Intelligent Environments 2020
C.A. Iglesias et al. (Eds.)

doi:10.3233/AISE200067

Representing Human Mobility Patterns in Urban Spaces

Luís Rosa [a], Fábio Silva [b,a] and Cesar Analide [a]
[a] *ALGORITMI Center, Dep. of Informatics, University of Minho, Braga, Portugal*
[b] *CIICESI, ESTG, Politécnico do Porto, Porto, Portugal*

Abstract. Human mobility is important in understanding urban spaces. Citizens interact with urban spaces using the available infrastructures, not just in the mobility sector but in public services, and in Information and Communications Technology (ICT) services, that simultaneously record their footprints. Besides, the number of mobile users is increasing very rapidly in the Internet of Things (IoT) era. These additional devices will produce a great amount of data and create new big challenges for network infrastructure. Because of this new connectivity platform, and the fast growth of wireless communication, it's important to discuss the arrival of 5G systems. They will have a large impact on coverage, spectral efficiency, data rate of global mobile traffic, and IoT devices, and in turn it will be possible to analyze the lifestyle and understand the mobility of people, such as the most frequently visited urban spaces. Therefore, this paper is relevant in the context of smart cities and will allow for an easy connection between citizens and technology innovation hub, acquiring detailed data on human movements. Based on the analysis of generated data we try to widen this view and present an integrated approach to the analysis of human mobility using LinkNYC kiosks and 311 Service Requests in New York city.

Keywords. Human Mobility Pattern, Device Network Dataset Collection, Smart Cities, Deep Neural Network

1. Introduction

Mobility is a term with multiple connotations and one of them relates to the sense of mobility as freedom of movement across physical space. Being ”mobile” refers to the movement of individuals or groups from place to place, job to job, or one social class or economic level to another. Nowadays, mobility is not only a social phenomenon, but also includes technologies and services that enable people to move around more freely[19]. As the world keeps changing at an incredibly fast pace, so is the way in which people get from point A to point B. 50% of the world's population lives in urban environments and, by 2040, more than 70% is expected to live in cities. Moreover, we are transitioning into a ”sharing” economy: sharing a ride, a home or an office are just examples of what's becoming the norm, instead of being exceptions. Along with this change, we are also witnessing a parallel increase in the use of mobile devices. On the other hand, the social and spatial behavior of individuals is affected by two main conditions. While we are constrained by time, cognition, age, the need for food, etc., each one of us is characterized by personality traits that make us if perhaps not unique, at least different from many others. Personality psychology conjectured a long time ago that a set of personality traits underlie all aspects of human behavior [13, 18].

These psychological factors along with the reasons which make mobility easier have contributed to the growing interest of the scientific community in studying the existing movement patterns of individuals. Some research projects integrate activities of members of the ALGORITMI Center, Department of Informatics, University of Minho (Braga, Portugal), that are focused on the various problems associated with the idea of ”Smart City”. The variety of research areas has been treated as an important added value of the smart human mobility. For example, using the potential of ubiquitous computing, the PHESS platform notifies the community about incorrect driving practices. From monitoring and building of communitiy road maps based on driving analysis, this service aims to drive behavioural change and offer intelligent planning to the users [20]. Still related to traffic, researchers forecasted traffic flow in data-scarce environments using ARIMA and LSTM Network models [7]. In smart mobility area, a project tackles the Vulnerable Road Users’ (VRUs) problem (non-motorised road users, such as pedestrians and cyclists as well as motor - cyclists and persons with disabilities or reduced mobility and orientation) and provides a proof of concept on crowd sensing for urban security in Smart Cities [8]. And, recently presented in a Doctoral Symposium on Artificial Intelligence, we apply three different Machine Learning techniques such as Convolutional Neural Network (CNN), Long-Short Term Memory (LSTM), and a combined architecture, called CNN-LSTM, to the data generated by LinkNYC Kiosk devices, based on the city of New York, and compare results in predicting human mobility [9]. Based on the experience learned from the studies, we put together a comprehensive tutorial about state of art in mobility research, which introduces and discusses popular crowdsensed data types, different human mobility subjects, and analysis methods.

Nowadays, people’s expectations, decisions, habits and life experiences change continuously, influenced by ongoing technology innovations and improvements in the infrastructures of cities. In this new ”urban innovation hub”, in many cases human movement can be followed digitally. Using empirical mobile phone datasets, this paper attempts to explore spacial distribution and human mobility patterns, as well as the interrelationship between them at the community level. To be more precise, when we analyze these large segments of New York City’s population, we will be able to gather knowledge on how and when people are moving around the city.

2. Human Mobility: Opportunities in smart cities

The study of human mobility has intensified in the last decades because to its potentiality. In fact, if applied to socio-economic studies in smart cities, understanding macroscopic mobility patterns can potentially better support decisions and eventually improve quality of life. Our study contemplates a large population of free-willed and autonomous decision-making individuals, but it also takes into account their interaction with mobile devices. However, other factors can be summarized to help us understand how human mobility patterns are established.

2.1. Activity-Based Human Mobility Patterns

Understanding the dynamics of daily mobility patterns is essential for the planning and management of urban facilities and services. For the same reason the movements can also be instrumental in understanding and assessing individual and collective mobility patterns. In this paper, aspects of these factors are analysed based on their importance and on how they affect mobility patterns.

2.1.1. Individual Mobility Behavior

Identifying individual mobility patterns is of fundamental importance to understand the socioeconomic dynamics behind spatial and human movements. Indeed, early measurements on monkeys, albatrosses and marine predators [16, 21] suggested that animal trajectory is approximated by a Lévy flight model. Human trajectories are explained by random walk and diffusion models such as Lévy Flight (LF) [15], discovered by P. Lévy in the beginning of this century, and Continuous Time Random Walks (CTRW), introduced by Montroll and Weiss in 1965. While in the first the movements of an individual are calculated by random walk processes with a Lévy probability distribution of step lengths that is heavy-tailed, in CTRW the step lengths and the waiting times are distributed according to a Lévy distribution.

Some researchers have extended the use of the term ”Lévy flight” to include different human moving contexts. On the basis of these behaviors, assume that the flight distance is well approximated to a power-law function (Eq. 1), $f(x;a,\alpha)$, of the form:

$$\tilde{f}(k;a,\alpha) =< \exp(ikx) >= \exp(-|ak|^{\alpha}) \tag{1}$$

This distribution model is defined for $0 < \alpha < 2$, otherwise $f(x;a,\alpha)$ is not normalized $(\alpha < 0)$ or gets negative values and cannot be a probability distribution function (PDF) $(Var(f(x;a,\alpha)) = 0)$. Here a represents the units of length. Moreover, it can be thought of as the limit distributions for random walks with steps of sizes distributed (N) or the width of the PDF $f(\frac{x}{N};a,\alpha)$. Therefore, in context of the Lévy Flight (Eq. 2) the width is the flight length:

$$x(N) \sim N^{\frac{1}{\alpha}} \tag{2}$$

Now, we introduce the concept of a random waiting time from a one-sided probability distribution. To address random waiting times in the context of CTRW, we begin with Laplace transform where $f(t;a,\alpha)$ (a represents the units of time) $\tilde{\phi}(s;a,\alpha)$, $f(t;a,\alpha)$ is non negative for $0 < \alpha < 1$. For $\alpha = 1$ and $\alpha = 2$ we get two known special cases, Lorentzian and Gaussian[2] in Eq. 3.

$$f(x;a,1) = \frac{a}{\pi(a^2+x^2)} \quad f(x;\frac{\sigma^2}{2},2) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp(-\frac{x^2}{2\sigma^2}) \tag{3}$$

For anomalous diffusion we assume the waiting times PDF $\tilde{\phi}(s;\tau,\beta)$, and step length PDF $\tilde{f}(k;\lambda,\alpha)$. Expand it for small k and s and put it into the Montroll-Weiss equation and taking the first order terms gives the Eq. 4.

$$\tilde{f}(k,s) \sim \frac{\tau s^{\beta-1}}{\tau s^{\beta} + \lambda k^{\alpha}} \tag{4}$$

With these equations, three distinct hypotheses can be presented to explore the statistical properties of the population mobility pattern. Firstly, each individual follows a Lévy flight trajectory with jump size distribution. Then, there is a population heterogeneity based on individual distribution, corresponding to the inherent differences between individuals, and it coexists with individual Lévy trajectories.

2.1.2. Collective Mobility Behavior

Shifting the attention to the collective level, the scientific community has identified urban moving types based on communities. In 2012, Jiang found that the population can be clustered into several representative groups based on the kind of activity[10]. In its turn, Kang explored human movements in Singapore based on taxicabs usage data and observed a much higher number of trips within a community than trips across different communities [11]. In all of them there are connections between the social and spatial behavior of groups of individuals, and they are used to produce predictive models of mobility.

Although these comprehensive projects have the potential to revolutionize human mobility research, until now our knowledge of the correlation, and aggregation among data has been extremely limited. Using extensive individual mobility records resolved in both time and space can be aggregated for the study of individual traveling between different zones of an urban area. This study can be organized in the framework of Origin-Destination (OD) matrices to provide a relevant high-level picture of human mobility[1]. Therefore, in OD matrix (ODM), for each spatiotemporal trajectory T_k, there is a flow between the areas (i, j), noted $F(i, j)$ if the two following conditions are met:

1. $p_1 \subset$ i OR ($p_2 \subset$ AND $p_1 \in$ neighbors list of p_2)
2. $p_n \subset$ j OR ($p_{n-1} \subset$ jj AND $p_n \in$ neighbors list of p_{n-1})

where T_k is composed by a set of n consecutive points, noted T_k = p_1, p_2, ..., p_i,...,p_{n-1}, p_n, where: p_i = (x_i, y_i, t_i) is a record point having a spatial location (x,y) at moment t, $t_1 < t_2 < ... < t_{n-1} < t_n$, i= 1..n represents the number of points composing the T_k.

From conventional (e.g., based on car parking sensors) to technological (e.g., mobile devices), several methods have been explored for the estimation of ODMs. Anyhow, each method collects data that can be spatially and temporally analysed to identify human mobility patterns.

2.2. Mobile Network Data

For human mobility studies across many disciplines, mobile network data serves as a primary source of human footprints with georeferenced and time-stamped records of human communication activities [4]. They are increasing due to the growth of intelligent, and interconnected mobile devices as we shift into the Internet of Things (IoT) era. The shift to an IoT world means that mobile devices such as smartphones, smartwatches, and people are always connected. These personal mobile devices generate a great amount of data about the activities of individuals on a daily basis. This new data source can shed light on everyday human activity both at an individual level and at a collective level. Taking advantage of this approach, called mobile crowd sensing, mobility projects are designed to address the spatial nature of human mobility, to remain independent of social characteristics, and to be comparable across geographic regions and time.

Thanks to the massive adoption of mobile devices the IoT is focusing on operational efficiency and automation. This increasingly connected society together with sophisticated sensor systems has changed not only the data transfer speed but also the scalability, connectivity, and energy efficiency of mobile networks. Thus, for the development of the digital society and to make the IoT work, we need a powerful wireless network that connects different types of devices and transfers large volumes of data fast. This is where the 5G mobile technology can be useful [14].

The move to 5G mobile networks is an action in response to the growth of IoT, improving spectrum efficiency, reducing latency, better supporting mobility and high connection density. This network densification combined with advanced wireless transmission technologies enables a higher throughput per individual user and makes the IoT more available to all [12]. The effect of 5G will be felt on almost every industry like information technology, manufacturing, health etc. Hence, many institutions, technological companies and research organizations are collaborating to form committees towards the research of 5G Technology. For example, supported by the European Commission, the mission of the European 5G Observatory provides updates on all market developments, including actions undertaken by the private and public sectors, in the field of 5G. Since the launches of the first 5G commercial services in late 2018, many commercial 5G networks and services have emerged in Europe. Consequently, a total of 15 players have launched commercial 5G services as of year-end 2019 [6]. The French IoT networking company SigFox has set a target to secure ”global” coverage by setting up in China, India, and Russia, finally, and plugging three major gaps in its existing footprint [3]. Together, these projects are essential to explore the information that can be extracted to improve human lives. Based on it, the final goal is being able to transform acquired knowledge into new services for citizens.

3. Experimental Case Studies

In this section, we first contextualize two datasets of our study, followed by a description of them. Then, explain the aggregation of all data generated by human interaction with digital devices to explore individual and collective behaviors.

3.1. Contextualization

Human mobility patterns reflect many aspects of life, from the urban planning to daily commute patterns. In recent years, the prevalence of technologies and methods, such as Wi-Fi systems, 3G, 4G, 5G and Bluetooth channels that are fundamental for real-time communication of location-based data, has driven efforts to collect human mobility data and to mine patterns of interest within this data in order to promote the development of services. In this paper, we mine significant patterns in pedestrian mobility data using neural network technique, and therefore, we survey different approaches that analyze and learn human mobility patterns using Long Short-Term Memory (LSTM) method. But first, in data mining phase, we transform the LinkNYC Kiosks and NYC-311 service datasets into knowledge. This knowledge is expected to be a useful tool in decision-making processes. For an easier analysis and interpretation of knowledge we use the visualization process. Finally, based on the results of the population density in New York City obtained from data mining and visualization processes, given any location of urban area, time period or observed human mobility, we can model and forecast data trends to predict movements in the near future.

3.2. Data Sources

LinkNYC is a communication network that brings the fastest available free high-speed Wi-Fi, nationwide calling, a dedicated 911 button, charging ports for mobile devices, and access to selected websites to millions of New Yorkers. The dataset lists locations where a LinkNYC kiosk has been installed, including sites that have replaced public pay telephones (PPT). The census column displays the number of users that connect to LinkNYC devices on a daily basis. Moreover, we also include the weather condition attributes such as temp, precipitation, snowfall, snow/ice and depth.

The NYC-311 service requests aim to provide the public with quick, easy access to non-emergency government services and information, covering issues such as broken traffic lights, noise complaints, parking law enforcement, and potholes [5]. For this service New York City's authorities added two new attributes. The inclusion of the ”Intake Channel” attribute, when available, will indicate on which channel a service request was filed by a customer and Borough Block-Lot (BBL), providing parcel numbers identifying the location of buildings or properties. As well as in the study of LinkNYC Kiosks, our attention was also focused on one dataset between 2016 and 2018 and includes the weather condition attributes previously mentioned.

3.3. Data Pre-processing

For both datasets being studied, the Knime Platform is used to get a selection of suitable mining techniques. First, data is aggregated to then perform data cleaning. Aggregation is combining various attributes into the ”date” attribute (row data separated by a range of days). This environment helps identify missing values as well as values and events that occur repeatedly in the dataset to form certain patterns. While for the LinkNYC Kiosk dataset we replaced missing attribute values with the mean value for that attribute in the database, on 311 Service Requests a frequent pattern is applied to data because it doesn't take into account the sequence of events, but the frequency with which an item occurs. In addition, we merge the weather conditions dataset. This merge is only possible because three datasets of our work use the date as a primary key. A sample of collected data can be seen in Table 1.

Once the data mining process was chosen, and the appropriate dataset was accessed, extracted and prepared, input data must be provided in the amount, structure, and format suited to the modeling algorithm. Basically, data exploration will be used in understanding the prediction results of the data mining process. However, the section 3.4 already allows for the visualization and comprehension of the data.

Table 1. LinkNYC Kiosks and NYC-311 Service Requests datasets.

Date	Census	Temp (°C)	Humidity (%)	Precip	Wind (MPH)
2016-02-12	646	12	50.3	30.3	2
2017-01-01	888	-1	61.3	29.9	3
...	...	...	...	...	...

3.4. Gaussian Distribution

To extract maximum information from our data, we calculate the Gaussian distribution (also known as normal distribution). This distribution is a bell-shaped curve, and it assumes that any measurement census will follow a normal distribution with an equal number of measurements above and below the mean census. To verify that, using Eq. 3 the normal distribution of our data is shown in Fig. 1.

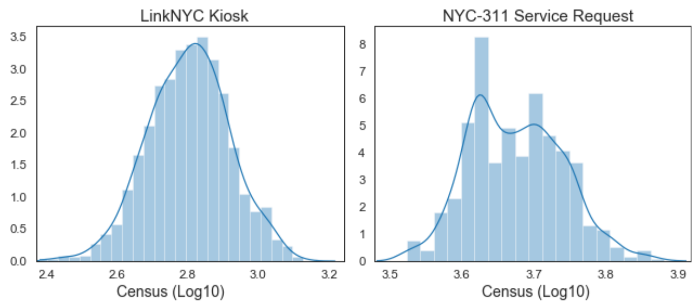

Figure 1. Normal distribution of data: density of population in city.

To determine if the data distribution strays from the normal distribution, we calculate the kurtosis and skewness. While the LinkNYC Kiosk data presents a kurtosis and skewness of 0.5 and 0.6, respectively, in the NYC-311 Service, these values are 0.2 and 0.6 respectively. Based on the results, the first dataset displays more satisfactory values. It indicates a Normal Distribution symmetrical in Kurtosis (e.g, values between -0.5 and +0.5) and a moderate skewness (e.g, values between -1 and -0.5 or 0.5 and 1). Therefore, unlike the distribution on the right side, the left side follows the normal distribution.

3.5. Model for predicting mobility

After the process of recognizing patterns, we train the data using a deep learning algorithm, for predicting and simulating human mobility. One of the best learning algorithms for time series data is Long Short Term Memory network (LSTM) [17]. To build it layer after layer, the Sequential class is used whereas Dense refers to a Deep multilayered neural network. Dropout Regularization is used to prevent overfitting or reduce complexity by randomly zeroing out a few units of different layers in a neural network based on its probability of retention. Moreover, the MixMaxScaler is used to scale the data to speed up the learning algorithm. While EarlyStopping is used to stop the Gradient at a point where it starts to increase again and prevent measure against overfitting. Before fitting the data, we normalize it and split into the train and test set. The model is trained over 100 epochs, number of batch size 70, and 50 LSTM neurons. To improve the performance of model we used k-fold cross-validation such as k=10 becoming 10-fold cross-validation. The outcome is the train and test loss of LinkNYC Kiosk and NYC-311 Service Request datasets.

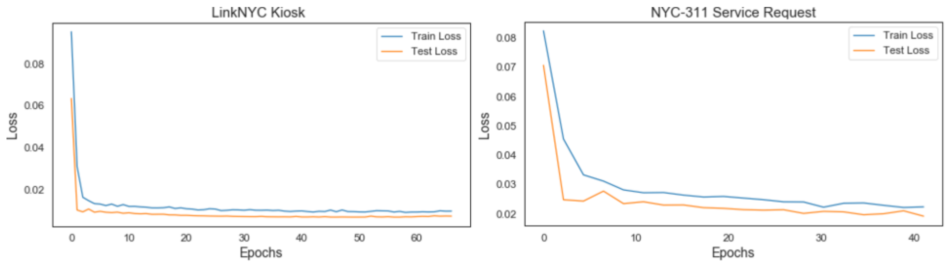

Figure 2. Data loss in the train and test sets.

From LSTM training, followed by running a few experiments, we can compare the performance between validation loss and training loss. Based on the results, the training loss is greater than validation loss for Fig. 2. This means that underfitting has occurred, respectively. Additionally, at the deepest level, we can quantify the loss function from Root Mean Squared Error (RMSE) and Root Mean Absolute Error (RMAE) values. They are used to evaluate the qualitative and quantitative trends in the data during validation and testing process. For the LinkNYC dataset the results of these errors are 0.18 and 0.14, but in the NYC-311 Services Requests dataset they are 0.15 and 0.12, so we classify the loss of both datasets as ”good”. Because, as loss measures the ”goodness” of a model, the smaller the loss, the better the model accuracy.

3.6. Results

Each mobile devices interaction dataset has near five hundred thousand rows of data. In this case, to predict the sequence of values we also use a deep multi-layered Long Short Term Memory network. This model's settings are the same as the ones used for the train and test loss function. The three-year dataset (2016–2018) was used for validation purposes, and the one-year dataset (2017) was used for testing purposes. Once the neural network is trained, we predict and test data for a one-day forecasting model.

LynkNYC Kiosks and NYC-311 Service Requests show no correlation between the observed and predicted values for forecasting one day for each scenario. In Fig. 3, there are essentially no significant differences between predicted cases. The results of the validation process of both datasets also indicate that the constructed model can achieve equally impressive performance even if parameters have been changed. For example, if the epochs vary from 70 to 100 or the number of units varies from 80 to 150. However, all of these changes do not significantly improve the forecast results. Additionally, they illustrate the weather condition does not influence the forecasted result of the population density interacting with devices (e.g, mobile devices and kiosks).

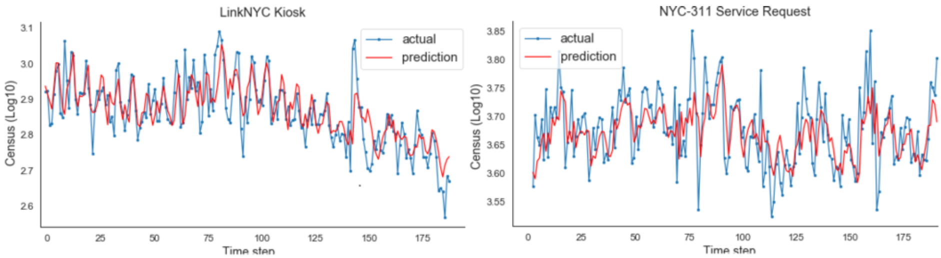

Figure 3. Actual and prediction data.

4. Conclusions

Human mobility has a significant impact within cities. With this paper we demonstrate that technology such as mobile phones, Wifi, and RFID devices can be used to reduce this impact. Moreover, due to its increasing sophistication, these technologies are currently considered one of the major trends. That's why, in the last decade, New York authorities have made a substantial effort to provide interactive devices to users. In response to this growth of available mobile technologies, a new wireless communication and network infrastructure that connect different types of devices operating at high speeds must be taken into account. Therefore, the move to 5G systems brings significant progress to mobility support and high connection density.

Based on these hyper-dense heterogeneous networks, we propose a Data Mining approach that analyses LinkNYC Kiosks and NYC-311 Service Requests data. Besides the spatial, temporal and particular attributes they also take into account information about weather conditions. Both problems were modeled as Long-Short Time Memory algorithm, where the aim was the prediction of the census in multivariate time series. Moreover, several different epoch and batch size configurations in deep learning model and feature selections (using distinct combinations of spatial, temporal and meteorological variables) were tested. In this step, we have concluded that after 100 epochs the outcomes showed no changes, that is, the predictive accuracy did not change.

To our knowledge, this is the first time these datasets are predicted using only meteorological based data and further exploratory research is required. As argued, predicting with this size of dataset was a challenging task. To improve it, we believe that additional information (not available in this study) is required, such as the type of events and reasons that attract people to city centers. In the future, we intend to consider a dataset from a larger timeframe. The proposed model could improve the human mobility prediction accuracy if it analyzed more than three years of data.

References

[1] Françoise Bahoken and Ana-Maria Olteanu Raimond. Designing Origin-Destination Flow Matrices from Individual Mobile Phone Paths The effect of spatiotemporal filtering on flow measurement. *26th International Cartographic Conference*, 2013.

[2] Peter Becker-Kern, Mark M Meerschaert, and Hans Peter Scheffler. Limit theorem for continuous-time random walks with two time scales. *Journal of Applied Probability*, 41(2):455–466, 2004.

[3] James Blackman. Sigfox sets 'Zero-G' strategy to underpin 5G/IoT system, secure global network, 2019.

[4] Guangshuo Chen, Aline Carneiro Viana, and Marco Fiore. Human Trajectory Recovery via Mobile Network Data.

[5] City of New York. Service request status, 2014.

[6] Commission European. 5G is really ON in Europe – 5G Observatory, 2020.

[7] Bruno Fernandes, Fábio Silva, Hector Alaiz-Moretón, Paulo Novais, Cesar Analide, and José Neves. Traffic Flow Forecasting on Data-Scarce Environments Using ARIMA and LSTM Networks. In *Advances in Intelligent Systems and Computing*, 2019.

[8] Bruno Fernandes, Fábio Silva, Cesar Analide, and José Neves. Crowd sensing for urban security in smart cities. *Journal of Universal Computer Science*, 2018.

[9] 19th EPIA Conference on Artificial Intelligence. https://epia2019.utad.pt/, 2019.

[10] Shan Jiang, Joseph Ferreira, and Marta C. González. Clustering daily patterns of human activities in the city. *Data Mining and Knowledge Discovery*, 25(3):478–510, 2012.

[11] Chaogui Kang, Xiujun Ma, Daoqin Tong, and Yu Liu. Intra-urban human mobility patterns: An urban morphology perspective. *Physica A: Statistical Mechanics and its Applications*, 391(4):1702–1717, feb 2012.

[12] Taras Maksymyuk, Mykola Brych, and Volodymyr Pelishok. Stochastic Geometry Models for 5G Heterogeneous Mobile Networks. *The Smart Computing Review*, pages 89–101, 2015.

[13] Walter Mischel. Toward a cognitive social learning reconceptualization of personality. *Psychological Review*, 80(4):252–283, 1973.

[14] Rupendra Nath Mitra and Dharma P. Agrawal. 5G mobile technology: A survey. *ICT Express*, 1(3):132–137, 2015.

[15] P. Lévy. Théorie de l'Addition de Variables Aléatoires. (Gauthier-Villars, Paris). 1937.

[16] Gabriel Ramos-Fernández, José L. Mateos, Octavio Miramontes, Germinal Cocho, Hernán Larralde, and Bárbara Ayala-Orozco. Lévy walk patterns in the foraging movements of spider monkeys (Ateles geoffroyi). Technical Report 3, Behavioral Ecology and Sociobiology, 2004.

[17] Tara N. Sainath, Oriol Vinyals, Andrew Senior, and Hasim Sak. Convolutional, Long Short-Term Memory, fully connected Deep Neural Networks. In *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, 2015.

[18] Ernst Schachtel. Personality. A Psychological Interpretation. *Zeitschrift für Sozialforschung*, 7(1):245–245, 1938.

[19] Ted Serbinski. Defining Mobility for the Automotive Industry, 2016.

[20] Fábio Silva and Cesar Analide. Ubiquitous driving and community knowledge. *Journal of Ambient Intelligence and Humanized Computing*, 2017.

[21] David W. Sims, Emily J. Southall, Nicolas E. Humphries, Graeme C. Hays, Corey J.A. Bradshaw, Jonathan W. Pitchford, Alex James, Mohammed Z. Ahmed, Andrew S. Brierley, Mark A. Hindell, David Morritt, Michael K. Musyl, David Righton, Emily L.C. Shepard, Victoria J. Wearmouth, Rory P. Wilson, Matthew J. Witt, and Julian D. Metcalfe. Scaling laws of marine predator search behaviour. *Nature*, 451(7182):1098–1102, feb 2008.