UNIVERSITÀ
POLITECNICA
DELLE MARCHE

Facoltà di Ingegneria
Corso di Laurea in Ingegneria Informatica e dell'Atuomazione

A Decision Support System for social action inspection on social networks multimedia data streams

Sistema di supporto alle decisioni per l'ispezione ed analisi dei flussi di dati multimediali provenienti dai social network

Candidate:
Simone ONORI

Advisor:
Prof. Emanuele FRONTONI

Coadvisor:
Dott.ssa Marina PAOLANTI

Academic Year 2019-2020

UNIVERSITÀ POLITECNICA DELLE MARCHE

Facoltà di Ingegneria
Corso di Laurea in Ingegneria Informatica e dell'Atuomazione

A Decision Support System for social action inspection on social networks multimedia data streams

Sistema di supporto alle decisioni per l'ispezione ed analisi dei flussi di dati multimediali provenienti dai social network

Candidate:
Simone ONORI

Advisor:
Prof. Emanuele FRONTONI

Coadvisor:
Dott.ssa Marina PAOLANTI

Academic Year 2019-2020

Università Politecnica delle Marche
Facoltà di Ingegneria
Corso di Laurea in Ingegneria Informatica e dell'Atuomazione
Via Brecce Bianche – 60131 Ancona (AN), Italy

Ai miei genitori Mauro e Mirella
che mi hanno sempre sostenuto e appoggiato
a mio fratello Davide
alla mia famiglia
a tutti quelli che mi hanno accompagnato in questo mio percorso

Sommario

Una rete sociale, o più semplicemente social network, consiste in un qualsiasi gruppo di individui connessi tra loro da diversi legami sociali che vanno dalla conoscenza casuale, ai rapporti di lavoro, ai vincoli familiari.
Nello specifico, al giorno d'oggi, i social network sono ormai parte integrante della sfera sociale della maggior parte degli individui e costituiscono un aspetto fondamentale della vita di tutti i giorni. Dai propri hobby, alla musica, passando per i film preferiti, i profili social vengono riempiti di informazioni personali al fine di condividerle con la propria cerchia di amici, follower o utenti. Estrapolando queste ed altre informazioni, i social stessi sono in grado di proporci nuove amicizie con le quali abbiamo una maggiore affinità, proporci delle notizie alle quali potremmo essere interessati o, semplicemente, mostrarci le pubblicità che più ci interessano. Un sistema di supporto che automatizzi il processo di raccolta dei dati e faciliti il processo di interpretazione degli stessi risulta, perciò, di elevato interesse.
A questo scopo esistono già numerose tecnologie sia per l'attività di data mining che per l'analisi dei social network. Lo scopo di questa tesi è, quindi, quello di analizzare i dati provenienti dai tre social più diffusi in Italia (Facebook, Instagram e Twitter) e di sviluppare un'applicazione in grado di analizzare in maniera efficace ed efficiente il flusso di dati multimediali ottenuti e di interpretarli attraverso l'utilizzo di metriche specifiche.
In questo elaborato viene presentata l'applicazione web SocMint2020 per l'analisi dei dati multimediali provenienti dai social.
In particolare, è stato scelto ed analizzato come caso di studio un evento politico: le elezioni del nuovo governatore della regione Marche avvenute a Settembre 2020. L'approccio proposto è stato testato su un dataset dinamico contenente un ingente numero di post, tweet ed elementi multimediali. I dati, estratti tramite l'utilizzo delle API di Facebook, Twitter e Instagram, coprono un periodo di 3 mesi che va dal primo Luglio del 2020 fino a fine Settembre 2020 in modo da poter includere nell'analisi anche le due settimane successive alle elezioni.
Per migliorare ulteriormente la tecnologia presentata, in futuro, saranno valutati diversi approcci e un diverso set di metriche necessarie all'analisi in maniera da consentire una stima e un'interpretazione più approfondita della situazione sociale in esame.
In conclusione, questa tesi dimostra le elevate potenzialità dell'utilizzo di un'efficiente analisi dei social media nell'interpretazione del contesto politico e sociale di tutti i giorni.

Abstract

A social network consists of any group of individuals connected to each other by different social ties ranging from casual acquaintance, to work relationships, to family ties.
Specifically, nowadays, social networks are now an integral part of the social sphere of most individuals and constitute a fundamental aspect of the life of all days. From your hobbies, to music, passing through your favourite movies, social profiles come filled with personal information in order to share it with your circle of friends, followers or users. By extrapolating this and other information, the social networks themselves are able to offer us new ones friendships with which we have a greater affinity, propose news to which we could be interested or, simply, show us the advertisements that interest us most. A support system that automates the data collection process and facilitates the process of data interpretation is therefore of high interest.
For this purpose, there are already numerous technologies for both data mining and analysis of social networks. The purpose of this thesis is, therefore, to analyze the data from the three most popular social networks in Italy and to develop an application capable of analysing in a manner effective and efficient flow of multimedia data obtained and to interpret them through the use of specific metrics.
This report presents the SocMint2020 web application and the metrics used in it for the analysis of multimedia data from social networks in the context of the regional elections of Marche.
The proposed approach was tested on a dynamic data-set containing a large number of posts, tweets and multimedia elements. The data, extracted through the use of the Facebook, Twitter and Instagram APIs, cover a period of 3 months from 1 July 2020 until the end of September 2020 in order to include in the analysis also the two weeks following the elections.
To further improve the technology presented, in the future, different approaches and a different set of metrics necessary for the analysis will be evaluated in order to allow for a more in-depth estimate and interpretation of the social situation in question.
In conclusion, this thesis demonstrates the high potential of using an efficient social media analysis in the interpretation of the everyday political and social context.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

In recent years, the use of social networks has been a constant in the daily life of all of us. Now both young people and adults consider it a necessary tool to be able to communicate with their acquaintances, make new ones, learn about job offers and much more. Social networks are a dynamic and constantly expanding environment, enriching themselves with new features in a short time, thus influencing the choice of users in reference to the social network to use.
Thanks to the great diffusion and the role they assume in contemporary society, nowadays, social media have attracted the interest of researchers, journalists, companies, movements and governments. The growth in the ease of access to information that the network makes available and the opportunity to communicate with a large number of users at virtually no cost have given way to interpretations that place social media at the center of a new digital revolution.
The most interesting aspect of this change is not only linked to the possibility of promoting political participation and activism. The real `social revolution` is the one that invests the lives of every single individual through the freedom to express oneself, to have one's own space in which to be yourself, or what one would like to be, with very few limits and almost non-existent barriers. The real revolution is to be able to tell one's emotions, to present one's opinions, especially to the circle of acquaintances, through an interaction that generates mutual openings on their respective worlds and browsing, in a more or less indiscreet way, in the lives of others. All this happens, paradoxically, while we live in a society in which it is increasingly difficult to know the names of one's neighbours, and in which the right to privacy is exploit by anyone, only to then communicate to the whole world and strictly `online`, anything: new loves, holidays, travels, daily troubles. This is because on social media, or more precisely, on those social media that are also social networks, you get to tell the whole, or almost, of your life: from anger for a failed exam, to happiness for a new job, to amazement for a news, up to the choice of vote made in the secret of the voting booth.
It is not surprising that we have begun to discuss how to best use this information since the data present on the network, if properly collected and analyzed, allow not only to understand and explain many complex social phenomena, but even to predict them. The prediction, both that made in real time and that relating to future events,

is in fact one of the most interesting frontiers in the `social world`.
In general, any social network we use this inevitably impacts our lives. The positive aspects of social networks are numerous: they allow each user to express their opinion, connect people all over the world, give the opportunity to meet people who share our interests and who support causes common to us.
Social networks today are therefore the protagonists of a social change that is constantly growing and that has led users to have completely new behaviours, ways of thinking and acting.

To understand what we mean by `social media`, to differentiate them internally and, consequently, to identify the particular sub-class of social media on which we will focus in this and the next chapters, it is useful to start our research by introducing a much more general concept: that of "*social networks*".
The general idea of construction of society is the basis for the social network definitions. Every researcher defines social network in different forms, stated as follows:

In 2014, Rabade, Mishra and Sharma defined it as "*a society is not just a simple collection of individuals; it is rather the sum of the relationships that connect these individuals to one another. So the social network can be defined as the finite set of nodes (actors) and connecting edges (relationships)*" [1].
In 1994, Wasserman and Faust [2] proposed a very sociological approach, which defines:

- Actor: An actor is a discrete individual, corporate or collective social unit.
- Relation: A set of ties of a specific type; a tie is a linkage between a pair of actor.
- Social network: The finite set or sets of actors and one or more relations defined on them.

In 2006, Yang, Dia, Cheng, and Lin [3] proposed a very formal way, which defines as:

- Actor: A node in a graph; each node represents a customer.
- Relation: The undirected, unweighted edges in the graph; each edge represents the connectedness between two nodes.
- Social network: An unweighted and undirected graph.

By social network we refer to any structure, formal or informal, comprising a set of people or organizations, together with their respective relationships [4].

1.1 Context

In this paragraph we will discuss the main types of social networks, which can be classified into two main groups:

- topology-based social networks;
- function-based social networks.

Based on the network topology, we can further divide social networks into online social networks (*ONS*) and mobile social networks (*MSN*). Online social networks are virtual communities in which people are connected and interact with each other based on a specific topic or simply by communicating online.
With the development of the World Wide Web (*WWW*) in the nineties and thanks to the development of new information technologies, `ONS` have experienced exponential growth among users around the world. This rapid growth has attracted the attention of a large number of researchers to analyze and explore this popular service on a large scale.
Subsequently with the significant development of communications and wireless networks, web-based social networks began to migrate from online networks to mobile networks.
There are three main topologies of `MSNs`.

The first is represented by the `centralized` Mobile Social Networks, which are a mobile extension of the Online Social Networks. Using a specific mobile application or mobile browser, users can connect to a centralized server to exchange and share information.

The second is made up of `distributed` Mobile Social Networks, thanks to which users can communicate and access information through an *ad hoc* method without the need for other infrastructures.

Finally, there are the `hybrid` Mobile Social Networks, which are simply a combination of the two previous topologies. Users have access to and share information using either the cellular infrastructure or through an *ad hoc* mode.

Thanks to countless advantages, such as simplicity and low distribution costs or, for example, portability, social mobile applications have attracted a lot of attention and studies from important research bodies.

Depending on their functionality, social networks can be divided into numerous categories. In this section, we will briefly describe the most common and best-known types nowadays.
Let's start with social news, which allow users, through dedicated websites, to send and vote new stories, articles or simple links, which will be then classified according to their popularity so as to determine what will be published based on its evidence and its relevance. In some cases, users can comment on these posts and be in turn classified, such as happens on `Reddit`.
Another category is the one that include the social networks of Instant Messaging, an online chat platform that provides its users with real-time communication of text

messages through the Internet. The main differences with respect to e-mail or other types of chat reside in the brevity of the messages, in the speed of their delivery and in the fact that the exchange of messages takes place in synchronous mode.
In instant messaging systems, in fact, sending a message is only possible when the recipient is also online and usually communications are not automatically stored by applications. A LAN-based messaging app operates in a very similar way but on a local area network, such as `Tencent QQ`.
A completely different category is that of microblogging, which allow users to exchange small contents on the net in the form of short messages, single images, short videos or links. This contents are posted on a social networking service, visible to everyone or only to people in your community. An example of this type is `Twitter`, a microblogging platform that allows registered users to publish small texts, images, links and short videos.
Similar, but not quite the same are forums, an online discussion site where people can have a conversation in the form of posted messages. Many forums require user registration before you can post messages and in some cases even to read them.
They differ from chats inasmuch the messages are usually longer than a single line of text, and are archived for a long term, furthermore while chats are a synchronous communication tool, that is in which communication occurs at the same time, the forums are asynchronous, as writing and replying can happen at different times.
The Media sharing networks, unlike the social networks described so far, allow users to upload, view, react, comment or simply forward specific media, be they video, images or audio. They are therefore characterized by the almost total absence of written posts since the type of information exchanged is based precisely on multimedia files. `YouTube`, for example, is a video sharing website and `Instagram` one of the most used image sharing networks.
Finally, we have social networks, also known as social networking services, or more simply social media, Internet services, typically usable through browsers or mobile applications, which, relying on the relative platform, allow the management of social relationships, communication and sharing of ideas, posts, images, activities, events and interests with people on your network, such as `Facebook`.

Facebook, Twitter and Instagram are currently among the most popular social media (Figure 1.1), and for this reason they are the three social networks on which our analysis was performed..

Facebook is the "oldest" of the three, having been founded way back in 2004. In its early years, Facebook was mainly limited to American college students until October 2006, when it was finally made accessible to all users of Internet. From that moment on, Facebook has recorded constant growth reaching, already in August 2008, the figure of 100 million users to reach, in January 2020, over 2 billion active users monthly (Figure 1.1).

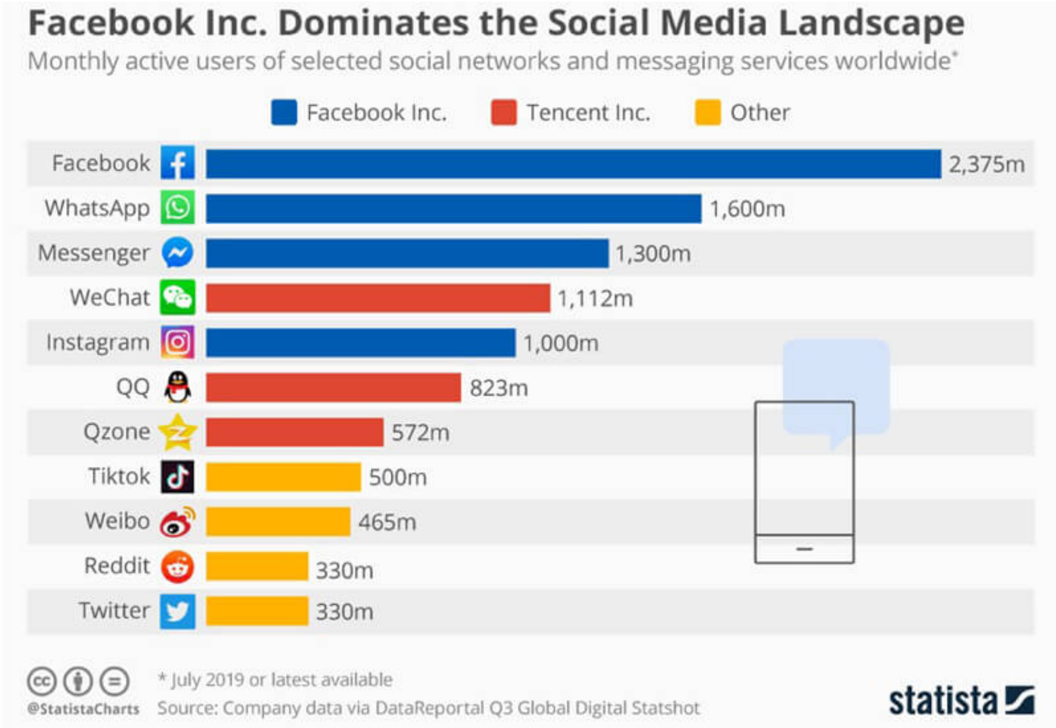


Figure 1.1: Most popular social media.

With regard to permitted social interactions, a Facebook user must become a "*friend*" of another user, being accepted as such by the latter in order to be able to access the information published by this second user and to interact with it. The only exceptions are the profiles or functions that users decide to make "public". In the latter case, the published information becomes freely accessible to all Facebook users as well as to those who do not have a social media account, making any type of interaction with this "public" profile available.
Finally, a Facebook user can express his liking or interest, in the activities of other users or in corporate or institutional brands or profiles, using the simple "*like*" feature or through the use of a predefined set of emoticons.

Twitter was born two years later, in 2006. Unlike Facebook, on Twitter each user can only share short text messages, up to a maximum of 140 characters, called "*tweets*", shown on the user's profile page and which are forwarded automatically also to all those who have registered to receive them, that is to the "*followers*" of this user.
Given the constant production of ideas and textual content, which may also include images, links, and short videos, Twitter is considered as a social network that generates microblogging. On Twitter, unlike Facebook, a user can "*follow*" another without their authorization, unless the latter has decided to make their account private. Users, if they deem it appropriate, have the possibility to

"*block*" some of their followers, preventing them from continuing to follow them. Furthermore, a user can freely address a public message to any other Twitter user, regardless of the relationship between the two users, by simply adding, in the message, the name corresponding to the account of the user to whom the tweet is to be addressed preceded by the at sign (@). In this case, the user whose account is included in the tweet will receive an automatic notification from Twitter with the relevant text.
On Twitter there is also the possibility to forward the tweet of another user to all his followers, through the "*re-tweet*" function. Messages posted on Twitter can then be tagged with the use of one or more "hashtags", that is words or combinations of words linked together, preceded by the hash symbol (#).
Tagging a post with a hashtag creates a hyperlink to all other posts, recent or not, that use the same tag. A Twitter user can thus easily access a large number of posts, regardless of whether these messages come from users they are following or not.
Although the user base of Twitter is only a fraction of that of Facebook, this social media, due to its characteristics, is becoming an extremely influential real-time news source both for the entire network and for traditional media.

Instagram, finally, is the youngest of the three, having been launched by Kevin Systrom and Mike Krieger in October 2010 and then joined the Menlo Park giant two years later.
Compared to other social media, Instagram was originally intended for sharing images in `1:1 format` for the iPhone's display width at the time. Only in 2012, the app was distributed on the play store for Android devices and then definitively loosened any restrictions in 2015, with an increase to 1080 pixels and the subsequent landing on `Windows 10 Mobile`. The service also added messaging features, the ability to include multiple images or videos in a single post, as well as the *stories*, a mechanism very similar to its main competitor `Snapchat`, that allows users to post photos and videos to a sequential feed making them accessible for only 24 hours.
The follower system, identical on Twitter and equivalent to friends on Facebook, can be divided into close friends and not, in order to limit the viewing of some contents to a small number of followers. To add a new follower, in the same way as for Twitter, the authorization of the other user is not required, however each one, unlike Twitter, has the possibility to customize the information shared with the various contacts based on some privacy settings or by inserting certain followers to the circle of closest friends.
In this sense, the Instagram model is halfway between Facebook and Twitter as regards the constraints placed on the possibility of interacting with any other profile on the same social platform.

1.2 Objectives and main contributions

In light of the above, the first goal of this thesis was to develop a user guide for the `SocMint2020` application for analysing the influence on social media.
In order to allow even inexperienced users to fully use the features of the app, a guide was required that would explain each feature systematically. At first glance, in fact, the analysis of user sentiment may seem like a complicated process that can lead to a bad interpretation of the data if not viewed and read in an ideal way. For this reason, the correct use of our application is fundamental as it makes this procedure easier and faster to use even for users with non-computer specifications.
In particular, the guide was developed trying to explain all the features and properties of the application in the simplest and most understandable way, leaving room for future additions following new developments and software updates.
Subsequently, tests were carried out on the application in order to test the correct functioning of all the features developed and to identify improvements that can be made for subsequent versions of the software. For this reason, a report file was created containing the changes to be made to the `SocMint2020` application to improve the use of information and ease of use by end users. In addition, the problems that occurred during the use of the application and the anomalies found during the various tests were reported. Finally, changes were suggested to be made to correct the errors.
At this point, another goal achieved by this thesis was to carry out a process of collecting information from the application by standardizing the data by creating special weekly reports. The visualization of the data plays a fundamental role in the interpretation of the same, for this reason the inclusion within the reports does not simply have an aesthetic purpose of presenting the results, but gives access to information and interpretations that would otherwise not be quickly accessible.
The reports were structured simply as a word file containing the main information observed during the week in question such as:

- Increase of weekly posts;
- Study of Hashtags;
- Dates of publication concentration;
- Number of publications per single user.

In order to better illustrate the development of our case study in correlation with the main national political events, the Marche regional elections were analyzed in detail, integrating the reports with PowerPoint presentations containing graphics and multimedia content.
The work of this thesis allowed the development of new metrics for the analysis of the

influence on social media that would allow further development of the `SocMint2020` application. In order to further analyze the level of influence of certain users on social networks, in fact, a set of indicators was required that would integrate with the pre-existing ones. In fact, the analysis of user sentiment alone is not able to define exactly which the most important influencers within the network are.
To this end, various previous papers and studies were analyzed in the field of sentiment and influence analysis, highlighting in particular some sets of metrics on which to seek further insights and focus our research in subsequent studies.
In particular, the developed metrics were tested with the integration within the existing datasets and on which the related application is structured.
For this reason, the metrics have been developed in Python and integrated with the database through the use of `PyMongo` in such a way as to respect the development canons of `SocMint2020`, given that the machine learning algorithms used for the analysis of social media by our application store data via `MongoDB`: a non-relational, document-oriented DBMS. Classified as a `NoSQL` database, `MongoDB` moves away from the traditional table-based structure of relational databases to adopt a `BSON`-style "document" style, making the integration of data with our application easier and faster.
The aim of our project is to develop an innovative integrated system that allows an enrichment of the instrumentation available for social media analysis and that is able to automatically highlight the data to the user in order to ensure easy interpretation of the same, creating a new model of study of the influence on social media applicable in every field, based on what has been developed in this thesis.

1.3 Structure of the thesis

The thesis is organized in five Chapters, which describe the social media analysis techniques and the development of new metrics that will allow a greater interpretation of the data from the `DB`, in order to improve the existing application by adding new features, correcting some errors present and by improving the visualization and usability of the data present.
In the first chapter, we will define a collective scenario in order to be able to frame the situation in question within a context of social analysis. Subsequently, it will be necessary to have a "dictionary" of Social Analyses, in order to clarify the meaning of some terms that will be used in the rest of the paper. Once a reference vocabulary has been determined, the work will continue with a study of the main aspects in order to recreate a realistic overview of the context in question.
Chapter 2 examines the state of the art on the main issues addressed. The main directions of social media analysis and the issues still pending in this regard are discussed, and then the current situation regarding the use of social media to predict the result of certain events is explored by introducing the concept of influence analysis, an essential element in the analysis and prediction on social media.

Once the above topics have been dealt with, the related research in the field of social media analysis and sentiment analysis will be illustrated, then we will see in detail their application in reference to our case study and the elections of the Marche region.
Chapter 3 describes the materials and methods that allow the analysis of the social network in question, starting from a study of the dataset at our disposal and then going on to illustrate the metrics and the main `KPIs` adopted for the identification of the `top-k` nodes within of the network. Finally, the operation of the `SocMint2020` platform will be illustrated, where the data from the analysis of social media can be viewed and analyzed.
While in this chapter are described the methodologies and tools used for our analysis, in Chapter 4 the attention is focused on the field in which the previously described tools have operated, i.e. the elections of Marche region.
Chapter 5 then discusses the results obtained from the analysis of the regional elections, verifying the veracity of the predictions made and the very reliability of the analysis carried out on social media. Finally, we will be able to draw our conclusions and outline the areas towards which future research will be directed.
The objective of the analysis is to arrive at conclusions that are as objective as possible of the topic through a study of the most important metrics and, through statistics of real cases, to make considerations that are free of personal evaluations.
The thesis work aims at obtaining a general overview of the use of the SocMint2020 application and specifically to evaluate the performance of social networks in relation to regional elections and events that have influenced the political environment for the duration of our study. New metrics are also studied and analyzed for the analysis of the sentiment and social media, in order to be able to implement them in our project, obtaining an excellent starting point for new developments and subsequent analysis.

Chapter 2

State of art and Perspectives

Firstly, in order to give a better understanding of social influence analysis, we provide an overview on types of social networks. As social influence analysis requires accurate measurement of social influence, we then summarize and analyze the existing works on evaluation of social influence. After that, we categorize and compare a number of relevant research works on influence analysis and social media predictions. We then identify the social influence analysis. Our goal is to pave a solid foundation for various communities for further investigation of this promising topic.

2.1 Main directions of social media analysis

Precisely because of their growing diffusion, social media has been the subject of strong interest as an object of analysis both in the academic and commercial fields. With regard to scientific research, various areas and disciplines have produced studies based on the analysis of social media [5, 6]. They range from economics to marketing, from political science to education, passing through psychology and epidemiology, just to name some of the most widespread recent cases. More generally, the works that have dealt with discussing social media can be divided into two broad approaches, not necessarily complementary, which pose different research questions.

The first approach seeks to understand if, and to what extent, communication on social media is able to influence choices and behaviours on the part of its users. This framework includes, for example, the works that focus on the dissemination of information on the net [7], and on how this dissemination is able to generate, in certain circumstances, new news that becomes public domain.

More generally, we can distinguish between an external and an internal news diffusion. In the first case, the news is first produced by the mass media and only subsequently relaunched online. In this scenario, social media mainly operate as a sounding board for rumours that in any case have an external origin. On the contrary, in the case of an internal diffusion, social media manage to act as real `news-media`, anticipating traditional channels and spreading information earlier than newspapers and televisions do.

Obviously, it can be argued that the type of user who "launches" a tweet is decisive for generating the mechanism just introduced. Much research has thus focused on the effect played by the type of networks existing on the network and on the role played by the so-called most "influential" users, i.e. those with a greater number of followers, in the case, for example, of Twitter and Instagram, or friends, in the case of Facebook. However, other more recent analyses have shown that it is not the number of followers that makes the difference in itself, but the fact that the content of the information to be disseminated is in effect news worth relaunching [8]. In addition to being very effective for the dissemination of real news, social media are also the technological version of the oldest information dissemination tool: *word of mouth*.

However, the “digital word of mouth”, the so-called *eWOM* (electronic words-of-mouth), is very effective even in the commercial field. Consumers often freely express their opinion online, whether positive or negative, in relation to a particular product. This aspect becomes so important in evaluating the *brand reputation* of a brand or *customer satisfaction*, in relation to a service or a product.

The discussion is different regarding the impact of social media on other types of choices, in particular on political ones, where the theoretical evidence, for example, on how many "votes Twitter moves" is decidedly less [9].

However, recent surveys show that although the impact of social media, and Twitter in particular, on the preference votes a candidate gets is limited, that effect remains significant, particularly among those candidates who maintain a direct relationship and continue with their followers through social media [10]. In the case of "zombie" candidates, it means candidates who are present on social media but who do not use them, this effect disappears.

Precisely the possibility of affecting the choices of social media users has given way to a growing use, which we can define *top-down*, of social media directly by companies, brands, institutions as well as political parties and leaders.

The most common top-down uses by companies are the creation of *community* on products or brands, the sharing of information on news and relevant facts, the possibility of having immediate and timely contact with consumers, all factors that allow to consolidate and convey brand or corporate values. Social media are also valuable low-cost communication channels for gaining a better understanding of the customer profile.

Nevertheless, what happens for companies, even if with the necessary differences, also happens in politics, where *micro-targeting* is widely used during election campaigns. By micro-targeting we mean the use of *data mining* techniques for market segmentation together with direct marketing strategies aimed at transmitting personalized messages, in the case of the policy for each group of voters.

On the other hand, other analyses [10] have shown how political leaders, through social media, try to influence not only what traditional media talk about but also how they talk about it, that is, which aspects to privilege a news and what salience to give it. Consequently, to influence through traditional media how the public perceives certain issues that might otherwise end up being ignored.

2.2 Social media and predictions

The other great social media research approach instead adopts a substantially *bottom-up* perspective.
That is, social media as a modern `agora` to be studied in an appropriate way and from which extract information that can provide us a valuable help in understanding the evolution of complex social phenomena. An extensive source of data to understand the opinion of the writer and/or the change of this opinion following some external event. In short, like a public survey but with the ambition of capturing dynamics that change continuously at any moment.
This is what in the literature is also called *nowcasting* [11], i.e. the possibility of producing a prediction of the present, the very near future and the very recent past, identifying the dynamics of stages that are taking place in `real time`.
This line of research also includes the numerous studies that use social media to make real forecasting on future events, and that exploit the "widespread wisdom" present on Big Data. A wisdom that, precisely because it comes from often profoundly different points of view, is particularly precious.
As an indirect confirmation of this, projects funded directly by governments are starting to emerge and spread, with precisely these ambitions.
In Table 2.1 we report some examples of scientific analysis that have used social media to carry out *nowcasting* and *forecasting*. Although it does not have the ambition to provide an exhaustive explanation of such research, Table 1.2 nevertheless allows us to identify some macro-areas and, within them, the related issues on which this social media analysis approach has become more focused.
In the following chapters, issues relating to the analysis of influence and sentiment will be discussed, discussed in chapters `chap.2.4` and `chap.2.5`, and to the predictions in the electoral political field, a topic that will be addressed in detail in the next chapter.
A first major area on which the forecasts have focused concerns `economic` issues.

Table 2.1: Topics studied on social media related to the topic of "predictions"

Field	Prediction	Source
Economy	Stock market indices	Twitter; Google;
Epidemiology	Influence spread and other diseases	Google; Twitter
Marketing	Purchase/consumption of products	Blog; Google
Politics	Election Results	Twitter;Facebook
	Popularity of politicians	Twitter; Google
	Riots	Wikipedia; Twitter; Others
Psychology	Mood and states of mind	Twitter
Seismology	Identification of earthquakes	Twitter
Society	Winners of television contests	Twitter
	Sports results	Twitter

Several studies first relate sentiment, or the mood of people online, with economic parameters such as the performance of the stock market.
The tensions and concerns relating to the occurrence of important events measured through *sentiment analysis* would therefore allow us to estimate the trend over time of the stock markets.
Moving on to the `marketing` area, some studies have attempted to predict sales trends. Consumer tastes therefore seem to be predictable, as can the choices made by "consumer-voters", as previously stated.
In the same way, it is also possible to measure the preferences of "consumers-viewers", for example, to predict the winner of a television contest.
In a similar way, that is, based on the idea that social media may be able to function as a sort of "collective brain" capable of successfully aggregating widespread individual expectations and predicting, for example, the traces of actuality selected for the topics of the state exam, using only the *rumors* online by the students.
A similar type of analysis was also carried out in the `sports` field, where, for example, they tried to identify the results of the 2010 World Cup via Twitter [12]. A fact confirmed also in the case of the 2012/13 Serie A football championship, whose ranking for the top positions was predicted using only the expectations of Twitter users published in the middle of the transfer market in August 2012 [12].
In the `political` field, there was no lack of studies aimed at trying to predict the outbreak of riots or the possibility of coups, as done for example by Kalev Leetaru, using the news archive of the New York Times together with *open source* data from Wikipedia and other web sources. Using in a similar way a mix of sources, social media and news published in online newspapers, he was able to predict the revolts in Tunisia, Libya and Egypt, and the substantial stability of Saudi Arabia, till identify with a radius of 200 kilometres the hideout of Osama Bin Laden [13].
A final area in which social media is used to make predictions is linked to `medical and natural sciences`, such as epidemiology, biology and psychology, and goes so far as to estimate the damage caused by natural phenomena such as earthquakes through the network.
Recently, analyses were carried out using data collected on Twitter [14], monitoring not only the spread of the flu in the United States but also the evolution of sentiment regarding the danger of the pandemic produced by the COVID-19 virus.
These works, which in turn have implications for medical sciences and psychology, also in a preventive key, give the opportunity to possible wider uses, linked to the assessment of the degree of well-being of citizens and to the development of public policies aimed at improving their happiness.

2.2.1 Election results predictions

The exponential growth of social media is starting to play an increasingly central role in the political life of democratic regimes and beyond. Social media have been used to involve individuals in various forms of mobilization on public issues and to build social movements or real political parties, such as the "*Movimento 5 Stelle*" in Italy, all of which use the network to precisely define their programmatic lines and to select the respective candidates.
On the other hand, the diffusion of social media, at the level discussed in `chap.1.4`, has also revealed a strong interest in analysing the network to directly explore the political preferences of its users [15], as well as the popularity of political leaders [16] or the consensus with respect to specific initiatives taken by governments [17]. Several reasons lead us to believe that the analysis of social media during an election campaign can be used in conjunction with traditional polls. Beyond being decidedly economic, at least when compared with the costs related to the organization, administration and data collection of surveys, the analysis conducted online allows you to constantly monitor, day by day, hour by hour, the evolution an election campaign.
Starting from these premises, the prospect of carrying out a *nowcasting* exercise, in this case with respect to the dynamics of an electoral campaign, becomes a concrete possibility. Through social media, it would thus be possible to observe trends and identify any sudden changes in public opinion, long before what could happen by relying on traditional polls. If only for technical reasons: between the administration of a survey, the collection of data and their analysis, it can take days. In the case of network analysis, hours.
In the previous chapter we discussed how social media can help in trying to predict various social phenomena, from the trend of the stock markets to the spread of diseases. In the wake of these studies, social media can be used to predict election results [15]. This is at the same time intriguing but also particularly difficult because when we talk about elections we are confronted, by definition, not only with electoral polls, but above all with the vote of the voters. Electoral forecasts are therefore one of the few fields within the social sciences in which it is possible to really speak of "forecast", precisely because there is a final datum on the basis of which to parameterize the success or failure of the prediction made.
Studies that have used social media to predict election results are certainly not lacking. Most of these works use some of the simpler techniques, and therefore rely on popularity indicators such as for example, the number of friends and likes on Facebook, or followers on Twitter, or the volume of data, limited to count the mentions, i.e. the number of times a candidate or party is mentioned online.
Others have tried to go further than measuring the volume of data and have focused on sentiment as measured by the us of dictionaries.
However, both of these lines of analysis have failed, in various circumstances, in an attempt to predict the outcome. A first criticism refers to the fact that none

of the previous analyses can be defined as a real "prediction", given that none of these studies was published before the vote but only ex-post, after elections. On the other hand, in this case the real risk is to overestimate the goodness of the result: in other words, there is the possibility that the authors make public only those results considered satisfactory, neglecting instead those analyses in which no relationship has been found between online and off-line behaviour.
A second problematic aspect is given by the difficulty in grasping sarcasm, double meanings, and strategic statements contained in the texts analyzed. The only successful strategy should be to implement more sophisticated analysis algorithms that, by using *machine-learning* techniques, are able to understand the true meaning of the language used on the network and to recognize the contents produced by "bots" for propaganda purposes.
On the other hand, this method is able to remedy the spam problem, identifying tweets produced by both real users and "bots". Regardless of this aspect, however, the existence of online propaganda is not necessarily a problem, as the accounts that are created only to spread the message, function as a "megaphone" only when there is a broad consensus around the content of the posted message. On the contrary, when the spam attempt goes to spread political proposals little shared by the network, the few hundreds or thousands of posts produced by bots or party press offices will end up getting lost in the hundreds of thousands of posts published by voters and analyzed by the system.
Finally, and from a more strictly statistical point of view, as shown in the following chapters, the *sentiment analysis* (SA) method makes it possible to produce much more reliable estimates than the estimates that are based on a classification of texts at an individual level. This is obviously crucial when the purpose of the analysis is to predict a result that is aggregated by definition, that is, the percentage of votes of the different parties or candidates, and where the difference between the contenders is usually a few points.
In short, for all the characteristics of the SA method discussed so far, the relationship between social media, votes and forecasts would seem able to develop on a potentially more promising level than what happened in previous attempts using other techniques.

2.3 Influence analysis

Social networks (SNs), such as Instagram, Facebook, and Twitter, have emerged and tightly connected web users all over the world. "By analysing and mining social networks, we can gather information on the comments people make with respect to a particular subject". For this reason "social influence analysis is becoming an important part of social networks" [] (Peng et al., 2017a).
Graph theory is probably the main method in social network analysis in the early history of the social network concept. The approach is applied to social network analysis in order to determine important features of the network such as the nodes

and links, or in our case influencer and follower. Influencers on social network have been identified as users that have impact on the activities or opinion of other users by way of followership or influence on decision made by other users on the network. In centrality measure, argument that will be explained in the following chapters, "the graph theory was used to inspect the representation of power and influence that forms clusters and cohesiveness on social network" [18].
As a result, social influence analysis in social networks has important social significance and application value. However, giving the significant increasing of social networks in both scale and volume, we are facing numerous challenges in measuring social influence for a given user [](Peng et al., 2017a) . We believe there are many barriers in this case, and list a few here. Firstly, we do not have a mathematical definition and measurement of social influence. Secondly, it is very difficult to decide the major factors for a specific case to social influence modelling. Thirdly, there is no effective ways to properly integrate various factors for influence measurement [7].
Moreover, social influence analysis is a cross discipline domain. "As the research on social networks is still in her infant stage, the study of social influence analysis interconnects with the other features of social networks, such as influence properties, evaluation metrics of influence, collection and processing social networking big data, and selection of most influential k-top nodes" [19]. Therefore, it is necessary to investigate social influence analysis in large scale social networks under the global environment of big data.
To do that, we have to deeply understand social influence in terms of social networks. Up to now, our understanding of social influence is shallow in many aspects. In terms of representation, social networks itself can be treated as an abstract network of interconnected nodes, but researchers generally adopt a graph or matrix to describe this type of networks.

2.4 Social influence

In this section, to give a better understanding, we provide the definitions of social influence and the framework of social influence analysis.
“Social influence refers to the case when individuals change their behaviors under the influence of others. The strength of social influence depends on the relation among individuals, network distances, time effect, characteristics of networks and individuals” [20].
“Social Influence is roughly defined as follows: Given two individuals u, v in a social network, u exerts the power on v, that is, u has the effect of change the opinion of v in a direct or indirect way” [21].
Basically, we interpret social influence as a level of uncertainty, and the basic understanding of social influence can be summarized as follows:

- "Social influence is a relationship established between two entities for a specific action. In particular, one entity influences the other entity to perform an action. Usually, the first entity is called the influencer, the second entity is called the influencee" [7].

- "Social influence is a function of uncertainty. In particular, if the influencer believes that the influencee will perform the action for sure, the influencer fully “affects” the influencee to perform the action and there is no uncertainty; if the influencer believes that the influencee will not perform the action for sure, the influencer does not “affects” the influencee not to perform the action, then there is no uncertainty either; if the influencer does not have any idea of whether the influencee will perform the action or not, the influencer does not influence on the influencee. In this case, the influencer has the highest uncertainty" [22].
- "The level of social influence can be measured by a continuous real number, referred to as the social influence value. Social influence can also be represented with uncertainty" [22].
- "The influencer may have different social influence values on the same influencee for the same action. Social influence is not necessarily symmetric. The fact that A influences B does not necessarily means that B also influences A, where A and B are two entities“ [7]

2.4.1 Framework of social influence

Once we understand the meaning of social influence, we can see in detail what is the "standard" process for its analysis. To do that, the framework of social influence analysis is shown in (Figure 2.1).

1. Collection of big data from social networks: It is a very important basis of social influence analysis. With the availability of big data it is much easier and convenient for us to collect raw data from both online and offline sources. Nevertheless beware, a large amount of data does not always correspond to an equally large amount of information, on the contrary, often the opposite can happen, that is, not being able to extrapolate any useful information.
2. Big data preprocessing: In order to improve the performance and the convenience of processing, it is necessary to remove the irrelevant information on the social influence analysis to not have the problem mentioned above. In order to protect the privacy of users, it is also necessary to filter out the sensitive information related to privacy protection from the collected data sets.
3. Selection of evaluation metrics: It is very important to extract a set of evaluation metrics to measure social influence of each user. As these evaluation metrics are helpful to quantify social influence of each user, and to find easily the most

influential *top-k* nodes. In this work, the utilised evaluation metrics include degree centrality of each node, betweenness centrality, closeness centrality and other evaluation metrics based on the flow of entropy.

4. Measuring social influence: According to the extracted evaluation metrics, evaluation model and computing equations are provided to a specific social network. Thus, social influence of each user can be measured, by integrating the computing equations into the collected real-world data sets.

5. Design of influence maximization algorithm: The influence maximization algorithm is designed to find the most influential top-k nodes [19]. In our specific case, no one of this kind of algorithm is used.

6. Performance analysis on related algorithm or model: Simulation is made to validate performance of the proposed algorithms based on a specific propagation model. To have an excellent algorithm, it is needed a large influential range and low computational complexity.

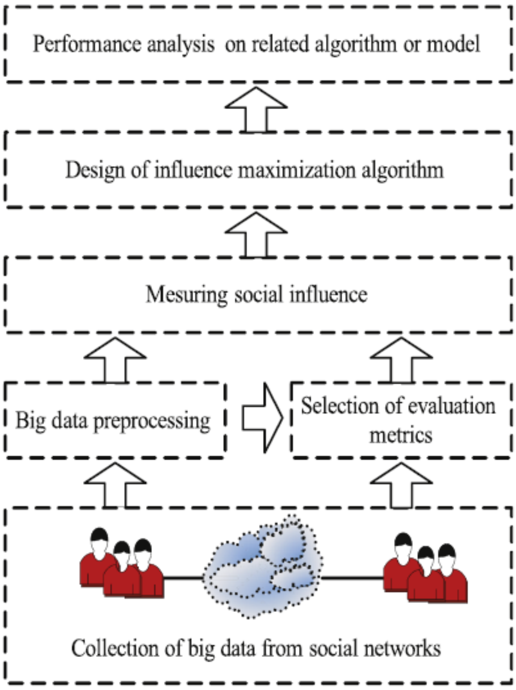


Fig. 1. Framework of social network analysis.

Figure 2.1: Framework of social network analysis

2.5 Sentiment analysis

Once we have introduced the concept of social influence and defined a structure for the study of it, we will define another fundamental concept in the context of a social

media evaluation: sentiment analysis.
This can be defined as discovering and acknowledging the expression of positive or negative opinions by people on various topics of interest. The opinions expressed by users of social networks are often convincing and these indicators can be used to form the basis of choices and decisions made by people who have sponsored certain products and services or who have supported political candidates during the elections. Furthermore, it is essential to analyze the influence of a particular user since it adds further information to it. A user with a high influence could generate, as a matter of fact, a high negative sentiment. At that point, the influence generated should also be interpreted as a negative data since it would mean that the users reached by the information have developed a negative interest in the user who generated it.
For example, in the case of political elections, a high level of influence by opposition parties could be instrumental in a change in government leadership. On the contrary, the same level of influence, which however is associated with an equally high level of negative sentiment, could instead simply mean that the opposition parties are not well regarded by the public, thus obtaining a meaning diametrically opposite to the previous one.
Consequentially it has become necessary to analyse sentiment expressed on social network with data mining techniques in order to generate a meaningful frameworks that can be used as decision support tools.
The purpose of sentiment analysis on social network is to recognize potential drift in the society as it concerns the attitudes, observations, and the expectations of stakeholder or the populace.
It is important to translate sentiment expressed to useful knowledge by way of mining and analysis.
Having given an overview of sentiment analysis on social network, an overview of some of the data mining tools used for sentiment analysis on social network are discussed in subsequent sections of the survey.

Chapter 3

Materials and Methods

This chapter is divided into the various parts which, when combined, form an innovative method for monitoring the social influence and sentiment generated by social media.
Firstly, will be illustrated the structure of the database on which the entire application rests on.
Subsequently, the metrics developed to analyze the data extracted from social media and stored in the DB, during the Marche regional elections, will be introduced.
Finally, the functioning of the SocMint2020 web application will be illustrated, highlighting how the three technologies presented work in synergy for the analysis of social media.

3.1 Database

A database is an organized architecture for the storage of information that can be used to manage a huge amount of data without loss of performance on data search and data organization.
The database chosen for our application is a document-oriented database managed by a non-relational *Database Management System* (DBMS). To be clear, a DBMS is a software system designed to allow the creation, manipulation and efficient querying of the databases, therefore also known as "database manager or engine". In our case, the chosen DBMS is of the `NoSQL` type. MongoDB, in fact, moves away from the traditional structure of relational databases, where the data is organized in tables that represent both the entities and the relationships between them, in favour of documents, making the integration of data with our application faster and easier respect to a graph based Database. In fact, a graph based DB, that represent data in a form of graph, it is useful to represent the relation between users on a social network but it has a problem of performance to store the raw data that are extracted with an hashtag based search and represent an entire post. This problem disappears with a document based `NoSQL` database. Furthermore, released under the GNU General Public License, MongoDB is free and open source software.
Document-oriented databases like MongoDB, unlike relational databases, do not store data in tables with uniform fields for each record, but each record is stored as

a document that possesses certain characteristics and that can be categorized in a collection. Any number of fields with any length can be added to the *JavaScript Object Notation document* (JSON) with dynamic schema, called *BSON* (Binary JSON). Compared to JSON, BSON is designed to be efficient for both the space required by the data and the search speed. The extended elements, in a BSON document, are equipped with a prefixed extension length field, in order to facilitate the search for information. Documents are addressed within the database using unique keys. In our case, the key is a simple alphanumeric string, in some cases a `URI` or a `path` could be used instead.
One of the special features of a document-oriented database is that, in addition to simply searching for the document by key, the database also offers `APIs`, or a simple query language, which allows for content-based retrieval. In our case, MongoDB has ad hoc queries and supports searches by fields, ranges and regular expressions. Queries can return document specific fields and include user-defined functions in JavaScript. This, for example, allows you to retrieve documents based on the value of a certain field.
In our case, the database has a structure that reflects the description of the categories obtained from the social API. The main documents of our database are three: posts, raw posts, users. In our database, within the `posts` document, which is stored in a posts collection, other fields relating to the publication of a post are described.
A post, for the *Spatial Decision Support System* (SDSS), is a JSON document containing heighten keys, which can be categorized in post keys, user keys, media keys and analysis keys. The post keys are the keys that contains the data from a raw post as, for example, the date of publication, the text and the language of the post and the `URL` needed to see the original data on the source platform.
The user keys are the key that contains information regarding the user that has created the post including also the information about the location of the user.
The media keys, simply, are a group of JSON keys that contains a value that is useful to obtain a media from a post, be it an image or a video, redirecting directly to it by the media `URL` key.
Finally, the analysis keys, the most useful set of keys utilized for the analysis of the post. These keys are chosen based on the analysis system described below and to keep trace of the information about the post on some data helpful for the final user of our platform. An example of these keys are:

- the engagement value, representing how many interactions a determined content has obtained;
- the sentiment value, an integer number representing of the overall sentiment associated at the entire post, including the post text, the text extracted from image and the image;
- the sentiment label associated to the sentiment value, describing if is a positive,

negative or neutral value of the sentiment;

- the list of hashtags presents in the post;
- the list of user of the same social media mentioned in the post.

Similarly, through the JSON document dedicated to users, the main attributes of a social user are described. An example of this is, in addition to the number of followers, followed and published posts, `user id` and `username` which, although often coincide, respectively indicate the unique identifier of the user and his name on the relative social network average. Finally we find the raw posts document very similar to posts. It contains posts from Facebook, Twitter and Instagram that are still being processed, for sentiment and engagement calculations. These posts will then be included in the `rawposts` collection together with superfluous data, which are not used by our application and for this reason will not be saved within the collection posts, but which could be useful and used in future analyses. As previously specified, data storage through mongoDB contains only textual information, thus managing to keep the DB as small as possible, effectively reducing the amount of data necessary for our application. In fact, through this choice the database presents 36 thousand documents for the collection of users in just `50MB`, 148 thousand documents for the `rawposts` collection with `1.2GB` of occupied memory and 148 thousand documents for the collection of posts in `190MB` of memory. Moreover, due to the absence of binary objects needed to include media in the database, the query response performances are very high with 10ms latency between formulation and response from the DBMS. This aids to reach real-time performances. This can be possible thanks to the use of the image link instead of downloading it.

3.2 Key Performance Indicators (KPIs)

In this section we will fundamentally analyze three metrics, choosing how to apply them to the inspection and analysis of multimedia data flows from social networks and adapt them in the context of politics, with reference to the case under consideration of the Marche regional elections.
On the fourth metric in question we will subsequently choose whether to develop it further or not, considering it as having little influence in our case study.
Two of the three metrics under consideration are based on the concept of centrality. *Centrality* is an important concept in studying social networks. Conceptually, centrality measures how central an individual is positioned in a social network [7]. The commonly used tools are graph theory and network analysis. Diverse centrality measures [1] have been proposed in the literature to widely be used in social influence analysis, including degree centrality, closeness centrality, betweenness centrality, eigenvector centrality, and Katz centrality [7].
The first metric based on the concept of centrality that we went to analyze was

degree centrality. Two definitions are given below.
"The *degree centrality* of any node in a graph or network is defined by the count of edges that are incident with it, or the count of nodes adjacent to it. Degree centrality of node in case of directional networks is given in two ways as in-degree and out-degree" [1]. "The *degree centrality* is defined by counting the number of links incident upon a node, namely, the number of edges that a node possesses" [7]. The degree centrality can be interpreted in terms of immediate risk of a node for catching whatever is flowing through the network.
"For a directed network, we usually use two metrics for degree centrality: in-degree, which is a count of the number of edges directed to the node, and out-degree, which is the number of edges that the node directs to others" [7].
In this case, this division was not possible because on the dataset at our disposal it is not possible to make this distinction between incoming and outgoing links with respect to a node belonging to the Facebook platform. It was therefore considered simply the number of links that each user has, without making any distinction between followed and followers (Figure 3.1). Finally, the degree centrality can be calculated as:

$$\boldsymbol{C_D(i)} = \sum_{j=1}^{n} a_{ij} \tag{3.1}$$

Where in Eq. 3.1 a_{ij} is 1 in the binary adjacency matrix A if a link from node i to j exists, or it is 0 if there is no connection between the two nodes, as shown in (Figure 3.1). That is to say, in our case, simply count the total number of followers of an Instagram or Twitter account, the number of subscribers to a page of a public figure or more simply the number of friends in the case of a simple Facebook user.

The second metric based on the concept of centrality that we went to analyze was closeness centrality. Two definitions are given below.
"Closeness centrality is defined as the average distance from one vertex to the other vertices in a network. It can be seen like the efficiency of each node in terms of spreading information in the network. The larger the closeness centrality of a node, the better positioned the node is in the social network" [7].
"Closeness centrality is a measure that identifies how fast it will take to flow information from node to all other nodes sequentially, it is defined as the average distance from one vertex to the other vertices in a network. Nodes occupying a central position with respect to closeness are very productive in distributing information to the other nodes" [1]. In influence analysis, we can treat it a metric of efficiency of each node in terms of spreading information in the network.
The larger the closeness centrality of a node, the better positioned the node is in the social network.

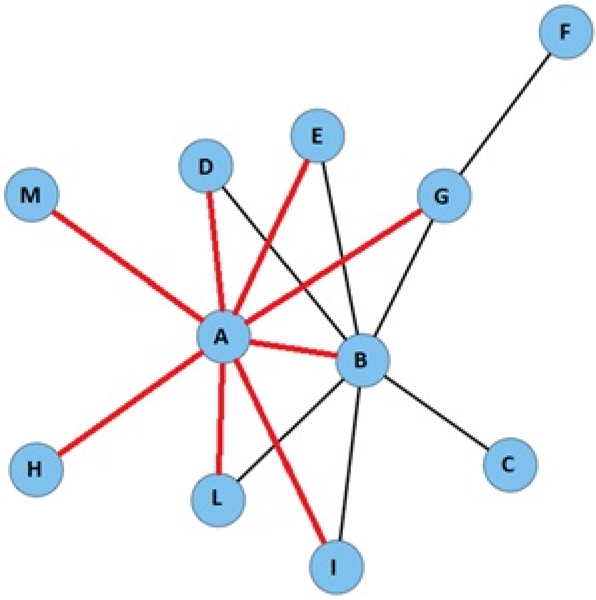


Figure 3.1: Degree centrality of the node A

Closeness centrality can be measured as:

$$\boldsymbol{C_C(i)} = \frac{1}{\sum_{j=1}^{n} d(i,j)} \tag{3.2}$$

Where in Eq. 3.2 $d(i,j)$ represent the distance between node i and j, which is the measured minimum length of any path connecting i and j.
Since it is not possible for us to directly measure the distance between two nodes i and j within our social network, we have tried to review this metric, no longer from the point of physical distance, but by measuring how easily the information passing from the node i can reach node j.
To do this, we have given a different weight to each interaction that user i can make on a post, image or tweet that is then displayed in user j's feed. In this way, not very invasive interactions will have a minor influence on user j, as likes, comments and shares will affect in a different way the feed of the users connected to us (Figure 3.2).

For example, a like on a post will have a lower chance of making the content appear on a friend's wall on Facebook, compared to a comment or direct sharing which will surely make the post appear on Facebook feed of everyone friend of the user.
By doing so, these data can be interpreted as a valid measure of the distance between i and j, a like to a post is having less chance of being reached by the user j than a direct share, in fact corresponds to a greater "virtual" distance between the two nodes.

The other category of metrics examined falls within the scope of metrics that use

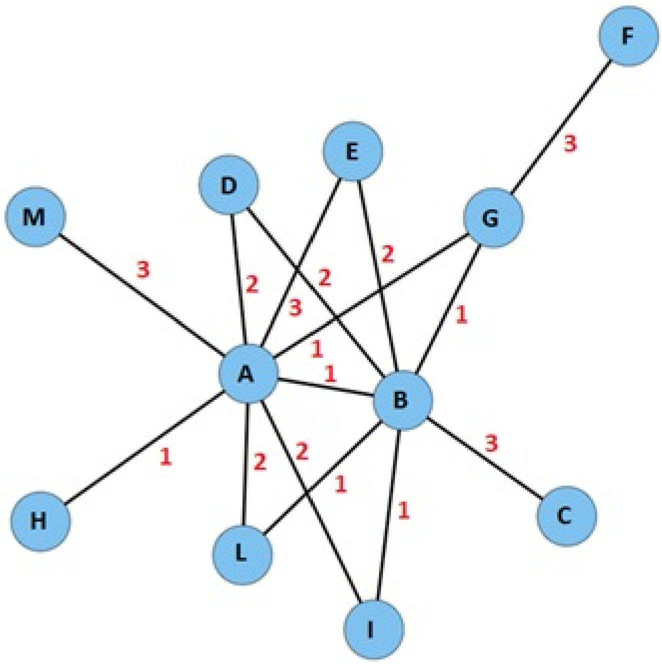


Figure 3.2: Closeness centrality of the node A

the measurement of `entropy` as a benchmark.
`Entropy` (from the ancient Greek $\epsilon\nu$ en, "inside", and $\tau\rho o\pi\eta$ tropé, "transformation") is, in statistical mechanics, a quantity that is interpreted as a measure of the disorder present in any physical system included, as a limit case, the universe.
Information entropy is a relatively recent concept and was firstly introduced by Claude Elwood Shannon, an American mathematician, electronic engineer and cryptographer, in the work of "A Mathematical Theory of Communication" in 1948.
"It is a concept from information theory. It tells how much information there is in an event. In general, the more uncertain or random the event is, the more information it will contain" [20]. According to Shannon's theory, if a random variable X represents a set of possible events x_i whose probabilities of occurrence are p_i, $i = 1, ..., n$, then a measure $H(X)$ of the uncertainty of the outcome of an event given such distribution of probabilities should have the following three properties [20]:

- $H(x_i)$ should be continuous in p_i;
- if all probabilities are equal, it means that $p_i = \frac{1}{n}$, then $H(X)$ should be a monotonically increasing function of n;
- if a choice is broken down into other choices, with probabilities c_j, $j = 1, ..., k$, then $H(X) = \sum_{j=1}^{k} c_j H_k$, where H_k is the value of the function $H(X)$ for each choice.

Thus, Shannon proved that the only function that satisfies all three properties is given by:

$$\boldsymbol{H(X)} = -\sum_{i=1}^{N} p(x_i)\log_2 p(x_i) \tag{3.3}$$

Since entropy is an effective tool to describe the complexity and uncertainty of social influence, it has been widely used in social networks. Steeg and Galstyan [23] measured social influence in social networks by using transfer entropy. In order to identify peer influence, He [24] introduced transfer entropy in online social networks. Peng [20] invented two concepts, friend entropy and interaction frequency entropy, to measure social influence in mobile social networks. He presented a novel entropy-based model to evaluate social influence of users. "The purpose is to design a general model, which shows what the social influence of each individual is on a given information entropy" [20].
To measure the social influence of the users, is enough to compute the entropy of friend nodes and the entropy of interaction frequency of the users.
Firstly, the entropy computation among friend nodes can be defined as follows.
"Let $N_i(t)$ be the number of friend nodes of node i in time t. In general, the size of $N_i(t)$ is an important factor to measure the influence of a node in a social network. Thus, the entropy of friend nodes $I_i^f(t)$ for node i is described as follows" [20].

$$\boldsymbol{I_i^f(t)} = -\frac{1}{N_i(t)+\theta}\log_2\frac{1}{(N_i(t)+\theta} \tag{3.4}$$

Where θ is an adjusted parameter used to satisfy the monotony characteristics of expression. In our case, it assumes the value of 1 because we are assuming it as the minimal value of followers or friends needed to let the function be monotonic increasing.
Secondly, we can introduce the entropy computation on interaction frequency among friend nodes as follows.
"Let $C_{ij}(t)$ be the number of interactions between node i and node j in time t. In general, the size of$C_{ij}(t)$ is also an important factor to measure the influence of a node in a social network. Thus, the entropy of interaction frequency $I_i^c(t)$ for node i is described as follows" [20].

$$\boldsymbol{I_i^c(t)} = -\frac{1}{\sum_{j=1}^{N_i(t)} C_{ij}(t)+\lambda}\log_2\frac{1}{\sum_{j=1}^{N_i(t)} C_{ij}(t)+\lambda} \tag{3.5}$$

Where λ is an adjusted parameter used to satisfy the monotony characteristics of expression. In our case, it assumes, as for θ, the value of 1 because we are assuming it as the minimal value of interaction between two users needed to let the function be monotonic increasing. The interactions taken into account are for Twitter, comments, likes and retweets, while for Facebook and Instagram they are comments, reactions and shares.
Finally, we can give the entropy-based definition of total social influence of node as

follows.
"According to the above analysis, the influence of i on its friend nodes is described as follows"[20].

$$\boldsymbol{I_i(t)} = \alpha I_i^f(t) + \beta I_i^c(t) \tag{3.6}$$

Where α and β denote $I_i^f(t)$ and $I_i^c(t)$ of weight, respectively, and $\alpha + \beta = 1$.

3.3 SOCMINT2020

Our web application has a very simple and intuitive dashboard, mainly characterized by four buttons (Figure 3.3) that allow navigation between the main functions. The dashboard displays the list of campaigns by default, identified by the name and by the list of hashtags present in the tracker's search query. Using the four icons located at the top right (Figure 3.3) we can:

- access the list of application owners;
- view the list of campaigns;
- change the keys related to your user;
- log out.

Figure 3.3: Navigation icons

The most important feature is certainly the one that allows viewing the list of campaigns. Using the `plus` button, located under the four main commands, will be possible to access the form for inserting a new campaign (Figure 3.4). By clicking on one of the chosen campaigns will be possible to expand the panel and visualize two buttons (Figure 3.4):

- the first allows access to the campaign data resulting from the search set;
- the second allows you to access the settings and modify the search parameters of the campaign.

After clicking on the `view` button we will be redirected to the dashboard to view the data (Figure 3.6) and five new buttons will appear at the top right (Figure 3.5):

- `Sentiment` for viewing data relating to engagement, time of publication, user name and the social platform of origin of the post;
- `Media` for viewing the campaign media;

Figure 3.4: Campaign list

- `Hashtag` for viewing hashtags and mentions in the form of bubble charts;
- `Influencer` for viewing the media influence that the campaign has had on users;
- `Map` to view the provenance of the campaign data on a regional basis.

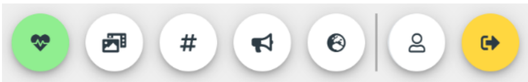


Figure 3.5: Visualization Icons

In the header, regardless of the selected category, there is a toolbar containing the main filters (Figure 3.6):

- `Date`: allows you to filter the contents based on the publication date by selecting a time window;
- `Platforms`: allows you to filter posts by selecting one or more social platforms of origin;
- `Sentiments`: allows you to filter only the posts that have obtained a positive, neutral or negative reaction;
- `Hashtags`: allows you to filter posts containing a specific hashtag. By clicking on `AND` it will be possible to enable the filter using more than one hashtag;
- `Mentions`: allows you to filter the contents based on the mention of a specific `userID`;
- `Users`: allows you to filter posts based on the user who posted them;
- `Aggregation`: allows you to view the contents and related graphs by setting a daily or hourly scale.

In the filter section below, after selecting one of the five categories, the complete list of posts corresponding to the search parameters and filters set will be displayed.

In addition, depending on the category chosen, the appropriate graphs and related multimedia files will be displayed, necessary for a greater interpretation of the data collected in the database (Figure 3.6).

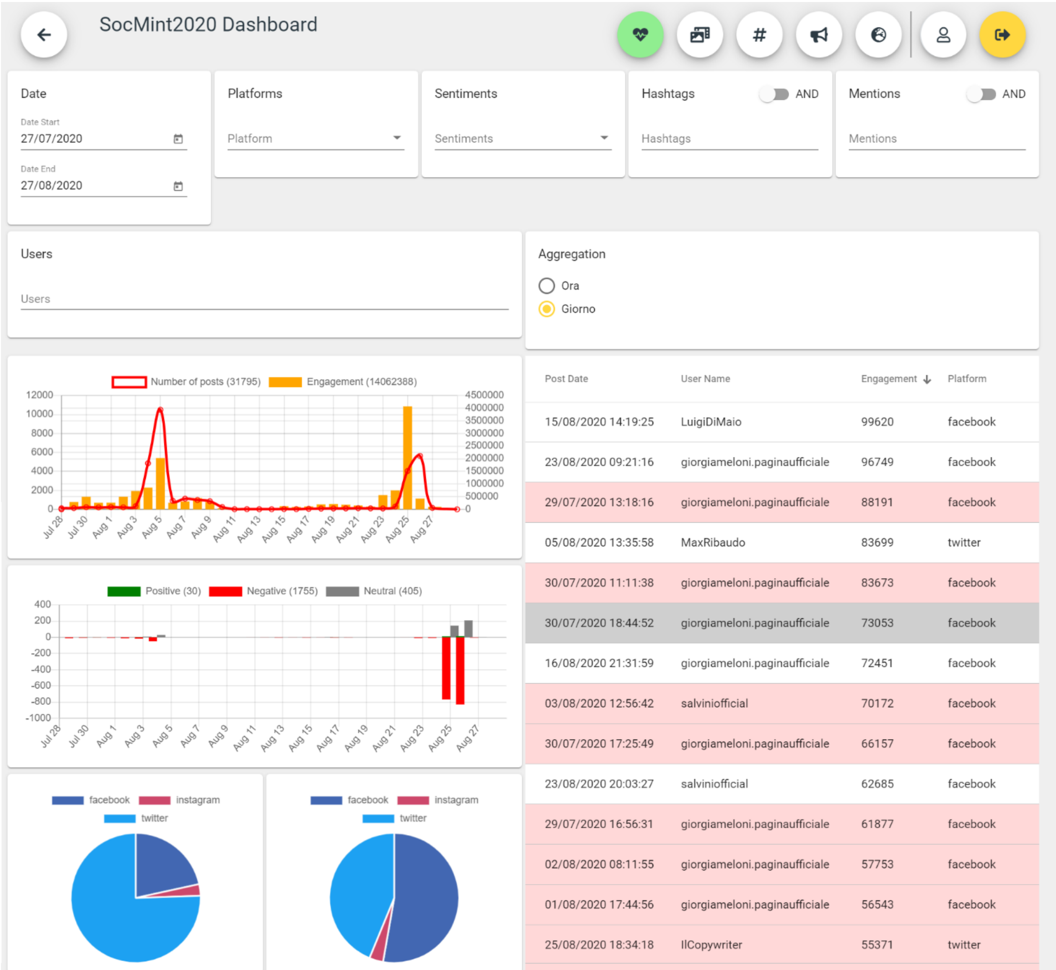


Figure 3.6: SocMint2020 Sentiment Dashboard

Chapter 4

Results and Discussions

4.1 Use cases Results: test Marche

This section presents the results obtained following the analysis of social media during the regional elections of the Marche.
The comparison within the network was mainly carried out on the two leading candidates of the elections: Maurizio Mangialardi of Partito Democratico and Francesco Acquaroli of Fratelli d'Italia and candidate of the center-right coalition.
The analysis and data collection period began in July 2020 and ended the week following the elections for a total of three months. During the period under review, there were various peaks regarding the number of posts published in correspondence with some national events such as, for example, the release of the Prime Ministerial Decree of 5 August or the regional elections and the constitutional referendum of 21 and 22 September.
Interestingly, the activity of both the candidates is more focused on content production of than in sharing or reposting. In fact, retweets represent almost 14,6% of all posts from the media and the right wing community, while in the case of the center-left community it is 34,5%.
In the first week of September in the comparison between the candidates for governors, concerning issues relevant to the future of the Marche region, the engagement was higher for the center-right (Table 4.1), while the sentiment was generally positive. During the analysis, improvements were found in the presentation graphics and the quality of the images of the events, consequently generating an increase in engagement. Another fact that emerged from the first week of analysis was a very positive impact of Francesco Acquaroli's tourism issues, resulting in an increase in engagement, generation of interactions from new profiles and an increase in the numbers of the young target on the candidate's Instagram profile. Quite the opposite, a negative impact was found on the health issue and the single hospital of Mangialardi.
Thanks to the use of multiple images with a higher photographic definition, the quality of the center-left candidate's posts proved to be superior, guaranteeing greater dissemination in terms of engagement and sharing of multimedia content, thus leading Mangialardi to greatly improve its performance on social media compared to the differences in engagement noted in August (Table 4.1).

Comparing our statistics with the official TECNE polls of the week in question (Figure 4.1), the advantage of center-right of Francesco Acquaroli is confirmed over center-left of Maurizio Mangialardi.

In the second week of September, the social analysis revealed a high level of en-

Figure 4.1: Comparison between SocMint2020 and surveys for the first week of September

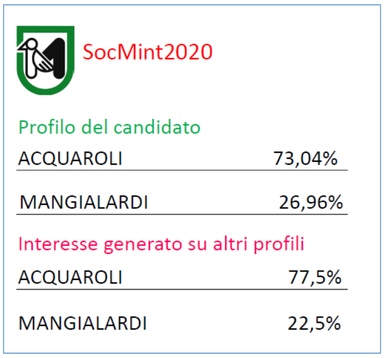
gagement by Francesco Acquaroli (Table 4.1), thanks above all to the photos of the events and meetings organized with the president of Fratelli d'Italia, Giorgia Meloni. The posts that preceded the event and those made during it allowed the candidate to involve 73% of the social community following the elections. Furthermore, as the quality of the images of the events has increased, the impact on engagement has increased. Compared to the differences in engagement of the previous week, as illustrated above, the underlining of the errors of past management has generated excellent results, in terms of impact, increasing the advantage of Acquaroli in engagement and leading the candidate to greatly improve his performance social thanks above all to the ability to skilfully play on the theme.

Despite this, the quality and engagement of posts on event programs of Mangialardi remains superior, ensuring greater dissemination in terms of engagement and content sharing. In addition, interactions with national level personalities generate the greatest impact for Mangialardi despite the fact that it was the event of his rival, with Giorgia Meloni, that generated the greatest engagement.

In both cases, the Instagram scenario remains low-key even during the period under review although the previous week there were slight improvements from the center-right candidate.

Also this week the advantage of center-right of Francesco Acquaroli over center-left of Maurizio Mangialardi is reconfirmed thanks to the comparison of our statistics with the official TECNE polls of the week in question (Figure 4.2).

During the third week of September, the meetings and comments on the post-`COVID-19` return to school had an excellent impact in terms of engagement for both candidates, as well as what happened for the news regarding the fire that took place at the port of Ancona on September 16. Both have in fact commented on the



SE SI VOTASSE OGGI , QUALE PRESIDENTE E QUALE PARTITO VOTEREBBE?

Candidato	%	margini della stima		Lista	%	margini della stima	
Francesco Acquaroli	51,5	49.5	53.5	Lega Nord	24,0	22.0	26.0
				Fratelli d'Italia	17,5	15.5	19.5
				Forza Italia - Civici per le Marche	6,0	4.0	8.0
				Movimento per le Marche	2,0	1.0	3.0
				Unione di Centro	1,0	0.5	1.5
				Civici con Acquaroli	1,0	0.5	1.5
Maurizio Mangialardi	36,0	34.0	38.0	Partito Democratico	21,0	19.0	23.0
				Mangialardi Presidente	7,0	5.0	9.0
				Italia Viva - PSI - DemoS - Civici Marche	3,0	2.0	4.0
				Le Nostre Marche e Il Centro con Mangialardi	2,5	1.5	3.5
				Marche Coraggiose	1,5	0.5	2.5
				Rinasci Marche	1,0	0.5	1.5
Gian Mario Mercorelli	9,0	7.0	11.0	Movimento 5 Stelle	9,0	7.0	11.0
Altri candidati	3,5	2.5	4.5	Altre liste	3,5	2.5	4.5
				COALIZIONE ACQUAROLI	51,5	49.5	53.5
				COALIZIONE MANGIALARDI	36,0	34.0	38.0

NOTA METODOLOGICA
Campione di 1.000 casi, rappresentativo della popolazione maggiorenne italiana, residente nelle Marche, articolato per sesso, età, provincia
Estensione territoriale: territorio regionale
Interviste effettuate il 2 settembre 2020 con metodo cati - cawi.
Margine di errore: 3,1%
Totale contatti: 6.047 (100%) rispondenti: 1.000 (16,5%) - rifiuti/sostituzioni: 5.047 (83,5%)
Soggetto che ha realizzato il sondaggio: Tecnè srl
Committente: Corriere Adriatico
Il documento completo sul sito:

Figure 4.2: Comparison between SocMint2020 and surveys for the second week of September

event but the engagement resulting from the two posts was higher for Mangialardi (Table 4.1). The center-left candidate continued his campaign with several meetings while Acquaroli expressed his closeness to the fire that occurred the day before, increasing engagement.
Both the televised debate that took place in the middle of the week and the public confrontation organized by Confindustria saw as the winner, in terms of engagement, the candidate from the center left who closed the gap compared to the previous week. Nonetheless, in terms of engagement, Acquaroli is once again the winner over Mangialardi this week.
In the last week under examination, the arguments of the candidate Acquaroli were mainly focused on the end of the electoral campaign and the subsequent victory of the elections.
In particular, the post of greatest interest was the one that portrayed the candidate together with the group leader Giorgia Meloni, generating a high engagement.
Even for Mangialardi the topics of the week are centered on the end of the campaign and the results of the elections, for the center-left candidate, the closure of the campaign met with a high level of interest which, however, aroused negative sentiment on social media; a fact also confirmed by the little-appreciated post-election comments. On the contrary, the comments on the victory of the candidate Acquaroli were appreciated by the majority who elected him (Table 4.1).

To conclude the analysis on the regional elections, we report statistics about the result of the elections related to the estimated data from the application (Figure 4.3).

Table 4.1: Engagement and sentiment data collected from social media analysis

	Mangialardi	Acquaroli
August	42%	58%
First week of September	31,69%	68,31%
Second week of September	29,96%	73,4%
Fire in Ancona	68,5%	31,5%
Television debate	59,43%	40,57%
Confindustria public debate	58,65%	41,35%
Third week of September	45,10%	54,90%
End of the campaign	59,84%	43,16%
Vote of the candidate	43,47%	56,53%
Election results	9,19%	90,81%

The red color indicates a negative sentiment

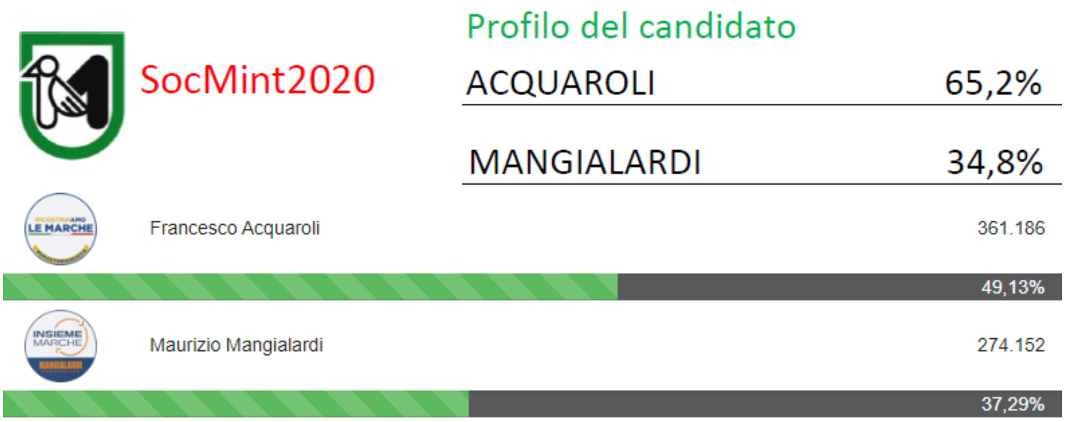


Figure 4.3: Comparison between SocMint2020 and election results

Chapter 5

Conclusions and Future Works

5.1 Discussions

In this paper, the `SocMint2020` web application was presented, integrated into a social network and designed with the aim of being a support tool for social media analysis and monitoring of social media influence by means of machine learning algorithms capable of analyze online sentiment and engagement.
To improve the analysis, new metrics have been introduced to provide new information regarding the centrality of a node within a social network and the ability of this node to spread information.
The proposed technology could represent an excellent solution for monitoring social networks, in a simple and intuitive way, allowing you to extrapolate an increasing number of information from network users.
The metrics illustrated in the Chapter 3.2 were chosen based on compatibility with our system and ease of integration with the existing dataset.
The dataset considered, as explained in Chapter 4.1, includes three main documents: posts, rawposts and users, generated by the analysis of the three social media in question.
During the analysis of the regional elections, constant monitoring of the network was carried out; starting from 1 July and which lasted until 30 September, thus also including the week following the elections.
Monthly reports were then produced, in the case of July and August, in which the evolution of the election campaign of the presidential candidates of the Marche region was further analyzed. Starting from September, the reports became weekly, allowing us to analyze in detail, and constantly, the performance of the two main candidates to win the elections: Maurizio Mangialardi and Francesco Acquaroli.
As reported in the previous chapter, very interesting results were achieved and not far from the actual figures of the scrutiny. Each report, in fact, highlighted the advantage of the center-right coalition, however, going from week to week to also underline the attempt to recover of the other candidate. The final gap between the two candidates, highlighted by the last analysis, would seem to emphasize that the data achieved were not as positive. The cause of this gap is actually given by a poor analysis of the influence of other competitors, leading to slightly overestimate the

data relating to the influence of the center-right candidate.
In both cases, however, a broader analysis would resolve this disparity.
Another possible reason could be the following: having found during the analysis a greater involvement of the young target by Francesco Acquaroli and his greater use of the Instagram platform compared to his political opponent, it could have significantly increased the gap detected by our application. In addition, in this case a broader analysis, including a greater number of candidates, would be enough to achieve superior results.

5.2 Thesis contributions

In light of the above, the first goal of this thesis was to develop a user guide for the `SocMint2020` application for analysing the influence on social media.
In order to allow even inexperienced users to fully use the features of the app, a guide was required that would explain each feature systematically. At first glance, in fact, the analysis of user sentiment may seem like a complicated process. For this reason, a correct use of our application is fundamental as it makes the social media analysis easier even for users with non-computer competences.
Subsequently, tests were carried out on the application in order to try the correct functioning of all the features developed and to identify improvements for subsequent versions of the software.
At this point, another goal achieved by this thesis was to carry out a process of collecting information from the application by creating special weekly reports. The visualization of the data plays a fundamental role in the interpretation of the same giving access to information and interpretations that would otherwise not be quickly reachable.
The work of this thesis leaded to the development of new metrics for the analysis of the influence on social media that would allow further development of the `SocMint2020` application. In order to further analyze the level of influence of certain users on social networks, in fact, a set of indicators was required that would integrate with the pre-existing ones. In fact, the analysis of user sentiment alone is not able to define exactly which the most important influencers within the network are.
Our project developed an innovative integrated system that allows an enrichment of the instrumentation available for social media analysis and that is able to automatically highlight the data to the user in order to ensure easy interpretation of the same, creating a new model of study of the influence on social media applicable in every field, based on what has been developed in this thesis.

5.3 Future Works

The results achieved in this thesis show how the use of the SocMint2020 application, integrated with the creation of weekly reports, can be an important and innovative

analysis tool in forecasting the trend of an electoral campaign, simplifying the analysis process and making efficient study of a given social phenomenon.
Nonetheless, this work represents only a small step towards new methods for social media analysis. Further research in this field is desirable, as it would make this system more effective and even more efficient.
Among the possible future developments, the possibility of developing artificial neural networks for the analysis of data from social media stands out. This would allow a more in-depth analysis of the information to the detriment, however, of a longer learning period. The necessary iterations, within neural networks, depend on many factors: such as the number and complexity of the input variables, algorithms used and much more. In fact, important progress has been made in this area, and it is reasonable to hypothesize that in the future the learning period may be further reduced.
In addition, to make the monitoring phase even more complete, in addition to the application of deep learning algorithms, reinforcement-learning techniques, the last frontier of machine learning, could be developed. A reinforcement algorithm, in fact, learns by trial and error with the aim of achieving a clear goal, which in our case could be the analysis of a particular user.
Moreover, some simple changes could be made to the handwriting interface, in order to improve the use of information and ease of use of the software by end users. Furthermore, other types of data could be integrated into the application like those coming from the analysis of increasingly popular content on social networks: like stories and short videos
Analysing these data in their entirety could lead to valid benefits in terms of forecast accuracy, by analysing new features that today are the basis of online social influencing.
This would therefore represent a fundamental tool for the analysis of social networks in the political sphere, and beyond, to be able to conduct studies on other themes and trend predictions of various kinds.

Bibliography

[1] Sanjeev Sharma Roshan Rabade, Nishchol Mishra. Survey of Influential User Identification Techniques in Online Social Networks. *Advances in Intelligent Systems and Computing*, 235:V–VI, 2014.

[2] Emmanuel Lazega, Stanley Wasserman, and Katherine Faust. Social Network Analysis: Methods and Applications. *Revue Française de Sociologie*, 36(4):781, oct 1995.

[3] Wan Shiou Yang, Hung Chi Cheng, Jia Ben Dia, and Hsing Tzu Lin. Mining social networks for targeted advertising. In *Proceedings of the Annual Hawaii International Conference on System Sciences*, volume 6, 2006.

[4] Scott. Social Network Analysis: A Handbook - John Scott, John P. Scott - Google Buku, 2000.

[5] Na Li and Denis Gillet. Identifying influential scholars in academic social media platforms. In *Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, ASONAM 2013*, pages 608–614, 2013.

[6] Zoha Rahman, Kumaran Suberamanian, Hasmah Zanuddin, Sedigheh Moghavvemi, and Mohd Hairul Nizam Md Nasir. Social media engagement metrics analysis - "Study on fan page content". *Journal of Telecommunication, Electronic and Computer Engineering*, 8(8):71–76, 2016.

[7] Sancheng Peng, Yongmei Zhou, Lihong Cao, Shui Yu, Jianwei Niu, and Weijia Jia. Influence analysis in social networks: A survey. *Journal of Network and Computer Applications*, 106(July 2017):17–32, 2018.

[8] Sitaram Asur and Bernardo A. Huberman. Predicting the future with social media. *Proceedings - 2010 IEEE/WIC/ACM International Conference on Web Intelligence, WI 2010*, 1:492–499, 2010.

[9] Clay Shirky. The political power of social media: Technology, the public sphere, and political change. *Foreign Affairs*, 90(1):28–41, 2011.

[10] Niels Spierings and Kristof Jacobs. Getting Personal? The Impact of Social Media on Preferential Voting. *Political Behavior*, 36(1):215–234, mar 2014.

[11] Marta Banbura, Domenico Giannone, and Lucrezia Reichlin. Now-Casting. (1275), 2011.

[12] Naushad UzZaman, Roi Blanco, and Michael Matthews. TwitterPaul: Extracting and Aggregating Twitter Predictions. 2012.

[13] Kalev Leetaru. Culturomics 2.0: Forecasting large-scale human behavior using global news media tone in time and space. *First Monday*, 16, 2011.

[14] May Oo Lwin, Jiahui Lu, Anita Sheldenkar, Peter Johannes Schulz, Wonsun Shin, Raj Gupta, and Yinping Yang. Global Sentiments Surrounding the COVID-19 Pandemic on Twitter: Analysis of Twitter Trends. *JMIR Public Health and Surveillance*, 6(2):e19447, 2020.

[15] Michael D. Conover, Bruno Gonçalves, Jacob Ratkiewicz, Alessandro Flammini, and Filippo Menczer. Predicting the political alignment of twitter users. *Proceedings - 2011 IEEE International Conference on Privacy, Security, Risk and Trust and IEEE International Conference on Social Computing, PASSAT/SocialCom 2011*, pages 192–199, 2011.

[16] Andrea Ceron, Luigi Curini, Stefano M Iacus, and Giuseppe Porro. Every tweet counts? How sentiment analysis of social media can improve our knowledge of citizens' political preferences with an application to Italy and France. *New Media & Society*, 16(2):340–358, mar 2014.

[17] Brendan O'Connor, Ramnath Balasubramanyan, Bryan R. Routledge, and Noah A. Smith. From tweets to polls: Linking text sentiment to public opinion time series. *ICWSM 2010 - Proceedings of the 4th International AAAI Conference on Weblogs and Social Media*, pages 122–129, 2010.

[18] Farid Ghareh and Mohammad Saniee. A Survey of Data Mining Techniques for Steganalysis. *Recent Advances in Steganography*, 2012.

[19] Sancheng Peng, Guojun Wang, and Shui Yu. Mining mechanism of top-k influential nodes based on voting algorithm in mobile social networks. In *Proceedings - 2013 IEEE International Conference on High Performance Computing and Communications, HPCC 2013 and 2013 IEEE International Conference on Embedded and Ubiquitous Computing, EUC 2013*, pages 2194–2199, 2014.

[20] Sancheng Peng, Jian Li, and Aimin Yang. Entropy-based social influence evaluation in mobile social networks. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, volume 9528, pages 637–647. Springer Verlag, 2015.

[21] Dumitru-Clementin Cercel and Stefan Trausan-Matu. Opinion Propagation in Online Social Networks. In *Proceedings of the 4th International Conference on*

Web Intelligence, Mining and Semantics (WIMS14) - WIMS '14, pages 1–10, New York, New York, USA, 2014. ACM Press.

[22] Sancheng Peng, Aimin Yang, Lihong Cao, Shui Yu, and Dongqing Xie. Social influence modeling using information theory in mobile social networks. *Information Sciences*, 379:146–159, feb 2017.

[23] Greg Ver Steeg and Aram Galstyan. Information-theoretic measures of influence based on content dynamics. In *Proceedings of the sixth ACM international conference on Web search and data mining - WSDM '13*, page 3, New York, New York, USA, 2013. ACM Press.

[24] Saike He, Xiaolong Zheng, Daniel Zeng, Kainan Cui, Zhu Zhang, and Chuan Luo. Identifying Peer Influence in Online Social Networks Using Transfer Entropy. pages 47–61. 2013.

Ringraziamenti

Mi è doveroso dedicare questo spazio del mio elaborato alle persone che hanno contribuito, con il loro instancabile supporto, alla realizzazione dello stesso e senza le quali questo lavoro di tesi non esisterebbe nemmeno.
In primis, un ringraziamento speciale al mio relatore Prof. Emanuele Frontoni, alla mia correlatrice Dott.ssa Marina Paolanti e al Dott. Marco Mameli, per la loro immensa pazienza, per i loro indispensabili consigli e per le conoscenze trasmesse durante tutto il percorso di stesura dell'elaborato. Grazie per avermi permesso di lavorare ad un progetto avvincente e interessante come questo; a cui ho avuto la fortuna di poter partecipare e contribuire nel mio piccolo. Grazie a voi aggiungo alla triennale un nuovo capitolo ad un già ricco bagaglio di conoscenze e competenze.
Ringrazio infinitamente i miei genitori che mi hanno sempre sostenuto, appoggiando ogni mia decisione, fin dalla scelta del mio percorso di studi. Senza di voi, sicuramente non ci sarei io e soprattutto non sarei diventato la persona che sono oggi. Ringrazio anche mio fratello Davide, che dir se ne voglia, anche se a volte finiamo in cagnara, ti voglio un mondo di bene, ed ogni giorno mi ricordi quanto sia bello condividere quasi tutto con un fratello, dalle partite di calcetto a quelle alla PlayStation, passando per un'infanzia piena di ricordi insieme.
Ringrazio poi i miei zii e i miei cugini, Noemi, Nicholas, Martina, Emanuele e Maria Chiara, perchè "nu sette" insieme, ne potremmo raccontare all'infinito. Sono immensamente felice e fiero di poter vantare di una famiglia così grande e così unita, nella quale siamo cresciuti insieme. Spero qualsiasi cosa accada di non perdervi mai.
Infine ringrazio i miei nonni, anche quelli che non ci sono più, per avermi viziato e voluto bene come solo i nonni sanno fare. Siete e sarete sempre una delle cose più preziose che ho.
Un grazie di cuore ai mie colleghi Elia, Riccardo e Paolo con cui ho condiviso l'intero percorso universitario. Ho voluto scrivere colleghi perché fa molto figo, se scrivevo compagni di avventure, di esami, di disperazione e anni di fuoricorso avrei sminuito quella che è la nostra amicizia e la nostra esperienza in ambito ingegneristico. È grazie a voi che ho superato i momenti, gli esami (tranne quelli che ho fatto in Polonia) e i progetti più difficili. Senza voi e i vostri consigli, non so come e quando ce l'avrei fatta.
È arrivato il momento di scrivere due parole anche per i miei amici, che hanno continuato a sostenermi ed hanno vissuto con me quasi ogni istante di questa mia triennale. Se fate parte di questa categoria vuol dire che ormai mi conoscete, chi più chi meno, e sapete già tutto ciò che ho da dirvi, ma, soprattutto, sapete anche che

questa è l'unica parte di tutto "sto lavoro" che andrete a leggere.
Comincio con il ringraziare i ragazzi di via Costa 11. "Tommà come te lo devo dire?". Con una delle frasi più iconiche del Fabaria rally; inizio dal ragazzo con cui ho condiviso la camera per quasi tutti i miei anni universitari. Arrivato il mio secondo anno, mi sono accaparrato la mia prima matricola e non l'ho lasciata più. Sicuro di poterti avere sempre al mio fianco in ognuna delle più disparate mattate da fare o in ognuno dei momenti più seri della mia vita. Io e te non parliamo spesso di cose serie, ma quando lo facciamo, ci capiamo subito senza bisogno di spiegarci.
Rimanendo in tema cameretta come non menzionare Michele, per me Mike all'inglese. Il vero uomo di via Costa in tutto e per tutto e non solo per la maggiore esperienza e per i veri consigli di vita; ma soprattutto per le opere d'arte contemporanea che riuscivi a creare con la tua montagna di vestiti. O per la brillante idea, appoggiata ed eseguita dal sottoscritto, di dare fuoco ad un lavandino, con rischio di incendio e di esplosione dell'appartamento, per, diremmo oggi, sanificare l'ambiente. Grazie davvero Mike perché se c'è uno che ho sempre ammirato in silenzio e al quale vorrei somigliare sotto alcuni aspetti, sei tu.
L'altro storico coinquilino, dal quale sono stato ingiustamente separato per un anno sei tu Loré. O Amico che dir si voglia. Tu mi hai letteralmente cresciuto, la tua matricola, il tuo sacco da box o scientificamente parlando la tua creazione. Dai derby visti insieme, sponde rigorosamente opposte, ai dilemmi amorosi, passando per consigli e confronti politici ma senza mai abbandonare l'enorme mole di cazzate dette e fatte insieme. Spero di non perderti mai di vista e di rimanere sempre noi nonostante tutto, anche perché dove lo trovo uno con un campetto da calcetto ed un lago tutto suo? Aspetto ancora di andare a pesca, "cojone".
Ste e Michele, anche voi fate parte degli irriducibili di via Costa, della spina nel fianco al dominio incontrastato dell'Impero Galattico, parte fissa dell'Alleanza Ribelle. Se qualcono ha mai cercato di ottenere "poooooteeereeee illimitato" io ringrazio voi per l'esatto opposto; per la semplicità nel vivere e per la vostra schiettezza e sincerità nei miei confronti.
Non ringrazio invece Spiga per non aver mai creduto nella resistenza e per essere passato al lato oscuro della forza, ma si sa, ogni storia per essere avvincente necessita del buono che tradisce tutti diventando un temibile cattivo. Anche se un temibile cattivo in realtà per me non lo sei mai stato, anzi, il contrario, uno dei ragazzi più buoni e semplici che io abbia avuto la fortuna di incontrare. Grazie per avermi fatto compagnia durante il lockdown ed aver mantenuto caldo l'asse Cracovia-Tolentino; mi hai davvero dimostrato di essere un vero amico. A volte abbiamo idee e pareri contrastanti, ma sono sicuro che sai che molte cose in realtà le dico soprattutto per farti incazzare e non perché le penso davvero, o quasi. Spero di poter passare ancora parecchio tempo insieme a te, soprattutto per la miriade di cose che abbiamo in comune e che capiamo solo noi.
Ringrazio infine gli ultimi, anzi i primi di via Costa, Quartino, Lo zio Tommaso, Gioè lo goffo e Marco, primi ad accogliere una matricola all'inizio della sua avventura.

Passiamo poi ad un altro pezzo da e del 90. Genius grazie per tutto; non so da dove cominciare. Altra guida che ho avuto durante il mio percorso, baluardo ed esempio della fede. La tua capacità di riflessione sui temi più reali che la vita ti pone e la semplicità con la quale esponevi il tuo pensiero a riguardo mi hanno sempre disarmato. Grazie soprattutto per avermi sfidato e fatto diventare uno dei migliori arbitri della sezione di Ascoli Piceno, non te ne sarò mai grato abbastanza. Grazie per avermi fatto da OA personale e per avermi sempre dato i consigli giusti in campo. Con te non serve sperare, so che la distanza non è abbastanza per minare un amicizia così grande.
Ultimo degli anconetani, ma non per ordine di importanza, ringrazio Mattia. Con te so di non dover aggiungere altro, ogni parola sarebbe superflua. Solo noi sappiamo quanto abbiamo condiviso ad Ancona prima e in Polonia poi. Una spalla su cui piangere, un degno avversario a calcetto, un amico vero, ma soprattutto un ragazzo in tutto e per tutto uguale a me. Peccato anche tu abbia i tuoi difetti ma ti capisco, con il tuo carattere e il tuo cuore non potevi che innamorarti della pazza di Milano. Sicuro di poter condividere con te nuove avventure, ti aspetto per il nostro prossimo viaggio a Cracovia.
Finisco con il ringraziare tutti gli amici di Ancona, purtroppo non ho abbastanza tempo e spazio per scrivere ad ognuno di voi, altrimenti qualehe riga in più ve le avrei dedicate.
Grazie Annalisa, Martina, Mariachiara, Elisa, Chiara, Teresa per essere state le mie giocatrici per una stagione intera. È stato un piacere allenarvi e condividere del tempo insieme.
Clara, l'unica a prendere le difese della povera matricola ascolana. Grazie per i discorsi eterni in macchina insieme, grazie anche per quando dormivi o quando ti ci sentivi male. Grazie per la tua dolcezza e per la tua onestà disarmante "scojattolinni".
Giulia, grazie per essere rimasta mia amica anche se ti do sempre fastidio e se so essre davvero stronzo a volte; mi dispiace che tue tecniche non abbiano mai funzionato. Grazie per avermi fatto studiare, ma soprattutto per essere stata una vera amica in facoltà.
Sara, grazie per il tuo silenzio. Non sto scherzando, pochi sanno cosa c'è veramente dietro e a noi sta bene così. A volte vale più di mille parole, a maggior ragione se abbinato ai tuoi sguardi cattivi o alle tue temibili minacce.
Caterina, Marilia grazie per i pomeriggi di studio insieme, per il discorso su come rifondare il sistema governativo italiano e instaurare una nuova monarchia, gli scherzi, i pranzi, ma soprattutto per il vostro affetto incondizionato e gratuito.
Grazie Elena, Sara, Elisabetta, Chiara, Bea, Giulia Pia, Sofia, Anna, Annalisa e Laura a tutte le ragazze di Monte Dago per avermi permesso di condividere con voi una parte del mio cammino.
Grazie Riccardo, Paul, Matteo, Giovanni, Ferro, Francesco, Pietro, Tappo, avete tutti lasciato un segno indelebile in questi anni.
Rimanendo sul settore maschile dei ringraziamenti, volevo dedicare due parole ai nuovi

arrivati nella mia vita. Per primi impossibile non nominare i quattro moschettieri. Non sapete la gioia, di avervi avuto per sei mesi incredibili a Cracovia. Mattia, il mio coinquilino, solo noi sappiamo cosa significa vivere 6 mesi in ulica Mazowiecka piętnaście, dom trzynaście. Madonna che delirio e quanti ricordi inutile elencarli tutti. Sei stato un signor coinquilino.
Matte, il mio clone, grazie per tutto, per essere un punto di confronto fisso, visto soprattutto la tua somiglianza caratteriale. Michael o Maicol che dir si voglia, lu pescià, il primo ad incontrare sul volo alla partenza, il primo a rivedere dopo essere atterrato. Serve aggiungere altro? Compagno di allenamenti e di stempie e soprattuto ultimamente di calcetti. Non ti ringrazierò mai abbastanza per avermi invitato all'ultimo calcetto che abbiamo fatto insieme.
Siappiamo tutti che continueremo a rivederci e scriverci su WhatsApp, non ci ha fermato un lockdown figurati un pochino di distanza, prossima tappa Istambul da Osman a farci il trapianto.
Ringrazio poi l'altro settore di Cracovia, quello sportivo e quello del secondo semestre. Grazie Giorda per avermi insegnato a gioca a calcio come solo a Terni sanno fare. Grazie Simo per la gita a Danzica, per Bipino e per le tue pizzate memorabili. Grazie Riza, per le serate più assurde e soprattutto per le corse kilometriche insieme. Ringrazio poi le mie due flatmates, con cui ho condiviso il la parte più drammatica del lockdown, la reclusione nello stesso appartamento con un uomo dai versi animaleschi e dallo stile di vita di Bob l'aggiusta tutto. Che dire, grazie per le risate, i pranzi insieme, l'avermi rassicurato sul mio arrivo in piena pandemia, le cazzate e i problemi amorosi. Una soap-opera spagnola in confronto ai nostri sei mesi è niente.
Dulcis in fundo la *old school.* I due Andrei, ora anche coinquilini, miei amici da una vita, finalmente ci siamo. Prima o poi vi raggiungo tra i lavoratori sottopagati; aspettatemi ancora un pochino. Vero, Chiara, Ari e Paola grazie per essere ogni giorno testimonianza vera di una grande amicizia che è iniziata oramai dodici anni fa e che non vedrà mai la fine.
Non mi sono scordato di voi affatto. Aurora, Arianna, Chiara e Elise per ora vi ringrazio di aver condiviso insieme a me questi ultimi mesi, fatti di tanti calcetti, di molte di più cene e di innumerevoli bevute. Con la speranza che siano solo i primi di una lunga serie insieme.
Ad Maiora Semper

Ascoli Piceno, Ottobre 2020

Simone ONORI