Human behavioural patterns: 
A reality mining study

Daniel Monsivais Velázquez
Human behavioural patterns: A reality mining study

Daniel Monsivais Velázquez

A doctoral dissertation completed for the degree of Doctor of Science (Technology) to be defended, with the permission of the Aalto University School of Science, at a public examination held at the lecture hall AS2 of the school on 5 October 2018 at 13:00.

Aalto University
School of Science
Department of Computer Science
Complex Systems
Abstract

**Author**
Daniel Monsivais Velázquez

**Name of the doctoral dissertation**
Human behavioural patterns: A reality mining study

**Publisher** School of Science

**Unit** Department of Computer Science

**Series** Aalto University publication series DOCTORAL DISSERTATIONS 184/2018

**Field of research** Computational Science

**Manuscript submitted** 18 June 2018 **Date of the defence** 5 October 2018

**Permission to publish granted (date)** 22 August 2018

**Language** English


**Abstract**

Mobile phone communication is a source of information for studying human behavioural patterns. A mobile phone can collect information of its usage, communication events and data captured by integrated sensors, and this information has been used for studying mobility, epidemics, health and depression, and information diffusion. Particularly, the call detail records have been used to study different social features, like the network structure, people's sleeping patterns, and response to natural disasters. They provide useful insights about the behaviour of the people involved in the calls. This dissertation is based on four research articles in which a huge data set containing call details records of around 3 million users over a 12-month period in 2007 is analysed, to study the dynamics of the human daily resting periods and human social focus over the life course.

Each day, the calling activity of the mobile phone users follows two different circadian rhythms, each one synchronised to a different clock. On one hand there is the clock of social time, marked by social activities of the daily routine, in which the working and schooling times, opening times of offices, etc, set a specific social schedule to follow. On the other hand, some human physiological processes, like the human sleep-wake cycle, follow a natural 24h cycle, entrained to a biological clock. The calling pattern shows the struggle of living between these two clocks. It follows a specific schedule (it peaks and decreases twice each day, showing a strong dependence on social time). The location and size of the peaks of activity change over the year, by expanding during the summer and shrinking during the winter, thus indicating a seasonal dependence. Moreover, people living in the same time zone but at different locations, are found to start (or cease) their activity at different times, with a difference given by their local sun transit times, thus people living eastward in the time zone have earlier schedules than those living westward.

The emotional closeness between users and their contacted alters can be determined based on their communication pattern. The level of interaction between a mobile phone user and the alters in his/her egocentric network is different, having a dominant interaction with the romantic partner. The features of the ego-centric network and the social focus invested on alters depends on the age and gender of the user, showing clear changes as the users go through different life course stages. Younger people contact more alters and more frequently but this changes noticeably as the egos cross the parenthood stage, in such a way that when egos reach old age, the size of the egocentric network has considerably decreased and it is mainly populated by alters younger than the ego. At the grand-parenting age, an important gender difference appears, when females (probably crossing menopause) show a strong change in social focus towards their daughters, who are in the reproductive stage, whereas males remain focused mainly on their romantic partners, providing supporting arguments for the grand-mothering hypothesis.

**Keywords** Reality-Mining Human-Behaviour Mobile-phone-communication SleepWake-Cycle


**ISSN (print)** 1799-4934 **ISSN (pdf)** 1799-4942

**Location of publisher** Helsinki **Location of printing** Helsinki **Year** 2018

**Pages** 142 **urn** http://urn.fi/URN:ISBN:978-952-60-8200-4
In memory of Ken Lueders;
to my parents Graciela and Raúl, and my siblings Graciela and Raúl;
to Satu.
This four year-long journey on complex systems in Finland started actually five years ago (2013), when my former adviser Professor Rafael Barrio, in an unexpected way, asked me if I was interested in going to Finland to do a PhD on Complex Systems. In that moment, only random thoughts about European capital cities and geography came to my mind, and I just babbled some incoherent answer. I do not know how the answer was interpreted by Prof. Barrio but he just replied “Think about it”. I have to confess my lack of knowledge at that time about Finland and Finnish culture, so after a thorough review about the research on Complex Systems done by the group lead by Professor Kaski and, truth be told, some hours of “googling” queries including the word Finland, I was able to give a proper, positive answer. Five years later and with a doctoral training completed, I still know almost nothing about Finland and its culture, but curiously I understand many things about Finnish people. This apparent contradiction can be understood from a funny sentence written in the recommendation letter that Prof. Barrio gave me when I was applying for the PhD position, to read "About his personality, he is a reserved and silent person very reflexive and honest. These traits make him very adaptable to a Scandinavian society". I can say that, at least, the last part is true, and I am very pleased of having come to this “corner” of the world.

During these years at Aalto University, I enjoyed to learn many things about complex systems, state-of-the-art computational and analysis tools, but the one I have enjoyed the most is to realise that human behaviour, despite its complexity, can be described and even predicted. I will always remember the conversations I had with my workmates, trying to decipher what the reasons behind the generated graphs and observed quantities could be. Many of those discussions and “discussers” contributed to the completion of my doctoral training. I want to thank Asim Ghosh, for offer-
ing me a helping hand and support on difficult times, and for those long discussions about work and common life things. Same to Kunal Bhattacharya, who also always shared interesting and useful comments and solutions on any arising research problem and from which I have learned so many things. I will always remember the group dynamics I had with those two Indian post-docs, even if they didn’t respect social manners by excluding me from their conversations by talking in Bengali. I also want to thank Talayeh Aledaood, for all the times she helped me and offered me support, and for all those conversations and mutual bullying we have had. A special acknowledgement to Professor Rafael Barrio, who guided through the complex systems world before coming here and building the linking bridge that brought me to this endeavour in Finland.

I would like to thank Professor Albert Diaz Guilera for agreeing to act as my opponent during the defence, to Professors Esteban Moro and Michael Szell, for acting as pre-examiners of my thesis, providing invaluable comments on the document. I want to acknowledge the scientists with whom I have coauthored the articles included in this dissertation: Professors Janos Kertesz, Robin Dunbar, and Anna Rotkirch.

My most sincere acknowledgement and gratitude to my supervisor Professor Kimmo Kaski, for providing always useful and novel ideas on how to approach and tackle any research problem, and for the warm support and guidance during these 4 years of doctoral training, ensuring that every place and project on which I have been involved here has been the best possible.

Finally, I want to express my gratitude and love to my parents Graciela Velázquez and Raúl Monsivais and to my siblings Graciela and Raúl Monsivais, for all the experiences we have shared over my life and for being there each time I have needed them. Also, my gratitude to my grandmothers, Carmen who left us sometime ago, and Dolores who sometimes wants to leave us. Lastly, special thanks to Satu (Karoliina) Vepsäläinen, for filling my life with love and happiness.

Espoo, Finland, September 6, 2018,

Daniel Monsivais
# Contents

Preface vii  
Contents ix  
List of Publications xi  
Author’s Contribution xiii  
List of Figures xv  

1. Introduction 1  
1.1 Human social behaviour ..................... 2  
1.2 Digital footprints ........................... 3  
1.3 Social networks ............................ 4  
1.4 Objectives and scope ........................ 6  

2. Computational Social Science 9  
2.1 Big Data .............................. 10  
2.2 Reality Mining ........................... 13  
2.3 Social network analysis ...................... 14  
2.4 Summary .............................. 16  

3. Human resting periods and reality mining 19  
3.1 Mobile phone data as a tool for reality mining ........ 19  
3.2 Periods of low activity in humans ................ 21  
3.2.1 Resting patterns: social time vs. biological clock ... 23  
3.2.2 Geographical factors influencing mobile phone communication patterns ... 26  
3.3 Summary .................................. 29  

4. Adaptation, Evolution, and Social Focus 33
This thesis consists of an overview and of the following publications which are referred to in the text by their Roman numerals.


Author’s Contribution

Publication I: “Tracking urban human activity from mobile phone calling patterns”

Carried the analysis of the data. Designed the experiment and methodology. Main author of the manuscript.

Publication II: “Seasonal and geographical impact on human resting periods”

Carried the analysis of the data. Designed the experiment and methodology. Main author of the manuscript.

Publication III: “Communication with family and friends across the life course”

Major role in data processing, visualization and analysis. Contributed to writing and editing the manuscript.

Publication IV: “Sex differences in social focus across the life cycle in humans”

Contributed to the analysis of the data. Involved in designing the project and the preparation of the manuscript.
List of Figures

2.1 Popularity of three words over the time .................. 12

3.1 Distribution of outgoing calls along the day inside a city ... 24
3.2 Yearly evolution of different quantities characterising the period of low calling activity .......................... 26
3.3 Width of the period of low calling activity for 12 different cities during one year ............................. 27
3.4 Temporal shift of the conclusion and onset times of the calling activity along different longitudes ............... 28
3.5 Fitting of the probability distribution of outgoing calls 29
3.6 Yearly dynamics of the afternoon break period compared with the corresponding ambient temperature .... 30

4.1 Layers in the ego-centric network .......................... 34
4.2 Network of mobile phone contacts generated from the CDRs data set ....................................................... 36
4.3 Number of contacted alters of an ego over different life stages 37
4.4 Age distribution of alters in the contacts network of egos of different age cohort ........................................ 38
4.5 Age dependent call frequency of most frequently called alters as functions of the ego’s and alter’s ages . . 40
4.6 Age-dependent phone communication patterns of egos with close alters .................................................. 43
1. Introduction

In our current society, the modern technology was propelled and is a consequence of the scientific advances in physics, mathematics and chemistry over the past 500 hundred years. After the second half of the 19th century, the development lead to the invention of the transistor and subsequently to the whole electronics technology, giving rise to completely new ways of communications and interactions. Radio and television broadcasters propagate transmissions in one-to-many fashion, allowing to immediately spread information widely in an easy way. With the invent of computers and digital technology giving rise to Information Communication Technology or ICT for short, humans became immersed into a rapidly developing digital world and being able to stay connected and exchange information with each others through an Internet network.

Nowadays no one could deny that societies are experiencing structural changes driven by the opportunities that the digital world provides. Easy and fast access to Internet have reduced considerably the time and effort required to establish long distance communication as well as information search and retrieval that would otherwise be very difficult to achieve. Examples of these today are social network services, video calls, cloud-services, online banking, and online encyclopedias.

The nature of the interaction involved in a telephone call has practically remained the same, namely one-to-one through a bidirectional channel, since its invention in the second half of the 19th century. However, the usability of and accessibility to telephone services have experienced
huge changes, particularly when telephones became hand-held, easy-to-carry devices, and not being restricted to a wired connection. The use of mobile phones was already extended in the 90’s, and when smart-phones with Internet-access capabilities appeared a decade later, mobile phones became an important part of the modern societies with regard to daily communication among people.

Recent technologies, particularly those related to Information and Communication, have been appropriated by the societies and assimilated by their members as well as becoming more and more an integral part of the daily lives of people. ICT and in particular web searching through the Internet, and the use of smart-phones and wearable devices, could therefore work also as indicators of human daily activities. This is due to the fact that the periods when the device owner is active corresponds in general with periods where these technologies are intensively used. Nowadays, almost every ICT device is capable of keeping logs of its usage given by the device user either by storing the information or sending it. These in turn leave traces or digital footprints and contain a lot of useful information, which provides a rather new and unique framework to study human behaviour.

1.1 Human social behaviour

Evolution has given us, millions of years after the appearance of the first primate and more than 300,000 years after the first modern Homo Sapiens, undeniable biological properties that are reflected in the consistent behavioural patterns of the daily lives of humans. There is a vast diversity of old and modern cultures, which demonstrates this consistency in a number of different societies. In spite of the possible differences, social humans show common behavioural patterns among different cultures, like organisation in complex social structures, establishment of social relationships (consanguinity and affinity), philosophy and self-reflection, and complex languages. Particularly the development and use of complex languages have been crucial in human history. Its importance has triggered a continuous development in the ways of communication and as a rule of thumb, any new communication technology can be quickly assimilated by societies.

There is an interplay between language and culture. Language can influence the way that its speakers interpret and process the information
from the outside world, thus contributing to the construction of their own culture, and in turn, culture defines and shapes the language in the course of time. The way that individuals communicate has been used to understand, study and infer human behaviour, as well as the ways the individuals and societies view and perceive the world.

Communication between individuals reflects some level of mutual interest and it may contain information about the (emotional) relevance of the bonding link between them [45]. For each individual (or ego) the intensity and frequency of interactions with his/her socially bonded individuals or alters change over the life course [92]. These changes can be classified in terms of life events, grouped in two main classes, the normative events that are highly expected events and driven by biological or social factors given the age of the individual; and the non-normative events related to unexpected situations, happening to only few or in unpredictable times. Puberty, marriage and parenthood, job entry and retirement, and loss of spouse are examples of the former [11] and relocation, divorce and death of a relative are of the latter.

The changes of the social network over the life course have been described mainly by two different theories. On one hand, the socio-emotional selectivity [14], which considers that humans become more selective in regard to social bonds as the perceived remaining life time of each individual decreases. On the other hand, social convoy theory [45] states that individuals have social bonds of different stability: an inner circle of close and stable relationships which individuals keep over the life, and an outer shell, where bonds with non-emotionally close alters are less stable. Both theories describe or rely on the fact that changes in social bonds depend on the age and social circumstances. Changes in individuals' social network could be classified accordingly to the life events which trigger the change.

1.2 Digital footprints

Today we live in a world, being connected in many ways. Modern telecommunication technologies have brought a number of new ways for connecting people and information access that can not be compared to any previous ways and channels of communication. The Internet and digital technologies have changed the world, boosted the development of other different technologies, and represent a new paradigm in human communica-
ination. In previous times, communication channels did not necessarily keep records of the communication events or of the information flowing through the corresponding channel, as the storing capacity imposed strong restrictions in the amount of information that could be kept and tracked. In the digital era, things are very different. Nowadays, the storage capacity of a common portable micro SD memory ($\approx 16$ GBytes) is enough to keep the 11 million Wikipedia articles (as a reference, if Wikipedia articles were printed in the format of a classical encyclopedia collection, say the *Encyclopedia Britannica* with $\approx 20$ volumes and 1000 pages per volume, around 7000 volumes would be printed [2]). To be able to store and transfer huge amount of information has many advantages to scientists. From high energy physics to bioinformatics, scientists have faced the problem of dealing with enormous data sets and it was until recent years when technology was able to provide enough capacity to store that kind of data. In digital telecommunications, every possible digital interaction or event can be recorded, and every user of a digital device or service has been leaving traces of the performed activities, in other words “digital footprints” in the digital world.

Digital footprints have two opposing sides concerning the use that the handler/administrator of the stored information gives to it. Cyber-security is one of the main issues and concerns in every networked system, as there are plenty of examples of misuse of digital information. The most recent case was the extraction and selling of large number of personal information from Facebook accounts and its use for personalised advertisement to the voters in the US presidential election, with the purpose of manipulating their preferences (Cambridge Analytica case). On the other hand and for non-adversary purposes, digital communication records information has been used by the scientific community to study individual social behavioural patterns [65, 47], mental health problems [82], human mobility [17] and urban spatial structure [52], and migration [35].

1.3 Social networks

The beginning of the 21st century was marked by the boom of online social networks, where now of the order of billions of users interact and share information about their lives, activities, opinions, etc. Massive events can be organised using online social network. It is frequent to find cases where a multitudinous social dynamics or even social movements have emerged
from an online social network, self-organised from local interactions taking place inside the social network and without of a central coordination. The most straightforward example of the social network boom is Facebook and the explosion of its number of users, passing from less than 100 million in 2008 to 2.2 billion in 2018 (April). Other platforms have also a huge number of users, like Instagram (800 million) and Twitter (300 million) and they all are increasing. In addition, online social network service platforms offering network-based messaging have been widely used in the last years, with Whatsapp (1.5 billion) and Facebook messenger (1.3 billion) having the largest number of users.

During the twentieth century many of the current theories and methodologies about social networks were established [34]. The characteristics of social networks, given by the linking bonds between the actors (nodes) belonging to the networks spans over different scales, which reflects different features of the social interaction. At the micro level, each actor interacts in his/her local network (ego-centric perspective), and concepts like that of homophily (similarity with a linked actor) and triads (possible interaction between members of the ego’s local network) are fundamental for understanding the interactions. At mesoscale, the concept of communities emerges, as set of connected actors that share similar properties and connected multiple triadic closures inside the community are present. At macro- or global level, the structural aspects of the whole social network is analysed. Micro and mesoscale studies of social networks were the main focus of attention by social scientists during most of the twentieth century due to the fact that collecting information of the social network was done mainly by surveys and in-place studies, and the difficulties associated with collecting or analysing information about the whole set of connections and actors involved.

The “Small world” problem studied by Milgram [84], included around 296 letters, from which only 64 reached their destination. This experiment, which was an experimental way of measuring the average path length of a network of unknown structure, is an example of the limitations that social networks analysis at macro level faced. The digital era changed this situation for better, giving access to larger amounts of data containing digital footprints of human social interactions. Data on online/digital social networks has been used to understand and describe social networks at global level, and to verify or at least provide supporting arguments of different social and anthropological theories, to mention
the Granovetter’s Weak ties hypothesis [37, 65], the grand-mothering effect [38] and the above mentioned small world problem [26].

1.4 Objectives and scope

This dissertation revolves around the study of human social behaviour from a perspective based on human communication patterns. Human behaviour displays characteristic features, with specific and expected activities as well as social interactions taking place at certain times and with certain peers. Using a large mobile phone communication data set as a proxy for describing human interactions, different social phenomena and relations are explored to understand human social behaviour. Four publications are included in this dissertation and can be classified into two research areas, human resting periods and human social focus.

The research questions and objectives of this dissertation as well as the publications which the thesis is based on, are described as follow:

- Modern humans are entrained to two different circadian rhythms dictating their daily activities: a biological clock and a socially-driven time, and it is not clear to what extent these two ‘clocks’ influence human activity schedules. Despite the fact that each individual has personal habits and timings, there are collective behavioural patterns reflecting different levels of social synchronisation, the so-called “social time”, which determines specific schedules for specific activities (working and schooling times, opening and closing times, etc). The urban mobile phone call activity is described in Publication I and Publication II, and used to determine the periods when people’s calling activity falls to a minimum. Those periods of low calling activity are used to infer the dynamics of human resting patterns, as well as the influence that environmental factors like daylight and ambient temperature have on resting periods.

- Different relations are included in the egocentric network of any individual. It is known that for each individual, the features associated with his/her egocentric network change over the life course. In addition the social focus that the individual as an ego dedicates to each alter of the network reflects the social role that the alters plays in the egocentric network. The intensity and frequency of the mobile communication between individuals is used in Publication III
and Publication IV to describe the changes of the social focus over
the life course, as well as to give insight into human social effort
investment and human evolution.
2. Computational Social Science

Big data is like teenage sex: everyone talks about it, nobody really knows how to do it, everyone thinks everyone else is doing it, so everyone claims they are doing it

---

Dan Ariely

Isolated entities, as many objects have been studied since the beginning of modern science, are not present in our universe. The state and dynamics of a single particle can be mathematically described by the Laws of Physics but only as a model of thought experiment, as there is nothing like an isolated particle. In any scientific field studying any possible object, there is set of other and possibly different objects interacting with it, and in this sense, any object belongs to or is part of a bigger system. The difficulty associated with studying the whole system of interacting objects is one of the main reasons why the so called reductionist approach is a good way to start. In reductionist thinking permeated scientists’ minds, and concepts like “the whole is the sum of the parts”, established clear insight into scope of each analysis. The higher the resolution of an analysis tool is, the more fundamental the extracted information about the building blocks of the system would be. This way of thinking, enhanced by the causal determinism (ultimately summarised in the Laplace’s demon articulation) dominated scientific thinking for many centuries, establishing clear limits for the reach and scope of any research work, with scientists trying to include in their study as much elements as possible without exceeding the capabilities of their analysis. In some way, the complications associated with studying a system depend on the size of the system, and as the analysis capabilities of the researcher and the current state of the art limits the number of systems that could be studied. Deterministic chaos brought a paradigm shift in scientific thinking, showing that for
some small-size systems, non linear effects of the interactions between the elements introduce unpredictability in the long-term evolution of the system. In addition, the knowledge of the individual dynamics of the building blocks of a system did not explain the appearance or emergence of collective global behaviour of certain systems, e.g. flock of birds or neurons in a brain.

The complication associated with the study of a system was introduced by Weaver [87], which states that problems range between two extremes, problems of simplicity on one side and problems of disorganised complexity on the other. Both extremes could be studied with the analysis tools present at that time, the former with basic techniques of analysis and measurements created during the “period in which physical science learned variables”, and the latter using statistical mechanics tools. However, Weaver stated that there were intermediate problems, of organised complexity, which included a number of variables big enough for not being tackled by the basic analysis tools, but small compared to the size of a classical statistical mechanics problem. These problems in the middle of the spectrum, exhibits in general organisational features, which can not be decipher from the simplicity approach, or can be filtered out by a statistical analysis. This class of problems, “involve dealing simultaneously with a sizeable number of factors which are interrelated into an organic whole”, and includes a vast catalogue of phenomena: cell-signalling [32], protein folding [85], brain processes [13], ecosystems [50], societies [76] and economies [6], to name a few, and nowadays are known as complex systems.

Human societies are, in many ways, a complex system. They contain a set of interacting elements or constituents, i.e. individuals, living in different environments and reaching different levels of organisation, in such a way that collective behaviours emerge from their interactions. Language, culture, and economies, are examples of self-organised behaviours emerged from human interactions, and in the last decades, they have been studied from a complex systems perspective, bringing a new branch in the social science research, the complex social systems area.

2.1 Big Data

The origin of the concept of “Big Data” is not well known, but it has been traced to Silicon Graphics premises in the mid 1990’s. Without a clear
agreement of who coined the term, it started to be popularised by John Masley, more for marketing-related purposes on mind, and used, for the first time in an academic context by Sholom Weiss and Nitin Indurkhya, in their book *Predictive Data Mining* [88, 24]. Since then, its usage has been growing continuously in many academic and non-academic contexts (see Fig. 2.1), in such a way that nowadays its a very popular term. However, there is no standard definition agreed by everyone about what Big Data is. Essentially, the definitions include concepts like huge digital datasets, data-collecting systems, special techniques to handle and analyse datasets, impact and relevance in society, among others. In a paper specifically focusing on the definition of the term, De Mauro *et al* proposed that “Big Data is the Information asset characterised by such a High Volume, Velocity and Variety to require specific Technology and Analytical Methods for its transformation into Value.” [22], and listed three main concepts that a Big Data entity should include, namely:

- the three V's (“Volume”, “Velocity” and “Variety”) related to the nature of the information included,
- “Technology” and “Analytical Methods” referring to the tools required to properly handle and analyse the information contained in the set,
- “Value” associated with conversion of the information contained in the data into a valuable product for companies and societies.

The perspectives about the use, nature and definition of Big Data rely on the background and activity of the people dealing with it. In computational social science, social scientists interact with data-scientists, and their respective methodologies and concepts about what Big Data is, in general very different. In his book "Bit by bit, Social Research in a digital age" [74], Salganik exposes this difference starting from the nature of the data itself, and the rest of this section is based on the definitions and ideas embodied in his book. The (Big) Data is in general, *re-purposed* data, meaning that initially it is generated and used for different purposes than research, requiring an exceptional challenge of adapting the data to new purposes, which it was not initially designed for. From social scientists side, as they are used to work with designed data, studying Big Data presents problems and challenges as it is not originally intended for research purposes. On the other hand, data scientists tend to focus on the advantages of the data and ignore possible weaknesses. A hybrid perspective would be the best choice to understand the characteristics of Big
Data, and following this marked line, Big data could be described in terms of different features, grouped into two classes:

- helpful for research: big size, always-on (continuous collection of data), and non-reactivity (unawareness of the subjects that they are being observed).

- problematic for research: inherently incomplete, inaccessibility, non-representative, drifting (the population and/or the data-acquiring system changes with time), algorithmically confounded (induces behavioural patterns introduced by the systems generating the data, disturbing the “natural” behaviour of individuals), dirty (full of junk data), and personal data sensitivity.

Given the previous general features of Big Data, Salganik describe three main strategies (or “research recipes”) to deal with Big Data problems, namely:

- Counting things, with the fundamental implicit requirement of asking “what does deserve the effort of being counted?” with an answer that it must be important and interesting for the research field

- Forecasting and now-casting, which consists of an immense amount of data available to feed predictive models. The particular interest
Computational Social Science

is forecasting, to predict the current state of a system, termed now-casting

• Approximating experiments, to investigate possible causal effects between different variables of the studied system but without doing in-place experiments, estimating the effects from the non-experimental (big) data.

2.2 Reality Mining

Reality mining, in words of the people which coined the term, is “not only about analysing Big Data, but about ensuring that the analysis reflects the reality of the situation and the people involved, while being consistent with a conscientious data-collection approach” [28]. It is a data collecting and analysis technique based on the idea of gathering information acquired by the sensors of mobile devices, to analyse behavioural patterns of the individuals using or interacting with the devices. The origins of reality mining comes from a set of experiments done by Pentland and Eagle from MIT in 2005 by analysing information collected by mobile phones given to 100 volunteers for a period of 9 months [31]. From the status of the phone, the usage of mobile applications and the location of the mobile phone (given as the position of the cell tower), these researchers where able to predict the behaviour of some of the volunteers [29] given the initial state of each volunteers’ day.

Different kind of information of variable size can be generated with reality mining, and it depends on the activity of the collecting device owner/user of the device and on the interest of the reality “miner”. Mobile phones offer a wide variety of sources for data collection, from sensors and communication logs to applications and network usage. GPS and location services are used by mobile applications, web browsers, and marketing companies to predict and offer personalised options to the mobile phone owner to improve user experience. Mobile phones sensors, like accelerometers and microphone, can be used to detect, evaluate and predict physical and mental health. Another important source of information that can be extracted from a mobile phone is communication logs, in the form of calling, text-messaging, electronic mail and social network services in order to infer, for example, people’s sleeping patterns [16, 3].

Reality mining is not only intended to understand and predict individual
behaviour. Digital information is generated all the time from many different sources, and if offers an unique opportunity to infer and study collective dynamics. A straightforward application of Reality Mining is traffic. Using location information from GPS inside vehicles and mobile phones of passengers, and from traffic sensors or police reports, real-time traffic information can be used for many purposes, from optimising the commuting times and traffic prediction, and to improve allocation of transportation resources [28].

Another important use of reality mining is in crime prediction. Historically, information about crimes has been collected by police officers, and these helped to make crime maps and generate crime statistics, which in turn could be used in crime prevention. Nowadays, reality mining offers a set of tools to improve crime prevention, to mention the use of modern algorithms for data analysis, analysis of real-time video data from street cameras, more precise information about current police location, and social network analysis of gang bands [28].

At a bigger scale than the previous examples, one important application of the reality meaning is related with the concept of trend. Twitter is an example of a framework where the trend of a topic can be studied and predicted. In almost real time overall trends, believes and sentiments can be sensed from the twitter activity. A very interesting example of reality mining applied at big scale is the diffusion of true and fake-news, disseminating through the Internet, mainly by the Facebook and Twitter. Logs of communication events are also rich resources where reality mining can extract information. Diverse studies using mobile phone activity logs of million of users have unveiled different social aspects, from mobility [35] and human response to emergencies [7], to differences in close relationships [67] and local and global structure of social networks [65].

2.3 Social network analysis

A network, in simple terms, is a description of a set of linked objects. The objects or nodes in the network could be physical in nature ranging from sub-atomic particles to galaxies or more conceptual entities like individuals in a society or trading companies, and the links represent various kinds of interactions taking place between the nodes of the network. In nature, these networks can be found everywhere, and in particular, any possible collection of interacting living organisms could be described
in terms of a network, from cells and neurons to ecosystems. Humans, among many other species, live in societies, and the interactions and dynamics occurring between the members of the society is an expression of (human) social behaviour. In social sciences, the study of human social behaviour comprises of the analysis of the interactions between the individuals in the social network, and this *structural* approach is the so-called social network analysis [34].

The origins of the social network analysis stems from the work of Comte in the nineteen century, with his own efforts to define a science, namely sociology, focused merely on unveiling the laws of society [34], pointing out the importance of a structural perspective in the study of “the laws of action and reaction of the different parts of the social system”. During the first decades of the twentieth century, the foundations of social network analysis were developed, in particular some of the tools used in sociometry (a quantitative method of analysing human interactions), introduced and developed by social scientists like Simmel, Radcliffe-Brown, Moreno, Hennings, and Warner, among others [34].

During the second half of the twentieth century social network analysis became an attractive field between social scientists, with important contributions from Harrison White and his former students at Harvard University, e.g. Mark Granovetter to mention one. White, a former physicist, and his colleagues developed many formal tools for analysing social networks, and established the basis of the current social network analysis. The use of mathematical and analytic tools for studying social networks boosted the development of the area. By the end of the twentieth century, many physicists and mathematicians started to get involved in this research area, with well-known examples like Watts’ and Strogatz’s [86] famous work on the Milgram’s Small World phenomenon or that of Albert and Barabási on scale free networks [8]. Freeman points out in his book “The development of social network analysis”, that these two famous and seminal works in network science were previously studied in a similar way many years before in the social science area, but clarifies that Watts and Strogatz were unaware of most of the previous studies on small world (particularly that from Pool and Kochen [23]) while Barabási and Albert did not know about the work done by Price [69]. However, one of the main contributions of the above mentioned works is the that they brought the attention of the physicists, mathematicians and computer scientists into the field.
Modern social network analysis, according to Freeman [34] can be described by the following features:

1. The interactions between the social actors motivate and determine the structural analysis of the system
2. It is based on systematic empirical data
3. It strongly depends on the use of graphic imagery
4. It strongly depends on the use of mathematical and computational modelling.

Graph theory has been a research area in mathematics for quite a while, in fact since Euler solved the famous problem of the 7 bridges of Könisberg, in 1736. Since then, graph theory machinery has been used in different fields, from pure maths to chemistry. However it was until the middle of the twentieth century when it was applied to a social context [63, 37], and until the second half of the century it was applied to biological networks. By the end of the twentieth century, with the advances in the study of networks, particularly in telecommunications, biological and social networks, a new field of research, the network science appeared. This field is growing fast, boosted by the continuously increasing computing power and data storage of modern computers. The appearance of massive online social networks, and huge (big)data sets containing human-generated information, the social network analysis has experienced a boom in the last two decades.

2.4 Summary

Cultures, cities, and economies are examples of self-organised behaviour emerging from human interactions, which is one of the key characteristics of the complex social systems. From this perspective their study has brought up a new branch in the social science research, namely that of Computational Social Science. The digital world has expanded the possibilities of interaction within human societies, and given the advances in telecommunication technologies and computers, resulting in huge amount of human-generated information or data is being collected, marking the beginning of the Big Data era. Due to its size and non-reactivity the information that could be extracted from the Big Data is very useful for social analysis, but it is problematic as it is not originally intended for research
purposes and it should be re-purposed. Each digital interaction leaves a digital footprint somewhere. Mobile phone and wearable devices are examples of digital systems collecting information about human activity and behaviour, in a non-intrusive way. Reality mining is an analysis tool that generates, extracts, and analyses the information stored in the digital footprints, indeed for the purpose of studying human social behaviour. In particular, the mobile phone call details records have been used to study different human aspects, from social networks to sleeping patterns. Social network analysis, studying the interactions between individuals in the social network, is a natural framework to understand human behaviour in the digital world. Empowered by graph theory machinery, by the richness of information included in BigData, and by the current computational power, Computational Social Science has been proven effective in corroborating theories of human societies.
3. **Human resting periods and reality mining**

If you torture the data enough, nature will always confess

Ronald Coase

Due to the fast development of mobile phone technology and reduction in its production costs over the past two decades, the use of mobile devices have dominated the telecommunication market, with an estimated of 66% of the world population using a mobile phone in 2018 (and 80% in Western Europe and North America) [1]. Nowadays, mobile phones provides a novel source for collecting human-related data. A common mobile phone can collect information of its usage, from communication events (calling and text messaging records), to data captured from some of the integrated sensors (accelerometers, GPS and compass, ambient light detector), and in the case of smart-phones from the interactions with user applications. Given the richness of information contained and produce in mobile phone, in the recent years several mobile phone data sets have been analysed, trying to unveil different social features and human behaviour. In an extensive review paper by Blondel et al [10], a number of important results in the field of Computational Social Science have been described, all of them based on mobile phone data analysis, such as revealing the social network structure, personal mobility, geographical partitioning and urban planning.

3.1 **Mobile phone data as a tool for reality mining**

Each time when a communication event is established between two phones (a call or a text message), information about the interaction is kept in a log by the mobile phone service provider. This log, usually denoted as Call Detail Records (CDRs) may contain an identifier associated with the phone.
numbers of the participating devices, the time and duration of the event, the initiator of the call, the cost that the event incurred, and descriptors of the accessing and routing points handling the communication. Worldwide, telephone is used on regular basis as means of communication by almost anyone in modern societies [10], which translates into huge CDRs including billions of entries. As CDRs contain information about human interactions, they can be used to study human behaviour.

In every publication included in this dissertation a unique CDRs data set is analysed from various perspectives. The data set was given by a mobile phone service provider offering services in a southern European country. Due to a non-disclosure agreement, both the name of the country and the name of the provider cannot be disclosed. It contains records of every call and text message of the mobile phone service subscribers between January 1st and December 31st, in 2007, and each phone number was indexed by the service provider to keep identity of the subscribers unknown. It contains of the order of 3 billions calls and text-messages between more than 50 million different identifiers. From the whole set of identifiers, around 10 million are associated with the subscribers of the company and the remaining belong to either subscribers of other mobile phone service providers or to land-lines phones. Each entry in the CDR contains:

- the time and date when the event happened
- the identifier of the caller
- the identifier of the callee
- the type of event (call or text-message)
- the duration of the event (if applicable)
- the associated cost.

In addition, the data set includes demographic information, extracted from the user contracts. Each user contract entry contains the age, gender, starting and ending date of the contract (if applicable), zip-code of the address the phone bill is sent, and the geographical coordinates (Latitude and Longitude) of the location of the most accessed cell tower by the subscriber. In some entries of the subscribers contract file some fields are missing, such that the number of subscribers from which all the fields are known is around three million.
Since 2007, this data set has been analysed to study different characteristics of human interaction from the perspective of the complex networks. If the mobile phone users appearing in the data set are considered as nodes and a link joining them is established if communication events have occurred between them, a (social) network could be constructed and subsequently studied. There is a number of works based on this concept using the above mentioned data set, reporting different results. In [65], the relation between the link strength and network structure is analysed, showing that the global integrity of the network is more sensitive to removal of weak links than strong links, and that strong links tend to have smaller betweenness centrality. In [43, 67, 46], the influence of demographic characteristics of the service subscribers are studied using the subscriber contract information available. It is found that the characteristic calling pattern depends strongly on the age and gender of each element participating in the link, with marked differences between different cohorts and between geographically spatially separated groups.

The calling pattern, described in terms of the calling activity and the demographic features of both individuals participating in the communication have been proved to be a fruitful approach to describe some features of the human behaviour [10]. In the next sections, the context of the research done in Publication I and Publication II is provided.

### 3.2 Periods of low activity in humans

Humans living in modern societies are entrained to different circadian clocks, which are 24-hour long periods defining their diurnal rhythms. There are essentially two main rhythms that dictate human daily activities: one representing the biological side, entrained to environmental and sun-based cues which in turn may be subjected to seasonal variations across the year; and another, specified by a local (social) time, where the coordination of social and economical activities impose restrictions to the schedules when these activities should be performed. Humans are required to synchronise their daily life between these two clocks, and it is well known [48, 79] that changes in the timing or period length of these clocks have a strong influence on many different facets of human daily life.

However, it is not clear to what extent human activity in urban, artificially-lighted environments are still influenced by the environmental cues [21,
93, 94], and one way to look into this question is through the mobile phone calling activity. From the analysis of CDRs it can be seen that mobile phone users tend to perform their calling activity within specific time intervals. Similarly, there is a specific time period, namely the nighttime, when the calling activity at population level falls to a minimum. Each day, urban calling activity shows a characteristic and well defined pattern, with a rise of activity from around 9:00am till the noon, after which the activity falls to a local minimum around 4:00pm. After, the calling activity rises again until 8:00pm when it reaches its daily maximum, and then steadily decreases towards the daily minimum, around 3:00am (see Fig. 3.1). This pattern shows consistency for each day of the week and over the whole year, and it corresponds with other human behavioural patterns, like the increase in human alertness [78], human performance [91], and as extreme example, the incidence of working and traffic related accidents [57]. The daily calling activity described by these four periods follows a complex dynamics across the year and along different geographical zones, such that it can be used to provide insight into the dynamics of human resting or sleeping pattern.

From the CDRs, the urban calling activity could be described by constructing a probability distribution function $P_{all}(t, d)$ of finding an outgoing call at time $t$, during the day $d \in (1, 365)$ inside a particular city. In Fig. 3.1, $P_{all}(t, d)$ (green curve) is shown, for two different days ($d = 46, 214$ counted from the beginning of the year 2007), for a specific city with a population size over one million. To take advantage of the observation that the “natural” starting and ending point of the calling activity falls around 3:00am, in the following a ‘day’ is defined from 4:00 am to 3:59 am on the next calendar day.

One finds that $P_{all}(t, d)$ shows two modes, first one corresponding to the calls made during the morning and a second mode being related with the calls during the evening. The two peaks of this (bi-modal) distribution can be separated by two one hour periods when the activity decays considerably, as one is happening between 3:00 pm and 4:00 pm, which could be related to the time when people have lunch, and the other coinciding with the definition of ‘day’, between 3:00 am and 4:00 am, i.e. in the middle of the sleeping time. The first delimiting period will be denoted as the diurnal calling gap $g_d$ and the second as the nocturnal calling gap $g_n$.

The daily calling activity is split into two non-overlapping periods, ‘morning’ and ‘night’, delimited by $g_n$ and $g_d$, each one lasting eleven hours. The
‘morning’ calling activity period is the time period between 5:00 am and 3:59 pm, while the ‘night’ calling activity period happens between 5:00 pm and 3:59 am on the following calendar day.

In order to study the human resting period from their mobile phone communication activity, one should take into consideration that only the rise and fall of the calling activity, i.e. the time when calling activity starts and ceases, respectively. To study the rise of the calling activity during the morning period, one should consider only the time at which the first call is made by each subscriber, and calculate the associated probability distribution of the time of the first call \( P_F(t, d) \). For the cease of calling activity during the night, the corresponding probability distribution for the time of the last call \( P_L(t, d) \) it could be calculated, taking into account only the last call made by each subscriber. It should be emphasised that in this study we include only the calls made by the subscribers (outgoing calls), excluding all the received calls, because these may not depend on the activity pattern of the subscribers. Fig. 3.1 shows a comparison between the probability distribution \( P_{all}(t, d) \) and the corresponding distributions for the last call \( P_L(t, d) \) and \( P_F(t, d) \) for the time of the first call, for a city with a population size around 600,000, during two of consecutive days in the winter (days 46 and 47) and in the summer (days 214 and 215).

3.2.1 Resting patterns: social time vs. biological clock

Artificial lighting, heating technologies, urban environments, and many other technological achievements have provided safer environments to humans for living [90, 21]. Particularly, the artificial lighting is expected to interfere with the natural rhythm of one of most important daily cycles on humans, namely the Sleep-Wake Cycle, whose dynamics is influenced by environmental cues, like daylight and ambient temperature, as well as physiological rhythms, like the body temperature regulation cycle. Studies of modern humans living in pre-industrial, isolated communities have shown that the Sleep-Wake Cycle of humans not exposed to artificial-lighting carries seasonal changes drawn by the yearly variation of the daylight, but, surprisingly the total sleeping time is similar to those found in urban environments [93].

As the sleep-wake cycle plays a fundamental role in many human biological, social, and economical processes [51, 48, 66, 33, 59, 20, 75], it has been in the focus of various studies over the years from very different perspectives [89, 40]. Recent studies on the length, onset and termina-
Human resting periods and reality mining

Figure 3.1. Probability distribution of finding a call at a time $t$, shown for sets of consecutive days in 2007. (Green) Distribution when all calls are included, $P_{\text{all}}$. Distribution when only the last call in the night is included, $P_L$. (Blue) Distribution when only the first call in the morning is included, $P_F$. From $P_L$ and $P_F$, we calculate their means, $\bar{t}_L$ and $\bar{t}_F$, and their standard deviations $\sigma_L$ and $\sigma_F$, respectively. We define the period of low calling activity (PLCA) to be the region bounded by $\bar{t}_L$ and $\bar{t}_F$, and calculate its width $T_{\text{low}}$ as the time interval between $\bar{t}_L + \sigma_L$ and $\bar{t}_F - \sigma_F$. On day 46 (middle of February), $T_{\text{low}} \approx 10.5$ hours, whilst on day 214 (early August), $T_{\text{low}} \approx 9.5$ hours. $P_{\text{all}}$ is delimited by the nocturnal calling gap $g_n$ (between 4:00 am and 5:00 am). $P_L$ lies between the diurnal calling gap $g_d$ (between 4:00 pm and 5:00 pm) and $g_n$, whilst $P_F$ is bounded by $g_n$ and $g_d$. Taken from [61].

The study of the sleeping period (chronotype) has been focused on analysing the results of experiments on individuals under controlled situations and in information gathered via fill-in questionnaires about the sleeping habits of volunteers [66, 20, 73, 49]. This approach has generated a good understanding of the Sleep-Wake Cycle and had been proven effective to assess some of its salient features. Nevertheless, these approaches have the disadvantage of studying the individuals outside the subjects’ normal daily routines (i.e. controlled environment case) or being possibly biased by each individual subjectivity (fill-in questionnaires). These problems could be avoided to a large extent with the help of the Reality Mining, as it is a non-intrusive way of collecting information of human activity and human social behaviour.

On the other hand, huge techno-socially generated data sets have been produced till recently with increasing pace, which has in turn facilitated
the study of human behaviour from diverse perspectives applying so called Reality Mining techniques. CDRs are activity logs from which human behaviour could be studied and analysed in-vivo, as the subscribers are not aware of the upcoming studies on their calling pattern, nor they have to fill or provide any information about their behaviour. CDRs have been used to study intrinsic mental health [83], social networks [46, 30, 42], sociobiology [4, 18, 9, 5], as well as behaviour of cities [52, 80], all of which demonstrate this new line of research on human behaviour and sociality is not only very promising but also giving rise to a novel yet complementary perspective to studies of social systems and networks.

Using the calling time distributions defined in the previous section, an estimation of average waking and sleeping times can be calculated. Both, \( P_F \) and \( P_L \), describe the characteristic pattern when subscribers made their first and last call, respectively, and their associated mean time \( \bar{t}_F + \bar{t}_L \) represent the expected behaviour. Despite the fact that the expected time when each subscriber made the first (last) call does not coincide with the time when the subscriber wakes up (falls asleep), it represents an upper (lower) bound to that activity. In Fig. 3.2, the mean time of the first and the last calls for some chosen cities is shown, during the whole one year period. Surprisingly, they carried a strong seasonal variation.

From 3.1, one can see a long period of low calling activity (interpreted as resting time) bounded between the termination of the calling activity during night hours of one day, and the rise of calling activity during the morning on the next day. A possible estimation of the length of this resting period, is to measure the time interval between the decay of the time of the last call distribution (given by \( \bar{t}_L + \sigma_L \)), and the begin of the activity next day (\( \bar{t}_F - \sigma_F \)), where \( \sigma_F, \sigma_L \) are the standard deviations of the corresponding distributions. Defining this period of low calling activity as \( T_{low} \), then

\[
T_{low}(d) = 24 - (\bar{t}_L(d) + \sigma_L) + (\bar{t}_F(d + 1) - \sigma_F). \tag{3.1}
\]

In 3.3 the period of low calling activity \( T_{low} \) for 12 cities with a population size greater than 100,000 is plotted. For every city, \( T_{low} \) shown a seasonal variation, with a minimum around the summer solstice (day 171) and a maximum near the winter solstice (day 356), with the dynamics resembling the length of the night (24 hours minus the length of daylight). For some cities, the difference between \( T_{low} \) during the summer and during winter days is as long as one hour, showing that human resting periods (proxied by their calling activity) is still sensitive to the yearly
Figure 3.2. Yearly evolution of different quantities characterising the PLCA (top panel) compared against the yearly temporal variation of 3 different solar-based events (bottom panel). (Top panel-left column) $t_{\phi}^F$ – average of the mean time of the first call at different latitudinal bands ($\phi$) denoted by the different colours. (Top panel-central column) $t_{\phi}^L$ – average of the mean time of the last call. (Top panel-right column) $t_{\phi}^{LF/2}$ – average of the centre of the PLCA. In the bottom panel, times for sunrise (left), sunset (centre), and solar midnight (right) along the year are shown for the reference point located in the middle of the central latitudinal band. The shape of $t_{\phi}^{LF/2}$ resembles to some extent the solar midnight, coinciding with the two minima (days 130 and 302) and one of the maxima (day 210). Also, $t_{\phi}^L$ has a similar behaviour when compared to the solar midnight. For the case of $t_{\phi}^F$, the curve has a correspondence with the sunrise although in a lesser extent. The discontinuities introduced by the daylight saving is present in the curves, suggesting that the PLCA is not solely influenced by the socially-driven time, and are synchronised with an external (astronomical) event. The number of cities inside the band $\phi = 37^\circ 30'N$ (blue), $40^\circ 20'N$ (green), and $43^\circ 0'N$ (red), are 7, 6, and 8, respectively. Taken from [61].

variation of the daylight.

3.2.2 Geographical factors influencing mobile phone communication patterns

As it was mentioned in the previous section, the calling activity of people living in urban environments, described by the calling distribution $P_{all}$, has a characteristic, bi-modal shape, which is present for every day and every city big enough. Similarly, $P_L$ and $P_F$ have a specific shapes that are found on every city and every day. In spite of this preserved similarity of the shape of the distributions, their mid-points and widths change from day to day and from city to city.

An interesting observation from different calling distributions is that they are not centred around the same time, even though they have very
similar shapes (see Fig. 3.4). The characteristic mobile phone calling activity of different cities rise and cease at different times. There is a quantifiable time delay between the calling activities of different cities, related to the geographical constraints. For cities lying at similar latitude but at different longitude but still restricted to the same time-zone, the calling activity always starts/ceases earlier for those cities located at eastern longitudes, and the time difference between the onsets and terminations of the calling activities of two different cities correspond quite closely with the time difference between the sun transit times of the places, as it can be seen in Fig. 3.4. This fact is in agreement with the observation that the average mid-sleep time of people living at different longitudes but within the same time zone follow the East-West progression of the Sun [73], and the use of low calling activity periods as a proxy for studying human resting periods seems plausible.

Besides the low calling activity period taking understandably place during the night, there is a noticeable and consistent fall in the calling activity around 4:00pm, as can be seen in Fig. 3.5. This period, associated with after-lunch time, corresponds with the period when human perfor-
Figure 3.4. Temporal shift of the conclusion and onset times of the calling activity along different longitudes. Probability distributions of the time of last call $P_L(t,d)$ and of the first call $P_F(t,d)$ for 5 different cities lying at the same latitude but at different relative longitudes (provided in legend) from a reference point located at the second city from east to west in the band for two different days during the year. (Top) Original distributions. (Bottom) Distributions shifted by a time corresponding with the difference between their local sun transit times ($30.8, 15.6, 11.2,$ and $-12$ minutes for the cities located at $-7.7^\circ, -3.9^\circ, -2.8^\circ, $ and $3^\circ$ from the reference city, respectively). The collapse of the distributions onto the reference city’s distribution is evident when the longitudinal time shift is added. This collapse implies that these 5 cities cease (and begin) their calling activity in a synchronised way, with a temporal phase corresponding with the difference between their sun transit times. Taken from [60].

mance and alertness decrease [25, 58], causing rise in work and traffic accidents [57], and drowsiness and napping times are usual [12]. In reference [12, 93] the influence of ambient temperature on the resting periods has been discussed and shown that in pre-industrial societies ambient temperature influences the timing and length of the resting periods [93]. The length of this period can be estimated approximating the probability distribution of outgoing calls $P_{all}$ by a function containing two gaussians (as depicted in Fig.3.5), given by the following expression

$$F(x) = \frac{a_0}{\sigma_M \sqrt{2\pi}} e^{-0.5((x-\bar{t}_M)/\sigma_M)^2} + \frac{a_N}{\sigma_N \sqrt{2\pi}} e^{-0.5((x-\bar{t}_N)/\sigma_N)^2},$$

(3.2)

were $\bar{t}_M$ and $\sigma_M$ are the mean and the standard deviation of the noon-centred mode and $\bar{t}_N$ and $\sigma_N$ the corresponding values for the evening-centred mode. Therefore the length of the low activity period during the
Figure 3.5. Probability distribution of outgoing calls during the day 265 (Sunday, orange line) in a city (of label 2), fitted by a superposition of two normal (Gaussian) distributions (black line), the first centred round noon and the second one in the evening (8 p.m.). The fitting is done with the mean values $t_M$ and $t_N$, and the standard deviations $\sigma_M$ and $\sigma_N$, corresponding to the noon and evening-centred activity modes, respectively (see details in the text). The afternoon break period $T_{\text{break}}$ is defined as $(t_N - \sigma_N) - (t_M + \sigma_M)$. Taken from [62].

The afternoon ($T_{\text{break}}$) is estimated as the time period between the two activity peaks, expressed by the following formula

$$T_{\text{break}} = (\bar{t}_N - \sigma_N) - (\bar{t}_M + \sigma_M).$$

(3.3)

In Fig. 3.6 the seasonal changes of the low activity period $T_{\text{break}}$ for different cities over the period of one year, reaching its maximum around August. In addition, the daily maximum ambient temperature in the analysed cities is shown (with the hottest days during the summer being around August) depicting the influence of temperature to the afternoon period activity $T_{\text{break}}$, stretching it during the hot days.

3.3 Summary

Mobile phones constitute a rather unique source of human behavioural data as they can collect user information such as communication events, data from sensors, user applications, etc. Very large data sets of Call Detail Records from mobile phone communication, containing identifiers of the pair of individuals involved in the communication event, the time, duration, and possibly location of the event, are stored worldwide by the ser-
Figure 3.6. Yearly dynamics of the afternoon break period compared with the corresponding temperature time-series, for 4 different cities lying at different $37^\circ N$, $40^\circ N$ and $42.5^\circ N$ latitude. The lines with markers represent the length of the afternoon break resting along the year, whilst those without markers represent the max temperature in those cities. Each city is represented by the same colour in both time-series. Wednesdays and Sundays are shown. Taken from Publication I
patterns. The calling activity shows each day two periods of high activity, one around noon and another in the evening. The rise and decrease of the calling activity is described in terms of the mean time of the first and of the last call, respectively, and both quantities are used to bound the onset and termination of the nocturnal resting period. It is found that both mean times carry a strong seasonal variation. In addition, the characteristic nocturnal period of low calling activity, which delimits the nocturnal resting period, also experiences seasonal changes, closely following the yearly variation of the duration of the night. The rise and decrease of the calling activity, indicating the onset and termination of the nocturnal resting period, occur at different times in different cities that are located inside the same time zone. The nocturnal resting period occurs first at the Eastern-most cities and progressively advances towards the Westernmost places, with a time difference corresponding with the time delay between the sun transit times at the different places. Another period of low calling activity is found each day, occurring around 4:00pm. The length of this afternoon period of low calling activity also changes over the year but its dynamics is linked to the yearly variation of the ambient temperature, increasing its length during the hot, summer days.
4. Adaptation, Evolution and Social Focus

Theorems of Psychohistorical Quantitiveness:

- The population under scrutiny is oblivious to the existence of the science of Psychohistory.
- The time periods dealt with are in the region of 3 generations.
- The population must be in the billions ($\pm 75$ billions) for a statistical probability to have a psychohistorical validity.

– Hari Seldon

Social networks reflect social interactions between different actors occurring at different levels. For any individual belonging to a social network, there is a local, ego-centric network which includes all those actors (alters) of the social network which interact directly with the individual (ego). The interaction through these bonds may reflect the level of emotional closeness or social interest between the ego and the linked alters. The frequency and strength of interactions, particularly in terms of communication, have been used to describe the emotional intensity between the ego and the alters [71, 70]. It has been proposed that emotional closeness distributes the alters of the egocentric network in different layers, where the social investment and emotional closeness with an alter decreases as the extent of the layer increases [27, 39]. 6 layers are proposed [95], varying in size and importance to each individual, namely: support clique (3-5 alters), sympathy group (12-20 alters), band (30-50 alters), clan ($\approx 150$ alters), megaband ($\approx 500$ alters) and tribe ($\approx 1500$ alters), depicted in Fig. 4.1 Any individual has limited resources of time and effort to divide with the individuals of his or her egocentric network. The divi-
sion of the social investment is uneven, such that emotionally closer alters are give more attention, thus to study the way humans split their social focus may provide useful insights into human social relations and social behaviour. Mobile phone communication, particularly call detail records (CDRs), offers a powerful substrate for studying human social dynamics, as it allows to give a closer look into direct interactions between pairs of individuals [29], with reciprocal information about the pairwise interaction that traditional questionnaire-based studies do not offer [56]. In addition, the CDRs that have been used to study human behaviour from their communication patterns are massive sets of data [47, 56, 67], providing a larger number of samples and thus more confidence to statistical analysis.

Figure 4.1. Egocentric network surrounding an ego is hierarchically distributed in layers, showing the social investment and emotional closeness between an alter belonging to alter and the ego.¹

### 4.1 Social focus across the life cycle in humans

Humans are social animals, living in large social groups. For any individual, the intensity and emotional closeness with his/her social relationships depend on many different biological and socio-economical factors, with the homophily, i.e. the propensity of having social bonds with individ-

¹Some fragment copied from GNOME icon artists (https://commons.wikimedia.org/wiki/File:Gnome-stock_person.svg), under Creative Commons License 3.0
uals that share similar features, being a fundamental aspect [54]. Among these factors, age and gender play decisive roles when deciding with whom to invest time and effort in keeping social relationships [39, 72]. The effort invested on each social bond is subjected to cognitive and temporal constrains, which limits the size and intensity of interactions that each individual could have with his/her egocentric network [27, 81]. Communication is one of the mechanism to maintain and reinforce social bonds, and the strength and frequency of communication is a straightforward way of inferring emotional closeness in social relations [70, 67].

ICT has brought new options for human-human interaction, and particularly mobile phone communication plays an important role in social networks, allowing instant communication between people regardless of their spatial location. Given the current adoption of the mobile phone technology (up to 80% penetration in western European countries), the set of mobile phone contacts of an individual could be a good descriptor of his/her true ego-centric network. In Fig. 4.2, a small sample of the network of contacts of mobile phone users (taken from Palchykov et al [67]) is shown. Individuals (represented by circles) are connected by a link if they were involved in at least one communication event during a 7-month time period. Different ego-centric networks can be observed, as well as the strength of the interaction, measured as the frequency of events between each connected pair.

The ego-centric network of the 50-year old female appearing in Fig. 4.2, shows different features that are expected for any individual of her age and gender cohort. She has strong connections (in terms of frequency of events) with alters of different characteristics: one with a 56-year old male (blue circle), and with two younger alters (23 year-old female and 20-year old male). In terms of the natural, expected life course of the pinpointed female, the older male alter could be expected to be her romantic partner, and the younger alters to be her children. Other less-contacted alters could be from very different social environments (parents, school and workmates, neighbours, etc) but the previous assumptions about her closest family are justified. The pattern seen in her communication profile with those alters, indeed, are consistently found in other people belonging to the same cohort. At the individual level, the egocentric network of each mobile phone user can be unique and particular, but at cohort level a lot of similarities can be found between the individuals of the group, originated from the basic evolutionary aspects [68]. It is expected that social focus
Adaptation, Evolution and Social Focus

Figure 4.2. Network of mobile phone contacts generated from the CDRs data set. Egos are represented by circles (females on red, males on blue, and grey means that the information of the ego is not available), and the age of the age is shown inside the corresponding circle. If there has been at least one communication event between two egos then a link (line) is connecting them. The numbers over the lines represent the number of events between each pair of connected egos over the studied period. Taken from [67]

Changes over the life course, as a consequence of the different roles and social strategies of the individuals belonging to a society, and it should be possible to detect these changes from their communication patterns.

Analysing the CDR data set, various features of the ego-centric networks can be quantified, detecting characteristic differences between the ego-centric network associated to different age and gender cohorts. In Fig. 4.3 the size of the ego-centric networks are studied. Considering as significant contacts those alters with whom each ego has reciprocal calls during one month period, the average size of the egocentric network is around 15 (grossly corresponding with the second Dunbar’s layer in ones personal networks [81]). However, it can be seen that characteristic size depends on age, reaching a maximum for 25-year old egos, and steadily decreasing for older age cohorts, with a small plateau between 45- and 55-year old groups, where the network size stops decreasing. Curiously, the age distribution of the population (mobile phone subscribers) peaks around the age of 33 years, and the excess of individuals in that age range is not reflected in the characteristic size of the ego-centric network (where 25-year old cohort shows the maximum value, as mentioned before). Gender differences in the characteristic size of the egocentric networks are also visible from Fig. 4.3. At younger ages (<40 years), males are more connected, but the shrinking of the ego-centric network size is stronger as
they aged, in such a way that after the age of 40, the situation is reversed, and females have larger ego-centric network sizes.

![Figure 4.3](image-url)  
**Figure 4.3.** Number of contacted alters of an ego over different life stages. The monthly average number of alters with the age (in years) of the ego when ego’s gender is not taken into account (left) and including gender (right). Male and female egos are denoted by blue squares and red circles, respectively. The dashed lines in the background are used to demarcate the different regimes. Taken from Publication II

Also other features of the ego-centric network of each individual change with age, with new members entering to it, other leaving from it, as well as the intensity and emotional closeness with each social bond. An example of these changes can be seen in Fig. 4.4. The plots show the probability distributions of finding an alter of a given age in the egocentric network, for egos belonging to one of the six different age cohorts 19–21, 29–31, 39–41, 49–51, 59–61 and 69–71 years. These distributions have in general two most-likely values, i.e. showing bi-modality and corresponding to alters of similar age as the egos’ age, and another, representing alters either of the previous generation (parents) or of the next generation (children). For the age cohort 49-51 year old, the distributions have three peaks, representing the middle-age egos which are old enough to have adult children and young enough to have parents still alive.

From Fig. 4.4 it is clear how these peaks shift as the age of the associated ego cohort changes, with a clear disparity on the relative heights between the peaks. For younger cohorts, the most prominent peak corresponds to alters of the same age, that is, the egocentric network is mainly populated by alters of similar age as the ego. For age cohort 59-61 and older, the situation is reversed, and the egocentric network is more popu-
lated by alters of younger age, presumably egos’ adult children. The size and composition of the egocentric network changes with age, giving clues of the social focus dynamics over the life course. In Publication III and Publication IV the intensity of the ego’s social focus with their linked alters is described by different properties of their communication events, confirming the important role that age and gender play in social focus.

Figure 4.4. Age distribution of alters in the contacts network of egos of different age cohort. The plots correspond to the egos’ age cohort (a) 19–21, (b) 29–31, (c) 39–41, (d) 49–51, (e) 59–61 and (e) 69–71 years. In each case, the number of alters in each different age group is counted and normalise by the total number of alters (male and female). The different symbols and colours used to denote males (M) and females (F) in the ego–alter pairs are: upright orange triangles (F–F), red circles (F–M), blue squares (M–F) and downward green triangles (M–M). The distributions were calculated over a monthly time window and averaged over 12 months. Taken from Publication II.

4.2 Adaptation, evolution, and social focus

The stages of human life course follow similar dynamics [92] practically in every culture, and during each different stage, individuals establish, keep and change social links with other members of the society. In particular, peer related links (close friends and romantic partners) and family links (parents, siblings and offspring) play crucial roles in each individual’s life [53, 36, 77].

The interest and effort invested in the social relationship depend on the type of social links between an ego-alter pair. Parent-child bond plays a crucial role over the life course of a person, particularly during the childhood and adolescence periods [41, 44]. In the case of the social bond between the mother and her daughter, the former tries to maintain a closer relation when the latter has grown-up and became a mother herself [15].
During young adulthood, social focus with peer relations is particularly intense, as is expected that individuals start a family (union formation) and strongly rely in the relation on close friends and spouse [55]. Parenthood, after union formation, induces changes in social focus, and over the next life course stages a reordering and variation in the intensity of the social focus will happen, particularly for females [92, 64].

In Publication IV, the dynamics over the life course of some of the previous social links are studied in terms of their mobile phone communication. Given the age and gender of the mobile phone subscriber, he/she can be classified into an expected life course stage, and the corresponding egocentric network is explored looking for those alters who represent a possible close relationship for that particular age. Once the set of special alters are found, a statistical analysis can be done to find differences in the characteristic social focus that each age cohort invest in the close social relationships. In Fig.4.5, age distributions of the most contacted alters (female top, male bottom) for different ego’s age and gender cohorts are shown. It can be seen how, as the age of the ego progresses over the life course, the characteristic features of the most contacted alters also vary. For younger egos of both genders, the most contacted alter of the same and of the opposite gender have age similar to that of the ego. After the ego crosses the age of 45, the most contacted alters could be either one of the same age as the ego or one belonging to a younger generation, born around 25 years after the ego (probably a son or daughter). Interestingly, the chance that the most contacted alter belongs to, a younger generation becomes dominant at age 55 and older, but this effect is only present for female egos, indicating that after middle adulthood, male and female social foci take different course.

Given the age of an ego (in the mobile phone subscribers set the age range is taken from 18-80), a possible classification of the life course stages is as follows [19]:

i) Early young adulthood, 18-21 year old. Characterised by the sexual maturity, involvement in secondary or tertiary education, or entering to the labour market.

ii) Union formation, 22-28 year old. The majority of the individuals start a long-term strong romantic relation, with cohabitation expected.

iii) Middle adulthood, 29-45 year old. From the arrival of the first child
Figure 4.5. Age dependent call frequency of most frequently called alters as functions of the ego’s and alter’s ages. (top) Case for female egos. (bottom) case for male egos. As the ego’s age increases the alter’s age also changes, suggesting the presence of three distinctive family generations in the ego’s social network. Taken from David et al [19]

to the point when the last child reaches adolescence.

iv) Post-reproductive late adulthood, 46-55 year old. Start from the reach of their own children to union formation stage, their own parents reaching old age, and in the case of females crossing to the menopause.

v) Grand-parenting, 56-75 year old. Marked by the arrival of the first grandchild, the exit from the labour market, and the possible death of the own parents.

vi) Old age, 75-year old till death. Grand children start leaving childhood and the onset of old age illnesses.

For each ego, the different close alters are defined based on the age
difference with the ego, falling in three possible age generations: older, same and younger generation, centred around a difference of -25, 0 and 25 years, respectively. Six possible relationships are defined [19] as follows:

1. *father* and *mother*, corresponding to the older generation, no less than 25 years older than the ego.

2. *partner*, *i.e.* romantic partner of opposite gender than the ego. From females, the age difference with her romantic partner should be in the range $[-2, 5]$ years and for males in the range $[-5, 2]$.

3. *daughter* and *son*, of around 25 years younger than ego.

4. *best friend*, with same gender as the ego and no more than one year age difference.

In the previous definitions an important relationship is excluded, namely the ego-sibling link as it is difficult to detect from the analysis of the mobile phone network. From the egocentric network, it is not possible to determine with a high degree of certainty if an alter with age closer to the age of the ego (but not the same age) is a friend or a sibling, so they are not included in the closer alters to be analysed.

Different quantities describing the intensity of the communication link between the ego and the six possible close alters are shown in Fig. 4.6 for different age and gender cohorts. The number of calls, the average fraction of calling time spent by the ego with the alter, call initiation imbalance, and the average time per call with the alter are calculated over a 7-month period. The different plots contain a lot of information about the changes the ego’s social focus carry over the life course, as well as some gender differences about the importance that each close relationship has on the ego. Grouping them by different life course stage, some of the main observations are listed next.

- During the young adulthood (stages $i$ and $ii$), the number of calls and time per call rise between the ego and all the possible relationship, but the fraction of time spent with the partner increases while for the other relations (parents and best friend) is reduced. Despite the increase in the intensity of the bonds (in terms of the number of calls and duration), during this stage the romantic partner bond increasingly receives more attention (in terms of the fraction of time
Adaptation, Evolution and Social Focus

- During the middle adulthood (iii), the fraction of time dedicated to the romantic partner decreases, and correspondingly increases in other relations, but particularly strong with the best friend. The most important change is in the balance of call initiation with parents, which was overwhelmingly negative (mainly parents initiating the calls) during the previous stage, while in this case it is balanced or even becomes positive. It is in this stage when the arrival of children happens, and a possible requirement of support of the ego from the parents could be the explanation of the reversal in the call initiation balance.

- During the post-reproductive late adulthood (iv), grand-parenting (v), and old age (vi), there is continuous decrease in the number of calls, particularly with the romantic partner. The fraction of time spent with an alter shows an increase in the relationship with their children while it decreases the relationship with the spouse, which is comparatively strong in case of females. Only in the last stage (vi) the fraction of time spent with the spouse increases slightly again.

In addition, there is a strong gender disparity in some of the ego-alter relationships in the grand-parenting age cohort. Egos in the grand-parenting age, both females and males show a similar change in the social focus towards the son, with an increase of the social focus compared with the previous life course stage (middle adulthood). This change is also found in the social focus towards their daughter, but is considerably stronger in the case of females (mother-daughter). At the same time, there is a disproportionate decrease in the fraction of time dedicated by a grandparenting female to her spouse, while in the case of males there is only a small decrease. Both observation combined provide evidence to support the evolutionary theory of the grand mothering effect [38], which establishes that menopause, which overlaps with the expected time when women are becoming grandmothers, was an advantage in human evolution, giving grandmothers a chance to provide alloparental care to their childbearing daughters. This in turn allowed their daughters to dedicate more effort towards increasing the number of offspring. If grandmothers dedicated more social effort to their daughters than to their romantic partners, there should be a noticeably change in female social focus towards
these pairs when she crosses menopause.

![Figure 4.6](image)

**Figure 4.6.** Age-dependent phone communication patterns of egos with close alters, namely “mother”, “father”, “romantic partner”, “best friend”, “daughter”, and “son”, using four different measures: the number of calls, the average fraction of total phone call time, the balance between out and in calls, and the average length of time per call. Taken from David et al [19]

### 4.3 Summary

Using the mobile phone data set, different relations forming the egocentric network of the subscribers are studied to understand social focus in humans over the life course. Individuals have limited resources of time and effort to divide with the members their egocentric network, with emotionally closer alters receiving more attention, and the social focus spend on each social relationships depends on biological and socioeconomical factors, with age and gender of both, individuals and members of their ego-centric network playing a decisive role. However, social focus changes over the life course of any individual, and the possible changes could be detected from his/her communication patterns. Measured over one month, the average degree (size of ego-centric network) of mobile phone subscriber is around 15, but the size changes with age. 25-year old subscribers, have larger networks and the size decreases almost continuously until the old age. Males have larger networks than
females until the age of 38, after which the situation is reversed. Similarly, the features of the egocentric network change over individual’s life course. For younger people, the egocentric network is mainly populated by alters of his/her same age group, but as an ego grows older, alters belonging to a younger generation appear in his/her egocentric network, until it becomes the most populated group in his/her egocentric network. Over the stages of human life course, individuals establish, keep and change social links with other members of the society. In particular, peer related links (close friends and romantic partners) and family links (parents, siblings and offspring) play crucial roles on each individual’s life. Given the age of an ego (a mobile phone subscriber) a proposed classification of the life course stages is: early young adulthood, union formation, middle adulthood, post-reproductive late adulthood, grand-parenting and old age. Different quantities describing the intensity of the communication link between the ego and the six possible close alters, at different life course stages are studied during a 7-month period. Younger egos spend a bigger fraction of the calling time with their best ranked alter. After the twenties, the number of calls steadily decreases until old age, but the fraction spent with most contacted alter decreases only until the age of 40. After, it increases again to values similar as those of younger egos. In addition, it is shown that individuals distribute their talking time inside their corresponding egocentric network is very different for different age and gender cohorts. Older people tend to split their calling time more unequally, whilst comparing genders, females have more unbalanced splitting. A strong change in social focus towards their children is found for grand-parenting individuals, considerably strong in the female-daughter relation. In addition there is a disproportionate decrease in the fraction of time dedicated by a grand-parenting female to her spouse, situation not present in the case grandparent male to his spouse. Both observations provide supporting arguments to the theory of the grand mothering effect.
5. **Scientific Results and Conclusions**

This last chapter briefly describes the results obtained in the publications I-IV, and concludes with an extract of the main ideas presented in this dissertation. All the results reported in the publications were obtained from a mobile phone data set containing call detail records (calls and text messages) from more than 10 million subscribers of a service provider during one year (2007). From these, information about the age, gender, postal code and location of the most accessed tower of around 3 million subscribers is known.

5.1 **Human resting patterns**

In Publication I and Publication II the calling activity of urban people living in cities with more than 100,000 inhabitants has been analysed. The periods where the calling activity falls to a minimum are used as a proxy for studying the dynamics of human resting patterns. Each day, the calling patterns of the analysed urban populations consistently show two periods of high activity, one occurring around noon and another around 8:00 pm. These peaks of activity are bounded by two periods where the activity falls to a minimum, one occurring around 4:00 am and the other 12 hours later around 4:00 pm. The rise of the calling activity during morning is described in terms of the mean time of the first call done by the analysed mobile phone users during that period. Similarly, the decrease of the nocturnal calling activity is described in terms of the mean time of the last call done during the night. These two quantities are used to determine the onset and termination of the nocturnal resting period.

---

*Say it. The Universe is made of stories, not of atoms*

______________________________

Muriel Rukeyser
In Publication I, it is shown that during the night, the calling activity ceases at different times in different cities, occurring first in the eastern-most cities and last in the western-most cities. In this study all the cities lie in the same time-zone hence people in them ‘live’ in the same social time, but their termination of calling activity occurs with a time difference corresponding with the time delay between their sun transit times, that is, following the East-West progression of the Sun. This suggests that the sun still acts as a cue in the synchronisation of the sleep wake cycle. The mean times of the first and last call follow seasonal changes, and for the latter, its yearly variation resembles the dynamics of the solar midnight, likely using this as an external cue to synchronise with. In addition, the average chronotype of urban users is approximated by the time period between the mean time of the last call and the first call on the next day. It is shown that the chronotypes and mid-sleep times of urban people depend on the age and gender, as well as on the presence of social jet-lag (lack of sleeping hours during weekdays compensated with oversleeping during weekends).

In Publication II, seasonal and geographical differences of the resting periods are explored. The nocturnal resting period, delimited by the nocturnal period of low calling activity, changes over the year, and it closely follows the seasonal variation of the duration of the night (24 hours minus the length of the daylight), with both quantities being strongly correlated. It is shown that the difference between the length of the resting period between the summer and winter is not the same for all the analysed cities but depends on the latitude of the city, with the difference being larger for southern cities. In addition, the period of low calling activity occurring during the afternoon (between the peaks of activity around noon and 8:00 pm) is analysed. Contrary to its nocturnal counterpart, during this period the calling activity does not decay to almost zero, but to a persistent lower activity level. Nevertheless, it appears consistently in the daily calling activity in all the analysed cities. It is shown that the length of this afternoon period of low calling activity changes over the year, and this change is linked with the dynamics of the ambient temperature over the year. A rise in the daily ambient temperature induces an increase in the afternoon period of low calling activity, and surprisingly this effect is only present when the maximum ambient temperature of the day is above a certain threshold value, around 25°C. In addition, total daily resting period is calculated by adding up the afternoon and nocturnal resting periods, and
Scientific Results and Conclusions

it is found to be almost constant. This shows an interplay between the resting periods, where an increase in the afternoon resting is counterbalanced by a decrease of the nocturnal resting.

5.2 Communication and social focus over human life course

In Publication III and Publication IV, the social focus of humans is studied from their calling patterns. Using the mobile phone data set, different communication properties of the links forming an egocentric network of the subscribers are extracted and compared, to unveil the dynamics of the social focus over their life course.

In Publication III, it is shown that the average number of contacted mobile phone alters in a month (degree of the focal node) is around 15, which is similar to the expected number of alters inside the second layer of the face-to-face egocentric personal networks (Dunbar layers) during one month. The size of the egocentric network changes over the life course. It reaches its maximum when the egos are 25-year old, and then decreases almost continuously until the old age, but with a short period in the middle from 45- to 55-year old when the size of the ego-centric network seems to level off or stabilise, after which it continues decreasing. The dynamics of the size of the ego-centric network (the number of linked alters) shows strong gender differences. Before the age of 38 years, males have more alters and after that age the situation is reversed, with females having larger networks. The features of the egocentric network change over the ego’s life course. For younger egos, the alters mainly belong to the same egos’ age group, and to the lesser extent to the previous, older generation (born 25 years before). As egos grow older, alters of a younger generation appear in their egocentric network, such that for 60-year old egos, the number of alters belonging to the younger generation dominates the egocentric network. The communication activity of each ego with his/her alters, particularly with the most contacted alter changes with the ego’s age. Younger egos talk more time and spent a bigger fraction of that time with their most contacted alter. After the twenties, the number of calls by the ego steadily decreases until old age, but the fraction spent with most contacted alter decreases only until the age of 40 years. After that, it increases again to the values similar to those of younger egos. In addition, it is shown that individuals distribute their talking time inside their corresponding egocentric network very differently for different age and
gender cohorts. On the other hand older people tend to split their calling time more unequally, whilst comparing genders, females have more unbalanced splitting.

In Publication IV, the evolution of the social focus over the life course is studied, taking into consideration those alters of the egocentric network, which are emotionally close to the ego. Based on gender and age difference between each ego and the corresponding alters, six possible close peers are defined, namely mother, father, son, daughter, best friend (same age and gender) and romantic partner. In addition, six age cohorts are defined based on the expected life stage of an ego given his/her age, as follow: young adulthood (18-21 year old), union formation (22-28 year old), middle adulthood (29-45 year old), post-reproductive adulthood (46-55 year old), grand parenting (56-75 year old) and old age (76-year old and older). Alters are ranked based on the number of calls with the ego, and from the top alters the possible close relations are tracked given the age cohort to which each ego belongs. Different communication properties reflecting intensity and effort invested in the social bond (number of calls, average time per call, average fraction of time spend in that pair and call initiation balance) are quantified and compared for different gender and age cohorts. It is shown that during young adulthood egos call the most, and after the union formation the calling activity decreases continuously until the old age. After union formation the fraction of time spend with the romantic partner decreases whereas with other close alters increases, particularly strong with the best friend. During the middle adulthood, when the arrival of the children is expected, the call initiation with each parent becomes balanced and it is only with them (i.e. parents) whom the average time per call increases. This shows a reinforcement of these bonds, probably as a consequence of a required support from the egos’ parents regarding the egos’ children. When the egos reach post-reproductive adulthood and during the grand-parenting stage, there is a strong change in their social focus towards their children, particularly to the daughters, which take place in the expected reproductive age. The change in the social focus of parent-daughter is considerably stronger in females than in males. In addition, during the grand-parenting stage, females (having crossed the menopause) reduce considerably their invested social focus towards their romantic partner, and direct it towards their daughters (which are in reproductive age and require alloparental support). The most contacted alters for grand-mothers is not the romantic partner but
the daughter, yet this change is not present in grand-parenting males which still keep their main social focus on their romantic partner. This situation provides arguments to support the theory of grand-mothering effect.

5.3 Conclusions

Information Communication Technologies (ICT) have brought the whole new set of facilities and tools that have expanded human horizon and ways of interactions in a number of unprecedented ways. Using the Internet and digital technologies, fast and long-distance communications channels keep humans connected in real time, and offer access to huge diverse sources of information for improving their daily lives. Human interactions in the digital world leave traces, digital footprints of the activities performed in the networked world, such that in the recent years, given the amount of information that digital systems can store, huge (big) data sets containing human-generated information are being captured and gathered continuously. However, these advances have not modified the essence of the human communicative behaviour, as they just extend its possibilities, and the social features present in face-to-face communication world are simply adapted to the new framework. This offers a new way of studying human behavioural patterns, from the perspective of the social network by analysing these massive sets of data containing traces of human social interaction.

Despite the recent appearance of the mobile phone communication, nowadays is a widely used technology, representing one of the main communication channels. The use of the mobile phones for communication reflects various features of human behaviour, from daily routine to the strength of the social bonds linking a mobile phone user with his/her surrounding social network. A communication event (registered in the form of a call details record) says not much about human behaviour, itself. The true power in analysing call details records comes from the regularities and patterns found in the collection of calling events done by each user, as he/she does not use the device at random times nor call to random people. The effort that he/she invests on a particular mobile phone contact depends on the mutual interest and emotional closeness between them, and the whole set of ego-alter links serves as a reflection of some part of the social fabric of humans.
Each day, mobile phone users follow at least two different circadian rhythms, each one synchronised to a different clock. On one hand, (social) humans live immersed in a social time, where most of the social activities are scheduled, defining a daily routine. Working and schooling times, fixed opening and closing times of stores, offices, and public places, weekdays vs. weekend activities, among other examples, create a very specific day-to-day schedule to follow. On the other hand, biological rhythms mark our daily, natural cycle and many physiological processes follow a different, close-to 24-hour biological clock. In particular, sleep-wake cycle and resting patterns have been influenced by daylight and ambient temperature since ancient times, even nowadays in modern civilisations, with artificial lighting and controlled environments. The communication pattern that mobile phone users follow, shows the struggle of living between these two clocks. The calling activity shows some specific schedules, peaking and decreasing twice in a day, at different time for different days of the week, showing a strong component associated with the social time. However, the width and location of these periods of high and low calling activity changes over the year. During the summer the activity extends towards early hours in the morning and towards late hours in the evening, and with the opposite dynamics during the winter, showing strongly seasonal changes. In addition, people living in the same social time (i.e., living inside the same time zone), start and cease their activity at specific but different times. The difference coincides with the time difference of the sun transit times between their geographical location, therefore people living eastward have earlier schedules than those living westward but in the same time zone.

In addition, the intensity and emotional closeness of each mobile phone user with the contacted alters are reflected in their communication pattern. The level of interaction between each mobile phone user and each member (alter) of the egocentric network is different. The communication inside the egocentric network is dominated by the closest alter, in general the romantic partner which is of a similar age as the ego. The social focus, depends on the age and gender of the ego, showing very characteristic features as the egos visit different life course stages. Younger people tend to call more frequently and intensively to alters of their same age and of both genders. This pattern changes noticeably as the egos cross the parenthood stage, in such a way that for old age egos, the most contacted alters are probably their children, and the size of their social egocentric
network has shrunk. Gender differences in the calling patterns also appear at the grand-parenting age, around the same age when females are crossing menopause. For females at that age, there is a strong social focus towards their daughters, which are in reproductive stage, while males still keep their romantic partner as the main, closest alter.

As a final note Publication I and Publication II contribute to the understanding of the dynamics of the resting patterns and sleep wake cycle of humans, and shows that they are still strongly influenced by the yearly dynamics of the daylight, in spite of the controlled environments where they live. On the other hand Publication III and Publication IV provide insights into the social focus over the life course of each individual with respect to the members of his/her egocentric network, with clear age and gender differences over each different stage. The research done in four publications is an example of the versatility of the use of reality mining as a framework in studying human social behaviour.


Human behavioural patterns: A reality mining study

Daniel Monsivais Velázquez