Big Data in Smart Cities: From Poisson to Human Dynamics

Antonio J. Jara, Dominique Genoud, Yann Bocchi

Institute of Business Information Systems

University of Applied Sciences Western Switzerland (HES-SO)

Sierre (Switzerland)

jara@ieee.org, {dominique.genoud, yann.bocchi}@hevs.ch

Abstract—**Big Data is conceived as the powerful tool to exploit all the potential of the Internet of Things and the Smart Cities. Historically several of the human-related behaviours have been modelled with Poisson distribution, but a new dimension of understanding about the human behaviours is reached through all the gathered data in the emerging smart environment. This work analyses the data from the European Project SmartSantander. This work has correlated the traffic behaviour with respect to the temperature in the Santander City. This has been presented as the evolution of both flows present a similar behaviour. The traffic distribution, aggregated by temperature bins, follows up a Poisson distribution model. Thereby, allowing interpolate and predict complex behaviours based on simple measures such as the temperature. At the same time, this data presents a burst behaviour (human dynamics), when the data is analysed in sequence, instead of aggregated by temperature bins. Therefore, this work concludes that human-related behaviours can be described with both, Poisson and Human Dynamics distribution, depending on how the data is represented and aggregated.**

Index Terms—**Big Data; Internet of Things; Smart Cities; KNIME; Human Dynamics.**

I. INTRODUCTION

Internet of Things (IoT) [1], [2] is generating prodigious amount of data, increasing sophisticated analytic mechanisms and tools, that are providing insight that allow us to operate the machines in entirely new ways, more effectively, and in a collaborative way.

The power of the data provided by all the resources that are being connected to Internet will bring a new conception of the world, where the big data analysis is required to take advantage of its potential for high-level modelling and knowledge engineering.

The potential of this flow of data from physical resources towards the Future Internet facilities is what we want to analyse in this work through Big Data Analytic tools.

The challenge for Big Data is to understand the interactions between smart objects and humans. The origin of the Internet has been a human-human type interactions, since the content was defined by humans to be consumed by other humans, but with the IoT the content is being defined by objects. Therefore, the interactions and influence over our lives is an open issue, and this needs to be understood how the IoT will play a key role in our Smart Cities and Smart environments.

In particular, this work will analyses these interactions and potential from the perspective of the *Human Dynamics*, the potential of the *Big Data* and *Smart Cities* to increase our quantitative and qualitative understanding regarding the human behaviours.

Therefore, this work will evaluate the potential of the real-time data to build data-driven models and patterns based on human dynamics, provide some data-driven models as proof-of-concept of the potential of the Big Data Analytic tool for Smart Cities, determinate patterns based on urban information with the data from the environmental status (temperature, CO, noise, light), integral traffic management (car presence, traffic), and citizen activity (augmented reality and participatory sensing), and finally understand how people use the city infrastructure in terms of mobility (transportation mode), environmental impact (noise, pollution), and consumption.

This work is focused on the SmartSantander testbed, which is a port city in the north of Spain, which offers fixed sensors that record temperature, traffic, noise, etc. The benefits of big data for smart cities include real-time systems monitoring, management, and optimization. Some of the areas where this benefits can be applied include flow of traffic, gas, water, and electricity; monitoring the condition of pipes and bridges; planning new public transport routes and grids for utility distribution systems; monitoring public health; and managing emergency response [9].

We analyse the correlation between traffic, temperature, season, and working day. These inter-relations define a complex network.

Complex networks mean that many interacting parts, which behave according to simple and individual rules, produce a globally coherent behaviour, i.e., emerging properties and patterns.

The challenge by what big data is required and all the analysis of these interactions as a network, is due that the behaviour at the systems as a whole cannot be predicted from the individual rules only, such as stated by Aristotle *the whole is greater than the sum of its parts* [15].

One of the reasons because the inter-connections/inter-relations among different variables in complex networks such as smart cities look random is due to the fact that we are not yet able to understand how all of them are inter-connected. Following the last results in network sciences [6], [7], we are aware that not everything is random, since everything is connected through networks. Therefore, we are starting

IEEE computer society

to disclosure the hidden patterns of the universe and the universality among them [4], [5].

These networks are physical, social, biological, cultural, and human networks, that govern how the different parts of the world operate, and how they affect our lives.

The challenge, for the Big Data and IoT, is to take benefit of this insight to really understand how things are interconnected. Then, we will be able in a close future, not only to understand, else able to predict, act, manage, and prevent these situations. Thereby, evolving from areas partially overlapped such as Big Data, IoT, Cloud Computing, physical devices, and humans to a common ecosystem, that will be able to act/operate, enhance and fix based on all this emerging knowledge and understanding, i.e., insight. This fusion among areas to build solutions able to solve real problems and act/operate, is denominated Cyber Physical Systems.

II. SMARTSANTANDER TESTBED

The port of Santander on the northern coast of Spain is the most data-intensive city in Europe. SmartSantander has some 18000 stationary and mobile sensors of various types thought the municipality of around 180000 residents [12]. These sensors monitor air pollution, noise, and other environmental conditions. Sensors buried in the pavement detect open parking spaces and relay that information to digital displays mounted at the major intersections to help guide drivers.

This infrastructure is available for researches, at the same time that it is continuously being used by the city services.

SmartSantander also features a smart phone application, denominated *PaceOfTheCity*, that allows residents to use an augmented-reality solution to interact with over 2600 optical and wireless tags at tourist attractions, bus stops, shopping centres, and other locations throughout the city to readily obtain online information about those places.

SmartSantander has defined an ecosystem, which it is not only a testbed for wireless sensor networks and Future Internet architecture, it is also opening new opportunities to be interconnected with Big Data and Cloud Computing to provide intelligence, in order to be able to understand the behaviours, and even define actions according to the information captured by the smart objects that are able around the emerging smarter cities.

In particular, SmartSantander EU Project is providing one of the major Smart Cities and IoT testbed in Europe. This testbed is providing data from noise, traffic, temperature, power consumption, parking pots, smart labels, and other environmental monitoring.

Figure 1 presents the part of the testbed used for this experimentation, composed of 97 temperature sensors, and 38 traffic sensors. The Figure 1 depicts that the areas where the traffic and temperature are measured are divergent. The same happens with the noise area. Consequently, it cannot be directly correlated some measure data such as the influence of traffic with the noise, and the direct influence of the traffic density (heat radiation from cars) with the temperature in that area.

III. KNOWLEDGE DISCOVERY FROM DATA

Knowledge discovery from data (KDD) is composed of multiple stages, covering from data analytic, to data mining. It is the confluence of multiple disciplines such as pattern recognition, statistics, visualization, machine learning, database technology, and other specific disciples depending on the problem, such as human dynamics in the scope of this problem.

The challenges in the knowledge discovery from the data, are the management of big and real-time data integration, and analytic for urban information, in order to build models and patterns about performance, infrastructure usage, urban information, and human dynamics. Some examples of these models and patterns are to identify the nature and cause of changes in the different streams in order to predict traffic congestions, look for patters to explain logical connection of knowledge at various point of time.

The patterns and behaviour evaluations, offers the opportunity to identify the truly interesting patterns representing knowledge based on the interestingness measures from the Smart City.

Regarding the behaviour, such as described by A. Barabasi [6], [8], it is very focused on bursts, one proof of this is the high number of peaks followed by period of inactivity, even during the working hours. However, it can be modelled with Poisson, in order to get a more general estimation.

Therefore, this scenario is probing the A. Barabasi conclusions about bursts, but at the same time, this also shows us that Poisson is a good model for the traffic behaviour based on the temperature. Similar conclusions about the proper Poissonian explanation of burst behaviours have been also offered for the Barabasi models in [11].

In order to see this correlation in more details, the knowledge presentation is presented through the visualization in the Figures 2 of the traffic aggregated by temperature bins. Thereby, this aggregation by temperature bins offer the opportunity to see clearly what is the direct influence of traffic with respect to traffic density.

The correlation, and aggregation of the data, between the temperature and the traffic, presents as several of the human behaviours a Poisson distribution model [10]. The Poisson distribution model allows us to determinate the quantity of traffic through the interpolation of the λ value. For example, for this specific case in August 2013, we have that $\lambda = 22$, i.e., the maximum traffic happens between at 22 degrees. Knowing that the distribution is Poisson, and $\lambda = 22$ for August. Then, the level of traffic for the rest of temperatures can be interpolated for that period of time. At the same way, it can be also seen that in December is the maximum traffic happens between at $\lambda = 10.5$ degrees. And, we can see as in April the traffic is much more distributed in a major range between 7 and 12 degrees, since this season (spring) presents a warmer weather, and not so characterized by cold or hot as winter and summer respectively.

Thereby, even when we have seen in the Figures 2a and 2b that the behaviour was based on bursts, i.e., when we analyze
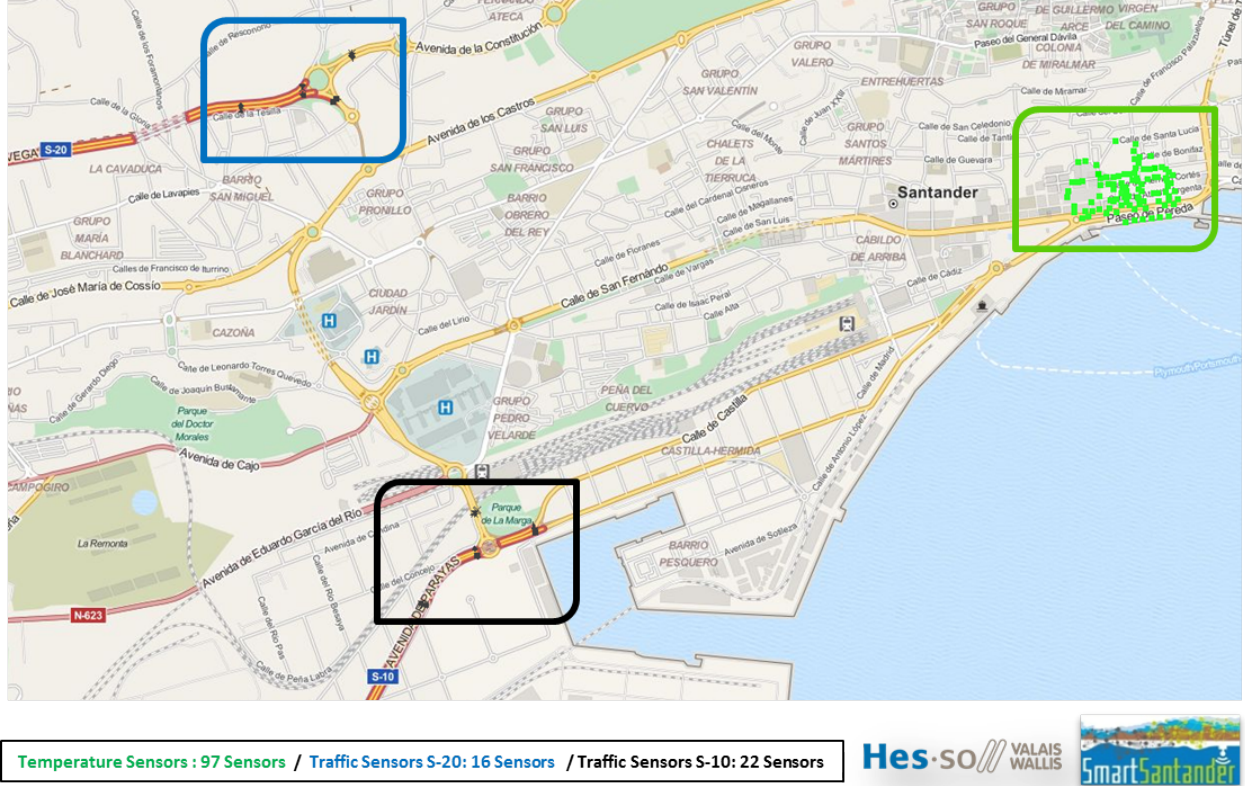
Fig. 1: SmartSantander TestBed (Traffic and Temperature Deployment).

the behaviour in detail, when we try to build a more general model, then we discover that such as defined historically, the Poisson is a good model to simplify and evaluate these variables.

Finally, the traffic density distribution have been also evaluated, where we can see in the plots from the Figure 3, that the majority of the time the traffic density is low, for that reason first column is always the higher, and that after that, the most common range of traffic is among 8% and 10%, which means a very fluid traffic. Finally, heavy tails are presented for high densities. This contrast among a long time of low activity, and heavy tails for high traffic density, is what present a bursts behaviour such as described by A. Barabasi in [6].

It is usual that these environments, where the time without activities, and with social influences such as time for lunch, and time to go to work, define the bursts, that are followed by long inactivities such as nights, and mid-term office hours.

IV. DISCUSSION

The human dynamics, and clear correlation among nature and our behaviours. Demonstrate that the nature is not a result of the chaos, and does not exist the lawlessness in the universe. If we see chaos or lawlessness, it is due to the lack of knowledge.

The lack of knowledge has been historically the main reason to does not understand the patterns, laws and complex networks inside of what we are just able to interpret as random.

The insight of the world will be one of the major success during the coming years, as a consequence of the new senses to see the world provided by the Internet of Things, the Big Data tools to extract knowledge from the data, and the advances in modelling and formalizing network sciences.

Parts of these synchronism between traffic and temperature, are biologically defined by the circadian rhythms. At the same time, it is clear that not only the temperature is an influence, else it is a complex network where all the people is influenced among them.

Another nature proofs of this are the fireflies that light up at the same time [14], ants colonies, neurons organizations, the evolution of species, are proofs of the existence of a universal connection, and collective intelligence. Such as mentioned by Aristotle in the 4th century BCE, *"the whole is greater than the sum of its parts"*.

This work has presented the analysis of the traffic and temperature correlations, where has been seen a Poisson distribution, even when the nature of the data has bursts behaviours, such as described in the Human Dynamics.
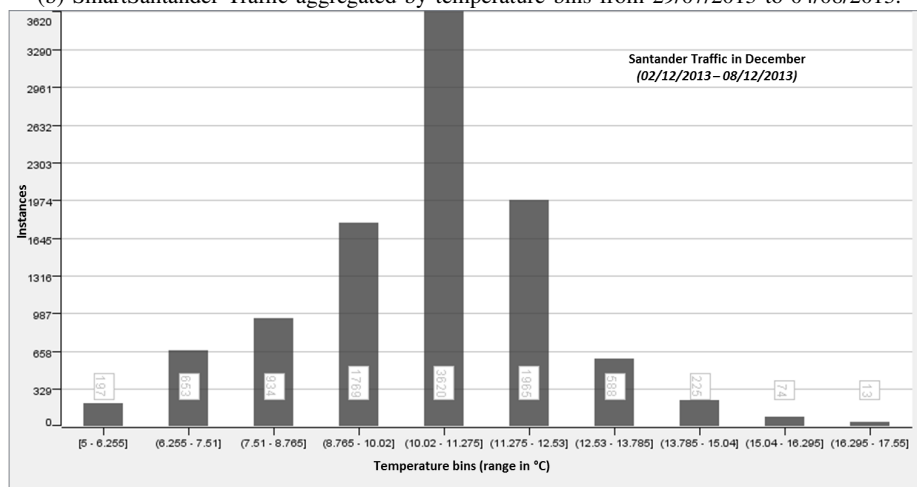
In order to present the bursts behaviour, it has also modelled the traffic density distribution, where it is evaluated that the
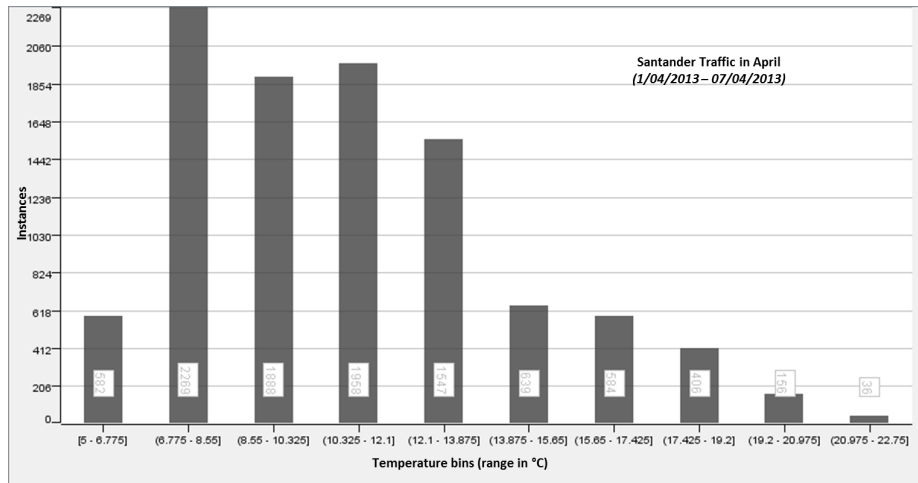
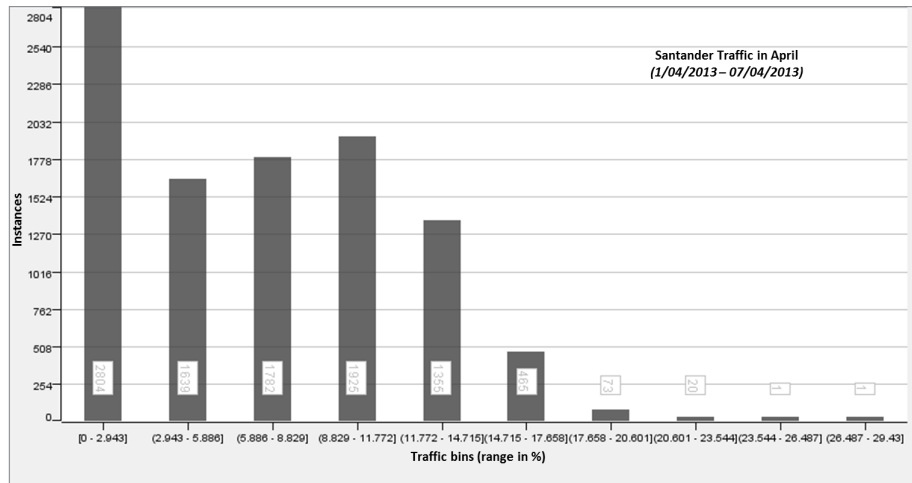(a) SmartSantander Traffic aggregated by temperature bins from 01/14/2013 to 07/14/2013.
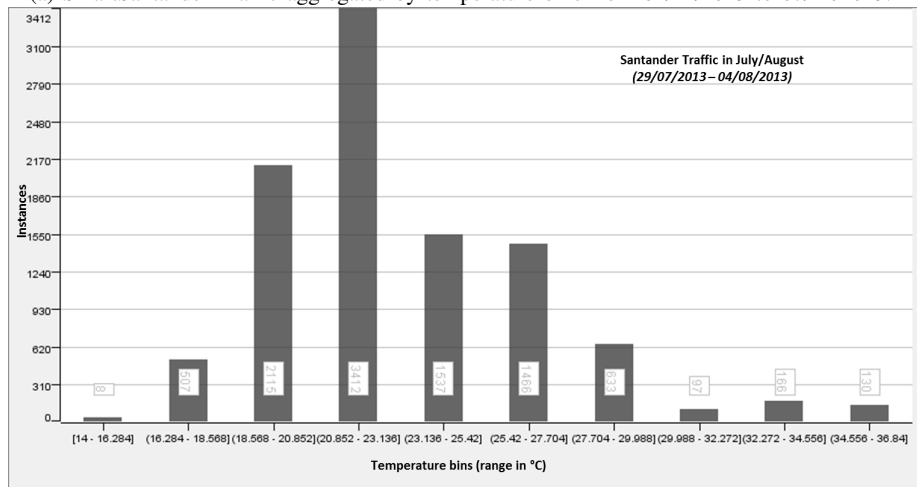
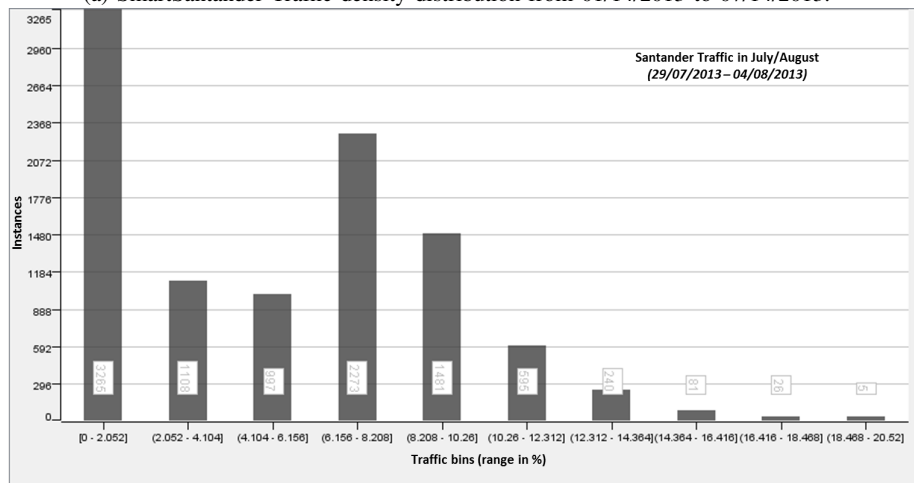(b) SmartSantander Traffic aggregated by temperature bins from 29/07/2013 to 04/08/2013.

(c) SmartSantander Traffic aggregated by temperature bins from 02/12/2013 to 08/12/2013.
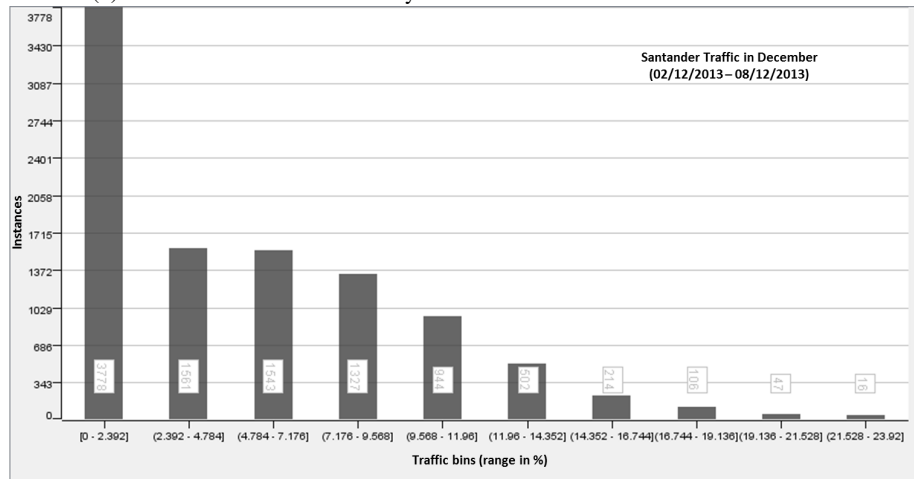
Fig. 2: Correlated SmartSantander Traffic with the temperature under different seasons.

(a) SmartSantander Traffic density distribution from 01/14/2013 to 07/14/2013.

(b) SmartSantander Traffic density distribution from 29/07/2013 to 04/08/2013.

(c) SmartSantander Traffic density distribution from 02/12/2013 to 08/12/2013.

Fig. 3: SmartSantander Traffic density distribution. The X-axis present the percent of traffic density, and the Y-axis the number of instances inside that density range.

majority of the time is presenting a low activity, due to the social cycles. Consequently, contrasting with the high activity periods (burst periods), such as early morning to move from home to the office, or lunch time.

Therefore, even when bursts behaviours are present, when it is analysed taking into account some of the influences in the complex network such as the temperature, a Poisson distribution is presented, offering therefore a modelling and simplification of the problem.

In definitive, it cannot be determinate that Poisson is not longer a valid model, since it is clearly modelling the correlation between temperature and traffic, but at the same time, we need to take into account the human dynamics and burst nature of the data.

V. Conclusions and future works

The entities in the world are not just able to act, else they are also able to sense. Smart cities, via IoT devices, are also being able to sense, in addition to act. Therefore, it is expected that, at the same way, that the rest of the nature, we can learn and synchronize, in order to optimize our energy, time, and life quality.

The recent changes in the world such as the lower cost for sensors, more capacity or data analysis, all the devices, machines, and appliances, communicating seamlessly with each other and with us, are allowing to build a world where information itself becomes intelligent, and comes to us automatically, when we need it, without having to look for it.

All this is impacting, since it is allowing us to shift to preventive condition-based maintenance, which means fixing machines just before they break, without wasting time, i.e., serving them on a fixed and planned schedule. Zero unplanned downtime, which means will not more power outages, no more flight delays, no more traffic congestion.

We are starting the new age of the human dynamics insight, and smart cities as SmartSantander are establishing the reference of infrastructure of what will be a mandatory need in a very close future.

This has been just an initial analysis of our experiences with Big Data and Smart Cities over European testbeds in IoT. Regarding the SmartSantander testbed, we have found the inconvenient of the lack of geo-correlations among the different sensors, which is limiting the direct correlation of noise, temperature, number of parking pots with traffic, etc.

For that reason, in addition to the SmartSantander testbed, the coming works are also promoting to gather more contextualized data through the crowdsensing of the temperature, and individual tracking informations from personal platforms such as smart phone. Thereby, it could be extrapolated to traffic forecast.

The ongoing work is focused on the data characterization, i.e., produce a description summarizing the characteristics of the human dynamic by traffic density, and other factors, in order to offer classes such as fluid and heavy traffic, which can be extrapolated from the difference between working day/weekend, temperature, season, and hour of the day.

This data characterization will allow to carry out data discrimination among classes, in order to offer a comparison of the general features of targeting classes of traffic and behaviours with the general features of other cities, which could act as contrasting classes.

Finally, it could offer the opportunity to develop new applications and solutions, to offer prediction of the traffic, and consequently prevention actions such as suggest usage of public transport, or recommend a time to depart in order to arrive on time to the office or for a meeting.

VI. Acknowledgement

References

[1] Jara, A. J., Varakliotis, S., Skarmeta, A. F., Kirstein, P. 2014. *Extending the Internet of Things to the Future Internet through IPv6 support*. Mobile Information Systems, 10(1), 3-17.

[2] Jara, A. J., Zamora, M. A., Skarmeta, A. 2012. *Glowbal IP: An adaptive and transparent IPv6 integration in the Internet of Things*. Mobile Information Systems, 8(3), 177-197.

[3] Albert, R., Barabasi, A. L. 2002. *Statistical mechanics of complex networks*. Reviews of modern physics, 74(1), 47.

[4] Gonzalez, M. C., Hidalgo, C. A., Barabasi, A. L. 2008. *Understanding individual human mobility patterns*. Nature, 453(7196), 779-782.

[5] Barzel, B., Barabasi, A. L. 2013. *Universality in network dynamics*. Nature physics, 9(10), 673-681.

[6] Barabasi, A. L. 2010. *Bursts: the hidden patterns behind everything we do, from your e-mail to bloody crusades*. Penguin.

[7] Vzquez, A., Oliveira, J. G., Dezs, Z., Goh, K. I., Kondor, I., Barabasi, A. L. 2006. *Modeling bursts and heavy tails in human dynamics*. Physical Review E, 73(3), 036127.

[8] Barabasi, A. L. 2002. *Linked: How everything is connected to everything else and what it means*. Plume Editors.

[9] Jara, A. J., Zamora, M. A., Skarmeta, A. 2009. *An ambient assisted living system for telemedicine with detection of symptoms*. Bioinspired Applications in Artificial and Natural Computation, Springer Berlin Heidelberg, 75-84.

[10] Haight, F. A. 1967. *Handbook of the Poisson distribution.*.

[11] Malmgren, R. D., Stouffer, D. B., Motter, A. E., Amaral, L. A. *A Poissonian explanation for heavy tails in e-mail communication*. Proceedings of the National Academy of Sciences, 105(47), pp. 18153-18158, 2008.

[12] Sanchez, Luis. 2011. *SmartSantander: The meeting point between Future Internet research and experimentation and the smart cities*. Future Network and Mobile Summit (FutureNetw), IEEE.

[13] Berthold, M. R., Cebron, N., Dill, F., Gabriel, T. R., Ktter, T., Meinl, T., Wiswedel, B. 2008. *KNIME: The Konstanz information miner*, pp. 319-326, Springer Berlin Heidelberg.

[14] Watts, D. J., Strogatz, S. H. 1998. *Collective dynamics of small-world networks*. nature, 393(6684), 440-442.

[15] Hofstadter, D. R., Godel, Escher. 1979. *An eternal golden braid.*