



Doctoral thesis

Universitat
de les Illes Balears

**The complexity of movement:
empirical data analysis and modelling of
dynamical processes**

Jorge Pablo Rodríguez García
2018



Doctoral thesis

Universitat
de les Illes Balears

**The complexity of movement:
empirical data analysis and modelling of
dynamical processes**

Jorge Pablo Rodríguez García

2018

Doctoral Program in Physics

Supervisor: Dr. Víctor Martínez Eguíluz

Tutor: Dr. Emilio Hernández García

Doctor in Physics by Universitat de les Illes Balears

List of publications

- JP Rodríguez, J Fernández-Gracia, M Thums, MA Hindell, AMM Sequeira, MG Meekan, DP Costa, C Guinet, RG Harcourt, CR McMahon, M Muelbert, CM Duarte, and VM Eguíluz. Big data analyses reveal patterns and drivers of the movements of southern elephant seals. *Scientific Reports*, 7(1):112, 2017.
- JP Rodríguez, F Ghanbarnejad, and VM Eguíluz. Risk of Coinfection Outbreaks in Temporal Networks: A Case Study of a Hospital Contact Network. *Frontiers in Physics*, 5:46, 2017.
- JP Rodríguez, YH Liang, YJ Huang, and J Juang. Diversity of hysteresis in a fully cooperative coinfection model. *Chaos*, 28(2):023107, 2018.
- AMM Sequeira, JP Rodríguez, VM Eguíluz, R Harcourt, M Hindell, DW Sims, CM Duarte, DP Costa, J Fernández-Gracia, LC Ferreira, GC Hays, MR Heupel, MG Meekan, A Aven, F Bailleul, AMM Baylis, ML Berumen, CD Braun, J Burns, MJ Caley, R Campbell, RH Carmichael, E Clua, LD Einoder, A Friedlaender, F Goebel, SD Goldsworthy, C Guinet, J Gunn, D Hamer, N Hammerschlag, M Hammill, LA Hückstädt, NE Humphries, MA Lea, A Lowther, A Mackay, E McHuron, J McKenzie, L McLeay, CR McMahon, K Mengersen, MMC Muelbert, AM Pagano, B Page, N Queiroz, PW Robinson, SA Shaffer, M Shivji, GB Skomal, SR Thorrold, S Villegas-Amtmann, M Weise, R Wells, B Wetherbee, A Wiebkin, B Wienecke, and M Thums. Convergence of marine megafauna movement patterns in coastal and open oceans. *Proceedings of the National Academy of Sciences*, 115(12):3072-3077, 2018.

Dr. Víctor M. Eguíluz, Científico Titular del Consejo Superior de Investigaciones Científicas

Jorge Pablo Rodríguez García

DECLARAN:

Que la tesis doctoral que tiene por título *The complexity of movement: empirical data analysis and modelling of dynamical processes* se ha realizado por Jorge Pablo Rodríguez García, ha estado dirigida bajo la supervisión del Dr. Eguíluz y que cumple con los requisitos necesarios para optar a los títulos de Doctor Internacional y Doctor Europeo.

Y para que quede constancia de ello firman este documento,

Víctor M. Eguíluz

Jorge Pablo Rodríguez
Palma, 12 de julio de 2018

Acknowledgements

“Amic, approfita la vida que passa”

Marià Aguiló, *Aubada*

A PhD is a long path, typically neither flat nor straight¹. In this trail, that through this thesis is arriving to its destination, I have not been walking alone, and the people next to me have been very important to help me success this challenge.

First of all, I need to acknowledge the two most important women in my life, Ginesa and Chia, mum and wife, as they are supporting me when things become difficult and sharing celebrations when summits are achieved. To them I need to attach Manuel, who is always there in his father role. Thank you because you are my family and the responsible for the happiness in my life.

Secondly, I want to dedicate some lines to my local research group in Mallorca. Thank you Víctor for all the knowledge that you have shared with me and, most importantly, for leading me to a more critical thinking. Paola and Juan, thanks for being there, sharing thoughts, ideas, laughs, moments in general. Thanks to Konstantin for enlightening discussion, especially in the models for ship traffic in Chapter 4, and for being a great colleague organising the Sociomeetings.

A great acknowledgement is for all the collaborators involved in my research along this PhD. Especially, to all that were hosting me in short and long research visits, in Germany (Fakhteh Ghanbarnejad), Australia (Ana Sequeira and Michele Thums), Taiwan (Jonq Juang and Yu-Hao Liang), Switzerland (Carlos Melián) and United Kingdom (Naoki Masuda). I would like also to show my

¹JPR acknowledges partial support from the FPU program of MECD (Spain)

gratitude to Carlos Duarte, Xabier Irigoien and the Office of Sponsored Research of KAUST for providing me with the data used in Chapter 4.

I cannot forget all the IFISC staff, highlighting the colleagues in the S07 room. I appreciate the great and dynamic work environment that we share.

Finally, I try to be cross-disciplinary, not only in research, but also in my life. When my work finishes everyday, there is a busy schedule including anthropology studies, sport and music. Specifically, I would like to thank all the members of *Coral Universitaria de les Illes Balears* and its conductor Joan Company for great musical moments.

Contents

List of publications	iii
Acknowledgements	v
Contents	ix
Abstract	1
Resumen	3
Resum	5
Preface	7
I Introduction	9
1 Introduction	11
1.1 Human tracking	14
1.1.1 Banknotes	15
1.1.2 Mobile phone users	18
1.1.3 Mobility networks	22
1.2 Animal movement	27
1.2.1 Lévy or not?	28
1.2.2 Data analysis of marine animals movement	29
1.3 Dynamical processes in systems of mobile particles	34
1.3.1 Epidemic spreading	34
	vii

1.3.2	Other dynamics	36
1.4	Conclusions	40
II	Marine movement analysis	43
2	Intraspecies analysis of movement: the case study of southern elephant seals	45
2.1	Analysis of displacements	47
2.2	Individual analysis	54
2.3	Spatial networks of flows	59
2.4	Conclusions	60
3	Interspecies analysis of marine megafauna movement	63
3.1	Data description	64
3.2	Global and individual convergence	67
3.3	Analysis of movement similarities	73
3.4	Conclusions	77
4	The global transit pattern described by vessel trajectories	79
4.1	Models of ship transit	81
4.2	Data description	83
4.3	Ports hierarchy	87
4.4	Clustering and highways	88
4.5	Conclusions	92
III	Dynamical processes in mobile networks	95
5	Disease spreading in temporal geometric graphs	97
5.1	Model	98
5.1.1	Topology: temporal random geometric graph	98
5.1.2	Dynamics: Susceptible-Infected-Recovered	101
5.2	Disease spreading in mobile networks	102
5.2.1	Infinite velocity limit	102
5.2.2	Finite velocities	104
5.3	Conclusions	108
6	Cooperative disease spreading in contact and mobile geometric graphs	109
6.1	Models of coinfection	110
6.1.1	Partial coinfection in mean field	110
6.1.2	Partial coinfection in complex networks	112

<i>CONTENTS</i>	ix
6.1.3 Full coinfection in mean field	116
6.2 Static random geometric graph	122
6.3 Mobile random geometric graph	125
6.4 Empirical contact network	127
6.5 Conclusions	132
IV Conclusions and appendix	133
7 General conclusions and future outlook	135
Appendix A: Fitting distributions	137
Bibliography	141

Abstract

The continuously growing amount of data, commonly referred as Big Data, represents two challenges for the scientific community. First, the large size datasets require the development of methods for a scalable analysis, like machine learning or statistical measurements, increasing the capacity for the recognition of the patterns that characterize the data. Secondly, in order to describe the underlying mechanisms in those patterns and increase the forecast power, it highlights the need for simple models that, with very few parameters, are able to capture the observed dynamical processes.

In this thesis, from the analysis of empirical data to the modelling of processes in mobility networks, we present five mobility works under the spirit of these two challenges.

Traditionally, animal movement researchers made remarkable efforts for developing their studies using small size datasets, as the data collection was expensive in terms both of human resources and devices. However, the development of new transmission devices, that are lighter and cheaper, is speeding up the data collection processes, leading to larger datasets. Complexity science has developed methods that are fast and computationally cheap for analysing human movement, and we argue that these methods will help in the study of animal movement. This aim is tackled in the second and the third chapters, where we study first the case of southern elephant seals, revealing a universal behaviour despite the observed idiosyncrasy among individuals, for finding later the relevant drivers of marine megafauna movement, in a joint dataset including fifty different species. These two studies were developed in the framework of the Marine Megafauna Analytical Program (<https://mmap.wordpress.com>), that

Abstract

promotes cross-disciplinary collaborations to analyse the movement of marine animals.

Mobility in the oceans is not only restricted to animals, with most of economic traffic being transported in vessels, both for goods, raw materials and fuel. Specifically, according to the 2017 Review of Maritime Transport, ships carry more than 70% of global trade value. In fact, characterizing the paths followed by vessels in the oceans is not only important for economical purposes, but also for the protection of the species living in this environment. This motivates the analysis of the spatial patterns described by vessel traffic in the ocean, in chapter 4.

The second part of the thesis focuses on dynamical processes and mobility models. Nowadays, transport networks play a key role in disease spreading in a global scale, allowing pathogens to travel, within their hosts, for thousands of kilometres in a few hours. This potential risk needs the formulation of simple models that reveal the basic mechanisms underlying the spreading in mobility networks. In chapter 5 we show how, for short interaction ranges, and counterintuitively, the mobility can have a detrimental effect on a contagion process, with the disease affecting a higher fraction of population when the network is static rather than in the mobile case. In chapter 6 we review recently introduced cooperative disease spreading models, for later studying them in static and mobile scenarios. While the static scenario allows us to match the short and long range interaction limits, which were leading to different behaviours, in the mobile case we show how, for short range interactions, the mobility-induced mixing can lead to the observed effects in a system with long range interactions. Finally, in an effort to bridge the gap between the two parts of the thesis, we propose a data-based model, in which we analyse the effects of cooperative disease spreading dynamics in an empirical contact network, finding that the temporal correlations and the specific activity pattern play a key role in our results.

Resumen

La cantidad creciente de datos, conocida como Big Data, representa dos desafíos para la comunidad científica. Primero, conjuntos de datos de gran tamaño requieren métodos para un análisis escalable, como el aprendizaje automático o las medidas estadísticas, aumentando la capacidad de reconocimiento de los patrones que caracterizan los datos. Segundo, para describir los mecanismos subyacentes a esos patrones y aumentar el poder de predicción, subraya la necesidad de utilizar modelos simples que, con pocos parámetros, son capaces de explicar los procesos observados.

En esta tesis, desde el análisis de datos empíricos hasta el modelado de procesos en redes de movilidad, presentamos cinco trabajos sobre movilidad bajo el espíritu de estos dos desafíos.

Históricamente, los investigadores de movimiento animal hicieron esfuerzos encomiables para desarrollar sus estudios utilizando pequeños conjuntos de datos porque su recolección era cara en términos de recursos humanos y dispositivos. Sin embargo, el desarrollo de nuevos transmisores, más ligeros y baratos, está acelerando la toma de datos, conduciendo a mayores conjuntos de datos. La ciencia de la complejidad ha desarrollado métodos que son rápidos y no requieren gran esfuerzo computacional para analizar el movimiento humano, y nosotros argumentamos que estos métodos ayudarán al estudio del movimiento animal. Este objetivo corresponde a los capítulos 2 y 3 donde, en primer lugar, estudiamos el caso de los elefantes marinos del sur, mostrando un comportamiento universal pese a la idiosincrasia observada entre los individuos, y después buscamos las variables que explican el movimiento de megafauna marina, en un conjunto de datos que agrupa cincuenta especies. Estos dos estudios se desarrollaron en el contexto del Marine Megafauna Ana-

Resumen

lytical Program (<https://mmap.wordpress.com>), que promueve colaboraciones interdisciplinarias para analizar el movimiento de animales marinos.

La movilidad en los océanos no está restringida a los animales, puesto que la mayoría del tráfico económico es transportada en barcos, tanto manufacturas, como materias primas y combustible. Según el 2017 Review of Maritime Transport, los barcos transportan más del 70% del valor del comercio global. De hecho, la caracterización de los caminos de los barcos en los océanos es importante no sólo por razones económicas, sino también para proteger las especies que viven en este entorno. Estos hechos motivan el análisis de los patrones espaciales descritos por el tráfico de barcos en el océano, en el capítulo 4.

La segunda parte de la tesis analiza procesos dinámicos y modelos de movilidad. Hoy en día, los transportes juegan un papel fundamental en la propagación de enfermedades a escala global, permitiendo a los patógenos viajar miles de kilómetros en pocas horas con sus portadores. Este riesgo requiere la formulación de modelos simples que revelen los mecanismos de propagación en redes móviles. En el capítulo 5 mostramos cómo, para interacciones de corto alcance, contra nuestra intuición, la movilidad puede tener un efecto contra los procesos de contagio, afectando a una mayor cantidad de población en el caso estático que en el móvil. En el capítulo 6 resumimos los modelos recientemente introducidos de propagación de enfermedades cooperativas, estudiándolos después en redes estáticas y móviles. Mientras que el caso estático nos permite interpolar entre las interacciones de corto y largo alcance, que conducen a comportamientos distintos, en el caso móvil mostramos cómo, para interacciones de corto alcance, la mezcla causada por la movilidad puede conducir a los efectos observados con largo alcance. Finalmente, en un esfuerzo por unir las dos partes de esta tesis, proponemos un modelo basado en los datos, analizando los efectos de la propagación de enfermedades cooperativas en una red de contactos, mostrando el papel clave de las correlaciones temporales y el patrón de actividad específico.

Resum

La quantitat creixent de dades, coneguda com Big Data, representa dos desafiaments per a la comunitat científica. Primer, conjunts de dades de grans dimensions requereixen mètodes per a una anàlisi escalable, com l'aprenentatge automàtic o les mesures estadístiques, augmentant la capacitat de reconeixement dels patrons que caracteritzen les dades. Segon, per descriure els mecanismes subjacents a aquests patrons i augmentar el poder de predicció, subratlla la necessitat d'utilitzar models simples que, amb pocs paràmetres, són capaços d'explicar els processos observats.

En aquesta tesi, des de l'anàlisi de dades empíriques fins a la modelització de processos en xarxes de mobilitat, presentem cinc treballs sobre mobilitat sota l'esperit d'aquests dos desafiaments.

Històricament, els investigadors de moviment animal van fer esforços encomiables per desenvolupar els seus estudis utilitzant petits conjunts de dades perquè la seva recollecció era cara en termes de recursos humans i dispositius. No obstant això, el desenvolupament de nous transmissors, més lleugers i barats, està accelerant la presa de dades, conduint a majors conjunts de dades. La ciència de la complexitat ha desenvolupat mètodes que són ràpids i que no requereixen gran esforç computacional per analitzar el moviment humà, i nosaltres argumentem que aquests mètodes ajudaran a l'estudi del moviment animal. Aquest objectiu correspon als capítols 2 i 3 on, en primer lloc, estudiem el cas dels elefants marins del sud, mostrant un comportament universal tot i la idiosincràsia observada entre els individus, i després cerquem les variables que expliquen el moviment de megafauna marina, en un conjunt de dades que agrupa cinquanta espècies. Aquests dos estudis es van desenvolupar en el context del Marine

Resum

Megafauna Analytical Program (<https://mmmap.wordpress.com>), que promou col·laboracions interdisciplinars per analitzar el moviment d'animals marins.

La mobilitat dins els oceans no es limita als animals, ja que la majoria del tràfic econòmic és transportada en vaixells, tant manufactures, com matèries primeres i combustible. Segons el 2017 Review of Maritime Transport, el vaixells transporten més del 70% del valor del comerç global. De fet, la caracterització dels camins dels vaixells en els oceans és important no només per raons econòmiques, sinó també per protegir les espècies que viuen en aquest entorn. Aquests fets motiven l'anàlisi dels patrons espacials descrits pel trànsit de vaixells a l'oceà, en el capítol 4.

La segona part de la tesi es centra en processos dinàmics i models de mobilitat. Avui dia, els transports juguen un paper fonamental en la propagació de malalties a escala global, permetent als patògens viatjar milers de quilòmetres en unes poques hores amb els seus portadors. Aquest risc requereix la formulació de models simples que revelin els mecanismes de propagació en xarxes mòbils. Al capítol 5 mostrem com, per interaccions de curt abast, contra la nostra intuïció, la mobilitat pot tenir un efecte contra els processos de contagi, afectant una major població en el cas estàtic que en el mòbil. En el capítol 6 es resumeixen els models recentment introduïts de propagació de malalties cooperatives, estudiant després en xarxes estàtiques i mòbils. Mentre que el cas estàtic ens permet interpolar entre les interaccions de curt i llarg abast, que condueixen a comportaments diferents, en el cas mòbil mostrem com, per interaccions de curt abast, la barreja causada per la mobilitat pot conduir als efectes observats amb llarg abast. Finalment, en un esforç per unir les dues parts d'aquesta tesi, proposem un model basat en les dades, en el qual analitzem els efectes de la propagació de malalties cooperatives en una xarxa de contactes, mostrant el paper clau de les correlacions temporals i el patró d'activitat específic.

Preface

“The world of humankind constitutes a manifold, a totality of interconnected processes, and inquiries that disassemble this totality into bits and then fail to reassemble it falsify reality [...]. If there are connections everywhere, why do we persist in turning dynamic, interconnected phenomena into static, disconnected things?”

Eric Wolf, *Europe and the people without history*

The anthropologist Eric Wolf, in this quote that could appear perfectly in the introduction to any complexity science book, highlights the importance of addressing problems from a holistic point of view, considering the whole and the connections between the different components of the systems in a dynamic approach. This holistic approach needs for cross-disciplinary work in order to join the knowledge from different fields to achieve a global comprehension of the system.

Statistical physics has developed some tools for addressing this challenge, understanding systems as a whole, giving birth to the complexity science. This complexity science is very useful for characterizing empirical data and revealing the mechanisms underlying the dynamical processes, but is also a source of questions for many other disciplines, encouraging the scientific community to increase the cross-disciplinary work.

Looking for connections everywhere, *i.e.*, considering the systems as a whole, was one of my motivations to start this PhD. These thoughts lead me to work in the two clear parts that are included in this thesis, integrating data analysis

Preface

and modelling. However, the connections are also present here, showing for example how disease spreading models can be applied to empirical mobility datasets.

Part I

Introduction

Chapter 1

Introduction

Our lives are built around spatial and temporal interactions. Oral communication has spatial constraints, as people separated by high distances cannot talk directly with each other and, as the level of the external noise increases, the maximum distance for a good transmission of the message decreases; specifically, the speech intelligibility scores, measuring the percentage of words correctly understood, increase with the speech-to-noise ratio, and show a remarkable improvement when the listener has visual contact with the speaker's facial movements [1]. Time is present as well in our social interactions: we follow circadian rhythms, typically with valleys of activity at night and peaks at noon. However, the individual temporal patterns of interactions are more complex than simply circadian curves shaping a general function, as the frequency of interactions has a high degree of complexity, with bursty periods characterized by very frequent interactions, combined with low activity lapses, where there are few interactions. This behaviour appears in the sequence of mobile phone calls, where the circadian patterns are observed in the collective behaviour, and the individual patterns reflect a characteristic burstiness [2, 3, 4]. In fact, the times taken to an individual between sending one letter or email and sending next follow a universal behaviour, with cascades in which many events happen in a brief period, shaped by circadian and weekly cycles [5].

The analysis of interactions becomes even more complex when the space and the time are combined, leading to the study of mobility. According to the mobility, we can divide the human history in three ages. In the first, starting when the early humans appeared, the nomad hunter-gatherers included two characteristic scales in their movement patterns: daily foraging in the search for food, and a seasonal migration, which implied a longer distance. These two scales are associated with the two mobility modes introduced by Binford [6]: while

CHAPTER 1. INTRODUCTION

in the residential movements all the members of a camp move, in the logistic movements smaller groups depart from the residential location, coming back after a period that typically is not larger than several days. Secondly, starting with the Agricultural revolution, some societies became sedentary, leaving the long distance trips for merchants, like the trips associated with the trade along the Silk Road. Finally, the Industrial revolution changed totally the movement patterns, facilitating faster displacements, first within more and more populated cities, and then across a highly interconnected world, where an individual is able to travel towards his antipodal in less than a day, if the suitable connections are taken. From the Industrial revolution, new means of transport have been developed, getting faster operational velocities that facilitate the transport of people and goods (Fig. 1.1). In fact, these means of transport have changed the daily distances travelled by people and the way they do, with a decrease in walked and train daily trip distances from the second half of 20th century, while those distances have increased for cars, buses and motorcycles, experiencing a much higher growth in air transportation (Fig. 1.2). With the generalization of leisure activities, the tourism has become a remarkable driver of human mobility, combining the movement oriented to working aspects, like migration and commuting, with tourist trips. In fact, tourism is changing the use of the urban space, as cities were typically divided into business and residential areas, but currently more and more space is occupied by coherent tourist areas, which imply high surfaces whose main activity is tourism [7]. The improvement of transportation velocities has a long list of advantages, but it requires for research in order to increase its efficiency. In the other hand, faster is not always better, because as we travel faster around our planet, carrying many pathogens within our bodies, we facilitate the spreading of infections that can lead to pandemics at the planetary scale.

If we have a look into the animal world, despite the social behaviour that is displayed in several species [8], the groups are restricted to small sizes, in contrast to human beings, whose group sizes coevolved with the neocortical cortex [9]. Then, having small group sizes may be one of the reasons for the observed animal movement behaviour, that can be associated only with the first stage of the human mobility history, including some species that both forage and migrate [10], and others that do not display long distance seasonal migrations [11], showing only foraging movement patterns, driven by the search of food, reproduction mates or shelter.

This chapter, which includes the background for the forthcoming chapters, is structured in four sections: firstly, we start with a summary of the different tracking methods for human movement; then, we continue with the analysis of animal movement; afterwards, we describe the influence of mobility on dynam-

ical systems; finally, according to the state-of-the-art reported in the first three sections, we make a list of the questions that will be addressed in the rest of the thesis.

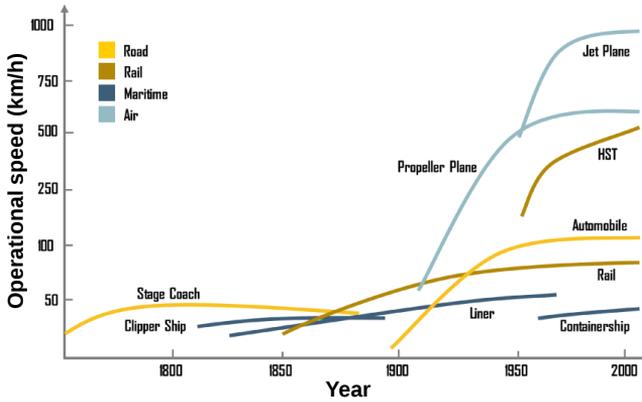


Figure 1.1: Time evolution of the operational speeds. Velocities are shown for different means of transport between years 1750 and 2010. Source: Ref. [12].

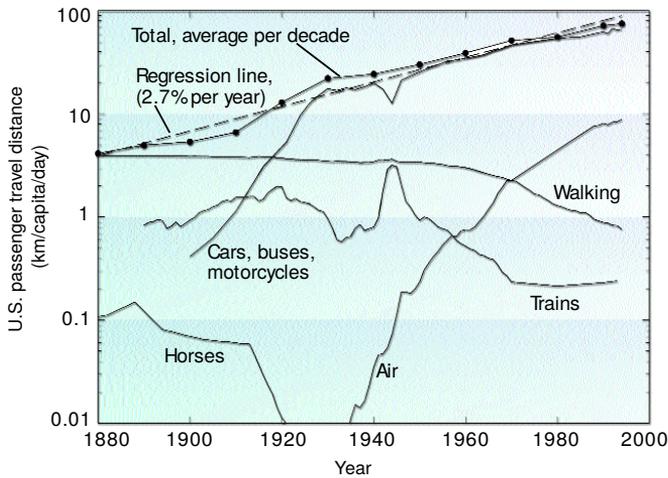


Figure 1.2: Time evolution of the average daily travelled distance per person in United States, for different means of transport. Each solid line corresponds to a different mean of transport, while the solid and dotted line represents the average and the dashed line is a linear fit. Source: Ref. [13].

1.1 Human tracking

While the early humans left just some footprints on the floor, which are difficult to survive for a long time (with few remaining samples, like the steps found in Laetoli, Tanzania [14], or those found in Goldcliff East in the Severn Estuary, United Kingdom [15]), the modern people leave continuously some clues of their presence in the places they visit. The census reflects the commuting mobility phenomena, happening when a person lives in one municipality and works in another, and transportation companies collect statistics about their services in terms of origin, destination, date and number of passengers. On the other hand, the moving humans are carrying banknotes, making or receiving mobile phone calls by connecting to the nearest antenna, taking money from ATM, or using smartphones that attach a location to our posts in social networks [16] (Table 1.1). Another approach to temporal interaction patterns in a group uses Radio Frequency Identification (RFID) devices that record who is close to whom [17], leading to empirical temporal networks known as contact networks. The Sociopatterns collaboration provides access to these networks, offering good material to assess the structure and the spreading processes in several environments, like conferences, hospitals, schools or offices [18]. For example, contact networks have been recently used to describe the social structure of hunter-gatherers in Philippines and Congo, revealing that different families were connected through few strong links, which illustrate how the information was shared [19].

Method	N	X	Reference
Banknotes	4.65×10^5	1.03×10^6	[20]
Mobile phone	10^5	1.63×10^7	[21]
Credit card transaction	5×10^{10}	NA	[22]
Geo-located posts	5.72×10^5	2.10×10^7	[23]
Public transport smart cards	2.03×10^6	1.12×10^7	[24]
Taxi GPS trajectories	10^4	2×10^{10}	[25]
Hospital contact network	75	3.24×10^4	[26]
Commuters from census	3.41×10^7	NA	[27]

Table 1.1: Comparative table of some examples of human movement analyses. The entries include the tracking unit method, the number of tracked units N , the number of analysed displacements X , and the reference. In the hospital contact network, X is the number of detected contacts.

1.1. HUMAN TRACKING

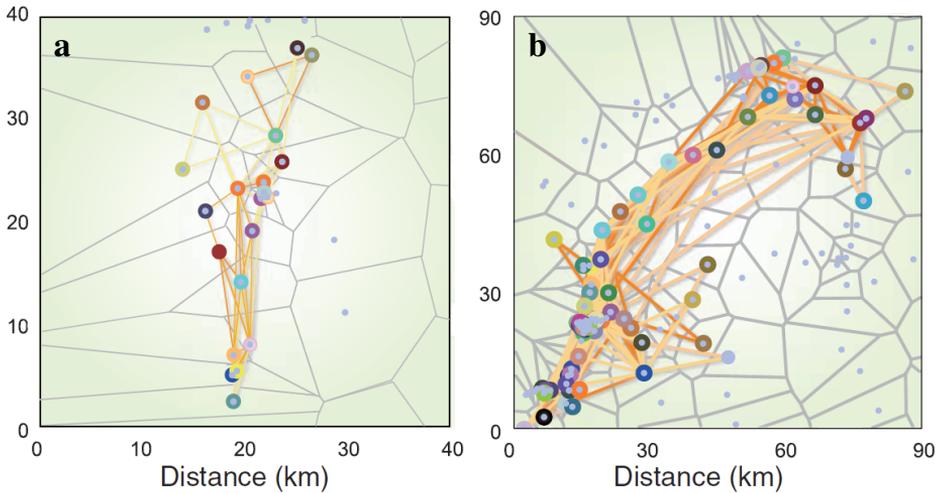


Figure 1.3: Trajectories from the mobile phone activity record. Region where users visit **a**, 22 and **b**, 78 locations. The position of an antenna (grey dots) defines a location, including in it all the points such that the antenna is the closest, leading to a Voronoi lattice whose cell limits are depicted with grey lines. Each colour corresponds to the trajectory of a different user. Source: Ref. [28].

However, there are issues for the measurement of the properties of the human mobility according to these tracking methods. For example, the mobile phone record includes the tower transmitting the information of a calling or texting event and the time associated with it. Nevertheless, the transmitting towers are not uniformly distributed in space, with strong differences between urban and rural areas, leading to different levels of accuracy on the position estimates [16] (Fig. 1.3). Apart from technical problems, human tracking has ethical issues, as the availability of a few points in a trajectory leads to a unique identification between this sample and the whole trajectory, affecting the privacy of the tracked users [29].

1.1.1 Banknotes

One of the items that humans carry in their daily lives is money. In contrast to coins, which are really difficult to track, every banknote has a serial number that identifies it. Then, its movement can be tracked if the citizens report when and where they found that specific banknote. This information is collected in the

CHAPTER 1. INTRODUCTION

web page www.wherisgeorge.com, offering a valuable dataset based not on the human displacements, but on the trips displayed by the currency that they used. When citizens find a banknote with a stamp that indicates that it participates in this tracking system, they can introduce the serial number of the bill, reporting the location where they found it.

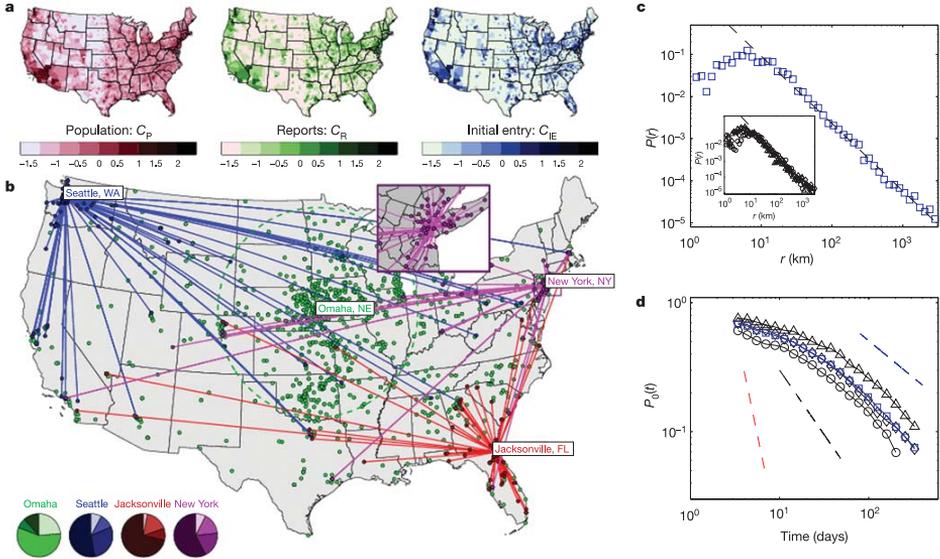


Figure 1.4: Dispersion of banknotes in the US. **a**, Comparison between population, reports and initial entry densities in the studied region. Values are represented in logarithmic scale after normalizing with the national average. **b**, Map of the displacements whose origin is located in Seattle (blue), New York (violet) and Jacksonville (red) happening with $T < 14$ days. **c**, Distribution of displacements r , aggregating the time lags with $1 < T < 4$ days, displaying a truncated power-law behaviour, $p(r) \sim r^{-\beta}$. The inset represents the distributions for initial locations in metropolitan areas (triangles), cities (diamonds) and small towns (circles). **d**, Fraction of secondary reports P_0 in locations that are at a distance smaller than 20 km from the initial location. Blue squares represent an average over 25375 initial locations, and the other symbols are associated with the same datasets as in the inset of **c**. Blue dashed line is a power-law fit $P_0 \sim t^{-\eta}$ with exponent $\eta = 0.60 \pm 0.03$, the black dashed line corresponds to 2D diffusion, and the red dashed line is the behaviour associated with a Lévy flight with spatial exponent $\beta = 1.59$. Source: Ref. [20].

1.1. HUMAN TRACKING

A dataset gathering all these reports in the US was analysed in Ref. [20], considering the distance r and the time T between two successive reports. The short-time trajectories ($T < 14$ days) whose origins were Seattle, Jacksonville and New York included a majority of short-ranged trips and a small number of long trips (Fig. 1.4b). The distribution of displacements, including all reports in the whole country except from those involving Hawaii and Alaska, for time lags $1 < T < 4$ days displayed a power-law behaviour between the short-range limit $L_{\min} = 10$ km and the typical distance between West and East coasts, $L_{\max} = 3200$ km. This power-law $p(r) \sim r^{-\beta}$ had an exponent $\beta = 1.59 \pm 0.02$ (Fig. 1.4c). This long tail appeared independently of the size of the initial location of the trajectories, as trajectories starting in metropolitan areas, cities and small towns showed the same behaviour for $r > L_{\min}$ (Fig. 1.4c inset). The power-law shown in the displacements distribution is compatible with the properties of the Lévy flight model.

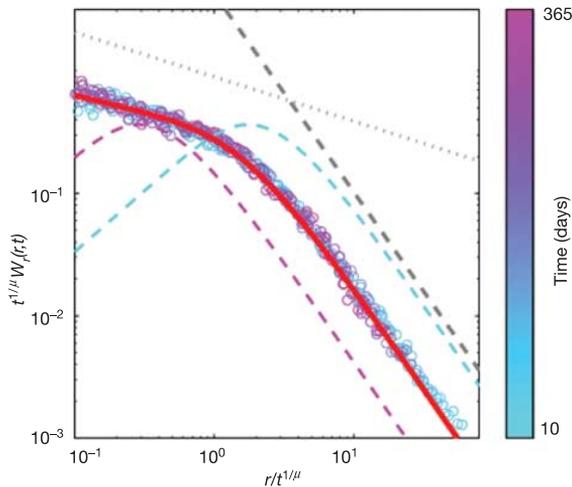


Figure 1.5: Rescaling of the movements data corresponding to banknotes. The displacements r are normalized dividing by a power of time $t^{1/\mu}$, with $\mu = \delta/\alpha$. Circles represent the probability density $W_r(r, t)$ of observing a bill at a distance r from origin at time t , with each colour associated with a different value of t , in the range 10-365 days. Dotted and dashed grey lines depict, respectively, the asymptotic behaviour for short and long distances. The behaviour associated with a pure Lévy flight with an spatial exponent $\beta = 1.60$ is shown in the blue ($t=10$ days) and magenta ($t=365$ days) dashed lines. Source: Ref. [20].

CHAPTER 1. INTRODUCTION

However, the first passage times were not those that could be expected in a movement following a Lévy flight. The time evolution of the fraction $P_0(t)$ of bills remaining in the initial region, which is the complementary cumulative of the first passage time distribution, had a tail that fitted to a power-law with exponent $\eta = 0.60$, which was different from the expected exponent for 2D ordinary diffusion ($\eta_{\text{dif}} = 1.0$), and different from a Lévy flight with spatial exponent $\beta = 1.59$ ($\eta_{\text{LF}} = \frac{2}{\beta-1} \approx 3.4$) (Fig. 1.4d). This behaviour can be modelled as a Continuous Time Random Walk (CTRW), where the walkers make jumps of distance x after a waiting time τ , with $p(x) \sim x^{-(1+\delta)}$ and $p(\tau) \sim \tau^{-(1+\alpha)}$, with $\delta = \beta - 1$ and $\alpha = \eta$. The equation describing the probability of being at time t at a distance r from the origin is given by

$$W_r(r, t) = t^{-\alpha/\delta} L_{\alpha, \delta}(r/t^{\alpha/\delta}) \quad (1.1)$$

According to Eq. (1.1), the distribution functions describing the displacements for different values of the time between reports will collapse in a single function $L_{\alpha, \delta}$ when the distance is normalized dividing by $t^{\alpha/\delta}$ and the distribution is multiplied by this quantity, such that it remains being a probability density function. For an interval of time between 10 and 365 days, the universal behaviour was confirmed as the curves collapse under this rescaling (Fig. 1.5).

1.1.2 Mobile phone users

The analysis of banknotes dispersion is a proxy for the study of human mobility [20]. However, the measurements around this dispersion process reflect more the collectively human-induced movement patterns than the individual ones, as the trajectory described by one banknote is associated with the movement of the different individuals that carry it for some time. For example, two consecutive reports may involve the trips performed by several individuals. In order to investigate the individual mechanisms leading to the observed collective pattern, other methods for individual tracking are needed. This issue was addressed in Ref. [21] using data based on cell phone activity, like the emission of reception of a call or a text message.

In contrast to money, mobile phones are devices that are carried individually by their users, such that the location of these devices can be used for determining their owner's locations. The study in Ref. [21] included the analysis of two datasets. First of all, the dataset D1 included the information about the location of the users when their phones were active, whether emitting or receiving a call or a text message, assigning to each user the location of the tower responsible of transmitting the information. Typically, the temporal pattern of calling activity

1.1. HUMAN TRACKING

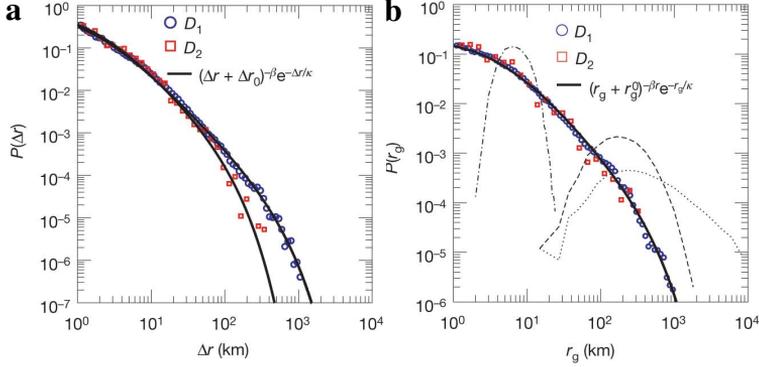


Figure 1.6: Human mobility patterns according to mobile phone data. **a**, Distribution of displacements Δr between two consecutive points. Blue circles are associated with the bursty activity pattern in dataset D1 and red squares represent the displacements observed in the regularly sampled trajectories in dataset D2. The solid lines are truncated power-law fits (Eq. 1.2) to each curve. **b**, Gyration radius r_g distribution. The solid lines indicate truncated power-law fits, while the dotted, dashed and dot-dashed curves are those associated with a random walk, a Lévy flight and a truncated Lévy flight, respectively. Source: Ref. [21].

is bursty, such that users have periods with many calls, one after another, and then a long period without any activity [3]. In order to check that the results coming from this temporal pattern were not affected by this burstiness, the second dataset (D2) was comprised of a regular sampling of the user location, reported every two hours for a week. The distribution of distances Δr between consecutive points was described by a truncated power-law:

$$P(\Delta r) = (\Delta r + \Delta r_0)^{-\beta} \exp(-\Delta r/\kappa) \quad (1.2)$$

where the fitting exponent was $\beta = 1.75 \pm 0.15$ (standard deviation) for both datasets, $\Delta r_0 = 1.5$ km, and the cut-off distances were $\kappa_{D1} = 400$ km and $\kappa_{D2} = 80$ km (Fig. 1.6). This exponent was similar to 1.59, which was obtained for the trajectories described by banknotes [20], and suggested the possibility of similar mechanisms underlying both processes. Note that $\kappa_{D1} > \kappa_{D2}$ due to the presence of long periods without activity in the bursty calling pattern in D1, which were longer than the uniform sampling pattern every 2 h in D2, allowing users to move further.

Three hypothesis were considered for explaining the mechanisms leading to the behaviour observed in Fig. 1.6a, which represented the aggregation of multiple

CHAPTER 1. INTRODUCTION

individual trajectories. First of all, every individual trajectory was described by Eq. (1.2), such that the aggregated distribution shows the same behaviour as the individual. Secondly, the aggregated distribution was a consequence of some idiosyncrasy, such that the shape of Fig. 1.6a emerged from analysing the aggregated behaviour of a heterogeneous population. Finally, the observed distribution was the aggregation of idiosyncratic Lévy flights, that would be the convolution of first and second hypothesis. In order to test the variability among individuals, it is useful to study the gyration radius r_g , which is an indicator of the spatial spread of the trajectory. First of all, the center \mathbf{c} was computed as the average position along the the trajectory:

$$\mathbf{c} = \frac{\sum_{i=1}^A \mathbf{x}_i}{A} \quad (1.3)$$

where i runs over the A locations \mathbf{x}_i describing the trajectory. Secondly, the gyration radius was the root mean square distance to the center of mass of the trajectory:

$$r_g = \sqrt{\frac{\sum_{i=1}^A d^2(\mathbf{x}_i, \mathbf{c})}{A}} \quad (1.4)$$

where $d(\mathbf{x}, \mathbf{y})$ was the geodesic distance between locations \mathbf{x} and \mathbf{y} . The distribution of the gyration radii fitted to another truncated power-law (Fig. 1.6b):

$$P(r_g) = (r_g + r_g^0)^{-\beta r} \exp(-r_g/\kappa) \quad (1.5)$$

with $r_g^0 = 5.8$ km, $\beta_r = 1.65 \pm 0.15$ and $\kappa = 350$ km. Although the gyration radius distributions arising from random walks (RW), Lévy flights (LF) and truncated Lévy flights (TLF) could reach some levels of heterogeneity, they could not explain the observed shape (Eq. (1.5), Fig. 1.6b). This suggested that the observed behaviour in Fig. 1.6a emerged from the heterogeneity among individuals.

The time evolution of the average gyration radius confirmed that the individual trajectories are not described by RW, LF or TLF. For these movement patterns, the average gyration radius scales as a power-law of time, $r_g(t) \sim t^{1/2}$ in a RW and $r_g(t) \sim t^{3/(2+\beta)}$ for LF and TLF. However, the measured average gyration radius grew logarithmically with time (Fig. 1.7a). Splitting the users according to their gyration radius, the users with larger r_g were able to arrive further (Fig. 1.7b inset), and all the displacements distributions collapsed into a universal curve when rescaling the displacements with the gyration radius: $P(\Delta r|r_g) \sim r_g^{-\sigma} F(\Delta r/r_g)$, with $\sigma = 1.2 \pm 0.1$ and $F(x) \sim x^{-\sigma}$ for $x < 1$

1.1. HUMAN TRACKING

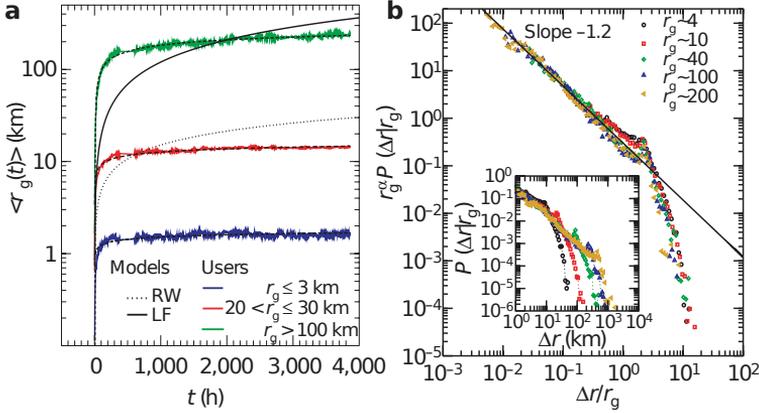


Figure 1.7: Idiosyncrasy and universality in human mobility patterns. **a**, Time evolution of the average gyration radius $\langle r_g(t) \rangle$. Curves are split in users for which $r_g(t = 6 \text{ months}) \leq 3 \text{ km}$ (blue), $20 \leq r_g(t = 6 \text{ months}) \leq 30 \text{ km}$ (red) and $r_g(t = 6 \text{ months}) > 100 \text{ km}$ (green), the dotted line is associated with RW, and the solid line with LF and TLF. Dashed curves are logarithmic fits to the data. **b**, Conditional distributions of displacements for users with different values of $r_g(t = 6 \text{ months})$. The curves collapse to a universal shape when the displacements Δr are normalized by the gyration radius. The solid line represents a power-law with exponent 1.2. Source: Ref. [21].

and decreasing sharply for $x > 1$ (Fig. 1.7b). Then, the individuals exhibited Truncated Lévy flights characterized by power-law distributions until they reached a characteristic distance related to their gyration radius, and further displacements were negligible. Finally, the relationship between the different introduced scaling laws was established through the defined exponents using $P(\Delta r) = \int_0^\infty P(\Delta r|r_g)P(r_g)dr_g$, leading to $\beta = \beta_r + \sigma - 1$, which agreed with the measured exponents. In conclusion, the collective pattern observed in Fig. 1.6a was the consequence of individual truncated Lévy flights and the heterogeneity in the characteristic travelled distance for each individual.

However, there are three human mobility features that the fitted Lévy flights [21] and the CTRW model [20] cannot explain, as highlighted by Song *et al.* in Ref. [30]:

- The number of visited locations S grew as a power of time, $S \sim t^\mu$, with $\mu = 1$ for Lévy flights and $\mu = \eta$ in a CTRW with a waiting time distribution

CHAPTER 1. INTRODUCTION

$p(t) \sim t^{-\eta+1}$. However, the reported exponent in Ref. [30] was $\mu = 0.6 \pm 0.02$, smaller than the expected for a CTRW, $\eta = 0.8 \pm 0.1$.

- The visitation frequencies to each location were distributed according to a power-law distribution, while in Lévy flights and CTRW they were expected to be uniform.
- The evolution of the average square displacements with time was slower than a power, argued to be a consequence of the tendency to return home.

Alternatively to the Lévy flight and CTRW models, Song *et al.* proposed a model that considered a power-law distribution of waiting times, but introducing one choice between two actions: a) exploration, in which the probability of moving towards a newly visited location decayed with time, or b) preferential return to the previously visited locations. Although this model did not capture the properties of the short-term behaviour, it was able to reproduce the long-term spatial and temporal scaling [30].

1.1.3 Mobility networks

Complex networks are an extremely useful tool to analyse dyadic interactions, which are interactions between two items. The complex networks formalism allows to study a broad diversity of systems, like those composed of human beings [31, 32], species [33], proteins [34], genes [35], airports [36], or web pages [37]. These interactions are described by a graph $G = (V, E)$, that includes the links E between the nodes V that compose the network. The simplest representation of the network is the adjacency matrix, whose entry A_{ij} has a non-zero value if the nodes i and j interact, and it is null otherwise. The adjacency matrices can include information about the strength of the interactions, which is called the weight. Unweighted networks have just binary values in their adjacency matrices. Moreover, the networks can reflect directionality for non-symmetric interactions. For example, in the network associated with Twitter following activities, one user can be follower of another user which does not necessarily follow her back. Another example occurs in neuronal networks, as the excitatory and inhibitory interactions are not reciprocal. These examples are directed networks, such that $A_{ij} \neq A_{ji}$. Interestingly, network theory has developed a formalism to study systems that include different kind of interactions; in those systems, represented in multilayer networks, each kind of interaction is associated with a layer, with some coupling between the layers [38]. For example, a multilayer network can contain three layers that represent the interactions with family, friends or colleagues. In fact, this approach is valuable to estimate the robustness of a system

1.1. HUMAN TRACKING

composed of several interconnected layers, like the coupling between the power grid network and the Internet communication network, which were involved in the electrical blackout affecting Italy in 2003 [39].

Weighted and directed networks represent a useful tool for studying movement. Considering each location as a node, with a suitable spatial discretization (far from the extremes in which whether one location includes all the visits, or all locations are visited once), the entry A_{ij} in the adjacency matrix will be the number of agents moving from i to j . In this case, the adjacency matrix is known as origin-destination matrix. Using an origin-destination matrix is a stationary approach, as it typically assumes that the fluxes in the system are time independent, focusing on spatial rather than on temporal patterns. The description and estimation of these fluxes are fundamental, and have a broad range of applications, like addressing more efficiently the public transportation demand or preventing epidemic spreading outbreaks that may be facilitated by the structure of these networks.

Fluxes of mobility have been used to describe the movement of people between the municipality in which they live and the one in which they work, *i.e.*, the commuting flux. There are two main theories that try to explain the difference on the commuting flux between different spatial locations. Firstly, the gravity model [40] states that the flux T_{ij} travelling between two locations i and j will be proportional to a power of their populations P_i and P_j and inversely proportional to a function of the distance d_{ij} between them:

$$T_{ij} \propto \frac{P_i^\nu P_j^\phi}{f(d_{ij})} \quad (1.6)$$

However, the disagreement in some orders of magnitude of the gravity model with empirical data (see for example the prediction of gravity law for the number of travellers between two counties in Utah in Fig. 1.8a) motivated the formulation of the radiation model [27]. This model, which was applied to calculate the fluxes of commuters between different counties in the US, considers that a potential commuter will choose her job in the closest location that provides her with a better benefit than that associated with the current location. In each county, there is a population n , and there is one job opening for n_{jobs} individuals, *i.e.*, all the counties have the same number of job openings per individual, which is $1/n_{\text{jobs}}$. Then, each job opening is assigned a number z , sampled from an externally given distribution $p(z)$, and the characteristic benefit for each county will be the highest among the benefits from the job openings at it (Fig. 1.8b). With this information, the commuter will decide to move to the closest county that gives a better benefit than that associated with her own county (Fig. 1.8c).

CHAPTER 1. INTRODUCTION

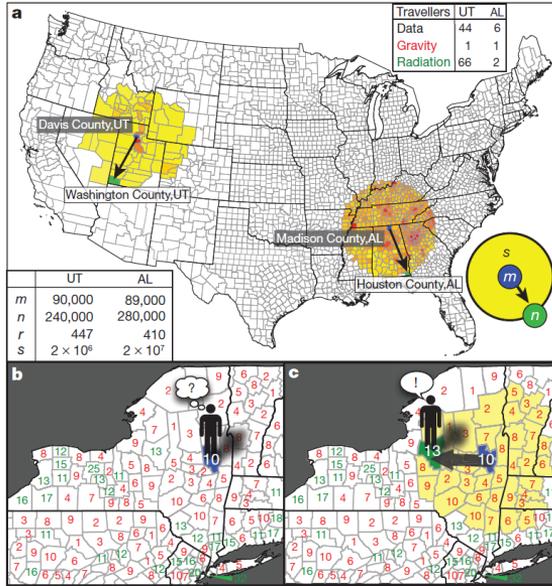


Figure 1.8: Examples of the radiation model. **a**, For a pair of counties, estimation of the amount of commuters predicted by gravity and radiation models, after fitting the models to the census data for year 2000. **b**, In the first step of the radiation dynamics, each county is assigned the highest benefit of its job openings, with the benefits sampled from an external distribution. **c**, Afterwards, the potential commuter chooses the closest county among those that have a higher benefit (with numbers written in green, in contrast to red numbers for lower benefits) than the benefit in the own county. Source: Ref. [27].

This model leads to a prediction for the average flux $\langle T_{ij} \rangle$ between counties i and j that is independent of the shape of $p(z)$ and n_{jobs} :

$$\langle T_{ij} \rangle = T_i \frac{m_i n_j}{(m_i + s_{ij})(m_i + n_j + s_{ij})} \quad (1.7)$$

where m_i and n_j are respectively the populations of the origin i and the destination j , s_{ij} is the population within the circle centred at i and with j in its border, and T_i is the number of commuters whose origin is i , calculated as the product of the total number of commuters in the system and the fraction of the total population living in i .

When the radiation model was compared with empirical data, in contrast to gravity model, its predictions showed more accurate results focusing, for example,

1.1. HUMAN TRACKING

on the commuting patterns for individuals departing from New York County (Fig. 1.9a). In fact, the gravity model considerably underestimated the long

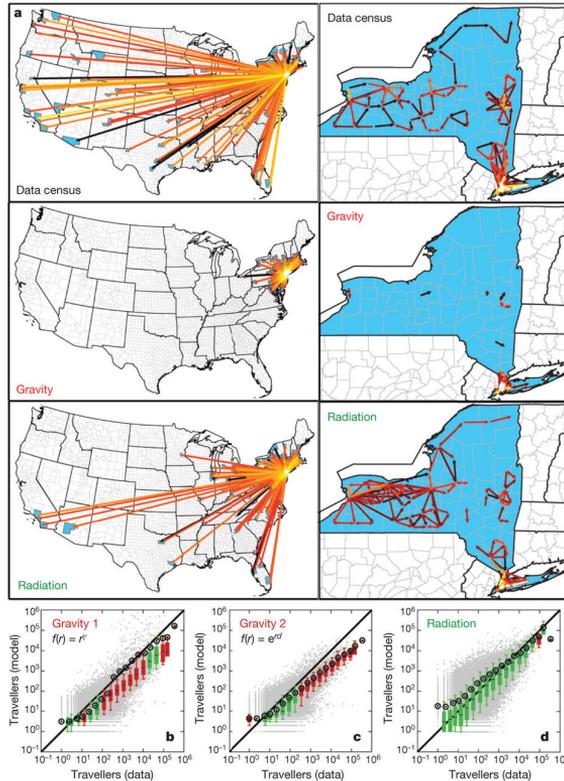


Figure 1.9: Comparison between the radiation and the gravity models. **a**, For the New York County, number of travellers for census data, and predictions arising from gravity law and radiation models. The colour of the link represents the intensity of the mobility flux, being black for 10 commuters and white for amounts higher than 10,000. The zoomed area represents includes the pattern within the state of New York. **b**, **c**, Predictions for gravity law for all the pairs of counties in the US with non-zero commuting flux, for two different spatial functions, compared with the empirical data. **d**, Prediction for radiation model. The grey points represent the scatter of all the possible pairs, with the diagonal solid lines showing the maximum accuracy. Black points represent the average, while the boxes are depicted in green (red) if the diagonal line falls in the interval between the 9th and 91st percentile. Source: Ref. [27].

CHAPTER 1. INTRODUCTION

distance commuting, with the longest connection being smaller than 400 km in this case. This particular example was extended to all the pairs of counties in the US with non-zero commuting flux, leading to a general underestimation (Figs. 1.9b,c), while the radiation model described more realistically the empirical data (Fig. 1.9d). However, another work showed that the gravity model described by Eq. (1.6) was making more accurate predictions than the radiation model in a dataset describing the commuting fluxes between the cities of England and Wales [41].

The radiation model represents the opportunity to find universal patterns in the processes that it describes, as a rescaling of the population in Eq. 1.7, making $m_i \rightarrow \lambda m_i$, $n_j \rightarrow \lambda n_j$, and $s_{ij} \rightarrow \lambda s_{ij}$, leaves the probability to go from i to j (*i.e.* T_{ij}/T_i) invariant. When the probability of, given an origin with population m , choosing a destination further than a population s , $p_s(> s|m)$ was plotted against s , there was a broad range of curves, each one associated with a different origin (Fig. 1.10a). However, these curves converged when they were plotted when s was normalized dividing by m , as the radiation model predicts $p_s(> s|m) = \frac{m}{m+s}$ (Fig. 1.10b), although they were still far from the curve given by $f(x) = \frac{1}{1+x}$. This issue was solved introducing an advantage to local people to find more job opportunities, increasing m to $m + \epsilon$, which for $\epsilon = 35,000$ lead to smaller discrepancies (Fig. 1.10c).

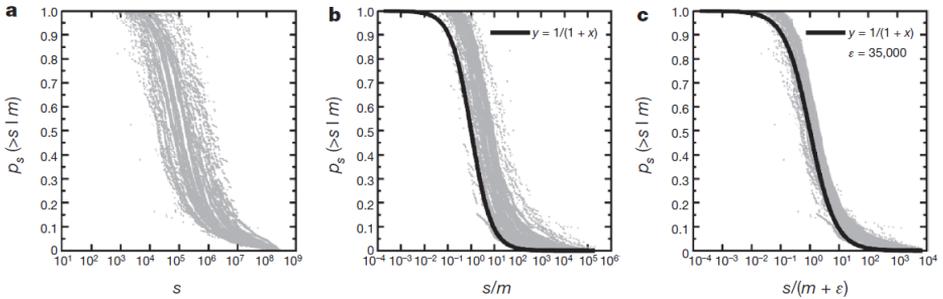


Figure 1.10: Universality of commuting patterns. **a**, Probability $p_s(> s|m)$ of, from an origin with population m , a commuter chooses a destination which is at a distance higher than the circle centred at the origin with a population s . **b**, When s is normalized by the origin population, the broadness of the empirical curves is reduced, but still they do not follow the function $f(x) = 1/(1+x)$. **c**, These curves collapse to the function $f(x)$ when a factor ϵ describing the advantage of local inhabitants to find jobs is added to the population m . Source: Ref. [27].

Other studies of mobility networks used data from geo-located posts in Twitter, in order to find the clusters describing a high internal level of flux. In fact,

1.2. ANIMAL MOVEMENT

network theory has developed some algorithms that detect community structure, searching for groups of nodes that are strongly connected between them in comparison with the connections that they have with the rest of the network [42], which are useful for describing the strongly connected locations through mobility. For example, Hawelka *et al.* grouped the countries in a global scale according to the flux that described the mobility patterns measured from geo-located posts [43].

These patterns are useful for describing the spatial features of social ties, which may or not be similar to those arising from mobility. Specifically, using the location of the the couples of users that shared a phone call, the fluxes of social interactions lead to a network in which a community search algorithm was used to draw the map of Great Britain according to this activity [44]. In fact, one model coupling the social ties and the mobility was able to reproduce several social and spatial features at a country level [45]. Another work compared the social and mobility spatial patterns, showing that, although the distributions of distances showed different behaviours, the distributions of fluxes between locations, given the distance between them, collapsed to a universal function for all the distances in both the social and mobility processes, after normalization dividing by the average flux [46].

Finally, community search methods have been applied for detecting spatial locations with similar temporal patterns. Network analysis for understanding human use of the space considers the temporal activity profile observed in different locations, connecting those that are highly correlated. These networks are called functional networks, connecting nodes that have similar activity patterns, in contrast to the physical networks, which represent those nodes that are connected physically. In fact, these two kind of networks, applied to urban studies, represent for example, the different uses of the space (functional network), or the places that are directly connected by a means of transport (physical network). This idea was recently applied to data reporting the transmitting tower for each mobile phone call emitted or received by a user, considering five Spanish cities [47]. A clustering method for the strongly connected locations in the functional network partitioned the system into zones whose temporal patterns were associated with leisure, business or residential activities.

1.2 Animal movement

The empirical measurement of animal trajectories addresses many challenges for the movement ecology community. Apart from technical problems related

CHAPTER 1. INTRODUCTION

with tagging processes and tracking devices, the sampling process is a remarkable issue, as the real trajectories are approximated by straight lines connecting the sampled locations. Following these concerns, Turchin defined a path as the spatio-temporal record of a tracked organism [48]. The paths have been described following two main approaches: for many species, the set describing the spatial coordinates when behavioural events occur, like a butterfly landing [49] or a wandering albatross in contact with water [50]; in contrast, other species display continuous trajectories exhibiting no clear break points, with their paths being described by the position of the tracked individual at regularly spaced times. Considering that several terminologies have been used in the literature, we introduce two key terms that will be used along this thesis. First of all, a displacement includes two consecutive points in a path composed of the positions at regularly spaced times. Secondly, a step includes two consecutive behavioural events. Both the displacement and the step lengths will be the distances between the two points that they include.

1.2.1 Lévy or not?

There are two traditional theoretical approaches to the study of animal movement: Brownian motion, characterised by a jump length distribution with well-defined mean and variance, and Lévy-like models, for particles displaying many short jumps, but a non-negligible probability for long jumps, leading to a jump length distribution whose second (and in some cases first) moment diverges, and distinguishing between the Lévy flight and the Lévy walk. In the Lévy flights, the moving particles display jumps between two consecutive position updates ($\Delta t = 1$), of a distance d sampled from a distribution $p(d) \sim d^{-\beta}$, with $1 < \beta \leq 3$ and $d > d_{\min}$, in a direction that is randomly chosen at each step from a uniform distribution; this distribution of jumps $p(d)$ is a heavy-tailed distribution and implies that the particles can travel with no limit velocity. In order to solve this issue, in the Lévy walk model the particles travel with a finite and constant velocity v (*i.e.*, $\Delta t = d/v$), from one location to another, again with a power-law distribution of distances, and when they arrive to the destination their direction is updated. Movement patterns described by the Lévy-like models have been reported to be optimal in the random search processes of sparsely and randomly distributed targets, with an optimal exponent $\beta_{\text{opt}} = 2$ [51, 52]. While the number of visited sites, defined from a discretization of the space, after n movement steps in a Lévy walk scales as n for $\beta < \text{dim} + 1$, with dim being the dimension of the space, for a random walk they scale similarly for $\text{dim} < 2$, but as $n^{\text{dim}/2}$ for $\text{dim} \geq 2$ [53]. For testing the optimality hypothesis, the number of found targets after a long time was studied for different movement patterns [53], reporting that a

1.2. ANIMAL MOVEMENT

Lévy-like movement pattern with $\beta \approx 2$ was optimal in the case of revisitable targets (*i.e.*, resources that appear in the same location after being detected), while in the case of non-revisitable targets (*i.e.*, resources that after being detected appear in a different location) the Lévy walks performed better than Brownian motion, but the optimal movement pattern was ballistic motion in a random direction [54]. In fact, other models, whose trajectories in some cases look quite similar to those arising from Lévy walks, have also been proposed as optimal strategies for resource search, like the intermittent motion, which combines a fast phase, with high velocity and in straight direction, in which there is no detection, with a search phase, characterized by smaller velocities and more frequent turns, when targets can be detected [53].

Optimal foraging theory states that if a movement strategy is optimal in search processes, it increases the fitness of the individuals displaying it, and hence it is selected by evolution [55]. Then, it is important to assess in which cases the Lévy walks or the Brownian motion are more efficient. In fact, the efficiency, measured in terms of encounter rates, does not only depend on the motion of the searcher, but also on that of the target. Specifically, Lévy walks are more efficient for finding targets that perform Brownian motion, and this efficiency increases for lower target densities, lower target velocities and lower target sizes (these two last features in comparison with the searcher) [56]. However, if the targets display Lévy walks, the advantages of searchers moving on a Lévy walk strategy are smaller, in some cases getting better performances with Brownian motion.

Then, according to optimal foraging theory, Lévy walks will be selected only when they represent the most efficient search strategy. This led to some controversy in the movement ecology community about the presence or not of Lévy-like movement patterns in empirical observations. Initially, these movement patterns were reported for wandering albatrosses [50], deer and bumblebees [51]. However, other studies have reported discrepancies with it [57]. In fact, an intermittent behaviour was observed in marine animals, with a clear switching between Brownian motion and Lévy patterns, depending on the habitat in which the animals were moving [58].

1.2.2 Data analysis of marine animals movement

In contrast to human mobility, the data collection for animal movement is expensive, in terms of both devices and human resources. This has been challenging the movement ecologists in the last decades, but the development of cheaper, more accurate and lighter position transmitter devices has facilitated the growth

CHAPTER 1. INTRODUCTION

of the available data for analysis, which currently is arriving to Big Data status, with some programs that promote the data collection, like Tagging of Pelagic Predators (TOPP) [59].

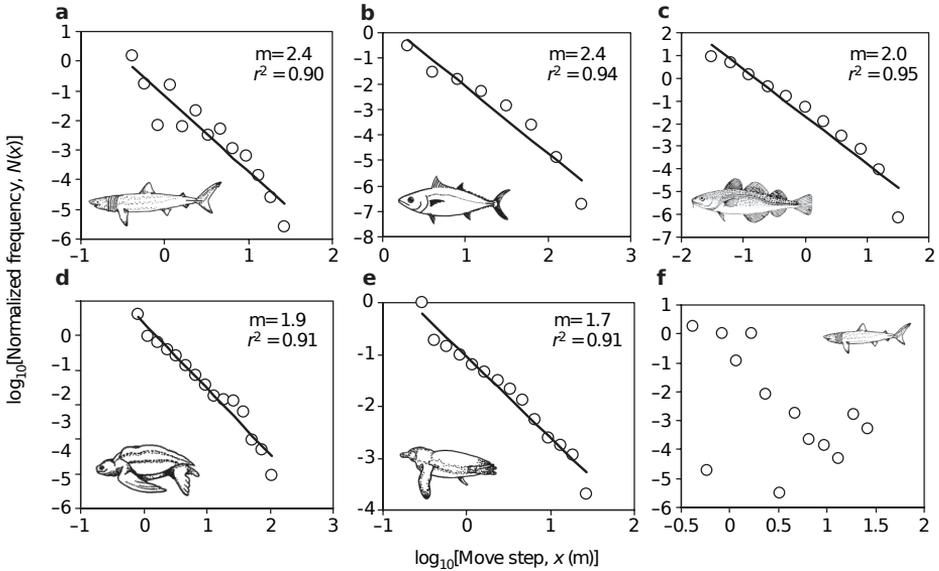


Figure 1.11: Vertical movement patterns of marine predators. Distributions of step length s for **a**, basking shark, **b**, bigeye tuna, **c**, Atlantic cod, **d**, leatherback turtle, **e**, Magellanic penguin, **f**, a juvenile basking shark. In the cases **a-e**, the distributions fit to a power-law $p(s) \sim s^{-\mu}$, while the juvenile basking shark displays a different behaviour. Source: Ref. [60].

The difficulties in the data collection processes are reflected in the analysis, typically addressed at individual scale or pooling the data arising from very few individuals. This is the case of a highlighted study on the movement patterns of marine predators, which analysed 31 individual displacement sequences corresponding to seven species [60]. In this article, the steps s in the vertical direction were analysed. The one dimensional data facilitates the measurement of steps, which are defined as the distances between two direction changes, *i.e.* the distances between local minima and maxima, and vice versa. The step lengths were distributed as power-law functions, $p(s) \sim s^{-\mu}$ for 24 individuals belonging to 5 different species, typically associated with Lévy-like movement patterns and characterized by the exponent μ (Fig. 1.11). These patterns, exhibited by marine predators, may be associated with optimal search strategies [61], emerging as

1.2. ANIMAL MOVEMENT

the animals have a deeper experience, in contrast to the pattern displayed by the juvenile basking shark (Fig. 1.11f).

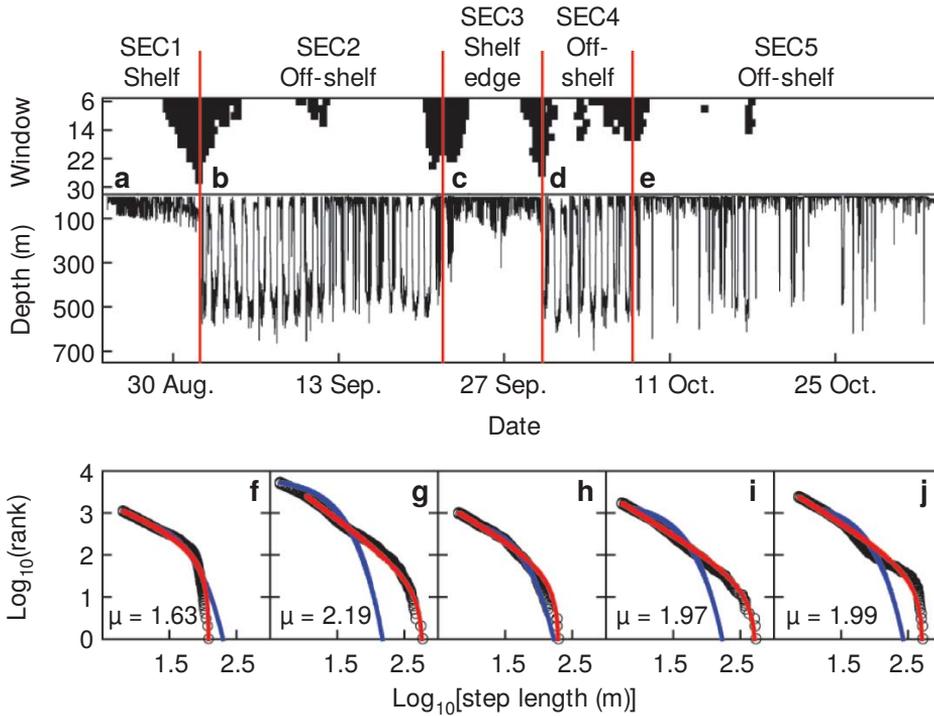


Figure 1.12: Switching movement patterns displayed by a blue shark. a-e, Time series of the observed depth and the window that facilitates splitting the trajectory in different behavioural sections. **f-j,** Step length distribution associated with the sections depicted in **a-e**. Red and blue lines indicate, respectively, truncated power-law and exponential fits. Source: Ref. [58].

However, power-law distributions do not explain all the observed regimes in the movement patterns displayed by marine animals. Specifically, Humphries *et al.* showed that Brownian movement, characterized by exponential distributions of step lengths, was displayed by marine predators when they were moving in regions with abundant prey, while the movement was described by a Lévy-like pattern in less productive waters, where resources were scarce [58]. In this study, they analysed the movement in the depth coordinate (vertical) and the influence of the properties of the location according to the other coordinates. *i.e.* the resource abundance in the ocean for a specific latitude and longitude. Each individual trajectory was partitioned in different movement regimes called

CHAPTER 1. INTRODUCTION

sections, and the steps associated with each section were analysed separately, as shown for the example of a blue shark (Fig. 1.12a-e). Specifically, the exponential model was able to approximate the tail of the distribution on the first section, although a truncated power-law was fitting all the range (Fig. 1.12f). In off-shelf environments, the distributions were truncated power-laws in off-shelf environments (Fig. 1.12g, i, j). However, at the shelf edge, the best fit was an exponential distribution (Fig. 1.12h). This observed behaviour was related with the abundance of resources in coastal habitats while they are scarce and randomly distributed in open ocean, highlighting that the marine predators display switching movement patterns according to the environment they are moving through.

When the measurement of the position is not performed at the detection of behavioural events, like the landing on the movement of a butterfly, these events need to be detected along the paths. Hence, for trajectories that are sampled continuously, using the step length method requires detecting the turning points, which are defined as a sudden change in the movement direction. In 1D, this definition is very clear, as the turning points are those locations where the velocity changes its sign. However, the trajectories have curvature in higher dimensions, such that it is harder to define these turning points. The most recent and accepted proposal suggests projecting the movement into lower dimensions, reaching at the end 1D movement and measuring the step lengths in the one dimensional projection [62]; in the context of Lévy-like movement patterns, this method is valid and it simplifies the problem: considering a 3D movement pattern, described by steps whose lengths are independently sampled from a power-law distribution, its projection into a lower dimension and the identification of turning points lead to a distribution of step lengths that keeps a power-law tail with the same exponent, although out of the range where the original power-law was defined the distribution may be different (Fig. 1.13). However, for other kind of movement patterns, the distribution may not keep the statistical properties after the projection into smaller dimensions. This problem highlights the need for new methods that overtake these difficulties.

1.2. ANIMAL MOVEMENT

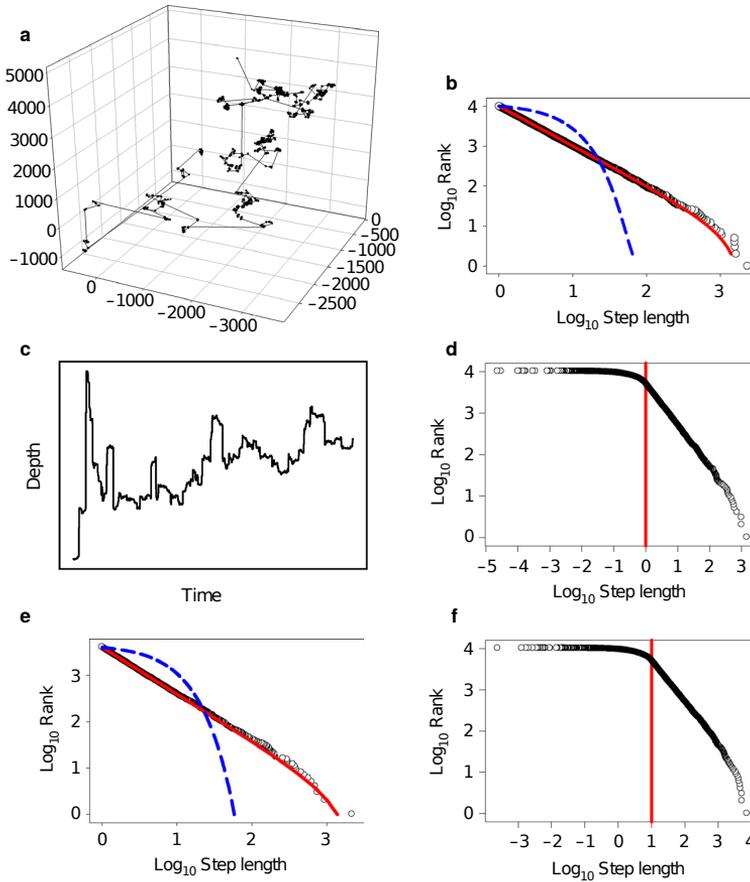


Figure 1.13: Example of the projection method applied to a 3D Lévy flight. **a**, 3D Lévy flight with step lengths x generated from the distribution $p(x) = x^{-\mu}$, with $\mu = 2$ and $x > x_{\min} = 1$. **b**, Rank as a function of the step length, fitting to a truncated power-law with exponent $\hat{\mu} = 1.996$. **c**, Time series of the one dimensional projection. **d**, Rank as a function of the step length for the one dimensional projection. **e**, Rank as a function of the step length for the one dimensional projection and $x > x_{\min}$, fitting to a truncated power-law with exponent $\hat{\mu} = 2.01$. **f**, Rank plot for the one dimensional projection of a Lévy flight generated with $\mu = 2$ and $x_{\min} = 10$. Solid red lines in **b**, **e** represent truncated power-law fits, while the dashed blue lines are exponential fits. Source: Ref. [62].

1.3 Dynamical processes in systems of mobile particles

Modelling the behaviour displayed by discrete systems of particles implies both considering a dynamical process and the topology specifying who interacts with whom. The proximity between particles facilitates their interactions in many real systems, which are described by spatial networks where each particle is assigned a position in space and the presence and strength of the interactions with other particles depend on the distance between them [63], *i.e.*, the element A_{ij} in the adjacency matrix that represents the network is a function of the distance d_{ij} between particles i and j , $A_{ij} = f(d_{ij})$. Spatial networks have been proposed to model processes such as the dispersal of pathogens on plant fields [64], biodiversity dynamics in a system of interconnected habitats [65], or synchronization in wireless networks [66].

There are two main approaches for addressing the study of dynamical processes in systems of mobile particles, related with the Eulerian description, focusing on points of the space, and the Lagrangian description, which analyses the trajectories. Many studies have described metapopulation approaches [67], where the system is considered as a set of patches of particles, with those patches distributed along the space. In such approach, the adjacency matrix describes the flows of particles between different patches. On the other hand, we can consider an agent-based mobility approach, following the specific trajectory of each particle and considering that the distance to the other particles changes with time, such that the interactions are described by a temporal network that joins a set of different spatial networks.

1.3.1 Epidemic spreading

One of the main mechanisms for biological contagion is airborne transmission, which requires physical proximity between a host and a healthy individual. This constraint in the contagion process suggests considering interactions that take into account the distance between the particles that compose the system. In fact, the interaction pattern of each individual changes in time due to its mobility, which can lead to a faster dispersal of the pathogens [68]. As introduced above, there are two main approaches to analyse the mobility influence on dynamical systems. In the metapopulation approaches, the spatial network, which describes the jumps between different patches, is constant and the populations of each patch change in time; within the patches, individuals are assumed to

1.3. DYNAMICAL PROCESSES IN SYSTEMS OF MOBILE PARTICLES

be well-mixed, such that there is an all-to-all interaction. In contrast, in the agent-based mobility approaches, the individuals follow different trajectories, leading to a spatial network that changes in time, and they only interact with those located in their proximity.

Early studies on disease spreading used percolation processes to analyse differences between massive and strategic vaccination campaigns on mobile populations [69]. In this work, Eubank *et al.* suggested the simulation of mobility patterns, from a synthetic population compatible with the census data, estimating its mobility from the land-use measurements. These simulated trajectories were analysed as a temporal network that connected the individuals to the locations that they visited at a given time. Specifically, that network had a bipartite structure, as each link connected two nodes belonging to different classes, that is, each link connected one individual with one location. In fact, the contact network was built from the projection of the bipartite network into the individual's space, connecting two individuals that were visiting the same location at the same time. However, the analysis performed in Ref. [69] did not take into account the temporal nature of this projected network, focusing on the properties of its static projection, which aggregated all the links and removed those that implied a total contact duration smaller than a threshold, and performing percolation analyses on it. Interestingly, both a vaccination campaign with priority for the nodes with highest degrees and a massive vaccination campaign, *i.e.* choosing randomly the vaccinated individuals, were not leading to the fragmentation of the network in several giant connected components, which would avoid the contagion between them.

Later studies on contagion processes considered global mobility flows, aggregating the flows that connect different scales: long-range mobility, represented by the airline passenger connections and short-range mobility reflected in the local commuting pattern (Fig. 1.14) [70]. After fitting the commuting flows to the gravity model, with the flows proportional to a power of the origin and destination populations and an exponential decay with the distance as introduced in Eq. (1.6), Balcan *et al.* created a synthetic commuting network representing the fit to the empirical dataset. This approach facilitated a comparison between the effects of long-range mobility and those arising from considering both short and long-range, *i.e.*, including multiple mobility scales. In fact, the comparison of the outbreak sizes between the cases considering only long-range mobility and multiple scales lead to small variations, although the commuting flows were one order of magnitude higher than those associated with flying passengers. Despite the outbreak sizes were not influenced by the short-range mobility, commuting was facilitating the synchronization of the epidemic spreading in local scales,

CHAPTER 1. INTRODUCTION

as dates when the maximum number of infected individuals was reached were more sparse when the mobility was only including long-range flows.

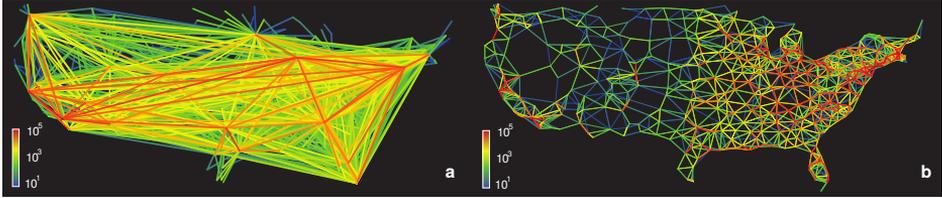


Figure 1.14: Mobility flows at different scales in continental US. **a**, Flows arising from airline passenger connections. **b**, Commuting flows. Link colours represent the number of travellers between the two connected locations. Source: Ref. [70].

The previous approach was performing data-driven simulations, using the flows of people between different locations [70]. However, it did not take into account that human-beings display a preferential return to very few specific locations [30]. This problem was analysed by Belik *et al.*, considering a system of individuals characterized by their base location, with the individuals whose base location is k being characterized by transition rates w_{mn}^k between two locations n and m [71]. If the transition and return rates w_{mn}^k does not depend on k , the disease spreading process on this system is described by a reaction-diffusion process, which leads to a propagating wave with speed $c_{RD} \sim \sqrt{\omega}$, with ω being the global transition rate. Belik *et al.* assumed movement between the base and other locations, with the users following star-like trajectories centred in the base location, *i.e.* $w_{mn}^k = 0$ for $k \neq n$ and $k \neq m$, setting a different return rate to the base location from than the leaving rate $w_{nk}^k \neq w_{kn}^k$. Considering a one-dimensional lattice, they studied the influence of the mobility of infected individuals leaving their base with a rate ω^+ and returning with a rate ω^- . In the case of symmetric movement, $\omega^+ = \omega^- = \omega$, the system displayed waves of infected individuals whose front moved with a speed $c = \frac{2\alpha\omega\sqrt{3\mathcal{D}}}{\alpha+3\omega}$, with α being the infection rate, $\mathcal{D} = l^2/2$ and l is the distance between two nearest neighbours. In contrast to the reaction-diffusion approach ($c_{RD} \sim \sqrt{\omega}$), this set-up was leading to a saturation of $c(\omega) = \frac{2\alpha\sqrt{3\mathcal{D}}}{3}$ as $\omega \rightarrow \infty$ (Fig. 1.15).

1.3.2 Other dynamics

The agent-based mobility approaches have been used more frequently for modelling other dynamics, like synchronization or game theory. For example,

1.3. DYNAMICAL PROCESSES IN SYSTEMS OF MOBILE PARTICLES

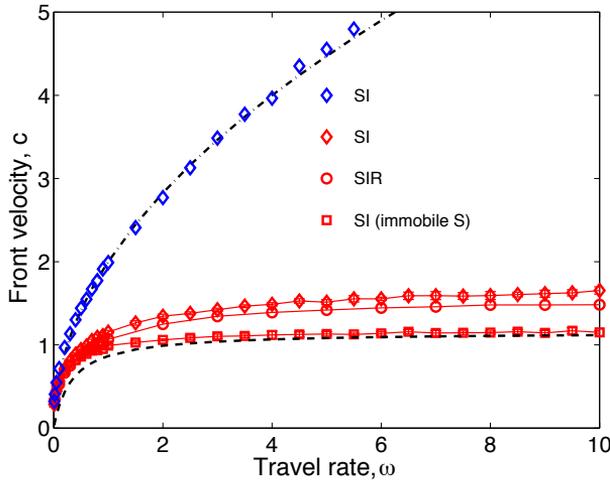


Figure 1.15: Disease spreading with reaction-diffusion and star-like trajectories. Front velocity c as a function of the travel rate ω for the stochastic simulation of different dynamics, for reaction-diffusion approach (blue symbols) and star-like trajectories (red symbols). Dashed and dashed-dotted curves represent the analytical results for star-like trajectories and reaction-diffusion, respectively. Source: Ref. [71].

Fujiwara *et al.* analysed a model of Kuramoto oscillators moving in a two-dimensional space [72]. Introducing two time scales, τ_P and τ_M , defined as the time between two consecutive phase and position updates, respectively, and setting $\tau_M = 1$, the influence of different τ_P was studied. The limit $\tau_P \rightarrow 0$ represented the absence of mobility; in this limit, there was global synchronization just for interaction ranges leading to a single connected component, as for short interaction ranges the system was composed of many connected components, each one synchronizing towards a different phase. However, the mobility ($\tau_P > 0$) was leading to global synchronization for all interaction ranges (Fig. 1.16). In fact, higher τ_P lead to lower synchronization characteristic times, approaching for $\tau_P \sim 10$ an asymptotic value. This asymptotic value represented the fast-switching approach, which considers that between two updates of the dynamics the network is changing such that the non-diagonal terms of the adjacency matrix can be approximated to the uniform probability of two particles being connected.

In a similar approach, Prignano *et al.* studied the influence of mobility in the characteristic synchronization time in a population of integrate-and-fire oscilla-

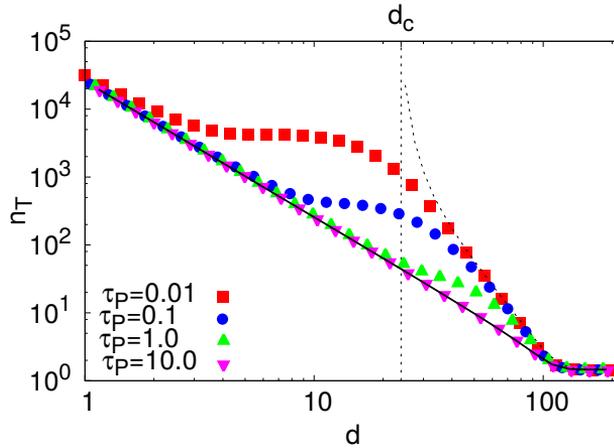


Figure 1.16: Synchronization of Kuramoto mobile oscillators. While in the static case (phase update time scale $\tau_P \rightarrow 0$) there is global synchronization only for interaction ranges d higher than the critical percolation distance d_c (dotted line), the mobility facilitates synchronization for all the values of the interaction range, and higher mobility (higher τ_P) leads to smaller synchronization characteristic times n_T , approaching asymptotically the fast-switching approach (solid line). Source: Ref. [72].

tors, setting $\tau_P = \tau_M = 1$ and varying the velocity v , defined as the displacement of each oscillator between two consecutive updates, and considering only interactions with the closest neighbour. For low velocities, the synchronization time was decreasing as a power of the velocity until it reached a minimum. Then, the synchronization time started growing and diverged for a finite range of velocities. Finally it decreased with velocity, reaching an asymptotic value for higher velocities (Fig. 1.17) [73].

The coupling between mobility and dynamics has also been studied for the voter model, which describes the evolution of a population that is characterized by a binary state, *i.e.*, 0 or 1, with a dynamics defined by updates in which an individual copies one of his neighbours' state [74]. As this model has typically been considered a model of social influence, its behaviour was compared with the results of US presidential elections [75]. The mobility was introduced from the commuting fluxes obtained from the census data, assuming that individuals interacted either at home or work. Considering a probability α of copying the state of one individual located in home location and $1 - \alpha$ for copying the state of one individual located in work location, Fernández-Gracia *et al.* set $\alpha = 0.5$

1.3. DYNAMICAL PROCESSES IN SYSTEMS OF MOBILE PARTICLES

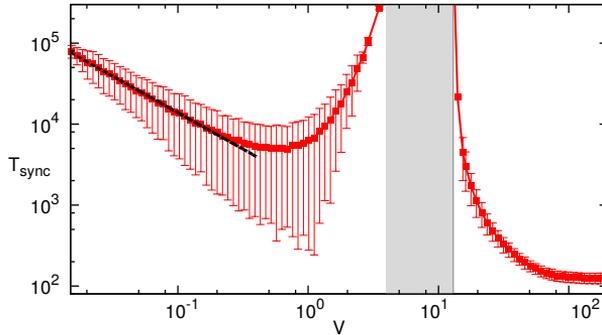


Figure 1.17: Synchronization of integrate-and-fire mobile oscillators. Characteristic synchronization time T_{sync} as a function of the velocity v . The dashed black line represents a power-law decrease, the shaded gray area depicts the range of velocities with a diverging synchronization time, and error bars correspond to the standard deviation. Source: Ref. [73].

(assuming an individual spending daily 8 h at work and 8 h sleeping and the rest at home), fitted the noise intensity $D = 0.03$ and the Monte Carlo (MC) step duration to the time between elections (10 MC step/election), finding that the vote-share fluctuations measured in the model were compatible with those observed in the empirical data, and showing that including long-range links was important in order to reproduce the empirical spatial correlations in the model.

Meloni *et al.* analysed the effects of mobility on an evolutionary model of the prisoner's dilemma [76]. Prisoner's dilemma dynamics considers an interaction between two agents that gives them a pay-off according to their state. Each agent has two possible states: cooperator (C) or defector (D). The maximum pay-off is T , obtained by a defector in the interaction with a cooperator, who gets the minimum pay-off S . Two cooperators have a higher pay-off R than two defectors, receiving P . Hence, $T > R > P \geq S$. An evolutionary version of the prisoner's dilemma considers that the agents can change their state if their neighbours have a better pay-off. Interestingly, while in an all-to-all connection the attractor is described by a global defector state, the complex networks topology can facilitate the survival of cooperative states [77]. Meloni *et al.* found that, while a static spatial network has attractors allowing the coexistence of cooperators and defectors, the mobility lead the system to only two attractors with a uniform global state, with all the individuals being whether cooperative or defective. Considering the agent's velocity v and the temptation to defect $T = b > 1$ as con-

CHAPTER 1. INTRODUCTION

tol parameters, the cooperative strategy survived only for low values of both parameters, leading to the global defector attractor otherwise (Fig. 1.18).

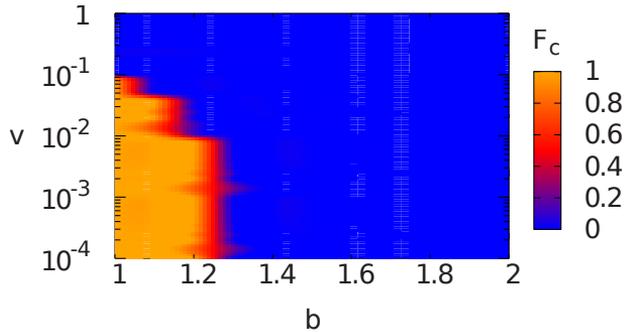


Figure 1.18: Evolutionary prisoner’s dilemma with mobility. Phase diagram showing the fraction F_c of cooperators as a function of the agent’s velocity v and the temptation to defect b . Source: Ref. [76].

1.4 Conclusions

According to this brief state-of-the-art on movement studies, both for human and animal scenarios, many questions remain open for the scientific community. Taking into account these ideas, we will pursue the following objectives:

- To develop methods suitable for the analysis of animal trajectories (Chapter 2).
- To approach the movement analysis from an aggregated point of view (Chapter 2).
- To compare trajectories associated with multiple species, identifying the common drivers of marine megafauna movement (Chapter 3).
- To describe the patterns exhibited by anthropogenic activities in the ocean. Specifically, analysing the transit pattern of vessels (Chapter 4).
- To combine models of dynamics and movement to understand the role of mobility, focusing on disease spreading processes (Chapters 5 and 6).
- To use empirical contact networks to simulate disease spreading, determining the features that play a key role in the dynamics and suggesting strategies for reducing outbreak sizes (Chapter 6).

1.4. CONCLUSIONS

This thesis is divided in two main parts, firstly analysing marine movement data and secondly studying the influence of mobility on dynamical systems. The tasks performed in these two parts are connected, as the data analysis is fundamental for obtaining general laws describing the empirical phenomena, and synthetic datasets following empirical observations can be used for modelling the interplay between the mobility and the dynamics. Data-driven simulations using large datasets would be the best approach to this problem, and we will illustrate how the empirical data can be used in the analysis of dynamical systems, studying the influence of mobility on a model of contagion dynamics. However, the costs associated with the tracking of a large number of individuals are high, making the synthetic trajectories, generated according to the observed general laws, a key tool for approaching realistic system sizes.

Part II

Marine movement analysis

Chapter 2

Intraspecies analysis of movement: the case study of southern elephant seals

The movement of individuals of the same species is the effect of a long list of causes: physiological needs, like food, shelter or reproductive mates, the interaction with the environment or even genetically programmed behaviours [58, 78]. The aggregated movement can reflect the characteristic individuals of a species, or it can emerge as a consequence of the idiosyncratic movement [21]. Hence, the analyses at different levels of aggregation facilitate the inference of characteristic behaviours for each scale. In this work, we perform an analysis of the movement at individual and aggregated scales of southern elephant seals (*Mirounga leonina*), a top predator living in Southern Ocean [79].

Previous works in animal movement have reported scale-free movement patterns, following the so-called Lévy models [80, 81]. These movement patterns were reported for several animal species [82, 83], or in modern and hunter-gatherer humans [84, 85]. The Lévy flight consist of a random walk in which the animals take steps from one location to other, separated by a step length s that is sampled from a power-law distribution $p(s) \sim s^{-\gamma}$, and taking a random direction. This model arises from the standard random walk paradigm, which assumes that the behaviour between two animal decisions is not abruptly interrupted, displaying trajectories can be described by discrete steps; in fact, each step starts when an individual chooses a duration, a speed and a direction for that step [48, 86]. In the Lévy flights, the time taken for every jump is the same, irrespectively of its length, implying in some cases extremely high velocities.

CHAPTER 2. INTRASPECIES MOVEMENT ANALYSIS

This is solved in the Lévy walks model, where the constant value is not the time taken for a jump, but the walker velocity. Interestingly, a walker displaying this movement pattern will optimally find randomly and sparsely distributed targets [51, 56]. When this Lévy-like behaviour is assumed as a hypothesis for the empirical data analysis, the identification of steps across the trajectories is required. This problem has been controversial in the movement ecology community. The steps are obtained after detecting turning points in the trajectories, and their lengths are obtained measuring the physical distance between consecutive turning points. In 1D movement, turning points are easy to identify, as they are identified as the points where the velocity changes its sign. However, this problem is more complicated in higher dimensions due to the presence of curvature. In two-dimensional movement, this issue has been addressed setting a threshold for the turning angles, such that the points in which turning angles higher than the threshold are considered turning points [87]. However, this method requires for an externally chosen threshold for the turning angles. The problem of the curvature is solved using one property of Lévy flights, which states that the projection of a d -dimensional Lévy flight into a dimension lower than d is still a Lévy flight with the same tail exponent. Then, the method for the analysis consists on projecting the movement into 1D for easily identifying the turning points. This allows to project a d -dimensional Lévy flight into a 1D movement [62], which can be easily analysed, and also facilitates the identification of spatial anisotropies after comparing projections across different directions when the projection direction is varied. It is important to highlight that this projection method leaves invariant the movement patterns that are described by independent samples from power-law distributions, but it is not the case for a generic distribution.

We propose a set-up in which we apply the methods introduced in statistical physics to animal movement analysis, which typically have been used for assessing huge datasets of human movement. Without defining turning points along the trajectories, we measure the distance between points separated by a time window T , computing the probability density function of finding a displacement d between two points that are separated by a time window T , $p(d|T)$.

This method for measuring the statistical properties of the trajectories does not assume a Lévy-like behaviour, as it does not need to determine which sampled points in the trajectories are turning points. Then, this approach is model-free, being able to measure the movement described by a generic distribution. We will apply this method to the analysis of southern elephant seals movement.

The dataset used for this analysis includes 550,537 location estimates along the trajectories of 272 southern elephant seals between 2004 and 2014, obtained with Advanced Research and Global Observation Satellite (ARGOS) platform trans-

2.1. ANALYSIS OF DISPLACEMENTS

mitting terminals (PTT), in the format (Id, Latitude, Longitude, Time). The raw data was processed with state-space models [88] to estimate the location points along movement paths with a temporal resolution of 2 h. The individuals were tagged in seven deployment locations: 15 at Macquarie Island, 11 at Campbell Island, 97 at Kerguelen Island, 29 at Elephant Island, 52 at Livingston Island, 24 at Casey Station and 44 at Davis Station (Fig. 2.1).

2.1 Analysis of displacements

We analysed the trajectories aggregating all the observed displacements for the 272 elephant seals. For a specific time window T , the displacement length $d_{i,T}(t)$, observed in individual i at time t , was the distance between the location of the seal at time t , $\mathbf{x}_i(t)$, and at time $t + T$, $\mathbf{x}_i(t + T)$:

$$d_{i,T}(t) = \text{dist}(\mathbf{x}_i(t), \mathbf{x}_i(t + T)) \quad (2.1)$$

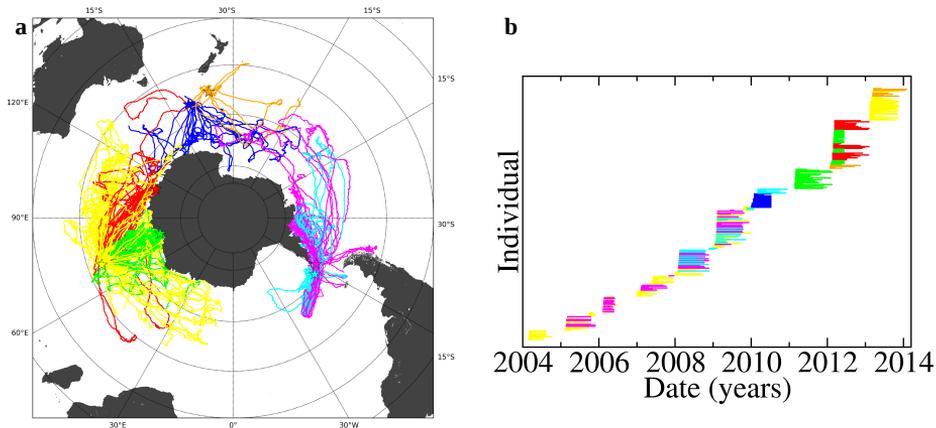


Figure 2.1: Movement data from southern elephant seals trajectories. **a**, Map of the trajectories described by the 272 tracked southern elephant seals. Star symbols represent the seven deployment locations, and the colours of the trajectories are associated with each deployment location: Kerguelen (yellow), Macquarie (blue), Campbell (orange), Livingston (magenta) and Elephant (cyan) Islands, Casey (red) and Davis (green) Stations, and land is shown in dark grey. Map generated with Matplotlib Basemap Toolkit. **b**, Temporal extent of the trajectories, with the color associated with the deployment locations introduced in **a**. Source: Ref. [79].

CHAPTER 2. INTRASPECIES MOVEMENT ANALYSIS

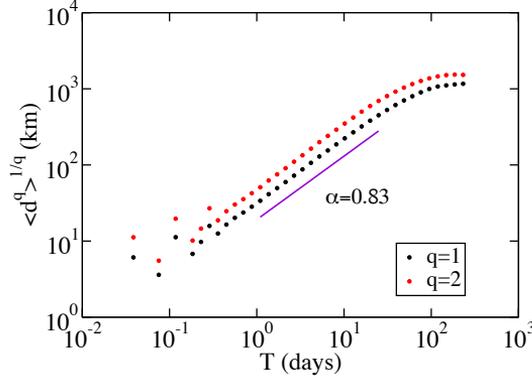


Figure 2.2: Scaling of the displacement lengths with time of southern elephant seals movement. The q -th root of the q -th moment, $\langle d^q \rangle^{1/q}$, averaging over the conditional distribution $p(d|T)$, is a power of time T , $\langle d^q \rangle^{1/q} \sim T^\alpha$. For $q = 1$ (black dots), this measurement represents the average displacement, while for $q = 2$ (red dots) it is the root mean square displacement (RMSD). Source: Ref. [79].

where $\text{dist}(\mathbf{x}, \mathbf{y})$ is the great-circle distance between locations \mathbf{x} and \mathbf{y} . Note that Eq. (2.1) can be used only when the position at $t + T$ is known; otherwise, $d_{i,T}(t)$ cannot be calculated and was not included in the analysis. For each individual, we measured the observed displacement lengths along its trajectory, and we aggregated these observations for all the individuals in our dataset. For a pair of points \mathbf{x} and \mathbf{y} on the surface of a sphere of radius R , whose coordinates (latitude, longitude) are given by $\mathbf{x} = (\phi_x, \lambda_x)$ and $\mathbf{y} = (\phi_y, \lambda_y)$, the haversine formula allows the calculation of $\text{dist}(\mathbf{x}, \mathbf{y})$ between them:

$$\text{dist}(\mathbf{x}, \mathbf{y}) = 2R \arcsin(\sqrt{h}) \quad (2.2)$$

where $h = \sin^2\left(\frac{\phi_x - \phi_y}{2}\right) + \cos \phi_x \cos \phi_y \sin^2\left(\frac{\lambda_x - \lambda_y}{2}\right)$.

First of all, we analysed the scaling of the moments of the observed distribution of displacement lengths d with the time window T , $\langle d^q \rangle(T) = \sum_{i,t} d_{i,T}^q(t)$, running over all the observations in one trajectory and over all the trajectories in our dataset. For intermediate values of T , these moments scaled as a power-law of T , saturating for high values of T due to the finite size of the regions that the seals explored. The moments for different values of $q > 0$ scaled with exponents that were linearly proportional to q , which suggested the absence of multifractality in the movement [89]. Then, the q -th root of the q -th moment had an exponent α that

2.1. ANALYSIS OF DISPLACEMENTS

did not depend on q , $\langle d^q \rangle^{1/q} \sim T^\alpha$ (Fig. 2.2). An ordinary least squares fit to these curves lead to the exponent $\alpha = 0.83$, which indicated that the movement was super-diffusive, as this exponent was higher than the characteristic for Brownian motion ($\alpha_{\text{dif}} = 1/2$).

Focusing on a specific time window T , the statistical properties of the displacement lengths, aggregating all the observations along a trajectory and all the trajectories, were studied through the probability density function describing them $p(d|T)$, which takes the value for the bin centred in d and representing the range (a, b) :

$$p(d|T) = \frac{\sum_{i,t} \int_a^b \delta(d_{i,T}(t) - x) dx}{A(T)(b - a)} \quad (2.3)$$

where $\delta(x)$ is the Dirac delta function ($\int_a^b f(z)\delta(z - z_0)dz = f(z_0)$ if $a < z_0 < b$ and 0 if $z_0 < a$ or $b < z_0$), $A(T)$ is the total number of observed displacements for the time window T and $b - a > 0$ is the length of the bin.

This distribution displayed a power-law scaling $p(d|T) \sim d^{-\gamma}$ for displacement lengths smaller than $d_L(T) \sim \langle d(T) \rangle$, followed by a sharp cut-off that appears for displacement lengths that represent the maximum effective velocities displayed by southern elephant seals, and this cut-off was higher for higher values of T (Fig. 2.3a). A maximum likelihood estimation (MLE) of the exponent in the power-law distribution (see Appendix) fitted this parameter to $\gamma = 0.60$. The behaviour was qualitatively the same for different values of T ; in fact, after normalizing the observed displacement lengths for a time window T , dividing them by the average displacement length, $D = \frac{d(T)}{\langle d(T) \rangle}$, the displacement length distributions for different T collapsed into a universal function $\tilde{p}(D)$ for a broad range of time windows (Fig. 2.3b). This universal function was obtained after applying the normalization transformation, which had the same properties as Eq. (1.1) taking into account that $\langle d(T) \rangle \sim T^\alpha$ and $p(d|T) \sim T^{-\alpha} p(d/T^\alpha)$, leading to:

$$\tilde{p}(D) = p(d|T) \langle d(T) \rangle \quad (2.4)$$

This collapse for a broad range of temporal scales (Fig. 2.3b), happening for time windows ranging between half a day and around a hundred of days, showed the temporal invariance of the observed movement pattern. In fact, the probability density distribution of displacement lengths was described by a power-law distribution, that exhibits self-similarity, showing the spatial invariance of the movement across several spatial scales. Then, we can conclude that the behaviour of southern elephant seals at a species level displays a spatial and temporal invariance.

CHAPTER 2. INTRASPECIES MOVEMENT ANALYSIS

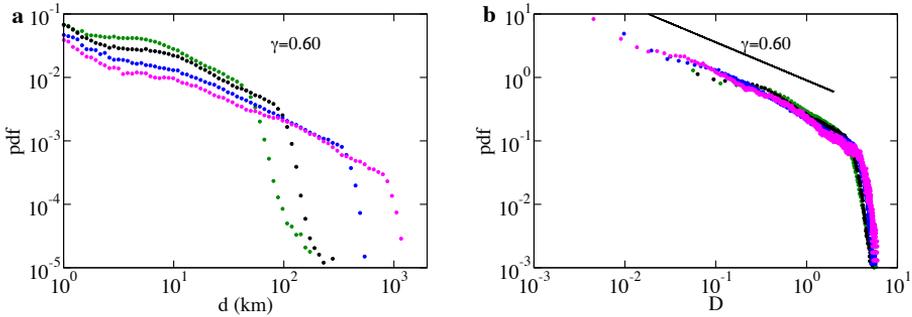


Figure 2.3: Displacements analysis of southern elephant seals movement. **a**, Distribution of displacement lengths for time windows T of 0.5 (green), 1 (black), 4 (blue) and 10 (magenta) days. **b**, After normalizing the distributions in **a**, the curves collapse into a universal function described by a power-law with exponent $\gamma = 0.60$ (solid black line) and a sharp cut-off. Source: Ref. [79].

Nevertheless, the measured exponent in the displacement length distribution was outside the Lévy stable regime, according to the central limit theorem, and it could have been artificially introduced through our measurement method. In order to check the robustness of the exponent, we used methods introduced in movement ecology for detecting steps, combining the projection into 1D trajectories [62] and rescaling techniques for different sampling frequencies that allow the assessment of different temporal scales [90]. For different sampling frequencies, all the distributions of step lengths curves collapsed in a region described by a power-law which fitted to an exponent 0.62, while for higher step lengths there was a power-law distribution whose exponent depended on the sampling frequency (Fig. 2.4).

The generalized central limit theorem states that a distribution with a power-law tail, described by the function $x^{-\theta}$, with $1 < \theta < 3$, is stable, meaning that, as the number of random numbers drawn from this distribution increases, its sum will converge to a stable distribution [91]. For $\theta > 3$, this sum will converge to a Gaussian distribution [92]. Our measured exponent $\gamma = 0.60 < 1$ was out of the stable regime, and this suggested that more ingredients were present in the seal's movement, like non-Markovian dynamics, which suggested memory as a driver of this pattern [93].

2.1. ANALYSIS OF DISPLACEMENTS

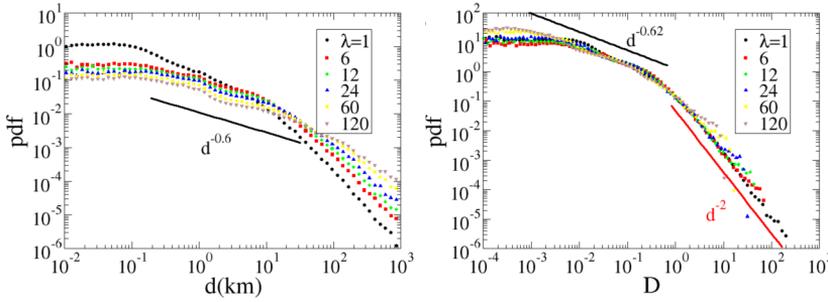


Figure 2.4: Step length analysis of the trajectories. **a**, Probability density function of the step length d for different temporal rescalings λ , such that the considered trajectories are those including the 1st, $(1 + \lambda)$ th, $(1 + 2\lambda)$ th ... locations from the original trajectories. The projection for the turning points identification is done in the latitude axis. **b**, When the step lengths are normalized dividing by the average step length, there are two power-laws, with the power-law associated with smaller displacements collapsing to the same curve, while for higher step lengths the power-law exponent depends on λ . Source: Ref. [79].

The central limit theorem

Consider $\mathcal{X} = \{X_1, X_2, X_3, \dots, X_n\}$ a set of random variables that are independent and identically distributed, with finite average μ_n and finite variance σ_n^2 . Then, the sum of the variables $S_n = \sum_{i=1}^n X_i$ converges to a Gaussian distribution as $n \rightarrow \infty$, with mean $\mu_n = n\mu$ and variance $\sigma_n^2 = n\sigma^2$ [91].

However, the central limit theorem cannot be always applied when the distribution of the random variables has the shape $P(x) \sim x^{-\theta}$, whose first and second moments are given by

$$\langle x \rangle = \int_0^{\infty} xP(x)dx = \frac{x^{2-\theta}}{2-\theta} \Big|_0^{\infty} \quad (2.5)$$

$$\langle x^2 \rangle = \int_0^{\infty} x^2P(x)dx = \frac{x^{3-\theta}}{3-\theta} \Big|_0^{\infty} \quad (2.6)$$

Eqs. (2.5) and (2.6) diverge for $\theta \leq 2$ and $\theta \leq 3$, respectively, such that the central limit theorem is valid for power-law distributions with $\theta > 3$. For $1 < \theta \leq 3$, the generalized central limit theorem states that $P(x)$ is a stable distribution, as the sum of the random variables converges to a distribution defined by a power-law with the same exponent.

CHAPTER 2. INTRASPECIES MOVEMENT ANALYSIS

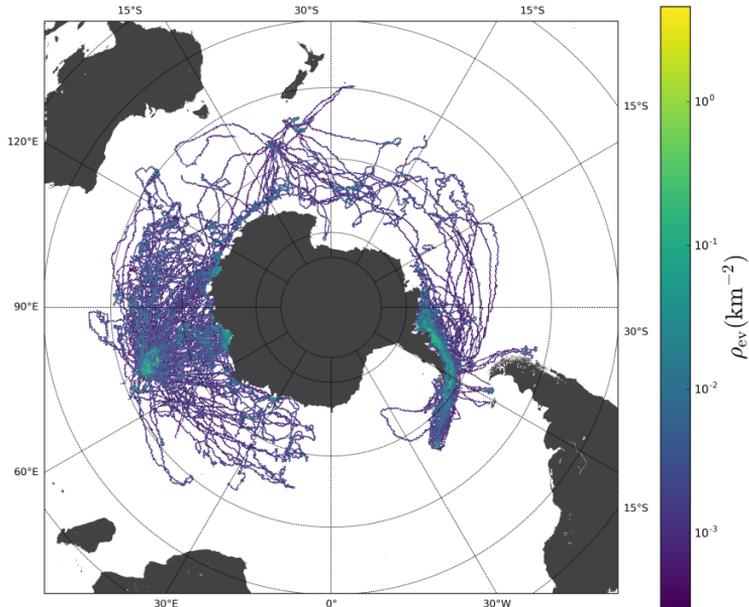


Figure 2.5: Spatial distribution of southern elephant seals. Map showing the occupancy ρ_{ev} of the grid cells visited by southern elephant seals. The colour scale is logarithmic, from violet (low occupancy) to yellow (high occupancy). Map generated with Matplotlib Basemap Toolkit. Source: Ref. [79].

In the following step, we analysed the use of space, according to the observed trajectories. We divided the space into grid cells of 0.25° both in latitude and longitude, calculating the occupation ρ_i of each cell i as the fraction of recorded visits to the cell, which is a proxy for the time spent on it, divided by the area of the cell. The general pattern of this occupation was described by a distribution that exhibited a long tail fitting to a power-law with exponent 1.88, revealing the presence of places with extremely high levels of occupation coexisting with many low-occupied grid cells. In fact, 80% of the observations occurred in the 23.7% most visited cells (Fig. 2.5). This broad range of values for the use of space and the exponent of the displacement lengths distribution that was out of the stable regime of the central limit theorem lead us to split the movement in two different stages. Hence, we divided the space into high occupancy cells, which were those located in the first third of the ranking according to occupancy, and low occupancy, which were the rest. After this split, we assessed the displacement lengths distributions according to where the displacements were observed: we separated the displacements originated in highly occupied regions from those

2.1. ANALYSIS OF DISPLACEMENTS

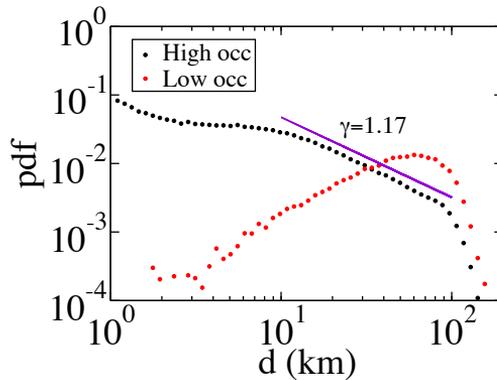


Figure 2.6: Movement regimes of southern elephant seals. Distribution of displacement lengths d for the time window $T = 1$ day, splitting the displacements according to the occupancy of the area in which they occur. Directed movement, with well-defined and high velocity is displayed in low occupancy regions (red points), while a slower but highly variable displacement lengths take place in highly occupied regions (black points). The solid violet line represents a power-law fit with exponent 1.17. Source: Ref. [79].

located in low-occupied regions, leading to remarkable differences between the behaviour exhibited in both stages. Specifically, the low-occupied areas were characterized by long displacements with well-defined characteristic average and variance, while the movement in highly occupied areas was described by a broad distribution of shorter displacements (Fig. 2.6). In fact, the distribution of displacements in the high occupancy regions had a power-law tail with a sharp cut-off, with an exponent $\gamma = 1.17$ that was in the stable regime of the central limit theorem (Fig. 2.6).

Summing up, southern elephant seals displayed a superdiffusive movement pattern, which was characterized by a distribution of displacement lengths that was out of the stable region of the central limit theorem. Considering this behaviour and the broad distribution of spatial occupancies, we split the space into highly and low-occupied regions, observing a remarkably different behaviour. Specifically, the elephant seals displayed fast and directed movements, which suggested the memory as a driver of the movement that connected highly occupied locations, in which the displacement lengths distributions were broad.

2.2 Individual analysis

The analysis of the movement patterns described by southern elephant seals revealed a universal behaviour that was consistent across a broad range of spatial and temporal scales. We wanted to know if these universal features arose as a consequence of aggregating a broadly diverse dataset or this pattern was reproduced by each individual seal. For this purpose, we analysed the movement of the elephant seals at an individual scale.

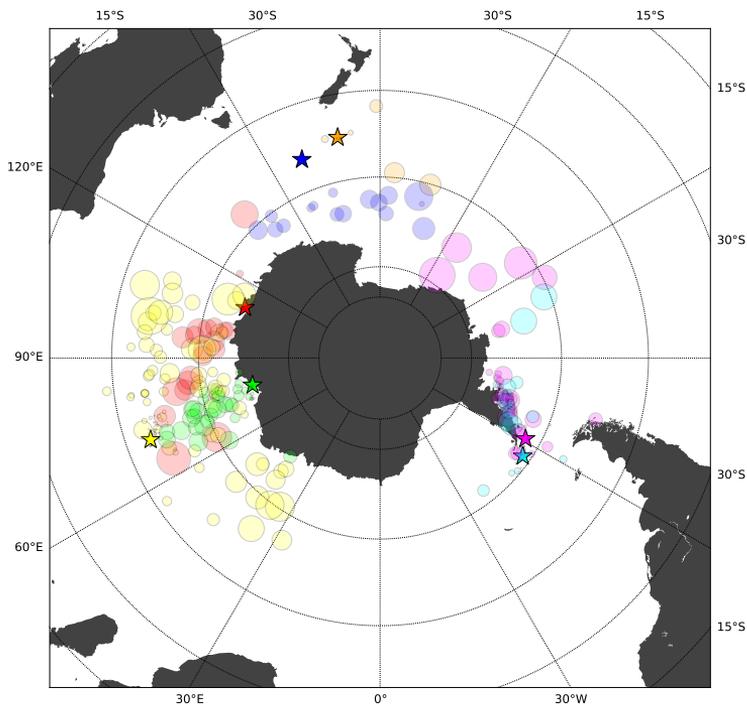


Figure 2.7: Centres and extensions of individual trajectories. Map of the centre of each individual trajectory, depicted with circles. The size of the circle is proportional to the gyration radius of the seal, and the color stands for the deployment location, which are depicted with star symbols. Map generated with Matplotlib Basemap Tool. Source: Ref. [79].

We measured the centre of the trajectories. This task was performed in two steps. First of all, we converted our original coordinates (latitude, longitude) into the three-dimensional Cartesian coordinates $\mathbf{x}_i(t) = (X_i(t), Y_i(t), Z_i(t))$ associated with

2.2. INDIVIDUAL ANALYSIS

the points in the surface of a sphere, setting our origin of coordinates in the Earth centre. Secondly, we obtained the average position along the trajectory of each seal i in Cartesian coordinates, $\langle \mathbf{x}_i \rangle_t = (\langle X_i(t) \rangle_t, \langle Y_i(t) \rangle_t, \langle Z_i(t) \rangle_t)$, and the centre of the trajectory was the projection of this point into the surface of the sphere with Earth's radius R_E :

$$\mathbf{c}_i = \langle \mathbf{x}_i \rangle_t \frac{R_E}{\sqrt{\langle X_i(t) \rangle_t^2 + \langle Y_i(t) \rangle_t^2 + \langle Z_i(t) \rangle_t^2}} \quad (2.7)$$

With the centre of each trajectory, we calculated the gyration radius for each seal, which indicates how sparse in the space its trajectory is, following Eq. (1.3). The gyration radius varied significantly across different individuals, indicating that the movement had a deep idiosyncratic character. Specifically, the gyration radii varied from 10 km to 2000 km (Figs. 2.7,2.8).

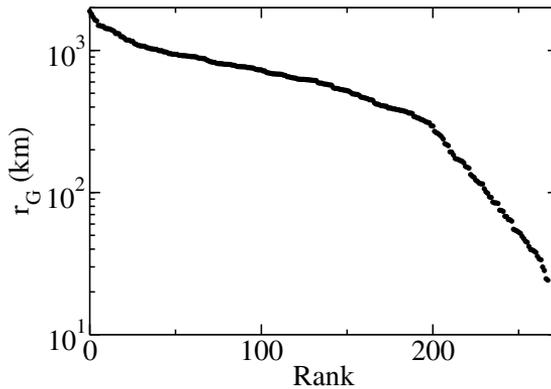


Figure 2.8: Ranking of the gyration radii of the individual trajectories. We observe a broad range of behaviours, with individuals whose gyration radii are of several thousands of kilometres and others in the order of the kilometre. Source: Ref. [79].

In the previous section, we showed that the collective movement pattern exhibited by southern elephant seals was directed and it could be split in two different modes. In the high speed and low occupancy mode, the movement was described by displacement lengths and turning angles that were consistent for long temporal scales, suggesting the presence of temporal correlations. This hypothesis was tested making a reshuffling of the trajectories, in which we kept the starting and final points of the trajectories, but reshuffled the order in which the displacements appeared (Fig. 2.9). In depth, we partitioned each trajectory

CHAPTER 2. INTRASPECIES MOVEMENT ANALYSIS

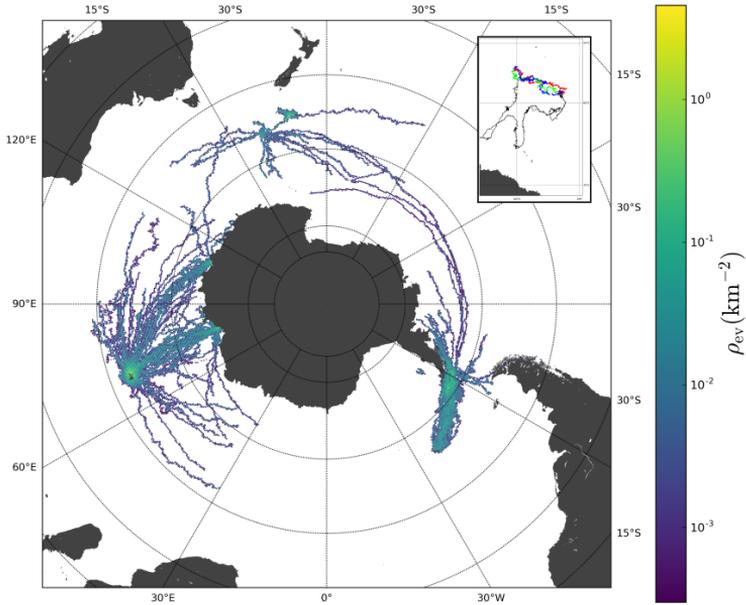


Figure 2.9: Occupancy pattern described by reshuffled trajectories. Inset: examples of three realizations of the reshuffling (red, blue and green trajectories), in comparison with the real trajectory (black). Map generated with Matplotlib Basemap Tool. Source: Ref. [79].

in a set of 2D displacements in which the time window was the resolution of our data ($T_{\min} = 2$ h), and then we randomized the sequence in which those displacements appeared. The first comparison between the original and the reshuffled data revealed that the original trajectories lead to a higher number of visited grid cells (Fig. 2.10, inset).

The entropy of the visitation pattern displayed by individual j , S_j can be calculated from $p_{i,j}$, the probability that individual j visits cell i (*i.e.* the fraction of locations in the trajectory of individual j that are associated to cell i):

$$S_j = - \sum_{i=1}^{M_j} p_{i,j} \log p_{i,j} \quad (2.8)$$

where M_j is the number of cells that individual j visits. This measurement is indicative of the fidelity of the individuals to the cells that they visit most frequently, where high values of the entropy correspond to uniform visitation

2.2. INDIVIDUAL ANALYSIS

patterns, and low values are associated with high fidelity to a few cells [22, 94]. In order to allow the comparison across different individuals, we normalized the entropy with its maximum value, obtaining $s_j = \frac{S_j}{\log M_j}$, which ranges between 0 (one visited cell) and 1 (cells are visited with uniform probability). The individual entropies had in most cases high values ($s > 0.6$), showing that the seals did not have fidelity to particular cells.

From the entropy S_j and the number of visited cells M_j , we calculated the limit of predictability Π^{MAX} , which is the maximum probability with which a predictive algorithm can guess the location of the individual [28], solving numerically the implicit equation for Π_j^{MAX} :

$$S_j = H(\Pi_j^{\text{MAX}}) + (1 - \Pi_j^{\text{MAX}}) \log(M_j - 1) \quad (2.9)$$

where $H(x) = -x \log x - (1 - x) \log(1 - x)$. The distribution of this bound for the predictability was peaked around $\Pi^{\text{MAX}} = 0.2$, decaying smoothly to $\Pi^{\text{MAX}} = 0.8$, with most individuals (60%) falling in the range 0.2-0.4 (Fig. 2.10). However, some individuals displayed highly predictable trajectories, with bounds that were approaching one. In fact, straight trajectories in which there were no repeated visits to the same cell lead to low values of predictability (Fig. 2.10, left lower panel), while as some cells were visited more frequently, the limit of predictability increased (Fig. 2.10, central lower panel), reaching maximum values when most of the points in a trajectory were located in a single cell (Fig. 2.10, right lower panel). In general, the limits of predictability measured from southern elephant seals trajectories were much smaller than the ones associated with human movement, which have a distribution peaked around $\Pi^{\text{MAX}} \approx 0.93$ [28]. However, despite this limit of predictability was relatively small, compared with the measurement for human movement, it was higher than the limit of predictability measured in the reshuffled trajectories (Fig. 2.10), which had an average bound for the predictability of 0.18 ± 0.07 (standard deviation), and displaying smaller predictabilities in the randomized trajectories for most individuals.

Recapitulating, the individual analysis of the movement of the seals revealed a high heterogeneity, first with a broad range of gyration radii, from some individuals that stayed close to the deployment location to others that displayed migrations of thousands of kilometres, and secondly through the use of space analysis, revealing in general low predictabilities, but in ranges that were higher and broader than in the randomized trajectories.

CHAPTER 2. INTRASPECIES MOVEMENT ANALYSIS

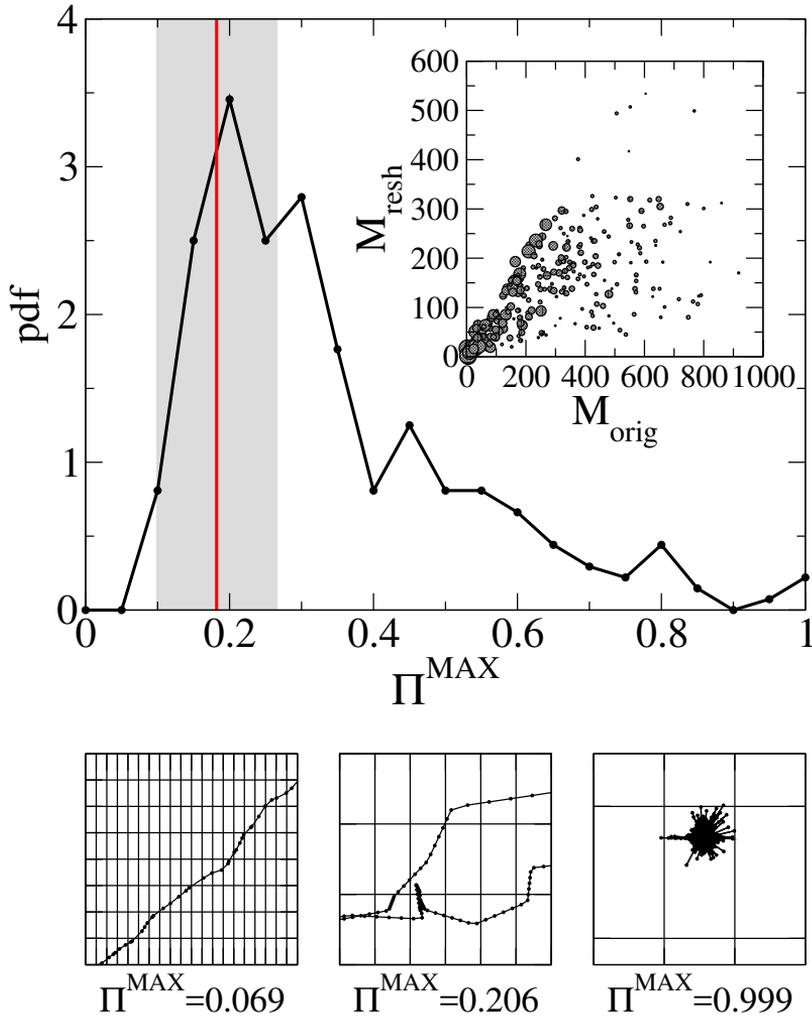


Figure 2.10: Predictability of the trajectories displayed by southern elephant seals. Distribution of the limits for predictability Π^{MAX} measured for each seal. The red line shows the average of Π^{MAX} for the reshuffled trajectories with a standard deviation (shaded area). Inset: number of visited grid cells for the reshuffled (M_{resh}) and original (M_{orig}) trajectories. Symbol size is proportional to the limit of predictability of the original trajectory. Lower panels show three examples of trajectories with low (left), average (centre) and high (right) limits of predictability. Source: Ref. [79].

2.3 Spatial networks of flows

After studying the displacements, the spread of the trajectories and the anisotropies due to the use of space, we considered another approach that has been used for the analysis of human mobility, building a transition probability matrix $\Omega(T)$, in parallelism with the Origin-Destination matrix. Setting a time window $T = 1$ day, the entry ij in the transition probability matrix is the fraction of visits to cell i leading to cell j after one day:

$$\Omega_{ij}(T) = \frac{\sum_{n=0}^N \sum_{t=t_0}^{t_{\max}-T} \delta_{c_{nt},i} \delta_{c_{n(t+T)},j}}{\sum_{n=0}^N \sum_{t=t_0}^{t_{\max}-T} \delta_{c_{nt},i}} \quad (2.10)$$

where n runs over individuals, c_{nt} is the cell in which individual n is located at time t , and $\delta_{i,j}$ is the Kronecker delta, taking value 1 if $i = j$ and 0 otherwise. Note that, as Ω_{ij} is a transition probability matrix, $\sum_j \Omega_{ij} = 1 \forall i$.

This matrix represents a weighted and directed network that includes the information of the flows in space described by the movement of southern elephant seals. Using network theory, we looked for communities, which are sets of nodes that are strongly connected between them, in comparison with the number of connections with the rest of the network. This provided us with the marine provinces defined by this movement pattern, which are the areas visited coherently. The community search in spatial networks has been applied to human-based datasets, using telecommunications networks [44], functional networks derived from activity patterns [47] or flows described by human mobility [95]. Infomap was the community search algorithm chosen for this analysis, as it studies the dispersion of random walkers along the network and then splits the communities according to the regions in which the random walkers stay for a long time [96]. This algorithm produces an output in which the communities are organized in hierarchical levels, from the most inclusive to the most local levels. In our analysis, we considered the two first levels of aggregation.

We found two communities at level 0, which had internal structure at level 1, giving rise to 6 different communities (Fig. 2.11a). At level 0, the deployment location of the seals was clearly correlated with the community associated with the cells that they visited, as seals tagged in Kerguelen Island, Casey and Davis Stations were visiting one community and seals tagged in Macquarie, Campbell, Elephant and Livingston Islands were associated with the other province (Figs. 2.1a and 2.11a,c). However, more structure was discovered at level one, finding

CHAPTER 2. INTRASPECIES MOVEMENT ANALYSIS

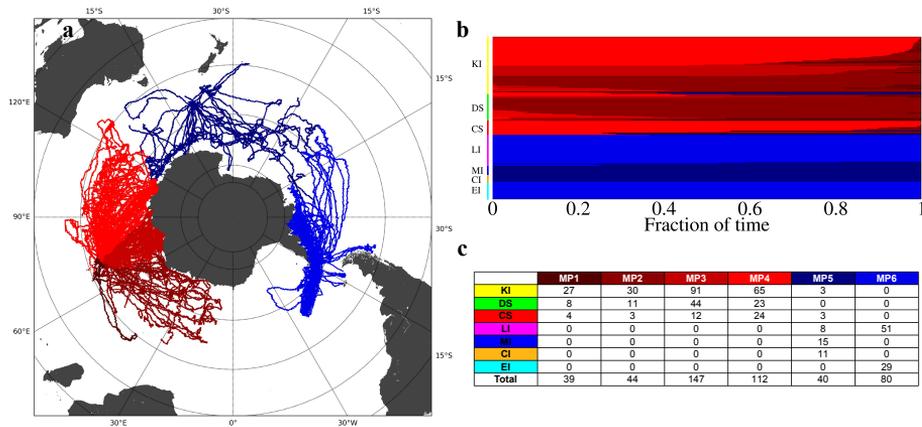


Figure 2.11: Marine provinces according to the movement of southern elephant seals. a, Map of the marine provinces detected by Infomap in the network defined by the transition probability matrix $\Omega(T)$. Red and blue colours correspond to the provinces at level 0, while the different levels of darkness separate provinces at level 1. Map generated with Matplotlib Basemap Tool. **b,** Fraction of time that each seal spent in each province. Seals are ordered according to their deployment location. **c,** Number of different seals, organized according to their deployment location, that visited at least once each province. Source: Ref. [79].

that most individuals spent most of their time in a single community. Specifically, 80% of the individuals were located in more than 80% of the tracked time in the cells belonging to a single community at level 1 (Fig. 2.11b).

2.4 Conclusions

The movement of southern elephant seals was described by universal laws that were invariant across several spatial and temporal scales, despite we detected high levels of idiosyncrasy at individual levels. The presence of a distribution of displacement lengths described by a power-law with a low exponent, which was out of the stable regime of the generalized central limit theorem, suggested that non-Markovian processes were present in these movement patterns. This was confirmed through the split of trajectories happening at highly occupied areas, in which the displacements were described by a distribution with a stable shape, and the trajectories in low-occupied regions, which were more directed,

2.4. CONCLUSIONS

with a well-defined and higher velocity, in which our conclusion was that the memory was driving the movement. This behaviour is compatible with the analysed trajectories, which were representing return journeys between breeding colonies and foraging grounds. In fact, southern elephant seals display a long-term fidelity to the foraging regions, visiting the same locations along scales of several years, with a higher probability of returning to the locations leading to higher levels of foraging success in the previous year [97]; this fidelity is also displayed in the diving behaviour, as elephant seals show preferred specific foraging depths [98].

The collective analysis of the transition probabilities in the space provided us with the marine provinces defined by southern elephant seals, which were consistently visited by the individuals, as most individuals spent most of their time in a single province. We also detected that, at the most inclusive level, the seals which were tagged in the same location were associated with the same community, making a split between three different populations: 1) the seals tagged in Kerguelen Island, Casey and Davis Stations, 2) the seals tagged in Macquarie and Campbell Island, and 3) those tagged in Elephant and Livingston Island. These provinces were consistent with previous knowledge of the regions explored by each population. For example, the seals from Kerguelen Island were travelling to either sub-Antarctic or high-Antarctic foraging grounds [99].

An analysis at individual level showed idiosyncrasy. This behaviour was reflected in a broad range of gyration radii. The seals were visiting a higher number of spatial grid cells than in a randomization of the trajectories due to the presence of low-occupied cells in which they moved directly. In general, the seals visited the regions of the space with a probability that was close to the uniform visitation, with high values of the normalized entropies. However, the random case was even more uniform, leading to smaller values of the predictability, although the number of visited cells in the reshuffled trajectories was smaller, and this could make the predictability rise. The general result was that the use of space displayed by elephant seals was highly unpredictable, especially compared to the one associated with human behaviour.

This case study was a sample of how these methods, used from the statistical physics community for the analysis of human mobility, can be applied to animal movement, specially when the availability of tracking data is increasing. In the next chapter, some of these methods will be applied to an interspecies analysis, identifying the general drivers of the marine megafauna movement.

Chapter 3

Interspecies analysis of marine megafauna movement

Satellite tracking has become the main data source in movement ecology [100, 101]. This community has made remarkable efforts to overcome the difficulties related with technical limits, like the price of the devices or the deployment process, focusing on studies that analyse the recordings of the presence of a species in a specific location, or the description of its movement in a confined area. These recordings were based on previous knowledge of the studied species. For example, the awareness that a region is visited by a species could lead to a project for assessing the frequency and the abundance of individuals visiting that area [102, 103]. The analysis of trajectories in well-defined areas is also an important research line, leading to the analysis of collective movement phenomena, as reflected in the studies of school fish [104], starling flocks [105], or sheep confined in a field [106]. In contrast, satellite tracking provides data describing the trajectories displayed by the individuals for long periods of time, leading to new discoveries about the movement in places where previous tracking techniques could hardly work. This is especially important for the study of marine megafauna movement [107], as it facilitates more accurate assessments of the movement behaviour in areas of hard accessibility, like open oceans [108], skies [109], or the polar regions [110]. Moreover, some sensors allow studying the diving behaviour, enhancing for the comparison between horizontal and vertical movement [111]. The availability of longer tracks for more and more species allows for multispecies analysis of animal movement [60, 59, 112].

In this chapter, we present the analysis of the movement of 50 marine species that flight, walk and swim [113]. Firstly, we characterize the observed movement

CHAPTER 3. INTERSPECIES MOVEMENT ANALYSIS

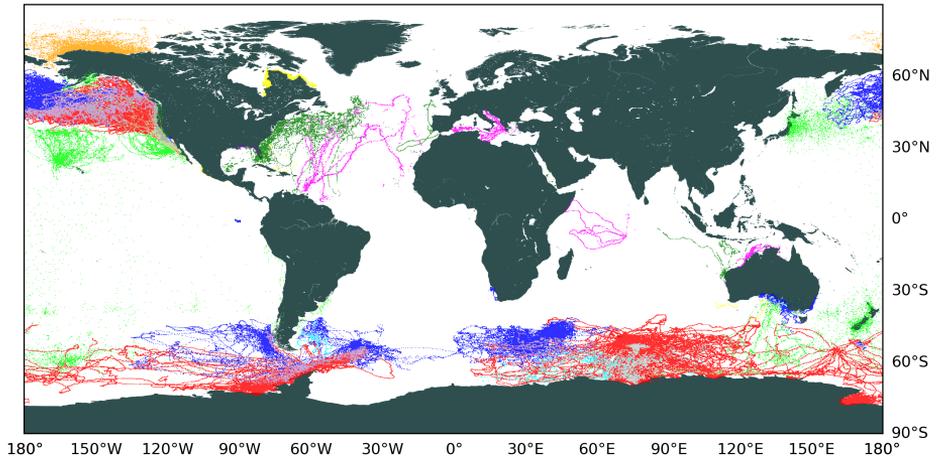


Figure 3.1: Global map of the gathered dataset. Colours represent each of the nine guilds with data: cetaceans (yellow), eared seals (blue), flying birds (green), penguins (cyan), polar bears (orange), sharks (dark green), sirenians (violet), true seals (red), and turtles (magenta). Mixed colours appear in locations visited by species belonging to different guilds. Map generated with Matplotlib Basemap Tool. Source: Ref. [113].

with a few measurements. Secondly, we use these measurements as response variables of the known features of the species, working at species and individual level. Our challenge is to detect the relevant features that explain the observed response variables, which are measured in the tracking datasets, that is, we search the general drivers of marine megafauna movement.

3.1 Data description

The original data used in this work is the result of collecting the data obtained across three decades (1985-2015), comprised of ~2.8 million locations corresponding to 2557 individuals that belong to 50 marine vertebrates, covering a significant fraction of the Earth's oceanic surface. The analysed species are grouped in nine different guilds: cetaceans, eared seals, flying birds, penguins, polar bears, sharks, sirenians, true seals and turtles (Fig. 3.1). The information in the datasets included the horizontal coordinates, (*i.e.*, latitude and longitude)

3.1. DATA DESCRIPTION

in which an individual i is located at position $\mathbf{x}_i(t)$ at time t . These datasets were coming from different tracking systems:

- Advanced Research and Global Observation Satellite (ARGOS) is a satellite-based system that uses Doppler effect for determining the position of the transmitter.
- Global Positioning System (GPS), which is supported by a network of 24 satellites orbiting around the Earth, and when the transmitter receives a signal from one of these satellites, it computes the distance to the satellite emitting that signal. If there are three or more signals, the location of the transmitter can be calculated from these distances.
- Global Location Sensor (GLS), which records the light level sequence, such that the latitude and the longitude of the sensor can be inferred.

There were some datasets based on processed information, while in others the information was raw. There were several species for which both ARGOS and GPS tracks were available, which allowed for comparison. After checking that the statistical measurements for trajectories obtained from ARGOS and GPS systems were compatible, we gathered the trajectories tracked with different tracking systems. A similar analysis was performed for raw and processed trajectories, obtaining no significant differences. However, the species that were tracked with GLS system could not be compared, so these species were initially discarded. After further pre-analysis, requesting a high number of observations such that the results were statistically significant, the final analysis was performed on the data describing the movement of animals belonging to 38 different species (Table 3.1).

The movement data was complemented with information about basic features of each species. Specifically, we initially considered taxonomy (taxa, family, group), allometry (body length and mass), life history traits (*e.g.*, breeding and foraging strategies), energy requirements, locomotion mode and the habitat region (polar, temperate, tropical). The main aim of this work was to connect these species features with the observed movement. For this purpose, we obtained the statistical properties of the movement data, and then we assessed the correlations between movement statistics and the species features, *i.e.*, the species features were used as explanatory variables of our response variables, which were the movement statistics.

CHAPTER 3. INTERSPECIES MOVEMENT ANALYSIS

	Common name	Scientific name	# Indiv	# Loc
Orange	Polar bear	<i>Ursus maritimus</i>	62	102499
Green	Western gull	<i>Larus occidentalis</i>	41	221209
Green	Laysan albatross	<i>Phoebastria immutabilis</i>	122	13545
Green	Short tailed shearwater	<i>Ardenna tenuirostris</i>	47	12649
Green	Black footed albatross	<i>Phoebastria nigripes</i>	72	11741
Blue	Macaroni penguin	<i>Eudyptes chrysolophus</i>	91	18318
Blue	King penguin	<i>Aptenodytes patagonicus</i>	8	13125
Blue	Little penguin	<i>Eudyptula minor</i>	102	5009
Yellow	Beluga	<i>Delphinapterus leucas</i>	31	37089
Yellow	Bottlenose dolphin	<i>Tursiops truncatus</i>	3	1681
Yellow	Southern right whale	<i>Eubalaena australis</i>	3	1130
Yellow	Humpback whale	<i>Megaptera novaeangliae</i>	12	707
Yellow	Pilot whale	<i>Globicephala macrorhynchus</i>	2	510
Red	Northern elephant seal	<i>Mirounga angustirostris</i>	321	1027839
Red	Southern elephant seal	<i>Mirounga leonina</i>	272	554003
Red	Weddell seal	<i>Leptonychotes weddellii</i>	96	62873
Red	Crabeater seal	<i>Lobodon carcinophaga</i>	30	12614
Blue	Australian sea lion	<i>Neophoca cinerea</i>	236	85507
Blue	Northern fur seal	<i>Callorhinus ursinus</i>	158	80863
Blue	California sea lion	<i>Zalophus californianus</i>	75	52333
Blue	New Zealand sea lion	<i>Phocartos hookeri</i>	26	25256
Blue	Long-nosed fur seal	<i>Arctocephalus forsteri</i>	102	23502
Blue	Southern sea lion	<i>Otaria flavescens</i>	20	17448
Blue	Galapagos sea lion	<i>Zalophus wolfebaeki</i>	64	14896
Blue	Australian fur seal	<i>Arctocephalus pusillus doriferus</i>	8	1342
Blue	Cape fur seal	<i>Arctocephalus pusillus</i>	6	1066
Green	Mako shark	<i>Isurus oxyrinchus</i>	16	16153
Green	Tiger shark	<i>Galeocerdo cuvier</i>	70	10430
Green	Whale shark	<i>Rhincodon typus</i>	16	3350
Green	Blue shark	<i>Prionace glauca</i>	6	1755
Green	Bull shark	<i>Carcharhinus leucas</i>	26	1301
Green	Great hammerhead shark	<i>Sphyma mokarran</i>	20	1059
Purple	West Indian manatee	<i>Trichechus manatus</i>	10	121668
Purple	Dugong	<i>Dugong dugon</i>	6	3728
Pink	Leatherback turtle	<i>Dermochelys coriacea</i>	32	72805
Pink	Green turtle	<i>Chelonia mydas</i>	20	13901
Pink	Flatback turtle	<i>Natatos depressus</i>	11	12205
Pink	Loggerhead turtle	<i>Caretta caretta</i>	6	5863

Table 3.1: Information about the 38 datasets used for this analysis. Colours represent the guilds introduced in Fig. 3.1.

3.2 Global and individual convergence

First of all, we analysed the scaling properties of the displacement lengths $d(T)$ with the time window T [48, 114]. Secondly, we characterized the distribution of displacement lengths observed for a fixed time window T , and the distribution of turning angles connecting two consecutive displacements.

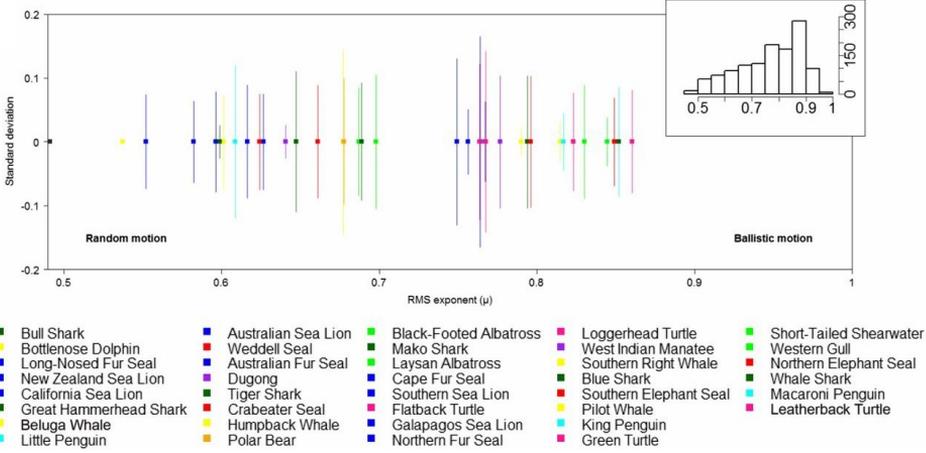


Figure 3.2: Scaling of the root mean square displacement length with time.

The exponent μ_i for each species i was obtained after fitting to a power with the shape $d_{RMS} \sim T^\mu$, where d_{RMS} is the root mean square displacement length and T is the time window. We obtained μ averaging over individuals, and the y -axis depicts the standard deviation of that average. Inset: histogram of the number of individuals with a specific value of μ . Source: Ref. [113].

The movement of the species was associated with a superdiffusive regime, as the root mean square displacement length $d_{RMS}(T) = \sqrt{\langle d^2(T) \rangle}$, where $d(T)$ was measured according to Eqs. (2.2) and (2.3), scaled as a power of time, $d_{RMS} \sim T^\mu$, with an exponent μ that for most of individuals fell in the range between 0.5 and 1. If the exponent $\mu_{i,j}$ was the scaling exponent associated to individual i of species j , the average exponent for species j was:

$$\mu_j = \frac{\sum_k \sum_i \mu_{i,k} \delta_{kj}}{n_j} \quad (3.1)$$

where δ_{kj} is the Kronecker delta, taking value 1 when $k = j$ and 0 otherwise, and n_j is the number of individuals of species j in the dataset. We averaged the

CHAPTER 3. INTERSPECIES MOVEMENT ANALYSIS

exponents over individuals belonging to a species, rather than aggregating all the observed displacements at the species level, in order to weight equally all the individuals, avoiding with this average the biases to the values corresponding to individuals with longer tracks.

The two limits of the observed exponents are characteristic for the ordinary Brownian ($\mu = 0.5$) and uniform motion ($\mu = 1$). We observed a great variability among species, with bull sharks having the closest exponent to the Brownian motion exponent, and leatherback turtles, which displayed a migration in the Atlantic Ocean, the closest to the uniform motion (Fig. 3.2).

Setting a time window $T = 1$ day, assumed to be a natural scale for movement analysis due to the circadian periodicity, we computed the displacement length as the great circle distance between the location of one individual at time t and at time $t + T$, following Eq. (2.2). For raw data, which was not uniformly sampled, we considered a resolution $\Delta t = 0.05$, such that the end location of a displacement starting at t was the location measured at the closest time to $t + T$ in the set of observations in the range $(t + T - \Delta t, t + T + \Delta t)$. After measuring the displacement lengths, we computed the probability density function (PDF) describing them, considering one PDF for each individual. In order to compare these PDFs, we introduced the coefficient of spread $CS = \frac{\langle d^2 \rangle}{\langle d \rangle^2}$, which is a dimensionless measurement of how broad the distribution is, with high values of CS associated with broad distributions with heavy tails, such as power-law or lognormal. For example, for an exponential distribution $PDF_{\text{exp}}(d) = \lambda e^{-\lambda d}$, $CS_{\text{exp}} = 2$, while for a lognormal distribution $PDF_{\text{ln}}(d) = \frac{1}{d\sigma\sqrt{2\pi}} e^{-\frac{(\log d - \mu)^2}{2\sigma^2}}$, whose moments are $\langle d^n \rangle = e^{n\mu + n^2\sigma^2/2}$, $CS_{\text{ln}} = e^{\sigma^2}$.

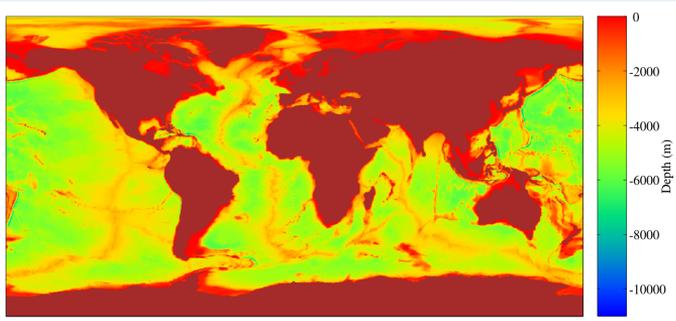
Boosted regression trees (BRT) [115], are a machine learning method that estimates the influence of multiple explanatory variables. This method consists of two algorithms. Firstly, the decision trees divide the explanatory variable space in regions with observed homogeneous observations in the response variable, and consider branching events that explain the connections between those regions. Each split point is related uniquely with one of the explanatory variables, and it represents a binary choice, *i.e.*, the split point is a question that has two answers, each one associated with one of the branches leaving from it. Secondly, the boosting method improves the accuracy of the models as, after one fit to a decision tree, it fits a new decision tree to the residuals of the previous fit, which are the changes in the response variable that were not explained by the previous decision fit. At the end, the boosted regression tree model will be a linear combination of all the obtained decision trees. In our measurement, BRT detected the species and the coastal affinity as the highest ranked explanatory variables for the individual CS , with relative importance of 74.0 and 21.5%, re-

3.2. GLOBAL AND INDIVIDUAL CONVERGENCE

spectively, identifying an interaction between these two variables (Fig. 3.3a). Specifically, individuals in open oceans were moving with patterns leading to low CS, which are associated with well-defined velocities, while we found a high variability in the movement patterns in the coast. Following this result, we grouped species in three main groups, according to their coastal affinity, and compared their displacement lengths PDF, finding that in general species with high coastal affinity had broader displacement lengths distributions than those with mixed and low coastal affinities (Fig. 3.3c). Finally, the species with mixed shelf affinity contained enough statistics both in coast and open oceans, such that we could compare the coefficient of spread associated with each environment, confirming the higher spread observed for movements located in open ocean (Fig. 3.3b).

Measuring the coastal affinity (CA)

We measured the coastal affinity of each individual as the fraction of observed displacements whose origin and destination were located in the coastal ocean. We considered coastal oceans as the land and the areas whose depths were smaller than 150 m. The depth was obtained from the General Bathymetry Chart of Oceans (GEBCO, <http://www.gebco.net>).



Global bathymetry. Land is depicted as dark red.

This analysis was complemented studying the turning angles displayed by the tracked individuals. The triangle formed by the origin and end points of two consecutive displacements defines a turning angle (Fig. 3.4a). Setting the same time window as in the displacement length analysis ($T = 1$ day), we analysed the distribution of angles. This distribution was peaked for straight and return angles, with significant differences between the behaviour displayed in open oceans and in the coast (Fig. 3.4b). While in open oceans the individuals were more likely to move in directed trajectories, with very few turning angles far

CHAPTER 3. INTERSPECIES MOVEMENT ANALYSIS

from zero degrees, close the coast the angles were distributed more uniformly. In fact, the peak for straight angles was higher than the return peak in the open ocean, while the return probability was higher than the straight displacement probability in the coastal environments (Fig. 3.4b). Despite the observed high variability, this difference was also detected in the exponents of the root mean square displacements, which were closer to uniform motion for open oceans (Fig. 3.4c). In fact, these differences between movement in open and coastal oceans were also observed in the southern elephant seals movement after splitting it between highly and low-occupied areas [79].

3.2. GLOBAL AND INDIVIDUAL CONVERGENCE

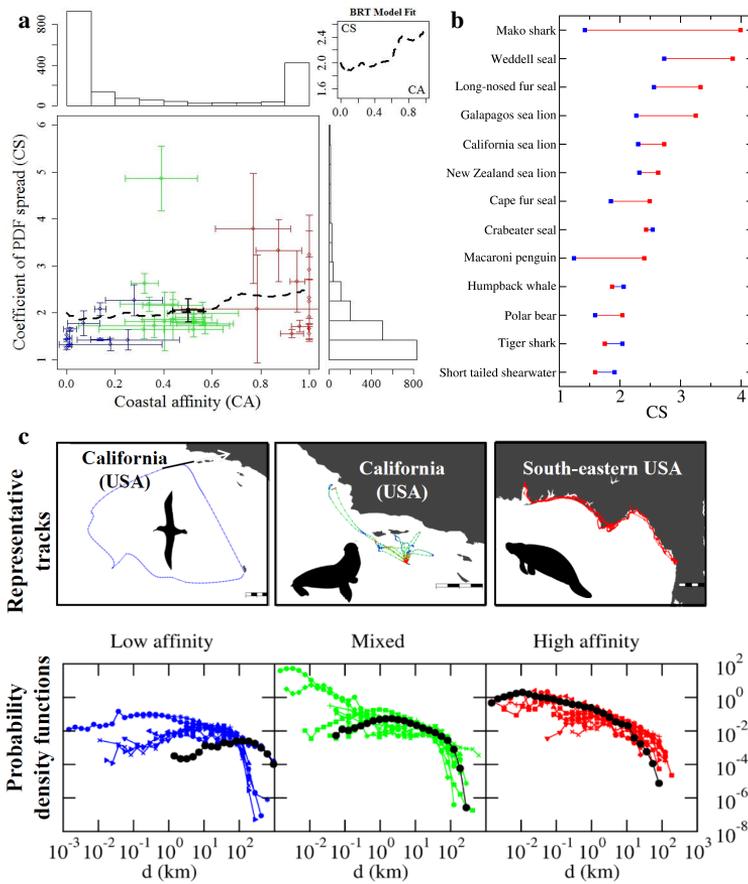


Figure 3.3: Observed displacement patterns in open and coastal oceans. a, Average values over individuals of the coefficient of spread for each species, for a given average coastal affinity. Error bars indicate standard deviation, while the colors indicate species with coastal affinities classified as low (blue), mixed (green) and high (red). **b,** Coefficients of spread measured for movement happening in the coast (red) and open oceans (blue), measured for species with mixed coastal affinity. **c,** Examples of the tracks for species with low (laysan albatross, blue), mixed (California sea lion, green) and high (West Indian manatee, red) coastal affinities, and displacement length distributions for the species in each group according to their coastal affinity. The distributions coloured in black are those associated with the species whose track is represented in the example. The tracking points are coloured in red and blue for points in the coast and in open ocean, respectively, and the black and white bar shows a scale of 100 km. Source: Ref. [113].

CHAPTER 3. INTERSPECIES MOVEMENT ANALYSIS

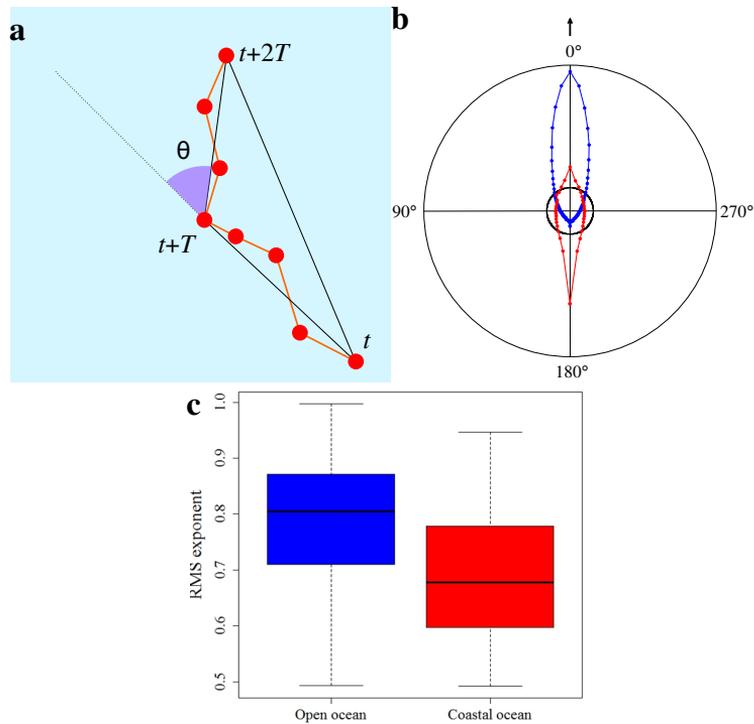


Figure 3.4: Global patterns on open and coastal oceans. **a**, An example of a track and the turning angle defined by two consecutive displacements. **b**, Distribution of turning angles for coastal (red) and open (blue) oceans. The black circle represents the uniform distribution and the arrow indicates the straight movement direction. **c**, The exponent of the root mean square displacement length displays a high variability, but it leads to higher values for individuals moving most of time in open ocean (blue), than for those moving in coastal ocean (red). Source: Ref. [113]

3.3 Analysis of movement similarities

The differences in the coefficient of spread of the displacement lengths could be explained firstly from the species to which the individuals belong to and, secondly, from the coastal affinity of the individual, identifying an interaction between these two variables. Next, we addressed the analysis at a species level, comparing the displacement lengths distributions across all the species.

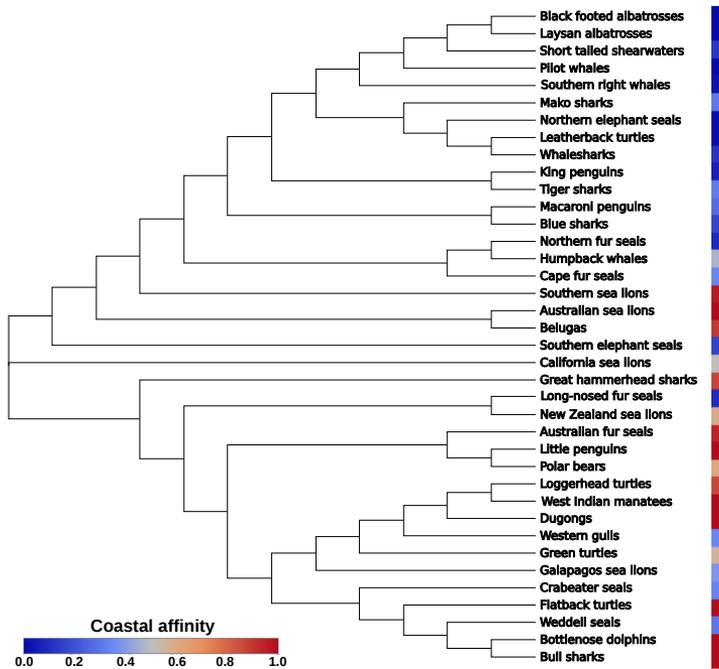


Figure 3.5: Analysis of the similarities between displacements observed in different species. Dendrogram obtained from the distance matrix after applying the neighbour joining method. The colors indicate the coastal affinity measured for each species. Source: Ref. [113].

We computed the distance between the PDFs describing the displacement lengths for each pair of species. This distance was the Kolmogorov-Smirnov d_{ij}^{KS} distance that, for a couple of set of samples i and j for the variable x , is the maximum distance between the cumulative density functions ($CDF(x)$), defined as probability

CHAPTER 3. INTERSPECIES MOVEMENT ANALYSIS

for an observation being smaller than x) associated with them, *i.e.*, $CDF_i(x)$ and $CDF_j(x)$:

$$d_{ij}^{KS} = \max | CDF_i(x) - CDF_j(x) | \quad (3.2)$$

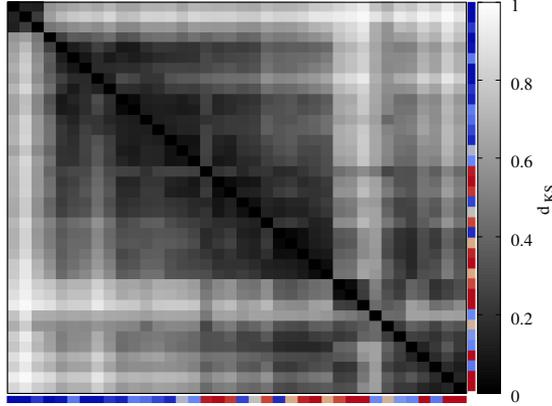


Figure 3.6: Kolmogorov-Smirnov distances between the sets of displacement associated with each species. The species have been ordered according to their positions in the dendrogram (Fig. 3.5). Source: Ref. [113].

This distance measurement, which is bounded between 0 and 1, takes the value 1 for distributions whose nonzero probabilities are associated with regions that do not overlap, and 0 for the distance between a distribution and itself, can be mapped to a similarity measurement, following $S_{ij} = 1 - d_{ij}^{KS}$. For measuring the Kolmogorov-Smirnov distance between species, we considered displacements in the tracked individuals belonging to each species. When this distance was calculated for all the pairs of species, we built a distance matrix, with the aim of grouping the species that had more similar displacement distributions. Several methods are available for inferring structures from distance matrices [116, 117]. We used neighbour joining [117], which iteratively selects the connection ij that represents the shortest distance and the elements i and j into a new element, recomputing the distances between the other elements k and the new one as the average of d_{ik}^{KS} and d_{jk}^{KS} ; this process is performed until there is just one remaining element in the matrix. These iterations can be visualized in a dendrogram, in which each iteration is a split point giving rise to two branches. Using the distance matrix whose entry ij was d_{ij}^{KS} , the neighbour joining method lead to a dendrogram that showed a split between two main branches, one being

3.3. ANALYSIS OF MOVEMENT SIMILARITIES

representative of the low coastal affinity and the other including most species that have intermediate and high coastal affinity (Fig. 3.5). Specifically, we did not identify a relationship between the phylogeny of these species and the shape of the tree, highlighting the importance of the coastal affinity as an explanatory variable. The Kolmogorov-Smirnov distance matrix, after ordering the species according to the dendrogram, with the most different in the extremes, showed three main blocks of species with low d^{KS} , with the first one associated with low coastal affinity flying birds, which were black footed albatrosses, laysan albatrosses and short-tailed shearwaters (Fig. 3.6), clustering together the pairs of nodes that have closer common roots.

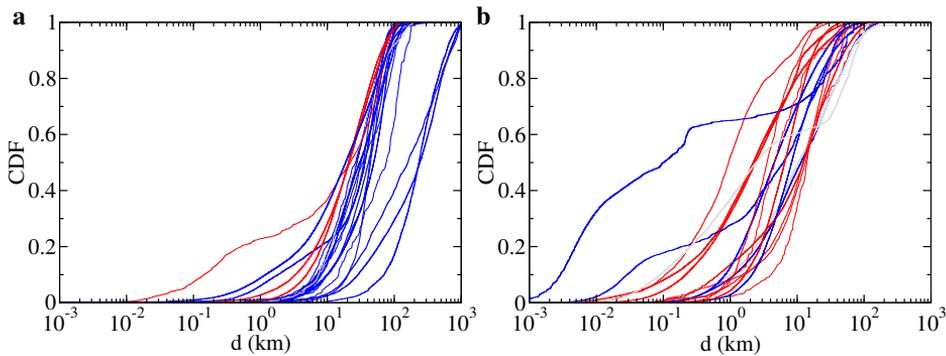


Figure 3.7: Comparison between the cumulative density functions of displacement lengths for species a, in the highest branch of Fig. 3.5 and b, in the lowest branch. The cumulative density functions corresponding to species with high coastal affinity ($CA \geq 0.6$) are depicted with red curves, grey for middle coastal affinity ($CA \in [0.4 - 0.6]$) and blue for low coastal affinity ($CA < 0.4$). Source: Ref. [113].

Finally, the plot of the cumulative distributions showed the differences between the two main branches, including displacement lengths that were typically high, with a low degree of spread, and associated with low coastal affinity in the species associated with the highest branch of the dendrogram (Fig. 3.7a), while smaller displacement lengths, with broader distributions, were observed in species moving in the coast (Fig. 3.7b).

From this analysis at species level, we focused on an analysis of movement similarities at individual level, that confirmed the species as the main driver of movement. We computed the Kolmogorov-Smirnov distances between the sets of displacement lengths observed for all the pairs of individuals, obtaining the similarity between their movement patterns. After setting a threshold on the

CHAPTER 3. INTERSPECIES MOVEMENT ANALYSIS

similarity, keeping the similarities higher than 0.5, we plotted the network of similarities, observing a high amount of high similarity links across individuals of the same species, and also confirming the influence of the environment, with many links across individuals that belong to different species but exhibit similar coastal affinities (Fig. 3.8).

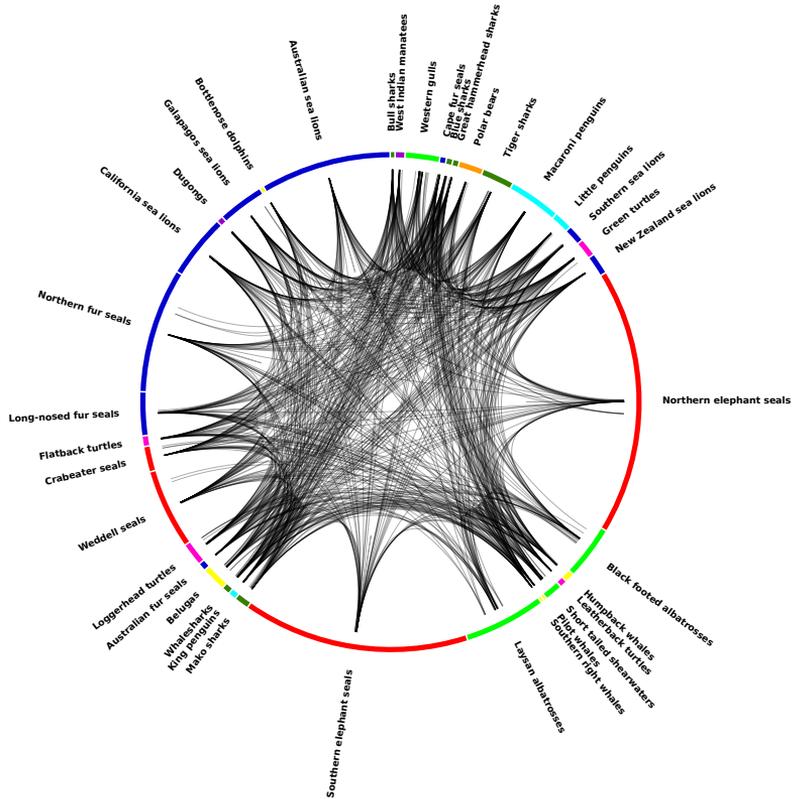


Figure 3.8: Network of individuals according to the similarity of their displacement lengths pattern. The links connect the pairs of individuals i and j with a similarity $S_{ij} > 0.5$. Individuals are grouped by species, with color depicting the nine guilds: cetaceans (yellow), eared seals (blue), flying birds (green), penguins (cyan), polar bears (orange), sharks (dark green), sirenians (purple), true seals (red), and turtles (pink). Figure generated with Circos [118].

3.4 Conclusions

Thousands of years ago, Aristotle wondered about the drivers of animal movement. Now, the availability of movement tracking datasets covering a significant part of the oceans and the development of both big data techniques and machine learning algorithms facilitate the aim of codifying the general drivers of marine animal movement. Specifically, we addressed this challenge following the suggestions of Nathan *et al.* for the development of a unified paradigm of animal movement research [119].

We collected a dataset including species from nine different guilds, exhibiting different locomotion modes, and living in regions with strong climatic difference, and detected two explanatory variables as the main drivers of movement, with a significant interaction between them. First of all, the species, as an internal driver, with a relative importance of 74%, explained the observed differences. Secondly, the environment was detected as another explanatory variable (22% of relative importance), showing that movement displayed in the coast was much more highly variable than movement observed in open oceans, where the displacement lengths distributions were narrower and the motion direction was strongly biased to the straight motion. We suggest that these features are related with the complexity of the habitats where the animals move: high complexity in the coastal habitats lead to highly variable displacement lengths, while the movement displayed in open oceans was described by a narrow range of displacements. In fact, the open oceans are a more uniform habitat, where the movement is constrained by oceanographic features, like fronts, eddies and currents [120, 121]. Another approach to the analysis of movement patterns was showing that the movement behaviour was leading to different modes related with the environment: while the environments with more abundant resources were leading to Brownian movement, the animals moving on regions with a lower resource availability displayed Lévy-like movement patterns [58].

One key question for the marine habitats conservation is how the anthropogenic impact on the coastal environments affects the movement behaviour of marine species. In fact, the human activities have effects in all the oceanic environments, and 41% of the oceanic area is strongly affected by anthropogenic impact [122]. For example, the routes described by vessels in the Arctic region are overlapped with the habitats of species like polar bears, beluga whales or narwhals, leading to possible disturbances, collisions or acoustic impact [123]. From our work, there are two relevant points to address how this issue may influence animal movement behaviour. On the one hand, the individuals show a capacity to adapt to the environmental condition despite the species they belong to, shifting

CHAPTER 3. INTERSPECIES MOVEMENT ANALYSIS

from frequently turning and highly variable movement patterns in the coastal oceans to more directed and well-defined velocity movement when they are in the open ocean. On the other hand, endangered coastal habitats may lead to the loss of key movement behaviours for the stability of the marine ecosystems.

Chapter 4

The global transit pattern described by vessel trajectories

Nowadays, more than 70% of the global trade value is transported through the sea, mainly by cargo and tanker vessels [124]. This trade requires suitable infrastructure to deal with the delivery processes in an efficient way, in terms of time and safety of the routes, leading to considerable amounts of money invested on it. From other point of view, vessels are moving through an ecosystem in which thousands of species live, some of them endangered, and these routes may affect deeply their populations [123, 125]. Hence, not only for economical reasons, but also for conservation, it is fundamental to describe the spatial patterns described by vessels in the seas and the oceans.

Previous studies have described the cargo ship movement with a complex network approach, considering every harbour as a node of the network, and the direct trips from one port to another as directed links (origin-destination), with repeated trips considered in the weight of each link [126]. This network, with 951 ports connected through 36,351 links, had an average degree (number of routes arriving and departing from one port) $\langle k \rangle = 76.5$, with an average clustering coefficient $\langle C \rangle = 0.49$ and an average shortest path length $\langle L \rangle = 2.5$, with the maximum shortest path length found between Skagway (Alaska) and Lampedusa (Italy) ($L_{max} = 8$). The degree distribution was reported to be broad, but not fitting to a power-law distribution (Fig. 4.1a), while the link weight distribution did, $P(w) \sim w^{-\mu}$, with $\mu = 1.71 \pm 0.14$ (Fig. 4.1b). The strength s of a node, computed as the sum of the weights into and out from it, scaled as a power law too, $P(s) \sim s^{-\eta}$, $\eta = 1.02 \pm 0.17$ (Fig. 4.1c). Finally, it was reported a power-law scaling between the strength of a node and its degree, with the average strength

CHAPTER 4. VESSEL TRANSIT PATTERNS

for a given degree scaling as $\langle s \rangle \sim k^{1.46 \pm 0.01}$ (Fig. 4.1d). These results represent a broad range of behaviours in the system: the power-law distribution of the link weight implies that there are very few highways, connecting pairs of ports very frequently, and then having a huge activity, and many connections that are less frequently used. In fact, the harbours show a broad level of hierarchy, including a few ports that register most of the trips, *i.e.* with a big strength, and most ports with a very low strength.

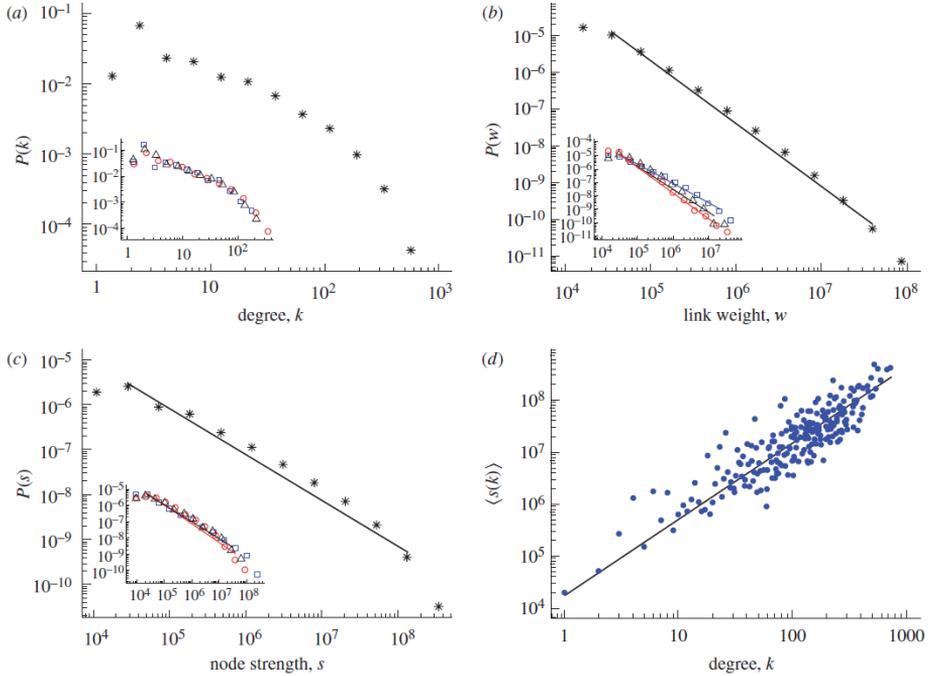


Figure 4.1: Properties of the global cargo ship movement network. **a**, Degree distribution $P(k)$. **b**, Link weight distribution $P(w)$, scaling as a power-law with exponent $\mu = 1.71 \pm 0.14$. **c**, Strength distribution $P(s)$, fitting to a power-law with exponent $\eta = 1.02 \pm 0.17$. **d** Average node strength $\langle s \rangle$ as a function of the node degree k , scaling as a power $\langle s \rangle \sim k^{1.46 \pm 0.01}$. The insets of **a**, **b**, **c** represent the distributions for container ships (squares), bulk dry carriers (circles) and oil tankers (triangles). Source: Ref. [126].

The broad variability of levels of activity in the harbours, measured as their strength, showed a hierarchy between them and suggests the presence of highways connecting them. In this chapter, from the spatial occupation patterns in

4.1. MODELS OF SHIP TRANSIT

the ocean, we will assess the presence of these highways, their properties and the hierarchy of harbours, comparing our results with null models of ship transit.

4.1 Models of ship transit

We introduce a simple model of ship transit in 2 dimensions, setting a source in the origin of coordinates $(0,0)$ and a set of destinations, which can be understood as the set of ports. We will consider two scenarios for the distribution of destinations in space, being located in a) a straight line from $(1,0)$ to $(1,1)$, and b) a circle of radius $R = 1$. In each trip from the source, a destination port is chosen at random. We assume that the particles are travelling following a straight line from the source to the destination, and our aim is to calculate the distribution of densities.

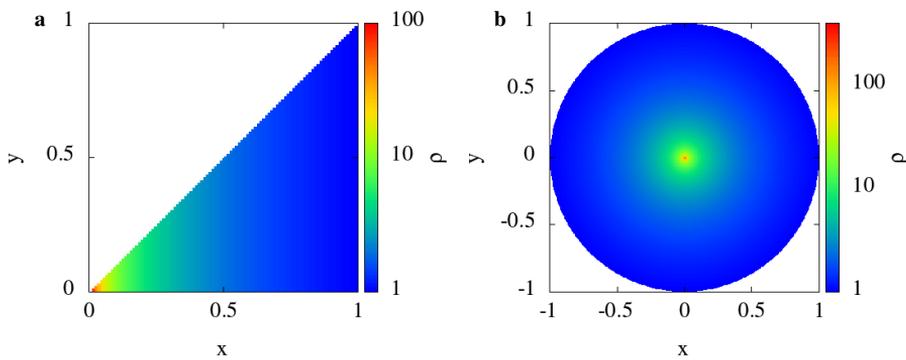


Figure 4.2: Null models of spatial spreading in two dimensions. The particles follow a straight line from the source, located in the origin of coordinates, to the destination, which is randomly located in **a**, a point at $(1, y)$, with $y \in (0, 1]$ and **b**, a circumference of radius $R = 1$.

In first scenario, we describe the positions in the space with Cartesian coordinates (x, y) , and the density $\rho(y|x)$, such that $\rho(y|x)dy$ is the probability of finding a trajectory crossing the vertical line between (x, y) and $(x, y + dy)$, and $\int_0^{\infty} \rho(y|x)dy = 1$. We consider a uniform probability for the destinations at $x = 1$, i.e., $\rho(y|x = 1) = 1$. As the vessels follow straight lines, for a given x , the maximum y for a trajectory is that associated with the vessels whose destination is $(1, 1)$, such that the locations with a non-zero density are those placed between

CHAPTER 4. VESSEL TRANSIT PATTERNS

$y = 0$ and $y = x$. Hence, the density takes the value $\rho(y|x) = \frac{\rho(y|x=1)}{x} = \frac{1}{x}$ for $y \leq x$ and 0 otherwise (Fig. 4.2a). The complementary cumulative density function Q for the density ρ , defined as the probability of finding an element with a density higher than ρ , is the ratio between surface covered by regions with density higher than ρ and the surface covered by regions with nonzero density:

$$Q(\rho) = \frac{A[\rho(y|x) > \rho]}{A[\rho(y|x) > 0]} \quad (4.1)$$

These surfaces can be calculated using the Heaviside step function $H(z)$, that takes the value 1 for $z > 0$ and 0 otherwise, leading to $A[\rho(y|x) > 0] = \int_0^1 \int_0^1 H(\rho(y|x) - 0) dx dy$ and, as $\rho(y|x) > 0$ between $y = 0$ and $y = x$, the Heaviside function modifies the limits of the integral: $A[\rho(y|x) > 0] = \int_0^1 dx \int_0^x dy = 1/2$. Analogously, we obtain the surface covered by regions with density higher than ρ : $A[\rho(y|x) > \rho] = \int_0^1 \int_0^1 H(\rho(y|x) - \rho) dx dy$ and, taking into account that $\rho(y|x) = \frac{1}{x}$, the values of x such that $\rho(y|x) - \rho > 0$, range between 0 and $1/\rho$, leading to $A[\rho(y|x) > \rho] = \int_0^{1/\rho} dx \int_0^x dy = \rho^{-2}/2$. Substituting in Eq. (4.1), we find $Q(\rho) = \rho^{-2}$. From $Q(\rho)$, the probability density function (pdf) is obtained taking the derivative, leading to $\text{pdf}(\rho) = 2\rho^{-3}$, with ρ ranging between 1 (at the destination) and ∞ (at the origin).

In the second scenario, we consider that the destinations are uniformly distributed along a circumference at a distance $R = 1$ from the origin. In this scenario, the density at a given point (x, y) depends on the polar coordinates $r = \sqrt{x^2 + y^2}$ and $\phi = \tan^{-1}(y/x)$, as the contours of equal densities have a constant value of r . At $R = 1$, the density $\rho(x, y)_{r=1} = \frac{1}{2\pi}$, such that the integral of the density along the circumference that includes all possible destinations is 1. In fact, $\rho(x, y) = 1/(2\pi r)$, such that integral of $\rho(x, y)$ in the circumference of radius r is 1. In this case, the area of the region with $\rho > 0$ is the area of a unit radius circle, $A[\rho(x, y) > 0] = \pi$, and the area of the region with a density higher than ρ is $A[\rho(x, y) > \rho] = \int_{\phi=0}^{2\pi} \int_{r=0}^1 H(\rho(x, y) - \rho) r dr d\phi$; the Heaviside function modifies the limits of integration, leading to r ranging between 0 and $1/(2\pi\rho)$: $A[\rho(x, y) > \rho] = \int_{\phi=0}^{2\pi} \int_{r=0}^{1/(2\pi\rho)} r dr d\phi = 1/(4\pi\rho^2)$. Hence, $Q(\rho) = 1/(4\pi^2\rho^2)$, that leads to $\text{pdf}(\rho) = \rho^{-3}/(2\pi^2)$, for ρ ranging between $1/(2\pi)$ (at the destination) and ∞ (at the origin).

These models provide us with several conclusions that may be useful for the analysis of the vessel transit empirical data: 1) uniformly distributed destinations and the transit to them in a two-dimensional space lead to occupancy distributions scaling as power-laws with exponent 3; 2) as we would obtain the same transit pattern if we set the origins in the line connecting (1,0) to (1,1) and

4.2. DATA DESCRIPTION

the destination in (0,0) in model a), we expect an increase of density close to the narrowing locations in the ocean, *i.e.*, canals, straits and regions connecting different oceans; 3) from model b), the density in the regions located close to the sources scales as the inverse of the distance to the source location.

4.2 Data description

The data used for this analysis reports the monthly number of unique vessels in grid cells of 0.25° both in latitude and longitude, from July 2010 to May 2015. For this analysis, we will aggregate the data corresponding to the recorded months and divide it by the number of months. This data also contains the category the vessels belong to, split in fishing, passenger, cargo, tanker and other. We analyse the distribution of densities ρ , which are calculated normalizing the number of events reported for a cell with the cell area (Fig. 4.3). The distributions of ρ for different categories are broad, with heavy tails that fit to power-law scaling with exponent α , $\text{pdf}(\rho) \sim \rho^{-\alpha}$, fitting to exponents that are smaller than those calculated for our models for uniform spatial transit (Fig. 4.4a, Table 4.1). Interestingly, the highest deviation from uniform spreading arises in passenger category, with a fitting exponent $\alpha_{\text{passenger}}=1.70$, which could arise from the different nature of this category, in contrast with other categories, like cargo. In fact, the exponents, which are smaller than those associated with the uniform spreading, can be related with the presence of highways that join the routes communicating the oceans, with an increase in the density close to the straits or man-made canals (Fig. 4.3). However, when the density is normalized, dividing by the average density in each category, there is a functional shape that describes all the categories (Fig. 4.4b). We analyse the same density distributions for each month, finding three different regimes: the most frequent has an exponent $\alpha \approx 2.3$, while the other two have $\alpha \approx 3$ and $\alpha \approx 2.0$, with this one being frequent in the summer months for Northern hemisphere (Fig. 4.5).

Normalizing the cell density ρ_i^c of cell i for category c with the average over all the cells, $\langle \rho^c \rangle$ for each category, $\tilde{\rho}_i^c = \rho_i^c / \langle \rho^c \rangle$, we select the category which has the highest relative density for detecting the main activity in each cell. This leads to coherent fishing areas, that are out of the main oceanic highways, allowing to split those highways into cargo, tanker and passenger highways, with a strong presence of the vessels categorized as other in the coast (Fig. 4.6).

CHAPTER 4. VESSEL TRANSIT PATTERNS

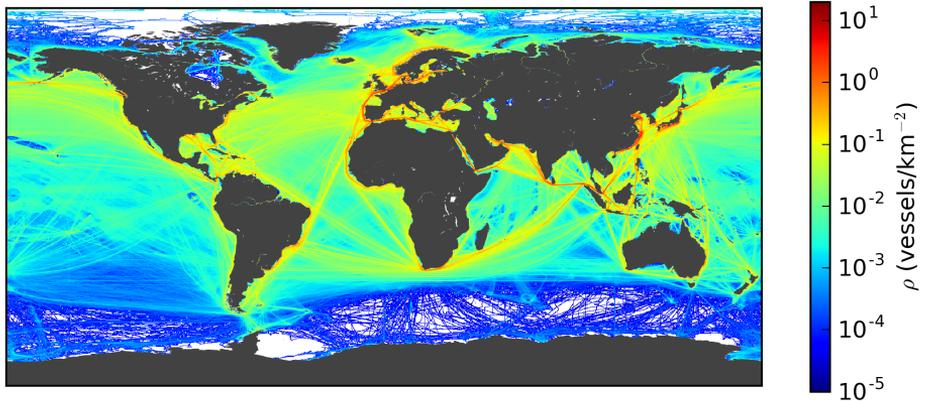


Figure 4.3: Spatial pattern of vessel transit in oceans.

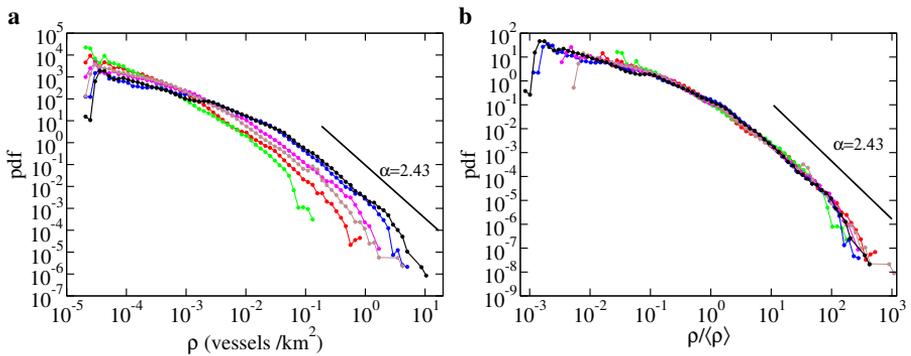


Figure 4.4: Distributions of vessel densities in the oceans. **a**, Density distributions for different categories. **b**, When normalized with the average density $\langle\rho\rangle$, the distributions collapse to a general shape. Colours represent different vessel categories: fishing (red), passenger (green), cargo (blue), tanker (magenta), and other (brown). The total number of vessels in a cell is represented in the black curve.

4.2. DATA DESCRIPTION

Category	Exponent (α)	ρ_{\min} (vessels/km ²)
Total	2.43	0.19
Fishing	2.66	0.026
Passenger	1.70	1.5×10^{-4}
Cargo	2.42	0.12
Tanker	2.22	0.029
Other	1.98	3.9×10^{-3}

Table 4.1: Fits to the distribution of densities for different vessel categories. The values are obtained from a power-law fit to the data (see Appendix A), fitting the minimum density ρ_{\min} for observing a power-law exponent and the exponent α .

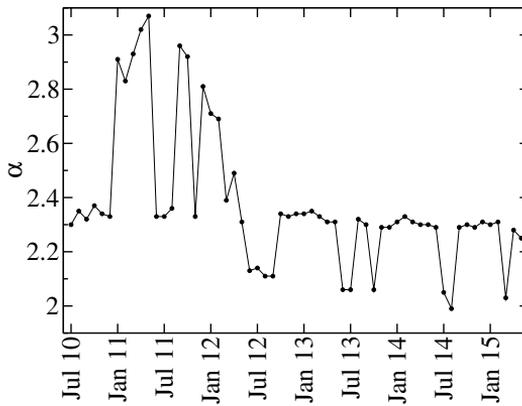


Figure 4.5: Scaling of the density distributions for different months. Exponents α obtained for the fit $\text{pdf}(\rho) \sim \rho^{-\alpha}$.

CHAPTER 4. VESSEL TRANSIT PATTERNS

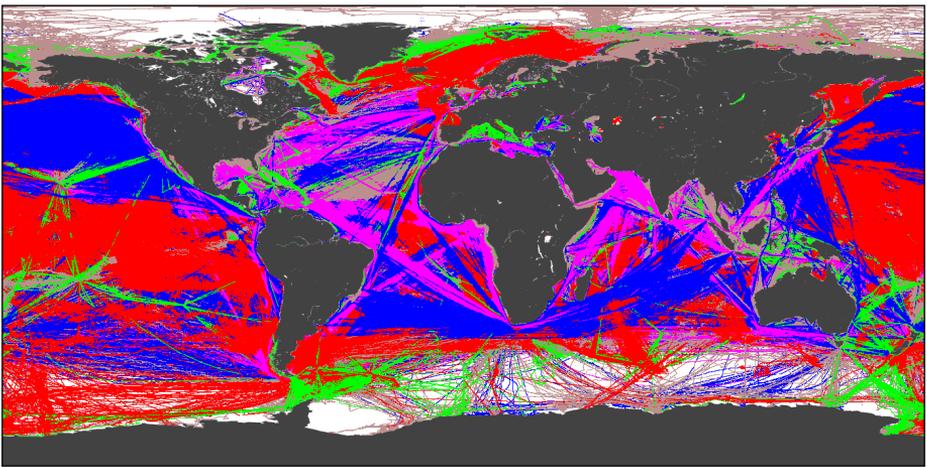


Figure 4.6: Strongest relative density for each cell. The colors indicate the category: fishing (red), passenger (green), cargo (blue), tanker (magenta) and other (brown).

4.3 Ports hierarchy

We characterize the activity distribution of ports, considering those included in the World Port Index¹, assigning to each port an activity A that is the measured density ρ in its cell, and considering the different ports located in the same cell as a single port (Fig. 4.7). The distribution of activities shows a broad variability,

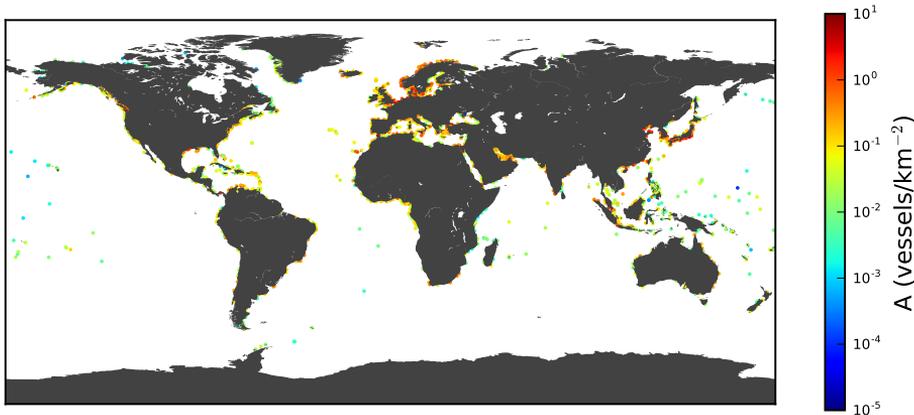


Figure 4.7: Spatial location of the ports with detected activity. Symbols are centred in ports coordinates, and the sizes are proportional to the activity, in logarithmic scale.

with the activity that joins all the categories (correspondent to the total number of vessels), having a heavy tail that follows a power-law $\text{pdf}(A) \sim A^{-\beta}$, with exponent $\beta = 2.18$. Similar exponents, around the value $\beta = 2$, are found for fishing, cargo and tanker vessels, while for passenger and the category other the exponents are around $\beta = 3$ (Table 4.2).

The comparison of these exponents with well-known processes that lead to power-law distributions suggests some mechanisms underlying these broad distributions. For exponents around $\beta = 2$, we consider the Zipf's law for cities, which states that the population P of a city of rank R behaves as $P(R) \sim R^{-1}$. For a high enough number of cities, this behaviour implies distribution with a power-law tail with exponent $\beta = 2$. The distribution scaling as $\text{pdf}(P) \sim P^{-2}$ arises as the steady distribution for a growth process in which all the cities have

¹World Port Index, from Marine Safety Information:
https://msi.nga.mil/MSISiteContent/StaticFiles/NAV_PUBS/WPI/PUB150.ZIP

CHAPTER 4. VESSEL TRANSIT PATTERNS

Category	β	A_{\min} (vessels/km ²)
Total	2.18	0.20
Fishing	1.61	3.1×10^{-3}
Passenger	3.01	1.9×10^{-2}
Cargo	2.13	0.15
Tanker	1.83	2.2×10^{-2}
Other	3.22	0.23

Table 4.2: Fits to the distribution of port activities. The fits consider a power-law tail, fitting the beginning of the power-law scaling, at A_{\min} , and the exponent β .

a growth rate, which at each time step is sampled from a universal distribution of finite average and variance [127]. However, this mechanism does not lead to the exponents for the categories other and passenger, in which other processes may be involved.

4.4 Clustering and highways

As stated previously, the pattern displayed by the vessels traffic in the oceans shows some highways, which are typically the shortest paths between the production centres and the destinations, crossing through interoceanic straits and canals (Fig. 4.3). First of all, we define a network in which the nodes are the cells that were visited at least by one vessel ($\rho > 0$), and connecting each cell with its four nearest neighbours (up, down, right and left). For our analysis, we consider the giant component of that network, disregarding the small connected components, for example the Caspian sea. Then, we define as control parameter the threshold density ρ_{th} , removing those cells with a density that is smaller than our threshold density. For this set-up, we find the critical point for the percolation process in the system, looking for the global maximum of both the susceptibility and the second largest connected component, which appears at $\rho_{\text{th}} = 2.88 \times 10^{-2}$ vessels/km² (Fig. 4.8a). At this point, the largest connected component splits in two, which represent the highways associated, in the largest connected component, with Indian and Atlantic Oceans, and in the second largest connected component, with the Pacific Ocean (Fig. 4.8c). In this critical point, the network has 724 connected components, whose areas S are distributed following a power-law pdf(S) $\sim S^{-\tau}$, with $\tau = 2.06$ from $S_{\min} = 1.80 \times 10^3$ km² (Fig. 4.8b). Then, the cluster sizes are distributed according to a power-law for sizes bigger than

4.4. CLUSTERING AND HIGHWAYS

$S_{\min} = 1800 \text{ km}^2$, a size that is approximately the size of Tenerife Island in the Atlantic Ocean (2034 km^2). The measured exponent is compatible with the site percolation critical exponent in 2D regular lattices, where the control parameter is the probability of occupying a site p , a cluster is defined as the set of cells that have at least one nearest neighbour connection within the set, and at the critical percolation point p_c the cluster sizes are distributed following a power-law with exponent $\tau_{2D} = 187/91 \approx 2.05$ [128].

Next, we study the pattern displayed by low density clusters instead of the high density highways, analysing the complementary percolation diagram that results from keeping in the network the cells with densities ρ satisfying $\rho < \rho_{\text{th}}$. In contrast to the ordinary percolation, in the complementary percolation the limits of low and high threshold densities lead to many small connected components and one large connected component, respectively. When the threshold density is decreased from the maximum density in the system, our system experiences a fragmentation transition, where the giant component is fragmented into several smaller components (Fig. 4.9a). In this case, the global maxima for the susceptibility and the size of the second biggest cluster do not correspond to the same value of ρ_{th} , and we choose for our study the highest threshold density among these two, $\rho_{\text{th},c} = 9.50 \times 10^{-2} \text{ vessels/km}^2$, as it represents the first fragmentation that the system experiences, which is a local maximum for the susceptibility and a global maximum for the second largest connected component. In this case, the largest connected component is fragmented in two clusters at the critical point, with the largest cluster including Pacific, Indian and Southern Atlantic Oceans, while the second largest connected component includes the Northern Atlantic and the Arctic Oceans (Fig. 4.9c). At $\rho = \rho_{\text{th},c}$, the cluster sizes are distributed according to a power-law for sizes $S > S_{\min} = 1.54 \times 10^3 \text{ km}^2$, $\text{pdf}(S) \sim S^{-\tau}$, with $\tau = 1.52$ (Fig. 4.9b).

CHAPTER 4. VESSEL TRANSIT PATTERNS

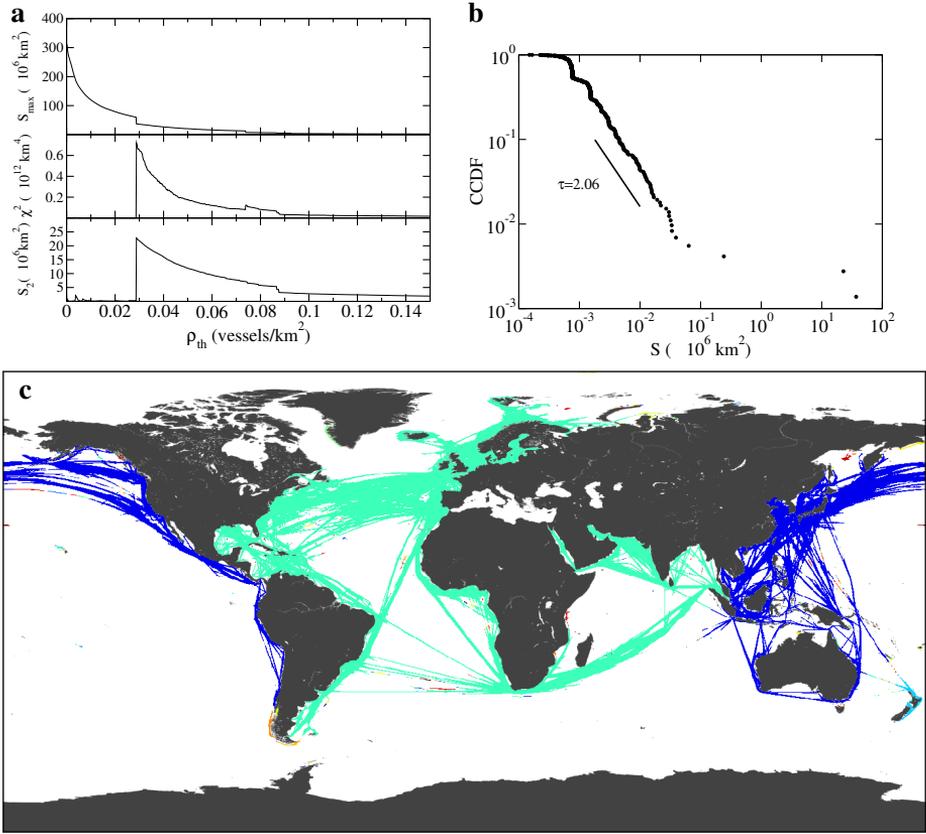


Figure 4.8: Percolation in the spatial network described by the vessels densities. **a**, Percolation diagram keeping cells satisfying $\rho \geq \rho_{th}$. The three graphs show, from top to bottom, the size S_{max} of the largest connected component, that is our order parameter, the square susceptibility χ^2 and the size S_2 of the second largest connected component, showing these last two a maximum that defines a critical point for $\rho_{th} = 2.88 \times 10^{-2}$ vessels/km². **b**, Complementary cumulative distribution function (CCDF) of cluster sizes S at $\rho_{th,c}$, described by power-law CCDF $\sim S^{-\tau+1}$ with $\tau = 2.06$ for $S > 1.80 \times 10^3$ km². **c**, Map of the clusters that result from keeping the cells with a density $\rho \geq \rho_{th,c}$, with different colours representing different clusters. The largest cluster includes the highways in the Atlantic and Indian oceans (light green), while the cells in the second largest component are located in the Pacific ocean (blue).

4.4. CLUSTERING AND HIGHWAYS

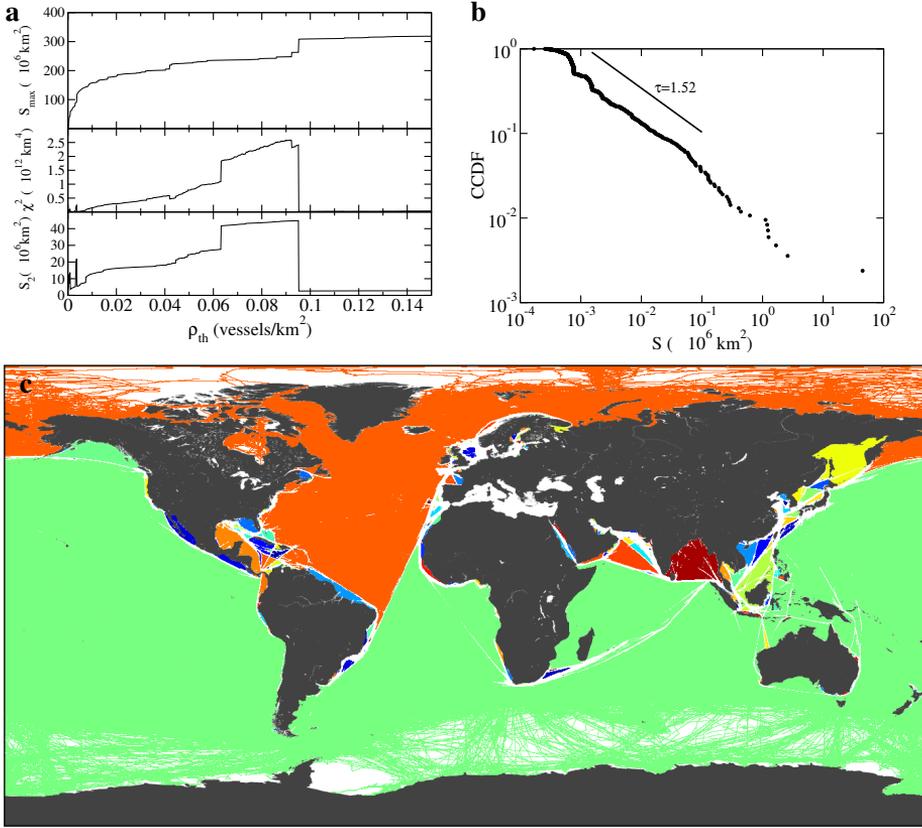


Figure 4.9: Complementary percolation in the spatial network described by the vessels densities. **a**, Percolation diagram keeping cells satisfying $\rho < \rho_{th}$. The three graphs show, from top to bottom, the size S_{max} of the largest connected component, that is our order parameter, the square susceptibility χ^2 and the size S_2 of the second largest connected component, showing these last two a maximum that defines a critical point for $\rho_{th,c} = 9.50 \times 10^{-2}$ vessels/km². **b**, Complementary cumulative distribution function (CCDF) of cluster sizes at $\rho_{th,c}$, described by power-law CCDF $\sim S^{-\tau+1}$ with $\tau = 1.52$ for $S > 1.54 \times 10^3$ km². **c**, Map of the clusters that result from keeping the cells with a density $\rho < \rho_{th,c}$, with different colours representing different clusters. The largest cluster includes Pacific, Indian and Southern Atlantic oceans (green), while the second largest includes cells in the Northern Atlantic and Arctic oceans.

4.5 Conclusions

In this study, we characterized the global spatial pattern emerging from the vessels movement. First of all, we analysed the spatial patterns associated with two models of regular spatial spreading, leading to a density distribution that decayed with the inverse cube density, with this exponent 3 related with the dimensionality of the problem. The pattern that we analysed displays a smaller exponent, $\alpha = 2.43$, which is smaller due to the presence of low dimensionality highways connecting the most important harbours in a hierarchy that is also distributed according to a power-law. However, several features extracted from the uniform spreading model were compatible with the empirical data, like the increase of density when in the narrow transit areas (see for example the regions around the Cape of Good Hope and the Strait of Malaca in Fig. 4.3), or the increase of density next to the sources, as observed in Southeast Asia. We identified the kind of traffic that is going through each region in a remarkable way, after normalizing the density associated with each category with the average density for that category, finding some coherent regions in which one activity is dominant, like for example the remarkable tanker traffic around the main oil reserves, or the passenger routes connecting islands with mainlands. Finally, we studied the structure of the highways and the clusters between them, describing them with a percolation and complimentary percolation analysis, respectively. The highways display a pattern with a distribution of cluster sizes compatible with 2D ordinary percolation, which recently has been also reported in the scaling of the lake sizes in the Earth [129], where it is proposed as a mechanism the asymmetry between area and abundance. In our case, we consider a distribution of densities in a 2D surface and that mechanism of asymmetry may work, as the bigger that two different clusters covering a fraction of Earth's surface are, the more likely to collide and become a single cluster they are.

The study by Kaluza *et al.* [126] introduced the analysis of networks describing the movement of ships. Here, although in our dataset the specific trajectories described by the vessels were not specified, we studied the spatial properties of the transit pattern described by vessels, classifying the regions according to the category with strongest relative density. We find hierarchies of ports, and also detect the routes with the highest transit in the oceans, with special importance for future assessments of how the marine animals are influenced by the marine traffic. We also show that the percolation transition points are different if we perform the usual percolation, *i.e.*, keeping the cells with $\rho \geq \rho_{th}$, or the complementary ($\rho < \rho_{th}$). We understand this different behaviour considering that our system is composed of highways, which are the sets of cells with highest density, including paths that are followed consistently, *i.e.* the

4.5. CONCLUSIONS

density inside a highway does not significantly vary. Hence, if we start looking at high values of ρ_{th} , *i.e.*, fragmenting for $\rho < \rho_{th}$ and assembling for $\rho \geq \rho_{th}$, the clusters which include parts of a highway are easy to be fragmented, leading to a soon fragmentation. However, the highways have some hierarchy, such that not all the highways are visited with the same frequency, such that the system needs to include cells with smaller densities such that the ordinary percolation ($\rho \geq \rho_{th}$) leads to a percolating cluster.

Part III

Dynamical processes in mobile networks

Chapter 5

Disease spreading in temporal geometric graphs

The development of modern means of transport, especially the air transport system, allows passengers to travel along distances of thousands of kilometres in a few hours, increasing the connectivity of the global network describing human mobility [130]. These means of transport provide societies with faster levels of diffusion of information, ideas, goods and help, but also they provide pathogens with highways to propagate with velocities considerably higher than those associated with natural propagation of an infection in space. In recent years, several infections have been able to spread at threatening velocities and, although they were controlled after several months, it was impossible to keep them confined in small areas, as infected individuals may remain asymptomatic, and then they cannot be effectively isolated before they travel, leading to a fast human-induced trip of the infection.

There are simple mathematical models that reproduce processes like disease spreading [131, 132], facilitating the inference of the mechanisms driving the disease spreading phenomena, allowing for forecasting and estimating the influence of variables like the connectivity or the level of vaccination in the most influential areas in the spreading. At population level, several disease spreading models have been studied in empirical interaction networks [133, 134].

Studies of disease spreading in complex networks help us determine who are the influential spreaders and then allow for more effective vaccination campaigns, giving priority to these individuals [135]. The most intuitive way of selecting which individuals should be vaccinated with priority is the number of connections that they have, which in network theory is known as their degree. However,

CHAPTER 5. DISEASE SPREADING IN MOBILE NETWORKS

the global structure of the network is very important, and other measures, like the number of shortest paths connecting each pair of nodes and going through an individual, known as betweenness centrality, may be more relevant to avoid effectively the disease spreading.

The combination of mobility and disease spreading leads to a reaction-diffusion process, where the disease spreading is associated with the reaction, while the mobility is responsible for the diffusion. In this framework, several studies have assessed the role of transport networks on disease spreading [70, 136, 137], considering the traveller flux between two locations. These studies are based on metapopulation approaches. In a metapopulation model, the nodes are locations on space, *i.e.*, cities, towns and villages, which contain populations performing trips to other nodes according to the traveller fluxes. Recently, interesting phenomena has been reported for both abstract and empirical mobility networks [138], showing the counter-intuitive effect of mobility being detrimental for disease spreading.

Motivated by the continuous effort made on understanding the mechanisms that drive disease spreading on systems allowing for individual mobility, we study in this chapter this reaction-diffusion process in an agent-based model where the agents are allowed to move in space, assessing both the critical point for the outbreaks to occur and the influence of the velocity on the outbreak size. Hence, we will follow the individual specific trajectories instead of focusing on the fluxes that arise from them. First of all, we will explain the analysed model in section 5.1, introducing the topology and the dynamics and highlighting the relevant features reported in the literature. Secondly, we will present our results on this model, studying the limit of infinite velocity and then focusing on finite velocities.

5.1 Model

5.1.1 Topology: temporal random geometric graph

The topology in which the disease will spread is defined by a network that is both spatial [63] and temporal [139]. Specifically, we consider N nodes that have a position in a square space of side L , and they interact with the nodes that are located at a distance smaller than an externally fixed interaction distance d ; then, each frame is described by a random geometric graph (RGG) [140, 141]. After interacting, the nodes move with a velocity \mathbf{v} , which has module v and its direction is randomly set and fixed as initial condition. The sequence of all the

5.1. MODEL

frames defined by RGGs will be the temporal random geometric graph that, due to the geometrical definition of the interactions, is spatial, and also temporal due to the time evolution.

The interactions in the temporal random geometric graph are reflected in the adjacency matrix $\mathcal{A}(t) = \{A_{ij}(t)\}$ whose entries are defined by

$$A_{ij}(t) = H(d - d_{ij}(t)) \quad (5.1)$$

where $H(x)$ is the Heaviside step function, taking value 1 for $x \geq 0$ and 0 otherwise, and $d_{ij}(t) \geq 0$ is the Euclidean distance between the locations of nodes i and j at time t . For the measurement of the distance, we will consider periodic boundary conditions. The expected average degree $\langle k \rangle$ of a RGG is calculated from the expected number of nodes that a randomly picked node will find in the circular region of radius d :

$$\langle k \rangle = (N - 1) \frac{\pi d^2}{L^2} \quad (5.2)$$

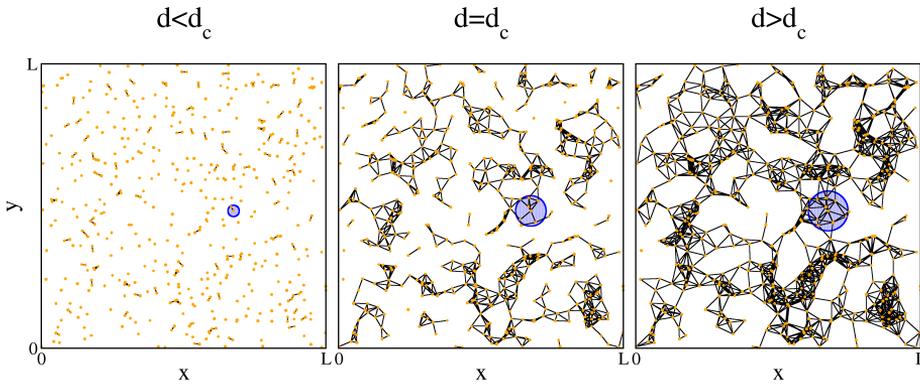


Figure 5.1: Examples of random geometric graphs. Orange dots represent the nodes, while black lines are the links and the blue dotted surfaces show the interaction area of a specific node. For the same spatial distribution of the nodes, the RGG is composed of many infinitesimally small connected components ($d < d_c$), several connected components whose sizes are distributed according to a power-law distribution ($d = d_c$) and a finite giant component ($d > d_c$).

In a static RGG ($v = 0$), several scenarios arise as d is varied. For $d \geq d_{\max} = \frac{L}{\sqrt{2}}$, the system is described by a fully connected network, in which every node interacts with every other [140]. As d is decreased, the interactions become

CHAPTER 5. DISEASE SPREADING IN MOBILE NETWORKS

local, and for lower values of d the network starts breaking apart in several connected components due to a percolation transition (Figs. 5.1,5.2). Finally, there is a critical interaction distance d_c below which the size of the largest component tends to zero in the thermodynamic limit. In two dimensions, the critical connectivity for a RGG with periodic boundary conditions is $\langle k \rangle_c = 4.52 \pm 0.01$ [140], which in our system, with $N = 4096$ and $L = 1280$ (setting the same density of particles $\sigma = \frac{N}{L^2} = 2.5 \times 10^{-3}$ as in Ref. [72]), substituting in Eq. (5.2), corresponds to $d_c \approx 24.0$.

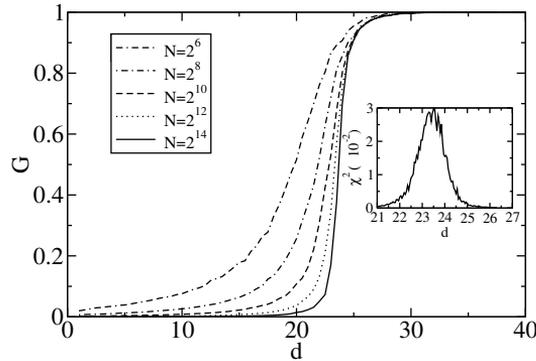


Figure 5.2: Percolation diagram in a two-dimensional random geometric graph. Size of the largest connected component G as a function of the interaction radius, for different system sizes, keeping the same density of nodes $\sigma = \frac{N}{L^2}$. Inset: the square susceptibility $\chi^2 = \langle G^2 \rangle - \langle G \rangle^2$ has a maximum at the critical interaction radius.

We use a simple mobility model with a single parameter, the module v of the velocity. Every node moves a distance v between two consecutive steps according to the following equations:

$$\begin{aligned} x_i(t+1) &= (x_i(t) + v \cos \theta_i) \bmod L \\ y_i(t+1) &= (y_i(t) + v \sin \theta_i) \bmod L \end{aligned} \quad (5.3)$$

where the pair $(x_i(t), y_i(t))$ is the position of node i at time t , $x_i(t=0)$ and $y_i(t=0)$ are random numbers uniformly distributed in the interval $[0, 1)$, and θ_i is a uniform random number in the interval $[0, 2\pi)$. Interesting phenomena have been reported for this simple mobility model in dynamics associated with synchronization [73] and game theory [76] like, for example, the divergence of the characteristic synchronization time for intermediate velocities.

5.1.2 Dynamics: Susceptible-Infected-Recovered

We will consider the classical model Susceptible-Infected-Recovered (SIR) with synchronous update [132]. An infected node transmits the disease to its susceptible neighbours with probability p , and it goes to the recovered state with probability p_r that, in our model, we will set to 1, meaning that the nodes remain in the infected state for one time step. Then, the probability that a susceptible node j becomes infected at time t is given by

$$p(S \rightarrow I)_{jt} = 1 - (1 - p)^{\sum_k A_{jk}(t-1)I_k(t-1)} \quad (5.4)$$

where $I_k(t) = 1$ if node k is in the infected state at time t and 0 otherwise. In our simulations, we will consider an initial condition where one node is in the infected state and the rest are in the susceptible state.

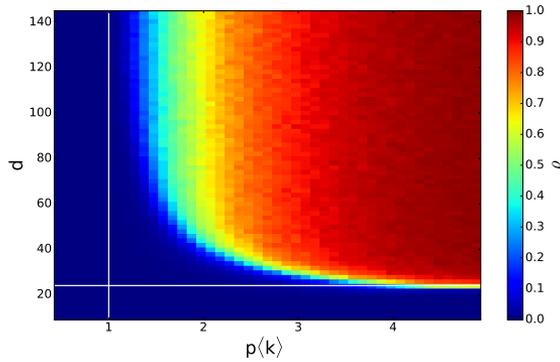


Figure 5.3: Disease spreading in a static random geometric graph. Diagram showing the fraction ρ of recovered at the end of the dynamics for different values of $p\langle k \rangle$ and d . The white horizontal line represents d_c , while the white vertical line depicts the limit $p\langle k \rangle = 1$.

Random rectangular graphs are similar to RGG, but they consider a rectangular space instead of square [142]. In random rectangular graphs, there is an epidemic critical point in the infection probability p_c separating the disease-free and the endemic phases, which increases linearly with the inverse of the average degree [143]. Considering that a RGG is one particular case of the random rectangular graph, in which the size lengths of the rectangle are the same, we expect a similar behaviour in both topologies; in fact, for RGG with $d \gg d_c$, the critical point is $p_c = 1/\langle k \rangle$, as reported in Ref. [143]. However, close to the critical percolation distance, $d \gtrsim d_c$, we observe that the critical point p_c is different from the inverse average degree (Fig. 5.3).

CHAPTER 5. DISEASE SPREADING IN MOBILE NETWORKS

The mean-field solution is valid when we assume a homogeneously mixed system [144]. In a network with average degree $\langle k \rangle$, the following equations describe the time evolution of the fraction of susceptible $S(t)$, infected $I(t)$ and recovered $R(t)$:

$$\begin{aligned}\dot{S} &= -p\langle k \rangle SI \\ \dot{I} &= -p_r I + p\langle k \rangle SI \\ \dot{R} &= p_r I\end{aligned}\tag{5.5}$$

This system of differential equations can be solved for an initial condition in which a small fraction of nodes is infected and the rest are susceptible nodes: $I(t=0) = \epsilon$, $S(t=0) = 1 - \epsilon$ and $R(t=0) = 0$, with $\epsilon \ll 1$. From this initial condition, setting $p_r = 1$, the fraction of susceptible nodes will decay following $S(t) = e^{-p\langle k \rangle R(t)}$. Using $S(t)$, and considering that at the end of the dynamics $I_\infty = 0$ and $S(t) + I(t) + R(t) = 1$, the fraction of recovered for $t \rightarrow \infty$, which is our order parameter ρ , is given by

$$\rho = 1 - e^{-p\langle k \rangle \rho}\tag{5.6}$$

The disease-free state $\rho = 0$ is always a solution of Eq. (5.6), while there is a non-zero solution for $p > \frac{1}{\langle k \rangle}$, arising from a second order phase transition at $p_c = \frac{1}{\langle k \rangle}$.

5.2 Disease spreading in mobile networks

5.2.1 Infinite velocity limit

In the infinite velocity limit, the time scales of the movement are much smaller than the time scales associated with the disease spreading dynamics. In this regime the topology is described by a temporal network that, at each time step, has the same statistical properties (number of nodes, dimensions of the space, interaction distance), but is uncorrelated with the network in the previous time step, *i.e.*, the positions of the nodes are randomly rearranged.

5.2. DISEASE SPREADING IN MOBILE NETWORKS

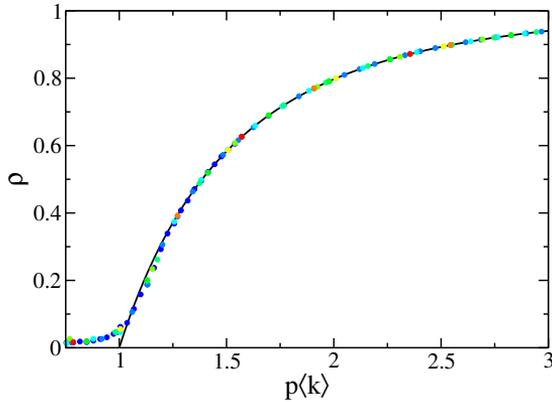


Figure 5.4: Disease spreading in the infinite velocity regime. Fraction of recovered nodes ρ as a function of the normalized transmission probability $p\langle k \rangle$. The dots depict the results of the simulations for interaction distances that range from $d = 20$ (blue dots) to $d = 100$ (red dots), with colours varying continuously between these limits. The curves associated with different values of d collapse to the homogeneous mixing solution, shown with a solid line.

In this case, for a given p , the fraction of recovered at the end of the dynamics ρ displays a universal curve for all the different values of $d > d_{\min}$, when the transmission probabilities p are normalized dividing by the inverse of the average degree (Eq. 5.2). The critical point is $p_c = \frac{1}{\langle k \rangle}$ and the values that $\rho(p)$ takes are described by the mean-field solution, associated with homogeneous mixing, given by Eq. (5.6) (Fig. 5.4).

This result is in agreement with the reported results for annealed networks, which are temporal networks described by a sequence in which, at every time step, the network has the same statistical properties, but is uncorrelated with the network in the previous time step, showing that the relevant parameter for calculating the epidemic threshold in the transmission probability is the average degree. In our case, it highlights the influence of the interaction distance and the density of particles in the epidemic threshold, as these are the variables that determine the average degree (Eq. 5.2).

While in the static case it is not possible to find macroscopic outbreaks for interaction distances that are smaller than the critical interaction distance $d_c = 24.0$, in this case it is possible to find it for smaller distances due to the particle mixing effects, which also lead to a smaller critical point than in the static case for $d > d_c$. In fact, there is a critical p_c such that for $p > p_c$ there are macroscopic

CHAPTER 5. DISEASE SPREADING IN MOBILE NETWORKS

outbreaks if $\langle k \rangle \geq 1$ is satisfied, taking into account that p is a probability and then it is bounded, $0 \leq p \leq 1$. This sets a boundary for the interaction radius, d_{\min} , such that for $d \leq d_{\min}$ there are no outbreaks, independently of the value of v , and for $d > d_{\min}$ there is a velocity above which the outbreak sizes start growing. Specifically, using Eq. 5.2, $d_{\min} = \frac{L}{\sqrt{(N-1)\pi}} \approx 11.3$.

Summing up, in the range $0 < d \leq d_{\min}$ the disease does not spread, while for $d_{\min} < d \leq d_c$ there is a critical value for v such that there are macroscopic outbreak sizes for $v > v_c(d)$, and for $d > d_c$ there are macroscopic outbreaks $\forall v$. With this general idea, we will analyse the scenario for finite velocities in the next section.

5.2.2 Finite velocities

We analyse the effect of finite velocities in our system, with the particles moving in random directions according to Eq. (5.3). First of all, we characterize the behaviour of the system for the limit case $p = 1$ (Fig. 5.5a). As introduced in the previous section, for $d < d_{\min}$ there are no macroscopic outbreaks for any particle velocity. However, for $d_{\min} < d < d_c$, the outbreak sizes grow with the velocity from a threshold velocity $v_c(d)$. Interestingly, the small velocities lead to a detrimental effect on the outbreak size for $d > d_c$, first decreasing to a minimum and then growing towards the limit of infinite velocity.

Although we expected that slow motion leads to larger outbreaks than in the static case, we found the opposite behaviour for $p = 1$, in a region of the phase diagram where $d \gtrsim d_c$ and the velocities are small. The phase space $v - p$, for an interaction radius $d = 30$ that is slightly higher than d_c , shows that the critical point p_c increases with low velocities, and the outbreak size decreases (Fig. 5.5b). For high velocities, $p_c \langle k \rangle$ approaches 1 as $v \rightarrow \infty$, matching the results shown in the previous section.

Interestingly, the counterintuitive regime, which we also observed in 1D mobile RGG, is a consequence of the changes in the shape of the spreading front, which are facilitated by the mobility. Setting $p = 1$, in the static case, and infecting initially a randomly chosen particle, the border of the region that includes the infected and the recovered nodes is always composed of infected nodes, without links between recovered and susceptible nodes. However, in the mobile case there is mixing in the front. That may change the spreading process.

We quantify the mixing measuring the number of links of the infected nodes: the links between infected and susceptible nodes (IS links), infected and infected (II links) and infected and recovered (IR). In fact, the IS links are those through

5.2. DISEASE SPREADING IN MOBILE NETWORKS

which the infection is transmitted, so we will compare the number of IS links with the number of IR links. When the network is static, there is an initial transient, after which there is a stationary period in which the ratio between the number of IS links and the sum of IS and IR links is slightly higher than $1/2$, a result that is expected for an expanding front: if the infected nodes are located at D from the original seed, the number of nodes at $D + d$, which will be in susceptible state and will have IS links with the nodes in the front, are slightly higher than those at $D - d$, which will be in recovered state and will have IR links with nodes in the front. However, when the particles move, this stationary value is smaller, indicating that when the nodes are moving with small velocities, the infected nodes are more likely to be connected to recovered nodes, rather than to susceptible (Fig. 5.6a). When the velocity is further increased, the global mixing effects appear, making the initial transient longer, with most of the nodes being infected in this transient.

A simple example illustrates the mechanism underpinning this dynamics. Focusing on a 1D system, we consider a spreading front in a system of particles in which half of them do not move and the rest move with velocity $2v$ (Fig. 5.6b, centre). At $t = 1$, there are two options: 1) if the particle that is now at the border between the front and the set of susceptible nodes move with $v = 0$, the epidemic will spread as in a static network, 2) if that particle is moving, it will approach the front, having a smaller number of links with the susceptible nodes that have $v = 0$. Hence, comparing the static and the mobile systems in these two updates, the mobile system (Fig. 5.6b, bottom) will infect a smaller number of nodes than the static one (Fig. 5.6b, centre). Specifically, if d is above but close to d_c , this effect may lead to an earlier end of the dynamics, reducing the size of the outbreak.

CHAPTER 5. DISEASE SPREADING IN MOBILE NETWORKS

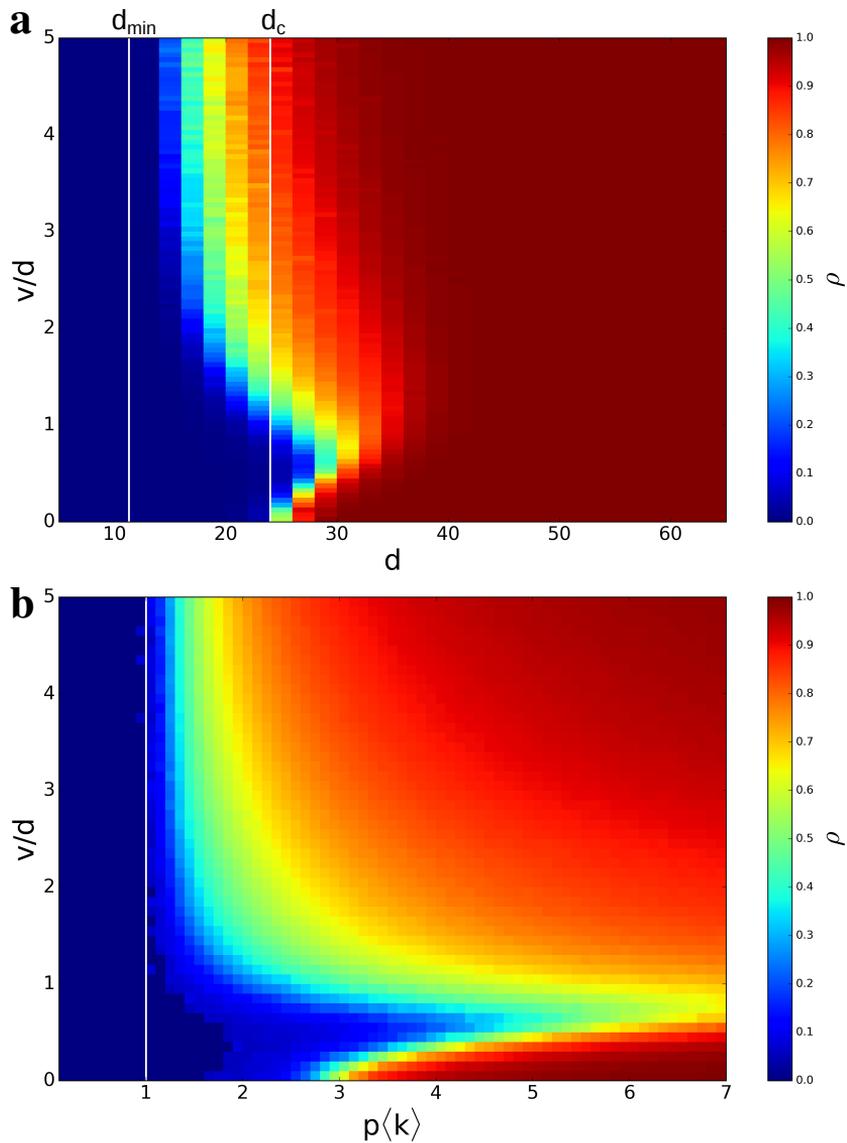


Figure 5.5: Disease spreading in mobile networks with finite velocity. **a**, Phase diagram $v - d$ for $p = 1$. The white lines indicate d_{\min} and d_c . **b**, Phase diagram $v - p$ for $d = 30$. The white line shows the limit $p\langle k \rangle = 1$.

5.2. DISEASE SPREADING IN MOBILE NETWORKS

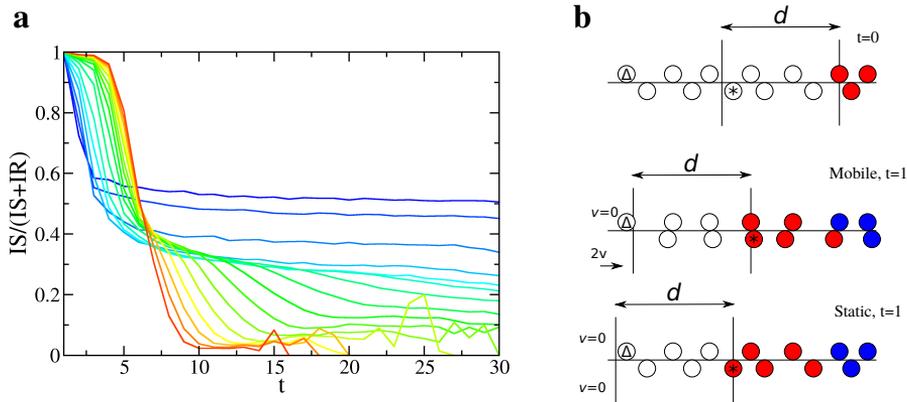


Figure 5.6: Mechanism for a regime with a detrimental effect for the contagion caused by the mobility. **a**, Time evolution of the ratio between the number of Infected-Susceptible (IS) links and the sum of IS and the number of Infected-Recovered (IR) links. Colours represent different velocities, from the static network ($v = 0$, blue) to an intermediate velocity $v = d$ (cyan), with $\Delta v = 0.2d$, and from $v = d$ to $v = 5d$ (red) with $\Delta v = 0.4d$. This ratio has been measured on simulations with $d = 30$ and $p = 1$. **b**, Example showing how the low velocities lead to smaller amounts of infected with $p = 1$ in 1D, showing susceptible (white), infected (red) and recovered (blue) nodes. At $t = 0$ (top), the node labelled with a star is the furthest susceptible that can be reached by the infection. At $t = 1$, if the network is static (bottom), all the nodes between the starred and that labelled with a triangle are infected, including this last one. However, if the nodes move with a small velocity (centre), with those depicted over the axis being static and those under it moving with velocity $2v$, the starred node will not infect the triangled.

5.3 Conclusions

Several works have matched empirical data of infection processes with reaction-diffusion models, like disease spreading processes in populations of mobile individuals [145, 146]. Other studies have focused on temporal networks, showing how they can lead to a slowing down on diffusion processes, in comparison with the corresponding aggregate network [147]. However, the role of the velocity has not previously been studied. We have shown that increasing the velocity, in some cases, leads to a detrimental effect on the disease spreading process, affecting to smaller fractions of the population. In fact, there is a range of velocities in which the effects of mobility on the outbreak sizes can strongly change. Specifically, the main difference between a static network and a mobile network with low velocity is the shape of the spreading front, which in the mobile case includes some links between recovered and susceptible nodes that, in the case of an interaction radius being above but close to the critical radius, will facilitate breaking the spreading network.

The infinite velocity regime is another point of strong interest, as this is the typical regime that we would expect for human interactions, with very few people located at a distance smaller than the interaction radius, which represents the maximum contagion distance, and the chance of moving much further than the interaction distance in the characteristic recovery time scales. This qualitative features, translated to our model, imply $d < d_c$, $v \rightarrow \infty$ and $p_r = 1$, parameters that are covered by our study, showing how, although we are below the percolation threshold, the disease is able to spread, infecting finite fractions of the population for high enough velocities.

Chapter 6

Cooperative disease spreading in contact and mobile geometric graphs

After studying a single infection spreading in mobile random geometric graphs, the complexity of the problem can be increased taking into account the presence of several diseases in our dynamical system. In fact, multiple interacting diseases lead to a broad variety of scenarios [148], analysing competitive [149] or cooperative effects [150], *i.e.*, cross immunity and host weakening, respectively.

In this chapter, we will focus on the scenario that considers two cooperative diseases, considering a recently introduced model that includes two infection processes, where the secondary infections, which are those experienced by hosts that suffered in the past or are currently suffering from one infection, and are exposed to the other infection, are more likely to occur than primary infections [150]. In fact, the cooperative interaction between diseases, which is known as coinfection, has been observed in several pairs of diseases, for example in the interaction of tuberculosis with Spanish flu [151], or between HIV and other diseases [152]. Firstly, we will review several models of coinfection explaining the phenomena they lead to and the underlying mechanisms. Secondly, we will analyse the basic model in static random geometric graph (RGG), and then we will study it in the mobile framework, focusing on the role of velocity. Finally, for concluding the chapter, we will bridge the two main parts of this thesis (data analysis and modelling), analysing the effects of data-driven simulations of coinfection processes in an empirical contact network.

6.1 Models of coinfection

6.1.1 Partial coinfection in mean field

In this section we will review the coinfection model introduced by Chen *et al.* [150]. This model considers two diseases, \mathcal{A} and \mathcal{B} , that are spreading in a system, with both diseases displaying Susceptible-Infected-Recovered (SIR) dynamics [132]. Then, the particles can have nine different states (Table 6.1).

		\mathcal{A}		
		S	I	R
\mathcal{B}	S	<u>S</u>	A	<u>a</u>
	I	B	AB	aB
	R	<u>b</u>	Ab	<u>ab</u>

Table 6.1: Possible states for a system with two spreading diseases \mathcal{A} and \mathcal{B} under SIR dynamics. Underlined states are those allowed at the end of dynamics, when none of the nodes are infected neither with disease \mathcal{A} nor disease \mathcal{B} .

This model assumes symmetric transmission rates between both diseases. The susceptible individuals become infected with a rate α , getting recovered with a rate that is set equal to 1, and then they can get a secondary infection with a higher rate $\beta = C\alpha$, being $C \geq 1$ the relative coinfection factor [150]. The case $C = 1$ implies independent spreading, while $C > 1$ is associated with cooperation between both diseases. Setting an initial condition in which the densities of infected nodes are symmetric ($[A]_0 = [B]_0 = \epsilon/2$), and assuming a symmetric behaviour of the diseases, we define S as the fraction of susceptible nodes, $P = [A] + [a] = [B] + [b]$ as the fraction of nodes recovered or infected from/by one disease, *i.e.* those whose future infections will be secondary, and $X = [AB] + [Ab] + [A] = [AB] + [aB] + [B]$ as the fraction of nodes that are infective. The evolution of these three variables is described by the following dynamical system:

$$\begin{aligned}
 \dot{S} &= -\alpha SX \\
 \dot{P} &= (\alpha S - \beta P)X \\
 \dot{X} &= (\alpha S + \beta P)X - X
 \end{aligned}
 \tag{6.1}$$

6.1. MODELS OF COINFECTION

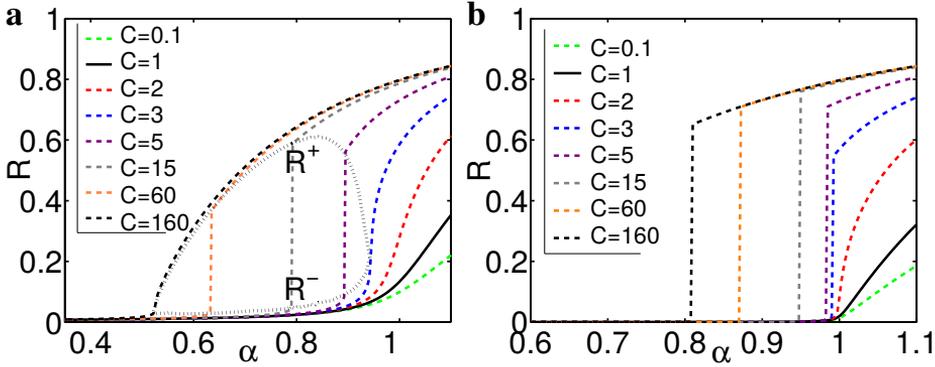


Figure 6.1: Partial coinfection in mean field. Fraction of recovered nodes (from \mathcal{A} , \mathcal{B} or both) at the end of the dynamics as a function of the primary infection rate α , with the initial condition **a**, $\epsilon = 0.005$ and **b**, $\epsilon = 10^{-4}$. The curves R^+ and R^- represent the two values of R at the epidemic threshold $\tilde{\alpha}(C)$ at the highest and lowest branch, respectively, such that if $R^+ - R^- > 0$ there is a jump at the epidemic threshold. Source: Ref. [150].

In the case $\beta = \alpha$ (*i.e.*, $C = 1$) the system is composed by two diseases that are independently spreading, and there is a critical transmission rate $\alpha_c = 1$, such that for $\alpha > \alpha_c$ the outbreak grows, leading to a finite fraction of the population that is recovered at least from one disease at the end of the dynamics. This fraction, defined as $R = 1 - S_\infty$, will be the order parameter. The numerical integration of Eq. (6.1) with an initial condition $S_0 = 1 - \epsilon$, $X_0 = P_0 = \frac{\epsilon}{2}$, shows that for $C < 2$ the behaviour is qualitatively similar to that displayed in SIR dynamics for one disease, with the outbreaks growing for $\alpha > \alpha_c = 1$ and with a scaling $R \sim (\alpha - \alpha_c)$, in a second order phase transition (Fig. 6.1). For the case $C = 2$, the transition still happens continuously at $\alpha_c = 1$, but the scaling is $R \sim (\alpha - \alpha_c)^{1/2}$. For $C > 2$, there are discontinuous jumps, with the transition point $\tilde{\alpha}$ ranging in $1/2 < \tilde{\alpha} < 1$. It has been discussed if these jumps represent first order phase transitions or not, as there is no coexistence of two different phases in the final state. In fact, it has been reported that this transition is not a first order but a spinodal transition, where, at the spinodal point $\tilde{\alpha}$, a locally stable branch (the branch with outbreaks of size going to zero) disappears when it collides with an unstable branch, with the system jumping to the high branch, which now will be globally stable [153]. Finally, the jump length tends to zero as $\tilde{\alpha}$ approaches $1/2$, having a continuous transition again. This limit of $\tilde{\alpha} \rightarrow 1/2$ for $\beta \gg \alpha$ represents a process in which the diseases grow together, such that

CHAPTER 6. COINFECTION IN MOBILE AND CONTACT NETWORKS

the fraction of primarily infected at one time become immediately secondarily infected, being able to transmit two diseases and then reducing α_c in a factor 2.

This cooperative framework leads to a rich scenario, where there is a minimum relative coinfection factor C , such that for $C > 2$ the epidemic threshold becomes smaller than 1, leading to a finite jump in the order parameter. The cooperation between diseases considered in this model has two basic reproductive ratios, one applying when the infinitesimally small fraction of infected, *i.e.*, when the probability of one individual getting a secondary disease is negligible and then the cooperation does not act, and the other that includes the cooperative effects, being higher than the first basic reproductive ratio, which will describe the system with a macroscopic fraction of infected. For $\alpha \rightarrow \tilde{\alpha}^-$, the fraction of infected nodes is not big enough such that the cooperation effects appear. However, for $\alpha \rightarrow \tilde{\alpha}^+$, the outbreak initially grows with the smaller basic reproductive ratio, but when it has a macroscopic size the basic reproductive ratio will be higher and then the disease outbreak will grow further.

6.1.2 Partial coinfection in complex networks

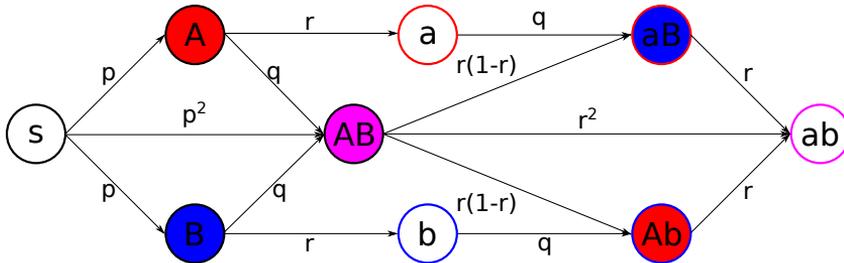


Figure 6.2: Scheme of the coinfection dynamics. Probabilities of transition, under the exposure to the suitable infective states, between different states are depicted in the center of the arrows connecting states. Primary infections, *i.e.*, when a state S gets infected, happen with probability p , secondary infections, which are those from active/recovered for/from one infection to active for the other, happen with probability q , and recovery, from active to inactive states, happens with probability r . Filling colors indicate active states, while stroke colors stand for recovered states (red for infection \mathcal{A} , blue for infection \mathcal{B} , magenta for both).

In analogy to the mean field model described by Eq. (6.1), a two-disease model was introduced and analysed by Cai *et al.* [154, 155] to study the effects of co-

6.1. MODELS OF COINFECTION

operation between two diseases that are spreading in complex networks. In this model, the control parameter is the primary infection probability p , setting externally the secondary infection probability q . While $q > p$ represents a cooperative interaction between the diseases, the case $q = 1$ represents the maximum cooperation scenario, implying that a previous infection with one disease strongly affects the host's immune system, such that she gets the secondarily infected when she becomes exposed to the other disease. In contrast, the case $q = p$ represents two independent diseases spreading in the network, and $q < p$ stands for a competitive scenario. The SIR dynamics, as in the previous section, leads to 9 different states (Table 6.1), and the transitions, under the suitable expositions, are depicted in Fig. 6.2.

We will show the results arising from this spreading dynamics in two models of networks. First of all, the Erdős-Rényi (ER) networks [156] of size N are random networks with an average degree $\langle k \rangle$. ER are built selecting each link with a probability $\frac{\langle k \rangle}{N-1}$. Secondly, we will analyse the cooperative disease spreading in 2D, 3D and 4D regular lattices.

In ER networks, the dynamics with maximum cooperation ($q = 1$) leads to a hybrid transition [154]. While in the spinodal transition there was no bistability, in the hybrid transitions there is coexistence between two branches, with the highest appearing discontinuously; however, the probability of a realization ending up in the highest branch is continuous and has system size effects. At the critical point p_c , the distribution of sizes of the infected clusters follow a broad distribution for low sizes, with an additional peak of higher size (Fig. 6.3a). The density of doubly recovered nodes at the end of the dynamics ρ_{ab} , considered as an order parameter, shows a broad region of coexistence between a high branch, which represents a macroscopic cluster of doubly recovered nodes, and a low branch, in which the fraction of doubly recovered is infinitesimally small (inset of Fig. 6.3b). The discontinuity of the transition is shown with the linear scaling of the size of the largest cluster of doubly recovered nodes m_{peak} with N , disregarding system size effects (Fig. 6.3c). However, if we consider another order parameter, which is the probability p_{ab} that the dynamics leads to the formation of a giant cluster of doubly infected nodes, *i.e.*, the probability of the dynamics ending up in the high branch, the system displays continuous growth from zero and has scaling with system size (Fig. 6.3b).

For regular lattices, the scenario changes with the dimensionality of the graph. On 2D lattices, for $q = 0.99$, the system exhibits a continuous phase transition for both p_{ab} and ρ_{ab} with the number of infected nodes $n(t)$ scaling as $n(t) \sim t^{0.5843}$ in the critical point, which corresponds to $p_c = 0.4503(1)$ (Fig. 6.4a) [154]. The scaling of the order parameters p_{ab} and ρ_{ab} corresponds to that associated with ordinary percolation in 2D lattices [157], with $p_{ab} \sim \rho_{ab} \sim (p - p_c)^\beta$, and $\beta \approx 5/36$.

CHAPTER 6. COINFECTION IN MOBILE AND CONTACT NETWORKS

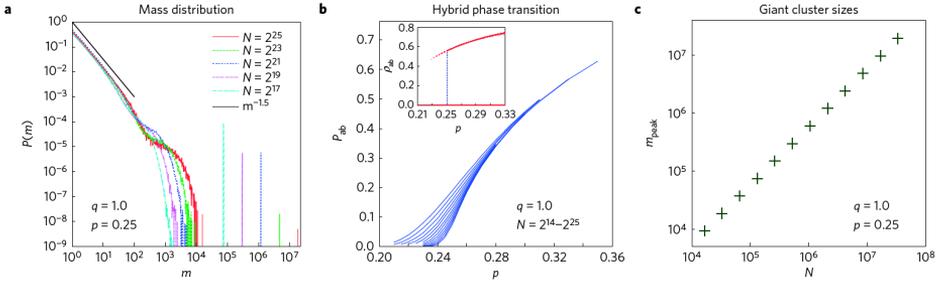


Figure 6.3: Cooperative disease spreading in Erdős-Rényi networks for average degree $\langle k \rangle = 4$. **a**, Distribution of sizes of the clusters composed of infected nodes at the end of the dynamics with $q = 1$, in the surroundings of the critical point $p_c = 0.25$. **b**, Probability p_{ab} for a giant cluster of doubly recovered being formed. Inset: fraction of doubly recovered nodes ρ_{ab} at the end of the dynamics. The dotted points represent finite size effects with realizations leading to the highest branch for $p < p_c$, that are less likely to appear as we approach the thermodynamic limit. **c**, Scaling of the size of the giant cluster of recovered nodes m_{peak} with system size N . Source: Ref. [154].

However, for 4D lattices the observed transition is hybrid, with ρ_{ab} undergoing a discontinuous transition, while p_{ab} grows continuously. In analogy with 2D lattices, the transition point is obtained analysing the time evolution of the number of infected nodes, which for $p \leq p_c$ decreases faster than any power-law, while for $p > p_c$ it initially decreases until it arrives to a minimum and then starts increasing due to the coinfection effects (Fig. 6.4b). Interestingly, in 3D lattices the behaviour depends on the updating rule, with asynchronous update leading to a continuous transition, while the dynamics with synchronous update displays a discontinuous jump in the order parameter ρ_{ab} .

This change of behaviour, between the scenario in 2D and 4D regular lattices, is argued to be related with the ratio between short and long loops in the network [154]. The proposed mechanism for the bistability between low and a high ρ_{ab} states considers that the infections follow different paths in the network and, after a long time they are able to meet, namely a "bottleneck"; since they meet, the coinfection effects start to appear, leading to avalanches of secondary infections (Fig. 6.4c). In fact, after the long time in which diseases follow different paths, the relative abundance of nodes with secondary infections increases (Fig. 6.4d). Hence, the order parameter p_{ab} is an indicator of the probability of both diseases meeting after a independent spreading bottleneck.

6.1. MODELS OF COINFECTION

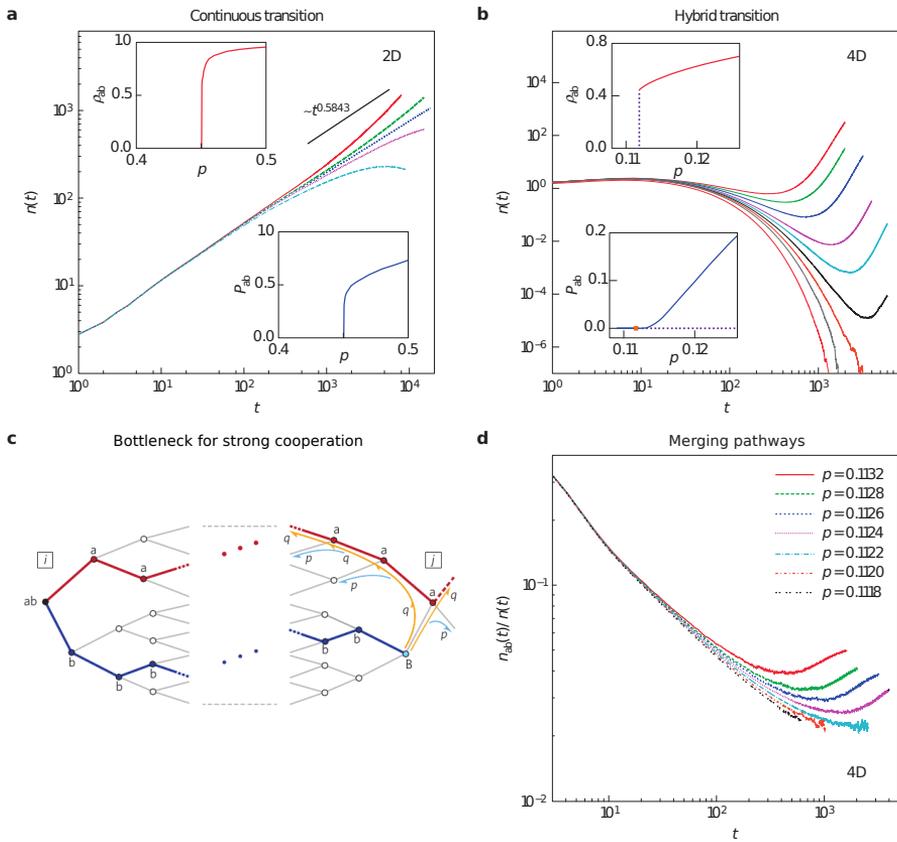


Figure 6.4: Coinfection dynamics in regular lattices. **a**, Time evolution of the number of infected nodes $n(t)$ in a 2D lattice, with a power law scaling at $p_c = 0.4503(1)$. Insets: behaviour of the two order parameters as a function of p , showing a continuous transition. **b**, $n(t)$ in the 4D lattice, showing a clear bottleneck after which, in the supercritical regime, the $n(t)$ grows. Insets: the order parameter ρ_{ab} shows a discontinuous phase transition, while p_{ab} grows continuously. **c**, Example of the bottleneck in complex networks leading to avalanche outbreaks. Both diseases spread along different paths and, after a long time, they are able to meet and generate long avalanches of secondary infections. **d**, Time evolution of the ratio between the number of secondarily infected nodes n_{ab} and the total number of infected nodes n in a 4D lattice. After the bottleneck, the fraction of secondarily infected nodes amongst the set of infected grows. Source: Ref. [154].

CHAPTER 6. COINFECTION IN MOBILE AND CONTACT NETWORKS

Under this bottleneck mechanism, the difference between 2D and 4D lattices is the ratio between short and long loops, which is higher for low dimensionality lattices. Moreover, this ratio is small in both 4D lattices and Erdős-Rényi networks, which both exhibit a hybrid transition under this dynamics. When there are many small loops, diseases are able to meet after short paths, producing microscopic coinfection effects and then leading to a second order phase transition. However, if they are able to spread along long paths without meeting, macroscopic differences will appear between the cases in which they meet or not, with the meeting probability growing continuously with the primary infection probability p .

Topology	Transition
Erdős-Rényi	H
2D lattice	C
3D lattice (Async)	C
3D lattice (Sync)	H
4D lattice	H

Table 6.2: Nature of the observed transitions for cooperative disease spreading in different topologies. H stands for hybrid transitions, while C stands for continuous.

6.1.3 Full coinfection in mean field

In the previous sections, it was considered an interaction in the diseases such that the infected/recovered population with/from \mathcal{A} (\mathcal{B}) was more likely to get the disease \mathcal{B} (\mathcal{A}) when it was exposed to it. However, from a theoretical point of view, a full picture in the interaction space will include interactions between the diseases for the infection capacity of the transmitter, considering that the doubly infected population transmits the diseases with a higher infection probability. Following the Susceptible-Infected-Susceptible (SIS) coinfection dynamics introduced in Ref. [158], we proposed a fully cooperative coinfection model [159], considering interaction effects both in the secondary infection of the primarily infected population and in the infection from the secondarily infected population (Fig. 6.5a).

Defining our control parameter as the primary infection rate α when the disease is transmitted from a susceptible node, we observe three kinds of hysteresis. The hysteresis cycles occur between an eradication rate $\alpha_e \leq 1$ and the epidemic

6.1. MODELS OF COINFECTION

threshold α_0 . Considering that there is a transcritical bifurcation at $\alpha = 1$, the observed hysteresis (Fig. 6.5b) are:

- Hysteresis of C type. This happens for $\alpha_e < \alpha_0 = 1$. There is a discontinuous phase transition at $\alpha = 1$ when we increase α , but for eradicating the disease, once the outbreak has been produced, it is necessary to lower the control parameter to α_e .
- Hysteresis of S_l type, with $\alpha_e < 1$ and bistability for $1 < \alpha \leq \alpha_0$. When α is increased, the outbreak starts growing at $\alpha = 1$, and there is a discontinuous phase transition at α_0 , while for eradicating that outbreak α needs to be decreased to α_e .
- Hysteresis of S_r type. Both α_e and α_0 are higher than 1, such that the eradication transition does not lead to a disease free state.

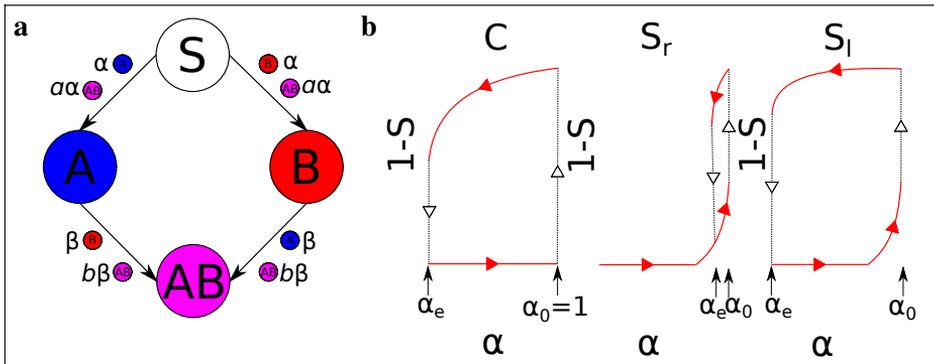


Figure 6.5: Scheme of the fully cooperative coinfection model and observed hysteresis. **a**, Primary infection probabilities α are modified by the factor a if the transmitter is doubly infected, while secondary infection probabilities $\beta = c\alpha$ are modified by the factor b when the transmitter is doubly infected. **b**, Observed hysteresis of the types C, S_r and S_l . The red continuous lines indicate the stable branches, while the black dotted lines represent the discontinuous jumps in the transition points, and arrows to the right (left) represent changes by increasing (decreasing) the control parameter. Source: Ref. [159].

In the SIS model, the susceptible nodes become infected with a rate α when they interact with the infected, and they recover with a rate, set to one without loss of generality. This model has a transcritical bifurcation at $\alpha_c = 1$, where the two fixed points of the system collide, with the disease-free branch $I^* = 0$ becoming unstable, leading the dynamics to the branch $I^* = 1 - \frac{1}{\alpha}$.

CHAPTER 6. COINFECTION IN MOBILE AND CONTACT NETWORKS

In this study, we consider a two-disease coinfection model, which leads to four different states: susceptible, singly infected with \mathcal{A} , singly infected with \mathcal{B} and doubly infected, and the fractions of the population in each state are denoted as S, I_A, I_B and I_{AB} , respectively. The equations describing these dynamics are:

$$\begin{aligned}\dot{S}(t) &= -\alpha S(I_A + aI_{AB}) - \alpha S(I_B + bI_{AB}) + I_A + I_B, \\ \dot{I}_A(t) &= \alpha S(I_A + aI_{AB}) - \beta I_A(I_B + bI_{AB}) + I_{AB} - I_A, \\ \dot{I}_B(t) &= \alpha S(I_B + bI_{AB}) - \beta I_B(I_A + aI_{AB}) + I_{AB} - I_B, \\ \dot{I}_{AB}(t) &= \beta I_A(I_B + bI_{AB}) + \beta I_B(I_A + aI_{AB}) - 2I_{AB}.\end{aligned}\tag{6.2}$$

where α is the infection rate for the interaction between a the susceptible and the singly infected population, a and b are the relative coinfection factors for doubly infected and singly infected transmitters, $\beta = c\alpha$ is the infection rate for secondary infections (Table 6.3 and Fig. 6.5), and the recovery rate has been set to 1.

Reaction	Rate
$S \xrightarrow{A} A$	α
$S \xrightarrow{B} B$	α
$S \xrightarrow{AB} A$	$a\alpha$
$S \xrightarrow{AB} B$	$a\alpha$
$A \xrightarrow{B} AB$	$\beta (= c\alpha)$
$B \xrightarrow{A} AB$	$\beta (= c\alpha)$
$A \xrightarrow{AB} AB$	$b\beta (= bc\alpha)$
$B \xrightarrow{AB} AB$	$b\beta (= bc\alpha)$

Table 6.3: Reactions allowed in this fully cooperative coinfection dynamics.

We analyse this system looking for the fixed points of Eq. (6.2), which we denote as $E^* = (S^*, I_A^*, I_B^*, I_{AB}^*)$. The non-trivial fixed points, defined such that $I_A^*, I_B^*, I_{AB}^* \neq 0$ satisfy

$$\begin{aligned}I_A^* = I_B^* &= \frac{a\alpha S^*}{1 + (2\alpha - 1)\alpha S^*}(1 - S^*) \\ I_{AB}^* &= \frac{1 - \alpha S^*}{1 + (2\alpha - 1)\alpha S^*}(1 - S^*)\end{aligned}\tag{6.3}$$

with $0 \leq S^* < 1$ and $0 \leq S^* \leq 1$, and S^* is the solution of

$$f(S, \alpha) = a(b-a)c\alpha^3 S^3 + [\alpha a(a-b)c - abc + 2a - 1]\alpha^2 S^2 + (\alpha abc - 2a + 2)\alpha S - 1 = 0\tag{6.4}$$

6.1. MODELS OF COINFECTION

In fact, for $a, b \geq 0, c > 0$, each S the solution of $f(S, \alpha) = 0$ leads to a unique $\tilde{\alpha}(S)$ in the interval $(0, 1/S]$. Then, the function $\tilde{\alpha}(S)$ is indicative for the equilibrium points of our system. Making the change of variable $T = \alpha S$ in Eq. 6.4 leads to

$$\tilde{\alpha}(T) = \frac{a(a-b)cT^3 + (abc - 2a + 1)T^2 + 2(a-1)T + 1}{acT[(a-b)T + b]} \quad (6.5)$$

with $0 < T \leq 1$. Defining $c_2 = 2/a$, for $a, b \geq 1$, the behaviour will be the following

- If $b \leq b_c = a + 1/2$, $\tilde{\alpha}(T)$ is concave upward on $(0, 1]$. In this case, for $c \leq c_2$, $\tilde{\alpha}(T)$ is strictly decreasing on $(0, 1]$, implying that the curve defining the equilibrium points in the $(\alpha, 1 - S)$ plane is strictly increasing, without any hysteresis. For $c > c_2$, $\tilde{\alpha}(T)$ has a local minimum, associated with an eradication point in α_e , leading to hysteresis of C type (Fig. 6.6a).
- For $b > b_c$, we introduce two variables:

$$c_1 = \frac{[a^{2/3}(2b-1)^{1/3} + (b-a)^{2/3}]^3}{ab^3} \quad (6.6)$$

$$c_{12} = \frac{2ab - 2a + b + 2\sqrt{a(2b-1)(b-a)}}{ab^2} \quad (6.7)$$

In the range $c \leq c_1$, $\tilde{\alpha}(T)$ is strictly decreasing on $(0, 1]$, such that there is no hysteresis in our system. However, $\tilde{\alpha}(T)$ the system has a minimum α_e and a maximum α_0 for $c_1 < c < c_2$. In fact, $\alpha_e > 1$ if $c < c_{12}$ (hysteresis of S_r type), $\alpha_e = 1$ at c_{12} , and $\alpha_e < 1$ for $c_{12} < c < c_2$ (hysteresis of S_l type in this range) (Fig. 6.6b). Finally, $\tilde{\alpha}(T)$ has a minimum, but no maximum in $(0, 1]$, for $c > c_2$, corresponding to hysteresis of C type.

The behaviour of the system will be completely characterized by the stability analysis of the equilibrium points, that are represented in the function $\tilde{\alpha}(T)$ in the interval $(0, 1]$:

- The disease-free state, which corresponds to $(S, I_A, I_B, I_{AB}) = (1, 0, 0, 0)$ is stable for $\alpha \leq 1$ and unstable otherwise.
- The trivial endemic fixed points (one disease with finite fraction of infected while this fraction is 0 for the other) are unstable for $\alpha > 1$.
- If a point is associated with the increasing part of the curve representing the endemic equilibrium points in the $(\alpha, 1 - S)$ plane, it is locally stable, and unstable otherwise.

CHAPTER 6. COINFECTION IN MOBILE AND CONTACT NETWORKS

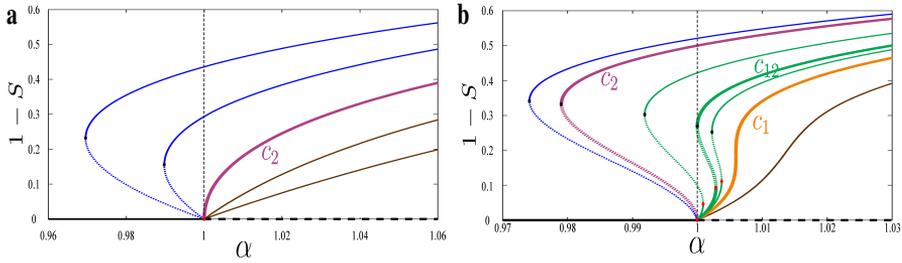


Figure 6.6: Bifurcation diagram in the fully cooperative coinfection model. a, Bifurcation diagram for $b \leq b_c$, where the system exhibits hysteresis of C type for $c > c_2$. **b,** Case $b > b_c$, with hysteresis of S_r type for $c_1 < c < c_{12}$, of S_l type for $c_{12} < c < c_2$ and of C type for $c > c_2$. Brown curves represent no hysteresis cases, while green curves are associated with hysteresis loops of the S type, and blue curves with hysteresis of C type. Source: Ref. [159].

We analyse the phase diagram of the system using dynamical systems theory, firstly setting a constant $a = 1$ and analysing the (b, c) space. For $b \leq b_c$, there is a codimension-2 point at $c_2 = 2/a$, in which the transcritical bifurcation at $\alpha = 1$, which makes the disease-free solution unstable, and a saddle-node bifurcation collide, such that for $c > c_2$ the system has a saddle-node bifurcation at $\alpha \neq 1$, $0 < 1 - S < 1$. Then, c_2 represents the border between the strictly increasing regime and the hysteresis of C type. However, for $b > b_c$, two saddle-node bifurcations are feasible for $c > c_1$ (Eq. 6.6) (for $c = c_1$ there is a cusp), such that first there is hysteresis of S_r type, then there is a crossover at $c = c_{12}$, leading to hysteresis of S_l type, and finally the saddle node associated with higher α collides with the transcritical bifurcation, with just one feasible (*i.e.* located in $0 < 1 - S < 1$) saddle-node bifurcation in the system, associated with hysteresis of C type (Fig. 6.7a). The same phenomenology applies to the description of the phase diagram with a constant $b = 3$, in the (a, c) space, setting $a_c = b - 1/2$ (Fig. 6.7b). In this case, for $a > a_c$ and $c > c_2$, the system will have a single feasible saddle-node bifurcation, while for $a < a_c$ there will be two saddle-node feasible bifurcations.

We can understand quantitatively the behaviour of the system. First of all, let us consider a choice of parameters (a, b) such that we see no hysteresis, hysteresis of S type and of C type if c is varied, like in Fig. 6.6b. In this case, small values of c lead to a continuously growing outbreak, as the level of cooperation does not have big effects on the system. If we keep increasing c (hysteresis of S type), we have a small outbreak branch and, when there is enough infected population

6.1. MODELS OF COINFECTION

in the system, the cooperation appears, with the main infection process being that coming from the population that is doubly infected, *i.e.* in AB state. If we keep on increasing c (hysteresis of C type), the infection to the singly infected part of the population dominates (it does not matter if it is from doubly or singly infected, as $c \gg a, b$), with a long hysteresis cycle that, once the outbreak has started, promotes cooperation, such that α has to be strongly lowered down such that the outbreak disappears. Secondly, we want to explain the reason for which higher differences between b and a lead to the appearance of hysteresis of S type, taking into account that a is the cooperative effect for infections from doubly infected to singly infected, while b is the effect for infections from doubly infected to doubly infected. The phenomenology can be easily understood setting two limits: a) $b \gg b_c > a$ and b) $a = b < b_c$. In the first limit, secondary infections from doubly infected are much more likely to occur, such that in a realization there are some primary infections, and then the fraction of primary infected gets a secondary infection, but as $a \ll b$ the susceptible population is less frequently infected, and then the system stays in the low branch. If we continue increasing α , the fraction of doubly infected increases, such that they can access to these singly infected, where the system jumps to the highest branch. However, in the second limit, in one realization the process is more effectively ordered, such that at the beginning there are primary infections; then, when the infected population is big enough, the secondary infections happen, but also the primary infections continue happening, even more frequently due to the feedback effects related with $a > 1$, leading the system to the highest branch in a C type hysteresis cycle.

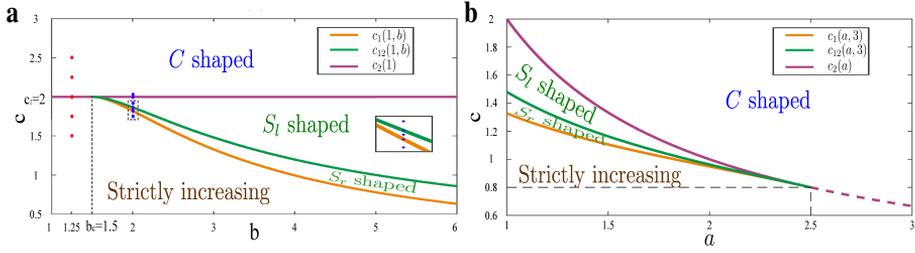


Figure 6.7: Phase diagram in the fully cooperative coinfection model. **a**, Phase diagram in the plane (b, c) , with $a = 1$. Red and blue dots correspond to the curves in Figs. 6.6a and b, respectively. **b**, Phase diagram in the plane (a, c) , with $b = 3$. The violet curve represents the parameters for which there is codimension-2 point in which a saddle-node and a transcritical bifurcation collide, the orange curve represents the parameters for which the system has a cusp, and the green curve represents the crossover between hysteresis of S_r and S_l types. Source: Ref. [159].

6.2 Static random geometric graph

In analogy to the previous chapter, we consider a random geometric graph (RGG) in two dimensions [140, 141], assigning to $N = 4096$ particles random positions in a square space of size $L \times L$, $L = 1280$, and linking a particle with all the particles located at an Euclidean distance smaller than a fixed interaction radius d , using periodic boundary conditions. In this network, our two-disease SIR spreading dynamics will be described by a probability of primary infection p , a probability of secondary infection q , and a recovery probability $r = 1$, following the scheme shown in Fig. 6.2, and we will analyse the dynamics with synchronous update.

According to the previous results, this dynamics in a 2D lattice leads to a continuous phase transition [154, 155], while mean-field ($d \sim d_{\max} = \frac{L}{\sqrt{2}}$) exhibits a first order phase transition [150]. In this section, we will use the parameter d to interpolate between these two limit cases, which display different behaviours.

For low interaction radius, and above the critical radius, $d_c < d \ll d_{\max}$, the system exhibits a continuous phase transition from the disease-free state to the endemic state (Fig. 6.8a). In the continuous phase transitions, the time evolution of the number of infected grows slower than a power of time in the subcritical regime, while at criticality it grows as a power of time, and faster than a power-law for the supercritical regime. The time evolution of the number of infected

6.2. STATIC RANDOM GEOMETRIC GRAPH

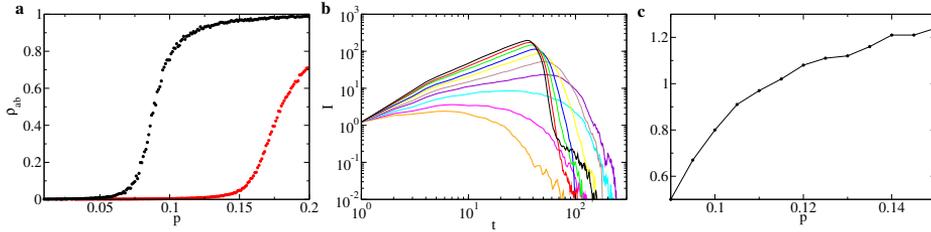


Figure 6.8: Cooperative disease spreading in a static RGG with low interaction radius, $d = 40$. **a**, Density of doubly recovered particles ρ_{ab} at the end of the dynamics as a function of the primary infection probability p . The outbreak appears at lower values of p for $q = 1$ (black) than for $q = p$ (red). **b**, Time evolution of the number of infected nodes, characteristic for a continuous transition. For supercritical values of p , there are Griffiths phases, in which the time evolution of the number of infected $I(t)$ scales as t^μ for a broad range of p . Different colours correspond to different values of $p = 0.06$ (orange), 0.07 (magenta), 0.08 (cyan), 0.09 (purple), 0.10 (brown), 0.11 (yellow), 0.12 (blue), 0.13 (green), 0.14 (red), and 0.15 (black). **c**, The power-law exponent μ changes continuously with p .

nodes $I(t)$ in coinfection dynamics has power-law scaling, $I(t) \sim t^\mu$, for a broad range of values of p , with an exponent $\mu(p)$ that is continuously changing (Figs. 6.8b,c). This phenomenon is known in the literature as Griffiths phases, and has been reported for complex networks in which two dynamical processes are happening, one in the subcritical regime and another in the supercritical. For example, this behaviour appears when a subcritical percolation process is combined with supercritical spreading [160]. Specifically, the subcritical percolation leads to a network that is fragmented in several connected components, with each one having a different characteristic time scale for the spreading process, which depends exponentially on the cluster size. The combination of the distribution of cluster sizes with the time scales growing exponentially with cluster sizes leads to a power-law scaling in the time evolution of the dynamics. In our case, although the system is in the supercritical regime of the percolation diagram $d > d_c$, the avalanches of secondary infections lead to long paths of nodes in doubly recovered state, which become barriers for the spreading, leading to the cluster formation, as the system is not far from the percolation critical point. Then, we have two processes in different regimes: while primary infections are subcritical, *i.e.*, $p_c(q = 1) < p_c(q = p)$, secondary infections are supercritical $q = 1 > p_c(q = 1)$, satisfying the requirements for having Griffiths phases, and

CHAPTER 6. COINFECTION IN MOBILE AND CONTACT NETWORKS

leading to a dynamical fragmentation of the network due to the role of doubly recovered nodes, which become barriers for the spreading.

Then, we study the limit of high d , setting $d = 700$, which corresponds to an average degree $\langle k \rangle \approx 3848 \sim N$. First of all, as highlighted in the previous chapter, for high $\langle k \rangle$, the behaviour of the network approaches the infinite velocity behaviour for independent disease spreading ($q = p$), such that the critical point for non-interacting diseases is the inverse average degree, $p_c = \frac{1}{\langle k \rangle} \approx 2.6 \times 10^{-4}$. In this case, for $q = 1 \gg p$, we observe a transition that is continuous. This continuous transition is in agreement with the mean-field results, which reported the recovery of a continuous transition when the secondary infection rates were much higher than the primary [150]. This can be qualitatively understood considering the characteristic time of both processes: as primary infections are much slower than secondary, the secondary infections, in a highly clustered topology (close to mean field), happen in the following step after primary infections, without the chance of having a bottleneck in which there are no secondary infections. Then, the diseases follow a growth characterized by the time scale of primary infections, leading to a continuous phase transition, and the influence of secondary infections, with $q \gg p$, is reflected in the critical point, which we estimate to be $p_c(q = 1) = \frac{p_c(q=p)}{2}$. This result is understandable taking into account that, at t the particles get one disease, spreading it at $t + 1$, when they get the second disease and can spread it at $t + 2$. In fact, a decrease in q leads to the appearance of a gap, showing the discontinuous phase transition that was expected for mean field dynamics (Fig. 6.9).

6.3. MOBILE RANDOM GEOMETRIC GRAPH

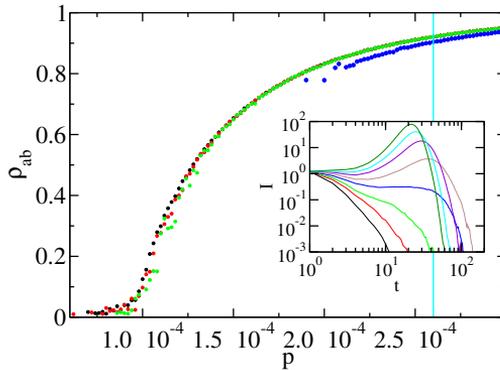


Figure 6.9: Cooperative disease spreading in a static RGG with high interaction radius, $d = 700$. Fraction of doubly recovered ρ_{ab} at the end of the dynamics as a function of the primary infection probability p , for several values of the secondary infection probability $q = 1$ (black), 0.5 (red), 0.25 (green), 0.01 (blue). The cyan line shows the predicted critical point for non-interacting diseases. Inset: time evolution of the number of infected nodes I for $q = 1$, with $p = (5, 6.5, 8, 9.5, 11, 12.5, 14, 15.5) \times 10^5$ from bottom right to the top left.

6.3 Mobile random geometric graph

We study the mobility scenario for networks in which the static case did not show a discontinuous transition, which is the case with low interaction radius. The high values of the clustering, associated with two dimensional topologies, and the short range of the interactions were avoiding the appearance of discontinuous transitions, which required a relatively low ratio between short and long loops. However, the velocity of the nodes may break the dynamical short loops, leading to the appearance of discontinuous transitions. Specifically, we set $d = 40$. The velocity v measures the distance between the position of a node at t and its position at $t + 1$.

For the maximum level of cooperation, *i.e.*, $q = 1$, and low v , the transition remains continuous, but the effect observed in previous chapter is present, with low velocities leading to higher critical points. However, further increasing the velocity, a gap appears between the disease-free state and the outbreak, leading to a discontinuous phase transition that, as velocity increases, approaches a limit that is associated with infinite velocity (Fig. 6.10). As reported for other net-

CHAPTER 6. COINFECTION IN MOBILE AND CONTACT NETWORKS

works, this discontinuous phase transition is hybrid, as the order parameter ρ_{ab} shows a discontinuous transition, while the probability of having macroscopic outbreaks p_{ab} , increases continuously (Fig. 6.11).

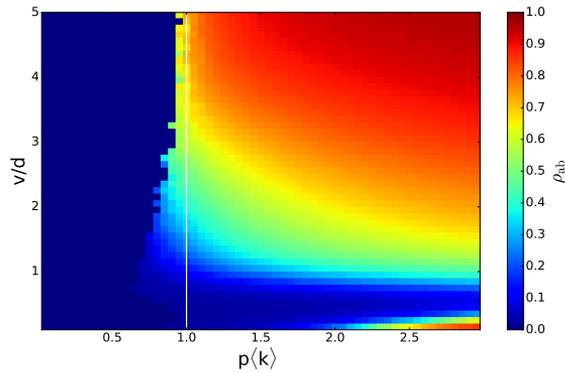


Figure 6.10: Cooperative disease spreading with changing velocity for low interaction radius. We observe the detrimental effect of mobility on outbreak size for small velocities, and for higher velocities a discontinuous transition appears. The white line indicates $p(k) = 1$, that is the predicted critical point for infinite velocity in the thermodynamic limit.

The ratio between short and long loops has been suggested to be the responsible of this behaviour. If diseases meet after short loops, the expected growth of the outbreak is continuous. This is the case of static and low dimensionality RGG, which have higher clustering coefficients. However, the velocity breaks this clustering, decreasing the probability that the two diseases meet in the first steps, and allowing them to take long paths before they meet, such that secondary infections start appearing after a bottleneck.

6.4. EMPIRICAL CONTACT NETWORK

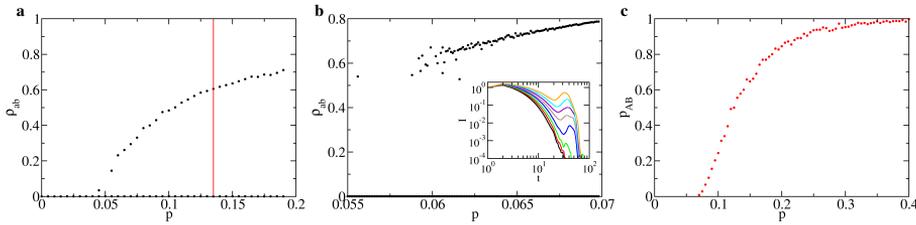


Figure 6.11: Cooperative disease spreading in a mobile RGG with $d = 40$. Density of doubly infected nodes ρ_{ab} at the end of the dynamics, for $q = 1$ and **a**, $v = 0.9d$. The red line shows the critical point for non-interacting diseases dynamics ($q = p$). **b**, $v \rightarrow \infty$. Inset: time evolution of the number of infected, for $p = 0.055$ (black), 0.057 (red), 0.059 (green), 0.061 (blue), 0.063 (brown), 0.065 (violet), 0.067 (cyan), 0.069 (orange). **c**, Probability of a realization giving rise to a macroscopic outbreak for $v = 5d$.

6.4 Empirical contact network

We run the cooperative disease dynamics model in empirical data describing a contact network [161]. This network was built after data collection in a hospital, where $N = 75$ people (46 health care workers and 29 patients) were carrying a wearable device that was reporting a contact when two individuals were located at a distance of 1–1.5 m, aggregating the contacts in intervals of $\Delta t = 20$ s [26]. A hospital represents a potential hub for infections to meet; in fact, some severe infections, like the one produced by *Clostridium difficile*, are typically hospital-acquired infections, and this infection has been reported to cooperate with others, like *Enterococcus* [162]. These facts motivate the analysis of the coinfection dynamics described in previous sections (Fig. 6.2) in an empirical network of contacts in a hospital.

The studied temporal network was composed of 32,424 contacts that were recorded in 17,375 time intervals ($t_{max} = 96.53$ h), meaning that the average number of contacts per unit time was 8.96 h^{-1} . A temporal adjacency matrix describes the network, where the entry $A_{ij}(t) = 1$ if a contact between individuals i and j at time t was detected. The activity pattern of the contact network was non-homogeneous in time, with peaks typically separated by a period of 24 h (Fig. 6.12a). Specifically, we observe some low activity valleys, which are associated with the night, indicating that our dynamics will die out at night if the recovery probability is not set properly, leading us to explore the dynamics with

CHAPTER 6. COINFECTION IN MOBILE AND CONTACT NETWORKS

a recovery probability $p_r = 10^{-3}$, such that the expected recovery time includes 1000 time steps (5.5 h). In fact, in a SI process, with transmission probability equal to one, there are no newly infected individuals at night periods, with the number of infected nodes remaining constant until the morning arrives and new infections occur (Fig. 6.12b).

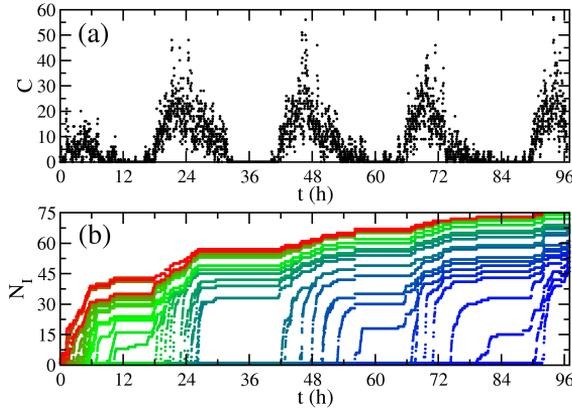


Figure 6.12: Temporal features of the empirical contact network. **a**, Number of contacts $C(t) = \sum_{i < j} A_{ij}(t)$, aggregated for time windows of 80 s. Low activity valleys are associated with night. **b**, Number of infected individuals N_I in a SI process for a single disease with infection probability 1. Each color corresponds to a different initially infected individual, with the colours associated with the first observed interaction for each individual (from sooner, in red, to later, in blue). Source: Ref. [161].

For coinfection dynamics, the control parameter will be the primary infection probability p , considering two cases for the secondary infection probability q : 1) $q = p$, that will stand for independent infections spreading in the network, and 2) $q = 1$, which corresponds to the maximum cooperation. We set an initial condition in which one randomly chosen individual is doubly infected, *i.e.*, its initial state is AB (Table 6.1), and consider a recovery probability $r = 10^{-3}$. We will consider three risk estimations:

6.4. EMPIRICAL CONTACT NETWORK

- The presence of an outbreak, which will be determined by a critical point separating a region in final disease-free state and a region where finite outbreaks are present.
- The fraction of population affected by the outbreak, characterized by the fraction of doubly infected nodes ρ_{ab} at the end of the dynamics.
- The distribution of ρ_{ab} , which will indicate the probability $\Pi_p(\rho_{ab})$ that, for a fixed value of p , a realization leads to a value of ρ_{ab} , indicating the risk of having a big outbreak when we focus in the upper branch of the $\rho_{ab} - p$ diagram.

The spreading dynamics leads to two different branches, both in the cases of independent infections ($q = p$) and strong cooperation ($q = 1$), with the highest reached discontinuously from the disease-free state, and the lowest growing continuously with p (Figs. 6.13a,b). This was a surprising result as, although we could expect a discontinuous transition in the cooperative scenario, we also found it in the non-interacting case. These results arise in contrast to a randomized temporal network. Specifically, this null model considers the same temporal activity pattern, *i.e.*, keeping $C(t)$, but at each time t we select $C(t)$ random contacts from the contact list, with the aim of having the same probability for a contact to happen, but breaking the temporal correlations that were present in the original adjacency matrix $A_{ij}(t)$. In contrast to the spreading in the empirical data, in the randomized case there is a single phase transition, both for $q = p$ and $q = 1$, growing continuously from the disease-free state (Figs. 6.13c,d). This suggests the discontinuous jump is the effect of the temporal correlations.

Although, qualitatively, the observed behaviour is the same for non-interacting and cooperative spreading infections (Figs. 6.13a,b), there are some quantitative differences that show the effects of the cooperation. For each ρ_{ab} and each p , we compute the difference between the probability of observing one realization leading to that point in the cooperative and the non-interacting scenarios $\Delta\Pi_p(\rho_{ab})$, showing how the probability of having big outbreaks (*i.e.*, of being in the higher branch), is higher for the cooperative case, and these outbreaks affect fraction of the population that is almost a 10% higher (Fig. 6.14). Moreover, the epidemic threshold in the continuous transition is smaller for the cooperative case.

In order to understand the mechanisms leading to the discontinuous transition that we observe even in the independent spreading case (Fig. 6.13a), we focus on specific realizations that lead to the two main branches, analysing the time evolution of the number of infected $N_I(t) = N_A(t) + N_B(t) + N_{AB}(t) + N_{aB}(t) + N_{Ab}(t)$ and the fraction of doubly infected with the two diseases N_{AB} (Fig. 6.15 and

CHAPTER 6. COINFECTION IN MOBILE AND CONTACT NETWORKS

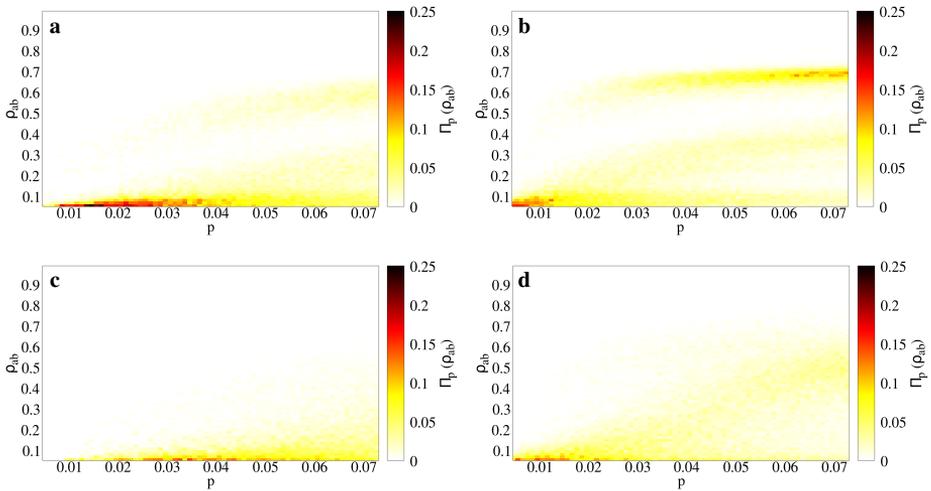


Figure 6.13: Spreading dynamics in a hospital contact network. Diagrams plotting in colour the probability $\Pi_p(\rho_{ab})$ that a realization leads to a density of doubly recovered ρ_{ab} for a transmission probability p . **a**, Independent diseases ($q = p$) spreading in the empirical network. **b**, Cooperative diseases ($q = 1$) spreading in the empirical network. **c**, **d** Independent and cooperative, respectively, diseases spreading in the randomized network. Source: Ref. [161].

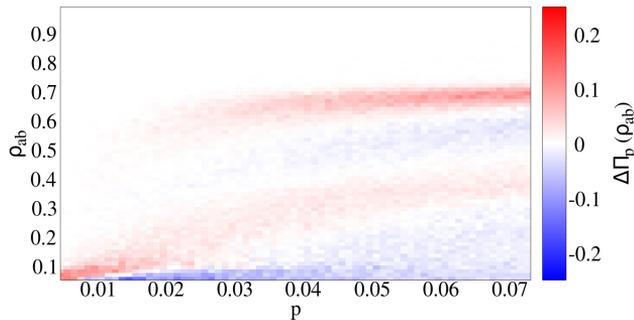


Figure 6.14: Quantitative differences between non-interacting and cooperative infections spreading in the contact network. The differences are calculated as $\Delta\Pi_p(\rho_{ab})$, showing that the three introduced risk estimators are favourable for the disease spreading in the cooperative case. Source: Ref. [161].

6.4. EMPIRICAL CONTACT NETWORK

inset). Initially, the number of infected particles grows, with most of infected nodes in state AB; this is understood taking into account the presence of temporal correlations, with contacts that were repeated very frequently, such that even if the diseases follow different paths, they are able to meet in a few steps before recovering, leading to high fractions of nodes in AB state. In fact, these repeated contacts are observed in the SI process, as the stationary values associated with night are reached before the night valley of activity arrives (Fig. 6.12). However, we have to combine these temporal correlations with another ingredient to explain how this discontinuous branch is created. In fact, when the night arrives, there are no new infections, and recovery processes dominate the dynamics (Fig. 6.15), with the infected population decreasing, but not disappearing, and the doubly infected population decreasing to zero. Then, the recovery in the low activity valleys allows the diseases to remain infective in different nodes. After the night, if the diseases are able to meet again, they will continue spreading together, reinforced in the cooperative case ($q = 1$), like in blue curve of Fig. 6.15, while if they do not meet the final density of doubly infected will be smaller, and associated with the continuously growing curve (orange and black curves in Fig. 6.15). Summing up, the combination of temporal correlations and a low activity valley associated with night is the mechanism underlying the presence of both a continuous and a discontinuous phase transition in the density of doubly recovered nodes ρ_{ab} .

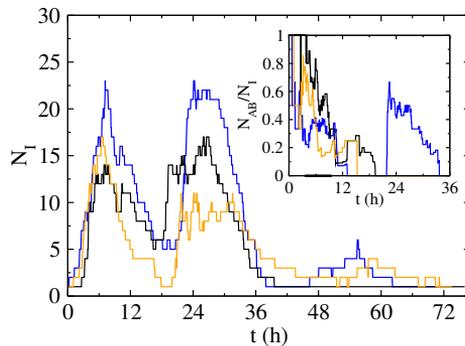


Figure 6.15: Specific realizations of the spreading process in the contact network. Time evolution of the number of infected individuals N_I for three realizations with $p = 0.06$ and $q = 1$, leading to a final number of infected of 56 (blue curve), 54 (black) and 52 (orange) individuals, while the final number of doubly infected is 50 (blue curve), 21 (black) and 14 (orange) individuals. Inset: ratio between the number of doubly active individuals N_{AB} and N_I . Source: Ref. [161].

6.5 Conclusions

In this chapter, we have analysed cooperative disease spreading, which is introduced through higher infection rates/probabilities for secondary contagions. Firstly, we reviewed recent results on a Susceptible-Infected-Recovered model, showing how this dynamics can lead to discontinuous transitions in mean field. Then, the extension of this model to complex networks showed that the discontinuous phase transitions were not always present. In fact, the mechanism giving rise to the discontinuous phase transitions was a bottleneck phenomenon: the diseases spread independently along the network and, after long paths, they met, giving rise to an avalanche of secondary infections. Specifically, the phase transitions were argued to be hybrid, with one order parameter showing a discontinuity but the other order parameter displaying a continuous growth and scaling with system size.

Afterwards, we focused on Susceptible-Infected-Susceptible dynamics, considering another kind of cooperation, which is that coming from higher infection rates when the infective agent is in doubly infected state. Interestingly, we reported three kinds of hysteresis for this system, with the presence of continuous transitions alone, discontinuous transitions alone and some scenarios in which the continuous and the discontinuous transitions were observed.

Motivated by the results of the partial coinfection models in complex networks, we analysed them in random geometric graphs, showing how for big interaction radii the transitions were discontinuous for intermediate values of the secondary infection probability, while for low interaction radii they were continuous. Interestingly, the dynamics in networks with low interaction radius scenario, close but above to the critical percolation radius, was leading to Griffiths phases, with the formation of clusters that were isolated due to the presence of nodes in doubly recovered state, with each cluster having a different characteristic time scale associated with this spreading process.

We wanted to assess the role of mobility, and the mobile random geometric graph was a natural model for this assessment, finding how an increasing velocity for short range interactions was giving rise to a gap, and characterizing the features of the different observed regimes.

Finally, in an effort to bridge the two main parts of this thesis, we analysed this cooperative spreading model in an empirical temporal network of contacts in a hospital. Surprisingly, we found discontinuous phase transitions even in the non-interacting diseases case, showing how these gaps appeared due to the temporal correlations and the activity pattern of the network.

Part IV

Conclusions and appendix

Chapter 7

General conclusions and future outlook

In the conclusions of chapter 1, we proposed some problems that we wanted to address in this thesis. Now, after five chapters, we recall these tasks and explain the conclusions from them:

- **To develop methods suitable for the analysis of animal trajectories.** We applied several approaches used for human movement analysis to the animal trajectories, learning that the memory is a driver of the movement of southern elephant seals.
- **To approach the movement analysis from an aggregated point of view.** In the case of southern elephant seals, we analysed one of the largest datasets describing their movement, finding a universal behaviour despite of the individual idiosyncrasies.
- **To compare trajectories associated with multiple species, identifying the common drivers of marine megafauna movement.** We found that the species and the habitat, split in open and coastal oceans, are the main drivers that explain the observed displacements and turning angles.
- **To describe the patterns exhibited by anthropogenic activities in the ocean. Specifically, analysing the transit pattern of vessels.** We characterised the general transit pattern, analysing the hierarchies between harbours and the shape of the vessel highways.
- **To combine models of dynamics and movement to understand the role of mobility, focusing on disease spreading processes.** In the one hand,

CHAPTER 7. GENERAL CONCLUSIONS AND OUTLOOK

we found a counterintuitive behaviour in which slow motion represents an obstacle for the disease spreading, which reaches higher fractions of the population if the hosts do not move. In the other hand, we matched two previously reported limits for coinfection in static networks, and addressed its study in mobile networks.

- **To use empirical contact networks to simulate disease spreading, determining the features that play a key role in the dynamics and suggesting strategies for reducing outbreak sizes.** We studied coinfection dynamics in an empirical contact network. We found a discontinuous transition that was appearing in the empirical network even in the case of two non-interacting diseases spreading along the network, and explained how the temporal correlations and the activity pattern lead to this behaviour.

However, the scientific knowledge is cumulative, so answering questions does not close the problem, but it opens new questions. After this work, we have some open problems, that we would like to address in the future. Some of them are the following:

- What can we learn from the movement data obtained with other methods? For example, acoustic receivers are located in fixed locations, detecting the individuals when they approach them.
- How the mobility patterns affect dynamics like disease spreading? The optimal foraging theory has argued that the patterns that minimize the search time represent a behaviour with positive fitness, such that they are selected by natural evolution [163]. However, these patterns could be selected not only in terms of the efficiency of resource searches, but also whether minimising disease spreading or maximising information transfer. This study could include both data-driven simulations (*i.e.*, Section 6.4), or models of mobility (*i.e.*, Section 6.3).
- How can we study the collective behaviour from empirical movement data? This question is interesting both for intraspecific behaviour in social species, and interspecific behaviour for species that share the same habitat.
- How the highways observed in the vessels transit patterns influence the movement of marine animals?
- How can we model the interactions between diseases that are described by different dynamics? What about interacting diseases spreading in different topologies, each one associated with a different contagion mode?

Appendix A

Fitting distributions

When a law is proposed to describe a physical phenomenon, a fit to the empirical data is needed to validate the results, getting the value of the parameters that best fit to the law, and measuring a quality parameter that informs us about the distance from our fit to the empirical data. Traditionally, the least squares method has been the most used for computing the regression of the curves. However, this method is not the most adequate to fit data that is distributed following a power-law [164] due to the following reasons:

- A binning process is required for computing a histogram, with the need for setting externally a bin width, which can influence the results.
- The distribution has a heavy tail that includes large fluctuations.
- It is hard to determine the range in which the power-law behaviour holds
- A regression line does not consider the normalization of the data, such that it fits to a distribution function.

Particularly, these methods are of growing interest in the complex systems community, as there has been a permanent discussion about the presence or not of power-laws in empirical data, like in the degree distribution of empirical networks [165]. In this appendix, we will follow the methods described by Clauset *et al.* [166] for fitting empirical data to power-law distributions with the shape $\text{pdf}(x) = Ax^{-\alpha}$ for $\alpha > 1$, that were extended to $\alpha > 0$ in Ref. [167].

APPENDIX A. FITTING DISTRIBUTIONS

We consider a continuous variable x that, whether on a histogram or a cumulative plot, seems to be distributed according to a power-law, *i.e.*, in a log-log plot it has linear scaling. The general shape of a power-law distribution follows the formula $\text{pdf}(x) = Cx^{-\alpha}$, where C is a normalization constant and $\alpha > 1$ is the power-law exponent. However, this probability diverges for $x = 0$, so it needs a lower bound x_{\min} , such that the distribution, after normalization, is described by

$$\text{pdf}(x) = \frac{\alpha - 1}{x_{\min}} \left(\frac{x}{x_{\min}} \right)^{-\alpha} \quad (\text{A.1})$$

Assuming that x_{\min} is known (later we will give some details of how to fit it), we can obtain a fit for α with the maximum likelihood estimation method. We introduce the likelihood, which is the probability for this data being described by the power-law function:

$$p(x|\alpha) = \prod_{i=1}^n \frac{\alpha - 1}{x_{\min}} \left(\frac{x_i}{x_{\min}} \right)^{-\alpha} \quad (\text{A.2})$$

where i runs over all the $x \geq x_{\min}$. The aim of this method looking for the $\hat{\alpha}$ such that this likelihood is maximum. Taking into account that the monotonicity of a function is the same as the monotonicity of its logarithm, their maxima will be located in the same value. Then, we introduce $\mathcal{L} = \log p(x|\alpha)$, that will simplify the optimization calculations:

$$\mathcal{L} = \sum_{i=1}^n \log \left[\frac{\alpha - 1}{x_{\min}} \left(\frac{x_i}{x_{\min}} \right)^{-\alpha} \right] = n \log(\alpha - 1) - n \log x_{\min} - \alpha \sum_{i=1}^n \log \frac{x_i}{x_{\min}} \quad (\text{A.3})$$

Imposing $\partial \mathcal{L} / \partial \alpha = 0$ for optimizing \mathcal{L} , we obtain the estimation $\hat{\alpha}$ of the power-law exponent:

$$\hat{\alpha} = 1 + n \left(\sum_{i=1}^n \log \frac{x_i}{x_{\min}} \right)^{-1} \quad (\text{A.4})$$

Once we have the method for fitting α , we need to fit the lower bound of the power-law behaviour in our empirical data. The simplest method is to test all the x_i in our dataset, fitting an estimator of $\hat{\alpha}$ for each x_i , and choose for x_{\min} the x_i such that the fitted distribution and the empirical are closest. For this purpose, a distance between distributions is needed, and we will use the Kolmogorov-Smirnov distance, which is defined in Eq. (3.2) as the maximum distance between the cumulative density functions of both distributions.

Note that we will always obtain from Eq. (A.4) estimators $\hat{\alpha} > 1$. In fact, we cannot fit a power-law distribution in the interval $[x_{\min}, \infty)$ with $\alpha \leq 1$, as the

APPENDIX A. FITTING DISTRIBUTIONS

decay is so slow that the distribution cannot be normalized. For these cases, the distribution needs to be bounded in the interval $[x_{\min}, x_{\max}]$. The probability density function describing a power-law, with $\alpha > 0$, $\alpha \neq 1$, in this bounded interval is given by

$$\text{pdf}(x) = \frac{1 - \alpha}{x_{\max}^{1-\alpha} - x_{\min}^{1-\alpha}} x^{-\alpha} \quad (\text{A.5})$$

The optimization of the likelihood function associated with this distribution leads to a non-linear equation:

$$\frac{1}{\alpha - 1} - \frac{1}{n} \sum_{i=1}^n \log x_i + \frac{x_{\max}^{1-\alpha} \log x_{\max} - x_{\min}^{1-\alpha} \log x_{\min}}{x_{\max}^{1-\alpha} - x_{\min}^{1-\alpha}} = 0 \quad (\text{A.6})$$

that, after setting x_{\max} and x_{\min} , can be solved numerically to obtain $\hat{\alpha}$. Note that this equation does not impose any restriction for all the values of α except for the case $\alpha = 1$, which would have a different value of the normalization constant. In fact, setting $x_{\max} \rightarrow \infty$ in Eq. A.6, we recover Eq. A.4.

After obtaining an equation for the estimator $\hat{\alpha}$, there are two additional parameters to find, x_{\max} and x_{\min} , which can be fitted together, fitting the distribution for different pairs of these two parameters, and selecting the pair of parameters associated with the smallest Kolmogorov-Smirnov distance between the fit and the empirical data.

Bibliography

- [1] WH Sumbly and I Pollack. Visual contribution to speech intelligibility in noise. *The Journal of the Acoustical Society of America*, 26(2):212–215, 1954.
- [2] J Candia, MC González, P Wang, T Schoenharl, G Madey, and AL Barabási. Uncovering individual and collective human dynamics from mobile phone records. *Journal of Physics A: Mathematical and Theoretical*, 41(22):224015, 2008.
- [3] HH Jo, M Karsai, J Kertész, and K Kaski. Circadian pattern and burstiness in mobile phone communication. *New Journal of Physics*, 14(1):013055, 2012.
- [4] J Saramäki and E Moro. From seconds to months: an overview of multi-scale dynamics of mobile telephone calls. *The European Physical Journal B*, 88(6):164, 2015.
- [5] RD Malmgren, DB Stouffer, ASLO Campanharo, and LAN Amaral. On universality in human correspondence activity. *Science*, 325(5948):1696–1700, 2009.
- [6] LR Binford. Willow smoke and dogs' tails: hunter-gatherer settlement systems and archaeological site formation. *American Antiquity*, 45(1):4–20, 1980.
- [7] SJ Page. *Urban tourism*. Routledge, 1995.
- [8] J Tinbergen. *Social behaviour in animals: with special reference to vertebrates*. Springer Science & Business Media, 2012.
- [9] RIM Dunbar. Coevolution of neocortical size, group size and language in humans. *Behavioral and Brain Sciences*, 16(4):681–694, 1993.

BIBLIOGRAPHY

- [10] ID Jonsen, RA Myers, and MC James. Identifying leatherback turtle foraging behaviour from satellite telemetry using a switching state-space model. *Marine Ecology Progress Series*, 337:255–264, 2007.
- [11] GA Breed, WD Bowen, JI McMillan, and ML Leonard. Sexual segregation of seasonal foraging habitats in a non-migratory marine mammal. *Proceedings of the Royal Society of London B: Biological Sciences*, 273(1599):2319–2326, 2006.
- [12] JP Rodrigue, C Comtois, and B Slack. *The geography of transport systems*. Taylor & Francis, 2016.
- [13] J Ausubel and C Marchetti. The evolution of transport. *Industrial Physicist*, 7(2):20–24, 2001.
- [14] DA Raichlen, AD Gordon, WEH Harcourt-Smith, AD Foster, and WR Haas Jr. Laetoli footprints preserve earliest direct evidence of human-like bipedal biomechanics. *PLOS ONE*, 5(3):e9769, 2010.
- [15] J Leary. *Past mobilities: Archaeological approaches to movement and mobility*. Ashgate Publishing, 2014.
- [16] H Barbosa, M Barthelemy, G Ghoshal, CR James, M Lenormand, T Louail, R Menezes, JJ Ramasco, F Simini, and M Tomasini. Human mobility: Models and applications. *Physics Reports*, 734:1–74, 2018.
- [17] C Cattuto, W Van den Broeck, A Barrat, V Colizza, JF Pinton, and A Vespignani. Dynamics of person-to-person interactions from distributed RFID sensor networks. *PLOS ONE*, 5(7):e11596, 2010.
- [18] A Barrat, C Cattuto, AE Tozzi, P Vanhems, and N Voirin. Measuring contact patterns with wearable sensors: methods, data characteristics and applications to data-driven simulations of infectious diseases. *Clinical Microbiology and Infection*, 20(1):10–16, 2014.
- [19] AB Migliano, AE Page, J Gómez-Gardeñes, GD Salali, S Viguier, M Dyble, J Thompson, N Chaudhary, D Smith, J Strods, R Mace, MG Thomas, V Latora, and L Vinicius. Characterization of hunter-gatherer networks and implications for cumulative culture. *Nature Human Behaviour*, 1:0043, 2017.
- [20] D Brockmann, L Hufnagel, and T Geisel. The scaling laws of human travel. *Nature*, 439:462–465, 2006.
- [21] MC González, CA Hidalgo, and AL Barabási. Understanding individual human mobility patterns. *Nature*, 453(7196):779–782, 2008.

BIBLIOGRAPHY

- [22] C Krumme, A Llorente, M Cebrian, A Pentland, and E Moro. The predictability of consumer visitation patterns. *Scientific Reports*, 3, 2013.
- [23] M Lenormand, B Gonçalves, A Tugores, and JJ Ramasco. Human diffusion and city influence. *Journal of The Royal Society Interface*, 12(109):20150473, 2015.
- [24] C Roth, SM Kang, M Batty, and M Barthélemy. Structure of urban movements: polycentric activity and entangled hierarchical flows. *PLOS ONE*, 6(1):e15923, 2011.
- [25] X Liang, X Zheng, W Lv, T Zhu, and K Xu. The scaling of human mobility by taxis is exponential. *Physica A: Statistical Mechanics and its Applications*, 391(5):2135–2144, 2012.
- [26] P Vanhems, A Barrat, C Cattuto, JF Pinton, N Khanafer, C Régis, B Kim, B Comte, and N Voirin. Estimating potential infection transmission routes in hospital wards using wearable proximity sensors. *PLOS ONE*, 8(9):e73970, 2013.
- [27] F Simini, MC González, A Maritan, and AL Barabási. A universal model for mobility and migration patterns. *Nature*, 484(7392):96, 2012.
- [28] C Song, Z Qu, N Blumm, and A L Barabási. Limits of predictability in human mobility. *Science*, 327(5968):1018–1021, 2010.
- [29] YA De Montjoye, L Radaelli, VK Singh, and A Pentland. Unique in the shopping mall: On the reidentifiability of credit card metadata. *Science*, 347(6221):536–539, 2015.
- [30] C Song, T Koren, P Wang, and AL Barabási. Modelling the scaling properties of human mobility. *Nature Physics*, 6(10):818, 2010.
- [31] R Guimera, L Danon, A Diaz-Guilera, F Giralt, and A Arenas. Self-similar community structure in a network of human interactions. *Physical Review E*, 68(6):065103, 2003.
- [32] WW Zachary. An information flow model for conflict and fission in small groups. *Journal of Anthropological Research*, 33(4):452–473, 1977.
- [33] J Bascompte, P Jordano, CJ Melián, and JM Olesen. The nested assembly of plant–animal mutualistic networks. *Proceedings of the National Academy of Sciences*, 100(16):9383–9387, 2003.

BIBLIOGRAPHY

- [34] S Li, CM Armstrong, N Bertin, H Ge, S Milstein, M Boxem, PO Vidalain, JD J Han, A Chesneau, T Hao, DS Goldber, N Li, M Martinez, JF Rual, P Lamesch, L Xu, M Tewari, SL Wong, LV Zhang, GF Berriz, L Jacotot, P Vaglio, J Reboul, T Hirozane-Kishikawa, Q Li, HW Gabel, A Elewa, B Baumgartner, DJ Rose, h Yu, S Bosak, R Sequerra, A Fraser, SE Mango, WM Saxton, S Strome, S van den Heuvel, F Piano, J Vandenhoute, C Sardet, M Gerstein, L Doucette-Stamm, KC Gunsalus, JW Harper, ME Cusick, FP Roth, DE Hill, and M Vidal. A map of the interactome network of the metazoan *C. elegans*. *Science*, 303(5657):540–543, 2004.
- [35] AHY Tong, M Evangelista, AB Parsons, H Xu, GD Bader, N Pagé, M Robinson, S Raghbizadeh, CWV Hogue, H Bussey, B Andrews, M Tyers, and C Boone. Systematic genetic analysis with ordered arrays of yeast deletion mutants. *Science*, 294(5550):2364–2368, 2001.
- [36] Roger Guimera and Luis A Nunes Amaral. Modeling the world-wide airport network. *The European Physical Journal B*, 38(2):381–385, 2004.
- [37] R Albert, H Jeong, and AL Barabási. Internet: Diameter of the world-wide web. *Nature*, 401(6749):130, 1999.
- [38] M Kivela, A Arenas, M Barthelemy, JP Gleeson, Y Moreno, and MA Porter. Multilayer networks. *Journal of Complex Networks*, 2(3):203–271, 2014.
- [39] SV Buldyrev, R Parshani, G Paul, HE Stanley, and S Havlin. Catastrophic cascade of failures in interdependent networks. *Nature*, 464(7291):1025, 2010.
- [40] GK Zipf. The $P_1 P_2/D$ hypothesis: on the intercity movement of persons. *American Sociological Review*, 11(6):677–686, 1946.
- [41] AP Masucci, J Serras, A Johansson, and M Batty. Gravity versus radiation models: On the importance of scale and heterogeneity in commuting flows. *Physical Review E*, 88(2):022812, 2013.
- [42] S Fortunato. Community detection in graphs. *Physics Reports*, 486(3-5):75–174, 2010.
- [43] B Hawelka, I Sitko, E Beinat, S Sobolevsky, P Kazakopoulos, and C Ratti. Geo-located Twitter as proxy for global mobility patterns. *Cartography and Geographic Information Science*, 41(3):260–271, 2014.
- [44] C Ratti, S Sobolevsky, F Calabrese, C Andris, J Reades, M Martino, R Claxton, and SH Strogatz. Redrawing the map of Great Britain from a network of human interactions. *PLOS ONE*, 5(12):e14248, 2010.

BIBLIOGRAPHY

- [45] PA Grabowicz, J J Ramasco, B Gonçalves, and VM Eguíluz. Entangling mobility and interactions in social media. *PLOS ONE*, 9(3):e92196, 2014.
- [46] P Deville, C Song, N Eagle, VD Blondel, AL Barabási, and D Wang. Scaling identity connects human mobility and social interactions. *Proceedings of the National Academy of Sciences*, 113(26):7047–7052, 2016.
- [47] M Lenormand, M Picornell, OG Cantú-Ros, T Louail, R Herranz, M Barthelemy, E Frías-Martínez, M San Miguel, and JJ Ramasco. “Comparing and modelling land use organization in cities”. *Royal Society Open Science*, 2:150449, 2015.
- [48] P Turchin. *Quantitative analysis of movement*. Sinauer assoc. Sunderland, 1998.
- [49] RB Root and PM Kareiva. The search for resources by cabbage butterflies (*Pieris rapae*): ecological consequences and adaptive significance of Markovian movements in a patchy environment. *Ecology*, 65(1):147–165, 1984.
- [50] GM Viswanathan, V Afanasyev, SV Buldyrev, EJ Murphy, PA Prince, and HE Stanley. Lévy flight search patterns of wandering albatrosses. *Nature*, 381(6581):413, 1996.
- [51] GM Viswanathan, SV Buldyrev, S Havlin, MGE Da Luz, EP Raposo, and HE Stanley. Optimizing the success of random searches. *Nature*, 401(6756):911–914, 1999.
- [52] GM Viswanathan, V Afanasyev, SV Buldyrev, S Havlin, MGE Da Luz, EP Raposo, and HE Stanley. Lévy flights in random searches. *Physica A: Statistical Mechanics and its Applications*, 282(1-2):1–12, 2000.
- [53] O Bénichou, C Loverdo, M Moreau, and R Voituriez. Intermittent search strategies. *Reviews of Modern Physics*, 83(1):81, 2011.
- [54] A James, MJ Plank, and R Brown. Optimizing the encounter rate in biological interactions: ballistic versus Lévy versus Brownian strategies. *Physical Review E*, 78(5):051128, 2008.
- [55] DW Stephens and JR Krebs. *Foraging theory*. Princeton University Press, 1986.
- [56] F Bartumeus, J Catalan, UL Fulco, ML Lyra, and GM Viswanathan. Optimizing the encounter rate in biological interactions: Lévy versus Brownian strategies. *Physical Review Letters*, 88(9):097901, 2002.

BIBLIOGRAPHY

- [57] AM Edwards, RA Phillips, NW Watkins, MP Freeman, EJ Murphy, V Afanasyev, SV Buldyrev, MGE da Luz, EP Raposo, HE Stanley, and GM Viswanathan. Revisiting Lévy flight search patterns of wandering albatrosses, bumblebees and deer. *Nature*, 449(7165):1044, 2007.
- [58] NE Humphries, N Queiroz, JRM Dyer, NG Pade, MK Musyl, KM Schaefer, DW Fuller, JM Brunnschweiler, TK Doyle, JDR Houghton, et al. Environmental context explains Lévy and Brownian movement patterns of marine predators. *Nature*, 465(7301):1066–1069, 2010.
- [59] BA Block, ID Jonsen, SJ Jorgensen, AJ Winship, SA Shaffer, SJ Bograd, EL Hazen, DG Foley, GA Breed, AL Harrison, JE Ganong, A Swithenbank, M Castleton, H Dewar, BR Mate, GL Shillinger, KM Schaefer, SR Benson, MJ Weise, RW Henry, and DP Costa. Tracking apex marine predator movements in a dynamic ocean. *Nature*, 475(7354):86–90, 2011.
- [60] DW Sims, EJ Southall, NE Humphries, GC Hays, CJA Bradshaw, JW Pitchford, A James, MZ Ahmed, AS Brierley, MA Hindell, D Morritt, MK Musul, D Righton, ELC Shepard, VJ Wearmouth, RP Wilson, MJ Witt, and JD Metcalfe. Scaling laws of marine predator search behaviour. *Nature*, 451(7182):1098–1102, 2008.
- [61] GM Viswanathan, EP Raposo, and MGE Da Luz. Lévy flights and superdiffusion in the context of biological encounters and random searches. *Physics of Life Reviews*, 5(3):133–150, 2008.
- [62] NE Humphries, H Weimerskirch, and DW Sims. A new approach for objective identification of turns and steps in organism movement data relevant to random walk modelling. *Methods in Ecology and Evolution*, 4(10):930–938, 2013.
- [63] M Barthélemy. Spatial networks. *Physics Reports*, 499(1):1–101, 2011.
- [64] JH Arias, J Gómez-Gardeñes, S Meloni, and E Estrada. Epidemics on plants: Modeling long-range dispersal on spatially embedded networks. *Journal of Theoretical Biology*, 453:1–13, 2018.
- [65] CJ Melián, O Seehausen, VM Eguíluz, MA Fortuna, and K Deiner. Diversification and biodiversity dynamics of hot and cold spots. *Ecography*, 38(4):393–401, 2015.
- [66] A Díaz-Guilera, J Gómez-Gardenes, Y Moreno, and M Nekovee. Synchronization in random geometric graphs. *International Journal of Bifurcation and Chaos*, 19(02):687–693, 2009.

BIBLIOGRAPHY

- [67] I Hanski. Metapopulation dynamics. *Nature*, 396(6706):41, 1998.
- [68] A Wesolowski, N Eagle, AJ Tatem, DL Smith, AM Noor, RW Snow, and CO Buckee. Quantifying the impact of human mobility on malaria. *Science*, 338(6104):267–270, 2012.
- [69] S Eubank, H Guclu, VSA Kumar, MV Marathe, A Srinivasan, Z Toroczkai, and N Wang. Modelling disease outbreaks in realistic urban social networks. *Nature*, 429(6988):180, 2004.
- [70] D Balcan, V Colizza, B Gonçalves, H Hu, JJ Ramasco, and A Vespignani. Multiscale mobility networks and the spatial spreading of infectious diseases. *Proceedings of the National Academy of Sciences*, 106(51):21484–21489, 2009.
- [71] V Belik, T Geisel, and D Brockmann. Natural human mobility patterns and spatial spread of infectious diseases. *Physical Review X*, 1(1):011001, 2011.
- [72] N Fujiwara, J Kurths, and A Díaz-Guilera. Synchronization in networks of mobile oscillators. *Physical Review E*, 83(2):025101, 2011.
- [73] L Prignano, O Sagarra, and A Díaz-Guilera. Tuning synchronization of integrate-and-fire oscillators through mobility. *Physical Review Letters*, 110(11):114101, 2013.
- [74] V Sood and S Redner. Voter model on heterogeneous graphs. *Physical Review Letters*, 94(17):178701, 2005.
- [75] J Fernández-Gracia, K Suchecki, JJ Ramasco, M San Miguel, and VM Eguíluz. Is the voter model a model for voters? *Physical Review Letters*, 112(15):158701, 2014.
- [76] S Meloni, A Buscarino, L Fortuna, M Frasca, J Gómez-Gardeñes, V Latora, and Y Moreno. Effects of mobility in a population of prisoner’s dilemma players. *Physical Review E*, 79(6):067101, 2009.
- [77] FC Santos and JM Pacheco. Scale-free networks provide a unifying framework for the emergence of cooperation. *Physical Review Letters*, 95(9):098104, 2005.
- [78] MR Heupel, CA Simpfendorfer, EM Olsen, and E Moland. Consistent movement traits indicative of innate behavior in neonate sharks. *Journal of Experimental Marine Biology and Ecology*, 432:131–137, 2012.

BIBLIOGRAPHY

- [79] JP Rodríguez, J Fernández-Gracia, M Thums, MA Hindell, AMM Sequeira, MG Meekan, DP Costa, C Guinet, RG Harcourt, CR McMahon, M Muelbert, CM Duarte, and VM Eguíluz. Big data analyses reveal patterns and drivers of the movements of southern elephant seals. *Scientific Reports*, 7(1):112, 2017.
- [80] WJ O'Brien, HI Browman, and BI Evans. Search strategies of foraging animals. *American Scientist*, 78(2):152–160, 1990.
- [81] GM Viswanathan, MGE Da Luz, EP Raposo, and HE Stanley. *The physics of foraging: an introduction to random searches and biological encounters*. Cambridge University Press, 2011.
- [82] G Ramos-Fernández, JL Mateos, O Miramontes, G Cocho, H Larralde, and B Ayala-Orozco. Lévy walk patterns in the foraging movements of spider monkeys (*Ateles geoffroyi*). *Behavioral Ecology and Sociobiology*, 55(3):223–230, 2004.
- [83] NE Humphries, H Weimerskirch, N Queiroz, EJ Southall, and DW Sims. Foraging success of biological Lévy flights recorded in situ. *Proceedings of the National Academy of Sciences*, 109(19):7169–7174, 2012.
- [84] I Rhee, M Shin, S Hong, K Lee, Seong J Kim, and S Chong. On the levy-walk nature of human mobility. *IEEE/ACM Transactions on Networking (TON)*, 19(3):630–643, 2011.
- [85] DA Raichlen, BM Wood, AD Gordon, AZP Mabulla, FW Marlowe, and H Pontzer. Evidence of Lévy walk foraging patterns in human hunter-gatherers. *Proceedings of the National Academy of Sciences*, 111(2):728–733, 2014.
- [86] F Bartumeus. Behavioral intermittence, Lévy patterns, and randomness in animal movement. *Oikos*, 118(4):488–494, 2009.
- [87] AM Reynolds, AD Smith, R Menzel, U Greggers, DR Reynolds, and JR Riley. Displaced honey bees perform optimal scale-free search flights. *Ecology*, 88(8):1955–1961, 2007.
- [88] ID Jonsen, M Basson, S Bestley, MV Bravington, TA Patterson, MW Pederesen, R Thomson, UH Thygesen, and SJ Wotherspoon. State-space models for bio-loggers: A methodological road map. *Deep Sea Research Part II: Topical Studies in Oceanography*, 88:34–46, 2013.
- [89] L Seuront and HE Stanley. Anomalous diffusion and multifractality enhance mating encounters in the ocean. *Proceedings of the National Academy of Sciences*, 111(6):2206–2211, 2014.

BIBLIOGRAPHY

- [90] RM Tromer, MB Barbosa, F Bartumeus, J Catalan, MGE da Luz, EP Raposo, and GM Viswanathan. Inferring Lévy walks from curved trajectories: A rescaling method. *Physical Review E*, 92(2):022147, 2015.
- [91] BV Gnedenko and AN Kolmogorov. *Limit distributions for sums of independent random variables*. Addison-Wesley Mathematics Series. Addison-Wesley, Cambridge, MA, 1954.
- [92] VV Uchaikin and VM Zolotarev. *Chance and stability: stable distributions and their applications*. Walter de Gruyter, 1999.
- [93] WF Fagan, MA Lewis, M Auger-Méthé, T Avgar, S Benhamou, G Breed, L LaDage, UE Schlägel, W Tang, YP Papastamatiou, J Forester, and T Mueller. Spatial memory and animal movement. *Ecology Letters*, 16(10):1316–1329, 2013.
- [94] T Takaguchi, M Nakamura, N Sato, K Yano, and N Masuda. Predictability of conversation partners. *Physical Review X*, 1(1):011008, 2011.
- [95] C Zhong, SM Arisona, X Huang, M Batty, and G Schmitt. Detecting the dynamics of urban structure through spatial network analysis. *International Journal of Geographical Information Science*, 28(11):2178–2199, 2014.
- [96] M Rosvall and CT Bergstrom. Maps of random walks on complex networks reveal community structure. *Proceedings of the National Academy of Sciences*, 105(4):1118–1123, 2008.
- [97] CJA Bradshaw, MA Hindell, MD Sumner, and KJ Michael. Loyalty pays: potential life history consequences of fidelity to marine foraging regions by southern elephant seals. *Animal Behaviour*, 68(6):1349–1360, 2004.
- [98] AC Dragon, A Bar-Hen, P Monestiez, and C Guinet. Horizontal and vertical movements as predictors of foraging success in a marine predator. *Marine Ecology Progress Series*, 447:243–257, 2012.
- [99] M Authier, I Bentaleb, A Ponchon, C Martin, and C Guinet. Foraging fidelity as a recipe for a long life: foraging strategy and longevity in male southern elephant seals. *PLOS ONE*, 7(4):e32026, 2012.
- [100] NE Hussey, ST Kessel, K Aarestrup, SJ Cooke, PD Cowley, AT Fisk, RG Harcourt, KN Holland, SJ Iverson, JF Kocik, et al. Aquatic animal telemetry: a panoramic window into the underwater world. *Science*, 348(6240):1255642, 2015.
- [101] R Kays, MC Crofoot, W Jetz, and M Wikelski. Terrestrial animal tracking as an eye on life and planet. *Science*, 348(6240):aaa2478, 2015.

BIBLIOGRAPHY

- [102] H Whitehead, J Christal, and S Dufault. Past and distant whaling and the rapid decline of sperm whales off the Galapagos Islands. *Conservation Biology*, 11(6):1387–1396, 1997.
- [103] AS Craig and LM Herman. Sex differences in site fidelity and migration of humpback whales (*Megaptera novaeangliae*) to the Hawaiian Islands. *Canadian Journal of Zoology*, 75(11):1923–1933, 1997.
- [104] C Becco, N Vandewalle, J Delcourt, and P Poncin. Experimental evidences of a structural and dynamical transition in fish school. *Physica A: Statistical Mechanics and its Applications*, 367:487–493, 2006.
- [105] A Cavagna, A Cimarelli, I Giardina, G Parisi, R Santagati, F Stefanini, and M Viale. Scale-free correlations in starling flocks. *Proceedings of the National Academy of Sciences*, 107(26):11865–11870, 2010.
- [106] F Ginelli, F Peruani, MH Pillot, H Chaté, G Theraulaz, and R Bon. Intermittent collective dynamics emerge from conflicting imperatives in sheep herds. *Proceedings of the National Academy of Sciences*, 112(41):12729–12734, 2015.
- [107] GC Hays, LC Ferreira, AMM Sequeira, MG Meekan, CM Duarte, H Bailey, F Bailleul, WD Bowen, MJ Caley, DP Costa, VM Eguíluz, S Fossette, AS Friedlaender, N Gales, AC Gleiss, J Gunn, RG Harcourt, EL Hazen, MR Heithaus, M Heupel, K Holland, M Horning, I Jonsen, GL Kooyman, CG Lowe, PT Madsen, H Marsh, RA Philips, D Righton, Y Ropert-Couder, K Sato, SA Shaffer, CA Simpfendorfer, DW Sims, G Skomal, A Takahashi, PN Trathan, M Wikelski, JN Womble, and M Thums. Key questions in marine megafauna movement ecology. *Trends in Ecology & Evolution*, 31(6):463–475, 2016.
- [108] BJ Le Boeuf, DE Crocker, DP Costa, SB Blackwell, PM Webb, and DS Houser. Foraging ecology of northern elephant seals. *Ecological Monographs*, 70(3):353–382, 2000.
- [109] MJ Carey, RA Phillips, JRD Silk, and SA Shaffer. Trans-equatorial migration of short-tailed shearwaters revealed by geolocators. *Emu*, 114(4):352–359, 2014.
- [110] SG Cherry, AE Derocher, GW Thiemann, and NJ Lunn. Migration phenology and seasonal fidelity of an Arctic marine predator in relation to sea ice dynamics. *Journal of Animal Ecology*, 82(4):912–921, 2013.

BIBLIOGRAPHY

- [111] S Bestley, ID Jonsen, MA Hindell, RG Harcourt, and NJ Gales. Taking animal tracking to new depths: Synthesizing horizontal–vertical movement relationships for four marine predators. *Ecology*, 96(2):417–427, 2015.
- [112] T Avgar, A Mosser, GS Brown, and JM Fryxell. Environmental and individual drivers of animal movement patterns across a wide geographical gradient. *Journal of Animal Ecology*, 82(1):96–106, 2013.
- [113] AMM Sequeira, JP Rodríguez, VM Eguíluz, R Harcourt, M Hindell, DW Sims, CM Duarte, DP Costa, J Fernández-Gracia, LC Ferreira, GC Hays, MR Heupel, MG Meekan, A Aven, F Bailleul, AMM Baylis, ML Berumen, CD Braun, J Burns, MJ Caley, R Campbell, RH Carmichael, E Clua, LD Einoder, A Friedlaender, F Goebel, SD Goldsworthy, C Guinet, J Gunn, D Hamer, N Hammerschlag, M Hammill, LA Hückstädt, NE Humphries, MA Lea, A Lowther, A Mackay, E McHuron, J McKenzie, L McLeay, CR McMahan, K Mengersen, MMC Muelbert, AM Pagano, B Page, N Queiroz, PW Robinson, SA Shaffer, M Shivji, GB Skomal, SR Thorrold, S Villegas-Amtmann, M Weise, R Wells, B Wetherbee, A Wiebkin, B Wienecke, and M Thums. Convergence of marine megafauna movement patterns in coastal and open oceans. *Proceedings of the National Academy of Sciences*, 115(12):3032–3037, 2018.
- [114] RS Schick, SR Loarie, F Colchero, BD Best, A Boustany, DA Conde, PN Halpin, LN Joppa, CM McClellan, and JS Clark. Understanding movement data and movement processes: current and emerging directions. *Ecology Letters*, 11(12):1338–1350, 2008.
- [115] J Elith, JR Leathwick, and T Hastie. A working guide to boosted regression trees. *Journal of Animal Ecology*, 77(4):802–813, 2008.
- [116] RR Sokal and CD Michener. A statistical method for evaluating systematic relationships. *The University of Kansas Science Bulletin*, 38:1409–1438, 1958.
- [117] N Saitou and M Nei. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution*, 4(4):406–425, 1987.
- [118] M Krzywinski, J Schein, I Birol, J Connors, R Gascoyne, D Horsman, SJ Jones, and MA Marra. Circos: an information aesthetic for comparative genomics. *Genome Research*, 19(9):1639–1645, 2009.
- [119] R Nathan, WM Getz, E Revilla, M Holyoak, R Kadmon, D Saltz, and PE Smouse. A movement ecology paradigm for unifying organismal movement research. *Proceedings of the National Academy of Sciences*, 105(49):19052–19059, 2008.

BIBLIOGRAPHY

- [120] CA Bost, C Cotté, F Bailleul, Y Cherel, JB Charrassin, C Guinet, DG Ainley, and H Weimerskirch. The importance of oceanographic fronts to marine birds and mammals of the southern oceans. *Journal of Marine Systems*, 78(3):363–376, 2009.
- [121] Emilie T Kai, V Rossi, J Sudre, H Weimerskirch, C Lopez, E Hernandez-Garcia, F Marsac, and V Garçon. Top marine predators track Lagrangian coherent structures. *Proceedings of the National Academy of Sciences*, 106(20):8245–8250, 2009.
- [122] BS Halpern, S Walbridge, KA Selkoe, CV Kappel, F Micheli, C D’agrosa, JF Bruno, KS Casey, C Ebert, HE Fox, R Fujita, D Heinemann, HS Lenihan, EMP Madin, MT Perry, ER Selig, M Spalding, R Steneck, and R Watson. A global map of human impact on marine ecosystems. *Science*, 319(5865):948–952, 2008.
- [123] DDW Hauser, KL Laidre, and HL Stern. Vulnerability of Arctic marine mammals to vessel traffic in the increasingly ice-free Northwest Passage and Northern Sea Route. *Proceedings of the National Academy of Sciences*, page 201803543, 2018.
- [124] D Barki and L Délèze-Black. *Review of maritime transport*. United Nations, 2017.
- [125] ASM Vanderlaan and CT Taggart. Vessel collisions with whales: the probability of lethal injury based on vessel speed. *Marine Mammal Science*, 23(1):144–156, 2007.
- [126] P Kaluza, A Kölzsch, MT Gastner, and B Blasius. The complex network of global cargo ship movements. *Journal of the Royal Society Interface*, 7(48):1093–1103, 2010.
- [127] X Gabaix. Zipf’s law for cities: an explanation. *The Quarterly Journal of Economics*, 114(3):739–767, 1999.
- [128] MB Isichenko. Percolation, statistical topography, and transport in random media. *Reviews of Modern Physics*, 64(4):961, 1992.
- [129] BB Cael and DA Seekell. The size-distribution of Earth’s lakes. *Scientific Reports*, 6:29633, 2016.
- [130] R Guimera, S Mossa, A Turtschi, and LAN Amaral. The worldwide air transportation network: Anomalous centrality, community structure, and cities’ global roles. *Proceedings of the National Academy of Sciences*, 102(22):7794–7799, 2005.

BIBLIOGRAPHY

- [131] WO Kermack and AG McKendrick. A contribution to the mathematical theory of epidemics. *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 115(772):700–721, 1927.
- [132] RM Anderson, RM May, and B Anderson. *Infectious diseases of humans: dynamics and control*, volume 28. Wiley Online Library, 1992.
- [133] LEC Rocha, F Liljeros, and P Holme. Simulated epidemics in an empirical spatiotemporal network of 50,185 sexual contacts. *PLOS Computational Biology*, 7(3):e1001109, 2011.
- [134] M Salathé, M Kazandjieva, JW Lee, P Levis, MW Feldman, and JH Jones. A high-resolution human contact network for infectious disease transmission. *Proceedings of the National Academy of Sciences*, 107(51):22020–22025, 2010.
- [135] M Kitsak, LK Gallos, S Havlin, F Liljeros, L Muchnik, HE Stanley, and HA Makse. Identification of influential spreaders in complex networks. *Nature Physics*, 6(11):888–893, 2010.
- [136] V Colizza, A Barrat, M Barthélemy, and A Vespignani. The role of the airline transportation network in the prediction and predictability of global epidemics. *Proceedings of the National Academy of Sciences*, 103(7):2015–2020, 2006.
- [137] V Colizza, A Barrat, M Barthelemy, AJ Valleron, and A Vespignani. Modeling the worldwide spread of pandemic influenza: baseline case and containment interventions. *PLOS Medicine*, 4(1):e13, 2007.
- [138] J Gómez-Gardeñes, D Soriano-Paños, and A Arenas. Critical regimes driven by recurrent mobility patterns of reaction–diffusion processes in networks. *Nature Physics*, page 1, 2017.
- [139] P Holme and J Saramäki. Temporal networks. *Physics Reports*, 519(3):97–125, 2012.
- [140] J Dall and M Christensen. Random geometric graphs. *Physical Review E*, 66(1):016121, 2002.
- [141] M Penrose. *Random geometric graphs*. Oxford University Press, 2003.
- [142] E Estrada and M Sheerin. Random rectangular graphs. *Physical Review E*, 91(4):042805, 2015.
- [143] E Estrada, S Meloni, M Sheerin, and Y Moreno. Epidemic spreading in random rectangular networks. *Physical Review E*, 94(5):052316, 2016.

BIBLIOGRAPHY

- [144] S Boccaletti, V Latora, Y Moreno, M Chavez, and DU Hwang. Complex networks: Structure and dynamics. *Physics Reports*, 424(4):175–308, 2006.
- [145] M Ajelli, B Gonçalves, D Balcan, V Colizza, H Hu, JJ Ramasco, S Merler, and A Vespignani. Comparing large-scale computational approaches to epidemic modeling: agent-based versus structured metapopulation models. *BMC Infectious diseases*, 10(1):190, 2010.
- [146] Q Zhang, K Sun, M Chinazzi, A Pastore y Piontti, NE Dean, DP Rojas, S Merler, D Mistry, P Poletti, L Rossi, M Bray, ME Halloran, IM Longini Jr., and A Vespignani. Spread of Zika virus in the Americas. *Proceedings of the National Academy of Sciences*, 114(22):E4334–E4343, 2017.
- [147] N Masuda, K Klemm, and VM Eguíluz. Temporal networks: slowing down diffusion by long lasting interactions. *Physical Review Letters*, 111(18):188701, 2013.
- [148] J Sanz, CY Xia, S Meloni, and Y Moreno. Dynamics of interacting diseases. *Physical Review X*, 4(4):041005, 2014.
- [149] B Karrer and MEJ Newman. Competing epidemics on complex networks. *Physical Review E*, 84(3):036106, 2011.
- [150] L Chen, F Ghanbarnejad, W Cai, and P Grassberger. Outbreaks of coinfections: The critical role of cooperativity. *Europhysics Letters*, 104(5):50001, 2013.
- [151] JF Brundage and GD Shanks. Deaths from bacterial pneumonia during 1918–19 influenza pandemic. *Emerging Infectious Diseases*, 14(8):1193, 2008.
- [152] SK Sharma, A Mohan, T Kadiravan, et al. HIV-TB co-infection: epidemiology, diagnosis & management. *Indian Journal of Medical Research*, 121(4):550–567, 2005.
- [153] HK Janssen and O Stenull. First-order phase transitions in outbreaks of coinfectious diseases and the extended general epidemic process. *Europhysics Letters*, 113(2):26005, 2016.
- [154] W Cai, L Chen, F Ghanbarnejad, and P Grassberger. Avalanche outbreaks emerging in cooperative contagions. *Nature Physics*, 11(11):936–940, 2015.
- [155] P Grassberger, L Chen, F Ghanbarnejad, and W Cai. Phase transitions in cooperative coinfections: Simulation results for networks and lattices. *Physical Review E*, 93(4):042316, 2016.

- [156] P Erdős and A Rényi. On the evolution of random graphs. *Publication of the Mathematical Institute of the Hungarian Academy of Sciences*, 5(1):17–60, 1960.
- [157] D Stauffer and A Aharony. *Introduction to percolation theory*. Taylor & Francis, London, 1985.
- [158] L Chen, F Ghanbarnejad, and D Brockmann. Fundamental properties of cooperative contagion processes. *New Journal of Physics*, 19(10):103041, 2017.
- [159] JP Rodríguez, YH Liang, YJ Huang, and J Juang. Diversity of hysteresis in a fully cooperative coinfection model. *Chaos*, 28(2):023107, 2018.
- [160] MA Muñoz, R Juhász, C Castellano, and G Ódor. Griffiths phases on complex networks. *Physical Review Letters*, 105(12):128701, 2010.
- [161] JP Rodríguez, F Ghanbarnejad, and VM Eguíluz. Risk of Coinfection Outbreaks in Temporal Networks: A Case Study of a Hospital Contact Network. *Frontiers in Physics*, 5:46, 2017.
- [162] RD Poduval, RP Kamath, M Corpuz, EP Norkus, and CS Pitchumoni. Clostridium difficile and vancomycin-resistant enterococcus: the new nosocomial alliance. *The American Journal of Gastroenterology*, 95(12):3513–3515, 2000.
- [163] F Bartumeus. Lévy processes in animal movement: an evolutionary hypothesis. *Fractals*, 15(02):151–162, 2007.
- [164] MEJ Newman. Power laws, Pareto distributions and Zipf’s law. *Contemporary Physics*, 46(5):323–351, 2005.
- [165] AD Broido and A Clauset. Scale-free networks are rare. *arXiv preprint arXiv:1801.03400*, 2018.
- [166] A Clauset, CR Shalizi, and MEJ Newman. Power-law distributions in empirical data. *SIAM Review*, 51(4):661–703, 2009.
- [167] A Deluca and Á Corral. Fitting and goodness-of-fit test of non-truncated and truncated power-law distributions. *Acta Geophysica*, 61(6):1351–1394, 2013.