Adaptative modeling of urban dynamics with mobile phone database

Suhad Faisal Behadili

▶ **To cite this version:**

Suhad Faisal Behadili. Adaptative modeling of urban dynamics with mobile phone database. Networking and Internet Architecture [cs.NI]. Université du Havre, 2016. English. NNT : 2016LEHA0017 . tel-01668513

HAL Id: tel-01668513

https://theses.hal.science/tel-01668513

Submitted on 20 Dec 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

THESE

Pour obtenir le diplôme de doctorat

Spécialité Informatique

Préparée au sein de l'Université du Havre

Adaptive Modeling of Urban Dynamics with Mobile Phone Database

Présentée et soutenue par
Suhad Faisal Behadili

Thèse soutenue publiquement le 29 Novembre 2016 devant le jury composé de		
M. A. Koukam	Full Professor, Université de Technologie de Belfort	Rapporteur
M. Christophe Cambier	HDR,UMI UMMISCO - IRD & UPMC	Rapporteur
M. Alexandre Berred	Professeur, Université du Havre	Examinateur
M. Bernard Elissalde	Professeur, Université de Rouen	Examinateur
MME. Rawan Ghnemat	Professeur, Université Princess Sumaya pour la technologie	Examinateur
M. Cyrille Bertelle	Full Professeur, Université du Havre	Directeur de thèse
M. Loay E. George	Professeur, Université de Bagdad	Codirecteur de thèse

Thèse dirigée par Prénom NOM Cyrille Bertelle, laboratoire LITIS- EA 4108

Dedicated to the memory of my father and grandmother

Acknowledgments

Foremost, I would like to thank my advisor Prof. Cyrille Bertelle for guiding and supporting me over the years. You have set an example of excellence as a researcher, mentor, instructor, and role model. I would like to express my sincere gratitude to my Co-supervisor Dr. Loay E. George for the continuous support, motivation, and enthusiasm over all PhD. years. I would like to thank my thesis committee members for all of their guidance through this process.

A special thanks to my friends: Ammami Mohamed Tahar and Adel Bellara, you are and always will be my best and great friends ever. I am lucky to have true friends like you. Thank you for always being with me, I will never forget the time that we were together and our boisterous discussions.

I would also like to thank my friends: Ennassiri Hamza, Mouheb Dhifli, Harik Elhoussein Chouaib, Masmoudi Nesrine, Mohammed Ali Benhamida, Mourad Kaddes, Riadh Saada, and all my fellow graduate students, research technicians, and collaborators in Le Havre university for their support and encouragements, which kept me smiling and inspired. I am very grateful to all of you.

I would especially like to thank my amazing family for their love, support, and constant encouragement I have gotten over the years. In particular, I would like to thank my mother, my brothers, my sister, and my brother in law. You are the salt of the earth, and I undoubtedly could not have done this without you.

Abstract

The communication networks (mobile phone networks, social media platforms) produce digital traces from their usages. This type of information help to understand and analyze the human mobility in very accurate way. By these analyzes over cities, it can give powerful data on daily citizen activities, urban planners have in that way, relevant indications for decision making on design and development. As well as, the Call detail Records (CDRs) provides valuable spatio-temporal data at the level of citywide or even nationwide. The CDRs could be analyzed to extract the life patterns and individuals mobility in an observed urban area and during ephemeral events. Whereas, their analysis gives conceptual views about human density and mobility patterns.

In this study, the mobile phone traces concern an ephemeral event called Armada in Rouen city. However, important densities of individuals are analyzed and are represented to extract the life patterns by classifying the most active regions during observed period in this urban area. Then, the collective mobility patterns are investigated in aggregated urban mobility patterns via extracting the universal mobility law (power-law distribution). This investigation explores the characteristics of human mobility patterns, and model them mathematically depending on substantial parameters, that are the inter-event time, traveled distances (displacements), and the radius of gyration.

The main purpose of this study is to determine the general pattern law of the population. And, its contribution is the resulting outcomes, which are revealed and visualized in both static and dynamic perspectives. They can be capitalized as guidelines to explore the urban pulse and life patterns. The numerical simulation results endorse the previous investigations. Hence, they found that the real system patterns almost follow an exponential distribution. Additionally, the experiments classified the mobility patterns into major classes as general, work, and off days.

Keywords : Complex systems, urban, mobility, CDRs, mobile phone, spatio-temporal, network, radius of gyration, individual trajectory, city pulse, simulation, power-law distribution.

Contents

List of Figures

List of Tables

List of Algorithms

Chapter 1

Introduction

Problem Statement

This thesis intends to study the human mobility inside urban environment by using mobile phone data during an ephemeral event. It analyzed the collected data from the mobile communication network. There are many questions may be raised when studying a specific event in the urban city such as:
How cities form the places of pulsation according to human mobility along the day? How individuals are moving inside the city for their daily activities? Do city structures are well adapted to these human mobilities ? Are we able to identify some common fluxes of individuals moving [C H14] inside the city revealing specific usages and how to be more efficient to support them? Does these collective mobilities are generating congestions or traffic jam? During an unusual event, how the city life pulsation is reorganized? How to manage some specific event like concerts or sportive manifestations in order to control the flow of human crowds? What are the fluctuations in dynamic patterns of city life during these specific events [And10]? How to have the right information at the right time in such event to deal with security? How finally better understand the human mobility behavior in usual time and in unusual events? [Dar13, Rob15].

All these questions need to analyze many aspects of city organizations and citizen behaviors. We will not answer to all these questions in detail in this work, but we will explain how a large amount of data produced by the traces of mobile phone can be interpreted to give relevant information to initiate a concrete reflection to answer to these complex questions. Hence, it will be easier for local administrations, and services appliances to analyze the events with high fluently responding. Therefore, any anomalous or emergent event in the city could be handled in appropriate time.

Mobile phones could be considered as a geolocalization devices for tracing each individual [Yua13]. They are considered as the most propagated and preferred communication devices in the world, also it is the close friend to almost people and accompanied them always. The mobile existence means the physical individual existence and mobile spatial behavior refers to the individual dynamic behavior at urban area. The mobile phone data have several aspects with a lot of indications for revealing human mobility patterns. These patterns are characterized by different scales of time and distances [Cha10a], which reflecting the daily circadian rhythm either in long or short distance/time scales to understand the global or local spreading of individuals, epidemics ...etc. [YAN11, Oll13, Chr13b].

Undoubtedly, the mobile phone data sets open the way to a new paradigm in urban planning, i.e. collecting the city data from real-time sensors (real-time cities), and highlighting the behavior analysis and spatio-temporal data mining. It became well known that urban structure has strong impact on urban-scale mobility patterns, indicating that different areas inside the city are associated with different inhabitants motion patterns.

There are many researches according to these concepts and perceptions, which analyze urban systems of cities, where the mobile phone data are the main material [And10]. The analyzing and tracing mobile phone activities are outlined in the geographical representation of human mobility in the cities [Dar13]. These data could be captured, processed, and analyzed in real time, because they are real temporal data and came from real infrastructure of the city space [Lin12]. Also, they could be analyzed as aggregated [San12] or disaggregated (collectively/individually) with different analysis patterns [Oll13, Mar14b]. Therefore, it would be a suitable depicting for future city characteristics and features.

The geographic and urban space representation could be depicted graphically, mainly to get a partial understanding of traffic dynamicity and uncover the activity patterns of the urban city, either for pedestrians or cars [Mar14a]. This kind of studies enables the understanding of individuals interaction with events, their reactions, responses, environmental effects on them in various city sectors, and revealing the transportation traffic. It would be easier to make decisions based on the impact of the simulation of all these effects. These studies are useful to uncover when and how individuals are arriving and leaving a specific area, when and how aggregation of crowd appears [And10], as well as the authority of their existence at this area. So, if there is any unpredictable event and emergency, hence the situation could be handled accurately as soon as possible [Dap15].

Objectives

The data of Information and Communications Technology (ICT) contribute to form an impression of the city dynamic evolution over time. With this intention, the social communication networks are a backbone tool for analyzing and understanding the individuals communication and city dynamics [P. 11]. This will reconstruct the comprehension of population and city management relative to the urban utilization [And10, Mar14b]. The produced acquaintances help researchers to obtain deep perception of individuals mobility and stream movement inside the urban area [Joh15]. This would help the decision makers and urban planners to improve the city life patterns sustainability, and to have active smart cities diagnosis [Chr10].

The study of human mobility patterns helps in deep understanding many of related disciplines like economy, urban planning, travel, health, crimes ... etc. [Yua13]. The researchers make their efforts to analyze the enormous data, and represent the acquired knowledge on different modes like statistical sheets, statistical diagrams, and the most effective way is the Geographical databases using a Geographical Information System (GIS) platform, in order to give more realistic and comprehensive perspective to the analyzed data. It is in fact characterizing the human behavior patterns in different daily activities [Dar13, Oll13, Chr13a, Mar09].

Several human mobility studies investigated the ephemeral events, especially the wide festivals and occasions (sportive competitions, public ceremonies,... etc.) by analyzing, modeling, and simulating human mobility of an observed event based on mobile phones traces analyzes [And10].

In this research the mobile phone traces are used in order to understand the human mobility. The investigated data concern important densities of individuals, that attended ephemeral event. These data concern 12 days of mobile phone activities during Armada festival, which is held periodically in Rouen city in the Upper Normandy region in the north west of France. The analyzed data called Armada DataBase (DB), which is a group of registered records known as Call Detail Records (CDRs), they are obtained from Orange company for mobile communications and services [Org16].

Contributions

Armada event could be considered a rich repository for miscellaneous human dynamic activities. Hence, human mobility patterns generated during this event would been estimated by investigating the Armada DataBase. It is analyzed using the SQL language, ArcGIS platform, Excel applications, MatLab platform, Gama platform, the geo-visualization, and statistical techniques. They are used in this thesis for studying the human mobility collectively and individually. The undertaken substantial parameters are the density, inter-event times, travel distances (displacements), and radius of gyration. They have been modeled and simulated using computing platform by integrating various applications for huge database management, visualization, analysis, and simulation.

Accordingly, general population pattern law is extracted from the data. The study contribution outcomes have revealed both the individuals densities in static perspective, and the individuals mobility in dynamic perspective with multi levels of abstraction (Macroscopic, Mesoscopic, and Microscopic). A synthetic diagram of the research as shown in figure 1.1.

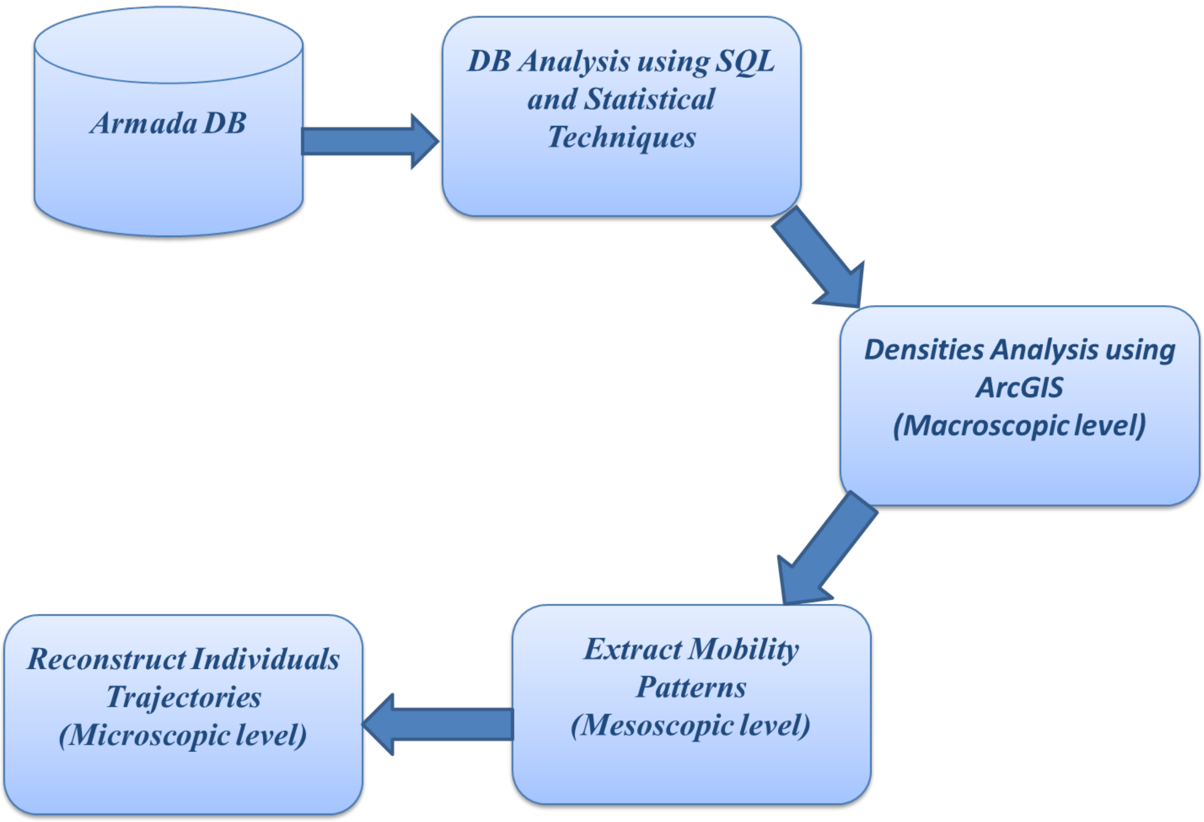


Figure 1.1 – Full project diagram.

The present work can be divided into two stages. The first stage deals with analysis, and acquire the activities densities, which indeed represent the individuals densities. The second stage is devoted to model and simulate the human mobility patterns of the observed data.

Analysis and Representation

The observed data are analyzed in the first phase, they require high computational resources and DB management techniques. Whereas, the available resources (in LITIS laboratory) were unable to accommodate them. However, these data are treated using an SQL query language to extract the information, then using the statistical application for drawing the histograms and diagrams to represent the individuals densities (activities). Then, the produced data sheets have been analyzed, and geographically represented using the ArcGIS platform. These data are lined upon the Rouen city map in association with the towers distribution all over the observed area around 30x30 Km. This stage is performed in the macroscopic level of abstraction, and is represented in figure 1.2.

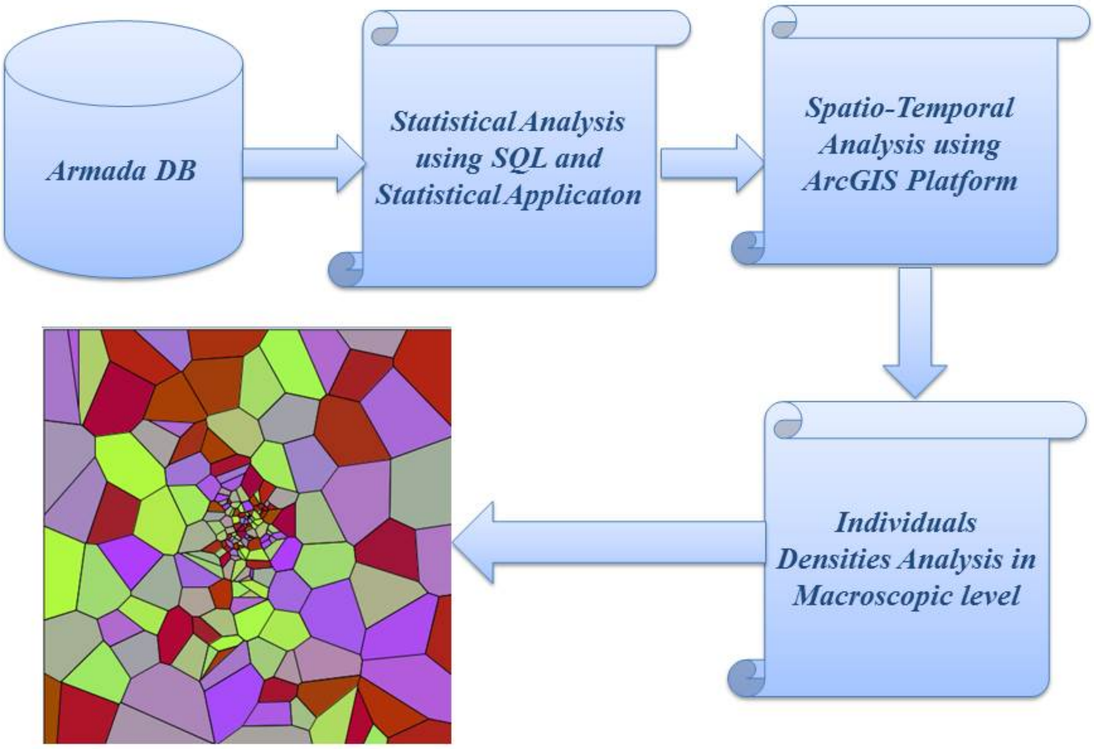


Figure 1.2 – Densities analysis.

The observed urban area is divided by Voronoi polygons technique, based on the spatial distribution of the mobile phone towers. Each Voronoi polygon is associated with the tower coverage area [Pu 09a, Rob15, Van12a], as in figure 1.3.

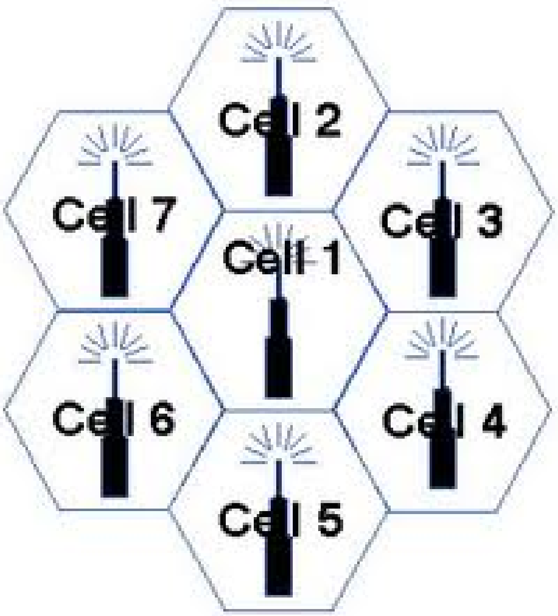


Figure 1.3 – Divide the coverage area using Voronio.

Modeling and Simulation

In the second stage, the life patterns and activities in the observed area are investigated during general, work days, and weekends of the observed period, i.e. determine the general mobility patterns law according to spatio- temporal mode.

Additionally, this stage is developed in mesoscopic and microscopic level of abstraction to model and simulate the individuals mobility. This is performed using statistical calculations by using sufficient parameters, which are the activities in inter-event times, traveled distances, and radius of gyration. However, the Matlab is used to model and simulate the observed data in the mesoscopic level. Whereas, the multi agent simulatoin platform MAS (Gama-Platform) is used to reconstruct the individual mobility with GIS coordination at the microscopic level, as represented in figure 1.4.

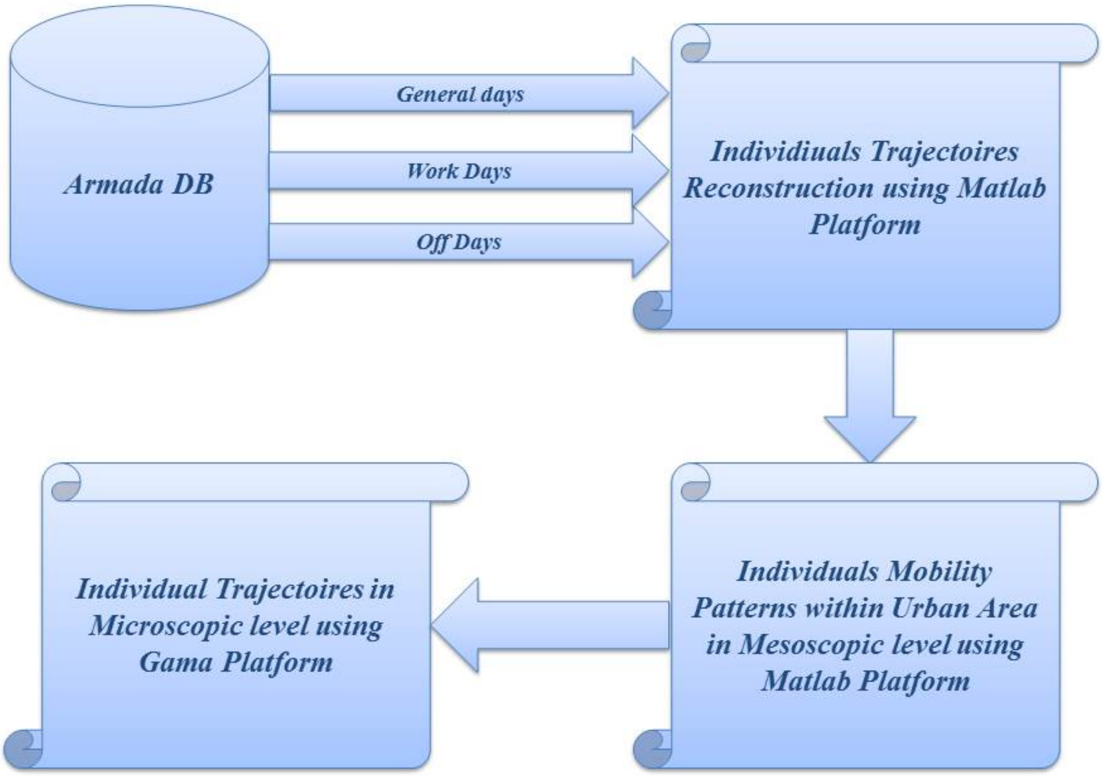


Figure 1.4 – Multi agent physical statistics.

Organization of the Thesis

The thesis is constructed of 7 chapters.
Chapter one gives an introduction to research problem definition, its objectives and methodologies. As well as, the used contributions with explanation of major two stages of the thesis. The first stage is the analysis and representation, and the second stage is the modeling and simulation.

Chapter two describes how to analyse urban mobility with mobile phone traces. The introduction of this chapter declares the relation between the urban mobility and mobile phone data, and the benefits of urban mobility studies. Then, the city is described as a dynamical complex system according to complexity theory. The mobile phone traces are explored as a tool for catching urban mobility patterns.

The mobile phone data management section explains the CDRs features, and the mobile networks characteristics with their common produced data. Finally, the conclusion of this chapter considers the mobile phone a good proxy for human mobility in urban, in spite of some limitations.

Chapter three describes how modeling in spatio-temporal approaches the urban mobility with CDR, MAS and GIS. In the introduction, the common urban modeling is investigated, espacially the approaches using CDRs, MAS, and GIS with respect to spatio-temporal properties. Then, the mobility modeling with GIS describes the data analysis and visualization in a way close to real time.

The basic models of attraction mobility in city with gravity and radiation models are presented as a kind of urban mobility modeling. Then, multi-scale modeling of human mobility describes the major used approaches, which are Macroscopic models, Mesoscopic models, and Microscopic models. Additionally, the section of modeling mobility dynamics and spatial constraints explores the common models that are respect to the physical laws of acceleration, velocity, and rate of change of direction.

The urban mobility patterns section presents the approaches that are used to capture the urban mobility using probability distributions of human activities, data mining, and some other approaches are mentioned. Then, integrated MAS and GIS platform explored the contribution between the MAS and GIS approaches to produce adequate models of spatio-temporal properties. Also, common agent-based platforms are presented.

Finally, the conclusion of this chapter endorsed the ABM to model the urban dynamics, and the major human mobility perspectives.

Chapter four describes the macro analysis of mobile phone traces based on the case study (the Armada event in the city of Rouen). The introduction gives an overview about the CDRs analysis phase and its contribution in human mobility and urban environment. Then, the Armada DB as case study is described in all aspects as a DB of ephemeral event, and as a periodic festival. The research methodology in analysis phase described the approach of data analysis using several DB management techniques, ArcGIS, and Gama platforms. As well as, data analysis and visualization in Marcoscopic perspective is explored with respect to the individuals densities using Voronoi technique and statistical manipulation. Finally, as a conclusion the density variation between several urban areas can explore the tendency of these areas, and verify the normal or anomalous events.

Chapter five describes the micro-modeling and simulation approaches based on reconstruction of human trajectories and human patterns. The introduction presents the modeling and simulation benefits, and the levels of system data analysis either individually or collectivelly. Then, data set section describes the CDRs as a discrete data, and explains their type as estimated geographic data. Describing individuals activities (Mesoscopic Perspective) section describes the population according to their activities in time using probability distribution. Also, the section of fixed inter-event time observations indicates the main parameters used in this thesis, that are inter-event time Δt, the travel distance Δr, and the radius of gyration r_g.

The section of trajectories in intrinsic reference frame presents the human trajectories characteristics into spatio-temporal mode, and their properties within r_g. Then, the real and simulated individual trajectory estimation is decribed in Microscopic Perspective. Finally, as a conclusion the CDRs provide common paramters can reflect the human mobility in multi level of abstraction.

Finally, last chapter describes conclusions and perspectives of the work presented in this PhD thesis.

Chapter 2

Analysing Urban Mobility with Mobile Phone Traces

Introduction

The recent spread of ubiquitous technologies, even in emerging economies population, is leading to a variety emergence of mobile phone based services for low-income populations in several areas like health, education, or banking [Alb13, Van12c]. This diversity is proved in 2012, when the mobile phone number is counted as 104.8 million Android, 26 million iOS smart phones all around the world [Lin12]. The massive spreading of the computing and sensor devices became a backbone of the modern daily life. Therefore, using their data to study the human mobility is a natural evolution for one of the most important life aspects in the new century, thanks to the fast progress in the embedded systems, and wireless networking technologies [Xia12, Van12b].

The infiltration of mobile phone networks around the world allows to capture data sets for millions of interactions, which are stored in real time [San12]. Such data could be taken from telecommunication companies, internet companies, or mobile networks for a wide geographic areas in national scale [Van12c]. These data do not contain detailed demographic or socio-economic factors for the networks subscribers [Mar10b, V. 11]. However, mobile phone networks could assist in providing a comprehensive vision for urban area usages through human mobility traces [Mar14b]. An efficient understanding could be reached with the help of GIS platform augmented by statistical and dynamical analysis, as will be mentioned later.

The urban term refers to the city, which is composed of many grouped items like number of buildings, streets, public or private educational organizations, shopping areas, residential buildings, ... etc., that are existing in predefined spatial space, so the open areas, and the streets are the linkages between these items. However, studying urban mobility is essential for urban planning. The analysis of citizen mobility enables the understanding of how individuals are evolving inside the city, and how they use the city services. This helps the city planners which have to deal with the cities development, and to be sure that the services like transportation are efficient and able to meet the citizens needs [Yih11].

The advantages of studying human mobility are to put the plans for emergency events prior to their existence, to verify and estimate the evacuation strategies by simulating the event and evaluating its rescue efficiency, to control and manage the crowd movements to avoid the potential problems, and to support and interconnect multi scientific fields (traffic engineering, architecture, socio-psychology, safety science... etc.) [Nir08, Mar13, Dap15].

Furthermore, all needed information about human mobility either as a walker, in a car [Xia12], behavior, and motion has become more affordable in a short time. As a consequence, the researchers have now rich data repositories for making their studies. Starting from extracting knowledge of social networks [P. 11, Bal13], dynamics of population [Alb13], human mobility, gender identification as performed in [Van10], till the analysis of different urban environment events in relation with land use [Jam12, C H14].

Therefore, the clear insights which extracted from identifying the urban areas give the planners useful information to manage many cases, like developing transport system, determine the specific times that demands special kind of services [Yih11], and verifying the areas that have fast spreading of panic diffusion in collective crowd behavior, so based on urban areas studies, the planners can do the necessary to deal with every situation according to predetermined strategies. With respect to characterize urban environments, it is very important to identify the dense areas [Alb13], which have high population at specific period of times. This is a principle aspect, because it gives a precise description of where and when high density of individuals takes places? [Pu 09a, Mar14a]. The former researches on the identification of dense areas have been carried out by using three main approaches as density-based clustering techniques, detecting dense fixed-size grids as spatio-temporal data, and spatial-based techniques to detect local maxima areas [Mar10b].

Whereas, the recording and analyzing of human mobility for a limited time in traditional methods was difficult and costly, since it needs to do interviews, observations for period of times for collecting data [Dan09]. However, the new technologies made this recording and analyzing easier and more flexible [Xia12, Jam12]. Nevertheless, there is a biggest challenge facing this evolution of capturing the huge data, it is the subscribers privacy that uses different network types, and their rights to be fortified against any interloper or unwanted monitoring [Van10, Oll13].

This chapter is organized as in the following sections: The first section is the city as dynamical complex system section describes the city as a complex system according to the complex system theory, and differentiate between the complex and complicated systems. Then, the next section is the catching urban mobility with mobile phone traces, it describes the mobile phone data utilization in urban mobility studies. However, this section is branched into two subsections as: Individuals as sensors and the individual mobility patterns. The first one explores the highly penetration of mobile phone and its accurate acquired spatio- temporal data from it, this make each individual carried mobile phone as a sensor. The second one describes the capability of mobile data to extract the human mobility patterns.

The mobile phone data management section gives a general overview of the mobile phone networks architecture. It is branched into three subsections as: The CDRs properties, the mobile networks coverage, and the mobile phone traces localization. The first one elaborates the complex nature of CDRs, and their ability to reflect human dynamic behavior in real time. Then, the second one illustrates the main elements of the mobile phone network, as well as the utilization of Voronoi technique to divide the urban under the covered area. Finally, the third one describes the procedure of the communication within the mobile phone network.

Then, the resulted tracing data from this kind of networks are classified into two main types as: Active cell phone localization and passive cell phone localization. Then, the technical and communication data section presents the location traces from Mobile Switching Center data(MSC), and other data collection methods like the Global System for Mobile Communication (GSM). Hence, the billing data section describes the data of formal type (CDRs) that are used for billing purposes, which are related to the mobile user activities specifically. Finally, the chapter conclusions consider the mobile phone as a continuous spatial monitor in real time property. Whereas, the urban area division is considered as a challenge issue.

City as Dynamical Complex System

The complexity theory becomes a common aspect in most of research fields that are aimed to understand systems like social and human developments, that have the features as nonlinear behavior, self-organization, irreducibility, and emergent properties [Yik13].

The complex systems are used to describe natural environmental systems for understanding, analyzing, and modeling them. Such systems are dynamical interaction systems which could be organized or disorganized during their evolution [Raw08a]. One of the main propertiy of complex systems consists of dealing with the feedback processes over their own components, as well as dealing with processes which emerged from their own entities interaction. These systems have different outputs that could be traditional results or new processes as in figure 2.1 [Raw09f, Arm10].

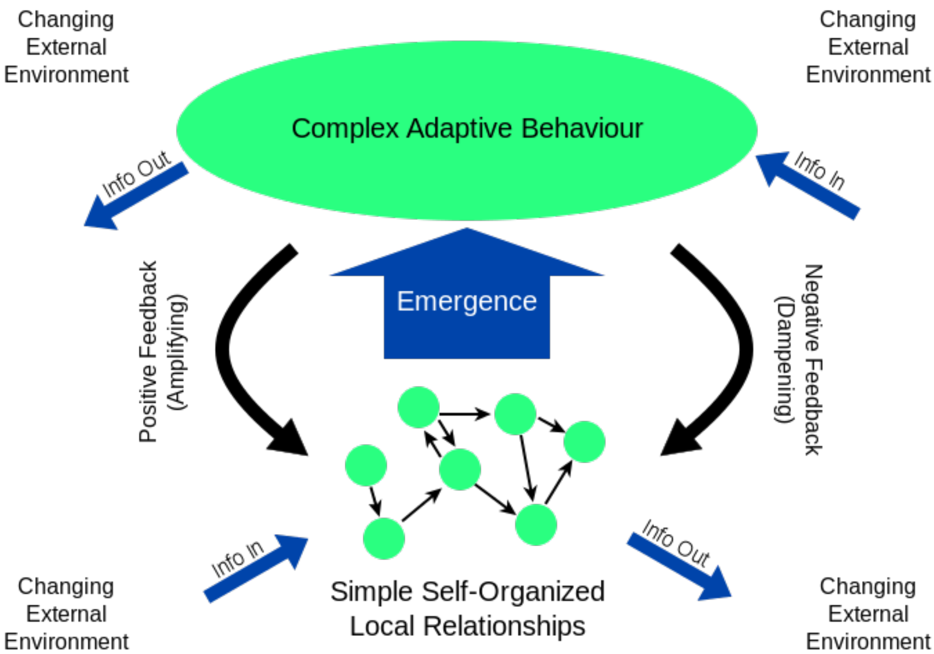


Figure 2.1 – Complex adaptive system [Wik14].

The difference with complicated systems will be explained to better understand the meaning of complex systems. Complicated systems can be decomposed to be studied with each part separatly. But complex systems cannot be studied in that way because every entity affects the whole system reactions and results. They have adaptive mechanisms as emergent process from their interaction among entities. These adaptive processes correspond to responses of the system as reactions to dynamic usage. Another charateristic of complex systems are the feedback processes as new processes distribution according to system evolution, because even the resulted events could be returned back to the system as new activities [Raw09f, Hak08]. However, complex systems deal not only with emergent organization processes from the interaction of its own entities, but also with the feedback processes of the organization over its own components, to represent complex multi- center, and multi-criteria self-organizations.

[Mic94] considers cities as complex systems, because there are comprehensive entities interacting at infinite variety. The urban evolution is either bottom-up [Raw08a] or somehow top-down approach. Whereas, [Mic07] considers urban development as a temporal dynamic, i.e. the urban evolution includes multiple individuals with different behavioral patterns at spatio-temporal manner, and indicates dynamic interactions between socio-economic [Van13, Van12c], and environmental effects. Hence, Urban environment and city could be regarded as a complex system, which is composed of different interacted (subsystems) parameters, that are affected by land use strategies, population growth, market behavior, and transportation systems.

The human mobility researches tackled the small scale data like origins, destinations, and travel patterns of daily individual mobility, as well as the large scale ones through long travel patterns (air-travel for example). These two kinds are the large and small scale data, which have distinct indications in the level of details, where the small scale provides a detailed description at the individual level, but not for population, whereas the large scale once describes the conceptual view of the population without details of individual mobility characteristics.

Catching Urban Mobility with Mobile Phone Traces

The Information and Communication Technology (ICT) has good reputation in human mobility studies, especially when looking at its most prevalent product "Mobile Phone", since it gives the researchers very speedy method for collecting real time data with cheap cost in opposite to old days approaches [Yua13]. Many researches in the field of network tracing emphasize on common techniques, like the mobile carriers [Lib04], which are used for collecting the time and location information (momentary location, spatio- temporal). It is determined by giving the closest mobile tower whenever the individual uses his/her mobile phone, however when no mobile phone activity, it means that no information will be available, thus the users' location is unknown. Also, the available information type is time-dependent location [Cha10b, Cor13].

The researchers explored several methodologies to trace the human mobility, which are particularly pivoted on the CDRs analysis [Chr13a, Rob15]. They emphasized on duration extension of the collected data, tracing the mobile phone locally by inside tracer, and exploring the network data (bills, network signaling). All these items are very useful in human mobility studies, that depend principally on the CDRs [Zbi13, Cor13, Mar10b].

However, there is another technique of active mobile positioning which is used in real time tracking using either Mobile Positioning System (MPS) or Assisted-Global Positioning System (A-GPS) Network. The positioning data are useful in several application areas, like direct travel and transport behavior, traffic flows, and urban engineering. The GPS is a system of 24 satellites that constantly orbit Earth, when the hardware inside the mobile phone receives signals from at least four of these satellites, then the handset can calculate its latitude and longitude with an error margin of 10 meters. The drawback of GPS technology is the system not fully operable when objects block its access. The other method of determining the location of mobile phone is by triangulation, a process by which mobile phone location can be determined from its communication with local cell towers [Rei10, Ant05, Yu-05, Mar14a].

Individuals as Sensors

Many ubiquitous sensors are prevalent in the recent decade, one of the most known is the mobile phone. These ubiquitous sensors are providing instantaneous sensing [Mar13], where spatio-temporal data can be acquired from mobile networks, and become affordable directly with high accuracy. Therefore, mobile user would be sensed "easily traced" due to the provided spatio-temporal data. Also, the individuals can be considered as human sensors per se, since they give important georeferenced investigations as independent perceptions [Xia15].

The mobile communication networks offer wireless communication services, therefore the individuals can communicate in multi locations all over times, which means the acquired data of this ubiquitous sensors would provide dynamic perception on this service users, so the individuals dynamic behavior will be sensed "exploited" inside urban environment in high scale level, with regard to the wide urban area under mobile networks coverage [Mar14b]. The term "in situ" is the opposite of "remote", which is used for the tightly closed sensor or the direct contact sensor with the event/phenomena being sensed. Herein, the mobile phone considered as in situ sensor according to its closeness for both the mobile user and the sensed event. This combination gives a user generated traffic inside mobile networks. Therefore, the georeferenced social media data could be considered as a proxy for the collective human behavior, which considered social in situ sensor data [Gun14, San12, Van12b].

Thus, the individuals can act as a human sensors by giving their investigations according to each individual perception, these perceptions are gathered from them as geographic information for example on Web 2.0 and social media. Then a huge spatio-temporal data will be acquired in high level of accuracy and in situ sensor environmental/ social, which reflects human dynamic behavior inside urban space with large-scale activity and mobility [Gun14, Den13, Van12b, V. 11].

The varied data nature could be unified in spatio-temporal mode to make more relevant analysis, the diverse data nature combined of temporal and geospatial components, which imposed the contribution of interdisciplinary methods to collaborate with Geographic Information Science theory and Applied Geoinformatics [Xia15]. This kind of researches focus on geo-spatial and temporal components of mobile phone data in acquisition and analysis of three main perspectives, that are the human-centered data acquisition via mobile phone, human behavior relationships in the environmental context, and spatio-temporal patterns of human behavior acquired from mobile phone data [Gun14, Oll13, Jam09].

Additionally, to obtain data by profiting from the fast rise of smartphone penetration, this implies human- centered approaches for data acquisition. The mobile networks create insight on human dynamics in social environments [San12], also people could be considered as sensors by expressing their perceptions as measurements for their situations like air quality, weather, and emotional conditions through their mobile phones. The integration between these perceptions measurements and the technical sensor networks gives well indicated data

As well as, some researchers introduce the approach of geo-monitoring using geo-spatial dimension to use the mutual context-awareness during analyzing and monitoring dynamic geographic phenomena. Mobile phones as ubiquitous geo-sensors are in general investigating human and environmental dynamics [Gun14, Van12a]. The challenges of this field are confined in deeper investigations of human and environmental dynamics using the ubiquitous mobile geo-sensors, where the user-generated data have different nature with different representativeness and semantic expressiveness, the generated involuntary from mobile network traffic data represent significant portion of the social groups population. However, this kind of data has only huge quantity without any semantics of them, otherwise social media data and Volunteered Geographic Information (VGI) are classified generated into specific sub group of the population, and own semantic value.

However, the environmental phenomena could be sensed directly or indirectly. Since, mobile networks enable their customers to communicate everywhere at any time, hence log files generated in their backend could be considered as indicator of human behavior patterns, this indication represents millions of mobile users. Therefore, any interested phenomena could be analyzed in terms of spatio-temporal dynamics [Joh15]. However, human behavior influenced by many different factors, one of the strongest presences is the environmental factor, but of course it is not the sole, other influencing contexts are the mobility, activity, and social interactivity [Gun14, Arm10].

Individual Mobility Patterns

Ubiquitous mobile communication technologies are capable of sensing a few aspects of human behavior either directly or indirectly. Space-time patterns of human dynamic behavior could be derived from mobile phone data. The digital traces of mobile usages can be considered as a reflection to part of the human behavior, that are individuals continuously leave behind them whenever they used mobile network, these traces could be considered as social sensor data, that serve as a proxy for individuals activity and mobility [Mar09, Jul08b, Mar08, Rea09, Gun14].

Many spatio-temporal analyses are done on social field by using the proxy data, hence through the extraction of more depth insights into the urban systems field of the complex nature [Arm07, Joh15]. It could be seen in [G12] which used the Visual Analytics approach and handovers (the signal of mobile network that happened when mobile switched its radio cell during mobile user transitions) to exploit the complex urban mobility patterns. The patterns visualization and analysis approaches supporting real-time events monitoring, due to their approximated indications on the individuals move and their density.

The traditional urban studies have many approaches that divide the observed geographical region into zones, where the population of individuals are moving among them [Mar10b]. As well as, Almost researchers use data aggregation to deal with the excessive mobile data in order to understand and reveal intra-urban mobility patterns. In furtherance of understanding urban mobility patterns, some manipulations such as [Fra12] used a map of the predefined home zone to individuals as mobile user or vehicular driver, then make statistical computations to find the average daily trip for them, and finally make association for the computed results with the urban environment and demographic home zone. It compares mobility features of mobile phone traces with mobility features of odometer readings from annual safety inspections of all private vehicles registered in the Boston Metropolitan Area. It focuses on human mobility traces and their usage in planning to transportation system.

Whereas, the recent researches prove that mobile profiles (in/out calls, in/out SMS) are useful data source as a proxy for exploiting human activity patterns. Wherefore, the Geo-visual Analytics (Geo-VA), Self-Organizing Maps (SOM), and Local Indicators of Spatial Association (LISA) approaches are emphasized on the intensity and similarity of collective human activity during the analysis of huge data amount of user-generated mobile network traffic. [Gun14] elaborates space-time patterns of mobile communication Udine city of Northern Italy for three months period in 2009. It shows similar patterns for the 5 variables (in/out calls, in/out SMS, internet) as shown in figure 2.2 [Gun14].

The arising insights improve the comprehension of daily pulse urban movements in the city, which informs extra information for several domains for instance public transportation [Fra12], event management, urban planning, ...etc. It is obvious that clear correlation exists between city functional configuration and human activity patterns [Rob15]. In spite of all features of the analysis approaches, there is a lack in semantic understandings, which is the answer of why behavior/patterns look like this? However, they are the base of full prospect in the future [Gun14, Arm10].

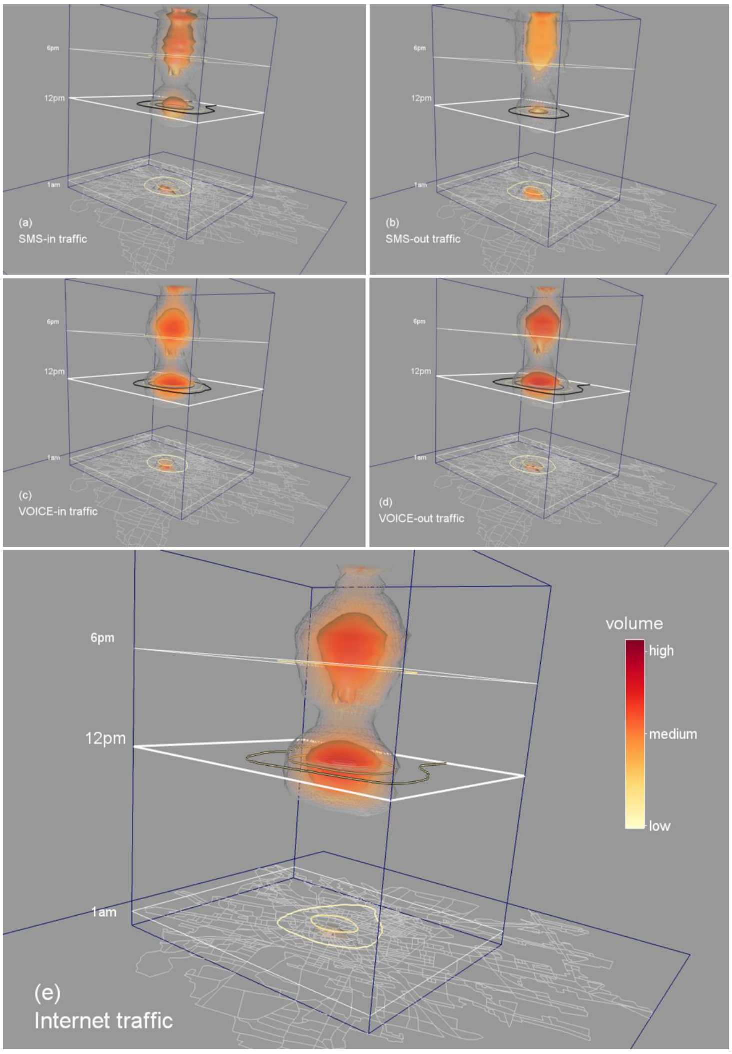


Figure 2.2 – Space-time visualization of mobile communication patterns: incoming SMS (a), outgoing SMS (b),incoming voice calls (c), outgoing voice calls (d) and overall data traffic (e) [Gun14].

Mobile Phone Data Management

CDRs Properties

The mobile data have a special feature, because they can reflect human mobility in real time. They give spatial and temporal indications simultaneously (spatio-temporal data). Each consecutive activities between two towers could be considered as displacement (mobility) from coverage area to another, which is the main point to capture the transition from one location to another [Ale14].

The CDRs have a complex nature and very rich environment to obtain human life elaboration from multi points of view [Rob15, Joh15]. The mobility studies benefit from mobile phone data with concentration on the regions of the highest mobile phone network coverage. These regions are characterized as mature, stable, and most developed. Thus, these models explore the human mobility, and social interactions of the wealthy, industrialized, and well developed regions. Therefore, it is considered as the models that elaborate just one-third of the world population, and ignore the rest two-thirds of the population, which lives in not-well conditions.

The ability to detect and characterize mobility patterns using CDRs provides huge information for spatio- temporal data, that could be understood in aggregated urban mobility patterns [Rob15], therefore it opens the door to a wide range of applications ranging from urban planning to crime or virus spread. Nevertheless, the proposed spatio-temporal query systems so far can not express the flexibility, that such applications require.

The telecommunication operator gives the huge volume of data in the form of CDRs, which contain very valuable spatio-temporal information at different levels of granularity (e.g. citywide, statewide, or nationwide). This information is relevant not only for the telecommunication operator, but can also be the base for a broader set of applications with social connotations like commuting patterns, transportation routes, concentrations of people, ...etc. The key to the smart cities development are the efficient search of spatio-temporal patterns to query CDRs databases and use of statistical techniques. Nevertheless, the available commercial systems of telecommunication operators today cannot handle this kind of spatio-temporal processing. One possible way to analyze such patterns is to perform sequential scanning of the whole database or call records and and check them using subsequence matching like algorithm against the query pattern.

However, it is computationally expensive due to the massive data to be processed, and there is no information about the temporal dimension (e.g. between two given days or between two given hours) or spatial properties (e.g. in a given neighborhood, near given spot, or intersecting given area), which are considered to process the database. Although, the CDRs have well known popularity in analysis, but it has two limitations, where there are no data when there is no communication activities (mobile phone idle) , as well as the aggregated spatial data of each tower region (coarse spatial granularity), since the spatial data are not the precise location of active mobile, but they determine the approximate one within tower coverage area [Mar10a, V. 11, Enr14].

Mobile Networks Coverage

The mobile phone networks are built using a set of Base Transceiver Stations (BTS), that are in charge of communicating mobile phone devices with the cell network [Van12c]. Each BTS has one or more directional antennas (typically two or three, covering 180 or 120 degrees, respectively), that define a cell and a set of cells of the same BTS define the sector, that defines sector and all the sectors of the same BTS define the cell. At any given moment in time the mobile phone is covered by one or more antennas. Depending on the network traffic, the mobile phone selects the BTS to connect to within the geographical area covered by the cell, which depends mainly on the power of the antennas. Also, pivoting on the population density, the area covered by the cell ranges from less than $1Km^2$ in dense urban areas to more than $5Km^2$ in rural areas. Each BTS has latitude and longitude, which indicating where is located. The area covered by BTS can be approximated with Voronoi diagrams as in figure 2.3 [S.F09, Pat13, Van12a, Enr14].

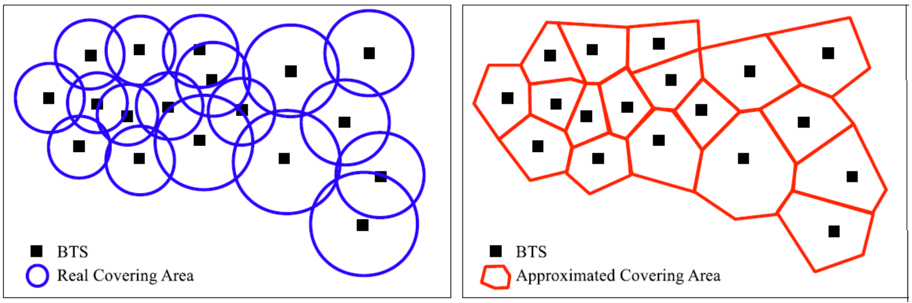


Figure 2.3 – Left: The original coverage areas of BTSs. Right: The approximation of coverage areas by Voronoi diagram [Abh03].

To summarize the dynamic mobility patterns in different urban areas, that requires dividing the study area into sub-areas. One option is to divide the study area into grid cells. However, it is difficult to decide on the appropriate cell size. Moreover, it is highly possible that the number of base towers in each cell varies, resulting in higher mobility in areas with higher tower density.

Therefore, the study area could be divided into Voronoi polygons, based on the spatial distribution of mobile phone towers, and then to summarize the hourly phone call frequencies for each polygon. Each Voronoi polygon is associated with time series to represent its hourly phone call frequency pattern [Yih11, Pu 09a]. For further extract the number of individuals (i.e. active mobile phone users) in each cell, and eliminate the repeated phone calls made by the same individual [Cor13, Mar10b].

Mobile Phone Traces Localization

The mobile network is a radio network composed of Base Transceiver Stations (BTS), one or several BTSs are held on each antenna, these distributed antennas cover specific areas of small regions (cells), where each cell has cell identifier (ID) to achieve wide radio coverage area. The mobile network updates the mobile location to establish the communication among mobile phones, which are under the network coverage area. The communication procedure inside the mobile network is established as follows, and as illustrated in figure 2.4 [Zbi13, Cor13, Van12a, V. 11]:

1 The connected mobile phones are using cell identifier (ID).

2 The cell ID mapped into BTS geographical coordinates or the center of cell area, that gives approximate geographic position of mobile phones than its precise position.

3 During mobility among cells, the network guide the mobile in switching from cell to the next cell, avoiding drop-ping the call and preserving the communication continuity.

4 The mobile HandOver (HO) registers all active cells, which the mobile phone passes through communication duration, having like micro-trajectory.

5 The network coverage area is divided into Location Areas (LA), which are larger geographical areas to organize the mobile phone (individual) mobility, so any mobility from LA to another will create Local Area Update (LAU) by giving cell ID of each new LA, regardless of the mobile phone condition either connected or not (active/ inactive) for long terms of hours periodically.

6 The mobile phone approximate location is required to establish the communication demand immediately, so the Global System for Mobile Communication GSM records the LAU in all times.

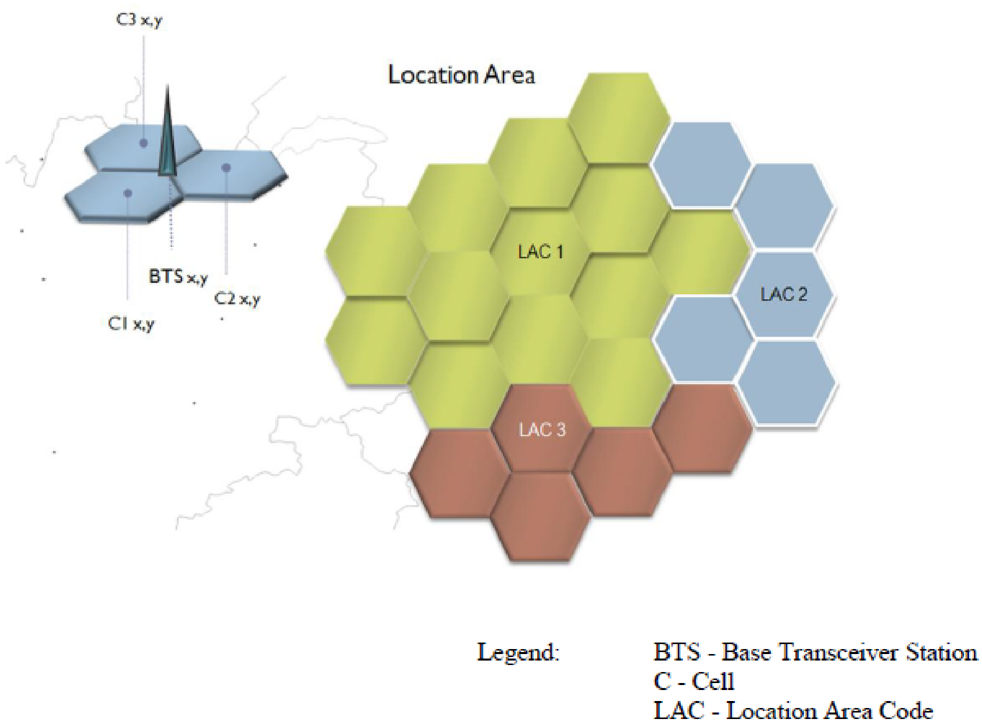


Figure 2.4 – Global system for mobile network communications architecture [Zbi13].

There are two main types of mobile network data, passive and active localization tracing [Zbi13, Cor13]:

1 Active cell phone localization

In this type of mobile localization there is mobility-aware software, for instance (Symbian operating system) incorporated in the mobile device or obtain the spatial location from mobile network to pick up the successive cells, that are changed during mobility. In spite of the imprecision of the mobile exact location, where the location of region Local Area Code (LAC), which the mobile user crossed, hence the spatio-temporal recorded data would be like personal diary extracted from mobile, which in turn acts like personal sensor for registering mobile user trajectory [Lin12].

2 Passive cell phone localization

In this type of mobile localization, the localization source of data is obtained from the mobile data, that are collected for billings/ network management purposes. These data are related to significant number in millions of a given mobile network provider. The mathematician and physician researchers are interested and agreed upon the use of these electronic personal data for many purposes, for instance mobile census to estimate the urban dynamics, studying the social networks, human mobility analysis, and renewing the complex and social sciences [C H14]. The passive data collection technique is used to record data on wide spectrum of population for long time slices, but it misses the semantic information (reasons for travel, modes of transportation, stops, and personal characteristics). This type of data is extracted from two sources as follows [Dan09]:

The Technical and Communication Data

These data are collected by the mobile network, they could be used for mobility analysis, when the cell grouping, which is named Location Area (LA) controlled by the Mobile Switching Center (MSC), then the middle layer of the network management is established. The network emitting signals for the mobile authorization and securing communication for subscribed mobile in the network in order to benefit from its services. The MSC registers the records of signaling events in a database. These data explore active devices under the coverage area (on its territory), they also record the common signaling events (common activities), which are the calls, handovers, and SMSs as events information, also the attachment and detachment which are the information of mobile switching (on/off) events, and LA updates which are bypassing LA boundaries and refreshing the inactive mobile position. The mobility analysis data of specific area are obtained from the MSCs in that area as shown in figure 2.5, it explores the location traces from MSC data (in dotted line), and other data collection methods like the GSM based trajectory (in solid line), it depicted the sms received, call starting with the cells handovers and LA updates the location points according to center precision.

However, the location is limited to the cell center only, i.e. it is not the exact mobile geographic location [Mar14b]. On the other hand, the radio propagation model could be used to calculate it, since the CDRs have accurate spatial shortcomings, so the data fusion could solve this shortage by integrating different data sources like geographic data, mobile networks data, and socioeconomic data. The socioeconomic data could be a population density, economic activities... etc., which are the base of extracting semantic information

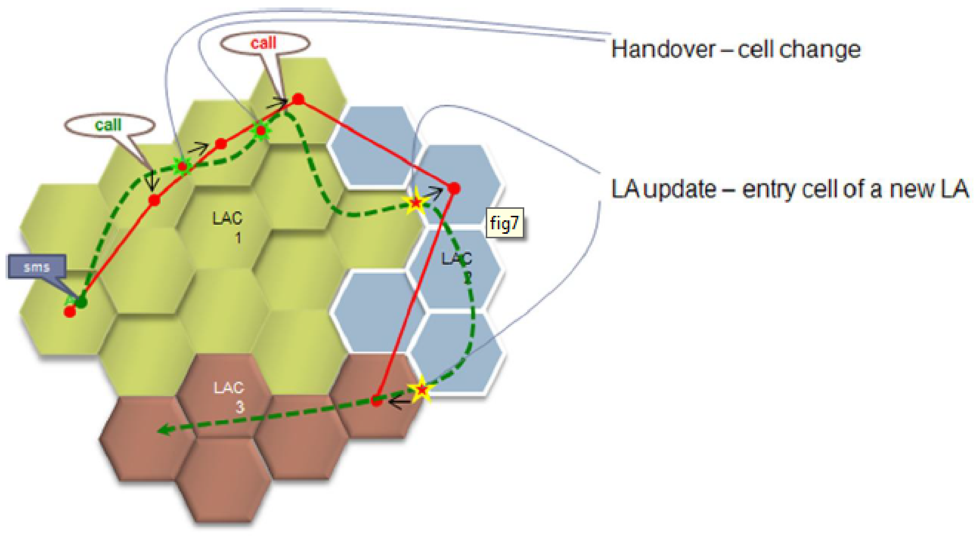


Figure 2.5 – Location information using MSC (dotted line) and GSM (solid line) [Zbi13].

to define and validate assumptions of semantic information with geographic considering [Van13, Van12c], this is achieved by following the procedure as shown in figure 2.6 [Zbi13, Cor13, Mar14a]:

A Model the spatio-temporal characteristics which represent the human mobility, for instance information sources and spatio-temporal trajectories.

B Calculate individual trajectories, then project them geographically using GIS (topographic/ raster) data.

C Merge the socioeconomic data to explore the territory (lands) and increase semantic information of human trajectories [P. 11].

D Analyze the human mobility (behavior) via the data mining methods.

Billing Data

This data category is mobile network data in common known as CDRs, they are records that register the whole individuals communications history for months, which are predetermined for billing purposes, but the researchers used it as very huge repository to the events (call, sms, and internet connection) of the mobile networks, and the mobility behavior of the network users, the CDRs are records registered in formal secured DB with standard style. Because of their standard format, they are easily extracted and manipulated, but the weakness of them is in the localization information, which is obtained from the mobile user activities only, as in figure 2.7 [Zbi13, Mar13, Rob15].

[Van12b] used CDRs to reveal the travel behavior domain, it extracted the origin-destination flows OD matrices from CDRs, which are used to realize the varying travel demands over time, meanwhile the classical surveys explore static information only. The automatic

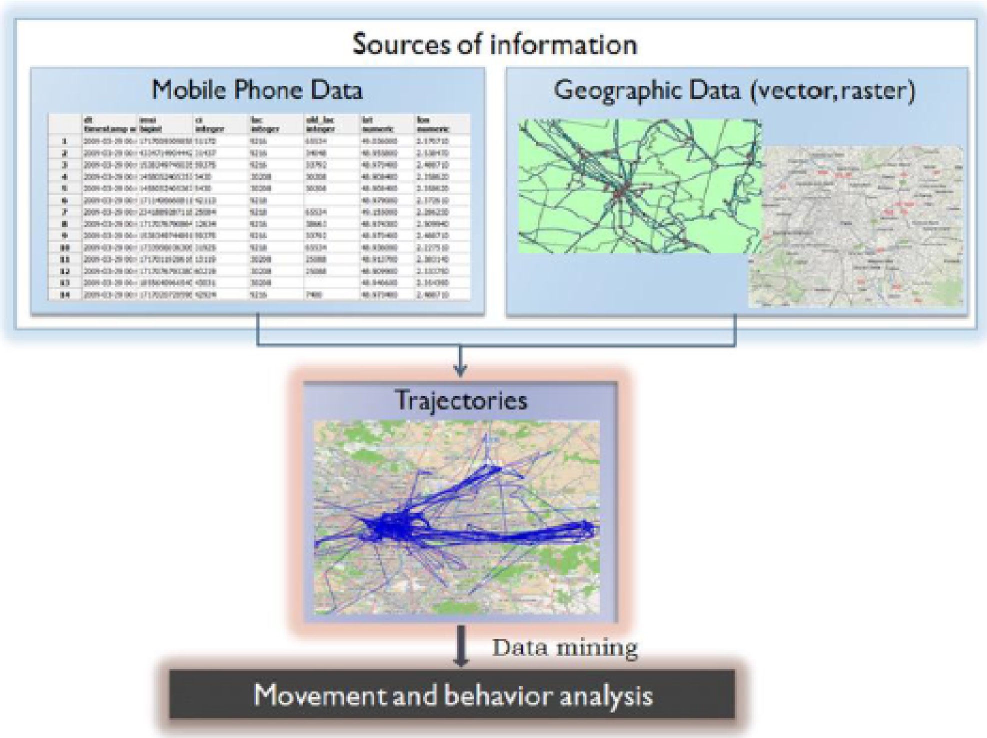


Figure 2.6 – Extract human mobility trajectories from fusion of CDRs and GIS [Zbi13].

data recording provides the geolocation, which means that mobile user has weak privacy, hence the anonymization would be no more useful for privacy [San12]. The data manipulation methods depend on the perspectives of the study, for instance sampling by short time window (daily/ weekly) [Tho11b], and geographical aggregation of individual localizations [Zbi13, Cor13, Chr08, Enr14]. The CDRs details are the timestamp, call duration, type of call (sms, voice call, data), the cell code of active cells during network service events, handovers (cells exchanging during the services used), also first and last cell used during each call [Zbi13, Rob15, V. 11].

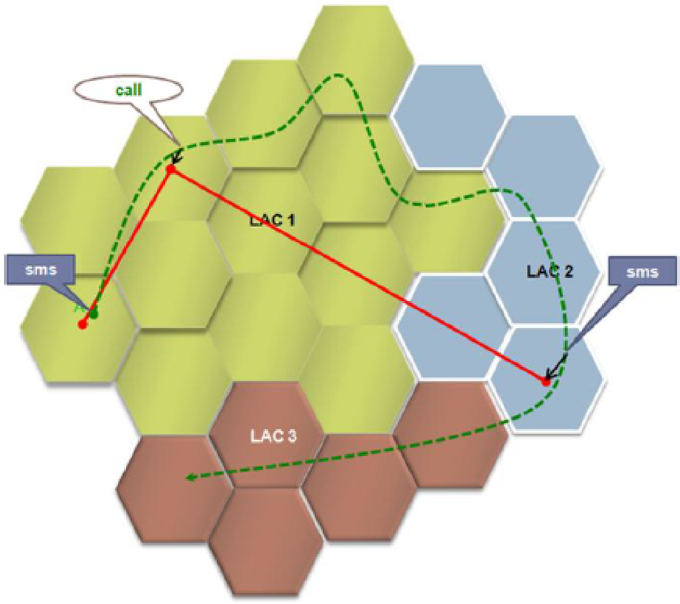


Figure 2.7 – Localization information using CDRs, in dotted line the real trajectory and in solid line the GSM based trajectory [Zbi13].

Conclusions

The mobile phones play a key role as sensors of human behavior [Van12b], because they are typically carried by their owners often and during their mobility. The CDRs could be considered as georeferenced data for human mobility [Van12a]. Hence, it is no surprise that most of the quantitative data about human mobility have been gathered via CDRs of mobile phone networks [Jam12, Mar10a, Enr14].

Also, there is a strong correlation between the individual which moves actively and his/her usage to mobile phone [Tho11b] rather than fixed phone (land line), since nowadays almost the populations hold the mobile phone permanently, it acts like a personal sensor, especially the developed mobiles now play the role of special purpose machine either as communication device or as full-fledged personal computers. So, the individuals connection to mobile networks will be for different purposes as internet browsing, reading Emails, entertaining, updating applications... etc., in addition to the periodical updating of the mobile networks for the mobile phones locations, which means continuous spatial monitoring in real-time property, with chronological archives of data records for multi periods (days, weeks, months, and years).

Additionally, urban mobility studies have some limitations such as the need to predetermine the number of areas to be identified automatically, this is imposed to predetermine the threshold of the minimum density to consider. And they identify dense areas by overlaying fixed grid on the geographical region, this might not correspond to the real shape of the observed dense area [Enr13, Mar10b].

Chapter 3

Spatio-temporal Models based on CDRs, MAS and GIS

Introduction

Modeling and simulation are the most effective tools to study the urban environment [Raw09c], its evolution (growth) [Raw09d] and urban pulse (mobility patterns) to make beforehand strategies, and planning for future management either in normal or abnormal conditions (daily life patterns, catastrophic cases, or crowded cases), and also to evaluate the current urban system. Modeling could be conceptual, symbolic, or mathematical depending on the scientific purpose of it [Yik13, Arm10, Raw08a].

The fast population growth gives highly urbanization development [Dan15], especially when comparing the number of megacities (cities of more than 10 million inhabitants). In 1970, the world enlisted only two megacities, but this number is expected to grow to 37 in 2025. The population growth implies changes in the urban system [Dan15], and these changes could cause either upward or outward city expansion [Yik13]. However, in order to reach an effective city planning and describe the organized complexity of cities, hence the human mobility has to be investigated in time and space by using CDRs at the cell tower level, so for example to find clusters of locations that have similar activity [Jam12, Joh15].

Many researches have been focused on extracting aggregated patterns in different urban areas from mobile phone data, such as hotspots [Sah14], clusters [Jam09], and Points Of Interest (POIs) [Lin12]. However, there has not been sufficient researches on characterizing and classifying mobility patterns in different urban areas from a dynamic perspective, i.e. analyzing these patterns with respect to time [Yih11, Mar10b, Mar14a].

Also, in common methods the clusters with high numbers of objects in specific geographical area are using the spatial properties of the denser regions data [Yih11]. Traditional clustering techniques, like k-means are used [Jam09] for grouping points in space with similar values of density [Dar13]. As well as, these methods demand determining the numbers of clusters or estimating distributional assumptions of the data [Tho13], which are hard to estimate [Jam12, Enr13, Siq11, Jam09, Mar14a].

However, GIScience is evolving with continuous attempts to solve some research concepts such as knowing where is everything along time, the impact of individuals either as a consumer or producer of the geographical information, the technology of dynamics, and the great influence in education [Yik13]. The GIS is a system designed to capture, store, manipulate, analyze, manage, and present all types of geographical data. The acronym GIScience is sometimes used for Geographical Information Science or Geospatial Information Studies to refer to the academic discipline or career of working with geographic information systems [Raw09f, Raw09b, Eli12]. Its features make it effictive tool to investigate the spatio-temporal data.

This chapter is organized as in the following sections: The first section is the mobility modeling with GIS explores the GIS contribution in analyzing and presenting the spatio-temporal data visually. Then, the spatio- temporal analysis tools section presents Voronoi

Tessellation technique as a tool to analyze the urban area, and shows some researches, which extracts human mobility patterns via spatio-temporal property.

The basic models of Attraction mobility in city with Gravity and Radiation section, it explores the common models, which simulate the fluxes among cities. Their simulation is in the marco level of abstraction. The section of multi-scale modeling of human mobility classifies the simulation models in several perspectives. Also, it presents some of related researches. As well as, it gives more details about the Macroscopic, Mesoscopic, and Microscopic models, in addition to the related researches for each of them.

The next section is modeling mobility dynamics and spatial constraints, which explores the mobility models with respect to the physical laws, in addition to presents some related researches. As well as, classifies the mobility models with respect to dynamics and spatial constraints. Then, in urban mobility patterns section the common used parameters are presented, which investigate the mobility patterns, and several related researches are presented.

The integrated MAS and GIS platform section illustartes the Multi-Agent System (MAS) models, and their collaboration with GIS platform for urban simulation. As well as, it mentions the common Agent-Based simulation platforms. Finally, the conclusions endorse that urban dynamics can be modeled effectively using ABM.

Mobility Modeling with GIS

Modeling human mobility has become an important research question in various fields, such as Geographic Information Science, Transportation, and Physics [G12, Enr14]. So, in order to represent the mobility information on real earth location, then GIS or any of its applications can be used [Raw09b, Eli12].

GIS contributes to the huge spatio-tempral data to be analyzed and visualized as collective dynamics in somehow close to a real time issue [Eli12]. GIS influences several scientific fields such as urban planning, traffic management, sustainable infrastructure establishment, and tourism. The combination of the social sensor data with data mining techniques [Bal13], and GIS could give deep global perception of the urban dynamics, towards socio-technical conceptualization for the city by elaborating the implicit patterns and behavioral interactions [G12, Mic11, Dap15].

In the simplest terms, the GIS is the merging of cartography, statistical analysis, and database technology. For extracting any additional information about specific spatial data, GIS could be used as data model, which has several modes according to layer based approach for representing geographic information in the map as in figure 3.1 [Seo16, Raw09f].

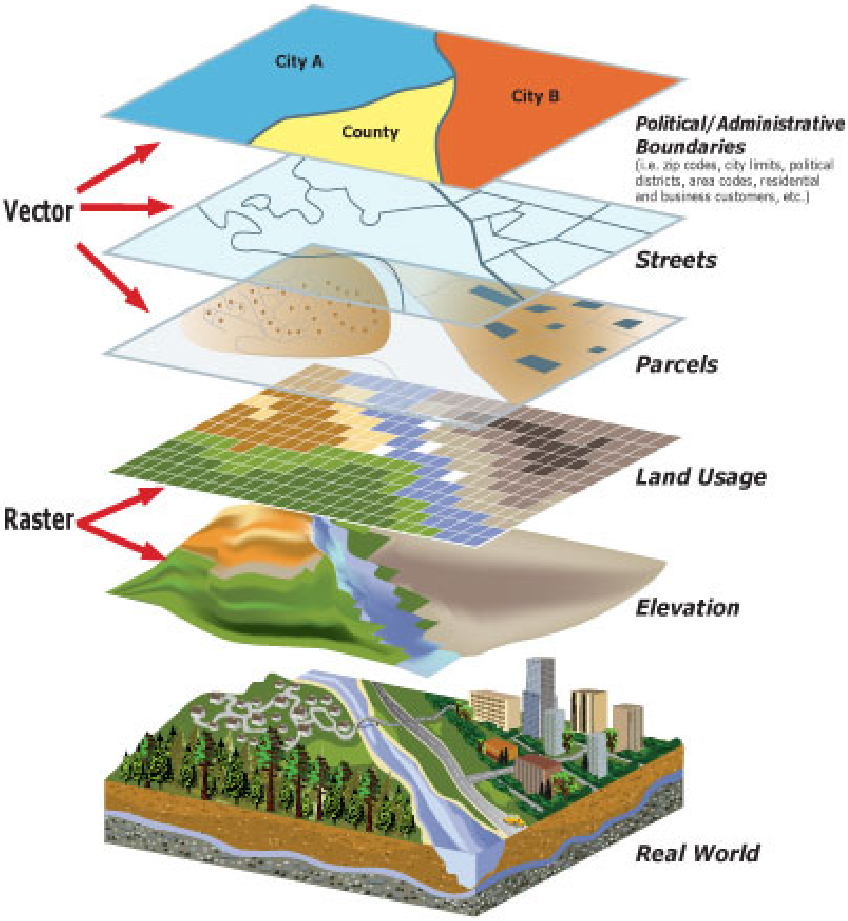


Figure 3.1 – GIS information layers [Seo16].

The communication devices such as the mobile networks or social media platforms produce digital traces for their users either voluntarily or not. This type of collective data can give powerful indications that are affecting the urban systems design and development. For understanding the collective human behavior of urban city, hence the geo-visualization techniques could be used [G12, Enr14].

Spatio-temporal Analysis Tools

The urban analysis could be achieved by using the spatial property indication of CDRs with contribution of the most known tool (Voronoi tessellation), where the specified region is divided into polygonal partitions, each tower is correlated with a polygonal region, which is divided according to Voronoi cells, so each Voronoi cell is covered map points that are most closely to specific tower [Van13].

However, using Voronoi cells makes it possible to project the temporal data on their related geographical location [Joh15, Pu 09a], hence the rhythm (pulsation) of the city will be followed and represented easily, either according to week days, or hours of the day. Depending on the mass amount of data the details can be extracted to represent the observed region activities, for instance the peak/ off-peak hours of work, the peak hours density, the high spot regions ...etc. All these consequences can give aggregated patterns of city life along the observed time [Jul08a, Rob15, Mar10b, Van12a].

There are many types of researches that extract human mobility patterns, they manipulate the mobile data from different views. However, there are some anomalous events that could be emerged, which are deviated from the mean normal activities patterns, so they need to be showen immediately on the maps in its real time existence. For instance the most known system called Wireless Phone-based Emergency Response (WIEPR), that is used for monitoring any catastrophe, terrorist attacks or natural crisis, then plot these events on GIS maps for the concerned region, in order to take the reaction by simulations software [Jul08a, Mic11].

The researches work on very large data sets as in [Van12c], which makes an analytical model to approximate census variables using the CDRs, the area covered by BTS is approximated by voronoi diagrams to get the geolocation of the connected mobile network using BTS. It uses Detection Residential Algorithm, hence presents the ability to build predictive model from the CDRs. It proposes an analytical approach that combines large- scale datasets of mobile phone records with country wide census data to reveal findings at national level. Main results show correlations between socio-economic levels and social network or mobility patterns [Van12c, V. 11].

[Xia12] concentrates on human mobility patterns and taxi transportation in urban. It proposes a prediction method based on a simulator, that predicts the spatial-temporal variation of passengers in a specific hotspot. It uses GPS as trace application, and uses patterns of pick up quantity. This work focuses on predicting human mobility from discovering patterns. The number of passengers from urban hotspots.

Another researchers focused on simulating positioning as in [S.F09], it works on E-OTD positioning method for measuring time required for BTS signals to arrive between two distant geographical points. However, Enhanced Observed Time Difference (E-OTD) method is one of the promising and fairly developed positioning technologies, that has been standardized for GSM systems.

However, this method has been utilized to locate the mobile equipment position. Whereas, [Van10] aimed to characterize and automatically identify gender using only behavioral, social, and mobility information obtained from CDRs. In the context of developing economies, it is particularly relevant, since pre-paid clients for whom there is no specific gender information account for the large majority of mobile phone users.

Whereas, [Mar10a] used the Spatio-Temporal Patterns System (STPS), which is a regular expression query language used to extract the spatio-temporal patterns in huge CDRs databases. The design of the language take into its consideration layout of the areas being covered by mobile towers, as well as the areas that label places of interested (e.g. neighborhoods, parks...etc.), and topological operators.

Basic Models of Attraction Mobility in City with Gravity and Radiation

Urban mobility system is a complex system, and several models attempt to model and simulate the urban dynamics based on individuals transitions among urban regions. The gravity and radiation models are common models, that are used to mimic the attracted fluxes of individuals among cities. They are based on the population size and distance between these cities.
Such models deal with individuals transitions from point to point or from region to region like Gravity model [Bal13, Fil12], which is invented by Ravenstein's laws of migration, hence according to individuals transitions from source to destination, as well as the distance between the source and destination, could be said from i to j, then the formula would be as in equation (3.1) [Ale14, Mic, C H14]:

$$T_{ij} = \frac{(m_i^{\alpha} n_j^{\beta})}{(r_{ij}^{\gamma})} \tag{3.1}$$

Where, i is the source and j is the destination locations, with populations of m_i and n_j; r_{ij} is the distance between them, whereas α, β, and γ are adjustable parameters to regulate the data.

Another model called Radiation Model [Ale14, Fil12] which acts as parameter free mobility, in which people are mobile and interact, this model pivoted on population density regions and the surrounding regions [Mic, C H14]. The average flux between two regions i, j could be formulated as in equation (3.2):

$$T_{ij} = T_i((m_i n_j)/((m_i + s_{ij})(m_i + n_j + s_{ij}))) \tag{3.2}$$

Where, s_{ij} represents the total population in the circle of radius r_{ij} centered at i, excluding the source and destination population, and Ti is the total outgoing flux at region i forward to region j. However, this kind of models had been used to predict people mobility (commuters) between small and big cities, based on population density and traveled distance.

These models give the macro influence of population repartition inside cities, but not how people are evolving individually or collectively, hence other models are developed to overcome this limitation as will be presented in the following sections.

Multi-scale Modeling of Human Mobility

Human mobility studies have to be highlighted in order to improve current modeling tools, and to have more realistic simulation results with flexible skills, hence this will influence greatly the precision and the accurate vision to the observed system. In general, these studies are done basically in several concepts such as Macroscopic, Microscopic, and Mesoscopic level [Nir08, Arm07].

The works in [Hen71] and [Fru71] are focused on the individuals supporting during emergency conditions, they used macroscopic level models. Thereafter, the recent advancements in computer sciences gave a strong push to this field, so the level of studies conducted in this field are developed to be more efficient via the concept of modeling in microscopic level, as in the initial works of [Min51] and [Qua57]. The authors elaborate the effect of socio-psychological factors during contingency conditions, whereas they explain the crowd behavior in different situations (panic/normal).

In [Qua57] the new emerged behavior as a consequence of individual interactions had been highlighted. Then the work in [Sim95] focused on engineering issues like entry, and exit of individuals. Meanwhile, the psychological studies extracted the individuals motivations, and their effect on crowd behavior [Dap15]. Then, afterwards the attempts to understand human behavior in specific situations, the urban human mobility emerged as a gradual progression in researches extension.

The simulation models of human mobility are classified according to their space representation, which could be continuous, grid based, or network structure [Mic11]. Also, They could be verified according to their intent either to be specific or general, or the level of abstraction as macroscopic, mesoscopic, and microscopic. Many scientists attempt to classify urban models and list them in extensive classification schemes based on several features, methodologies, application fields and modeling concepts. The considered models Macroscopic, Mesoscopic, and Microscopic are classified with regard to the level of abstraction, as will be explained in the following sections [Nir08, Yik13, Arm07].

Macroscopic Models

The Macroscopic models focus on the clustering or aggregation process to the individuals dynamicity of the crowd through associations among the flow (Q), density (K), and speed (V), founding the inverse relationship between average speed and density is given by equation (3.3):

$$Flow(Q) = Speed(V) \times AverageDensity(K) \tag{3.3}$$

As well as, [Tek02] developed a microscopic model to simulate individuals mobility, and considered the parameters movements as the flow, average speed and area module. Area Module (M) which is the density reciprocal in order to relate the individual factors with their flow space as shown in equation (3.4):

$$Flow(Q_p) = \frac{AverageSpeed(V_p)}{AreaModule(M)} \tag{3.4}$$

[Daa04] submitted SimPed model for simulating passenger flows in public transport by exploring individuals density and speed association. This type of modeling is well suited for the huge population and large crowd systems. The insight to the pedestrian is important individuals behavior (unthinking/intelligence), hence it will give different behavior for each element, then affect the whole crowd attitude (behavior), according to the situation which governs the event, however it will be considered as thinking fluid.

Mesoscopic Models

These models are located between Macroscopic and Microscopic models according to their level of abstraction, they have the concept of aggregating the individuals/agents in smaller groups than in Macroscopic models. This would produce more detailed analysis for the observed system. The mesoscopic model is proposed by [Han03] as an online pedestrian simulation flow in public buildings. Whereas, each group of individuals has its own behavior strategies. Hence, the model would have transited groups between network nodes. Similar work can be found in [Tol04], which used this issue to simulate pedestrian traffic in public buildings, also they accomplished prototype simulator to test performance and accuracy of their simulator. [Han03] emphasized on this issue to be effective for simulating large population as individual groups [Arm07].

Microscopic models

These models consider each individual of the population is an agent that occupies a specific space at a specific time, which is capable of giving elaboration for population behavior [Nir08, Yik13, Hé16]. The researches [Tek02, Sti00] highlighted the interaction between individuals. In spite of their limitations related to analytical manipulation and computation effort and cost with considering the level of detail, but they are sufficient to submit a more realistic individuals mobility. These models can be classified into four groups:

Physical based models

In these models, the crowd is composed of individuals which are represented as physical particles with specific properties. The interaction between individual particles are obtained by models based on physical field theory. One of the most popular model of this category is the model of social force proposed by [Hel02].

Cellular based models (Cellular Automata (CA) or matrix-based system)

These models consider the event space if group of cells represents the event area or obstacles, which contains one individual, group of individuals, or regions with attributes, the individuals are moving from cell to cell under occupancy rules as in [Xia06, Blu99], where the authors build a CA micro-simulation model for bidirectional pedestrian walkway. As well as [Sch02] submits a cellular automaton model to generate collective effects and self-organization phenomena in individual traffic. However, not all simulation attempts were successful in reflecting the mobility, where [Xia06] attempts to simulate pedestrian crowed mobility, but the individuals mobility appeared to look like hopping (skipping not moving smoothly) across cells. Also, [Pab15] develop a CA-based model for simulating the urban growth of Madrid city.

Queueing network models

These models simulate the individuals flow in specified environment such as building within predetermined scenario like rescue scenario. They are networks of individuals queues that are flow in a their specified environment. Each individual can flow independently in its own queue within the network and can interact with the other objects. They are accomplished to model the rescue from buildings. [Gun94] built a discrete event based model, the probability issue is used to represent the individuals mobility towards a goal under priority rules of the their interaction.

Multi agent models (MAS)

The multi-agent simulations are very suggestive for spatio-temporal dynamics, since they elaborate the relationships between micro-level individual actions and emergent macro-level phenomena [Bat99]. The MAS is composed of Agent-Based Models (ABM), which in turn composed of multi-agents interact among each other in simulated environment. These models build artificial environment composed of agents, which have the ability to interact in intelligence and adaptability with each other [Arm07, Chr13a, Hé16]. In these models the agents are acting based on some strategies. The agent is an atomic unit in the computer program and it is goal-directed [Raw08a]. However, the interaction of agents are based on predetermined mobility conditions like leader, follower, and inhibition agents.

This kind of simulation is very effective for large scale rescue scenario and complex systems, or modeling crowd behaviors. [Xia06] accomplished multi-agent system framework, which model emergent human social behaviors (competitive, queuing, and herding) at the microscopic level.

However, they have some limitations to simulate the dynamic city, because the city is very complicated environment, so the level of details might be either as abstract model with lack of some features, or very complicated model with more details. Also, human have irrational and subjective decisions, which make it hard to be simulated, calibrated or quantified. The most influenced point is the big size model, which make it needs high resources to simulate.

The Multi-Agent System MAS adapts more complicated of agents interactions to be simulated effectively as could as possible [Raw09c, Arm07]. They had been proposed to micro-simulate land-use changes, contributing environment, transport, and economic models to obtain complex urban systems [Yik13]. [Raw08b] used the swarm algorithm for modeling the urban dynamic system based on ant system, depending on its development on geographical system environment (attractive areas).

MAS is an effective tool for building the complex systems as multi agent systems, where each agent has its own dependency, in spite of its basic common properties with the other system agents [Raw09e, Raw09d]. As agent- based models, ILUTE and UrbanSim projects [J. 05] are examples of the models that deal with the pedestrians concepts in the urban environments, which give more details (disaggregated), and represent a radical transformation in the model quality from the static to dynamic modeling [Mic12].

The agent-based modeling provides a high variety of representation which is very important facility. The agent can represent any kind of structures either as individual agent or aggregated ones in the reference system, and it coordinates with the spatio-temporal scales in any abstraction level as modeler desires. All this pivoted on the modeling targets of the reference system or its concepts to be shown. This ability is very useful, especially when respecting the changes that could be appear during simulation process, like new events or structures, agents emergence as in pheromone trail built by ant, evolution social clusters in any population. However, Gama platform is a practical tool to model and simulate with GIS and Agent concept. This platform makes possible to deal with dynamical changes in the simulation process, and gives a name "$emergentagents$" to the new emerged agents during the system life cycle [Pat12c, Hon13].

As well as, [Gre08] simulates large scale microscopic evacuation, where every evacuee is considered as an individual agent that optimizes its personal evacuation route. Additionally, researchers in [Chu09] build an agent-based modeling and simulation using Repast platform, the simulator represents secure crowd evacuation in the environment suffering a fire, this project tests several parameters of agents to explore their effect.

Modeling Mobility Dynamics and Spatial Constraints

The mobility models represent the movement of mobile users, and how their location, velocity, and acceleration are changed over time. Such models are frequently used for simulation purposes, especially when new communication or navigation techniques are investigated [Mar05]. Most popular tracing mobility models are in two major categories: The Statistical-Based Model and the Constrained-Based Topology Model [Min06].

Physical laws of acceleration, velocity, and rate of change of direction are also included within mobility models. The current velocity of mobile node may depend on its previous velocity, it is called in this case Temporal Dependency of velocity (Gauss-Markov, Smooth Random) Mobility models. Whereas, in Spatial Dependency models, various nodes velocities are correlated in space, like Reference Point Group, Column, Pursue Mobility model, and Nomadic model. The geographic restrictions are considered in the mobility models, where for example the motion of vehicles is bounded to the freeways or local streets. Pedestrians mobility may be blocked by buildings or obstacles, the common proposed models are Pathway and Obstacle Mobility model.

According to the assumption that each mobile phone can be considered as an individual existance, so this assumption makes it possible to use the mobility models in both human/ network modeling and simulation, interchangeably. Inspite of, the characteristics variance between networks and human. For example, Random Waypoint Model [Mar05] proposed by [Dav96], where the nodes move independently to chose randomly the destination with randomly selected velocity. It is considered the benchmark mobility model to evaluate Mobile Ad hoc Network routing protocols, because it is simple and the sources are publicly available.

The relevant mobility models for dynamics and spatial constraints could be classified as in figure 3.2 [Pat13].

Urban Mobility Patterns

The human activities are differentiated from individual to another, this can be noticed from their mobile usage, where some of them have high frequency of mobile activities, in contrast to others, which have less usage with long inter-event distribution [Tho11b]. However, the individuals groups activities can be distinguished according to their mobile activities frequency (calls/SMS total number) distribution [Jam09], this is done using the probability density function $p(\Delta t)$ [Jul08a, Mar09]. As well as, another distributions could be estimated like $p(\Delta r)$ or $p(\Delta r|r_g)$, this will be explained in chapter 5.

The works in [Zbi13, Jam09, Xia10, Mar09] explore the modeling of human mobility, and they conclude from their studies that human trajectories clarify high degree of spatio-temporal regularity. The individuals are characterized by non time-dependent travel distances, they are also characterized by high probability of returning to few previously passed locations. The

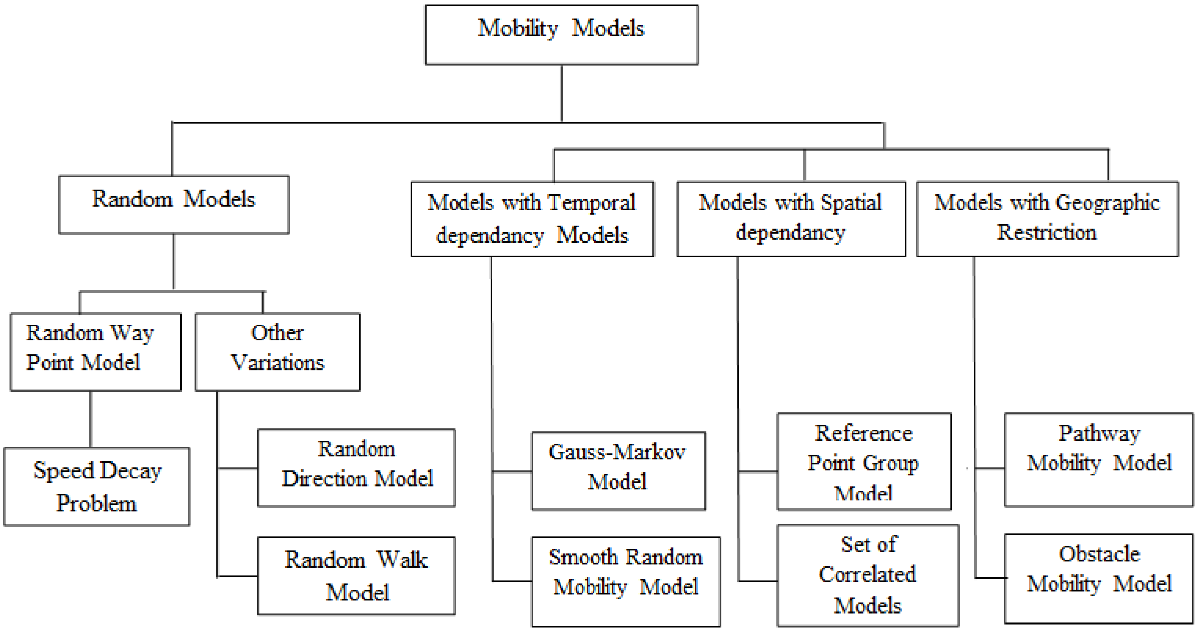


Figure 3.2 – Mobility models relevant for dynamics and spatial constraints [Pat13].

mobility patterns of the individuals have almost the same patterns in predictability average ratio of 93% although of the varying in dynamic histories and social features due to human behavior regularity.

The visited locations are connected by mobility trips and constructing movement oriented networks. These networks are named mobility motifs, which give a view of the daily dynamic activities. These motifs could be assumed as general human mobility characteristics, then to be used in the travel time analysis for urban activity modeling and simulation. The mobile network data have main benefits in revealing human mobility, because the mobile phone is ubiquitous now, so their data (active/passive) would cover wide range areas, and pick up huge population number, which makes it the most suitable media for human mobility studies [Sah14, Sah13].

The authors in [G12] used the handover density to highlight the most dynamic regions in an urban environment. Also, the correlation coefficient is used to explore the relation between human behavior and the spatio-temporal. Hence, it elaborates the average patterns of collective human behavior at appointed environments in macro scale level (the more general activity and mobility patterns). Many researches used CDRs to study the collective and individual human mobility [Mar08], the segmentation of urban spaces [Rea09], and understand the events effect on the people attraction [Fra10]. [Phi10] suggested the activity-aware map based on POI using mobile data to uncover the dynamic of inhabitants. As well as, [Rat10] deal with large telecommunication data (landline calls) of Great Britain in order to explore human interactions, it is emphasized on the highly interactions correlation with administrative regions for urban planning and transportation purposes

[Cal11] invented the co-location perspective, which explores the behavior of the interacted telecom network users, which they call each other frequently and share same spatio-temporal features in the city. Hence, examining the relation between calls and physical location. Whereas, [Gir08] presented the digital footprints (left traces/ user-generated electronic trails), which are setting by individuals during their mobility in urban space, CDRs and georeferenced images of Flickr are used in order to trace the tourists presence and mobility in Rome city. Also, [Sev10] emphasized that urban mobility is highly correlated with individuals behavior during mobile phone usage. The used data are the activity patterns of the network cells.

Data mining methodologies contributed in representing human mobility. The authors in [Yua13] attempt to represent human mobility patterns depending on mobile spatial indication by building human mobility models, and then using data mining methodologies for acquiring the desired information, which are important to explore the whole image of human mobility inside his environment. They also acquire the impacts of this relation on urban environment development and the readiness to face any unusual events according to predetermined understanding and prepared planning. This research emphasized on important nature of the mobile data, which is the uncertainty and inaccuracy, due to the nature of human life patterns [Lin12, Sah14, Sah13], therefore it will be reflected on mobility models and extracted traces for human mobility as in figure 3.3. The northeast China was used as case study for this project. Considering it as high populated country with speedy developed environment, so it is easy to relate between human mobility and applied communication tools like mobile. After analyzing and understanding these relations, the authors emphasized on the geographic representation in order to have a sufficient description of the extracted mobility patterns on the map as in figure 3.4. While in [Lin12] the researcher suggests to elaborate anomalous individual trajectories, and to verify the anomalous events.

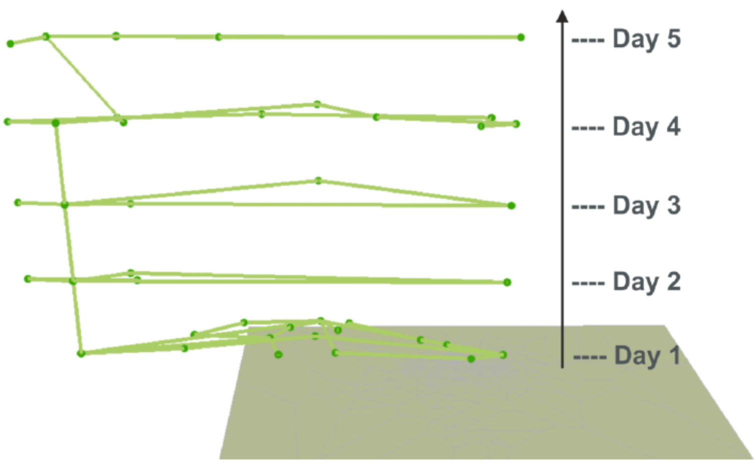


Figure 3.3 – Trajectory of mobile user during several days [Yua13].

The most common indication on the mobility is the handovers average, population (activities) density of each network coverage region, and the In/Out visits to each of the aforementioned regions. Many researches are performed on the mobility modelling, which are focused on how to model any system, and be as close as possible to real life scenarios [Pat13, Siq11].

[Yih11] investigated mobility patterns by identifying the aggregated mobility patterns, using hourly time series to extract and represent the dynamic mobility patterns in different urban

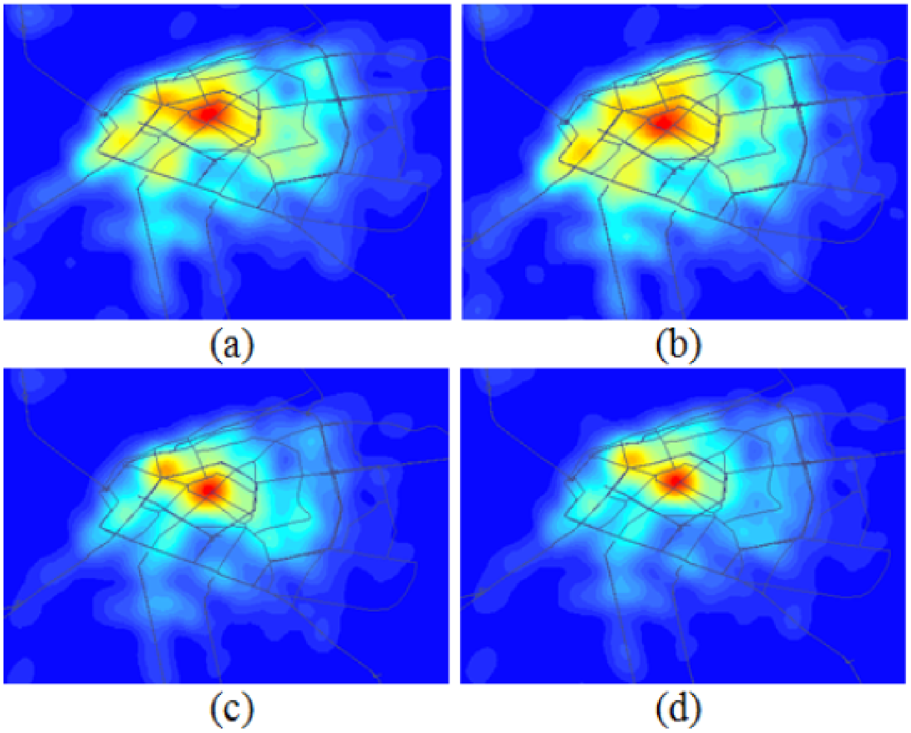


Figure 3.4 – Extracting urban mobility and activities from georeferenced CDRs [Yua13].

areas. Dynamic Time Warping (DTW) algorithm is applied to measure the similarity between these time series, which also provides input for classifying different urban areas based on their mobility patterns. It is well-developed algorithm in the field of speech recognition and signal processing for matching two time series. It has been used for classifying dynamic mobility patterns in different urban areas using hourly time series, taking the urban perspective for analyzing the data. It also uses outlier detection algorithm for extracting the abnormal activities, and uses Voronoi polygons to divide the study area.

[Enr13] uses an automatic identification method to identify a dense area automatically. It investigates the CDRs of urban dense area. Also, uses the natural tessellation technique (produced by Voronoi) of the spatial domain to construct a graph. The Delaunay triangulation, and Optimizing spatio-temporal query system are used to develop Dense Area Discovery (DAD) method.

Other researchers intend to detect the dense area as in [Alb13], where they propose novel technique called Adaptive Dense Area Discovery (AdaptiveDAD), in order to detect dense areas that define the concept of density using the infrastructure provided by mobile phone network. They evaluate and validate this approach with real dataset containing CDRs of fifteen million individuals. The adaptive and non-parametric method for identification of dense areas is based on using the ubiquitous infrastructure provided by mobile phone network [Alb13].

Integrated MAS and GIS Platform

The MAS composed of ABM models, which are constructed to imitate a specific system, each model composed of autonomous, heterogeneous, and interacting components called "*agents*" [Koh13, Duc12]. These agents could be static, dynamic, or alternated between these two states, depending on predetermined model rules. The interactions among agents will produce new states or even new subsystems inside the entire system, in addition to the possibilities of dealing with the outside world of the system, and the space which conserve the agents named environment.

The environment that holds the agents with their processes could be in simple sense as a space that contains the system agents. This space could be continuous (spatial space), discrete (grid cells), or connective social network [Mic12]. Thus, if the environment has spatial nature, then agents must have the coordination to this environment, whether they are static or dynamic, so the geometrical metrics will be most important in this type of models. Therefore, the interleaving between the GIS and the agent-based models is the high spot point in most simulation models even if the environment is only a small geometric space [Mic12, Raw09b].

The recent great tendency is to deal with catastrophic disasters, cities planning, diseases spreading [Mic], traffic forecasting, decision making for rescue people with disabilities in emergencies ... etc. [Koh13]. Any of these possible events needs to work on huge data and geographic data for granting the simulation model more reality, and since it is almost the simulation platforms are dealing with only a grid to be the background of the model, so the resulted model will be far away from reality [Pat, Pat12b, Pat12c]. Therefore, the need for a simulation platform that integrates a real GIS data emerged, as the GAMA platform that is an agent-based simulation platform, which deals with all obvious needs.

There are several agent-based simulation platforms, but not all of them have the capability to deal with complex environment of GIS data [Pat, Pat12b, Pat12c, Pat12a]:

1 Swarm: Its original version does not allow the integration of GIS data. However, this is resolved by adding a library called Kenge that allows loading layers of GIS data. It doesn't has spatial primitives or able to store the resulted environment.

2 Netlogo: The environment is made of 'patches' grid, though agents have their coordinates defined in continuous space. The GIS extension is then added, but much more advanced spatial analysis operations are not offered.

3 CORMAS: It is dedicated to the modeling in ecology and especially the natural resources management essentially with space representation and interaction.

4 CORMGIS: The GIS integration is done through a data-connection to ArcGIS. Hasn't GIS primitive (union, intersection, ...etc.).

5 Repast J: The GIS integrated using the OpenMap library. It covers only little primitives of GIS: importing/exporting shapefiles and raster data, some geometrical operations, access to data attributes, ...etc.

6 Repast Symphony (Repast S): Is the up-to-date version of the Repast toolkit. It is based on Geotools, and provides additional GIS primitives. It allows to directly model a network of lines as a graph and to compute the shortest paths from one point to another. It allows as well visualizing and managing 3D data. However, the number of GIS operations available is limited.

7 GAMA platform is a multi-agent simulation platform that builds complex model integrated with GIS data, it benefits from OpenGL to simulate in 3D, and it is still in progress for more achievements to improve it's capabilities. The almost known simulation platforms are very limited when working with GIS data, require high level programming skills. But GAMA overcomes this difficulty, by enabling automatically model GIS data as agents and optimized agents dynamic representation.

Conclusions

The urban environment is considered as a complex system, hence the ABM is very effective tool to model the urban dynamics. It supports the researches on the network science, social networks, and human geographical systems, with their capability to connectivity and producing new developed complex network systems, during the period of time with relevant locations (spatial space). These models manipulate the system by dis-aggregate in the level of agents (individuals), and give high degrees of explanation.

However, they have limitations in their reflection of the system reality, which is depending on the degree of abstraction (quantity of the system variables and parameters), and each model is constrained in imitating only the environment that is modeled (constructed) for it, as well as the model researchers are restricted by the toolkits, platforms efficiency in studying their entire system [Mic12, Raw09b].

The common two major perspectives in exploring human mobility patterns, which are extracted from mobile phone data, that are: The individual perspective, where the individual trajectory patterns are related to the theme of pattern recognition in Physics and Computer Science [Yih11, Cor13, YAN11]. And, the urban perspective, where the cities can be considered complex systems, that are constituted by different processes and elements [Yih11, Arm10].

Chapter 4

Mobile Phone Traces Analysis: Case Study Armada

The chapter has been partially published in [Beh14]

Introduction

There exist a lot of aspects of mobile networks, which need different categories of studies, and provisions such as privacy, management, tracing, development, topology, protocols, hardware devices, user's psychology, user's requirements, society ethics, laws, and norms. The deployment of mobile devices made data transfer spread rapidly in several modes such as the wired, wireless, mobile devices carried from place to place [Aug05], the information flow and its side effects on the whole society all over the world opens new argument regions such as the human dynamic behavior [Jin04]. Using the CDRs gives huge data, but they are sometimes ambiguous, since the raw data would be meaningless if not analyzed or verified. Therefore, useful features must be extracted to transform these data into information.

In order to solve this ambiguity, special techniques have to be found, for extracting the additional required data, such as cells connections and their directions at specific times. Whenever the CDRs are analyzed, and events of the mobile network are extracted as directed mobility and densities, then the CDRs will become an interesting material to discover individuals density and mobility [Chr13a, Mar10b]. They help to find out a lot of information about individuals points of interests (POI) and mobility [Lin12], which are very important for scientists of city planning, disaster avoiding strategies, understanding human life patterns, discover most POI in any region...etc.

This chapter is organized as in the following sections: The Armada DB as case study section, it explores the data material of this thesis. It is branched into two subsections as: Armada of Rouen, and CDRs of Armada. The first one presents the Armada festival, which is our observed event. Then, the second one elaborates in detail the DB of the observed event. The next section is the research methodology in analysis phase, which presents the main used techniques and platforms, which are used to analyze the observed data.

The data analysis and visualization section presents the results of individuals densities analysis, it investigates the observed data in Macroscopic level of abstraction. As well as, it is branched into six subsections as: The area sectors, the center sector, the east sector, the west sector, the north sector, and the south sector. Each one of the mentioned subsections explores a sub-area of the observed total area, and it shows its analysis results, as in [Beh14].

Then, the individuals densities over Voronoi diagrams section, which presents densities variation along one day, and explores the used technique and platform. Finally, the section of conclusions, it endorses that densities analysis gives indications about the life patterns of urban.

Armada DB as Case Study

The essential material used in this study is Armada database generated by the given mobile phone operator (Orange Company, the historical main operator in France) in the form of CDRs [Org16].

Armada of Rouen

Armada is the name of the famous marine festival occurring every 4 to 5 years periodically, it takes ten days period [Arm14]. Large sailing ships, private yachts, naval ships, and even military frigate schooners are gathering from twenty different countries in a world free spectacle in Rouen city (capital city of Upper Normandy in France) [Mad14]. This case study is about the fifth edition of this event in 4^{th}- 15^{th} July 2008 [Ara14]. The activities of Armada start from from 10:00 to 20:00 every day with miscellaneous activities. The last day of this event coincides with 14^{th} July (French national day). It attracts 8 million visitors in addition to Rouen citizens [Die12], which is estimated to 460,000 in 2008 [Roa14, Uoa14].

Many artists and professionals participate in this occasion by their presence with their art works, this festival like the Olympic Games held one time each four years. In this event, Armada Rouen 2008 sponsored "Sail Training International", the association which organizes tall ship races around the world, prologue in Rouen city on the banks of the Seine Maritime heading towards Liverpool [Arm14, Mad14, Roa14, Uoa14].

Miscellaneous activities are held like organizing tall ship races, learn diving, flying balloon, get an egg into a bottle... etc., in addition to the entertainments such as the bands, dances, and traditional parade of teams accompanied with fireworks. The last day of the event was in on Monday 14^{th} July at 9:30 AM. People arrive from all over the world, it attracts 8 million of visitors during the 10 days of the festival [Arm14, Ara14, Daa14, Roa14, Uoa14].

Armada Editions History

Armada festival is a successful maritime festival, where the number of visitors reached approximatelyand six million in 1994, and 9 million visitors in 2003, in addition to city citizens and 6000 foreign sailors came from 15 different countries [Daa14]. The Tall Ship Races are organized by a subsidiary company of sail training international, which is specialized by youth people development in sail training from diverse countries with various cultural mixtures [Arm14, Ara14]. This event has several editions as the following:

1 The first edition (1989) named Freedom of Sails. It commemorated the bicentenary of the French Revolution, held from 9 to 16 July 1989.

2 The second edition (1994) named Armada of freedom. It commemorated the fifth anniversary of the Allied landing in Normandy, held from 10 to 17 July 1994.

3 The third edition (1999) named Armada of the century. It commemorated the end of the second millennium, held from 9 to 18 July 1999. Postage stamps of France were released on 12 July on the occasion, in the Youth Collection, and represents ten participants' sailboats.

4 The fourth edition (2003) named Armada Rouen 2003. It was held from 28 June to 6 July 2003.

5 The fifth edition (2008) named Armada 2008. It was held from 5 to 14 July 2008.

6 The sixth edition (2013) named Armada 2013. It was held from 6 to 16 June 2013.

CDRs of Armada

The city pulsation and life patterns has to be explored and visualized using the observed area. In this phase, the Armada data are treated as any DB system, where the manipulation uses query language to extract the information of the raw data, and then show the results in an understandable and indictable graphs, charts, and maps. The DB is described in table 4.1, which gives its abstract view and main components. Armada DB covers 4-15 July 2008, so its days are classified in two categories, days out of Armada, which are 4^{th} and 15^{th} of July, As well as, days within Armada which fall in 5^{th}-14^{th} of July.

Table 4.1 – Armada Data Base Description.

Property	Value
No. of Days	4-15 (12 Days)
No. of Hours	12*24= 288 but available actually= 273
No. of Towers	190
No. of Cells	587
No. of Alias	615711
No. of Events	50982274
Types of Events	7

During this event period, there are 5 days considered as off days, which are weekends and vacation that fall in the dates 5, 6, 12, 13, and 14. Whereas, the other days are work days that fall in the dates 4, 7, 8, 9, 10, 11, and 15. The Armada mobile network DB has classical CDRs that contain:

- Mobile IDs (alias).
- Towers IDs and positions, which are georeferenced by coordinates (x, y).
- Number of cells on each tower and cells IDs.
- Mobile activities type (call in/out, SMS in/out, mobile hand over, abnormal call halt, and normal call end).
- Date and time of mobile activity, which indeed represents a user activity.

In order to analyze the human mobility, it is supposed to interpolate individual position and to simulate the individuals mobility from the traces by reconstructing his trajectories [Sah14]. The manipulation of these data requires several phases to get the required analysis results, and obtain precise expected knowledge about the individuals densities [Mar10b] and mobility patterns [Eli11, Luc13, Luc11, Rob15]. The common considered assumption is that each mobile represents an individual, as well as its occurrence and mobility. Hence, as a consequence the individuals density and their mobility patterns could be obtained, which concerns the observed area [Chr13a].

This kind of analysis has main limitations, consists in that the data given by mobile phone traces are incomplete to describe the mobility because individuas may move inside the city without using their mobile phones. But the propagation of a mobile phone usage today allows to produce an acceptable approximation of mobility. The mobile phone traces are only caught by the towers, which are located at some specific locations in the city. Hence, the extracted data will be relevant to these locations only without more details about the exterior of their coverage areas. In order to describe the human mobility, it is needed to simulate individuals behavior from these traces, and reconstruct human trajectories [Enr13].

Research Methodology in Analysis Phase

The basic concept of this research phase is to obtain the most active regions along time in the observed urban area. It is performed by the analysis, and representation of individuals concentration within the observed area. In order to integrate and analyze the initial raw database, hence PHP server and SQL query language are used. The data visualization is accomplished using graphs, representing regions densities corresponding to individuals activities. ArcGIS platform is used for geographic representation, and to extract the aggregated data of five classified sectors, that are corresponding to specific spatial sub-areas. The combination of these tools allows lining up the spatio-temporal data on the city map.

The observed area has been divided into sub-areas to summarize the spatial patterns, Voronoi polygons is the division technique as in [Rya06, Ale14], where each polygon is associated with tower, and depicts the area under the main influence of this tower (its coverage area), hence each Voronoi cell centered on one tower, that corresponds to all the closest possible locations of the individual, which are observed by this tower coverage area [Van13]. The aggregated adjacent polygons formulate five sectors, they are described as follows and in as in figure 4.1:

1 The Center sector is around Seine river banks, the place of Armada event.

2 The Eastern sector to the Center.

3 The Western sector to the Center.

4 The Northern sector to the Center.

5 The Southern sector to the Center.

Data Analysis and Visualization (Marcoscopic Perspective)

The individuals densities are captured as the overall activities of the observed area (30x30 Km) along time, this could be elaborated according to the periodic total hours, the following formula is used: $HourSeries = (day - 4) * 24 + hour$.

Figure 4.2 describes the densities for the 12 days period lined up as hour series. Each curve represents one day activities. It is noted that day 9 has lack in its data about 15 hours, its

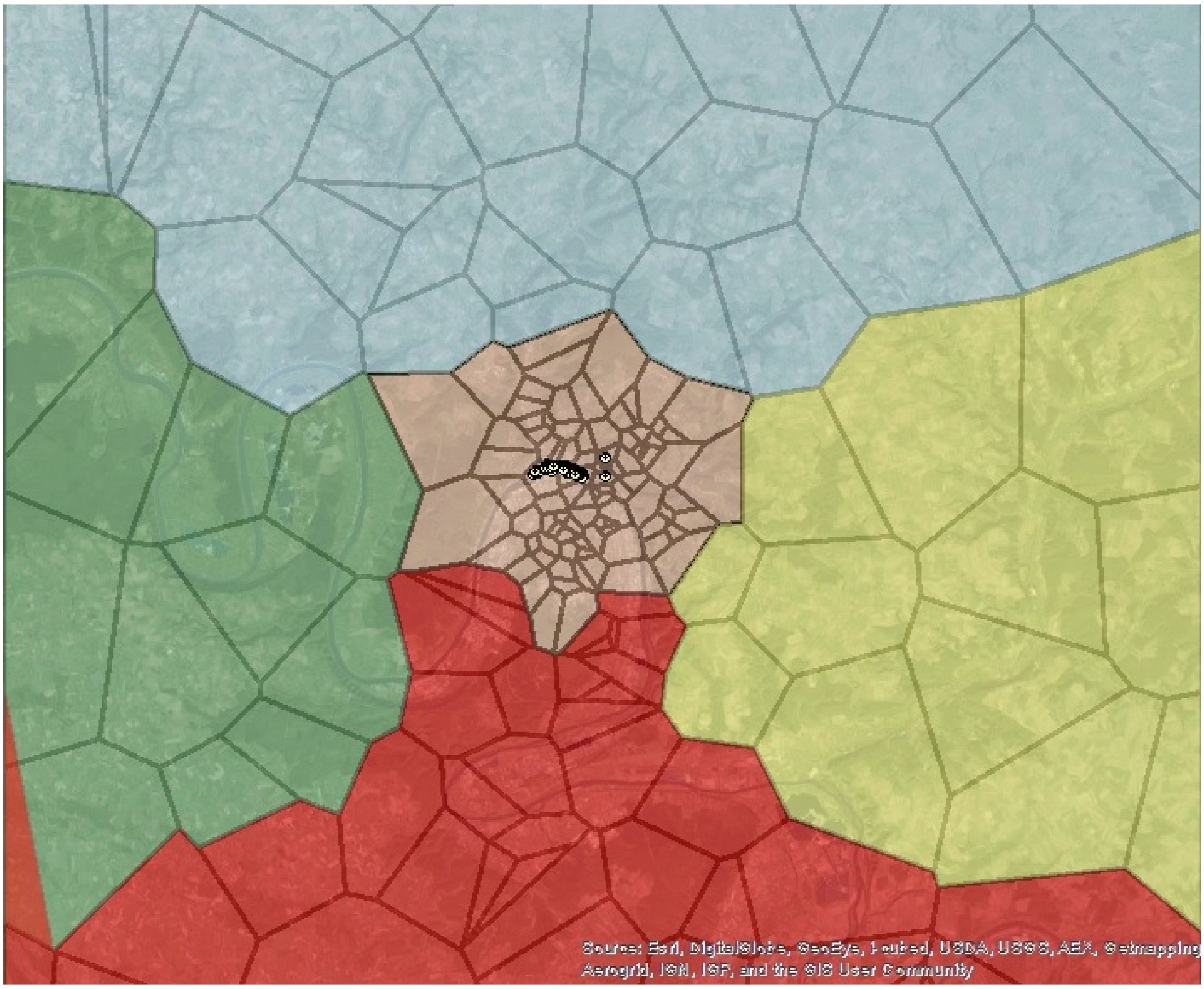


Figure 4.1 – ArcGIS output from Armada DB. Each Voronoi cell is centered on each tower. The 5 sub-areas are described by 5 color zones, they are built by grouping Voronoi cells. Black anchors symbols represent places along the Seine river platforms, where boats are situated during Armada event.

available data are within time series 120:00-135:00 only, this leads to an irregular curve at this time.

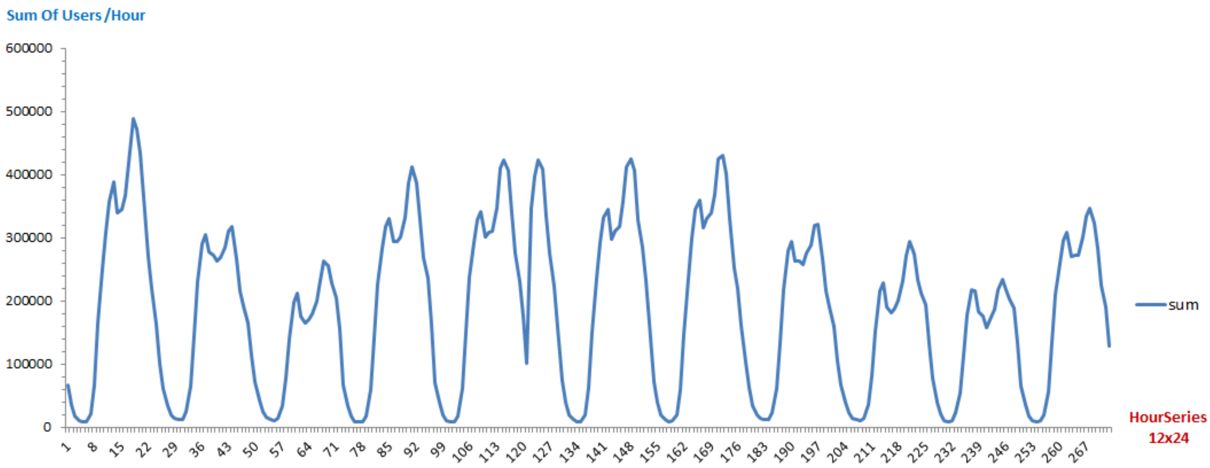


Figure 4.2 – Activity patterns according to people density during days hours over the observed area.

The obtained results are as follows:

1 The patterns curves are closely similar, which means the daily activities patterns are almost similar.

2 The Patterns are regular, except in curve of day 9, where clearly appears there is a lack of its values ($hourseries = 121 - 134$ is exceptional). The city life at this period is stable, since there aren't any abnormal events.

3 As it will appear in figure 4.4, it could be seen that each day curve is increasing from 08:00 to 12:00, then decreasing in the lunch time, then returns to increase again from hour 17:00 afternoon, then to decrease approximately to zero.

4 All work days (4, 7, 8, 10, 11, 14, 15) are similar to each other, otherwise the off days (5, 6, 12, 13) are similar to each other too, so the differences between them are shown in figure 4.3.

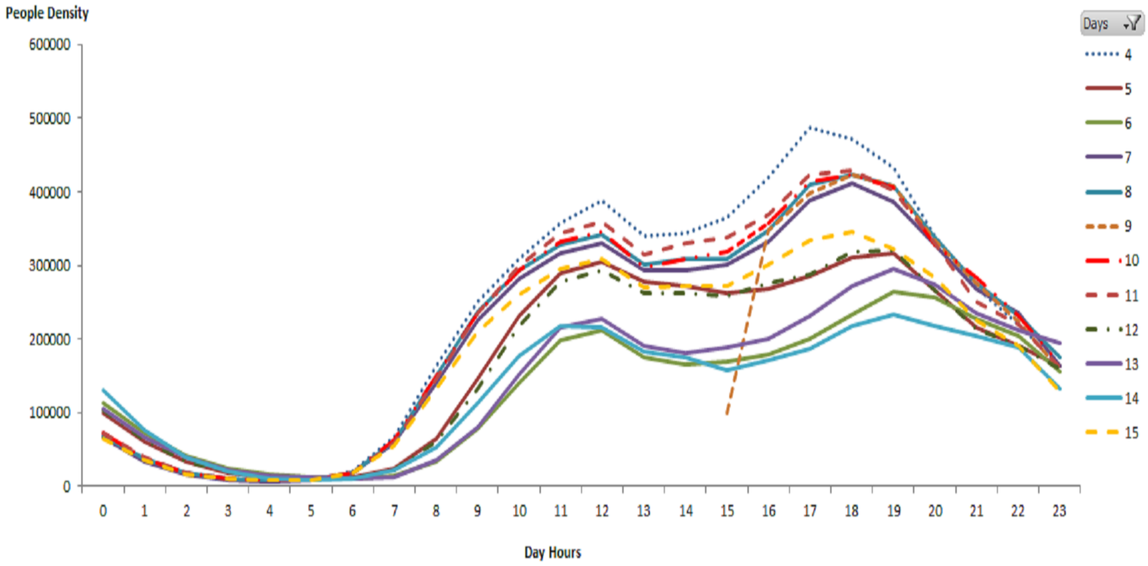


Figure 4.3 – Activity patterns of individuals densities according to days hours for all over the whole region.

The Area Sectors

This analysis phase is an attempt to investigate the life patterns of the observed data, with regard the alteration of the individuals densities within the sectors over different time intervals. However, that doesn't mean absolute indication to pure individuals activities, because they can be present in a location out of coverage area, so this probability makes individual exist actually, but doesn't necessarily appear on the CDRs. Nevertheless, mobile activities reflect the individuals density in somehow, also CDRs don't give indication to the path direction of individuals mobility. The drift of human activity in different sectors and times would be calculated using simulation platform.

The density analysis is performed along the daily hours, so it had been realized that all sectors have asymptotic activities patterns. However, the north sector is coming in the moderate rank of activities and patterns, which is close to the calm and stable sectors that are west and east sectors respectively, rather than active sectors are north and south sectors respectively. The analysis of daily activities ratios for the 5 sectors, according to the daily hours as in the following, and as shown in figure 4.4.

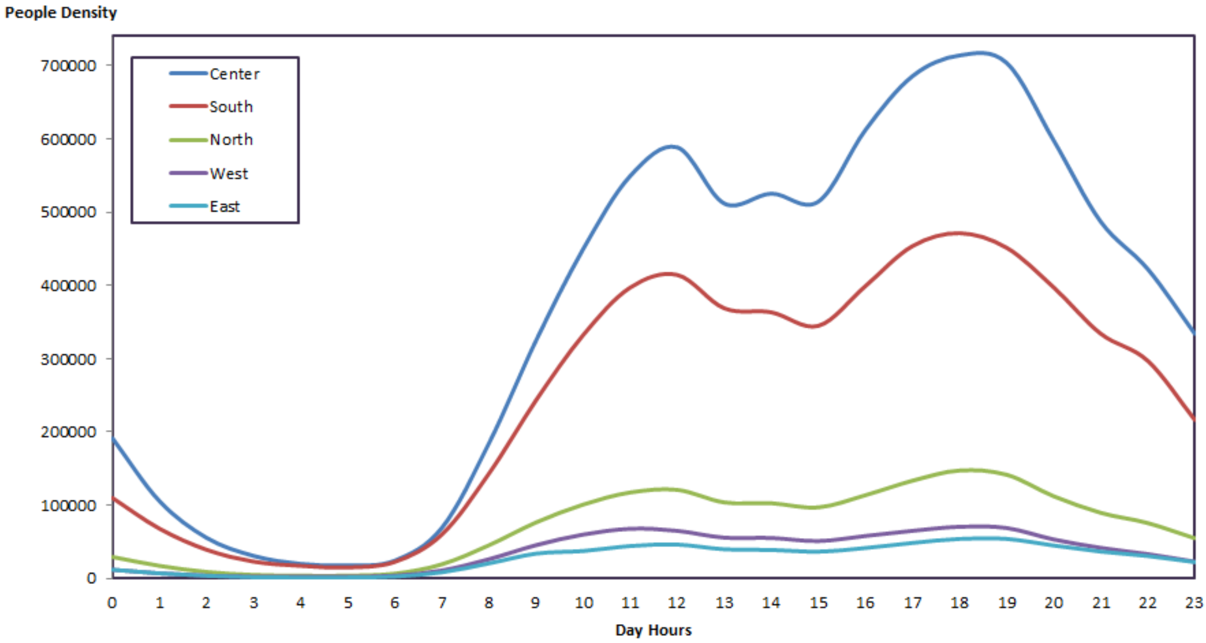


Figure 4.4 – Daily analysis of the 5 sectors along 24 hours with average over the days period of Armada.

1 All sectors have similar patterns, but in variant ratios. They could be ranked in descending order the center, south, north, west, and east.

2 Highest activities ratios for all sectors are along time intervals (00:00, 06:00-20:00), otherwise the activities ratios are descendin along time intervals (01:00-06:00) and (21:00-23:00).

3 The lowest activities ratios for all sectors are along time intervals (02:00-06:00, 15:00, 23:00).

4 All sectors have peak values in common 12:00 and 18:00, which are lunch hours and end work hours respectively, where the social activities are in their top rank.

5 The most active sectors are the center and the south along the observed period of 12 days.

6 There is high variance in activities ratios among sectors, so it could be ranked in descending order as the center then south as the most active sectors, followed by the north, west, and south sectors respectively.

The sectors activities varies along all observed days. Day 4 has the highest activities, the Armada influence is obvious here, since it's the former day to the Armada event, where arrangements are underway, and visitors start to arrive at the event place. Then the activities are smoothly descending in days 5 and 6 respectively, where they are off days/ weekend days. The days fall in period 7-11 record high activities, since they are work days, which reach the peak in day 11. However, the activities pattern is descending regularly in the days period 11-14, where the minimum descending is clear in day 14, which is the Armada event final day, therefore start to leave the place.Then area sectors are back to gradual ascending of activities in 15th July, which coincides with a working day. The analysis of daily activities ratios for the 5 sectors along the 12 days is shown in figure 4.5 and as follows:

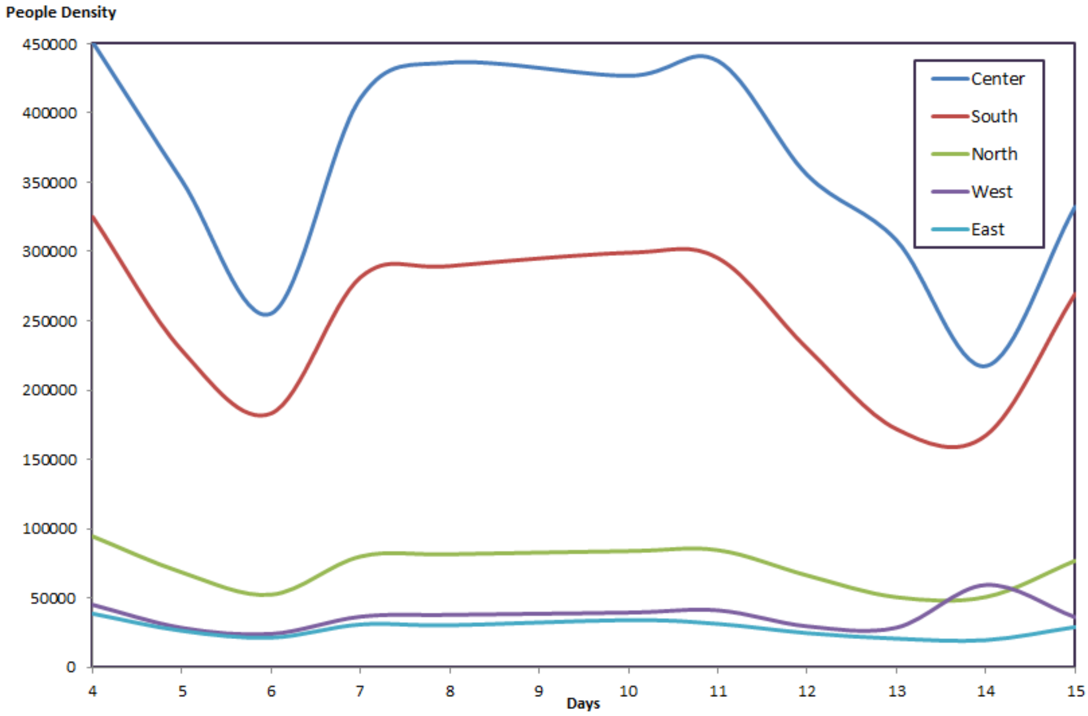


Figure 4.5 – Daily analysis of the 5 sector along the 12 days.

1 All sectors have high activities ratios in 4th July, since it is former to the Armada event, hence the influence of all arrangements is clear.

2 All sectors are decreasing in their activities ratios in off days (6, 12, and 13), that means the individuals are in their lowest activities during off days.

3 All sectors have their highly activities ratios in the work days interval (7-11).

4 The anomalous event is in day 14th, where all the sectors are in their lowest activities ratios except the west sector, where its activities ratios mark peak values, may be influenced by France National Day, where the ceremonial is performed in this sector.

The Center Sector

The center sector has activities densities indications according to the all observed period, as in figure 4.6 and the following:

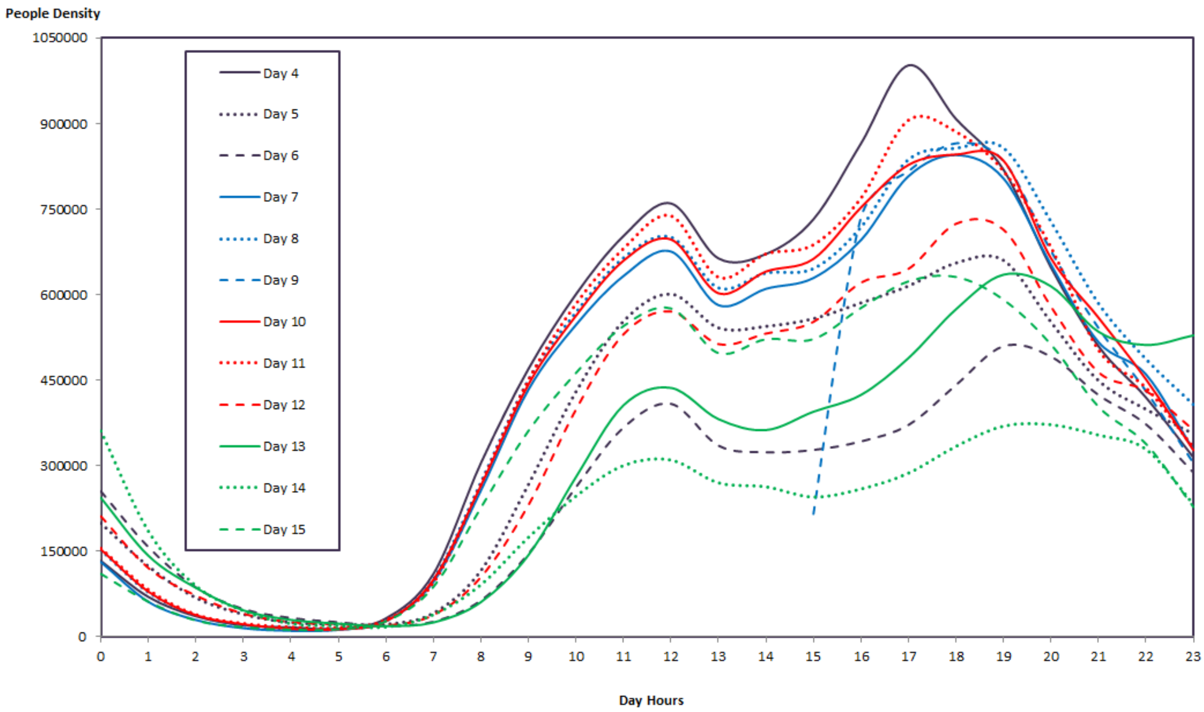


Figure 4.6 – The centre sector activity analysis according to the average of individuals density along daily hours.

1 The activities ratios are decreasing along time interval (0:00-06:00), where the lowest once are in 05:00.

2 The activities ratios are increased along time interval (06:00-20:00), and then decreasing gradually along time interval (20:00-23:00), where the peak activities ratios fall in the times (12:00, 7:00-18:00).

3 Day 4th July clearly has the highly activities ratios, the preparations for start and visitors arrival, then followed by days 11, 8, 10, 7, 5, 15, 13, 6, and 14 in descending order according to their activities ratios.

4 This sector has very similar patterns, sometimes interleaved patterns, however the lowest activities ratios were in day 14th July, which is France National Day (national holiday day).

5 This sector is active in almost days, since it is approximately the center of city in addition to Armada event occurrence.

The East Sector

The east sector has activities ratios indications according to the 12 days period, as shown in figure 4.7 and the following:

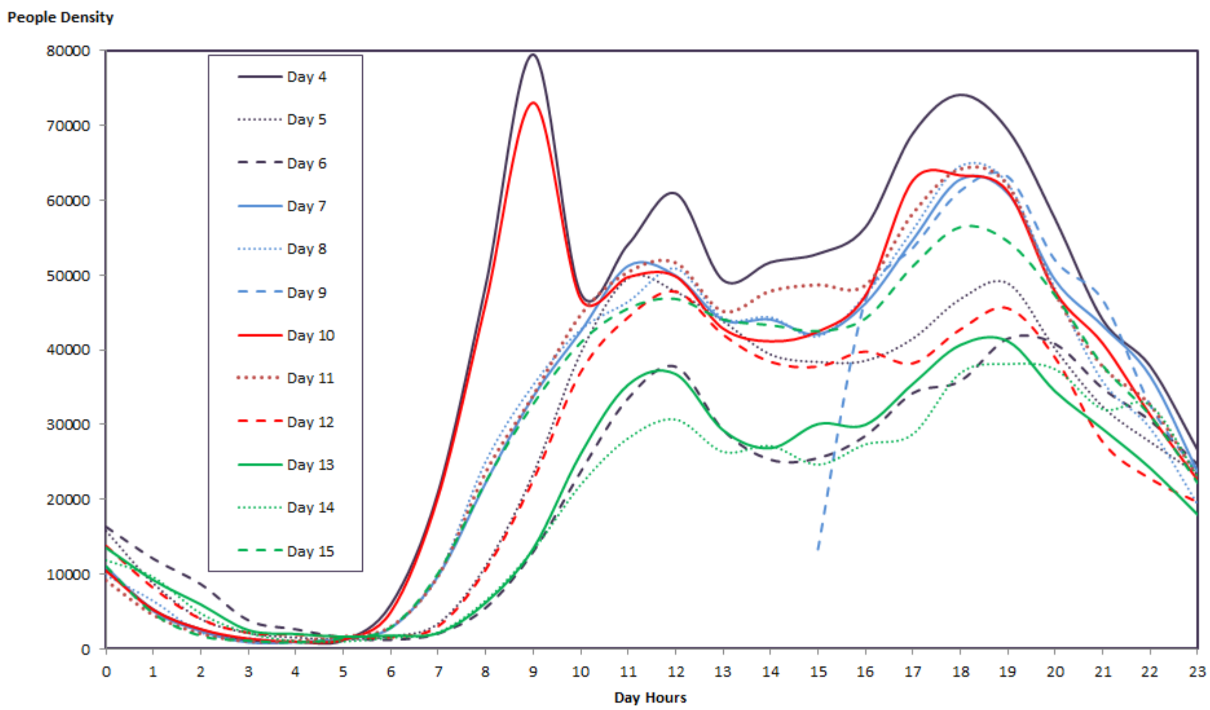


Figure 4.7 – The East sector analysis according to the average individuals densities along hours days.

1 The activities ratios are decreasing along time interval (0:00-06:00), where the lowest activities ratios along the time interval (04:00-05:00).

2 The activities ratios are increased along time interval (06:00-20:00), where the peak activities ratios along times (09:00, 12:00, 18:00), which are the times of work start, lunch time, and work end.

3 Days 4 and 10 have anomalous activities ratios, where the peak is in hour 09:00, these events may be due to ships arrival, so attract the individuals to see them. However, day 4 still has the highest activities.

4 The days activities have very similar patterns, except the days 4 and 10.

5 The activities ratios in almost days could be ranked in descending order as follows 4, 10, 11, 8, 7, 15, 12, 13, 6, and 14, where most of them are off days, so it could be an entertainment or commercial region.

6 Day 9 has lost data, but according to the provided interval (15:00-23:00), it is close and interleaved with other work days.

7 Day 14 has the lowest activities ratios, because it is France National Day as mentioned above.

The West Sector

The west sector has activities densities indications according to the 12 days period, as in figure 4.8 and the following:

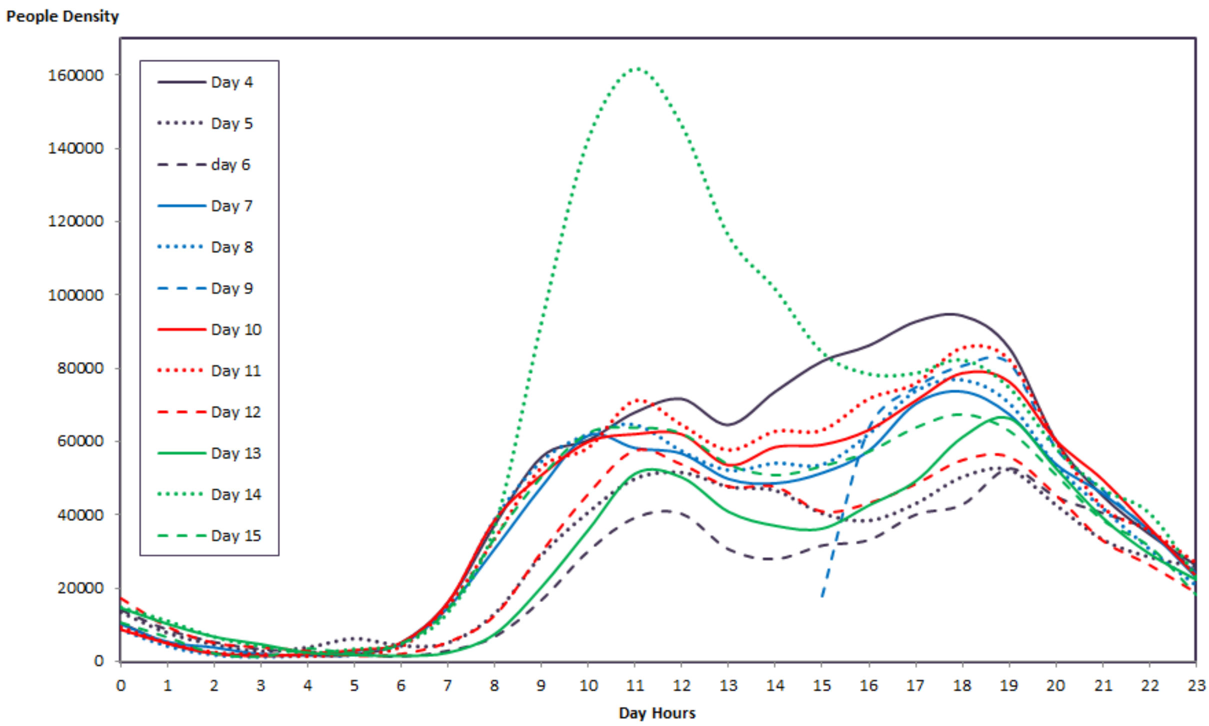


Figure 4.8 – West sector activity analysis, according to average of people density along day hours.

1 This sector marks different patterns with highest activities ratios in day 14, which is the most influenced sector by National France Day and Armada effects, the activities ratios increasing along time interval (08:00-15:00) with peak hours at time interval (00:11-19:00).

2 All days activities ratios are decreasing along time interval (00:00-07:00), then they are increasing along time interval (08:00-19:00).

3 Days activities ratios are arranged in descending order as 14, 4, 11, 10, 8, 15, 12, 5, 13, and 6, where almost days in this sector have an approximated ratios activities except day 14.

The North Sector

The north sector has activities densities indications according to the 12 days period, as in figure 4.9 and the following:

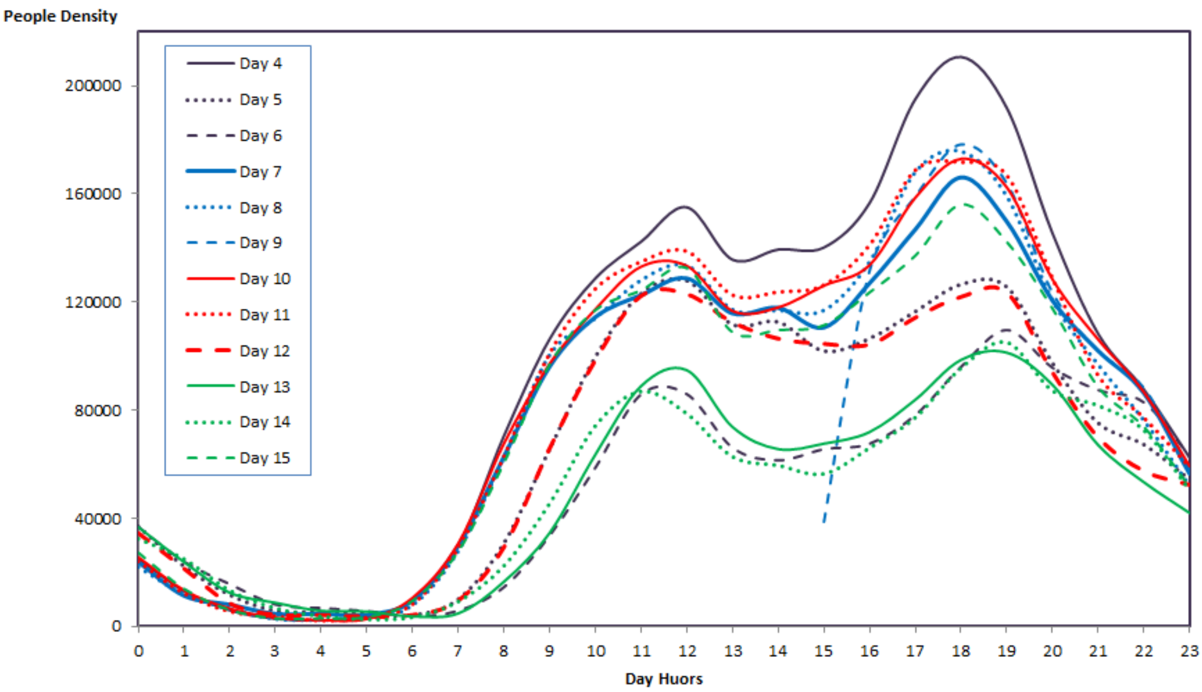


Figure 4.9 – The north sector activities analysis according to the average of individuals densities along day hours.

1 This sector has similar patterns, with the highest activities ratios at day 4.

2 All days of this sector have similar patterns to previous sectors, where activities ratios are decreasing along time interval (00:00-06:00), where the lowest ratios along time interval (04:00-05:00).

3 The activities ratios are increasing over time interval (06:00-00:19), with peak hours (12:00 and 18:00), then the activities are decreasing along time interval (20:00-23:00).

4 The most active days in this sector could be ranked according to their activities ratios in descending order as 11, 10, 8, 9, 15, and 12. All these days are work days except day 12 is off day, whereas 14, 6, and 13 are off days.

The South Sector

The south sector has activities densities indications according to the 12 days period, as in figure 4.10 and the following:

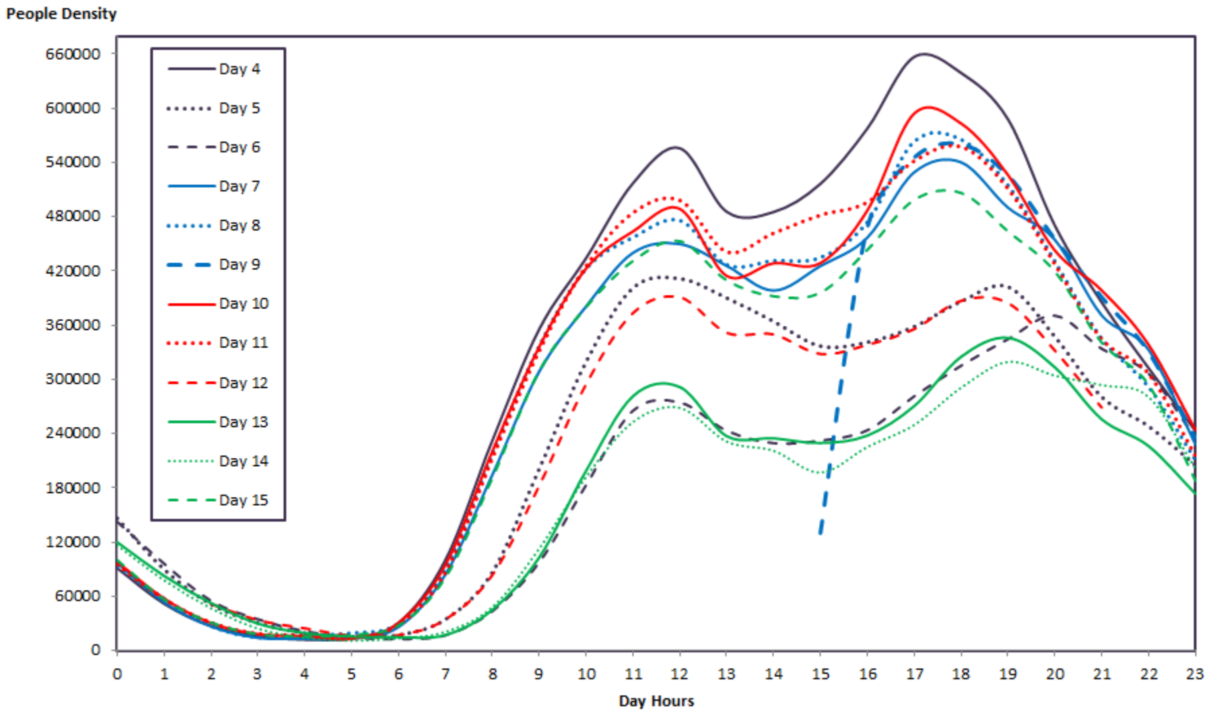


Figure 4.10 – The south sector activities analysis according to the average of individuals densities along day hours.

1 All days on this sector have similar patterns with the most highly activities ratios in day 4.

2 The activities ratios are decreasing along time interval (00:00-06:00).

3 The activities ratios are decreasing along interval time (06:00-20:00), where peak hours (12:00 and 18:00), which are lunch time and after work time, respectively.

4 The activities ratios are decreasing along time interval (20:00-23:00).

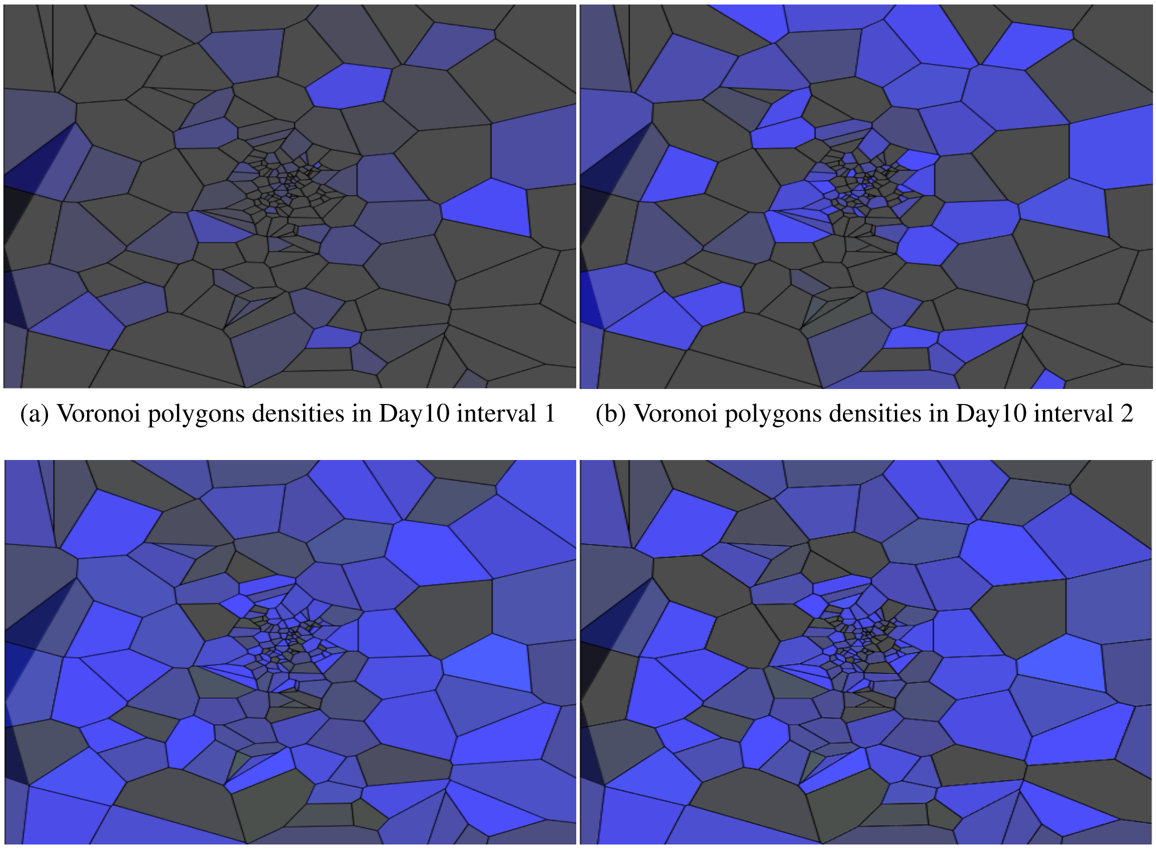
5 The most active days in this sector could be ranked according to their activities ratios, in descending order as 4, 11, 10, 8, 7, 15, 5, and 12. The days 5 and 12 are off days, but they have remarkable activities ratios, where the lowest ratios are in the off days 13, 6, and 14 days.

Note: day 9 has an average activities ratios, in spite of it has loss of data over time interval (00:00-15:00) on all over sectors.

Individuals Densities over Voronoi Diagrams

The densities in each Voronoi polygon could be represented [Ale14] according to 6 intervals along day, hence any anomalous event could appear within its specified region and time interval. The data are divided into six basic time intervals: P1(2:00-6:00), P2 (6:00-10:00), P3(10:00-14:00), P4(14:00-18:00), P5(18:00-22:00), P6(22:00-23:00) and (0:00-2:00). This division is performed according to the general people life patterns intervals as sleep, wake up, go to work, end of work, and entertainment hours. The P1 and P2 intervals have the lowest activities densities, it is normal because they are the the sleeping interval in the city, whereas other intervals have the most active and varied activities densities. This variance gives a view perspective about the wake up and sleep periods for the observed area.

This representation is performed using Gama-platform [Pat12a]. Only one day has been analyzed due to Gama limited resources compared with the observed data size. The densities are computed and represented on each Voronoi polygon for all 6 intervals of day 10, the polygons are colored to represent there densities independently, the blue color and its shades represent the high activities, and the gray color represents the low activities as shown in figures 4.11 and 4.12 respectively.



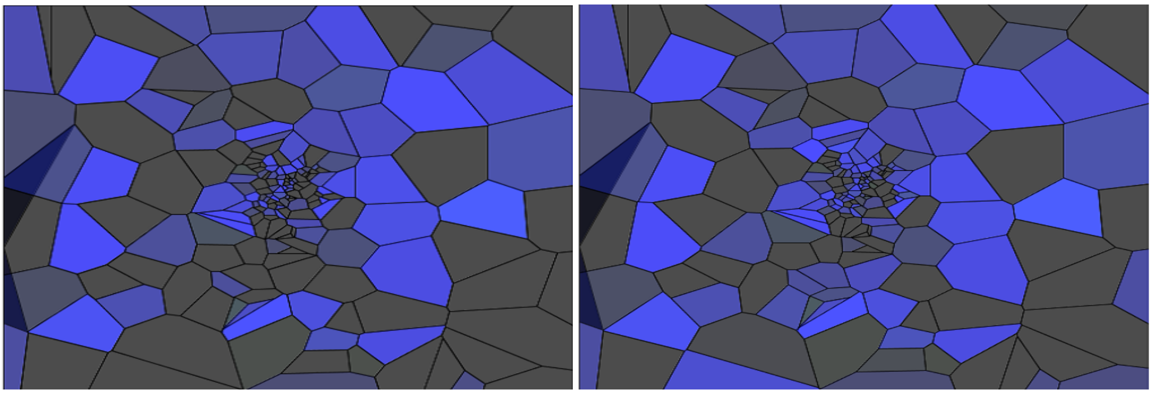
(a) Voronoi polygons densities in Day10 interval 1

(b) Voronoi polygons densities in Day10 interval 2

(c) Voronoi polygons densities in Day10 interval 3

(d) Voronoi polygons densities in Day10 interval 4

Figure 4.11 – Voronoi polygons densities of day 10 intervals: P1, P2, P3, and P4. The blue color and its shades represent the high acitivities, and the gray color represents the low activities



(a) Voronoi polygons densities in Day10 interval 5 (b) Voronoi polygons densities in Day10 interval 6

Figure 4.12 – Voronoi polygons densities of day 10 intervals: P5 and P6. The blue color and its shades represent the high acitivities, and the gray color represents the low activities

Conclusions

The spatio-temporal datasets, that are acquired from mobile phone data have huge size due to the millions of subscribers, and it holds many events either anomalous or normal events [Lin12, Luc13]. The normal events give a good reflection to the daily life pattern in the observed area [Mic11, Rob15, Tho13].

The population urban analysis is a wide range area has a multi-disciplinary subject to obtain density [Dar13], community detection, and crowd movements, spatial coordination. It is a major improvement produced by the new technologies, since it guides to a lot of sciences like: cities planning, smart cities, decision making, facing disasters, social networks analysis, diseases spreading [Mic], ecosystems evolution, human attitudes, human body inter-relations, networks understanding, society cultural analysis... etc. So, the need to deal with both GIS and huge data are emerged depending on the type of results that are needed, which have the type of socio economic [V. 11], land use management [Jam12], urban planning or social networks analysis [Fos13], with respect to the complexity property, which is the most effective face in these networks [Raw09a, Arm10, Joh15]. Additionally, the anomalous events data are very useful in detecting the disasters or catastrophes immediately (in real time) with the spatial and temporal indications, so the evacuation, emergency reactions could be taken as soon as possible according to any alarm.

The most remarkable result of the mobile phone database that we have used, is that all sectors all over the observed period have peaks ranging between 10:00 Am-20:00 Pm, in fact this interval is the Armada festival activities, and it is seen that the sectors could be arranged in descent order according to their activities densities as center, south, north, west, and east, as well as the most remarkable thing is the rank order related to the Seine river, as we move away from the river the density of activities are descending.

Additionally, the GSM data have in general good synchronization, and then network analysis techniques could be implemented on it [Lin12, Tho11b]. Therefore, individuals density and dens variance between several areas could be estimated in order to explore the tendency or the orientation at these areas.

Chapter 5

Simulation and Dynamics Reconstruction

The chapter has been partially published in [Suh16d, Suh16f, Suh16e, Suh16b, Suh16c, Suh16a]

Introduction

The digital environment development makes the scientists able to acquire more scientific, and heterogeneous details about any case study (system model) [Mic12]. Understanding, modeling, simulating the pedestrian behavior and movements (individuals mobility) among urban regions is very challenging effort, since it is very important in rescue situations for many kinds of events, either in the indoor events like evacuation of buildings, stadiums, theaters, ships, aircraft or outdoor ones like public assemblies, open concerts, religious gatherings, community evacuation ...etc. [Nir08, Jam09].

The simulation is preferred to analyze any system instead of working on it in real life, since it is easier, safer, cheaper to make all the experiments on the simulated model. The most important and difficult issue in modeling and simulating any system is the determination of the probability distribution and parameters, and hence to model the uncertainty of the system input variables. However, the purpose of simulation analysis is to acquire and analyze the results in well conceptual vision. In order to give high indications for decision makers the simulation is pivoting on two events kinds, the discrete events, and the frequent (continuous) events [Sou14, Vin15].

As well as, the system data analysis could be performed in two levels [Geo12, Hei12]:

1 Individual level: The data of space, time or both of them are point to the individual items.

2 Collective (group) level: The data of aggregated space or aggregated time intervals or both of them are point to the aggregated data items [Mar14b].

Additionally, from the system theory, it could be concluded that any system (ecological, physical, informational, or social) has items, and could be considered as a population of agents (cities, humans ...etc.). They are interacted and affected each other, then affect the whole system consequently, and consist the system environment. They could affect the outer world or not, but the internal processes and relations inside the system certainly are affected by the outside world (feedback). All these aspects cause the system evolution during system lifetime at a specific spatial in the world. However, the models complexity that are produced by the interactive processes inside /outside the system environment, is ranging between the aggregated level and the disaggregated level, which gives more effective description for the spatio-temporal properties [Mic12].

This chapter is organized as in the following sections: The first section is the data set, which explores the observed data in details, and presents the lack of information in this data. And, it propose a compensation and estimation issues in order to solve the data incompleteness. As well as, it presents the practical and technical limitations.

The next section is the describing individuals activities (Mesoscopic Perspective), it elaborates the statistical techniques. And, the probability of inter-event time (ΔT) is computed. Then, the fixed inter-event time observations section explores the main parameters used in this thesis, which are: Δt, the travel distance Δr, and the radius of gyration r_g. As well as, presents the results and its algorithm to compute the probability distributions of activities and displacements by inter-event times, and as in our contributions [Suh16c, Suh16f, Suh16b].

The trajectories in intrinsic reference frame section explores the statistical models that are used to reconstruct the human trajectories. This section is branched into three subsections as: individual trajectory characteristics, trajectories within radius of gyration, and real and simulated individual trajectory estimation (Microscopic Perspective). The first subsection presents the general physical characteristics. Also, it elaborates the results and algorithm of computing the radius of gyration for the observed data, as in our contribution [Suh16b, Suh16d, Suh16e]. Then, the second subsection explores the probability distribution and results of the trajectories within radius of gyration, and as in our contribution [Suh16a]. The third subsection explores the estimated individual trajectory in the Microscopic level, and presents its algorithm and the produced results.

Then, the output section presents the modeling and simulation results of day 4 data according to the previous mentioned parameters. Finally, the conclusions section endorses that the used parameters are capable of investigating the human mobility patterns.

Data Set

The investigated data base of the Armada case study is composed of 51,958,652 CDRs (entry records) for 615,712 subscribers. It contains individuals occurrence in discrete (irrelevant) mode only, means that any mobile individuals activity is recorded at start/end time, but there is a lack (lost information), since it didn't indicate the individuals occurrence during inactive case (mobility without any mobile phone activity). There are no data, meanwhile the mobile phone is idle, i.e. inactive or doesn't make any communication activities neither calls nor SMS activities. As well as, the available spatial data are only the towers (X, Y) coordinates, hence it would be considered to estimate individual's transitions from position to another (from tower to tower), hence the positions would be determined approximately with regard to the tower coverage area and its signal strength.

Also, the basic issue, which is used to compute individual speed wouldn't be accurate, but rather it is estimated, since it uses the distance between each two successive transition locations, that are recorded on CDRs. Then, the traveled distance would be divided by the time difference of each transition, to acquire the speed of individual's mobility, as in equation (5.1).

$$EstimatedSpeed = \frac{TraveledDistance}{ElapsedTime} \tag{5.1}$$

The Traveled Distance is the displacement between two successive locations in the Elapsed Time, which the time required to move from location to another. Because of the non-deterministic and discrete nature of these data, and each individual could be disappeared for a while from the DB records, which makes individual tracing is unworthy, without significant indications on the individuals mobility in the city. Therefore the collective behavior would be the effective approach to analyze and simulate them [V. 11].

Armada DB is a huge one, so it was difficult to manipulate the total observed data in comparing with the limited laboratory resources, where each day has large records ranging between 2,715,077-5,661,428 records as shown in table 5.1.

Table 5.1 – Armada DB records are classified according to days.

Day4-15	records
Day 4	5,661,428
Day 5	4,223,202
Day 6	3,257,343
Day 7	5,007,466
Day 8	5,206,989
Day 9	2,715,077
Day10	5,227,176
Day11	5,291,250
Day12	4,155,966
Day13	3,531,929
Day14	3,221,719
Day15	4,459,119

Practical and Technical Limitations This research faced limitations and technical complexity as in the following:

1 The enormous data size enforces the use of a remote server (virtual machine), which needs RAM (30 GB) and processors (12 CPU), which represents the half resources of the LITIS laboratory, even though it is not efficient enough.

2 Generating shape files via ArcGIS is halted when trying to convert big .csv files to .shp (shape) files.

3 Matlab platform is used to manipulate data of each day independently, hence the manipulation ranges between 21-45 hours for each day. However the numerous records enforce to make some additional changes on Matlab environment, by using the '-v7.3' flag to save MAT files over that 2GB threshold. As well as, coding improvements take an important role to deal with this kind of treatment.

Describing Individuals Activities (Mesoscopic Perspective)

The mobile communications data have heterogeneous nature, since the individuals are varied in their usage of the mobile phone, they are ranged between (rarely-frequently) usages during observed period. The individuals are grouped according to their total activities (activities in the following refer to the mobile phone events). The probability of waiting time (inter-event time ΔT) of each consecutive activities has been computed for each individual, so the individuals are grouped with regard to their activities. Figure 5.1 shows that the long waiting times are characterizing the individuals of less activities. The distribution of the average inert-event time ΔT is estimated by equation (5.2).

$$P(\Delta T) = (\Delta T)exp^{-\Delta T} \tag{5.2}$$

Whereas, the power-law exponent ΔT is the average waiting time of the total population. In order to formulate simplified models with their quantitative analysis parameters, hence computing the probability function to get the system universal behavior (pattern) according to consecutive inter-event time [Siq11, CS315]. The individuals (population) probabilities computations are performed using the statistical techniques in following [Siq11, Ily13]:

1 Compute probability function to get the system universal behavior (pattern), this is done according to consecutive inter-event time.

2 Compute probability density function (PDF) to get each agents sample probability [Sta15b].

Fixed Inter-Event Time Observations

The modeling and simulation process starts by computing the probability model, which is modeled the data and simulate the responses of the model by using one of the well-known functions, it is the PDF [Sou14, Vin15, Mar09].

The main used parameters in this research are the inter-event time Δt, the travel distance Δr [Siq11], and the radius of gyration r_g. The samples means of the distributions are drawn from an exponential distribution. Then, compute the mean of all means. However, understand the behavior of sample means from the exponential distribution, in order to have the universal system pattern (general population law).

The computations are performed as in the following order and considerations:

1 Manipulate all the 12 observed days data (each day independently) for all individuals (population) in spatio-temporal mode.

2 Eliminate the individuals of only one occurrence in the CDRs, since they didn't have a significant indication on the mobility.

3 Sort the data by time, in order to have the real sequence of positions transitions for the individuals trajectories.

4 Classify the data according to individuals activities (sampling)[Cor13], this done by computing the inter-event time Δt, which is the time elapsed between each successive activities of each individual. It is ranged between $15-1440$ $minutes$, this assumption is pivoted on the logical intuition, since the 15 $minutes$ is the minimum time that can give mobility indication, and the 1440 $minutes$ (24 hours) is considered as the highest elapsed time to travel inside the city.

5 Compute ΔT_a, which is the average inter-event time of all individuals (population).

6 Classify the minimun and maximum samples of the activities scores (activities densities).

7 Sort the samples of activities according to the inter-event time.

8 Compute the inter-event time of all individuals Δt and ΔT_a for each sample, where $\triangle t/\triangle T_a$ is the average inter-event time of the total population.

9 Compute exponential distribution probability for the total data, hence identify the universal population (general) pattern law.

The exponential distribution (probability distribution) as in equation (5.3) is capable of modeling the events happened randomly over time. In this case, it is able to describe the inter- event time and the average of inter-event time of individuals activities using the Probability Density Function PDF [CS315, Sta15a, Sou14, Mar09]. The cutoff distribution is defined by the maximum observed inter-event time, at which the individual can wait to make any mobile activities, where it is $\triangle t = 1431$ minutes.

$$f(x \mid \mu) = \frac{1}{\mu} e^{\frac{-x}{\mu}} \tag{5.3}$$

It is estimated according to algorithm 1 of the complexity $O(n^3 + 2n^2 + n)$, which is implemented with Matlab platform. Where the data matrix composed from 4 columns corresponds to time (TimeCol), geographic coordinates (CoorCol_x and CoorCol_y), and the individual ID (PersonCol) respectively. The results are explored as in the figures 5.1, 5.2, and 5.3 for the total, work days, and off days periods respectively.

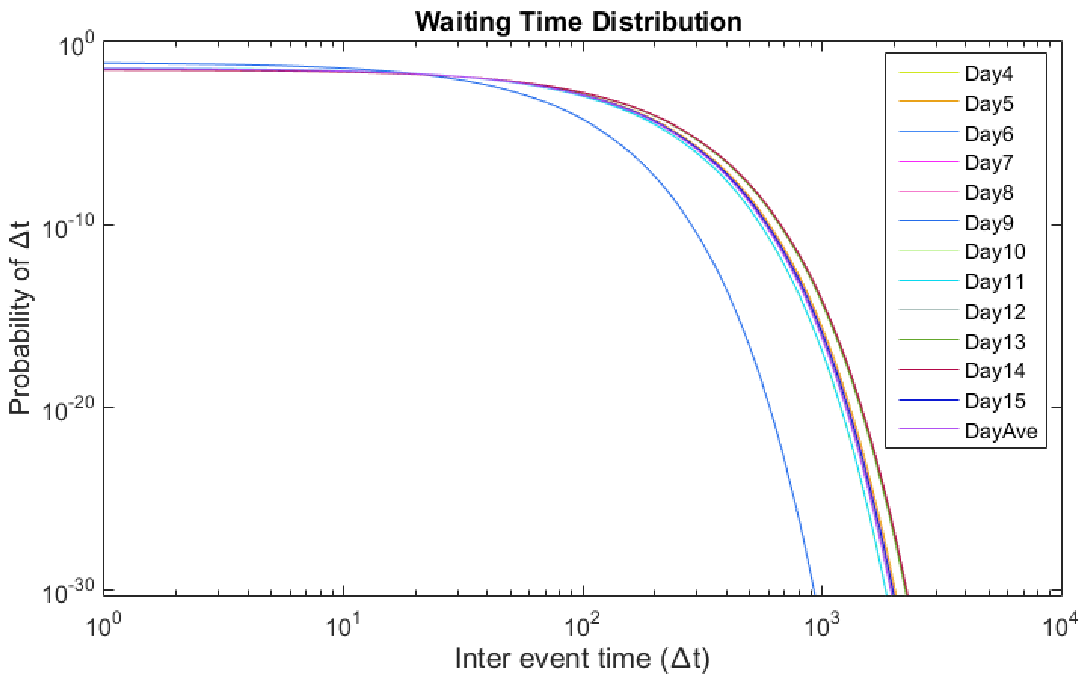


Figure 5.1 – Waiting time distribution $p(\triangle T)$ of mobile activities, where $\triangle T$ is the spent time between each two successive activities. The legend symbols are used to distinguish the individuals groups according to their activities ratio, and the curve of their mean $(Day_m oy)$ for the total population during $total$ period.

Algorithm 1: Probability distributions of activities and displacements by inter-event times _Part 1

Data: File of events sorted by time
Result: Activities and displacement distributions in inter-event time

```
data ← (Rows, TimeCol);
data ← (Rows, CoorCol_x);
data ← (Rows, CoorCol_y);
data ← (Rows, PersonCol);
UnrepeatedPersons ← unique[data(PersonCol)];   // The unique function
    returns data with no repetition based on PersonCol
L ← size(UnrepeatedPersons); Val ← data(1, PersonCol);
debi ← 1; Per ← 1;
while Per < L do
    ExactPerson ← data(debi, PersonCol);
    ExactPersonRepition ← size(find(data(:, PersonCol) == ExactPerson));
    while i<= debi do
        idi = data(i, PersonCol);
        debj ← i + 1;
        for j<=debj to size(data) do
            idj ← data(j, PersonCol);
            if idi=idj then
                index ← (index + 1);
                CptVector ← CptVector + 1;
                Δt(index) ← (data(j, TimeCol) − data(i, TimeCol));
                P1 ← (data(j, CoorCol_x) − data(i, CoorCol_x)^2;
                P2 ← (data(j, CoorCol_y) − data(i, CoorCol_y)^2;
                Δr(index) ← sqrt(P1 + P2);
                D(CptVector, 1) ← (△r(index));
                D(CptVector, 2) ← (△t(index);
                debi ← j;
                i ← debi;
            end
        end
        break;
    end
    if (index+1)= ExactPersonRepition then
        for k=1 to size(data) do
            if data(k,PersonCol)=Val then
                data(k, PersonCol) ← 0 ;          // Tagging treated data
            end
        end
    end
    Per ← (Per + 1);
    S(Per, 1) ← (∑_{i=1}^{size(Δt)} (Δt)/size(Δt));
    S(Per, 2) ← size(Δt);
    for n=1 to size(data) do
        if data(n, PersonCol) ≠ 0 then
            debi ← n;
            Val ← data(n, PersonCol);
        end
    end
end
```

Algorithm 1: Probability distributions of activities and displacements by inter-event times _Part 2

```
/* Remove the only one occurrence in data, where the ~isnan function
   selects just non zero values                                      */
for i=1 to size(S(:, 1)) do
    if (~ isnan(S(i,1))) then
        T(i,1) = S(i,1);
        T(i,2) = S(i,2);
    end
end
S = T ;
c1 ← zeros ; c2 ← zeros;
c3 ← zeros; c4 ← zeros;
c5 ← zeros; c6 ← zeros;
r1 ← zeros; r2 ← zeros;
r3 ← zeros; r4 ← zeros;
r5 ← zeros; r6 ← zeros;
r7 ← zeros;
/* Classify S into activities samples according to Δt              */
for i=1 to size(S) do
    if S(i, 2) > 5 then
        c1(i) = S(i,1);
    end
    if 50 > S(i, 2) ≥ 5 then
        c2(i) = S(i,1);
    end
    if 100 > S(i, 2) ≥ 50 then
        c3(i) = S(i,1);
    end
    if 500 > S(i, 2) ≥ 100 then
        c4(i) = S(i,1);
    end
    if 1000 > S(i, 2) ≥ 500 then
        c5(i) = S(i,1);
    end
    if S(i, 2) ≥ 1000 then
        c6(i) = S(i,1);
    end
end
```

Algorithm 1: Probability distributions of activities and displacements by inter-event times _Part 3

```
/* Classify D into samples according to △r during time intervals */
```
for *i=1 to size(D)* **do**
  if *21* $\leq D(i,2) \geq 19$ **then**
    $r1(i) = D(i,1)$;
  end
  if *41* $\leq D(i,2) \geq 39$ **then**
    $r2(i) = D(i,1)$;
  end
  if *59* $\leq D(i,2) \geq 61$ **then**
    $r3(i) = D(i,1)$;
  end
  if *119* $\leq D(i,2) \geq 121$ **then**
    $r4(i) = D(i,1)$;
  end
  if *239* $\leq D(i,2) \geq 241$ **then**
    $r5(i) = D(i,1)$;
  end
  if *479* $\leq D(i,2) \geq 481$ **then**
    $r6(i) = D(i,1)$;
  end
  if *1339* $\leq D(i,2) \geq 1341$ **then**
    $r7(i) = D(i,1)$;
  end
end
```
/* Compute and draw the Exponential probability distribution using
   makedit and expfit matlab functions */
```
$PopT \leftarrow S(:,1)./mean(S(:,1))$;
$PopR \leftarrow D(:,1)./mean(D(:,1))$;
$p(\triangle t) \leftarrow makedit(expfit(PopT))$;
$p(\triangle r) \leftarrow makedit(expfit(PopR))$;
Draw $p(\triangle t)$ distribution of activities using $c1, c2, c3, c4, c5, c6$ Samples;
Draw $p(\triangle r)$ distribution of traveled distances using $r1, r2, r3, r4, r5, r6, r7$ Samples;

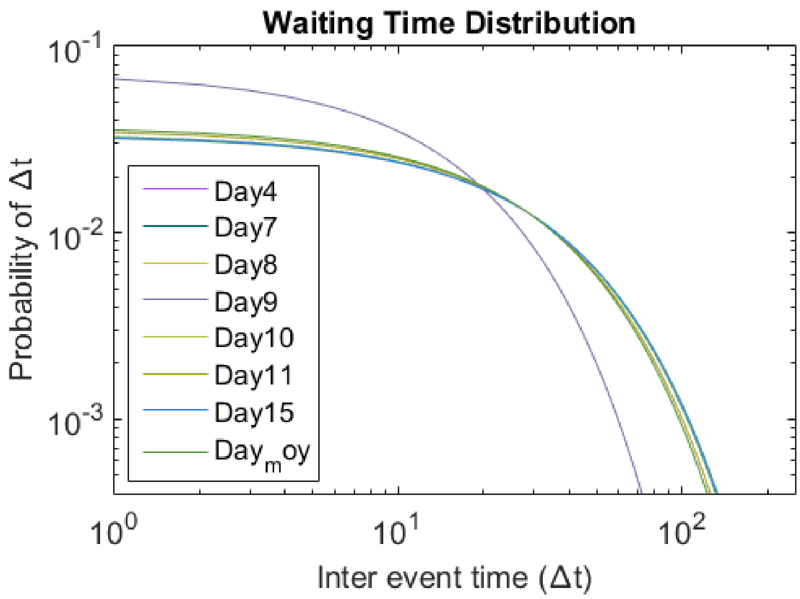


Figure 5.2 – Waiting time distribution $p(\triangle T)$ of mobile activities, where $\triangle T$ is the spent time between each two successive activities. The legend symbols are used to distinguish the individuals groups according to their activities ratio, and the curve of their mean (Day_moy) for the total population during $work$ days period.

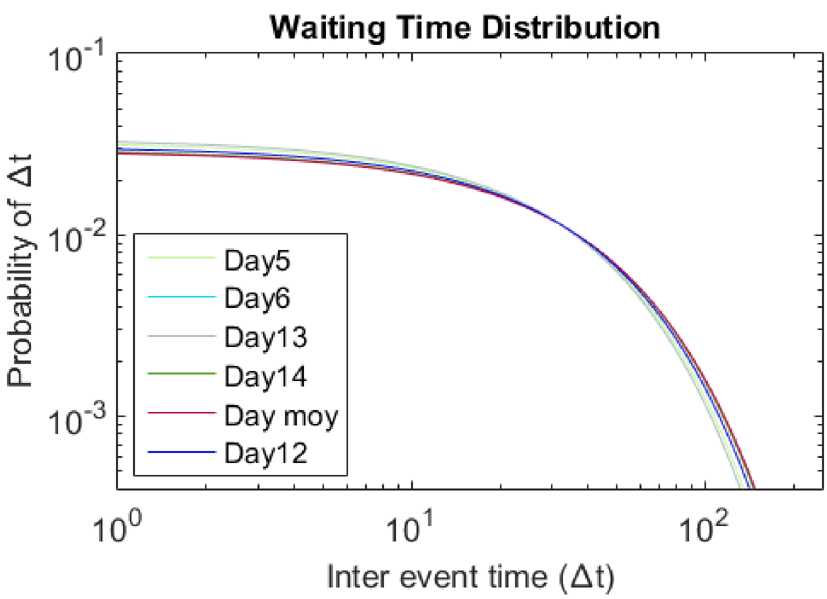


Figure 5.3 – Waiting time distribution $p(\triangle T)$ of mobile activities, where $\triangle T$ is the spent time between each two successive activities. The legend symbols are used to distinguish the individuals groups according to their activities ratio, and the curve of their mean (Day_moy) for the total population during off days period.

The displacement distribution $p(\triangle r)$ [Ale14] has been computed for the total individuals (population) during the observed period. As in figures 5.4, 5.5, and 5.6 for the total, work days, and off days period respectively. However, the $\triangle r$ is the covered distance between each two successive activities during time

Hence, the cutoff distribution is determined by the maximum observed distance at which the individual can travel, where it is $\triangle r = 7.32 \times 10^4\ m$ along day hours, since the maximum time slice couldn't exceed $24\ hours$ with regard to the observed region. The consequence is the $p(\triangle r)$ distributions for different $\triangle t$ follows truncated power law, as in equation (5.4).

$$P(\triangle r) = (\triangle r)exp^{-(\triangle r)} \tag{5.4}$$

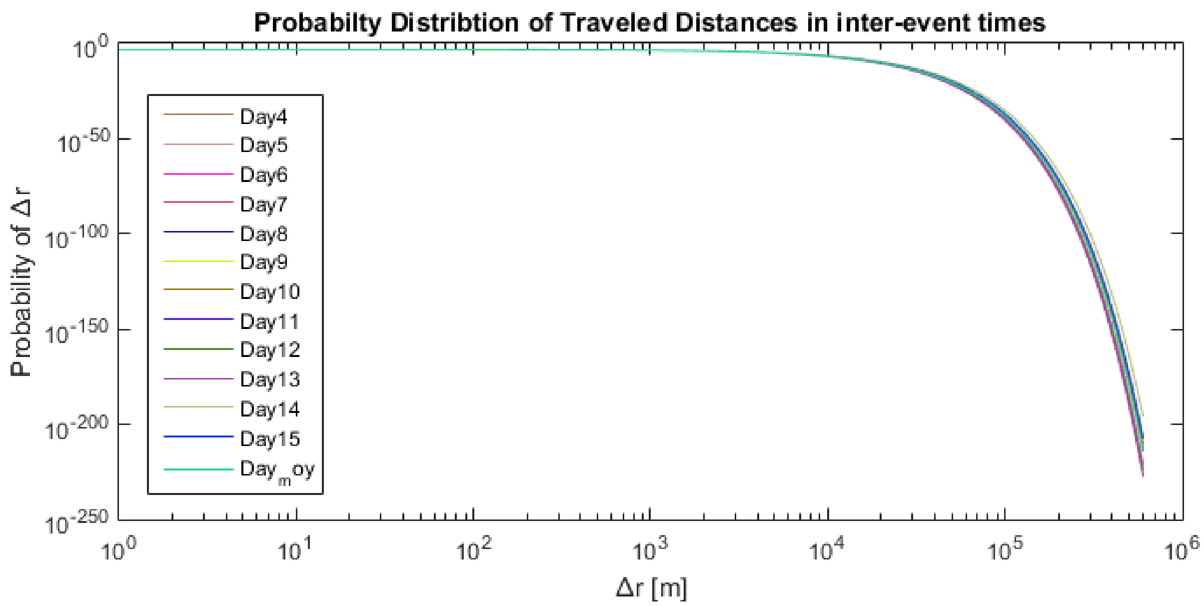


Figure 5.4 – Distance (displacement) distribution $p(\triangle r)$ for waiting times (Inter-event times). The cutoff distribution is determined by the maximum distance traveled by individuals for specific inter-event-times during the *total* period.

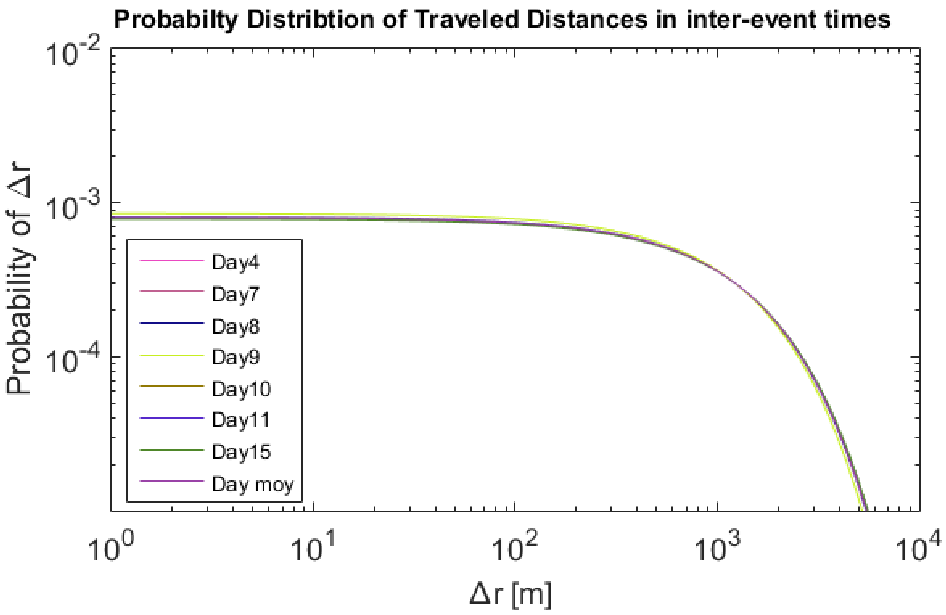


Figure 5.5 – Distance (displacement) distribution $p(\triangle r)$ for waiting times (Inter-event times). The cutoff distribution is determined by the maximum distance traveled by individuals for specific inter-event-times during $work$ days period.

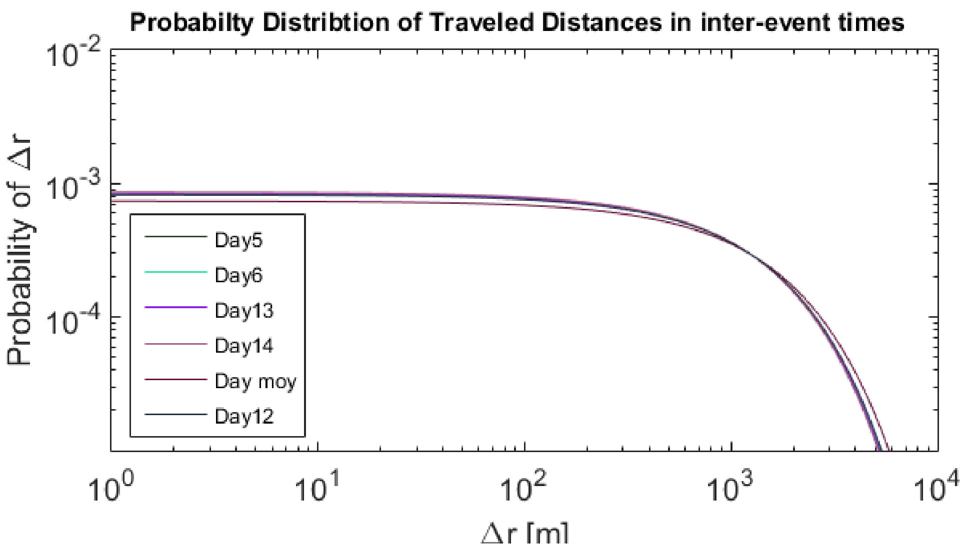


Figure 5.6 – Distance (displacement) distribution $p(\triangle r)$ for waiting times (Inter-event times). The cutoff distribution is determined by the maximum distance traveled by individuals for specific inter-event-times during off days period.

After all, the total days period obtained averages of the waiting time distribution are $MAX\Delta t = 1431\, minute$, and $MIN\Delta t = 0\, minute$, whereas of the displacement distribution are $MAX\Delta r = 7.229515e + 04\, meter$, and $MIN\Delta r = 0\, meter$.

Trajectories in Intrinsic Reference Frame

The significant importance of revealing human trajectories enforces the tendency to build the statistical models. The human trajectories have random statistical patterns, hence the tracing of human daily activities is the most challenging issue [Mar09]. In spite of data sources variance (billing system, GSM, GPS), but the common characteristics are the aggregated jump size Δr, and waiting time Δt distributions [Fos13]. Where, the Δr gives an indication on the covered distances by an individual during Δt for each two consecutive activities, and the Δt is the time spent by an individual between each two consecutive activities [Cha10a]. The individual trajectory (mobility) is considered as the microscopic level of mobility abstraction, which is constituted of sequenced coordinates positions a long time i.e. the agent displacement in spatio-temporal unit, as shown in figure 5.7 [Cor13, Chr13b, Xia10, Zah15].

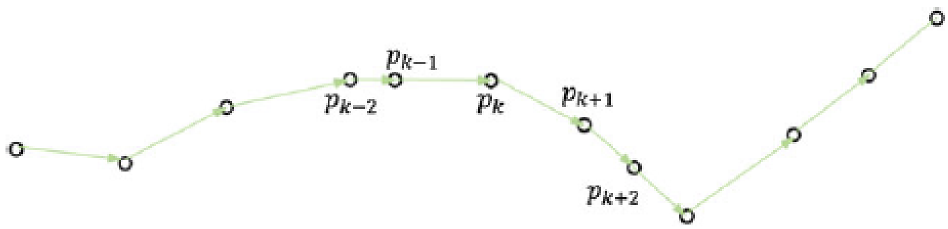


Figure 5.7 – The individual trajectory composed of consecutive coordinate points [Cor13].

Individual Trajectory Characteristics

The mobile phone data are very heterogeneous, since the mobile users could be very active (have so many calls/SMSs), or inactive (have few calls/SMSs), so sampling the individuals can be pivoted on their activities number [Mar08].

The sparseness of individual activities causes incomplete spatial information, therefore the dynamic individual has some general physical characteristics, that can be used to compensate the lack of data (when no mobile activity recorded) [Pu 09b]. They are useful to build a mathematical model of human mobility patterns, in order to verify the dynamic behavior and life patterns using the most common mobility characteristics [Mar08]. The general physical characteristics as follows [Far13, Chr13b, Mar09]:

1 Center of mass c_m: It's the most visited positions by individual, as in equations (5.5) and (5.6).

$$x_{cm} = \sum_{i=1}^{n} x_i/n \tag{5.5}$$

$$y_{cm} = \sum_{i=1}^{n} y_i/n \tag{5.6}$$

Where x_i and y_i are the coordinates of the spatial positions, n is the number of spatial positions, that are recorded in the CDRs.

2 Radius of Gyration r_g: It is the average of all individual positions, it determines visited area by the individual (the traveled distances during a time period). In this research the distribution of r_g uncovers the population heterogeneity, where individuals traveled in $p(r_g)$ in (long/short) distances regularly within $r_g(t)$, it is formulated in equation (5.7) [Jam09].

$$r_g^a(t) = \sqrt{\frac{1}{n_c^a(t)} \sum_{i=1}^{n_c^a} (\overrightarrow{r_i^a} - \overrightarrow{r_{cm}^a})^2} \tag{5.7}$$

Where $\overrightarrow{r_i^a}$ refers to $i = 1, ..., n_c^a(t)$ positions recorded for individual a, and $\overrightarrow{r_{cm}^a} = 1/n_c^a(t) \sum_{i=1}^{n_c^a} \overrightarrow{r_i^a}$, which refers to the center of mass of the individual trajectory [Sah14, Sah13]. The algorithm 2 computes the radius of gyration distribution, with a complexity of $O(n^2 + 4n)$, it is implemented using Matlab platform. Where the data matrix composed from 4 columns corresponds to time (TimeCol), geographic coordinates (CoorCol_x and CoorCol_y), and the individual ID (PersonCol) respectively.

The resulted power law distribution of gyration radius is formulated in equation (5.8), and the results are explored as in figures 5.8, 5.9, and 5.10 for total days, work days, and off days period respectively. The $p(r_g)$ distribution investigates the aggregated traveled distances.

In figure 5.8 the $p(r_g)$ is computed for total individuals (population) along total observed period hours series. This classifies the r_{g_s} evolution along the time series, the figure shows the r_{g_s} variance along time and the approximate similarity in their patterns, also the highest r_{g_s} are the small once range between [5; 20] Km, while the range between [25; 30] Km are the lowest once. It indicates that the individuals have a tendency to travel in small r_{g_s} and in stable patterns along total period.

The r_{g_s} distributions are computed for the total days, work days, and off days period respectively as shown in figures 5.8, 5.9, and 5.10, it is clear that r_{g_s} could be aggregated in 3 main ranges ranked orderly the $10 \le r_g \le 15$, $r_g \le 3$, and $20 \le r_g \le 30$. The r_{g_s} evolution over time $r_g(t)$ is the same for all classes of time series during all periods(total, work, and off) days period. Generally, the individuals have high tendency to travel within $10 \le r_g \le 15$. The r_{g_s} are evolved in same patterns during all verified pariod classes, also they are stable and ranked orderely as mentioned ealier.

$$P(r_g) = (r_g)exp^{-(r_g)} \tag{5.8}$$

Algorithm 2: Probability distribution for the radius of gyration _Part 1.

```
Data: File of events sorted by time
Result: Radius of gyration distribution
data ← (Rows, TimeCol, CoorCol_x, CoorCol_y, PersonCol);
Val ← data(1, PersonCol);
debi ← 1;
ExactPerson ← data(debi, PersonCol);
Per ← 1;
Counter ← 0;
UnrepeatedPersons ← unique[data(PersonCol)];      // The unique function
 returns data with no repetition based on PersonCol
L ← size(UnrepeatedPersons);
for i<=debi to L do
    ExactPersonRepition ← size(find(data(:, PersonCol) == ExactPerson));
     // Count individual occurrence
    idi ← data(i, PersonCol);
    debj ← i + 1 ; index ← index + 1;
    XY(index, 1) ← data(i, CoorCol_x);
    XY(index, 2) ← data(i, CoorCol_y);
    for j<=debj to size(data) do
        idj ← data(j, PersonCol);
        if idi=idj then
            index ← index + 1;
            XY(index, 1) ← data(j, CoorCol_x);
            XY(index, 2) ← data(j, CoorCol_y);
            P1 ← (data(j, CoorCol_x) − data(i, CoorCol_x)^2;
            P2 ← (data(j, CoorCol_y) − data(i, CoorCol_y)^2;
            XY(index, 3) ← sqrt((P1)^2 + (P2)^2);
            XY(index, 4) ← data(j, TimeCol);
            debi ← j ;
        end
    end
end
/* Tagging treated individuals                                   */
if index >= ExactPersonRepition then
    for k=1 to size(data) do
        if data(k, PersonCol )= Val then
            data(k, PersonCol) ←= 0 ;
        end
    end
end
Per ← (Per + 1);
X_cm ← (mean(XY(:, 1));
Y_cm ← (mean(XY(:, 2));
N ← size(XY(:, 1));
Sum ← 0;
```

Algorithm 2: Probability distribution for the radius of gyration_Part 2.

```
for i=1 to N do
    Sum ← Sum + ((XY(:,1) − x_c m)^2 + (XY(:,2) − y_c m)^2));
    R_g(Counter,1) ← sqrt(Sum/N);
    Counter ← Counter + 1;
    R_g(Counter,2) ← idi;
    R_g(Counter,3) ← Sum(XY(:,3))/n;
end
for n=1 to size(data) do
    if data(n, PersonCol) ≠ 0 then
        debi ← n;
        Val ← data(n, PersonCol);
    end
end
dR_g ← 0;
dRg ← R_g(:,3)./R_g(:,1);
c1 ← 0;
c2 ← 0;
c3 ← 0;
cp1 ← 1;
cp2 ← 1;
cp3 ← 1;
for i=1 to size(R_g(:,1)) do
    if R_g(i) < 10 then
        c1(cp1) = R_g(i);
        cp1(cp1) = cp1 + 1;
    end
    if 10 ⩽ R_g(i) ≤ 20 then
        c2(cp2) = R_g(i);
        cp21(cp2) = cp2 + 1;
    end
    if R_g(i) > 20 then
        c3(cp3) = R_g(i);
        cp3(cp3) = cp3 + 1;
    end
end
/* Classify r_gs into samples using c1,c2,c3 and compute p(r_g) distribution
   of radius of gyration using makedit and expfit matlab functions   */
p(r_g) ← makedit(expfit(R_g(:,1)));
Draw p(r_g) distribution using c1, c2, c3 Samples;
```

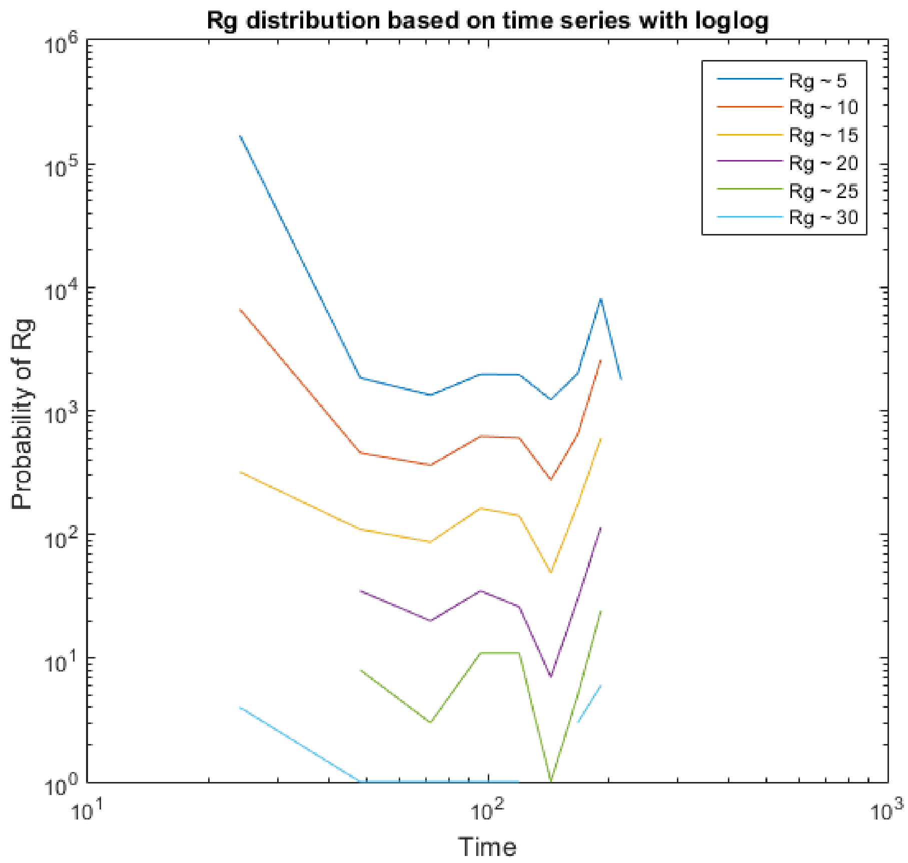


Figure 5.8 – The Rg distribution based on time series with loglog during *total* observed period.

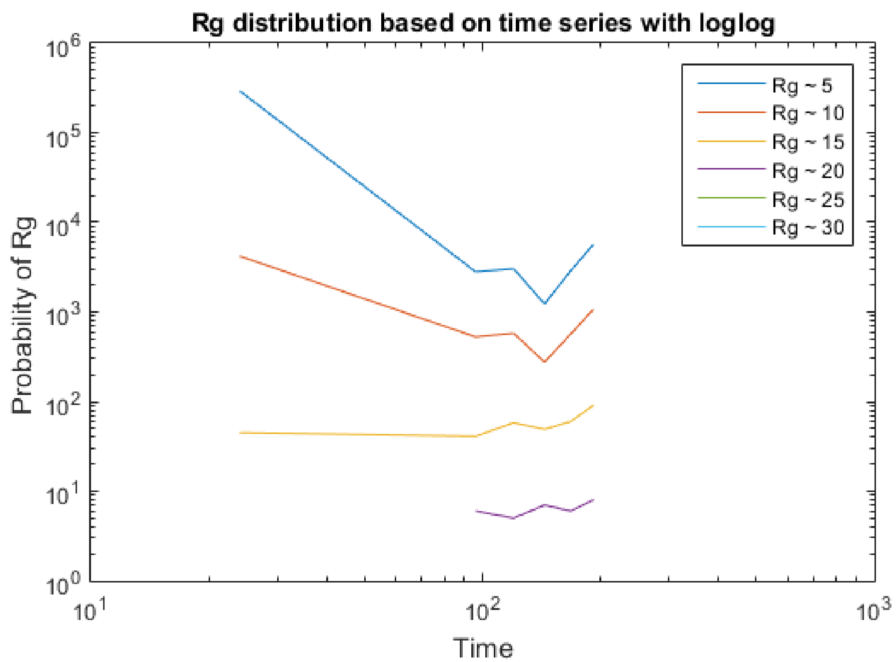


Figure 5.9 – The Rg distribution based on time series with loglog during *work* days period.

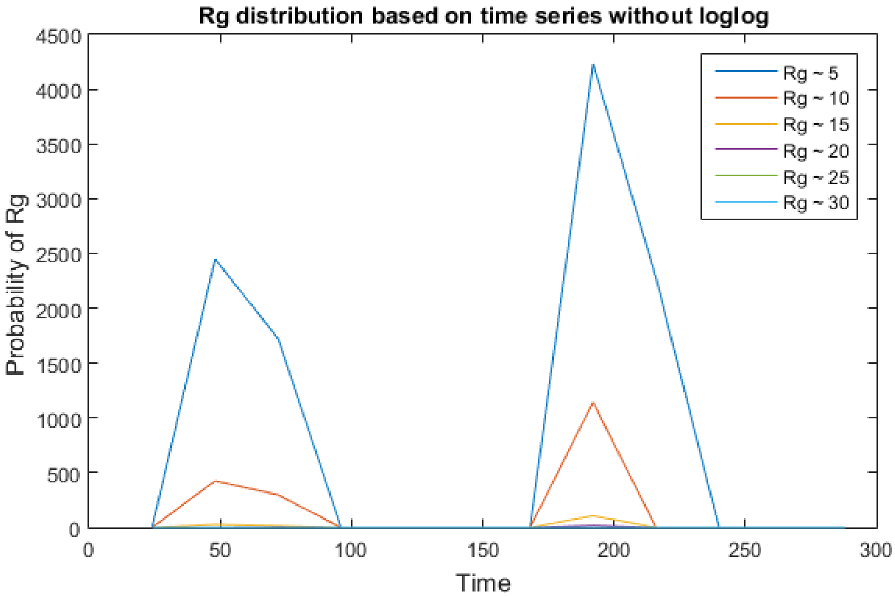


Figure 5.10 – The Rg distribution based on time series with loglog during off days period.

3 Most frequent positions to uncover the individual tendency, according to visited locations.

4 Principal axes θ (moment of inertia): This technique permits to study the individuals trajectories in a common reference frame by diagonalizing each of trajectory inertia tensor, hence compare their different trajectories. Moment of inertia to any object is obtained from the average spread of object mass from a given axis. This could be elaborated using two dimensional matrix called tensor of inertia. Then, using the standard physical formula to obtain the inertia tensor of individual trajectories, as in the following equations (5.9)- (5.19) respectively:

$$I = \begin{pmatrix} I_{xx} & I_{xy} \\ I_{yx} & I_{yy} \end{pmatrix} \tag{5.9}$$

$$I_{xx} = \sum_{i=1}^{n} y_i^2 \tag{5.10}$$

$$I_{yy} = \sum_{i=1}^{n} x_i^2 \tag{5.11}$$

$$I_{xy} = I_{yx} = -\sum_{i=1}^{n} x_i y_i \tag{5.12}$$

The tensor I is symmetric, so it would give a set of coordinates that make I be diagonal, these coordinates are called tensor principle axes $(\hat{e}_1, \hat{e}_2)$, this set of coordinates formulate I as in the following equation (5.13):

$$I_D = \begin{pmatrix} I_1 & 0 \\ 0 & I_2 \end{pmatrix} \tag{5.13}$$

Where, I_1 and I_2 are the principal moments of inertia, and they are match the eigenvalues of I. However, they can be computed from the original points as in the following equations (5.14), (5.15), and (5.16):

$$I_1 = \frac{1}{2}(I_{xx} + I_{yy}) - \frac{1}{2}\mu \tag{5.14}$$

$$I_2 = \frac{1}{2}(I_{xx} + I_{yy}) - \frac{1}{2}\mu \tag{5.15}$$

$$\mu = \sqrt{4I_{xy}I_{yx} + I_{xx}^2 - 2I_{xx}I_{yy} + I_{yy}^2} \tag{5.16}$$

Where, the individual trajectory as follows: $(x_1, y_1), (x_2, y_2), ..., (x_n, y_n)$, where $n =$ is the number of positions passed by the individual.

The principal axes (symmetric try axes of a trajectory) $(\hat{e}_1, \hat{e}_2)$ are obtained from the eigenvectors. Then the individual trajectory rotation could be obtained by transforming principle axes of the individual $(\hat{e}_1, \hat{e}_2)$ to common intrinsic reference frame $(\hat{e}_x, \hat{e}_y)$. Then, computing the angle between $\hat{e}_x$ and $\hat{e}_1$ would be as in equation (5.17):

$$\cos(\theta) = \frac{\overrightarrow{v_1}.\hat{e_x}}{|v_1|} \tag{5.17}$$

Where, v_1 represents eigenvector, which is related to the eigenvalue I_1 as in equation (5.18).

$$\overrightarrow{v_1} = \begin{bmatrix} -\frac{I_{xy}}{1/2I_{xx} - 1/2I_{yy} + 1/2\mu} \\ 1 \end{bmatrix} \tag{5.18}$$

obtaining the equation (5.19).

$$\cos(\theta) = -I_{xy}(1/2I_{xx} - 1/2I_{yy} + 1/2I_{\mu})^{-1} \frac{1}{\sqrt{1 + \frac{I_{xy}^2}{(1/2I_{xx} - 1/2I_{yy} + 1/2I_{\mu})^2}}} \tag{5.19}$$

Rotation by θ produced a conditional rotation of $180\,^{\circ}$ as the most frequent position lies in x>0.

5 Standard Deviation: Verify the horizontal and vertical coordinates of individual mobility in the intrinsic reference frame. However, trajectories are scaled on intrinsic axes using standard deviation of the locations for each individual a, as in equations (5.20), (5.21), and (5.22).

$$\sigma_x^a = \sqrt{\frac{1}{n_c^a}\sum_{i=1}^{n_c^a}(x_i^a - x_{cm}^a)^2} \tag{5.20}$$

$$\sigma_y^a = \sqrt{\frac{1}{n_c^a}\sum_{i=1}^{n_c^a}(y_i^a - y_{cm}^a)^2} \tag{5.21}$$

Then, obtain the universal density function using equation (5.22). Using the spatial density function is to aggregate the individuals with the common r_{g_s}

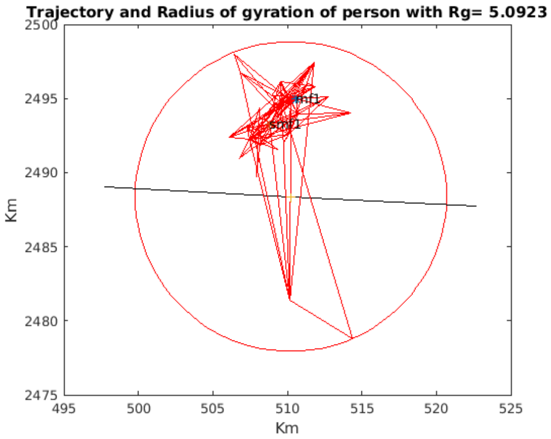
$$\widetilde{\Phi} = (x/\sigma_x, y/\sigma_y) \tag{5.22}$$

The potential trajectories of the individuals have been explored to visualize and analyze the mobility of 3 unique individuals along the period of the total 12 observed days.

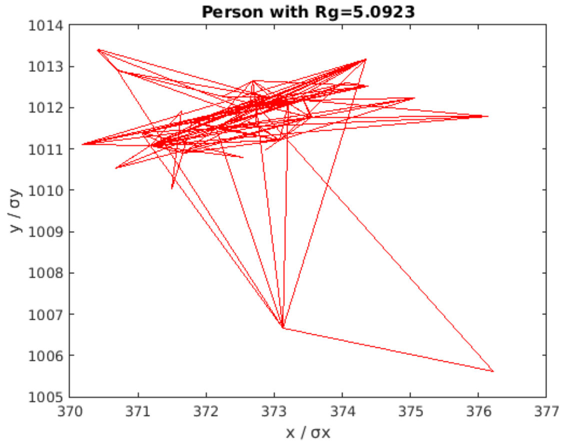
These individuals are chosen according to their pertinence of r_g, where each one is elected randomly from unique r_g sample, then the potential trajectories of the individuals $user_1$, $user_2$, $user_3$ are chosen from r_{g_2} where r_g=2Km, r_{g_9} where r_g=9Km, and $r_{g_{15}}$ where r_g=15Km respectively. They are computed and simulated as in the following figures 5.11, 5.12, and 5.13 for total period, work days period, and off days period respectively.

Afterwards, the choosing of individuals from different classified groups with regard to their radius of gyration, then the trajectories would be rescaled for these individuals according to the physical mobility characteristics mentioned above.
It's concluded that even in the microscopic level of abstraction the individuals trajectories are related to their r_g for any period class. Whereas, the trajectories shapes of work days period are variant from that of the off days period, and the patterns of work period are more isotropic, whereas that of the off days period are less isotropic. So, the individuals have the same tendency of the population, as concluded in previous sections of the population trajectories. Also, it is clear that the individuals are travel in many short trajectories during work days, but they are travel in long trajectories during off days within large r_g.

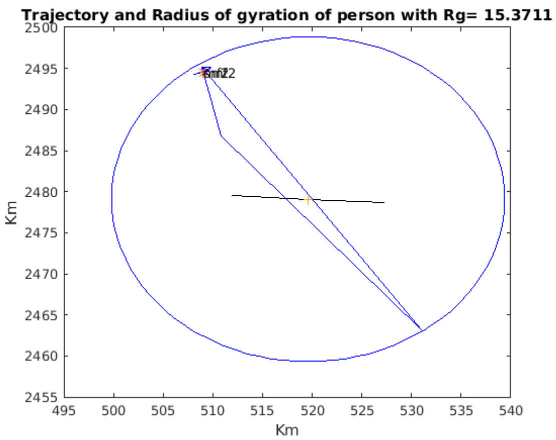
The frequent positions show that individuals have regular pattern, since they are so related to some places my be the home/work places, and they return to some positions regularly. As well as, the trajectories are stable in the middle r_{g_s}, but variant in the large and small ones.



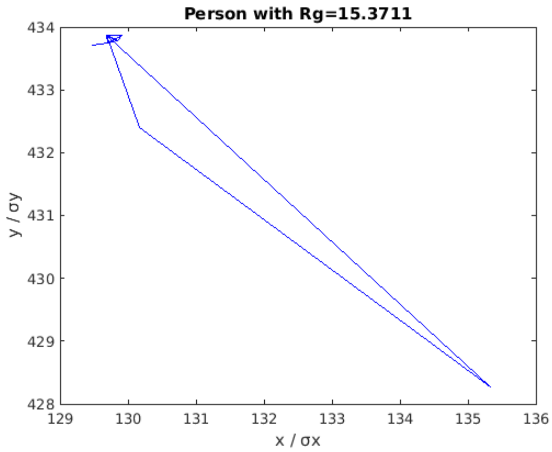
(a) Individual $user_1$ chosen from r_{g_2}



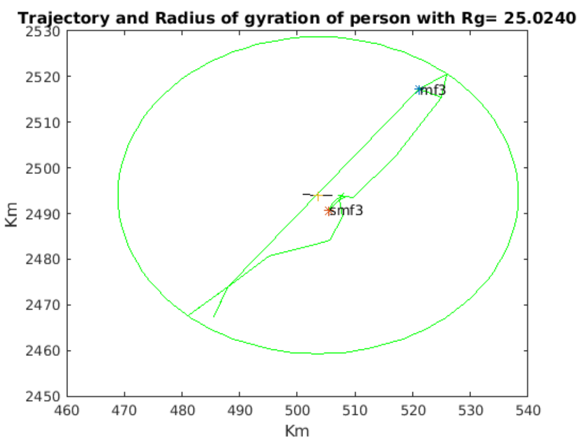
(b) Individual $user_1$ chosen from r_{g_2}



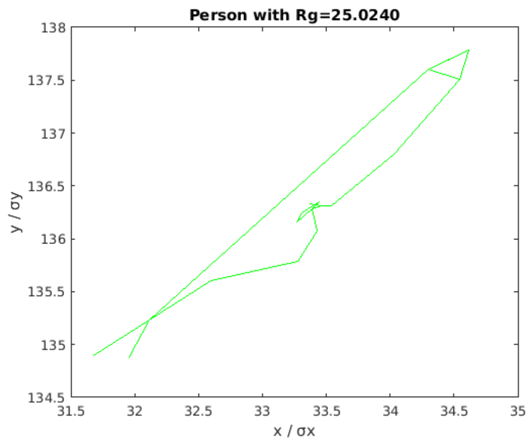
(c) Individual $user_2$ chosen from r_{g_9}



(d) Individual $user_2$ chosen from r_{g_9}

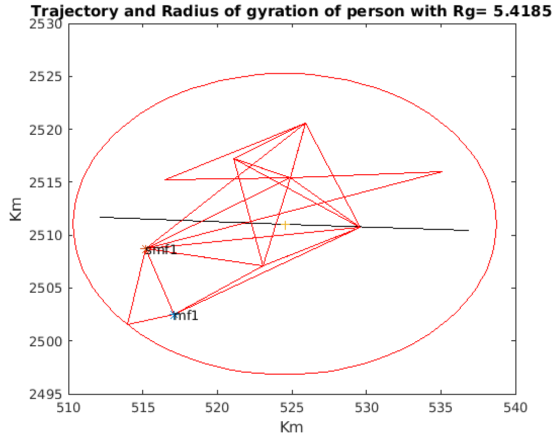


(e) Individual $user_3$ chosen from $r_{g_{15}}$

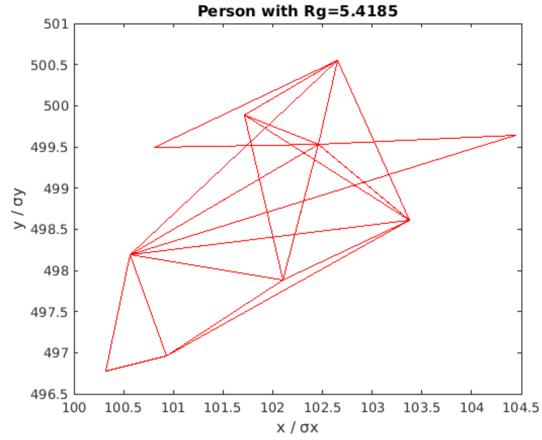


(f) Individual $user_3$ chosen from $r_{g_{15}}$

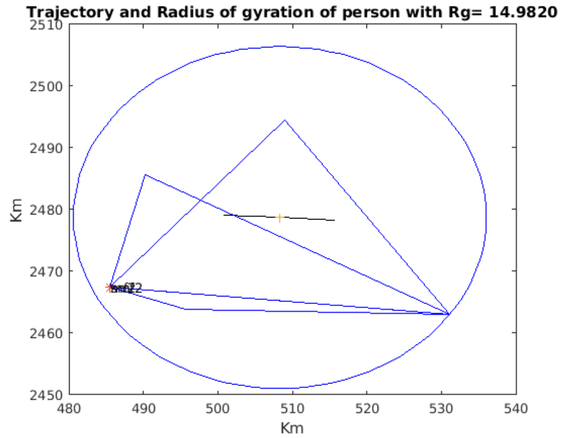
Figure 5.11 – Estimated individuals trajectories within intrinsic reference frame in microscopic perspective level during the $total$ days with red circle refers to the most frequent position: a: Individual trajectory chosen from r_{g_5}, b: Individual scaled trajectory chosen from r_{g_5}, c: Individual trajectory chosen from $r_{g_{15}}$, d: Individual scaled trajectory chosen from $r_{g_{15}}$, e: Individual trajectory chosen from $r_{g_{25}}$, f: Individual scaled trajectory chosen from $r_{g_{25}}$.



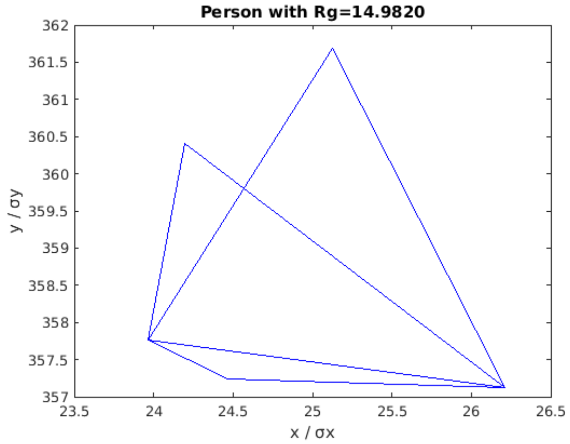
(a) Individual $user_1$ chosen from r_{g_2}



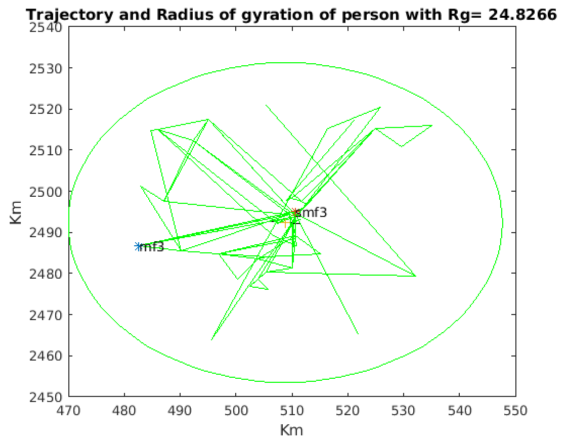
(b) Individual $user_1$ chosen from r_{g_2}



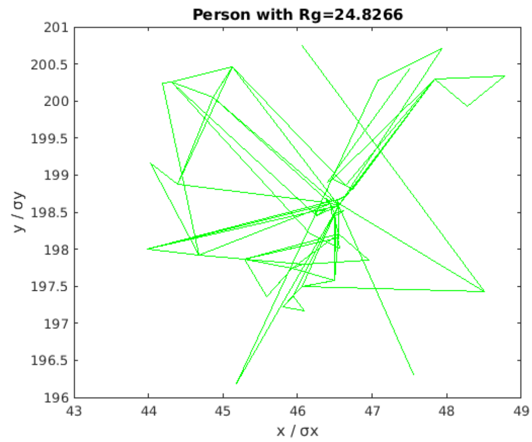
(c) Individual $user_2$ chosen from r_{g_9}



(d) Individual $user_2$ chosen from r_{g_9}

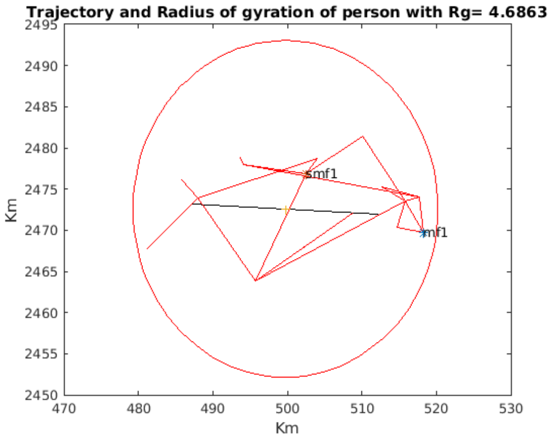


(e) Individual $user_3$ chosen from $r_{g_{15}}$

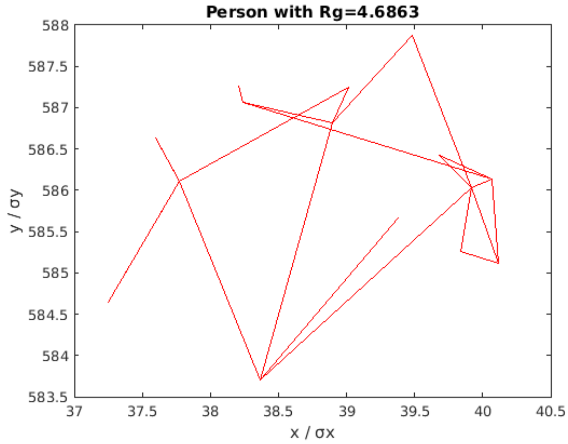


(f) Individual $user_3$ chosen from $r_{g_{15}}$

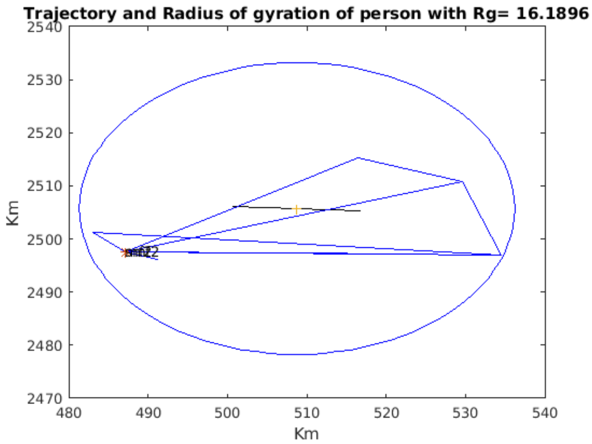
Figure 5.12 – Estimated individuals trajectories within intrinsic reference frame in microscopic perspective level during $work$ days with red circle refers to most frequent position: a: Individual trajectory chosen from r_{g_5}, b: Individual scaled trajectory chosen from r_{g_5}, c: Individual trajectory chosen from $r_{g_{14}}$, d: Individual scaled trajectory chosen from $r_{g_{14}}$, e: Individual trajectory chosen from $r_{g_{24}}$, f: Individual scaled trajectory chosen from $r_{g_{24}}$.



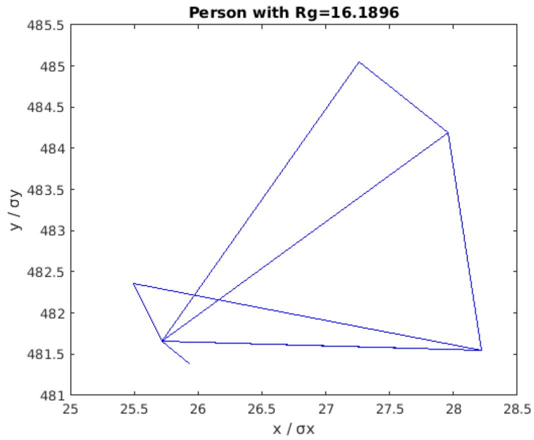
(a) Individual $user_1$ chosen from r_{g_2}



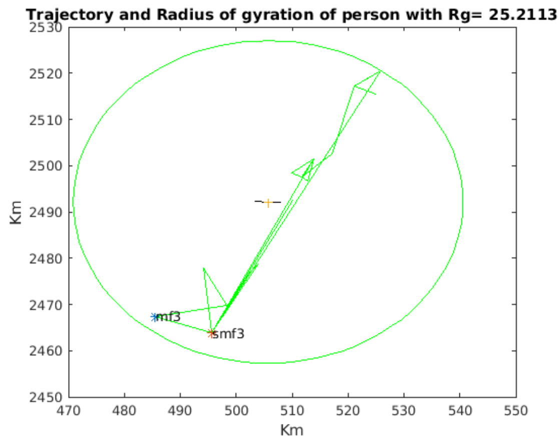
(b) Individual $user_1$ chosen from r_{g_2}



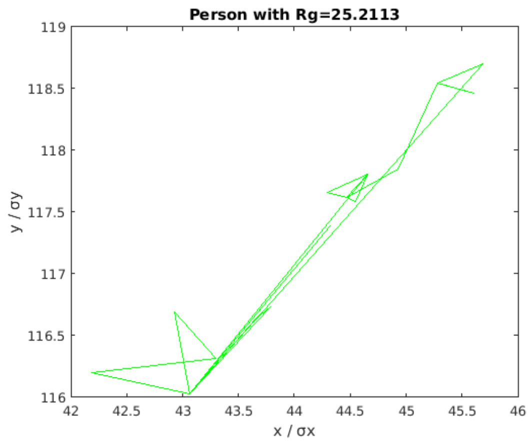
(c) Individual $user_2$ chosen from r_{g_9}



(d) Individual $user_2$ chosen from r_{g_9}



(e) Individual $user_3$ chosen from $r_{g_{15}}$



(f) Individual $user_3$ chosen from $r_{g_{15}}$

Figure 5.13 – Estimated individuals trajectories within intrinsic reference frame in microscopic perspective level during off days with red circle refers to most frequent position: a: Individual trajectory chosen from r_{g_4}, b: individual scaled trajectory chosen from r_{g_4}, c: Individual trajectory chosen from $r_{g_{16}}$, d: Individual scaled trajectory chosen from $r_{g_{16}}$, e: Individual trajectory chosen from $r_{g_{25}}$, f: Individual scaled trajectory chosen from $r_{g_{25}}$.

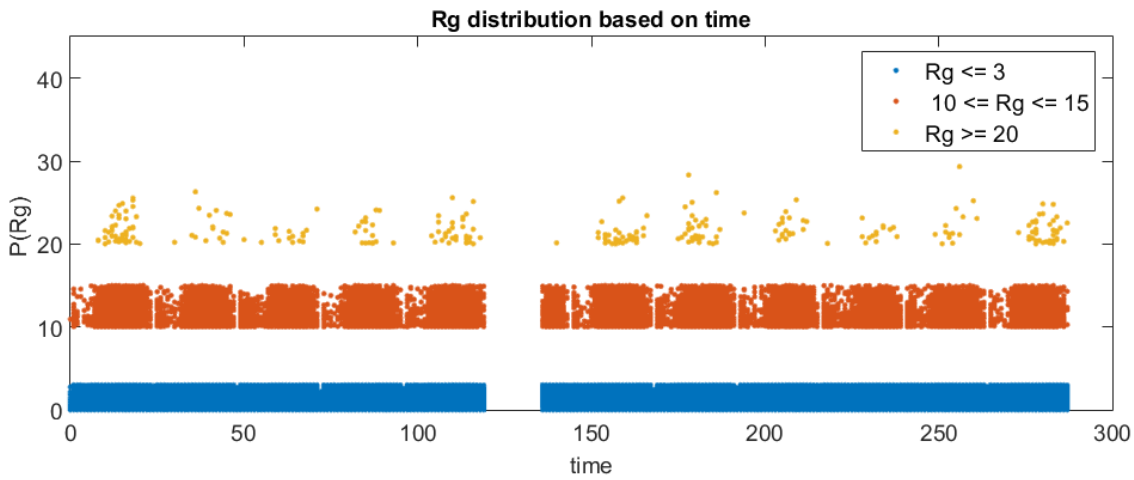
Trajectories within Radius of Gyration

The radius of gyration (r_g) is considered to be the most dedicated feature that is capable of characterizing the periodic trajectories of individuals [Sah14, Sah13]. However, the travel distance distributions showed that Δr_s of individuals are almost identical, so they are invariant. As well as, the activities distributions are uncovered the similarities of regular patterns, during the time evolution of radius of gyration (r_g).

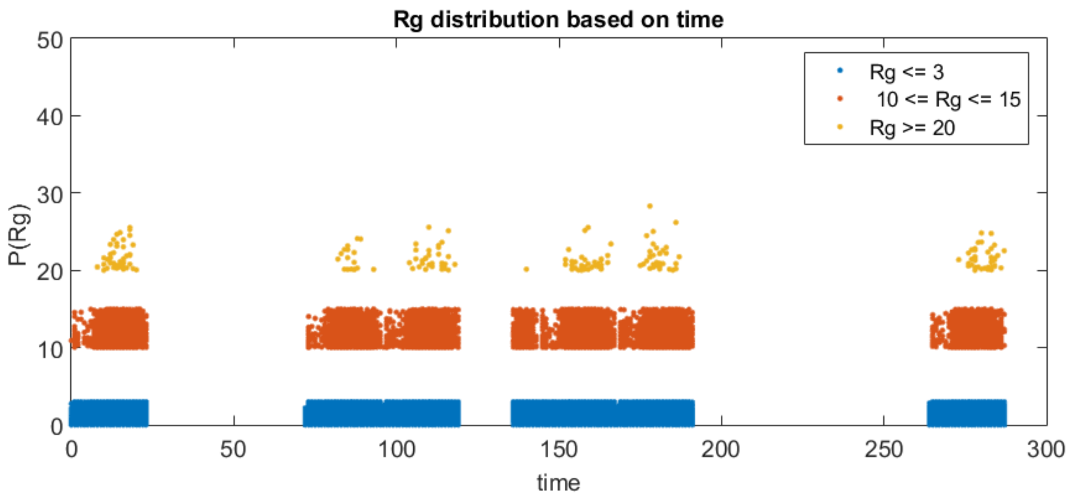
The experiments revealed the possibility of classifying individuals activities into some patterns samples [Cor13]. Almost individuals have similar activities patterns, and these activities in general could be varied depending on the days either working or off days. The individuals are sampled according to their relevance $r_g(t)$, where each group has similar patterns (similar asymptotic). So, the distribution $p(\Delta r|r_g)$ is computed as shown in figure 5.14, where the $p(\Delta r|r_g)$ is presented in function of Δr. The distribution is computed for the total individuals during total, work, and off days period. It's clear that individuals of small r_{g_s} have short Δr, on the other hand individuals of large r_{g_s} have mix of short and long Δr_s. Therefore, it could be said that r_g and Δr are correlated to each other, and capable of giving impressive view of individuals mobility patterns. This distribution shows that the individuals are traveling in Δr_s that are bounded by their r_g, the short distances are included within small r_{g_s}. However, the large r_{g_s} have the mix of short and long Δr_s, which give them the heterogeneity property.

As well as, the PDF of each $p(\Delta r|r_g)$ is verified for each sample for individuals travel within the r_{g_s}, as shown in figure 5.15, also its probability distribution is modeled by the equation (5.23). The individuals are sampled using PDF, this sampling is achieved according to their bounded r_g. It is performed for the total days, work days, and off days period. Hence, it is clear that the small r_g is the most prevalent pattern for all period classes (total, work, off). Whereas, in off days period there is some change where the $10 \leq r_g \leq 15$ is increased a little bit, which means the individuals change their regular patterns during off days and vacations. However, the short Δr_s still the dominant pattern, whenever small (r_g) are get close to value zero, hence the pattern get to be irregular.

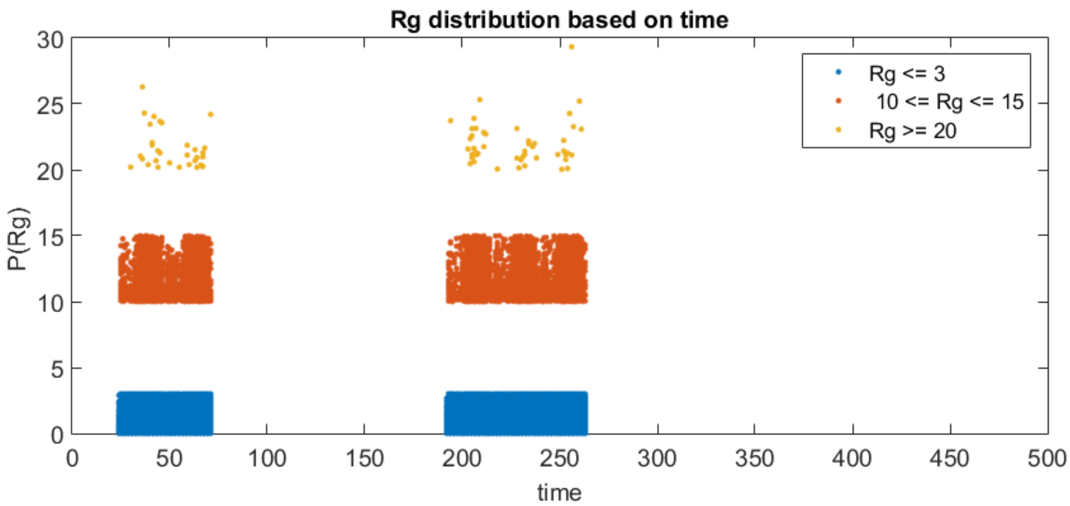
$$P(\Delta r|r_g) = (\Delta r|r_g)exp^{-(\Delta r|r_g)} \tag{5.23}$$



(a) $r_g(t)$ during $total$ observed period

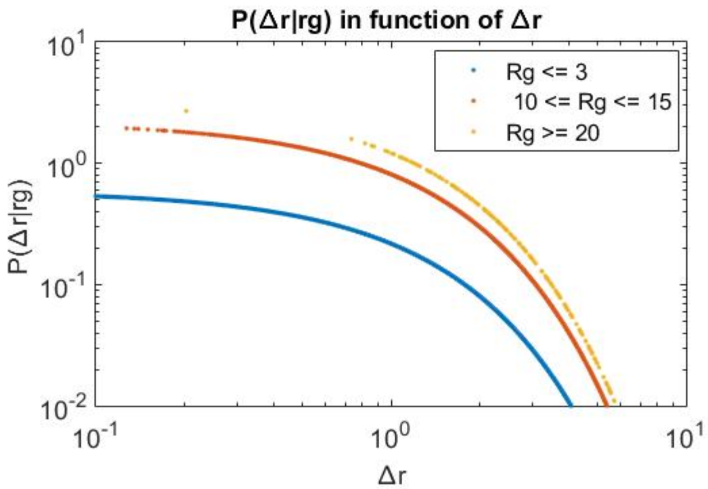


(b) $r_g(t)$ during $work$ days observed period

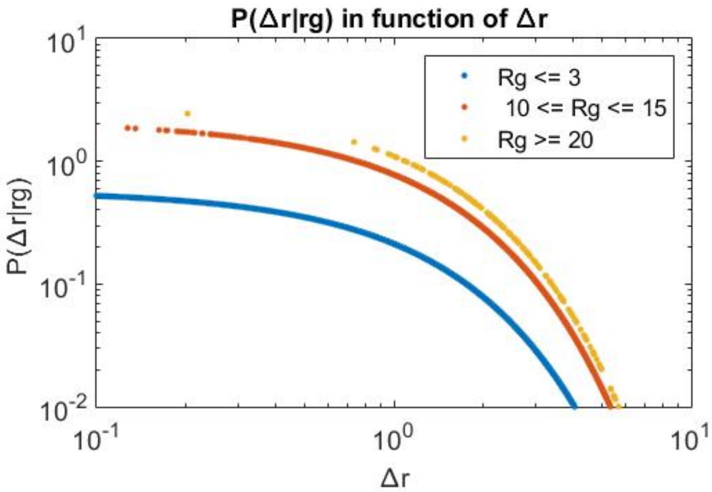


(c) $r_g(t)$ during off days observed period

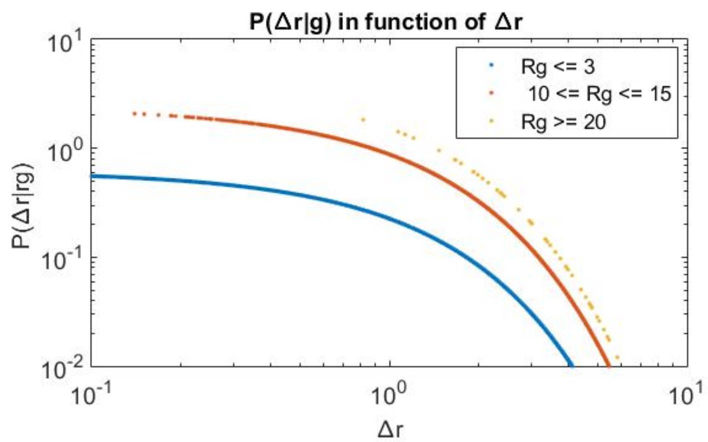
Figure 5.14 – Distribution of gyration radius in function of time series $r_g(t)$ for all individuals, the r_{g_s} evolution sampled in 3 main groups as in the figure legend, the measuring units for time unit is hours and for r_{g_s} unit is Km, a: The $r_g(t)$ during $total$ observed period, b: The $r_g(t)$ during $work$ days observed period, c: The $r_g(t)$ during off days observed period.



(a) $p(\Delta r|r_g)$ during $total$ observed period

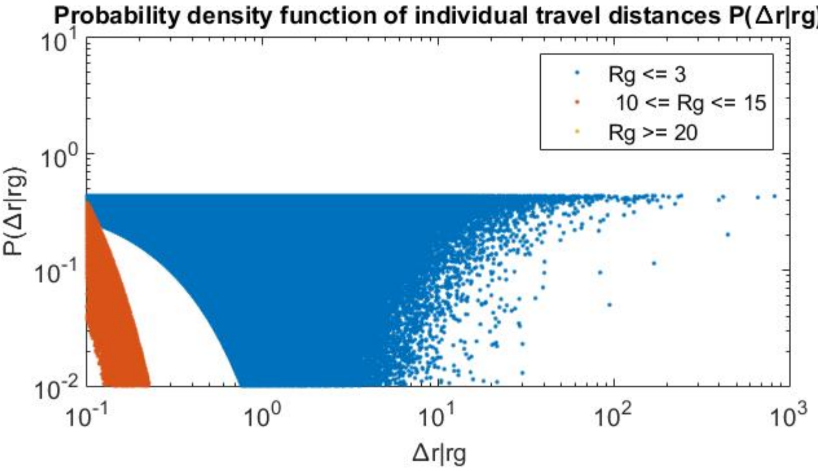


(b) $p(\Delta r|r_g)$ during $work$ days observed period

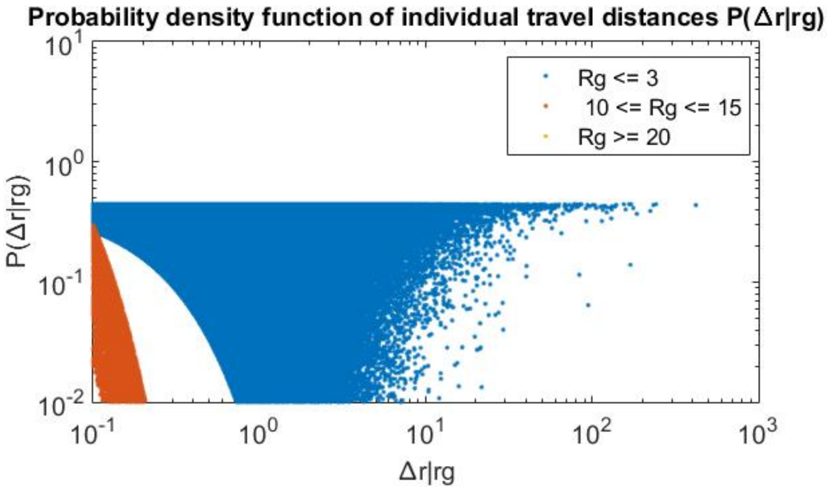


(c) $p(\Delta r|r_g)$ during off days observed period

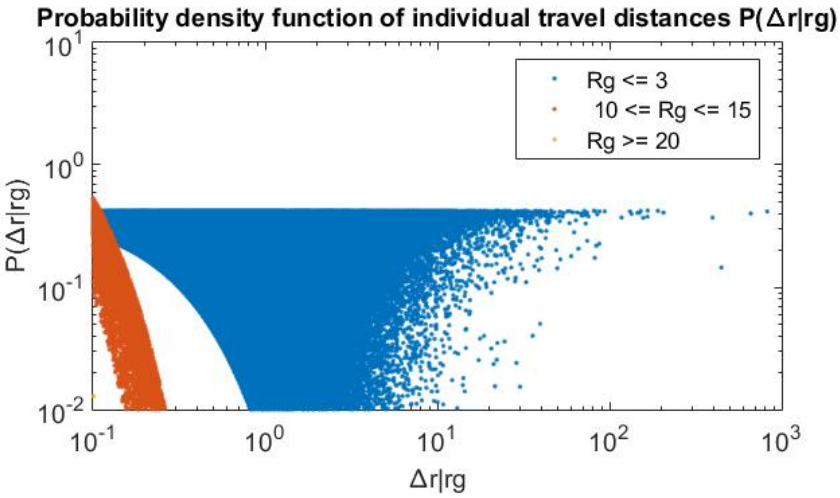
Figure 5.15 – Radius of gyration distribution $p(\Delta r|r_g)$ in function of (Δr) for the individuals travel distances that bounded by their relevant r_{g_s}, the small r_g bounded short travel distances Δr_s and the bigger ones have mix of short and long Δr_s, also small and medium r_{g_s} are the dominant in the three periods, a: $p(\Delta r|r_g)$ during $total$ observed period, b: $p(\Delta r|r_g)$ during $work$ days observed period, c: $p(\Delta r|r_g)$ during off days observed period.



(a) $pdf(\Delta r|r_g)$ during $total$ observed period



(b) $pdf(\Delta r|r_g)$ during $work$ days period



(c) $pdf(\Delta r|r_g)$ during off days period

Figure 5.16 – The PDF of individuals in function of their travel distances within relevant r_{g_s}, that classifies the individuals samples, a: PDF of individuals travel distances within relevant r_{g_s} during $total$ observed period, b: PDF of individuals travel distances within relevant r_{g_s} during $work$ days observed period, c: PDF of individuals travel distances within relevant r_{g_s} during off days observed period.

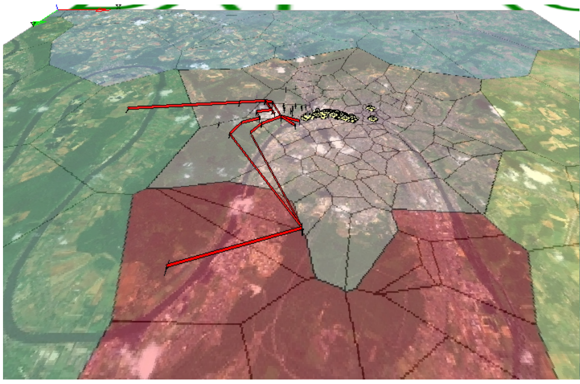
Real and Simulated Individual Trajectory Estimation (Microscopic Perspective)

The CDRs data of each individual have many lack of data followed a bursty pattern, since the recorded data are not continuous instead it is discrete (irrelevant locations), because it is limited by the individual activities (just when makes mobile phone activity), whenever a mobile phone is inactive there is no data, hence this would produce an irregular activities pattern.

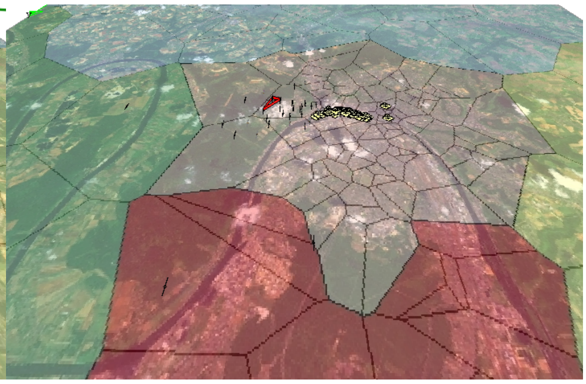
Hence, this lack of data need some estimations to compensate them, the individual trajectory has time- invariant properties, however the trajectories estimations need to be computed according to common individual mobility characteristics,as mentioned earliear.
The algorithm 3 is to compute an individual trajectory, with complexity $O(n)$, it is implemented using Gama-Platform. The figures 5.17, 5.18 are to simulate one individual trajectories during the 12 days. Hence, his dynamic behavior could be explored from the variance between his trajectories of each day independently, notice that day 9 has lack of data so his trajectory was incomplete and very limited. Whereas, his trajectory patterns of other days are approximately similar, and restricted to specific regions.

Algorithm 3: Real and estimated individual trajectory.

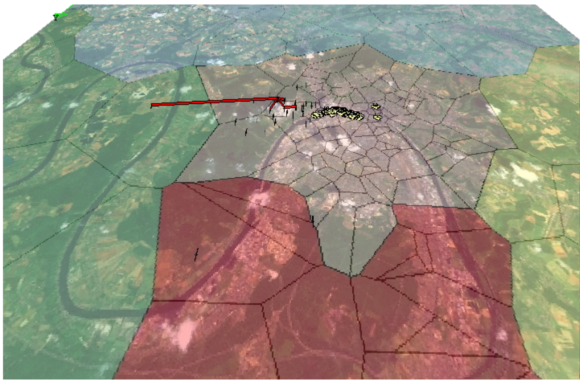
Data: Shape File of Individuals Sorted by Time
Result: Real and Estimated Individual Trajectory
$ReadData \leftarrow Shapefile$;
$AgentAttrib \leftarrow (time, position(x, y), personID)$;
$Create \leftarrow AgentSpecies(personID)$;
$AgentAttrib \leftarrow Sort(AgentAttribTime)$;
while *not eof(File)* **do**
    $PointsList \leftarrow position(x, y)$;
    $TimeList \leftarrow time$;
    for *j=1 to eof(PointsList)* **do**
        if $PointsList(i + 1) \neq PointsList(i)$ **then**
            $Distance \leftarrow PointsList(i + 1) - PointsList(i)$;
            $TimeDiff \leftarrow TimeList(i + 1) - TimeList(i)$;
        end
        $Speed \leftarrow Distance/TimeDiff$;
        $Trajectory \leftarrow Trajectory + PointsList(i)$;
    end
    Draw $(Trajectory)$;
end



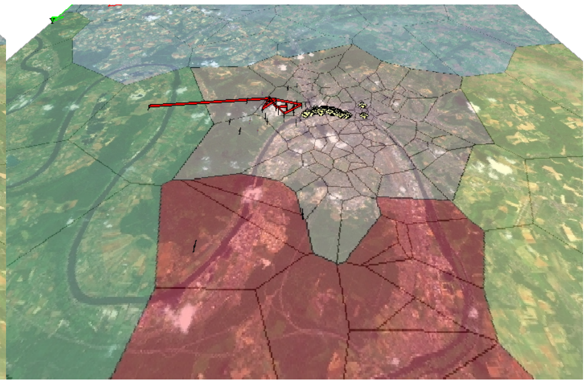
(a) Alias519 potential trajectory for day 4



(b) Alias519 potential trajectory for day 5

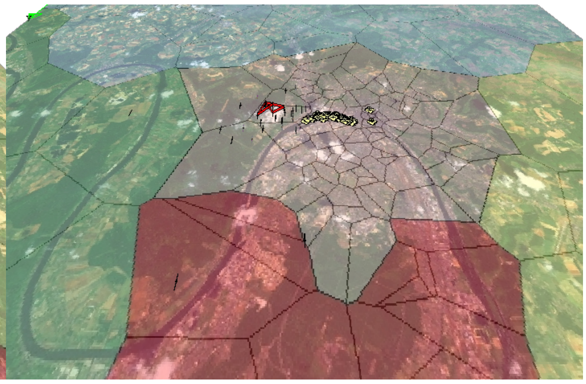


(c) Alias519 potential trajectory for day 6



(d) Alias519 potential trajectory for day 7

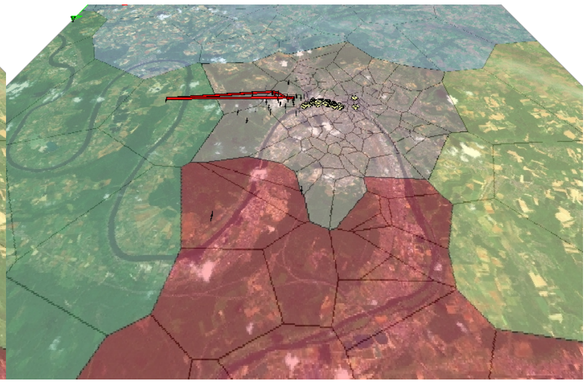
(e) Alias519 potential trajectory for day 8



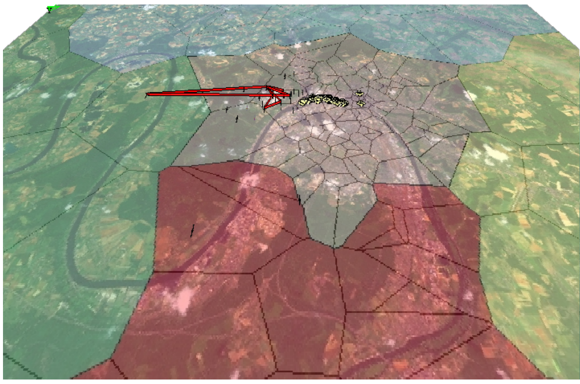
(f) Alias519 potential trajectory for day 9

Figure 5.17 – Alias519 potential trajectory for *total* period (12 days).

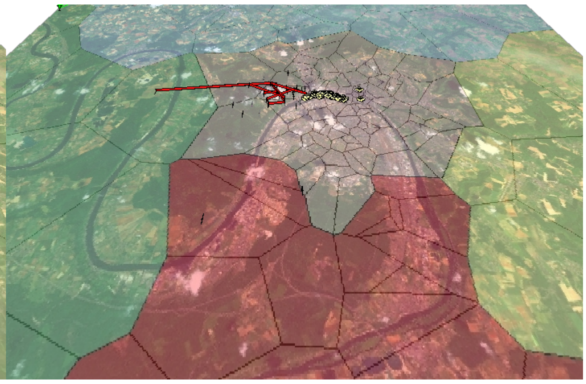
(a) Alias519 potential trajectory for day 10



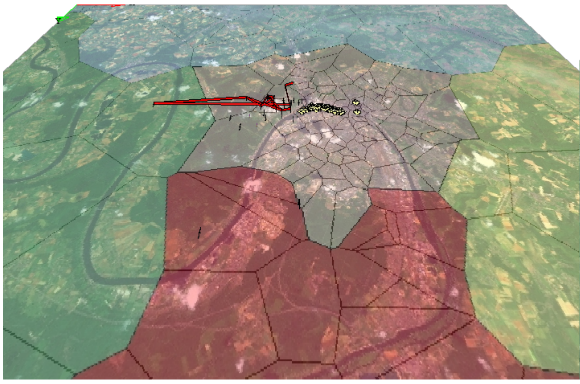
(b) Alias519 potential trajectory for day 11



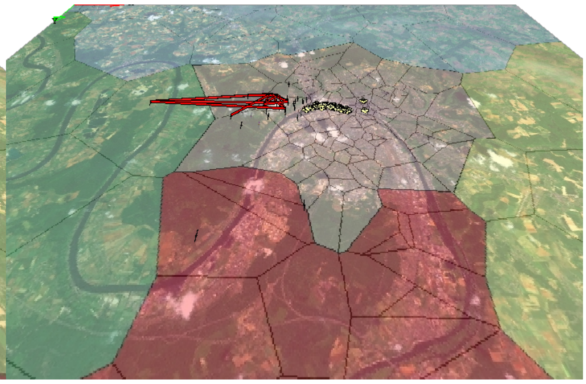
(c) Alias519 potential trajectory for day 12



(d) Alias519 potential trajectory for day 13



(e) Alias519 potential trajectory for day 14



(f) Alias519 potential trajectory for day 15

Figure 5.18 – Alias519 potential trajectory for $total$ period (12 days).

Output

The general distributions of activities and displacements during inter-event time are estimated for all the observed days period. As well as, there are other distributions, which are obtained for each day, The results of Day 4 are represented, the distributions and histograms are explored for Day 4 figures 5.19- 5.29, whereas the other days results are placed in the Appendix A of figures.

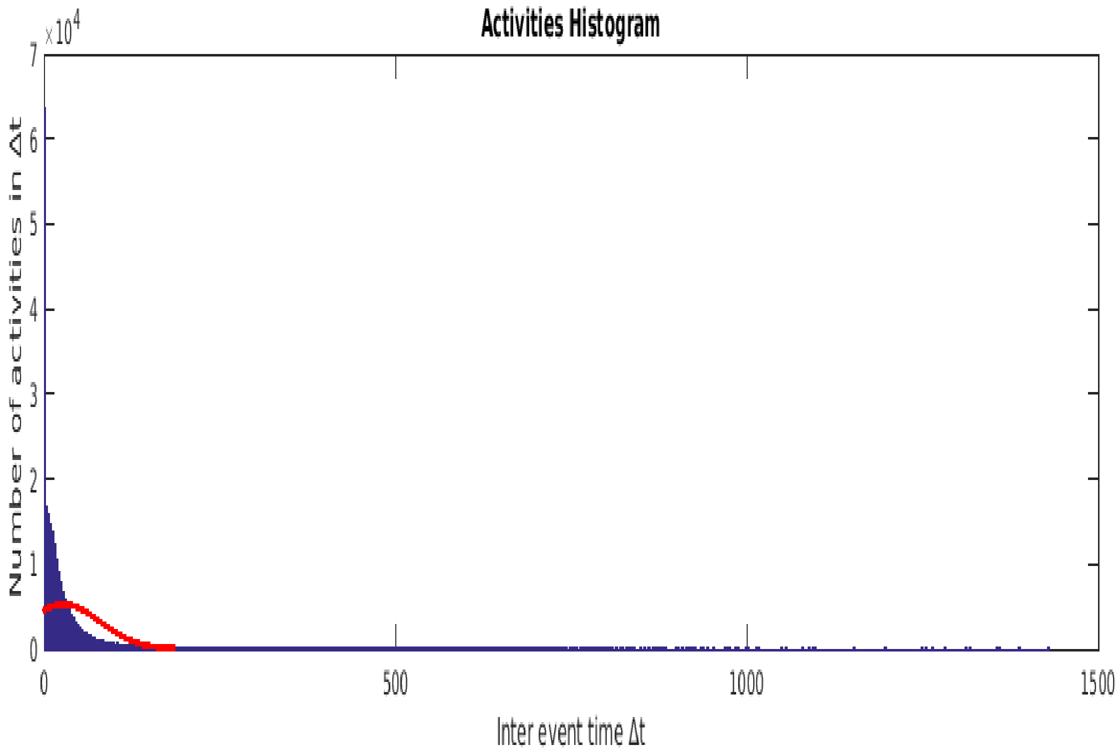


Figure 5.19 – Day 4 activities histogram.

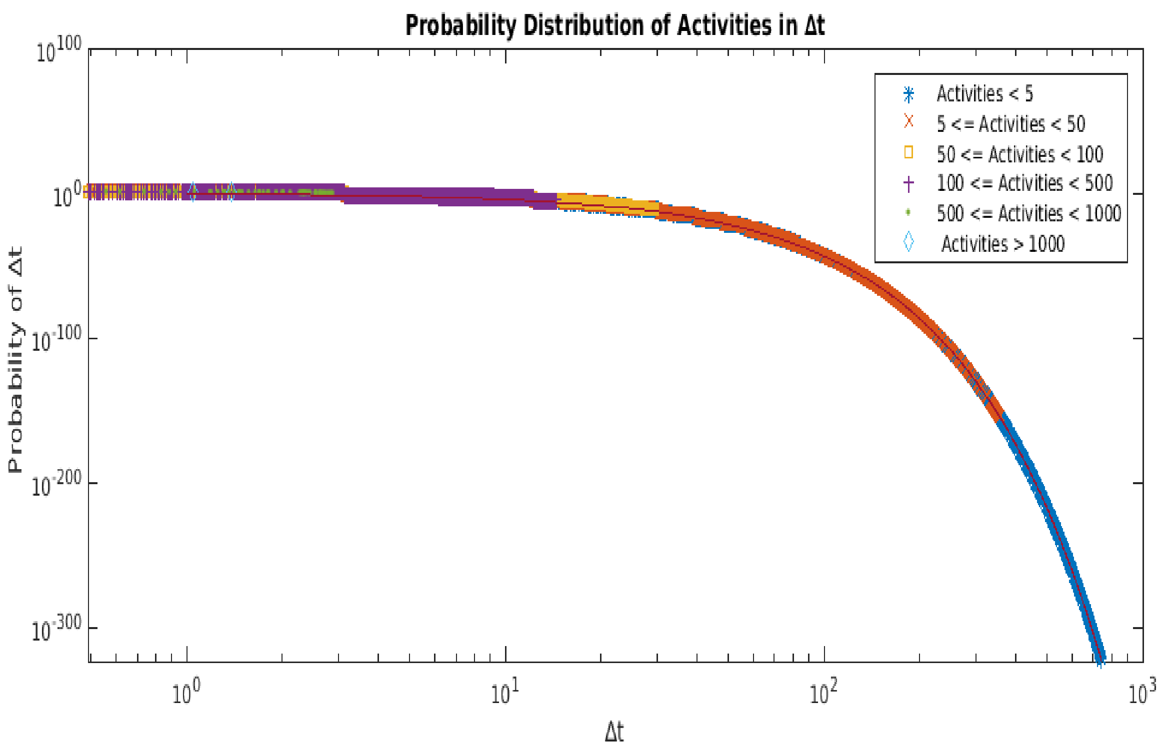


Figure 5.20 – Day 4 activities by inter-event time.

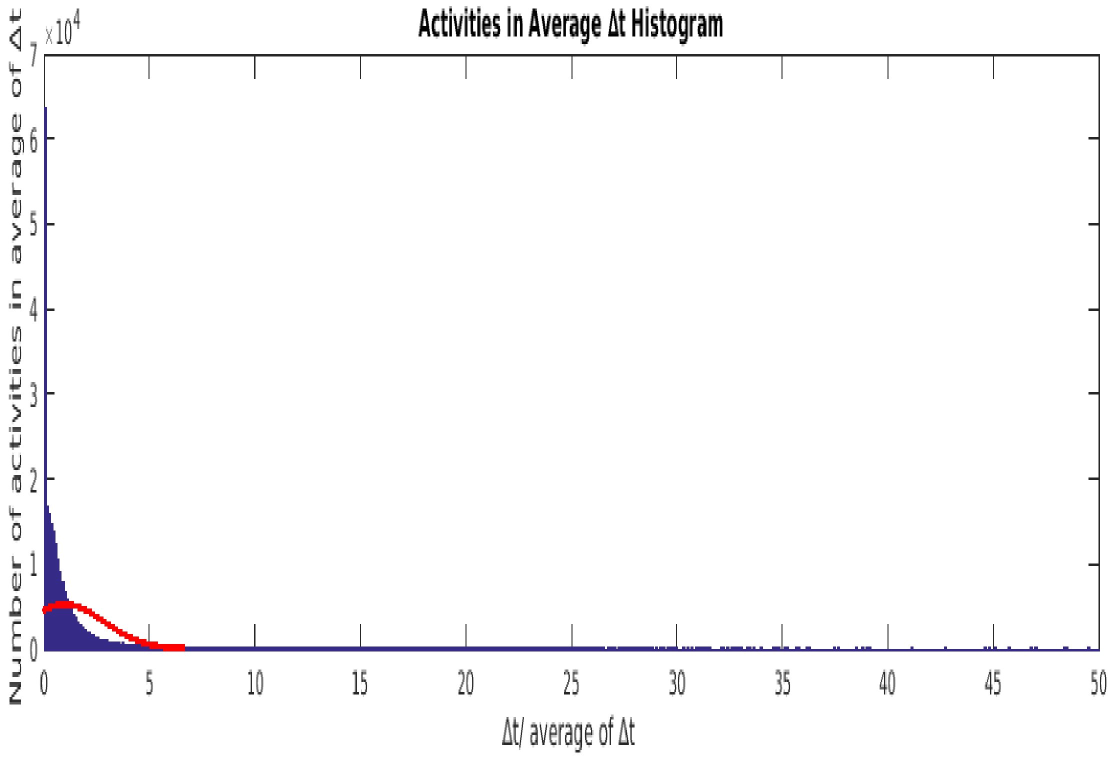


Figure 5.21 – Day 4 activities in average time histogram.

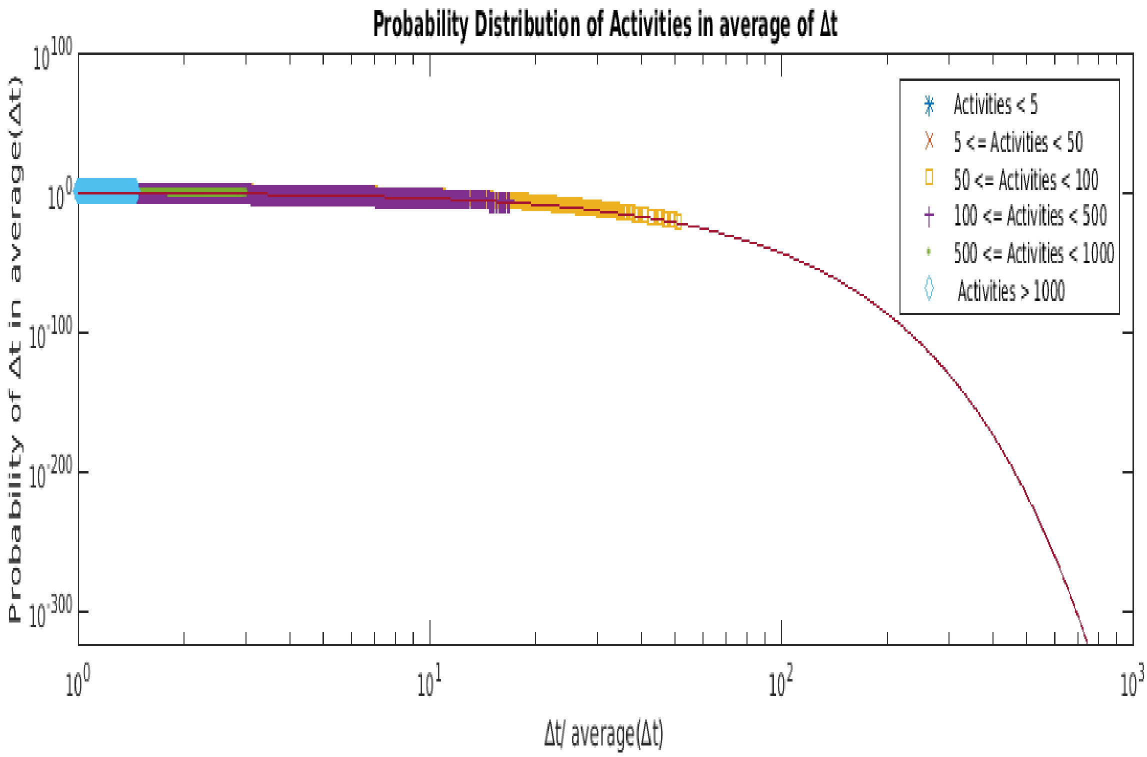


Figure 5.22 – Day 4 activities in average inter-event time.

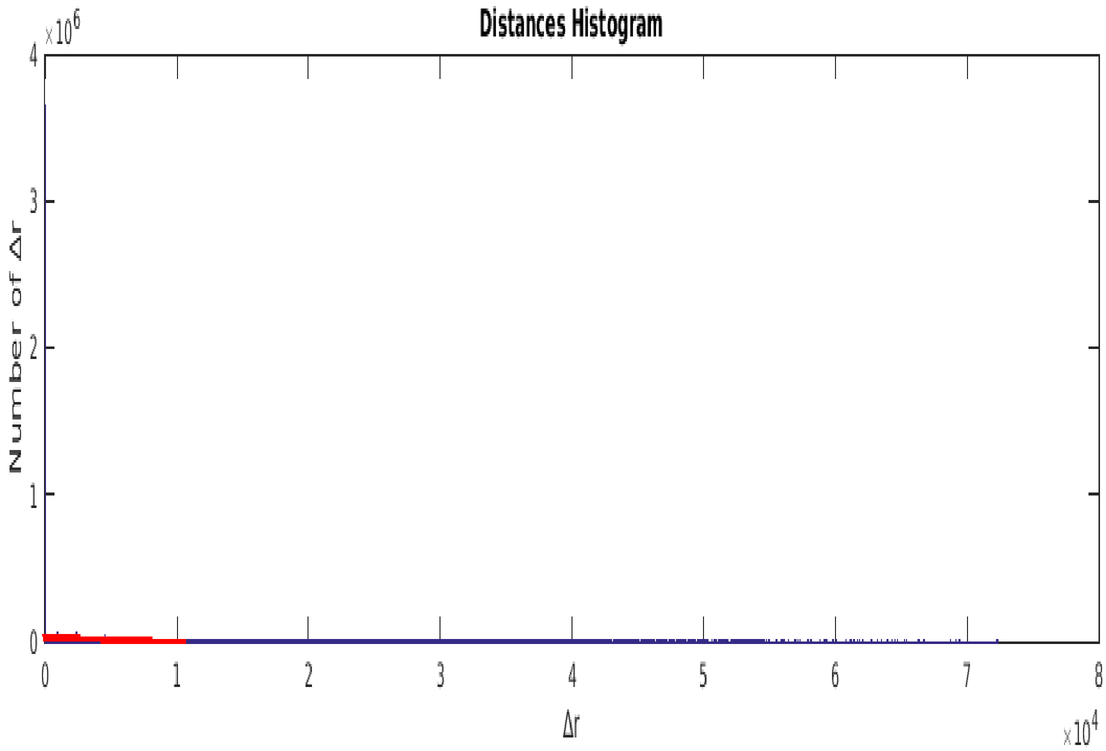


Figure 5.23 – Day 4 distances histogram.

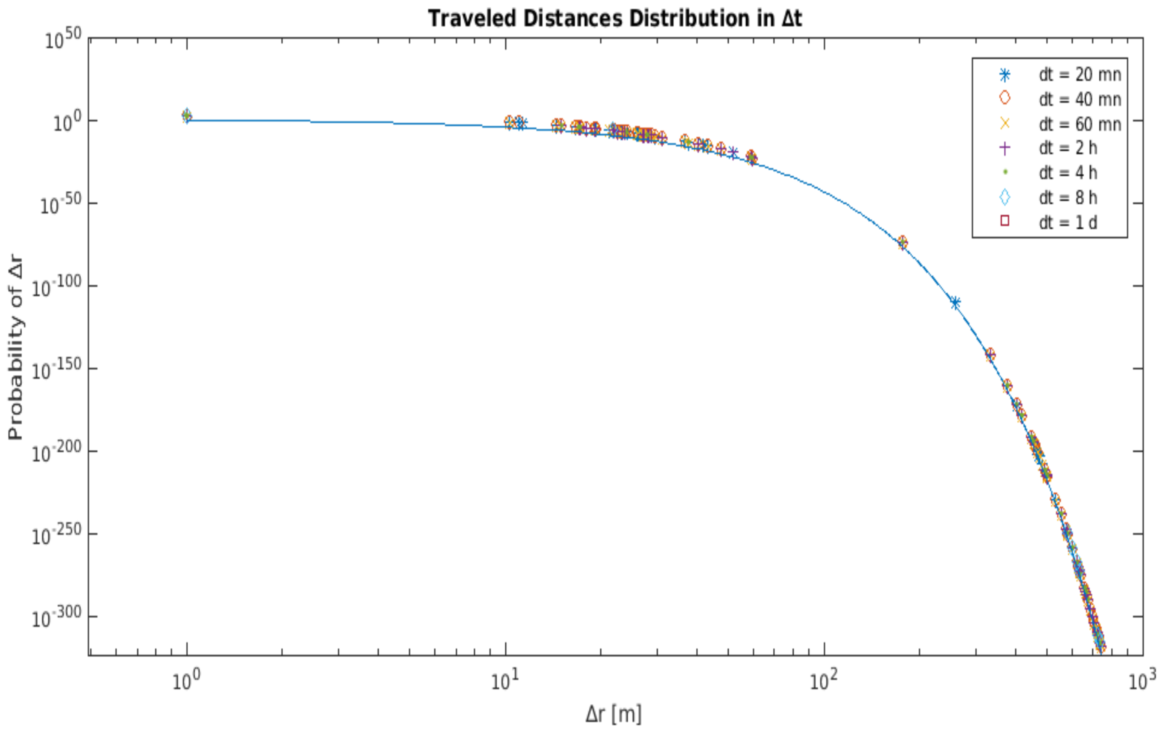


Figure 5.24 – Day 4 traveled distances by inter-event time.

Figure 5.25 – Day 4 radius of gyration histogram.

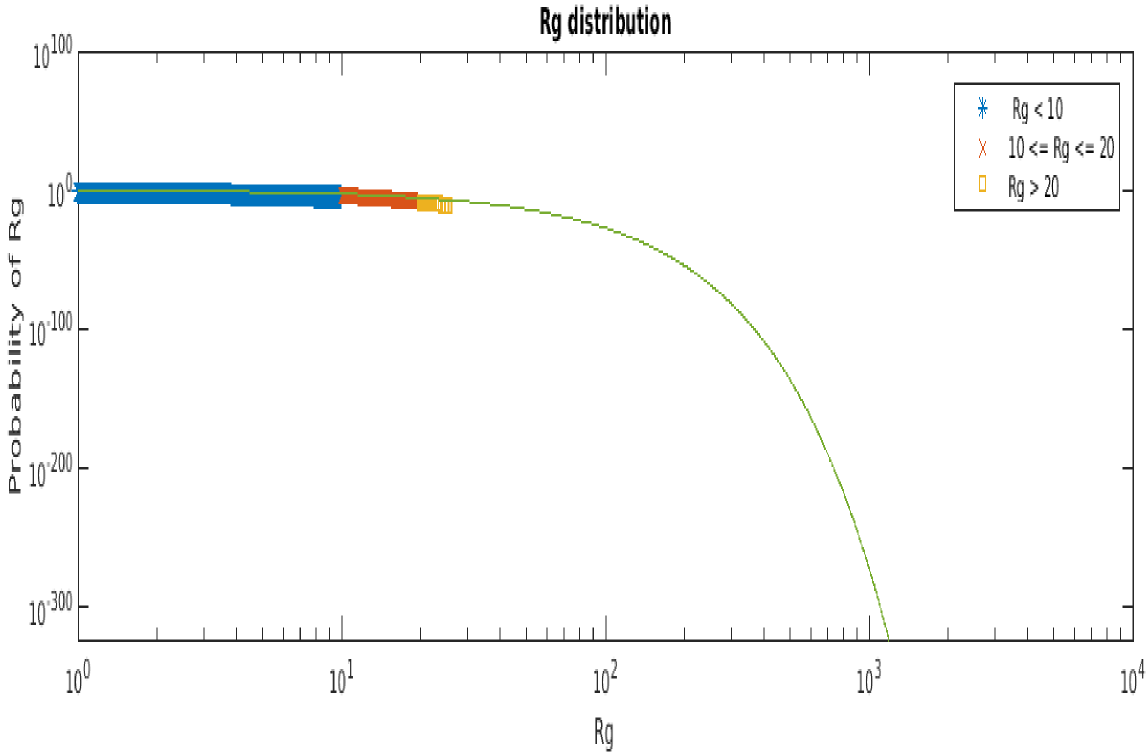


Figure 5.26 – Day 4 probability distribution for the radius of gyration.

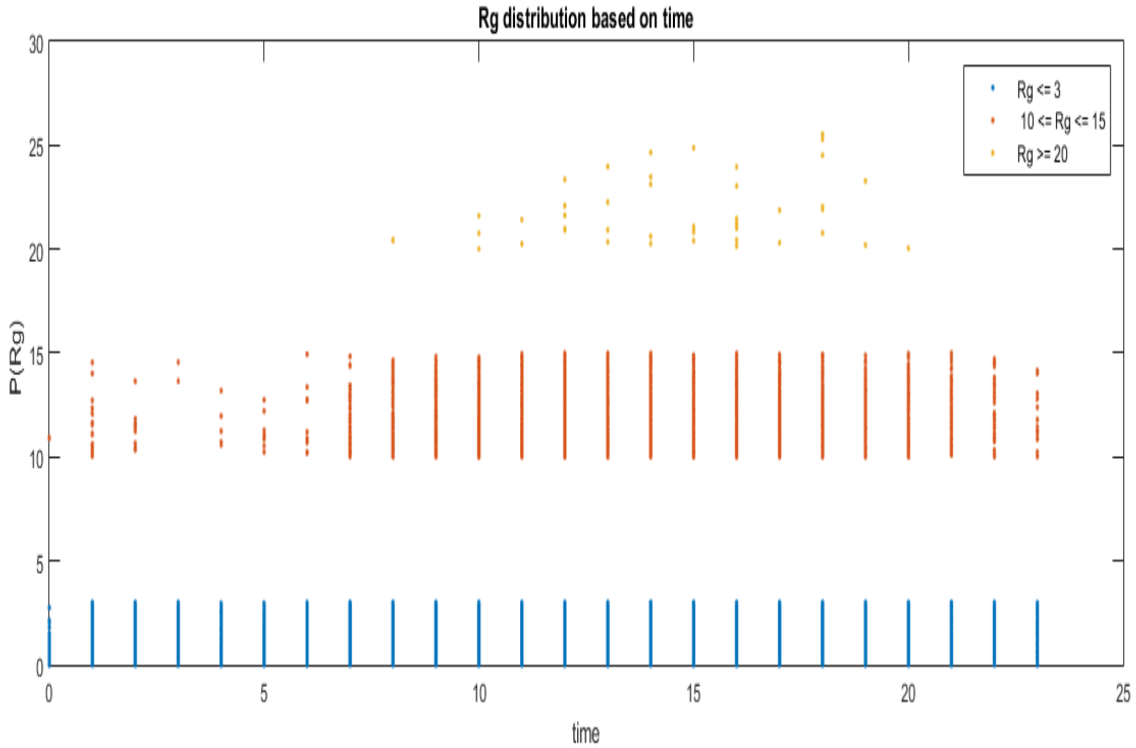


Figure 5.27 – Day 4 probability distribution for the radius of gyration based on time.

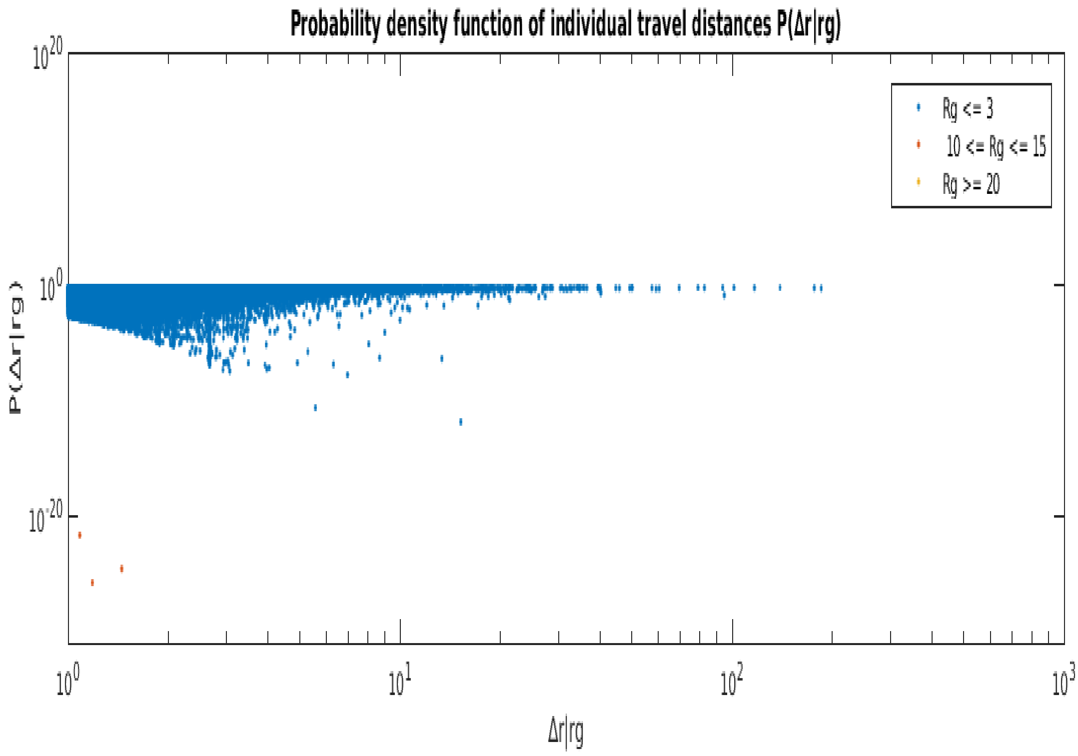


Figure 5.28 – Day 4 probability density function of individual travel distances.

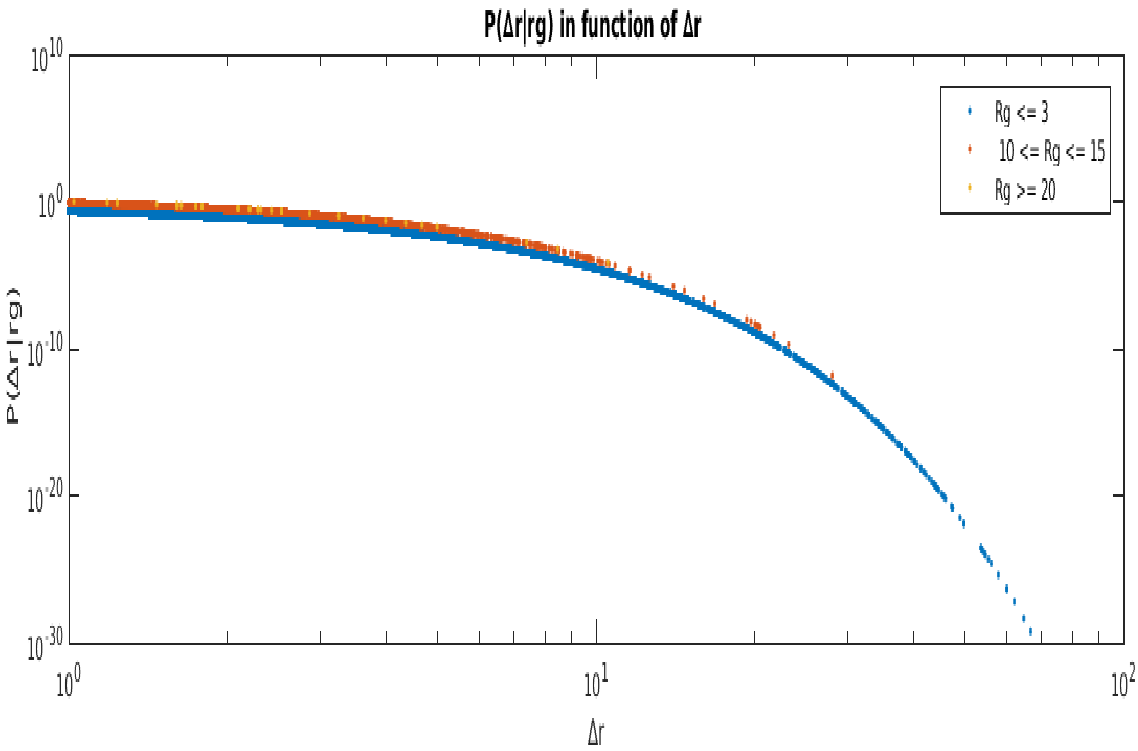


Figure 5.29 – Day 4 probability distribution of $\triangle r|\triangle g$.

Conclusions

The simulation aims to study and build the models in order to mimic the real system or any real-life phenomena either existing or contemplated inception one. The aim of this study is to explore the characteristics of human mobility, eventual real effects conditions, and actions of the specified system. This is achieved by modeling them mathematically.

This research aimed to understand human mobility using the CDRs. The investigation endorsed that mobile phone activities can reflects individuals density, also could act as a feature of grouping the total number of mobile network users, hence each group will have its own feature. the individuals density and dense variance between the sub-areas could be estimated in order to explore the tendency or the orientation of individuals in these areas, but almost in a static manner.

As well as, it is concluded that the most common parameters of modeling human mobility (inter-event time Δt, travel distance (displacement) Δr, and radius of gyration r_g) are represented by the power-law distribution, whereas, the most of real systems are almost followed an exponential distribution. These parameters can reveal mobility patterns of the evolved population. Also, this study confirmed that radius of gyration (r_g) is the most common quantity, which is associated with human mobility trajectories, due to its capability in measuring the how far the mass from the center of mass [Sah13].

The r_g gradually increases at the beginning, but it settles down versus time. It has a key effect on the travel distance distributions. The traveled distance distributions are collapsed or overlapped for groups, however the r_g is considered to be a more dedicated feature that capable of characterizing the travel distances (Δr_s) of individuals. The distributions showed that the individuals travel activities are almost identical, where the periodic trajectories are invariant. As well as, the activities distributions are uncovered the regular patterns and dynamic behaviors similarities during the time evolution of r_g. The experiments analyzed the relationships among the r_g, Δr and the r_{cm}, the experiments showed that almost individuals have similar activities, but these activities in general are classified into two variant types, which are: working days and holiday days.

Conclusions and Perspectives

Conclusions and Perspectives

The human mobility and communication networks draw the attention of the scientists in recent decades, where the studies are accomplished to understand the individuals mobility and analyze their life patterns via the communication data [Mar14b]. However, the common data were collected from the transportation surveys of long term mobility data, then it is sought-after for another efficient automatic datasets systems such as the CDRs [Tho11a, Zbi13, Ric09].

Nevertheless, using CDRs as an effective material in these studies have several problems such as the mass size of these data, so they need sampling [Tho11b, Cor13] and accurate manipulations to preserve their accuracy, and avoid any misuse or misunderstanding of their reality, by other hand the geographical representation of these datasets is difficult, because of the hardware/software limitations that could be encountered during the processing operations. However, CDRs datasets overcome the acquisition problems (financially/ time consuming), but their projection/ trajectories tracing still time consuming [Tho11a, Dar13, Mar13]. As well as, the privacy problem, since the information of the mobile network subscribers will be vulnerable, therefore it is anonymized (hidden) to protect the subscribers' privacy [San12].

Essentially, there are two data types, the communication data such as mobile phone data, and the communication independent data (Itinerancy data). The mobile phone datasets considered as the most powerful media to analyze and discover the human mobility either individually or collectively (rally) [Jul08a, Alb11]. This makes the mobile phone data are the highly elected tools to uncover and understand the human life patterns, as well as the anomalous events to the investigated data [Tho11a].

The communication activities (events density) are considered as individuals density. They characterize the observed population by grouping the total number of mobile network users, and each group will have its own feature. Nevertheless, the individuals density and dense variance among several observed regions could be calculated, in order to explore the tendency or the orientation of individuals in these regions, but almost in static manner.

The results of this research show that the inter-event time between each two successive activities has bursty pattern, since there is long period without activities, this gives an indication about the population heterogeneity. Also, the human trajectories follow a power-law distribution of travel distance (step size), and they are modeled using the displacements and waiting time distributions.
As well as, the mobility patterns are shown that there are many short distances in contrast to the few long distances. The mobile phone data is used to reconstruct the individuals trajectories. However, the patterns of all days are very similar, and they have approximately identical curves of spatio-temporal features.
In this research the mobility patterns are classified with regard to the work or off days for more understandings of the life patterns. Whereas, the radius of gyration would be considered as a significant modelling parameter to give the model more reality with focusing on patterns regularity features.

The radius of gyration is computed for all individuals to estimate the mobility patterns, which is followed the power-law, in this case it is endorses the research [4]. This law gives the indication of population heterogeneity, which characterized the patterns of the individuals trajectories. It is explored that there is a variance in individual mobility patterns, as well as the individuals trajectories are bounded by their radius of gyration, which is the most common quantity, which is associated with human mobility trajectories, due to its capability in measuring the how far the mass from the center of mass. It is appeared that gradually increases at the beginning, but it settles down versus time. It has a key effect on the travel distance distributions.

The traveled distance distributions are collapsed or overlapped for groups, however the radius of gyration is considered to be a more dedicated feature that capable of characterizing the travel distances of individuals.
The distributions showed that the individuals travel activities are almost identical, here the periodic trajectories are invariant, as well as the sedentary individuals have small radius of gyration , in opposite to the mobile individuals which have large radius of gyration.
These parameters can reveal life patterns of the evolved (dynamic) population, hence the dynamic perspectives of the population would be almost reflected. Also, the activities distributions are uncovered the regular patterns and behaviors similarities during the radius of gyration evolution in time.
The experiments analyzed the relationships between the activities during inter-event time, radius of gyration, and travel distance, and shows that all individuals have almost similar activities and grouped in common relevant. However, they can be classified generally into two variant classes, which are: working days and off (holiday) days.

It is recommended that further research be undertaken as future perspectives as in the following suggestions, perform more detailed mobility analyzing by extracting new parameters to find individuals behavior, use more statistical techniques to construct social networks via CDRs, analyze more CDRs data for the same region, but in other times (the region without Armada event), then after, compare the two cases studies to elaborate Armada effects on the city, make more analysis to find the hotspot regions (dens) in order to find the most attractive regions, develop more techniques to manipulate the big data (CDRs) in order to find and achieve more flexible tools for data analysis, compute city rhythm using the average displacements and average inter- event time, and the individual trajectory could be elaborated more to obtain his behaviour deeply, since human mobility open the door in front of human behavior.

Bibliography

Bibliography

[Abh03] Abhishek Roy, Soumya K. Das Bhaumik, Amiya Bhattacharya, Kalyan Basu, Diane J. Cook, Sajal.Das. Location Aware Resource Management in Smart Homes. Pervasive Computing and Communications, 2003. (PerCom 2003). Proceedings of the First IEEE International Conference, 2003.

[Alb11] Albert-Laszlo Barabasi. From Network Science to Human Dynamics. *The first of CEE's DES4:Distinguished Engineering and Science Speaker Seminar Series, Northeastern University, Harvard Medical School*, 2011.

[Alb13] Alberto Rubioa, Angel Sanchezb, Enrique Frias-Martinez. Adaptive Non-Parametric Identification of Dense Areas using Cell Phone Records for Urban Analysis. *Engineering Applications of Artificial Intelligence*, January 2013.

[Ale14] Alexander Amini, Kevin Kung, Chaogui Kang, Stanislav Sobolevsky, Carlo Ratti. The Impact of Social Segregation on Human Mobility in Developing and Urbanized Regions. *EPJ Data Science*, January 2014.

[And10] Andrea Vaccari, Mauro Martino, Francisca Rojas , Carlo Ratti. Pulse of the city: Visualizing Urban Dynamics of Special Events. *Association for Computing Machinery*, 2010.

[Ant05] Anthony LaMarca, Yatin Chawathe, Sunny Consolvo, Jeffrey Hightower, Ian Smith, James Scott,Tim Sohn, James Howard, Jeff Hughes, Fred Potter, JasonTabert, Pauline Powledge, Gaetano Borriello, Bill Schilit. Place Lab: Device Positioning Using Radio Beacons. the Wild. Proceedings of the 3rd International Conference on Pervasive Computing, 2005.

[Ara14] http://rouen-armada-2013.soft112.com/, April 2014.

[Arm07] Armando Bazzani, Massimiliano Capriotti, Bruno Giorgini, Giuseppina Melchiorre, Sandro Rambaldi, Graziano Servizi, Giorgio Turchetti. *The Dynamics of Complex Urban Systems An Interdisciplinary Approach*. Physica-Verlag Heidelberg, Laboratory of Fisica della Citt, University of Bologna, Italy, 2007.

[Arm10] Armando Bazzani, Bruno Giorgini, Sandro Rambaldi. Modeling Urban Mobility for E-Governance with Low Energy Complexity. *Green Engineering*, 2010.

[Arm14] http://www.armada.org/en/armada/historique/2008-l-armada-rouen-2008, April 2014.

[Aug05] Augustin chaintreau, Pan Hui, Jon Crowcroft, Christophe Diot, Richard Gass, James Scott. Pocket switched Networks: Real-World Mobility and its Consequences for Opportunistic Forwarding, technical report no.617, 2005.

[Bal13] Balzs Cs. Csjia, Arnaud Browetc, V.A. Traagc, Jean-Charles Delvennec, Etienne Huensc, Paul Van Doorenc, Zbigniew Smoredae, Vincent D. Blondelc. Exploring the Mobility of Mobile Phone Users. 2013.

[Bat99] Batty M., Jiang B. Multi-agent simulation: New approaches to exploring space-time dynamics within GIS, 1999.

[Beh14] Behadili Suhad Faisal, Cyrille Bertelle, Loay Edwar George. Visualization of People Attraction from Mobile Phone Trace Database: A Case study on Armada 2008 in French City of Rouen . Proceedings of the 1st International Engineering Conference On Developments in Civil & Computer Engineering Applications (IEC2014), Ishik University, Erbil, KRG, Iraq, November 2014.

[Blu99] Blue V.J., Adler J.L. . Cellular automata micro simulation of bidirectional pedestrian flows. *Transportation Research Board*, 1999.

[C H14] C Herrera-Yag, CM Schneider, Z Smoreda, T Couronn, PJ Zufiria, MC Gonzlez. The Elliptic Model for communication fluxes. 2014.

[Cal11] Calabrese F., Smoreda Z., Blondel V. D. , Ratti C. Interplay between Telecommunications and Face-to-Face Interactions: A Study Using Mobile Phone Data. *Plos One*, 2011.

[Cha10a] Chaoming Song, Tal Koren, Pu Wang, Albert-László Barabási. Modeling the Scaling Properties of Human Mobility. 2010.

[Cha10b] Chaoming Song, Zehui Qu, Nicholas Blumm, Albert-Laszlo Barabasi. Limits of Predictability in Human Mobility. *Science*, February 2010.

[Chr08] Christian Licoppe, Dana Diminescu, Zbigniew Smoreda, Cesary Ziemlicki. Using Mobile Phone Geolocation Data for the 'Socio-Geographical' Analysis of Patterns of Coordination, Urban Mobilities and Social Integration Patterns. December 2008.

[Chr10] Christian Saad, Bernhard Bauer. Data-Flow Based Model Analysis. *IRD*, 2010.

[Chr13a] Christian M. Schneider, Christian Rudloff, Dietmar Bauer, Marta C. Gonzlez. Daily Travel Behavior: Lessons from a Week-Long Survey for The Extraction of Human Mobility Motifs Related Information. UrbComp '13 Proceedings of the 2nd ACM SIGKDD International Workshop on Urban Computing, 2013.

[Chr13b] Christian M. Schneider, Vitaly Belik, Thomas Couronnè, Zbigniew Smoreda, Marta C. Gonzàlez. Unravelling Daily Human Mobility Motifs. 2013.

[Chu09] Chuanjun Ren, Chenghui Yang, Shiyao Jin. Complex Sciences. Springer Berlin Heidelberg, 2009.

[Cor13] Corina Iovan, Ana-Maria Olteanu-Raimond, Thomas Couronn, Zbigniew Smoreda. *Geographic Information Science at the Heart of Europe, Part of the series Lecture Notes in Geoinformation and Cartography pp 247-265*. Springer International Publishing Switzerland, 2013.

[CS315] CS302 Lecture notes. CS302 Lecture notes, Simulating Random Events. http://web.eecs.utk.edu/~leparker/Courses/CS302-Fall06/Notes/PQueues/random_num_gen.html, 2015.

[Daa04] Daamen Winnie. *Modelling passenger flows in public transport facilities*. PhD thesis, Delft University of Technology, Faculty of Civil Engineering and Geosciences, Department Transport and Planning, Netherlands, 2004.

[Daa14] http://www.armada.org/en/armada/historique/naissance-d-une-idee, April 2014.

[Dan09] Dana Diminescu, Christian Licoppe, Zbigniew Smoreda, Cezary Ziemlicki. *The Reconstruction of Space and Time through Mobile Communication Practices*. Transaction Publishers, 2009.

[Dan15] Dani Broitman, Eric Koomen. Residential density change: Densification and urban expansion. *Computers, Environment and Urban Systems*, November 2015.

[Dap15] Dapeng Li, Thomas J. Cova, Philip E. Dennison. A household-level approach to staging wildfire evacuation warnings using trigger modeling. *Computers, Environment and Urban Systems*, 2015.

[Dar13] Darakhshan J. Mir, Sibren Isaacmany, Ramon Caceres, Margaret Martonosi, Rebecca N. Wright. DP-WHERE: Differentially Private Modeling of Human Mobility. IEEE BigData, 2013.

[Dav96] David B. Johnson, David A. Maltz. Mobile Computing. chapter Dynamic Source Routing in Ad Hoc Wireless Networks, pages 153–181. Kluwer Academic Publishers, 1996.

[Den13] Denise Brehm. Study Shows that People Organize Daily Travel Efficiently. 2013. http://phys.org/news/2013-05-people-daily-efficiently.html.

[Die12] Dieter Van Cauteren, Nathalie Lucas. Surveillances sanitaire et environnementale de l'Armada de Rouen, édition 2008. http://opac.invs.sante.fr/doc_num.php?explnum_id=8115, April 2012.

[Duc12] Duc-An Vo, Alexis Drogoul, Jean-Daniel Zucker, Tuong-Vinh Ho. *A Modelling Language to Represent and Specify Emerging Structures in Agent-Based Model*. Springer, 2012.

[Eli11] Elissalde B., Lucchini F., Freire-Diaz S. Mesurer la ville èphèmère. *Donnèes Urbaines, 6, Pumain D., Mattei M-F., Anthropos, Economica*, 2011.

[Eli12] Elijah Meeks. Possibility and Probability in Geospatial Information Visualization, 2012.

[Enr13] Enrique Frias Martinez , Vanessa Frias Martinez. Method For An Automatic Identification Of Urban Dense Areas From Cell Phones Records. *United States Patent 8,577,380*, November 2013.

[Enr14] Enrique Frias Martinez, Vanessa Frias Martinez. Method and Computer Programs for The Construction of Communting Matrices using Call Detail Records and a use for Providing user's Mobility Information. *US 20140024389 A1*, 2014.

[Far13] Farnaz Fathali, Hatairatch Charoenkulvanich. Human-Mobility Modeling for Cellular Networks. 2013.

[Fil12] Filippo Simini, Marta C. González, Amos Maritan, Albert-László Barabási. A Universal Model for Mobility and Migration Patterns. *Nature*, 2012.

[Fos13] Fosca Giannotti, Luca Pappalardo, Dino Pedreschi, Dashun Wang. *Mobility Data Modeling, Management, and Understanding*. Cambridge University Press, 2013.

[Fra10] Francesco Calabrese, Francisco C Pereira, Giusy Di Lorenzo, Liang Liu, Carlo Ratti. Pervasive Computing. 2010.

[Fra12] Francesco Calabrese, Mi Diao, Giusy Di Lorenzo, Joseph Ferreira Jr., Carlo Ratti. Understanding Individual Mobility Patterns from Urban Sensing Data: A Mobile Phone Trace Example. *Transportation Research Part C: Emerging Technologies*, September 2012.

[Fru71] Fruin, John J. Designing for pedestrians: A level of service concept. 50th Annual Meeting of the Highway Research Board, Washington District of Columbia, United States, Highway Research Record 355, pp. 1-15, 1971.

[Geo12] Geoff Jacquez, Leah Estberg, Dunrie Greiling, Leah Estberg, Andrew Long. *ClusterSeer Software for The Detection and Analysis of Event Clusters, User Manual Book1, version 2.5*. BioMedware geospatial research and software, 2012.

[Gir08] Girardin F., Blat J., Calabrese F., Dal Fiore F., Ratti C. Digital Footprinting:Uncovering Tourists with User-Generated Content. *Pervasive Computing*, 2008.

[Gre08] Gregor Lammel, Marcel Rieser1, Kai Nagel. *Pedestrian and Evacuation Dynamics*. Springer-Verlag Berlin Heidelberg, 2008.

[G12] GSagl, Bernd Resch, Bartosz Hawelka, Euro Beinat. From Social Sensor Data to Collective Human Behaviour Patterns-Analysing and Visualising Spatio-Temporal Dynamics in Urban Environments. GI_Forum 2012: Geovizualisation, Society and Learning. © Herbert Wichmann Verlag, VDE VERLAG GMBH, Berlin/Offenbach, ISBN 978-3-87907-521-8., 2012.

[Gun94] Gunnar G. Lovas. Modelling and simulation of pedestrian traffic flow. *Transportation Research, Part B: Methodological 28 (6)*, 1994.

[Gun14] Gunther Sagl, Bernd Resch. *Encyclopedia of Mobile Phone Behavior*. IGI Global, 2014.

[Hé16] Héctor A. López-Ospina, Francisco J. Martínez b, Cristián E. Cortés. Microeconomic model of residential location incorporating life cycle and social expectations. *Computers, Environment and Urban Systems*, 2016.

[Hak08] Hakima Kadri-Dahmani, Grard H. E. Duchamp, Rawan Ghnemat, Hatem Hadj Kacem, Cyrille Bertelle. Emerging Decision Support Systems for Geographical Information Systems. 2008. http://www.researchgate.net/publication/228549847_EMERGING_DECISION_SUPPORT_SYSTEMS_FOR_GEOGRAPHICAL_INFORMATION_SYSTEMS.

[Han03] Hanisch A., Tolujew J., Richter K., Schulze T. . Online simulation of pedestrian flow in public buildings. Winter Simulation Conference, USA, 2003.

[Hei12] Heidi Durbeck, Dunrie Greiling, Leah Estberg, Andrew Long, Geoff Jacquez, Yanna Pallicaris, Susan Hinton. *ClusterSeer Software for The Detection and Analysis of Event Clusters, User Manual Book 2, version 2.5*. BioMedware geospatial research and software, 2012.

[Hel02] Helbing D., Farkas I.J., Molnar P., Vicsek T. Pedestrian and Evacuation Dynamics. 2002.

[Hen71] Henderson L. F. The statistics of crowd fluids. *Nature*, 1971.

[Hon13] Hong Van Truong, Elise Beck, Julie Dugdale, Carole Adam. Developing a Model of Evacuation after an Earthquake in Lebanon. ISCRAM Vietnam 2013 Conference,Institut de la Francophonie pour línformatique (IFI), Ha Noi, Vietnam, October 30th November 1st, 2013.

[Ily13] Ilya Narsky, Frank C. Porter. *Statistical Analysis Techniques in Particle Physics, Fits, Density Estimation and Supervised Learning*. Wiley-VCH, 2013.

[J. 05] J. D. Hunt, D. S. Kriger, E. J. Miller. Current operational urban land-use-transport modelling frameworks: A review. *Transport Reviews*, 2005.

[Jam09] James P. Bagrow, Tal Koren. Investigating Bimodal Clustering in Human Mobility. IEEE, International Conference on Computational Science and Engineering, 2009.

[Jam12] Jameson L. Toole, Michael Ulm, Marta C. Gonzlez, Dietmar Bauer. Inferring Land Use from Mobile Phone Activity. UrbComp'12 Proceedings of the ACM SIGKDD International Workshop on Urban Computing Pages 1-8, 2012.

[Jin04] Jing Su, Alvin Chin, Anna Popivanova, Ashvin Goel, Eyal de Lara. User Mobility For Opportunistic Ad-Hoc Networking. Proceedings of the 6th IEEE Workshop on Mobile Computing Systems and Applications, 2004.

[Joh15] John Steenbruggen, Emmanouil Tranos, Peter Nijkamp. Data from Mobile Phone Operators: A tool for Smarter Cities? *Elsevier, Telecommunications Policy*, 2015.

[Jul08a] Julian Candia, Marta C González, Pu Wang, Timothy Schoenharl, Greg Madey, Albert-Laszlo Barabasi. Uncovering Individual and Collective Human Dynamics from Mobile Phone Records. *Physics*, May 2008.

[Jul08b] Julian Candia, Marta C González,Pu Wang, Timothy Schoenharl, Greg Madey, Albert-Laszl Barabasi. Uncovering Individual and Collective Human Dynamics from Mobile Phone Records. *J. Phys. A: Math. Theor. 41 224015*, 2008.

[Koh13] Kohei Arai ,Tran Xuan Sang. Decision Making and Emergency Communication System in Rescue Simulation for People with Disabilities. *(IJARAI) International Journal of Advanced Research in Artificial Intelligence*, 2013.

[Lib04] Libo Song, David Kotz, Ravi Jain, Xiaoning He. Evaluating Location Predictors with extensive WiFi Mobility data. IEEE, 2004.

[Lin12] Lin Sun. *Enabling Pervasive Applications by Understanding Individual and Community Behaviors*. PhD thesis, Universite Pierre & Marie Curie -TELECOM SudParis, December 2012.

[Luc11] Lucchini F., Bernard Elissalde, Sylviano-Freire-Diaz S. From the Event to the Ephemeral City. mesm 2011 conference, Amman, Jordan, Nov. 14-16, 2011, pp 88-95, 2011.

[Luc13] Lucchini F., Elisssald B, Freiré-Diaz S. Caractȉriser l Áttractivitè des Quartiers Urbains par les Donnèes de la T l'phonie Mobile. *Information Gògraphique*, 2013.

[Mad14] http://www.thefreelibrary.com/NORMAN+ARMADA%3B+TOM+Bodden+at+the+Tall+Ships+festival+in+Rouen, April 2014.

[Mar05] Marvin MeNett, Geoffrey M. Volker. Access Mobility of Wireless PDA users. *ACM SIGMobile Computing and Communications Review*, April 2005.

[Mar08] Marta C. González, Cesar A. Hidalgo, Albert-Laszlo Barabasi. Understanding Individual Human Mobility Patterns Supplementary Material. *nature*, 2008.

[Mar09] Marta C. Gonzlez, Cesar A. Hidalgo, Albert-Laszlo Barabasi. Understanding individual Human Mobility Patterns. *nature*, 2009.

[Mar10a] Marcos R. Vieira, Enrique Frias-Martinez, Petko Bakalov , Vanessa Frias-Martinez, Vassilis J. Tsotras. Querying Spatio-Temporal Patterns in Mobile Phone-Call Databases. 2010 Eleventh International Conference on Mobile Data Management, 2010.

[Mar10b] Marcos R. Vieira, Vanessa Frias-Martinez, Nuria Oliver, Enrique Frias-Martinez. Characterizing Dense Urban Areas from Mobile Phone-Call Data: Discovery and Social Dynamics. Social Computing (SocialCom), 2010 IEEE Second InternationalSocial Computing (SocialCom), IEEE Second International Conference, 2010.

[Mar13] Martin Wirz, Tobias Franke, Daniel Roggen, Eve Mitleton-Kelly, Paul Lukowicz, Gerhard Tröster. Probing Crowd Density through Smartphones in City-Scale Mass Gatherings. *EPJ Data Science*, December 2013.

[Mar14a] Mario Boffi, Matteo Colleoni. Human behaviour and GIS Methodologies for analysing urban mobility in a GIS environment. *Networks and Communication Studies, NETCOM*, 2014.

[Mar14b] Markus Schl, Luis M. A. Bettencourt, Sebastian Grauwin, Mathias Raschke, Rob Claxton, Zbigniew Smoreda, Geoffrey B. West, Carlo Ratti. The Scaling of Human Interactions with City Size. July 2014.

[Mic] Michele Tizzoni, Paolo Bajardi, Adeline Decuyper, Guillaume Kon Kam King, Christian M. Schneider, Vincent Blondel, Zbigniew Smoreda, Marta C. González, Vittoria Colizza.

[Mic94] Michael Batty, Paul Longley. *Fractal Cities: A Geometry of Form and Function*. Academic Press Limited INC., 1994.

[Mic07] Michael Batty. *Cities and complexity: understanding cities with cellular automata, agent-based models, and fractals*. The MIT press, 2007.

[Mic11] Michel Nabaa. *Morphodynamique De Reseaux Viaires Application Au Risque*. PhD thesis, Le Havre University, 2011.

[Mic12] Michael Batty. *Agent-Based Models of Geographical Systems*. Springer, Science Business Media B.V., 2012.

[Min51] Mintz A. Non-adaptive group behaviour. *Abnormal and Social Psychology*, 1951.

[Min06] Minkyong K, David K, Songkuk K. Extracting a Mobility Model from Real User Traces. IEEE INFOCOM'06, April 2006.

[Nir08] Nirajan Shiwakoti, Majid Sarvi, Geoff Rose. Modelling Pedestrian Behaviour Under Emergency Conditions State-of-the-Art and Future Directions. 2008.

[Oll13] Olle J, Rein Ahasa, Frank Witloxb. Understanding Monthly Variability in Human Activity Spaces: A Twelve-Month Study Using Mobile Phone Call Detail Records. 2013.

[Org16] http://www.orange.com/en/home, September 2016.

[P. 11] P. Roy, A.J. Martínez, G. Miscione, M.H.P. Zuidgeest, M.F.A.M. van Maarseveen. Using Social Network Analysis to Profile People Based on their E-Communication and Travel Balance. *Transport Geography*, 2011.

[Pab15] Pablo Barreira-González, Montserrat Gómez-Delgado, Francisco Aguilera-Benavente. From raster to vector cellular automata models: A new approach to simulate urban growth with the help of graph theory. *Computers, Environment and Urban Systems*, 2015.

[Pat] Patrick Taillandier1, Alexis Drogoul. From GIS Data to GIS Agents, Modeling with The GAMA Simulation Platform. http://www.irit.fr/TFGSIM/DOCS/PatrickTaillandier_GAMASimulationPlatform.pdf.

[Pat12a] Patrick Taillandier, Alexis Drogoul. From Grid Environment to Geographic Vector Agents, Modeling with The GAMA Simulation Platform. 2012.

[Pat12b] Patrick Taillandier,Duc-An Vo, Edouard Amouroux,Alexis Drogoul. *Principles and Practice of Multi-Agent Systems*. Springer Berlin Heidelberg, 2012.

[Pat12c] Patrick Taillandier,Vo Duc An, Edouard Amouroux,Alexis Drogoul. GAMA: bringing GIS and Multi-Level Capabilities to Multi-Agent Simulation. 2012.

[Pat13] Patrik Dahlstrom. *Geo-based Mobility Control for Mobile Traffic Simulators*. PhD thesis, Doctoral Thesis in Geodesy and Geoinformatics with Specialization in Geoinformatics Royal Institute of Technology Stockholm, Sweden, September 2013.

[Phi10] Phithakkitnukoon S., Horanont T., Di Lorenzo G., Shibasaki R., Ratti C. Activity Aware Map: Identifying Human Daily Activity Pattern Using Mobile Phone Data. *Human Behavior Understanding*, 2010.

[Pu 09a] Pu Wang, Marta C. González. Understanding Spatial Connectivity of Individuals with Non-Uniform Population Density. *Philos Trans A Math Phys Eng Sci.*, 2009.

[Pu 09b] Pu Wang, Marta C. González, Cesar A. Hidalgo, Albert-Laszlo Barabasi. Understanding the Spreading Patterns of Mobile Phone Viruses, Supporting Online Material. *Science Express*, 2009. http://mobilephonevirus.barabasilab.com/paper/som.pdf.

[Qua57] Quarantelli E. The behaviour of panic participants. *Sociology and Social Research*, 1957.

[Rat10] Ratti C., Sobolevsky S., Calabres F., Andris C., Reades J., Martino M., Claxton R., Strogatz S. H. Redrawing the Map of Great Britain from a Network of Human Interactions. *PLoS One*, 2010.

[Raw08a] Rawan Ghnemat, Cyrille Bertelle, Grard H.E. Duchamp. Adaptive Spatial System Emergence from Community Swarm Optimization. 2008. https://hal.archives-ouvertes.fr/hal-00431233/document.

[Raw08b] Rawan Ghnemat, Cyrille Bertelle, Grard H.E. Duchamp. Agent-Based Modeling Using Swarm Intelligence in Geographical Information Systems. Innovations'08, Al-Ain:United Arab Emirates 2008, 2008.

[Raw09a] Rawan Ghnemat, Cyrille Bertelle, G'erard H.E. Duchamp. A Methodology for Urban and Land-Use,Management Simulation Using Spatial Self-Organization Processes. *Dynamics of Continuous,Discrete and Impulsive Systems-series B16,4,2009,501-513*, 2009.

[Raw09b] Rawan Ghnemat, Cyrille Bertelle, Gérard H.E. Duchamp. *Complex Systems and Self-Organization Modelling*. Springer, 2009.

[Raw09c] Rawan Ghnemat, Cyrille Bertelle, Gérard H.E. Duchamp. Swarm Intelligence Engineering For Spatial Organization Modelling. 2009. https://hal.archives-ouvertes.fr/hal-00430562/document.

[Raw09d] Rawan Ghnemat, Cyrille Bertelle, Grard H.E. Duchamp. Urban Dynamics Modelling Using Ant Nest Building. 2009. https://hal.archives-ouvertes.fr/hal-00431231/document.

[Raw09e] Rawan Ghnemat, Françoise Lucchini, Cyrille Bertelle. Urban Cultural Dynamics Modelling Using Swarm Intelligence and Geographical Information Systems. 2009. https://hal.archives-ouvertes.fr/hal-00430566.

[Raw09f] Rawan M.O. GHNEMAT. *Adaptive Modeling for Spatial Emergence within Complex Systems*. PhD thesis, University of Le Havre , LITIS, 2009. http://hal.archives_ouvertes.fr/docs/00/43/12/33/PDF/ghnemat4sias2008.pdf.

[Rea09] Reades J., Calabrese F., Ratti C. Eigenplaces: analysing cities using the space-time structure of the mobile phone network. *Environment and Planning B: Planning and Design*, 2009.

[Rei10] Rein A, Margus T, Erki S, Christophe D. Mobile Telephones and Mobile Positioning Data as Source for Statistics: Estonian Experiences. *Department of Geography, University of Tartu*, 2010.

[Ric09] Rich Ling, Scott W. Campbell. *The Reconstruction of Space and Time through Mobile Communication Practices, Introduction Chapter*. Transaction Publishers, 2009.

[Roa14] http://www.meretmarine.com/fr/content/le-port-de-rouen-se-prepare-accueillir-larmada, April 2014.

[Rob15] Roberto Trasarti, Ana-Maria Olteanu-Raimond, Mirco Nanni, Thomas Couronne, Barbara Furletti, Fosca Giannotti, Zbigniew Smoreda, Cezary Ziemlicki. Discovering urban and Country Dynamics from Mobile Phone Data with Spatial Correlation Patterns. *Telecommunications Policy archive*, 2015.

[Rya06] Ryan Bravo. Research technical paper. *summer research*, August 2006.

[Sah13] Sahar Hoteit, Stefano Secci, Stanislav Sobolevsky, Guy Pujolle, Carlo Ratti. Estimating Real Human Trajectories through Mobile Phone Data. Mobile Data Management (MDM), IEEE 14th International Conference, Volume 2, Pages 148-153, IEEE, 2013.

[Sah14] Sahar Hoteit, Stefano Secci, Stanislav Sobolevsky, Carlo Ratti, Guy Pujolle. Estimating Human Trajectories and Hotspots through Mobile Phone Data. *Computer Networks*, 2014.

[San12] Santi Phithakkitnukoon, Zbigniew Smoreda, Patrick Olivier. Socio-Geography of Human Mobility: A Study Using Longitudinal Mobile Phone Data. July 2012.

[Sch02] Schadschneider A. Cellular automaton approach to pedestrian dynamics-theory. 2002.

[Seo16] http://www.srh.noaa.gov/bmx/?n=gis, June 2016.

[Sev10] Sevtsuk A., Ratti C. Does Urban Mobility Have A Daily Routine? Learning From the Aggregate Data Of Mobile Networks. *Urban Technology*, 2010.

[S.F09] S.F. Shaukat, M.I. Ansari, R. Farooq, U. Ibrahim, Muhammad Faisal. Mobile Phone Location Determination in Urban and Rural Areas Using Enhanced Observed Time Difference Technique. *IDOSI Publications*, 2009.

[Sim95] Sime J.D. Crowd psychology and engineering. *Safety Science*, 1995.

[Siq11] Siqi Shen, Niels Brouwers, Alexandru Iosup. Human Mobility in Virtual and Real Worlds:Characterization, Modeling, and Implications, November 2011.

[Sou14] Sougata Biswas. Simulation of Exponential Distribution using R. https://rstudio-pubs-static.s3.amazonaws.com/26693_e1151035722942b2813c0063c6b220ae.html, August 2014.

[Sta15a] Stat. Chapter 10, Simulation: An Introduction. http://www.stat.nus.edu.sg/~zhangjt/teaching/ST2137/lecture/lec%2010.pdf, 2015.

[Sta15b] Stat414. Probability Theory and Mathematical Statistics, STAT 414/415. https://onlinecourses.science.psu.edu/stat414/node/97, April 2015.

[Sti00] Still G. K. *Crowd Dynamics*. PhD thesis, University of Warwick, UK, 2000.

[Suh16a] Suhad Faisal Behadili, Cyrille Bertelle, Loay E. George. Adaptive Modeling of Urban Dynamics during Armada Event using CDRs. *International Journal of Information Technology and Computer Science(IJITCS)*, 2016.

[Suh16b] Suhad Faisal Behadili, Cyrille Bertelle, Loay E. George. Adaptive Modeling of Urban Dynamics during Ephemeral Event via Mobile Phone Traces. *Informatics Engineering, an International Journal (IEIJ)*, June 2016.

[Suh16c] Suhad Faisal Behadili, Cyrille Bertelle, Loay E. George. Human Mobility Patterns Modelling using CDRs. *International Journal of UbiComp (IJU)*, 2016.

[Suh16d] Suhad Faisal Behadili, Cyrille Bertelle, Loay E. George. Human Trajectories Characteristics. International Conference on Urban Planning, Transport and Construction Engineering (ICUPTCE'16), Pattaya (Thailand), January 2016.

[Suh16e] Suhad Faisal Behadili, Cyrille Bertelle, Loay E. George. Modeling of City Pulse using Radius of Gyration via Mobile Data. 2nd International Conference on Communication Systems and Computing Application Science (CSCAS2016), Jeju Island, South Korea, March 2016.

[Suh16f] Suhad Faisal Behadili, Cyrille Bertelle, Loay E. George. Modelling Dynamic Patterns using Mobile Data. Sixth International Conference on Computer Science, Engineering and Applications (CCSEA 2016), Dubai, UAE, January 2016.

[Tek02] Teknomo K. . *Microscopic pedestrian flow characteristics: Development of an image processing data collection and simulation model*. PhD thesis, Tohoku University, Japan, 2002.

[Tho11a] Thomas Couronn, Zbigniew Smoreda, Ana-Maria Olteanu. Individual Mobility and Communication Patterns, Analysis of Mobile Phone Datasets and Networks. NETMOB conference, MIT, Cambridge, MA, 2011.

[Tho11b] Thomas Couronné, Zbigniew Smoreda, Ana-Maria Olteanu. Chatty Mobiles: Individual Mobility and Communication Patterns. Analysis of Mobile Phone Datasets and Networks, MIT, Cambridge, MA, 2011.

[Tho13] Thomas Couronné, Valery Kirzhner, Katerina Korenblat, Zeev Volkovich. Some Features of the Users' Activities in the Mobile Telephone Network. August 2013.

[Tol04] Tolujew J., Alcal, F. . A mesoscopic approach to modelling and simulation of pedestrian traffic flows. 18þEuropean Simulation Multi-conference, Magdeburg, Germany, 2004.

[Uoa14] http://normandinamik.cci.fr/11482-armada-2008-40-grands-voiliers-deja-attendus, April 2014.

[V. 11] V. Soto, V. Frias-Martinez, J. Virseda and E. Frias-Martinez. Prediction of Socioeconomic Levels using Cell Phone Records. 19th International Conference, UMAP, Girona, Spain, July 2011.

[Van10] Vanessa Frias-Martinez, Enrique Frias-Martinez, Nuria Oliver. A Gender-Centric Analysis of Calling Behavior in a Developing Economy Using Call Detail Records. Association for the Advancement of Artificial Intelligence (www.aaai.org), 2010.

[Van12a] Vanessa Frias-Martinez, Alberto Rubio, Enrique Frias-Martinez. Measuring the Impact of Epidemic Alerts on Human Mobility using Cell-Phone Network Data. Second Workshop on Pervasive Urban Applications Pervasive1́2, Newcastle, UK, 2012.

[Van12b] Vanessa Frias-Martinez, Cristina Soguero, Enrique Frias-Martinez. Estimation of Urban Commuting Patterns Using Cellphone Network Data. UrbComp1́2 Proceedings of the ACM SIGKDD International Workshop on Urban Computing Pages 9-16, 2012.

[Van12c] Vanessa Frias-Martinez, Jesus Virsesa. On The Relationship Between Socio-Economic Factors and Cell Phone Usage. ICTD '12 Proceedings of the Fifth International Conference on Information and Communication Technologies and Development, Atlanta, GA, USA, Pages 76-84, March 2012.

[Van13] Vanessa Frias-Martinez, Jesus Virseda,Alberto Rubio, Enrique Frias-Martinez. Towards Large Scale Technology Impact Analyses: Automatic Residential Localization from Mobile Phone-Call Data. Proceedings of ICTD 2010. 4th ACM/IEEE International Conference on Information and Communication Technologies and Development. London: IEEE, May 2013.

[Vin15] Vincent Zoonekynd. Probability Distributions. http://zoonek2.free.fr/UNIX/48_R/07.html, April 2015.

[Wik14] https://commons.wikimedia.org/wiki/File:Complex_adaptive_system.svg, July 2014.

[Xia06] Xiaoshan Pan. *Computational modelling of human and social behaviours for emergencyegress analyses*. PhD thesis, Stanford University, U.K., 2006.

[Xia10] Xiao-Yong Yan, Xiao-Pu Han, Tao Zhou, Bing-Hong Wang. Exact Solution of Gyration Radius of Individual's Trajectory for a Simplified Human Mobility Model, 2010.

[Xia12] Xiaolong LI, Gang PAN, Zhaohui WU, Guande QI, Shijian LI, Daqing ZHANG, Wangsheng ZHANG, Zonghui WANG. Prediction of Urban Human Mobility Using Large-Scale Traces and its Applications. *Frontiers of Computer Science in China*, February 2012.

[Xia15] Xiaolu Zhoua, Chen Xub, Brandon Kimmonsc. Detecting tourism destinations using scalable geospatial analysis based on cloud computing platform. *Computers, Environment and Urban Systems*, 2015.

[YAN11] YAN Xiao-Yong, HAN Xiao-Pu, ZHOU Tao, WANG Bing-Hong. Exact Solution of the Gyration Radius of an Individual's Trajectory for a Simplified Human Regular Mobility Model. *Chinese Physical Society and IOP Publishing Ltd*, 2011.

[Yih11] Yihong Yuana, Martin Raubala. *Geographic Information Science, Lecture Notes in Computer Science, Volume 7478, 2012, pp 354-367*. Springer Berlin Heidelberg, 2011.

[Yik13] Yikang Rui. *Urban Growth Modeling Based on Land-use Changes and Road Network Expansion*. PhD thesis, Doctoral Thesis in Geodesy and Geoinformatics with Specialization in Geoinformatics Royal Institute of Technology Stockholm, Sweden, 2013.

[Yu-05] Yu-Chung Chengyz Yatin Chawathey Anthony LaMarcay John Krumm. Accuracy Characterization for Metropolitan-scaleWi-Fi Localization. *Microsoft Corporation*, January 2005.

[Yua13] Yuan, Yihong, Raubal Martin. *Characterizing Human Mobility from Mobile Phone Usage*. PhD thesis, University of California, Santa Barbara, 2013.

[Zah15] Zahedeh Izakiana, Mohammad Saadi Mesgaria, Ajith Abrahamb. Automated clustering of trajectory data using a particle swarm optimization. *Computers, Environment and Urban Systems*, 2015.

[Zbi13] Zbigniew Smoreda, Ana-Maria Olteanu-Raimond, Thomas Couronnè. *Transport Survey Methods: Best Practice for Decision Making*. Springer, Emerald, 2013.

Appendix A

List of Publications

International Conferences

[1] Suhad Faisal Behadili, Cyrille Bertelle, Loay E. George, Visualization of People Attraction from Mobile Phone Trace Database: A Case study on Armada 2008 in French City of Rouen, Proceedings of the 1st International Engineering Conference On Developments in Civil & Computer Engineering Applications (IEC2014), November 24-26, 2014, Ishik University, Erbil, KRG, Iraq.

[2] Suhad Faisal Behadili, Cyrille Bertelle, Loay E. George, Human Trajectories Characteristics, Proceedings of International Conference on Urban Planning, Transport and Construction Engineering (ICUPTCE'16), Jan. 2-3, 2016, Pattaya (Thailand).

[3] Suhad Faisal Behadili, Cyrille Bertelle, Loay E. George, Modelling Dynamic Patterns using Mobile Data, Proceedings of Sixth International Conference on Computer Science, Engineering and Applications (CCSEA 2016), January 23-24, 2016, Dubai, UAE.

[4] Suhad Faisal Behadili, Cyrille Bertelle, Loay E. George, Modeling of City Pulse using Radius of Gyration via Mobile Data, Proceedings of 2nd International Conference on Communication Systems and Computing Application Science (CSCAS2016), March 19-20, 2016, Jeju Island, South Korea.

International Journals

[1] Suhad Faisal Behadili, Cyrille Bertelle, Loay E. George, Human Mobility Patterns Modelling using CDRs, International Journal of UbiComp (IJU), Vol.7, No.1, January 2016, DOI:10.5121/ iju.2016.7102.

[2] Suhad Faisal Behadili, Cyrille Bertelle, Loay E. George, Adaptive Modeling of Urban Dynamics during Ephemeral Event via Mobile Phone Traces, Informatics Engineering, an International Journal (IEIJ), Vol.4, No.2, June 2016, DOI : 10.5121/ieij.2016.4204, ISSN: 2349-2198.

[3] Suhad Faisal Behadili, Cyrille Bertelle, Loay E. George, Adaptive Modeling of Urban Dynamics during Armada Event using CDRs, International Journal of Information Technology and Computer Science(IJITCS), Global Impact Factor 0.716, 2016.

Appendix B

Appendix of Figures

This appendix have modeling and simulating results, which are about the waiting time, travel distance, and radius of gyration distributions. They are selected for some particular days. According to the previous classification in chapter 5, hence the days are in two main classes the work days and off days. The selection is pivoted on the highest and lowest activities of each class of observed days. Where, the day 5 and day 7 represent the highest activities number of the work and off days respectively. Whereas, the day 14 and day 15 represent the lowest activities number of the off and work days respectively. The results are in the following figures.

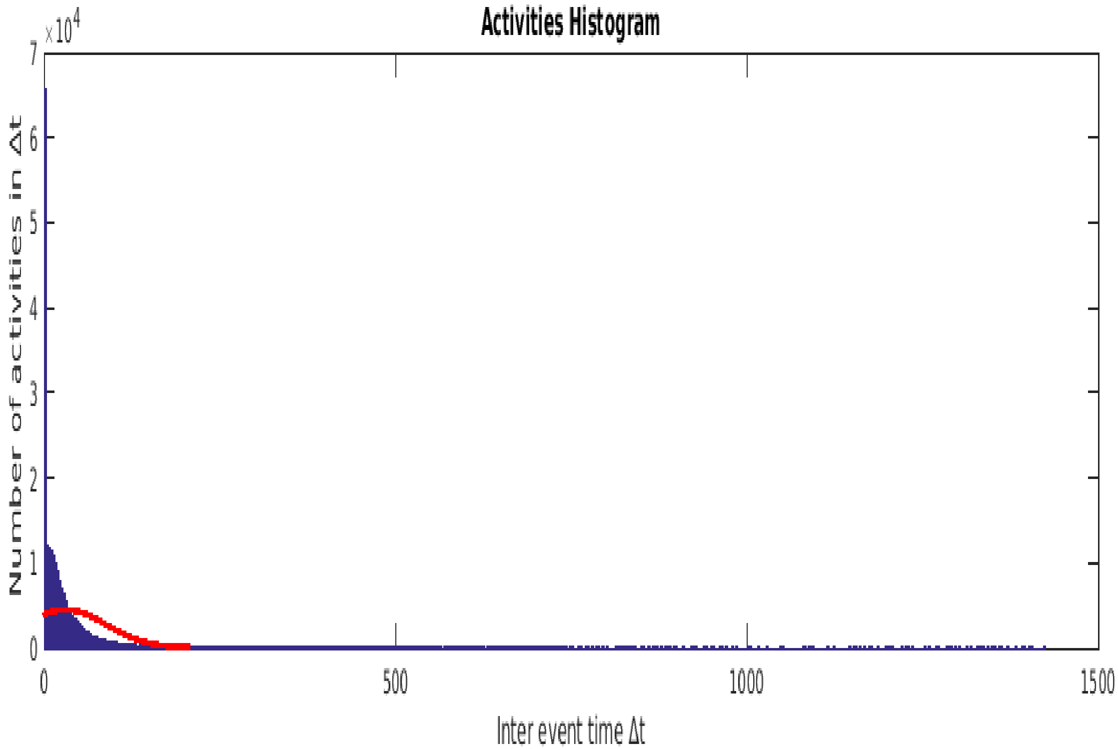


Figure B.1 – Day 5 activities histogram.

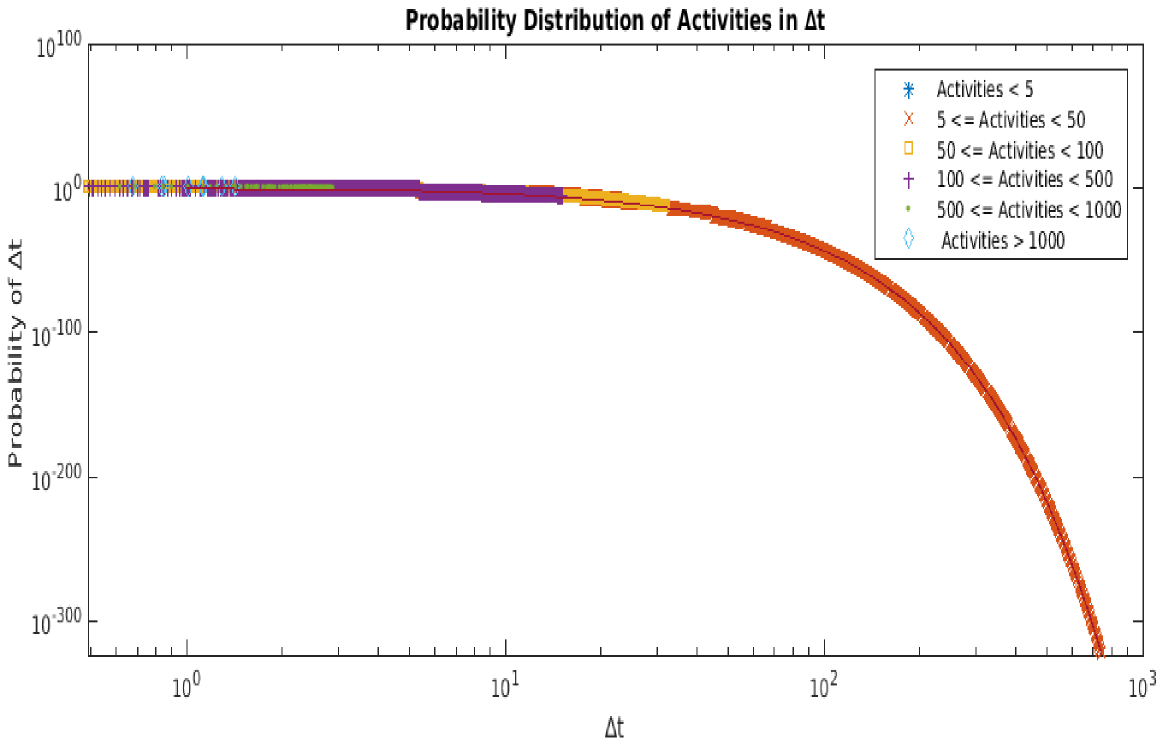


Figure B.2 – Day 5 activities by inter-event time.

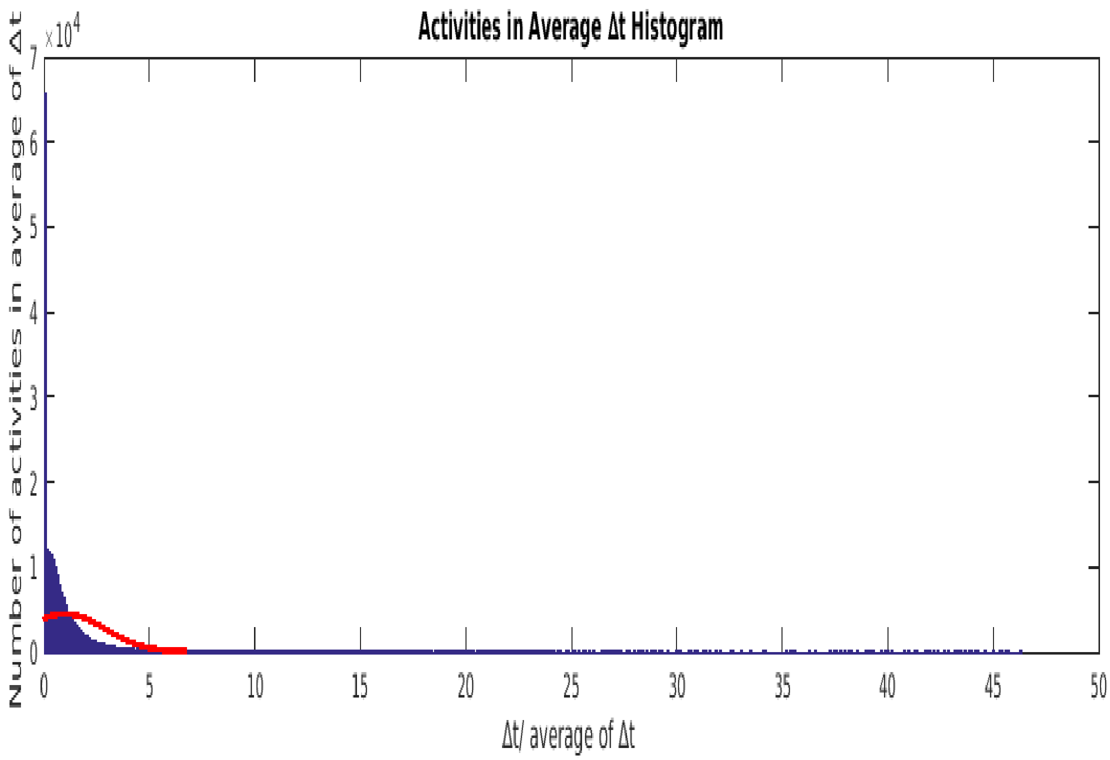


Figure B.3 – Day 5 activities in average time histogram.

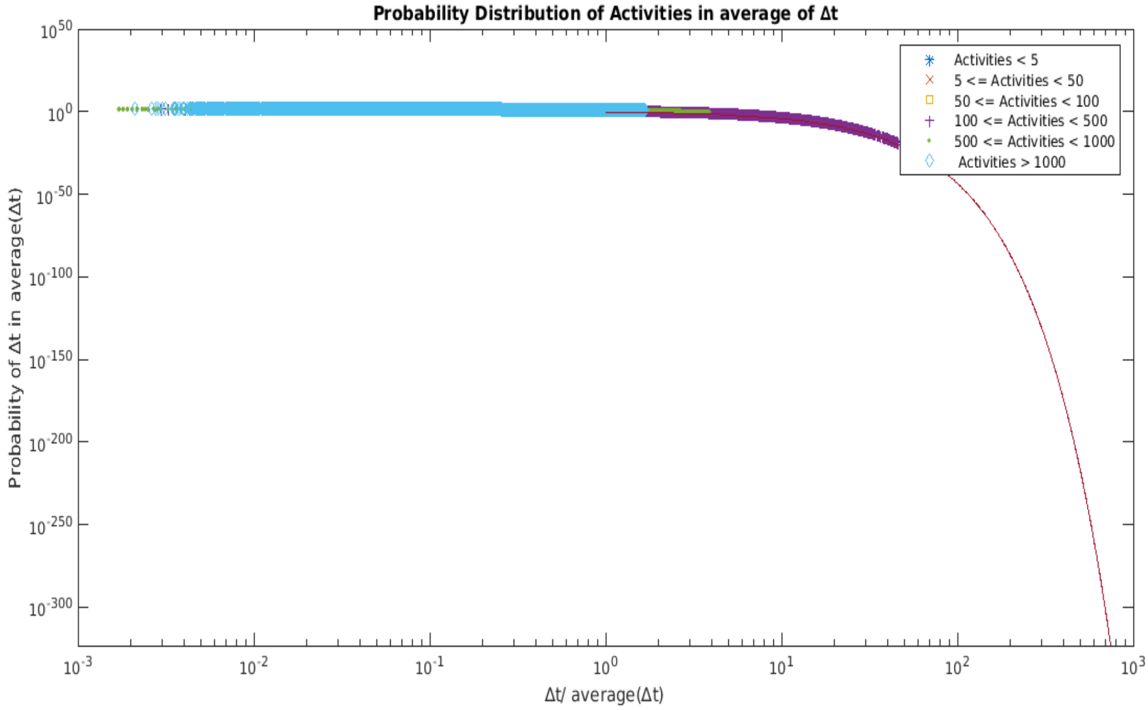


Figure B.4 – Day 5 activities by average inter-event time.

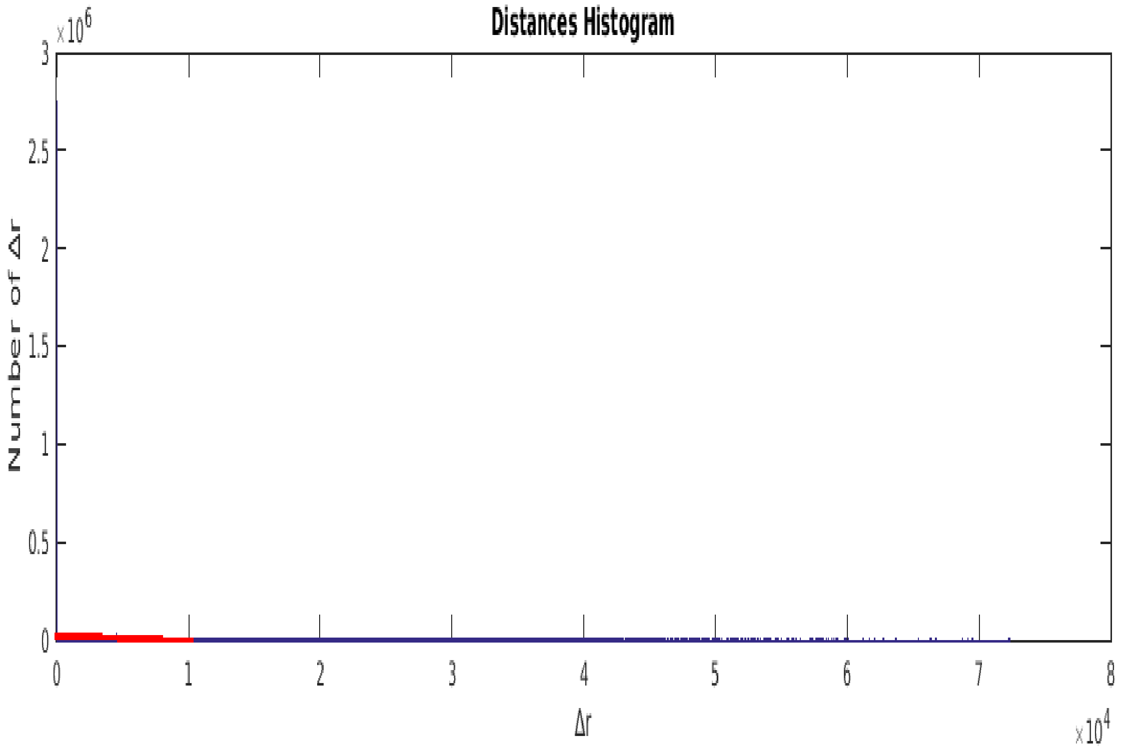


Figure B.5 – Day 5 distances histogram.

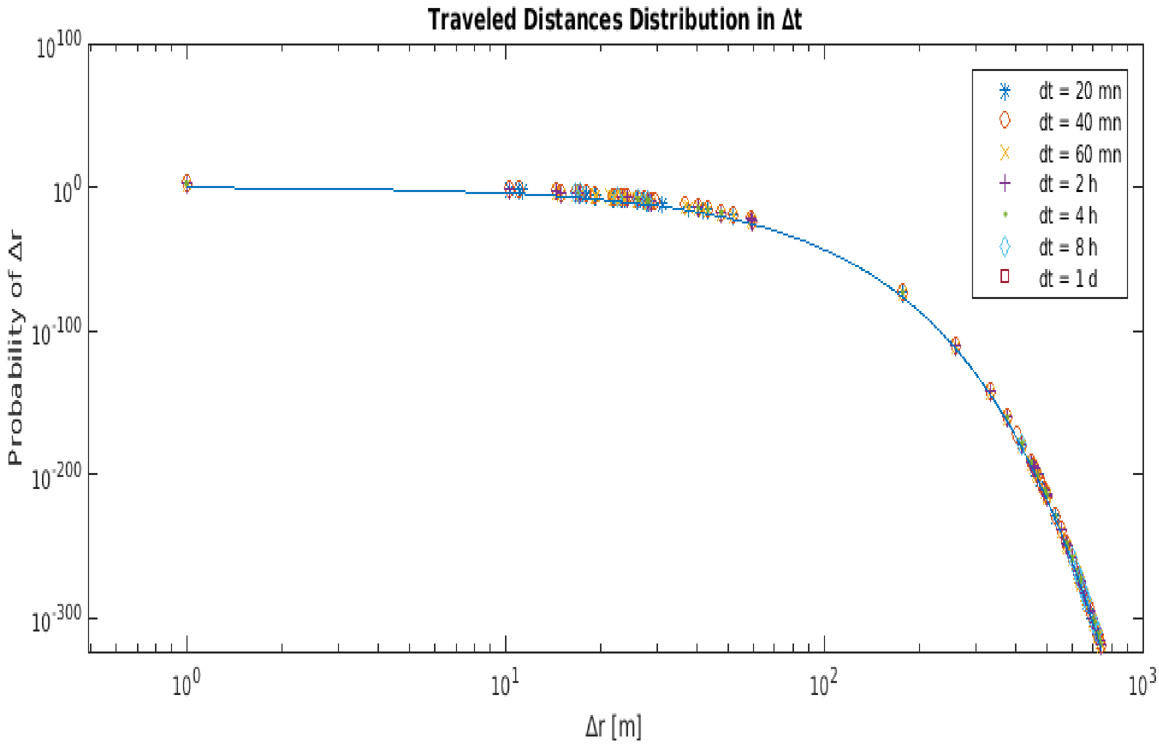


Figure B.6 – Day 5 traveled distances by inter-event time.

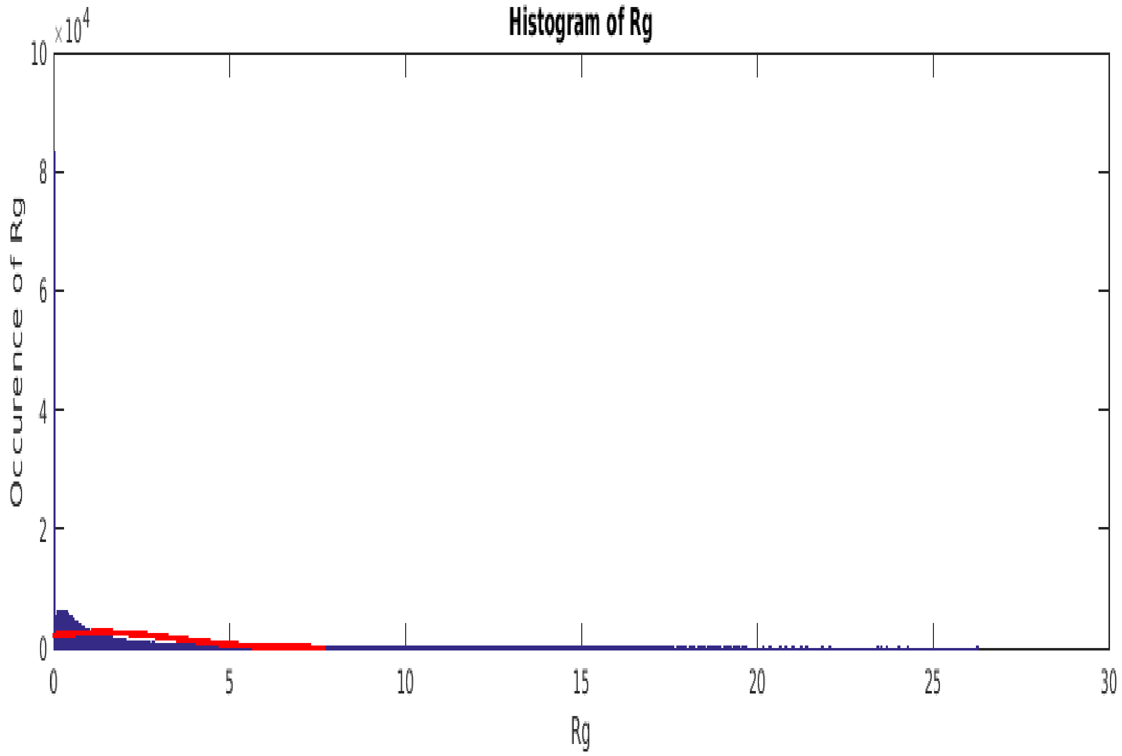


Figure B.7 – Day 5 radius of gyration histogram.

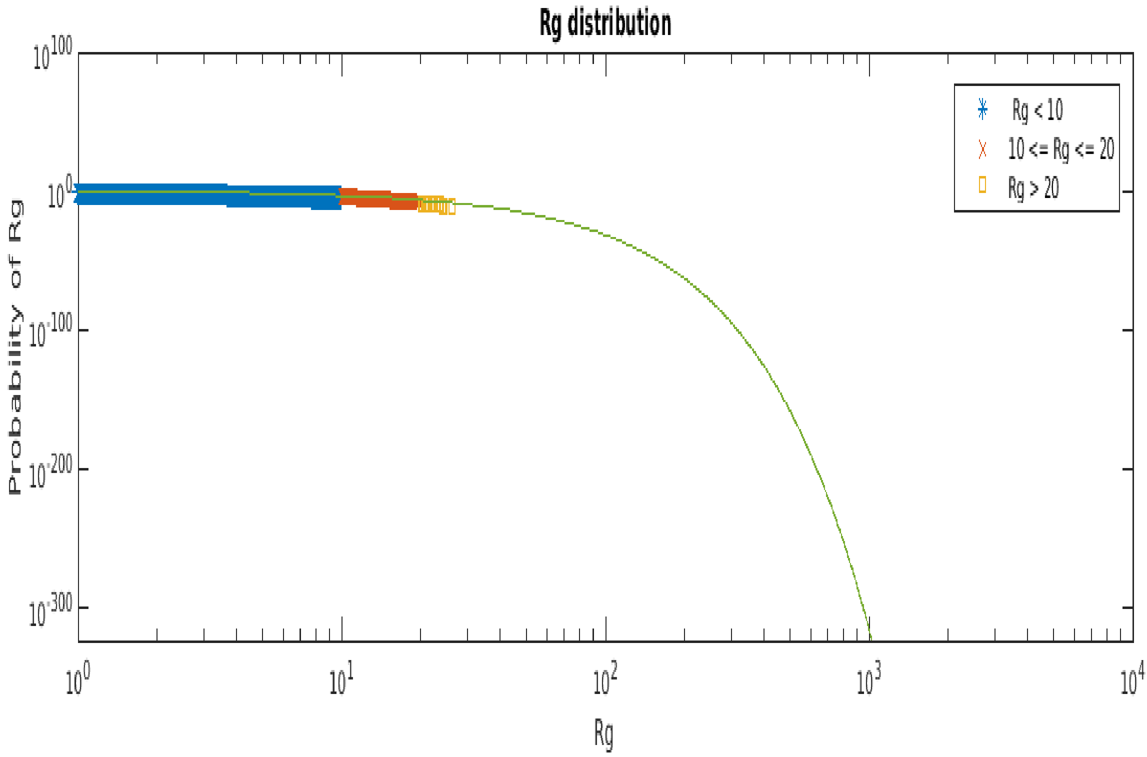


Figure B.8 – Day 5 probability distribution for the radius of gyration.

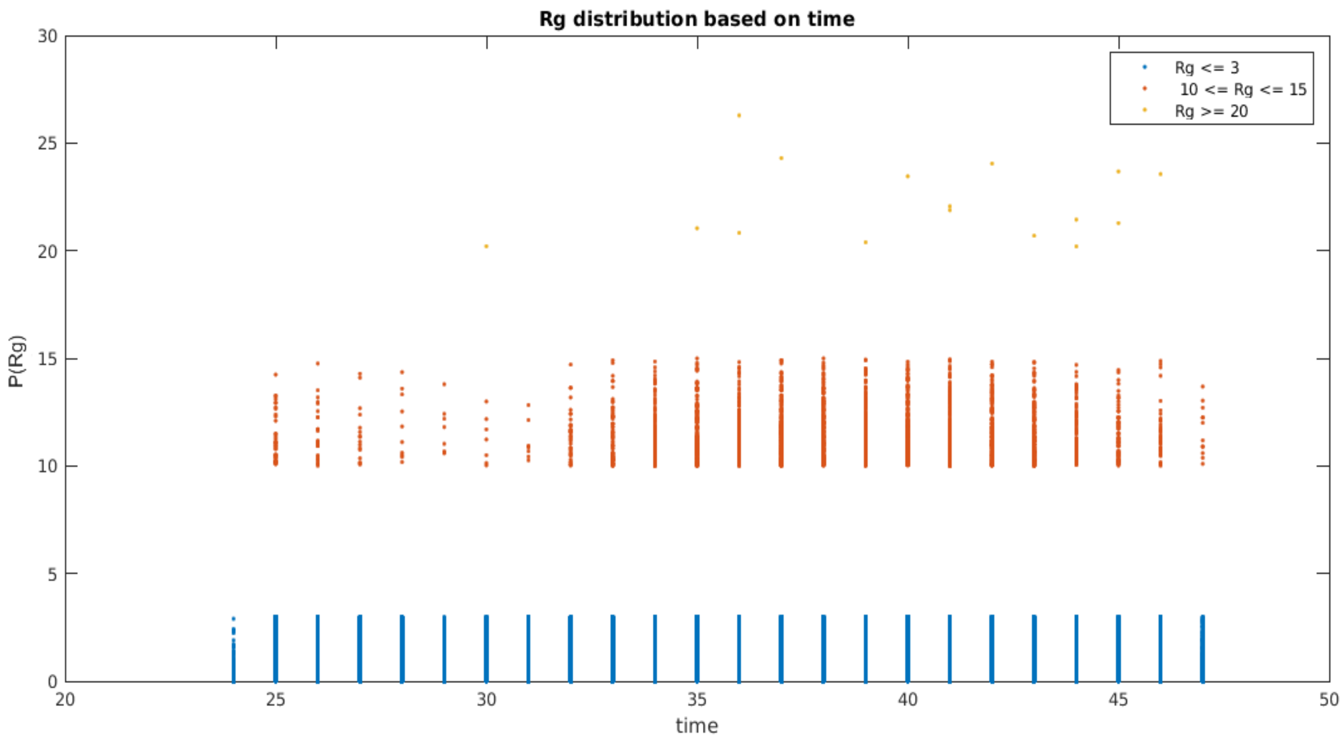


Figure B.9 – Day 5 probability distribution for the radius of gyration based on time.

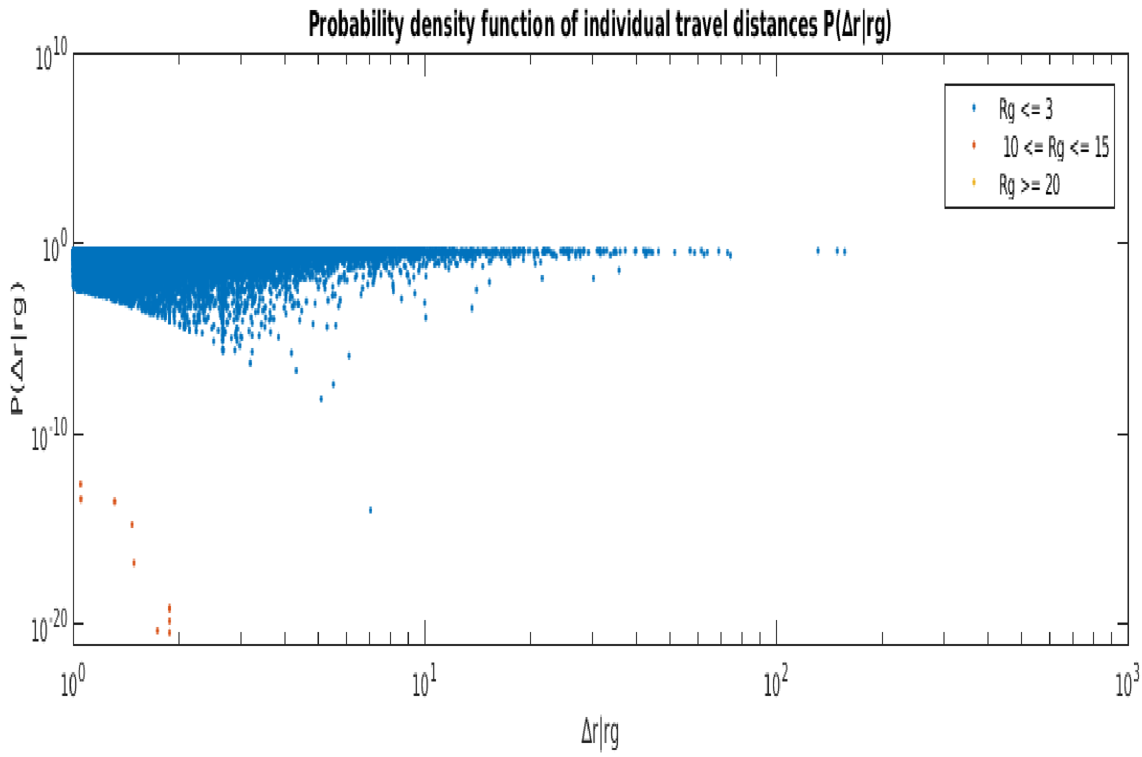


Figure B.10 – Day 5 probability density function of individual travel distances.

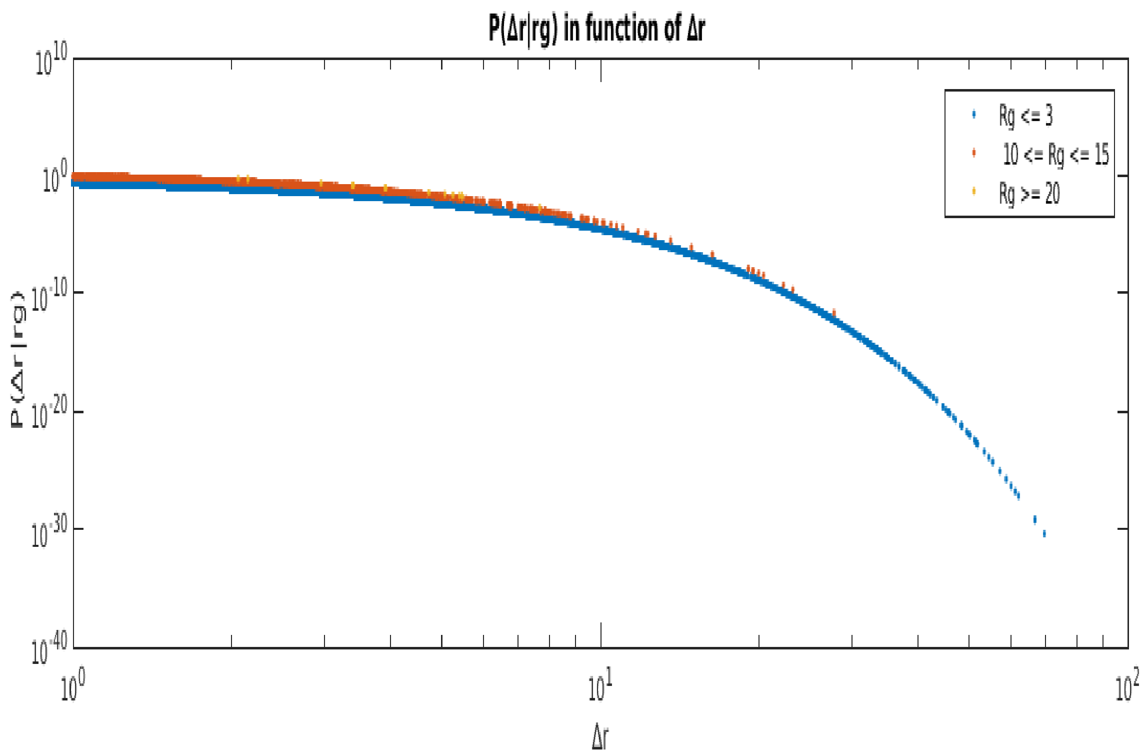


Figure B.11 – Day 5 probability distribution of $\triangle r|\triangle g$.

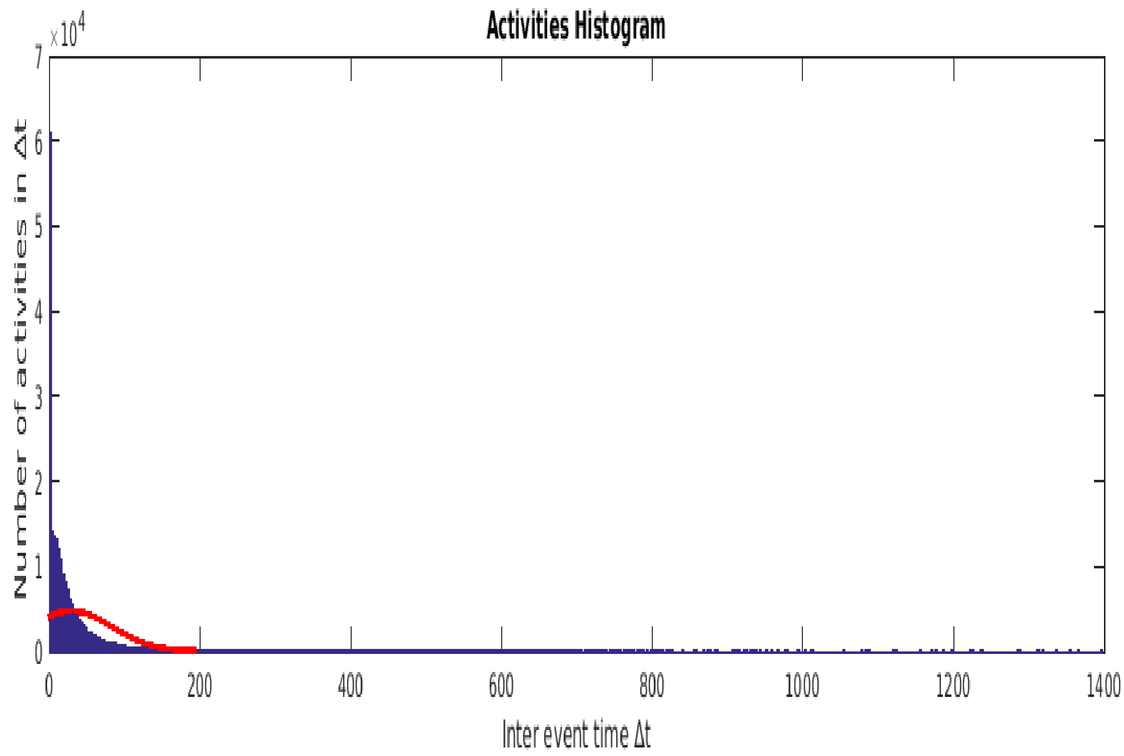


Figure B.12 – Day 7 activities histogram.

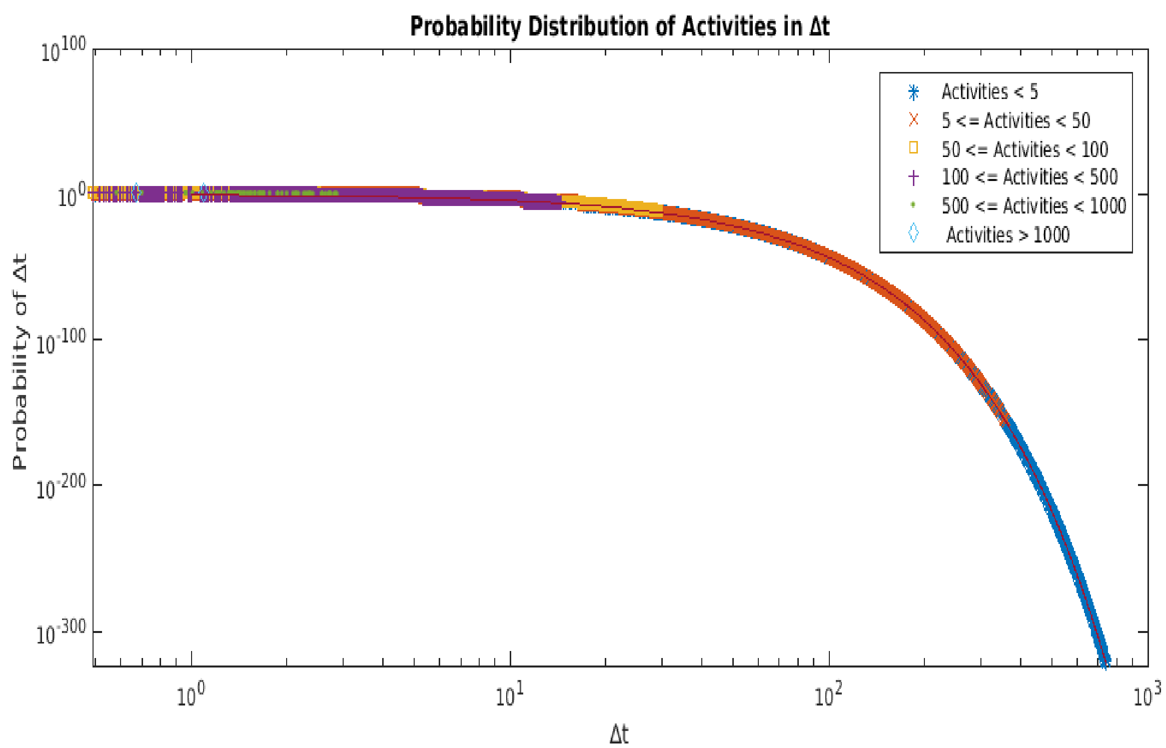


Figure B.13 – Day 7 activities by inter-event time.

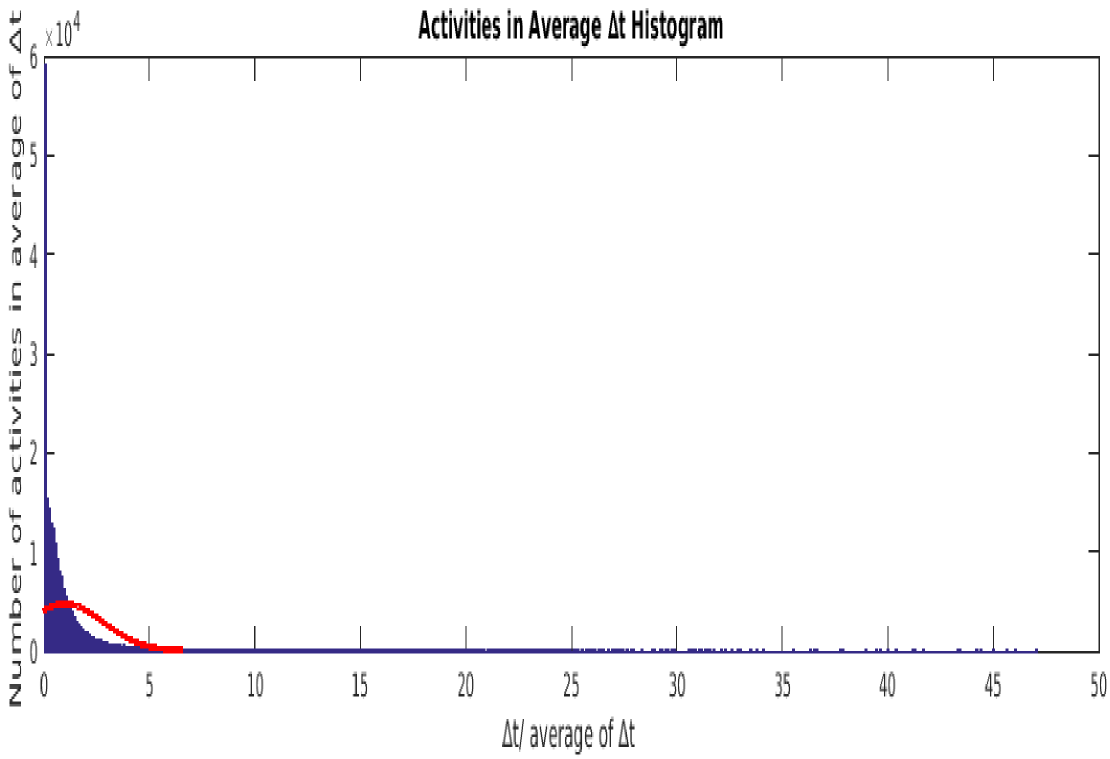


Figure B.14 – Day 7 activities in average time histogram.

Figure B.15 – Day 7 activities in average time.

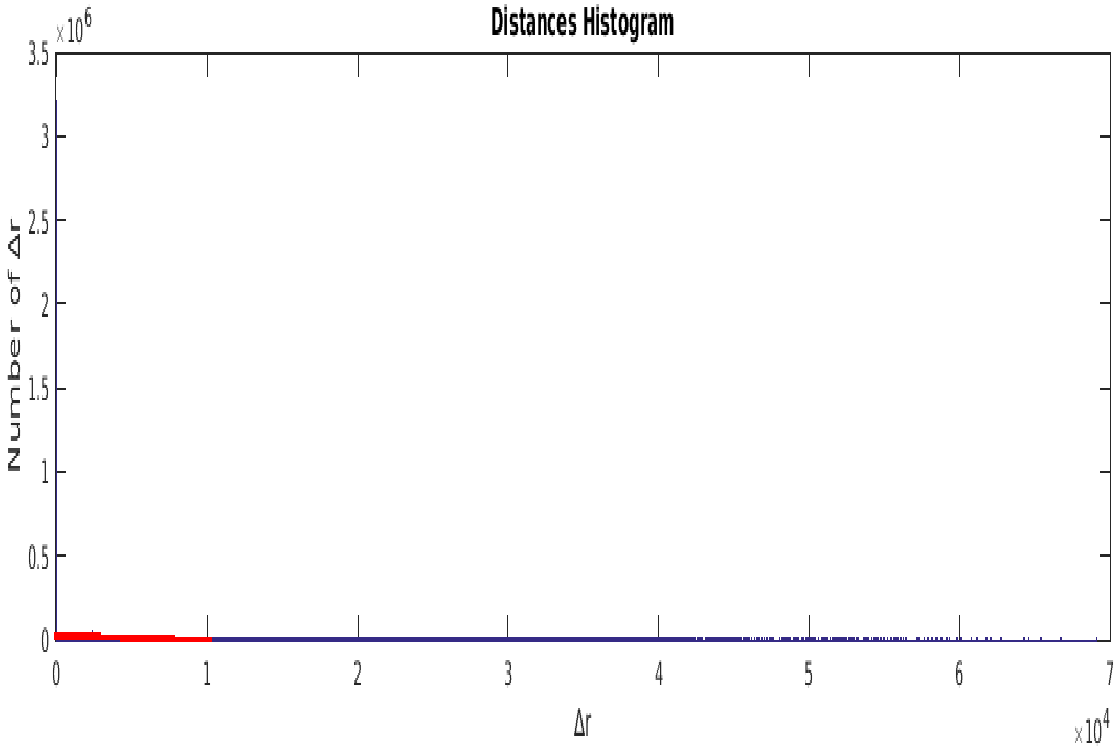


Figure B.16 – Day 7 distances histogram.

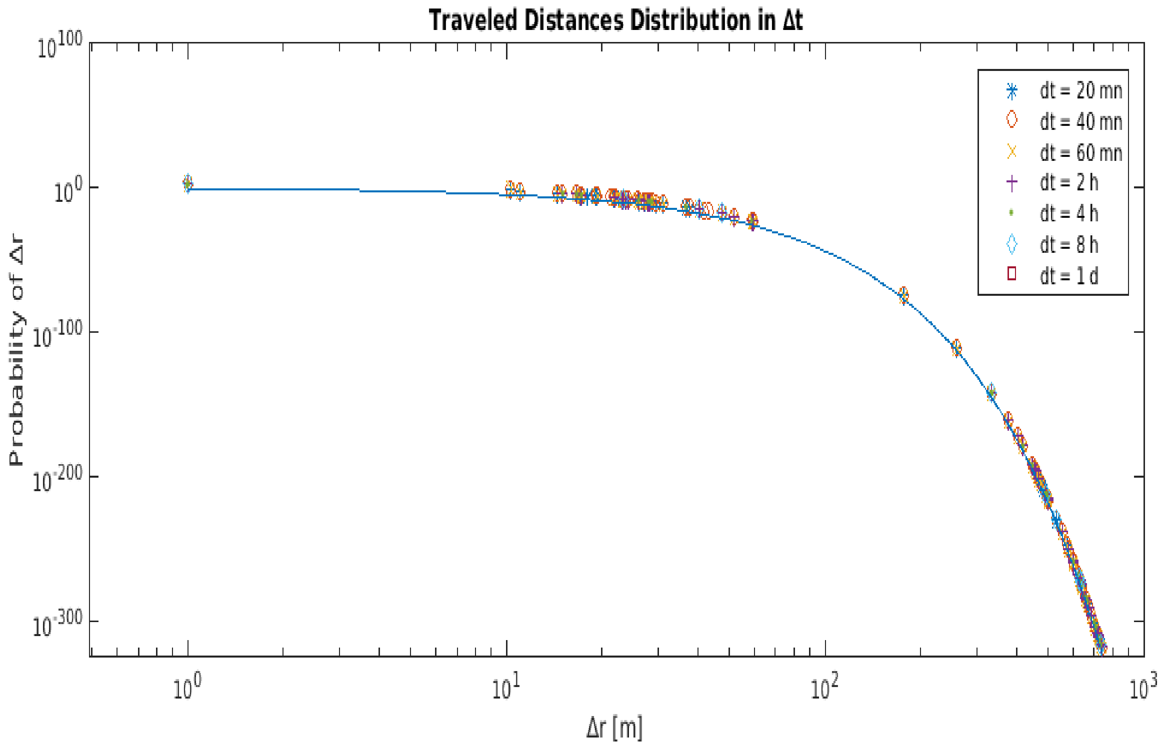


Figure B.17 – Day 7 traveled distances by inter-event time.

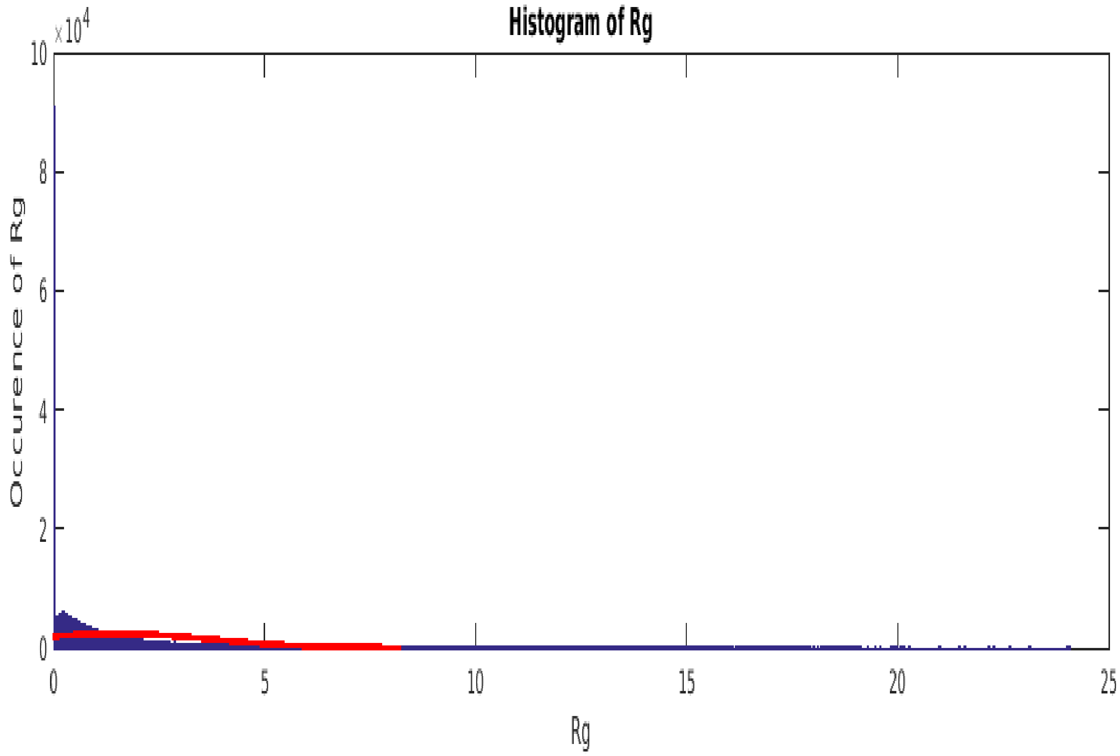


Figure B.18 – Day 7 radius of gyration histogram.

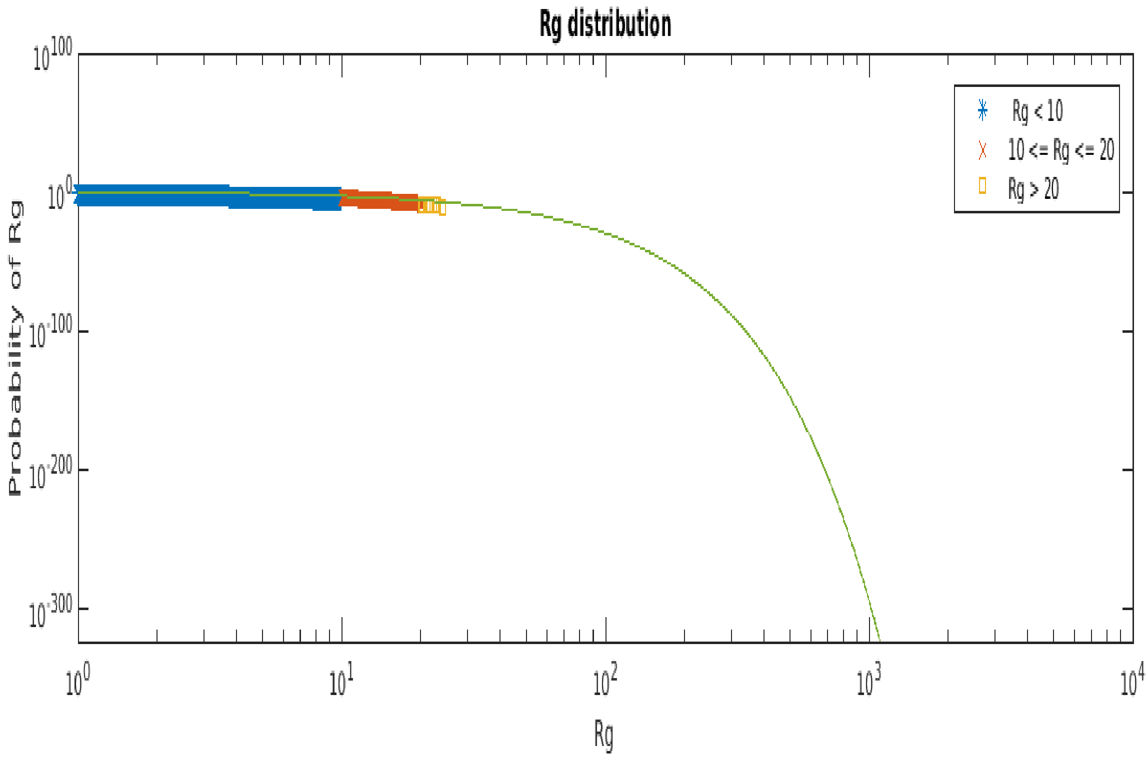


Figure B.19 – Day 7 probability distribution for the radius of gyration.

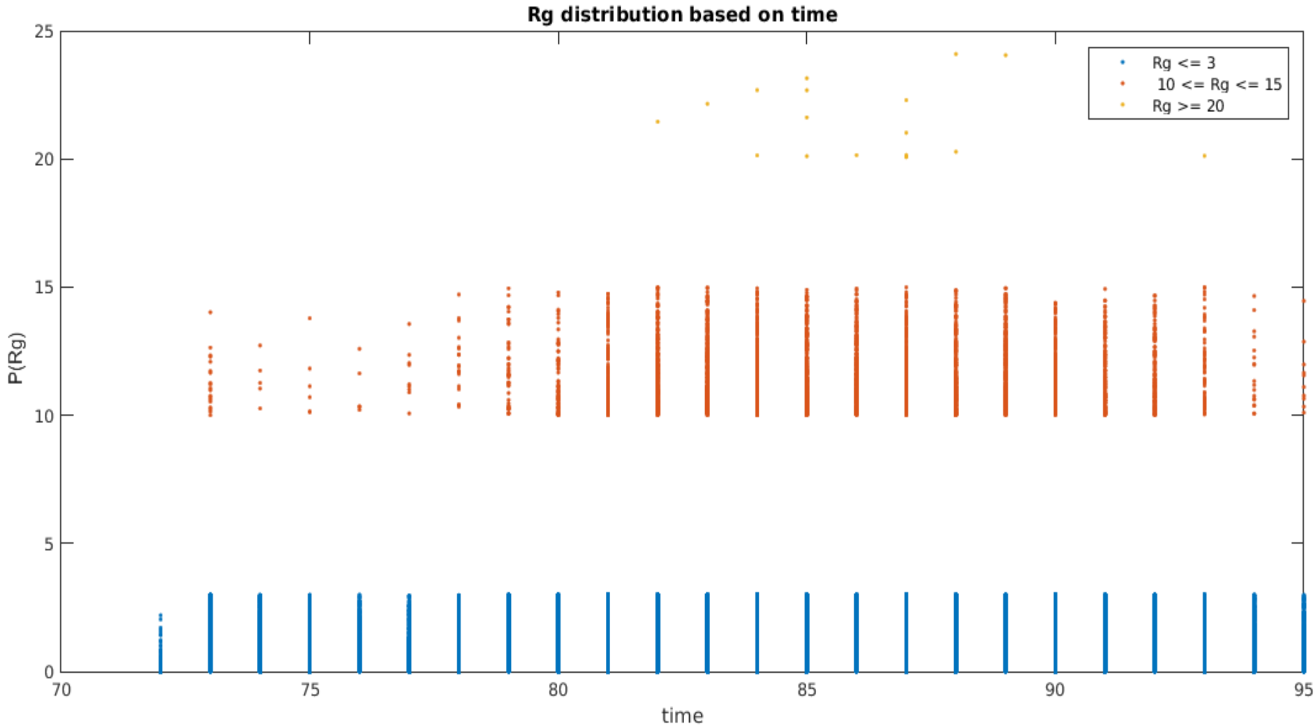


Figure B.20 – Day 7 probability distribution for the radius of gyration based on time.

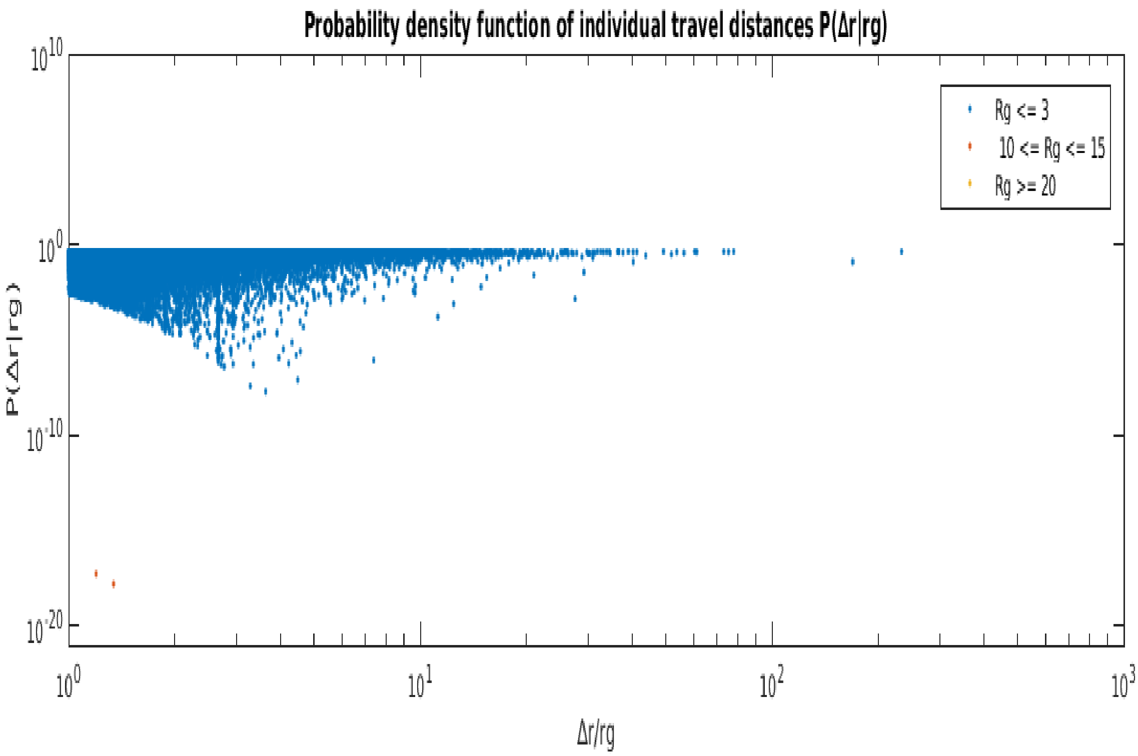


Figure B.21 – Day 7 probability density function of individual travel distances.

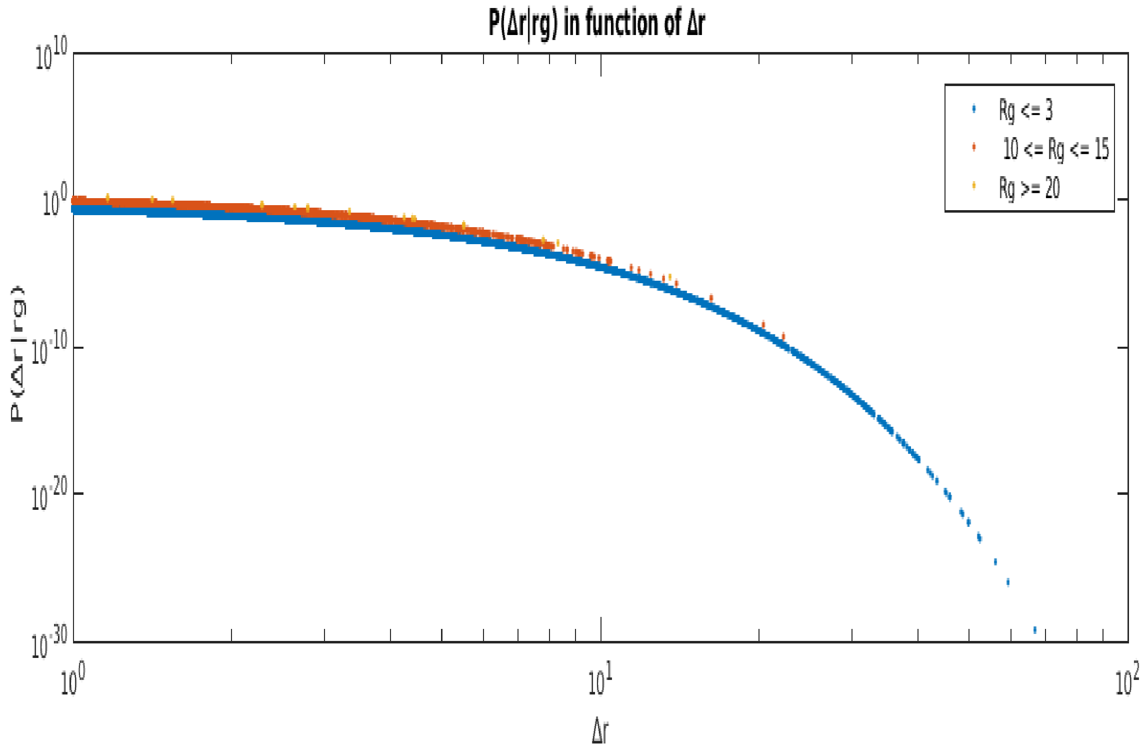


Figure B.22 – Day 7 probability distribution of $\triangle r|\triangle g$.

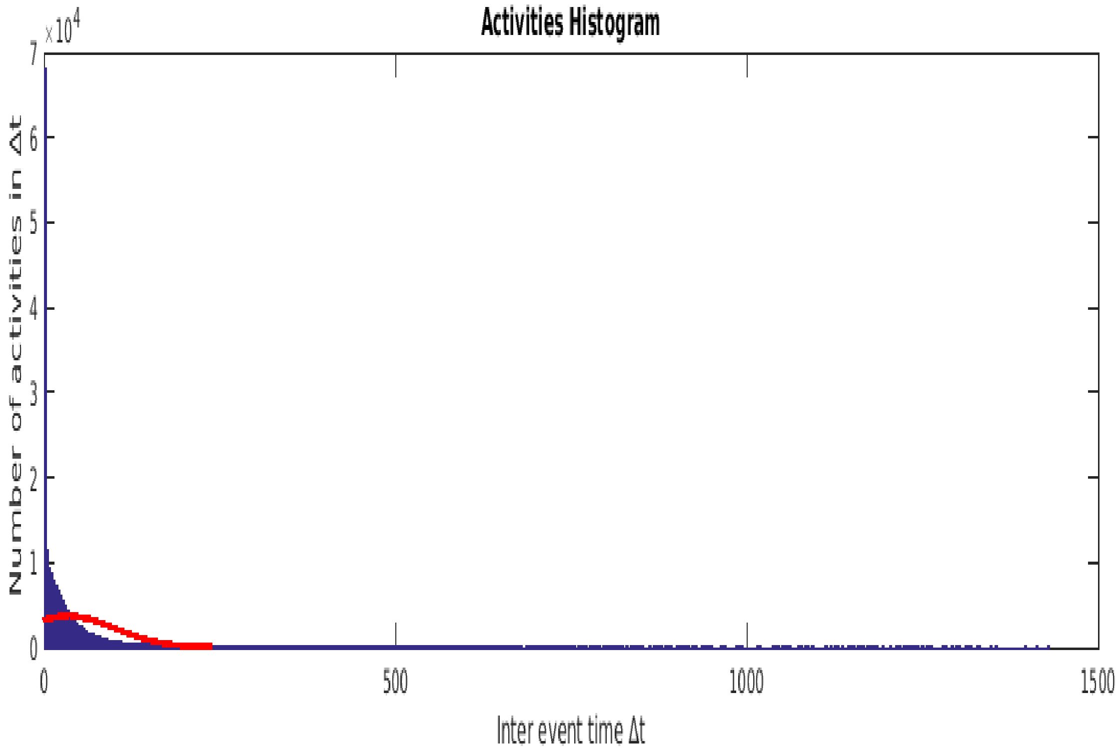


Figure B.23 – Day 14 activities histogram.

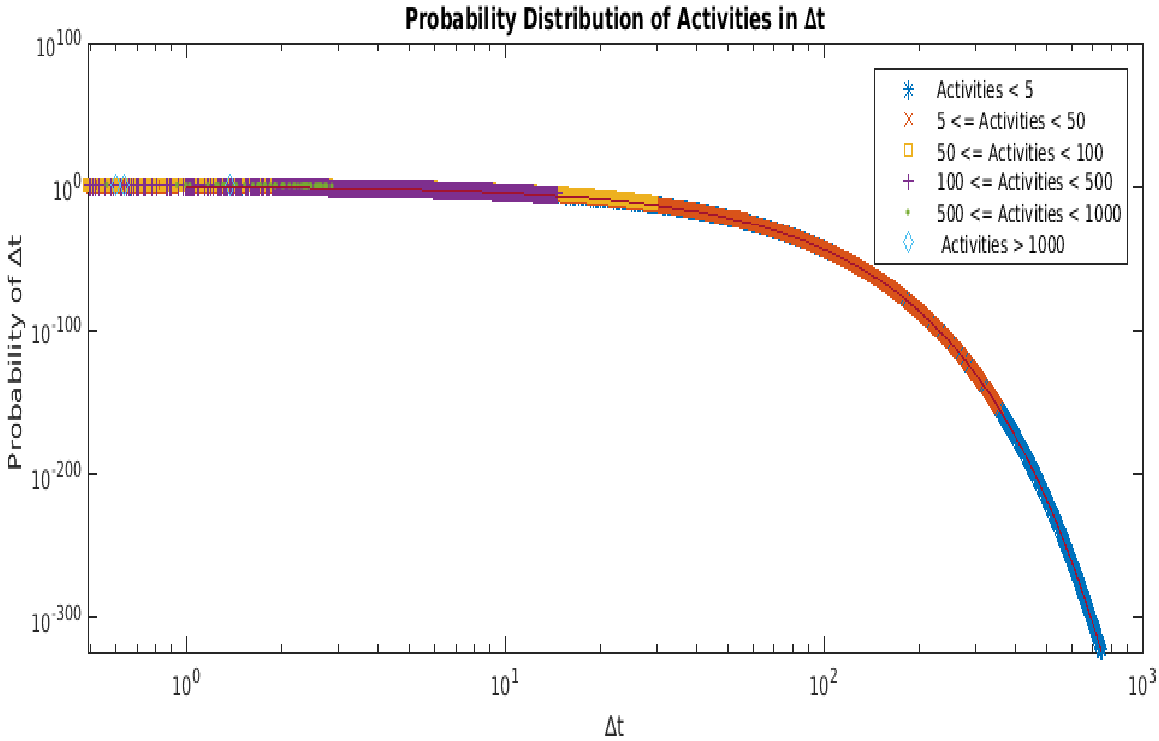


Figure B.24 – Day 14 activities by inter-event time.

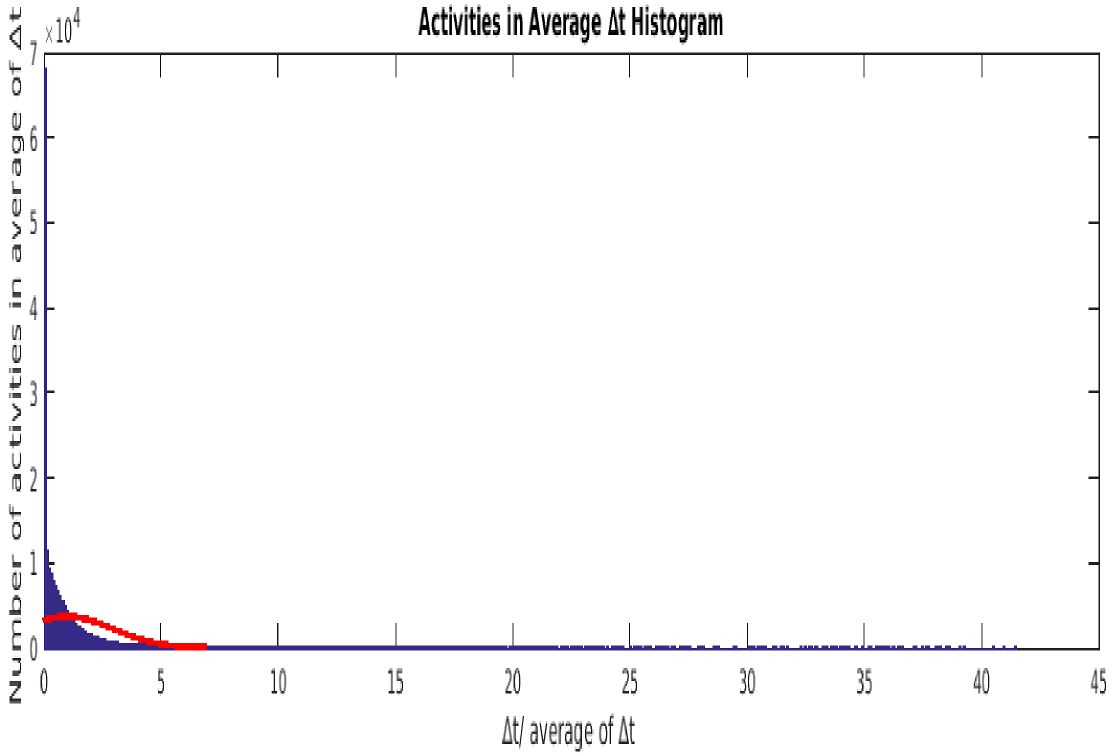


Figure B.25 – Day 14 activities in average time histogram.

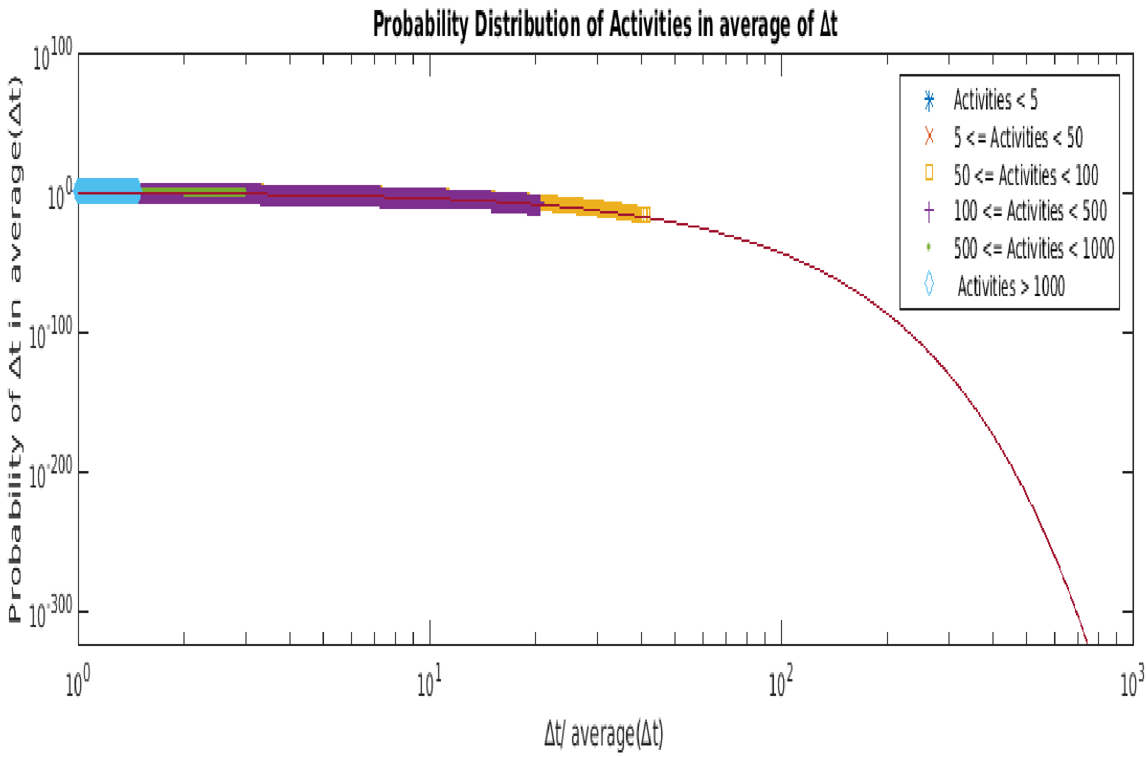


Figure B.26 – Day 14 activities in average time.

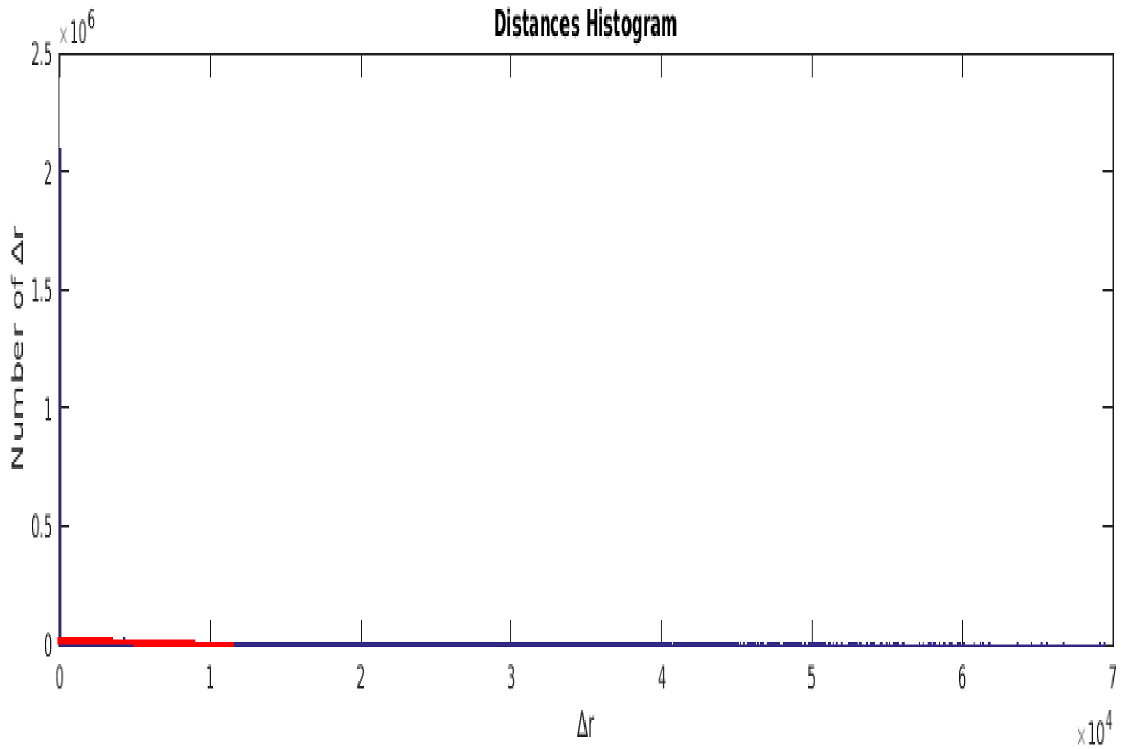


Figure B.27 – Day 14 distances histogram.

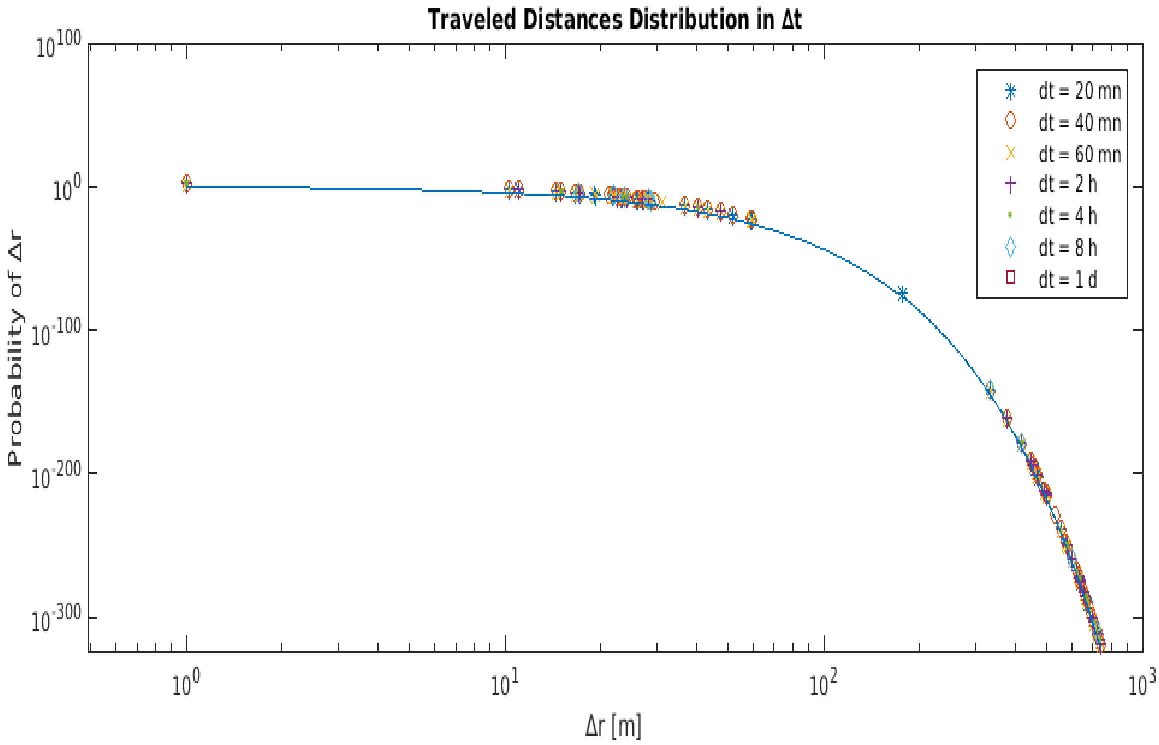


Figure B.28 – Day 14 traveled distances by inter-event time.

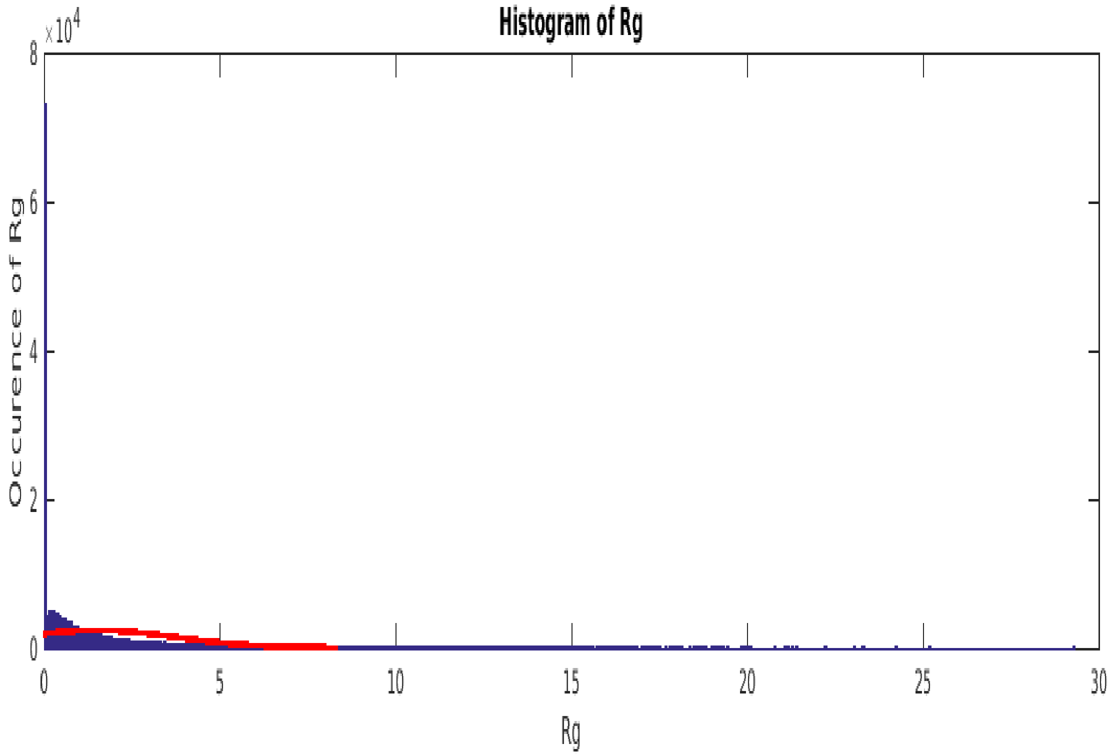


Figure B.29 – Day 14 radius of gyration histogram.

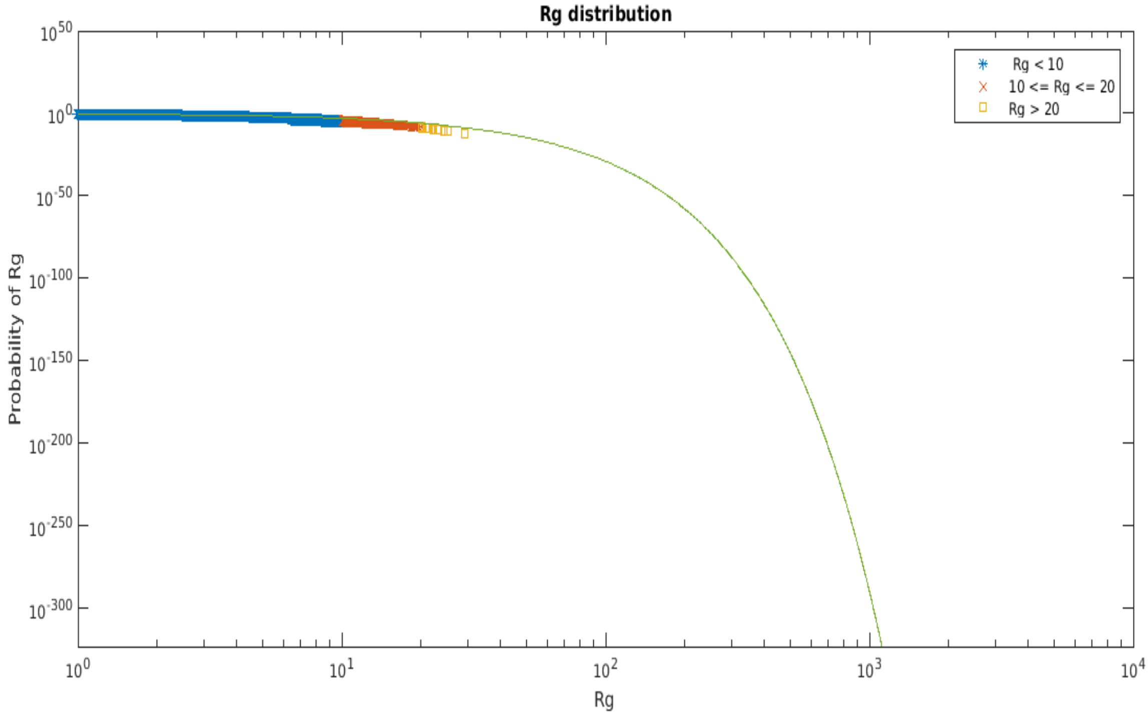


Figure B.30 – Day 14 probability distribution for the radius of gyration.

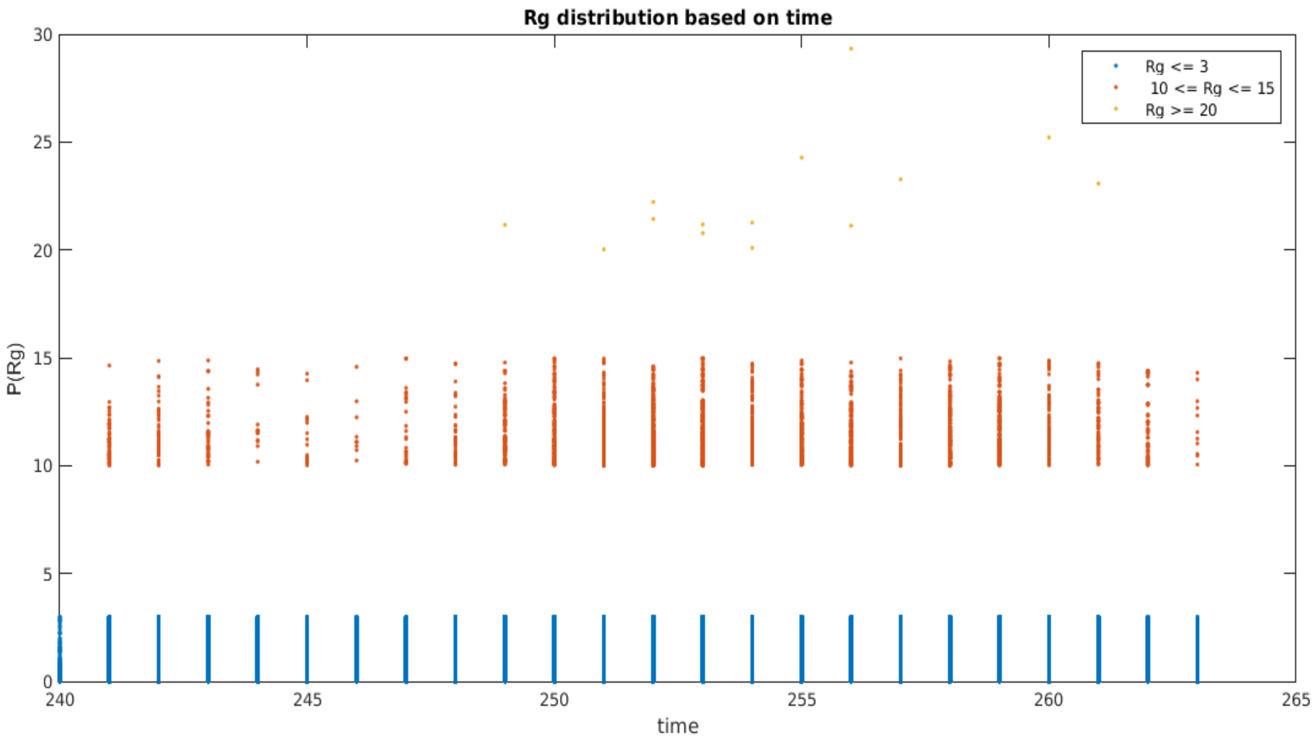


Figure B.31 – Day 14 probability distribution for the radius of gyration based on time.

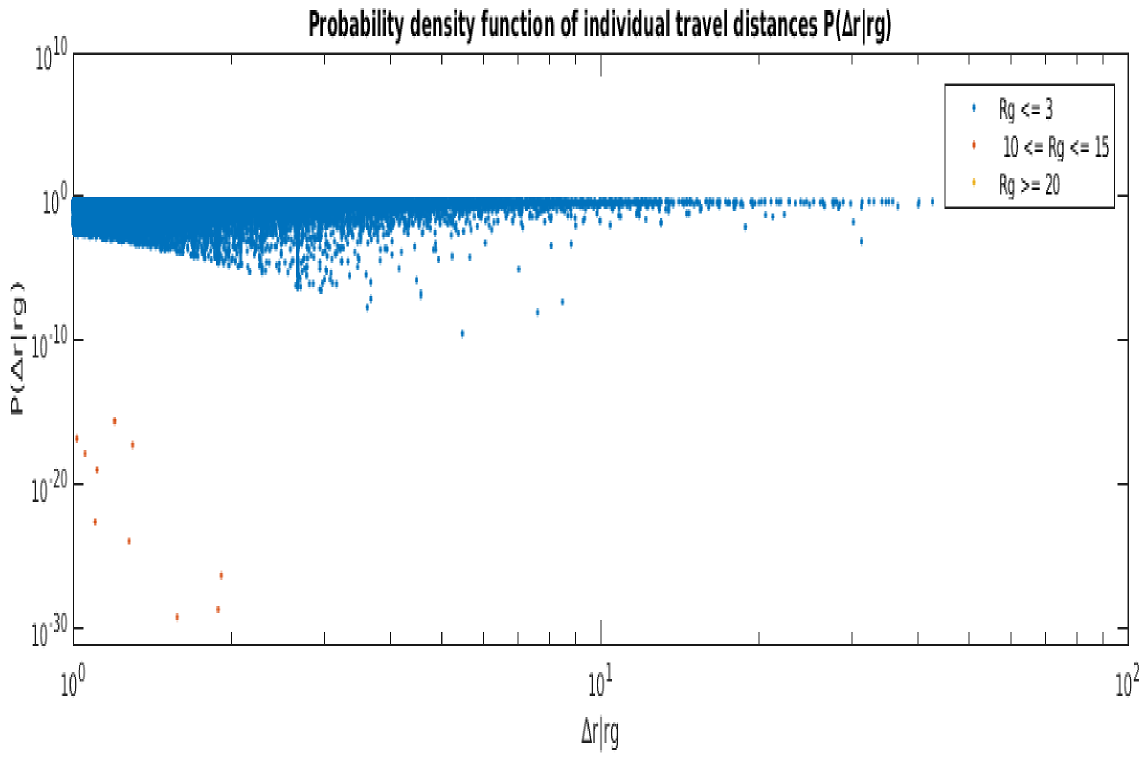


Figure B.32 – Day 14 probability density function of individual travel distances.

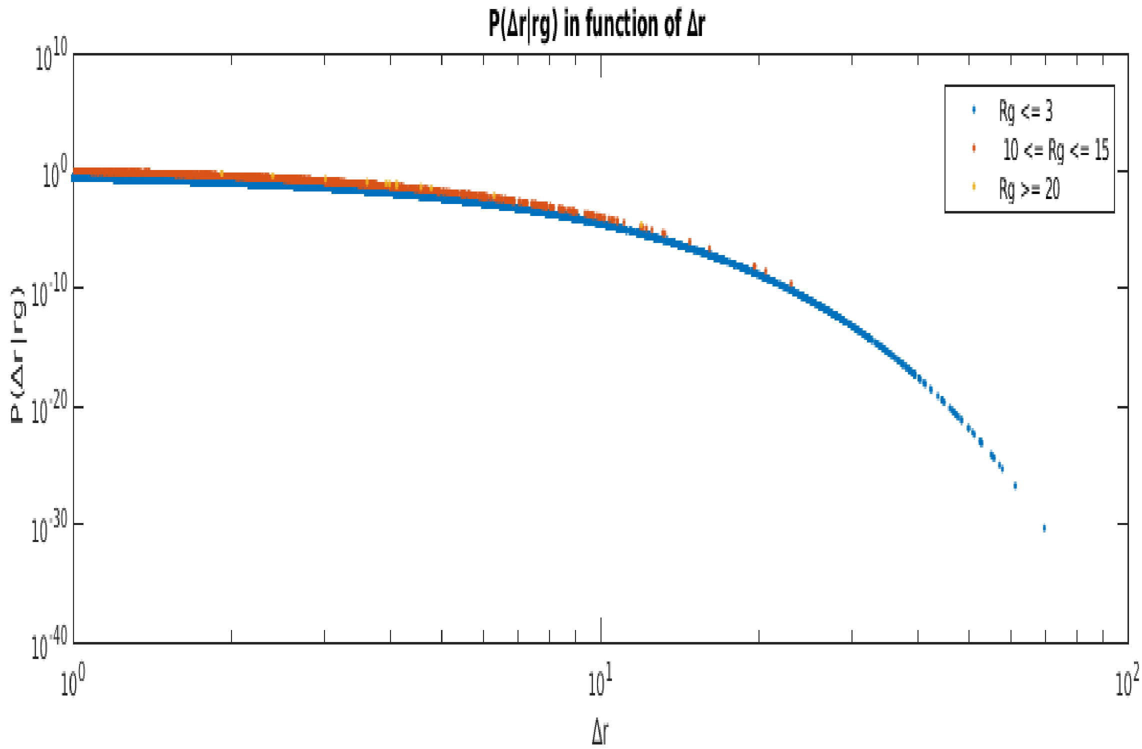


Figure B.33 – Day 14 probability distribution of $\triangle r|\triangle g$.

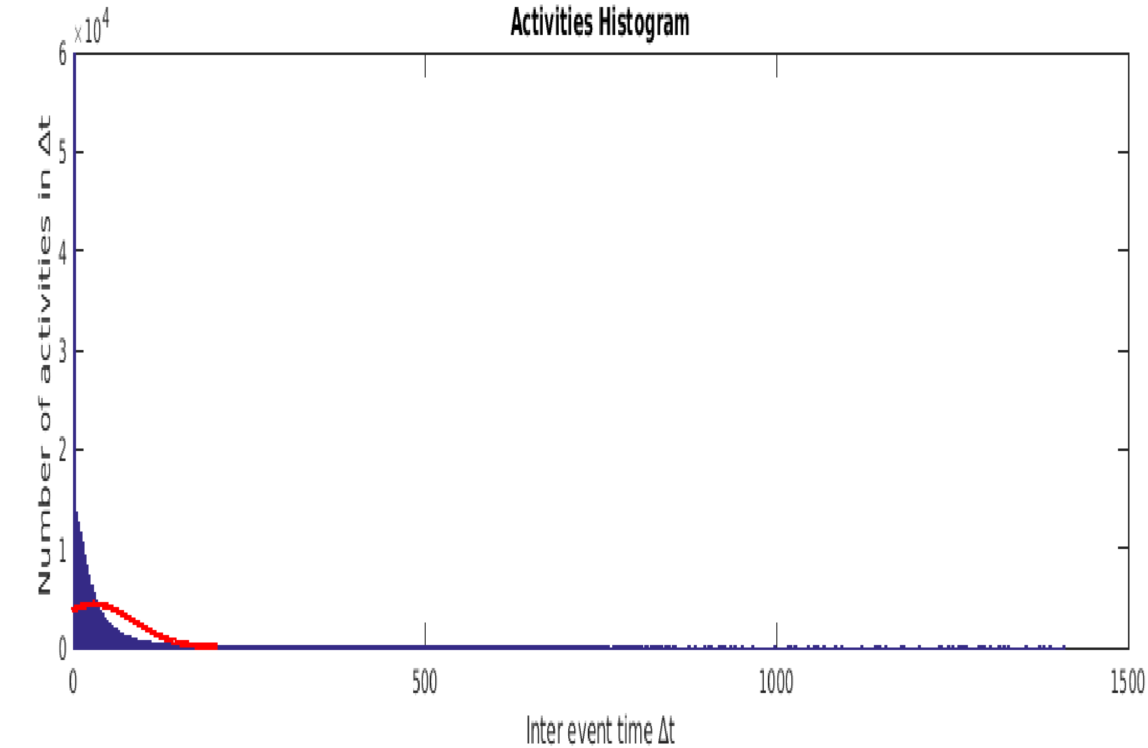


Figure B.34 – Day 15 activities histogram.

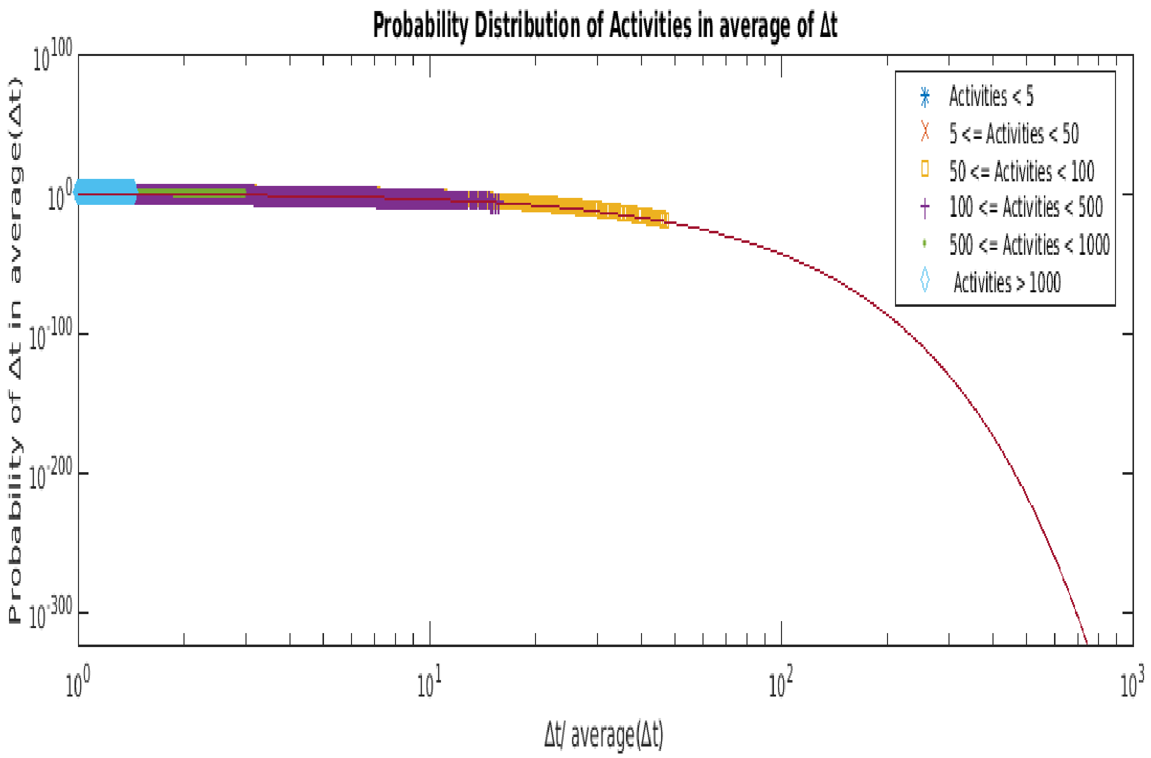


Figure B.35 – Day 15 activities by inter-event time.

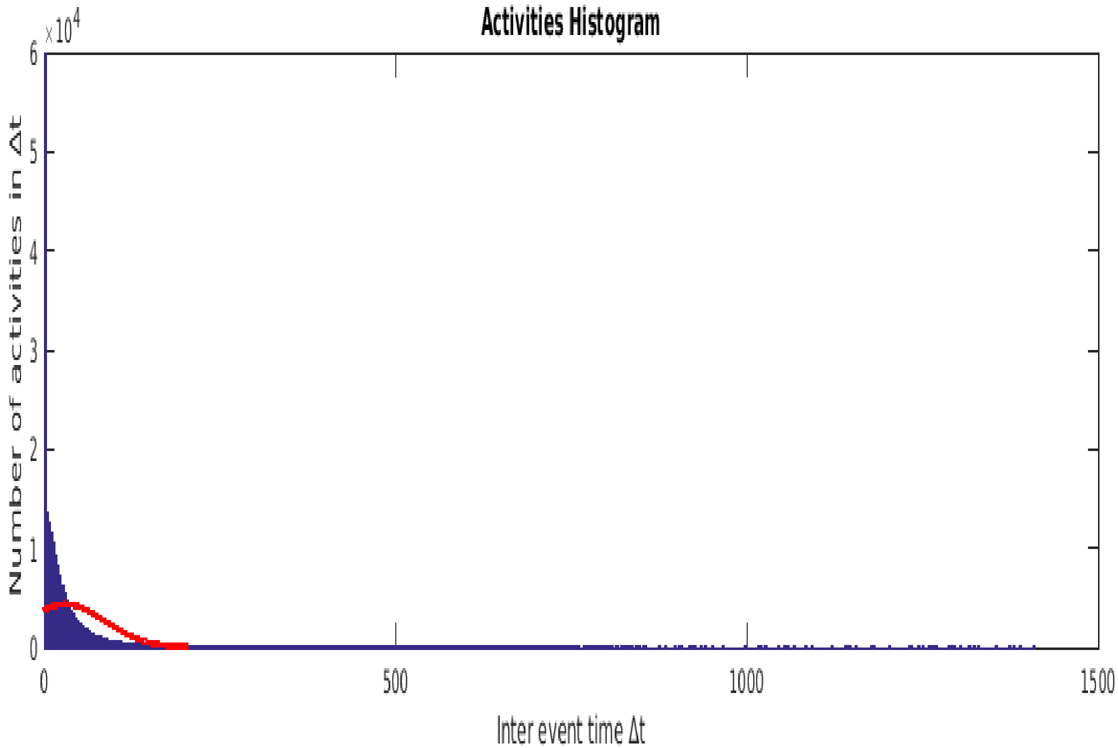


Figure B.36 – Day 15 activities in average time histogram.

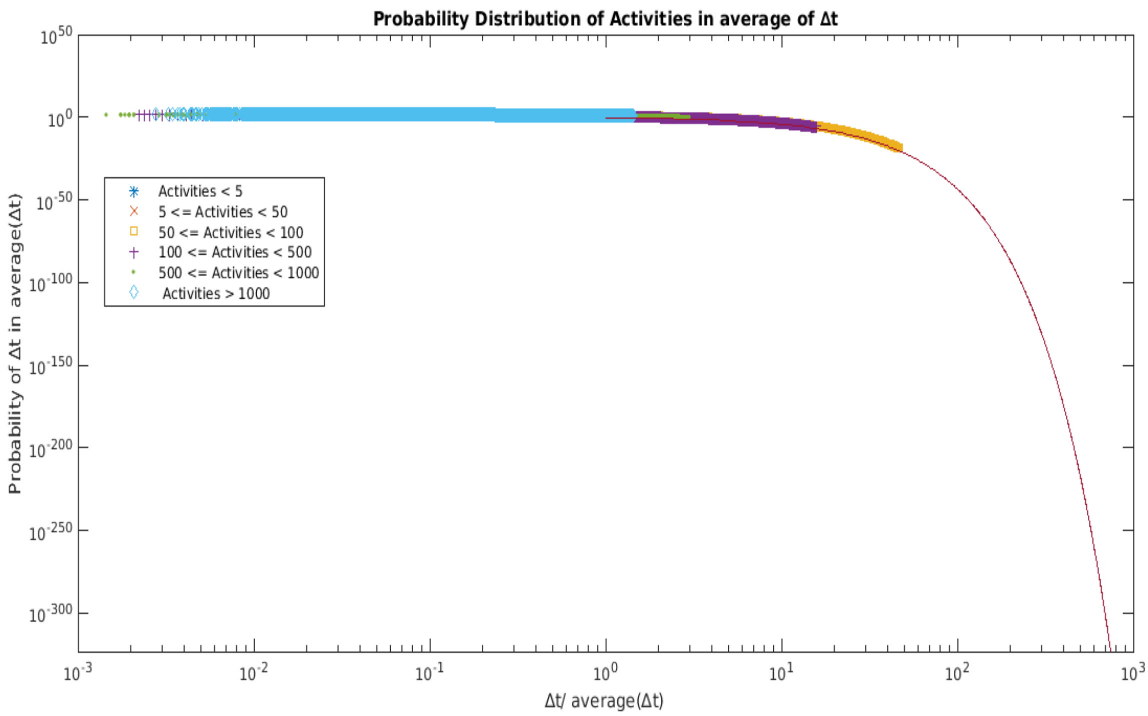


Figure B.37 – Day 15 activities in average time.

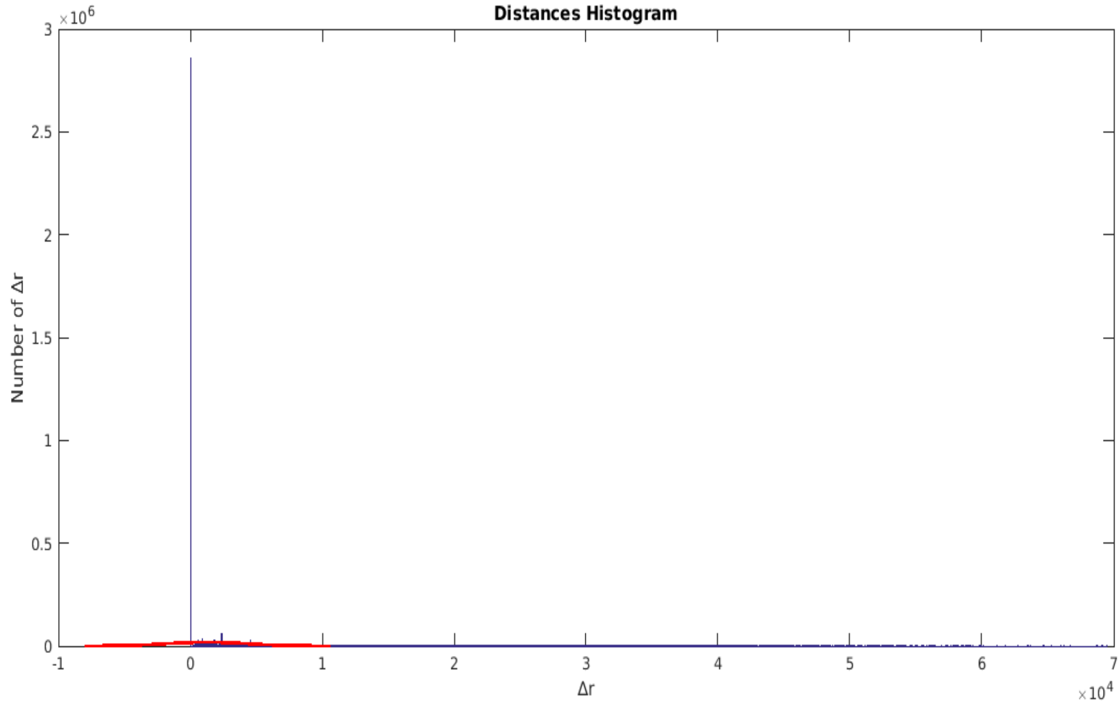


Figure B.38 – Day 15 distances histogram.

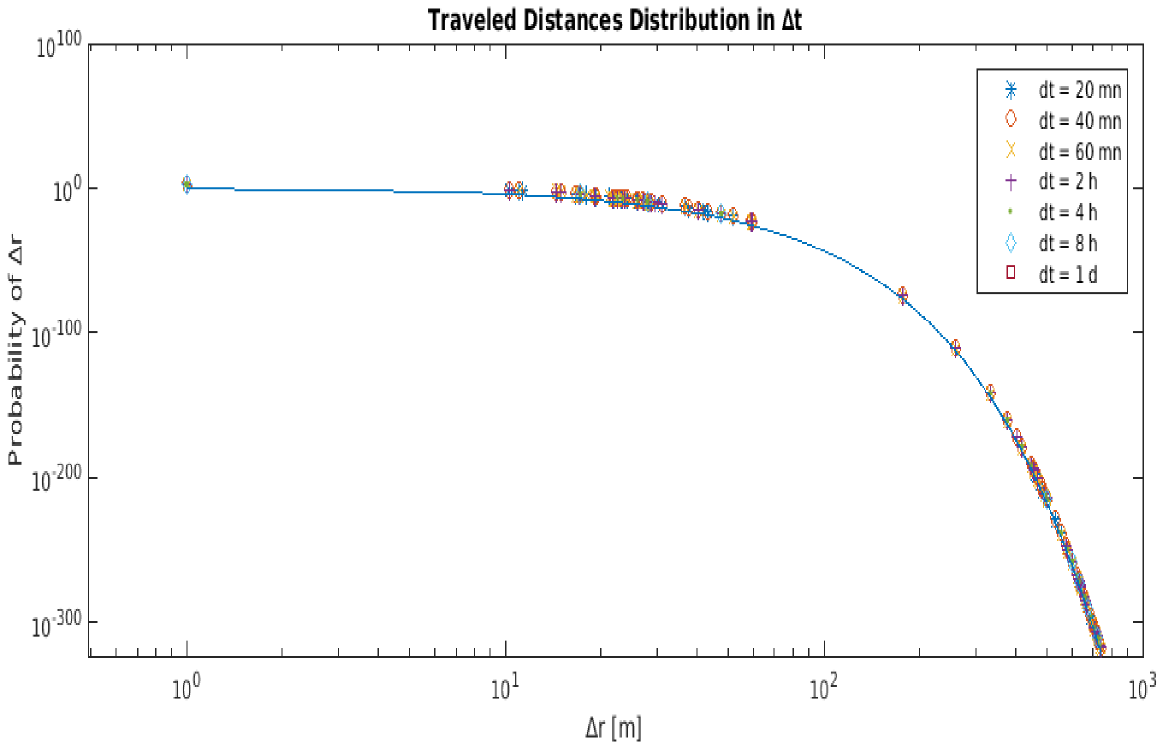


Figure B.39 – Day 15 traveled distances by inter-event time.

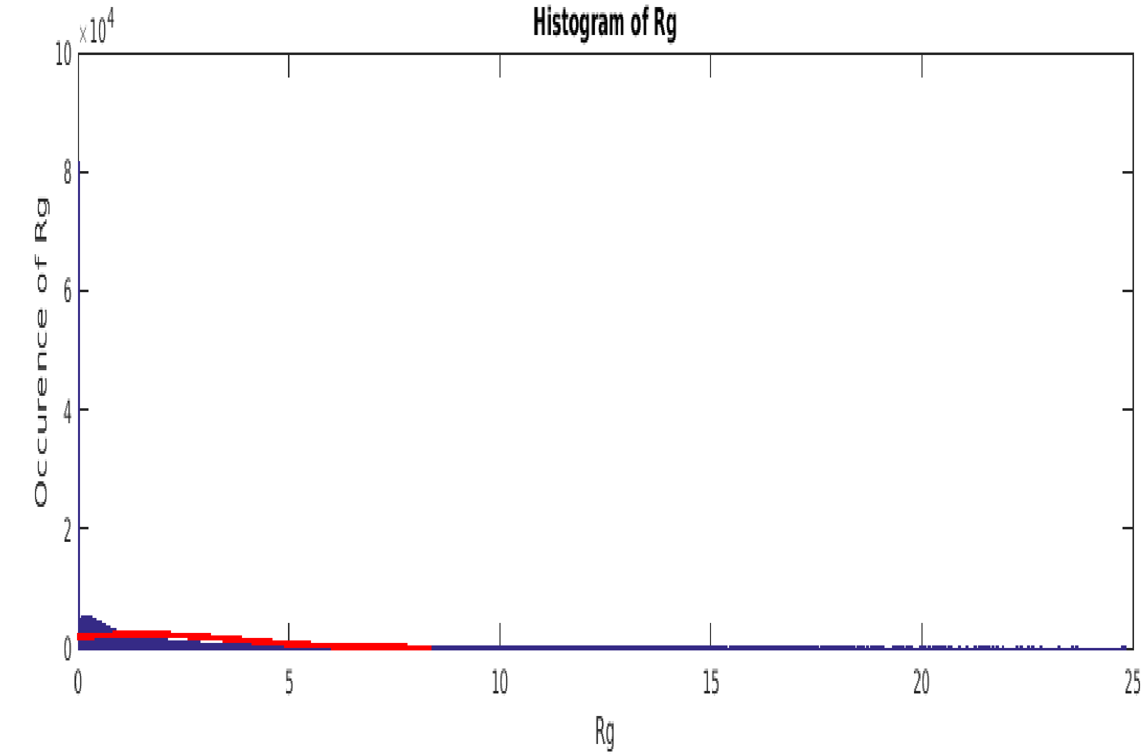


Figure B.40 – Day 15 probability distribution of radius of gyration histogram.

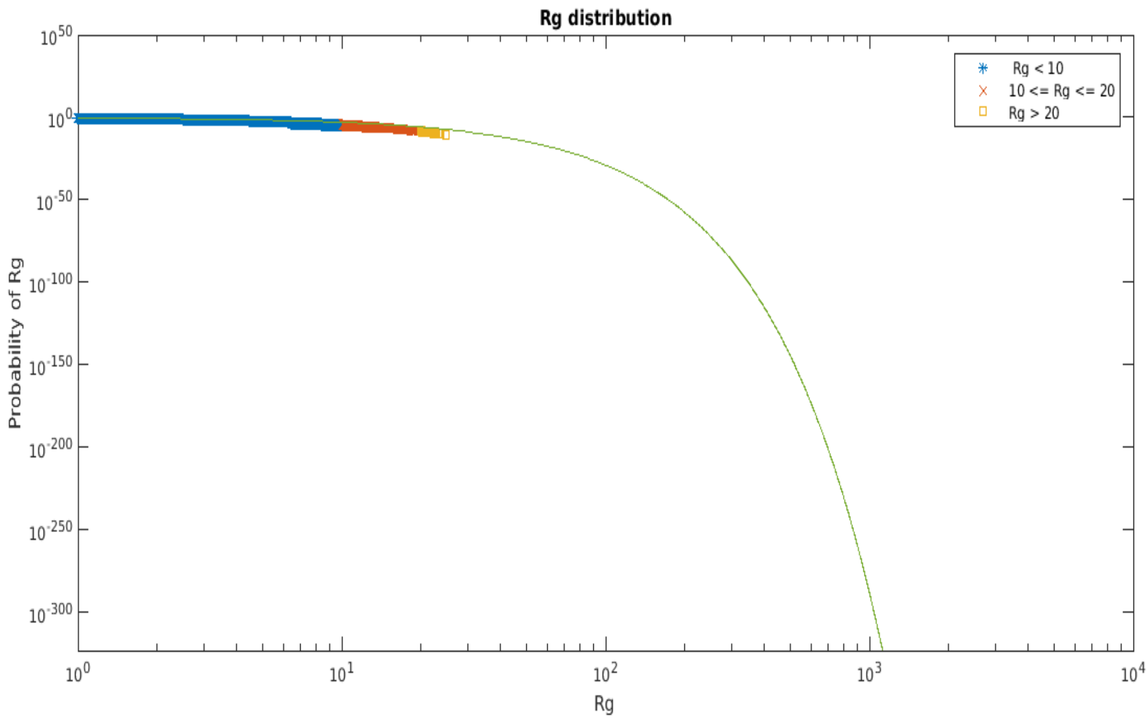


Figure B.41 – Day 15 probability distribution of radius of gyration.

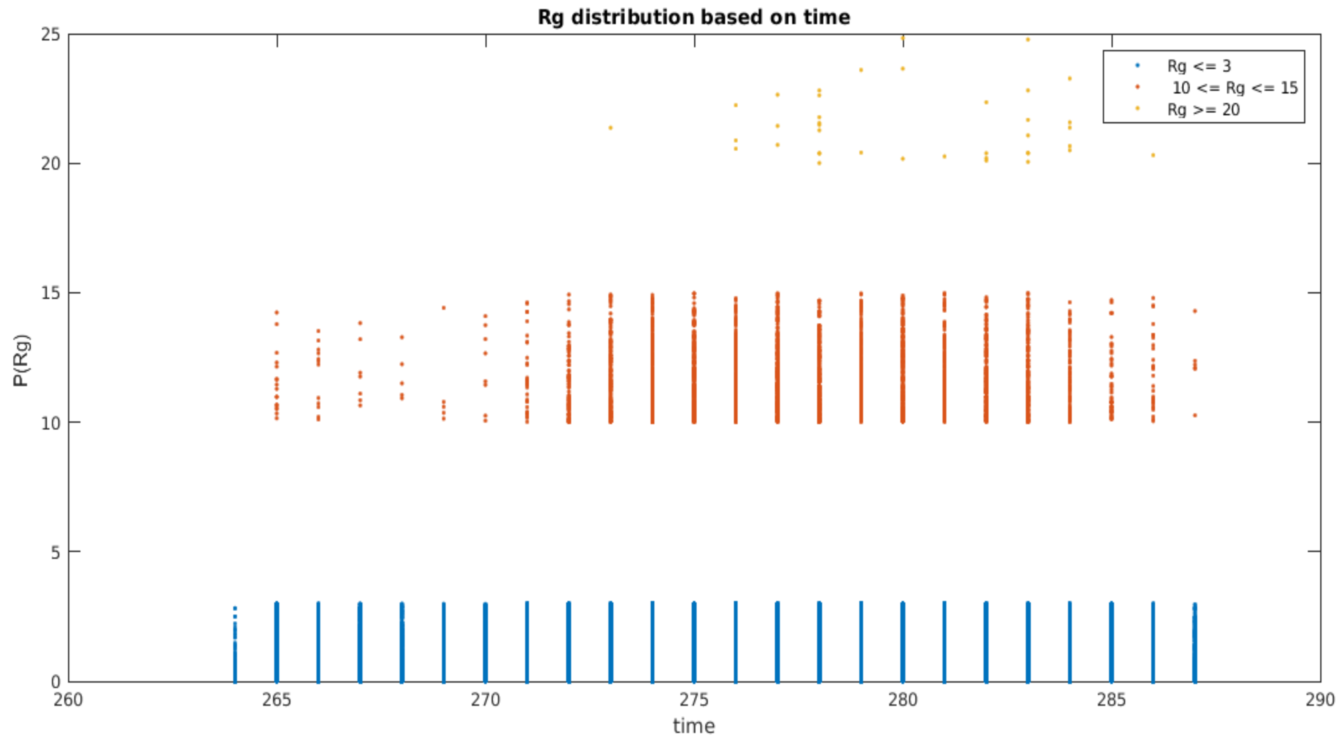


Figure B.42 – Day 15 probability distribution for the radius of gyration based on time.

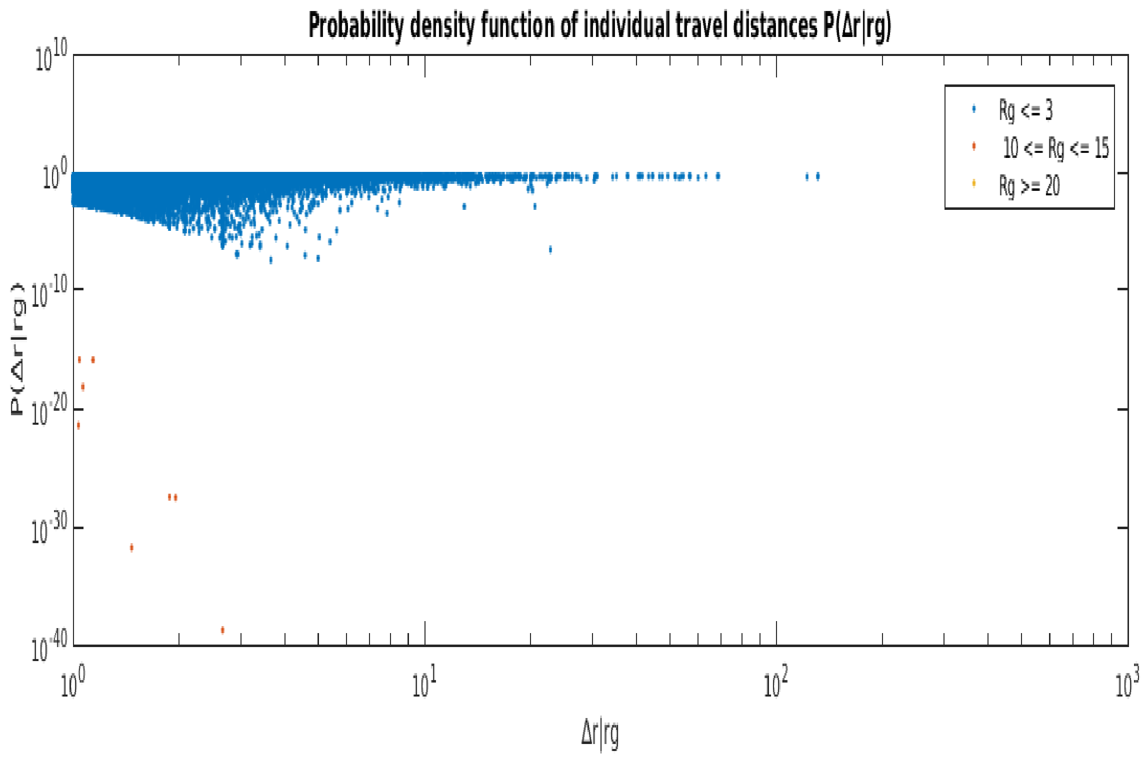


Figure B.43 – Day 15 probability density function of individual travel distances.

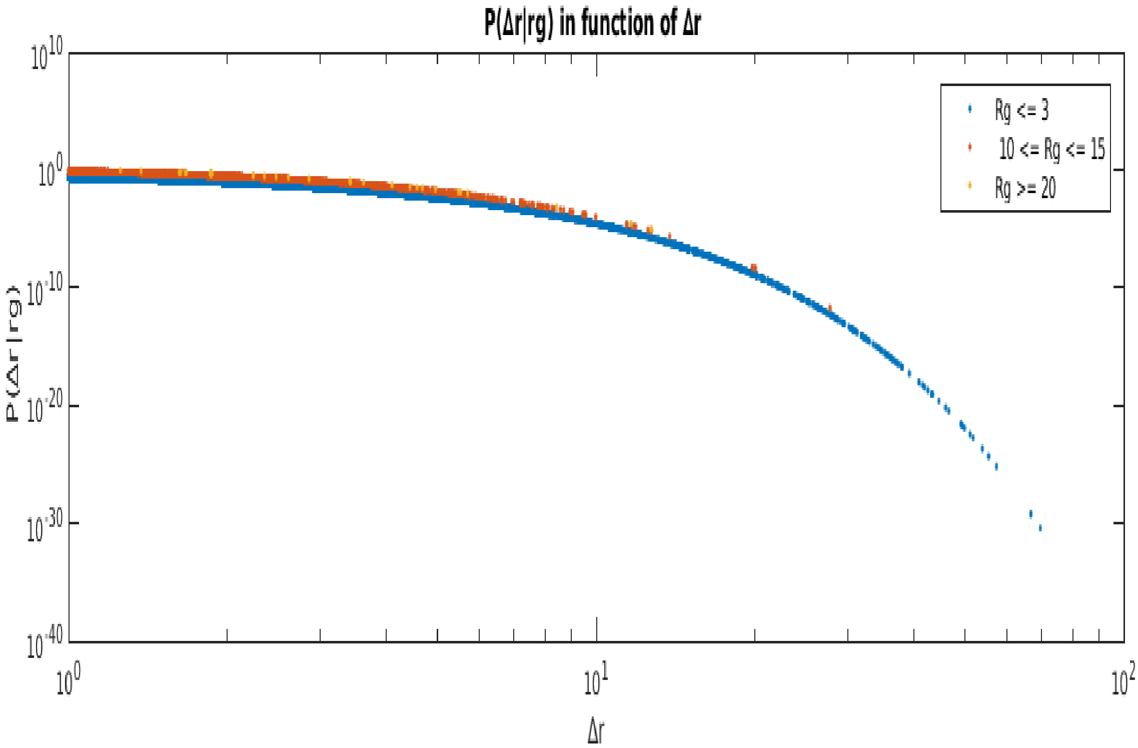


Figure B.44 – Day 15 probability distribution of $\triangle r|\triangle g$.