UNIVERSITÀ DEGLI STUDI DI PADOVA

DIPARTIMENTO DI SCIENZE ECONOMICHE ED AZIENDALI "M. FANNO"

CORSO DI LAUREA MAGISTRALE IN

BUSINESS ADMINISTRATION

TESI DI LAUREA

"Big Data for smart cities and citizen engagement: evidence from Twitter data analysis on Italian municipalities."

RELATORE:
CH.MA PROF. SSA SILVIA RITA SEDITA

LAUREANDO: GOBBO EDOARDO

MATRICOLA N. 1178903

ANNO ACCADEMICO 2019/2020

To my parents, my sister,
my aunt Mariangela and my uncle Giancarlo,
who instilled in me the virtues of perseverance, commitment
and relentlessly encouraged me to strive for excellence.

Il candidato dichiara che il presente lavoro è originale e non è già stato sottoposto, in tutto o in parte, per il conseguimento di un titolo accademico in altre Università italiane o straniere.
Il candidato dichiara altresì che tutti i materiali utilizzati durante la preparazione dell'elaborato sono stati indicati nel testo e nella sezione "Riferimenti bibliografici" e che le eventuali citazioni testuali sono individuabili attraverso l'esplicito richiamo alla pubblicazione originale.

Firma

____

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Executive summary

Changing times are often a sign of opportunities. The notion of Big Data has recently been attracting an increasing degree of attention. Big Data refers to the large sets of information that grow at ever-increasing pace. It encompasses the volume of information, the *speed* at which it is produced and stored. Big Data often comes from multiple sources and in multiple formats, and its applications are almost unlimited. Moreover, the capillary diffusion of wireless technologies and the entire network infrastructure, allow the detection and collection of large amounts of data that can be used to understand patterns and innovative interpretative models.
New trends are thus spreading across firms and governments: Big Data analytics and Open Data platforms are realities that cannot be ignore anymore. Companies innovation strategies and governments decisions are already data- driven made, with both increasingly relying on outside data and collaborations to develop new patterns.

Big Data along with Information and Communication Technologies (ICT) aim to increase the wellness of cities and our way of living. Thus, the concept of *smart city* is introduced, which consists in exploiting the modern Information and Communication technologies in operating public affairs, making a better use of public resources, increasing the quality of services offered to citizens and in turn the quality of life, hopefully reducing the operational costs for public administrations. Nevertheless, for achieving such goals governments and municipalities have to be involved, not just from a financial point of view, rather letting citizens contributing to build that pattern giving them the instruments for achieving it.
Open Data represents a critical factor for public policy development and service delivery and can be really valuable for everyone: sharing data with communities of citizens and setting up collaborations with them might not just increase the usage of data and its application, but it might also increase the relationships between them. Understanding how citizens are effectively collaborating with governments and local governments represents a crucial variable in assessing whether cities are effectively moving toward *smart cities* patterns.
Open Data, E-government and citizen engagement represent key concept towards the achievement of the *smart city*.

One of the most known application of Big Data is Social Media which played an important role in modern society and had a critical success across different businesses. Over the past decade, governments have also adopted this tool as a new method of communication and engagement with their citizens: social media offer opportunities for dialogic communication where for relatively low cost a large amount of data reaching a wide audience can be published in real time (Bonsón, Royo, &Ratkai, 2015).

On the basis of what just presented above, the purpose of our dissertation is to contribute to the existing scarce literature on Big Data applied to *smart cities*, E- government and citizen engagement. We will start presenting the topics from a theoretical point of view and secondly, we will empirically examine, through a quali- quantitative analysis how citizens engage with their municipalities in Italy. This research for the purpose of our analysis focuses on Twitter, which is considered the largest micro-blogging site and one of the largest social media platforms used in business.

Specifically, the dissertation is organized as follows. In the next section, we will provide a detailed picture of the main theoretical aspects of Big Data, Internet of Things and smart cities. Afterwards, we will explore how governments are implementing Big Data and in particular social media for interacting with their citizens and then, the concept of citizens engagement will be introduced.

In the third and last section, we will describe our empirical analysis made on 28 Italian municipalities on Twitter. We will describe also our empirical conceptual model based on content and text mining analysis conducted with Python and RStudio. In conclusion, we will discuss our results and the implication of our research.

1. CHAPTER

Big Data for smart cities

Nowadays, the world is inundated with data generated every minute of every day, with the growth rate increasing approximately 10 times every five years (Del Vecchio, Di Minin Petruzzelli, Paniello, Pirri 2017). According to the Industrial Development Corporation (IDC) and EMC Corporation (IDC, 2014), the amount of data generated by 2020 will be 44 times greater than in 2009 and by 2020 there will be 5,200 Gigabytes of data for every person on earth, resulting in more than 40 ZB (Del Vecchio et al., 2017).

Big Data may come from a variety of sources, especially sources outside the usual boundaries of organizations, and it represents an interesting and emerging opportunity for sustaining and enhancing the effectiveness of the so-called Open Innovation paradigm (Del Vecchio et al., 2017). Sustaining Big Data through Open Innovation strategies and collaborations will help in better shaping our future, our cities and our technologies.

The development of new technologies and new ways of communication, alongside with an the increasing population density in urban centers, represents a unique opportunity for applying Big Data in developing smart cities opportunities: through Big Data city management and citizens are given access to a wealth of real-time information about the urban environment upon which to base decisions, actions and future planning in a collaborative atmosphere.

In the following chapter the paper will aim at giving an overview on Big Data technologies and analytics, on the Internet of Things; it will also introduce the concept of Smart City together with what it implies.

1.1 Defining Big Data

Social networks, downloaded and advertising applications, digital images and videos and many more helped to accelerate data generation during the past years.

Big Data represents a promising frontier in the future agenda of researchers and scholars in the field of business management and information systems.

Big Data can be seen as the result of an evolutionary process in the field of IT that started in the 1960s, by moving from the phases of data processing to information about new applications

(1970s-1980s) and then from the emergence of decision support models (1990s) to data warehousing and mining (2000s).
The possibility to have access and to store data and programs over the internet instead that from a personal computer's hard drive was crucial for the development of Big Data (Del Vecchio et al., 2017). This process is widely known as cloud computing, which is currently one of the best ways to store data and programs easily, cheaply, from any device and at any time.
The decrease in storage costs and the large diffusion of cloud solutions have positively influenced the adoption of Big Data technologies and approaches. The availability of cloud computing solutions for Big Data management is an opportunity for all companies, especially SMEs, often limited by scarce financial and organizational resources (Del Vecchio et al., 2017).
Big Data definitions have evolved rapidly throughout the years, hence there is not a general and commonly adopted definition for it.
Some of the main definitions of Big Data found in literature are summarized below:

- "Big Data is high-volume, high-velocity and high-variety information assets that demand cost-effective, innovative forms of information processing for enhanced insight and decision making." (Gartner It Glossary, n.d.) (Gandomi & Haider, 2015, p.2).
- "Big Data is a term that describes large volumes of high velocity, complex and variable data that require advanced techniques and technologies to enable the capture, storage, distribution, management and analysis of the information." (Tech America Foundation's Federal Big Data Commission,2012) (Gandomi & Haider, 2015, p.2).
- SAS: "Big Data is a popular term used to describe the exponential growth, availability, and use of information, both structured and unstructured" (Michalik, Stofa, Zolotova, 2014, p.331-334).
- IBM: "Data, coming from everywhere; sensors used to gather climate information, posts to social media sites, digital pictures and videos, purchase transaction record, and cell phone GPS signal to name a few" (Michalik, Stofa, Zolotova, 2014 p.331-334).
- "Big Data is defined as a large set of data that is very unstructured and disorganized" (Khan, Uddin, Gupta, 2014, p.1-5).
- "Big Data is a form of data that exceeds the processing capabilities of traditional database infrastructures or engines" (Khan, Uddin, Gupta, 2014, p.1-5).
- "Big Data is a new generation of technologies and architectures that is designed to extract value from large volumes of a wide variety of data by enabling high-velocity capture, discovery and analysis" (Villars, Olofson, Eastwood, 2011, p.2-6).

- “Regarded as the most representative synthesis of the complexity characterizing the current socio-economic scenario and its configuration as a knowledge-based economy, Big Data refers to any set of data that, with traditional systems, would require large capabilities in terms of storage space and time to be analyzed” (Kaisler, Armour, Espinosa, Money, 2013, p.995-1004).

A set of critical dimensions arose as suitable perspectives for the comprehension and management of the Big Data paradigm (Del Vecchio et al., 2017).
The main feature that comes to mind considering the question “what is Big Data?” refers to the size of firms (or simply players of the market); however, on a first stage, other characteristics of Big Data have recently emerged: commonly known as the 3Vs of Big Data, Volume, Variety and Velocity are the three dimensions of challenges in data management. The 3Vs have emerged as a common framework to describe Big Data (Gandomi & Haider, 2015).
The Three Vs are described below.

- Volume: volume refers to the magnitude of data. Definitions of Big Data volumes are relative and vary by factors, such as time and the type of data (Gandomi & Haider, 2015). The concept of volume might be commonly understood as the quantity of generated and stored data. The size of data matters, Big Data are usually reported in multiple terabytes and petabytes. Moreover, the term “Big”, which as written above refers to the magnitude of data, depends also on the type of data used: two datasets of the same size may require different data management technologies based on their type, hence definitions of Big Data also depend upon the industry (Gandomi & Haider, 2015).

- Variety: variety refers to the structural heterogeneity of the dataset (its nature and its type). Technological advances allow firms to use various types of structured (constituting only 5% of all existing data, referring to tabula datasets (Cukier,2010)), semi-structured and unstructured data (e.g. text, images, audio and video) (Gandomi & Haider, 2015).

  The emergence of new data management technologies and analytics, which enables organizations to leverage data in their business processes, is the innovative aspect (Gandomi & Haider, 2015). Using Big Data analytics, even small and medium- sized enterprises (SMEs) can mine massive volumes of semi-structured data to improve websites designs and implement effective cross-selling and personalized product recommendation systems (Gandomi & Haider, 2015).

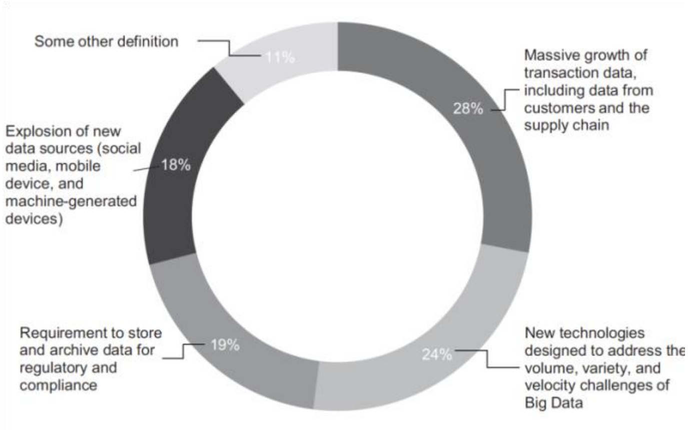
- Velocity*:* refers to the rate at which data are generated and the speed at which it should be analyzed (Gandomi & Haider, 2015). The proliferation of digital devices such as smartphones and sensors has led to an unprecedented rate of data creation and it is driving a growing need for real-time analytics and evidence-based planning (Gandomi & Haider, 2015). Data coming out from mobile devices and flowing through mobile apps produces torrents of information that can be used to generate real-time, personalized offers for every customer (Gandomi & Haider, 2015).

These data can provide crucial information about customers, such as geospatial location, demographics, and other patterns representing important data which can be gathered and analyzed by firms or entities for exploiting the customer's value. However, on the other hand, such an enormous quantity of data may result difficult to handle through traditional data management systems. Here is where data technologies come into play: enabling firms to create real-time intelligence from high volumes of "perishable" data (Gandomi & Haider, 2015).
Given the necessity to develop new data management systems, a second called generation of Vs has been identified, which represents the managerial side of Big Data: Veracity, Variability, Value.

- Veracity: commonly known as the fourth V, it represents unreliability in certain sources of data. For instance, customers' emotions in social media are uncertain data by nature, since they involve human judgment, however they contain crucial and precious information. Therefore, Big Data needs to deal with imprecise and uncertain data, which represents a huge challenge that can be faced adopting new tools and analytics management for uncertain data.
- Variability: variability refers to the variation in the data flow rates. Big Data are often generated through a myriad of sources. This imposes a critical challenge: the need to connect, match and transform data received from different sources (Gandomi & Haider, 2015).
- Value: Big Data are often characterized by relatively "low value density". That is, the data received in the original form usually has a low value in relation with its volume. However, a high value can be obtained by analyzing large volumes of such data.

The Big Data volumes' relativity can be applied to all dimensions. Thus, widespread benchmarks do not exist for volume, variety and velocity that clearly define Big Data. The defining limits depend on size, sector and location of the firm and these limits evolve over time (Gandomi & Haider, 2015). Moreover, it is crucial to outline that these dimensions are clearly not independent from one another: as one dimension changes, the likelihood that another dimension will also change as a result, increases (Gandomi & Haider, 2015).

Fig.1. 1: Definitions of Big Data based on an online survey of 154 global executives in April 2012.



Source: *A. Gandomi, M. Haider / International Journal of info. Management 35 (2015) 137-144..*

1.2 Privacy, policy and regulation- EU perspective

Big Data can be a really useful tool, however some big concerns can consequently be brought up. As information and data are consistently increasing and shared around the world, questions on how this data will be processed by companies are boosted. In recent years institutions and governments have tried to figure out issues on removing the privacy risks, "anonymizing" the data, deleting personal identifiers, names, account numbers. They thought that simply through this process of anonymization they would have been able to solve the problem and to protect users against any data breach. This however, remains a pale comparison to the current cyber security climate.

The amount of data that is currently flowing through the market is too broad for a singular entity to regulate and manage, companies can often re-identify and deanonymize the information hidden in an anonymized dataset. It is not an exaggeration to say that the population is just a collection of data to most of the institutions (Kuner, Cate, Millard, Svantesson, 2012).

Sellers are not even asking their customers whether they want to opt in or out, since they are able to track them anyway through their cell-phone signals, print-fingers recognition, social networks access or wi-fi.

Re-identification combines datasets that were meant to be kept apart (Ohm, 2010), and thanks to that companies can aggregate a huge amount of data and they are able to provide tailored products to their customers, based on their interests or desires. As Paul Ohm stressed on his work: “Data can be either useful or perfectly anonymous but never both” (Ohm, 2010).
This part of the paper will aim at providing an overview on how the European Union is facing the Big Data issue, focusing on the directives and regulations mainly applied in the last three years.

1.2.1 The General Data Protection Regulation (GDPR)

In 2016, the European Parliament and the European Council made the decision to apply a new regulation for data protection, which operates both for businesses and individuals.
The General Data Protection Regulation (GDPR) became effective on the 25th of May 2018. It took almost 2 years, since the EU gave the necessary time to companies for implementing the regulation.
Before 2016, data protection regulations were fragmented throughout the EU, with each country adding different clauses based on the main directive of 1995. Thus, on one hand some countries imposed consistent fees upon companies which breached or stole data, on the other hand, countries such as France did not inflict almost any sanctions at all. Over the years, thanks to the globalization process, the world has become “smaller” and businesses operate across different regions all over the world, therefore dealing with different legislations and interpretations of data protection. This is the reason why the EU felt the need to apply a new general data protection regulation in order to endorse a unique system which would be effective for the whole of Europe.
Apparently, there were two main issues regarding the data protection directive of 1995: the directive itself was not a mandatory tool and it belonged to an “ancient” technological era.
Indeed, back to 1995 just 1% of the world’s population was using Internet, but today it is almost ubiquitous across the EU; nowadays the vast majority of information is produced and consumed electronically, making it harder for the legislator to protect it (Tankard, 2016).
The new regulation is directly binding for all the companies which are operating in the EU and some which also extend their branches outside the European Union are affected by that. Indeed, given that the European market represents the biggest market in the world, tech companies or institutions which are operating worldwide cannot ignore the new regulation applied in the EU and other governments will be under pressure to raise their data protection standards in order to allow their economies to get access to the digital single market of the European Union

(Albrecht, 2016). Thus, the introduction of the GDPR might be useful not just for European countries, but even for those outside Europe.

The main purpose of the regulation is clearly that of harmonizing all the laws of the countries in the Union. The GDPR expands the scope of data protection so that anyone or any organization that collects and processes information related to EU citizens must comply with it, disregardful of where they are based or where the data is stored (Tankard, 2016)therefore even those companies which are established e.g. in the USA or China have to conform with that.

The regulation founds its framework on exacting that companies write down their privacy policies in a clear, straightforward language. Focusing on transparency, therefore clearly informing users about such transfers of data, businesses have to inform the user whether the decision is automated or not. Stronger rights have been given to individuals in case of harmful data breach. The user has the right to get access and also a copy of the data a business has about him/her. Furthermore, the European Union has the right to adopt binding decisions and to impose fines up to 20 million euro or 4% of a company’s worldwide revenues.

In addition to all the above mentioned measures, the EU decided to add a controversial right, named: “Right to be forgotten”. There has been a significant debate in Europe about the “right to be forgotten”, since it theoretically gives individuals the right to demand companies to delete personal data upon request. Conceptually speaking, the idea seems virtually impossible to achieve (Jim Remo, 2012). It would require the ability to track the data as it goes from place to place on the web (Jim Remo, 2012). When something has been uploaded or registered online it is rarely possible to delete it and it will “stay” on the internet forever. Massive institutions such as the European Union, or those companies who work for them, might hide sensitive information to the majority of users, but they will not be able to hide them from IT companies which collect huge amounts of data. These huge amounts of data are critical assets for those firms, because nowadays companies know how to monetize that asset, and they are able to gain as much as they can from it. Moreover, the “right to be forgotten” may also conflict with the fundamental right of freedom of speech (Jim Remo, 2012).

In the end, thanks to the GDPR there has been a huge change on the data protection approach, but as just seen it appears to be only the beginning of a long process that needs to be concluded for ensuring users about the complete protection of their data.

Due to new regulation applied, companies are facing huge changes for complying with it.

From the companies' side, better personal data protection measures will mean, first and foremost, more legal requirements and as a consequence, more time and effort expected, introducing for example new or more comprehensive processes (Freiherr, Zeiter, 2016).

The GDPR does not require any investments in new technologies to achieve the goals of the regulation; the GDPR rather states that organizations need to implement appropriate technological and operational safeguard measures for securing data, including putting strong privacy controls in place (Tankard, 2016).

Companies have to design and deal with privacy issues and data protection at the very beginning of the planning process. Nevertheless, the overall costs linked to collection, processing and storage of data will increase significantly, meaning that companies will be forced to delete far more data than necessary during the process. This might affect the value of relevant information about data pools which are worth billions of dollars.

In this new "era of consent", consumers seem to have much more power against firms, therefore many marketers fear the potential loss they might face in future periods. This will happen mainly to marketing firms which focused their core capabilities in collecting and processing data in order to help firms to tailor their products matching the customer's needs. The GDPR affects this process and might probably cause a decreasing in their revenues and performances. Relying more on macro data and less on more accurate data might affect the capabilities of marketing and tech companies, regressing back their strategy of years, offering less personalized products and seeing a decrease in fidelity by their consumers. Thus, the GDPR could clearly affect e-commerce market and even those free services like e-mails that send ads and cookies from which companies gain a lot of money. Hence, firms will be forced to find out new ways for extrapolating information from the users adopting new innovative ways for doing it, perhaps providing greater value to customers.

Another big issue that the GDPR might bring is the potential decrease in trust that might undermine the very process of Open Innovation which has been observed in the last years.

Companies might be scared to share data and information anymore, due to the new laws imposed by the new regulation. Being able to collect and store less data than in the past years, firms could regret to share crucial data with others, thereby decreasing their access to third-party data and this, consequently, may affect the whole development of Open Innovation.

The GDPR on one hand tries to give back to users their "privacy power" and let them trust the Internet again; on the other hand, as just written above, the new regulation might affect companies' results and incomes undermining trust and collaboration between firms.

1.3 Defining Big Data analytics

After analyzing Big Data, the dissertation will continue presenting how firms, entities, governments and communities, through different technologies, are using such a huge and varied amount of data and how all these players are starting being conscious that collaborations can help them to achieve greater results.

Big Data analytics represents the right tool to extract all the potential value that is hidden inside those data. This potential can be unlocked only if leveraged to drive decision making.

As highlighted by Vashisht and Gupta, Big Data has no value, unless used to gain insights for helping people in taking better strategic choices and decisions (Vashisht & Gupta, 2015). The method of processing big data in order to gain insights includes five stages where each stage comprises two-main sub-processes (Fig. 1.2):

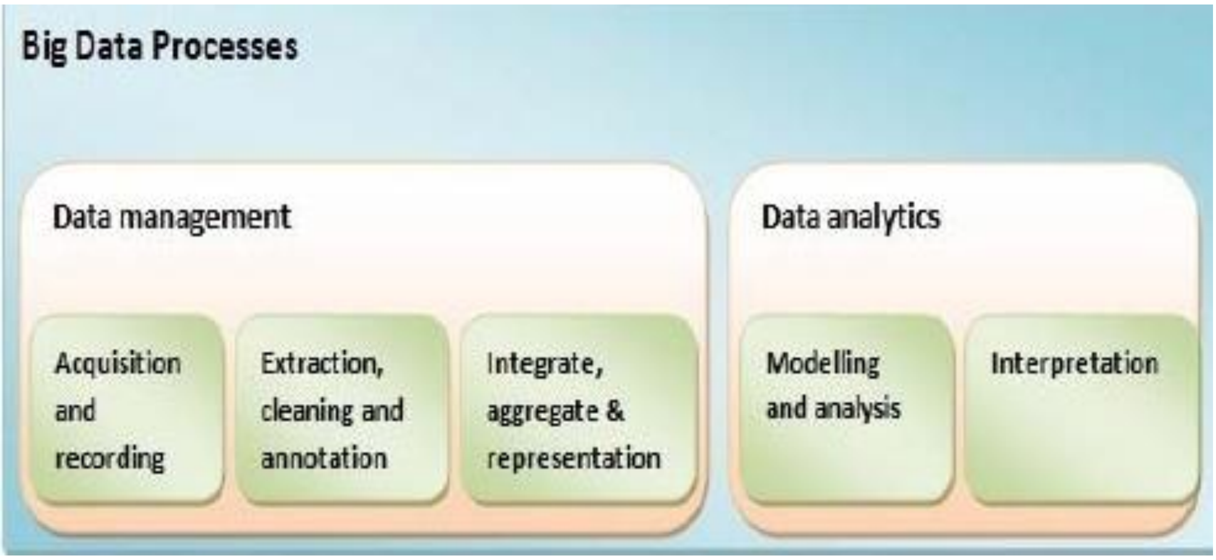
- Big Data management, which includes: acquiring, storing and preparing for the analysis.
- Big Data analytics, which includes: analyzing and acquiring intelligence from the data.

Data management involves processes and supporting technologies able to acquire, store and prepare data for being analyzed. On the other hand, analytics refer to techniques used to analyze and acquire intelligence from Big Data (Gandomi, Haider, 2015). Hence, Big Data analytics can be widely perceived as a sub-process in the overall process of “insight extraction” from Big Data (Gandomi, Haider, 2015).

The existing techniques used by traditional Business Intelligence products are modeled to perform well with highly structured data, however they are not sufficient for managing Big Data which compromises of structured, unstructured and semi-structured data (Vashisht, Gupta, 2015).

When dealing with Big data it is important to know that those data are aggregated from different sources, both internal and external. Moreover, verifying how trustworthy those external sources are is essential (Vashisht, Gupta, 2015).

Fig.1. 2: Processes for extracting insights from Big Data



Source: *Vashisht & Gupta, (2015), p. 265.*

1.3.1 Big Data analytics techniques and methods

In the following sections, the paper will review the main Big Data analytical techniques for analysing structured and unstructured data:

- *Text Analytics:* text analytics, or text mining, refers to the process of analysing and extracting information from textual data. Examples of textual data held by organizations are: social network feeds, emails, blogs, online forums, survey responses, corporate documents, news and call center logs. Analysis of text analytics involves statistical analysis, computational linguistics and machine learning. Text analytics enables businesses to convert large volumes of human generated text into meaningful summaries, which support evidence-based decision-making (Gandomi, Haider, 2015). Text analytics can be understood, for example, as a useful tool to predict stock market based on financial news. Examples of text analytics methods follow:
  Information extraction (IE): techniques extract structured data from unstructured text (Vashisht, Gupta, 2015), for instance, drug name and dosage from medical prescriptions. These techniques produce text summarization which follows two approaches: the extractive approach and the abstractive approach. Systems using abstract approach generate more coherent summaries, but extractive systems are easy to adopt for Big Data (Vashisht, Gupta, 2015).

- *Audio Analytics:* audio analytics or speech analytic techniques are used to analyze and extract information from unstructured audio data and are mainly adopted in customer call centers and healthcare institutions. Since these technologies have mostly been applied to spoken audio, the terms audio analytics and "speech analytics" are often used interchangeably (Gandomi, Haider, 2015). Call centers technologies help improve customer experience, evaluate agent performance and enhance sales turnover rates. These systems can be designed to analyze live calls in order to design the cross-selling recommendation system based on the past and present interactions of a customer and provide real-time feedbacks to clients (Vashisht, Gupta, 2015). Audio analytics uses two approaches: Transcript-based approach and Phonetic-based approach.

- *Video Analytics:* video analytics or VCA has a variety of techniques which monitor, analyze and extract meaningful information from video streams. The video analysis demand is growing quickly as the use of closed-circuit television (CCTV) cameras is increasing. Furthermore, also websites platforms adopted for sharing videos are getting popular in the video analytics sector. Although video analytics is still in its infancy compared to other types of data mining, various techniques have already been developed for processing real-time as well as pre-recorded videos (Gandomi & Haider, 2015). Big Data technologies are trying to turn this challenge into opportunity. Obviating the need for cost-intensive and risk-prone manual processing. Big Data technologies can be leveraged to automatically examine and draw intelligence from thousands of hours of video. As a result, the Big Data technology is the third factor that has contributed to the development of video analytics (Gandomi & Haider, 2015). Security and surveillance systems have been the first application of video analytics in recent years. The data generated by CCTV cameras in retail outlets can be extracted for business intelligence; indeed, the main areas in which those data can be useful are marketing and operations management. For example, smart algorithms can collect demographic information about customers, such as age, gender and ethnicity. On the other hand even retailers can count the number of customers, measuring the time spent inside the store and also detecting their movement patterns. Valuable insights can be obtained by correlating this information with customer demographics to drive decisions for product placement, price, assortment optimization, promotion design, cross-selling, layout optimization and staffing.

- *Social Media Analytics:* social media analytics refer to the analysis of structured and unstructured data, gathered from social media channels. Social media is a broad term encompassing a variety of online platforms that allow users to create and exchange content (Gandomi & Haider, 2015). Social Media can be divided into: Social Networks (f.i. Facebook, Twitter, LinkedIn, Instagram), blogs, social news, social bookmarking, media sharing, Q&A websites and review sites. The key characteristics of social media analytics is its data-centric nature (Gandomi & Haider, 2015). The research on the social media analytics spreads across psychology, sociology, anthropology, computer science, mathematics, physics and economics.

Marketing has been the main application of social media analytics so far. This can be explained by the fact that a large portion of consumers, worldwide, have adopted social media, mainly social networks. There are two sources of information in social media analytics which can be categorized into two groups: *Content-based analytics*, which focuses on data posted by users on social media platforms, such as customer feedback, product reviews, images and videos (Gandomi & Haider, 2015) and *Structure-based analytics* which mainly refers to social network analytics, synthesizing the relationships among participants.

Various techniques have been implemented to extract information from social networks (Fig. 1.3).

Community detection, also known as community discovery, extracts implicit communities within a network. Community detection helps to summarize huge networks, which then facilitates uncovering existing behavioral patterns and predicting emergent properties of the network (Gandomi & Haider, 2015). Community detection has found several application areas including, as written above, marketing and the World Wide Web (Gandomi & Haider, 2015).

Social influence analysis, refers to certain techniques that evaluate the influence of actors and connections in a social network: indeed it is widely known that the behavior of an actor in a social network is affected by others' behaviors (Gandomi & Haider, 2015). Social influence analysis techniques can be leveraged in viral marketing to efficiently enhance brand awareness and adoption (Gandomi & Haider, 2015).

Link prediction refers to the issue of predicting future linkages among nodes. Indeed, the structure of social networks is not static and continuously grows through the creation of new nodes and edges, therefore, the main goal is to understand and predict the dynamics of the network (Gandomi & Haider, 2015).

Fig.1. 3: Social Network Analytics techniques and Methods

Names of the techniques	Methods used by the techniques	Applications
a. Community detection (extracts the communities within a network)	CD using DBSCAN algorithm, CD with edge connection in social networks, CD using new-man Girvan algorithm, CD method using Bayesian network and expectation maximization technique, graph mining technique etc [18]	Marketing, world wide web for e.g. building product recommendation systems.
b. Social influence analysis(evaluates the influence of the actors and connections in the network)	Computational models for social influence analysis :learning influence probability, influence maximization model, influence action and dynamics, influence and interaction [19]	For viral marketing generally for brand awareness and adoption.
c. Link prediction(used to predict link and associations between the nodes)	Similarity-based algorithms(local and global similarity index), maximum likelihood methods(hierarchical structure model, stochastic block model) and probabilistic methods(probabilistic relational model, stochastic relational models) [20]	Link predictions in biological networks, for security by identifying criminal or terrorists networks

Source: *Vashisht & Gupta,(2015), p. 268.*

1.4 Big Data: an "Open" approach

Big Data refers to those data that, with traditional systems, would require large capabilities in terms of storage space and time to be analyzed (Del Vecchio et al., 2017). Big Data, as already mentioned above, comes from different sources and represents an interesting opportunity for sustaining and enhancing the effectiveness of Open Innovation strategies and Open source data platforms. Summarizing, the Open Innovation paradigm, according to Chesbrough (Chesbrough, 2003) is seen as "the use of purposive inflows and outflows of knowledge to accelerate internal innovation and expand markets for external use of innovation , respectively" (Chesbrough, 2003), while the Open data concept could be summed up as the idea that certain data should be available for public use without imposing copyright or patents restrictions.

As stressed by Del Vecchio et al., in this environment, Business Intelligence and analytics, and the related field of Big Data analytics, have seen their importance increase in academic and business communities as a way to analyze Big Data for better insights which can help organization in making decisions and effective strategies. For this reason, an "Open approach" is necessary to manage the huge amount of data generated by various sources.

According to the Open Innovation paradigm, companies, governments and other entities can share external as well as internal data, ideas, and knowledge to develop new products and services (Del Vecchio et al., 2017).

Open source data platforms are principally developed to help Big Data developers in designing Hadoop[1]-powered Big Data applications on a common platform. Hadoop is considered the most important platform for managing and exploiting large data sets across computer clusters (Del Vecchio et al., 2017). The Hadoop framework demonstrates how "open-collaborations" and "co-creation" represent key factors in looking at and establishing new business opportunities by watching on the community, also facing and solving common problems.

The process described above, and the possibility to bring other entities, stakeholders and organizations into a common virtual platform, represents a logical connection between the Open Innovation paradigm, Big Data analysis and Big Data analytics (Del Vecchio et al., 2017).

This strong connection with the process developed through the Hadoop platform, itself born "open", represents an ideal set of strategies to leverage and enrich an Open Innovation and knowledge-sharing process (Del Vecchio et al., 2017).

The organizations' value and competitiveness depend on the development, use and distribution of knowledge-based competencies, in order to find valuable collaborations, to exploit the right information and develop a strong knowledge network is becoming crucial.

Moreover, as stressed by Tom Rosamila (senior Vice President at IBM Systems and Technology Group) (Del Vecchio et al., 2017): "Big Data accelerates the opportunity for new discovery while at the same time magnifying the challenge scientists face … the current approach to computing presumes a model of data repeatedly moving back and forth from storage to processor in order to analyze and access data insights, a process that is unsustainable with the onslaught of Big Data because of the amount of time and energy that massive and frequent data movement entails."

Rosamila continues highlighting the issue, outlining that the Big Data challenge can only be solved through Open Innovation strategies. Some large organizations are collaborating in co-creating more value by leveraging communities of web users to develop new technologies, products and services, hence reaching economies of scale and lowering their costs in serving customers (e.g., Wikipedia, which is one of the most known open source platforms on the web). Thus, implementing Open Innovation strategies through the creation of communities of customers, citizens and society might be a great source of value (e.g., markets of ideas, crowdsourcing, etc.) allowing firms and governments to gather new ideas and insights from a larger and more committed pool of users (Del Vecchio et al., 2017).

[1] Hadoop: the Apache Hadoop software library is a framework that allows for the distributed processing of large data sets across clusters of computers using simple programming models. It facilitates the users in solving problems that involve massive amounts of data and computation. It provides a software framework for distributed storage and processing of Big Data using the Map Reduce programming model.

This trending attention to the use of Big Data for Open Innovation strategies is pushing governments, firms and communities of citizens to collaborate in order to develop a better environment to live in. Given that "co-creation" is a two-way process, especially firms are called to provide feedback to stimulate continuing participation and commitment. Hence, this implies several additional activities that have to be managed by all the players (particularly governments and firms), such as coordinating the community of users, communicating with most of the users in the community or rewarding the deserving users (Del Vecchio et al., 2017). Of course, there are costs in terms of time, money and effort that have to be taken into consideration, but the outcomes that might be derived from these kinds of opportunity are huge. In conclusion, using Open Innovation strategies, through the development of open source platforms, for handling and enhancing Big Data represents a great opportunity for governments and firms for the creation of an ecosystem.

Several large corporations have already taken advantage of "open co-creation" of value -based on collaborations with governments, entities, citizens and customers- which generate large amounts of data that can be used as a source of value (Del Vecchio et al., 2017).

This approach can be supported by providing citizens or customers (whether the subjects are the firms or the governments) with feedback to stimulate continuing participation, discussion and commitment. Indeed, significant trends in developing "customer or citizen based applications" or "decision based applications" implemented for companies or governments also exist (Del Vecchio et al., 2017).

1.5 The Internet of Things

One of the main opportunities in Big Data for creating an ecosystem where firms, governments and citizens can create collaborations comes from the use of technologies or devices characterizing the Internet of Things (IoT).

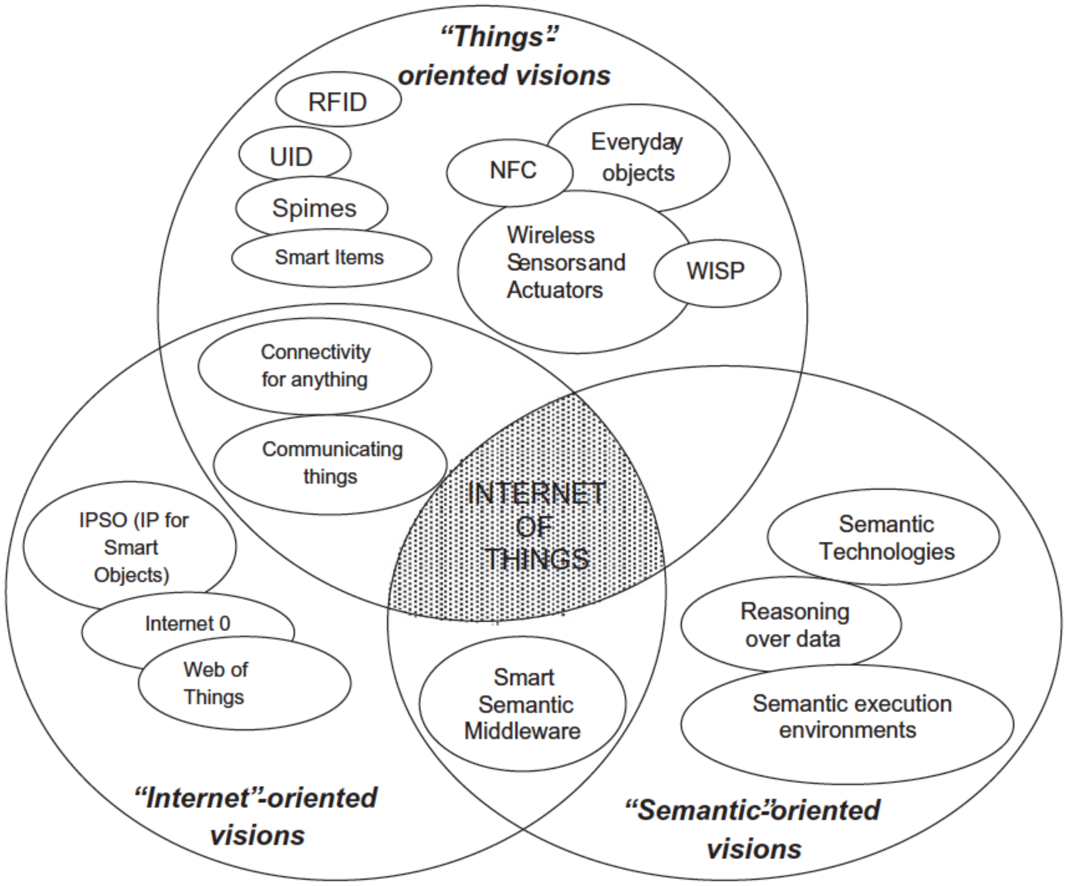
The IoT is a new paradigm of information networks with the aim of expanding the potential of the Web (Atzori, Iera & Morabito, 2010).

There are various approaches among researchers, stakeholders, firms and governments in giving a real definition of IoT, depending on their specific interests, aims and backgrounds; however, two schools of thought that have emerged for describing the Internet of Things: "Internet oriented" or "Things oriented".

The rationale behind the IoT is in its denomination: "Internet" and "Things". The first term reflects a network- oriented vision of communication, which entails the use of hardware, standards and protocols characterizing the Web 2.0, while the second tends to move the focus to physical objects, rather than end users (Atzori, et al., 2010). It shall not be forgotten that the words "Internet" and "Things" assume a meaning which introduces a disruptive level of innovation into today world (Atzori, et al., 2010). Put together, the IoT can be semantically summed up as a "world-wide network of interconnected objects uniquely addressable, based on standard communication protocols" (Del Vecchio et al., 2017).

The unique object addresses the representation and the storing of exchanged information, which is one of the most challenging issues nowadays, thus introducing the third IoT perspective: "Semantic oriented" (Atzori, et al., 2010).

Fig.1. 4: "Internet of Things" paradigm as a result of the convergence of different visions.

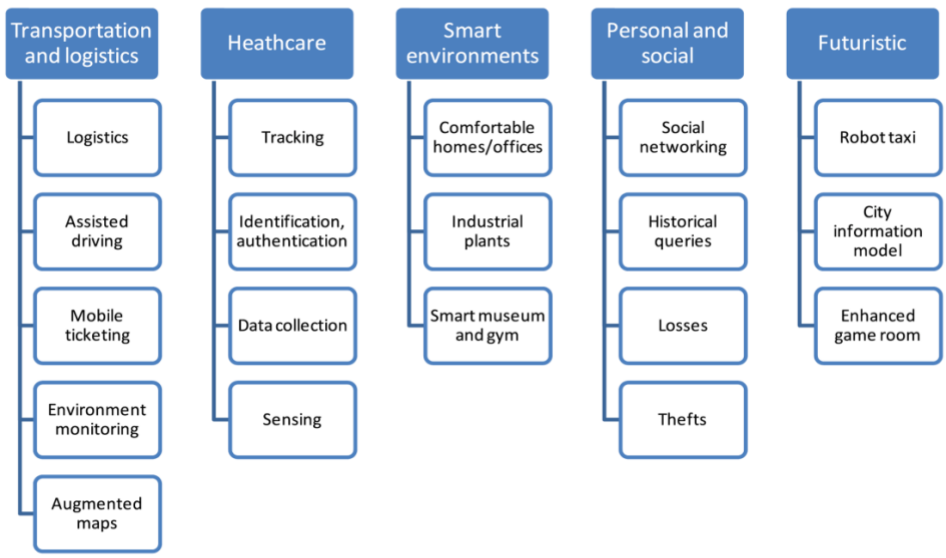


Source: L. Atzori et al./ Computer Networks 54 (2010) 2787-2805.

In Fig. 1.4 the main concepts, technologies and standards have been divided and classified based on the IoT visions that have contributed to best characterize the whole paradigm. It clearly appears that the IoT paradigm is the result of the merging of the three visions described above. IoT provides and will provide instant and remote access to information about physical objects, thereby leading to innovative networked systems with higher efficiency and productivity; therefore, multiple devices are used for web-connected applications, hence also increasing the amount of data collected and open data source platforms will gain benefits from that (Del Vecchio et al., 2017).

According to Atzori et al, the IoT process can offer huge potentialities for developing a huge number of new applications; there are indeed many domains and environments where new applications would likely improve the quality of everyone's life.
Nowadays, there are objects or technologies that are able to elaborate the information perceived from the surrounding environments, thus coordination and communication among them are important aspects for exploiting new applications.
The following domains represent the main areas for developing these applications (Atzori, et al., 2010):
-Transportation and logistics
-Healthcare
-Smart environment (home, office, plant)
-Personal and social
In the following subsections the paper will provide a general overview of the short and medium term applications for each of these categories and an arrangement of futuristic applications; however, a clear and precise description of technologies will not be provided by this paper, as it is beyond the objectives of this research.

Fig.1. 5: Application domains and relevant major scenarios



Source: L. Atzori et al./ Computer Networks 54 (2010) 2787-2805.

1.5.1 Main Internet of Things applications

Advanced cars, trains, buses as well as bicycles along with roads and/or rails are becoming more “structured” thanks to the introduction of sensors, actuators and processing power (Atzori, et al., 2010); roads and transported goods too are equipped with sensors that are sending important information in order to track the traffic flows and vehicles in order to better route the traffic, provide citizens with appropriate transportation information and monitor the status of transported goods. Moreover, vehicles equipped with those sensors and technologies might help drivers and passengers in increasing safety and navigation. Governments would also benefit from more accurate information about road traffic patterns for planning purposes (Atzori, et al., 2010). Hence, crucial information about the vehicles transporting goods and information regarding the type and status of goods could be integrated to provide new data about delivery time or delivery delays, for instance.

Use of IoT has also been applied to the payments sector. Payments made through mobile devices are already widely used and large corporations such as Apple, PayPal and other entities in the payments and financial sectors have adopted and developed contactless payments applications or other systems and objects able to process cashless payments (f.i. ApplePay or PayPal apps).

A very common and useful IoT application refers to the augmented maps. Tourist maps can be equipped with tags that allow phones to search for a place on the web and automatically call web services providing information about hotels, restaurants, monuments and events related to the area of interest for the user (Atzori, et al., 2010). There is a collection of Physical Mobile Interaction (PMI) techniques that can be employed to augment the information of the map (Atzori, et al., 2010).

One of the main sectors where Big Data and Internet of Things technologies have been adopted is the Healthcare domain. Applications in this sector can be grouped into: tracking of objects and people (staff and patients), identification and authentication of people, automatic data collection and sensing (Atzori, et al., 2010).

Big Data, through IoT technologies, has the potential to create significant value in healthcare by improving outcomes while lowering costs, but really sensitive data about patients are involved (Roski, Bo-Linn, Andrews, July 2014). Nowadays, an estimate has been made that through Big Data the healthcare sector may potentially generate $300 billion annually (Roski, Bo-Linn, Andrews, July 2014).

Big Data has already helped the health care sector through the development of new clinical technologies, such as for instance: using mobile devices to tailor diagnostic and treatment decisions, delivery of personalized medicines based on patient's risk profile and introduction of automated analysis of x-rays, computed tomography (CT), scan images and magnetic resonance imaging (MRI) (Roski, Bo-Linn, Andrews, July 2014). Even though some investments for implementing these new technologies have already been made, such as using digitized patients' records, other infrastructures are required for applying Big Data. Moreover, to be successful with Big Data organizations need to develop processes and policies that accommodate new protocols, time requirements, risk factors, and mandates for managing data, mainly in the area of privacy and security (Roski, Bo-Linn, Andrews, July 2014).

Indeed, the main concerns regarding the adoption of IoT in the Healthcare sector are linked to the privacy side of the issue. Current policies and laws need to be revised and checked by policy makers and healthcare institutions. For achieving the Big Data potential, large amounts of data need to be stored, involving very sensible data about individuals. Moreover, the data involved may not be owned or controlled by a single organization (Roski, Bo-Linn, Andrews, July 2014).

Other applications of IoT may refer to the smart environments domain which aims to create an "employment environment" made easy and comfortable for everyone thanks to the intelligence of contained objects such as: comfortable offices and homes, new industrial plants and smart museums and gyms.

The applications described above are realistic and most of them could already be implemented over a short/medium period since the required technologies are already available. However, the most interesting applications of the IoT is represented by the "Futuristic applications" which rely on some technologies that are either in the development process or whose implementation is still too complex (e.g., robot taxi, city information model, enhanced game room) (Atzori, et al., 2010).

These technologies are even more interesting regarding the required research and impact.

1.5.2 Internet of Things for smart cities

The IoT concept, as described above, aims at making the Internet even more pervasive than it is already, enabling easy access and interaction with a wide variety of devices. The Internet of Things shall be able to transparently and seamlessly incorporate a large number of end systems, while providing open access to selected sets of data for the development of digital services (Zanella, Bui, Castellani, Vangelista, Zorzi, 2014).

The IoT will help the development of applications that use the enormous quantity of data generated by such a great number of objects to provide new services to citizens, companies and public administrations. This paradigm indeed finds applications in many different domains as cited above in the paper; however, such a heterogeneous field of application makes the identification of solutions capable of satisfying the requirements of all possible application scenarios a great opportunity (Zanella, et al., 2014).

The application of IoT paradigm to an urban context is particularly interesting, as it responds to a strong push by many national governments to adopt solutions in the management of public affairs, hence aiming to realize the *Smart City* concept (Zanella, et al., 2014).

There is not yet a formal and worldwide accepted definition of Smart City, however the main goal of the Smart City project is to create a better use of public resources, increasing the quality of the services offered to citizens, while reducing the operational costs of public entities and governments. An urban IoT might bring great benefits to the management and optimization of traditional public services such as transport and parking, lighting, surveillance and maintenance of common public areas (Zanella, et al., 2014).

1.6 Smart city: a new vision leveraged by Big Data

As introduced in the previous subsection, governments are considering to adopt the Smart City project in their cities and implementing Big Data applications that support Smart City components to reach the required level of sustainability and improve the living standards (Al Nuaimi, Al Neyadi, Mohamed & Al-Jaroodi, 2015).

Many governments are developing strategies for becoming Smart Cities, based on many possible innovative usages of network infrastructures and Internet-based applications, through the engagement of Open Innovation ecosystems. A crucial part of these strategies is based on developing new types of innovation approaches in urban areas, which are generally characterized by a high level of citizen engagement in co-creating internet-based applications, towards elaborating also new forms of collaboration among local governments, research institutes, universities and companies (Komninos, Pallot & Schaffers, 2012).

This might involve a reduction in costs and in resources consumption as well as actively engage citizens in making decisions.

One of the most powerful technologies used for enhancing smart cities services is Big Data analytics. As digitization has become an integral part of everyday life, data collection has resulted in the accumulation of huge amounts of data that can be used in various beneficial application domains (Al Nuaimi, et al., 2015).

Current work and research projects on this field have dramatically risen during the last years, thus outlining the centrality of the issue and the crucial importance it has. Most of the relative literature highlights the importance of Big Data in supporting smart city applications and services and in exploiting the general requirements for the design and implementation of Big Data.

Analyzing literature, there is no evidence of an official or universally agreed definition of smart city. However, the majority of definitions highlights common features and aspects that may describe smart cities, for instance: enhancement of life quality, infrastructures, transports, logistics, healthcare, public safety, etc.

In *Table 1.1,* different definitions applied to the smart city concept that have been found in literature are depicted. From these definitions it appears clear that the smart city is seen as an integrated living solution that links many life aspects such as power electricity, transports and buildings in a smart and efficient manner to improve the quality of citizens' lives (Al Nuaimi, et al., 2015).

Tab. 1. 1: Definitions of Smart City in the literature.

Authors	Definitions of Smart City concept	Area of focus
Aguilera, Galan, Campos & Rodriguez	"Smart city is a very broad concept, which includes not only physical infrastructure but also human and social factors"	Includes the social aspects and agrees that smart city has a broad focus.
Neirotti, De Marco, Cagliano, Mangano & Scorrano	"The concept of Smart City (SC) as a means to enhance the life quality of citizen has been gaining increasing importance in the agendas of policy makers. However, a shared definition of SC is not available and it is hard to identify common global trends"	Policy makers are an additional aspect of the smart city definition. Consents to the lack of a shared definition of smart cities.
Su, Li & Fu	"Smart city, the important strategy of IBM, mainly focuses on applying the next-generation information technology to all walks of life, embedding sensors and equipment to hospitals, power grids, railways, bridges, tunnels, roads, buildings, water systems, dams, oil and gas pipelines and other objects in every corner of the world, and forming the "Internet of Things" via the Internet"	Addresses the technological aspect of smart cities and focuses on how next-generation information technology is the key.

Chourabi, Nam, Walker, Gil-Garcia, Mellouli, Nahon & Scholl	"A city well performing in a forward-looking way in economy, people, governance, mobility, environment, and living, built on the smart combination of endowments and activities of self-decisive, independent, and aware citizens"	Views a smart city as a futuristic model of collaborative components.
Kitchin	"A city that monitors and integrates conditions of all of its critical infrastructures, including roads, bridges, tunnels, rails, subways, airports, seaports, communications, water, power, even major buildings, can better optimize its resources, plan its preventive maintenance activities, and monitor security aspects while maximizing services to its citizens"	Focuses on the integration of infrastructures and systems that monitor and control the resources to achieve sustainability as the main aspect of a smart city.
Chourabi, Nam, Walker, Gil-Garcia, Mellouli, Nahon & Scholl	"Connecting the physical infrastructure, the IT infrastructure, the social infrastructure, and the business infrastructure to leverage the collective intelligence of the city"	A more generic view that puts together all main aspects of a smart city to achieve the goal. Seems to be the most comprehensive definition of a smart city.
Chourabi, Nam, Walker, Gil-Garcia, Mellouli, Nahon & Scholl	"A city striving to make itself "smarter" (more efficient, sustainable, equitable, and livable)"	General definition, does not specify how a city will get smarter.
Khan, Anjum & Kiani	A smart city is "… a city which invests in ICT enhanced governance and participatory processes to define appropriate public service and transportation investments that can ensure sustainable socio-economic development, enhanced quality-of-life, and intelligent management of natural resources"	Views the smart city as specific, and narrow, set of resources/services working together to achieve a better life.

Source: *Al Nuaimi, Al Neyadi, Mohamed & Al-Jaroodi, (2015).*

Given the concepts presented above regarding the smart city, the whole process could be thought as a large organic system connecting many subsystems and components like the ones described above. Indeed, several researchers and experts consider a smart city as the organic integration of various systems, usually heterogeneous.

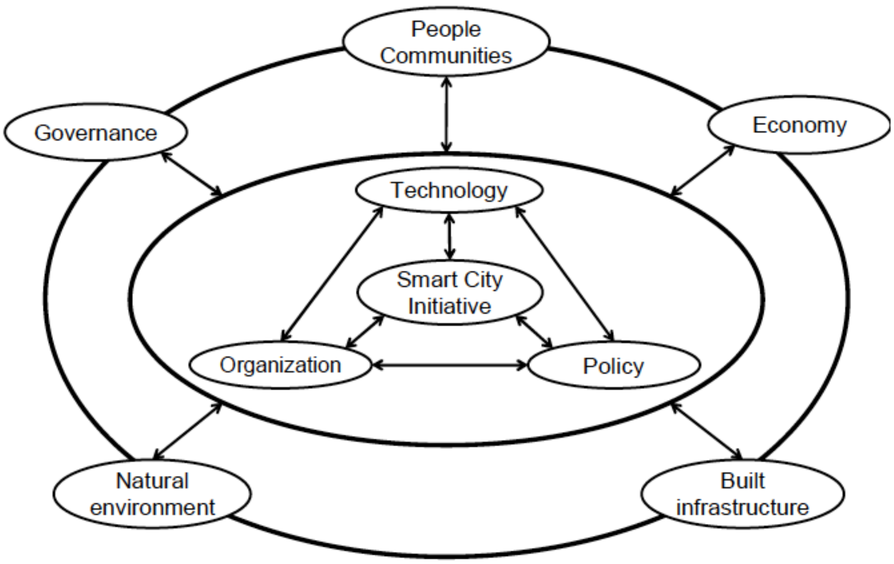
The interrelationship between a smart city's core systems is taken into account to make "the system of systems smarter": no system operates in isolation (Chourabi, Garcia, Pardo, Nam, Mellouli, Scholl, Walker & Nahon, 2012).

While systems in industrial cities were mostly skeleton and skin, postindustrial smart cities are like organisms that develop an artificial nervous system, which enables them to behave in intelligently coordinated ways (Chourabi, et al., 2012). The new forms of cities' intelligence

reside in the increasingly effective combination of digital telecommunication networks, embedded intelligence, sensors, tags and software (Chourabi, et al., 2012).
Moreover, with urbanization breaking records over records and with an expected 70% of population that will live in cities and surrounding regions by 2050, it is important to understand the demand for services profiles in order to increase the efficiency of city management (Jin, Gubbi, Marusic & Palaniswami, 2014). Currently, only few municipalities have platforms or systems for live monitoring and inferring of urban process parameters. Data collection exercises are often costly and difficult to replicate, however there is an increased demand on municipalities for incorporating smart technologies that collect the required data and analyze them (Jin, et al., 2014). Through advanced sensing and computational capabilities, data are gathered and real-time evaluated to extract the information, which is further converted to usable knowledge: the whole process will help enhancing the decision making of city management and citizens to turn the city into a smart city (Jin, et al., 2014).

Chourabi et al. have developed an integrate framework (Fig. 1.6) to explain the relationships and influences between some factors that, in their opinion, interact together in the same environment and are crucial for developing intelligent cities, in order to pursue smart city initiatives.

Fig.1. 6: Smart City initiatives framework.



Source: *Chourabi, Garcia, Pardo, Nam, Mellouli, Scholl, Walker & Nahon, (2012), p.2294.*

These factors provide a basis to compare how cities envision their smart initiatives, implement shared services and related experiments (Chourabi, et al., 2012). The above displayed graph is

also presented as a very interesting tool to support understanding of the relative success of different contexts and for different purposes.
According to the research made by Chourabi et al., in order to reflect the different levels of impact, the factors proposed are represented through two different levels of influence: “outer factors” and “inner factors”. Outer factors (governance, people and communities, natural environment, infrastructures and economy) are more influenced than the inner factors (technology, management and policy) (Chourabi, et al., 2012).

1.6.1 Benefits and opportunities of Big Data applications in smart cities

In this section, the paper will aim at giving a look at the main applications adopted by some cities for evolving towards smart cities; it will also aim at examining which the main benefits and opportunities are, mostly regarding Big Data analytics applications, that may help in taking the decision to convert and redesign a city.
Some of the main benefits that may be gained in developing a smart city are (Al Nuaimi, et al., 2015): efficient resources utilization, better quality of life, higher levels of transparency and openness.
Achieving these benefits requires high levels of sophistication and involvement in terms of applications, resources and citizens’ engagement (Al Nuaimi, et al., 2015). The opportunities for reaching these benefits and goals are concrete, however they require huge investments in technology and an effective use of Big Data by the governments which have the responsibility and the need of ensuring data accuracy and security and using data documentation standards to provide guidance on the content and use of datasets (Al Nuaimi, et al., 2015).
Big Data applications have the potential to serve many sectors in a smart city. Deploying Big Data applications requires the support of a good information and communication technology (ICT)[2] infrastructure (Al Nuaimi, et al., 2015). ICT supports smart cities because it provides useful solutions and also unique solutions that may not be possible without it. Adopting ICT, Cloud Open source platforms and Big Data technologies will help to address many issues such as providing storage and analysis tools, reaching innovation stages and encouraging collaborations and communication between the different entities of a smart city (Al Nuaimi, et al., 2015).

[2] ICT: Information and Communications Technology is an extensional term for information technology (IT) that stresses the role of unified communications and the integration of telecommunications and computers, as well as necessary enterprise software, middleware and storage that enable users to access, store, transmit and manipulate information.

This process can be achieved by building Big Data communities to work as one entity to foster collaborative and creative solutions in addressing applications for areas like education, health, energy, law, transportation, environment, safety and governance.

Some of the main benefits discovered through the adoption of different Big Data domains used in smart cities are summarized below:

-Smart Healthcare (Al Nuaimi, et al., 2015):

- Allow healthcare providers and practitioners to gather, analyze, and utilize patient information, which can also be used by insurance companies and by government agencies.
- Increase the amount of real-time data gathered from certain patients through smart devices, which are connected to various monitors for an accurate and timely response to health issues and for a comprehensive patient history record.

-Smart Energy (Al Nuaimi, et al., 2015):

- Facilitate decision-making related to the supply levels of electricity in line with the actual demand of the citizens.
- Allow forecasting in a near real-time manner through efficient analysis of the data collected.

-Smart Transportation and Logistics (Al Nuaimi, et al., 2015):

- Recognize traffic patterns by investigating real time data.
- Reduce main city roads congestion by predicting traffic conditions and adjusting traffic controls. Through Big Data, smart cities will be able to reduce traffic and accidents by opening new roads, enhancing infrastructures on the basis of congestion data and collecting information about car parking patterns and usage.
- Reduce supply chain waste by associating deliveries and optimizing shipping movements.
- Enable data streaming software to process and communicate traffic information collected through sensors, smart traffic lights and on-vehicle devices directly to drivers via smartphones or through other communication devices.

-Smart Environment (Al Nuaimi, et al., 2015):

- Provide more accurate weather information that will improve the country’s agriculture.
- Adopt sensors and other devices which can help farmers providing them with smart solutions in order to reduce crop failures, introducing advanced farming systems such as mobile irrigation management systems or farm management systems. Moreover,

combining new information with historical data, farmers can improve the decisions making process thus reducing any waste.

-Smart Safety (Al Nuaimi, et al., 2015):

- Provide detailed, spatial and temporal geographic area maps.
- Help predict future environmental changes or natural disasters like earthquakes or tsunami waves, thus savings lives, cities and the surrounding environment.

-Smart Governance (Al Nuaimi, et al., 2015):

- Support the integration and collaboration of different government agencies and combine or streamline their processes. This will result in more efficient operations, better handling of shared data and stronger regulation management and enforcement.
- Establish new policies for the benefits of data owners (citizens) and producers (government agencies).
- Help governments focus on citizens’ concerns related to health and social care, housing, education, policing, and other issues.

1.6.2 Challenges and requirements

Many are the challenges that cities face in implementing Big Data applications, due to the dynamic and evolving environments in which they operate. The main challenges rely on the availability of Big Data tools, real-time analytics tasks, accuracy, representation, cost and accessibility of data. Such issues may affect the results and performances of smart city applications and services which rely on Big Data (Al Nuaimi, et al., 2015).

Sharing data and information among different city departments represents a huge challenge: each government agency or department usually has its own “data warehouse” with public or confidential information, and they are reluctant to share what might be considered reserved data. As written above (section 1.2), the challenge is to make sure not to cross the line between collecting and using Big Data and ensuring citizens their rights to privacy (Al Nuaimi, et al., 2015). Smart cities need to find out new and alternative ways to reduce or even to prevent any kind of barriers in sharing information and data among different entities; however, there is the need to assure privacy and security for the citizens involved.

Nevertheless, one of the major challenges is represented by the costs in setting up these kind of projects: huge investments by public and private entities are needed, however, due to the difficulties that might occur in implementing such technologies, results may not be easily achieved and cities can be negatively affected.

Furthermore, even population size has a great effect on Big Data. Indeed, as the population grows, the size of generated data increases becoming massive and difficult to manage: a rapid growth would generate traffic congestion, pollution and social inequality, alongside with increased urbanization which raises a variety of technical, social, economic and organizational problems that tend to jeopardize the economic and environmental sustainability (Al Nuaimi, et al., 2015). Hence, smart city applications need to evolve quickly and to extend efficiently to handle the growing volume and variety of Big Data involved. They will also need to try to avoid any negative issue (Al Nuaimi, et al., 2015).

When considering smart city applications based on Big Data, it is crucial to address several requirements that stem from the special nature of smart city needs and Big Data features (Al Nuaimi, et al., 2015). In the following lines, several requirements based on the type of Big Data applications for smart cities are discussed: some of them are technological while others are related to citizens' awareness and governments' roles.

- Big Data Management: in order to ensure a useful usage of the huge quantity of data accumulated in smart city applications, it is necessary to set up a suitable and effective data management.

  Big Data management includes the development, architectures, policies, practices and procedures that properly manage the full data lifecycle. As the data come from different sources with different formats, there is a need for advanced data management features that will lead to recognize different formats and sources of data, structuring, managing, classifying and controlling all these structures (Al Nuaimi, et al., 2015).
- Smart network infrastructure: most Big Data applications for smart cities require smart networks connecting their components including residents' equipment such as cars, smart house devices and smart phones. These networks should be capable of transferring collected data from their sources to where Big Data is collected and processed and then to transfer responses back to different entities in the smart city (Al Nuaimi, et al., 2015).
- Open Standard Technology: applying Big Data applications in smart cities involves working with large systems and data software, therefore it is convenient to follow open standard models in order to design and implement such solutions. This process will lead to an increase in flexibility for upgrading, maintaining and adding new application features for smart cities and facilitating the integration of Big Data components (Al Nuaimi, et al., 2015).
- Citizen awareness: citizens must be aware of how to correctly and safely use ICT solutions for smart city. Their active participation in providing information related to

various issues would help in enhancing the quality of collected data and the performance of these applications (Al Nuaimi, et al., 2015).

- Government Role: governing entities of smart cities must establish guiding principles of openness, transparency, participation and collaboration to keep Big Data exchange and flow under control (Bertot & Choi, 2013). Moreover, governments must review and recalibrate information and data policies as necessary by focusing on privacy, data reuse, data accuracy, data access, archiving and preservation, hence having well-defined data documentation and codebooks to ensure informed use of datasets (Bertot & Choi, 2013).

  In order to support the adoption of Big Data applications, governments should balance the valuable uses of data against individuals' privacy concerns by addressing some of the fundamental concepts of privacy laws (Al Nuaimi, et al., 2015).

2. Chapter 2

E-Government for citizens' engagement

E-government, through Internet of Things technologies, represents a unique opportunity for helping citizens to improve societies' engagement with governments. Governments are accountable to citizens and society at large, since they are responsible for safeguarding the interests of the general public (Bonsón, Royo & Ratkai, 2015).

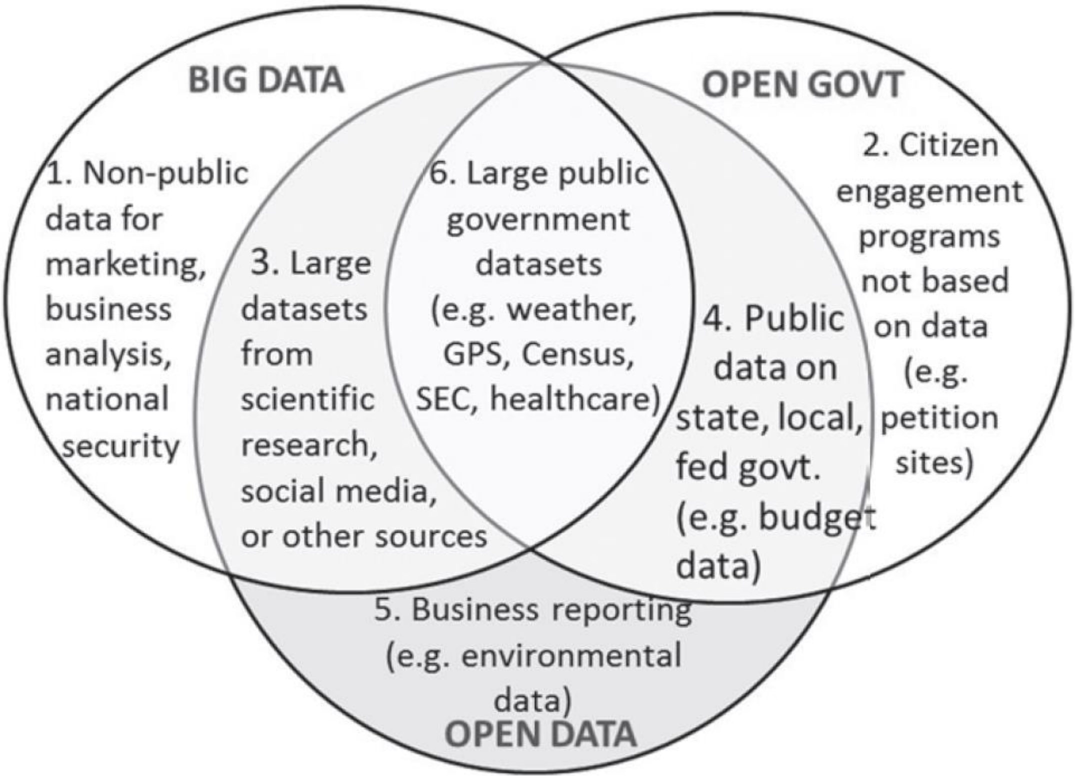
The concept of Information Technologies (or e-government) in government has changed throughout the years. The interactions between government and citizens (G2C), government and businesses (G2B), and inter- agency relationships (G2G) are nowadays more effective, democratic and transparent (Mellouli, Luna-Reyes & Zhang, 2014).

Internet and social media represent a relatively low cost source of value, being able to produce a huge amount of information and data, as pointed out in the first chapter. Thus, the use of Information and Communication Technologies (ICTs) in local governments is ranked as one of the preferred ways of communications by certain stakeholder groups (politicians, administration and civil society) after e-mails. Johannessen, Flak and Sæbø (2012) stressed that effective municipal communication and e-participation require the use of different media and that 70% of respondents prefer some kind of digital communication.

The idea of smart community refers to the adoption of information and communication technologies by local governments and cities in order to better interact with their citizens, deriving advantages from all available data to solve critical issues. However, there are two main components to take into consideration: the extensive use of technology by governments (e-government), and the extensive use of technology by citizens to interact with governments, which refers to citizens' engagement (Mellouli, Luna-Reyes & Zhang, 2014). Citizens' engagement is a new form of democracy in which citizens are part of the decision- making process with regard to the development of their society: it represents an interactive process among citizens themselves, and between citizens and governments in order to contribute in making and taking decisions in the most transparent and responsible manner (Mellouli, Luna-Reyes & Zhang, 2014).

Citizens' engagement introduces new ways for governments to formulate decisions by integrating citizens' point of view: citizens need to be empowered through useful, relevant and complete sets of information from governments, developing Open data platforms that can be interpreted and re-used by populations (Fig. 2.1).

Fig.2. 1: Open data framework.



Source: *Guri, p.73,2014.*

2.1 Leveraging Big Data technologies to approach an open government path

As highlighted in the previous chapter, governments are pushing their cities into the future, approaching new paths to create new ecosystems.

As it happens in businesses while implementing structural changes, communication results as one of the most critical aspects also for governments in empowering citizens and in convincing them of the possible outcomes that can derive in implementing such strategies.

Thus, given the rise of the Web 2.0[3] in the past years and driven by rising citizen expectations and the need of government innovation particularly after the global crisis in 2008, social media (SM) have become a central factor for e-government.

The main benefits that the Web 2.0 offers to public sector entities are the enhancement of transparency and citizen participation. The Web 2.0 has favored the emergence of citizen-

[3] Web 2.0: refers to websites that emphasize user-generated content, participatory culture and interoperability for end-users.

created content that enriches socio-political debates and that increases the diversity of opinions, the free flow of information and freedom of expression (Bonsón, Torres, Royo, & Flores, 2012). Over the past decade, governments have indeed adopted SM as a new method of communication and engagement with their citizens (Guillamón, Ríos, Gesuele & Metalo, 2016).

These new applications provide an opportunity for dialogic communication where, at a relatively low cost, a large amount of information can be published real-time, arguing that social media may not only enhance the transparency of public administration but may also improve policy making, the provision of public services and facilitate knowledge management in local governments (Bonsón, Perea, & Bednárová, 2019).

Open data leads to two important assumptions about government: firstly, it represents an assumption of the readiness of public agencies referring to the opening process which considers influences and exchanges as constructive and welcomes opposing views. Secondly, Open data leads to the assumption that government is going to give up control, at least to some extent, demanding considerable transformations of the public sector (Janssen, Charalabidis & Zuiderwijk, 2012). Instead of reinforcing current processes, Open data should result in open government, in which the government acts as an open system interacting with its environment. Moreover, not only should data be published, but they should easily be obtained by everyone in order to being able to actively contribute in improving the government processes (Janssen, Charalabidis & Zuiderwijk, 2012).

This outlines that governments and institutions should enable and embrace Open data.

The outcomes stemming from the enactment of technology are difficult to predict because of multiple and unanticipated effects influenced by rational, social and political logics (Orlikowski, 2000).

In sharing Open data to the public, managers and politicians might reach crucial advantages in understandings trends and needs of both customers and citizens, nevertheless incurring in lack of control: indeed, outside the boundaries of government, command and control mechanisms cannot be used.

No one has an overview of what is done with the Open data, and even having such an overview might mine the basic idea behind Open data itself (Janssen, Charalabidis & Zuiderwijk, 2012).

As stressed by Janssen et al., whether the opening of data will unambiguously lead to a more transparent, interactive, open, and, hence, accountable government it represents anyway a massive challenge.

Adopting new technologies in order to support, enhance and expand democratic practices does not represent a recent trend: the history of the last century media has demonstrated that the introduction of new communication technologies routinely gives rise to intense speculation about their impact on the processes and practices of democracy (Harrison, Guerrero, Burke, Cook, Cresswell, Helbig, Hrdinovà, Pardo, 2012). In the case of computer-mediated communication and information technologies, as emphasized by Harrison et al. (2012), speculation has been particularly intense, and has been applied to broad processes of democratic decision making and e-democracy as well as to more targeted forms of government action as e-government.

Therefore, the common idea of “Open government” is animated by optimism over what can be accomplished politically through the adoption of new forms of technology, mainly referring to the use and development of open source data that can be accessible to the citizens through online clouds (Harrison et al., 2012).

The whole open source movement aims at being characterized by three main features such as transparency, participation and collaboration that can be reached through Big Data technologies, but these terms also represent political values with a substantial history in democratic theory, directly relevant to broad processes of citizen action related to public policy choices (Harrison et al., 2012).

Transparency does not represent an empty purpose for people: citizens may desire their government to be transparent in order to being able to scrutinize information and actions and thus assessing how concrete outcomes are going. Similarly, participation for the mere sake of participating is an empty and alienating exercise (Harrison et al., 2012).

Citizens are willing to participate in order to help governments in developing, gathering data and producing government actions that respond to their needs.

Collaboration is meaningful when all the participants can contribute and can make practical and substantial decisions.

2.1.1 Open data benefits in open governments

Janssen, Charalabidis and Zuiderwijk (2012) recognized the potential benefits of open data, through their research, by making interviews to managers, politicians and citizens,.
The basic assumption is that open data itself creates and generates more value than the selling of data sets (Janssen, Charalabidis & Zuiderwijk, 2012).
The benefits, as summed up in the table below, coming from the introduction of open data clouds can be clustered in:

- Political and Social benefits
- Economic benefits
- Operational and Technical benefits

Opening data is expected to create benefits such as stimulating innovation and promoting economic growth (Janssen, Charalabidis & Zuiderwijk, 2012). However, there is no way to predict and calculate the return of investment (ROI) for the issues. Thus, one of the main challenges is that Open data has no value in itself, though it starts producing real value only when used.
One of the main benefits explored by Janssen, Charalabidis and Zuiderwijk (2012) about opening data is the ability to tap into the collective intelligence of the public. The basic idea is that under the right circumstances, groups can generate better alternatives and make better decisions. In the table below (Tab. 2.1) the main benefits that derive from the adoption of Open data are summed.

Tab. 2. 1: Open data benefits.

Category	Benefits
Political and Social	-More transparency -Democratic accountability -More participation and self-empowerrment of citizens (users) -Creation of trust in government -Public engagement -Scrutiny of data -Equal access to data -New governmental services for citizens -Improvement of citizen services -Improvement of citizen satisfaction -Improvement of policy-making processes -More visibility for the data provider -Stimulation of knowledge developments -Creation of new insights in the public sector -New (innovative) social services
Economic	-Economic growth and stimulation of competitiveness -Stimulation of innovation -Contribution toward the improvement of processes, products and/or services -Development of new products and services -Use of the wisdom of the crowds: tapping into the intelligence of the collective

	-Creation of a new sector adding value to the economy -Availability of information for investors and companies
Operational and Technical	-The ability to reuse data/not having to collect the same data again and counteracting unnecessary duplication and associated costs (also by other public institutions) -Optimization of administrative processes -Improvement of public policies -Access to external problem-solving capacity -Fair decision-making by enabling comparison -Easier access to data and discovery of data -Creation of new data based on combining data -External quality checks of data (validation) -Sustainability of data (no data loss) -The ability to merge, integrate, and mesh public and private data

Source: *Janssen, Charalabidis & Zuiderwijk, 2012, p.261.*

2.1.2 Open data barriers in open governments

Throughout their studies, Janssen, Charalabidis and Zuiderwijk (2012) also outline which the main impediments and barriers for implementing Open data clouds at public level are (Tab. 2.2). The barriers were clustered and divided into different groups, as can be noticed from the table below, depending on the different levels:

- Institutional
- Task Complexity
- Use and Participation
- Legislation
- Information Quality
- Technical

Usually, these barriers are strictly linked to either data providers or data users.

For instance, the institutional level concerns barriers from the data providers' point of view, whereas task complexity and use and participation are seen as barriers from the user's perspective (Janssen, Charalabidis & Zuiderwijk, 2012).

The remaining categories might be considered significant for both.

Tab. 2. 2: Open data barriers.

Category	Barriers
Institutional	-Emphasis of barriers and neglect of opportunities -Unclear trade-off between public values (transparency vs. privacy values) -Risk-adverse culture (no entrepreneurship) -No uniform policy for publicizing data -Making public only non-value-adding data -No resources with which to publicize data (especially small agencies) -Revenue system is based on creating income from data -Fostering local organizations interests at the expense of citizen interests -No process for dealing with user input -Debatable quality of user input
Task Complexity	-Lack of ability to discover the appropriate data -No access to the original data (only processed data) -No explanation of the meaning of data -No information about the quality of the Open data -Apps hiding complexity -Duplication of data, data available in various forms, or before/after processing resulting in discussions about what the source is -Difficulty in searching and browsing due to no index or other means to ensure easy search for finding the right data -Even if data can be found, users might not be aware of its potential uses -Data formats and datasets are too complex to handle and use easily -No tooling support or helpdesk -Focus is on making use of single datasets, whereas the real value might come from combining various datasets -Contradicting outcomes based on the use of the same data -Invalid conclusions

Use and Participation	-No incentives for the users -Public organizations do not react to user input -Frustration at the existence of too many data initiatives -No time to delve into the details, or no time at all -Having to pay a fee for the data -Registration required before being able to download the data -Unexpected escalated costs -No time to make use of the open data -Lack of knowledge to make use of or to make sense of data -Lack of the necessary capability to use the information -No statistical knowledge or understanding of the potential and limitations of statistics -Threat of lawsuits or other violations
Legislation	-Privacy violation -Security -No license for using data/Limited conditions for using data -Dispute and litigations -Prior written permission required to gain access to and reproduce data -Reuse of contracts/agreements
Information Quality	-Lack of information -Lack of accuracy of the information -Incomplete information, only part of the total picture shown or only a certain range -Obsolete and non-valid data -Unclear value: information may appear to be irrelevant or benign when viewed in isolation, but when linked and analyzed collectively it can result in new insights -Too much information to process and not sure what to look at -Similar data stored in different systems yields different results
Technical	-Data must be in a well-defined format that is easily accessible: while the format of data is arbitrary, the format of data definitions needs to be rigorously defined -Absence of standards -No central portal or architecture -No support for making data available -Lack of meta standards -No standard software for processing open data -Fragmentation of software and applications -Legacy systems that complicate the publicizing of data

***Source**: Janssen, Charalabidis & Zuiderwijk, 2012, p.262-263.*

2.2 Citizen coproduction and citizen engagement in the age of social media

New real and tangible trends have emerged in recent years that sustain Open government concepts and they have gone far beyond any experimentation and theory.

The Obama Open Government Initiative aimed at increasing the transparency, participation and collaboration of citizens in the government decision-making processes, developing and setting up an Open Government Data (OGD) to facilitate those relationships. Even though Obama's directive on open data is widely considered the most famous example of "openness" to an active citizen collaboration, it does not represent a unique example. The British government's Big Society program, which aimed to do "more with less" by decentralizing and devolving power, or Singapore's "Government- with You" e-government strategy, that tried to "facilitate a collaborative government that co-creates and connects with the people", (Linders, 2012) provide further evidence.

These ones outline the re-emergence of citizen coproduction, in which citizens perform the role of partner rather than customer in the delivery of public services.

Linders (2012) defines the following framework typology for citizen participation:

- Citizen Sourcing (Citizens to government or C2G) in which the government has primary responsibility, but citizens do have the power to influence the outcomes, improving the government's situational awareness.
- Government as a Platform (Government to Citizen or G2C) in which government makes its knowledge and IT infrastructure available to the public that paid for their development.
- "Do It Yourself Government" (Citizen to Citizen or C2C) in which citizens are able to effectively self-organize themselves in approaching new opportunities for citizen-to-citizen coproduction, thus potentially substituting traditional government responsibilities.

In the following paragraphs the paper will discuss each one of these categories, explaining also the dramatic impact that social media and open collaboration platforms are having on them.

2.2.1 Citizen Sourcing

In recent years governments have increased their efforts in collecting input from the public through e-participation and e-rulemaking, therefore being able to collect citizen preferences with far greater levels of sophistication than periodic-binary votes initiatives (Linders, 2012).
As widely known, politicians, institutions and other establishments already rely on Facebook, Instagram and Twitter not to only interact with citizens during political campaigns but also to consult the public while in office (Linders, 2012). This institutional adoption of government-to-citizen online interaction indeed opens up to powerful new problem-solving mechanisms: they embrace citizens to use their skills and expertise to solve government challenges, importing innovation from social entrepreneurs and from outside the government.
According to Linders (2012), governments allow for closer, deeper and more frequent collaboration between citizens through crowdsourcing, co-delivery and online platforms.

For instance, even the NASA launched a “micro tasking platform” to coordinate the volunteer participants’ activities in examining huge amounts of satellite imagery of map craters on Mars at lower costs and increasing the speed of the analysis (Linders, 2012). This outlines that opening up large open data clouds to the public in order to achieve mutual goals or challenges can be a great source of value from which huge benefits can be derived. Big Data represents a powerful new channel for improving the government’s situational awareness by enabling citizens to efficiently and conveniently share knowledge with government (Linders, 2012). Such channels are more easily exploited rather at a local level through direct citizen reporting systems. For instance, indeed, apps such as SeeClickFix provide mobile and online reporting platforms: they give citizens the ability to report situations such as potholes or even crimes using their smart phones and to take pictures as evidence.

2.2.2 Government as a Platform

This new wave of digital information and public databases enables governments to deliver highly personalized information to support citizens’ personal decisions. For example, government data mining could notify users about relevant health risks or about useful government programs for which they qualify (Linders, 2012). Such data mining and dissemination help citizens make more informed, socially-responsible decisions while reducing the need for government intervention (Linders, 2012). The high Big Data potential might enable governments to become a stronger part of the social ecosystem. Indeed, governments do not

usually take part in the business of co-developing personal or private goods, but they can create an environment more conducive to private, peer and personal production by “embedding” their capabilities into the wider ecosystem (Linders, 2012).
Moreover, giving that digital information and computerized services can often be shared at near zero marginal costs, it follows that institutions can take the chance to give those resources to the public for their own use almost without additional costs (Linders, 2012).
As outlined by Linders (2012), the internet also opens up to a range of opportunities for civil servants entrepreneurship, actively contributing to an online community of practice thanks to the government ‘s vast stores of knowledge, talent and infrastructure that could be used by the public.
These advancements in data management and data analysis have equipped individual citizens and civil society organizations with the capability to sift through vast amounts of government data: a new level of open and transparent government has been enabled (Linders, 2012).

2.2.3 “Do It Yourself Government”

Social media and online collaboration platforms might represent great opportunities to let citizens easily and effectively self-organize themselves. As stressed by Flanagin, et al. (2001) in literature, the Information Age, as called by the author, has reduced citizens’ reliance on formal, professionalized institutions of collective action (political parties, mass media, etc.); nevertheless on the other hand the increasing power that these online communities have accomplished has brought to a reduction in coordination costs, which are now considered less necessary (Flanagin & Metzger, 2001). Internet platforms, for instance, make it far easier for groups to make collective decisions by collapsing time, space and hierarchy (Linders, 2012).
According to Linders (2012) these new platforms offer new channels for communicating and coordinating community-based collective action with Big Data and IT technologies, in that they replace government as the intermediary and facilitate direct citizen-to-citizen assistance. Social networks such as Facebook or mainly Twitter represents perfect examples of this trend: citizens effectively rely on these new forms of communications for having a direct knowledge on the latest political, economic and social news which regards their local public administrations.
According to the “Do It Yourself Government” category and to the mass coordination through social media, citizens are provided with a vehicle/ to self-perform functions of government that institutions are no longer able or have refused to provide.
Called “self-service for citizens” according to literature, this trend has a great appeal in countries which are suffering from shortcomings in government service provision: citizens can

take matters into their own hands by using social media to coordinate their actions when governments are not able to act or do not want to act (a phenomenon which is widely spread in China) (Linders, 2012).

2.3 From e-government to we-government: the impact of social media on citizen engagement

As illustrated in the previous paragraphs, embracing an e-government framework does not only represent a useful tool for internal modernization, reporting information and the development of public services, but it rather represents a new successful communication channel provided to citizens in order for them to actively participate in democratic institutions and political processes which can create value for both individuals and institutions.

According to Bonsón, et al. (2015) one of the greatest concerns in public sectors is represented by the agency problem. Following his research, Bonsón outlined that stakeholders cannot ensure that the government is acting in the best interest for the citizens, nevertheless by adopting a sophisticated social media plan this conflict could be reduced by embracing this process of disclosing information. The implementation of social media by governments, mainly local municipalities, might represent a great promotion of transparency engaging citizens and building trust: social media and open platform technologies allow people to be integrated in administrative and policy processes as shareholders (Bonsón, et al., 2015).

As outlined by Ellison and Hardey (Ellison & Hardey, 2014), in an era where levels of citizens engagement are low, municipalities need to actively adopt social media in order to involve residents properly in all the aspects of local governance.

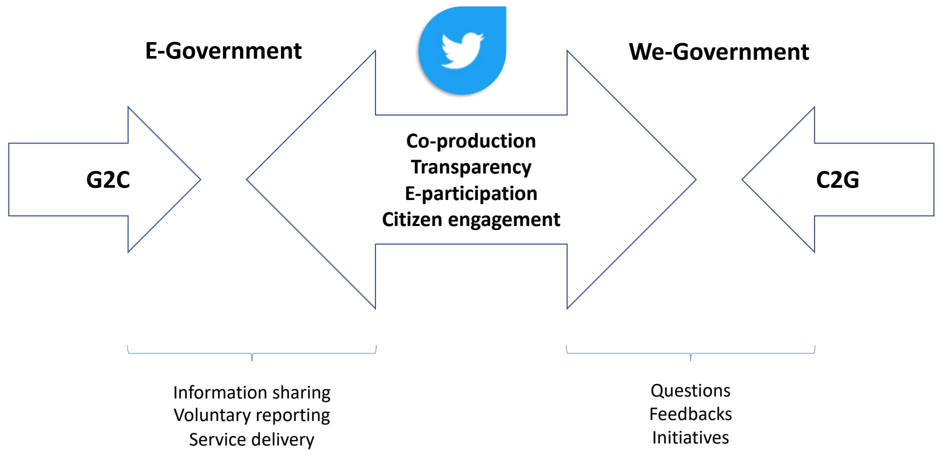
According to Linders (2012), this type of governance is called "we-government" where there is the need both to transform governments by co-operating and to increase citizen e-participation.

Surely, social media users might not always represent the general public opinion, but it opens towards great opportunities to improve G2C interactions, in order to inform about policy development or about other public crucial news.

Therefore, having a Twitter account can be considered as a sign of transparency and accountability (Bonsón, et al., 2015). By now, it is widely recognized that social networks are more immediate and flexible than other forms of communication.

The figure below (Fig. 2.2) shows the role of Twitter on G2C and C2G communication. As stressed once again by Linders (2012), online communication channels make data mining possible for governments, helping both citizens and governments to be more informed and able to make socially responsible decisions as partners.

Fig.2. 2: Role of Twitter on both G2C and C2G communication.



Source: Own re-elaboration of *Bonsón, et al.,2015, p.54.*

2.4 Practical evidences of citizen engagement: FixMyStreet, SeeClickFix and Comuni- Chiamo

As outlined in literature by Linders (2012), one of the main G2C issues is represented by a lack of trust in the society, whereas trust should be gained through transparency and effective activities. Big Data technologies through apps, social media and other forms of open collaboration might reduce the gap between governments and citizens. Social media and open collaboration apps offer public institutions the opportunity to integrate information and opinions in the policy-making process or in facing challenges and problems.

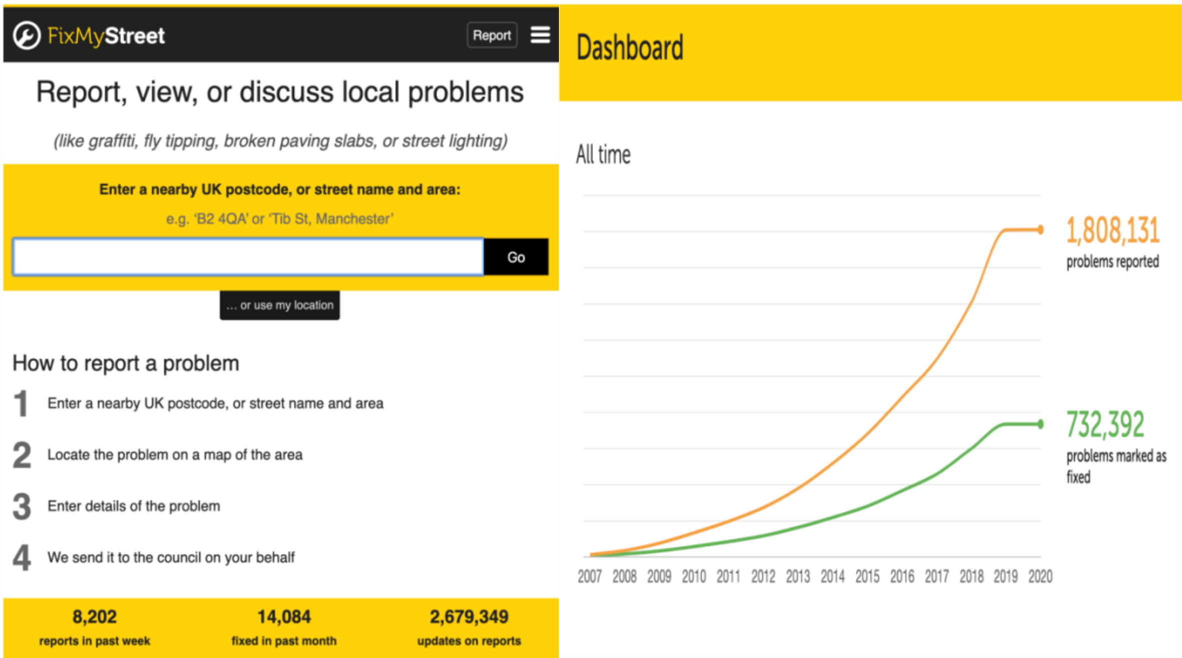
During the past years, numerous applications and technologies have been developed by both citizens and institutions in order to find new solutions for increasing engagement, communication and active participation of people.

Three examples on how citizen engagement and active citizen participation might solve problems and create benefits for the society in general are presented below.

FixMyStreet is an app developed by mySociety, an e-democracy project by the UK-based registered charity named UK Citizens Online Democracy, which helps people in the United Kingdom inform their local authority about issues which need their attention, such as potholes, broken paving slabs, graffiti etc.

All the reports are published anonymously on the website, where everyone can see them. On one hand most local councils have their own reporting systems, on the other hand FixmyStreet solves this kind of obstacles between citizens and local entities. Indeed, by using the mySociety's open source cloud called MapIt software, FixMyStreet matches users' postcodes, categorizing the problems and forwarding them to the correct local authority.
By simply entering a nearby UK postcode or even a street name and area, then locating the problem on a map and entering the details of the issues, FixMyStreet will directly send the reports to the council on the reporter behalf.
FixMystreet has had a critical success so far, being awarded numerous prizes and having a great impact in the UK.
As may be noticed from the picture below, FixMyStreet currently has great numbers. It is the website itself that provides information on the status of the reports.
Having a look at the graph, it may be noticed that since the boost of Big Data technologies, social media and smart phones, starting from 2012, the slopes of the orange line have drastically risen up outlining its success. It also means that people do believe that participating in the governments' daily routine and processes is the right choice for having a better engagement.

Fig.2. 3: FixMyStreet website homepage and dashboard.



Source: https://www.fixmystreet.com/#main-nav.

One of the most successful websites and mobile apps that improved citizens' participation in the US is SeeClickFix (Fig. 2.4).
SeeClickFix is a global digital communications system company co-founded by Ben Berkowitz and based in Downtown Haven, Connecticut.

The website and app help users communicate with local governments about non-emergency neighborhood issues.

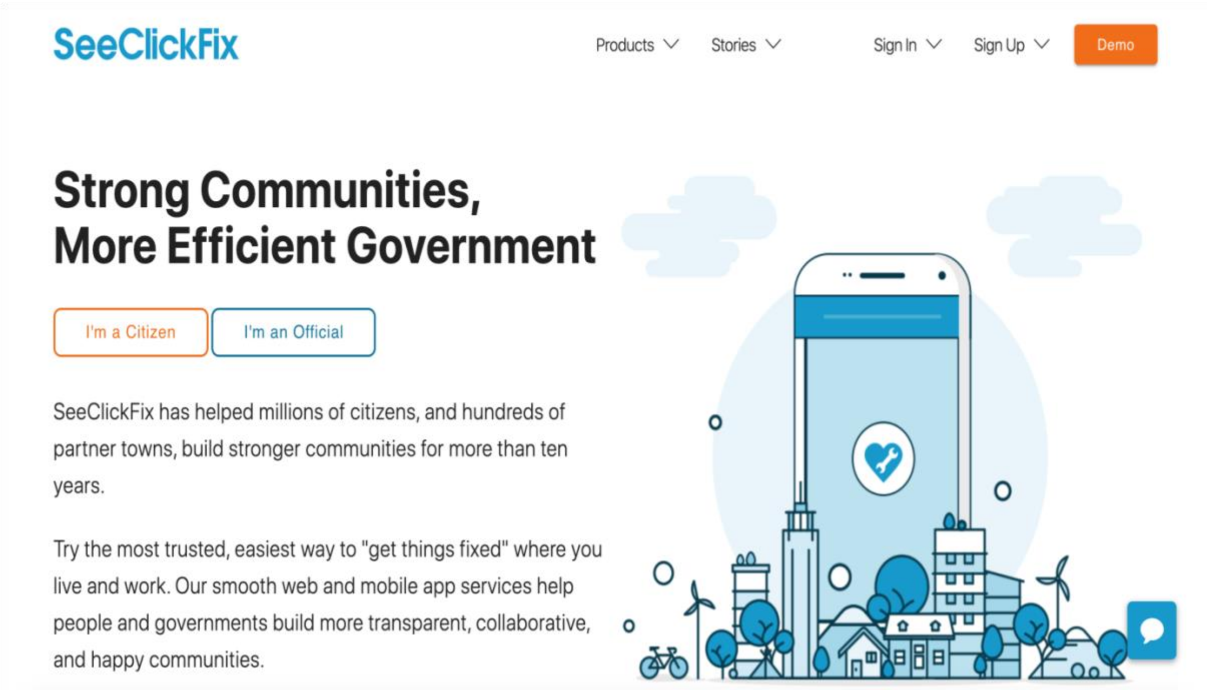
Through the mobile app users may add comments, pictures and videos as a documentation.

Currently spread over 300 cities in the US (e.g., Houston and Detroit) and on international scale, it also covers more than 25,000 towns and 8,000 neighborhoods. Those local municipalities usually pay the company in order to manage the whole flow of reports and also to act as back office support function for managing the orders. Through SeeClickFix more than 5 million issues have been fixed.

Users may anonymously report any problem, a condition which, according to the company, should encourage participation. Moreover, users can receive notifications based on selected areas and keywords, being perfectly conscious of what is happening in their own town or city.

As presented in websites through the slogan "get things fixed", SeeClickFix services help people and governments build more transparent, collaborative and trustable communities.

Fig.2. 4: SeeClickFix website homepage.



Source: https://seeclickfix.com/

Comuni-Chiamo is an Italian website and app which has had a critical success during the last 9 years, since its launch.

Nowadays Comuni-Chiamo serves 103 municipalities spread all over the country, more than 4 million citizens have actively used the website and app and there have been more than 5 million interactions between government and citizens; 233,000 issues have been solved by the municipalities thanks to the reports provided by the website and there are more than 7,000 municipalities' employees online.

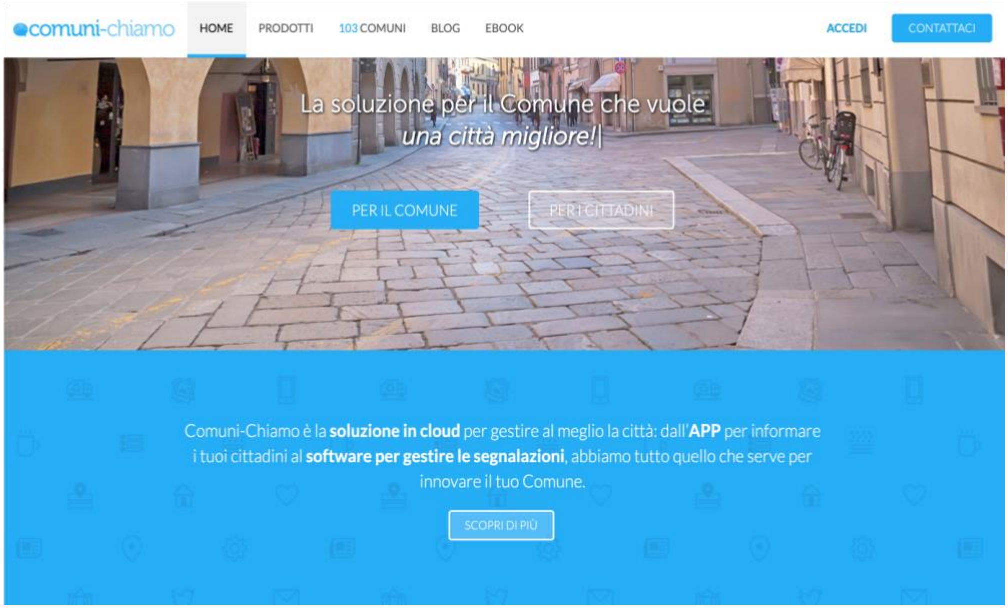
It represents the Italian version of FixMyStreet and SeeClickFix: citizens can actively inform governments about any issues which need their attention, moreover they can also suggest new ideas and offer different viewpoints from which the government can extrapolate crucial information.

On the other side, municipalities offices can also add all the reports that come through calls, emails and private messages to the platform. Thus, both players can share their information and issues for reaching a common goal: increasing the relationship between citizens and government, increasing transparency, improving benefits, services and welfare.

Other than gathering, managing and then forwarding all the reports coming from the citizens, Comuni-Chiamo also communicates the latest events and news directly linked to the municipalities' websites.

The open cloud platform is a free tool for citizens while governments and municipalities pay a fee for having access to the website. Such a form of collaboration may help governments to save money, to increase the flexibility of the processes, to reduce bureaucracy through an externalization of some activities and thus to better manage the territory, consequently improving the quality of people life. Indeed, thanks to the data provided by the Comuni-Chiamo website, the municipalities which have adopted it have noticed a 30% time reduction in solving issues.

Fig.2. 5: Comuni-Chiamo website homepage.



Source: https://comuni-chiamo.com

Twitter content analysis on a sample of Italian municipalities

3.1 An overview of the recent empirical literature on citizen engagement

Before introducing our analysis, the thesis provides a brief overview of the main empirical studies that have recently been conducted.

In order to select the academic contributions that best fitted our area of interests we decided to limit our research only to those papers able to empirically assess and measure the relationship between citizens and municipalities. The existing literature lacks in providing a significant amount of papers able to analyze the correlation between citizen engagement and municipalities' activity.

However, by deeply inspecting the most known scientific databases and selecting the most recent publications among business and economics journals related to the topic (such as Government Information Quarterly) we found out two publications made by Enrique Bonsón a financial economics and accounting professor from the University of Huelva (Spain).

After carefully examining the papers written by the professor Bonsón we then decided to replicate and to, possibly, implement them exploiting citizen engagement in the Italian municipalities, which, to our best knowledge, represents something never analyzed in such a manner, so far.

Bonsón through his studies provided a general overview of social networks usage by local municipalities and developed a mathematical formula to assess citizen engagement and how citizens are actively participating with governments, through Facebook and Twitter.

In the following lines the dissertation provides an overview of the studies made by Bonsón, outlining in more depth the empirical evidences and results achieved.

The first academic contribution that we have selected comes from Enrique Bonsón, Sonia Royo & Melinda Ratkai. Published in 2015, *"Citizens' engagement on local governments' Facebook sites: The impact of different media and content types in Western Europe"* (Bonsón, et al., 2015) aims at providing an overview of the Facebook usage by Western European municipalities considering two aspects: citizens' engagement and municipalities' activity on Facebook.

Data from 75 local governments spread all over 15 countries in Europe were analyzed and tested by the authors for both government use and citizens' engagement: 50 posts from each municipality were coded and divided into 16 different content types and 5 different media types, respectively. As explained in the paper, large-sized local governments were chosen due to their tendencies in adopting new technologies and implementing the Web 2.0 framework.

Through this paper the authors propose a new methodology in order to measure the citizen engagement on social media platforms: metrics of popularity, commitment, virality and engagement were computed for each post. Indeed, to measuring the citizens' engagement on Facebook's accounts, the authors examined the number of likes, comments and shares in order to assess three different metrics such as popularity, commitment and virality, along with an additional aggregated metric defined as engagement.

Then, the engagement metrics were computed among media, content types, public administration styles and whether the municipalities allowed posts by stakeholders in their home section on Facebook.

For the purpose of our analysis we adopted the new methodology introduced by the authors for assessing the citizen engagement on Italian municipalities, as it will be explained later in the chapter.

This paper offers a tremendous contribution to the citizen e- participation and government 2.0 literature demonstrating that the content and media types have an impact on stakeholders' engagement on Facebook. Results show that citizens tend to better engage with their local authorities on content related to the municipal management because closer to their everyday lives. The authors assert that by disclosing relevant information voluntarily discrepancies between local governments and population might be avoided. Furthermore, the study demonstrated that citizens tend to engage with more frequencies when a photo/video is involved in Facebook's posts, outlining how much citizens nowadays are more sensitive and interested on media types.

Moreover, engagement levels by the population and the influence of the different topics related to citizens' engagement seem to be dependent upon the public administration style, confirming that the institutional setting has an important influence on e-participation and citizen engagement (Bonsón, et al., 2015).

The second selected paper written by Enrique Bonsón along with David Perea & Michaela Bednárová (2019) titled *"Twitter as a tool for citizen engagement: An empirical study of the Andalusian municipalities"* (2019), represents the milestone of our analysis.

Indeed, we adopted the empirical procedures and methods written by the Bonsón, Perea & Bednárová to analyze our case study.

This paper, published in the *Government Information Quarterly* business and economics journal, provides a general overview of the way local governments use Twitter as a communication tool to engage with their citizens, identifying factors associated with both the activity channels and citizen engagement in order to understand the relationship between media type and citizens and whether different content types generate different levels of engagement (Bonsón, et al., 2015). Citizen engagement is understood to include social capital, civic engagement and political participation, and thus Big Data technologies through social media can be seen as a tool to boost citizen engagement through communication, discussion and the coordination of public and social activities.

The analysis was performed on 29 municipalities belonging to the Andalusia region, which represents the first study ever made providing a general overview of Twitter usage by Andalusian local governments. Tweets from the official accounts of the local governments were downloaded and scraped through the Python software. Then a text mining analysis was performed, through a statistical open source software called R Studio. Different content types and media types categories were identified and along with Twitter metrics such as the number of followers and the citizens' engagement retweets, replies and likes the authors were able to assess which were the main content and media with the highest engagement. Retweets represented the most frequent way for citizens to interact with the city council account, even though engagement appeared not to be related to the population of a municipality. The findings of the study also show that photos and videos generate more engagement than any other media (Bonsón, et al., 2015).

The authors, however, found no significant relationship between the population of a municipality and its citizens' engagement and between activity and engagement.

For the sake of simplicity, below are summarized in the Tab. 3.1 the main contribution discussed above.

Tab. 3. 1: Review of the main academic contributions.

Author(s)	**Objective and focus of the research**	**Type of analysis and data source**
Bonsón, Royo & Ratkai, (2015)	Investigation which aims at providing an overview of the Facebook usage by Western European municipalities considering two aspects: citizens' engagement and municipalities' activity on Facebook.	Quatitative analysis on 75 local governments in the Western Europe. Data includes 3750 posts on Facebook analyzed.
Bonsón, Perea & Bednárová, (2019)	Overview on how local governments adopt Big Data and Twitter as communication tools to engage with their citizens.	Quantitative content analysis on 29 municipalities in Andalusia (Spain). 345,960 tweets scraped through Python. Text mining analysis performed through RStudio.

Literature on Twitter linked to the usage by local governance, appears to be scarce, given that the most recent studies focus on national level (Sobaci & Karkin, 2013). Despite the presence of few publications on Twitter and its usage in local governments, none of them cover the citizen engagement topic, rather they focused in analyzing the social network platform as a tool for election campaigns pools and predictions purposes (Xenos, Macafee & Pole, 2017; Mambrey & Dörr, 2011).

Examining the literature, we can affirm that there is a lack of studies analyzing citizen engagement on Twitter and in advanced most of the researches investigated on US cases, thus it represents a lack of the literature in exploring this phenomenon in other regions such as Europe.

3.2 Hypotheses definition

As described in the previous chapters, the wide usage of Big Data technologies helps achieving new frameworks and paths for the development of smart cities, for instance recurring to the adoption of new technologies.

Smart cities initiatives are no longer just about optimizing traffic patterns, parking management, efficient lighting, improvements to public works and etc., rather citizen engagement is considered critical to accomplish a smart city pattern (Bonsón, Royo & Ratkai,2015; Bonsón, Perea & Bednárová,2019; Gartner, 2018).

Existing academic contributions have devoted great attention to the technological development of smart cities, with the introduction of the Web 2.0 and the implementation of crucial technologies in our daily lives, nevertheless just few studies have been conducted in understanding how local governments effectively interact with their citizens through the usage of such technologies.

The main purpose of our study is to fill this gap in the literature by providing a general overview of how municipalities in Italy use social networks, in our case we decided to choose Twitter, as a social media able to interact with their citizens. Moreover, the research aims to find out which factors influence citizens engagement considering: municipalities' size, Twitter activity, audience, content and media types. Hence:

RQ1: Which are the main areas of interactions between local governments and citizens?

RQ2: How do municipalities, based on citizens engagement, face the transition towards smart cities?

3.3 Empirical setting: some considerations about Italian municipalities

Before introducing our research methodology we would like to provide some consideration about the Italian municipalities relationships with the web and social networks.

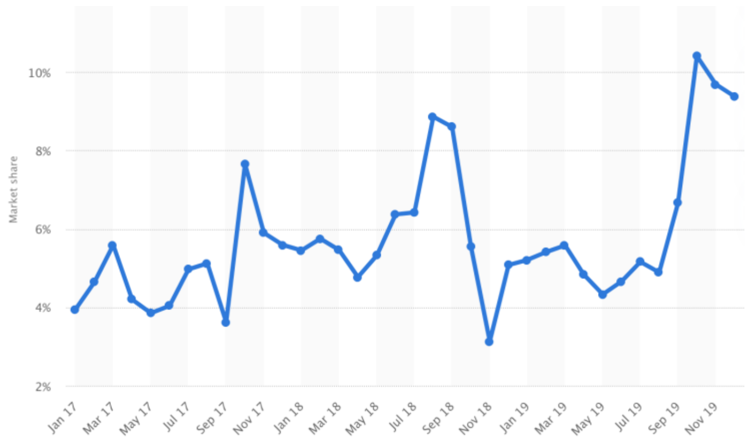
At global level, the United States are no longer leaders in the use of social media, indeed regions such as EMEA (Europe, Middle East and Africa) have gained much ground (Bonsón, et al.2019). Furthermore, those differences between regions might be explained by Hofstede's cultural dimensions' theory, which have been observed also within regions belonging to the same country (Hofstede, 2007).

Regarding European countries, 64% of Italians habitually use social networks whilst the European average stands at 65% (European Commission, 2018).

According to the FPA Digital 360 (Orlando, Puelli, Vicentini,2019) 100 Italian municipalities out of 107 (99 municipalities in 2018) have at least one social media account, while on average one out of two municipalities have at least two social media accounts. The most adopted social network by Italian municipalities is Facebook, followed by YouTube which drastically increased during 2019 compared to 2018 and then Twitter which has a market share of roughly 10% according to the European Commission studies, as displayed below in Fig. 3.1.

Twitter is also considered the most reliable and serious social networks by Italians municipalities according to the FPA's research, moreover even citizens tend to rely much more on Twitter rather than Facebook for being informed about the latest news from their local governments.
Despite the character limitation, which is 280 characters per single tweet, might be seen as a drawback, Kim, Chun, Kwak & Nam (2014) stressed that short messages can be published more frequently and that their content might reach a wider audience (Bonsón, et al., 2019).
Twitter is an innovative tool with open and horizontal networks (Honey & Herring, 2009), which is able to give access to the users to an incredible number of followers and diversified content. Hence, Twitter appears to be the right tool for studying citizens engagement.
Nevertheless, according to Bonsón et al., (2019) the passive presence on social media is not enough for understanding citizens' needs: local governments should clearly define their online communication strategies enhancing citizen engagement.

Fig.3. 1: Social network market share held by Twitter in Italy from January 2017 to December 2019.



Source: https://www.statista.com/.

3.4 Research Methodology

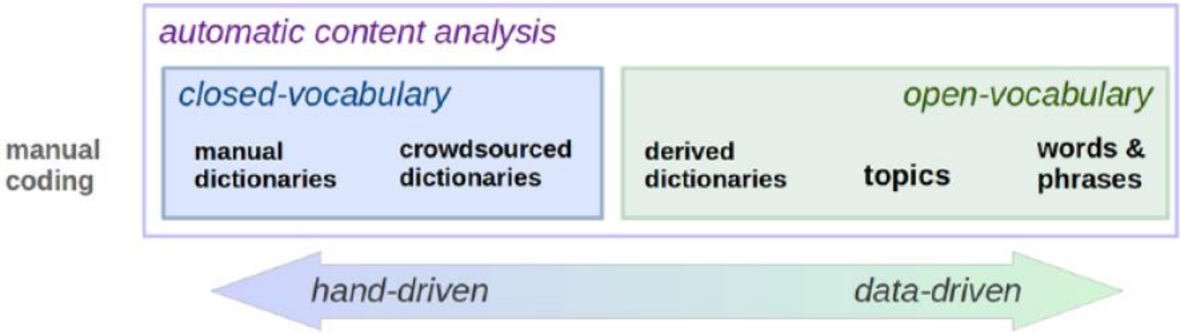
3.4.1 Content analysis

The proliferation of social networks, such as Twitter, offers new and alternative measurement approaches: for instance, automated content coding and the statistical power to do open-vocabulary exploratory analysis (Schwartz & Ungar, 2015).
In our research, through content and text mining studies we analyzed all the tweets coming from different municipalities for understanding the rate of citizen engagement between the Italian population and their municipalities, as already stressed above. For doing so, we proceed a content analysis, which is a quantitative analysis of the words and concepts expressed in texts, that has been largely used across the social sciences to analyze people's communications (Schwartz & Ungar, 2015).

Through social media there is access to enormous samples, via Facebook and Twitter for instance, and it changed the way in which content analysis can be used to understand people: indeed through tweets or status updates on Facebook, people post their daily activities and thoughts, then researchers began leveraging these data for a wide range of applications including monitoring health diseases or for predicting the stock market or simply by understanding people's feelings (Schwartz & Ungar, 2015).

For the purpose of our research we needed to set automatic vocabularies in order to assess the different content categories of words (Fig. 3.2). Setting automatic vocabularies represents a crucial phase of the content and text mining analysis: according to Schwartz & Ungar, 2015 automatic content analysis techniques range from "hand- driven closed- vocabulary" to more "data-driven open- vocabulary", indeed several factors tend to influence the choice (accessibility, sample size requirements, scientific approach and transparency). Hand- driven techniques tend to be more accessible, theory driven and abstract, whilst data driven ones tend to be more transparent, capturing more connections (Schwartz & Ungar, 2015).

Fig.3. 2: categorization of content analysis techniques.

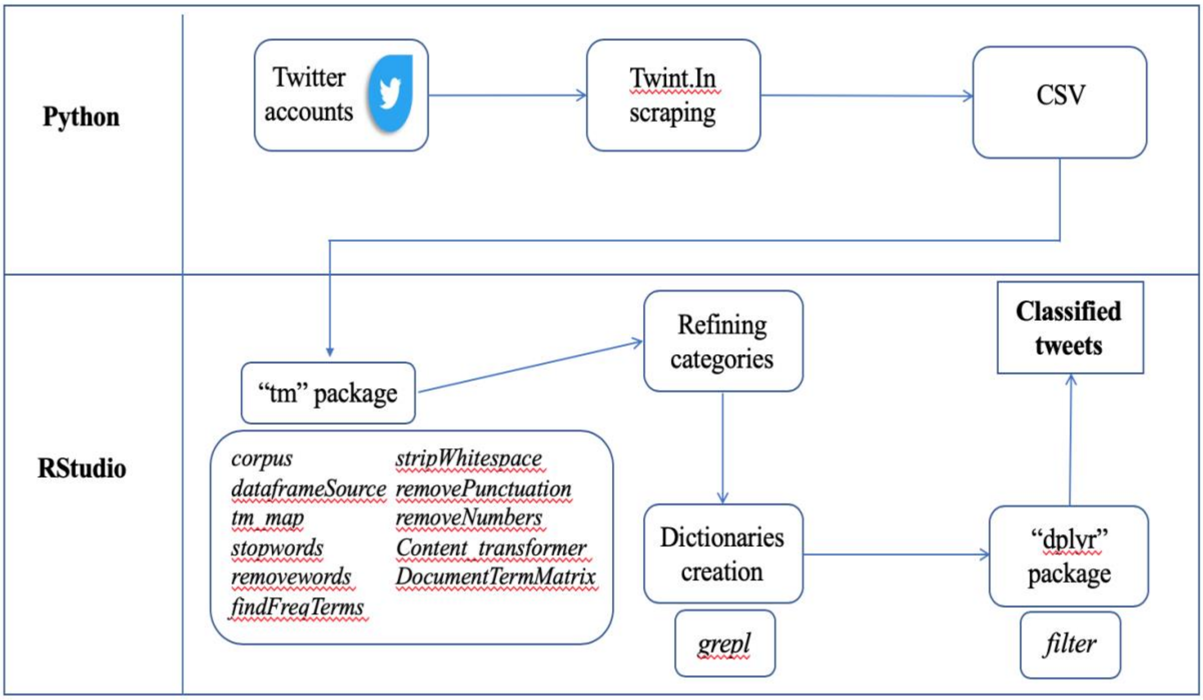


Source: Schwartz & Ungar, 2015, p.80.

The process of content analysis has been studied by several authors (Erlingsoon & Brysiewicz, 2017; Kohlbacher, 2006; Neuman, 2014), nevertheless for our research we decided to take the approach suggested by Bonsón, et al., 2019 who refined the content analysis process in three different steps: sampling and data collection, coding and analyzing the content.

In the figure displayed below (Fig. 3.3) is summed up the whole process of content analysis that we conducted.

Fig.3. 3: flowchart of the content analysis process.



Source: adaptation of Bonsón, et al., 2019, p484.

3.4.2 Sampling and data collection

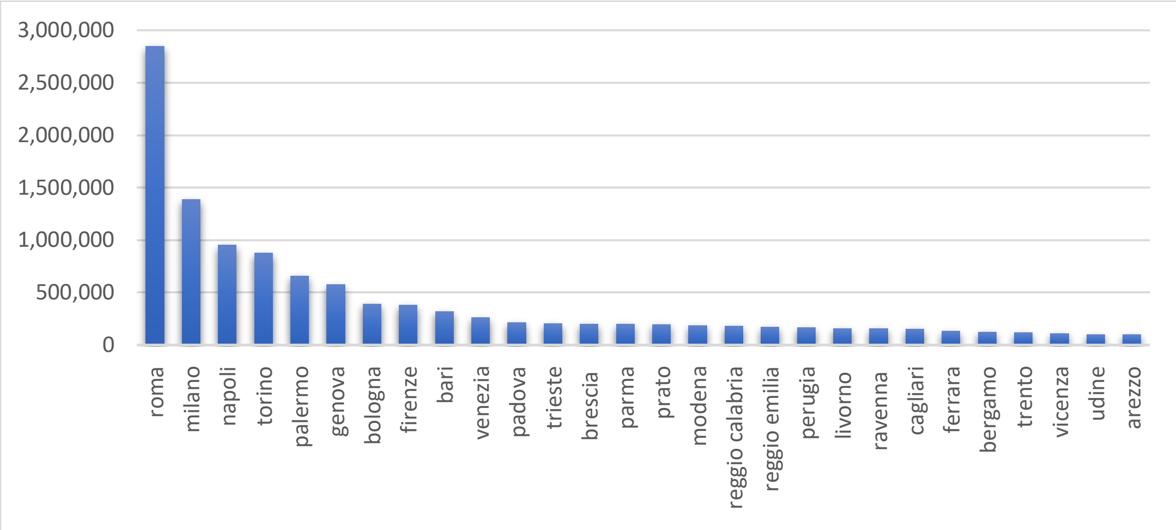
3.4.2.1 Sampling

For the creation of our sample we decided to define certain thresholds in order to deal with a uniform and reliable set of data. As anticipated above, our analysis starts with a selection of the largest Italian municipalities ranked by population. The sample of this study is taken from the Instituto Italiano di Statistica (INSTAT) which owns all the data related to the Italian population.From the data found in the INSTAT platform we selected only those municipalities exceeding 99,000 inhabitants which results in 48 cities. We then checked whether the municipalities had an official online website: the results showed us that 46 out of 48 municipalities had an official online website, we then took out the two municipalities which lacked of an official online website from our sample.

In the second step the Twitter account of each municipalities was collected by clicking the icon link on the official website previously found or by manually searching the Twitter platform to identify the verified account. From the 46 municipalities left we found out that 10 municipalities do not have an official Twitter account and thus we deleted them from our set of data.

The last step before starting the download process of the tweets was then performed: this phase of the process regards the activity of the single account, indeed we selected only those accounts which issued at least one tweet in the previous 6 months (the threshold limit was then set on April 2019, given that the data collection started in October 2019). Moreover, we included an additional threshold: each account needed to have issued at least 1,000 tweets since joining Twitter. We set these thresholds in order to avoid biases due to inactive accounts or fake accounts that could have affected our analysis. We then obtained our sample: composed by 28 municipalities among the largest Italian cities (Fig. 3.4) with an active Twitter account.

Fig.3. 4: number of inhabitants per municipality.



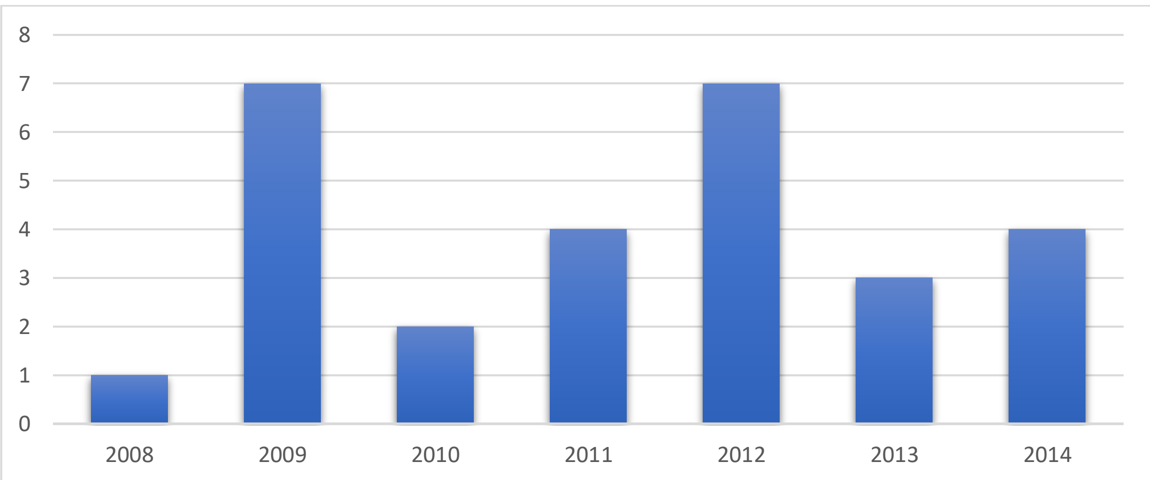
Source: own elaboration.

In the next lines, features and settings of our sample will be explained, in order to give a broader understanding of the topic to the reader.

A characteristic to consider is the "age" of the Italian municipalities Twitter accounts, i.e. when the *join in* on the official social network was performed, an information we have retrieved from the official Twitter accounts of each city. Municipalities represent a homogenous group in terms of date of *join in.*

As can be observed from the figure 3.5 (Fig. 3.5), which shows the number of *joining* for each year, the first official Twitter account was opened by an Italian municipality (Torino) in 2008. According to the data, we know that the mean is 2011 and the standard deviation is 2,16, meaning that on average an account deviates from the mean of about 2,16 years, outlining the short temporal window in which the accounts were created.

Fig.3. 5: Twitter's accounts joining date.



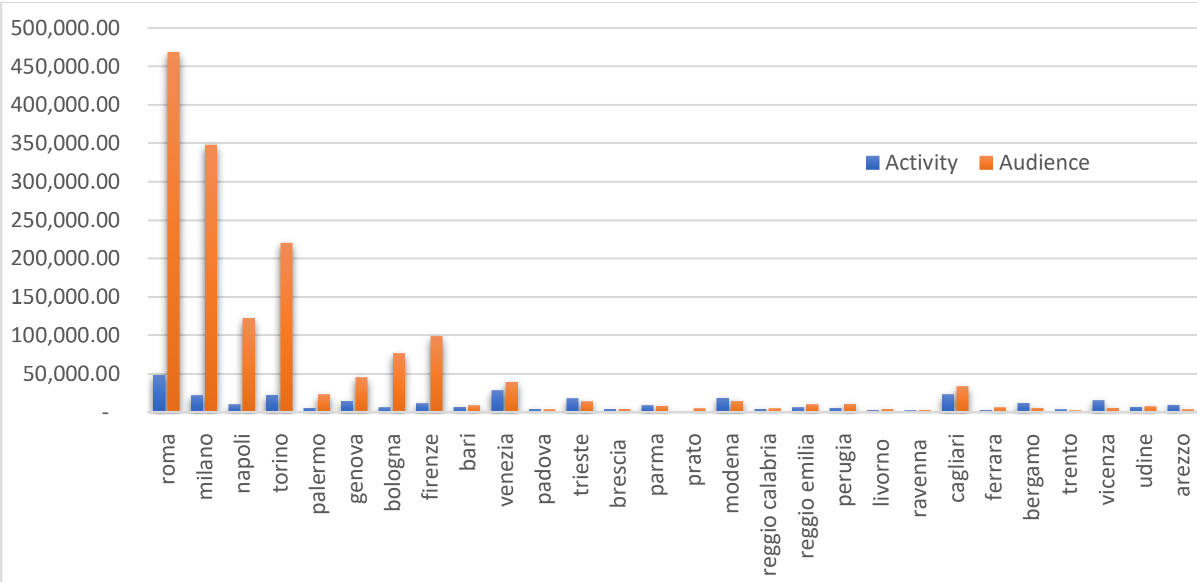
Source: own elaboration.

The graph attached below (Fig. 3.6) shows an heterogenous situation on Twitter by the largest Italian municipalities: volume of tweets (activity) and number of followers (audience) seem to drastically vary from one municipality to the other.

The most active accounts on Twitter are: Roma (48,402 tweets), followed by Venezia (27,921 tweets), Cagliari (22,666 tweets), Torino (22,199 tweets) and Milano (21,720 tweets).

On regards of the number of followers, the accounts belonging to the largest Italian cities registered the higher audiences: Roma (468,600 followers), Milano (347,800 followers), Torino (220,400 followers) and Napoli (122,100 followers).

Fig.3. 6: activity (total tweets) and audience (total followers) of the Italian municipalities (ranked by population).

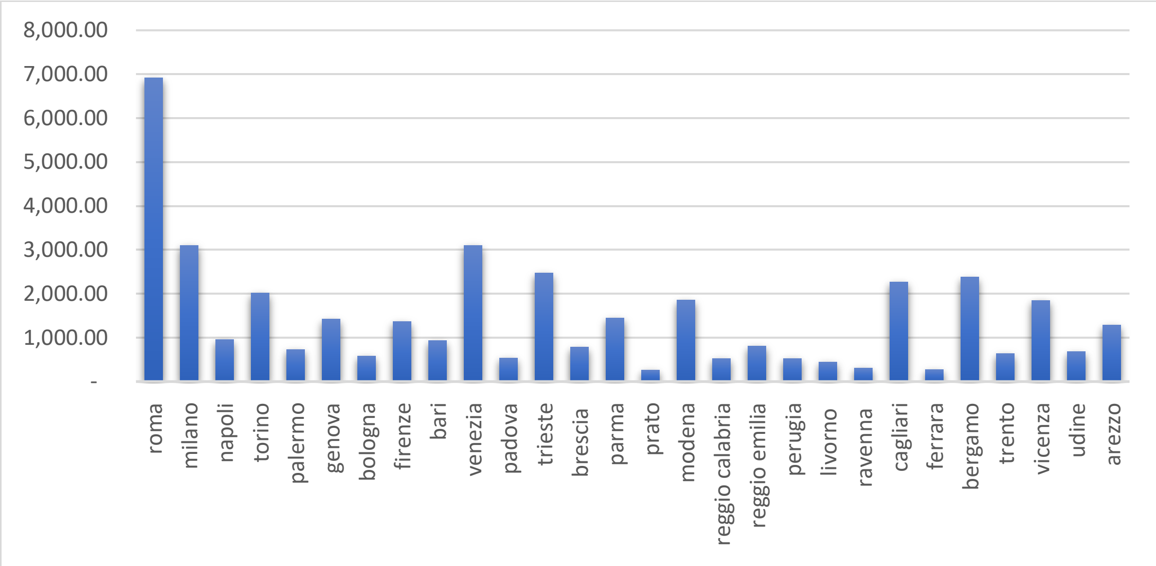


Source: own elaboration.

In the graph displayed in Fig. 3.7, we measured, instead, the rate between the sum of the activity (total tweets) leveraged over the number of years, ranked by descending population (from the largest to the smallest), since each municipality joined Twitter.

The results show that throughout the years Roma has kept his account far more active compared to other municipalities, more than doubling the results obtained by cities such as Milano, Napoli and Torino. On the other hand, the graph shows us some unexpected evidences: cities such as Bergamo, Cagliari, Trieste and Venezia, despite their lower number of inhabitants compared to other municipalities, have been really active through the years.

Fig.3. 7: average activity per year since joining Twitter by Italian municipalities (ranked by population).



Source: own elaboration.

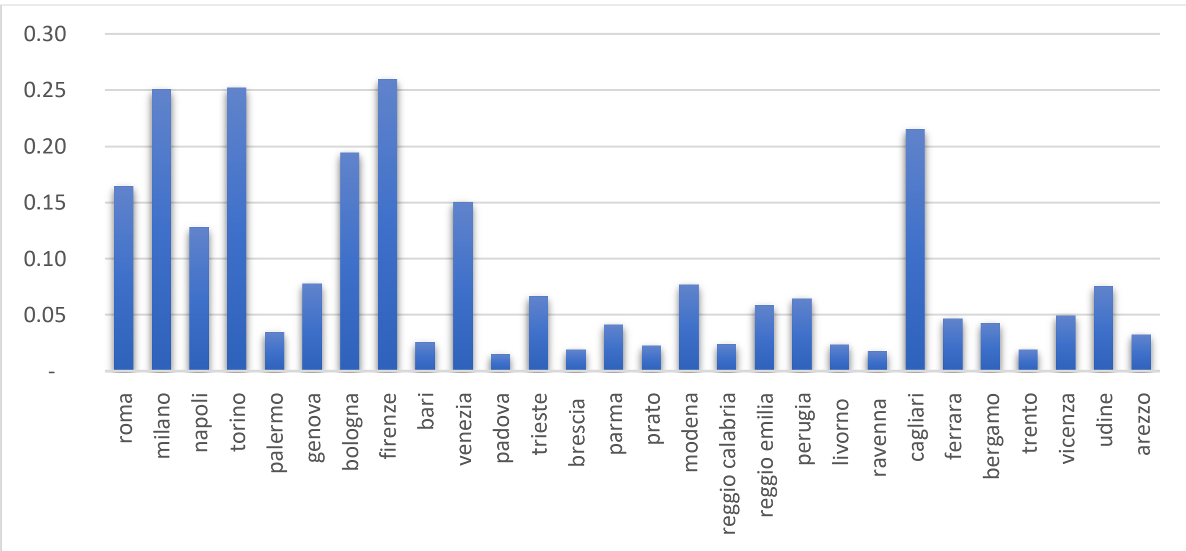
Figure 3.8 (Fig. 3.8) wants to make a comparison between the audience and the number of inhabitants per municipality. This percentage does not represent the number of inhabitants who own a Twitter account, nevertheless it represents the number of followers, belonging to the same municipality, who actually *follow* the official municipality Twitter account.

Through the graph attached below in figure 3.8 (Fig. 3.8), we want to provide a proxy of the potential magnitude of followers a municipality can have compared to its own population.

Firenze represents the municipality with the highest rate (25,9%), followed by Torino (25,2%), Milano (25,1%), Cagliari (21,5%) and Bologna (19,4%).

As we will see later on, this rate and generally the audience of each account do not imply a higher citizen engagement.

Fig.3. 8: percentage of followers over number of inhabitants per municipality (ranked by population).



Source: own elaboration.

3.4.2.2 Data collection

The data collection took place between October and November 2019 using Twint.In a Python library which allows to perform Twitter scraping (Fig. 3.9).

Python is an interpreted, object- oriented, high-level programming language with dynamic semantics. It is extremely attractive in the field of Rapid Application Development because it offers dynamic typing and dynamic binding options. Through a Python library called Twint.In we scraped every tweet from each account. Twint.In is an advanced Twitter scraping tool written in Python language that does not use Twitter's API, allowing you to scrape a user's followers, following, tweets and much more while evading most API limitations. Twint.In utilizes Twitter's search operators to let scraping tweets from specific users.

The tweets then were automatically scraped in a csv format and stored in different folders. In the end, the total number of downloaded tweets amounted to 316,801 coming from 28 different municipalities.

Fig.3. 9: Python code adopted for scraping the tweets..

```
The default interactive shell is now zsh.
To update your account to use zsh, please run `chsh -s /bin/zsh`.
For more details, please visit https://support.apple.com/kb/HT208050.
MacBook-Pro-di-Dodo:~ edoardogobbo$ ls
Applications            Dropbox                 Pictures
Creative Cloud Files    Dropbox (Precedente)    Public
Desktop                 Library                 PycharmProjects
Documents               Movies                  Untitled.R
Downloads               Music                   iCloud Drive (Archivio)
MacBook-Pro-di-Dodo:~ edoardogobbo$ cd Desktop/E
MacBook-Pro-di-Dodo:E edoardogobbo$ ls
comuni.txt      script.py
MacBook-Pro-di-Dodo:E edoardogobbo$ python
comuni.txt  script.py
MacBook-Pro-di-Dodo:E edoardogobbo$ python
comuni.txt  script.py
MacBook-Pro-di-Dodo:E edoardogobbo$ python script.py
```

Source: own elaboration.

3.4.3 Coding and analyzing the content

Before starting to analyze and coding the tweets, the content categories of the tweets were identified and defined and the metrics for assessing citizen engagement were established.
The identification of the main content categories was based on the lists of local services provided by Torres & Pina, 2001 and then revised by several authors such as Bonsón, et. Al, 2015 and Martí, et al., 2012. We also revised and adapted our list according to the most frequent words appeared in the tweets after the analysis was accomplished, as it will be later explained (Tab. 3.2).

Tab. 3. 2: content categories.

Content type	Tweet content
Cultural	Cultural activities and events.
Employment and Education	Employment and training schemes. Public education.
Environment	Environmental concerns.
Security and Health	Citizen Protection and security. Health service.
Sport	Sport activities and events
Transport and Public Works	Public and private transport. Public works in the city. Town Planning.
Others	

Source: Bonsón, et al., 2019, p. 484.

As stressed in the previous lines, for the purpose of our analysis, in order to assess the citizen engagement per municipality, we utilized formulas and metrics frameworks developed by Bonsón, et al., 2019.As can be detected from the table below (Tab. 3.3), for assessing the citizen engagement the authors introduced three different metrics, such as: "Popularity" which is based on the number of "likes" received by each municipality, "Commitment" which is based on the number of "replies" and "Virality" which refers to the number of retweets obtained.
Once obtained the three different metrics per 1,000 followers, as shown in the table (Tab. 3.3), the "Engagement" can be easily achieved by summing the components up.

Tab. 3. 3: metrics for citizen engagement.

Metrics	Code	Calculation
Popularity	P1	Number of tweets favorited/total tweets
	P2	Total number of times favourite/total tweets
	P3	(P2/number of followers)*1000
Commitment	C1	Number of tweets commented/total tweets
	C2	Total number of comments/total tweets
	C3	(C2/number of followers)*1000
Virality	V1	Number of tweets retweeted/total tweets
	V2	Total number of retweets/total tweets
	V3	(V2/number of followers)*1000
Engagement		P3 + C3 + V3

Source: Bonsón, et al., 2019, p. 484.

Before coding started, we "cleaned" our dataset with numerous functions provided by the statistical software RStudio: for instance, "NAs" value in our dataset were identified and deleted, useless columns and information for the purposes of our study were taken off (such as "cashtags", "quote_url", "geo", "source", "user_rt", "retweet_id", "retweet_date") and dummy variables were applied.

In order to proceed with our research, the RStudio package "tm" was installed and its library launched. The "tm" package is a text- mining tool which offers some powerful functions which aids in text processing steps (Bonsón, et al., 2019) (Fig. 3.10).

Fig.3. 10: RStudio script for text mining analysis.

```r
library(tm)
data_frame <- data.frame(doc_id=doc_ids,
                         text=txt,
                         stringsAsFactors=FALSE)

(df <- VCorpus(DataframeSource(data_frame),
               readerControl = list(language = "ita")))

removeURL <- function(x) gsub("http[^[:space:]]*", "", x)
corpus <- tm_map(df, content_transformer(removeURL))

#clean text
corpus<-tm_map(corpus, content_transformer(tolower))
inspect(corpus[1:5])

corpus<- tm_map(corpus, removePunctuation)
inspect(corpus[1:5])
corpus<- tm_map(corpus, removeNumbers)
inspect(corpus[1:5])
corpus <- tm_map(corpus, stripWhitespace)

cleanset <- tm_map(corpus, removeWords, stopwords("italian"))
stopwords("italian")
inspect(corpus[1:5])
cleanset<- tm_map(cleanset, stemDocument)
stopwords("italian")

#document term matrix
dtm <- DocumentTermMatrix(cleanset)
dtm
inspect(dtm)
```

Source: own elaboration.

The main structure for managing documents in tm is a so-called Corpus, which represents a collection of text documents: a corpus is an abstract concept and its default implementation is the so-called VCorpus (Volatile Corpus) (Feinerer, 2019).
For inspecting such a great number of tweets at the same time, DataFrameSource function was used which is able to handle a directory or a vector interpreting each component, in our case each tweet, as an independent document. Then the "transformations" phase of our study began. Once the Corpus was defined, we wanted to modify the documents getting rid of all those "useless" information contained in the tweets. Transformations of the Corpus were done via the tm_map function which applies a function to all the elements of the corpus: meaning that all tranformations work on single text document and tm_map applies them to all the documents within the corpus (Feinerer, 2019).
Hence, transformation such as removePunctuation, stemming, stripWhitespace removal and content_tolower were applied. In advanced, two of the most powerful functions at disposal, such as stopwords and removewords, which are able to identify and delete all the most common words (articles, verbs, etc.), were used (Fig. 3.10).
The corpus was then transformed in a DocumentTermMatrix which employ sparse matrices for the corpus. Through a DocumentTermMatrix which transforms the corpus in a sort of matrix, various quantitative functions for text analysis, such as removeSparseTerms and findFreqTerms, can be performed.
Afterwards, the process of storing the words into categories previously defined (Tab. 3.2) started: words that appeared with a frequency equal to 4500 were identified, tabulated, mutually compared and adjusted. A dictionary was created for each category, where the words that identify each content were found (Bonsón, et al., 2019).
The creation of the dictionaries was done through the function of the base package of R called "grepl". Performing this classification automatically reduces subjectivity nevertheless implies that there are tweets that share more than one category, hence giving an overlap error (Bonsón, et al., 2019). Such changes were included into the content analysis path and the final coding was accomplished.

After the coding phase, the tweets were analyzed and classified automatically (Tab. 3.3) using the RStudio library "dplyr". This library represents a powerful package able to transform and summarize tabular data with rows and columns. The package contains a wide set of functions that perform common data manipulation operations such as filtering for rows, selecting specific columns, re- ordering rows, adding new columns and summarizing data (Bonsón, et al., 2019).

Moreover, in order to classify the tweets matching them to the exact category previously set, the "filter" function was used.

3.5 Descriptive statistic results

As anticipated in the previous paragraphs, the account with the largest audience had 468,600 followers, while the account with the lowest audience had just 1,316 (Tab. 3.4).
Due to the high standard deviation numbers for both the variables, there are enormous differences regarding the total number of tweets and the total audience. The average activity (number of tweets for single municipality) amounts to 11,314, whilst the average amount of followers per municipality is 56,795. The average amount of audience is clearly biased by the number of followers belonging to the largest cities in Italy (Roma, Milano, Torino and Napoli).

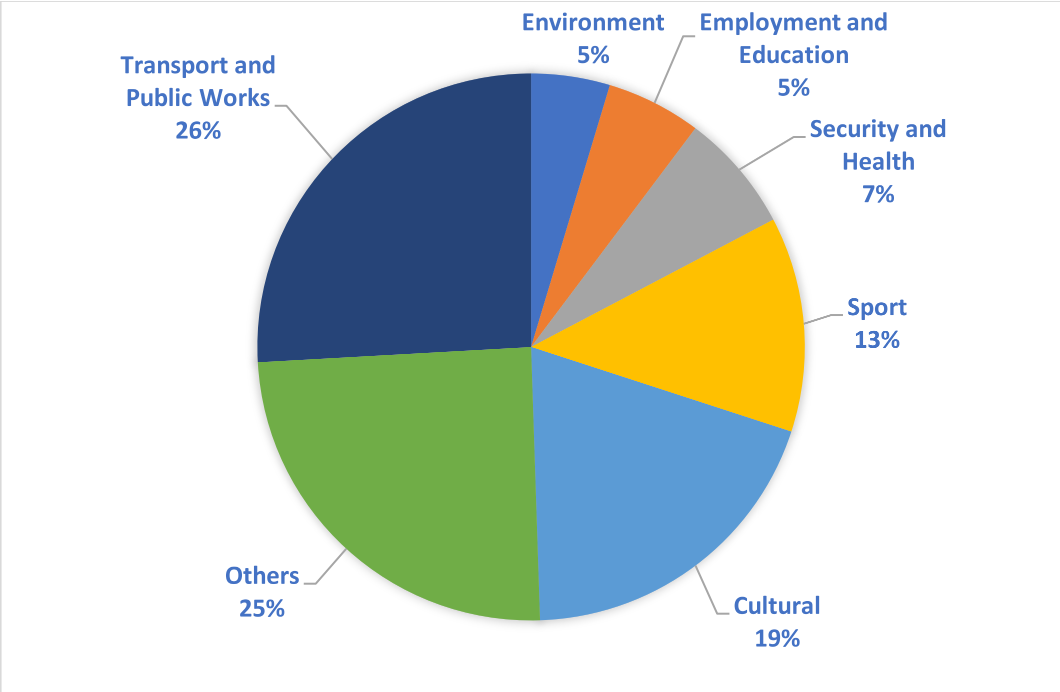
Tab. 3. 4: activity and audience empirical results.

	Activity	**Audience**
Average	11.314,36	56.795,57
Median	7.733,50	9.169,00
Maximum	48.402,00	468.600,00
Minimum	1.316,00	2.220,00
Std. Deviation	10.321,33	111.446,45

Source: own elaboration.

The results regarding the most content types used among the largest municipalities in Italy are displayed in the graph attached below (Fig. 3.11).
The results show that municipalities tend to adopt Twitter account as a channel of communication in order to inform the population about the transports, public works and new town planning ideas, given that 25,89% of the tweets fall under this category.
The second content category type, represented under the name "others" (24,63%), covers a diversity of topics that had not been included in our category selection. Cultural (19,47%) and Sport (12,72%) are the other categories above the 10% threshold. In particular, cultural activities and events seem to appear as crucial topics used by the municipalities for engaging with the population.

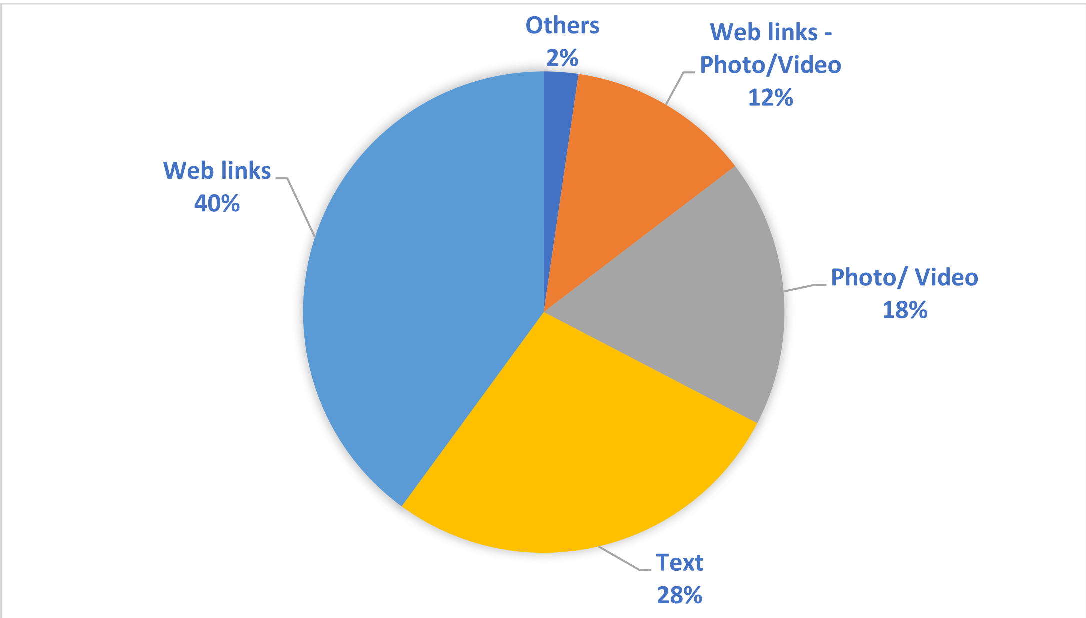
Fig.3. 11: content categories results.



Source: own elaboration.

Figure 3.12 (Fig. 3.12) shows instead that the preferred media type adopted by the Italian municipalities is the website link (39,95%), which nowadays is roughly included in one out of two tweets produced. Although, the massive presence on the social network of photos and videos, the plain text still seems to be one of the most preferred way of communication adopted by municipalities. Photos and videos or the combination of web links- photos and videos appear with percentage of 18,01% and 12,32 %, respectively.

Fig.3. 12: media type results.



Source: own elaboration.

As previously stressed, in order to assess the citizen engagement between the Italian municipalities and the population, the metrics proposed by Bonsón, et al., (2015) were adopted. The results are shown in the table 3.5 (Tab. 3.5).

Tab. 3. 5: *metrics for citizen engagement.*

	Max	Average	Min	Std.deviation
Popularity				
P1	0,86943824	0,42809801	0,14086629	0,188947599
P2	11,8008553	2,16548154	0,16705465	2,568666337
P3	0,75714314	0,14798961	0,013935	0,165891054
Commitment				
C1	0,49788214	0,11194724	0,01363485	0,111304662
C2	0,91118785	0,17493106	0,01407826	0,232629249
C3	0,03635865	0,00988196	0,00181481	0,007941579
Virality				
V1	0,93790275	0,46104294	0,0287071	0,214761821
V2	5,9154787	1,52004359	0,1354617	1,435520691
V3	0,45039497	0,12046044	0,01233172	0,118683037
ENGAGEMENT ITALY (P3+C3+V3)	1,24389676	0,27833201	0,02808153	0,292515669

Source: own elaboration.

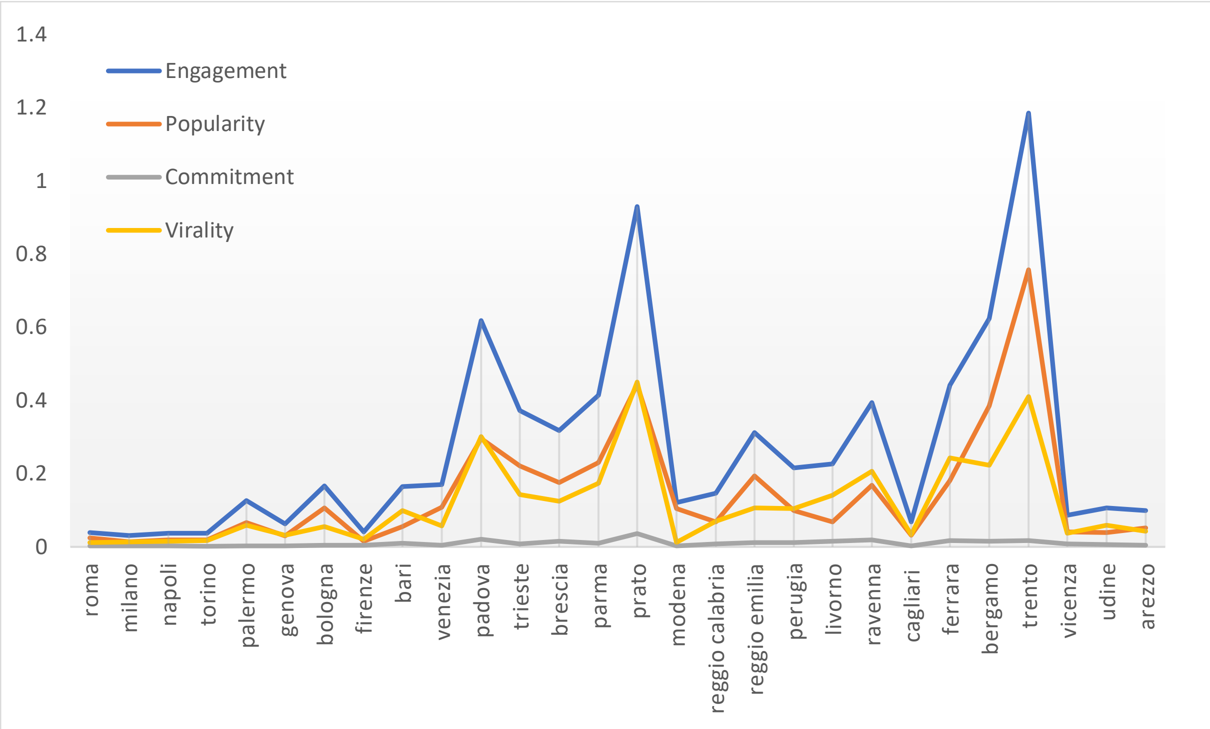
As shows in the attached table, on average roughly 42% of the tweets are liked as represented by P1, 11% of the tweets are replied (C1) and 46 % of the tweets are retweeted (V1) by citizens. Moreover, as can be noticed looking at P2, C2 and V2 the average number of likes and retweets per tweet are pretty high, 2,17 and 1,52 respectively, while the average number of replies per tweet is really low (0,17). According to the methodology proposed by Bonsón, et al. (2015), dividing P2, C2 and V2 by the number of followers and multiplying it by 1,000, we obtained the average number of likes, replies and retweets per tweet per 1,000 followers.

The results obtained through the last step shows us that among three different ways to interact with municipalities, citizens tend to prefer likes more often than replies and retweets as a means of engaging with municipalities (Bonsón, et al., 2019).

In the end, by summing P3 (14,79%), C3(0,0098%) and V3(12,04%), we obtained the average citizen engagement (E) Italian index of 27,83%.

By applying the methodology proposed by Bonsón, et al., (2015) to each one of our sample's municipalities, we obtained the following results, as can be noticed from the graph (Fig. 3.13). The municipalities with the highest levels of citizen engagement are Trento, Prato, Bergamo, Ferrara and Parma. Moreover, by looking at the shaping of the lines, it appears that smaller municipalities have higher levels of citizen engagement. Perhaps, this unexpected result is due to the higher levels of closeness to the municipality that citizens from smaller cities tend to show.

Fig.3. 13: citizen engagement results per municipality (ranked by population).



Source: own elaboration.

3.5 Statistical analysis

For the purpose of our study we were also interested in identifying the factors associated with citizen engagement.

Indeed, we wanted to understand whether the audience (number of followers) and the activity (number of tweets) of municipalities might have an impact on the engagement and thus, if the municipality size was somehow linked with it. The great numbers showed by certain municipalities regarding audience and activity might bias the idea of citizen engagement as it is, indeed as already anticipated in the previous paragraph, smaller municipalities and more generally small villages, tend to care much more about their cities compared to those living in the largest ones. Engagement, once again, stands for collaborating with the local governments in enhancing old services and providing new ones for reaching a common wellness.

As can be noticed from the table 3.6 (Tab. 3.6) attached below, evidences show significant negative relationships among citizen engagement, which is our dependent variable, activity, population and audience which represent our independent variables. For understanding the relationships among the variables the Spearman's coefficient was applied.

The Spearman's correlation between two variables is equal to the Pearson's correlation between the rank values (statistical dependence between the rankings of two variables) of those two variables; while the Pearson's correlation gives linear relationships, the Spearman's correlation assesses monotonic relationship. In other words, the Spearman's correlation between two variables will be higher when the observations have similar rank and lower when observations have a dissimilar rank between two variables.

Tab. 3. 6: Relationship between activity, population, audience and citizen engagement.

Dependent variable	**Independent variable**	**Spearman's coefficient**	**Significance**
Engagement	*Activity*	-0.6551724**	0.0002183
	Population	-0.4499179	0.01716
	Audience	-0.7192118**	0.000002701

***** Significant at $p < 0.01$ (2-tailed)***

Source: own elaboration.

After assessing the negative relationships among citizen engagement, activity, audience and population we wanted to deeply analyze the relationships between media types and citizens reactions on Twitter (likes, replies and retweets).

In the table 3.7 (Tab. 3.7) are summarized the descriptive statics regarding the media types and the citizens engagement.

Tab. 3. 7: Descriptive statistics of media types and citizens engagement.

Media Type		Likes	Replies	Retweets
Web links	Mean	0.63	0.09	2.01
	Std.Deviation	3.07	0.61	4.01
Text	Mean	0.68	0.78	0.61
	Std.Deviation	2.90	0.19	3.53
Photo/ Video	Mean	4.09	0.28	4.12
	Std.Deviation	13.34	0.90	8.69
Web links - Photo/Video	Mean	3.01	0.25	2.39
	Std.Deviation	4.46	0.98	8.92
Others	Mean	3.90	0.59	2.49
	Std.Deviation	16.11	1.45	11.16
Total	**Mean**	**12.31**	**1.99**	**11.62**
	Std. Deviation	**39.88**	**4.13**	**36.31**
	N	**936,713**	**77,342**	**732,192**

Source: own elaboration.

This table provides a really interesting overview of the ways in which citizens tend to engage. According to the results, all those tweets that contain photos or videos generate much more likes and retweets compared to the other media types. This outlines a common global trend, indeed due to the massive usage of smartphones and other devices, photos and videos seem to be a more direct way for interacting with the others: social networks widely used, such as Instagram and Snapchat that are based on sharing photos and videos represent two of the greatest example of that trend.

While, when dealing with plain text tweets citizens tend to reply with a higher frequency compared to other media types: this can be explained by the fact that this media type helps in starting open directly dialogues more easily with municipalities and thus increasing the engagement rate. However, the frequency appears really low and thus, once again, suggests that Twitter users do not use the reply button as much as they do with the like one and the retweet one, probably due to its own feature of not being as “immediate” as likes and retweets.

Hence, municipalities pursuing similar studies might understand and might shape their communications strategies by looking at results such as those summarized above in table 8. Understanding which media types generate more engagement could be a really useful tool in order to take out the right information from citizens. For instance, if a municipality needs feedbacks or wants to understand people’s feelings about certain topics, issuing a plain text tweet could generate more replies, on average, than issuing tweets through other media types.
On the other hand, if a municipality wants to share a certain message for reaching the highest number on citizens, it should issue tweets including photo, video and web links. Indeed, as already seen, media types that fall into the categories “web links” and “photo/video” generate much more likes and, in particular, retweets which represent the main tools for drastically increasing the popularity and the virality of contents posted online.

We also analyzed whether different content categories generate different levels of citizens engagement.
Table 3.8 shows us that each tweet generates on average 14.21 likes, 1.71 replies and 12.41 retweets. In Italy citizens tend to engage with municipalities by pushing on the like button while scrolling Twitter, rather than using the reply button and the retweet one, even if the average number of retweets per tweet is considerably high. These results disagree with the ones provided by Bonsón, et al., (2019) on the Andalusian (Spain) region, in which the authors demonstrate that citizens tend to engage much more through the retweet button.
From our evidences, the *Sport* category is the content that generates more likes and retweets, in accordance with the results provided by Bonsón, et. al. (2019), *cultural* is the content that obtains more replies.
Through the table provided below (Tab. 3.8) we found out that different content types generate different levels of engagement among citizens, thus a relationship seems to exist between them. Municipalities, once again, should use these crucial information for understanding citizens’ needs and they should not miss out the opportunity to listen to citizens’ voices for increasing the community’s common wellness. In advance, implementing such a study, as the ones proposed by us and Bonsón, et. al (2019), could help municipalities in savings money avoiding them to take useless and unprofitable decisions.

Tab. 3. 8: Descriptive statistics of content types and citizens engagement.

Content Type		Likes	Replies	Retweets
Transport and Public Works	Mean	2.03	0.24	1.23
	Std.Deviation	10.89	0.61	2.07
Others	Mean	2.01	0.32	2.22
	Std.Deviation	5.79	0.54	6.98
Cultural	Mean	2.33	0.14	2.79
	Std.Deviation	6.65	1.78	6.72
Sport	Mean	3.78	0.21	2.87
	Std.Deviation	11.48	0.43	8.42
Security and Health	Mean	0.75	0.24	0.21
	Std.Deviation	5.63	0.56	4.89
Employment and Education	Mean	1.22	0.11	1.06
	Std.Deviation	6.78	0.66	4.98
Environment	Mean	2.09	0.51	2.03
	Std.Deviation	6.32	0.71	5.95
Total	**Mean**	**14.21**	**1.71**	**12.41**
	Std. Deviation	**53.54**	**5.29**	**40.01**
	N	**1,180,445**	**97,092**	**774,69**

Source: own elaboration.

3.6 Presentation of our results

Through our research, just presented in the third chapter, we wanted to give a new contribution to the existing literature regarding citizens engagement applied to Big Data leveraging the usage of Twitter.

To our best knowledge, there are no performed studies in Italy on citizens engagement usage through a Twitter analysis, so far. This study brings to light some evidences on the factors that influence citizens engagement on Twitter, thus these findings might be used to build a Twitter commitment model for the municipalities to explain the relationship among factors, the commitment that occurs and the behavioral effects on the users.

This is the first study exploring municipality communication and citizen's engagement via Twitter through an analysis of all the tweets coming from the largest Italian municipalities with active accounts. Moreover, this represents the first study, applied to Italian municipalities, that adopted the methodology and the metrics suggested by Bonsón et al. (2015) and it represents the first study who adopted the citizens engagement variable as a proxy for understanding how municipalities are moving towards smart cities, provided by Gartner, Inc. (2018).

The first finding that emerged from our study is that no significant relationships was found between municipality size and the Twitter activity: being a large municipality does not lead to a more active presence on Twitter. This result disagrees with the first publication published by Bonsón et. al, (2017) on Western European municipalities on Facebook, which outlined that bigger municipalities tend to pursue better online communication strategies and then being more active on social media compared to smaller municipalities. Whereas, the other study published by Bonsón et al., (2019), made on Andalusian municipalities Twitter accounts, provided results very similar to ours. Also, by applying the metrics proposed by Bonsón et al., (2015) (see Tab. 3.5 and Fig. 3.13) we were able to assess the citizens engagement per each municipality and then we calculated the average citizen engagement Italian index, which equals to 27,83%.

The municipalities with the higher levels of citizen engagement through our Twitter analysis resulted in Trento, Prato and Bergamo, whereas, unexpectedly, the largest Italian municipalities tended to present really low levels of citizen engagement. This finding agrees with previous research (Bonsón et al., 2019; Ma,2013) which showed similar results: in particular, Ma, (2013) argues that citizens living in smaller municipalities tend to express higher feelings of closeness and are encouraged to participate in discussions about common topics in more active ways.

In order to demonstrate this result, we calculated the relationship between Twitter activity (measured by the number of published tweets), the audience (measured by the number of followers) and the citizen engagement: a negative relationship was found (see Tab.3.6), thus supporting our findings.
Another finding that we discovered through our study is that among three different ways for interacting with municipalities on Twitter (likes, replies and retweets) citizens are more inclined to choose likes rather than retweets and replies, showing us different results compared to the ones provided by Bonsón, et al., (2019) in which the retweet option was the preferred.
Possible reason for this result could be the simplicity of clicking on the like button, which is easier and faster compared to the other options and it enables citizens to express their opinions without typing any comment.
Regarding the most used media types (see Fig.3.12 and Tab. 3.7), our findings show that website links (39,95%) were the most frequently used. This result might be explained by the Twitter's characters limitation which allows tweets up to 280 characters to be published (Bonsón, et al., 2019). Municipalities due to that tend to post tweets with an attached weblink in order to direct citizens to the web page which includes the targeted information (f.i. events, job offers, sports activities, etc.). We also identified which media types generate greater engagement (taking into consideration the number of like, replies and retweets). Photos/videos was the category that generated the highest rates of likes and retweets, then showing more citizen engagement. The results agree with previous studies: Bonsón, et al. (2019) and Zavattaro, et al., (2015).
Although, photos and videos transmit information in a quick and visual way that generates more interactions (Bonsón, et al., 2019), plain text generated the highest response rate levels (replies), which might help in starting a dialogue with the municipality.

We have also examined which areas of interactions generated the higher levels of engagement between citizens and municipalities.
First of all, the most common content type category resulted in Transport and Public Works (25,89%) (see Fig.3.11), followed by Others (24,63%), Cultural (19,47%), Sport (12,72%), Security and Health (7,04%), Employment and Education (5,56%) and Environment(4,65%). The result obtained by Transport and Public Works, might appear unexpected and weird, but knowing that more than half of the total tweets of our sample belong to the largest cities of Italy, and knowing the issues that Italians municipality face in managing public transports and public works (f.i. Roma), the results seem reasonable.

The content that generates the higher level of engagement falls into the Sport category, which generates the higher numbers of likes and retweets and thus it guarantees more popularity and virality for each sport issued. The Sport category is followed by Cultural an Environment categories. In advanced, the Environment category results to be the one with the highest number of replies per post, following a global trend that outlines how people nowadays care about such an issue. Moreover, as outlined by Hofmann, Beverungen, Räckers and Becker (2013) and Bonsón, et al., (2019) there is no topic that guarantees success in terms of citizen engagement, given the subjectivity of the issue.

As outlined by Gartner, Inc. (2018), citizen engagement is a critical factor to the success of smart cities: initiatives, indeed are no longer just about optimized traffic patterns, parking management, efficient lighting and improvements to public works. Thus, being able to assess the citizen engagement is considered as a new and crucial starting point to understand whether a city is effectively approaching a smart city's pattern. Through the methodology and metrics developed by Bonsón, et al., (2015), we were able to assess in a quantitative way an index of the citizen engagement for each municipality of our sample and then an average index of the Italian citizen engagement (27,83%), as already stressed.

Given the very broad concept of smart cities, in which an enormous number of variables should be taken in consideration (technologies, transports, electronic payments and etc), we are not assuming that our citizen engagement index represents an index of the development of the Italian smart cities, rather we considered it as one of the variables to take into account when discussing about smart cities.

Nevertheless, the low value of the Italian citizen engagement index, which is 27,83%, permits us to affirm that the process towards smart cities of the Italian municipalities, based on the citizen engagement, is proceeding slowly.

Conclusions, limitations and future researches

Our research has attempted to deepen understand Big Data technologies world, and its applications to smart cities through the usage of social media and how it effects the relationship between local municipalities and citizens. Providing an overview on how governments and local municipalities are implementing such technologies for enabling citizens to actively participate in day- life activities and issues. We introduced the newly concept of E-government, which embodies a still hybrid and abstract topic for most of our readers. E-government, through Internet of Things technologies, represents a unique opportunity for helping citizens to improve societies' engagement with governments. We then tried to explain how powerful Big Data and the data mining process can be in improving and shaping citizens engagement in Italian municipalities. Referring to a sample of 28 Italian municipalities.

To perform our studies, a content analysis was executed on 316,801 tweets coming from the official Twitter accounts of the Italian municipalities contained in our sample. To the best of our knowledge this is the first study exploring municipality communication and citizen's engagement via Twitter.

A content analysis has allowed us to move beyond classical statistical analyses on social media, understanding how citizens react to certain area of interactions and how municipalities and local governments, extracting those data can be able to shape their decisions leveraging citizens' needs. The inspiration of our analysis went from the publications made by Bonsón, et al., (2015) and Bonsón, et al., (2019), which performed studies on citizens engagement on Western European municipalities (through Facebook) and Andalusian municipalities (through Twitter) that represents the milestone of our research.

To the best of our knowledge, no such studies have been performed so far on Italian municipalities using the methodology suggested by Bonsón, et al., (2015) and Bonsón, et al., (2019). This is the first study exploring Italian municipalities' communication and citizens' engagement via Twitter. Therefore, our research has provided a theoretical and practical contribution to the Big Data and citizen engagement fields of studies offering an interesting overview of the Italian perimeter.

The results show us an heterogenous situation of the local governments in Italian municipalities, regarding audience (total followers) and activity (total tweets). Regarding citizen engagement, likes are the most frequent way for Italian citizens to interact with their local city council account, on the other hand the percentage on replies per tweet appear really low, outlining how nowadays citizens tend not to spend much time in giving feedbacks to municipalities on different contents and topics.
The Italian citizens engagement index, found through the methodology introduced by Bonsón, et al., (2015), was 27,83%.
Citizen engagement does not seem to be related to the number of inhabitants of municipality, yet there are negative relationships regarding audience (total followers) and activity (total tweets). Our findings show that the media type category that generates more engagement than other media types (likes and retweets) is Photos/Videos. Regarding content analysis, Transport and Public Works and Cultural result the content categories most frequently tweeted, whereas the content type category with the highest levels of engagement is related to Sports.

A series of limitations together with recommendations for future research have to be acknowledged before concluding. The first limitation regards the size of the sample: 28 Twitter accounts cannot generalize the results, extending it to a higher sample, perhaps including all the Italian provinces. Moreover, pursuing such a study also on Facebook, given that it represents the most used social network in Italy, would add new interesting insights about the topic.
Furthermore, the usage of an automated content analysis, classifying the tweets through a dictionary of the most frequent words, might reduce subjectivity, but the classifications are biased to an error if there are tweets that share more than one category. A suggest for future research might be to design and develop more accurate techniques in order to reduce this issue.
A limitation comes out while analyzing media types, since both photos and videos were assigned to the same category, thus developing an automated way for distinguishing these two categories might be really useful.
One of the most critical limitation of our analysis is related to the way in which citizen engagement was assessed, using likes, replies and retweets: they were considered as proxies for citizen's engagement, but they might be used differently.
An interesting suggestion that we would like to stress for future research regards the usage of hashtags and its effects on citizens commitment, which represent an interesting insight on the topic. In the end, an interesting future study might investigate on the reason leading Twitter

users in interacting with local governments, assessing and developing a commitment model based on social media.

References

Al Nuaimi, E., Al Neyadi, H., Mohamed, N., & Al-Jaroodi, J. (2015). Applications of big data to smart cities. *Journal of Internet Services and Applications*, *6*(1), 25.

Anderson, M., & Jiang, J. (2018). Teens, social media & technolog

Ashton, K. (2009). That 'internet of things' thing. RFID journal, 22(7), 97-114.

Atzori, L., Iera, A., & Morabito, G. (2010). The internet of things: A survey. Computer networks, 54(15), 2787-2805.

Baum, S., Ma, J., & Payea, K. (2013). Education pays 2013. *The College Board*.

Bertot, J. C., & Choi, H. (2013, June). Big data and e-government: issues, policies, and recommendations. In *Proceedings of the 14th Annual International Conference on Digital Government Research* (pp. 1-10). ACM.

Boer, H., & During, W. E. (2001). Innovation, what innovation? A comparison between product, process and organizational innovation. *International Journal of Technology Management*, *22*(1-3), 83-107.

Bonsón, E., Perea, D., & Bednárová, M. (2019). Twitter as a tool for citizen engagement: An empirical study of the Andalusian municipalities. *Government Information Quarterly*.

Bonsón, E., Royo, S., & Ratkai, M. (2015). Citizens' engagement on local governments' Facebook sites. An empirical analysis: The impact of different media and content types in Western Europe. *Government Information Quarterly*, *32*(1), 52-62.

Bonsón, E., Torres, L., Royo, S., & Flores, F. (2012). Local e-government 2.0: Social media and corporate transparency in municipalities. *Government information quarterly*, *29*(2), 123-132.

Carpini, M. X. D., Cook, F. L., & Jacobs, L. R. (2004). Public deliberation, discursive participation, and citizen engagement: A review of the empirical literature. Annu. Rev. Polit. Sci., 7, 315-344.

Chesbrough, H. W. (2003). *Open innovation: The new imperative for creating and profiting from technology*. Harvard Business Press.

Chesbrough, H., Vanhaverbeke, W., & West, J. (Eds.). (2006). *Open innovation: Researching a new paradigm*. Oxford University Press on Demand.

Chourabi, H., Nam, T., Walker, S., Gil-Garcia, J. R., Mellouli, S., Nahon, K., ... & Scholl, H. J. (2012, January). Understanding smart cities: An integrative framework. In *2012 45th Hawaii international conference on system sciences* (pp. 2289-2297). IEEE.

Comuni-Chiamo è la soluzione in cloud per gestire al meglio la città: dall'APP per informare i tuoi cittadini al software per gestire le segnalazioni, abbiamo tutto quello che serve per innovare il tuo Comune, 2018. [online]. Available at: https://comuni-chiamo.com/ [Access: November 2019]

Cukier, K. (2010). Data, data everywhere: A special report on managing information. Economist Newspaper.

Damanpour, F., & Schneider, M. (2006). Phases of the adoption of innovation in organizations: effects of environment, organization and top managers 1. *British journal of Management*, *17*(3), 215-236.

De Lange, M., & De Waal, M. (2017). Owning the city: New media and citizen engagement in urban design. In Urban land use (pp. 109-130). Apple Academic Press.

Ellison, N., & Hardey, M. (2014). Social media and local government: Citizenship, consumption and democracy. *Local Government Studies*, *40*(1), 21-40.

Feinerer, I. (2013). Introduction to the tm Package Text Mining in R. *Accessible en ligne: http://cran. r-project. org/web/packages/tm/vignettes/tm. pdf*.

Flanagin, A. J., & Metzger, M. J. (2001). Internet use in the contemporary media environment. *Human communication research*, *27*(1), 153-181.

Gandomi, A., & Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. International journal of information management, 35(2), 137-144.

Gaventa, J., & Barrett, G. (2010). So what difference does it make? Mapping the outcomes of citizen engagement. IDS Working Papers, 2010(347), 01-72.

Grover, P., Kar, A. K., & Davies, G. (2018). "Technology enabled Health"–Insights from twitter analytics with a socio-technical perspective. International Journal of Information Management, 43, 85-97.

Guillamón, M. D., Ríos, A. M., Gesuele, B., & Metallo, C. (2016). Factors influencing social media use in local governments: The case of Italy and Spain. *Government Information Quarterly*, *33*(3), 460-471.

Gurin, J. (2014). Open governments, open data: A new lever for transparency, citizen engagement, and economic growth. SAIS Review of International Affairs, 34(1), 71-82.

Hadoop, 2017. *The Apache™ Hadoop® project develops open-source software for reliable, scalable, distributed computing. [online]. Available at:* http://hadoop.apache.org/ [Access: September 2019]

Harrison, T. M., Guerrero, S., Burke, G. B., Cook, M., Cresswell, A., Helbig, N., ... & Pardo, T. (2012). Open government and e-government: Democratic challenges from a public value perspective. *Information Polity*, *17*(2), 83-97.

Hendriks, C. M. (2012). The politics of public deliberation: Citizen engagement and interest advocacy. Springer.

Hernández-Muñoz, J. M., Vercher, J. B., Muñoz, L., Galache, J. A., Presser, M., Gómez, L. A. H., & Pettersson, J. (2011, May). Smart cities at the forefront of the future internet. In The future internet assembly (pp. 447-462). Springer, Berlin, Heidelberg.

Hofmann, S., Beverungen, D., Räckers, M., & Becker, J. (2013). What makes local governments' online communications successful? Insights from a multi-method analysis of Facebook. *Government information quarterly*, *30*(4), 387-396.

Honey, C., & Herring, S. C. (2009, January). Beyond microblogging: Conversation and collaboration via Twitter. In *2009 42nd Hawaii International Conference on System Sciences* (pp. 1-10). Ieee.

INSTAT, 2011. *15° CENSIMENTO DELLA POPOLAZIONE E DELLE ABITAZIONI 2011. [online] Available at:* https://www.istat.it/it/censimenti-permanenti/censimenti-precedenti/popolazione-e-abitazioni/popolazione-2011. [Access: October 2019]

Janssen, M., Charalabidis, Y., & Zuiderwijk, A. (2012). Benefits, adoption barriers and myths of open data and open government. *Information systems management*, *29*(4), 258-268.

Jin, J., Gubbi, J., Marusic, S., & Palaniswami, M. (2014). An information framework for creating a smart city through internet of things. *IEEE Internet of Things journal*, *1*(2), 112-121.

Johannessen, M. R., Flak, L. S., & Sæbø, Ø. (2012, September). Choosing the right medium for municipal eParticipation based on stakeholder expectations. In *International Conference on Electronic Participation* (pp. 25-36). Springer, Berlin, Heidelberg.ISO 690

Kaisler, S., Armour, F., Espinosa, J. A., & Money, W. (2013, January). Big data: Issues and challenges moving forward. In *2013 46th Hawaii International Conference on System Sciences* (pp. 995-1004). IEEE.

Kaisler, S., Armour, F., Espinosa, J. A., & Money, W. (2013, January). Big data: Issues and challenges moving forward. In 2013 46th Hawaii International Conference on System Sciences pp.995-1004). IEEE.

Khan M, Uddin MF, Gupta N. Seven V's of Big Data understanding Big Data to extract value. In American Society for Engineering Education (ASEE Zone 1), 2014 Zone 1 Conference of the IEEE; 2014. pp. 1–5.

Khan M, Uddin MF, Gupta N. Seven V's of Big Data understanding Big Data to extract value. In American Society for Engineering Education (ASEE Zone 1),2014 Zone 1 Conference of the IEEE; 2014. pp. 1–5.

Khan, Z., Anjum, A., & Kiani, S. L. (2013, December). Cloud based big data analytics for smart future cities. In *2013 IEEE/ACM 6th International Conference on Utility and Cloud Computing* (pp. 381-386). IEEE.

Kim, D., Chun, H., Kwak, Y., & Nam, Y. (2014). The employment of dialogic principles in website, Facebook, and Twitter platforms of environmental nonprofit organizations. *Social science computer review*, *32*(5), 590-605.

Kitchin, R. (2014). The real-time city? Big data and smart urbanism. *GeoJournal*, *79*(1), 1-14.

Kohlbacher, F. (2006). The use of qualitative content analysis in case study research. In *Forum Qualitative Sozialforschung/Forum: Qualitative Social Research* (Vol. 7, No. 1, pp. 1-30). Institut für Qualitative Forschung.

Kourtit, K., & Nijkamp, P. (2012). Smart cities in the innovation age. *Innovation: The European Journal of Social Science Research*, *25*(2), 93-95.

Letaifa, S. B. (2015). How to strategize smart cities: Revealing the SMART model. *Journal of Business Research*, *68*(7), 1414-1419.

Linders, D. (2012). From e-government to we-government: Defining a typology for citizen coproduction in the age of social media. *Government Information Quarterly*, *29*(4), 446-454.

Macintosh, A. (2004). Using information and communication technologies to enhance citizen engagement in the policy process.

Mambrey, P., & Dörr, R. (2011, May). Local government and social networking technologies in Germany: The example of Twitter. In *Conference for E-Democracy and Open Government* (p. 249).

Matos, N. A., Carballo, N. I. M., Miyares, V. M., de la Cruz Mejías, Y., & Pimienta, A. V. (2013). Indicadores del bajo peso al nacer en el área de salud del Policlínico Docente "José Martí Pérez" en 2012. *Medisan*, *17*(8), 3029-3035.

Mellouli, S., Luna-Reyes, L. F., & Zhang, J. (2014). Smart government, citizen participation and open data. *Information Polity*, *19*(1, 2), 1-4.

Michalik P, Stofa J, Zolotova I. Concept definition for Big Data architecture in the education system. In Applied Machine Intelligence and Informatics (SAMI), 2014 IEEE 12th International Symposium on 2014. pp. 331–334.

Miorandi, D., Sicari, S., De Pellegrini, F., & Chlamtac, I. (2012). Internet of things: Vision, applications and research challenges. Ad hoc networks, 10(7), 1497-1516.

Neirotti, P., De Marco, A., Cagliano, A. C., Mangano, G., & Scorrano, F. (2014). Current trends in Smart City initiatives: Some stylised facts. *Cities*, *38*, 25-36.

Ojo, A., Curry, E., & Zeleti, F. A. (2015, January). A tale of open data innovations in five smart cities. In *2015 48th Hawaii International Conference on System Sciences* (pp. 2326-2335). IEEE.

Orlando, Puelli & Vicentini, 2019. *FPA: La sfida per il futuro? Potenziare le competenze per interagire con i cittadini*. [online] Avilable at: https://forumpacitta2019.eventifpa.it/event-details/?id=8972 [Access: January 2020]

Orlikowski, W. J. (2000). Using technology and constituting structures: A practice lens for studying technology in organizations. *Organization science*, *11*(4), 404-428.

Paskaleva, K. A. (2011). The smart city: A nexus for open innovation?. *Intelligent Buildings International*, *3*(3), 153-171.

Pina, V., & Torres, L. (2001). Analysis of the efficiency of local government services delivery. An application to urban public transport. *Transportation Research Part A: Policy and Practice*, *35*(10), 929-944.

Reddel, T., & Woolcock, G. (2004). From consultation to participatory governance? A critical review of citizen engagement strategies in Queensland. Australian Journal of Public Administration, 63(3), 75-87.

Report, view, or discuss local problems, 2017. [Online] Available at: https://www.fixmystreet.com/ [Access: November 2019]

Roski, J., Bo-Linn, G. W., & Andrews, T. A. (2014). Creating value in health care through big data: opportunities and policy implications. *Health affairs*, *33*(7), 1115-1122.

Schaffers, H., Komninos, N., Pallot, M., Trousse, B., Nilsson, M., & Oliveira, A. (2011, May). Smart cities and the future internet: Towards cooperation frameworks for open innovation. In *The future internet assembly* (pp. 431-446). Springer, Berlin, Heidelberg.

Schwartz, H. A., & Ungar, L. H. (2015). Data-driven content analysis of social media: a systematic overview of automated methods. *The ANNALS of the American Academy of Political and Social Science*, *659*(1), 78-94.

SeeClickFix has helped millions of citizens, 2017. Strong Communities, More Efficient Government [online]. Available at: https://seeclickfix.com/. [Access: Novemeber 2019]

Shetty, S., 2018. *Citizen Engagement Is Critical to the Success of Smart Cities [online]. Available at:* https://www.gartner.com/en/newsroom/press-releases/2018-03-07-gartner-says-citizen-engagement-is-critical-to-the-success-of-smart-cities [Access: November 2019]

Sobaci, M. Z., & Karkin, N. (2013). The use of twitter by mayors in Turkey: Tweets for better public services?. *Government Information Quarterly*, *30*(4), 417-425.

Statista, 2019. *Social network market share held by Twitter in Italy from January 2017 to December 2019.* [online] Available at: https://www.statista.com/ [Access: December 2019]

Su, K., Li, J., & Fu, H. (2011, September). Smart city and the applications. In *2011 international conference on electronics, communications and control (ICECC)* (pp. 1028-1031). IEEE.

Susmitha, K., & Jayaprada, S. (2017). Smart Cities using big data Analytics. *Int. Res. J. Eng. Technol.(IRJET)*, 1617.

Tuomi, I. (2002). *Networks of Innovation: Change and Meaning in the Age of the Internet* (Vol. 249). OUP Oxford.

Van de Werfhorst, H. G., & Hofstede, S. (2007). Cultural capital or relative risk aversion? Two mechanisms for educational inequality compared 1. *The British journal of sociology*, *58*(3), 391-415.

Vashisht, P., & Gupta, V. (2015, October). Big data analytics techniques: A survey. In 2015 International Conference on Green Computing and Internet of Things (ICGCIoT) (pp. 264-269). IEEE.

Villars, R. L., Olofson, C. W., & Eastwood, M. (2011). Big data: What it is and why you should care. *White Paper, IDC*, *14,* 1-14.

Xenos, M. A., Macafee, T., & Pole, A. (2017). Understanding variations in user response to social media campaigns: A study of Facebook posts in the 2010 US elections. *new media & society*, *19*(6), 826-842.

Xia, F., Yang, L. T., Wang, L., & Vinel, A. (2012). Internet of things. *International Journal of Communication Systems*, *25*(9), 1101.

Zanella, A., Bui, N., Castellani, A., Vangelista, L., & Zorzi, M. (2014). Internet of things for smart cities. *IEEE Internet of Things journal*, *1*(1), 22-32.

Zavattaro, S. M., French, P. E., & Mohanty, S. D. (2015). A sentiment analysis of US local government tweets: The connection between tone and citizen involvement. *Government information quarterly*, *32*(3), 333-341.

Appendix A

CITY	*OFFICIAL NAME*	*WEBSITE*
Arezzo	Comune di Arezzo	http://www.comune.arezzo.it/
Bari	Comune di Bari	http://www.comune.bari.it/
Bergamo	Comune di Bergamo	https://www.comune.bergamo.it/
Bologna	Comune di Bologna	http://www.comune.bologna.it/
Brescia	Comune di Brescia	http://www.comune.brescia.it/Pagine/default.aspx
Cagliari	Comune di Cagliari	https://www.comune.cagliari.it/portale/
Ferrara	Comune di Ferrara	https://www.comune.fe.it/
Firenze	Comune di Firenze	https://www.comune.fi.it/
Genova	Comune di Genova	https://smart.comune.genova.it/
Livorno	Comune di Livorno	http://www.comune.livorno.it/
Milano	Comune di Milano	http://www.comune.livorno.it/
Modena	Comune di Modena	https://www.comune.modena.it/
Napoli	Comune di Napoli	http://www.comune.napoli.it/home
Padova	Comune di Padova	http://www.padovanet.it/
Palermo	Comune di Palermo	https://www.comune.palermo.it/
Parma	Comune di Parma	https://www.comune.parma.it/homepage.aspx
Perugia	Comune di Perugia	https://www.comune.perugia.it/
Prato	Comune di Prato	https://www.comune.prato.it/
Ravenna	Comune di Ravenna	http://www.comune.ra.it/
Reggio Calabria	Comune di Reggio Calabria	https://turismo.reggiocal.it/
Reggio Emilia	Comune di Reggio Emilia	https://www.comune.re.it/
Roma	Comune di Roma	https://www.comune.roma.it/web/it/home.page
Torino	Comune di Torino	http://www.comune.torino.it/
Trento	Comune di Trento	https://www.comune.trento.it/
Trieste	Comuned di Trieste	http://www.comune.trieste.it/social-media-policy
Udine	Comune di Udine	https://www.comune.udine.it/
Venezia	Comune di Venezia	https://live.comune.venezia.it/it
Vicenza	Comune di Vicenza	https://www.comune.vicenza.it/

Appendix B

Python's scripts for processing Twitter scraping process

```
#!/usr/bin/python3

import subprocess

f = open("comuni.txt").readlines()

for comune in f:
        out = "%s.csv" % comune.strip()
        comune = comune.strip()
        subprocess.call(["twint", "-u", comune, "-o", out, "--csv"])
```

```
The default interactive shell is now zsh.
To update your account to use zsh, please run `chsh -s /bin/zsh`.
For more details, please visit https://support.apple.com/kb/HT208050.
MacBook-Pro-di-Dodo:~ edoardogobbo$ ls
Applications            Dropbox                 Pictures
Creative Cloud Files    Dropbox (Precedente)    Public
Desktop                 Library                 PycharmProjects
Documents               Movies                  Untitled.R
Downloads               Music                   iCloud Drive (Archivio)
MacBook-Pro-di-Dodo:~ edoardogobbo$ cd Desktop/E
MacBook-Pro-di-Dodo:E edoardogobbo$ ls
comuni.txt      script.py
MacBook-Pro-di-Dodo:E edoardogobbo$ python
comuni.txt  script.py
MacBook-Pro-di-Dodo:E edoardogobbo$ python
comuni.txt  script.py
MacBook-Pro-di-Dodo:E edoardogobbo$ python script.py
```

Appendix C

Document Term Matrix

```
<<DocumentTermMatrix (documents: 316801, terms: 248987)>>
Non-/sparse entries: 2989546/78876341041
Sparsity           : 100%
Maximal term length: 79
Weighting          : term frequency - inverse document frequency (tf-idf)
```

Appendix D

Correlation between citizen engagement and audience

```
        Spearman's rank correlation rho

data:  relationship$Engagement and relationship$Audience
S = 6282, p-value = 2.701e-05
alternative hypothesis: true rho is not equal to 0
sample estimates:
       rho
-0.7192118
```

Correlation between citizen engagement and population

```
        Spearman's rank correlation rho

data:  relationship$Engagement and relationship$Activity
S = 6048, p-value = 0.0002183
alternative hypothesis: true rho is not equal to 0
sample estimates:
       rho
-0.6551724
```

Correlation between citizen engagement and activity

```
        Spearman's rank correlation rho

data:  relationship$Engagement and relationship$Activity
S = 6048, p-value = 0.0002183
alternative hypothesis: true rho is not equal to 0
sample estimates:
       rho
-0.6551724
```