SOFTWARE – PRACTICE AND EXPERIENCE
*Softw. Pract. Exper.* (2014)
Published online in Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/spe.2274

# Big data for smart cities with KNIME a real experience in the SmartSantander testbed

Antonio J. Jara*,†, Dominique Genoud and Yann Bocchi

*Institute of Information Systems (IIG), University of Applied Sciences Western Switzerland (HES-SO), Technopole, 3, Sierre, Switzerland*

SUMMARY

Big data techniques are conceived as the powerful tool to exploit all the potential of the Internet of Things and the smart cities. A new dimension of understanding about the human behaviours is expected to be reached through all the gathered data in the emerging smart environment. The described potential, so-called Human Dynamics, pursues to describe in real-time the human behaviours and activities. This work presents our experiences for big data analytics in smart cities, in terms of sensors data management, data fusion and knowledge discovery from the data. The data used is from the European Project SmartSantander, where the traffic behaviour has been correlated with respect to the temperature in the Santander City. The evolution of both flows present a similar behaviour, in detail, a fine grain correlation is discovered. On the one hand, the traffic distribution, aggregated by temperature bins, follows up a Poisson distribution model. The Poisson modelling allows to interpolate and predict complex behaviours based on simple measures such as the temperature. At the same time, on the other hand, the isolated traffic density distribution, without taking into account the temperature-based aggregation has been analysed. The traffic distribution has presented a burst behaviour, which presents a closer model to the human dynamics. Therefore, this work presents as the smart cities data can be modelled as Poisson or Human Dynamics (burst models). Finally, reference data analytics process, data sets and models are offered for the Open Source Data analytics platform Konstanz Information Miner (KNIME). Copyright © 2014 John Wiley & Sons, Ltd.

Received 5 March 2014; Accepted 28 April 2014

KEY WORDS: big data; data fusion; knowledge discovery from data; KDD; smart cities; SmartSantander; Internet of Things; human dynamics; KNIME

## 1. INTRODUCTION

Future Internet is the engine to reach the next generation of infrastructure, services and solutions to facilitate the sustainable development of its industry, buildings and citizens. Initial proofs of the potential of the Future Internet are found in the integration of the real-world in Internet through the smart objects.

Smart objects are being extended to support and enable an extended range of solutions based on cellular infrastructure and wireless sensor networks. Some examples of smart objects in our daily life are watches, clinical devices, building automation sensors, security networks, access control systems, smart phones, tablets, smart TVs and cars.

Their flexibility and capabilities are being extended, nowadays, with the infrastructure capacities to provide an Internet access to the cellular networks (M2M), and wireless sensors networks through technologies such as IPv6 Low Power Wireless Personal Area Networks (6LoWPAN) [1–4].

These new Internet-based features are making them more ubiquitous in the different environments and accessible by the systems and users. Consequently, the number of sensors continues growing

*Correspondence to: Antonio J. Jara, Institute of Information Systems, University of Applied Sciences Western Switzerland, HES-SO, Technopole, 3, Sierre, Switzerland.
†E-mail: jara@ieee.org

exponentially, estimating that by 2020 between 50 to 100 billion of devices will be connecting all the things around the user [5].

These billions of devices are destined to improve our quality of life, security and performance. This will occur thanks to the potential from the interconnection of information, objects and people through the deployment of Internet of Things (IoT)-based solutions such as smart cities and the usage of all these data through big data analysis.

The challenge is to define and understand the interactions between smart objects and humans. The origin of the Internet has been a human-human type interactions, because the content was defined by humans to be consumed by other humans, but with the IoT/Web of Things, 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 analyse 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.

The following subsections describe, in detail, each one of these enablers.

### *1.1. Internet of Things*

Internet of Things is the marriage of minds and machines, that is, the union of big data running on the cloud computing platforms and physical devices/things.

Several enablers have also influenced to make possible this relationship. First, Internet has been a crucial enabler to reach this end-to-end connection [6]. Second, Web of Things defines common protocols to interact among them and third, semantics to offer a language among them in order to reach an understanding [7]. Initial consequences and benefits of this marriage have been the reduction of development costs, because the simplification of interaction among systems through the Internet and REST-based architectures, the integration of heterogeneous systems, the cooperation among them, re-use of the data, devices and extending functionality to new solutions, never before imagined.

In other words, IoT is generating prodigious amount of data, increasing sophisticated analytics 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 the 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 analytics tools.

This work will analyze 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 Analytics tool for smart cities, determinate patterns based on urban information with the data from the environmental status (temperature, CO, noise and light), integral traffic management (car presence and 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 and pollution) and consumption.

### *1.2. Smart Cities*

The concept of smart city has gained popularity in Europe during the last years; however, it was coined at the beginning of 90s [8].

A smart city means how the urban infrastructure is evolved through the Information and Communication Technologies. The goal of promoting the use of innovative technologies from the Information and Communication Technologies area is to satisfy the challenges of the cities in terms of sustainability (e.g., water, gas and energy), to the social demand for real time information



DOI: 10.1002/spe

(e.g., parking, public transport and weather) and to the emergence of the Future Internet-related technologies.

The motivation to deploy smart cities is to make the use of resources more efficient, to increase services quality, to identify new needs, to provide information to citizens in real time and to create a more sustainable path of economic and social development.

Smart cities need to employ the emerging technologies to optimize the performance of their public services, infrastructure (i.e., streets, car parks and shopping centres) and culture (i.e., information points, tourist office, museums and highlight places).

This optimization and efficient usage of the resources in a smart city is not independent of their citizens. For that reason, they require to carry it out in an interactive way, that is, involving to the citizens. The challenge, consequently is to enable its infrastructure with solutions to facilitate user interaction with a wide range of urban elements and information. This interaction can be carried out with infrastructure elements, but also through applications integrated into hotspots, mobile phones or available through the Web.

In addition to the technologies deployed to interact with the people, the ecosystem is also composed of sensors to gather information, actuators to react against events and the backend managing big datasets that make you understand what is happening in the city and allow you to provide some information and services to citizens more efficiently.

In detail, we will address the traffic monitoring correlation with the temperature, in order to evaluate the potential in the future of information to the citizens about predictions of traffic status, in order to prevent delays and congestion, some examples of interaction with citizens are presented in [9], but these solutions are at the road level, the objective is to predict to prevent even before that departing through personal devices.

The traffic monitoring is one of the most extended infrastructures in cities [10]. From traffic in urban roads to highway traffic all requires an analysis and integration of the data, in order to model and predict the evolution of the traffic. Some existing works are presented in [11, 12].

### *1.3. Big data*

The analysis of large data sets (big data) has gained considerable interest over the past decade.

IoT has enabled to record vast amounts of data that need to be analysed in order to reveal interesting relationships. Methods, techniques and tools from diverse disciplines have been combined to help analyse such datasets [13].

Specifically, the approach carried out for this work is based on applying insights from one domain (big data algorithms) to data from smart cities based on IoT.

From this fusion between smart cities and big data has been defined the multidisciplinary field urban informatics [14], where the difference with the general term big data, in this scope, is that big data needs to focus on enhancing the dynamics of the cities.

For this purpose, several cities around the world are deploying a wide range of stationary and mobile sensors.

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 and so on.

In addition to the fixed infrastructure for specific purpose, the existing infrastructure such as the proliferation of cameras are already monitoring the movement of pedestrians and vehicles should be considered, in conjunction with the opportunistic data from crowd-sourced sensing can be provided by the mobile phones, personal sensors that record the location, activity and physiology of individual citizens, and the social networks data sources such as Twitter feeds and news, should be also considered.

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.



DOI: 10.1002/spe

The challenge is not to build a complex general climate model based on the temperature and humidity values from the hundreds of sensors deployed in the Santander City, it is not either to build a complex model about the transportation, pollution, noise levels and so on. The challenge is to combine all of them together to build a good predictive model, which are data-driven solutions that allow us to understand much better the dynamics of the society, and most importantly, to provide back all this data converted into knowledge to the citizens in order to enhance their behaviours and help them to enhance their quality of life.

For example, real-time flight data can be used to better manage taxi traffic, agenda organization, reduction of stress through preventive planing and finally other environmental factors such as pollution.

Regarding this work, our goal is to extend big data, IoT and Smart cities potentials through the facilitation for end-users from a wide range of sectors to be benefited and make use of these technologies.

The first step is to offer a mechanism through which they express their problems. Thereby, IoT, big data and smart cities integration will be driven by real problems, and for that purpose, we need to facilitate tools for the expression (modelling), evaluation (execution) and deployment (operational model) phases.

For this purpose, we are using KNIME platform [15]. We have chosen this platform because it is an open source big data analytics, reporting and visualization tool. This offers a visual graphical user interface, based on workflows, to design and debug the data analytics processes. In addition, this platform is integrated with Weka, R, Hadoop, and can be also linked to any third party tool. This offers a very powerful, intuitive and extensible usage for other institutions, researchers and citizens.

This paper will explore the smart cities data, since on the one hand, it is the common connection environment in our society, therefore, it is an ideal environment to explore universe rules, and on the other hand, they are an emerging source of open and multi-parametric data.

Finally, the motivation of this work is to obtain an experience of apply big data techniques to smart city data.

## 2. HUMAN DYNAMICS

This work will 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, that is, 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* [16].

Complexity used to be the result of the lack of understanding about the underlay laws, that is, simple rules, which drive the system. For that reason, the methodology is based on simplifying complexity to analyse complex patterns and discover correlations that can lead us to an explanation. Some techniques for this purpose are, on the one hand, fusion-fission dynamics based on regularly split and merge data in different subgroups to look up correlation among them, on the other hand, long-term association patterns by applying techniques with network analysis.

One of the reasons why 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 [17, 18], we are aware that not everything is random, because everything is connected through networks. Therefore, we are starting to disclose the hidden patterns of the universe and the universality among them [19, 20].

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 or able to predict, act, manage and prevent these situations. Thereby, evolving from areas partially



DOI: 10.1002/spe

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, that is, insight. This fusion among areas to build solutions able to solve real problems and act/operate, is denominated Cyber Physical Systems.

Following the statement of Rene Descartes, *The world is written in mathematical language* [21], and consequently, as an instance of this statement. The cities and human behaviours in the city are also written in mathematical language.

Specifically, we need a scientific theory of cities. This means quantifiable, relying on generic principles that can be transferred into models and finally, into predictive frameworks.

Cities are just a physical manifestation of your interactions, our interactions and the clustering/grouping of individuals.

## 3. SMARTSANTANDER TESTBED

The port of Santander on the northern coast of Spain is the most data-intensive city in Europe. SmartSantander has some 18 000 stationary and mobile sensors of various types thought the municipality of around 180000 residents [22]. 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. 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 that 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.

However, it is offering opportunities to carry out more general correlations among the traffic based on the mean temperature of the city, because it is one homogeneous variable among different areas. Following this statement, some variables such as temperature are also offering a big opportunity for data aggregation.

The coming section, about knowledge discovery from data, gives into details the data aggregation carried out over the temperature, and the correlations and conclusions reached among the temperature, traffic and human dynamics.

### *3.1. SmartSantander traffic sensors deployment*

SmartSantander offers a total of 38 traffic sensors. These sensors have been allocated in the center of the city, specifically in the main two nerves of the city. On the one hand, 22 sensors around S-10 highway (bottom subfigure in Figure 2), and on the other hand, 16 sensors around S-20 highway (top subfigure in Figure 2). The deployment covers multiple lanes in each direction.



DOI: 10.1002/spe


Figure 1. SmartSantander testBed (traffic and temperature deployment).

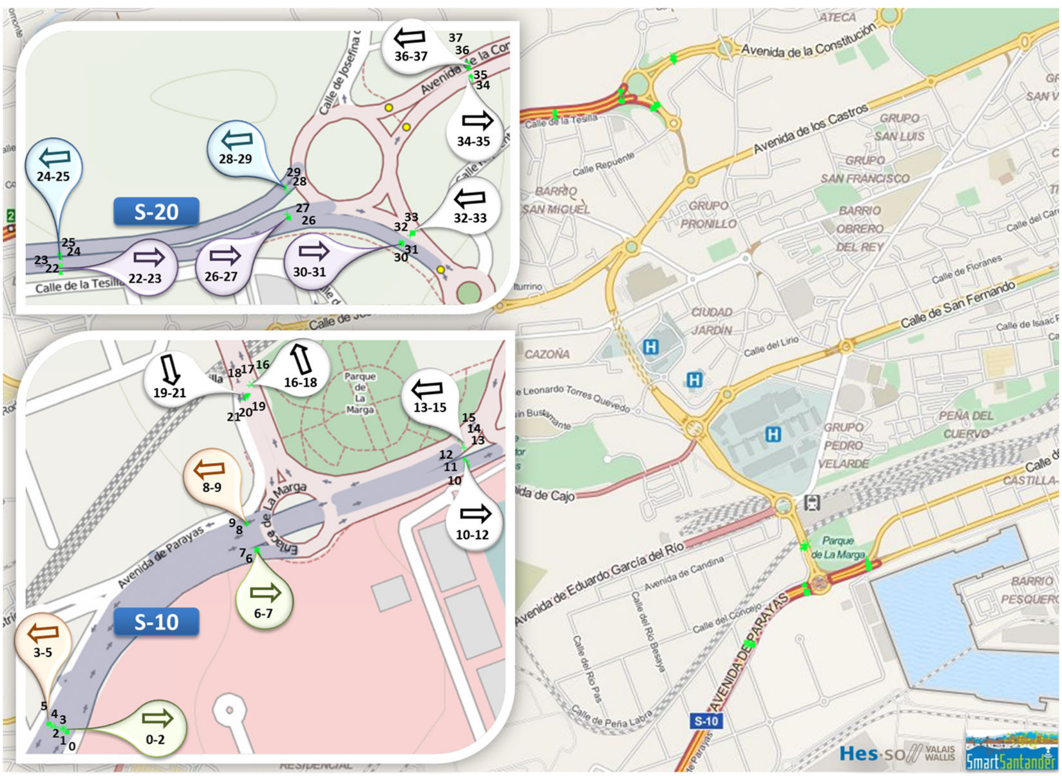



Figure 2. SmartSantander traffic sensor deployment.



DOI: 10.1002/spe


Figure 3. SmartSantander traffic sensor deployment.

### *3.2. SmartSantander temperature sensors deployment*

SmartSantander offers a total of 97 sensors for temperature monitoring. Figure 3 presents the area covered by these sensors. Because the sensors are highly redundant, they will be aggregated in order to simplify the data analysis.

## 4. KNOWLEDGE DISCOVERY FROM DATA

Knowledge discovery from data (KDD) is composed of multiple stages, covering from data analytics 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 analytics 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 following subsections present our experiences for each one of the stages from the data manipulation to the knowledge discovery, going through the data analytics.

### *4.1. Data reading, cleaning, selection, transformation and integration*

The first big challenge for big data management in smart cities is the huge data manipulation required. Figure 4 presents the workflow in KNIME with the KDD process[‡]

[‡]It has started reading the data from files, although it is also supported direct query to REST Services, but since SmartSantander services are limited to 10 000 records, it is required to store them for long-term analysis.



DOI: 10.1002/spe


Figure 4. KNIME workflow for knowledge discovery from data process during the first week of December 2013.

The data cleaning process is focused on removing noise, inconsistent data and format adaptation. KNIME offers multiple nodes to rename columns, change type (e.g., string to number, string to date) and other more useful ones such as the sorter data.

For example, in the SmartSantander testbed, the *String to Date* node has been used to round off the time, because every sensor is monitored every minute, but they are not synchronized at the same second. Therefore, this offers different dates for data that belong to the same minute range. For that reason, this node has been used in order to make feasible the data aggregation of all the measures from the different sensors for each sampling period.

In addition, for the data selection, complementary to the basic operations, KNIME offers *Java Snippets* nodes. They allow to manipulate the data at a row level, it allows to build new columns based on the data from other ones and remove some rows based on some rule. For example, we have used this resource to remove out of range data and for the event detection when the traffic density and temperature are over a pre-defined threshold. Thereby, we have carried out the data selection to the data relevant to the analysis tasks, and also generation of events that can analysed to model behaviours.

Regarding the data transformation, KNIME offers a very powerful *JPython script* node, which offers all the simplicity from Python over the Java objects that define and describe the data. Thereby, this node, as an extension to the *Java Snippets* node, allows to manipulate the data table, that is, iterating between different rows, in order to aggregate, correlate and build new data. This is highly useful for complex event detection over time series, and for aggregation of data from multiple sources. In the SmartSantander testbed, the data from each sensor at each minute is a new data entry, that is, all the temperature data of 1 min in the city is offered in 97 data entries (rows), at the same way for traffic (offering 38 data entries). This data transformation node has allowed us to aggregate all the data entries per minute, that is, a single data entry with the information of 97 temperature sensors, and another one with the 38 traffic sensors, where each sensor is a different column. This aggregation allows at the same time to calculate the mean, standard deviation and other statistics per minute, thereby, making easier the analysis tasks.

Finally, the data integration stage is where multiple data sources may be combined. For the SmartSantander testbed, the data from traffic and temperature are combined.

### *4.2. Data analytics and visualisation*

As a result from the previous stages, the data from the traffic has been aggregated for each highway (S-10 and S-20), offering the mean value of the traffic in S-10 and the traffic in S-20 for each minute.



DOI: 10.1002/spe

In addition, because both highways presented a very similar behaviour, they were aggregated in *meanGlobal*; the global traffic from the city. At the same way, the temperature has been aggregated in *Mean*.

Figure 5 presents the traffic for S-10, S-20 and *meanGlobal* traffic. This shows that traffic in S-10 and S-20 present a very similar behaviour, with a higher density in S-10. This correlation among both traffic flows offer the opportunity to detect accidents or anomalies.

In addition, Figure 5 presents the plot of the mean temperature (*Mean*) during the first week of April, in conjunction with the mentioned traffic. This plot depicts as the evolution of the traffic and temperature follows up the circadian rhythms based on the sun influence, i.e, the cycle of the day. In 2008, we started to analyse the influence of circadian rhythms [23] in the human behaviours.

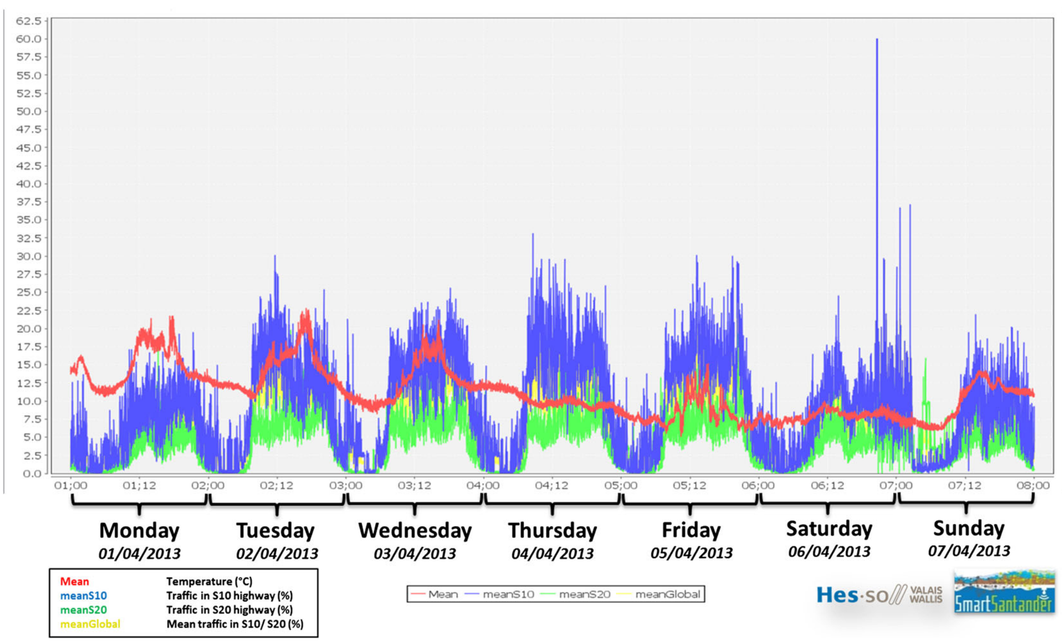



Figure 5. First week of April plot of the traffic for S-10 and S-20 highways, mean temperature and mean traffic.

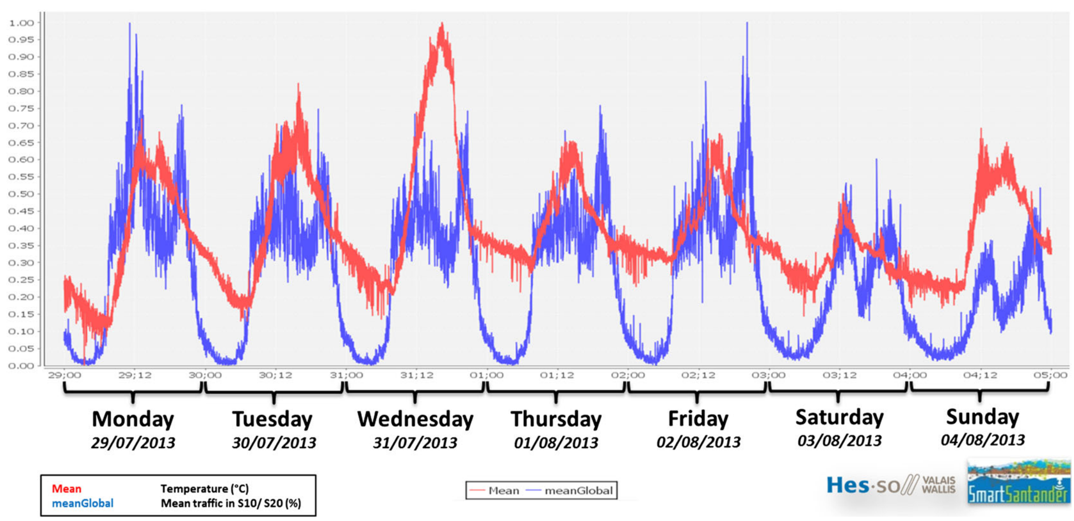



Figure 6. First week of August plot of mean temperature and mean traffic (normalized).



DOI: 10.1002/spe

Therefore, the first conclusion that we can reach is that the human behaviour is synchronized with the solar cycle, and consequently the temperature is synchronized with the solar cycle too.

We can see as also the working time between 7:00 and 21:30 is the most active, at the same time that also is the higher temperature, during the working days, that is, from Monday to Friday.

The situation for Saturday is very interesting, because this presents a clear decreasing after 14:00, this demonstrate that several places are just working half-journey, and that also is an important date for lunch meetings, because people, as a difference with respect to the other days, reduce significantly his activity.

Regarding, the rest of the weekend, that is, Saturday afternoon and Sunday is more homogeneous, with lower traffic, because it also has a lot of influences of family that goes out of the city, or stay at home. At the same time, it is clear as the influence of lunch time is highly significant.

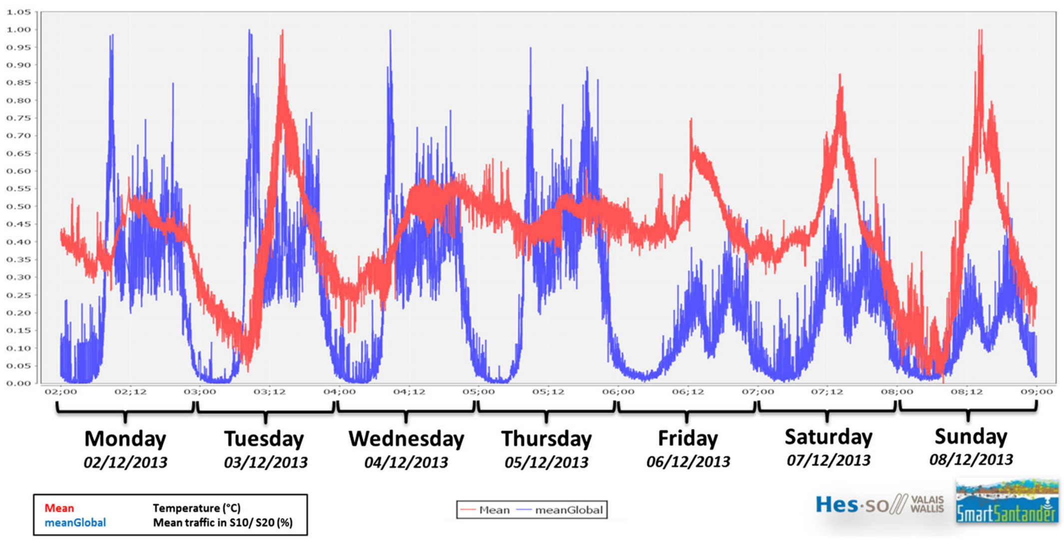



Figure 7. First week of December plot of mean temperature and mean traffic (normalized).

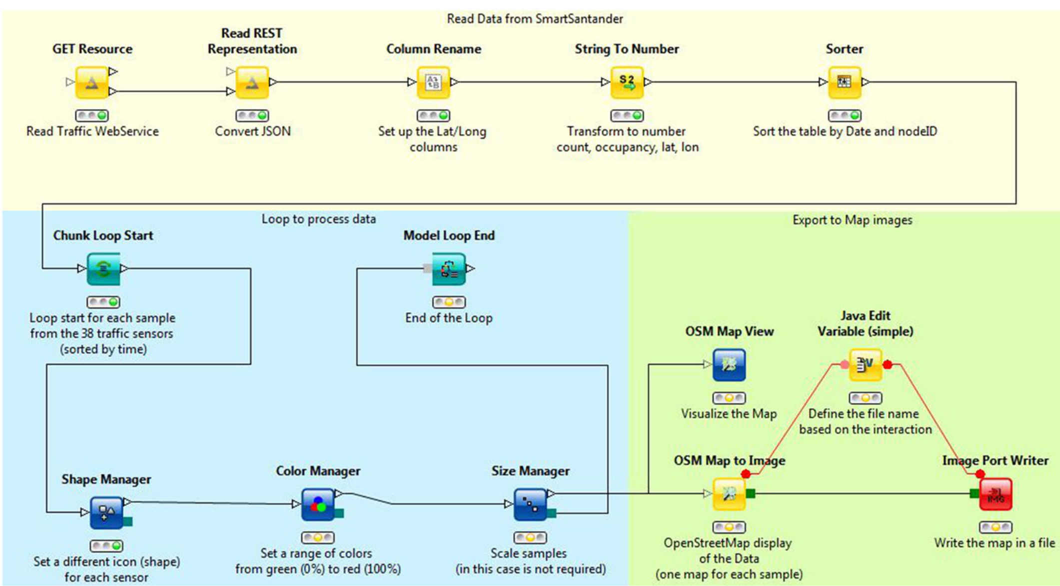



Figure 8. Workflow in KNIME to build frames with the status of the traffic status on Open Street Map (one frame per minute).

Figure 6 presents the plot of the mean temperature (*Mean*) during the first week of August, in conjunction with the mean traffic of the S-10 and S-20 highways (*meanGlobal*). Both of them have been normalized in order to correlate them. The flows present a very similar behaviour.

In addition, to confirm the initial conclusions about the human dynamics of the people, we can also see that from the numerical point of view, it is presenting a correlation of 57.4% in a fine grain level.

We can see in Figure 7 a similar correlation and behaviour during December.

Finally, Figure 8 presents the KNIME workflow to export the data through visualization on Open Street Maps. This offer nodes for shaping, define color and sizes of the representation of the nodes in the map. This example also presents the potential of KNIME to build loops, and also the flexibility to transfer *flow variables* among different nodes (red rows between *OSM Map to Image*, *Java Edit Variable*, and *Image Port Writer* nodes), in order to exchange data that is not directly present in the data tables. For example, in this case, we are transferring the image number, in order to build a new image for each iteration. After this, a video can be build with all the images, in order to see in motion the evolution of the traffic in the city.

The next section presents in more details the analysis of the correlations and behaviours determination.

### *4.3. Knowledge discovery*

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 [17, 24], 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 [25].

In order to see this correlation in more details, the knowledge presentation is presented through the visualization in Figure 9 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.

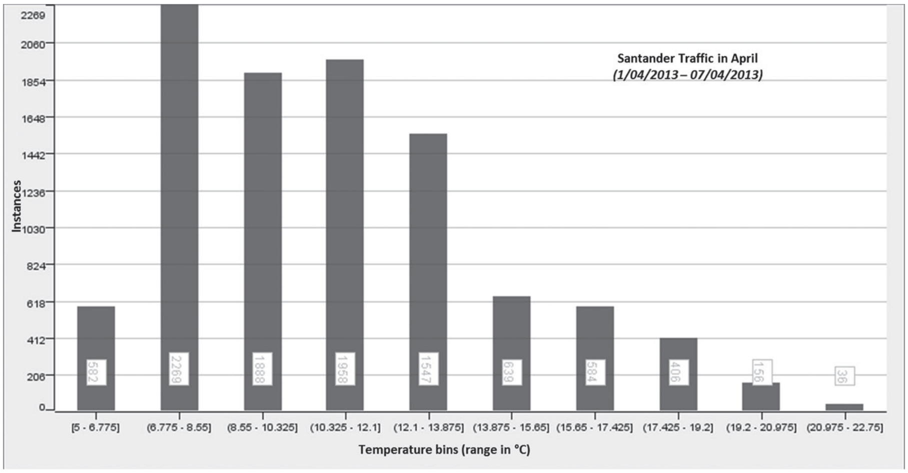
The correlation, and aggregation of the data, between the temperature and the traffic presents as several of the human behaviours a Poisson distribution model [26]. The Poisson distribution model allows us to determinate the quantity of traffic through the interpolation of the $\lambda$ value. For example, for this specific case in August 2013, we have that $\lambda = 22$, that is, the maximum traffic happens between 22°. 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 $\lambda = 10.5°$. And we can see as in April the traffic is much more distributed in a major range between 7 and 12°, because 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 9a and 9b that the behaviour was based on bursts, that is when we analyse 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.

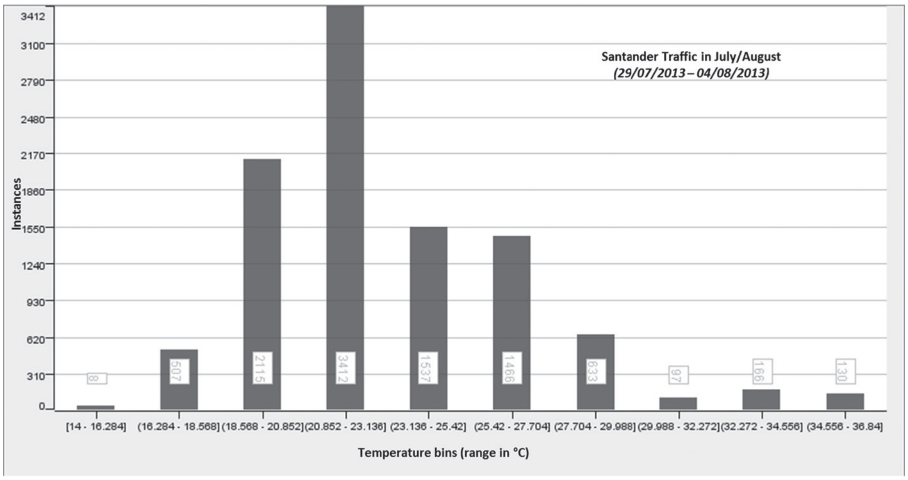
Finally, the traffic density distribution have been also evaluated where we can see in the plots from Figure 10 that the majority of the time the traffic density is low, for that reason first column is always the higher, and that after that, the most common range of traffic is among 8% and 10%, which 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 [17].



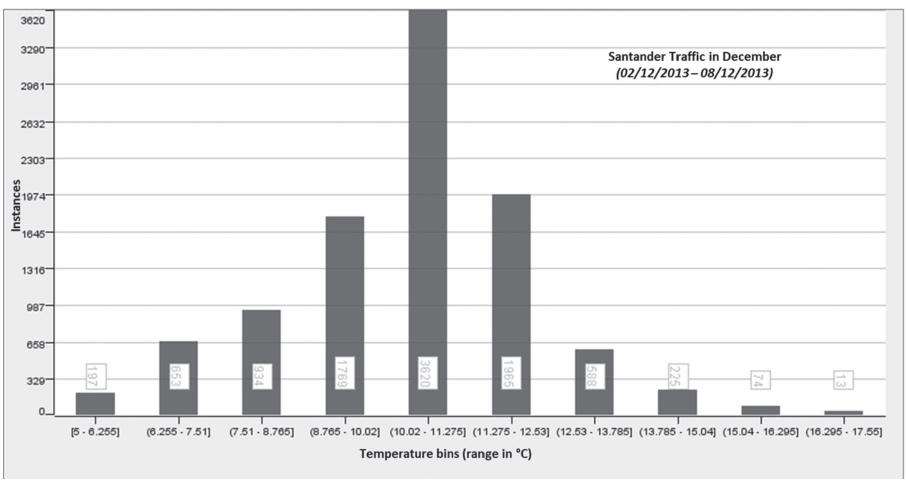
DOI: 10.1002/spe




(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.

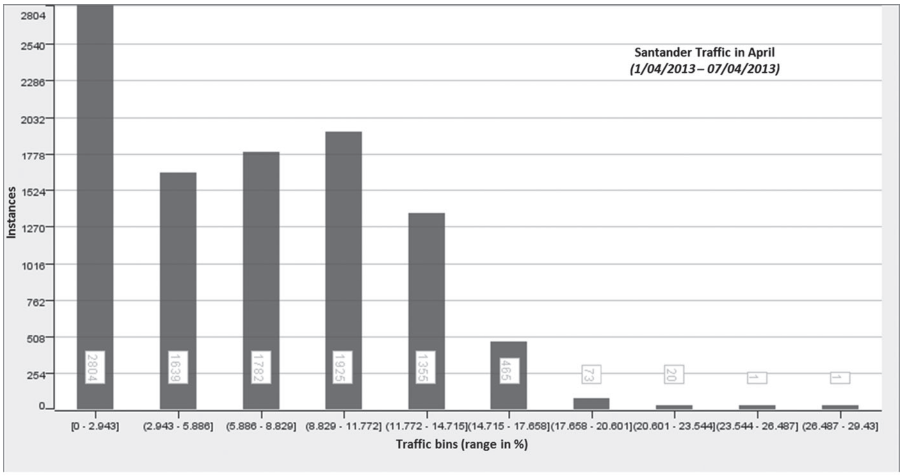



(c) SmartSantander Traffic aggregated by temperature bins from 02/12/2013 to 08/12/2013.

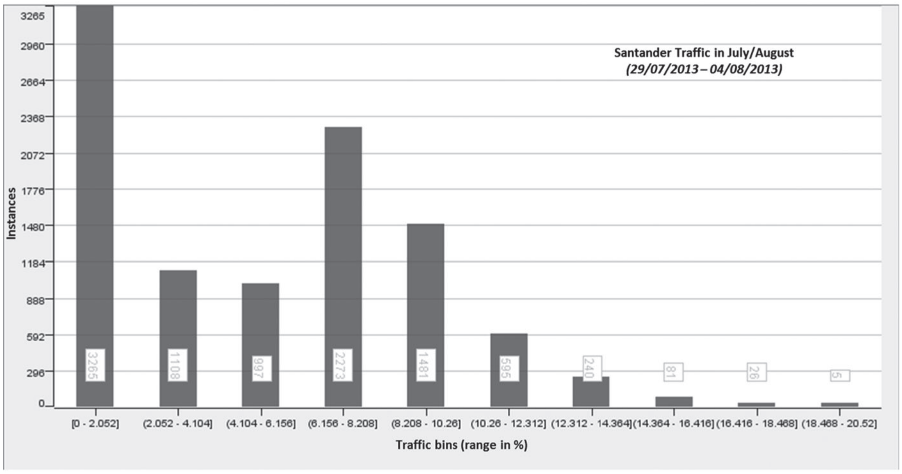
Figure 9. Traffic aggregated by temperature bins during 2013.



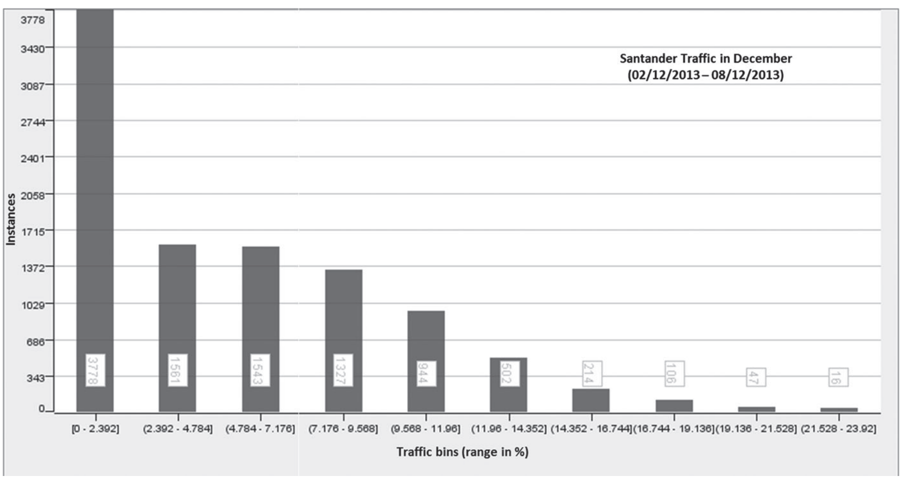
DOI: 10.1002/spe




(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.

Figure 10. Traffic density distribution in SmartSantander during 2013.



DOI: 10.1002/spe

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 midterm office hours.

## 5. 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 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 IoT, 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, or 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 [27], ants colonies, neurons organizations, the evolution of species, are proofs of the existence of a universal connection, and collective intelligence.

This work has presented the analysis of the traffic and temperature correlations where a Poisson distribution has been seen 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 majority of the time is presenting a low activity, because of 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.

## 6. CONCLUSIONS AND FUTURE WORKS

The entities in the world are not just able to act, 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, because it is allowing us to shift to preventive condition-based maintenance, which means fixing machines just before they break, without wasting time, that is, 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



DOI: 10.1002/spe

lack of geo-correlations among the different sensors, which is limiting the direct correlation of noise, temperature, number of parking pots with traffic and so on.

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 information from personal platforms such as smart phone. Thereby, it could be extrapolated to traffic forecast.

The ongoing work is focused on the data characterization, that is, 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.

## REFERENCE MODELS AND EXPERIENCES FOR KNIME

The models and data sets from our experiences can be downloaded from https://www.dropbox.com/sh/wi97e419xrk4kkh/y0zAnv7DWP


## ACKNOWLEDGEMENTS

The authors would like to thank to the HES-SO and the Institute of Information Systems funding and support. Special thanks to the SmartSantander EU project team, and in particular to *Luis Muñoz* from the University of Cantabria for this great opportunity to work with the SmartSantander data. Finally, the authors would like also to thank to The European Project 'Universal Integration of the IoT through an IPv6-based Service Oriented Architecture enabling heterogeneous components interoperability (IoT6)' from the FP7 with the grant agreement no 288445.


## REFERENCES


1. Hui J, Thubert P. Compression Format for IPv6 Datagrams over IEEE 802.15.4-Based Networks. *RFC6282* 2011.
2. Jara AJ, Zamora MA, Skarmeta A. Glowbal IP: an adaptive and transparent IPv6 integration in the internet of things. *Mobile Information Systems* 2012; **8**(3):177–197.
3. Yan Z, Zhang S, Zhou H, Zhang H, You I. Network mobility support in PMIPv6 network. *Proceedings of the 6th International Wireless Communications and Mobile Computing Conference*, Caen, France, ACM, 2010; 890–894.
4. Jara AJ, Ladid L, Skarmeta A. The internet of things through IPv6: An analysis of challenges, solutions and opportunities. *Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications (JoWUA)*; **3**(4):97–118, ISYOU.
5. Emmerson B. M2M: the internet of 50 billion devices. *WinWin Magazine* 2010:19–22.
6. Jara AJ, Varakliotis S, Skarmeta AF, Kirstein P. Extending the internet of things to the future internet through IPV6 support. *Mobile Information Systems* 2014; **10**(1):3–17.
7. Jara A, Olivieri A, Bocchi Y, Jung M, Kastner W, Skarmeta AF. Semantic Web of Things: An analysis of the application semantics for the IoT. *International Journal of Web and Grid Services (IJWGS)* 2014; **10**:1741–1106.
8. Gibson DV, Kozmetsky G, Smilor RW. *The Technopolis Phenomenon: Smart Cities, Fast Systems, Global Networks*. Rowman and Littlefield: Lanham, Maryland, USA, 1992.
9. Kumar P, Ranganath S, Huang W, Sengupta K. Framework for real-time behavior interpretation from traffic video. *IEEE Transactions on Intelligent Transportation Systems* 2005; **6**:43–53.
10. Valverde S, Sole RV. Networks and the city. *Architectural Design* 2013; **83**(4):112–119.
11. Zhang M, Yu T, Zhai GF. Smart transport system based on the internet of things. *Applied Mechanics and Materials* 2011; **48-49**:1073–1076.
12. Lin H, Zito R, Taylor M. A review of travel-time prediction in transport and logistics. *Proceedings of the Eastern Asia Society for Transportation Studies* 2005; **5**:1433–1448.
13. Berthold M, Hand DJ. *Intelligent Data Analysis: An Introduction* (2nd edn). Springer: Berlin, Germany, 2003.
14. Kramer D. Smart Cities will need big data. *Physics Today* 2013; **66**(9):19–20. DOI: 10.1063/PT.3.2110.

15. Berthold MR, Cebron N, Dill F, Gabriel TR, Ktter T, Meinl T, Wiswedel B. In *KNIME: The Konstanz Information Miner*. Springer: Berlin Heidelberg, 2008; 319–326. Available from: http://www.knime.org/ [last accessed 14 May 2014].
16. Hofstadter DR, Godel E. An eternal golden braid, 1979.
17. Barabasi AL. Bursts: the hidden patterns behind everything we do, from your e-mail to bloody crusades. *Penguin* 2010.
18. Vzquez A, Oliveira JG, Dezs Z, Goh KI, Kondor I, Barabasi AL. Modeling bursts and heavy tails in human dynamics. *Physical Review E* 2006; **73**(3):036127.
19. Gonzalez MC, Hidalgo CA, Barabasi AL. Understanding individual human mobility patterns. *Nature* 2008; **453**(7196):779–782.
20. Barzel B, Barabasi AL. Universality in network dynamics. *Nature Physics* 2013; **9**(10):673–681.
21. Davis PJ, Hersh R. *Descartes' Dream: The World According to Mathematics*. Macmillan Publishers Limited: London, UK, 2005.
22. Sanchez L. SmartSantander: The meeting point between Future Internet research and experimentation and the smart cities. *Future Network and Mobile Summit (FutureNetw)*, Warsaw, Poland, IEEE, 2011; 1–8.
23. Jara AJ, Zamora-Izquierdo MA, Skarmeta AF. A wearable system for Tele-monitoring and Tele-assistance of patients with integration of solutions from chronobiology for prediction of illness. *Ambient Intelligence and Smart Environments, Ambient Intelligence Perspectives* 2008; **1**:221–228. DOI: 10.3233/978-1-58603-946-2-221.
24. Barabasi AL. *Linked: How Everything is Connected to Everything else and What it Means*. Plume Editors: New York, USA, 2002.
25. Malmgren RD, Stouffer DB, Motter AE, Amaral LA. A Poissonian explanation for heavy tails in e-mail communication. *Proceedings of the National Academy of Sciences* 2008; **105**(47):18153–18158.
26. Haight FA. *Handbook of the Poisson Distribution*. Wiley: New York, USA, 1967; 036127–036146.
27. Watts DJ, Strogatz SH. Collective dynamics of small-world networks. *Nature* 1998; **393**(6684):440–442.



DOI: 10.1002/spe