Signed Network Construction and Sentiment Analysis Enhancement in the Twitter Social Network

Jaime Zornoza

School of Science

Thesis submitted for examination for the degree of Master of Science in Technology.

Espoo 27.6.2019

Thesis supervisor:

Prof. Aristides Gionis

Thesis advisor:

Bruno Ordozgoiti

Aalto University
School of Science

AALTO UNIVERSITY
SCHOOL OF SCIENCE

ABSTRACT OF THE
MASTER'S THESIS

Author: Jaime Zornoza

Title: Signed Network Construction and Sentiment Analysis Enhancement in the Twitter Social Network

Date: 27.6.2019 Language: English Number of pages: 4+79

Data Mining Group

Professorship: -

Supervisor: Prof. Aristides Gionis

Advisor: Bruno Ordozgoiti

The aim of this work is to build a signed network from Twitter data using sentiment analysis, analyze that network, and enhance the performance of the sentiment analysis using information extracted from it. To construct the network the interactions or tweets with mentions addressing other users will be analyzed using a variant of traditional sentiment analysis known as targeted sentiment analysis. A signed network will be built between users who post a tweet and those users who are mentioned in the tweet, predicting the relationship between such users using natural language processing on the tweet's text. After the network is built, it will be shown that using networks similar to it the performance of the sentiment analysis can be significantly improved. Using different approaches, the accuracy of the initial targeted sentiment analysis model is improved by 10%.

Keywords: Social Network Analysis, Signed Network, Sentiment Analysis, Targeted Sentiment Analysis, Graph Theory, Twitter.

Preface

Going from the specific to the more general picture, I first want to thank my supervisor Aristides Gionis and my advisor Bruno Ordozgoiti for their support, advice, and trust during these months, and for introducing me to the topic of Social Network Analysis, which I have found deeply fascinating. In this period I have learned more than I could have imagined, and for that I can only be grateful.

As this work is possibly the last milestone of my academic life, I also want to thank all the people that have been there all along the way in these past six years, supporting, and putting up with me in the difficult times, and helping me celebrate when the occasion came. To my family and friends, thanks for being by my side all this time.

Otaniemi, 24.5.2019

Jaime Zornoza

Contents

Abstract ii

Preface iii

Contents iv

1 Introduction 1

2 Background 4
2.1 Social Network Analysis 4
2.2 Graph Theory 5
2.3 Twitter research 8
2.4 Sentiment analysis 10
2.5 Machine Learning and Graph Theory 15

3 Extraction of information from raw Twitter Data 17

4 Targeted Sentiment Analysis 25
4.1 Why Targeted Sentiment Analysis is needed 25
4.2 Challenges of Sentiment Analysis on Twitter data 26
4.3 Previous targeted Sentiment Analysis approaches for Twitter data . . 29
4.4 Chosen solution overview and evaluation 31

5 Graph Construction and analysis 36
5.1 Signed graph construction 36
5.2 Signed graph analysis 37

6 Sentiment Enhancement using the Signed Graph 49
6.1 Introduction and influencing works 49
6.2 Proposed Methodologies 53
6.2.1 Heuristic Approach - Global Sentiment 53
6.2.2 Heuristic Approach - Relationships 57
6.2.3 Heuristic Approach - Combination of Global Sentiment and Relationships 62
6.2.4 Heuristic approach - Weight and Threshold explanation 63
6.2.5 Machine Learning Approach 66
6.3 Error analysis 70
6.4 Sentiment Towards Influential users 71

7 Summary, Discussion and Future lines of work 72
7.1 Summary and Discussion 72
7.2 Future lines of work 73

References 74

1 Introduction

Social Network Analysis (SNA) is the use of different techniques for investigating social structures and extracting useful information from them. It maps and measures the relationships and flows of information between people, communities and organizations, and in the last decades with the rise of many online social networks it has created a lot of research interest as it has proven to have a great role in understanding and improving social computing applications.

On the other hand, Graph Theory is the branch of mathematics that studies networks. It's strength comes from its abstraction capabilities, as it is able to model any kind of structure with pairwise relationships between objects, like computer networks, transport networks or social networks like in the specific case of this work. Social networks are one of the most clear and beautiful examples of everyday interaction with graphs. When these social networks are used to construct graphs, most times the users are represented as different nodes, and edges represent some sort of relationship or interaction in between the users. These graphs are then used to get a deeper insight into the social network and extract meaningful knowledge. Because of this, graph theory is one of the most useful tools in Social Network Analysis.

Most of the research on this area has focused on the construction of graphs with positive relationships between individuals (friends, followers, etc...), however if the interactions in these social media sources are carefully analyzed, many non-positive or negative interactions can be discovered, and incorporating these interactions on the analysis can provide a much deeper insight towards understanding the social network [1].

One reason for not taking into account negative links is the difficulty to explicitly label negative interactions in social media, as most of these applications allow users to make friends, follow, and like, but do not explicitly allow to label negative interactions with others. The main objective of this work is to create a signed network from social media by analyzing the text underlying the interactions between the different users, and infering if such text depicts a positive or negative intention towards the user that it is directed to.

Out of all the possible social networks, Twitter has been chosen for this work. Twitter is a very popular micro-blogging service where users create short messages called "tweets", that generally express opinions about different topics [2]. Over the past years Twitter has become a very popular network, and as a result a lot of work has been done to construct graphs from it and analyze its structure. However, as far as we are aware, there have not been many works that attempt to infer a signed network from Twitter. This is done by collecting tweets, extracting the ones which have mentions (a mention looks like *" I am mentioning @MentionedUser"* in the Twitter text), which are a way to address another Twitter user, and then analyzing the rest of the text to infer the sentiment towards such mentioned user.

Many important personalities like politicians, actors, and journalists, use Twitter frequently to exchange their opinions on certain topics, celebrate events, or to just keep their followers aware of their activities. Extracting who are the most influential or active users on a topic, and infering the sentiment of the rest of the users towards those influential personalities can be of great use, like shown in [3], where an over-date sentiment analysis of the three top candidates to the Indian elections is done, predicting with success the outcome of such elections. In our work we will specifically analyze the sentiment towards the most influential players of the Brexit Panorama.

For this Machine Learning techniques are used, specifically Sentiment Analysis (SA), which is a branch of Natural Language Processing (NLP) that tries to infer the sentiment of a given sentence or document. However, as the goal is to infer the sentiment towards a specific mentioned user, traditional sentiment analysis is not enough, and a more thorough process has to be carried out to achieve this classification, which is a task that is made even more challenging by the peculiarity of Twitter‘s data, as the text most times is unstructured, has misspellings, abbreviations, and slang.

Because of the difficulty of targeted sentiment analysis (even human annotators find it hard to agree, and only do so about 80% of the time) and these extra challenges derived from the intrinsic structure of Twitter text, a method to increase the quality of the sign prediction will be leveraged also, by combining the Natural Language Processing techniques with insights taken from the social structures intrinsic to the network. A heuristic approach and a Machine Learning approach will be implemented to improve the performance of the targeted sentiment analysis by itself, incorporating some previous knowledge about the sentiment towards the discussed influential users and knowledge about the relationships in the social graph.

Different Twitter data sets about popular topics will be collected, and previous to any kind of Machine Learning or Graph Making step, this raw data will be analyzed to also show how much useful information can be retrieved from it without performing any kind of exhaustive analysis.

The main parts of this Master Thesis are:

- 1. Collecting raw Twitter data and extracting information from it.
- 2. Performing a Targeted Sentiment Analysis on the data to obtain the building blocks necessary to construct the signed graph.
- 3. Signed graph construction.
- 4. Graph analysis
- 5. Sentiment Analysis Enhancement using the social relationships embedded in the graph.

In the path of reaching the final goal of this work, which was making the signed graph using the sentiment analysis, and then further trying to improve this technique by using the graph itself, various difficulties were found, which had to be analyzed and solved before being able to progress any further. These milestones will be discussed in depth in their specific sections. This project tries to prove how a signed graph can be built by carefully analyzing the text inherent to the interactions belonging to certain kind of social networks, and how these signed graphs better represent the inherent structure of such networks. Also, it will try to show how using this signed graph, the performance of the sentiment classifier can be enhanced by not only taking into consideration the natural language processing approach, but also some information extracted from the graph itself. Figure 1 depicts the Pipeline of this process for improving sentiment labels.

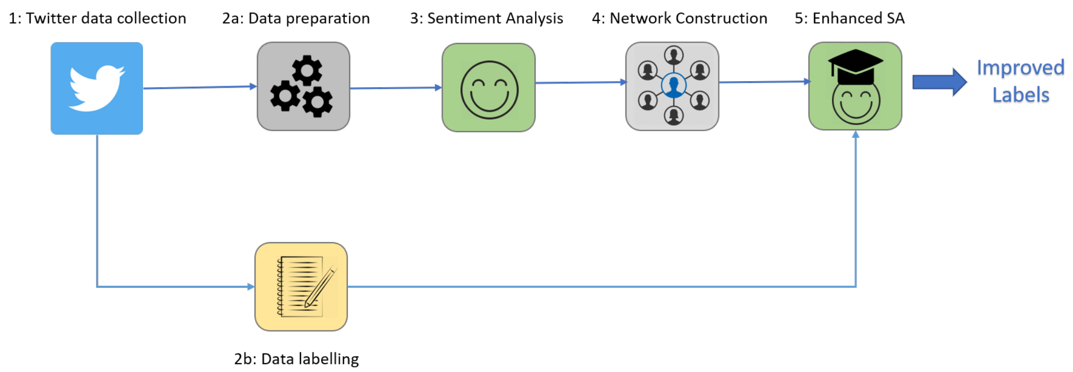

Figure 1: Pipeline of the Sentiment Analysis enhancement.

The rest of this work is organized in the following manner: chapter 2 describes the background of social network analysis, graph theory, previous construction of signed graphs from social networks, and sentiment analysis and its variants for targeted sentiment classification. Chapter 3 describes the different methods for obtaining data from Twitter, the challenges, and the information that can be obtained straight away by a very lightweight processing of this data, before any kind of Machine Learning or Graph analysis algorithms are applied to it. Chapter 4 outlines the way in which the links for the graph will be obtained: using Natural Language Processing and a target specific variant of Sentiment Analysis, while discussing some of the peculiarities that make Twitter text a challenging source of data for NLP. Chapter 5 discusses the Graph Construction process and the properties obtained from these graphs along with the normal properties of graphs extracted from social networks, as well as the different insights obtained by applying Graph Theory techniques to them. Chapter 6 discusses various ways to improve the quality of the sentiment analysis using social graphs. Lastly, chapter 7 provides a summary of the work done, and future lines of work that could be addressed.

2 Background

This work, belonging to the broad world of Social Network Analysis, mixes exploratory data analysis of Social Networks, Machine Learning, and Graph Theory. This section will consist of a small introduction to each of these fields, as well as a description of the previous related works within them that have been taken into account and used in order to reach the goal of making the signed graph and extracting information from it.

2.1 Social Network Analysis

Social Network Analysis or SNA has its roots in the 20th century, where social networks were used to connote relationships between different entities of social systems. Back in the 1930s, it emerged as one of the main building blocks of social theory, whose approach to Social Network Analysis was referred to as "sociometry" and was carried out at first among prison inmates and school students. At the same time studies were carried out at Harvard by anthropologist W. Lloyd Warner [4].

From this period until the 1970s many centers for social network research appeared, each studying different applications of social networks, and even belonging to different social science fields. However, none of these centers succeeded at providing a general approach for social science research focused on social networks.

In the 1990s researchers from other fields started joining the world of SNA, like physicists, bringing new potential into the field and making it evolve [4].

In the last decades, with the creation of the internet and the uprising of many online social applications, social network analysis has gathered more interest than ever before, as it is leveraged as one of the most powerful tools for extracting information out of this medialized world that we live in. It has evolved to englobe a broad variety of areas and disciplines related to sociology, like community analysis and detection, communication science, or social anthropology.

Social Network Analysis comes from three main insights:

1. Social relations are more important than individual attributes in understanding societies.

2. The structure of social relationships is more important than their content.

3. Social relationships can be expressed in the terms of graphs, which can then be visualized or analyzed using different the different methods and theorems of Graph Theory.

While social network theory can be readily applied in theoretical research and qualitative empirical studies, there is a general emphasis on the use of software to analyze and visualize network data once it has been collected, which probably comes from the source of that data in most cases: the internet. Because of this, social network studies involve social, mathematical, statistical and computer sciences.

Some applications of Social Network Analysis are:

1. User grouping and community detection.
2. Trend Identification in online social networks.
3. Fake news detection.
4. Assistance in marketing decision and planning [5].
5. Content and information flow and diffusion.

A Social network is built by representing social relationships or interactions in between individuals (single persons) or groups (like businesses or communities). The relationships can represent friends or followers in a social network, business relationships like partners, or authorship, co-authorship in the case of writers for example. The interactions can be messages or correspondence exchange between different users, acquisitions in the case of businesses or citations in the last case of the writers.

These ties, by nature, appear to be positive or even neutral, however negative ties can also exist, derived from hostility or conflict in interactions. As intuitive as this might seem, not many social network graphs are built using negative interactions, leading to incomplete versions of these networks, where some very valid information is not included. This deficit makes up for the main goal of this work: making a signed network from a social media site, including both positive and negative interactions.

2.2 Graph Theory

As mentioned earlier Graph Theory is the study of graphs, which represent objects and the relationships and connections in between them, using nodes and edges. In social networks most times the users are represented as nodes and the edges consist on some sort of relationship or interaction in between those users. Much work has been done in the construction and analysis of graphs from social networks like [6] where different community detection algorithms are performed, evaluated and compared on different types of graphs. specially social ones. The word community actually comes from the social context, referring to the way individuals tend to form groups. A similar work related to clustering could be [7] that studies the application of spectral clustering and visualization to graphs, showing that spectral machine learning methods can be extended to networks using specifically designed matrices.

Graph theory however, is a very powerful tool, which can be used for a lot more than community detection or clustering. Theorems of graph theory can be used in the field of nanotechnology and biology, being widely used in ADN-like structured being created, or in analyzing biological networks [8]. Graph theory is also at the heart of our navigation systems [9], and is fundamental in the functioning of our communication networks [10]. This work is focused on the application of graph theory to social networks, which is an area of research that has gained a lot of interest in the past years, leading to a lot of investigation efforts.

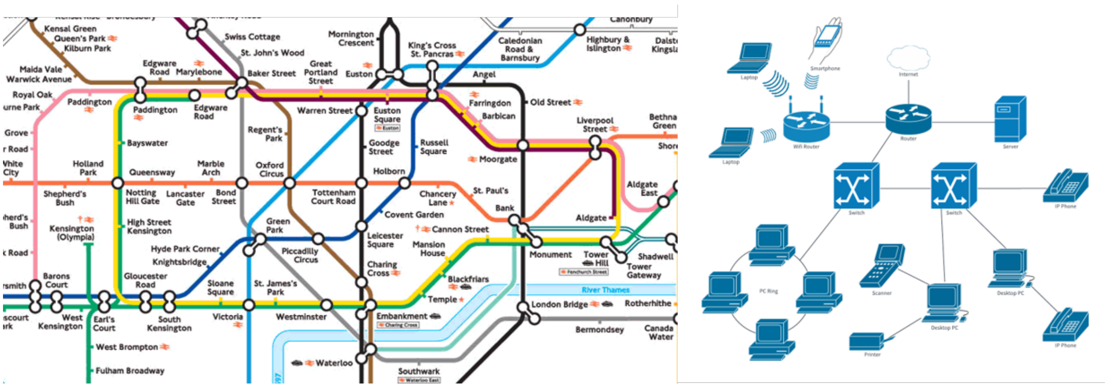

Figure 2: Picture of different day to day networks: a transportation network where the nodes are metro stations and the connections are the paths in between them, and a computer network where the nodes are the different devices.

The goal to build a signed graph comes not only from the will to fully represent the social relationships within an online social network, where negative interactions are fundamental, but also to the theoretical possibilities that constructing a signed network gives from the graph theory point of view. Building a signed graph allows us to apply new theories which could not be used in a normal weighted positive graph. The main theories that we can use in this manner are the Social Balance Theory, first proposed by Fritz Heider [11], and Status Theory [12].

Heider's basic hypothesis is that there is a tendency to achieve balanced states in a network where the relationships or interactions are depicted as signed edges, and that there is a pressure towards reaching such states. *"If no balanced state exists, then forces towards this state will arise"*. Works done around this theory indicate that the tendency towards balance is a significant factor of cognitive organization, and that it is also important in interpersonal relations. In his later work Hararay incorporates mathematical and graph notions to Heider's theory, introducing the concept of signed graph or *s-graphs*, which can have positive and negative signs and therefore can be have balance theory applied to its triads or cycles. A graph is balanced when all of its cycles are positive, and cycles are positive when the product of all its edges is positive. This balance theory will later be applied in the signed graph obtained from Twitter data and is a very powerful tool, that can be used, for example, for sign prediction.

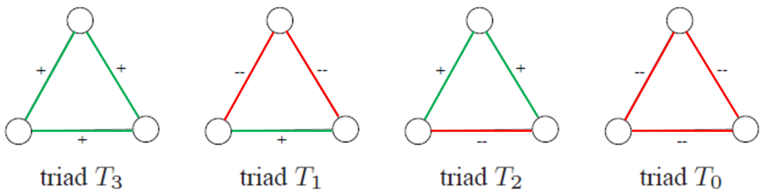

Figure 3: Signed Triads. Based on the number of positive edges the triads are *balanced* (T3,T1) or *unbalanced* (T2,T0).

While Balance Theory is predominantly defined for un-directed networks, Status Theory is defined for directed ones. This theory suggests that nodes with a low status will have a positive link towards nodes with a higher status and vice-versa. In directed networks there exist two types of triads: cyclic, where all the links follow the same direction, and un-cyclic, when they don't. In a signed network, combining edge sign with its direction leads to four types of cyclic triads and eight different types of un-cyclic triads.

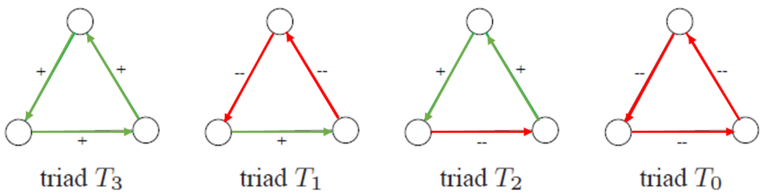

Figure 4: The four possible cyclic triads in a directed signed network.

This theory, like Balance Theory, can also be used for sign prediction in signed networks, but in this case when the edges in these networks have direction. Balance theory can also be applied to directed networks by not taking into consideration the direction of the edges

Previous works have been done in the literature of SNA about building signed graphs from social network applications like [13], that also apply Balance Theory to the resulting networks, however, these works infer the negative signs from applications where a negative interaction can be explicitly labelled by giving likes and dislikes, or bad ratings or reviews. These applications are Wikipedia admins site, Slashdot (which allows to tag other users as friends or foes) and Epinions product review site. Epinions is a very popular site for research on signed networks, like shown in [14], where the authors try to predict signed ties between users starting with an unsigned network and some additional information from the application, and [15], which uses Epinions to predict missing links on the user network. While the resulting networks are robust and well built, using only these kind of platforms greatly limits the applications from where signed social networks can be extracted, derived from the lack of a explicit notion of negative relations in most social applications.

Despite of this, some work has been done trying to extract negative relationships from social applications without any negative interaction inference mechanism. The way to do this is by analyzing the text that is intrinsic to most social applications, which amounts for a large portion of the interactions on them, and using Natural Language Processing techniques to extract the sentiment of these interactions and therefore the positive or negative emotions in between the users taking part in them. In [1] threads from discussion forums between two individuals are manually annotated using Amazon Mechanial Turk, and analyzed in order to infer the sentiment between the two users who are engaging in the discussion. It does this by first identifying polarized words to see if the sentence shows any attitude, and then training a classifier using lexical items, polarity tags of words, and POS tags to classify the sentiment of the interaction. Once this sentiment is obtained, the network is built by connecting the individuals based on the sentiment of their interactions. Once the network is built Balance Theory is applied to it and compared to what is obtained in a random network, showing that the network built is highly complaint with Social Balance Theory. Also, [16] creates a signed network by using sentiment analysis on user posts, dividing the users into five simple categories: Highly Positive, Positive, Neutral, Negative and Highly Negative. However, as mentioned before, Twitter presents some very peculiar characteristics that make the sentiment analysis of the text very challenging. Also it is not clear how to create a signed network from Twitter, as we have users that are linked via retweets, follower/followee relationships or mentions, leading to a wide variety of options to represent the edges of the graph.

2.3 Twitter research

Twitter is a very popular microblogging platform that was created in 2006 and has it's main headquarters in California. It's functioning is very simple: users post text messages of up to 140 characters (although this limit can be extended to 280), alongside with URLs, images, videos or GIFs. Inside this text the user can add keywords using hashtags (a hashtag looks like: *#Hashtag*) that relate to a relevant topic or event, and they can mention other users from the network (a mention looks like *@mention*) to address them. The way the content shows up to the users is either by a user being mentioned by another user so a notification appears on the firsts users interface, or by tweets from related users of the first one showing up on his time-line. The way these relationships are established is via a follower/followee mechanism, in which users can follow any other user whose tweets they want to receive in their feed. Generally users can not put restrictions on the accounts that follow their profile, so anybody can follow anybody without any kind of request, like it is needed in other popular social networks like Facebook. This fact, and the relaxed nature of the content (short sentences discussing a certain topic or event don't require as much work to post as content on other kinds of social networks) has made Twitter explode in popularity, reaching 300 Million users in 2016 [17].

Twitter has become over the last years one of the most important news sources for collecting quick hints of information. As users unambiguously post their opinions or feelings directly in the platform, Twitter is a great way of receiving information from journalists, celebrities or politicians, and getting an insight of what they are saying themselves, instead of through the delay/filter that's intrinsic to some other kind of news sources like newspapers. Due to this, and the increasing popularity of the platform, Twitter has gained a lot research interest in the last decade, leading up to a wide variety of published works. Twitter data being publicly available, unlike in many other social networks is another of the reasons why Twitter has attracted such a high volume of research interest.

Despite the richness of the interactions in the Twitter network, it is true that by only following those users that are similar to us we are very proclive to get stuck inside a so called *filter bubble* or an *echo chamber*, where the same kind of information gets repeated and reinforced, having information from sources with different points of views or ideologies weighted down or discarded. Because of this there has been a lot of research work oriented towards homogenizing the information that reaches users in social networks (specially Twitter) like [18], where certain topics are picked from the content published on Twitter, characterizing the different viewpoints on these topics by classifying them onto *campaigns*. It then leverages information propagation techniques to be able to pick new users from opposing campaigns in order to maximize the number of other users for which that topic is relevant which are exposed to both campaigns, and therefore not anchored in their own filter bubble. Following this line of research, works have also been done to explore when a topic can be regarded as *"controversial"* like [19] where the *conversation graph* on a certain topic is studied, as this graph for a controversial topic should have a clustered structure, derived from the fact that individuals with the same point of view on a certain topic tend to endorse and amplify each others arguments, creating a dense network of interactions. It also proposes and describes different metrics for evaluating if a topic is controversial in a quantitative manner.

In [20] the authors model user behaviour in Twitter to capture the emergence of new trending topics by investigating the retweet graph, where the nodes are the users making retweets about a certain topics and the original users who posted those tweets, and an edge is created between two nodes if one of them has retweeted a tweet from the other one. Their findings show that the retweet graph for a new arising trending topic tends to have a relatively dense Largest Connected Component (LCC). There are also many works that try to detect influential users in the Twitter network and quantifying the influence of each of them. [21], proposes a set of metrics to quantify influential users in Twitter, that range from the well known centrality to measures obtained using the Twitter API, to metrics derived from complex mathematical models. This research will be relevant to the work presented in this project, as it will later be exposed.

Twitter has also been used to study the user behaviour during relevant events like political campaigns or elections, like in [22], where a large and complex dataset of users who participated actively in this topic is collected and analyzed. Also in [3] the authors build a graph from Twitter data collected around the topic of the presidential elections in Pakistan and India, infer the sentiment on a certain target-topic pair (being the target a candidate for the presidential elections) and use diffusion models to learn how support for each candidate spreads through the network, calculating the outcome of the elections for each candidate using a sentiment score and successfully predicting who would be the winner.

In line with one of the main topics on this project, which is Twitter Sentiment Classification, various recent works have studied the detection of harassment or bullying on Twitter. Automatic abuse detection is a hot topic right now for most social media platforms, as to keep their users from actively harassing each other, if mechanisms like these were implemented those users that are heavily disrespectful towards others could be automatically banned. [23] uses a traditional Machine Learning approach to detect hate speech on Twitter using publicly available Twitter datasets, and reporting the best performance using n-grams from 1 to 3 units long and Logistic Regression with Tf-Idf features. [24] uses a similar approach but this time manually collecting the tweets and labelling 25000 of them. The performance in this case achieves accuracies of more than 90%, which is a very impressive result for a text classification task of this kind.

To end this sub part of Twitter research, works that are more closely related to this one will be superficially discussed. [25] takes a granular approach to sentiment analysis, performing an overall sentiment analysis of the users that mention a certain technological account (this could be used for example to infer if the feeling towards a new product or marketing campaign is positive or negative) and then performs this same sentiment analysis in a more granular way by dividing the users in the initial set into different communities and performing the per-community sentiment analysis for the content created over a certain period of time. By doing this they obtain the evolution of the overall sentiment towards the targeted user and also the evolution of the sentiment of individual communities towards that account. [26] studies the structure and characteristics of different kind of Twitter networks like the previously commented Retweet Network, or the follower/followee network. There is also an important battery of works that focus on performing sentiment analysis on Twitter, however most of them solely rely on a Machine Learning approach, analyzing the plain text intrinsic to the twitter interactions but not with the goal to build any kind of network. These kind of works will be described in the next sub section.

2.4 Sentiment analysis

Sentiment Analysis (SA) is a specific branch of Natural Language Processing (NLP), which in turn is a branch of Machine Learning (ML). Its goal is to given a certain text as an input, classify such text as having a positive, negative or neutral feeling or emotion. At its start, sentiment analysis was regarded like another text classification task, of the sort of topic classification or spam detection, however the initial works showed that sentiment analysis required a more thorough approach than these other tasks [27]. This is derived from a lot of facts, like the use of sarcasm, vague sentiments or nonexistent sentiment, and the difference between polarized and depolarized expressions. As sentiment analysis started being regarded as its own NLP task, the results of the works with the goal of performing this task on different kind of texts started increasing in performance as more and more techniques were proposed and evaluated.

These first works showed that the *dictionary* approach, where certain keywords are looked for in a sentence in order to classify it as positive or negative, was not enough for accurate sentiment classification. A very well known example sentence that reflects this is:

“Jane Austen’s book madden me so that I can’t conceal my frenzy from the reader”

While the words in this sentence like *frenzy* or *madden* could express a negative sentiment, by reading the whole sentence we can see that it shows a positive feeling towards Jane Austen and her work. It is because of this kind of problems that Machine Learning was leveraged as a solution to the Sentiment Analysis task, resulting in a wide variety of works, performed in all kinds of text, like movie reviews, articles or Tweets.

One of the main issues that makes sentiment analysis so challenging is inferring if the sentence actually reflects any kind of opinion, that is to say, if its subjective or not. Subjective sentences are relevant and present user’s attitude, view, or belief, while objective sentences are irrelevant and present information which is factual. In most sentiment analysis models this "objective" category is incorporated by adding a third classification output: 1 or positive, -1 or negative and 0 or neutral/non opinionated.

The Machine Learning model that is generally used as a baseline for sentiment analysis due to the trade-off between performance and simplicity is Naive Bayes. It is a probabilistic model that comes in various shapes (mainly Multinomial and Bernuilli for the text classification task) and that makes an independency assumption from Bayes Theorem to make it computationally effective while still keeping its fundamentals and performance [29]. These models calculate the probability of a sentence belonging to a certain class (positive, negative or neutral in the case of sentiment analysis) as the product of the prior probability of such class times the joint probability of the document and the class, which in the end, taking the document as a sequence of independent words (due to the independence hypothesis previously mentioned) its just the prior probability of the class times the probability of each word in the sentence belonging to such class. After computing this for all the classes, the given document will be assigned to the class with the highest probability score.

Mathematically this is expressed as:

$$P(c|d) = \frac{P(d|c)P(c)}{P(d)} \tag{1}$$

Where the term on the left is the probability of a class given a certain document, which is calculated for all the classes, the numerator on the right is the probability of the document given the class, which can later be expressed as the product of the probabilities of each of its words belonging to such class, and the prior probability of the class. The denominator is called the predictor probability, and most of the times it is cancelled out, as it is the same for all the classes.

More complex and powerful models that can be used for sentiment analysis are Logistic Regression and Support Vector Machines, like in [30] where SVMs combined with feature selection techniques are used to classify movie reviews as positive or negative achieving better results than the traditional Naive Bayes approach performed in the same dataset.

Support vector machines or SVMs are supervised machine learning models that are mostly used for classification, and can perform non-linear classification by the use or kernels, mapping their non-separable linear inputs to a higher dimensional space where they can be effectively separated. They are the last evolution of the maximal margin classifier, which is a classifier that in a p dimensional spaces tries to create a hyper-plane that can perfectly separate the differently labeled data instances. The equation of such hyper-plane has the following structure:

$$b0 + b1X1 + b2X2 + ...bpXp = 0 \tag{2}$$

After we get a new data point (X) and input its coordinates into this equation, the sign obtained (positive or negative) tells us in which side of the hyper-plane the point lies, and therefore what class it belongs to. Out of all the possible hyper-planes, the optimal one is the one that maximizes the margin to the closest training data point, that is, assuming we have many hyper-planes which perfectly separate our data, we take the one that is the furthest away from the training observations. This is a very natural way of performing classification if a separating hyper-plane exists, however most times this is not the case, and the different classes in which we want to classify our data can not be perfectly separated. This issue leads to the next step in the improvement of the maximal margin classifier, which is called a support vector classifier or soft-margin classifier. This classifier allows a certain number of training observations to violate the margin or even the hyper plane when choosing this hyper plane, in order to perform better in the future when new data is evaluated and to prevent over-fitting.

In the case that the data can't be separated at all in a linear manner, there is one last turn around to these methods, which is the use of kernels. This way we arrive at the previously mentioned Support Vector Machines, which are an extension of the support vector classifier that results from enlarging the feature space in a specific way to accommodate non-linear boundaries, by the use of *kernels*. These kernels or kernel functions map our low dimension feature space to a high dimensional feature space where the data points can be separated by a hyper-plane. The following figure gives a very clear representation of this mapping.

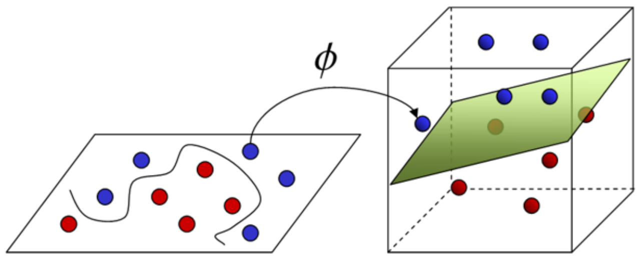

Figure 5: Mapping from a low dimensional initial space to a high dimension feature space where our data points can be separated by a hyper-plane.

Naive Bayes was introduced here because it is the most basic Sentiment Analysis approach, and SVMs because they will be periodically mentioned throughout the text. Logistic regression and different types of Neural Networks will be explained in the following chapters, when we talk about their application in the particular use case of this work.

The area of sentiment analysis on Twitter has been largely explored, reporting a large amount of results that bring to evidence how Twitter's text peculiar structure makes sentiment analysis even more challenging. For traditional Machine Learning methods of sentiment analysis, the features that are fed to the classifier involve vector representations of the words, part of speech tags, and information derived from syntactical parsers. In a properly structured sentence, most of the words would have a vector representation derived from the vocabulary that the NLP model was trained with. Also, the words would easily be recognized as nouns, verbs, adjectives or adverbs. Finally, using an average performance syntactical parser, the relationships between the different words could be easily obtained. In Twitter however because of abbreviations, slang, hashtags and non-coherent sentence structures, these assumptions do not always hold true. Consider the sentence:

"btw if u think that #PeoplesVote, would mak a diference now, stop thinking: it wont #Brexit"

The abbreviations (*btw*), misspellings (*mak* and *diference*) and the hashtags being embedded in the sentence and used as words (*#PeoplesVote*) make it very hard for the previously discussed features to be efficient in a Twitter Sentiment Analyzer. These characteristics will be futher explored in the chapter dedicated to Targeted Sentiment Analysis. Because of them, the State Of the Art techniques for Sentiment

Analysis on Twitter, use Neural Networks and Word embeddings to automatically learn features for this task. Word embeddings are fixed length n vector representations of each word, built in a way so that similar words or words with similar meanings have a close vector representation in the n-dimensional space where the vectors are represented. The classical example of how word embeddings work is represented in the following image:

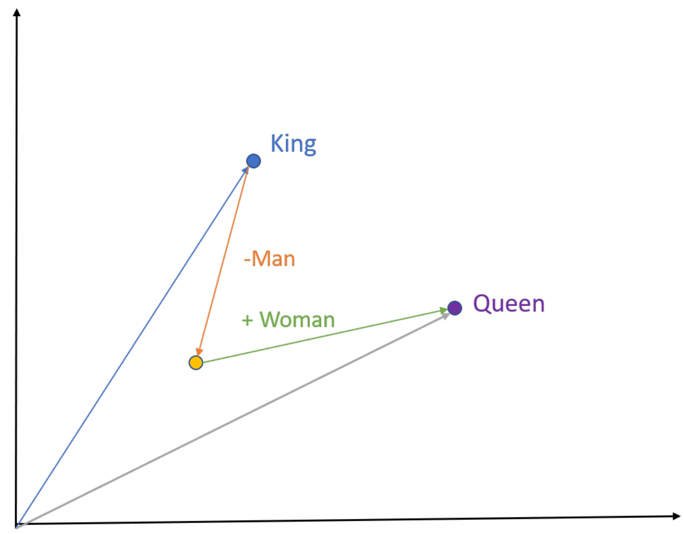

Figure 6: If from the vector representation of the word *King* we subtract the vector of *Man* and add the vector of *Woman*, we are left with the vector representation of *Queen*.

There are many different kinds of word embeddings, being some of the most known GloVe from Stanford University [31] and the Word2vec models [32]. These representations have greatly contributed to increase the performance of just about any text classification task, and specially sentiment analysis. More about them will be discussed in the chapter dedicated to this topic.

For the goal of constructing a signed social graph from Twitter interactions however, normal sentiment analysis is not enough, as we do not specifically want the sentiment of a whole sentence, but rather the sentiment of that sentence with regards to a specific target (as the way that users have to address other users in the Twitter text is by mentioning them with *@mention*). Because of this, a branch of sentiment analysis denominated *Targeted Sentiment Analysis* will be used. In this variant the inputs to the classifier are a sentence, and a target entity within that sentence, and the output like before is a sentiment (positive, negative or neutral) but in this case not concerning the whole sentence but rather the sentiment towards the user. Consider the sentence:

"I have to say, the Oscars ceremony was horrible, lousy presenters, boring commentators and edgy nominations, however @emiliaclarke looked so amazingly beutiful <3 #Oscars"

In this sentence the overall sentiment is negative, however the sentiment towards @emiliaclarke (Emilia Clarke from Game of Thrones) is positive. We would want a classifier that can differentiate these two types of classification and that given a certain target it will output the sentiment towards that target, and not the overall sentiment of the sentence. Different algorithms for this kind of task will also be discussed in the chapter dedicated to Targeted Sentiment Analysis.

2.5 Machine Learning and Graph Theory

By now the individual backgrounds of sentiment analysis and of Graph Theory have been explored, seeing how useful each of them can be if used separately. However, there are some works in the literature that discuss the combination of both techniques to achieve results with surpass the ones obtained by applying any of them individually. [33] combines Support Vector Machines and the graph-like structure inherent to proteins to propose new algorithms for protein structure prediction. [34] discusses the different methods (supervised and unsupervised) that use graphs for learning, with the posterior purpose or clustering or classification.

More relevant to this project are various works that explore the mixture and combination of graph theory and Sentiment analysis, in order to achieve results that outperform any of the individual techniques. Balance Theory, which has been introduced before, is a well known method for predicting non existing edges on a network by completing triangles with edges of signs that make them comply with balance theory, like in [35]. Also, sentiment analysis can be used to predict the sign of the interactions in between users of a network and infer the edge in between them based on the output of this classification, like in [1].

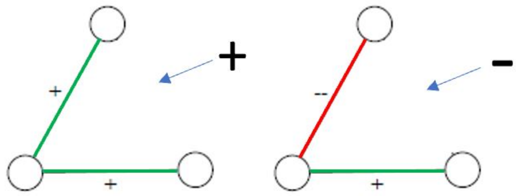

Figure 7: Sign prediction using balance theory. For the triangle on the left the missing edge should be positive in order to comply with balance theory, and negative for the one on the right.

This previous work is similar to what this project wants to achieve, however the interactions analyzed belong to a messaging exchange platform, where the sender and receive of the messages are intrinsic to the application, meaning that straightforward sentiment analysis can be used, without the need for the targeted version. In [36] sentiment analysis and balance theory are used to predict the edges of a network (the Wikipedia admin network) and then the final set of edges is chosen such that the error of incorrect edges is minimized, by dividing such error into a term that comes from the sentiment's analysis confidence and a term that comes from the number of non balanced triads in the network. The results show that by combining sentiment analysis with social insights and balance theory the performance of the edge prediction is increased from the performance of each of the individual methods.

In [37], acknowledging that the task of sentiment classification for Twitter presents many challenges, a method to enhance the performance of sentiment analyzers is proposed by taking advantage of the rich social structure in which the Twitter text is embedded. By making some assumptions about the patterns in which users react to specific target-topic pair opinions from other users in the same social network, a method is build to improve the performance of the plain sentiment analysis by using sentiment aggregates of target-topic sentiment, and user-tweet aggregates.

Lastly in [38] the follower/followee and mentions network (built by connecting the users who mention somebody in the tweet text to that user who is being mentioned) are used together with standard sentiment analysis to predict new edges in a Twitter network for a certain topic. First a set of gold-label users with a very strong feeling towards the topic are chosen and labeled as positive or negative towards that topic. New tweets on that topic are then collected, and a traditional sentiment analysis technique is used on them to determine the first tier sentiment. However, if the user who is making the new tweet follows any of the users from the gold-label set, the sentiment of their tweets is changed to whatever sentiment the user from the gold set has, independent of the sentiment derived from the Machine Learning approach. This approach is backed up by the *Homophily* concept, which is is the tendency of individuals to associate and bond with similar others. Because of this, the authors assume that if a user follows another one whose position on a certain topic is known, the first user will have the same opinion about such topic as the second one.

These three previous works are the main sources of the inspiration for the sentiment analysis improvement proposed in this project, so they will be covered more in depth in chapter 6. By mentioning them we finish describing the background of the previous works that have inspired this one, and therefore continue to the first chapter of practical work: Analyzing how raw Twitter data, with no machine Learning or Graph analysis can already provide a large amount of useful insights and information.

3 Extraction of information from raw Twitter Data

Twitter is a social network created in 2006 that unites millions of users around the new and popular concept of microblogging, allowing them to post text messages up to 280 characters long that can include images, urls or other kinds of content. Users also can "follow" other users without mandatory reciprocity [39]. Its laid back and simple principals have made it a very popular platform to share and consume information about current events.

This kind of blogging encourages a faster mode of communication, lowering the user's requirements of time and thought investment for content generation, thus increasing the frequency with which post are created, having microbloggers post many different updates in the same day.

The popularity of Twitter, along with the peculiar characteristics of the social network underlying the application, have made it a very powerful tool in a wide range of tasks, from trend modelling to marketing or even predicting outcomes of presidential elections [3]. These characteristics have also awakened a lot of research interest, leading a large corpus of work being dedicated to analyzing the Twitter graph structure [40] [20] [41] community detection on Twitter [42], or just plain analysis of the inherent properties of the application [43].

All of this work is also motivated by the good quality of service provided by the Twitter APIs, that despite having some limitations, let users download almost all the information regarding the application that can be of some interest. There are two Twitter APIs that can be used, the Streaming API, for downloading content that is being produced in real time, allowing for users to filter the requested tweets to contain specific words or hashtag, come from specific locations or from certain accounts, and the Twitter REST API, that allows users to download historical data like user timelines, retweeters of a certain tweet, or followers and friends of a certain account [44].

In this work, the Twitter Streaming API was used to collect data on four different popular topics. These topics are:

- The Oscars, that happened on the 25th of February. Data was collected from a week before and for a week after. In this period 1.6 Million Tweets were collected.
- Brexit, which had a very busy period from the 11th to the 29th of March. Almost 900.000 tweets were collected for this topic.
- Venezuela, which was also a very active topic constructed a dataset of 327557 tweets.
- Game of Thrones, which last seasons premier episode aired on the 14th of april. For this dataset data was collected from the day before the episode to one day

after, collecting 72312 tweets.

To do this, the Streaming API was queried in the planned dates for Tweets containing the following hashtags: #Brexit or #brexit for the Brexit dataset, #Oscars and #oscars, for the Oscars dataset, #Venezuela for the Venezuela dataset, and #GOT, #GoT, #Gameofthrones and #GameOfThrones for the Game of Thrones dataset. To expand the size of the datasets, and not limit the tweets to these specific hashtags, a script was built to dynamically increase the list of hashtags that were being queried, so that taking into account that tweets most times include various related hashtags, for every 5000 tweets, a list was built with each hashtag collected and the number of times it had appeared, and then the top hashtag that had not previously been included in the queried hashtag list was added.

In this manner we end up with hashtags like #PeoplesVote on the Brexit dataset being the second most used hashtag, ahead of #brexit, which was included in the initial list of queried hashtags, and also enlarged the mentioned datasets.

The Twitter Streaming API returns a JSON object which contains a lot of valuable information for later analysis like the tweets text, user who produced the tweet, a timestamp, source, location, if its a retweet or not, and much more. The goal of this chapter is to exploit this information, without any kind of Machine Learning or Graph Theory techniques, to see the large amount of useful insights that can be taken from Twitter data with a very straightforward analysis.

The following figure depicts part of the structure of the JSON object returned by the Twitter Streaming API.


```
{
 "created_at" :  "Thu Apr 06 15:24:15 +0000 2017" ,
 "id_str" :  "850006245121695744" ,
 "text" :  "1\/ Today we\u2019re sharing our vision for the future of the Twitter API platform!\nhttps:\/\/t.co\/XweGngmx
 "user" :  {
   "id" :  2244994945 ,
   "name" :  "Twitter Dev" ,
   "screen_name" :  "TwitterDev" ,
   "location" :  "Internet" ,
   "url" :  "https:\/\/dev.twitter.com\/" ,
   "description" :  "Your official source for Twitter Platform news, updates & events. Need technical help? Visit https:\
 } ,
```


Figure 8: Figure showing part of the JSON object returned by the Twitter Streaming API

Using Tweepy [46], a Python Library for interfacing both Twitter APIs, and Pandas, Matplotlib and other well known Python Libraries, it is very easy to get a lot of valid information out of this JSON object without any complex analysis.

First we will explore the different types of Tweets that are returned by the Streaming API. These Tweets can be grouped into 3 main categories: Retweets, which are tweets posted by one user that are forwarded by another user, tweets with mentions, which are a way to address another twitter account using their username, and tweets that are just plain text. The following are examples of each of these, in the format returned by the Twitter Streaming API:

1 *"RT @TIME: Here are all the moments that got fans shipping Brienne of Tarth and Tormund Giantsbane on #GameofThrones https:t.co6wTUBwUGvq"*

2 *"Here it is! The #Brexit Party. And @Nigel_Farage is back!"*

3 *"Protecting against the risk of #Brexit in your construction contract #construction https://t.co/myGYDHfe5ZX"*

Tweet one is an example of a retweet. These tweets are provided by the Streaming API with the format *RT @OriginalAccount: text* in their twitter text, where *@OriginalAccount* is the initial account that posted the tweet, and to see the user who actually retweeted it we would have to look at the JSON object "user" category. Tweet 2 is a tweet that contains a mentioned towards the user @Nigel_Farage, and Tweet 3 is a normal plain text tweet. The Figure 10 shows the percentages of each tweet tweet type for the different collected datasets.

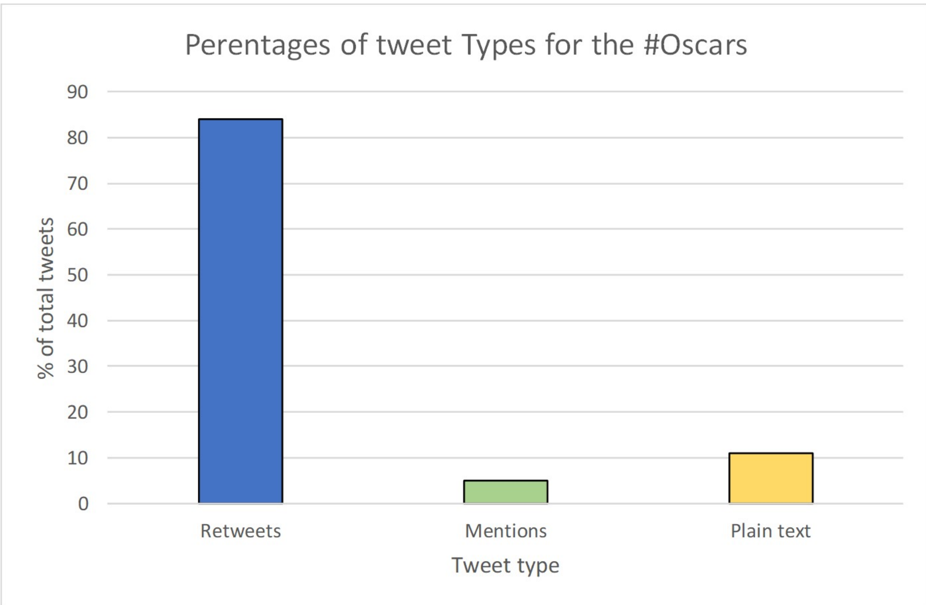

Figure 9: Figure showing the different tweet types for the #Oscars dataset: Retweets, mentions and plain text tweets

The percentage of retweets in comparison to the rest of the other types of tweets shows that most people retweet posts from other users rather than creating their own content. For example, in the Oscars dataset, there are 1655669 collected tweets, but out of these there are only 317057 different ones, reflecting that there are a lot of retweets (different users retweeting the same tweet). The following figure depicts each type of tweets for all the datasets: From looking at this table, it looks like the percentage of retweets in the topics that are not politically related (#Oscars and #GoT) is larger than in the ones that are politicaally related (#Brexit and

Oscars		
Retweets	Mentions	Plain text
84%	5%	11%
Brexit		
Retweets	Mentions	Plain text
74%	10%	16%
GoT		
Retweets	Mentions	Plain text
82%	4%	14%
Venezuela		
Retweets	Mentions	Plain text
78%	10%	12%

Figure 10: Figure showing the different tweet types for each dataset: Retweets, mentions and plain text tweets

#Venezuela). This could mean that for this last category of topics, users forward less messages and create more of their own content, sharing their personal opinion than in the "amusement" topics like TV Shows or events, where they Retweet more of the content that is being published by other accounts.

From this raw Twitter data, using text matching patterns or regular expressions, we can extract who are the most mentioned Twitter accounts from the different datasets. The following image depicts a bar chart and a wordcloud representation of the most mentioned accounts in the Oscars dataset:

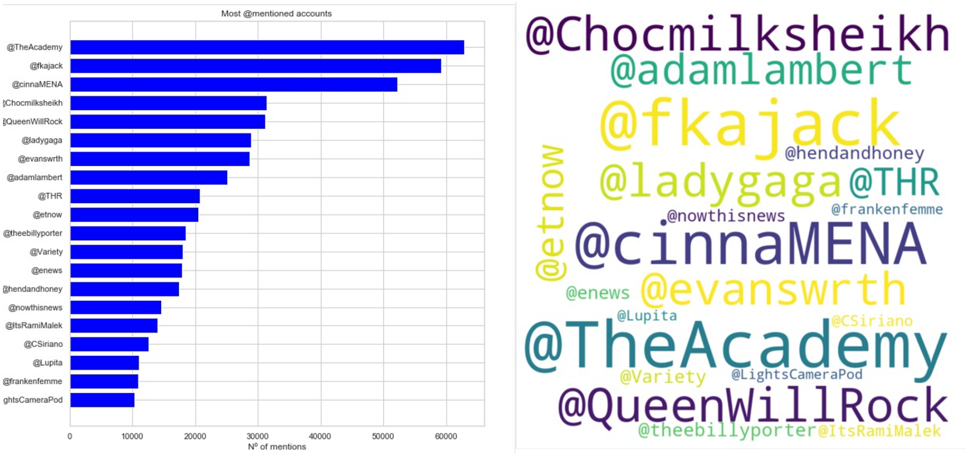

Figure 11: Bar chart and wordcloud representation of the most mentioned accounts in the Twitter Oscars Dataset

This figure shows that the most mentioned account in the #Oscars topic was theacademy, which is the official account for the Oscars. By doing this we can see who are the most popular subjects on a certain topic.

For the Venezuela dataset, the previous figure looks like:

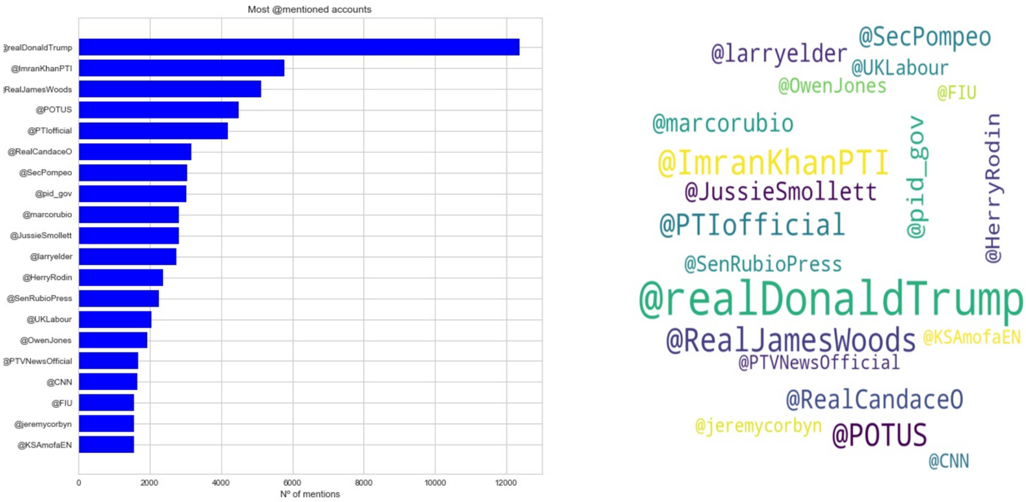

Figure 12: Bar chart and wordcloud representation of the most mentioned accounts in the Twitter Venezuela Dataset

Its interesting to see from this dataset that most of the mentioned accounts are related to politics and that there are actually many accounts shared with the #Brexit dataset, that come from British politics like *@UKLabour*, *@jeremycorbyn* or *@Realjameswoods*. Also, in both of these datasets with political taste *@RealDonaldTrump*, the official account for US president Donald Trump, is one of the most mentioned accounts.

The same thing can be done with the hashtags included in the tweets of each dataset, to see find popular trends or related topics to the ones that are being specifically searched. In Figure 13 we can see that the top posted hashtag is #Oscars, which is the main hashtag for the event, but also by only analyzing this image we can see for example that the hashtags #BackPanther and #BohemianRhapsody, that are hashtags representing the movies *Black Panther* and *Bohemian Rhaposody* are present in the top 15, illustrating how these were the movies that were given awards and therefore created the most discussion.

Another useful insight that can be obtained, is who are the most Tweeting users for a certain topic. The following list depicts the users of the Oscars who have produced the most tweets:

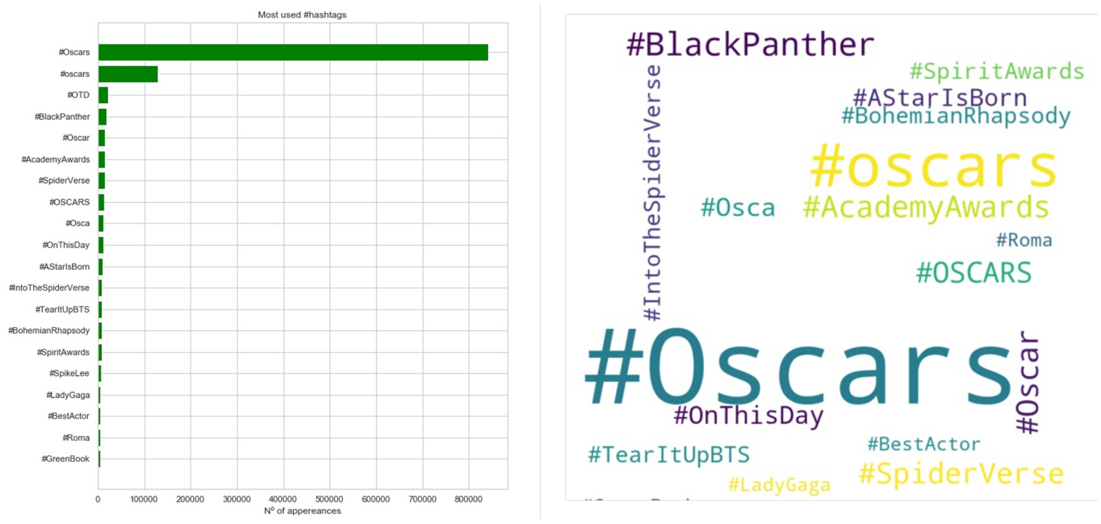

Figure 13: Bar chart and wordcloud representation of the most used hashtags in the Twitter Oscars Dataset

User	**Tweets**
gadbadshetty	3343
bbluesky922	647
MMandOscar	638
THR	592
ronniehowlett3	574
dodleee6doo	540
MRoopchand143	472
CelebBlogNews1	442
nolimonfoo	435

It is even possible to analyze, among these top tweeting users, if any of them are Bots that are automatically posting tweets. This can be done using the Botometer API [45], or their online website. Using this API, out of these users the user *@gadbadshetty* gets a chance of 87% of being a bot, and *@CelebBlogNews1* of 88%, suggesting that they are possibly bots. All the other accounts get scores under 30%.

Another interesting insight that can be discovered from this data is the amount of Twitter traffic per day on a certain topic. In this way we can see the evolution of tweet production inline with events regarding the topic. Figure 14 shows the time series of tweet production for the #Oscars topic. From this image we can see that in the days previous to the Oscars ceremony the traffic starts going up, to reach its maximum in the night of the event, the day after there is still many tweets about it, and then the number of tweets produced gradually goes down.

The Twitter Streaming API also tells us the device where the tweet was posted from, if its known. The Figure 15 shows the number of posted tweets from different categories of devices for the #GameOfThrones dataset. In this figure web refers to

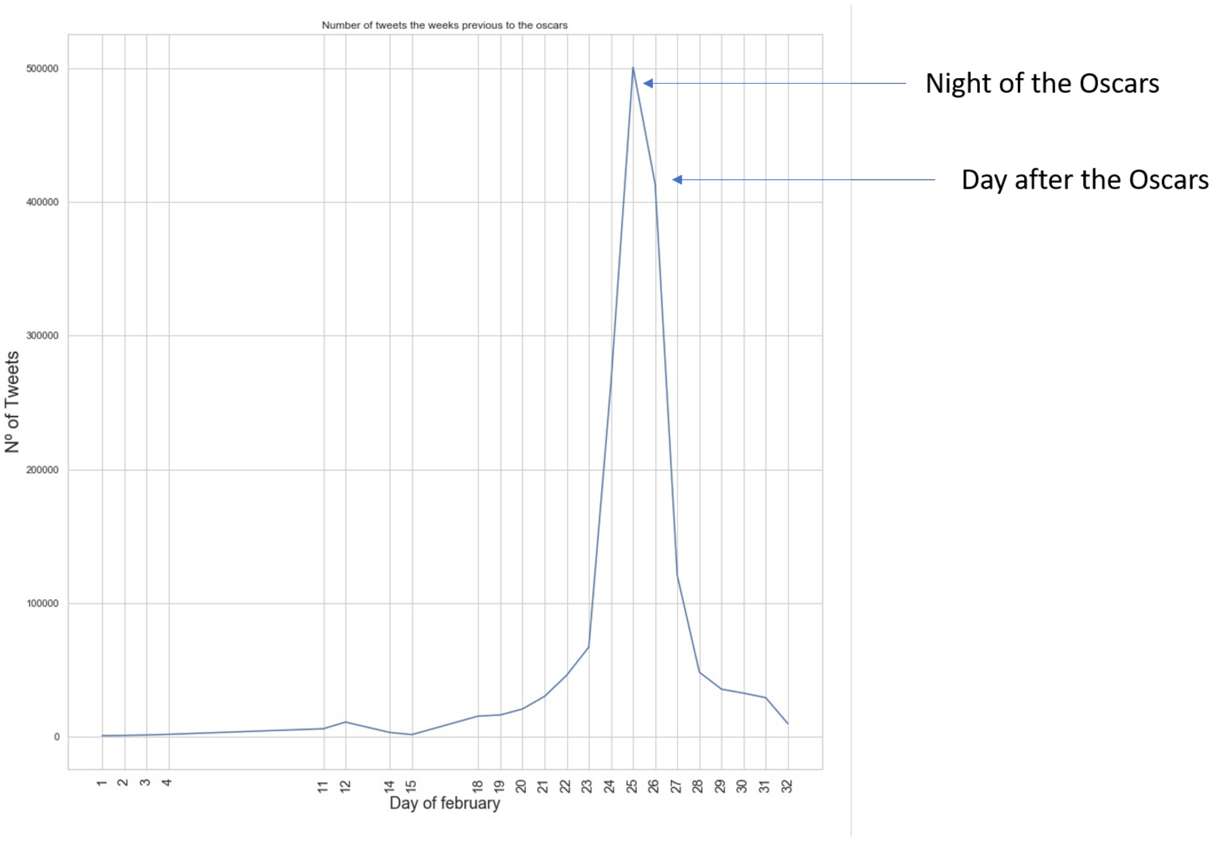

Figure 14: Evolution of the tweets produced for different days of the Oscars topic collection process.

the Twitter official website, and iphone and android to the applications installed in either kind of device. As we can see most tweets are posted from Smartphones, specifically Iphone and Android, with iPhone being the most used medium to post tweets.

The following table shows other kinds of information that could be relevant for each dataset:

Dataset	Mean Tweets/user	Mean Tweet lenght
Oscars	1.91	112
Brexit	2.44	141
Venezuela	1,966	143
Game of Thrones	1.3	108

Table 1: Additional information that can be collected from the Twitter API for each of the datasets.

From observing this table it looks like the political topics (#Brexit and #Venezuela) lead to the production of longer tweets than the topics regarding TV Shows or events.

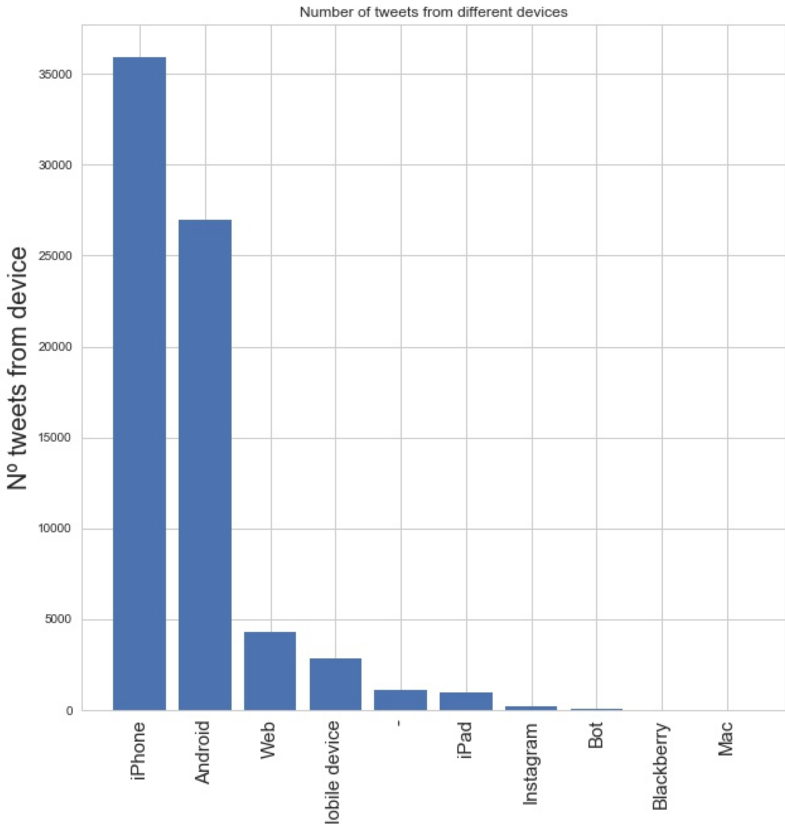

Figure 15: Tweets posted from different devices in the #GameOfThrones dataset.

The Tweet length can be larger than 140 characters, as there exists the possibility to post "extended tweets" which can be up to 280 characters, twice as long as a regular tweet.

After having analyzed the available information that can be readily extracted from Twitters Streaming API, the main parts of this work will be addressed, explaining how to build a signed graph using Twitter data, and how to use graphs to enhance the performance of the Sentiment Analysis needed to do this.

4 Targeted Sentiment Analysis

As introduced Chapter 2, Sentiment Analysis is the task of given a sentence, outputing its sentiment: positive, negative, or neutral. As the Twitter social network does not include any kind of explicit mechanism from which a negative relationship could be infered between two users, the only way to proceed for creating a signed network is to analyze the text intrinsic to the application. However, some further tough has to be put into how to analyze the tweet text in order to build this signed network.

Every tweet is produced by a specific Twitter user, but just classifying the sentiment of each tweet using a sentiment analysis model, would only give us the polarity of the tweets of such user, leading to a dead end in the construction of the signed network. Tweets however, sometimes contain mentions, which are direct referrals to another user of the Twitter network with the form *"Hello @MentionedUser, how are you?"*.

Considering the tweets with mentions as an interaction between the user who produces the tweet and the user who is mentioned, a signed network could be formed by analyzing the sentiment of said tweet. However, as mentioned in chapter 2, analyzing the sentiment of the whole tweet does not always lead to a correct sentiment classification for the target that is being mentioned. Consider the following Tweet:

"Again, another horrible #GameOfThrones ep simply a disaster full of plotholes, mistakes, and breakin the storyline, however there were some top performances from @emiliaclarke #MotherofDragons Bad bad writing"

This Tweet, classified by the online sentiment analyzer TextBlob [47] as negative with a polarity score -0.71, carries a very strong overall negative sentiment, however the sentiment or opinion towards the mentioned target (*@emiliaclarke*) is positive. Because of this, a distinct approach from plain sentiment analysis has to be used, in which the task is not to analyze the sentiment of a whole tweet or sentence, but rather to infer the sentiment towards a given entity within the tweet or the sentence. This variant of traditional sentiment analysis is called *Targeted Sentiment analysis*, and its what this chapter will be dedicated to.

In the framework of this work, the targeted entities would be the users mentioned within the tweets, building in that way a network of *@mentions*, where a connection is created between two users when one of them mentions the other one, and the sign of such connection is infered by the targeted sentiment analysis of the tweet towards the mentioned user.

4.1 Why Targeted Sentiment Analysis is needed

If the example shown above with *@emiliaclarke* is not clear enough of why traditional sentiment analysis is not fit for the job we want to do, in this subsection a couple more examples will be shown to highlight why the targeted approach is needed. Consider the following tweets, taken from our #Brexit dataset:

1 *"#Eggboy is my hero. THATS what we need more of. Absolutly love him. Not the #FraserAnning and @RealJamesWoods of the world that are racist pieces of crap that are scared of change because they are irrelevant now."*

2 *"Tonight I could not have more respect for @maitlis on #NewsNight verbally kicking all levels of shit out of every utterly deluded member of parliment inteviewed for triggering A50 leaving us #2weeks till utter diaster. Fuck the lot of you."*

Tweet 1 is directed towards *@RealJamesWoods* with a negative sentiment, however, if given as an input, either to TextBlob or to The Pararel Dots AI API for sentiment analysis [48], its gets classified as positive with 66% and 69% confidence respectively, as they perform plain sentiment analysis over all the sentence with no specific target. The targeted sentiment analysis model that will be explained shortly, classifies this tweet as negative towards *@RealJamesWoods* with a confidence of 1, which is the highest possible confidence output.

Tweet 2 on the other hand gets classified as negative by TextBlob and the Paralel Dots AI API with a confidence of 57% and 63% respectively, while it cleary shows a positive feeling towards the mentioned user *@maitlis*. This Tweet is classified as positive towards @maitlis with the highest possible confidence by the Targeted sentiment analysis that has been implemented for this work.

These examples highlight the need for a tool that doesn't focus specifically on analyzing the sentiment of a whole sentence, but rather the sentiment expressed in such sentence towards a target entity within it. The different approaches on the State of the Art for targeted sentiment analysis will be discussed in this chapter, but first we will explore some of the difficulties of performing Sentiment Analysis on Twitter.

4.2 Challenges of Sentiment Analysis on Twitter data

Twitter text, due to the relaxed and easy going nature of the network for content creation, leads to a high amount of data being posted, but at the cost of this data sometimes resembling more an SMS with abbreviations and shortcuts than a properly constructed sentence.

Most NLP approaches represent phrases as a vector of words using the *"Bag of words"* approach. In this approach a certain vocabulary set of size N is created by using a set of documents similar to the documents that are going to be analyzed afterwards, (tweets in our case) and every word instance extracted, becoming known to the algorithm and part of the vocabulary. Now, when a sentence is going to be analyzed for whatever purpose, it is represented as a vector of size N with either a

1 or 0 in each spot of the vector depending on whether the word that corresponds to that spot is present or not, or by a number representing the word count in the phrase. In either way, for a word to be able to have a significant effect on the NLP task at hand, it needs to be included in the vocabulary. For Twitter, where a lot of different expressions, abbreviations and slang are used, it is hard to build a vocabulary that contains all the possible words that will be found in the following tweets to be analyzed, and if so, the vocabulary size becomes incredibly big, dragging down the performance of most algorithms.

Also, many NLP algorithms use features as the *Part of Speech Tags* or relationships derived from syntatic parsers, which again, for Twitter, with its particular text structure, dont work very well. Consider the following tweet, extracted from out dataset:

.@maitlis's is #mood rn xD #Brexit

If this tweet is analyzed for Part of Speech tags, the following results are returned:

Word	**POS Tag**
.@maitlis	PUNCT
's	PROPN
is	VERB
#	SYM
mood	NOUN
rn	NOUN
xD	SYM
#	SYM
Brexit	PROPN

As we see, for this sentence that presents almost all the complications of Twitter data (abreviations with *"rn"*, emoticons with *"xD"* and metions and hashtags), the POS Tagger from Spacy, a very popular python NLP library [49], does not know what to do with the mention, caracterizes the abreviation *"rn"* for *"right now"*, which are two adverbs as a noun, and characterizes the *xD* as a symbol. For this specific symbol is not very relevant, but symbols like *:) =) :(* or *=(* can be very good indicators of sentiment, and for an NLP model to recognize them in that manner would greatly help the classification performance.

Figure 16 shows the dependency relationships between the words. As it can be seen from the image the syntactic parser does not perform very well, due to not understanding the mention, the abbreviation and the emoticon.

Another recurrent issue found in Twitter data that limits the effective application of sentiment analysis models is the use of URLs and pictures together with the twitter text to infer sentiment. The following figure shows an example of the image case.

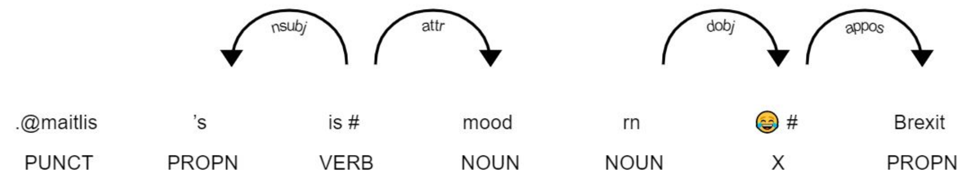

Figure 16: Dependency parsing on the previously shown tweet

Figure 17: Two tweets from the dataset which include images which are essential for identifying the sentiment of the tweet.

These two examples clearly show that in some cases, only using the tweets text is not enough to infer the sentiment, as the pictures that go along with this text can be used to express an opinion that is not so evident in the text. The tweet on the left part of Figure 17 criticizes the news covered by the account *@BBCNews* using a combination of the picture and the text. The image on the right illustrates the same behaviour, as the one on the left posting an image alongside the text mocking the policy of *@UKLabour* and therefore showing a negative opinion towards them.

Another issue is the use of URL links that re-direct to an external article or post. The following tweet

"@bbclaurak Take a look to see how much more business #EU will gain from #UK https://t.co/pNEOeMEZMt"

which is a post to an article from The Telegraph, a British Newspaper, describing the economical issues of Brexit. This tweet could be a neutral tweet just linking to some information, or it could be an answer to a statement the mentioned user has made, to correct it or show an contrary opinion to the mentioned user. To know this, the article, the context, and the tweet's text would have to be analyzed in order to correctly asses the sentiment towards the mentioned user.

This brings us to the next issue to be discussed: the context of the tweet, understanding by context the actions from the actors that have led to the publishing of such tweet. Consider the tweet

"@Nigel_Farage Marching to Leave Nige?! Great Idea!! When you reach the channel, keep going....... #Brexit #MarchToLeave https://t.co/9JkhTvUzWu".

This tweet refers to a walk in Sunderland organized by Nigel Farage to protest towards leaving the EU and meet up with his followers, which was widely controversial and lead to many jokes and sarcastic content being published. This tweet shows mockery towards the initiative, however, without knowing the context it would have been impossible to know. Also, even knowing the context of the post, this tweet is sarcastic, which is another one of the problems to be faced by any sentiment analysis tool.

The final issue for the sentiment analysis itself is the excesive use of Hashtags, and the use of these as actual words in the sentence, not just references to a topic. For example in the tweet

"@Jacob_Rees_Mogg I hope so Mogg, because the alternative is #democracy no longer carries any weight whatsoever in this country which means civil disobedience. These #filibustering #Remoaner #MPs need to stop and realise the country comes first, not themselves or their #EU utopian dreams."

there are a lot of hashtags that are replacing normal words are therefore could interfear in the functioning of the sentiment analysis model.

Aside from issues intrinsic to the nature of the application and its text, there are also some other difficulties that should be reported. The first is the absence of a large body of datasets for Twitter Sentiment Analysis, and even more so for Target Sentiment Analysis [50]. Many times the semi-supervised or noisy labeling approach is leveraged to rapidly label tweets for sentiment analysis using emoticons for example [2]. This is also due to the difficulties of labelling the tweets because of the previously mentioned issues.

4.3 Previous targeted Sentiment Analysis approaches for Twitter data

Targeted sentiment analysis, as mentioned earlier, is a variant of sentiment analysis that does not have the goal of predicting the sentiment of a whole sentence, but rather of predicting the sentiment towards a given entity within such sentence. Works in this field show that straightforward sentiment analysis for this task doesn't work out well, as the relationships between the target and the rest of the sentence are not taken into account [51].

However, incorporating features that model the relationship between the target and the rest of the sentence, like POS tags and relationships derived from syntactic parsers have shown to help in this task, despite the difficulties of using these tools in Twitter data that were described before. These features were then fed to different classifiers, like Recurrent Neural Networks, along with the tweet's text to achieve acceptable results of about 65% accurracies using adapative recurrent neural networks, but still using features that depend on the syntactical structure of the sentences to perform the classification [52]. These results however, could be better if the features derived from POS tags and syntactical parsers were not necessary for the Target-dependant sentiment analysis.

Because of this, works in the field have advanced towards creating features for targeted sentiment analysis that do no depend on syntax. Some of these works use sentiment lexicons and include slang to infer the sentiment of sentences using consistency relationships between expressions and optimization algorithms [53]. This work, that tries to do a three class classification differentiating between positive, negative and neutral tweets, shows how challenging the task of identifying polarized expressions is, compared to infering the polarity of the expression once its known whether it is polarized or not. They achieve an accuracy of 59% of this task of identifying polarized tweets and about 63% accuracy in the classification of sentiment towards targets.

The most recent works on this topic use word embeddings to model the relationship between the different words in the sentence, and therefore solving the need to use POS tags or syntactic parsers to infer the interaction between the target word and the rest of the sentence. In this way the relationships between the target word and the rest of the words in the tweet are automatically captured, without the need to manually craft the features for the NLP model [54]. Also, these works split the sentence to the left and to the right of the target word, and then use Long-Short-Term Memory recurrent neural networks or Gated Neural networks to perform the classification using the word embeddings of the two different parts of the sentence: left and right of the target entity [55][56].

Figure 18 depicts the structure of the LSTM of one of these works.

This image shows a recurrent trend in the state of the art works for targeted sentiment analysis: splitting the sentence between the right and left context of the target word, and using two LSTMs, fed with the word embeddings of each word of the sentence, to get a combined context vector and then perform the classification.

After having discussed the evolution reported from the literature about sentiment analysis, the work whose strategy was used will be explained in detail in the following subsection.

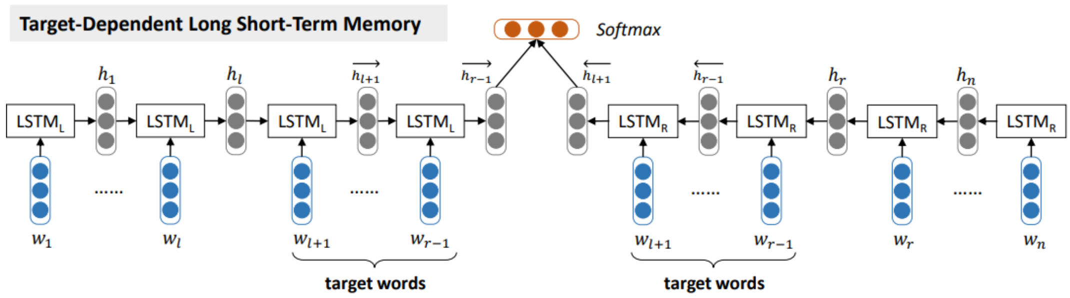

Figure 18: The Target Dependent LSTM of Tang et al. In the figure it can be seen that the sentence is split it two: the right and the left contexts of the target words, and then the outputs of each LSTM are concatenated and fed to a softmax layer to obtain the classification output.

4.4 Chosen solution overview and evaluation

To perform the targeted sentiment analysis on the collected Twitter data, the work of Liu and Zhang has been chosen [57]. At the moment of investigating the different methodologies for targeted sentiment analysis, this seemed like the work that had the best performance and that had used the most varied datasets available for this topic, which will be described shortly. Also, this work made its source code public, so only a few changes had to be made to this code in order to obtain the desired sentiment labels and their confidence. The process for adapting our tweets to the format required for this algorithm will be explained too.

The success of the previously described works, suggests that using word embeddings and deep neural network structures can automatically exploit the syntactic and semantic structures without the need to handcraft the features. The work of Liu and Zhang follows the same line of work with regards to the methodology.

The main contribution from the work of Liu and Zhang, aside from their results, which are the best reported to date, is the modelling of the importance of each individual word on the sentiment polarity of the sentence towards the target. The previous neural models use different neural network structures to model the relation between context (left or right) and target, but they don't model the importance of each word in the sentiment polarity.

The following example is shown in the paper to illustrate this:

"#nowplaying lady gaga - let love down"

This tweet is neutral for the target *lagy gaga*, where the contribution of the word *love* is little, besides this word being positive. To do this, and calculate the contribution of each word in the sentence towards the sentiment Liu and Zhang use an attention mechanism.

This work splits the sentence into left and right context of the target word, like described before, and the attention mechanism [58] is applied to left and right context to calculate the weight of each word in the sentiment towards the targeted entity, showing that individual word contribution is highly useful for targeted sentiment analysis. The following image, also from this paper, gives a clear example of the attention mechanism, and the influence of the different words towards the sentiment.

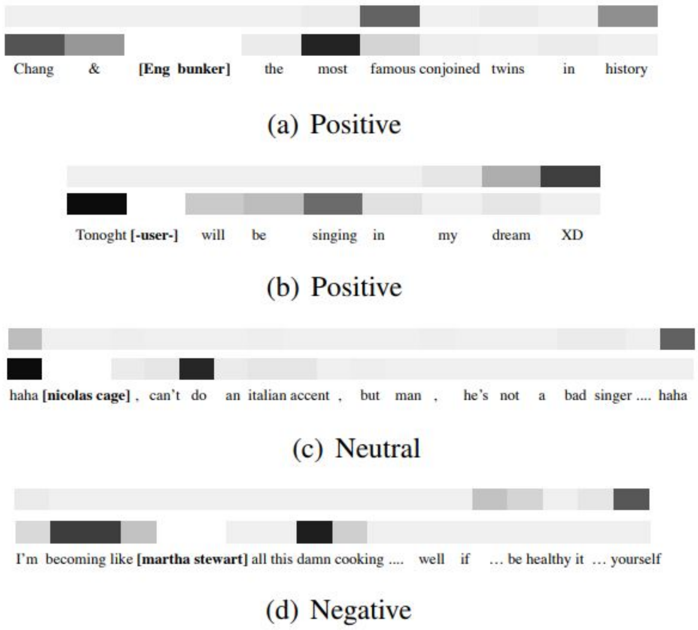

Figure 19: Four examples of the contextualized attention of Liu and Zhang. The words in bold represent the targets of the targeted sentiment analysis, and the contribution of each word is represented by the level of gray.

From this example it can be seen that the words *"most"*, *"famous"*, *"history"*, and *"xD"* lead to a positive label and the word *"damm"* leads to a negative one.

The paper trains their LSTM and evaluates the results of the targeted sentiment analysis using two public targeted sentiment analysis datasets, the Z dataset of [56] and the T dataset of [55]. To train the bidirectional LSTM attention model that will be used to classify the collected Tweets about Brexit, the Z dataset was chosen because of the greater size of the dataset and the more similar structure the tweets in that dataset have to ours than in the T dataset. The Z dataset is a combination of three other datasets from the Targeted Sentiment Analysis Literature: The dataset of [52], in which the Twitter API is queried for specific keyrwords like *"bill gates"*, *"taylor swift"*, *"xbox"*, *"windows 7"*, *"google"* and then the sentiment of these tweets towards those targets is manually labeled, conforming a dataset of more than 7000 tweets. The second part of this dataset is from the MPQA corpus [59], which has almost 1500 data instances, and the corpus of [60] which has about 3300 labeled entities.

The previously described data however, does not have exactly the same format as ours, as for us the targeted sentiment analysis is done towards the mentions in each tweet, increasing the difficulty for various reasons. The targets of the previous dataset, are entities that are addressed in a normal manner, and play no role in the application. Mentions, however, can be used as a way to reply to somebody's tweet so that it shows up in his feed, as a way to address someone so they will read the content of the tweet that is being posted, or used normally, in the same way as the targets on Liu and Zhang's work. Despite of this, without the notion of a dataset with exactly the same format and goal that would be preferred, the Z dataset was the one to resemble it the most, so it was adopted for this work. The following show various examples of the tweets in this dataset:

– *"lmao okay im making myself feel sick now. i did not watch never say never and i do not love justin bieber . i just want more followers lmao"*. This tweet is negative towards *justin bieber*

– *"sky tv through xbox seems pretty good"*. This tweet is positive towards *xbox*

– *"anyone out there want another contact on google wave ? trying to see how useful this will be"*. This tweet is neutral towards *google wave*

The parameters for the Sentiment Analysis performed in this work are the following:

Parameters	**Value**
Word dimmension	200
LSTM hidden dimmension	150
attention hidden dimension	100
dropout probability	0.5

Table 2: Hyperparameter values for the Bidirectional LSTM Attention model

Before feeding the data to the algorithm we do some light processing of the data to make it adequate for the algorithm and for our own purposes. First we extract the mentioned user from the Tweets text and find its index, as this index or position of the mention within the sentence is one of the inputs to the Targeted sentiment Analysis algorithm. Then urls are replaced by -url- and the mentions are replaced by -user-, as done by Zhang et al. All the text is put to lowercase, and double whitespaces are removed.

After this light pre-processing, the data is fed to the algorithm and a label (positive, neutral or negative) is the output, along with a confidence for the prediction of that label, ranging from 0 to 1. For the sentiment analysis improvement that will be described in chapter 6, some of the tweets from the Brexit dataset were manually labeled. These tweets can be used to asses the initial performance of the Sentiment

Analysis model of Zhang and Liu straight out of the box, trained with the Z dataset, in our Brexit data, with the mentions of each tweet as the target entities for the targeted sentiment analysis.

For 100 Tweets of confidence 1 (the highest possible confidence value) the percentage of correctly labeled tweets from the sentiment analysis model, was of 85%, showing that the tweets labeled with the highest confidence value are very accurate. Out of the errors from these tweets 77% consisted of neutral tweets being labeled as positive, confirming another issue with the works that include neutral tweets, which is the hardness of the task of infering subjectivity reported by [53].

Another issue with these tweets of high confidence, is the sarcastic tweets, that get classified by the targeted sentiment analysis as positive tweets with a confidence of 1, but that should actually be negative. The following tweets illustrate this behaviour:

– *"So good to see @theresa_may being SO current with her choise of catchphrase... #Simples #Brexit https://t.co/vuwk8o2jb"*

– *"@theresa_may just responded 'Simples'. 'Im loving it' Brexit"*

As it can be seen, it is really difficult for a sentiment analysis algorithm to detect this sarcasm, which limits its performance in these kind of contexts.

The overall accuracy of this sentiment analysis on our data, evaluated on 1500 hand labelled tweets is of 57%. This is due to the particular structure of our targeted sentiment analysis dataset, which uses the mentions as the targets for the targeted sentiment analysis, with the previously mentioned difficulties. Chapter 6 will explain various methodologies for improving this performance using information derived from the signed graph that will be described in the next chapter.

Figure 20 shows the evolution of the accuracy as the confidence of the tweets from the sentiment analysis increases.

This figure shows the percentage of correctly classified tweets when taking tweets only above or equal to the confidence value. For all tweets ranging from confidence 0.5 to the maximum value (those tweets with a confidence of 1), the percentage of correctly classified tweets is of 57%, and the tweets sum up to an total of 1500 (all the tweets to be hand labeled where taken to be above this confidence level). As the confidence increases the number of tweets involved goes down at the same time that the accuracy for the left over tweets goes up, finishing with 85% accuracy when only considering the tweets with a confidence equal to 1.

After having discussed the difficulties of sentiment analysis on Twitter data, explained the reason behind the need to use a more complex version of sentiment analysis than the traditional one, and seeing some of the results of applying it to our

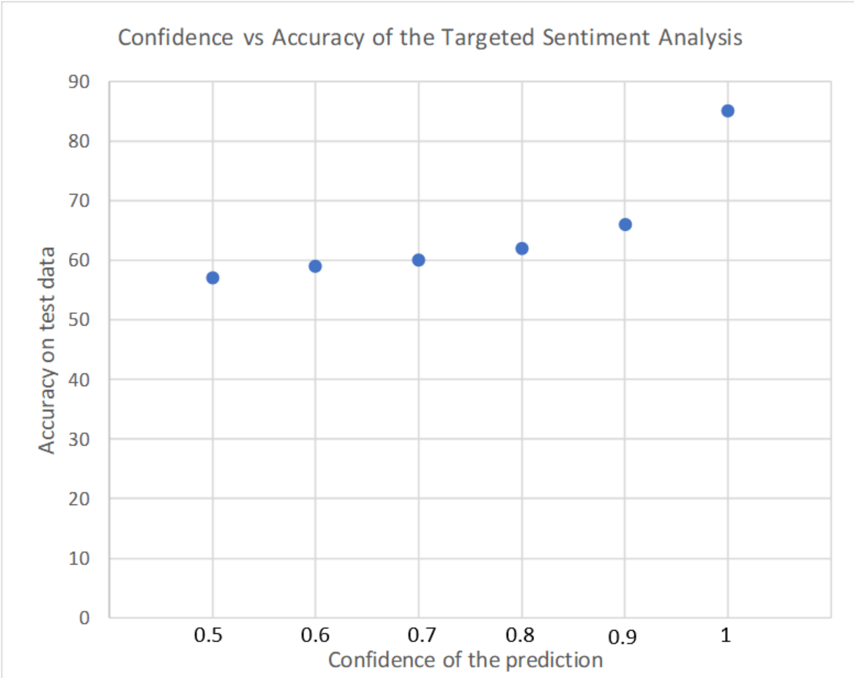

Figure 20: Evolution of the percentage of correctly classified tweets as the confidence of the sentiment analysis increases.

Brexit data, how to make a Signed Graph out of this data using the aforementioned techniques will be explained.

5 Graph Construction and analysis

Initially, the main purpose of this work was to build a signed network from the Twitter social application. In the previous chapter some hints have been given as to how this can be done, so in this chapter the construction of the signed graph will be fully described, along with its parameters and different insights derived from it.

5.1 Signed graph construction

First, the procedure for constructing the signed graph will be fully explained, and afterwards the graph will be analyzed. Two different signed graphs will be constructed, each with a different number of nodes, and compared. The steps for building this signed graph are the following:

1 As the initial step, the set of 835552 Tweets from the #Brexit dataset is filtered to keep only those which have a mention (*@MentionedUser*) in their text and are not retweets. This results in a set of 76404 tweets produced by 42751 different twitter users.

2 From these tweets with mentions, we keep those that have a single mention, as the sentiment analysis algorithm described in the previous chapter is not prepared to analyze the sentiment towards multiple targets. This leads to a set of 41877 Tweets, produced by 28407 different users, and mentioning 15608 different Twitter accounts.

3 From these Tweets we take away those which have the mentioned target at the end like *"I think all hope is lost, time to go back and think #Brexit @MentionedUser"*. For these tweets it is hard to infer the sentiment, as the mention at the end is just used most of the time as a tag for the *@MentionedUser* to read the text written by the user posting the tweet, but is not necessarily directed towards him. After removing these tweets we are left with just bellow 40.000 tweets.

4 These Tweets are then fed to the Targeted Sentiment Analysis model by Liu and Zhang described in the previous chapter, using the *@MentionedUser* as the target of the sentiment analysis. In this way we can predict the sign of the interaction between the user who is posting the tweet and the user who is mentioned.

5 To make a robust and trustworthy graph, a certain threshold is picked, and only the tweets that are output of the sentiment analysis with a confidence higher than said threshold are used. This is done because there is a lot of noise in the network due to personal messages, jokes, sarcasm, etc, so in order to build a network structure that is robust only tweets that surpass a certain confidence are considered.

6 Lastly, to build the signed network using the tweets filtered by the confidence, for every tweet, if they are not already in the network, two new nodes get

added, one representing the user that is posting the tweet, and one representing the user that is being mentioned. The sign of the edge in between these two nodes is chosen using the label from the sentiment analysis. If this label is positive, then a positive edge is created in between the users, and if the label is negative a negative edge is created. Tweets labeled as neutral are discarded.

Two different signed graphs will be built and analyzed, a graph taking only those tweets that are classified by the targeted sentiment analysis with a confidence of 1, which is the highest possible confidence, and a graph taking those tweets classified with a confidence or 0.9 or more. The main difference between these graphs is the number of nodes on each of them.

5.2 Signed graph analysis

First we will explore the graph with 0.9 confidence or more, which is the larger of the two. This graph is made from a set of 9376 tweets coming from 7837 different users and mentioning 4985 ones. The graph has 12564 nodes and 8931 edges. The following table shows the main parameters of this graph. The Average degree is

Nodes	12564
Edges	8931
Average Degree	1,4217
Density	0.0001131

Table 3: Table showing the main properties of the network constructed using the tweets with a confidence threshold of 0.9.

calculated as the number of nodes divided by the number of edges, and the Density as the total number of edges divided by he possible edges:

$$k = \frac{2E}{V} \tag{3}$$

$$d = \frac{2E}{V(V-1)} \tag{4}$$

Formula (3) represents the formula for the average degree of the network, where k is the average degree, V is the number of nodes in the network and E is the number of edges on the network. Formula (4) is the equation for the density of the network. The density of a network is a measure of the portion of possible ties which are present between the members of a network. As it can be see from the very low density of our graph, using the Twitter mentions as an interaction mechanism in between two users to create a network leads to this network being quite sparse. This issue has already been reported in the literature by [40].

This is mostly due to the creation of what from now on will be called hubs, created around central users which are influential accounts inside of the Twitter network within the topic being discussed (#Brexit), that are recurrently being mentioned. In graph theory, the influence of a node in a network can be infered by the degree of the node, or the number of connections it has. The following list depicts the top 15 most influential users in this network, along with their degree, which represents the number of connections of such node, and degree centrality, which represents the fraction of nodes in the network that it is connected to.

Username	**Degree**	**Centrality**
theresa_may	314	0.0249
realDonaldTrump	180	0.0143
jeremycorbyn	154	0.0122
UK_Labour	109	0.0086
Anna_Soubry	96	0.0076
Nigel_Farage	62	0.0049
bbclaurak	51	0.00405
eucopresident	49	0.0039
BBCNews	44	0.00350
maitlis	42	0.00334
SkyNews	39	0.00310
Jacob_Rees_Mogg	37	0.00294
Keir_Starmer	37	0.00294
mrjamesob	37	0.00294
Conservatives	35	0.00278

As we can see from this list, the influential or central users of the network can be divided into four main groups: politicians (Theresa May, Donald Trump, Jeremy Corbyn, Anna Soubry, Nigel Farage, Donald Tusk, Jacob Rees Mogg and Keir Starmer), political parties (UK Labour and Conservatives) journalists and reporters (Emily Maitilis, James O'Brien, and Laura Kuenssberg) and news sources (BBC News and Sky News), forming a varied and heterogeneous group of accounts.

One of the key insights that can be discovered from any social network, and specially from Twitter, are the main actors of such network. By using the degree of the nodes in the network a pretty good idea of whom the most influential individuals in this network are can be obtained, however for Twitter we can use some additional features to confirm that such users are undoubtedly influential and important [21]. One of the ways to do this is by exploiting the information available from the Twitter API about the followers and friends of each user.

Traditionally users that are highly followed are personalities and institutions whose influence and reputation is superior to users who subscribe to a large number of accounts without themselves being widely followed. Users who follow a large number of accounts but that do not have very many followers, are users who use Twitter as a monitor, without necessarily creating much content. As Kwak et al reported in [61] Twitter is one of the most used media for news monitoring: about 85% of the Twitter content is news related.

Users with the opposite pattern, that is, who are followed a lot more than they follow, are usually some kind of special personality or actor within a certain field or topic. Most of the interactions on the network address one of these users. [21] proposes a wide variety of metrics, aside from the centrality one that has already been discussed, to infer the influence of a node within a Twitter network. Two of these metrics are the *Twitter Follower/Followee ratio* and the *Follower rank*.

$$TFF = \frac{Followers}{Followees} \tag{5}$$

$$FRank = \frac{Followers}{Followers + Followees} \tag{6}$$

In the equations above *TFF* is the *Twitter Follower/Followee ratio* and FRank is the *Follower Rank*. The following list shows these two metrics for the influential users listed before, along with their followers and friends/followeers, ordered by their degree in the network.

Username	**TFF**	**FRank**	**Followers**	**Followees**
theresa_may	29006	0.9999	841177	29
realDonaldTrump	1312633	0.9999	60381119	46
jeremycorbyn	800	0.998	1961127	2451
UK_Labour	56	0.982	676449	12008
Anna_Soubry	272	0.996	176970	649
Nigel_Farage	2828	0.9996	1312346	464
bbclaurak	665	0.9985	1000786	1503
eucopresident	2741	0.9996	1049972	383
BBCNews	98752	0.9999	9974037	101
maitlis	130	0.9923	153640	1177
SkyNews	17300452	0.9999	4991765	26
Jacob_Rees_Mogg	1787	0.99944	277093	155
Keir_Starmer	472	0.997	167583	355
mrjamesob	207	0.9951	474656	2291
Conservatives	222	0.9955	375237	1688

From this table it can be seen that the relative ratio of followers to followers and followees represented by the *FRank* measure is very high for all the users in the list, confirming that they are all accounts that are largely followed, however the *TFF* varies a lot to the greater dependency it has on the number of users that are followed by each of these accounts, which go from 29 (Theresa May) to 12008 of the UK Labour party.

Another characteristic of the constructed network is the power-law degree distribution that it follows. This basically means that there is a very small number of nodes with a very large number of connections (the hub centres or influential users) and a very large number of nodes with very little connections. Figure 21 depicts this

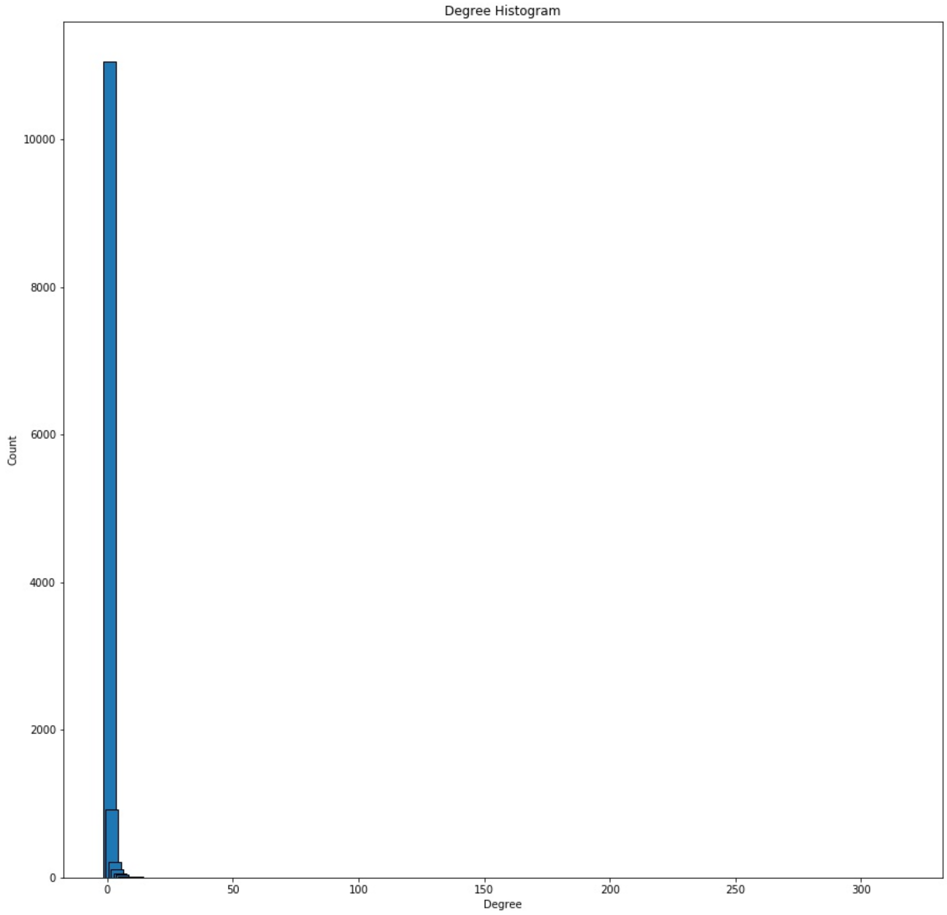

Figure 21: Degree histogram of the nodes belonging to the signed graph extracted from the Tweets with confidence higher than 0.9.

power-law in the form of a histogram representing the node degrees by their quantity.

What this figure is showing, is that there are a large number of nodes with degree 1 (specifically 11045, which is 87% of the nodes in the network), 925 nodes with degree 2, and less than 500 total nodes with degrees in between 3 and 13, and from there, the network has a very small portion of nodes that have a very high degree (1 node with degree 314; Theresa May, 1 node with degree 180; Donald Trump, and so on). Low density Twitter networks built around a certain topic have a large number of nodes which lack any connections at all, called *isolates*, as popular topics attract tweets that are not part of a conversation, but rather an opinion on the topic at hand [26].

For our network, as it is essentially built from user-user interactions, it is impossible for any node to be un-connected. However we will extend the isolate term defined above to all nodes that have a degree of 1 and are not connected to any of the influential targets or large connected components: that is either a user who has mentioned a certain target who has not been mentioned by any other user, or such mentioned user.

Some other network properties are shown in the Table 4. The transitivity, which computes the fraction of all possible triangles present within a network being so low, means that the network can not be easily analyzed in the terms of Balance Theory to see if the edges predicted by the sentiment analysis are correct, as it does not contain many triangles (The network formed by these mention interactions only contains 5 Triangles). Also, the clustering coefficient is quite low, as it is closely related to the Transitivity.

Partitions	3958
Connected Components	3932
Transitivity	0.00011
Clustering Coefficient	0.000235

Table 4: Table showing more properties of the constructed signed network

$$T = 3\frac{Triangles}{Triads} \tag{7}$$

$$Cu = \frac{2T(u)}{deg(u)(deg(u)-1)} \tag{8}$$

Formula (7) is the equation for the transitivity of a network, represented by the number of existing triangles out of all the triads in the network. Formula (8) represents the clustering coefficient for node u, where *T(u)* are the number of triangles which include node u, and *deg(u)* is the degree of node u.

Table 4 showed the number of partitions in the network, which were formed trying to maximize the network modularity; that is, maximizing the number of connections within each cluster while minimizing the connections in between the clusters. By doing this we get 3958 partitions, out of which only 12 have more than 80 nodes and 3800 partitions which have less than 5 nodes. All of these 12 partitions with more than 80 nodes contain one of the influential or highest degree nodes, and consists on their most part of nodes with a degree of 1 that are connected to the central user, representing individual accounts who have mentioned a well known actor of the network, but have not interacted in any other way.

From all of this we can see that our network is a large-scale network (follows a exponential degree distribution with many nodes having a low degree and very little nodes having a really high degree) and has a repeated star-like network structure, where the influential nodes that are being mentioned the most have a lot of connections to degree 1 nodes, and nodes that do not mention one of these influential users are sparsely distributed forming isolates. Figure 22 depicts an example of a portion of the network.

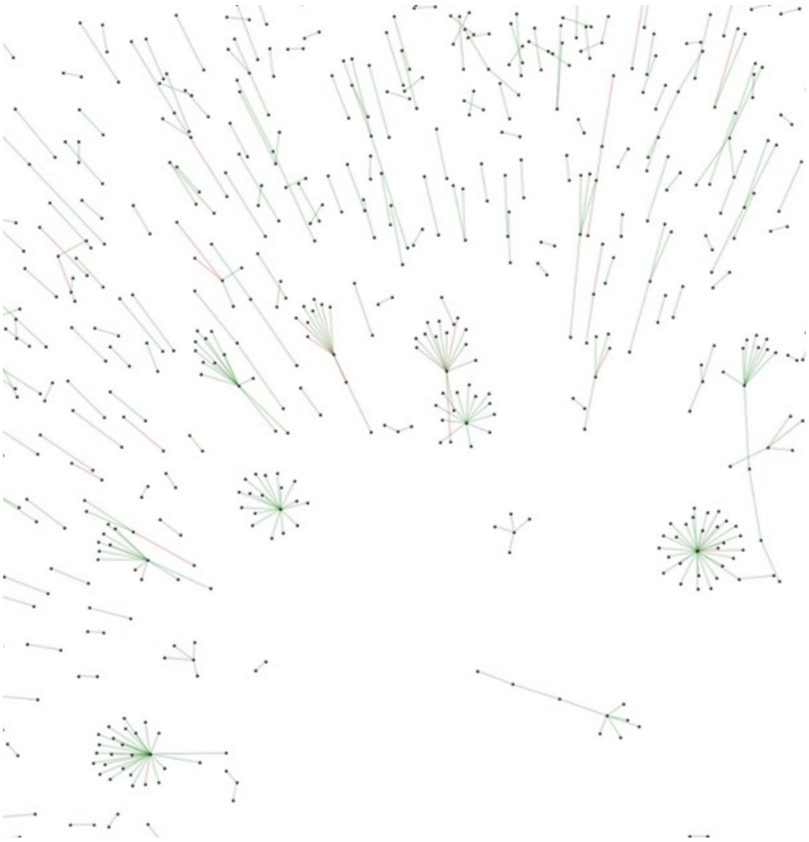

Figure 22: Graphical representation of the network structure. We can see how some minority of the users gather a lot of connections creating the so called stars, and many users form a single connection either to one of the central nodes of these stars or to some other random node.

As this figure shows, there are various star-shaped clusters, representing the main personalities of the network, and then very many isolates, connected to one or two other nodes. The color of the edges in the network represents the sign of such edge: green edges represent a positive relationship between the nodes that are connected by such edge, and red edges represent a negative relationship.

From the signs of the edges of this network, we can infer how many positive and negative connections are made to the most influential users. The following list shows this:

Username	Positive edges	Negative edges
theresa_may	189	127
realDonaldTrump	151	29
jeremycorbyn	105	50
UK_Labour	64	45
Anna_Soubry	84	12
Nigel_Farage	43	19
bbclaurak	44	7
eucopresident	42	7
BBCNews	27	17
maitlis	32	10
SkyNews	25	14
Jacob_Rees_Mogg	28	9
Keir_Starmer	33	4
mrjamesob	35	3
Conservatives	15	20

From looking at this list, it seems like a majority of the created links are positive, which is what normally happens in signed networks from social media. We can clearly see this, as out of all the top 15 users, only one (*Conservatives)* has more negative links than positive links. Also, it could be possible that the Targeted Sentiment Analysis is biased towards predicting positive edges in between targets. This could be due to the algorithm classifying neutral tweets as positive, like it was explained in the previous chapter, the also previously mentioned difficulty of sentiment analysis on Twitter data, and the extra challenge of going one step further from traditional sentiment analysis towards targeted sentiment analysis. In the network there are 7624 positive edges and 1307 negative ones, which account for 15% of the total edges, showing the higher tendency of links to be positive rather than negative. In the next chapter a method to increase the performance of the sentiment analysis in order to get a more accurate estimate of the aggregated feeling towards the important personalities will be described and evaluated.

After having described this network, the same parameters for the network but with just the tweets being labeled with a confidence of 1 will be exposed, and then some further insights into these kind of networks will be discussed.

For this graph, made with only the Tweets labeled as positive or negative with a confidence of 1 the parameters are presented in Table 5. After filtering the Tweets to keep only those with a confidence of 1, we are left with 1391 different Tweets made by 1286 different users with 1052 different mentions.

This graph, because of the limited number of Tweets it uses, presents a slightly different set of central users than the graph constructed from the tweets with 0.9 confidence, however most users in this Top 15 remain the same. We take as more valid the user set from the initial graph (0.9 confidence graph) as it includes more tweets, and more opinions, and despite of the opinions of this graph being labeled with less

Nodes	2331
Edges	1363
Average Degree	1,1695
Density	0.000502

Table 5: Table showing the main properties of the network constructed using the tweets with a confidence of 1.

confidence, they would have still had the same mentioned target independently of the label, so it better reflects which are the important users of the network. As more and more tweets get added, the more sure we can be that the nodes with the highest number of connections are actually the most influential users. The following list shows who are the 15 users with the highest degree in this network, along with their Twitter Follower/Followee ration and their FollowerRank, and with their number of followers and followees, ordered by their degree.

Username	**TFF**	**FRank**	**Followers**	**Followees**
theresa_may	29006	0.9999	841177	29
Anna_Soubry	272	0.996	176970	649
realDonaldTrump	1312633	0.9999	60381119	46
jamesAcaster	151	0.993	351872	2325
jeremycorbyn	800	0.998	1961127	2451
jessphillips	75	0.986	200431	2642
EuAvObservatory	68	0.985	2817	41
DGrylls4	5	0.986	552	109
TheRogueEnergy	37	0.974	34064	908
OttawaMommyClub	3	0.74	21653	7358
UK_Labour	56	0.982	676449	12008
ByDonkeys	-	1	208718	0
RedLetterDaysUK	6	0.849	28007	4977
mrjamesob	207	0.9951	474656	2291
POTUS	662737	0.99	25846746	39

By observing this table it can be seen why some of the users that are included in it drop out when making the same list using more tweets. The user *EUAvObservatory* has a very small number of followers compared to some of the accounts belonging to real influential users, and so does the user *DGrylls4*. The user *OttawaMommyClub* has a decent enough amount of followers (altough far less than the users with the most followers of this list), however, the *TFF* is very low, suggesting that this amount of followers is a consequence of the high number of accounts that the user follows. Something similar, but to a lesser degree happens with the user *RedLetterDaysUK*.

Table 6 shows the rest of the parameters for the network constructed with only those tweets with the highest possible confidence The transitivity and the clustering coefficient were not included, as they reported insignificantly small values. It shows how the number of partitions is equal to the number of connected components. Out of these, only 12 partitions/connected components have more than 10 nodes, and they are all star-like clusters with one of the most influential nodes in the middle.

Partitions	972
Connected Components	972

Table 6: Table showing more properties of the constructed signed network with confidence 1

In this graph there are 1254 positive edges and 109 negative ones, which make up to 8% of the edges in the network. Figure 23 depicts a graphical representation of this network, where the green lines represent positive edges in between the nodes, and the red lines represent negative edges in between them (the previous network of 0.9 or above confidence was not plotted because it is very challenging to get a neat representation of a network with so many nodes). It can be observed how there are a large amount of node pairs or isolates, that surround start like structures of groups of connected nodes. If we remove these isolates to be able to get a cleaner look of the biggest clusters, the graph looks like in Figure 24.

From this figure it is easier to observe at the star like structure from our graphs. After taking at look at this structure, and exploring the different parameters of the graph derived from tweets of confidence 1, some common characteristics to all the graphs of this sort will be explained.

The first one is the reason for the low density of these kind of graphs. As it was mentioned before, networks derived from Twitter, and specially those constructed from interactions and not from follower relationships, generally contain hubs which are order of magnitudes larger in degree than the rest of the nodes.

When increasing the tweets that these networks are formed from, and therefore the number of involved users, the number of nodes and edges in the network tends to grow linearly, however the number of possible connections in the network grows quadratically with the number edges, leading to a reduced density as the number of nodes gets increased. This is true for almost every social network, not just Twitter.

Figure 25 shows the evolution of the graph density as more nodes (tweets) are added to the graph by decreasing the confidence threshold. As it can be seen from this image, starting at the least number of possible nodes (which corresponds to only taking those tweets with a confidence of 1) and gradually increasing the number of nodes (decreasing the confidence) the density of the graph decreases as described before.

The other characteristic that will be described is the *preferential attachment* phenomenon that happens in these kind of graphs. Preferential Attachment expresses the

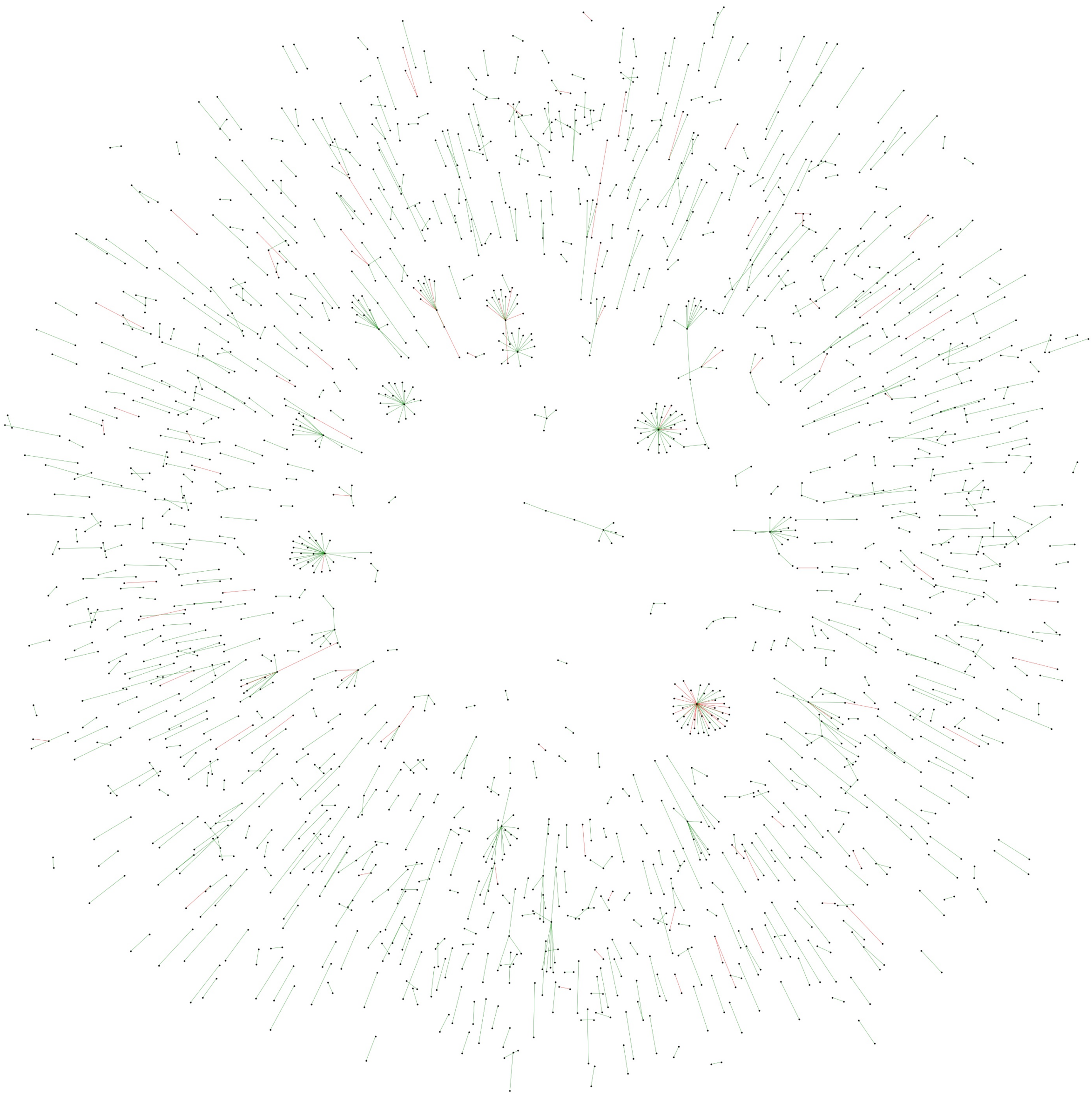

Figure 23: Graphical representation of the network built with tweets of confidence 1 using a force-directed plotting algorithm.

tendency for nodes that are already central in the network to gain more connections at a greater rate than those who are not. For our network this would mean that as more nodes get added, there is a higher chance of those nodes connecting to one of the largest components or hubs, than connecting to a small connected component or creating a new isolate.

From the Twitter application point of view this would mean that as more tweets are produced, they are more likely to be mentioning one of the top users from the #Brexit topic, than to be mentioning some other random user. The following figure

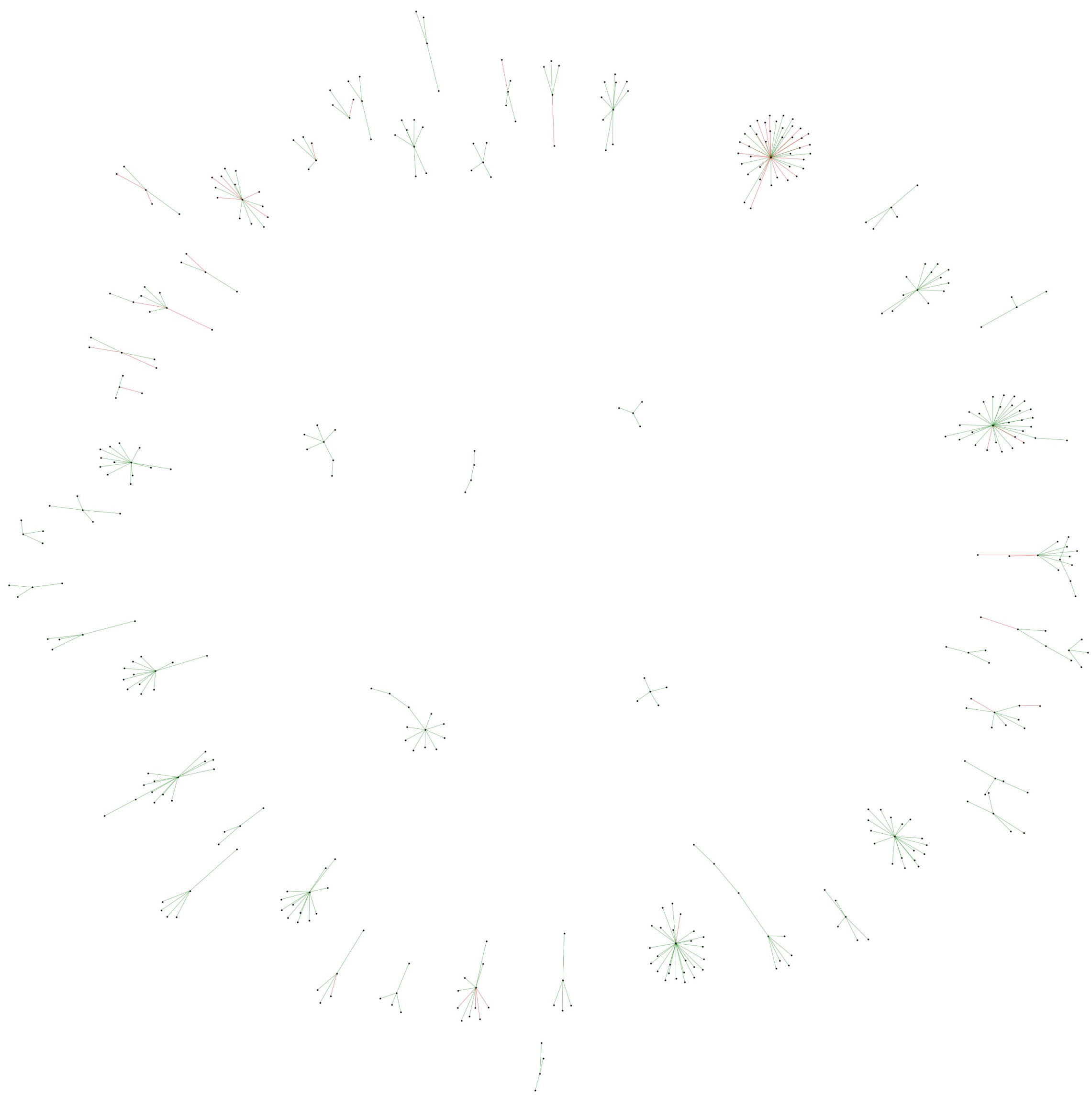

Figure 24: Graphical representation of the network built with tweets of confidence 1 using a force-directed plotting algorithm taking away the connected components of size less than 3 nodes. In this figure the star-like structure of the largest connected components of our graph can be clearly observed.

confirms this behaviour in the signed mentions network: as we add more nodes to the network in the same way that it was done for the previous figure, the percentage of network nodes that are attached to one of the top 15 highest degree users increases.

After having described the process of making the graph and acquired some insights about its structure and properties, we will explore how using graphs like these, the performance of the Targeted Sentiment Analysis described in Chapter 4 can be increased by taking using the social relationships embedded in these kind of graphs,

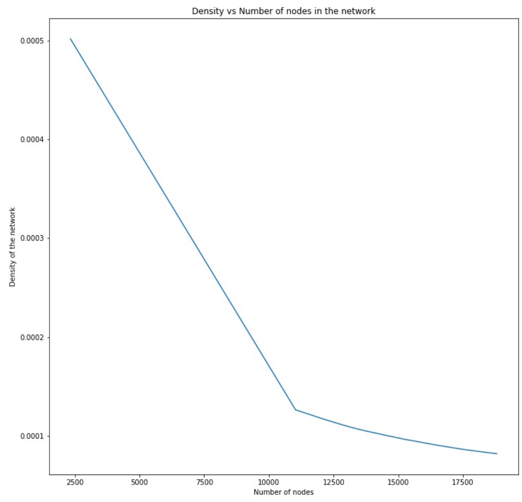

Figure 25: Graph showing the evolution of the density of the graph as more nodes are added to the network.

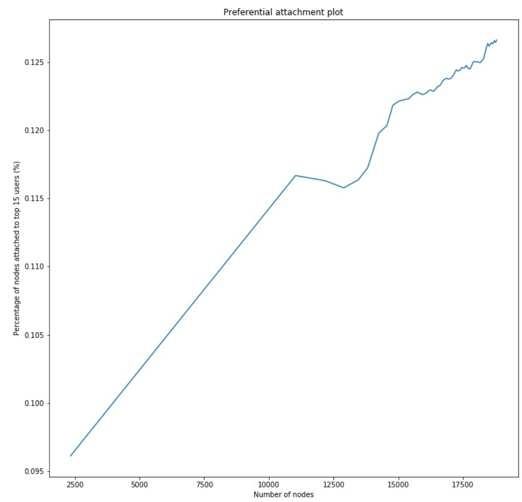

Figure 26: Plot showing the increase of the percentage of nodes that are attached to the largest connected components. phenomenon.

and including new kinds of relationships intrinsic to the Twitter network.

6 Sentiment Enhancement using the Signed Graph

In this section we will address the problem of enhancing the performance of the sentiment analysis classifier by using information coming from the various graphs that can be made with the data that is available. In this way we can not only infer more precisely the sentiment of the tweets towards a certain target, but we could also make a more robust graph as more users would get added by having their sentiment confidence increased.

6.1 Introduction and influencing works

For this task, the contributions from three main papers, briefly mentioned in Chapter 2 of this work were used. The first of these works is the work by West et al [36]. This work uses Sentiment Analysis, along with Social Balance Theory to predict the signed edges on the Wikipedia admins network dataset. For a person to become a Wikepedia administrator (the people who check that the editions to the content are legitimate and correct) a *request for adminship* (RFA) must be created, and any Wikipepdia member can cast a supporting, opposing or neutral vote towards such request. These interactions induce a signed network in which the nodes represent Wikipedia members and the edges represent the votes (positive edge for supporting request vote and negative edge for an opposing vote). In addition to the vote itself, the members can post a short comment explaining the reason of such vote. Using this text, the power of Sentiment Analysis will be used to leverage edge sign, together with the edges created by the different votes.

The problem is formulated as a network $G = (V, E, x)$, where the vertices V represent the Wikipedia members, the edges E the relationships in between them, and the sign vector x the polarity of the edges. The structure of this network (V and E) is assumed to be fully observed, whereas x is not completely known. Also, a sentiment model that outputs for each edge e a label and a confidence pe in such label is assumed. The task then, is to predict the sign of the unobserved edges based on this information, so that the new signs agree with the results of the sentiment analysis and with the results obtained to applying balance theory on those triads for which there are 2 known signs. As it is not always possible to achieve this, a compromise between predicted edge signs by sentiment analysis and by balance theory has to be found. This is defined as a combinatorial optimization problem that seeks to reduce the cost or error derived from both NLP edge predictions and triangle costs. The mathematical description of this problem is the following:

$$x^* = \underset{x \in \{0,1\}^{|E|}}{\arg\min} \sum_{e \in E} c(x_e, p_e) + \sum_{t \in T} d(x_t).$$

where x^* are the predicted edges, the first term of the formula is the cost corresponding to how much the sign of an edge xe deviates from the probability of the sentiment model predicted for that edge, pe and the second term represents the number unbalanced triads *xt* present in the network. One key takeaway from this work is that it divides the confidence range *pe* into a series of bins, and assigns a different weight for each bin, in order to obtain a solution that doesn't penalize equally those predictions which are not very confident and those who are. This insight will later be used for our heuristic approach to enhancing the sentiment analysis.The next image depicts a perfect example of the intuition behind this work. While the

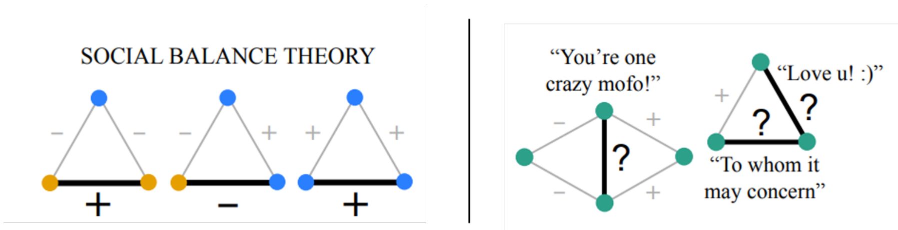

Figure 27: Predictions using Balance Theory (left) and sentiment analysis (right) of the darkened edges

three triads on the left part of the image correspond to a normal edge prediction using balance theory, the right part of the image contains some very illustrative examples of the power of combining both Balance Theory and Sentiment Analysis. The darkened edge with the question mark to the left, if only sentiment analysis was used by looking at the text embedded in that interaction *"You're one crazy mofo!"*, would seem negative. However, a negative edge would violate balance for both of the triads included in the image, so using the confidence of the sentiment analysis and the number of violated triads this edge would be labeled as positive. For the triad on the far right, knowing only one positive edge, balance theory can not be used to predict the other 2, as there is not enough information. Also, the phrase *"To whom it may concern"* does not carry a clear sentiment. However, the other sentence *"Love u! :)"* expresses a positive sentiment, and therefore if the edge corresponding to that interaction was positive, it would mean that for the triad to be balanced the bottom edge would also have to be positive, infering the interaction *"To whom it may concern"* as a positive interaction.

The results are evaluated by removing certain edges from the network and predicting their sign using the aforementioned technique, and show a considerable increase in the edge prediction using it in relation to the individual network theory or sentiment analysis.

The second work to influence this enhancement is [37]. This work, like ours, was carried out in Twitter data, and builds a dataset for political topics (specifically the 2012 American presidential elections). For each tweet, much like this work, it performs sentiment analysis, but using standard NLP features (not word embeddings) and Support Vector Machines. There is some very heavy initial preprocessing to remove URLs, hashtags, replace user mention by defined patterns, normalize slang words and correct spelling mistakes. Also, a lot of different features are used, like N-grams, POS tags, Word clustering, and syntactic parsing relationships. By doing all this, and optimizing their sentiment model to the maximum, they are able to achieve an accuracy of just under 80% on some labeled tweets by just using the sentiment analysis. As high as this accuracy might seem, the heavy previous work to the SA has to be taken into account, and also the fact that it is straightforward sentiment analysis, and not the targeted variation that this work needs to use. Also, they focus on a binary classification problem, not taking into account that the tweets can show a neutral sentiment.

For further improvement of these metrics, the authors assume some hypothesis derived from social cognitive theories, and translate them to their practical implementations. These assumptions are:

– *The sentiment for a particular target is globally consistent across users because of the target's stance on some particular issue.*

– *The sentiment for a particular target is globally consistent when he or she is compared with another particular target*

– *One user's sentiment toward one target or his/her stance on one issue tends to be consistent during a short period of time.*

Out of these three, the hypothesis which is particularly relevant to this work is the first one. After seeing the star structure of the graph that was described in the previous chapter, It was clear that one of the most evident tasks that these kind of graphs could be used for is for determining the aggregate feeling of the users of the network towards another specific user who is some kind of influential or important person in the topic that is being discussed. This hypothesis assumes that a certain target (the mentioned user in this work) issue (collected hashtag, which for this work is #Brexit) pair are consistently associated with a particular sentiment across most users. This means that given one of the relevant personalities of our dataset (one of the users associated with the nodes with the highest degree from the graph), the sentiment towards them for the topic *#Brexit* will be consistent for most of the users that are mentioning them.

In the paper this is reflected in the following manner: Tweets for a certain topic (hashtag) are collected, specifically for the 2012 American presidential elections. This tweets are noisily labeled by using hashtags like #Obama2012 or #GOP2012 for the positive labels and #Obamafail or #GOPfail for the negative labels. Sentiment Analysis using Support Vector Machines is used to asses the performance of using only the NLP approach by splitting the obtained data into training and test data. Then, for each unique target/issue pair in the training data, the number of positive and negative instances are counted, and a confidence is calculated using the following formula:

$$c = \frac{max(fp, fn)}{fp + fn} \tag{9}$$

where c is the confidence, *fp* are the positive instances and *fn* are the negative instances. If this confidence is above a certain threshold (0.8 in the paper) the target is assumed to have a strong general overall feeling, and then in compliance with the previously discussed hypothesis, all of the tweets for that target get relabeled to the strongest sentiment. To illustrate this procedure at work, consider the following Tweets:

1 *#Obama rebuilding America using Chinese workers! http://t.co/Pk4HvtL*

2 *But we had to rush #Obamacare thru? In the pipeline? Obama has it both ways on a controversial plan http://t.co/rb65LIx3*

3 *Small business owners confirm #Obamacare is a job killer: http://t.co/If7yNqVo*

The sentiment analysis approach missclassified the first tweet as positive, but correctly classified the other two as negative with high confidence. As the target has an overall negative feeling higher than the threshold, using the relabeling approach the first tweet was correctly re-classified as negative. For this work, a similar approach has been taken, using an overall global sentiment for each of the influential users of our dataset, however, the strategy implemented does not re-classify all the tweets to the most frequent sentiment label, but rather shift the sentiment output of the NLP sentiment analysis towards this most frequent sentiment. This approach will be fully explained after discussing the third work that served as an inspiration for this part of the project.

This last work is [38], which improves sentiment analysis by using information extracted from social relationships. This paper enhances the results of sentiment analysis on Twitter data using the relationships embedded in the Twitter follower network of the collected data. In this work tweets for 5 different topics (Obama, Lakers, Fox News, Sarah Palin and Glenn Beck) are collected and analyzed. Also, for each of the topics (the topics are identified by collecting tweets containing certain words) the follower/followee network of the users involved on that topic is built. The authors first engage in a small investigation to see if the social relationships and the sentiment labels towards a certain topic follow any kind of correlation. Their results show that the probability of two connected users sharing the same sentiment for a concrete topic is much higher than chance, and also that two users are more likely to be connected if the share an opinion (sentiment) on a topic than if they don't.

After this, for each new tweet on a certain topic, the probability of the each label (positive or negative) is calculated taking into account the output of the sentiment analysis and the sentiment labels of the neighboring nodes (followers and friends) of the user producing the tweet within the social network. For estimating the weights of each of the components of this probability two approaches are taken: first a heuristic approach using some pre-labeled data, and the second way is by applying Sample-rank in a semi-supervised setting, which is an algorithm for estimating parameters in complex graph models [62]. Their approach achieves better performance than the baseline NLP labeling, showing that social relationships can improve the performance of the sentiment analysis when it doesn't achieve very good performance because of a) the amount of available data, b) the characteristics of this data, or c) the complexity of the task. The main takeaways from this work are that social structures can help overcome the poorness of textual information coming from short tweets, and also the lack of a large amount of labeled data.

6.2 Proposed Methodologies

After having discussed these three papers, it is time to fully explain our approach, how it is similar to these previous works, but also where it differs from them. The goal of this approach, following the line of the previously exposed works is to improve the sentiment analysis performed on the tweets using additional information obtained from social graphs. Our approach is different from the previous ones, despite being inspired by them because of the following reasons:

1 First, the sentiment analysis performed is not normal topic sentiment analysis, but targeted sentiment analysis, which is a harder task and therefore doesn't achieve as good results as the straightforward sentiment analysis performed in the previous papers. Also, using a Twitter mention as the target of the targeted sentiment analysis adds an extra difficulty, as it was noted previously in the targeted sentiment analysis chapter. Furthermore, our work takes into account the existence of neutral tweets, which don't show any sentiment towards the target, rather than focusing only on a binary classification problem between positive and negative.

2 Secondly, the structure of our graph is different from the structure of the graphs in the previous works. Our graph has a different star-like, mostly disconnected clustered structure, as it represents a mentions graph, where very few influential users are being mentioned most of the time. The graphs in the previous works are follower graphs, or graphs taken from other contexts, which makes their structure different.

3 Third and last, only some parts of each of the previous works have been chosen to take part in this one, as they are quite consistant with the data and graph we have, and the approach we want to take. Also, we ought to incorporate parts of all of the previous works to arrive at a final solution which integrates the best fitting parts of all of them.

Two different approaches have been taken to improve the performance of the Sentiment analysis: a heuristic approach, using some of the elements from the previous papers, and a Machine Learning approach, using the output of the sentiment analysis and features extracted from the signed graph. First the heuristic approach will be discussed.

6.2.1 Heuristic Approach - Global Sentiment

The goal of both approaches is to enhance the performance of the sentiment analysis towards the important personalities within the #Brexit Twitter topic. The important

Figure 28: Some of the Twitter accounts of the most influential users on the graph

personalities are the users identified in the previous chapter, extracted from the graph made with a confidence threshold of 0.9. This list of users is compromised of polititians (Theresa May, Nigel Farage, Jeremy Corbyn, Donald Trump, Donald Tusk, Keir Starmer), journalists (Emily Matilis, James O'Brien), accounts from News Sources (SkyNews, BBCNews), and accounts from Political Parties (UKLabour, Conservatives), consisting on a varied and rich set of users who have been frequently mentioned on the *#Brexit* topic.

For each of these users 100 Tweets which mentioned them were labeled, having a total count of 1500 labeled tweets. Then, half of these tweets were used to obtain a global sentiment value, in the manner shown in Formula (9) and half of them were used to test the proposed heuristic. Similarly to [37], if this global sentiment value is above a certain threshold, we use it to assist the classification, but unlike this work, we do not change every tweet to the most frequent sign, but rather use this sentiment value together with the output of the sentiment analysis (confidence and label) to produce a new label for each tweet.

To do this, we first calculate a score using a weighted version of the sentiment, where the weight depends on the confidence of the classification from the sentiment analysis, and the global sentiment feeling towards the target calculated using 50 labeled tweets that are different from the tweets used to asses the performance of the approach. The mathematical representation of this would be:

$$S = W(cf) \cdot Se + Gs \tag{10}$$

where S is the calculated score, *W(cf)* is the weight which depends on the confidence of the sentiment analysis, and Gs is the global score calculated using (9) but with a sign dependent on which label is more frequent for the target: if positive is the most frequent label then Gs will be positive, and if negative is the most frequent label for the user then Gs will be negative. As for the weight W, the following table shows its values depending on the confidence of the Sentiment Analysis: For Tweets with a high confidence ($Cf = 1$) of the opposite sign of that of the global sentiment, a very high weight is used, so that even if the score of those tweets gets shifted down by the global sentiment, they will be still recognized as having the initial sentiment.

Table 7: Table showing the different weights given to the sentiment analysis depending on the confidence of such analysis.

Weight	Confidence	Sign(label) = Sign(Gs)
2	1	No
1	1	Yes
0.75	[0.9,1)	-
0.5	[0.5,0.9)	-
0.25	<0.5	-

This is one of the elements that [37] is missing: by classifying every tweet to the sentiment of the most frequent label if this global score is above a certain threshold, it misses out on strong supporters of people who are badly regarded, and on haters or people that don't agree with users that are highly regarded, miss-classifying the tweets of such users.

Consider the following Tweets:

1 *Against so many in Labour Party itself and pretty much ALL of the mass media @jeremycorbyn has stood firm as a magnificent. leader...#JC4PM*

2 *When Corbyn did put forward a motion for a #PeoplesVote on January 29th 2019 @Anna_Soubry voted against it!!!! Hypocritical, undemocratic, homophobe and racist appeaser!! By election now!!*

The first tweet is labeled as positive towards *@jeremycorbyn* by the sentiment analysis with a confidence of 1. *@jeremycorbyn* however, using the 50 tweets labeled to extract the global sentiment, has a negative overall score of 0.75 out of 1. As this score surpasses the threshold used in this work to asses if there is a strong feeling towards a certain target (this threshold, set at 0.7, will be explained shortly), this tweet would have incorrectly been missclasified as negative. Using our heuristics it stays positive while other positive tweets with less confidence get correctly reclassified as negative, like it will be shown later. In a similar manner, tweet number 2, directed towards *@Anna_Soubry* is labeled as negative with a confidence of 1 towards such user. Anna Sourbry however, has a positive overall sentiment value of 0.73 out of 1. Therefore, this tweet would have been incorrectly classified as positive.

For all the other tweets, this weight decreases as the confidence of the sentiment analysis does. After calculating this score for the 50 test tweets for every influential user, it is evaluated, and if it is above or below an experimentally set threshold, the tweet is classified as positive or negative, otherwise the tweet is classified as neutral. In a simple manner, what this heuristic is doing, similarly to [37] is relabeling tweets targeting a certain user in a user/topic framework towards the most frequent sign for said user. In our case, however, not all the tweets are relabeled, as some tweets with high confidence and opposite sign of the most frequent label stay with their initial label, tweets with confidences that are not too low can get their sign toggled or re-classified as neutral, and tweets with low confidence of the opposite sign to the most frequent label get re-classified to such label. Tweets of the same sign as the most frequent label always stay with that label. The reason as for the values of the weights and these thresholds will be described after the algorithms are explained.

The following tweets show some examples of this behaviour:

1 *INSIGHTFUL point by @Jacob_Rees_Mogg So much for the House of Commons representing the will of the voters. Too many are focused on corporate interests, but if MPs have integrity they will have to allow the No Deal Brexit now! exitVote #NoDealBrexit*

2 *The way it looks right now that'll probably need a permission slip from the #Brussels? They've been sold out by @theresa_may #Brexit*

3 *I could not agree more with @Anna_Soubry hear bloody hear! #BrexitShambles #Remain #PeopleVote*

4 *yes, but with @jeremycorbyn I fear we are sunk. Young people will not forgive #Brexit*

Tweet number 1 is another example of what has already been shown before: A Tweet with a high positive confidence, for a target with a negative overall global sentiment (Jacob Rees Mogg has an overall negative sentiment of 0.74 out of 1), stays positive because of the very high confidence output of the sentiment analysis.

The output of the sentiment analysis for Tweet 2 is positive with a small confidence (0.653). Using the global score of negative 0.81 that Theresa May has, this tweet gets correctly re-classified to negative.

Tweet 3 has a negative label output from the sentiment analysis with confidence 0.52. As the overall sentiment towards Anna Sourbry is positive of score 0.73, it gets correctly re-classified as positive.

Lastly, tweet number 4 gets initially classified as negative with a confidence of 0.6 towards Jeremy Corbyn. The ground truth for this tweet is negative. Taking into account the negative global sentiment towards this user, this tweet does not get re-classified, but rather the confidence that is negative gets increased, adding the negative value of the global score to the negative output of the sentiment analysis.

For the users whose global sentiment does not surpass the set threshold of 0.7 (4 out of 15 in our data) this heuristic is not used, and the assigned label is the label given by the sentiment analysis. The following table compares the results obtained from the individual sentiment analysis, the work of Li et al, and our approach in the 750 test tweets: As it can be seen, both the proposed algorithm of Li et al,

Sentiment Analysis	57%
Li et al	65%
Global Sentiment	67%

Table 8: Table comparing the accuracy on the test data of each of the different approaches discussed to this point

and our approach improve the performance of just the sentiment analysis. Also our approach achieves a slight performance increase over Li et al, and appart from this, it maintains some of the data integrity by not re-classifying every tweet towards a specific target to the same sentiment label. Next we will discuss the following part of our heursitic approach individually, before detailing the combination of both and exploring the results.

6.2.2 Heuristic Approach - Relationships

This part of the approach is inspired by [38]. In this paper, which was previously described, social relationships are used to improve the performance of sentiment analysis on Twitter data. This work doesn't use a signed graph where the edges represent relationships between users, but rather a follower/followee graph, where an attribute of each node is their sentiment toward a specific topic. It performs Sentiment Analysis using SVMs, altough it does not give any details of their NLP approach. The motivating intuition behind this work is that "connected" users will tend to hold a similar opinion with regards to a certain topic.

As our goal is to improve the sentiment classification towards influential users, and acknowledging the star-like structure of our graph, some tough had to be put into how to incorporate an insight like this one to this work. The task here would be to gather users from the dataset that have mentioned one of the influential accounts that are being analyzed so that those gathered users would have a high probability of being followed by other users mentioning the same influential accounts. After considering the available data, and reading some works on Twitter user filtering like [63] the following procedure was decided:

1 First the users of the dataset were filtered to keep only those whose Twitter description contained words such as Brexit, brexit, politics, journalism, news, and more. The Twitter description or biography is a short text associated with the Twitter account which describes such account.

2 Then from these users all those with less than 2000 followers were discarded, keeping only those with a large sum of followers to maximize the chance that a new user would follow them.

3 Lastly, out of these users, only the ones who had tweets mentioning one of the 15 influential accounts which are being analyzed were kept.

This led to a collection of 314 Tweets, created by 127 different Twitter accounts. All these Tweets were then hand labeled to asses the sentiment of each of these *gold users* towards the central user that they are mentioning. The following image shows some of the accounts of these gold users, along with their description. Now,

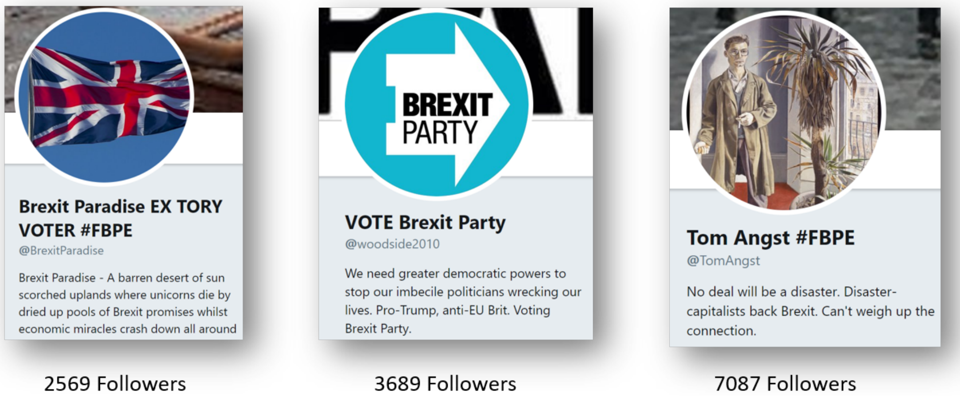

Figure 29: Twitter Profile of three of the accounts included in the gold user set, along with their number of followers.

for every one of these gold users we know the influential user or users that he has mentioned and their sentiment towards them. For every tweet towards one of these influential users, we will see if the user who posted such tweet follows any of the gold users who have mentioned the same influential user, and if so, the score of the new tweet will be affected, shifting the weighted score from the sentiment analysis towards the sign of the feeling of the gold user that is being followed towards the target.

Sometimes, the user who is making the new tweet follows various of the gold users for the mentioned target. If so, the sentiment scores (positive or negative) of all the gold users that the new tweeting user follows are added, and the final relationship sentiment will be either positive or negative depending on this sum. If the new user follows the same number of gold users with positive feeling towards the central as the gold users with negative feeling towards the same target then the sentiment analysis score will not be affected by the relationships of the user. Figure 30 depicts a diagram of this procedure:

This diagram starts with the following tweet by the user *@Santilespr*: *"@theresa_may and Parliament's biggest mistake - triggering #A50 - laid bare"*, which is classified as positive by the sentiment analysis with a confidence of 0.626, while the ground truth

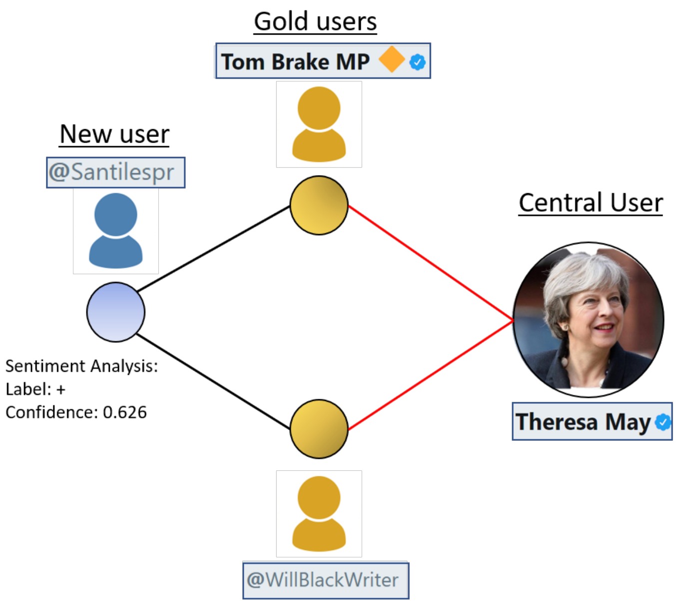

Figure 30: Diagram showig the relations between a new tweeing user *@Santilespr*, two users from the gold user set *@thomasbrake* and *@WillBlackWriter*, and a central user *@theresa_may*

is a negative sentiment towards Theresa May. From the set of gold users who have mentioned Theresa May, the new tweeting user *@Santilespr*, follows 2: *@thomasbrake* and *@WillBlackWriter*, both of whom have a negative sentiment towards Theresa May (obtained from labelling the tweets produced by the gold users), represented by the red connections from these users towards her in the diagram.

Using the weighted score from the sentiment analysis in function of the confidence like in the previous section, and the overall relationship sentiment of the gold users who are followed by the tweeting user towards the target entity we arrive at the following formula:

$$S = W(cf) \cdot Se + Re \tag{11}$$

where all the terms are the same as in formula (10) except the Global Sentiment term, which has been replaced by a relationship term which acts in the same way, shifting the results from the sentiment analysis towards the aggregated feeling of the users from the gold set who the new tweeting user follows. The term Re derived from the relationships of the user can take the values -1 if the agreggated feeling of the gold users is negative, 1 if its positive and 0 if the tweeting user does not follow any gold users who have mentioned the target of the tweet or if he follows the same number of users with a positive feeling towards that target than users with a negative feeling towards it.

Like in the previous subsection, after this score is calculated, we proceed to evaluate if its above the positive threshold or bellow the negative one to infer if the final sentiment of the tweeting user is positive, negative or neutral to the target entity, taking into account the sentiment analysis and the relationships embedded in the Twitter follower network. For the example shown in the above figure, the tweet gets correctly re-classified to negative and therefore a negative connection is constructed between the user *@Santilespr* and Theresa May. The following image depicts the final result of this process, resulting in the creation of a negative link in between them. In a similar manner to how it was done in the previous subsection, if the

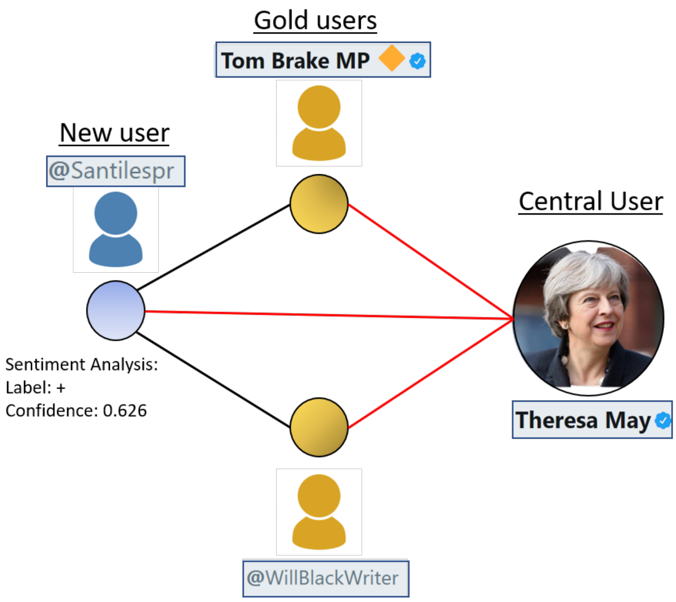

Figure 31: Diagram showing the relation new relationship made by the user *@Santilespr* to the central user *@theresa_may* by using the follower/followee relationships with two users from the gold set.

tweet in question that is being analyzed has a very strong confidence on the sign of its prediction from the sentiment analysis, even if the user who made such tweet follows various users from the gold set who have the opposite sentiment from the NLP approach, the Sentiment Analysis will prevail. However, in the case of the relationships, this situation is not very frequent, as reported by [38] users who follow each other actually tend to have the same view on the topic that is being discussed.

For the 750 Tweets of our test dataset, produced by 621 different users, 26% percent of these users (whose tweets all mentioned one of the top 15 most influential users) followed some user belonging to the gold user set who has mentioned the same central user as them. Out of these Tweets from users who followed someone in the gold set mentioning the same target 82% had the same ground truth label as the sentiment from the gold user, 15% of them were labeled neutral while the user from the gold set had a positive or negative sentiment towards the target and only 3% of them had a different ground truth label to the sentiment of the gold user.

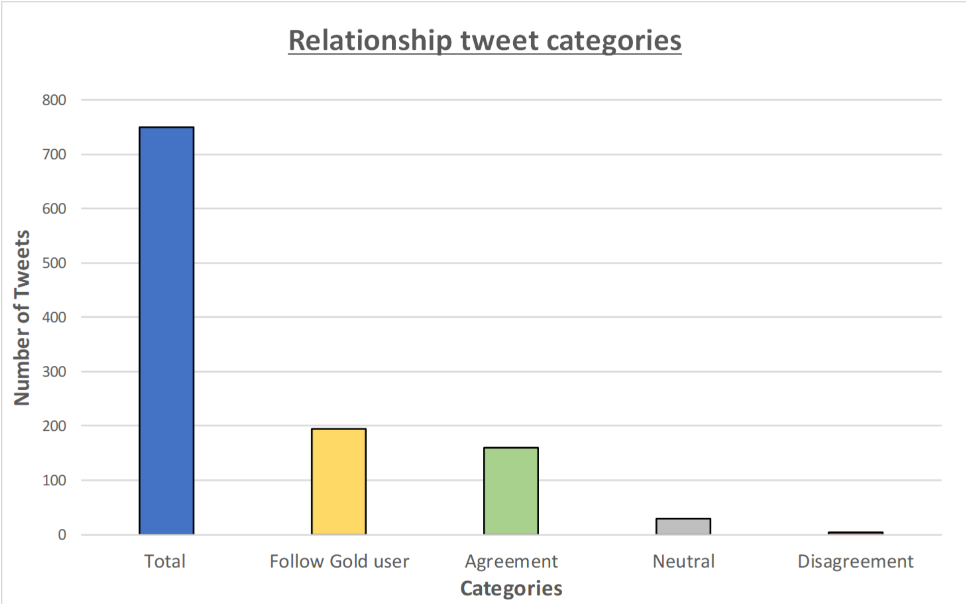

Figure 32: Graph showing the percentages mentioned in the lines above. Out of all the test tweets 26% are tweeted by a user who follows a gold user who has mentioned the same target as them. Out of these tweets 82% share the same label wit the gold user, 15% are neutral and therefore can't agree with the labels from the gold users, which are only positive or negative, and 3% disagree.

This confirms what has previously demonstrated in [38]: the probability of two connected users sharing sentiment is a lot higher than chance, therefore making this approach evidently useful in networks where we want to infer the sentiment of a user towards a topic or towards another user, and know the opinions of some of the users he follows towards that same topic or user.

The following table compares the results of the sentiment analysis, the classification using the global score, and the classification using the relationships:

Sentiment Analysis	57%
Global Sentiment	67%
Relationship	62%

Table 9: Table comparing the accuracy on the test data of each of the different individual approaches proposed in this paper and the baseline sentiment analysis

The Global Sentiment approach seems to perform better on the test data than the relationship approach, although this approach still improves the results from just the sentiment analysis, but by a smaller margin. This is probably due to the fact that the global score heuristic affects every tweet targeting a specific user if the global sentiment is above the threshold (in our dataset for 15 central users 11 got affected by the global sentiment heuristic approach which accounts 73% of the tweets towards these targets), however as we have seen before, the tweeting user-gold user relationships only affect about 1 out of every 4 users, so the effect of the relationship is not as pronounced. In the following subsection the combination of both approaches will be discussed.

6.2.3 Heuristic Approach - Combination of Global Sentiment and Relationships

After explaining each approach separately and comparing their individual performances, we will proceed to evaluate how they perform when used in combination. The reasoning and procedures for each of the parts (Global Sentiment and Relationships) apply in the same way in the combined approach than they did individually, making the same assumptions.

For the global sentiment we proceed as before, we calculate it in the same way with formula (9), and if it is above the threshold, we use to it to influence the classification. For the relationships, exactly the same procedure as before is used: for every tweet towards one of the main personalities, we see if the tweeting user follows any of the users from the gold set who have mentioned the same important personality. If so, the score of that tweet will be affected by the sentiment of the users of the gold set who the tweeting user follows towards the influential target. This time the equation defining the heuristic would be a combination of formulas (10) and (11):

$$S = W(cf) \cdot Se + Gs + Re \tag{12}$$

All the terms from this equations have been previously discussed, so they will not be explained again. The weights for the sentiment analysis are maintained. The following table shows the results for all of the previous considered approaches on our test data:

Sentiment Analysis	57%
Global Sentiment	67%
Relationship	62%
Both	67%

Table 10: Table comparing the accuracy on the test data of each of the different approaches discussed of this work, individually and combined, against the baseline Sentiment Analysis

As it can be seen from this table, both individual approaches enhance the performance of the sentiment analysis. Also, this table shows that using the combined approach has the same results as using the best of the two approaches (The global sentiment) individually. This can be explained by the fact that most tweets which have the same sentiment towards one of the influential targets as the global sentiment will have already probably been correctly re-classified by the global sentiment if they had initially been missclassified by the sentiment analysis. The users producing those tweets, taking into account the principle expressed in [38], that two users are more likely to follow each other if they share the same feeling towards a target/topic, will then probably follow a user from the gold set with the same feeling towards the target mention, and therefore push the score of those tweets even more towards the sentiment that is more common for that target, but not re-classify them. This confirms the findings that two users with same opinion on a topic are more likely to follow each other, and that two users who follow each other are more likely to have the same opinion on a topic than if they did not share a relationship in the network.

6.2.4 Heuristic approach - Weight and Threshold explanation

The goal of this heuristic approach is to do something similar to [37], but in a less drastic way, while also incorporating relationship information like in [38]. As mentioned earlier, the full relabeling of tweets towards an individual if the aggregated sentiment of the tweets in the training set was above a certain threshold performed in this paper, leaves out any opinion that is contrary to the general belief. Also [37] does not take into account the existence of neutral tweets. Inspired by [36], the idea of giving different weights to the sentiment analysis depending on the confidence of it's classification was obtained. The established weights and thresholds for the heuristic approach were set taking into account these two factors (tweets with high confidence and label opposite to that of the global sentiment, and neutral tweets), in order for the algorithm to be able to keep and correctly classify these two types of tweet instances and achieve the highest possible accuracy.

First of all, [37] uses 0.8 as a threshold to relabel all the tweets to the label of the most frequent sentiment. This project relaxes that threshold to 0.7 so that more tweets get affected by this heuristic, taking into account that if almost 3/4 of the training tweets towards one specific user report a certain polarity label, then it is enough to infer that the global sentiment towards that user has said polarity, and also that with the previous considerations not all tweets will be relabeled. This is the threshold that is used for deciding if the global sentiment score will be used to assist the classification of the Sentiment analysis, and must not be confused with the positive and negative thresholds described in the following lines, that serve the purpose of classifying tweets with scores above or bellow them as positive, negative or neutral. This global sentiment threshold was set to such value by calculating the global sentiment scores for each of the influential users of our dataset, and investigating if incorporating such global score to the sentiment analysis would improve the classification or not. Figure 33 shows the global sentiment score calculated using the 50 training tweets, for most of the influential users in our dataset, where the red bars signify that this score is negative and the green bars that it is positive.

For every user with a global score greater than 0.7 (Anna Sourbry is the closest one wit a positive score of 0.707) the classification improved when incorporating

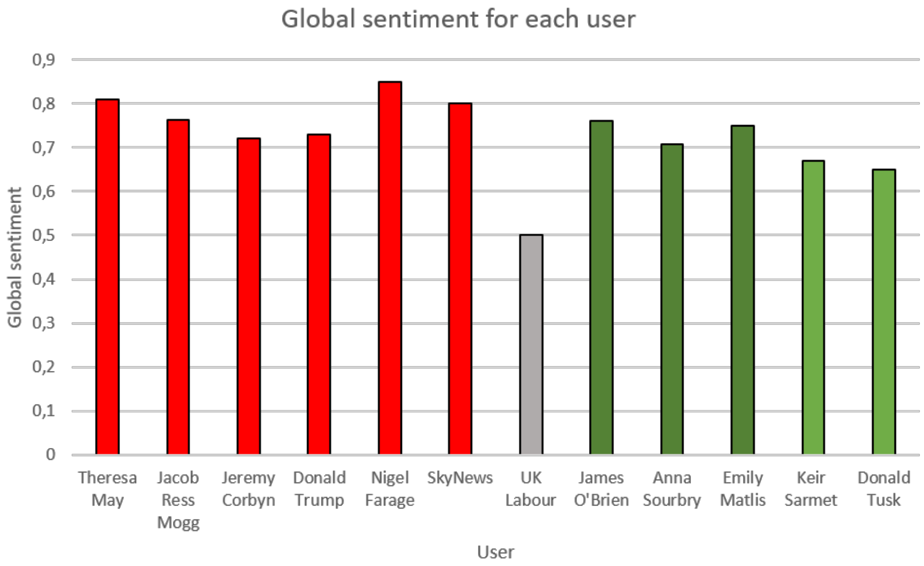

Figure 33: Global Sentiment Score for some of the users of our dataset. The red bars represent this score is negative and the green bars represent this score is positive.

the global sentiment. However, some users with scores under this threshold (Keir Starmer, Donald Tusk, UKLabour) the classification actually became worse if the global sentiment was incorporated, indicating that the shift towards a specific sentiment that incorporating the global score creates is not beneficial if the user does not have a heavy general opinion of such given sentiment. Because of this, the Global Sentiment Threshold, to decide whether or not to incorporate the global sentiment score to the sentiment analyis was set at 0.7.

As mentioned previously, to keep the first type of tweets (those that have a strong feeling of the opposite sign to the global sentiment) the weight for the sentiment analysis term for these tweets is set to a high value, in order to overcome the push or shift of the tweet's score to the direction of the global sentiment.

The rest of the weights and score thresholds are set in order to keep some neutral tweets, and not relabel all the tweets that are not included in the first category, or that already have such label to the most frequent label in order to maximize the accuracy of the algorithm.

The score thresholds are set in such a way that if the global sentiment is positive, if the score is greater than 0.2 then the assigned label is positive, and if the global score is negative and the score is smaller than -0,2 then the label is negative. For the first case, the negative threshold is set at -1, to allow tweets in between 0.2 and negative 1 to be classified as neutral, and for the second case the positive threshold is set at 1 to allow any tweet with a score value between -0.2 and 1 to be classified as neutral. The classification rule is described in figure 34.

Table 11: Table showing the different weights given to the sentiment analysis depending on the confidence of such analysis.

Weight	Confidence	Sign(label) = Sign(Gs)
2	1	No
1	1	Yes
0.75	[0.9,1)	-
0.5	[0.5,0.9)	-
0.25	<0.5	-

If the Score < -negative threshold → label as negative.

If the Score > positive threshold → label as positive.

Else → label as neutral.

Figure 34: Classification rule for the Heuristic approach

To fix the thresholds that would yield the best possible results, a broad range of positive/negative thresholds were swept over, and their performance was evaluated. For negative targets the positive threshold was fixed and then a range of negative thresholds were swept, calculating the accuracy, and then repeated for different positive thresholds. For the targets with a positive overall score the same was done but fixing the negative threshold instead and testing different positive thresholds for that fixed negative one. Figure 35 shows an example of this for negative targets, with a fixed positive threshold of 1, testing different negative thresholds.

Table 11 shows again the different weights assigned depending on the confidence. If the confidence of the classification is opposite to the one of the most frequent sentiment, and the confidence is between 0.9 and 1, then the tweet will either get re-classified as neutral, or to the sign of the most frequent label if the global score for that target has is very high. For tweets with a confidence of less than 0.9 and sign opposite to the one of the most frequent label for the target, they will all get relabeled to such sign.

While these weights were set taking into account the individual approaches (Global Sentiment and Relationships) they were maintained for the combined approach, as the only difference would be that for certain tweets, for which the tweeting user follows a user from the gold set with the same feeling towards the mentioned target as the global feeling, even the tweets with a high confidence value of the opposite

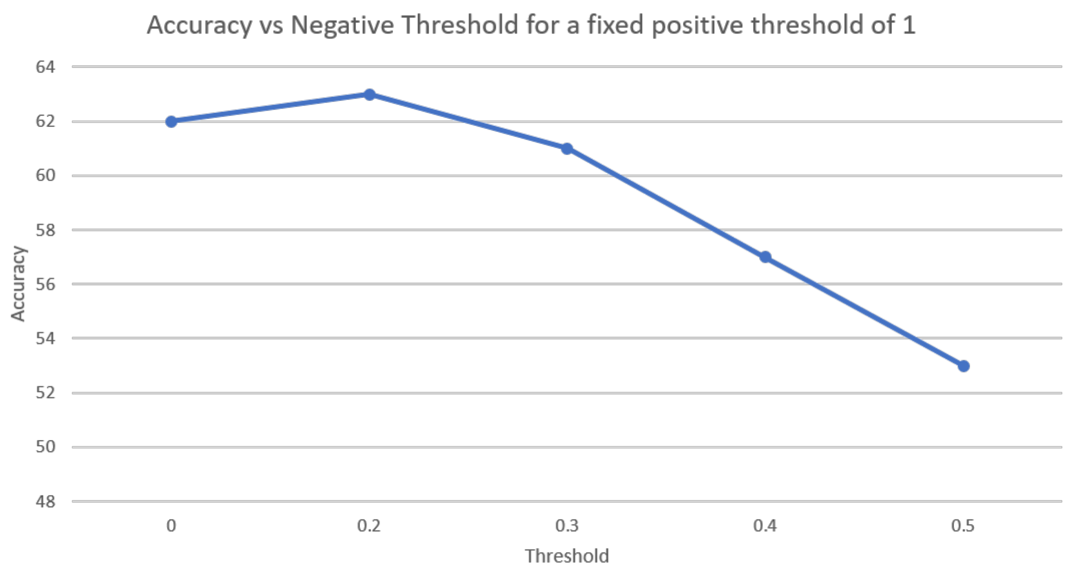

Figure 35: Accuracy evaluated for different negative thresholds for a fixed positive threshold of 1

sign would get reclassified either to that sign or as neutral. This is in line with the reasoning behind the relationship heuristic: if a user has a strong feeling opposite to the general belief, most of the time he will not follow people with the opposite opinion which are included in the gold set, and if he does then it probably means that the tweet was incorrectly missclassified by the Sentiment analysis, despite having a high confidence, which could be the case for example of sarcastic tweets, that get classified as positive with a very high confidence by the sentiment analysis, despite the ground truth being negative.

6.2.5 Machine Learning Approach

After exploring the different heuristic approaches individually and also combined, the last part of this work consists on trying Machine Learning approaches using features coming from the sentiment analysis (label and confidence), graph (global sentiment, connections to positive and negative gold users, centralitys, and more) and the actual text of the tweet, which when fed to the initial Sentiment Analysis algorithm was not exploited or preprocessed in any way, but could contain certain parameters that could be of use to determine the sentiment label (positive, neutral or negative) towards a target, like the position of the mentioned target, the count of URLs or the number of hashtags in the sentence. To do the initial sentiment analysis,the data was only adapted to fit the format the algorithm of Zhang et al required, but no further information was extracted from it.

As mentioned in the previous paragraph, we can divide the features fed to the Machine Learning model in three main groups:

1 Features extracted from the initial targeted sentiment analysis used: The algorithm of Zhang et al.

2 Features extracted from social graph structures.

3 High level features extracted from the Tweet text like number of Hashtags, position of mentioned target, number of URLs...

The features derived from the Sentiment Analysis are already known and need to explanation: the label predicted by the model and the confidence in such label.

For the features extracted from the graph, a more detailed explanation is needed. First we will use the same Global Score that was used in the previous heuristic approach and calculated from 50 hand labeled tweets for each of the influential targets. This time however, in contrast to what happened in the heuristic approaches, this global score will always have an influence in the classification, as even the global scores of those influential users for whom it does not surpass the threshold used in the previous section are added as features for the machine learning model. The second set of features derived from the graph is related to the relationships heuristic approach. These features are the number of connections to users from the gold set with a positive sentiment towards the mentioned target that is mentioned in the tweet to be classified, and the number of connections to users of the gold set of the user who is making the tweet to users of the gold set with negative sentiment towards the target that is being mentioned.

Also, from the graph both, the degree centrality which is a measure of how influential or important a node is in a graph, and the eigenvector centrality, which measures the quality of the connections of a node, are extracted for the 15 central users which are being analyzed and used as features as well.

Lastly, features from the tweet text itself were extracted. This is done because for the previous sentiment analysis, only the word embeddings of each word in the tweet were fed to the algorithm, leaving out information which could be of use for detecting the polarity of the tweet. These features are the number of URLs in the tweet text, the number of hashtags, if there are questions or not in the tweet, and the length of the tweet. Also, a feature was added to reflect if the Tweet contains the word *"via"* or not, as during the manual labelling of the tweets, many tweets where found with the following structure: *"The #Brexit hell about to get worse, via @SkyNews"*, which represents a tittle for a new or an article being posted by one of the news sources, and therefore does not indicate any kind of sentiment towards such target.

These features were then fed to 3 different machine learning classifiers in order to see if by using this kind of approaches the performance of the plain sentiment analysis could be improved. These approaches are the already described Support Vector Machines, Logistic Regression and AdaBoost. Before showing the results for each model and comparing them, Logistic Regression and AdaBoost will be briefly explained, as SVMs were explained in chapter 2.

Logistic regression is one of the most famous machine learning algorithms used for classification. It uses a sigmoid function to output a value between 0 and 1 for any input value, which can then be used to predict the label for that input value. In a binary classification problem with classes A and B, if the output value of the sigmoid function is above certain threshold then the new data point gets classified as class A, and if it is below such threshold it gets classified as class B. As Logistic regression is mostly used for binary classification (that is two class classification) and our problem is a multi-class classification problem, for our approach we will use a one versus all or one versus the rest strategy, which consist on training one classifier per class, with the instances of that class as positive and the instances of all the other classes as negative, and then choosing the solution which yields the highest probability of the data point belonging to such class. The following figure depicts an image of the sigmoid function used by Logistic Regression.

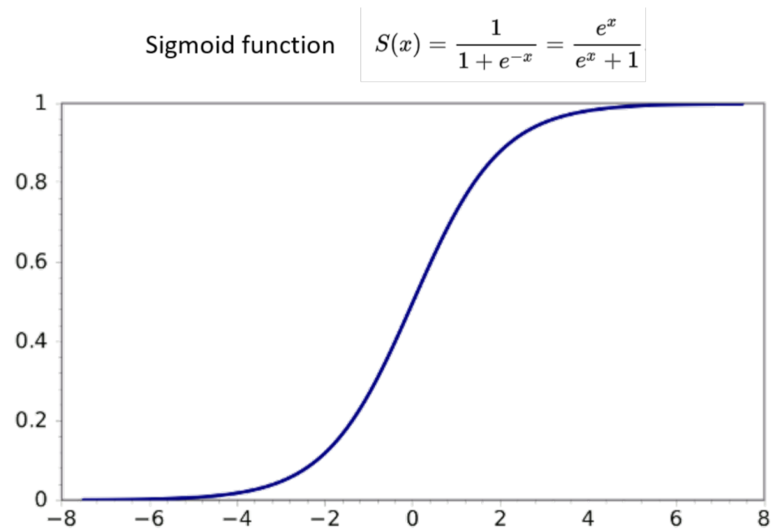

Figure 36: The sigmoid function used in Logistic Regression. For any input value, the sigmoid function outputs a value between 0 and 1.

The other Machine Learning Model that has been used, but has not yet been explained is AdaBoost, which is short for Adaptive Boosting. AdaBoost belongs to a family of algorithms called *ensemble learners*, which produce strong predictive models by the combination of many weak ones. In these models the output of the weak algorithms are combined in a weighted sum that represents the final output of the improved classifier. The Adaptive part of AdaBoost comes from the fact that the progression of weak models are trained by favouring those instances of the data that were missclasified by the previous weak models. A number of iterations T is set, and after each iteration the weights of the different data points get updated by giving more weight to the instances that were miss-classified by the classifier constructed in the previous iteration. All the points start with the same weight, and after each iteration, when a new classifier is trained, their weights get updated. After T iterations we have trained T classifiers, and their weighted sum constitutes the output of the final boosted model. These weights, assigned to each classifier depend on how well the classifier performed when it was evaluated.

After briefly explaining these algorithms, the following table shows their performance on the labeled tweets, with the features from the three categories mentioned previously on this section included, using a 10 fold cross-validation, and comparing it to the basic sentiment analysis and to the top performing approach from the heuristics; the combined approach of global sentiment and relationships:

Sentiment Analysis	57%
Combined Heuristics	67%
SVMs	62%
Logistic Regression	64%
AdaBoost	67%

Table 12: Table showing the accuracy of the different Machine Learning models used to enhance the sentiment analysis in comparison to the baseline SA and to the best performing heuristic model.

As it can be seen from the table, all the machine learning approaches perform similarly to each other, with SVMs having the worst performance and AdaBoost the best, and all perform better than the baseline sentiment analysis. Also, the top performing one shows the same result as the combined heuristics approach, which beats the baseline by 10% accuracy. Machine Learning models, because of the automatic weight setting, and the ability to incorporate additional features than the heuristic approach in an easier manner, should generally beat these kind of methods. However, because of the small amount of data available to train and test the Machine Learning algorithms, its very possible their its performance is not optimal, and that the results would be better if more labeled data was available.

Feature selection using the Mutual Information approach was used in order to see the features that are the most related to the target labels. Out of the 14 features used, the feature that was most related to the target label was the global sentiment for each of the influential users. The other features derived from the graph (number of connections to users within the gold set that mentioned the same influential target with a positive or a negative label and the centralities) showed some correlation with the label of the tweet but less than the global sentiment and the label and confidence features derived from the sentiment analysis.

6.3 Error analysis

For the best performing Machine Learning model (AdaBoost), 60% of the errors come from incorrect classification of neutral tweets as either positive or negative. This highlights the previously mentioned point of how hard it is to address the subjectivity of the tweets. This percentage however, as big as it might seem, has been reduced from the straightforward sentiment analysis, where for the tweets classified with a confidence of 1, 77% of the errors are neutral tweets being labeled either as positive or negative. This indicates that some of the features used in the Machine Learning classifiers with the intention to enhance the sentiment analysis are good for discriminating whether a tweet is polarized or not.

Also, it seems that the Sentiment Analysis performs worse at the classification of tweets towards a negative target than towards a positive one, which is possibly due to sarcasm, subtle negativity and the difficulties mentioned in chapter 4 about images and URLs infering sentiment along with the text, which most time is a negative sentiment. The following figure shows the correctly classified tweets out of the 50 test tweets for different influential users by the sentiment analysis and by the heuristic approach depending on whether they have an overall positive sentiment or a negative one.

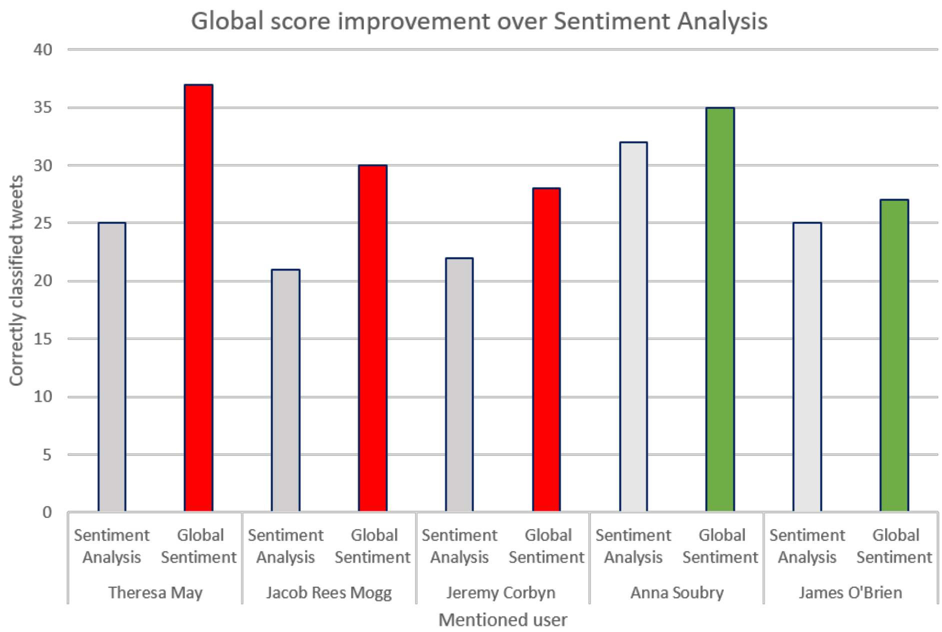

Figure 37: Comparison of the performance of the SA and the improvement by the heuristic approach on different targets

This figure shows how the sentiment analysis performs worse on those targets with an overall negative feeling than a positive one, and therefore the increase of performance from using any of the approaches to enhance this sentiment analysis is greater.

6.4 Sentiment Towards Influential users

With this improved Sentiment Analysis, we can now infer the sentiment towards the main actors of our network with a higher confidence than by just using the Sentiment Analysis. Chapter 5 presented a list of the sentiment towards each of the users evaluated using the 0.9 confidence graph, showing that the number of positive links for each of the users was higher than the number of negative links for all the users except 1. Taking the same tweets and applying the combined heuristic approach, and removing the neutrals, we can see out of these remaining tweets how many are positive and how many are negative, in order to correctly infer the overall sentiment towards each of them and compare the results. The following figure shows a comparison between the labels obtained by the sentiment analysis on the tweets that were used to construct the 0.9 confidence graph, and the results obtained by the combined heuristic approach on those same tweets, removing the neutral labels obtained.

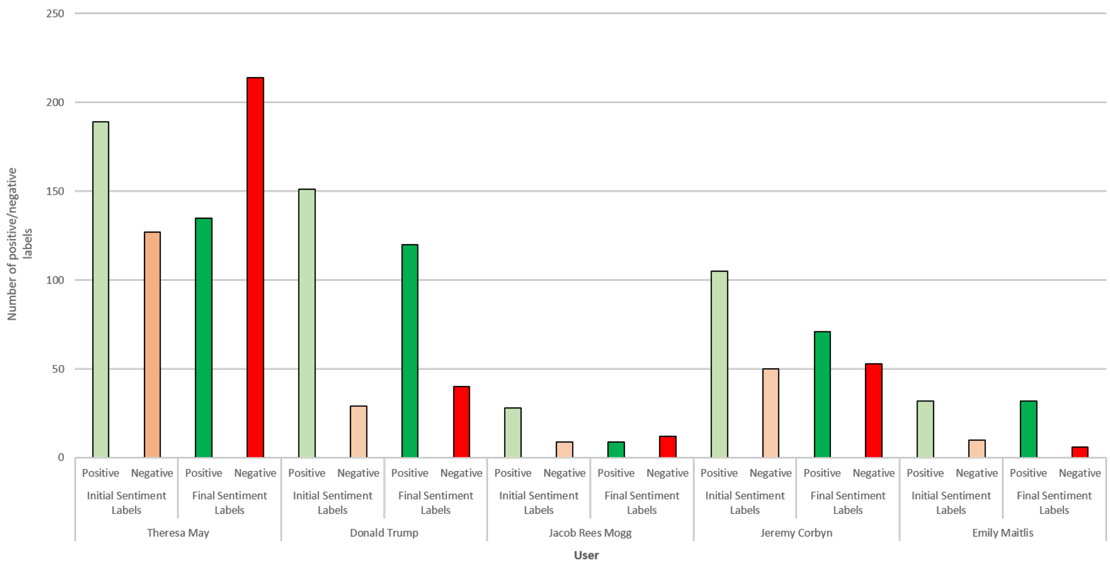

Figure 38: Initial sentiment labels output by the Targeted Analysis and Sentiment Labels on those same tweets achieve by the combined heuristic approach

It can be seen that some users who were initially regarded as having more positive connections than negative ones (Theresa May, Jacob Rees Mogg), get their aggregated sentiment changed from positive to negative, while having a higher accuracy on this classification.

After having presented the different strategies for improving the sentiment analysis, the next chapter will make a summary of the project, and discuss some future lines of work that could be tackled.

7 Summary, Discussion and Future lines of work

This chapter closes this project by making a short summary of the work done and discussing some of the results, future lines of work and possible improvements.

7.1 Summary and Discussion

This work has addressed several issues:

First the extraction of useful information from Twitter data without any preprocessing, showing that a lot of valuable information can be obtained without the need to use complex algorithms.

Then, a methodology for creating a signed graph from Twitter, where the nodes of the graph are Twitter users and the edges of the graph represent a positive or negative relationships in between those users, was developed and implemented, showing that a meaningful signed graph can be constructed, even if the social application that is being analyzed has no explicit way to create negative interactions. This is possible by using sentiment analysis, or in the case it is needed its targeted variant, for extracting the sign of user relationships out of the text embedded in the social applications.

The structure of this graph was analyzed, and from it a possible use case was derived: infering the sentiment towards popular actors in the networks, represented by the nodes with highest degree, which are the Twitter accounts that have been mentioned the most on a certain topic. Different ways for evaluating the influence of users in the network were debated, and once the top influential users were discovered a whole framework for calculating the sentiment of the rest of the users of the network towards them was developed and implemented.

As the sentiment analysis needed to construct this graph is a very challenging task, various approaches for improving this sentiment analysis were developed and implemented, improving the performance of the baseline sentiment analysis, and showing that using the social structure intrinsic to the applications where this text is extracted from can greatly improve the performance of the plain Sentiment analysis. The two best approaches developed, which have approximately the same results, show a 10% increase in the classification over the plain sentiment analysis. Out of these, the Machine Learning approach could be further improved, looking for more complex features from the graph to incorporate, and using a larger amount of training data.

Various previously discovered topics have been confirmed in this work. First, the greater difficulty of classifying neutral or subjective content from that carrying a certain polarity. Secondly, that Twitter follower/followee relationships are a good indicator of shared sentiment on a topic, entity or topic/entity level between the two users who share such relation. Third, that the hypothesis that the overall sentiment for a topic/entity is consistent across most individuals has been confirmed, by the percentage of influential users in our dataset that have an overall global sentiment larger than the established threshold, and the improved classification with regards to the sentiment analysis for the users with those global sentiment scores.

Lastly and most important, it has been demonstrated that despite the difficulties of sentiment analysis for Twitter data, which were thoroughly described in Chapter 4, if this sentiment analysis is combined with social information derived from the connections inside of the social network, its performance can be greatly increased. This work addressed the issue on Twitter data, but the same principles could be applied to any other social network with text interactions among users and some kind of friendship mechanism in between the users.

7.2 Future lines of work

One of the reported issues that limited the performance on the sentiment analysis, is the use of sarcasm within the tweet's text. Specially for targets with an overall negative sentiment, sarcastic tweets were classified as positive tweets with a very high confidence, while they should have been negative, and therefore none of the proposed approaches could successfully reclassify them. To tackle this issue several things could be done. The first solution would be to research for features in the tweet's text that are indicative of sarcasm and incorporate them to the Machine Learning model that has already been constructed. The second solution would be to use a model that is specifically trained for recognizing sarcasm after the first sentiment analysis like [64], and toggle the sign of those tweets that have a high confidence from the sentiment analysis and have been recognized as sarcastic by the sarcasm detection algorithm.

Another possible improvement, specifically aimed at the Machine Learning approach would be to label more data, in order to see if with a larger training dataset the algorithm would perform better, as it is very possible that it doesn't reach its optimal performance because the small availability of training data.

A heavier preprocessing step could be done before feeding the initial un-labeled tweets to the Targeted Sentiment Analysis algorithm. Despite this work using word embeddings to automatically extract the features for the Bi-directional neural network attention model, a slight cleaning of the tweet's text might increase the performance of the sentiment analysis.

Lastly, in chapter 3, some differences were found between the datasets belonging to political topics (#Brexit and #Venezuela) and the ones belonging to amusement events (#Oscars and #GameOfThrones), like the percentages of retweets and the mean length of the tweets. Doing a more throught investigation of this would be very interesting, to confirm these results and get some insights into why this happens. Also, it would be interesting to construct signed graphs using the sentiment analysis improvement for this other topics, in order to be able to compare them.

References

[1] Hassan, A., Abu-Jbra, A., Radev, D., *Extracting Signed Social Networks From Text*, In Proceedings of the TextGraphs-7 Workshop at ACL, 2012.

[2] Go, A., Bhayani, R., Huand, L., *Twitter Sentiment Classification using Distant Supervision*, Stanford University, 2009.

[3] Kagan V., Stevens A., Subrahmanian V., *Using Twitter Sentiment to Forecast the 2013 Pakistani Election and the 2014 Indian Election*, in IEEE Intelligent Systems, vol. 30, no. 01, pp. 2-5, 2015.

[4] Freeman, L. *The Development of Social Network Analysis*, 2004.

[5] Antoniadis, I., Charmantzi, A., *The Application of Social Networking Analysis in Marketing: A Case Study of a Product's Page in Facebook* Proceedings of the 2nd International Conference on Contemporary Marketing Issues (ICCMI), 2014.

[6] Fortunato, S., Castellano C., *Community Structure in Graphs*, Encyclopedia of Complexity and System Science, 2008.

[7] Kunegis, Jérôme Schmidt, Stephan Lommatzsch, Andreas Lerner, Jürgen De Luca, Ernesto Albayrak, Sahin, *Analysis of Signed Graphs for Clustering, Prediction and Visualization*, Spectral. Proc SDM. 559, 2010.

[8] Gao, Wei Wu, Hualong Siddiqui, Muhammad Baig, A., *Study of Biological Networks Using Graph Theory*, Encyclopedia of Saudi Journal of Biological Sciences. 25, 2017.

[9] Even-Tzur, G., *Graph Theory Applications to GPS Networks*, GPS Solutions. 5, 2001.

[10] Deswall, S., Singhrov, A., *Application of Graph Theory in Communication Networks*, International Journal of Application or Innovation in Engineering Management (IJAIEM), 2012.

[11] Heider, F., *The Psychology Of Interpersonal Relation*, 1958.

[12] Lim, Y Rubineau, B., *The Psychology Of Interpersonal Relation*, 2013.

[13] Girdhair, N., Bharadwaj K., *Signed Social Networks: A Survey.*, 2017.

[14] Yang, S., Alexander .B, Long, B., Chang, Y., *Friend or frenemy? Predicting signed ties in social networks*, SIGIR'12 - Proceedings of the International ACM SIGIR Conference on Research and Development in Information Retrieval, 2012.

[15] Leskovec, J., Huttenlocher, D., *Predicting Positive and Negative Links in Online Social Networks*, Proceedings of the 19th International Conference on World Wide Web, WWW, 2010.

[16] Pushkar, S., Singh, K., Virajita, .S, Debasis, .D, *Algorithm for Prediction of Links using Sentiment Analysis in Social Networks*, 2016.

[17] Wikipedia contributors. *"Twitter." Wikipedia, The Free Encyclopedia.* Wikipedia, The Free Encyclopedia, 19 May. 2019. Web. 24 May. 2019

[18] Garimella, K., Gionis, A., Parotsidis, N., Tatti, N., *Balancing Information Exposure in Social Networks*, 2017.

[19] Gerimella, K., De Francisci Morales, G., Gionis, A., Mathioudakis, M., *Quantifying Controversy in Social Media*, in ACM Transactions on Social Computing. 1, 2015.

[20] Ten Thij, M., Ouboter, T., Worm .D, Litvak, N., Van Den Berg, H., Bhulai, S., *Modelling of Trends in Twitter Using Retweet Graph Dynamics*, 2015.

[21] Riquelme, F., Gonzalez-Cantergiani, P., *Measuring user influence on Twitter: A survey*, In Journal of Information Processing and Management, 52, 2016.

[22] Fraisier, O., Cabanac, G., Pitarch Y., Besanc, R., Mohand, B., *#Elysee2017fr: The 2017 French Presidential Campaign on Twitter*, In Proceedings of the Twelfth International AAAI Conference on Web and Social Media, 2018.

[23] Gaydhani, A., Doma, V., Kendre, S., Bhagwat, L,. *Detecting Hate Speech and offensive Language on Twitter using Machine Learning: An N-gram and TFIDF based Approach*, 2018.

[24] Davidson, T., Warmsley, D., Macy M., Weber, I,. *Automated Hate Speech Detection and the Problem of offensive Language*, 2017

[25] Deitrick, W., Valyou, B., Jones, .W, Timian, J., Hu, W., *Enhancing Sentiment Analysis on Twitter Using Community Detection*, In Communications and Network, 5, 2013.

[26] Himelboim, I., Smith, M., Rainie, L., Shneiderman, B., Espina, C., *Classifying Twitter Topic-Networks Using Social Network Analysis*, In Social Media + Society, 2017.

[27] Pang, B., Lee L., Vaithyanathan, S., *Thumbs up? Sentiment Classification Using Machine Learning Techniques*, in EMLP, 10, 2002.

[28] Jawad, K., Aftab, A., *Sentiment Analysis at Sentence Level for Heterogeneus Datasets*, 2016.

[29] Wang, S., Manning C., *Baselines and Bigrams: Simple, Good, Sentiment and Topic Classification*, In Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics, 2012

[30] Zainuddin, N., Selemant, A., *Sentiment analysis using Support Vector Machine*, I4CT 2014 - 1st International Conference on Computer, Communications, and Control Technology, Proceedings, 2014.

[31] Pennington, J., Socher, R., Manning C., *Glove: Global Vectors for Word Representation*, In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing EMNLP 14, 2014.

[32] Asgari-Chenaghlu, M., *Word Vector Representation, Word2Vec, Glove and many more explained*, 2017.

[33] Altun, G., W. Harrison, R., *Machine Learning and Graph Theory approaches for Classification and prediction of protein structure*, 2019.

[34] Latouche, P., Rossi F., *Graphs in Machine Learning: an introduction*, In Proceedings of the 23-th European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN 2015), 2015.

[35] Leskovec, J., Huttenlocher, D., Kleinber, J., *Signed Networks in Social Media*, Conference on Human Factors in Computing Systems, 2010.

[36] West, R., Paskov, H., Leskovec J., Potts, C., *Exploiting Social Network Structure for Person-to-Person Sentiment Analysis*, In Transactions of the Association for Computational Linguistics, 2014.

[37] Li, H., Chen, Y., Muresan, S., Zheng, D., *Combining social cognitive theories with linguistic features for multi-genre sentiment analysis*, In Proceedings of the 26th Pacific Asia Conference on Language, Information and Computation, PACLIC 2012, 2012.

[38] Chenhao, T., Lillian L., Jie T., Long, J., Ming, Z., Ping, L., *User-level sentiment analysis incorporating social networks*, In Proceedings of KDD 2011, 2011.

[39] Grandjean, M., *A social network analysis of Twitter: Mapping the digital humanities community*, 2016.

[40] Sadri, A., Hasan, S., Ukksuri, S., Suarez Lopez, J., *Analyzing Social Interaction Networks from Twitter for Planned Special Events*, 2017.

[41] Lerman, K., Ghosh, R., *Information Contagion: An Empirical Study of the stread of News on Digg and Twitter Social Networks*, in the proceedings of The International Conference on Weblogs and Social Media 2010, 2010.

[42] Kim, YH., Seo, S., Ha YH., Lim S., Yoon, Y., *Two Applications of Clusterng Techniques to Twitter: Community Detection and Issue Extraction*, on Discrete Dynamics in Nature And Society 2013, 1-8, 2013.

[43] Java, A., Song, X., Finin T., Tseng, B., *Why we Twitter: Understanding microblogging usage and communities*, In Proceedings of the 9th WedKDD and 1st SNA, 2007.

[44] Twitters Streaming and REST Apis, accessed 24 May, 2019, <https://developer.twitter.com/en/docs.html>

[45] Botometer API for Bot detection among Twitter users, accesed 24 May 2019 <https://botometer.iuni.iu.edu/!/>

[46] Tweepy, a Python Library to access Twitters APIs, accessed 24 May 2019, <https://www.tweepy.org/>

[47] TextBlob: Simplified text processing, accessed 24 May 2019, <https://textblob.readthedocs.io/en/dev/>

[48] PararelDots AI API For Sentiment Analysis, accessed 24 May 2019, <https://www.paralleldots.com/sentiment-analysis>

[49] Spacy: Industrial-Strength Natural LanguageProcessing in Python, accesed 25 May 2019, <https://spacy.io/>

[50] Saif, H., Fernandez, M., He, Y., Alani, H., *Evaluation Datasets for Twitter Sentiiment Analysis. A survey and a new dataset, the STS-Gold*, in Proceedings of the CEUR Workshop 1096, 2013.

[51] Jiand, L., Yu, M., Zhou, M., Liu, X., Zhao, T., *Target-dependent Twitter Sentiment Classification*, In Proceedings of The 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, 2011.

[52] Dong, L., Wei F., Chuanqi, T., Tang, D., Zhou, M., Xu, K., *Adaptive Recursive Neural Network for Target-dependent Twitter Sentiment Classification* in the Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics, ACL 2014, 2014.

[53] Chen, L., Wang, W., Nagarajan M., Wang, S., Sheth, A., *Extracting Diverse Sentiment Expressions with Target-Dependent Polarity from Twitter*, in the Proceedings of the 6th International AAAI Conference on Weblogs and Social Media, 2012.

[54] Vo, D., Zhang, Y., *Target-Dependent Twitter Sentimentt Classification with Rich Automatic Features*, Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence (IJCAI 2015), 2015.

[55] Tang, D., Qin, B., Feng, X., Liu, T, *Effective LSTMs for Target-Dependent Sentiment Classification*,In the Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics, 2016.

[56] Zhang, M., Zhang, Y., Vo D., *Gated Neural Networks for Targeted Sentiment Analysis*, In the Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence (AAAI-16), 2016.

[57] Liu, J., Zhang, Y., *Attention Modeling for Targeted Sentiment*, In the Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics, 2016.

[58] Bahdanau, D., Cho, K., Bengio, Y., *Neural Machine Translation by Jointly Learning to Align and Translate.* CoRR, 2015.

[59] MPQA Opinion Corpus Release Page, accessed 24th May 2019, <https://mpqa.cs.pitt.edu/corpora/mpqa_corpus/>

[60] Mitchell, M., Aguilar, J., Wilson T., Van Durme, B., *Open Domain Targeted Sentiment*, In the Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing, 2014.

[61] Kwak, H., Lee, C., Park, H., Moon, .S, *What is Twitter, a Social Network or a News Media?*, In Proceedings of the 19th International Conferenc on World Wide Web, 2010.

[62] L.Wick, M., Rohanimanesh, K., Bellare, K., Culotta, A., Mccallum, A., *SampleRank: Training Facotr Graphs with Atomic Gradients*, in the Proceedings of the International Conference on Machine Learning, 2011.

[63] Haustein, S., Costas, R., *Identifying Twitter audiences: who is tweeting about scientific papers*, ASIST SIG/MET Metrics 2015 workshop, 2015.

[64] Zhang, M., Zhang, Y., Fu, G, *Tweet Sarcasm Detection Using Deep Neural Network*, Published in COLING, 2016.