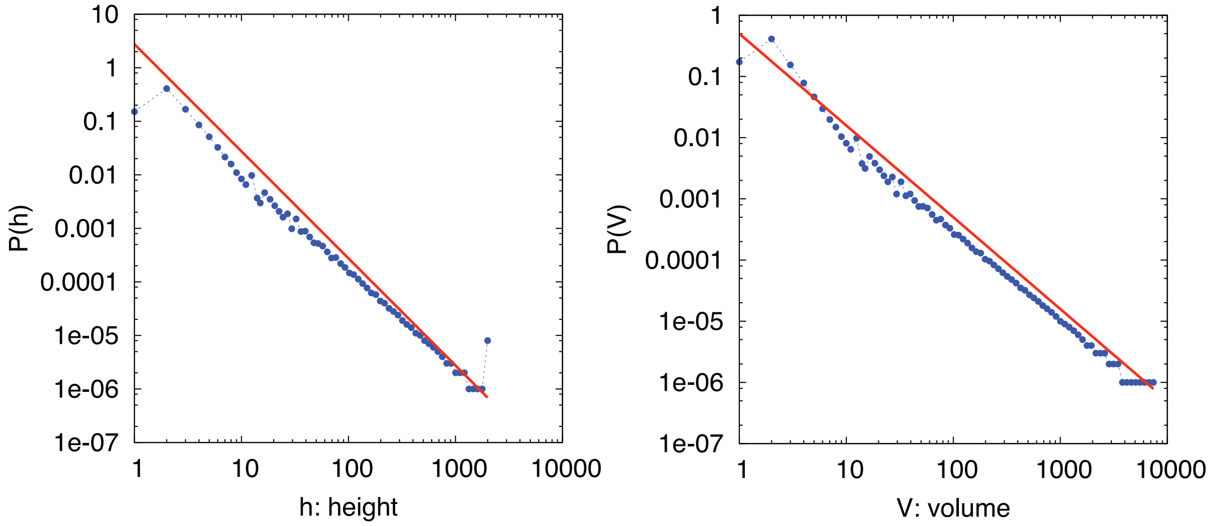


Figure 2.8: Comparison between analytical results and numerical simulations of the diffusion-aggregation model with injection in 2+1 dimensions. Left: probability distribution function of height, $P(h)$; the red line is the analytical result $P(h) \sim h^{-3}$. Right: probability distribution function of metabolic rate/volume (V), $P(V)$; the red line is the analytical result $P(V) \sim V^{-3/2}$.

the accord with data, and comprehends the diffusion-aggregation model of forest as a particular case, is presented in the next section.

## 2.4 General theory

As in earlier work (see, e.g. Ref. (*39, 38*) and references therein), we assume that the forest is in steady state and we do not consider the distinction between different tree species. We base our derivations on the following four hypotheses, of which the first two concern a single tree whereas the last two pertain to the whole forest:

1. **Tree shape**: For a tree of height $h$, the transverse extension or crown radius is *postulated* to scale as $r_{cro} \sim h^{H}$. Thus, the crown volume scales as $h^{1+2H}$. Quite generally, $H \leq 1$. $H = 1$ would imply an isometric tree shape, whereas $H < 1$ would result in a taller tree being more elongated than a smaller one. The scaling analysis predicts the dependencies of various exponents characterizing *the individual tree and the forest* on this shape exponent $H$ and thus provides links between exponents.

2. **Energy optimization of a tree**: The metabolic rate-mass relationship is obtainable by *maximizing the metabolic rate, $B$, for a given tree mass, $M$*. In agreement with empirical data, the metabolic rate, $B$, of a tree is assumed to be proportional to the number of leaves or to the tree crown volume, $B \sim h^{1+2H}$. This optimization is performed under the hypothesis that the average volume flow rate to the leaves is mass independent[2] (*49*).

3. **Energy optimization of the forest**: In order *to maximize the energy utilized by the forest, the leaves must fill the volume of the forest* which is

[2] This assumption follows from the observation that the leaves are supplied nutrients at a rate independent of tree height, facilitated by xylem tapering.

proportional to the product of the forest area, $A$, (or equivalently the total number of trees in the forest (*11*)) and the typical height of the tallest trees (we denote this by $h_c$). We show below that this allows one to deduce that the tree height probability distribution function (PDF), $p_h(h)$, of a forest is a power law, when $h < h_c$, characterized by an exponent that depends on $H$.

4. **Scaling**: We generalize the pure power law behavior that we will deduce from the previous hypotheses building on the finite size scaling approach (*50, 51, 52, 53, 54*). The power of the scaling framework is that it will allow us to carry out a collapse (see Appendix A) to deduce the range of parameters over which pure power law behaviour holds and determine the exponents. We consider a scaling form for the fraction of trees with height between $h$ and $h + dh$, given the typical height of the tallest trees in the forest is $h_c$, $p_h(h|h_c)dh = h^{-\alpha} f_h(h/h_c)dh$, where $\alpha$ is the familiar power law exponent. $h_c$, as shown below, is a measure of the average resource use per tree or equivalently per unit area. The scaling function $f_h(h/h_c)$ is postulated to tend to a constant value when $h \ll h_c$, thus leading to pure power behaviour, $p_h(h|h_c) \sim h^{-\alpha}$, and approaches zero when $h$ approaches $h_c$ from below or is greater than it.

The scaling theory, based on the last hypothesis, thus takes into account the resource limitation in an ecological community, which cuts off pure power law behavior. While all scaling relationships involve a power law portion, the presence of the scaling function and the inherent characteristic height arising from resource limitations typically result in the power law behaviour occurring over a limited range of scales.

We now proceed to deduce power law exponents and to a verification of the validity of finite size scaling (*53, 54*) with the available data from the BCI forest (*40*).
There are at least two distinct masses that one may attribute to a tree – the first is the mass of the fluid contained within the transportation network (i.e. the sapwood) and the second is the total tree mass including also the heartwood (which provides structural stability). The former mass is, of course, contained in the latter. We will assume, for simplicity, that the two masses scales isometrically, i.e. they are proportional to each other. Based on a general theorem pertaining to transportation networks (*1*), the maximum metabolic rate for a tree of mass $M$ is given by $B \sim M/h$ yielding $M \sim h^{2+2H}$ on using hypothesis 2. This result follows from the observation that efficient directed transport along the tree ensures that the mean distance from the source to all the leaves scales as $h$. Thus the metabolic rate-mass relationship takes the form $B \sim M^{(1+2H)/(2+2H)}$. We use the same optimization principle of maximizing the metabolic rate for a given mass to deduce that the optimal tree shape is characterized by $H = 1$, yielding the maximum value of the metabolic rate-mass exponent $(1 + 2H)/(2 + 2H)$ of $3/4$. This optimal case yields the classic Kleiber law (*19, 20*), $B \sim M^{3/4}$. Thus in the optimal case, $B \sim h^3$ and $M \sim h^4$. The total tree mass scales isometrically with the mass of the stem (*30*) and therefore $M \sim r^2h$, where $r$ is the stem diameter. Thus $r \sim h^{3/2}$, coinciding with the result obtained from considerations of buckling (*19*). This result justifies the simplifying assumption made above of the isometric scaling of the two distinct definitions of tree mass M. This relationship between tree diameter and height predicts the tapering of the tree trunk and leads to the pleasing result that the metabolic rate $B \sim r^2$ as empirically demonstrated (*27*).

From allometric theory (*19, 20*), the characteristic biological time scales as $M/B$. It is the length of time required for a non-feeding animal to exhaust its stored metabolic energy or for blood circulation to take place in an organism. Thus, the characteristic mortality rate is predicted to be proportional to $B/M$ and scales as $h^{-1} \sim M^{-1/4} \sim r^{-2/3}$, which is in accord with the empirical data presented by Enquist et al. (*38*) when $H = 1$.
A refinement of the previous argument allows one to bridge metabolic ecology (*22, 23, 24, 26, 28, 30, 31, 36, 55, 38, 39*) with demographic equilibrium theory (*32, 34, 37*).

The finite size scaling of the diameter distribution and the ontogenetic growth equation (*56, 57, 58*) yield a finite size scaling form for the growth rate, $g(r) \sim r^c G(r/r_c)$ and the mortality rate, $m(r) \sim r^b M(r/r_c)$, with $c = b + 1 = (2H - 1)/(2H + 1)$. The scaling functions are related by the following equation:

$$M(x) = 2G(x) + x\frac{dG(x)}{dx} + x\frac{d\ln f_r(x)}{dx} \tag{2.2}$$

thereby providing a quantitative link between metabolic ecology and demographic equilibrium theory through finite size scaling.
On using the ontogenetic growth equation with the generic finite size scaling assumption, one gets the growth rate, $g(r) \sim r^c G(r/r_c)$ with $c = (2H - 1)/(2H + 1)$. The mortality rate, $m(r)$, can be obtained, following demographic equilibrium theory, and requiring that

$$\frac{\partial}{\partial r}\left[g(r)p_r(r|r_c)\right] + m(r)p_r(r|r_c) = 0 \tag{2.3}$$

Inserting the finite size scaling equation for the diameter distribution, $p_r(r|r_c)$, in the previous equation leads to $m(r) \sim r^b M(r/r_c)$, with $b = c-1 = -2/(1+2H)$ and $M(x) = 2G(x) + xdG(x)/dx + x/f_r(x)\,df_r(x)/dx$. The choice $G, M =$

constant corresponds to the Muller-Landau et al. (*37*) postulate of pure power law behaviour, which is what happens here in the regime $r < r_c$, yielding $M/G = 2$, independent of $H$. In ref. (*37*) this ratio was predicted to be 5/3 based on the previous incorrect results of metabolic theory of ref. (*59, 60*). When $H = 1$, $c = 1/3$ and $b = -2/3$ and are consistent with empirical data (*40*) and agree with the predictions of ref. (*59, 60*).

To summarize, we have used a single optimization principle of maximizing the metabolic rate for a given mass to derive the shape and energy intake of a tree. Now, we utilize the *same principle at the level of a forest*, i.e. hypothesis 3, to determine the forest structure.

Let us first assume, consistent with hypothesis 4, that the PDF of the tree heights, $p_h(h)$, is zero both below some recruitment height, $h_0$, (lower cut-off) and above the typical height of the tallest tree, $h_c$, (upper cut-off) with $h_0 \ll h_c$. Using hypothesis 3, the total energy utilized by the whole forest is given by the alternative expressions in the two sides of the following equation

$$A\,h_c = A\int_{h_0}^{h_c} dh\,p_h(h)h^{1+2H} \tag{2.4}$$

$A\,h_c$ is the total volume at disposal of the forest whereas $A\,dh\,p_h(h)$ is the total number of trees with heights in the interval $(h, h+dh)$. Thus $A\,dh\,p_h(h)h^{1+2H}$ is the metabolic rate of (or volume occupied by) trees with their heights in that range. If $p_h(h)$ is a continuous function, the above equation readily implies that

$$p_h(h) \overset{H=1}{\propto} h^{-3}\Theta(1 - h/h_c), \qquad h > h_0 \tag{2.5}$$

where the $\Theta$ function is 1 if its argument is positive and zero otherwise. The distribution of stem radii follows from the relationship between $r$ and $h$ to be $p_r(r) \propto r^{-7/3}\Theta(1 - r/r_c)$ with the cut-off value $r_c \sim h_c^{3/2}$. The case for generic

$H$ is reported in Table 2.3 as well as the exponent values for the distribution of the crown height, the metabolic rate, the plant mass and other attributes using the standard rule for the change of variables. The scaling form postulated in hypothesis 4 for the tree height PDF (and the derived ones for other related variables) follows on substituting $\Theta(1-h/h_c)$ with a more general function $f_h(h/h_c)$.

| | $h$ | $r$ | $r_{cro}$ | $r_i$ | $d_s$ | $B$ | $M$ |
|---|---|---|---|---|---|---|---|
| $\omega$ | $1$ | $\frac{1+2H}{2}$ | $H$ | $H$ | $\frac{1+6H}{4}$ | $1+2H$ | $2(1+H)$ |
| $\omega\big\|_{H=1}$ | $1$ | $3/2$ | $1$ | $1$ | $7/4$ | $3$ | $4$ |
| $\alpha$ | $1+2H$ | $\frac{1+6H}{1+2H}$ | $3$ | $3$ | $\frac{1+14H}{1+6H}$ | $\frac{1+4H}{1+2H}$ | $\frac{1+2H}{1+H}$ |
| $\alpha\big\|_{H=1}$ | $1$ | $7/3$ | $3$ | $3$ | $15/7$ | $5/3$ | $3/2$ |

Table 2.3: Scaling relationships for tropical forests. Summary of the key predictions of the idealized scaling framework. The second row shows the exponent $\omega$ characterizing the scaling relationships of the form $y \sim x^{\omega}$, where $x$ is the tree height $h$ and $y = h, t, r_{cro}, \ldots$. The third row presents the $\omega$ values for the idealized case of $H = 1$. Our analysis predicts that the PDF of $y$ satisfies the scaling form $p_y(y|y_c) = y^{-\alpha} f_y(y/y_c)$ where $y_c \sim h_c^{\omega}$ and $f_y$ is a suitable scaling function as explained in Appendix A. The corresponding value of the exponent $\alpha$ is predicted to be equal to $1 + 2H/\omega$. As an example, the PDF of the distribution of the metabolic rate $B$ is predicted to be $p_B(B|h_c) = B^{-\frac{1+4H}{1+2H}} f_B(B/h_c^{1+2H})\big|_{H=1} = B^{-5/3} f_B(B/h_c^3)$.

The energy equivalence principle states that when trees are binned in discrete size classes, the total energy consumed within each class is the same. We find that this does hold when the size classes are based on tree height and not on tree radius as has been suggested previously (*26, 38, 39*).

Indeed, the total energy resources used in the forest is proportional to the total volume of the forest, $Ah_c$. This general result follows from the space filling condition and is independent of any scaling assumptions or the value of the $H$ exponent. Thus the characteristic height of trees, $h_c$, is a measure of the average energy use per unit area or equivalently per tree. Therefore if we express the energy equivalence with respect to a generic size variable $x \sim h^{\gamma}$, i.e. $p_x(x) \sim x^{-(1+2H)/\gamma}$ for $x < x_c \sim h_c^{\gamma}$, we obtain $E_{tot} \propto A \int x^{(1+2H)/\gamma} p_x(x|h_c^{\gamma})dx \propto Ah_c^{\gamma}$ and this is in contrast with the scaling of energy per tree derived above, unless $\gamma = 1$.

To summarize, we have used the same optimization principle at the tree level and at the forest level of maximizing the metabolic efficiency to derive relationships between, and the values of, exponents characterizing individual tree shape and forest structure. For the forest, the power law behavior of the distribution of tree heights or diameters is derived and not assumed *a priori*. However the range over which pure power law behavior is observed is limited. We turn now to the issue of how, in practice, one might determine the range over which pure power law behavior holds and estimate the values of exponents. We begin by introducing a new variable, the range of influence, $r_i$, defined as the distance from a given tree to the nearest tree having a larger diameter. Trees compete for light mostly with individuals of their size or greater, and so $r_i$ can be considered as the distance to the nearest significant competitor. The PDF of $r_i$ is predicted to be (see Table 2.3)

$$p_{r_i}(r_i|h_c) = r_i^{-3} f_{r_i}(r_i/h_c) \qquad (2.6)$$

with an exponent value of 3 independent of the specific value of $H$. The cut-off dependence arises because $r_i$ is expected to scale isometrically with $r_{cro}$ due to competition for space, implying $r_i \sim r_{cro} \sim h^H$. Figure 2.9 shows a plot that confirms the above prediction and, surprisingly, the power law behaviour holds up to the size of the forest that is much larger than the largest crown size and there is little need for the scaling function to provide the expected cut-off of pure power law behaviour. We will utilize an analysis of this quantity, which exhibits pure power law behaviour valid over a wide range, to deduce whether scaling holds in the BCI forest. Before we do that, let us ask why the exponent value is 3, independent of $H$. Assuming a random distribution of trees within the forest, which ought to be true for trees separated by a large distance, one can prove that the PDF of $r_i$ is a power law with the same decay exponent of 3. It is this harmonious matching of exponent values at short length scales (where the tree shape and the width of the crown matter) and the long length scale behaviour (where one may apply random distribution considerations), which leads to an almost perfect power law relationship extending over a wide range of $r_i$.

We now turn to an application of finite size scaling through a powerful scaling collapse procedure (*54, 53*). The conditional PDF of $r_i$, the range of influence, given the stem diameter $r$, is predicted to be,

$$P_{r_i}(r_i|r) = \frac{1}{r_i} F_{r_i}\left(\frac{r_i}{r^{\frac{2H}{1+2H}}}\right) \tag{2.7}$$

Eq. 2.7 represents the probabilistic generalization of the deterministic counterpart $r_i \sim r^{2H/(1+2H)}$. The pre-factor $1/r_i$ ensures that the average of $r_i$ scales as $r^{2H/(1+2H)}$ (more generally, the $n$-th moments of $r_i$ scale as $(r^n)^{2H/(1+2H)}$). The prediction from Eq. 2.7 is that a plot of the cumulative PDF (which incorporates the pre-factor on the right hand side) versus $r_i/r^{2H/(1+2H)}$ for various size classes, over which the scaling framework holds, must collapse (*54, 53*) on to a single plot. Figure 2.10 shows the collapse plot of $P_{r_i}(r_i|r)$ for the predicted exponent of $H = 1$. The collapse is optimal for trees with diameters in the range of 2.4 - 31.8 cm indicating that this is the correct range in which power law behaviour ought to be observed and measured. The collapse begins to break down for larger diameters due to resource limitations. The finite size scaling collapse is an objective method for estimating exponents when pure power law behaviour is not observed over a significant range of values of the variables being studied leading to a spread of exponent values. Figure 2.12 shows the results of two other tests of the consistency of the scaling theory and are compared with previous predictions

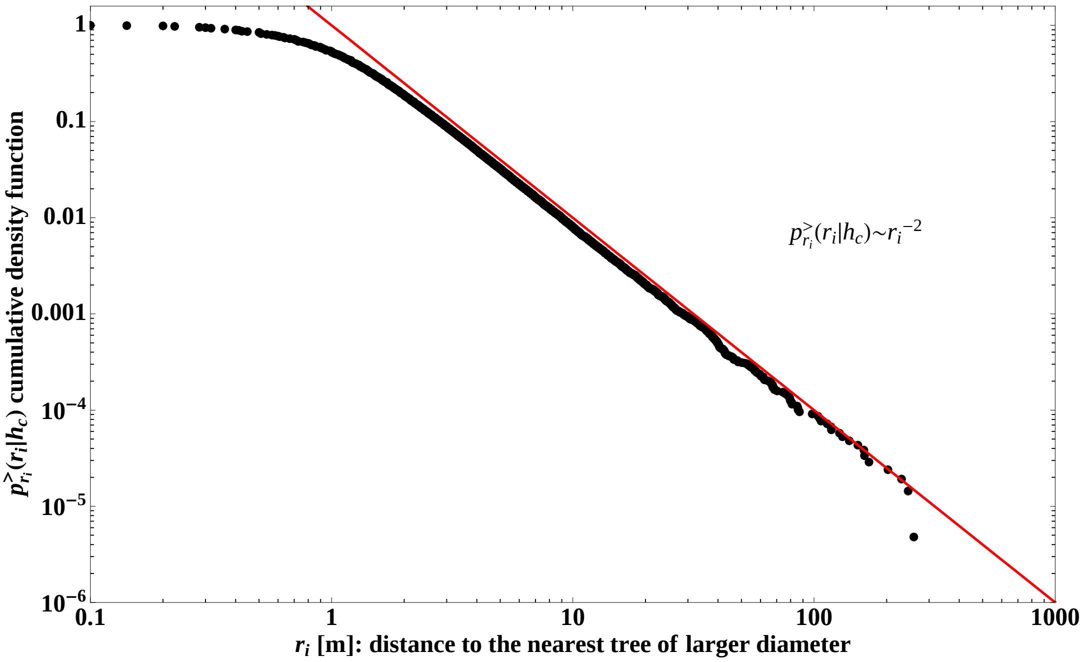


Figure 2.9: Cumulative probability distribution of the range of influence ri for the BCI dataset (1995) (*40*). $p^{>}_{r_i}(r_i)$ is the fraction of trees whose minimum distance to a tree of bigger size is greater than $r_i$ (measured in meters). The red line is a power law with exponent -2, equivalent to an exponent of -3 for the PDF. Surprisingly, the power law behaviour holds up to the size of the forest that is much larger than the largest crown size and there is little need for the scaling function to provide the expected cut-off of pure power law behaviour. This fact and that this exponent is independent of $H$ has a simple interpretation. Indeed by assuming a random distribution of trees within the forest, which ought to be true for trees separated by a large distance, one can prove that the PDF of $r_i$ is a power law with the same decay exponent of 3. It is this harmonious matching of exponent values at short length scales (where the tree shape and the width of the crown matter) and the long length scale behaviour (where one may apply random distribution considerations), which leads to an almost perfect power law relationship extending over a wide range of $r_i$.

of metabolic ecology (*22, 23, 24, 26, 28, 30, 31, 36, 55, 38, 39*). Figure 2.11 further demonstrates the validity of our hypothesis.

The scaling results hold exactly as predicted by our approach for the self-similar forest model presented below that satisfies all the hypotheses.

In two (three) dimensions, the forest is represented by a rectangle $L \times h_c$ (a volume $L \times L \times h_c$), where $L$ and $h_c$ represent the linear size of the forest and the height of the tallest trees respectively with $L \gg h_c$. At the zero-th step, we start to fill the ecosystem with the highest trees represented by triangles of height $h_c$ and crown extension $h_c$ with $H = 1$ (upside down pyramids of height $h_c$ and square base $h_c \times h_c$). The area (volume) occupied by a single large tree is a measure of the metabolic rate, $B = h_c^2/2$ ($B \sim h_c^3$). The number of the tallest trees is $\rho = L/h_c$ ($\rho = (L/h_c)^2$). In the next steps – labelled with index $t > 0$ – we introduce trees of a different shape: in two dimensions, the tree now has the shape of a rhombus of height $h(t) = h_c/2^{t-1}$ and transverse extension $h_c/2^t$ . As shown in Figure 2.13, they perfectly fill in the empty spaces between trees in the former levels. (In three dimensions, the shape of the tree is an upside down pyramid with a square base with height $h(t) \sim h_c/2^{t-1}$ and base side $\sim h_c/2^t$. A crown of variable shape is attached to the base so that the space between pyramids of two consecutive levels is completely filled, and each tree occupies the same volume proportional to $h(t)^3$. The three dimensional case is self-similar as well but is harder to visualize).

In two dimensions, the metabolic rate of a tree of level $t$ is $B(t) = (h_c/2^t)^2$ and there are exactly $N(t) = \rho 2^{t-1}$ trees in step $t$ and $N_>(t) = \rho 2^t$ total trees at step $t$. Thus the total number of trees with metabolic rate larger than $B$ is $N_>(B) = \frac{L\Theta(h_c^2/2-B)}{B^{1/2}}$, where $\Theta(x)$ is the step function equal to 1 when $x > 0$ and 0 otherwise. The PDF is given by $p_B(B|h_c) \propto -\frac{d}{dB}N_>(B) \propto \frac{L\Theta(h_c^2/2-B)}{B^{3/2}}$, yielding a probability density function of the form $p_B(B|h_c) = B^{-\phi} f_B(B|h_c^2)$ with a scaling function $f_B(B|h_c^2) \propto \Theta\left(1 - \frac{2B}{h_c^2}\right)$. The cut-off arises because the largest tree in the model has an area and thus a metabolic rate equal to $h_c^2/2$. Thus the power law $p_B(B) \sim b^{-\phi}$ holds only when $B < h_c^2/2$ –the range over which scaling is observed grows as $h_c$ increases. The other PDFs can be found in a similar manner. For example, $p_h(h|h_c) \propto \frac{\Theta(h_c-h)}{h^2}$ which is consistent with the PDF for $B$ because $B \sim h^2$. Note also that trees of the same size are uniformly distributed, a postulate used to derive the $d_s$ scaling shown in Figure 2.12 b. A similar analysis in the realistic three dimensional case yields $p_B(B|h_c) \propto \frac{1}{B^{5/3}}\Theta(1-\frac{cB}{h_c^3})$, where $c$ is a numerical constant. The exponent value of 5/3 is in accord with the results presented in Table 2.3. Other scaling laws and PDFs follow in a straightforward manner. Using Hypothesis 2, $B \sim r^2$, the above equation gives for the diameter PDF: $p(r|h_c) \propto 1/r^{7/3}\Theta(1-cr/h_c^{3/2})$. On introducing more realistic ingredients such as randomness in the plant position and size, one finds that the exponent of the power law is robust and does not change whereas the $\Theta$ function becomes a smooth function, $f_r(r/h_c^{3/2})$, with the characteristics described in hypothesis 4.

## 2.5 Perspectives

We have demonstrated that scaling provides a powerful framework for the analysis of forest data even in the absence of power law behaviour over extended scales and yields predictions in accord with data. We have shown that seemingly distinct patterns are all derivable from a single tree shape exponent, $H$, thus predicting links between them. The scaling results hold exactly as predicted by our approach for an exactly solvable self-similar model which satisfies all the hypotheses. For tropical forests, we have found that the maximum value of $H = 1$, corresponding to the optimization of tree metabolic rate, provides a good fit to data.

Our framework also allows to predict ecological features of a forest (like dimeter, height and mass distributions) from few experimental measures of the scaling relationships that hold in its trees, without an extensive and time-consuming survey on the entire forest. In fact we showed how the tree shape exponent $H$ uniquely determines both the scaling relationships between the parts of a tree (diameter, height, crown radius), and the exponents of the distributions of tree sizes in the forest. Once exponent $H$ is obtained by measuring diameter, height and crown radius of a tree and its branches, then it will simultaneously yield the exponents of the distributions, as indicated in Table 2.3.

Our framework is eminently suited for the study of forests across the globe (when detailed information pertaining to the locations of trees and their diameters, heights, and crown shape become available) to understand the steady-state conditions allowing the maximal use of resources, to elucidate the dependence of the value of the shape exponent on latitude and climate, and to understand the effect of disturbances on forest structure, carbon stock and sink.

The next step to achieve a full comprehension of resource partitioning in tree communities will be the study of size distribution in relation to biodiversity. As mentioned above, the use of resources among and within species is not a trivial issue because the distribution of size of a specie may be completely different from the one of another specie and of the entire forest. This indicates that the species are not equivalent with respect to energy and space use, and considered separately they do not follow the "average" trend of the forest. Probably the competition for the same resources (light and water) has led to a niche differentiation of the species, in such a way to minimize the overlap and competition of multiple species on the same size range. However, the reasons of coexistence of this niche differentiation with the high degree of biodiversity observed in tropical rainforests deserve further studies.

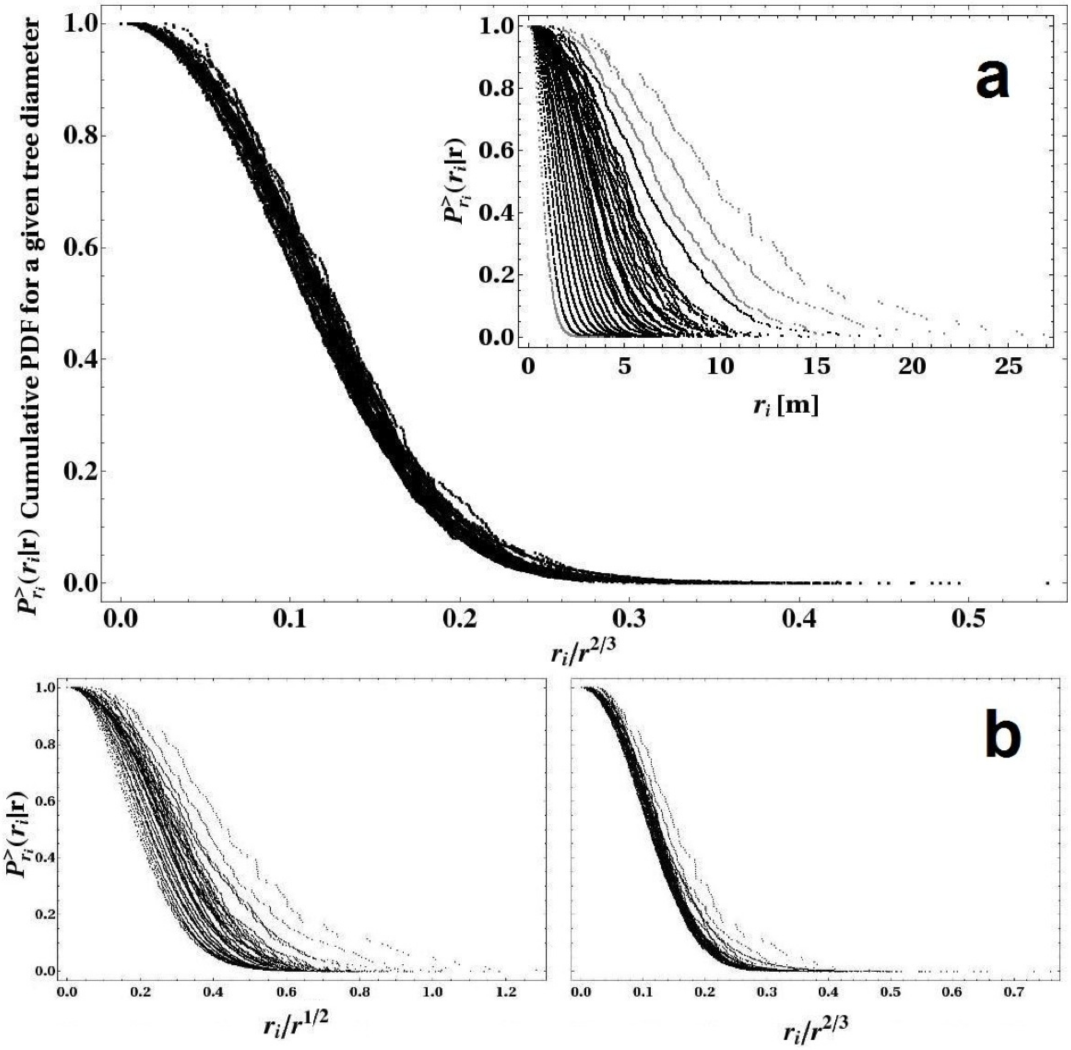


Figure 2.10: (a) Scaling collapse plot. The inset shows the probability of having a range of influence $\geq r_i$ for trees with diameter in the interval $(r, r + \delta r)$. We divided the BCI tree-diameter dataset in 1cm-size bins and for each bin we calculated the cumulative distribution of distances from the nearest neighbour tree of larger size $P^{>}_{r_i}(r_i/r)$. These distributions are shown for $r$ ranging between 1.4cm and $\sim$ 49.4cm in the inset. The main figure depicts a scaling collapse plot of the curves shown in the inset. The scaling framework predicts a collapse plot (see main figure) when the cumulative distributions of distances are plotted against the scaling variable $r_i/r^{2/3}$ when one is in the scaling regime. The grey curves in the inset do not collapse and the black curves define the range of tree diameters over which scaling is observed (2.4cm to $\sim$ 31.8cm).
(b) Comparison of scaling collapses corresponding to $H = 1/2$ (left) and $H = 1$ (right). These distributions are calculated for sets of trees grouped in 1cm-bins of diameter and are shown for $r$ ranging between 1.4cm and $\sim$ 49.4cm. $H = 1$ clearly provides a better collapse of the $P^{>}_{r_i}(r_i/r)$ distributions, allowing one to rule out the exponent value $H = 1/2$. When the scaling function is substantially constant, $H = 1$ leads to $p(r|h_c) \propto 1/r^{7/3}$ whereas $H = 1/2$ yields the prediction $p(r|h_c) \propto 1/r^2$.

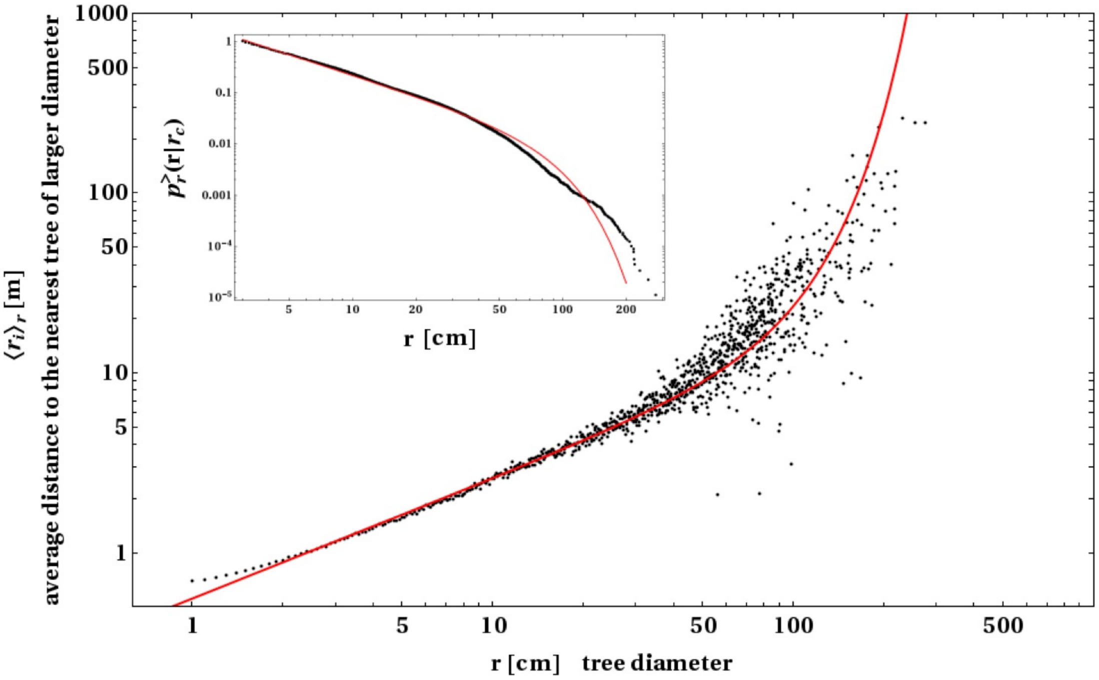


Figure 2.11: The main panel shows a plot of the mean range of influence $r_i$ versus the stem diameter $r$. The BCI diameter dataset were divided in 1 mm-bin classes and for each class we measured the average distance to the nearest larger tree. The inset depicts the cumulative PDF of tree diameters for the BCI dataset (1995). $p_r^>(r|r_c)$ is the fraction of trees with diameter greater than or equal to $r$. The red line shows a least squares fit with the function $p_r^>(r|r_c) = r^{-4/3}g(r/r_c)$, with $g(x) = \exp(-x^2/2)$. This functional form is consistent with our theory with a simple choice of a scaling function with $r_c$, the fitting parameter, equal to 86.6 cm. The precise fitting function $g$ is somewhat arbitrary and we have chosen it to be a Gaussian. This choice is based on simplicity and merely serves to demonstrate how scaling can be used to interrelate two distinct patterns. $r_c$ is the cut-off diameter and scaling is expected to hold for length scales much less than this cut-off value. This result is in accord with the direct determination of the scaling range ($\sim$ 2.4cm to $\sim$ 31.8cm) obtained with the scaling collapse of $P_{r_i}^>(r_i|r)$ (see Figure 2.10). The red line in the main figure shows the average range of influence of trees of a given stem diameter, $\langle r_i \rangle_r$, using the estimate $\langle r_i \rangle_r \propto 1/\sqrt{p_r^>(r|r_c)} = 1.2r^{2/3}g(r/r_c)^{-1/2}$ with exactly the same $g(x)$ used in the inset and the value of $r_c$ determined therein. The quality of the fit again demonstrates the validity of the scaling framework (see Appendix A).

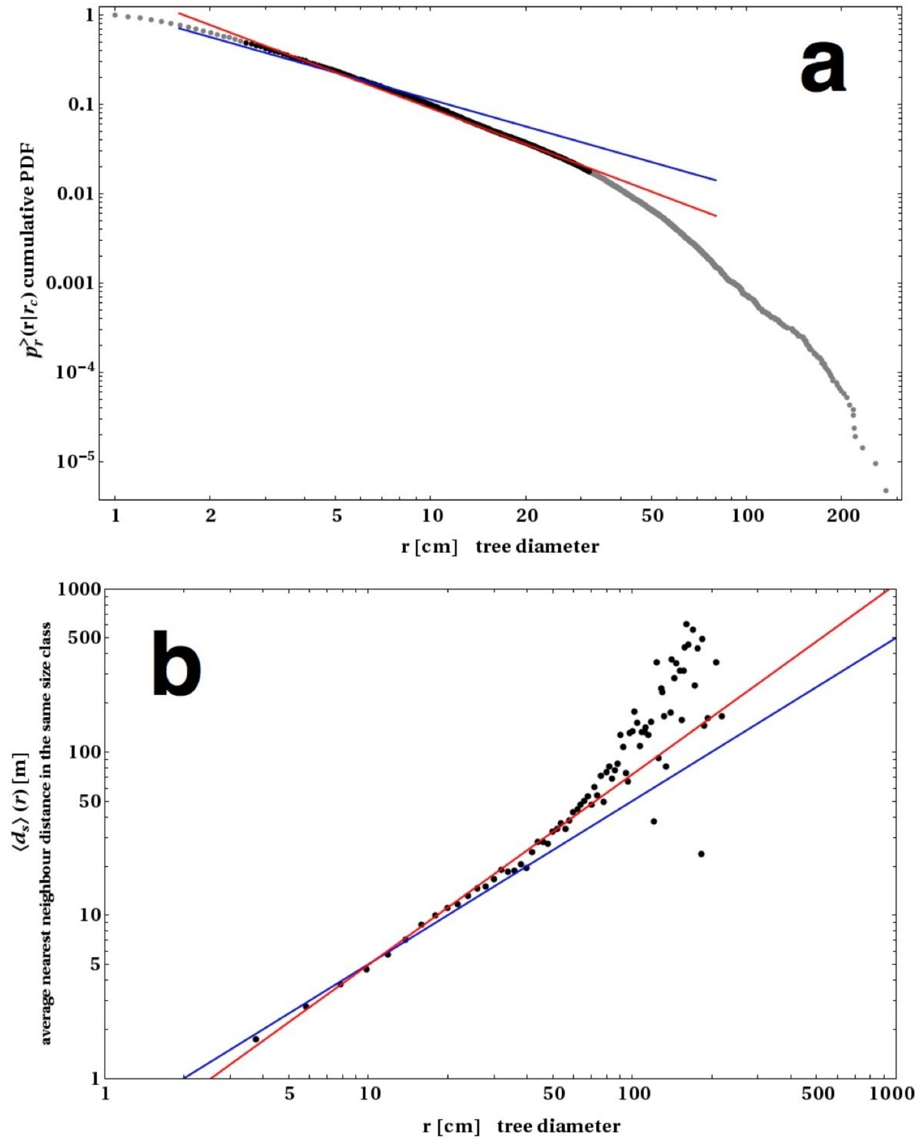


Figure 2.12: (a) The cumulative PDF of tree diameters in the BCI forest (1995). $p_r^{>}(r|r_c)$ is the fraction of trees with diameter greater than or equal to $r$. The black dots correspond to diameters in the interval from $\sim 2.4$cm to $\sim 31.8$cm (the range over which scaling is expected to hold from the scaling collapse plot in Figure 2.10). The red line indicates our predicted exponent of -4/3 (derived from a power law probability density with exponent -7/3). The blue line depicts the exponent of -1 (corresponding to a probability density with exponent -2) and is shown for comparison.

(b) Plot of the average distance from the nearest neighbour individual in the same size class, ds, versus the tree diameter. We divided the BCI tree-diameter dataset in 2 cm bin size and for each bin, we calculated the average distance between nearest neighbour trees belonging to the same bin. The red line shows the predicted power law behaviour with exponent 7/6, whereas the blue line is a power law with exponent 1. Our prediction follows from the assumption that the trees in a given size class are distributed uniformly across the forest, thus implying that $d_s \sim P_r(r|r_c)^{-1/2} \sim r^{7/6}$.

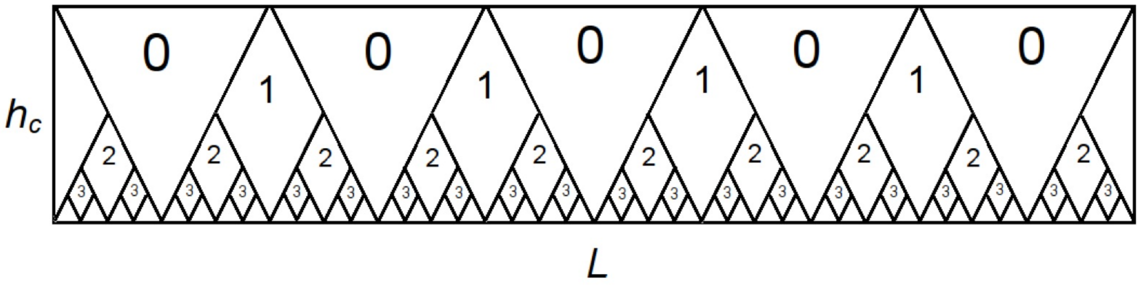


Figure 2.13: The two dimensional self-similar forest at step $t = 3$. The lines denote the boundaries of each tree, and the number inside each tree is the time step of its creation. The space without numbers is filled with higher generation trees. Note that it is not crucial to have complete space filling. For example, removing trees at the 0-th level does not change the results of our scaling analysis.

# Chapter 3

# Human Mobility

## Abstract

Understanding human mobility is important for problems related to urban planning, design of transportation systems and forecasting the diffusion of infectious diseases. Much of these applications are based on an empirical law often called the gravity law, capable of fitting the data on the human mobility in terms of few adjustable parameters. However a fundamental and robust modeling framework to explain the observed patterns is still lacking (*61*). Here we propose a simple stochastic model reflecting the human behavior/attitude in the choice of trips within a whole country. As input the model needs only the spatial population distribution, available in census data in most developed countries. Data for the number of trips toward a given location and distribution of travel lengths for three data-sets –the call records of cell-phone users and work-flows from US census 2000 and Portugal census 2001– are statistically indistinguishable from the model predictions. Analytical exact predictions can also be obtained for the case of ideal fractal distributions of population.

This chapter is organized as follows. After a brief summary about the previous studies on human mobility in section 1, in section 2 we present and motivate our stochastic model, the Nearest Better Than Me model. Analytical predictions about the distributions of mobility fluxes are also derived from it. Section 3 is dedicated to the derivation of the gravity law that follows from the model. A complete comparison between all model's predictions and the available data is provided in section 4. In section 5 we demonstrate the advantage of the stochastic approach in the modeling of mobility fluxes, showing that the probability to find a given mobility flow between locations obeys scaling. In the last section we discuss the results and perspectives of our work.

## 3.1 Introduction

Driven partly by data availability and also by the need for better urban planning and epidemic control, the study of human mobility has drawn increasing interest in the last years. So far, the framework used to describe the process has been the weighted network representation (*62, 63*), which represents the raw data as origin-destination tables (ODTs), describing a weighted network whose the nodes are the locations (e.g. municipalities or counties) and a link of weight $T$ between two nodes indicates a commuting flux of $T$ individuals between the two locations. The computation of typical quantities of network analysis (such as degree, strength, clustering coefficient, disparity of nodes) and their mutual relationships is useful to grasp some fundamental properties of the commuting process (like the relation between traffic and connections and the role of hubs in building the backbone pathways) and to obtain a quantitative measure of the relevance of each location in the transportation system.

However, it is not clear how the weighted network representation is sensible to the particular administrative subdivision considered: indeed, by properly choosing the subdivision of locations, it is possible to arbitrarily create a hub by grouping together a great number of small towns, or vice-versa, a hub can be split into many different small-degree nodes and disappear. How stable the results obtained from network representation are with respect to changes in the administrative subdivision has yet to be assessed.

Moreover, although providing a description to the available data, this approach is lacking a geographical and socio-demographic interpretation of the mobility process, as well as a general law/algorithm able to reproduce the observed patterns or generate them when data are not available.

Indeed, the mechanism most widely used to assign weights to the links, i.e. the gravity model, has several limitations. The gravity model consists of an empirical formula for the number of trips $T_{ij}$ between locations $i$ and $j$ (*17*). Generally $T_{ij}$ is assumed to depend on the number of people in the locations of origin $n_i$ and destination $n_j$ and the distance between them $r_{i,j}$. Using these variables, a function with up to six free parameters is arbitrarily chosen to estimate $T_{ij}$ from the data through a best fit procedure.

Until now there has not been any non-empirical justification or derivation of the specific form of the gravity model. It is unsatisfactory because it does not provide a predictive nor universal law. Indeed using this approach a previous knowledge of real flows is required in order to fit the parameters and calibrate the model. Moreover, several different functions of $n_i$, $n_j$ and $r_{i,j}$ have been chosen to fit the experimental fluxes, varying for example from $T_{ij} = \frac{n_i^\alpha n_j^\beta}{r_{i,j}^\gamma}$ (in (*14*)) to $T_{ij} = n_i^\alpha n_j^\beta e^{-r_{i,j}/d}$ (in (*16*)) and different values of the parameters have been found for different countries.

A simple and general model to pass from the static picture (population and distances of locations) to the dynamic picture (the commuting fluxes, i.e. the weights) has not yet been proposed.

A complementary approach (*18, 64*) is to focus on the motion of single individuals and then coarse-grain the microscopic trajectories to extract information about commuting flows. Actually, finding the basic laws of human mobility is challenging: the trips' lengths exhibit power law distributions and contrary to Lévy flights, they periodically return to few landmark locations. Trying to reproduce the human mobility dynamic at this fine level requires a huge effort which is not justified if the main interest is retrieving commuting fluxes at mesoscopic scale, like required in urban planning, design of transportation systems and forecasting the diffusion of human infectious diseases.

In this work we propose a stochastic model which at the coarse-grained scale (e.g. the municipality-county level) is able to predict the probability distribution of the commuting fluxes between locations once the spatial distribution of the population is known. This goes beyond the simple prediction of the commuting fluxes which in the present approach are obtainable as expectation values. The strength of the model is that it is free of any fitting parameter, allowing us to compare simulation results with flow patterns extracted from three different data-sets: the flow patterns of call records of cell-phone users, workflows from US census 2000 and Portugal census 2001. In all cases, despite the differences of spatial resolution (cell towers vs counties and municipalities), the model is in good agreement with real data.

This novel approach allows to generate commuting flows at medium-short length scales (in the range ∼10 to ∼1000 km) by employing only static information about the distribution of population (available from census data), in contrast with currently used methods (like, for example, the gravity model) which require a previous knowledge about commuting flows.

## 3.2 “The Nearest Better Than Me” model

The starting point of the model is the knowledge of the spatial distribution of population in the country. Usually there are many ways to retrieve this information; in our case we will use census data and the number of calls registered by a cell phone company for billing purposes.

These data will provide us with the two inputs we need in our model: first, the spatial pattern of population's distribution, second, the number of people present in each location. The spatial pattern is obtained considering the coordinates of the barycenters of the basic areas that are used to count people (e.g. cell-phone towers, zip codes, counties, municipalities, regions...). For each location the number of individuals is determined by counting the number of users who call from a given cell-phone tower, or the number of people resident in a given county.

In the model we utilize this information in the following way. A list of all possible destinations is own by each individual who assigns her/his personal *rank* to each of them, a real number, e.g. in the interval (0,1). We call this the list of preferences. We assume that each individual constructs her/his list of preferences by assigning the rank to each location by extracting it from a distribution, $p$, which is the same for every person, believing in an universal process in ranking procedure. Notice that we do not make any assumption on the form of this distribution.

We postulate that the trip's destination of each person is the closest location with rank higher than the rank she/he assigns to the location where she/he lives (i.e. the nearest more attractive location). We only consider trips between different locations (i.e. we exclude intra-location trips).

There is an issue caused by the inhomogeneities in the subdivision of a country into administrative regions which we must pay attention to in order to have consistent results. An idealized subdivision for our model would be the one in which the same number of people live in each region (in fact we can argue that the number of destinations in a location is directly proportional to the number of people present there). However, due to political and historical reasons, this is not the case and we observe large differences in population between locations. To overcome this problem we proceed as follows. We fix the density of destinations, i.e. one destination is assigned every $n_{dest}$ people). If location $j$ has $n_j$ people, $n_j/n_{dest}$ destinations are assigned to location $j$, assigning for each of them a different random number. Then we simply define the rank of location $j$ as the highest between them.

Our goal is to use the Nearest Better Than Me model to generate commuting trips. Although the home-to-work trip is a daily process, it is determined by a one-time choice, i.e. the job selection. We can illustrate the process of job selection proposed in the model step by step in an algorithmic way:

1. The individual applies for interviews all over the country.

2. He/She sorts the job offers in descending order from the one with biggest salary (here the salary is the rank).

3. The counties which offer a salary bigger than the salary of his/her county (the homeplace) are kept, and the others are discarded.

4. The job in the closest of the remaining counties is accepted.

**Analytical result for the distance of trips** It can be shown that the probability that the nearest more attractive location is at distance larger than $r$ is $P_{\delta}^{>}(r) \sim 1/r^{d_f}$, where $d_f$ is the fractal dimension of the set of available destinations, independently of the rank distribution.

We will prove this result for a generic system in $d$-dimensions.

The system we are considering is a distribution of $N$ points in a domain $M$ embedded in a $d-$dimensional space. On this space let us define a distance between any two points $\mathbf{x}$ and $\mathbf{y}$ as the Euclidean distance: $\mathsf{dist}(\mathbf{x}, \mathbf{y}) = \sqrt{(\mathbf{x}-\mathbf{y}) \cdot (\mathbf{x}-\mathbf{y})}$. The points are distributed according to a generic spatial probability distribution $\rho(\mathbf{x})$. We assume that this distribution, although varying with $\mathbf{x}$, has a fixed fractal dimension $d_f$, i.e. $\forall \mathbf{x} \in M$ the number of points in a $d-$dimensional hypersphere of radius $R$ centered in $\mathbf{x}$ scale as $R^{d_f}$.

Let us consider an individual at the point $\mathbf{x}$ who has assigned a rank $z_{\mathbf{y}}$ extracted from the distribution $p$, for each location $\mathbf{y}$. Note that the value $z_{\mathbf{y}}$ associate to point $\mathbf{y}$ is independent of the position of that point ($p(z)$ does not depend on $\mathbf{x}$). We define the distance from "the nearest better than me" location, $\delta(\mathbf{x})$, as the shortest distance from a location $\mathbf{y}$ whose rank, $z_{\mathbf{y}}$, is greater than $z_{\mathbf{x}}$, i.e.

$$\delta(\mathbf{x}) \equiv \min_{\mathbf{y} \in V_{\mathbf{x}}} \mathsf{dist}(\mathbf{x}, \mathbf{y}) \tag{3.1}$$

where $V_{\mathbf{x}} \equiv \{\mathbf{y} \in M : z_{\mathbf{y}} > z_{\mathbf{x}}\}$.

Under these assumptions, we will show that in the large size limit, $(N \to \infty)$, the probability to find a generic point $\mathbf{x}$ with a distance from "the nearest better than me" location greater than $r$ is $P_{\delta}^{>}(r) \sim \frac{1}{r^{d_f}}$, irrespective of the form of distributions $\rho$ and $p$. The proof is rather straightforward.

Let us first calculate the probability to extract a number greater than $z$ after exactly $n$ extractions: $P^{>}(n,z)$. Since two consecutive extractions from $p(z)$ are independent, we have:

$$P^{>}(n,z) = p^{>}(z)(1-p^{>}(z))^{n-1} = p^{<}(z)^{n-1} - p^{<}(z)^{n} \qquad (3.2)$$

where $p^{>}(z) \equiv \int_z^\infty d\zeta \, p(\zeta)$ and $p^{<}(z) \equiv (1-p^{>}(z))$. Notice that the probability to extract a number larger than $z$ in any number of extractions is $\sum_{n>0} P^{>}(n,z) = 1$

Having extracted a number from $p(z)$, the probability to extract a greater number at the $n$-th extraction is:

$$\begin{aligned} P_n &= \int_0^\infty dz \, p(z) \; P^{>}(n,z) \\ &= \int_0^\infty dz \, \frac{dp^{<}(z)}{dz} [p^{<}(z)^{n-1} - p^{<}(z)^{n}] \\ &= \left.\frac{p^{<}(z)^n}{n}\right|_0^\infty - \left.\frac{p^{<}(z)^{n+1}}{n+1}\right|_0^\infty = \frac{1}{n} - \frac{1}{n+1} \end{aligned} \qquad (3.3)$$

whereas the probability to extract a greater number not before the $n$-th extraction is:

$$P_n^{>} = \sum_{m\geq n} P_m \; = \; \frac{1}{n} \tag{3.4}$$

Notice that this probabilities are independent of the distribution $p(z)$ and it is correctly normalized, i.e. $\sum_{n\geq 1} P_n = 1$.

Suppose now that all $N$ points have been placed in $M$ according to $\rho(\mathbf{x})$ distribution, and that for each point of $M$ we have extracted a number $z$ from the $p(z)$ distribution. Consider a $d$-dimensional ball of diameter $l$ centered at point $\mathbf{x}$, $B_{\mathbf{x}}(l)$; the number of points in $B_{\mathbf{x}}(l)$ is $n(l|\mathbf{x}) = \int_{B_{\mathbf{x}}(l)} d^d y \, \rho(\mathbf{y}) = c(\mathbf{x}) l^{d_f}$, where $c(\mathbf{x})$ is a function of the distribution $\rho(\mathbf{x})$ and does not depend on $l$ (as assumed above).

For each point $\mathbf{x}$, one can calculate the probability to find a point with greater $z$ at distance greater than $l$ as

$$\begin{aligned} P_\delta^{>}(l|\mathbf{x}) &= \frac{1}{n(l|\mathbf{x})} \qquad \text{if} \;\; l > l_{\mathbf{x}} \\ &= 1 \qquad\qquad \text{if} \quad l < l_{\mathbf{x}} \end{aligned} \tag{3.5}$$

where $l_{\mathbf{x}}$ is defined as the distance such that $n(l_{\mathbf{x}}|\mathbf{x}) = 1$, that is the distance of the nearest-neighbor of $\mathbf{x}$. Averaging this result over a large region $M$ and for $r \equiv l$ large enough, $r \gg \max_{\mathbf{x}\in M} l_{\mathbf{x}}$) we finally get

$$P_\delta^{>}(r) = \int_M d^d x \, P_\delta^{>}(r|\mathbf{x}) \rho(\mathbf{x}) \sim r^{-d_f} \int_M d^d x \, \rho(\mathbf{x})/c(\mathbf{x}) \sim r^{-d_f} \; . \tag{3.6}$$

**Analytical result for the number of people in the destination** We now calculate the probability to have a trip to a destination with $n$ people for the Nearest

Better Than Me model.

Assuming that we know the starting position of every person (the resident location), we will divide the country with a square grid into squares of side $a$ and we associate to each square the number of people found in it, $n$. According to Zipf's law the distribution of $n$ is a power law: $P_0(n) \sim n^{-\alpha}$, with $\alpha \simeq 2$. It can be shown that the exponent $\alpha$ is related to the fractal dimension of the spatial distribution of people $d_f$[1].

The probability to end a trip in a location with $n$ individuals can be expressed as

$$P(n) = P_0(n) \sum_{m=1}^{\infty} P_0(m) \int_0^{\infty} dz\, P(> z|n) P(z|m) \tag{3.7}$$

$P(z|n) \equiv -\frac{dP(>z|n)}{dz}$ is the probability to have a rank equal to $z$ for a square with $n$ people; i.e. $P(z|n) = -\frac{d}{dz}\left(1 - (1 - p_>(z))^n\right) = -n(1 - p_>(z))^{n-1}\frac{dp_>(z)}{dz}$, where $p_>(z)$ is the cumulative distribution of ranks.

The term $\int_0^{\infty} dz\, P(> z|n)P(z|m)$ in eq. 3.7 is the probability to find a greater rank in a square with $n$ people, given that the starting square has a population of $m$ people, and it is equal to $m\left(\frac{1}{m} - \frac{1}{m+n}\right)$ independently of the rank distribution $p(z)$.

[1] Indeed, by definition of fractal dimension, the number of people in a disk of radius $R$ is proportional to $R^{d_f}$: $R^{d_c}\langle n \rangle_R \simeq R^{d_f}$, where $d_c$ is the fractal dimension of the counties' centers of mass ($d_c = 2$ for a uniform distribution), and $\langle n \rangle_R$ is the average number of people in the disk, i.e. $\langle n \rangle_R = \int_1^{n_M(R)} dn\, n P_0(n) \propto n_M(R)^{2-\alpha}$ if $1 < \alpha < 2$, $\propto \ln(n_M(R))$ if $\alpha = 2$ and constant if $\alpha > 2$. $n_M(R)$ is the maximum number of people in a square contained in the disk of radius $R$: $R^{d_c} \int_{n_M(R)}^{\infty} dn/n^{\alpha} \lesssim 1$ leading to $n_M(R) \sim R^{\frac{d_c}{\alpha-1}}$. Substituting this result in the previous equation we finally get $d_f = d_c \cdot \max(1, \frac{1}{\alpha-1})$. Note that when $d_c = 2$ and the spatial distribution of population follows the Zipf's law, i.e. $\alpha = 2$, then the fractal dimension is $d_f = 2$; when $1 < \alpha < 2$ the tail of the distribution is longer and the fractal dimension increases.

Inserting this result in the previous equation we obtain:

$$P(n) = \frac{1}{n^\alpha} \sum_{m=1}^{\infty} \frac{1}{m^{\alpha-1}} \left( \frac{1}{m} - \frac{1}{m+n} \right) \simeq \frac{1}{n^\alpha} \frac{n}{\alpha} {}_2F_1[1, \alpha, 1+\alpha, -n] \sim \frac{1}{n^\alpha} \tag{3.8}$$

where ${}_2F_1$ is the hypergeometric function: when $\alpha = 2$, $P(n) \propto \frac{1}{n^2} - \frac{\ln(1+n)}{n^3}$.

We performed numerical simulations to check the goodness of the two derived results for the distribution of distances of trips $P(r)$ and population in the destination $P(n)$. We divided the space with a square grid of $L$ sites per side (each site represents a county) and we assigned a population to each site by randomly extracting a number from a power law distribution $P_0(n) = 1/n^\alpha$. Then we decided the trip destination for every person according to the rules of the Nearest Better Than Me model described above and we calculated the two distributions $P(r)$ and $P(n)$. The results for various values of exponent $\alpha$ are presented in Figure 3.1; the simulations (points) prove to agree with the analytical predictions (lines).

## 3.3 Derivation of the gravity law from the NBTM model

Here we derive a gravity law from the Nearest Better Than Me model described in the previous section. Our gravity law is free of parameters, i.e. it doesn't need a previous knowledge of real fluxes to be calibrated.

The novel aspect of our gravity law is that we show that the number of trips between locations $i$ and $j$ does not depend on the number of people in the destination $n_j$, but only on $n_i$ and the distance between $i$ and $j$.

Notice that this is an illustrative example in order to show the versatility of the model also from the analytical point of view. In fact the calculation needs the input of the population distribution and so, of course, the result depends on it. The population distribution depends crucially on the counties/municipalities subdivisions, which are somewhat arbitrary and surely are different for different countries. This immediately implies that the gravity law cannot be universal. However, as observed above, the numerical implementation of the model bypass all the problems related to the subdivision.

**Derivation of the distance dependence in the Gravity Model** The gravity model is an empirical law that predicts that the number of trips between locations $i$ and $j$ is $T_{ij} = T(n_i, n_j, r_{i,j})$, where $n_i\,(n_j)$ is the number of people in location $i\,(j)$ and $r_{i,j}$ is the distance between location $i$ and $j$. To be specific we assume that $T(n_i, n_j, r_{i,j})$ is separable, i.e. $T_{ij} \propto n_i g(n_j) f(r_{i,j})$, where $f$ and $g$ are functions to be determined. In several studies (*16, 14*) different forms of the gravity law have been proposed; usually a power law is chosen for $g$ and for $f$ a power law or an exponential.

Within the gravity model the probability of a trip of length $r$ departing from a location $i$ with $n_i$ people, $P(r|n_i)$, can be obtained as:

$$P(r|n_i) = \frac{1}{T_i} \sum_j T_{ij} \delta(r - r_{i,j}), \tag{3.9}$$

where $T_i$ is the total number of trips from location $i$ and the function $\delta(x)$ is equal to 1 if $x = 0$ and to 0 otherwise. We have:

$$\begin{aligned} P(r|n_i) &= \frac{1}{T_i} \sum_j n_i g(n_j) f(r_{i,j}) \delta(r - r_{i,j}) \\ &= \frac{n_i}{T_i} f(r) \sum_j g(n_j) \delta(r - r_{i,j}) \simeq \frac{n_i}{T_i} f(r) \cdot r \left\langle g(n) \right\rangle_r \end{aligned} \tag{3.10}$$

where we have used that $\sum_j \delta(r - r_{i,j}) \sim r$ and $\sum_j g(n_j)\delta(r - r_{i,j}) / \sum_j \delta(r - r_{i,j}) \equiv \left\langle g(n) \right\rangle_r$ is the average of $g(n)$ over the locations $j$s at a distance $r$ from the given location $i$. The average number of people in a location at distance $r$, $\langle n \rangle_r$, is independent of $r$ after $\simeq 50$ km (the maximum distance of spatial correlation in population distribution). Thus, considering only the dependence on distance, we have $P(r) \propto rf(r)$.

The Nearest Better Than Me model we proposed states that the probability that the nearest more attractive location is at distance $r$ is $P(r) \sim 1/r^{d_f+1}$ and this implies that $f(r) \sim 1/r^{d_f+2}$.

**Derivation of population dependence in the Gravity Model** Here we derive the dependence on the population in the destination for the Gravity Model, i.e. the $g$ function.

The probability of a trip to a location with $n$ people, $P(n)$, can be obtained within the gravity model as:

$$P(n) \propto \sum_i \sum_j T_{ij} \delta(n - n_j) = g(n) \sum_j \delta(n - n_j) \sum_i n_i f(r_{i,j}) \tag{3.11}$$

The last sum in the previous equation, $\sum_i n_i f(r_{i,j}) \equiv N \langle f(r) \rangle_j$, is the average of $f(r)$ from location $j$ and can be rephrased in the following way:

$$\langle f(r) \rangle_j = \int_l^L dr\, f(r) r^{d_f-1} F\left(\frac{r}{R(n_j)}\right) \tag{3.12}$$

$r^{d_f-1}$ is proportional to the number of people at distance $r$ and $F(x)$ is a function such that $F(x) \xrightarrow{x \gg 1}$ constant and $F(x) \xrightarrow{x \ll 1} x^\gamma$. $R(n_j)$ is the minimum distance from a location with $n_j$ people, for which the global spatial distribution of population is recovered on the shell of radius $R(n_j)$. Let us assume that the populations of two neighbor locations are independent, i.e. there is not a correlation between them. This hypothesis, that is verified when a subdivision of the country in large administrative areas (like counties) is considered, is equivalent to set exponent $\gamma$ equal to 0, and thus $\langle f(r) \rangle_j$ is independent of $j$ and $n_j$, obtaining $P(n) \sim g(n) \sum_j \delta(n - n_j) \sim g(n)P_0(n)$ and consequently $g(n) \sim P(n)/P_0(n) \sim$ constant (cfr eq. 3.8).

This unexpected result reveals that the number of trips is independent of the number of people present in the destination, as long as we can neglect correlations between the population of neighbor locations:

$$T_{ij} \sim \frac{n_i}{r_{i,j}^{d_f+2}} \tag{3.13}$$

The fact that eq. 3.13 does not depend on $n_j$ has to be searched in the "information" already captured by the distance dependence. Indeed, as shown above, $d_f$ is related to the population distribution.

Notice that this gravity law has been derived assuming that the population distribution in the counties/municipalities, $P_0$, is a power law (cfr. equation 3.8). This might be a reasonable assumption if the country's administrative units have approximatively the same area, but this is usually not the case since the administrative subdivision is more related to socio-economic and historical rather than geographical factors. However it is important to stress that the model and the simulation results are completely general and do not depend on any of the specific hypotheses made in the above derivation.

## 3.4 Comparison of the model to real data

There is a caveat that one needs to be aware of when applying the model to real data. This caveat applies only to the cell-phones data-set, for which the spatial resolution (i.e. the density of cell phone towers) is higher than the smallest characteristic scale of a trip. To highlight the problems arising in the model when the spatial resolution is so high, let us consider the extreme situation of the following example. Suppose that a new safety regulation for cell phones antennas imposes a strict limit to their power, forcing the companies to increase their number and place one antenna every 100 meters. Clearly the average distance of trips generated by the model crucially depends on the spatial density of locations (here antennas) and the more dense they are, the shorter will be the average distance of trips. On the contrary, real trips have a characteristic length scale which is independent of the resolution that we use to measure it. Indeed, if the real average length of trips is 20 km it doesn't matter if we measure it with a resolution of 100 m, 1 km or 10 km: it will always be 20 km (the real one). On the other hand if the sensitivity of our measure is 50 km, a value greater than the characteristic trip length, both the model and the trip data can be safely compared for length scale greater than 50 km. Therefore in order to match the simulations with the measures we have to calibrate the model by adjusting the number of locations to be considered, $N_{\text{loc}}$. For the countries we have studied, we find that the results obtained with the model are in good agreement with real data when $\sqrt{A/N_{\text{loc}}} >$ 20 km (where $A$ is the area of the country).

**Comparison to the trips extracted from cell phone calls records** Here we show how we used an anonymized cell phone database to extract information about people's travels and calculate the distribution of trips' distances. The database is the collection of calls made and received by anonymized users (*65, 66, 67*). For each call we know the users involved, the time (date and hour) and the location of the tower that processed the call. Thus we can assign a location (i.e. the tower's position) to each user at the moment she/he makes or receives a call.

The operative definition of a trip is the record of the same user in different locations in two consecutive time intervals of length $\Delta t$ (if the same user is found in different locations during the same time window, he will be randomly assigned to one of them).

Figure 3.2 depicts the comparison between the data and the model's results about the distribution of distances of trips, $P(r)$, and the probability that a trip's destination is a location with a certain population, for $\Delta t = 4$ hours (time window 4-8 a.m./8-12 a.m.). We choose a 4-hours time window because it is the smallest time interval with a stationary distribution of distances; indeed increasing the time window from 1 hour to 2, 3, 4 hours we observe that the distribution of trips' distances has an increasing cut off (i.e. the maximum distance for a trip) due to the limited duration of the trip: for example, in 1 hour it is hard to travel more than 200km. For time windows of 4, 16, 24 hours the cut off remains unchanged, suggesting that 4 hours is enough time to reach the destination for the majority of trips.

**Comparison to the US and Portugal workflows data-set** The US census 2000 provide information about the home and workplace of people. For each pair of the 3129 counties, we have the number of people who are resident in the first county and work in the second. It is thus possible to extract the commuting flows between counties and to calculate the statistical properties of these trips.

Figure 3.3 depicts the comparison between the data and the model's results about the distribution of distances of trips, $P(r)$, and the probability that a trip's destination is a location with a certain population, for trips generated with the model described above.

In the same fashion, we compared the NBTM model to workflows between the 252 municipalities of continental Portugal. The results are presented in Figure 3.4.

To verify the goodness of the model and the gravity law we proposed, we perform direct comparisons between predicted and measured fluxes for each pair of counties/municipalities.

In Figure 3.5 we compare the real fluxes from census data to the fluxes generated by the Nearest Better Than Me model for United States and Portugal.

In Figure 3.6 we compare real and simulated fluxes to the prediction of the gravity law that we derived from our model.

## 3.5 Beyond gravity model: Probabilistic/Stochastic approach to mobility fluxes

As noticeable from figure 3.7 both the gravity model and the real data show a rather large standard deviation. This suggests that a more satisfactory approach has to be intrinsically stochastic. Indeed within our approach we can predict the mobility fluxes probability distribution, $p(k|n, m, r)$, from a location with $n$ individuals to a location with $m$ individuals at distance $r$. In this approach $T(n, m, r)$ is the average of $k$ with the probability distribution $p(k|n, m, r)$:

$$T(n, m, r) \equiv \langle k \rangle = \int_1^\infty dk\, k\, p(k|n, m, r) \tag{3.14}$$

An average value is representative only if it corresponds to the maximum of its distribution, like for example in a gaussian distribution, while it is less informative if the distribution is broad or fat-tailed, like a power-law. To understand what is the degree of uncertainty in the gravity law estimate, we calculated the cumulative distribution $p(> k|n, m, r)$ for 18 triplets of bins: $(r, r + \Delta r)$, $(n, n + \Delta n)$, $(m, m + \Delta m)$. The results are displayed in the two top panel of Figure 3.7, for the real data and for the NBTM model. The distributions of trips have a long tail without any peak and thus the gravity law in itself is inadequate to comprehend the whole complexity inherent to the mobility process.

However the average $T(n,m,r)$ becomes very important if all moments of the distribution depends on it. This happens if the distribution obeys the following scaling:

$$p(k|n,m,r) = \frac{1}{T(n,m,r)}\, f\left(\frac{k}{T(n,m,r)}\right) \tag{3.15}$$

that is $p(k|n,m,r)$ does not depends separately by the four variables $k\,n\,m\,r$ but rather only on the ratio $k/T(n,m,r)$. $f$ is the scaling function whereas the prefactor in eq.(3.15) is simply due to trivial dimensional analysis. Eq.3.15) implies that the scaling for the cumulative distribution is:

$$p(>k|n,m,r) = F\left(\frac{k}{T(n,m,r)}\right) \tag{3.16}$$

where $F(x) = \int_x^\infty \mathrm{d}z\ f(z)$. The bottom panels of Figure 3.7 and Figure 3.8 show that re-plotting the same data as in the above panels –both the data and the NBTM model– versus $k/T(n,m,r)$ leads to a collapse of all the eighteen curves in a single universal curve, $F(x)$. Thus NBTM model provides much more information than the gravity law; in fact it allows to predict the distributions $p(>k|n,m,r)$ from it and therefore have a complete information about the probability of trips and not just its average value. Furthermore it obeys scaling, as expressed by eqs. (3.15) and (3.16), as the data do.

## 3.6 Conclusions

We proposed a new model (the Nearest Better Than Me model) which is able to predict the distribution of the commuting trips between municipalities/counties of a country on short-medium length scale (from ∼10 to ∼1000 km). Unlike the current approaches, our model does not need any previous information about commuting flows but it generates the trips solely considering the spatial distribution of the population.

Our study shows that the spatial distribution of population contains the essential information to reproduce the coarse-grained stochastic dynamic of human motion, and that this process can be modeled with a simple mechanism reflecting the human behavior in the choice of trips' destinations.

From the assumptions of the Nearest Better Than Me model we derived a gravity law which expresses the number of trips between two locations as function of people in the origin and the distance from the destination. Our derivation of the gravity law is free of parameters. More importantly the NBTM model allows to calculate the probability distribution of human mobility. This allows not only to calculate the average mobility flux between any two locations and eventually derive the gravity law, but also any moment of the distribution. We have shown that the distribution of both the real and the predicted fluxes obey scaling and leads to universal behavior.

We successfully tested the predictions of our model comparing the simulated fluxes with real data from two distinct kind of data sets: the census surveys and a data-set of cell phones calls. We showed that these two sources of information about human mobility are consistent and provide a validation to the findings of our model.

For what concerns the practical applications, the model can be used to obtain commuting fluxes for those countries where data about home-workplace flows are not available. Moreover, considering that these kind of surveys are usually carried out once every 5-10 years, it can also yield up to date estimates on commuting fluxes based on the latest data on resident population.

The model can be directly applied to several subjects/studies where human mobility data are needed. For example it can be a useful tool to generate the human dynamic in the models that simulate the diffusion of human infectious diseases, or in the field of urban and transportation planning, it can be used to compare the capacity of the actual transportation system with the need of transportation indicated by the simulated fluxes and can thus be helpful in the design of the future and most urgent improvements to it.

We deliberately kept the model in its simpler form in order to stress the generality and accuracy of its results without the necessity to introduce any parameter. Anyway it can be made more sophisticated with the addition of more complex rules or features and obtain a better agreement with real data. For example, in USA and Portugal, the model tends to underestimate the number of people in the location of destination for highly populated destinations, i.e. highly populated areas attract slightly more people than what the model predicts. This bias could possibly be corrected introducing more complex mechanisms in the rank assignment.

**Acknowledgments** I acknowledge the kind hospitality at Center for Complex Network Research at Northeastern University, Boston. This work was carried out in collaboration with Marta C. González and Albert-Lászlo Barabási.

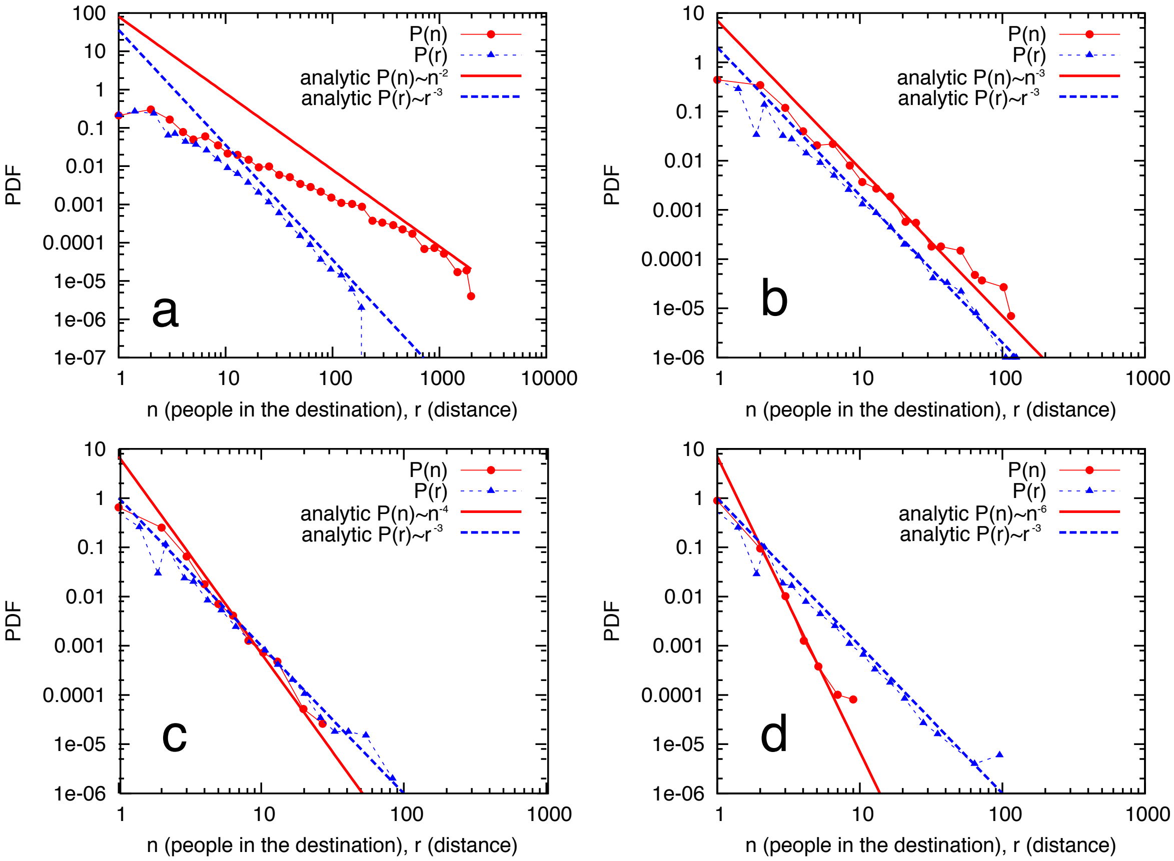


Figure 3.1: The distributions of distances of trips $P(r)$ (blue triangles) and population in the destination $P(n)$ (red circles) obtained from numerical simulations of the Nearest Better Than Me model on a square lattice for various values of exponent $\alpha$ ($\alpha$ is the parameter of population distribution: $P_0(n) \sim n^{-\alpha}$). The analytical predictions are $P(r) \sim r^{-(d_f+1)} = r^{-3}$ (dashed blue lines) and $P(n) \sim n^{-\alpha}$ (solid red lines). From top left to bottom right: panel 'a', side size $L = 200$ and $\alpha = 2$; panel 'b', $L = 150$ and $\alpha = 3$; panel 'c', $L = 150$ and $\alpha = 4$; panel 'd', $L = 150$ and $\alpha = 6$. For the case $\alpha = 2$ the agreement between numerical and analytical results is harder to observe because of logarithmic corrections.

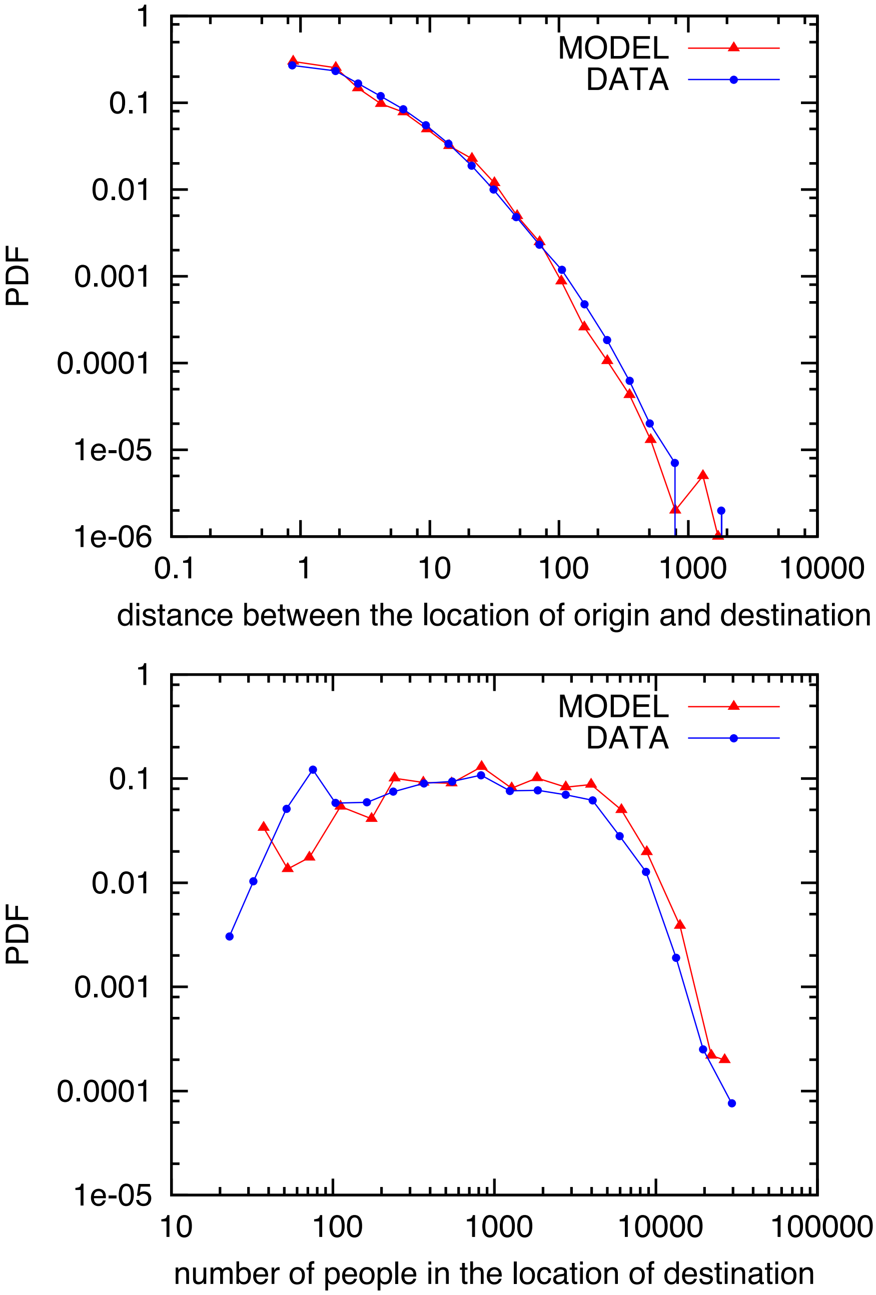


Figure 3.2: Above, the distribution of trips' distances $r$ for trips extracted from cell-phones database (blue dots) and generated with the parameter-free Nearest Better Than Me model (red triangles). Below, the distribution of the population in the location of destination (i.e. the probability that the population of the destination is $n$). The number of cell phone towers has been reduced of 12 times, following the procedure described in the text.

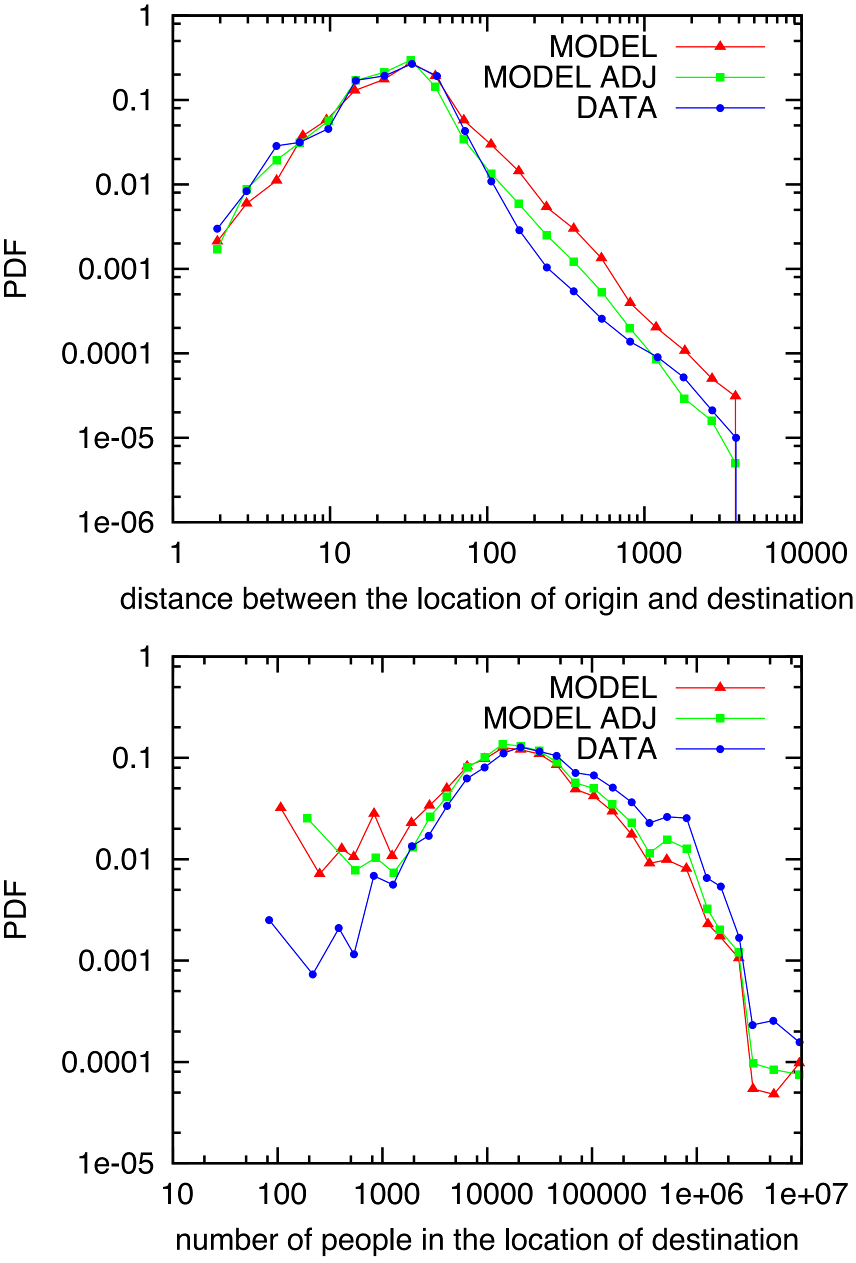


Figure 3.3: Above, the distribution of trips' distances $r$ for work flows given by US census 2000 (blue dots) and generated with the rank model (red triangles). The green squares correspond to the rank model when the number of trips from each location is the real one, obtained from workflows data, instead of using the resident population. Below, the distribution of the population in the location of destination (i.e. the probability that the population of the destination is $n$).

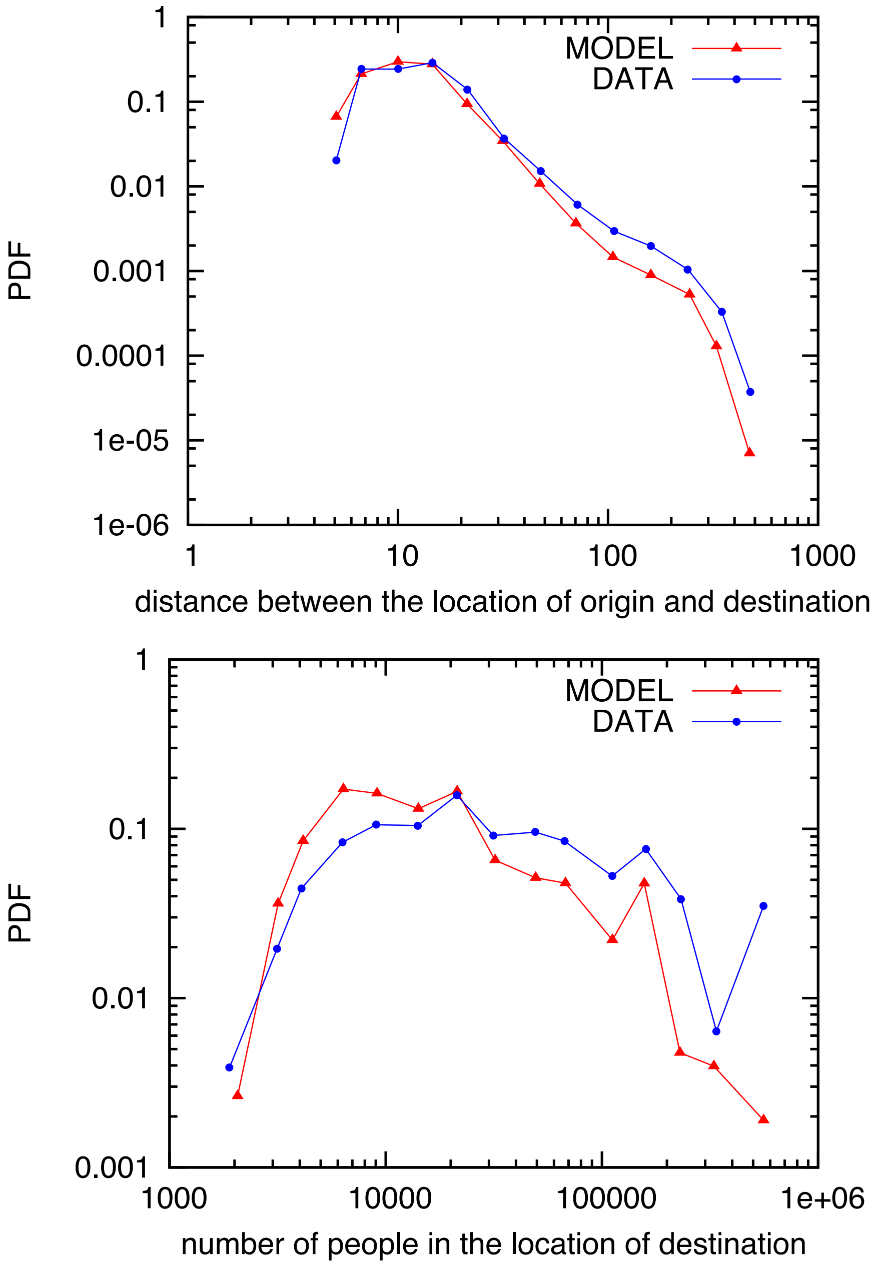
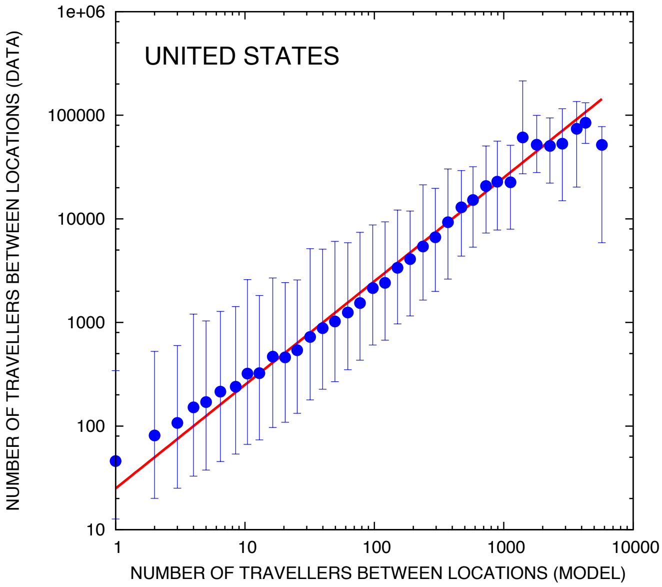


Figure 3.4: Above, the distribution of trips' distances $r$ for work flows given by Portugal census 2001 (blue dots) and generated with the rank model (red triangles). Below, the distribution of the population in the location of destination (i.e. the probability that the population of the destination is $n$).



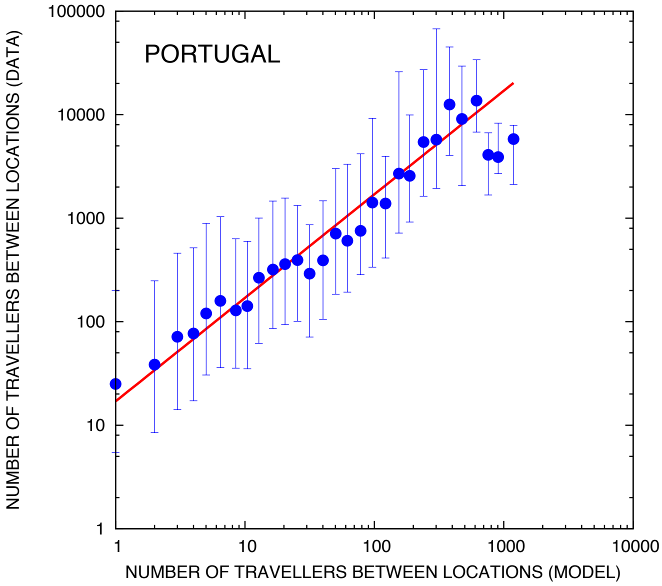


Figure 3.5: In this figure we compare directly the fluxes between locations generated by the model with the ones from the data. We draw the figure in the following way: for every pair of locations, on the x-axis is the flux obtained from the model while on the y-axis the corresponding flux from data is reported. Then we group together the points in logarithmic bins to obtain an average value and to measure the dispersion around the mean (i.e. the asymmetric error bars in the figures are the root mean squares for points above and below the mean). The points lay close to the red solid line $y \propto x$ meaning that, on average, the fluxes generated by the model are proportional to the real ones.

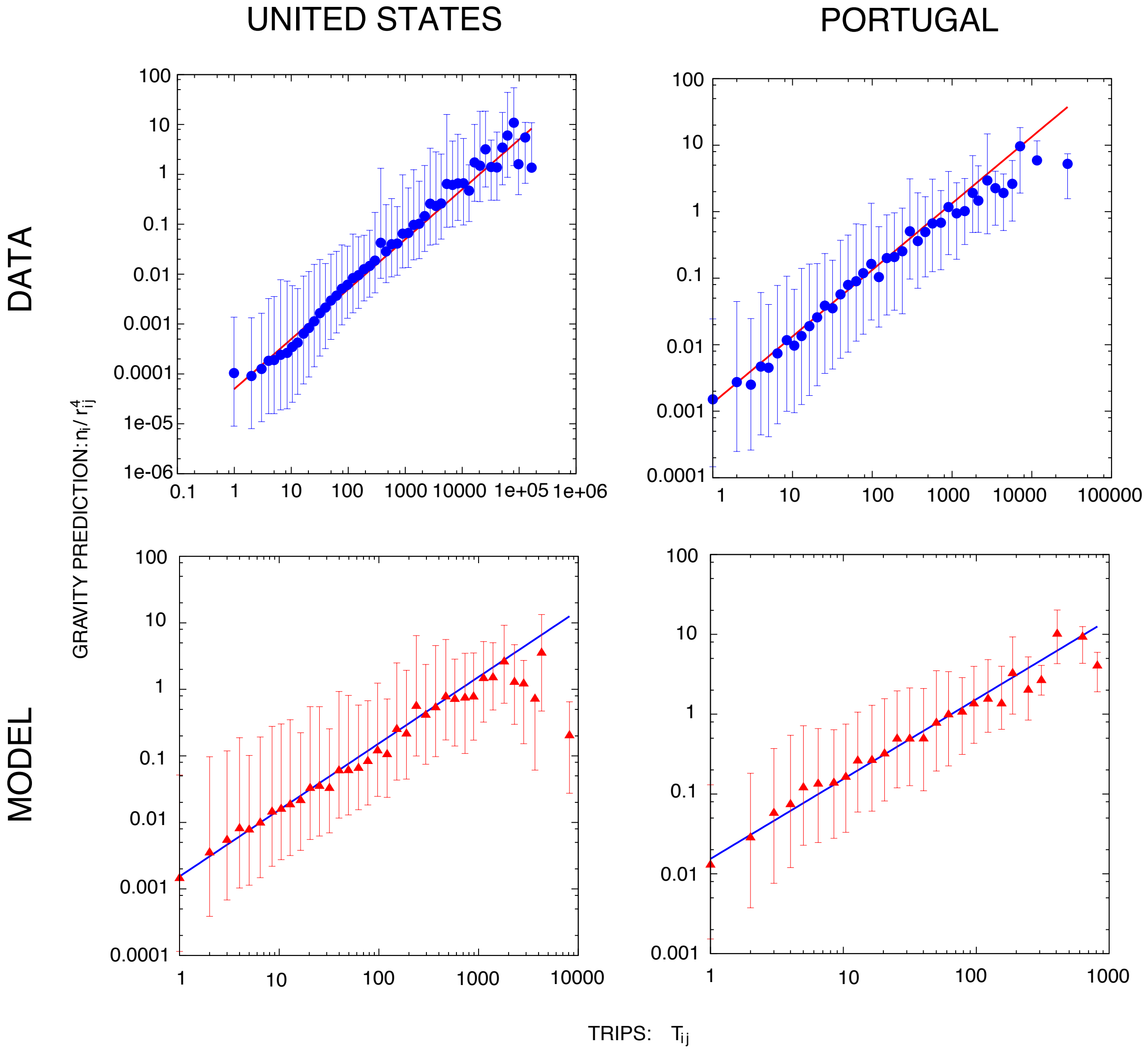


Figure 3.6: The test of the gravity law derived in the text for both data and model. We compare directly the fluxes between locations generated either by the model or the data with the predictions given by the gravity law (eq. 3.13). We draw the figure in the following way: for every pair of locations, on the x-axis is the measured or simulated flux while on the y-axis the flux predicted by the gravity law $n_i/r_{i,j}^{d_f+2}$ is reported. Then we group together the points in logarithmic bins to obtain an average value and to measure the dispersion around the mean (i.e. the asymmetric error bars in the figures are the root mean squares for points above and below the mean). The points lay close to the red solid line $y \propto x$ meaning that, on average, the fluxes measured or generated by the model are proportional to the prediction of the gravity law we derived from the Nearest Better Than Me model.

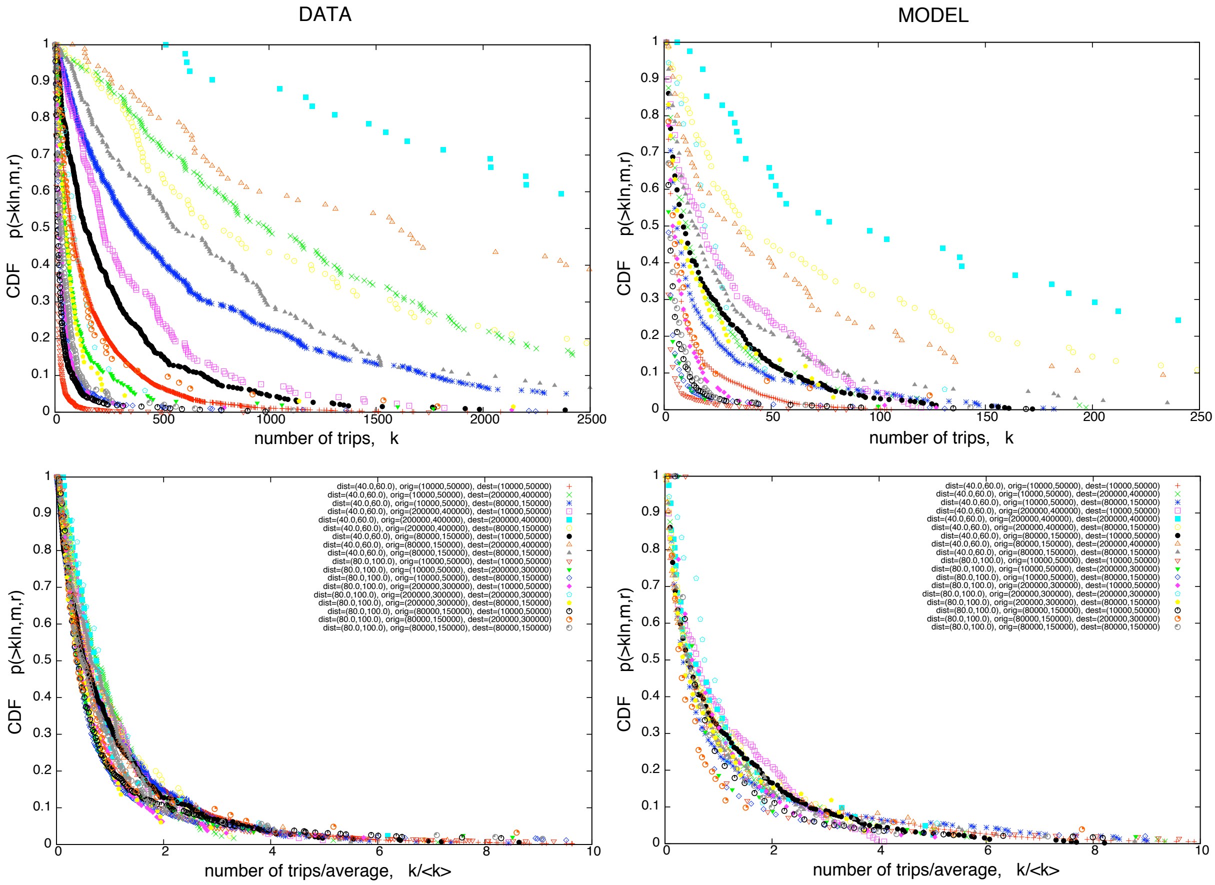


Figure 3.7: The cumulative distribution $p(> k|n, m, r)$, the probability to have more than $k$ trips, for 18 triplets $(r, n, m)$. We selected trips at two distances: $(r, r + \Delta r) = (40, 60)$km or $(80, 100)$km; originating from locations in three bins of population: $(n, n + \Delta n) = (10000, 50000)$, $(80000, 150000)$ or $(200000, 3-400000)$; and directed to destinations in the same three bins of population. In the above panel the results for the US census data (left) and the NBTM model (right) are shown, while in the panel below there are the respective collapses: the same data are plotted versus $k/\langle k\rangle$. The range of $k$ in the simulation is less than the range of data because only a fraction of trips have been generated. Nevertheless a similarity between data and model can be seen both in the shape of the distributions and in the relative positions of most of the curves. In particular, the cumulative scaling function $F(k/\langle k\rangle)$ (defined in eq. 3.16) obtained with the collapses is close to an exponential for both data and model.

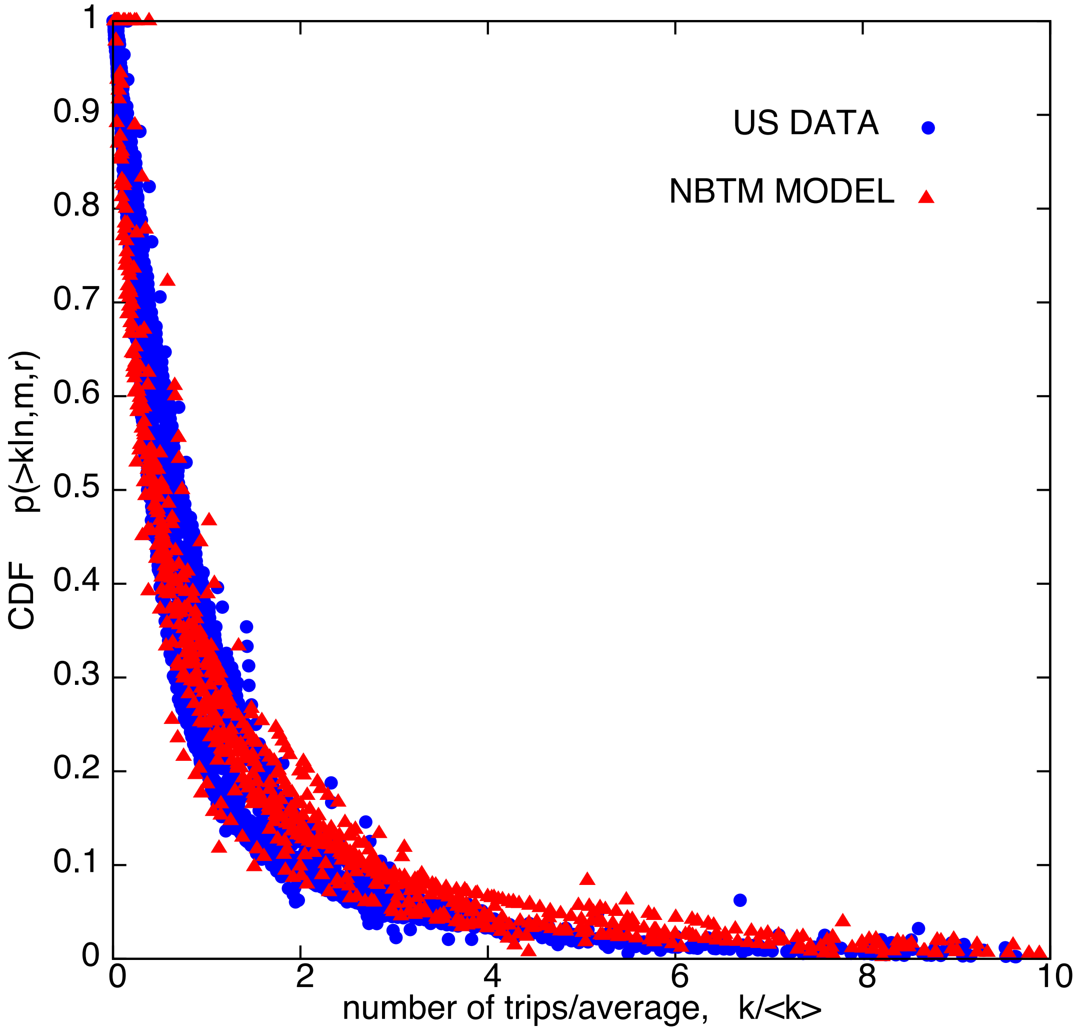


Figure 3.8: A comparison of the scaling collapses of bottom panels of Figure 3.7. The 18 triplets considered in the above figure are plotted versus $k/\langle k\rangle$ for the US census data (blue circles) and the simulations of the NBTM model (red triangles). The two collapses are superimposed, meaning that the cumulative scaling function $F(k/\langle k\rangle)$ (defined in eq. 3.16) generated with the NBTM model is the same of real data.

# Chapter 4

# Conclusions

For the two complex systems we analyzed in this work, tropical forests and human mobility, we were able to classify the empirical transportation networks as optimal configurations that derive from the definition of appropriate variational principles.
In particular, we showed that the tree size distribution in tropical forests is determined by the shape of each tree (i.e. the scaling relationship between height and crown extension), when resources are fully used. In turn, the shape of trees is the result of the optimization in their use of resources, and can be obtained maximizing the metabolic rate with respect to a given tree mass.
In the same fashion we were able to define the coarse-grained laws of human mobility with a simple optimization process. We observed that the commuting fluxes emerge by optimizing the destination's choice of each individual, balancing between attractiveness and distance of locations.

It is important to remark the main role and the utility of the scaling framework in our analysis.
In both systems we were able to identify scaling relationships between the important variables, and thanks to finite-size scaling we succeeded in linking together seemingly unrelated quantities. This provided new insights and helped in discovering universal relationships.

For example in tropical forests we managed to relate many different characteristics of a tree (height, diameter, crown radius, mass, metabolic rate, distance form competitors, characteristic biological time scales) using scaling relationships whose exponents depend on a *single* parameter, the tree shape exponent $H$. In addition we used the powerful scaling collapse procedure to deduce the range of parameters over which pure power law behaviour holds and determine the mutual consistency of the exponents.

In the analysis of human trips we discovered that the stochastic approach of our model is eminently suited to describe the probabilistic nature of human mobility process. Indeed one remarkable achievement of the scaling analysis is that the probability to observe $k$ trips from two locations with fixed populations ($n$, $m$) and at a given distance ($r$) does not depend separately on the four variables but rather only on the ratio $k/T(n, m, r)$. This allows to perform a scaling collapse of the probability distributions of trips for different values of $n, m, r$ in a single universal curve. Additionally, the scaling collapse provides a strong validation for our model, because the collapse for the simulations and real data is on the *same* scaling function.

Future generalizations and extensions of the results presented in this work are currently under consideration.

Two main problems are still open in the study of size distribution in forests. The first question is whether the framework we have developed for tropical forests might also apply to non-tropical climates, i.e. to temperate and boreal forests. Experimental data for cold-climate forests are fewer and with poorer statistic, but the size distributions apparently do not show the universality features of tropical forests. In particular these data would yield a smaller value for the shape exponent $H$, leading to a sub-optimal metabolic rate–mass ratio. The ability to account for

the causes and implications of these differences between temperate and tropical forests will be the main challenge to the framework we have developed.
The second issue deals with the study of size distribution *within* the species of a tropical forest. As observed above, the species are not equivalent with respect to the size of individuals. In fact there are species with a small average height, species characterized by tall individuals, as well as species with trees of all sizes. Which are the reasons for this differences? Is it a sign that the neutrality assumption does not hold in this case?

On the other hand, regarding human mobility, the opportunity of future improvements is boundless.
One of the first tasks is to test further the model predictions. We showed the agreement of the NBTM model to census data and cell phones data, and it would be important to test its predictions with more accurate measures, like GPS data, in order to explore in finer detail the temporal and spatial scales of its validity.
Then our model and its future generalizations could be a useful tool to analyze human mobility at the individual scale, and obtain informations on the distributions of the radii of gyration and of the waiting times between two consecutive trips.
Finally, there is the possibility to implement the NBTM algorithm in the simulations for applied studies, like epidemic spreading and transportation planning.

# Appendix A

# Scaling and self-similarity

Scaling and power law relationships are observed when the phenomenon being studied does not exhibit a characteristic length scale (*50, 53, 51, 54, 17*). Typically, there are both lower and upper cut-off scales for power law behaviour and if these are well separated (say, by several orders of magnitude), scaling could hold in an intermediate range. In physical systems, one can discern the scaling regime by increasing the upper cut-off scale or the correlation length by adjusting the temperature or the pressure closer to its critical value. No such tuning is possible in an ecological community. The diameter distribution of trees in a forest has a lower cut-off scale set by the size of the plant upon recruitment whereas the upper cut-off is necessarily less than the typical diameter of the largest tree in the forest. Determining the scaling regime and even verifying that a scaling description is valid in an ecological community can be a challenge.

Finite size scaling postulates that the PDF of tree diameters has the form $p_r(r|r_c) = r^{-\alpha} f_r(r/r_c)$ when $r$ is larger than a lower (unspecified) cut-off value. This scaling form is a power law decay $r^{-\alpha}$ characterized by an exponent $\alpha$, but modified by a scaling function $f_r(r/r_c)$, where $r_c$ represents the upper cut-off. Over a range of $r$ values, for which the scaling function is approximately constant, one obtains pure power law behaviour. The scaling function $f_r(r/r_c)$ has the property that it decays to zero rapidly when its argument $(r/r_c)$ becomes larger than 1 or when the tree diameter becomes larger than the cut-off value. In this regime, the PDF is dominated by a characteristic length, the cut-off scale, and pure power law behaviour is lost. This ensures that the PDF appropriately vanishes when the tree diameter becomes larger than its cut-off value. Indeed, power law scaling is expected to hold only when the diameter is much smaller than its cut-off value. The exponent $\alpha$ is expected to be universal and depends only on certain essential attributes, whereas the scaling function $f_r$ can depend on details such as the climate and the resource availability in a given forest. For another variable, such as the height $h$ or the crown radius $r_{cro}$, the exponent $\alpha$ of the PDF has to be determined using the transformation rules described below. Such a scaling form can be used to describe the PDF of variables beyond the range over which they exhibit pure power law behaviour.

When a scaling relation, $y \sim x^{\omega}$, exists between two random variables $x$ and $y$ (for example $x$ could be the height, $h$, and $y$ could be the tree diameter, $r$), it is meant that the conditional probability distribution of $y$ given $x$, $P_y(y|x)$, satisfies the relationship (we use $P$ for the conditional probability of two random variables whereas we use $p$ for PDF of a single random variable):

$$P_y(y|x) = \frac{1}{y} F_y \left( \frac{y}{x^{\omega}} \right) \tag{A.1}$$

This is the correct generalization of the deterministic relation $y = x^{\omega}$ to a more general case in which the $n$-th moment scales as $\langle y^n \rangle = c_n x^{\omega n}$, where the $c_n$s are constants. The deterministic case is obtained when $c_n = 1$. In the case finite size scaling holds for the PDF of random variable $x$, $p_x(x|x_c) = x^{-\alpha} f_x(x/x_c)$, the corresponding PDF for the random variable $y$, obeying eq. A.1, is $p_y(y|y_c) = y^{-(\alpha+\omega-1)/\omega} f_y(y/y_c)$ with the cut-offs transforming in the natural manner, i.e.

$y_c = x_c^\omega$, and two scaling functions $f_x$ and $f_y$ related through an integral equation involving the $F_y$ which appears in eq. A.1. The power law exponent is the one it would expected by the standard change of variable rule for PDF, i.e. $p_y(y) = p_x(x)|dx/dy|$ with $|dx/dy| = y^{(1-\omega)/\omega}/\omega$. As $x$ varies, in principle, one would obtain independent curves $P_y(y|x)$ versus $y$. However, if Eq. A.1 holds, all these curves can be collapsed on to a single curve if one plots $yP_y(y|x)$ (or equivalently the cumulative $P_y^>(y|x) = \int_y^\infty dy' P_y(y'|x))$ versus $y/x^\omega$. In other words, for a given $x$, the characteristic scale of $y$ is $x^\omega$. Viewed in this manner, all curves appear the same. An example where Eq. A.1 holds is shown in Fig. 2.10.

**Confirmation of the validity of finite-size scaling. Derivation pertaining to Figure 2.11** Let us define the probability density functions (PDFs) for two measurable variables, diameter $r$ and distance from the bigger nearest $r_i$:

$$p_r(r|r_c) = r^{-\alpha_r} f_r\left(\frac{r}{r_c}\right) \tag{A.2}$$

$$p_{r_i}(r_i|r_{i,c}) = r_i^{-\alpha_{r_i}} f_{r_i}\left(\frac{r_i}{r_{i,c}}\right) \tag{A.3}$$

and the conditioned probabilities for the same quantities:

$$P_r(r|r_i) = \frac{1}{r} F_r\left(\frac{r}{g^{-1}(r_i)}\right) \tag{A.4}$$

$$P_{r_i}(r_i|r) = \frac{1}{r_i} F_{r_i}\left(\frac{r_i}{g(r)}\right) \tag{A.5}$$

We do not assume that $g(r)$ is a power law. $g^{-1}(r_i)$ is the inverse of $g(r)$. When $g(r) \sim r^\omega$ the inverse is simply $g^{-1}(r_i) \sim r_i^{1/\omega}$. Eqs. A.4 and A.5 are a generalization of the scaling introduced above. We assume –and data perfectly confirm this as shown in Figure 2.9– that the PDF of $r_i$ has the following superuniversal form:

$$p_{r_i}(r_i|L) = r_i^{-3}\Theta\left(L - r_i\right)$$

where $L$ is the maximum length scale (plot size) and $\Theta$ is the Heaviside's step function: $\Theta(x) = 1(0)$ if $x \geq 0$ ($x < 0$).

We will show that our scaling approach allows one to estimate the distribution A.2 and the conditional probability A.5 in terms of a single function, $g(x)$. This means that we can, for example, fix the parameters defining the function $g$ by fitting the PDF of tree diameters, Eq. A.2 and then verify that the same parameters provide an equally good fit to the distance-diameter plot, Eq. A.5. This result can be used as a cross test to verify the validity of the scaling hypothesis (See Figure 2.11). The derivation uses the standard results for conditional probabilities:

$$\begin{aligned}
p_r(r|r_c) &= \int_0^\infty dr_i\; P_r(r|r_i)\; p_{r_i}(r_i|L) \\
&= \int_0^\infty dr_i\, \frac{1}{r}\, F_r\left(\frac{r}{g^{-1}(r_i)}\right) r_i^{-3}\; \Theta\left(L - r_i\right) \\
&= \int_0^{g^{-1}(L)} dy\, \frac{1}{r}\, F_r\left(\frac{r}{y}\right) g(y)^{-3}\frac{d}{dy}g(y) \\
&= \int_{r/g^{-1}(L)\ll 1}^\infty dz\, \frac{F_r(z)}{z^2} g(r/z)^{-3}\frac{d}{d(r/z)}g(r/z) \simeq -\frac{d}{dr}g(r/z_0)
\end{aligned} \tag{A.6}$$

In the 3rd line we performed the change of variable $y = g^{-1}(r_i)$, in the 4th line $z = r/y$. To justify the last step, observe that if the relation between $r$ and $r_i$ was "deterministic", then $P_r(r|r_i) = \delta(r - g^{-1}(r_i))$ ($\delta$ is the Dirac delta function), which is a special case of Eq. A.4 with $F_r(z) = \delta(z-1)$. In this case, the last step in the above derivation is exact. In the more general case, one expects that $F_r(z)$ has a maximum at $z = 1$ and so one can evaluate the above integral by expanding the integrand around $z = 1$. The approximation becomes better as the maximum sharpens.

The $g$ function also appears in the calculation of the moments of $P_{r_i}(r_i|r)$ distribution:

$$\langle r_i^n \rangle_r = \int_0^\infty dr_i\ P_{r_i}(r_i|r)\ r_i^n = \int_0^\infty dr_i\ F_{r_i}\left(\frac{r_i}{g(r)}\right)\ r_i^{n-1} = g(r)^n \int_0^\infty dy\ F_{r_i}(y)\ y^{n-1} \tag{A.7}$$

thus, $\langle r_i^n \rangle_r \sim g(r)^n$.

Combining the two results, we obtain a direct relation between Eq. A.2 and the first moment of Eq. A.5 in terms of the $g$ function:

$$p_r^>(r|r_c) = \int_r^\infty dx\, p_r(x|r_c) \overset{A.6}{\sim} g(r)^{-2} \overset{A.7}{\sim} \langle r_i \rangle_r^{-2} \tag{A.8}$$

which yields (see Figure 2.11)

$$\langle r_i \rangle_r \sim \frac{1}{\sqrt{p_r^>(r|r_c)}} \tag{A.9}$$

# Bibliography

1. Banavar J.R., Maritan A., Rinaldo A.; "Size and form in efficient transportation networks"; *Nature*; **399** 130–132 (1999).

2. Rodriguez-Iturbe I., Rinaldo A.; *Fractal river basins: chance and self-organization*; Cambridge Univ Pr (2001).

3. Banavar J.R., Colaiori F., Flammini A., Maritan A., Rinaldo A.; "Scaling, Optimality, and Landscape Evolution"; *Journal of Statistical Physics*; **104**(1) 1–48 (2001).

4. Simini F., Rinaldo A., Maritan A.; "Universal scaling of optimal current distribution in transportation networks"; *Physical Review E*; **79**(4) 46110 (2009).

5. Kleinberg J.M.; "Navigation in a small world"; *Nature*; **406**(6798) 845 (2000).

6. Li G., Reis S., Moreira A., Havlin S., Stanley H., Jr J.A.; "Designing optimal transport networks"; *Arxiv preprint arXiv:0908.3869* (2009).

7. Bianconi G., Pin P., Marsili M.; "How relevant are features for network structure?"; *Arxiv preprint arXiv:0810.4412* (2008).

8. Bonan G.B.; "Forests and climate change: forcings, feedbacks, and the climate benefits of forests"; *Science*; **320**(5882) 1444 (2008).

9. Canadell J.G., Raupach M.R.; "Managing forests for climate change mitigation"; *Science*; **320**(5882) 1456 (2008).

10. Korner C.; "ATMOSPHERIC SCIENCE: Slow in, Rapid out–Carbon Flux Studies and Kyoto Targets"; *Science*; **300**(5623) 1242 (2003).

11. Hubbell S.P.; *The Unified Neutral Theory of Biodiversity and Biogeography*; Princeton University Press (2001).

12. Azaele S., Pigolotti S., Banavar J.R., Maritan A.; "Dynamical evolution of ecosystems"; *Nature*; **444**(7121) 926–928 (2006).

13. Ferguson N.M., Cummings D.A.T., Fraser C., Cajka J.C., Cooley P.C., Burke D.S.; "Strategies for mitigating an influenza pandemic"; *Nature*; **442**(7101) 448–452 (2006).

14. Viboud C., Bjornstad O.N., Smith D.L., Simonsen L., Miller M.A., Grenfell B.T.; "Synchrony, waves, and spatial hierarchies in the spread of influenza"; *Science*; **312**(5772) 447–451 (2006).

15. Colizza V., Barrat A., Barthélemy M., Vespignani A.; "The role of the airline transportation network in the prediction and predictability of global epidemics"; *Proceedings of the National Academy of Sciences*; **103**(7) 2015 (2006).

16. Balcan D., Colizza V., Gonçalves B., Hu H., Ramasco J.J., Vespignani A.; "Multiscale mobility networks and the spatial spreading of infectious diseases"; *Proceedings of the National Academy of Sciences*; **106**(51) 21484 (2009).

17. Wilson A.G.; "The Use of Entropy Maximising Models, in the Theory of Trip Distribution, Mode Split and Route Split"; *Journal of Transport Economics and Policy*; pp. 108–126 (1969).

18. González M.C., Hidalgo C.A., Barabási A.L.; "Understanding individual human mobility patterns"; *Nature*; **453**(7196) 779–782 (2008).

19. Kleiber M.; "The fire of life. An introduction to animal energetics."; *New York* (1961).

20. Schmidt-Nielsen K.; *Scaling, why is animal size so important?*; Cambridge Univ Pr (1984).

21. Pacala S.W., Canham C.D., Jr J.S.; "Forest models defined by field measurements: I. The design of a northeastern forest simulator"; *Can.J.For.Res*; **23**(10) 1980–1988 (1993).

22. West G.B., Brown J.H., Enquist B.J.; "A General Model for the Origin of Allometric Scaling Laws in Biology"; *Science*; **276**(5309) 122 (1997).

23. Enquist B.J., Brown J.H., West G.B.; "Allometric scaling of plant energetics and population density"; *Nature*; **395** 163 (1998).

24. West G.B., Brown J.H., Enquist B.J.; "A general model for the structure and allometry of plant vascular systems"; *Nature*; **400**(6745) 664–667 (1999).

25. Damuth J.; "Scaling of growth: plants and animals are not so different"; *Proceedings of the National Academy of Sciences*; **98**(5) 2113 (2001).

26. Enquist B.J., Niklas K.J.; "Invariant scaling relations across tree-dominated communities"; *Nature*; **410** 655–660 (2001).

27. Meinzer F., Goldstein G., Andrade J.; "Regulation of water flux through tropical forest canopy trees: Do universal rules apply?"; *Tree physiology*; **21**(1) 19 (2001).

28. Niklas K.J., Enquist B.J.; "Invariant scaling relationships for interspecific plant biomass production rates and body size"; *Proceedings of the National Academy of Sciences*; **98**(5) 2922 (2001).

29. Allen A.P., Brown J.H., Gillooly J.F.; "Global biodiversity, biochemical kinetics, and the energetic-equivalence rule"; *Science*; **297**(5586) 1545 (2002).

30. Enquist B.J.; "Universal scaling in tree and vascular plant allometry: toward a general quantitative theory linking plant form and function from cells to ecosystems"; *Tree physiology*; **22**(15) 1045–1064 (2002).

31. Enquist B.J., Niklas K.J.; "Global allocation rules for patterns of biomass partitioning in seed plants"; *Science*; **295**(5559) 1517 (2002).

32. Coomes D.A., Duncan R.P., Allen R.B., Truscott J.; "Disturbances prevent stem size-density distributions in natural forests from following scaling relationships"; *Ecology Letters*; **6** 980–989 (2003).

33. Muller-Landau H.C., Condit R.S., Harms K.E., Marks C.O., Thomas S.C., Bunyavejchewin S., Chuyong G., Co L., Davies S., Foster R.; "Comparing tropical forest tree size distributions with the predictions of metabolic ecology and equilibrium models"; *Ecology Letters*; **9**(5) 589–602 (2006).

34. Kohyama T., Suzuki E.I., Partomihardjo T., Yamada T., Kubo T.; "Tree species differentiation in growth, recruitment and allometry in relation to maximum height in a Bornean mixed dipterocarp forest"; *Journal of Ecology*; pp. 797–806 (2003).

35. Kohyama T.; "Size-structured tree populations in gap-dynamic forest–the forest architecture hypothesis for the stable coexistence of species"; *Journal of Ecology*; pp. 131–143 (1993).

36. Niklas K.J., Midgley J.J., Enquist B.J.; "A general model for mass-growth-density relations across tree-dominated communities"; *Evolutionary Ecology Research*; **5**(3) 459–468 (2003).

37. Muller-Landau H.C., Condit R.S., Chave J., Thomas S.C., Bohlman S.A., Bunyavejchewin S., Davies S., Foster R., Gunatilleke S., Gunatilleke N.; "Testing metabolic ecology theory for allometric scaling of tree size, growth and mortality in tropical forests"; *Ecology Letters*; **9**(5) 575–588 (2006).

38. Enquist B.J., West G.B., Brown J.H.; "Extensions and evaluations of a general quantitative theory of forest structure and dynamics"; *Proceedings of the National Academy of Sciences*; **106**(17) 7046 (2009).

39. West G.B., Enquist B.J., Brown J.H.; "A general quantitative theory of forest structure and dynamics"; *Proceedings of the National Academy of Sciences*; **106**(17) 7040–7045 (2009).

40. Condit R., Hubbell S., Foster R.; "Barro Colorado forest census plot data"; *Online Dataset.URL http://ctfs.si.edu/datasets/bci* (2005).

41. Volkov I., Banavar J.R., He F., Hubbell S.P., Maritan A.; "Density dependence explains tree species abundance and diversity in tropical forests"; *Nature*; **438**(7068) 658–661 (2005).

42. Bond B.J.; "Age-related changes in photosynthesis of woody plants"; *Trends in plant science*; **5**(8) 349–353 (2000).

43. Koch G.W., Sillett S.C., Jennings G.M., Davis S.D.; "The limits to tree height"; *Nature*; **428** 851–854 (2004).

44. Rényi A.; *Selected Transl. Math. Stat. Prob.*; **4** 205 (1963).

45. Evans J.W.; "Random and cooperative sequential adsorption"; *Reviews of modern physics*; **65**(4) 1281–1329 (1993).

46. Iwasa M., Fukuda K.; "The exact probability distribution of saturating states in random sequential adsorption"; *Arxiv preprint arXiv:0810.5632* (2008).

47. Meakin P., Jullien R.; "Random-sequential adsorption of disks of different sizes"; *Physical Review A*; **46**(4) 2029–2038 (1992).

48. Takayasu H., Takayasu M., Provata A., Huber G.; "Statistical properties of aggregation with injection"; *Journal of Statistical Physics*; **65**(3) 725–745 (1991).

49. Anfodillo T., Carraro V., Carrer M., Fior C., Rossi S.; "Convergent tapering of xylem conduits in different woody species"; *New Phytologist*; **169**(2) 279–290 (2006).

50. Kadanoff L.P.; "Scaling laws for Ising models near Tc"; *Physics*; **2**(6) 263âĂŞ272 (1966).

51. Widom B.; "The critical point and scaling theory"; *Physica*; **73**(1) 107–118 (1974).

52. Wilson K.G.; "The renormalization group and critical phenomena"; *Rev.Mod.Phys.*; **55**(3) 583–600 (1983); URL `10.1103/RevModPhys.55.583`.

53. Fisher M.E.; "Critical Phenomena"; *ed. Green M.S.* (1971).

54. Stanley H.E.; "Scaling, universality, and renormalization: Three pillars of modern critical phenomena"; *Reviews of Modern Physics*; **71**(2) 358–366 (1999).

55. Brown J.H., Gillooly J.F., Allen A.P., Savage V.M., West G.B.; "Toward a metabolic theory of ecology"; *Ecology*; **85**(7) 1771–1789 (2004).

56. West G.B., Brown J.H., Enquist B.J.; "A general model for ontogenetic growth"; *Nature*; **413**(6856) 628–631 (2001).

57. Banavar J.R., Damuth J., Maritan A., Rinaldo A.; "Ontogenetic growth (Communication arising) Modelling universality and scaling"; *Nature*; **420**(6916) 626 (2002).

58. Stegen J.C., White E.P.; "On the relationship between mass and diameter distributions in tree communities"; *Ecology Letters*; **11**(12) 1287–1293 (2008).

59. Enquist B., West G., Charnov E., Brown J.; "Allometric scaling of production and life-history variation in vascular plants"; *Nature*; **401** 907–911 (1999).

60. Brown J.H., Gillooly J.F., Allen A.P., Savage V.M., West G.B.; "Toward a metabolic theory of ecology"; *Ecology*; **85**(7) 1771–1789 (2004).

61. González M.C., Barabási A.L.; "Complex networks: From data to models"; *Nature Physics*; **3**(4) 224–225 (2007).

62. Barrat A., Barthélemy M., Pastor-Satorras R., Vespignani A.; "The architecture of complex weighted networks"; *Proceedings of the National Academy of Sciences*; **101**(11) 3747–3752 (2004).

63. Montis A.D., Barthélemy M., Chessa A., Vespignani A.; "The structure of interurban traffic: a weighted network analysis"; *Environment and Planning B: Planning and Design*; **34**(5) 905–924 (2007).

64. Brockmann D., Hufnagel L., Geisel T.; "The scaling laws of human travel"; *Nature*; **439**(7075) 462–465 (2006).

65. Yook S.H., Jeong H., Barabási A.L.; "Modeling the Internet's large-scale topology"; *Proceedings of the National Academy of Sciences*; **99**(21) 13382 (2002).

66. Onnela J.P., Saramäki J., Hyvönen J., Szabó G., Lazer D., Kaski K., Kertész J., Barabási A.L.; "Structure and tie strengths in mobile communication networks"; *Proceedings of the National Academy of Sciences*; **104**(18) 7332 (2007).

67. Wang P., González M.C., Hidalgo C.A., Barabási A.L.; "Understanding the spreading patterns of mobile phone viruses"; *Science*; **324**(5930) 1071 (2009).