Master's thesis

NTNU
Norwegian University of Science and Technology

Kristian Sandøy Lyster

Twitter as a communication tool for the 2019 Spitzenkandidaten

Under which conditions do they gain interaction?

Master's thesis in European Studies
Supervisor: Pieter de Wilde
May 2019

NTNU
Kunnskap for en bedre verden

Kristian Sandøy Lyster

Twitter as a communication tool for the 2019 Spitzenkandidaten

Under which conditions do they gain interaction?

Master's thesis in European Studies
Supervisor: Pieter de Wilde
May 2019

Norwegian University of Science and Technology

NTNU
Norwegian University of
Science and Technology

Abstract

This thesis is a study of how the 2019 Spitzenkandidaten were able to gain interaction on Twitter prior to the European elections taking place in May 2019. By both qualitatively and quantitatively assessing tweets, this thesis looks at how the candidates used Twitter, typical traits of the public discussing the elections, and compares how the candidates gained interaction on Twitter. The chosen time periods were 15th December 2018 to 15th March 2019 for the tweets posted by the candidates and 20th to 28th April 2019 for the tweets posted by the public. The thesis finds that all candidates all had different personalities and strategies on Twitter, even the ones representing the same political party. They were all different when it came to language patterns, how frequently they would use hashtags, and to what extent they would reply to other users. Furthermore, the results showed that the people tweeting about the elections mostly represented Western European member states and that they were generally quite neutral and objective in their tweet sentiments. Finally, this thesis found that some of the Spitzenkandidaten's tweet topics such as Brexit and climate change were universally more likely to receive interactions than other topics. It also found that publishing tweets in English and frequently using hashtags were successful strategies for gaining interaction for most of the candidates. Despite these findings, the data was not sufficient to conclude whether other aspects of the candidates' Twitter usage affected the number of interactions they gained. With this, it seems that there is not yet a golden route for successfully gaining interaction as a Spitzenkandidat on Twitter and that their interaction rates were likely influenced by external factors as well.

Acknowledgments

First and foremost, I would like to express my sincere gratitude to my supervisor, Pieter de Wilde for your essential feedback and support during the writing process. Thank you for always staying positive and encouraging me to keep going.

Special thanks to Jakob Karevold Grønhaug for helping me create the scripts for this thesis. Thank you for staying patient with me despite my lack of knowledge about programming. I will forever be grateful for all the hours you put in to help me with this thesis.

It has been a joy to study with my fellow students in the Master's programme. Thanks for all the laughter, joy, and great company throughout the last two years. Special thanks to Ørjan Sivertsen and Tony Marås Back for your positive spirits and for keeping up with our workout routine.

I am also very grateful to the Historical Institute at NTNU. Thank you so much for funding my trip to Bamberg, Germany to take part in the ECPR Winter School. It was a great experience and it made gathering the data from Twitter a much simpler task.

I would also like to thank Lars Terje Husbyn for taking the time to proofread my thesis. Much appreciated. All errors in this thesis are mine alone. Finally yet importantly, I would like to thank my parents for your unconditional support throughout my studies and for always believing in me.

Kristian Sandøy Lyster

Trondheim, May 14th 2019

Table of contents

List of figures

List of tables

Abbreviations

ACRE	Alliance of Conservatives and Reformists in Europe
API	application programming interfaces
ALDE	the Alliance of Liberals and Democrats for Europe
EC	European Commission
ECR	European Conservatives and Reformists Group
ECSC	European Coal and Steel Community
EFA	European Free Alliance
EFDD	Europe of Freedom and Direct Democracy
EGP	European Green Party
EL	European Left
ENF	Europe of Nations and Freedom
EP	European Parliament
EPP	European People's Party
EU	European Union
NLTK	Natural Language Toolkit
PES	Party of European Socialists
SMA	social media analytics
SNP	Scottish National Party
SNS	social networking site
UK	United Kingdom

1. Introduction

"This time it will be different", was the slogan of the European Parliament (EP) when they released the Spitzenkandidaten process for the May 2014 European elections. With this new process, the political parties in the EP could select candidates for the European Commission (EC) presidency that would be directly elected by the people. We are now approaching the second European elections with the Spitzenkandidaten process at the end of May 2019, and this time we will have more data to determine to what extent the new process is making a difference. A great way to research and follow the process can be looking at social media. This thesis aims to do just that, and is an analysis of how the 2019 Spitzenkandidaten were able to gain interaction from other users on Twitter from 15th December 2018 to 15th March 2019. To understand this better, this thesis first looks at how the candidates used Twitter and the public who tweeted about both the European elections and the candidates' campaign slogans, before finally comparing how the candidates gained interaction. This makes it easier to differentiate between the candidates, and it gives insight into who the candidates were likely to gain interaction from as well. With this, the main aim of this present thesis is to see if there is a golden path to be a successful Spitzenkandidat in terms of gaining interaction on Twitter.

Having an active social media presence is important for politicians to gain attention from voters. This makes it possible to reach out to potential voters on a regular basis, and studies have shown that having a presence on Twitter can encourage political engagement (Park, 2013; Kruikemeier, 2014). This active social media presence could be more important than ever in European elections, as there has been a diminishing voter turnout ever since they began (Shackleton, 2017, p. 154). How did the 2019 Spitzenkandidaten use Twitter in the months before the elections? How did their usage of Twitter differ? And, is it possible to see a universal Twitter strategy to successfully gain interaction for the candidates?

1.1. Research question

Analysing discourse on Twitter is a complex matter. This is because there are many different areas and variables to consider when doing such research. Having an active Twitter account is today almost mandatory for politicians to stay relevant and interact with their voters. However, what does it mean to have a successful Twitter account? Which content sells well to the public, and how can the candidates successfully promote themselves on Twitter? Is there a common successful path for all candidates, or does this depend on the followers that the candidates have? This present thesis aims to look at how the Spitzenkandidaten 2019 gain interaction from

Twitter users. This is done by first looking at how the candidates used Twitter, who the typical Twitter users they reach out to were, and by comparing their most and least popular tweets in terms of interaction gained. The thesis thus seeks to find out whether there are any clear patterns between the candidates in how they were able to gain interaction. With this, the research question presented in this thesis is:

"Under which conditions do tweets from the 2019 Spitzenkandidaten gain interaction from Twitter users?"

It is interesting to look at how the candidates were able to gain interaction because recent studies conducted by Spierings & Jacobs (2014) suggest that interaction with politicians contribute to fulfilling citizens' desires to receive attention and stay updated on the political news picture. Because of this, interactivity can be a key factor for gathering people to vote for elections. Furthermore, Hsu & Park (2012) write that interactions that politicians gain on Twitter can help determine the importance and popularity of the politician. Seeing how the 2019 Spitzenkandidaten gain interaction can thus help us see if some twitter strategies can help the candidates increase their popularity. To answer this question presented above, the present thesis will look at original tweets from eight Spitzenkandidaten from the period 15th December 2018 to 15th March 2019, and original tweets using hashtags related to both the Spitzenkandidaten process and the European elections from the period 20th April to 28th April. In total, this accounted for 1566 tweets from the candidates and 5059 tweets from the hashtags from the chosen hashtags. As the Spitzenkandidaten process on social media is a scarcely researched topic, answering the research question can provide a solid foundation for further research about the topic.

1.2. Justification of the study

This thesis is justified on three (3) grounds. First and foremost (1), this thesis is relevant because it monitors and explains political communication on Twitter in 2019. How are these candidates using Twitter, and how does their usage correlate with political communication on Twitter during previous national or European elections? Is the 2019 Spitzenkandidaten tweeting in a similar, or different fashion than other national or European politicians? These candidates all have different backgrounds and different visions of how the future of the European Union should be. Because of this, it is interesting to see which topics they were focusing on in their tweets. Not only is the Spitzenkandidaten process a new election process, but it is also different and more complex process than other national elections. With this, conducting an analysis of the Spitzenkandidaten's Twitter behavior will contribute to a better understanding of what

campaigning for the European Commission top job means, as well as to a better understanding of the rhetoric of their respective political parties.

Secondly (2), it seeks to expand the existing literature about the public who are discussing elections on Twitter. The use of social media is in constant changing nature, and the same could be said about the political climate in the EU, emphasizing the importance of a more coherent study on this field. This thesis seeks to find answers to questions like "How are the current political climate in the EU shaping the discussions on the platform?", and "are those who are discussing the elections negative or positive towards European integration?". What categorizes the public discussing the elections, and how representative are they to the general population of the EU? By finding answers to this, the present thesis aims to gain insight into traits from the public discussing the European elections. This will give us a better understanding of who the public that the Spitzenkandidaten are gaining interactions from are.

Thirdly (3), this thesis is interesting because it seeks to add to the already existing literature about gaining interaction on Twitter. However, this study is especially interesting because the Spitzenkandidaten process is a new process that has not yet seen much attention from scholars. There have been conducted many studies regarding the European Union (EU) and democracy from before, and there exists a wide variety of theories explaining the European Parliament (EP) elections and the so-called 'democratic deficit' in the EU because of it. As the problem of the democratic deficit has been a problem in the EU for decades, it is interesting to see if the new Spitzenkandidaten process will have any impact on the extent the public engage with the elections. To enhance more engagement with the European elections, it could be beneficial to look at how European politicians gain attention and interaction from voters on social media. By looking at how these candidates gain interaction from other Twitter users, it will be possible to gain more insight in which topics sells to their followers and why their followers follow them. Park (2013) suggested in his study about Twitter use motivations and political engagement that Twitter opinion leadership are playing a crucial role in encouraging individuals to participate in public and political processes (Park, 2013, p. 1646). It is thus possible to assume that successful opinion leadership on Twitter could mobilize more engagement towards European elections, and this makes it meaningful to look at how European politicians gain interaction on Twitter.

1.3. Thesis outline

The present thesis is divided into five chapters. Chapter two first presents a historical background to the 2019 Spitzenkandidaten elections and lays out the theoretical framework for the thesis by looking at previous literature discussing how social media works. Chapter two then looks at how communication is done on Twitter, sentiments of tweets, differences between users on Twitter, and political communication on Twitter both by politicians and the public. Based on this literature, the thesis presents seven hypotheses for the 2019 elections on Twitter. The third chapter presents the methodological choices for this thesis and gives insight into how the information from Twitter was gathered and analysed. Following this, the fourth chapter analyses and presents the results gathered. This is divided into four parts showing how the 2019 Spitzenkandidaten used Twitter, typical traits of the public who discussed the European elections, to what extent the candidates were able to gain interaction on Twitter, and a final section comparing the results with the hypotheses created in chapter 2. Further, the fifth chapter concludes the study based on the results and discusses the strengths and limitations of the thesis. It also assesses the thesis' validity, reliability, and replicability. This thesis concludes that while there are some strategies that can increase the amount of interactions such as using hashtags frequently, it seems that there is not yet one golden route or strategy to successfully increase the number of interactions gained for the candidates'. With this,

2. Theoretical framework

This chapter introduces and discusses the theoretical framework of the analysis and presents assumptions for the analysis accordingly. The first section of this chapter gives an overview of the background to the 2019 EP elections, the Spitzenkandidaten process, who the candidates were, and their election campaigns for the 2019 elections. The second section gives an overview of relevant literature about social media and tweets, followed by a third section about interaction on Twitter. Following this, chapter two assess previous literature about tweet sentiments, and how these sentiments are shaping the discourse on Twitter. Furthermore, this chapter introduces previous literature about how political communication is done on social media. This is done by first looking at how politicians communicate on social media, how the public interacts with these politicians, and by looking at political communication during previous European elections. All of this creates the theoretical framework for this thesis, and the chapter ends with creating six hypotheses that will be tested in the results chapter.

2.1. Background to the European Parliament 2019 elections

The EP began as the Common Assembly when the European Coal and Steel Community (ECSC) was established in 1952 to provide a link with national parliaments of the member states (Burns, 2013, p. 160). Dinan (2014) writes that it later evolved into a directly elected European institution and held its first elections in June 1979. In the first election, the candidates were selected from national parties, and the results reflected the performance of the governments in the member states (Dinan, 2014, p. 165). The elections have since become more and more supranational, with their own election campaigns and candidates. Ever since the first election took place, the EP has suffered from a diminishing voter turnout. Because of this, there has been an ongoing argument about the democratic legitimacy of the EU (Burns, 2013, p. 160-62).

Research conducted about the elections to the European parliament indicates that these elections are being perceived by the voters as second-order national elections (Burns, 2013, p. 168-169). The last election in 2014 had a turnout of 42.6%, the lowest ever turnout since the voting process began. Despite the low turnout, the EP's powers have gradually increased in the EU's legislative procedures since the first election (Shackleton, 2017, p. 154). Nevertheless, the level of support has from the public has depleted in recent times as shown by the decreasing voter turnout. Because of this, Shackleton (2017) argues that this has led to the argument of the EU suffering from a democratic deficit becoming even more acute in recent times (Shackleton, 2017 p.154).

One recent initiative to combat the low voter turnout from the EP elections was introducing "the Spitzenkandidaten process" for the 2014 elections. Spitzenkandidaten is a German term that traditionally has been referring to the lead candidate of a political party (Westlake, 2017, p. 2). This new initiative for the 2014 elections allowed the voters to place their vote on a candidate for the post as a Commission President. Braun & Popa (2018) writes that the EP tried several measures to gain more engagement from voters with this new process. This was amongst other things demonstrated by the EP having posters on their building stating that 'this time it will be different' (Braun & Popa, 2018). With the new process, the EP did not only seek to improve the voter turnout, they also wanted to strengthen the European Commission's (EC) legitimacy by enabling voters to have a say in who their president should be (Dinan, 2015).

Despite the promising signs when the process was introduced, many scholars argue that the new process' influence was limited (Hobolt, 2014; Van der Brug et al, 2016) and that it decreased political party competition in the EP (Christiansen, 2016). In the end, Hobolt (2014) argues that while the Spitzenkandidaten process did play a limited role in determining the composition of the EP, it altered the election process and built a foundation for the election of future Commission Presidents (Hobolt, 2014, p. 1443). Van der Brug et al (2016) also back this argument. They argue that neither the Spitzenkandidaten process nor the increased politicization of the elections did change the elections from being viewed as second-order elections. However, as several scholars have pointed out negative effects associated with the process, recent studies by Schmidt et al (2015) suggest that the process did influence the public and that people who knew the lead candidates were more likely to vote in the elections (Schmidt et al, 2015 p. 364). The upcoming challenge could then be to get more people to know the lead candidates of the process.

The future of the Spitzenkandidaten process is still a debated topic. There has been raised questions about whether the process de facto has strengthened the democratic nature of the elections, and the European Council has raised concerns about not being able to pick the next Commission President (Politico, 2018). With the Council's reluctant stance on the process, there has been ongoing speculation prior to the elections that Chief Brexit Negotiator Michel Barnier could become the next Commission President, despite not being part of the Spitzenkandidaten process (Politico, 2019a). Because the Spitzenkandidaten process was introduced in 2014, it is not pressed in any of the treaties. The European Council has thus previously stated that it cannot be legally required to follow it (Politico, 2018).

2.1.1. Candidates and election campaigns

The upcoming elections for the European parliament featuring the Spitzenkandidaten process will be held in May 2019. At the time of writing this thesis, the European People's Party's (EPP) candidate Manfred Weber looks to be the clear favorite to win the elections (Politico, 2019b). Frans Timmermans, the candidate from the Party of European Socialists (PES) is regarded as Weber's main competitor, however he is not projected as a likely winner (ibid 2019). The Alliance of Liberals and Democrats for Europe (ALDE) nominated a list of candidates. Their candidates were Guy Verhofstadt, Sylvie Goulard, Margrethe Vestager, Cecilia Malmström, Hans van Baalen, Emma Bonino, and Violeta Bulc. Other candidates include Jan Zahradil from the Alliance of Conservatives and Reformists in Europe (ACRE), Ska Keller and Bas Eickhout from the European Green Party (EGP), Violeta Tomic and Nico Cue from the European Left (EL), Oriol Junqueras from the European Free Alliance (EFA), and Yanis Varoufakis for Democracy for Europe (DiEM25). The remaining political parties in the EP have not decided to nominate any candidates for the elections (europeelects, 2019).

All candidates represented their own party's election campaigns. For the EPP, Manfred Weber's campaign slogan was "Stronger Together for a Better Europe". He also arranged a listening tour where he visited all EU member states and listened to potential voters' ideas for a better Europe (EPP, 2019). Frans Timmermans campaign slogan for PES was "It's Time for A New Social Contract for Europe". Among his most prominent issues was combating climate change, fight for tax justice, manage migration better, and to guarantee security to the people of Europe (Party of European Socialists, 2019). The EGP nominated two candidates who campaigned for environmental protection, fighting racism and injustices, and taking responsibility with refugees. Both candidates used the EGP's campaign slogan "Lets Act. Together". Like Weber, they also toured Europe to contribute to election discussions (European Greens, 2019).

The EL also nominated two candidates for the Commission Presidency, Violeta Tomic and Nico Cue. They campaigned against the increasing gap between the wealthy and the poor, and focused on combating environmental change (European Left, 2019). ALDE nominated seven candidates consisting of mostly women and called the candidates "Team Europe". Their campaign slogan was #RenewEurope, and they campaigned for a more liberal Europe. Some of their most prominent topics were to embrace digitalization, protect the rule of law, and to pursue economic innovation (Team Europe, 2019). Further, Jan Zahradil's campaign slogan was "Retune The EU". As a self-declared Eurosceptic, the meaning of his slogan was to retune

the EU to get it back into harmony with its people. He decided to run for the presidency for the campaign despite being against the Spitzenkandidaten process. This was because he wanted to raise the publicity of the party (The Conservative, 2019, p. 22).

Representing the newly established party DiEM25, Yanis Varoufakis did not have a concrete campaign slogan related to his Spitzenkandidaten candidacy. His party positioned itself as the first transnational party in Europe, cooperating both with national parties and separately from national parties in member states (DiEM25, 2019). For the elections, they published a manifesto called "European Spring" where they aimed to re-shape the European project and to give a voice to all people in the EU (European Spring, 2019). The EFA listed controversial Catalonian politician, Oriol Junqueras as their Spitzenkandidat. According to the party, he was nominated to expose the regression in Spain and to show the regression of fundamental rights in the EU (Spain in English, 2019). At the time of his nomination, Junqueras was imprisoned in Spain because of his involvement with the 2017 Catalan independence referendum (europeelects, 2019).

2.2. Social media, Twitter and tweets

Social media is a concept that nearly all internet users are familiar with today, but it is at the same time a concept that has caused confusion amongst a vast amount of scholars. According to Kwak et al (2010), this confusion has happened because it is possible to question what should be included in this term, as most social media networks function differently (Kwak et al, 2010). Therefore, it is reasonable to have a clear definition of the term when looking to analyse social media. Kaplan & Haenlein, (2010) defines social media as *"a group of Internet-based applications that build on the ideological and technical foundations of Web 2.0 that allow the creation and exchange of user-generated content"* (Kaplan & Haenlein, 2010). With this, social media refers to sites where users can publish their own content, share content, cooperate, and socialize.

Furthermore, social media can be divided into different categories. Stieglitz & Dang-Xuan (2012) distinguishes from three main types of social media in their framework for analysing political communication on social media: microblogging sites, Social Networking Sites (SNS) and weblogs. A microblogging social network is a social network where the messages that people post are either restricted or typically very short. Larsson & Moe (2011) writes that microblogging can be looked upon as a diminutive version of blogging. With this, they explain that a microblogging site is a site where users publish small comments or posts that are usually delivered to a network of associates (Larsson & Moe, 2011). In the case of microblogging sites,

Twitter is known as the most popular as of 2019 in terms of active users (Lifewire, 2019). An SNS on the other hand, is a site where users create a public profile and form relationships with other users using that site. Facebook and Linkedin are examples of such social networking sites (Kushin & Kitchener, 2009). Lastly, weblogs are full blogs where people write longer texts for more interested readers. Examples of weblogs can be political blogs, business blogs and personal blogs (Herring et al, 2004).

As previously mentioned, Twitter is the most prominent microblogging service in terms of active users. Twitter is a microblogging service where users interact with each other with messages known as tweets. Kwak et al (2010) write that a tweet is a message about any topic and that there are many ways for users to interact with each other on the platform (Kwak et al, 2010, p. 591). There was registered to be 326 million monthly active users on Twitter in the third quarter of 2018. This is an increase of 55 million users since the second quarter of 2014 when the last elections to the EP were held (Statista, 2019). Twitter was known for its simplicity and a limit of 140 characters; however, this limit was doubled to 280 characters in 2017. This is still a lower limit than other popular social media platforms such as Linkedin and Instagram who has a limit of 700 and 2200 characters respectively (Gligorić et al, 2018).

2.3. Interaction aspects on Twitter

It is interesting to look at aspects of interaction on Twitter. Doing so makes it easier to understand which messages are more likely to gain interaction, and how people perceive the different ways of communicating on the platform. As previously mentioned in the introduction, Spierings & Jacobs (2014) write that gaining interaction can make a substantial contribution to winning the hearts of voters, making it an important measurement for politicians. There are several ways for users to interact with each other on Twitter. As previously mentioned, Twitter is a microblogging site where people write and publish simple messages about any topic. Users can post their tweets by writing manually on the site, directly from their smartphones, or by using other third-party sites (Kwak et al, 2010, p591-592). The simplicity of posting a tweet makes it easy to spread real-time information to several users (Mendoza et al, 2010). Kwak et al (2010) write that users on Twitter can follow other users and that this is a relationship that does not need to be reciprocated. A user who follows another user will receive their tweets in their feed and will thus be able to easily stay updated on what the person is sharing (ibid, 2010, p. 591).

In a study about how language is used to build communities on Twitter, Zappavigna (2011) writes that the main features for users to interact with each other on Twitter is by flagging topics

using hashtags, addressing other users, favouriting tweets and reproducing other user's tweets. In the first case, flagging using hashtags can be done by putting the '#' symbol in front of a keyword (Zappavigna, 2011, p. 791). With this, the topic(s) of tweets are defined, and the tweet can be referenced to, or found by other users if they use the Twitter search function. Hanteer et al (2018) write that one of the ideas behind using hashtags is to be able to reach out to users that are not following you. This is the case because using hashtags in tweets makes it easier to find the tweets by doing simple searches (Hanteer et al, 2018). Addressing other users is done by a so-called @-mention, which is done by putting a '@' symbol in front of a user's username. This is commonly referred to as replies (Kwak et al, 2010). In this way, followers will be able to see the original tweet that was replied to as a context. This function can also be used to tag, mention, or credit another user (Twitter, 2019a).

A third way to interact on Twitter is to favourite other tweets by clicking the heart icon located on the bottom of a tweet. Recent studies outlined by Meier et al (2014) suggest that favouriting tweets is a discrete way of interacting on Twitter, and that the reasons for favouriting are very heterogeneous. Motives from favouriting a tweet can be anything from strongly agreeing with the topic, finding the tweet funny, to just being able re-find the tweet easily (Meier et al, 2014, p. 346, 350). The fourth and final way to interact on Twitter is to reproduce other users' tweets using the retweet function. With this, Twitter users forward another user's tweet on their own page and the user's followers will get the retweeted tweet in their feed (Zappavigna, 2011). Boyd et al (2010) explain that retweets should not only be looked upon as copying and pasting other tweets, it should also be seen as a discursive community of voices that give rise to a sense of a shared conversational context. Because of this, users of Twitter both retweet other users and aim to gain retweets themselves (Boyd et al, 2010 p. 1).

2.4. Tweet sentiments

Users of social media have certain ways of expressing their emotions on Twitter, and many tweets thus have positive and negative associations to them (Liu, 2012). This is an important part of the discourse on social media because people often prefer hearing other opinions about a topic before making their own opinion (Stieglitz & Dang-Xuan, 2013). These ways of expressing emotions on Twitter are commonly referred to as tweet sentiments. Understanding this can help us determine to what extent a person or groups of people feel about topics (Yu & Wang, 2015, p. 393). Humour and tone are among the most important aspects of tweet sentiments. Previous research about the tone of tweets have differed between positive, negative, and a mixed tone consisting both negative as well as positive content (Diakopolous & Shamma,

2010; Zavattaro et al, 2015). Understanding the tone of tweets can sometimes be a tricky task as communication takes place in an asynchronous way. Using emoticons or exclamation marks are thus popular ways of helping people understand the tones of tweets intuitively (Zavattaro et al, 2015). Algorithms are consistently becoming better at understanding the tone of tweets, however there are still several limitations in this area today (Stieglitz & Dang-Xuan, 2013).

Humour in tweets is another aspect of using Twitter that should be taken into consideration. For instance, using humour on Twitter can be be posting funny videos or pictures, writing a joke, or replying sarcastically to other users (Castro et al, 2016). Raz (2012) writes that the use of humour has shown to not only influence human beliefs, but also affect the feelings of the audience. It has even shown to encourage activity and engagement when used on social media such as Twitter (Raz 2012). Holton & Lewis (2011) back up these indications. In their study, they found that using humour in tweets might contribute to creating stronger connections between the users posting tweets and their followers (Holton & Lewis 2011). With this, humour is an important aspect of the discourse happening on Twitter.

2.5. Political communication on Twitter

The addition of Twitter as a communication platform has added new ways for the public to interact with politicians and vice versa. The increase in popularity of social media has almost made it mandatory for politicians to have a presence on them, and it has given politicians new, interactive ways of reaching out to the public (Kessel & Castelein, 2016). Recent studies conducted by Dubois & Gaffney (2014) suggest that politicians can be opinion leaders in their networks on Twitter (Dubois & Gaffney, 2014, p. 1274). Despite this, there are real differences between the popularity of politicians on social media, and thus varieties in how many people they are able to reach out to (Vergeer, 2015). Politicians are also different when it comes to the amount of activity on Twitter. In the words of Lee & Shin (2012), the amount of activity among politicians on Twitter can be distinguished between high and low interactivity. Low interactivity refers to politicians mostly posting messages on his or her own, while high interactivity is when the politician is actively responding to his or her followers. (Lee & Shin, 2012). With this, the usage of Twitter by candidates varies largely. Recent research by Graham et al (2014) show following that most politicians follower clear tweet patterns, while a minority deviate strongly from the norms (Graham et al, 2014).

With parties and candidates, recently published research back up the fact that parties and campaigns in opposition tend to use Twitter actively (Vergeer & Hermans, 2013; Larsson & Kalnes, 2014). In general, earlier research indicates that politicians are more likely to create a

dialogue with other politicians rather than their public following (Hsu & Park, 2012; Larsson, 2015), however some scholars have found exceptions of this in recent years. Larsson & Ihlen (2016) found in their study about Party leaders on Twitter during the 2013 Norwegian elections that the party leader's tweets were mostly @-mentions towards other users (Larsson & Ihlen, 2016, p. 677). Tromble (2018) also found that large numbers of politicians in the United Kingdom and the Netherland were reciprocally interacting with the public (Tromble, 2018). She also found that politicians could gain considerable rewards for interacting with the public when it comes to trust (Tromble, 2018). This indicates that there are exceptions in how politicians use Twitter depending on elections and the strategy of the politician.

When it comes to the wording and the sentiments of politicians on Twitter, Heiss et al (2019) write that polarizing messages are more likely to receive interaction and attention than more neutral tweets (Heiss et al, 2019). Karkin et al (2015) also researched the usage of polarizing tweets from politicians on Twitter. In their study of politicians' usage of Twitter during the Gezi Park protests in Turkey, they found that politicians use more polarizing language during social uprisings. Their findings also suggest that using more polarizing language can help reconstruct a discoursive power for the politicians (Karkin et al, 2015). These findings suggest that it can be beneficial for politicians to write polarizing messages not only if the goal is to gain interactions and attention, but also if the politicians want to influence their followers. Recent studies also show that how political candidates behave on Twitter can be influenced by the media. Common ways of doing so can be interacting with journalist's Twitter accounts or by publishing their opinions about the coverage (Ekman & Widholm, 2015). Previous research also indicates that political twitter activity corresponds with political coverage in the national media, albeit it is not following it deterministically (Stier et al, 2018). With this, the press can affect what people discuss on Twitter.

It is important to note that the use of social media for political communication also has been a process that has its drawbacks. One drawback can be an information overload for the users as the internet offers few limitations to the amount of information that the user may gather (Rodriguez et al, 2014). Another drawback can be that many users use Twitter as one of their main sources of staying updated. Every user decides who to follow and who not to follow, and many users gaining a selective news feed on Twitter because of this. Because of this, Halberstam & Knight (2016) argue that users mostly gain information that confirms their point of view, and that they are thus less open to politicians with other views (Halberstam & Knight, 2016). Alcott & Gentzkow (2017) writes that the information bias on Twitter has given many

actors possibilities to spread fake news, especially in cases when people are exposed to isolated and biased information from Twitter. Many people are thus unsure what information to believe in from Twitter (Alcott & Gentzkow, 2017).

2.5.1. The public's interaction with politicians on Twitter

There have also been conducted several studies about how the public is using Twitter to interact with politics. Previous research in this field indicates that the Twitter users who are likely to engage with politics represent a small, political interested subgroup, that are typically young, and more likely to be ideologically left wing (Klašnja et al, 2016, p. 9). With this, the users tweeting about elections are often referred to as "the vocal minority", while the people who do not tweet actively are commonly referred to as "the silent majority" (Mejova et al, 2013). There is also a large variation in how intensely the vocal minority contributes to the debates; some contribute heavily, while others contribute scarcely (Mustafaraj et al, 2011). Considering the explanations of the public tweeting about the elections elaborated above, the public's interaction with politicians on Twitter can be summarized by citing Jungherr et al (2015): *"While Twitter may not offer a true picture of reality, it shows the attention, motives, and interests of specially interested Twitter users"* (Jungherr et al 2015).

Studies about the tweet sentiments from the public have found slight correlations between the sentiments of the tweets posted by the public during the campaigning season and the election results (O'Connor et al, 2010). Recent studies about the vocal minority's tweet sentiments that the tweets that these people post tend to be more for than against the politician, less sarcastic and humourous, and that they are more likely to use hashtags and post media content than other typical Twitter users (Mejova et al, 2013). Recent studies by Tromble (2018) indicate that the public is more likely to interact with politicians who have already shown willingness to respond to other users (Tromble, 2018). Furthermore, a study conducted by Yardi & Boyd (2010) indicates that users are more likely to retweet tweets that support their own beliefs, thus providing evidence to the confirmation bias when having the option to choose who to follow or not to follow. However, other interactions such as replies do not seem to fall under confirmation bias, as users have a tendency of replying to both messages they agree and disagree with (Yardi & Boyd, 2010).

2.5.2. Political communication in European Parliament elections

Campaigning for European elections on Twitter is a topic that has not gained much attention from scholars compared to election campaigns in national elections. Older research suggests that EP politicians in opposition begin their campaigning earlier and more actively compared

to politicians representing parties in power (Vergeer et al, 2011). Previous research conducted on the 2014 European elections on Twitter suggest that EP election candidates are more active during election years and become more inactive after the elections have taken place (Larsson, 2015; Nulty et al, 2016). Furthermore, according to Larsson (2015), politicians tend to use Twitter as a platform for interacting with similar associates, rather than the public (Larsson, 2015, p. 163). Studies thus suggest that politicians are more likely to interact with other politicians when it comes to both national and European elections

Because of its nuance, the Spitzenkandidaten process in a social media context is also a topic that few scholars have researched. Research on the matter suggests that anti-EU candidates are less likely to have an active Twitter account than pro-EU candidates, however anti-EU candidates present on Twitter were more likely to tweet more frequently (Nulty et al, 2016, p. 442). Reviewing the results of their case study, Nulty et al (2016) concludes that the political communication for the 2014 EP elections on Twitter politicizes the debate of European issues and institutions, rather than national issues (ibid, 2016, p. 443). With this, the little data currently available about the Spitzenkandidaten process in a social media context suggest that anti-EU candidates are more likely to tweet frequently and that European issues see a more frequent discussion than national issues.

2.6. Hypotheses

Based on the previous research elaborated in chapter two, this chapter presents six hypotheses for the present thesis. These hypotheses were created for each of the three main sections of the results chapter to correlate well with the three different parts of the results chapter. This makes it so that the theoretical framework covers all three sections of the results chapter: How the 2019 Spitzenkandidaten used Twitter, who the people discussing the elections were, and how the candidates were able to gain interaction on Twitter. With this, two hypotheses were created about the candidates' Twitter usage, one about who the public tweeting about the elections represented and three hypotheses about how the candidates were able to gain interactions from voters. These will be tested in the results chapter.

Research about politicians' Twitter usage suggests that politicians in opposition use Twitter more actively when in opposition in both national elections (Vergeer & Hermans, 2013; Larsson & Kalnes, 2014) and in European elections (Veerger et al, 2011). How does this relate to the new Spitzenkandidaten process? Based on the studies mentioned above, this study assumes that the Spitzenkandidaten from parties in opposition will use Twitter differently than the candidates representing political parties in power. Because of this, the first hypothesis is:

H1: The candidates representing the political parties in power use Twitter differently than the candidates representing the political parties in opposition

Secondly, the theoretical chapter about tweet sentiments has shown that humour and tone are important aspects of the discourse on Twitter. In the case of tweet sentiments and the Spitzenkandidaten process, recent studies outlined by Nulty et al (2016) showed that there was no clear pattern between left-right positioning and emotional tone during the 2014 Spitzenkandidaten elections. There was however a strong correlation between being pro-EU and writing positive tweets and being anti-EU and writing negative tweets (Nulty et al, 2016, p 442). Because this was a strong correlation, this thesis assumes that the same will be the case in this study. With this, the second hypothesis of this thesis is:

H2: There is a correlation between being more Eurosceptic and posting more negative and subjective tweets for the candidates

Over to the public who are likely to discuss elections, the theoretical framework has shown that the Twitter users who are likely to engage with politics represent a small, young, politically interested subgroup. This vocal minority are usually more supportive of the politicians than negative towards them. How does this compare with the public discussing more general hashtags? This thesis looks at both hashtags related to the candidates and the European elections. Since the vocal minority discussing the elections tend to be supportive towards politicians, this thesis assumes that this will be the case when looking at the hashtags discussing the candidates or their election campaigns. Therefore, the third hypothesis for this thesis is:

H3: The public who tweet about hashtags related to the Spitzenkandidaten's campaigns post more positive tweets than the public who tweet about the European elections

Finally, we are able to draw three hypotheses about how the candidates are able to gain interactions from these users. Are there any Twitter strategies that are likely to work better than others? As newer research shows that using hashtags can be a successful way to reach out to a larger amount of people (Hanteer et al, 2018), it is interesting to see if this is the case with 2019 Spitzenkandidaten as well. In light of this, the fourth hypothesis of this thesis is:

H4: Using hashtags more frequently increases the number of interactions gained by the candidates

When it comes to the wording and sentiments of politicians, the theoretical framework has presented two key findings. These are that tweets containing polarizing messages are more

susceptible to gaining interaction (Heiss et al, 2019) and that posting polarizing tweets can lead to a higher discursive power (Karkin et al, 2019). Since the findings presented in the theoretical framework show that polarizing tweets are more likely to gain interaction, this thesis assumes that this is also the case for the 2019 Spitzenkandidaten.

H5: There is a correlation between posting polarizing and subjective tweets and gaining more interactions on Twitter.

Finally, studies about politicians replying to other users are quite split in their conclusions. Older research suggests that replying to other users are not necessarily beneficial for politicians, while some newer research has begun indicating that replying to other users can be beneficial for politicians. The split conclusions in studies about the matter make this a very interesting topic to research. This thesis bases its assumptions on the newer studies and assumes that replying to other users will have a positive effect on interactions gained. Therefore, the sixth and final hypothesis of this thesis is:

H6: Replying to other users increases the number of interactions gained by the candidates

3. Methodology

This chapter presents the methodological choices made for this thesis, and how the material was gathered and analysed. The first section of this chapter discusses issues related to gathering data from social media and adds theoretical background to different ways of gathering data from Twitter. Section 3.2 introduces the sources and codebook used for this thesis and explains how the data was collected. Section 3.3 outlines the theoretical background used for gathering and analysing the data. The concepts of social media analytics (SMA) and sentiment analysis are discussed in this section. Finally, section 3.4 explains the candidates chosen for the analysis, the methodological choices for the results chapter, and the thought process behind the three sections of the results chapter.

3.1. Data access and computer-mediated communication

It is common for social scientists to deal with topics that are hard to observe and analyse and dealing with digital trace data is no exception. The two most common issues when conducting such analysis is problems regarding representability and sampling, and problems when analysing user behaviour on different platforms (Jürgens & Jungherr, 2016). Firstly, this relates to privacy concerns and limitations of accessing and analysing user data from Twitter (Morstatter et al, 2013). Twitter is thus not willing to give unlimited access to their data. Instead, researchers are usually provided with restricted access to a so-called Application Programming Interface (API), which is a set of functions and procedures that allow the creation of applications, which access the features of data of an operating system (Puschmann & Burgess, 2013).

Analysing different Twitter-users' behaviour is also problematic because the design of the platforms are limiting the ways users can express themselves. Researchers will have to translate the actions of users based on pre-made channels, interaction patterns and modes defined by their platform. In light of this, it is always important to take the rules and algorithms of the social media platform into account before making any conclusions (Gillespie, 2014). For instance, when it comes to Twitter, it is important that the researcher is well aware of the character limits, and how people use the platform.

Twitter provides two different API's that developers can use to gather data: Streaming APIs and Rest APIs (Twitter, 2019c). Streaming APIs delivers data to the researcher in real time and focuses on three parameters: words, geographical boundaries and user IDs (Morstatter et al, 2013). Rest API, on the other hand, returns the latest 3200 tweets from a user on Twitter, or the

3200 most recent tweets containing a specific phrase or word (Jürgens & Jungherr, 2016, p. 22). Twitter also offers different kinds of access to the APIs depending on how much the user is willing to pay. They differ between normal APIs (free) Premium APIs (monthly fee) and Enterprise APIs (expensive monthly fee). I therefore had to make a binding choice for how much information I would be able to gain for the thesis: Should I choose the free version of APIs and then possibly limiting the scope of the study, or would it be enough to use the normal version of APIs? As most of the data that would be of interest for this thesis would be recent, I did not see myself benefiting much from purchasing access to Premium APIs. Hardware specifications and disc space could also have been problematic if I had chosen a more expensive version, as processing the data could be very demanding for my computer. There were two relevant limitations with choosing Normal APIs: I would not be able to gain access to the number of replies that users gained and there would be a time limit of seven days when gathering tweets from a hashtag. While these factors could limit the findings to some extent, I figured that I would still get more than enough interesting information with just retweet count and favourite count. Reviewing these limitations and possibilities, I decided to use normal APIs as the primary source of data.

3.2. Sources and codebook

For this thesis, I decided to use Python as the program to both gather and to do most of the data analysis from Twitter. This inspired by the guidelines presented in the chosen model for conducting an SMA. First, the data was gathered from Twitter using a premade script for gathering Twitter data made public by Jurgens & Jungherr (2016), with some slight modifications made by Tahee Kim for a seminar held at ECPR Winter School 2019. With this I had a deductive approach with the coding for gathering the data for the thesis. This means that the codebook was determined before the actual coding began. The benefits of using a deductive approach to coding is that it is easy to replicate for other researchers (White & Marsh, 2006).

Python was chosen as the program for gathering the data for this present thesis because web scraping is well documented on the program (Russel, 2014; Mitchell, 2018), and because it is designed in a way that makes the data accessible and easy to read (Lubanovic, 2014 p. 10). The fact that it is easy to read also makes it easy to write, learn and remember which was important to me as I had little experience with programming from before. One of the drawbacks from using the program can be its simplicity. Python's simple nature makes it sometimes not fast enough for some more demanding scripts (Lubanovic, 2014, p 11). This was deemed not to be

a likely problem for the goal of this thesis, as the codes needed for scraping information and tweets from Twitter was accessible and well documented.

The analyses were done both in a quantitative and a qualitative matter. By combining both methods, I was able to process and analyse a large amount of data as well as gain a better understanding of the context of the tweets. Gaining a better understanding of the context that many tweets were written in, made it easier to understand the quantitative data. First, 25 randomly picked tweets were read and analysed for the first section of the result chapter. This was to get an overview of how the candidates tweeted, and the context of the tweets. This laid the foundation for the tweet categories used to categorize the candidates' most and least popular tweets in terms of interaction gained. Furthermore, qualitative assessments of the candidates 10 most and 10 least popular tweets were done, and categorized.

The quantitative analyses were done by running my own script that I wrote for analysing large amounts of data from Twitter. For the sake of this study's transparency, the codes for conducting the quantitative analyses of the data have been uploaded into a GitHub repository. More in-depth information about this can be found in the Appendix A. The uploaded files consist of several explanations of what has been done, with folders including all tweets analysed from candidates, most of the tweets gathered from the hashtags, codes for gathering data from Twitter, and the codes for analysing the data. The explanations are presented in lines using the #-symbol, as this indicates that the text is not used as a code in the script. All scripts should be easy to run for anyone with basic Python knowledge. As web scraping from Twitter is a popular area for conducting research, most of the information needed for expanding the script to do further research should be easily accessible on popular discussion sites such as Stackoverflow.com. It is important to note that the coding has been done in Python 3.6 and that this is not the newest version of the program. This was the case because some of the Python packages needed for analysing the data were not applicable to the newest version of the program.

3.3. Theoretical approaches to data analysis

Over the last years, the demand for collecting, analysing, visualizing and monitoring information from user-generated content on the internet has increased. This demand is not only for research purposes, it can also be used for advertising companies or social customer relationship management (Stieglitz et al, 2014). To have a better understanding of the data analysis, it is helpful to have a theoretical background to it. This makes it easier to critically

evaluate the purpose of the study and gives the researcher a more in-depth understanding of what is being done. This section of the methodology chapter explains the two theoretical approaches used for the analysis: Social Media Analytics (SMA) and sentiment analysis.

3.3.1. Social Media Analytics

This study has been inspired by the popular SMA research area. Bekmamedova & Shanks (2014) writes that SMA is a method to analyse and interpret social media data (Bekmamedova & Shanks, 2014). This is still a fairly new and emerging research field that can be applicable in many different fields. In computer science, SMA could be developing algorithms and tools for predicting changes in and analysing social networks, while in political science it aims to examine the impacts social media has for political participation (Stieglitz et al, 2014). There are several different models for conducting SMA on Twitter (Stieglitz et al, 2018; Stieglitz & Dang-Xuan, 2013; Chae et al, 2012), however for this present thesis, I decided to use a version of the model presented by Fan & Gordon (2016) because of its simplicity. This choice was made due to a lack of experience both with creating scripts for gathering data and with quantitative methods. It was thus important that I had a complete understanding of what I was doing to be as transparent with my results as possible. Fan & Gordon (2016)'s SMA model consists of three main phases illustrated in the figure below:

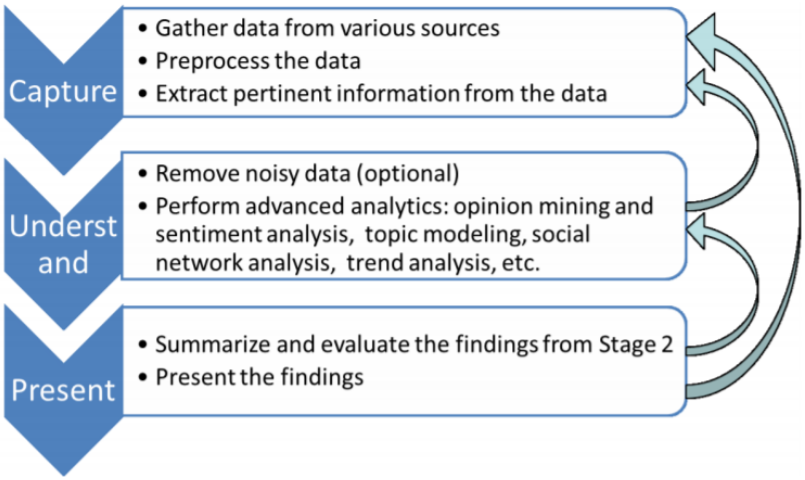

Figure 1: Model for Social Media Analytics (Fan & Gordon, 2016, p. 6)

The first phase of the social media analytics process is the capture phase. The capture phase usually consists of gathering a massive amount of data from various sources such as APIs or by looking at news feeds. The second phase is the understand stage, and this is seen as the core of the model. It is at this stage that the captured data is assessed and analysed. The first part of this

stage is to remove eventual noisy data to have a better background to perform a more meaningful analysis. From this point, the researchers have several options for analysing the data (Fan & Gordon, 2016, p. 7). The last stage of this SMA-model is the present phase. In this phase, the results are evaluated and presented, often with the help of various visualization techniques. This is commonly done by creating plots, diagrams, and wordclouds (ibid, 2016, p. 8).

Basing data analysis on SMA is a process that has both strengths and weaknesses. There are many challenges related to using the method because of its nuance (Stieglitz et al, 2018). One strength of the process is that it is a multidisciplinary research field that combines knowledge from multiple research areas, making it easy to use for both practical and research perspectives (Stieglitz et al, 2014). Being multidisciplinary can also be a limitation for the field because it can make the field complex with many disciplines focusing on different areas. Because of this, many of the people using the framework have been required to co-operate interdisciplinarily, and this has made SMA lack a theoretical core (ibid, 2014). With this, there are both strengths and weaknesses connected to the method that is important to keep in mind when conducting research about social media.

This present thesis is loosely based on the SMA model presented above. In the first phase, the data was captured using Python. I then had several files both containing tweets from both candidates and tweets that had used certain hashtags. Secondly, the files were filtrated meaning that noisy data such as retweets posted by the candidates were removed. Following this, several simpler analyses were conducted to get an overview of the characteristics of the tweets before more advanced sentiment analyses were done. This all correlates well with the understand phase. Finally, the findings were evaluated with the hypotheses presented in chapter 2 in mind. The findings were then ready for presentation.

3.3.2. Sentiment analysis

To be able to gain a better insight into publics' and the candidates' emotions, I decided to conduct a sentiment analysis for this thesis. This is regarded as a more advanced method of SMA (Fan & Gordon, 2014, p. 8). Liu (2012) writes that sentiment analysis is a way to rely on human coders to conduct content. Doing so on Twitter makes it possible to gain a quantitative insight into people's emotions, or opinions towards a subject (Liu, 2012). This was deemed to be an interesting addition to the thesis, as doing so will contribute to a better understanding of

how the public rate the European elections, as well as a better understanding of the candidates' subjectivity and neutrality.

The Python library Textblob was used to do sentiment analyses in this study. Textblob is a package on Python that makes it easy and accessible to process language (Planspace, 2015). When conducting a sentiment analysis, the program returns two forms of sentiment: Polarity and subjectivity. Polarity floats within the range of [-1.0, 1.0] and indicates to what extent statements is positive or negative. [-1.0] is a very negative message, while [1.0] indicates a highly positive message. Subjectivity, on the other hand, is ranged between [0.0, 1.0] where [0.0] is very objective and [1.0] is regarded as highly subjective (Textblob, 2018a). The word "great" can be used as an example of how the program works. This is a word that is both positive and subjective, meaning that it gives high positive numbers in both polarity and subjectivity. This could be further enhanced if a modifier word was found before the word. "very great" would thus have a higher polarity and subjectivity than great, while "not great" would give a lower number to these categories (Planspace, 2015). If Textblob sees a name that it does not know, it will simply ignore the word in its analysis (ibid, 2015). It does also not include words containing only one letter (Textblob, 2018a).

As Textblob has decided which words are positive and negative before starting its analysis, it limits possibilities to edit the weighting of words if it should be needed. This could arguably be regarded as a limitation of the library; however, it seems unlikely that the person doing the research with disagreeing with the classifications of the program. Another more influential limitation with the library is that Textblob does not account for other languages than English. Therefore, tweets written in other languages were not included in the analysis (Textblob, 2018b). While not perfect, Textblob was regarded as a simple package that should be able to give overall good indications of the sentiments from the tweets. Another option for conducting a sentiment analysis with python could have been using the Natural Language Toolkit (NLTK) package. This would allow the researcher to define how positive or negative different kind of words should be. This is done manually and would be more time consuming (nltk, n.d.). As categorizing word in NLTK is done manually, the researcher could risk not adding words that could have been beneficial to add to the sentiment analysis. I thus decided that Textblob would be the best-suited package for me when conducting the sentiment analyses after learning how the package works.

3.4. Selection

Because I had to hand in this thesis before the actual elections are held in May 2019, the thesis has had some limitations regarding the choice of the period for gathering data. As this thesis seeks to explain how the candidates gain interactions on Twitter, analysing all of their tweets since they registered to the platform would not be significant for the purpose of this study. I thus had to pick a relevant selection period for gathering the tweets. My thesis was due within the 15th of May 2019, and I thus had to make sure I was able to analyse the results in time. The chosen period therefore became from 15th December to 15th March, giving the results a three-month period. While many of these candidates had not yet been presented as their party's candidate(s), it was still assumed that many of their tweets would be related to the upcoming elections. A limitation with the time of writing this thesis is that it is not possible to yet determine whether the number of interactions gained on Twitter has any correlation with the election results. This does nevertheless seem like an unlikely correlation given the relative low follower count of most candidates. The main challenge when picking out material to study in a thesis is to have a selection that does not leave out text that provides vital context for the study. This is a difficult problem when analysing tweets as there are so many tweets to consider when doing so.

For the results to correlate with previous research about political communication and interaction on Twitter, this thesis' analysis is divided into three main parts: First, this thesis looks at what the chosen candidates are writing about, which hashtags they are using and other typical traits and patterns with their usage of Twitter. With this, this study gives an indication of whether or not the candidates' behaviour on Twitter correlates with previous research about political communication. This was a natural first step towards seeing how they gain interactions from other users. Secondly, this thesis looks at the public discussing the process. This could help us get an understanding of who are discussing relevant topics related to the elections, and more importantly, the typical user that the candidates are gaining interaction from. Lastly, this study looks at how the candidates gain interactions on Twitter. Which topics are selling the most, and to what extent is it possible to tell that there is a pattern in the tweets that gain interaction. It also looks at topics of tweets that did not receive any interaction at all.

3.4.1. Candidates analysed

The first step before starting the data gathering was to pick candidates for the analysis. A selection of eight Spitzenkandidaten was chosen for this study. Information about the candidates are found in the following table:

Candidate	Twitter username	Political party	Tweets during selected period	Followers
Manfred Weber	@ManfredWeber	EPP	236[1]	31 694[2]
Frans Timmermans	@TimmermansEU	PES	150	101 976
Guy Verhofstadt	@guyverhofstadt	ALDE	252	358 060
Margrethe Vestager	@vestager	ALDE	86	246 818
Bas Eickhout	@BasEickhout	Greens-EFA	184	18 142
Ska Keller	@SkaKeller	Greens-EFA	95	36 071
Jan Zahradil	@ZahradilJan	ACRE	377	6419
Yanis Varoufakis	@yanisvaroufakis	Diem25	182[3]	956 412

Table 1: The characteristics of the chosen candidates' Twitter profiles

The eight Spitzenkandidaten shown in the table above were chosen for the analysis. In total, they published 1562 tweets excluding retweets. Retweets were excluded because they were not written by the candidates themselves, and they would thus not contribute to a better understanding of how the candidates gained interaction from users. While there were more candidates who had announced their campaign to become the next Spitzenkandidat than the table above shows, the candidates were narrowed down to eight for the sake of this analysis. As ALDE nominated seven candidates, they would represent more than half of the candidates if they all were chosen for the analysis. Verhofstadt and Verstager were thus picked for the sake of this study because they were the most popular male and female candidates in terms of followers on Twitter. Both candidates from the EL were excluded from this analysis as Violeta Tomic had only published one tweet during the chosen period and Nico Cue did not have a public Twitter account. Oriol Junqueras from EFA were also excluded from this analysis due to all his tweets being in either Spanish or French, and this could lead to the researcher not understanding the context of the tweets. While many of these politicians were not officially presented as the parties' candidates until late in the selection or after the selection period was over, they were still representing their respective parties on Twitter. It was thus decided that they would nonetheless contribute to the purpose of the study.

3.4.2. Tweet categorization

It is helpful to have different categories when looking at large amounts of data. It makes it much easier to filter the tweets which make it easier to get a grasp of the data amount. As this study had gathered over a thousand tweets, categorizing them was regarded as a necessary step to

[1] The tweets from the candidates do not include retweeets
[2] All follower counts as of 20.03.2019
[3] If counting retweets, Varoufakis was tweeting the most active candidate with 632 tweets.

understand them. These categories were based on a qualitative look at 25 of the candidates' top tweets using Twitter's advanced search on their website. This gave me an overview of typical topics that the candidates would tweet about as well as a better understanding of the contexts of the tweets. With this, the following thirteen categories were created for the thesis:

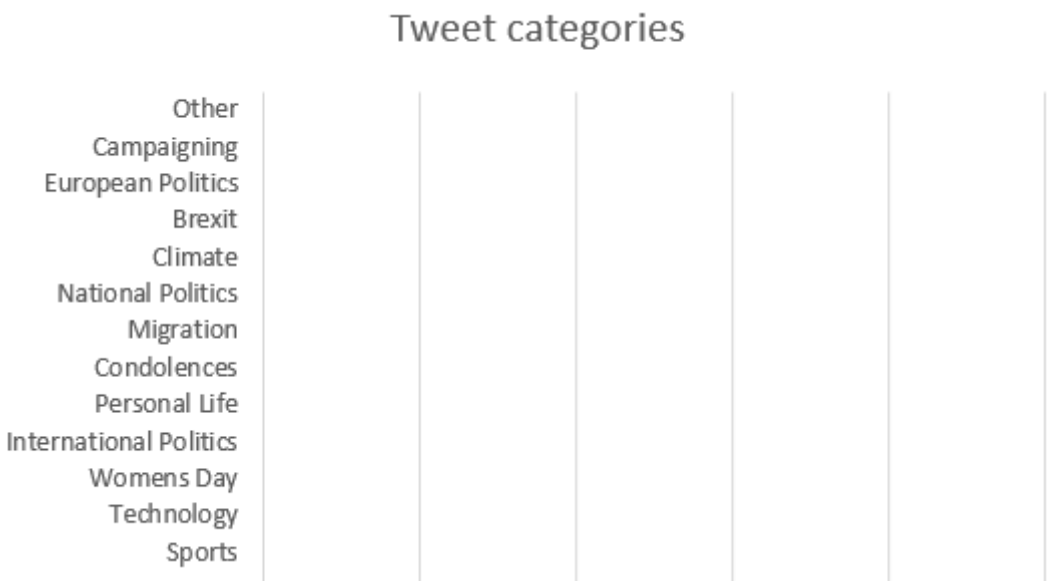

Figure 2: Tweet categories looked at in the thesis

The keywords shown in figure 3 were chosen for categorizing the candidates' tweets. These thirteen categories were chosen for categorizing the candidates most popular and least popular tweets in chapter 4.3. Campaigning is the largest category, and consists of anything related to the election campaigns such as campaign slogans, promoting meetings, or speeches. European politics consists of opinions and statements about the political developments in other countries than the candidates' national country. Since previous research mention that European issues are discussed more than national issues in these elections (Nulty et al, 2016), I decided to split them into their own categories. As Brexit one of the main topics on the agenda during the chosen time-period, I decided that it would be interesting to make it its own category to see how often it was mentioned and how it affected interactions from other users. The few tweets that did not fit the categories were placed in the "other" category.

3.4.3. The chosen hashtags for analysing the discussion of the European elections

To gain insight into how the public was discussing the elections on Twitter, I decided to gather relevant hashtags about the elections and analyse them quantitatively. A much shorter time-period had to be chosen when gathering the data from the hashtags. The main reason for this was the limits with the API version that I chose. Normal API's users can only gather tweets

from the last seven days (Twitter, 2019b), thus limiting the scope of tweets gathered. Nevertheless, this limitation was not regarded as a significant one, as seven days should still be enough time to be able to draw conclusions of the patterns from the tweets gathered. There was also a possibility there would be too much information to consider if looking at all hashtags used in the same three-month period as the candidates. It is therefore likely that a significantly shorter period would have been chosen for the hashtags even if I had access to another API version. First, the waters were tested in March to interpret the popularity of the hashtags. Seeing that the campaign slogan was scarcely tweeted about in March, I decided to wait until late April before gathering the tweets. This was because I assumed that the campaign slogans would be more frequently used closer to the elections in May. With this, the period that the tweets were gathered from became from the 20th of April to the 28th of April[4].

When choosing the hashtags for this part of the analysis, I wanted the chosen hashtags to represent both the European elections in general and the candidates. Doing so would make it possible to see if the same characteristics of the people who are discussing the European elections have the same characteristics as the people who are discussing the campaign slogans. With this, it is possible to see if the typical people who tweet about the Spitzenkandidaten process represent the same people that tweet about the European elections. Doing so is also necessary to test H3 which assumes that the public who tweet about post more positive tweets than the public discussing the European elections. Adding the general hashtags about the elections would also contribute to more tweets for this study as they were tweeted much more frequently about than the tweets related to the candidates, even when gathering them in a period closer to the elections taking place. The selection thus became two hashtags about the 2019 European elections and most of the campaign slogans related to each candidate. Therefore, the following hashtags were chosen for the study:

[4] Due to some technical difficulties, the tags #LetsActTogether, #RetuneTheEu, #BetterEurope and #Timmermans were gathered a day later than the others were.

Hashtag	Number of tweets gathered
#EP2019	2825
#EuropeanElections2019	1467
#BetterEurope	11
#Timmermans	605
#RetuneTheEU	12
#LetsActTogether	17
#RenewEurope	47
#EuropeanSpring	75
Total	5059[5]

Table 2: Relevant hashtags for the 2019 Spitzenkandidaten elections

The table above shows the hashtags that tweets were gathered from. Here, the two hashtags #EP2019 and #EuropeanElections2019 were chosen to get an overview of what the public wrote generally about the elections. These hashtags became included because they were frequently used when testing the waters in March. These hashtags were also deemed unlikely to represent other topics than the European elections.

To add more depth to the chapter about the public's political communication towards the elections, I wanted to see how frequently the campaign slogans were tweeted about. All the candidates' slogans were represented here except for the slogans of Varoufakis and Timmermans. Timmermans' slogan #ItsTime was not suitable for the analysis because it was a hashtag that was commonly used to discuss other topics. It was for instance used by the Scottish National Party (SNP) during the same time-period (SNP, 2019), meaning that the results would be representative of the Spitzenkandidaten elections. With this, #Timmermans was instead chosen as it was deemed to be a much more representative hashtag. As Diem25 did not have a concrete slogan for the elections, this section instead focuses on #EuropeanSpring as this was the name of their manifesto. Another methodological choice of note was that #BetterEurope was chosen instead of #StrongerTogether, despite #StrongerTogether being the main saying of Weber's slogan. This was because #StrongerTogether could relate to other topics than the Spitzenkandidaten elections. With all these exceptions in mind, I was ready to analyse the gathered tweets.

[5] None of these numbers count retweets or replies

3.4.4. Interaction variables

The last and most complicated part of this study was looking at how the candidates were able to gain interaction on Twitter. Counting this was more complicated because I had to both take favourite count and retweet counts into consideration. Reply count could not be added because of my API package not giving access to this feature (Twitter, 2019b). First, a qualitative approach was done to see which tweets gained the least and the most interactions. With this, I was able to get a thorough overview of typical tweets written by the candidates.

While the qualitative approach did give clear indications of which topics were the least and the most popular for each candidate, it did not give a good enough ground for comparison since it did not account for differences between favourite count and retweets. With these two variables of interaction, I had to decide if they should have the same weighting or if they should weigh differently for the results. In addition to this, all candidates had a different number of followers, which had to be kept in mind. This was very important, as there were cases of several hundred thousand followers differentiating the candidates, making it a somewhat unfair ground for comparison. With all these factors into consideration, the following formula was chosen for finding out which candidates gained the most interaction based on their potential:

$$r = \frac{70\% * retweets + 30\% * favourite\ count}{amount\ of\ followers}$$

Figure 3: Formula for calculating interaction rate per follower

Because favouriting tweets is a much more discrete way of interacting on Twitter than retweeting (Meyer et al, 2014), I decided to take this into account when gathering the results. Retweets thus count for 70% of the interaction gained, while favourite count counted for 30% of the interactions. This was then divided with the number of followers that the users had at the time of conducting this thesis to adjust the results. This was done to have a fairer comparison between the candidates. One retweet gained by Jan Zahradil with 6419 followers should count more than one retweet gained by Yanis Varoufakis with nearly one million. With the formula presented above in mind, I was then able to calculate a weighted interaction rate per follower for each candidate. With this, I now felt I had a more fair method of comparing the candidates, and were thus ready to run the analyses.

4. Results

This chapter presents the results from the script and represents the Present-phase of the SMA-model. In total 1562 tweets were analysed from eight Spitzenkandidaten Twitter accounts and 5059 tweets using the hashtags mentioned in chapter 3.4.3. The first part of this chapter looks at which topics the Spitzenkandidaten were tweeting about. This was done by looking at the different languages the candidates were tweeting in, qualitatively assessing tweets of the candidates, and by assessing their most frequently used hashtags. Chapter 4.1 also looks into the tweet sentiments of the candidates, and to what extent they replied to other users on Twitter. Secondly, this chapter looks at typical traits about the public who tweeted about the elections. To do this, this thesis looks at the language of the tweets and the locations of the users tweeting about the topics. Chapter 4.2 ends with a sentiment analysis of the gathered tweets using the hashtags to get a better understanding of the tone of the tweets posted by the public. Chapter 4.3 looks at how the candidates gained interaction on Twitter. This was done by first looking at which topics gained the most and least interactions from each candidate. The chapter then compares the candidates' interaction rates using the results from the formula presented in chapter 3.4.4. Finally, chapter 4.4 interprets the findings and discusses to what extent they correlate with the hypotheses presented in chapter 2.6.

4.1. The candidates' Twitter usage

This part of the results chapter looks at how the different candidates used Twitter. First, this chapter presents the different languages that the candidates used in their tweets and looks at typical content in the candidates' tweets. This chapter then looks at how the candidates behaved on Twitter, which hashtags they used, the mean subjectivity and polarity of the candidates and to what extent they replied to other users. With this, this chapter aims to provide an overview of the candidates Twitter usage that will be helpful when looking at how they gained interaction, as well as provide answers to H1 and H2.

As the different candidates represent a variety of nationalities, it is interesting to see what kind of languages they used on Twitter. Did they mainly aim to reach out to followers from their national countries, European followers, or a combination of both? Running a script to categorize which languages the candidates tweeted in found that the candidates tweeted in a large variety of languages, however many of these languages were not used frequently. I thus decided that there would not be necessary to mention all the languages used. With this, languages that would

frequent in less than five of the total original tweets from each candidate were excluded from this table. The table below shows the different languages that the candidates tweeted in:

Candidate	**Total original tweets**	**Top languages in percentage**
Manfred Weber	236	English (55%), German (34.3%), French (3.3%), Polish (2.9%)
Frans Timmermans	150	English (80%), French (8.6%), Spanish (4.6%)
Guy Verhofstadt	252	English (88.4%), French (3.5%), Spanish (3.1%)
Margrethe Vestager	86	English (79%), Danish (17.4%)
Bas Eickhout	184	Dutch (63%), English (35%)
Ska Keller	95	English (66.3%), German (32.6%)
Jan Zahradil	377	Czech (65.5%), English (33.6%),
Yanis Varoufakis	182	English (66.4%), Greek (24.7%), German (3.8%), French (3.3%)

Table 3: The candidates' tweet languages

After looking at the different languages, a qualitative assessment of 25 of the top tweets found in the Twitter advanced search option. Here, popular tweets during the selected time period were assessed. This showed 25 popular tweets written, however Twitter's algorithms did not range these tweets based on retweet count or favourite count. Twitter's algorithms also took the time the tweets, which resulted in a mixture of new tweets and tweets that were more popular. Doing this created a detailed overview of the content and the context of typical tweets posted by the Spitzenkandidaten. By doing this, it was clear that the candidates all had different approaches to their Twitter usage not only when it came to the different languages they used, but also in the contents of their tweets. The first assessed candidate was Manfred Weber. Representing the EPP, Weber frequently tweeted about campaigning for the elections, Brexit, and the political developments in Greece. Some of his most prominent topics to campaign about on Twitter was to reduce the gender gap, stand up against antisemitism and a more pro-

European Poland. Most of his tweets were written in English, however he also posted a significant amount of tweets in German. He was also the only candidate who tweeted in Polish. Out of the assessed tweet sample, the tweets written in Germans were about national politics. The following figure shows an example of a typical Manfred Weber tweet during the three-month period:

Figure 4: Manfred Weber tweet example

The tweet example above shows Weber both discussing Brexit and criticizing political decisions in Greece. Frans Timmermans was the next candidate in line after looking at Weber's tweets. Representing PES, Timmermans was on the other hand to a larger extent focusing on climate change, and campaigning to even out the economic differences in Europe. While Weber frequently tweeted about reducing the gender pay gap in Europe, Timmermans focused on this to a larger extent, and especially the international women's day received attention. A large majority of his tweets were posted in English, however he also tweeted a notable amount of tweets in Spanish and French. A typical tweet of Frans Timmermans can be seen in the following figure:

Figure 5: Frans Timmermans tweet example

The screenshot above is one of the tweets Timmermans posted about the international women's day. He shared a video showing highlights of the atmosphere of the celebration in Madrid and used the occasion to ask his followers to interact with his post. He almost reached 150 retweets in this tweet, however this seems like a low number considering that the video had over 16400 views. After looking at Timmermans, the next candidate to have a selection of his tweets assessed was Guy Verhofstadt. A qualitative assessment of Verhofstadt's tweets made it clear that Brexit was one of his main topics to tweet about. He also frequently tweeted about European identity, and commonly used the phrase "I am European". It also became clear that he is a very popular figure on Twitter, and that he gains much interaction in about every tweet. He was also the candidate with the highest percentage of tweets written in English. The only other languages to note was small percentages of tweets written in French and Spanish. As Verhofsatdt is a candidate from Belgium, his language usage suggests that he mostly wanted to reach out to European or international followers. A typical tweet from Verhofstadt can be seen in the following screenshot:

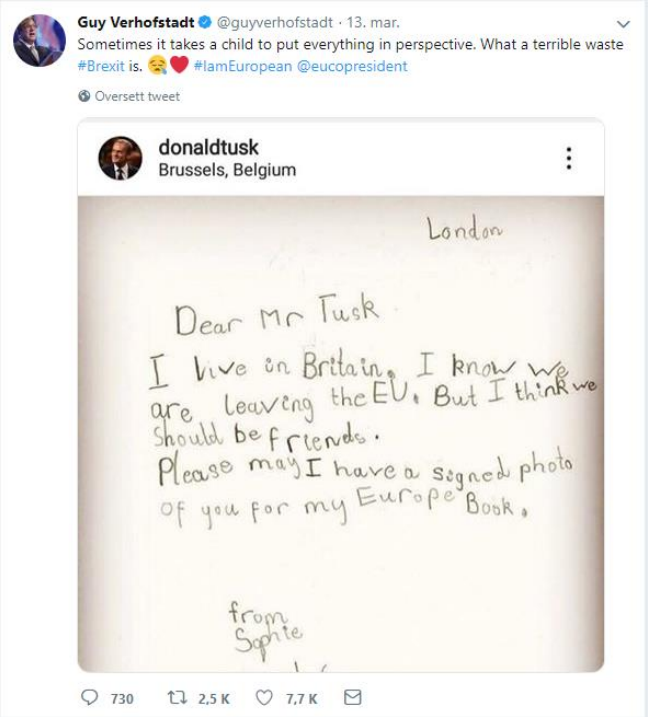

Figure 6: Guy Verhofstadt tweet example

In the example above, Verhofstadt's retweeted a popular tweet by Donald Tusk and used the opportunity to give his view of Brexit. Interestingly, this was not the only example where he would interact with other European politicians out of the 25 assessed tweets, suggesting that this was a somewhat frequent habit. Also representing ALDE, Vehofstadt's colleague Margrethe Vestager had a different approach with her Twitter usage. Instead of consistently tweeting about Brexit, or promoting European identity on Twitter, she instead focused on EU competition law, gender equality and climate change. She would also post photos from real life to greet her followers or update them where she was in the world. Like most candidates, the majority of Vestager's tweets were posted in English, however a notable amount of her tweets were written in her national language, Danish. A typical tweet posted by Vestager can be seen in the following figure:

Figure 7: Margrethe Vestager tweet example

In the tweet above, Vestager shared an article about the EC fining Mastercard. As she held a position as the European Commissioner for Competition and were also the only candidate to frequently post photos of natures or cities seen when traveling, the results from the tweet selection suggests that Vestager was largely tweeting as a private person rather than a Spitzenkandidat during the selected period. The first candidate assessed from the opposition was Bas Eickhout. Representing the EGP, Eickhout frequently tweeted about climate change. While other topics such as Brexit or European politics would be tweeted about every now and then, a vast majority of his tweets were about climate change. Interestingly, a majority of Eickhout's tweets were written in Dutch, suggesting that he either wanted to reach out to national voters, or was aware that many of his followers knew Dutch. An example of his tweets can be seen in the screenshot below:

Figure 8: Bas Eickhout tweet example

Figure 8 shows Eickhout tweeting about a majority of the EP calling for a rise of the EU's climate targets. Recall that Eickhout was the candidate with the second lowest follower count of all candidates. His relative high amount of retweets in this tweet suggests that many of Eickhouts followers agreed with this statement. Eickhout's tweets were more one-sided than the tweets of his companion, Ska Keller. While she also would frequently tweet about climate change, she tweeted more frequently about the gender gap or migration compared to Eickhout. This suggests that the candidates represented different areas of the EGP's politics. Furthermore, Keller would tweet in either English or German, and her English tweets accounted for about two-thirds of her total tweets during the three-month period. A typical tweet posted by Keller during the selected period can be seen in the figure below:

Figure 9: Ska Keller tweet example

The screenshot above shows a tweet where Keller praised Portugal for stepping up and taking a share of refugees. The next candidate from the opposition that was assessed was Jan Zahradil. Representing ACRE, Self-declared Eurosceptic candidate, Zahradil differed significantly from the other candidates. The most active Spitzenkandidat in terms of original tweets tweeted in both English and Czech, with Czech being his most frequently used language. Not only that, he was also found to be tweeting more frequently about national politics than his other candidates. Other than that, he would typically tweet about Brexit or criticize Emmanuel Macron in France. A typical tweet posted by Jan Zahardil can be found in the following figure:[6]

[6] Even though a more typical tweet for Jan Zahradil would be written in Czech, an English tweet was chosen as it would be a more readable example.

Figure 10: Jan Zahradil tweet example

In the example above, Zahradil tweeted about the UK's territorial integrity and posted an image of him holding a speech for the Ulster Unionist Party. Finally, the last Spitzenkandidaten this thesis assessed was Yanis Varoufakis. Representing the newly established party, Diem25, Varoufakis mostly focused on spreading the word about his new party. Other than that he would commonly tweet about EU-Greece relations and Brexit. He varied between tweeting in English, German and Greek. German and English tweets would typically be about the same topics; however, some of his Greek tweets were discussing Greece and its relationship with the EU. The figure below shows a typical example of Varoufakis' tweets.

Figure 11: Yanis Varoufakis tweet example

Varoufakis campaigns for Diem25 in the tweet example in figure 11. In that tweet, he shared an article about his party's manifesto for the elections: European Spring. In his attached article, he discussed the party's vision to rebuild Europe.

The results from a qualitative assessment of the candidates' tweets indicate that all candidates had different strategies for their twitter profiles. While some topics such as Brexit, climate change and reducing the gender gap were frequent topics for most candidates, it seems that all candidates had their own personal varieties. Most candidates posted a majority tweets in English, however two candidates representing the opposition (Eickhout and Zahradil) posted a majority of their tweets in their native language. Another finding of note was that all candidates tweeted in at least two languages, suggesting that this was a deliberate strategy. The following sections will give a more in-depth look of how the candidates differed in their Twitter usage

4.1.1. Hashtag frequency

Using hashtags in tweets can be a succesful strategy to reach out to more people. Doing so makes it easier for other users interested in a certain topic to find tweets discussing it (Hanteer et al, 2018). This chapter seeks to find out which hashtags were popularly used by the candidates to discuss the extent the 2019 Spitzenkandidaten were using this strategy to reach out to more people. First, this chapter gives an overview of the most common hashtags used by all candidates. Following this, this chapter looks at the top 5 most frequently used hashtags for each candidate to see how the candidates differed in their use of hashtags. First, the 10 most frequently used hashtags by each candidate were summarized together to see which hashtags were most commonly used by the candidates. The results were the following after running the script to see which hashtags were the most frequent:

Figure 12: The most frequently used hashtags used by all candidates

The figure above shows all the top 10 most frequently used hashtags by all eight candidates combined. Larger words indicate more frequently used hashtag. The most frequently used hashtags were about the EU and European elections, many of the candidates' campaign slogans, or about international crises/events such as Brexit. This correlates well with the general trends found when qualitatively assessing the tweets in the previous chapter. The figure above gives us a quantitative overview of the most commonly used hashtags by all the candidates, however it does not account for whether these hashtags were frequently used by all candidates, or how they were distributed. To gain a better understanding of who stood for which hashtags and how frequently they used hashtags, the same script was ran for each individual candidate. The results were the following:

Candidate	Top 5 most frequently used hashtags
Manfred Weber	#Strongertogether (49) Bettereurope (47) #DeinEuropa (29) Brexit (13) #Listeningtour19 (12),
Frans Timmermans	#Itstime (19) IWD2019 (6) #8m (6) #internationalwomensday (5) #HuelgaFeminista (3)
Guy Verhofstadt	#Iameuropean (28), #Brexit (20), #EP2019, (20), #Venezuela (14), #Generationeurope (7),
Margrethe Vestager	#EUcompdigit (4), #ligap3 (3), #next364 (2), #krakow (2), #eudialogues (2),
Ska Keller	#Europe (11), #Fridaysforfuture (6), #seawatch3 (6), EU (6), #Malta (5),
Bas Eickhout	#klimaatakkoord (1), #LetsActTogether (1), #Manfredonlyretweets (1), #Klimaatmars (1), #nordstream2 (1),
Jan Zahradil	#EU (60) #EP (27), #RetuneTheEU (17), Brexit (14), #Spitzenkandidat (8),
Yanis Varoufakis	#Rosaluxemburg (2), #Gießeneranzeiger (1), #Annewill (1), #Deineuropa (1), #Diem25 (1)

Table 4: Top 5 most frequent hashtags for each candidate

The table above illustrates the five most used hashtags and their frequency from all eight candidates. The results found very different patterns in which topics and hashtags the different candidates used during the three months period. Manfred Weber used the tags "Stronger Together", "Better Europe" and "DeinEuropa" most frequently. This correlates well with his usage of Twitter to campaign for his candidacy, as the slogan for his campaign was "Better Together for a Stronger Europe". Weber used #StrongerTogether slightly more than "BetterEurope" which indicate that both phrases were sometimes used separately as hashtags in his tweets. Frans Timmermans also used his campaign slogan "ItsTime" the most frequently but were also found discussing topical events such as the international women's day 2019. Interestingly, Timmermans tweeted far more frequently about the international women's day compared to the other candidates, indicating that this was a very important topic for PES' election campaign.

The candidates from ALDE used Twitter differently during the three-month period and were thus different in how frequently they posted tweets with hashtags. Guy Verhofstadt used hashtags very frequently. His most frequent hashtag was #IamEuropean, a phrase often used by him in debates or promotion campaigns. Vestager on the other hand, never featured the same hashtag more than four times. She was, for the most part, emphasizing hashtags related to

digitalization and combating climate change. Generally, Verhofstadt tweeted much more and used much more hashtags than his colleague Vestager. Their campaign slogan #RenewEurope was barely used by either candidate, but this is likely to be because ALDE announced the candidates for the elections in late March, and no tweets were gathered after 15th March.

The EGP's campaign slogan "Let's Act. Together" was only used once as a hashtag in their candidates' tweets. Instead, Ska Keller and Bas Eickhout used hashtags about topical events such as "FridaysForFuture" and "SeaWatch3". They were quite different in their use of hashtags. Keller commonly used hashtags about topics such as #FridaysForFuture and the migration crisis, while Eickhout never used the same hashtag twice in his original tweets. In the few cases where Eickhout used hashtags, they were usually about combating climate change or tweets where he criticized Manfred Weber on his stance on Victor Orban in Hungary. An example of a tweet where he criticized Weber can be seen in the screenshot below:

Figure 13: Bas Eickhout tweet example 2

Jan Zahradil used hashtags the most frequently out of all candidates. His most used hashtags were #EU, #EP and his campaign slogan #RetuneTheEU. As he had the lowest follower count of the candidates, his frequent hashtag use could be a deliberate strategy for reaching out to more people. Last, but not least, Yanis Varoufakis was the candidate who used the least number of hashtags in his original tweets. As he used so few hashtags, it is hard to find any pattern in his hashtag usage. Since more than 75% of his tweets were retweets, perhaps focusing on promoting his new party by retweeting tweets about it was a higher priority for him than promoting himself or discussing topics.

4.1.2. Sentiment analysis of the candidates

To get an understanding of the general wording of the candidates' tweets, a sentiment analysis was conducted for all candidates. This was done by using Textblob and the classifications used

by the program (see chapter 3.3.2.). The results after running the script are presented in the table below:

Candidate	**Mean Polarity**	**Mean Subjectivity**
Manfred Weber	0.091[7]	0.252
Frans Timmermans	0.092	0.250
Guy Verhofstadt	0.111	0.395
Margrethe Vestager	0.263	0.426
Ska Keller	0.071	0.263
Bas Eickhout	0.038	0.201
Jan Zahradil	0.032	0.176
Yanis Varoufakis	0.029	0.191
All tweets[8]	0.075	0.254

Table 5: Mean subjectivity and polarity of the candidates

These results show that the candidates were generally quite neutral in their language. Jan Zahradil and Yanis Varoufakis were the most neutral candidates, suggesting that they both used the least amount of modifiers and/or adjectives in their tweets. Margrethe Vestager was the candidate with the highest mean polarity, suggesting that she used more positive descriptions in her tweets than the other candidates. However, she was also the candidate who tweeted the least during the selected three-month period. It is thus not certain if the tone of her tweets would be regarded as this positive if the sentiment analysis was done on a larger sample of her tweets. Even though Verhofstadt did not differ much in terms of polarity compared to the other candidates, both candidates from ALDE interestingly had the highest mean polarity scores.

Zahradil and Varoufakis were also regarded as the most objective candidates, while Vestager and Verhofstadt were regarded as the most subjective candidates. With this, the two candidates who had the highest mean polarity also had the highest mean subjectivity, while the two candidates with the lowest mean polarity were also the most objective. This is an interesting finding, as H2 suggested that this should be the opposite. However, these results could partly be because of some limitations with Textblob. Textblob does not take other languages than English into account (Textblob, 2018b). As shown in chapter 4.1.1, Varoufakis wrote a notable of his tweets in German and Greek, while Zahradil wrote many tweets in Czech. These tweets were thus automatically left out of the program's analysis. While many of their tweets were

[7] Rounded down to three decimal places
[8] Sentiment analysis of all 1562 original tweets written by the candidates

written in English, their mean polarity and mean subjectivity should be taken with a grain of salt.

4.1.3. The candidates' usage of replies

Another important characteristic of the candidates was to what extent they were replying to other users. As indicated by Tromble (2018), replying to other users may have a positive impact on the extent the candidates are able to gain interactions (Tromble, 2018). This is important background knowledge to have when discussing how the candidates gained interaction as H6 assumes that the candidates were able to better their interaction rates by interacting with the public. The results were the following after running a script to filter the results:

Candidate	**Total original tweets**	**Number of replies**	**Percentage of tweets being replies**
Manfred Weber	236	29	12.2%
Frans Timmermans	150	80	53.3%
Guy Verhofstadt	252	9	3.5%
Margrethe Vestager	86	6	6.9%
Ska Keller	95	8	8.4%
Bas Eickhout	184	62	33.6%
Jan Zahradil	377	233	61.8%
Yanis Varoufakis	182	28	15.3%

Table 6: Number of replies posted by each candidate

The figure above shows each candidate's number of replies, and how much percentage of their original tweets the replies account for. These findings show that Jan Zahradil, Frans Timmermans, and Bas Eickhout were the candidates who replied the most frequently to other Twitter users during the selected time period. Zahradil and Eickhout were the candidates with the lowest number of followers, suggesting that there is some correlation between having a low follower count and replying to other users. This is further backed up by looking at the three candidates with the most followers (Vestager, Verhofstadt, and Varoufakis). While Varoufakis was somewhat of an exception with a reply percentage of 15.3%, Vestager and Verhofstadt had the lowest reply percentage of the candidates. On the other hand, Timmermans replied much more frequently than both Weber and Keller for instance, despite having more than twice as many followers than them. With this, there is not necessarily a correlation between the follower count that the candidates have and the extent that they reply to other users. Timmermans' example could thus suggest that the extent of replying to other users depends on the candidates' Twitter strategy rather than their follower count or position.

4.2.The public's interaction with the Spitzenkandidaten process on Twitter

The second part of the results chapter looks at what categorizes the public that discussed the Spitzenkandidaten process. This gives us insight into who the public who are likely to interact with the 2019 Spitzenkandidaten are. Whom do they represent, and how do they perceive the elections? This chapter analyses the gathered tweets from the eight hashtags presented in chapter 3.4.3 to gain knowledge about this. This was done by first looking at the different kinds of languages the tweets discussing the hashtags are written in, the locations of the public writing about these hashtags, and by doing a sentiment analysis of the public's tweets to see if they are positive or negative towards the elections.

4.2.1. Tweet languages of the public

The first way the tweets gathered from the hashtags were distinguished was by the language they were written in. The results were the following after running the script to distinguish the different languages:

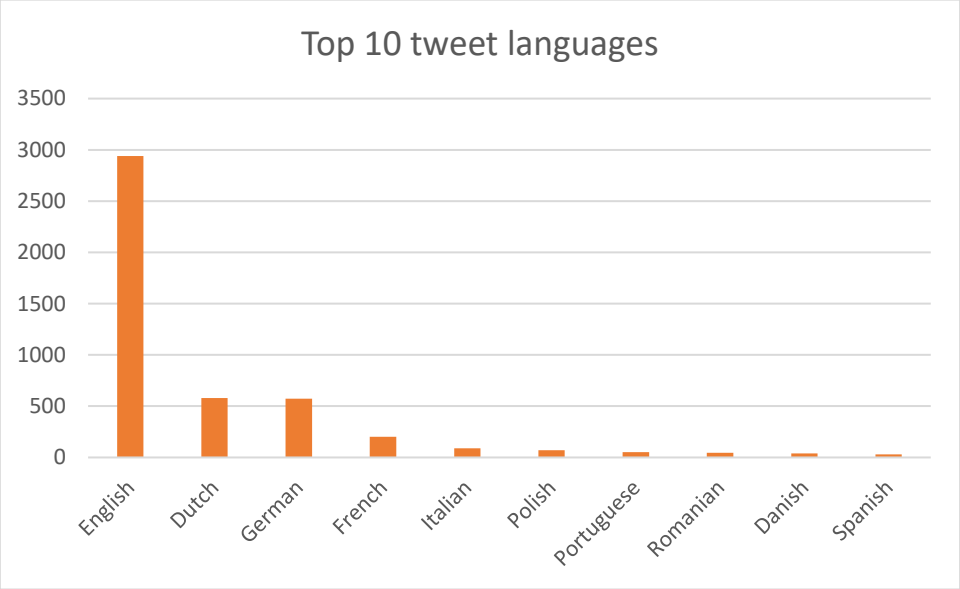

Figure 14: Tweet language from the public's tweets

The figure above shows the most frequently used languages from all tweets using the hashtags. Most tweets were written in English which is not unsurprising as Twitter is an international social media platform. The other most frequently used languages were mostly languages spoken in countries with high population counts except for Dutch. The more surprising findings were that Romanian was one of the top ten most commonly used languages and that few tweets were written in Spanish. This could either suggest that few people are tweeting about the process in Spain or that many Spanish users are tweeting in another language. With this, the results show

that the most frequently used languages were from Western European states, but that this does not necessarily represent the population speaking each language. To gain a better understanding of how the different languages were used for each hashtag, the hashtags were split up to see the top five languages used for each hashtag:

Hashtag	**Top five languages**
#EP2019	English [1406], German [461], Dutch [425], French [78], Italian [61]
#EuropeanElections2019	English [1205], German [56], French [47], Italian [16], Dutch [8]
#BetterEurope	English [8], Unspecified [2], German [1]
#Timmermans	English [250], Dutch [144], Polish [55], French [36], Romanian [29]
#RetuneTheEU	English [10], Latvian [1], Czech [1]
#LetsActTogether	Italian [7], English [6], French [2], German [1], Greek [1]
#RenewEurope	English [28], French [6], German [4], Dutch [4], Spanish [1]
#EuropeanSpring	German [16], French [16], English [12], Polish [2], Swedish [2]

Table 7: Twitter language per hashtag

The figure above shows the top five languages that the tweets gathered from each hashtag were written in. English and German were the two most frequently used languages for both #EP2019 and #EuropeanElections2019. As #EP2019 was a much more frequently used hashtag than #EuropeanElections2019, the tweets from the latter had a much higher percentage of tweets written in English. This suggests that #EP2019 were used more frequently in the different member states than #EuropeanElections2019, or that the users using the #EP2019 hashtag were more likely to have a national following rather than international. With this, there seem to be differences in which hashtag reaches out to whom, despite the hashtag representing more or less the same.

The hashtags directly related to the candidates were, quite similarly to the hashtags about the European elections mostly written in English. #LetsActTogether was a small exception, where the most commonly used language was Italian. #RetuneTheEU was the only hashtag who had Latvian or Czech as one of its most frequent languages. This is an interesting correlation with Zahradil being a Eurosceptic candidate, as these countries are usually regarded as more

Eurosceptic member states (Hobolt & De Vries, 2016). #EuropeanSpring was different from the other hashtags. In the case of this hashtag, the tweets were more frequently written in German or French than English. A reason for this could be that the public was reaching out to Diem25's affiliated parties on the national level and that it thus was more natural to tweet in these languages.

4.2.2. The geographical locations of the users tweeting about the elections

Now that we know which languages that are the tweets gathered from the hashtags were most frequently written in, another interesting way to distinguish the tweets is looking at the geographical locations of the users posting them. This helps us gain a better insight into where the people who were interested in the European elections and the 2019 Spitzenkandidaten process were from. To do this, lists were created of all the cities that the users using the selected hashtags had set on their Twitter profiles. The cities were then sorted by country. The results were the following after sorting the cities by country:

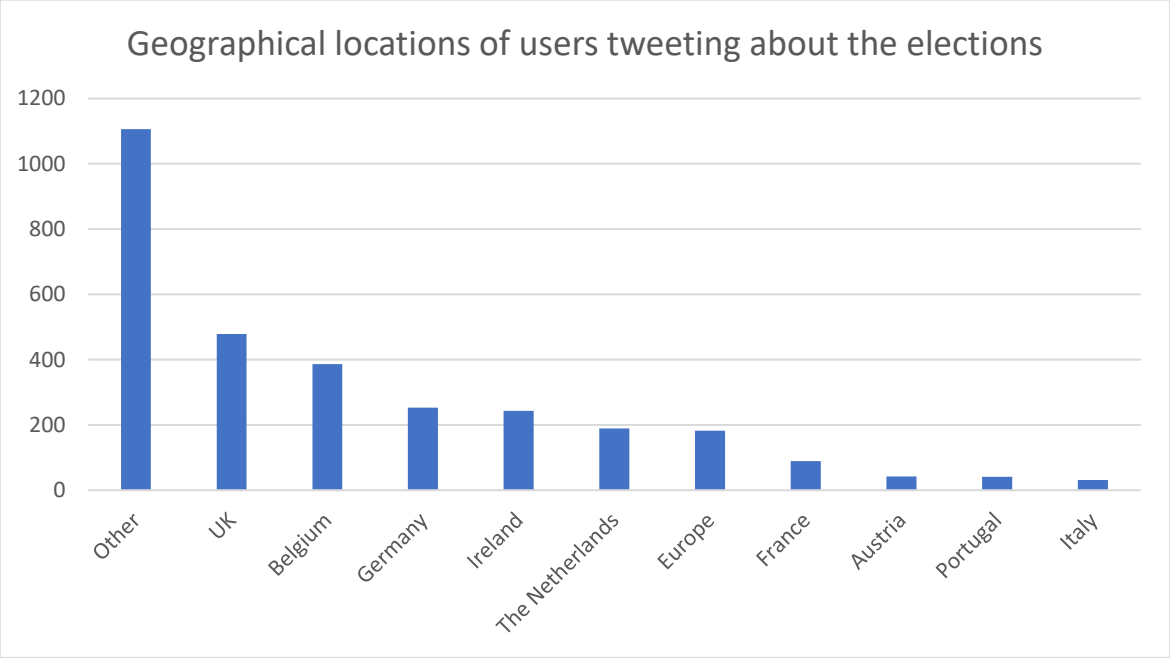

Figure 15: The geographical locations of the users tweeting about the process

The figure above shows the top ten user locations of the public who tweeted about the eight selected hashtags[9]. The term "other" is for users who had not added a specific location in their profile or for special cases such as "Nowhere near Brussels" and "planet earth". These results show that were fewer people located in France or Germany than there were tweets written in

[9] Some users only had their city meaning that there were hundreds of values. Because of this, values that were <5 were excluded from this graph

these languages. This suggests that many users tweeting in French or German either did not provide their locations on their Twitter profile, or that they were users located in "Europe" or "The European Union". The UK was not expected to take part in the elections until Brexit became extended in March 2019 (BBC, 2019). Despite this, the UK stood for the majority of the users who tweeted about the elections' location. This indicates that there is still much interest in the elections in the UK. Another interesting finding was that several users had their location set to Europe. This could be an indication that many of those tweeting about the elections feel connected to their European identity. Furthermore, we see that most of the countries participating in discussing the elections were from the more developed Western European countries. In order to gain a better understanding of which hashtag stood for which location, a classification of the top five locations for each tweet was done. The results were the following:

Hashtag	**Top five locations**
#EP2019	Belgium [242], Ireland [216], Germany [170], UK [167], The Netherlands [128]
#EuropeanElections2019	UK [421], Belgium [81], Germany [38], Europe [31], Ireland [26]
#BetterEurope	Other [3], Greece [2], Germany [1] Slovenia [1]
#Timmermans	UK [65], The Netherlands [45], Belgium [35], Europe [31], Romania [22]
#RenewEurope	Europe [21], Belgium [10], Germany [4], France [3] UK [2]
#LetsActTogether	Belgium [7], Italy [4], France [2], UK [1], Unspecified [1]
#RetuneTheEU	Unspecified [11], Czechia [1]
#EuropeanSpring	Other [24], Canada [16], France [16], Germany [5], UK [3]

Table 8: Geographical locations per hashtag

The figure above shows the locations of the users tweeting about each hashtag. The figure shows that the locations of the users using the two selected hashtags about the European elections was quite different. Tweets containing the #EP2019 hashtag were quite evenly spread out between Belgium, Ireland, and Germany. Out of these, Belgium was the most frequent location. This was unsurprising because the EP is located in Brussels, making it seemingly likely that many of those discussing the elections are residing in Brussels as well. There was no clear correlation between country population and the users discussing the elections on

Twitter. Instead, the users discussing the elections were more likely to be from smaller member states such as Ireland and the Netherlands. This indicates that the people discussing the elections are not representative of the population of the EU. Twitter users who used the #EuropeanElections hashtag were mostly from the UK, correlating well with the fact that a high percentage of these tweets were written in English. This hashtag was not evenly spread out between member states at all, consisting of mostly users from the UK. This

#BetterEurope differed from the other hashtags related to the individual candidates. The most common location of the users tweeting with the hashtag (except for the other category) was Greece. This is an interesting correlation with Weber frequently tweeting about Greece, perhaps indicating that his campaign has reached well out to Greek Twitter users. #Timmermans most common users were located in the UK, which was quite surprising considering he is a Dutch candidate residing in Brussels. This could be an indication that there was much interest in Timmermans' campaign in the UK. Furthermore, #Timmermans the only hashtag who had a notable amount of users located in Romania. This suggests that there was more interest in his campaign from Romanian Twitter users compared to the other campaigns. The most frequent locations of the users using #RenewEurope was Europe. This correlates well with Verhofstadt's Twitter usage where he frequently promoted European identity. This could be an indication that a majority of the public using this hashtag relate to a European identity. With this, the typical user who tweeted about #RenewEurope wrote their tweets in English and had their location set as Europe.

The users tweeting about #LetsActTogether were mostly located in Belgium. This is interesting because most of the tweets were written in Italian. This suggests that some of the tweets were written by Italian expats. Most of the Twitter users using this hashtag had not specified their location on Twitter. This was unfortunate because it made it impossible to get an overview of where the users tweeting about this hashtag were typically from. One tweet was written by a user located in the Czech Republic, indicating a somewhat correlation with Zahradil's tendencies of tweeting about national politics. However, one tweet is not a large enough example to create a conclusion about this. Surprisingly, several users using the #EuropeanSpring hashtag had set their Twitter location to Canada. The locations of the users using the hashtag represent mostly Western European countries, however the fact that users from Canada were this frequent could also indicate that some users are using this tag to talk about a European spring rather than the Diem25 party manifesto. The results from this hashtag should thus be taken with a pinch of salt.

4.2.3. Sentiment analysis of the hashtags

We now know some typical traits of the users discussing the elections. The next step is to now get an idea of how these users were perceiving the elections. To gain a better understanding of this, chapter 4.2.3 looks at the mean polarity and subjectivity found from each of the chosen hashtags. This will contribute to gaining an overview about how the people tweeting about the elections are perceiving them about a month prior to the elections. Using the same Textblob script as in chapter 4.1.2, the results from the sentiment analysis were the following:

Hashtag	**Mean Polarity**	**Mean Subjectivity**
#EP2019	0.071	0.196
#EuropeanElections2019	0.111	0.304
#BetterEurope	0.214	0.477
#Timmermans	0.054	0.158
#RenewEurope	0.115	0.211
#LetsActTogether	-0.008	0.090
#RetuneTheEU	0.087	0.225
#EuropeanSpring	0.117	0.165
All hashtags	0.082	0.223

Table 9: Mean Polarity and Subjectivity per hashtag

Based on the sentiment analysis of the hashtags chosen for this thesis, it is clear that there isn't much difference in the mean polarity between the hashtags. Neither of the hashtags about the European elections has a high mean polarity. This either suggests that the tweets about the hashtags are quite neutral, or that the tweets are polarizing with several tweets on both ends of the spectrum. The hashtag with the highest mean polarity was #BetterEurope, indicating that tweets using this hashtag had the most positive descriptions. As this hashtag is a part of Weber's campaign slogan, this could be an indication that the people supporting his campaign are feeling more strongly positive towards his campaign compared to the users using the other hashtags. However, the sample of tweets gathered using the hashtag was low, so one should be wary of drawing any early conclusions of this. Interestingly, #LetsActTogether was the only hashtag that was regarded as slightly negative by TextBlob. This does not necessarily indicate that the users using the hashtag were negative towards the elections or European integration though. A seemingly more likely assumption is that they were slightly negative towards the EU's climate or migration policy.

#EP2019 and #EuropeanElections2019 differed significantly in their mean subjectivity scores. #EuropeanElections2019 had the second highest mean subjectivity out of all hashtags, suggesting that the users tweeting about this hashtag had a somewhat subjective language. This is interesting as the majority of the people using this hashtag were located in the UK, indicating that the discourse about the elections in the UK is of a more subjective manner. The users using #BetterEurope were regarded as the users who posted the most subjective tweets, scoring much higher than any other hashtag. While there was not a large sample of tweets using this hashtag, this is nevertheless an interesting finding as it differs significantly from the other hashtags related to the campaigns. The tweets using this hashtag consisted thus of quite positive wordings and were very subjective. #LetsActTogether was the only hashtag who had a negative mean polarity, however it was also the hashtag with the least amount of subjective language. This indicates that the tweets using this hashtag was both very neutral and objective, as opposed to having extreme numbers on both ends. Generally, the users who tweeted about the two hashtags about the European elections during the 7 day period were slightly positive in wording and a bit subjective. The same can be said about all campaigning hashtags except #LetsActTogether, which on average were slightly negative.

4.3.How did the candidates gain interaction?

The third part of the results chapter looks at how the candidates were able to gain interaction from their Twitter usage. Interaction was in this case based on favourite count and the number of retweets each tweet gained because this was what my Twitter API version gave me access to. Based on the classifications presented in chapter 3.4.2., this chapter first looks at which topics from the candidates gained the most interaction for each candidate. Following this, the chapter looks at the topics that gained the least amount of interaction for each candidate. Finally, this chapter compares the candidates based on the weighted interaction rate presented in chapter 3.4.4.

4.3.1. Tweet topics that gained the most interaction

By following the classifications presented in chapter 3.4.2, I was able to get an indication of which topics were the most popular in terms of interactions from other users. After running a script to find the top 10 most popular tweets for each candidate, the tweets were then qualitatively assessed and classified based on the classification presented in chapter 3.4.2. The results were the following:

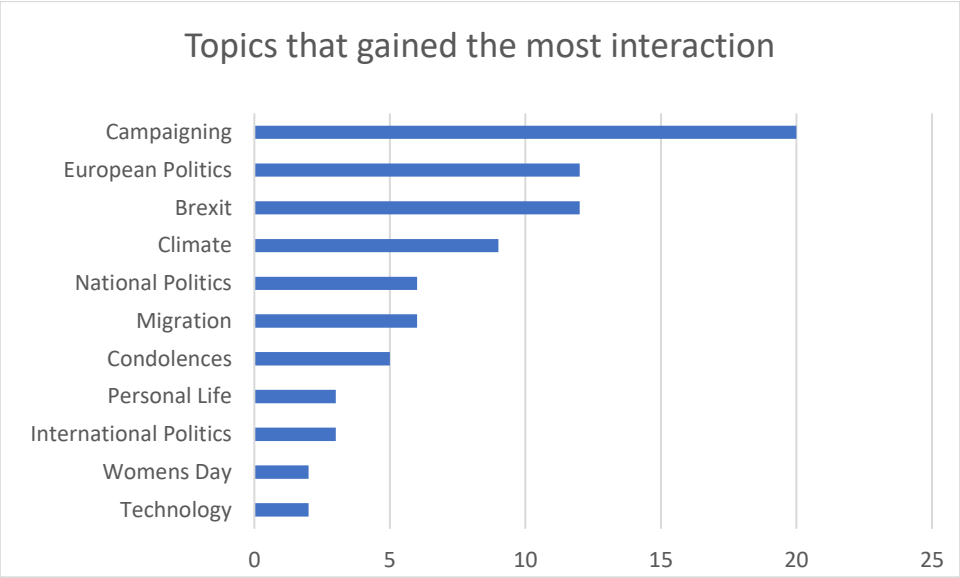

Figure 16: The most popular topics in terms of interactions gained

The table above (n=80) shows the 10 most popular topics in terms of interactions gained for all candidates combined. The findings of this indicate that topics such as European Politics and Brexit were political topics that were regarded popular for all candidates. Campaigning was the most popular category, but this is not as surprising as this was also the most common tweet category for the candidates. Topics such as International politics, women's day and technology were not among the most common topics that gained interaction, but this could be an indication that they were scarcely tweeted about, rather than an indication that they did not sell well. The category "sports" were not found in any of the tweets, suggesting that this is a topic that is not popular amongst the candidates' followers. To compare which topics were the most popular for each candidate, each candidate's top 10 most popular tweets were qualitatively assessed. The following table shows the results:

Candidate	10 most popular topics in terms of interaction
Manfred Weber	Campaigning [5], European Politics [2], International Politics [2], Brexit [1]
Frans Timmermans	Campaigning [4], Condolences [4], Brexit [1], Climate [1]
Guy Verhofstadt	Brexit [8], Campaigning [2]
Margrete Vestager	European Politics [3], Personal Life [3], Womens day [3], Technology [2], Climate [1],
Ska Keller	Migration [6], Campaigning [2], European Politics [1], Womens day [1]
Bas Eickhout	Climate [8], Campaigning [2]
Jan Zahradil	National Politics [4], European Politics [3], Campaigning [2], International Politics [1]
Yanis Varoufakis	Campaigning [3], Brexit [2], European Politics [2], National Politics [2], Condolences [1]

Table 10: The topics that gained the most interactions from each candidate

The table above shows that there were significant differences in which topics gained the most interactions for each candidate. Many of Weber's most popular tweets were about him campaigning for a better Europe, and him criticizing the political developments in Greece. Compared to other candidates, two of Weber's most popular tweets were about International Politics, suggesting that his followers were interested in his stances on the international agenda. Timmermans' most popular topics were quite different from Weber's. Despite tweeting a lot about both International women's day and his campaign slogan (ItsTime), none of the tweets about these fields of politics were the most popular. Instead, his most popular tweets were about campaigning for the elections and writing condolences to the city of Cracow after the mayor was stabbed in January 2019. Timmermans most popular tweet by far was a tweet where he cited a popular C.S Lewis quote and related it to Brexit. Another interesting example of his popular tweets was a campaigning related tweet he posted in January. This tweet is shown in the following example:

Figure 17: Frans Timmermans tweet example 2

In the tweet above he posted a video clip from one of his speeches with a strong message. He then asked his followers to retweet if they agreed. By looking at the high amount of retweets, it seems that many of his followers agreed with the message.

Based on the results it is clear that both of the candidates from ALDE had different tweet topics in their most popular tweets. Guy Verhofstadt was the candidate who gained the most likes and retweets in his most popular tweets by far. A large majority of his most popular tweets were about Brexit, suggesting that this is a topic that a majority of his followers were interested in. Like Timmermans, he also asked his followers to interact with him in his campaigning tweets. An example of this was "*How many of you can proudly say that I am European"* shown in the tweet example below:

Figure 18: Guy Verhofstadt tweet example 2

Other than that, an interesting note was that he interacted with other politicians in many of his most popular tweets about Brexit. Vestager had a different approach to using Twitter, and many of her more retweeted tweets were about topical events and wishing good fortune for her followers. This thus correlates well with her having the highest mean polarity out of all candidates. Interestingly, three of her most popular tweets were about international women's day. As these tweets were not as popular for Frans Timmermans despite him tweeting more frequently about it, this suggests that this topic engages Vestager's followers to a larger extent than Timmermans'. Vestager was also the only candidate who did not have any tweets about campaigning among her most popular tweets. By qualitatively looking at her most popular tweets, it is thus possible to assume that she was representing herself on a personal level more than her party or her role as a Spitzenkandidaten.

The candidates from the EGP's most popular tweets were also different. Keller's most popular tweets were mainly about migration. These tweets could be anything from criticizing Victor Orban's migration policies in Hungary, to taking part in the #SeaWatch3 rescue mission herself. Eickhouts most popular tweets, on the other hand, was mostly about climate change. These differences could suggest that both of these candidates had different groups of followers, or that

they had different policy areas that they were strategically focusing on. With this, it is possible that they intentionally were focusing on different areas of the EGP's politics

Jan Zahradil's most popular tweets were mostly sarcastic bites towards other politicians. This was both on a national and European level. Interestingly, his most popular tweet category was national politics, where he would for instance tweet his support of Czech Prime Minister Andrej Babis. This suggests that many of his followers are just as, if not more interested in his opinions about national politics. The fact that his best-selling tweets often were about national politics is also an interesting find, because he is the only self-proclaimed Eurosceptic Spitzenkandidat in this analysis. Yanis Varoufakis' most popular tweets were mostly about campaigning for DiEM25, opinions about Brexit and national politics where he discussed the debt of Greece. He had the largest following of all candidates, however he had no category that was significantly more popular than other categories. This could indicate that he has different kinds of followers. Some may follow him because of his new party, others may follow him because of his opinion about European Politics, while others may follow him because of his role as a national politician. If this is the case, it could make it more challenging for him to reach his full potential with his large amount of followers.

4.3.2. Tweet topics that gained the least interaction

Now that we have gained insight into what the most popular topics for the candidates were during the selected time period, a natural next step is to look at which topics were the least popular. The same method as in chapter 4.3.1 was conducted, but with the least popular tweets instead of the most popular tweets. The results were the following:

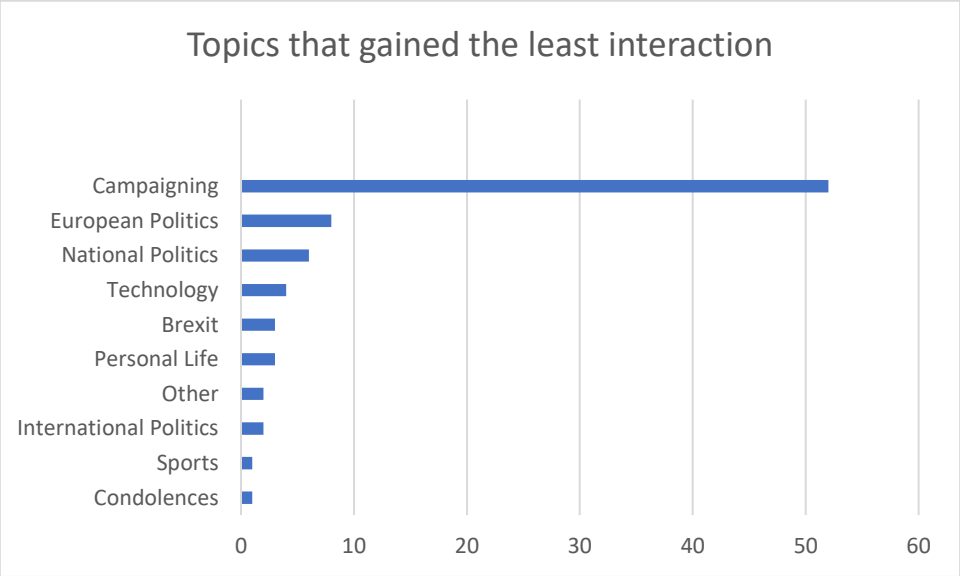

Figure 19: The least popular tweets in terms of interactions gained

The figure above (n=80) shows the 10 least popular topics in terms of interactions gained for all candidates combined. As campaigning was the highest scoring category in the previous chapter as well, this indicates that there are differences between successful and unsuccessful campaigning on Twitter. Popular topics such as Brexit and climate were much less frequently seen here, indicating that these were topics that were likely to sell well for most candidates. As in the previous chapter, the tweets were classified for each candidate. The results of this were the following:

Candidate	**10 most popular topics in terms of interaction**
Manfred Weber	Campaigning [7], National Politics [2], Technology [1]
Frans Timmermans	Campaigning [8], European Politics [2] Other [1]
Guy Verhofstadt	Campaigning [6], International Politics [2], Brexit [1], Condolences [1]
Margrete Vestager	Campaigning [5], Technology [2], National Politics [1], Other [1], Sports [1]
Ska Keller	Campaigning [6], National Politics [2], European Politics [1], Technology [1]
Bas Eickhout	Campaigning [7], Brexit [1], European Politics [1], Personal Life [1]
Jan Zahradil	Campaigning [5], European Politics [3], Brexit [1], Other [1]
Yanis Varoufakis	Campaigning [6], European Politics [2], National Politics [1], Personal Life [1]

Table 11: The topics that gained the least amount of interactions for each candidate

The table above shows each candidate's least popular tweet topics in terms of interactions gained. A qualitative assessment of these tweets indicates that Weber's least popular tweets were about election campaigning, either of him meeting EPP groups in Europe or the EPP. Many of his unpopular tweets were written in German or Polish, suggesting that tweeting in English gives him more interactions. His unpopular campaigning tweets were usually him presenting support from other central figures of the EPP. Similar traits can be seen with Timmermans, who also had a majority of his least popular tweets classified as campaigning. His least popular tweets were also commonly written in other languages than English (mostly Spanish or Italian), suggesting that tweeting in English increases the likelihood for him gaining more interactions as well.

Both candidates from ALDE were more similar to their unpopular tweets. Most of Verhofstadt's least popular tweets were classified as being related to campaigning, however his least popular tweets also included international politics. These tweets were about the political situation in Venezuela. This is interesting because tweets about the situation in Venezuela were among Weber's most popular. This suggests that Weber and Verhofstadt have different kinds of followers on Twitter. Like Timmermans and Weber, many of Verhofstadt's least popular tweets were written in another language than English. Most of Vestager's least popular tweets were also classified as being related to campaigning. These tweets were often talking about meeting other representatives in different countries. As such tweets were significantly less popular than tweets about her personal life, this could suggest that her following follows her because of other reasons than her position in ALDE.

Ska Keller and Bas Eickhout were also more similar when looking at their least popular tweets than they were with their most popular. Keller's least popular tweets were typically about campaigning for the Greens where she would often post highlights from Eickhout's speeches, or about German politics where she would typically write in German. Like Keller, Eickhout's least popular tweets were also classified as being related to campaigning for the most part. In these tweets, he would usually show highlights of himself in debates, or criticize Weber and the EPP for their stance on Victor Orban in Hungary. This suggests that most of his followers follow him because of his opinions about climate policy.

An interesting finding in the tweets from Zahradil with little interaction was that many of them were related to campaigning for ALDE. Compared to his most popular tweets where his most popular tweets were about national politics, this further suggests that many of his followers are interested in national politics. As opposed to the other candidates, many of Zahradil's most popular tweets were in Czech, while many of his least popular tweets were written in English. Surprisingly, one of Zahradil's least popular tweets was a tweet where he asked his followers to interact with him. In this tweet he wrote *"Where RU from"*, and can be seen in the following example:

Figure 20: Jan Zahradil tweet example 2

This tweet shown in the figure above did not receive any interaction at all. It is also the only tweet categorized as "other" in the graphs. This suggests that asking the followers questions is not always a successful recipe for gaining interactions, despite this being a successful recipe for both Frans Timmermans and Guy Verhofstadt. It might depend on the context of the tweet, the candidate, and the followers. Varoufakis' least popular tweet category was the "campaigning" category as well. Here he would usually post about Diem25, and many of these tweets were in Greek or German. This further suggests that there is a higher chance of gaining more interaction for tweets written in English.

With this, an interesting pattern in the tweets that received the fewest interactions was that several of them were written in another language than English. This indicates that tweeting tweets in English increases the chance of gaining more interactions for most Spitzenkandidaten, suggesting that their followers are from different countries. Zahradil was the only exception of this, as all of his most popular tweets were in Czech. This is an interesting correlation with the fact that his most popular tweet topics were national politics.

4.3.3. Weighted interaction rate per follower

The candidates chosen for this analysis were all very different when it came to the number of tweets they posted and how many followers they had. While we now know which topics gained the most and the least interaction for all candidates, we have not yet created a fair ground for comparison between the candidates. In order to gain a better understanding of how good the candidates were when it came to gaining interaction, the results were divided among the number of followers that they had to account for the differences in followers. Figuring this out would contribute to understanding which candidates used their Twitter account the best based on how many they were able to reach out to. In other words: How were they able to utilize their

potential? By following the "weighted interaction rate per follower" formula presented in chapter 3.4.4, this chapter assesses the candidates' performance based on their potential:

Candidate	**Weighted interaction rate per follower**
Guy Verhofstadt	0.0092
Manfred Weber	0.0040
Jan Zahradil	0.0029
Bas Eickhout	0.0025
Frans Timmermans	0.0024
Ska Keller	0.0018
Yanis Varoufakis	0.0005
Margrethe Vestager	0.0004

Table 12: The candidates' weighted interaction rate per follower

The table above shows the interaction rate of all candidates when taking their follower count into account. A higher number indicates better performance based on the number of followers that the candidates have. Somewhat unsurprisingly, Guy Verhofstadt was the candidate who scored the highest in the weighted interaction rate per follower. This was expected because he is one of the most known and prominent figures in the EP, and because his most popular tweets had more interactions than the other candidates' most popular tweets. He was also the most polarizing and subjective candidate as shown in chapter 4.1.2. Interestingly, his colleague Margrethe Vestager had the worst score of all candidates, so the chosen ALDE candidates were the opposite of each other in terms of weighted interaction rate per follower. Manfred Weber came second in interaction rate per follower, indicating that he was successfully able to engage his relative low follower count.

As previously mentioned, Jan Zahradil was the candidate who used the most two-way communication with his followers. With this, he scores well with his favourite count, but not necessarily retweets. If one were to not include retweets and only look at favorite count per follower, Zahradil would overtake the second place from Weber (the rest of the table would stay the same). At first look, these results indicate that Jan Zahradil is gaining many interactions by taking the time to reply to other users, thus making the most of his relative low follower count. He also was the candidate who used hashtags the most frequently, suggesting that this may have impacted his weighted interaction rate per follower positively. However, the correlation with having a high percentage of replies and gaining a higher amount of interaction is not seen when looking at Frans Timmermans, whose amount of replies accounted for more

than half of his tweets. Timmermans only had the fifth highest weighted interaction rate per follower and did slightly worse than Bas Eickhout who had the third highest percentage of replies. These findings thus indicate that there isn't necessarily a correlation between replying to other users and gaining a high interaction rate.

Surprisingly, Varoufakis had the second lowest interaction rate per follower. This was surprising because he is a well-known figure in European politics. While this is true, he did have the highest number of followers which could have affected his performance negatively. To gain insight into this, the code was rerun to not account for the number of followers. This alternative table can be found in Appendix B. A rerun of this code showed that even if one were to not divide interaction with the number of followers, Verhofstadt would still pass him in interactions gained, while Varoufakis would overtake the second place. This suggests that Varoufakis is not using his Twitter optimally to take advantage of his large following, or that there are different reasons and interests among his followers (as suggested in 4.3.2.). Another explanation could be that Varoufakis had many inactive followers making his high follower count a bad indication of how many people he was able to reach out to. Nevertheless, these findings indicate that Varoufakis was not using his Twitter account optimally during the three-month period when looking at the interaction rate per follower.

These results have shown that there is a significant difference in how the candidates are utilizing the potential of their followers on Twitter. Guy Verhofstadt is clearly the candidate who is able to create the most engagement from his followers. As shown in previous chapters, Verhofstadt was one of the least neutral candidates and also one of the least subjective. He was also consistently creating posts where he asked his followers to interact with him. Like Verhofstadt, Weber was also mixing tweets related to campaigning with taking a stance to both European and International politics. This strategy seems to have worked well for both of them and seems to have sold well for their target groups. Another factor that could have contributed to their high interaction rate is media coverage. As Verhofstadt is one of the most known figures in the EP and Weber is the candidate from the largest political party in the EP, it would not be unheard of to assume that this could have influenced their high interaction rate.

Neither of those most successful candidates in terms of interaction rate replied frequently to their followers, however this was a very frequent habit of Zahradil. He seems to have gained a lot of interactions from doing so, thus indicating that replying to other users can be a factor that increases the number of interactions gained, however this is a subject that should be studied

further. Jan Zahradil and Manfred Weber were also the candidates who used hashtags the most frequently, and both scored high on their interaction rate. Varoufakis and Vestager, on the other hand, used the fewest hashtags with no specific hashtag frequenting often in their most popular tweets. They had the lowest scores in interaction rate, suggesting that there is a correlation between using hashtags frequently and gaining interactions. The three highest scoring candidates (Verhofstadt, Weber and Zahradil) were also seen tweeting frequently about highly covered topics such as Brexit, suggesting that doing so can contribute positively to candidates' interaction rates. These results thus posting tweets in English, using hashtags frequently, and taking stances in highly covered political topics, all can be factors that can increase interactions gained on Twitter.

4.3. Interpretation of the findings

This chapter assesses the results found in this chapter with the hypotheses presented in chapter 2.6. By looking at the results presented in chapter 4, it is clear that all candidates had different strategies for using Twitter in the selected three-month period from 15th December to 15th March. This was even the case for the candidates who represented the same political party. This was seen in the languages they typically tweet in, the number of hashtags they used, and how frequently they replied to other users.

The results also showed that there were differences between the candidates representing political parties in power and the candidates representing political parties in opposition. As previously mentioned, the EP's coalition government from 2014 was EPP, PES, and ALDE, and the candidates from these parties mostly found to tweet less frequently than the candidates from political parties in opposition. The only exceptions for this were Guy Verhofstadt from the ruling parties and Ska Keller from the opposition. This correlates well with both previous research about national elections and European elections saying that the opposition posts more frequently on social media (Nulty et al 2016; Vergeer & Hermans, 2013; Larsson & Kalnes, 2014). Another interesting finding from the candidates representing parties in opposition was that they were more likely to tweet in other languages. There was no clear correlation between representing parties in opposition or power, and using more hashtags or replying more often during the time period. All of the findings of the candidates thus correlate well with **H1: The candidates representing the political parties in power use Twitter differently than the candidates representing the political parties in opposition**.

The present thesis has also looked at the sentiments of each candidate. Doing so, showed a smaller difference in the candidates' tweet sentiments than expected. Based on previous research**, H2 assumed that Eurosceptic candidates would be more negative and subjective than the pro-European candidates.** While the results showed that the most pro-European candidates from ALDE also were the most positive, there were never any extreme differences between the candidates. The candidates representing ALDE were also the most subjective candidates. The findings of this thesis thus suggest that there were not much correlation between the degree of Euroscepticism and tweet sentiments for the 2019 Spitzenkandidaten. Again, it is important to note that a majority of the tweets of Bas Eickhout and Jan Zahradil were not written in English, and that no candidates from the far left or far right were included in this analysis. There are thus some limitations when testing H2.

When it came to the public tweeting about the elections, the results from the showed that these people are usually from Western European countries, and that they were most likely to tweet in English. Furthermore, the results showed small differences when it came to the tweet sentiments of all hashtags. The tweets using the hashtags related to the Spitzenkandidaten process did not differ much from the tweets related to the European elections. Since #BetterEurope, #EuropeanSpring and #RenewEurope all had higher mean polarities than the hashtags about the European elections, there is a slight correlation with **H3: The public who tweet about hashtags related to the Spitzenkaniddaten's campaigns post more positive tweets than the public who tweet about the European elections.** It is important to keep in mind that the tweets using the hashtags #EP2019 and #EuropeanElections2019 were much more frequently used, and the sentiments from these tweets are thus more significant. Because of this, the findings were not deemed sufficient enough to confirm H3.

When it came to how the candidates were able to gain interaction from their followers, the results did not indicate that there was one clear path for successfully gaining interaction for the candidates on Twitter. However, one of the more clear correlations were found were that the candidates who used hashtags the most frequently had the highest interaction rate per follower. Recall that Manfred Weber, Guy Verhofstadt and Jan Zahradil were the candidates that used hashtags the most frequently. They were also the candidates with the highest interaction rates per follower. This thus suggests that the candidates are able to reach out to more followers using hashtags frequently (see Hanteer et al, 2018). This thus suggests a confirmation of **H4: Using hashtags more frequently increases the number of interactions gained by the candidates.**

Furthermore, **H5 predicted that there is a correlation posting polarizing and subjective tweets and gaining more interactions on Twitter.** However, the candidates were much less polarizing and much more neutral than expected. The most polarizing and subjective candidate was Margrethe Vestager, the candidate with the lowest interaction rate per follower out of all. However, Verhofstadt with the highest interaction rate per follower was the second most polarizing and second most subjective candidate. Despite Verhofstadt slightly correlating with H5, the results did not correlate well with H5 in general. These results should be taken with a slight grain of salt, as a majority of Jan Zahradil's tweets were written in Czech. These tweets were thus not counted in the sentiment analysis, and the sentiments of these Czech tweets could have contributed to his high interaction rate.

Finally, we have seen that the findings of this thesis partly correlate with **H6: Replying to other users increases the number of interactions gained by the candidates.** Jan Zahradil could be an example of this being the case as he had both a high interaction rate per follower and the highest reply count. He was also among the most neutral and objective candidates, suggesting that his high interaction rate was not a result of factors such as tweet sentiments. Timmermans and Eickhout also had a high percentage of replies to their followers, however they both had a notably lower interaction rate per follower. Because of this, it is not yet possible to determine if the percentage of replies had an effect on the interaction rate.

By interpreting these findings, it is clear that there is not yet a clear pattern for being a Spitzenkandidaten on Twitter. The candidates all seemed to have different strategies to their Twitter accounts, and the differences most and least popular Twitter topics indicate that there were different followings for each candidate as well. Despite this, some topics such as Brexit and climate change were more universally popular than other topics, and the results suggested that tweeting in English and using hashtags frequently is successful strategies to gain interaction for all candidates. The Spitzenkandidaten process is an election process that is still fresh, and it will be interesting to see if there will develop a more clear identity of being a Spitzenkandidat over time.

5. Summary

By looking at the findings from the chapter 4, we now know how the candidates used Twitter, typical traits of the public discussing the elections, and we have acquired knowledge in how the 2019 Spitzenkandidaten were able to gain interaction on Twitter during the chosen three-month period. This chapter summarizes and concludes this thesis based on the results and the discussion from chapter 4. This is done by first discussing the thesis' strengths and weaknesses by using the key terms validity, reliability, and replicability. Following this, this chapter summarizes the findings and concludes the thesis. Finally, this chapter gives suggestions for further research about the Spitzenkandidaten process on Twitter.

5.1. Strengths and limitations

This study has both strengths and limitations to take into consideration. Three central methodological concepts are used to discuss the strengths and limitations of this study: Validity, reliability, and replicability. Validity is an indication of how the study's findings actually answer the research question. To have a high extent of validity, the researcher must be transparent with his or her choices and discuss the consequences from them (Tjora, 2017, p 232). A study's replicability can be defined as the extent that new experiments with similar research questions will consistently produce similar results (Patil et al, 2016, 540). Finally, reliability in research means to what extent exists any bias with the researcher. One can never achieve complete objectiveness in research and some bias will thus always occur. Because of this, it is always important to be open about potential biases in the conducted research (Tjora, 2017, p. 235). By assessing the study's strengths and limitations with the key terms presented above, this chapter shows that this thesis has many strengths and limitations.

I argue that this study's validity is high because the present thesis consists of a detailed methods chapter explaining all methodological choices and their potential consequences from them. While some of the choices made for this study may not have been the objectively best choices, they were looked upon as good choices for me, as I was not familiar with using Python or analysing tweets. The most important thing for me was that I understood the programs I was using so I could be as open and transparent as possible with my results. No matter which methodological choices I made, there would always be both strengths and limitations depending on which API-package I chose, the program analysing tweets, and the libraries for sentiment analysis. Being open about these choices and leaves other scholars with better a better background to create stronger analyses in the future. Furthermore, I argue that the fact

that tweets were both qualitatively and quantitatively assessed strengthens this thesis' validity. This made it possible to both be able to analyse large amounts of data, while also making it possible to understand the context behind the data by qualitatively assessing tweets. The thesis' also answers the research question in a concrete manner by having hypotheses as a running theme throughout the thesis.

In addition to this, the replicability of this study is high because of the published codebook on GitHub. Other researchers with basic Python knowledge can easily run the same codes to find the same findings as this study. This is possible with just a few minor tweaks of the codes uploaded in the GitHub repository. Most of the exact tweets gathered from the hashtags and the candidates are available from the same repertoire, making it possible to run the same codes on the exact same files and thus gain the same results. The only files lacking from the repository were all tweets gathered from the #EP2019 and #EuropeanElections2019 hashtags, as these files were too large for the platform. Having made this repository public largely strengthens the replicability of this study. However, it is worth noting it would not be possible to get the exact same results if one were to gather tweets again from the same, or a similar time period. While it would be likely that one would find similar results by gathering tweets from that period, it would be unlikely that the exact same results would be found due to time lag. Time lag will always be a relevant problem, as some tweets might be deleted and some accounts may be deleted or made private (inaccessible from API's). Therefore, other researchers would never be able to get the exact same results if the tweets were gathered at a later point, thus limiting the replicability of this study to a small extent if this was chosen for a similar study. With this, one would be very likely to find the same results if one were to base the analysis on the same GitHub repertoire as this study.

I have tried to appear as reliable as possible for this study. All methodological choices are explained in detail, and all codes are uploaded to a public repository that will not be changed after May 15th 2019. The methods chapter is open about limitations with both the chosen API-package, Python as the program doing the analysis and the limitations connected with Textblob for sentiment analysis. While I am interested in the European elections, I do not have any connections to any political parties or any people affiliated with them. My interest in the elections may have influenced which findings I found the most interesting. However, the focus of this thesis has always been to answer the research question. With this, I argue that the findings of the thesis are reliable.

While I argue that this thesis has high validity, replicability, and reliability, there are some notable limitations to this thesis. One of these can be the chosen time-period. The period chosen was the 15th of December 2018 to the 15th of March 2019. This period limited some other possibilities for the research such as comparing the results of the thesis with the election results. The main reason for choosing this time-period was the hand in date, and I wanted to make sure that I contributed to filling a research gap. Despite these limitations, the period has given us a clear overview of how the Spitzenkandidaten 2019 process was discussed on Twitter, and could thus function as a framework for further research about this topic. There have previously been conducted studies about how European elections on Twitter, and how the public tweet about the Spitzenkandidaten process, however no analysis have yet analysed how the Spitzenkandidaten tweets or gain interaction on Twitter. The present thesis thus contributes to filling an important research gap about European Parliament elections on social media, despite the time-period perhaps not being optimal for researching this.

Another limitation is related to the selection that was available for the thesis. Several parties in the EP decided not to put forward a candidate for these elections, resulting in many political parties being left out of this study. This was especially unfortunate for the sake of this study with the case of EL lacking candidates with active Twitter accounts, as it would have been interesting to add examples from the far left to the study. The far right was also not represented in this study, as both Europe of Freedom and Direct Democracy (EFDD) and Europe of Nations and Freedom (ENF) did not present a candidate. With this, this study was not able to take far-right and far-left parties into consideration, which would have likely impacted the sentiment analyses of the candidates. The study also chose two out of seven candidates for ALDE, as picking all seven candidates would make it so that they would count for more than half of the chosen parties. With this, some interesting Twitter strategies may have been left out.

5.2. Conclusion

The present thesis has attempted to find out which conditions make the Spitzenkandidaten for the 2019 European elections gain interaction from other Twitter users. By looking at 1560 tweets from the candidates and 5059 tweets from related hashtags, this thesis has found several correlations with previous research. This thesis has first looked at how the candidates for the 2019 Spitzenkandidaten elections were using Twitter prior to the elections. This was done by qualitatively assessing tweets from the candidates, looking at which languages they used, the hashtags they were using, the candidates' tweet sentiments, and to what extent they replied to other users. Secondly, this thesis has looked at typical traits about the public who tweeted about

the elections. This was done by looking at the language of their tweets, their locations, and by looking at the sentiments of their tweets to see whether these users were tweeting positive or negative messages about the elections. Lastly, this thesis has looked at how the candidates were able to gain interaction on Twitter. This was done by qualitatively assessing both the most and the least popular tweets of the candidates. Following this, the candidates were compared based on weighted interaction rate per follower, and the results became interpreted to determine to what extent they correlated with the six hypotheses created from the thesis' theoretical framework

When it came to the candidates Twitter usage, this thesis found that Spitzenkandidaten in opposition differs from the Spitzenkandidaten representing political parties in power. In general, the candidates representing political parties in opposition (Bas Eickhout, Ska Keller, Jan Zahradil, and Yanis Varoufakis) tweeted more frequently than the candidates representing political parties in power. They also tweeted more frequently in other languages than English than the candidates from political parties in power. The topics of the candidates' tweets varied greatly, however some topics such as climate change, Brexit, and reducing the gender gap were common universal topics for most candidates during the selected time period. Surprisingly, this thesis did not find any clear correlation between posting more polarizing tweets and the extent the candidates were Eurosceptic. An explanation of this could perhaps be that this thesis did not include any hard Eurosceptic candidates. Another finding of note was that the candidates very rarely interacted with each other, nor mentioned the Spitzenkandidaten process much. This indicates that there was somewhat of an overlap with campaigning for the European elections and the Spitzenkandidaten process.

Over to the public who tweeted about the elections, the findings of this thesis show that a majority of the public tweeting about the elections and the candidates were from Western European countries. Surprisingly, the most frequent locations of the users posting tweets using the hashtags were from the UK. When looking at the hashtags individually, some notable findings were that most of the tweets using #RenewEurope were from users who had set their location to 'Europe' or 'Belgium', and that #EuropeanSpring had a high percentage of users located in France. Most of the tweets using these hashtags were posted in English and a notable amount of tweets was posted in Dutch and German. A surprisingly low number of tweets were posted in Spanish. The sentiments of these tweets were quite neutral and objective, and there were not many notable between the hashtags discussing the elections and the hashtags about the candidates. There were some exceptions where hashtags related to the candidates would be

more polarizing than the hashtags discussing the elections, but these hashtags were scarcely used compared to the other hashtags during the selected period.

When it comes to under which conditions the candidates' were able to gain interaction on Twitter, this thesis made some striking findings. First, a clear correlation between using hashtags frequently and gaining more interaction on Twitter was found for the 2019 Spitzenkandidaten during the three-month period. The results chapter has also shown that tweeting in English correlates better with gaining a high amount of interactions than in other languages. However, their choice of tweeting in other languages could contribute to them reaching out to followers from certain member states in a better way. As all candidates posted tweets in at least two languages during the selected time period, the findings suggest that this was a deliberate strategy for all candidates. Because of this, tweeting in other languages should not necessarily be reviewed as a poor strategy, however the results showed that this was an inferior way of gaining interactions compared to tweeting in English.

Furthermore, the thesis found a slight correlation between being polarizing and subjective on Twitter and gaining more interaction, however this was not sufficient for creating a conclusion about this matter. This was because the program for sentiment analysis did not account for other languages than English, and because this thesis did not analyse candidates from the far right or far left. In addition to this, the results suggested that taking stances in mediated topics could increase the number of interactions. Finally, the results of this thesis did not find a correlation between replying to other users and gaining more interaction on Twitter. Some findings such as Jan Zahradil's high interaction point towards this being the case, however both Frans Timmermans and Bas Eickhout had a significantly lower interaction rate per follower than Zahradil despite them also frequently replying to other users. As the previous literature assessed in this thesis were split in their opinion about this as well, this would be an interesting topic to investigate further

With the findings from the results chapter in mind, this thesis has found that the 2019 Spitzenkandidaten were all very different on Twitter. They all had different Twitter usage, different groups of followers, and their number of followers varied largely. While the findings of this study suggest that both using hashtags frequently and tweeting in English increases the number of interaction gained for the candidates, the findings were not clear in other areas. Because of this, it seems that there is not yet a golden route for successfully gaining interaction as Spitzenkandidat on Twitter. Another perspective to keep in mind is that external

factors could have contributed to Verhofstadt's and Weber's high weighted interaction rates per followers. We are now approaching the second European elections featuring the new Spitzenkandidaten process. As the candidates were this different in the way they used Twitter, it seems like there is not yet a standard for how one should act as a Spitzenkandidaten on Twitter. Perhaps the identities and standards of different Spitzenkandidaten on social media will be more clear as the process ages.

5.3. Further Research

This study has laid many foundations for further research about the Spitzenkandidaten process on Twitter. While the Spitzenkandidaten process generally has received attention from many scholars, EP elections on social media, and especially the Spitzenkandidaten process on social media are currently scarcely researched topics. As the Spitzenkandidaten process is still a significant research gap in the academic literature about social media, it is suggested that the results of this thesis should be looked upon as building blocks for further research instead of concrete answers. When it came to the findings of this thesis, some of the results correlated more clearly with previous research than others. For instance, the results of this gave a clear correlation between using hashtags frequently and gaining more interaction, while the results were not as clear when looking at correlations between replying to other users and gaining more interaction. With this, it is suggested that one should especially look at correlations between replying to other users and gaining more interactions on Twitter if conducting a similar analysis in the future.

In addition to this, this thesis has only looked at tweets from the candidates in a three-month period. Doing a later analysis with the results from the elections and future of the Spitzenkandidaten process determined could further contribute to an even better understanding of campaigning for the process. It could also help us see if it is possible to see a correlation between gaining interactions on Twitter and election results, and will there be a change in how many o what extent will the election results change the candidates' popularity. Will they revert to tweeting less frequently after the election results? Will there still be as much of a difference between the candidates from the ruling parties and the opposition after the elections have taken place? All of these questions would be interesting to look at for further research.

6. References

Allcott, H., & Gentzkow, M. (2017). "*Social media and fake news in the 2016 election. Journal of economic perspectives*", 31(2), 211-36.

BBC (2019) *"European elections: What can UK MEPs expect?"* Retrieved 05.05.2019 from: https://www.bbc.com/news/uk-politics-48037162

Bekmamedova, N., & Shanks, G. (2014). *"Social media analytics and business value: A theoretical framework and case study.*" In 2014 47th Hawaii international conference on system sciences (pp. 3728-3737). IEEE.

Boyd, D., Golder, S., & Lotan, G. (2010). *"Tweet, tweet, retweet: Conversational aspects of retweeting on twitter."* In 2010 43rd Hawaii International Conference on System Sciences (pp. 1-10). IEEE.

Braun, D., & Popa, S. A. (2018). *"This time it was different? The salience of the Spitzenkandidaten system among European parties."* West European Politics, 41(5), 1125-1145.

Burns, Charlotte (2013*) "The European Parliament"* in Cini, M., & Borragán, N. P.-S. (2013). *European Union Politics* (4th ed.). Oxford: Oxford University Press.

Castro, S., Cubero, M., Garat, D., & Moncecchi, G. (2016, November). *"Is this a joke? Detecting humor in Spanish tweets."* In Ibero-American Conference on Artificial Intelligence (pp. 139-150). Springer, Cham.

Chae, J., Thom, D., Bosch, H., Jang, Y., Maciejewski, R., Ebert, D. S., & Ertl, T. (2012, October). *"Spatiotemporal social media analytics for abnormal event detection and examination using seasonal-trend decomposition."* In IEEE VAST (pp. 143-152).

Christiansen, T. (2016). *"After the Spitzenkandidaten: fundamental change in the EU's political system?"* West European Politics, 39(5), 992-1010.

DiEM25 (2019) "*What is DiEM25?"* Retrieved 01.04.2019 from: https://diem25.org/what-is-diem25/

Dinan, D. (2014). *"Europe recast: a history of European Union"* 2nd edition Basingstoke: Palgrave Macmillan.

Dinan, D. (2015). *"Governance and institutions: The year of the Spitzenkandidaten."* JCMS: Journal of Common Market Studies, 53, 93-107.

Europeelects, (2019) "*2019 Spitzenkandidaten"* Retrieved 26.03.2019 from: https://europeelects.eu/2019spitzenkandidaten/

European Greens (2019) "*Leading Candidates"* Retrieved 01.04.2019 from: https://vote.europeangreens.eu/leading-candidates

European Left (2019) *"Two Candidates From The People For The People"* Retrieved 01.04.2019 from: https://www.european-left.org/campaigns/two-candidates-from-the-people-for-the-people/

European People's Party (2019) *"Manfred Weber. My Ideas"* Retrieved 01.04.2019 from: https://weber.epp.eu/my_ideas

European Spring (2019) *"European Spring is Europe's first transnational party"* Retrieved 01.04.2019 from: https://europeanspring.net/about-european-spring/

Gillespie, T. (2014). *"The relevance of algorithms."* Media technologies: Essays on communication, materiality, and society, 167.

Gligorić, K., Anderson, A., & West, R. (2018). *"How Constraints Affect Content: The Case of Twitter's Switch from 140 to 280 Characters."* In Twelfth International AAAI Conference on Web and Social Media.

Fan, W., & Gordon, M. D. (2014). *"The power of social media analytics."* Commun. Acm, 57(6), 74-81.

Fischer, E., & Reuber, A. R. (2011). *"Social interaction via new social media:(How) can interactions on Twitter affect effectual thinking and behavior?"* Journal of business venturing, 26(1), 1-18.

Hanteer, O., Rossi, L., D'aurelio, D. V., & Magnani, M. (2018). "*From interaction to participation: the role of the imagined audience in social media community detection and an application to political communication on Twitter*." In *2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)* (pp. 531-534). IEEE.

Herring, S. C., Scheidt, L. A., Bonus, S., & Wright, E. (2004, January). *"Bridging the gap: A genre analysis of weblogs."* In 37th Annual Hawaii International Conference on System Sciences, 2004. Proceedings of the (pp. 11-pp). IEEE.

Hobolt, S. B. (2014). *"A vote for the President? The role of Spitzenkandidaten in the 2014 European Parliament*" elections. *Journal of European Public Policy*, *21*(10), 1528-1540.

Hobolt, S. B., & De Vries, C. (2016). "*Turning against the Union? The impact of the crisis on the Eurosceptic vote in the 2014 European Parliament elections."* Electoral Studies, 44, 504-514.

Holton, A. E., & Lewis, S. C. (2011). *"Journalists, social media, and the use of humor on Twitter."* Electronic Journal of Communication, 21(1/2), 1-22.

Heiss, R., von Sikorski, C., & Matthes, J. (2019). *"Populist Twitter Posts in News Stories: Statement Recognition and the Polarizing Effects on Candidate Evaluation and Anti-Immigrant Attitudes."* Journalism Practice, 1-17.

Hsu, C., & Park, H. W. (2012). *"Mapping online social networks of Korean politicians."* Government Information Quarterly, 29(2), 169–181. doi:10.1016/j.giq.2011.09.009

Jungherr, A., Schoen, H., & Jürgens, P. (2015). *"The mediation of politics through Twitter: An analysis of messages posted during the campaign for the German federal election"* 2013. Journal of Computer-Mediated Communication, *21*(1), 50-68.

Jürgens, P. & Jungherr, A. (2016) *"A Tutorial for Using Twitter Data in the Social Sciences: Data Collection, Preparation, and Analysis"*. Available at SSRN: http://ssrn.com/abstract=2710146

Kaplan, A. M., & Haenlein, M. (2010). *"Users of the world, unite! The challenges and opportunities of Social Media."* Business horizons, 53(1), 59-68.

Karkın, N., Yavuz, N., Parlak, İ., & İkiz, Ö. Ö. (2015, May). *"Twitter use by politicians during social uprisings: an analysis of Gezi park protests in Turkey."* In Proceedings of the 16th Annual International Conference on Digital Government Research (pp. 20-28). ACM.

Klašnja, M., Barberá, P., Beauchamp, N., Nagler, J., & Tucker, J. (2016). *"Measuring public opinion with social media data. Handbook on polling and polling methods".*

Kruikemeier, S. (2014). *"How political candidates use Twitter and the impact on votes." Computers in human behavior*, *34*, 131-139.

Kwak, H., Lee, C., Park, H., & Moon, S. (2010) *"What is Twitter, a social network or a news media?".* In *Proceedings of the 19th international conference on World wide web* (pp. 591-600). AcM.

Larsson, A. O. (2015). "*The EU Parliament on Twitter—assessing the permanent online practices of parliamentarians."* Journal of Information Technology & Politics, 12(2), 149-166.

Larsson, A. O., & Ihlen, Ø. (2015). *"Birds of a feather flock together? Party leaders on Twitter during the 2013 Norwegian elections."* European Journal of Communication, 30(6), 666-681.

Larsson, A. O., & Kalsnes, B. (2014*). "'Of course we are on Facebook'": Use and non-use of social media among Swedish and Norwegian politicians."* European Journal of Communication, 29(6), 653–667. doi:10.1177/0267323114531383

Larsson, A. O., & Moe, H. (2012). *"Studying political microblogging: Twitter users in the 2010 Swedish election campaign."* New Media & Society, 14(5), 729-747.

Lee, E. J., & Shin, S. Y. (2012). *"Are they talking to me? Cognitive and affective effects of interactivity in politicians' Twitter communication."* Cyberpsychology, Behavior, and Social Networking, doi: 15(10), 515-520.

Lifewire (2019) *"What Is Twitter & How Does It Work?"* Retrieved 02.05.2019 from: https://www.lifewire.com/what-exactly-is-twitter-2483331

Liu, B. (2012). *"Sentiment analysis and opinion mining. Synthesis lectures on human language technologies",* 5(1), 1-167.

Lubanovic, B. 2014 "*Introducing Python: Modern Computing in Simple Packages*" O'Reilly Media

Mejova, Y., Srinivasan, P., & Boynton, B. (2013, February). Gop primary season on twitter: popular political sentiment in social media. In Proceedings of the sixth ACM international conference on Web search and data mining (pp. 517-526). ACM.

Mitchell, Ryan. 2018 *"Web Scraping with Python: Collecting More Data from the Modern Web"* 2nd edition O'Reilly Media

Morstatter, F., Pfeffer, J., Liu, H., & Carley, K. M. (2013). *"Is the sample good enough? comparing data from twitter's streaming API with twitter's firehose."* In Seventh international AAAI conference on weblogs and social media.

Mustafaraj, E., Finn, S., Whitlock, C., & Metaxas, P. T. (2011, October). *"Vocal minority versus silent majority: Discovering the opinions of the long tail."* In 2011 IEEE Third International Conference on Privacy, Security, Risk and Trust and 2011 IEEE Third International Conference on Social Computing (pp. 103-110). IEEE.

NLTK (n.d.) "*Sentiment Analysis"* Retrieved 22.04.2019 from: http://www.nltk.org/howto/sentiment.html

Nulty, P., Theocharis, Y., Popa, S. A., Parnet, O., & Benoit, K. (2016). *"Social media and political communication in the 2014 elections to the European Parliament*". Electoral studies, 44, 429-444. https://doi.org/10.1016/j.electstud.2016.04.014

O'Connor, B., Balasubramanyan, R., Routledge, B. R., & Smith, N. A. (2010). *"From tweets to polls: Linking text sentiment to public opinion time series."* In Fourth International AAAI Conference on Weblogs and Social Media.

Park, C. S. (2013). *"Does Twitter motivate involvement in politics? Tweeting, opinion leadership, and political engagement. Computers in Human Behavior",* 29(4), 1641-1648.

Party of European Socialists (2019) "*Frans Timmermans: #ItsTime for 'A New Social Contract for Europe*". Retrieved 01.04.2019 from: https://www.pes.eu/en/news-events/news/detail/Frans-Timmermans-ItsTime-for-A-New-Social-Contract-for-Europe/

Patil, P., Peng, R. D., & Leek, J. T. (2016). *"What should researchers expect when they replicate studies? A statistical view of replicability in psychological science."* Perspectives on Psychological Science, 11(4), 539-544.

Planspace (2015) *"TextBlob Sentiment: Calculating Polarity and Subjectivity"* Retrieved 22.04.2019 from: https://planspace.org/20150607-textblob_sentiment/

Politico (2018) *"EU leaders: We won't be bound by Spitzenkandidat process"* Retrieved 24.03.2019 from: https://www.politico.eu/article/jean-claude-juncker-spitzenkandidat-eu-leaders-we-wont-be-bound-by-spitzenkandidat-process/

Politico (2019a) *"Michel Barnier back pushing his vision for Eu's future"* Retrieved 20.04.2019 from: https://www.politico.eu/article/michel-barnier-back-pushing-his-vision-for-eu-future-epp/

Politico (2019b) "*The trials and tribulations of Manfred Weber"* Retrieved 20.04.2019 from: https://www.politico.eu/article/manfred-weber-trials-and-tribulations-european-commission-presidency-race/

Puschmann, C., & Burgess, J. (2013). *"The politics of Twitter data"* (HIIG Discussion Paper Series No. 2013-01). Retrieved from: http://papers.ssrn.com/abstract=2206225

Raz, Y. (2012). *"Automatic humor classification on Twitter."* In Proceedings of the 2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies: Student Research Workshop (pp. 66-70). Association for Computational Linguistics.

Rodriguez, M. G., Gummadi, K., & Schoelkopf, B. (2014). *"Quantifying information overload in social media and its impact on social contagions."* In Eighth International AAAI Conference on Weblogs and Social Media.

Russell, M. A. 2014 "*Mining the Social Web: Data Mining Facebook, Twitter, LinkedIn, Google+, GitHub, and More"* 2nd ed. Sebastopol, CA: O'Reilly Media

Schmitt, H., Hobolt, S., & Popa, S. A. (2015). *"Does personalization increase turnout? Spitzenkandidaten in the 2014 European Parliament elections."* European Union Politics, 16(3), 347-368.

Shackleton, M. (2017) *"The European Parliament: The power of democratic ideas"* In D. Hodson & J. Peterson (Ed.), The Institutions of the European Union (4th edition). Oxford: Oxford University Pres

SNP (2019) "*Polls show people want change from failed Westminster system"* Retrieved 29.04.2019 from: https://www.snp.org/polls-show-people-want-change-from-failed-westminster-system/

Spain in English, (2019) *"Junqueras chosen as EFA candidate to head EU Commission"* Retrieved 24.03.2019 from: https://www.spainenglish.com/2019/03/07/junqueras-efa-candidate-eu-commission/

Spierings, N., & Jacobs, K. (2014). Getting personal? The impact of social media on preferential voting. Political Behavior, 36(1), 215-234.

Statista, (2019) *"Number of monthly active Twitter users worldwide from 1st quarter 2010 to 3rd quarter 2018 (in millions)"* Retrieved 04.02.2019 from: https://www.statista.com/statistics/282087/number-of-monthly-active-twitter-users/

Stieglitz, S., & Dang-Xuan, L. (2012). *„Political communication and influence through microblogging--An empirical analysis of sentiment in Twitter messages and retweet behavior."* In 2012 45th Hawaii International Conference on System Sciences (pp. 3500-3509). IEEE.

Stieglitz, S., & Dang-Xuan, L. (2013). *„Social media and political communication: a social media analytics framework. Social network analysis and mining",* 3(4), 1277-1291.

Stieglitz, S., & Dang-Xuan, L & Bruns, A & Neuberg, C. (2014). *„Social Media Analytics: An Interdisciplinary Approach and Its Implications for Information Systems."* Business & Information Systems Engineering. Forthcoming. DOI: 10.1007/s11576-014-0407-5.

Stieglitz, S., Mirbabaie, M., Ross, B., & Neuberger, C. (2018). *"Social media analytics–Challenges in topic discovery, data collection, and data preparation".* International journal of information management, 39, 156-168.

Stier, S., Bleier, A., Lietz, H., & Strohmaier, M. (2018). *„Election campaigning on social media: Politicians, audiences, and the mediation of political communication on Facebook and Twitter."* Political communication, 35(1), 50-74.

Team Europe (2019) "*Our Manifesto"* Retrieved 01.04.2019 from: https://teameurope.eu/our-manifesto/

Textblob (2018a) *"Tutorial: Quickstart"* Retrieved 22.04.2019 from: https://textblob.readthedocs.io/en/dev/quickstart.html#sentiment-analysis

Textblob (2018b) "*API Reference"* Retrieved 22.04.2019 from: https://textblob.readthedocs.io/en/dev/api_reference.html#wordnet

Tjora, A. (2017). '*Kvaltative Forsknihjngsmetoder i praksis'* 3rd edition Gyldendal Norsk Forlag AS

Twitter (2019a) *'About replies and mentions'* Twitter help center. Retrieved 22.04.2019 from: https://help.twitter.com/en/using-twitter/mentions-and-replies

Twitter (2019b) "*Developer Tools and APIs built for innovation and scale"* Twitter Developer Centre. Retrieved 22.04.2019 from: https://developer.twitter.com/en/pricing.html

Twitter (2019c) *"Evaluate Twitter data to inform business decisions"* Twitter Developer Centre. Retrieved 05.03.2019 from: https://developer.twitter.com/en/use-cases/analyze

The Conservative (2019) "*The EU could be A Good Servant but is A Bad Master".* Retrieved 01.04.2019 from: http://www.theconservative.online/tCn-02.pdf

Tromble, R. (2018). *"Thanks for (actually) responding! How citizen demand shapes politicians' interactive practices on Twitter."* New media & society, 20(2), 676-697.

Van Kessel, S., & Castelein, R. (2016). *„Shifting the blame. Populist politicians' use of Twitter as a tool of opposition".* Journal of contemporary European research, 12(2).

Vergeer, M. (2015). *"Twitter and political campaigning."* Sociology compass, 9(9), 745-760.

Vergeer, M. R. M., Hermans, E. A. H. M., & Sams, S. (2011). *"Is the voter only a tweet away? Micro-blogging in the 2009 European Parliament elections."*

Vergeer, M., Hermans, L., & Sams, S. (2013). "*Online social networks and micro-blogging in political campaigning: The exploration of a new campaign tool and a new campaign style."* Party Politics, 19(3), 477–501. doi:10.1177/1354068811407580

White, M. D., & Marsh, E. E. (2006). Content analysis: A flexible methodology. Library trends, 55(1), 22-45.

Yu, Y., & Wang, X. (2015). *"World Cup 2014 in the Twitter World: A big data analysis of sentiments in US sports fans' tweets".* Computers in Human Behavior, 48, 392-400.

Zappavigna, M. (2011). *"Ambient affiliation: A linguistic perspective on Twitter. New Media & Society"*, 13(5), 788-806.

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

Appendix A: Codebook for the analyses

The scripts used for this study are available in the following GitHub repository: https://github.com/kristianlyster/twitter-json-tools. These are called twitter-json-tools and are a small set of functions and scripts intended to make processing and performing analysis on Twitter API JSON data easier. In this repertoire, you can see all changes, comments and the time/date the files were uploaded in. This can be regarded as proof that none of the files have been edited after the thesis was handed in. All lines include short descriptions of what they are doing. Some lines are highlighted with a #-mark. This is because several different scripts are included in the main.py with different objectives. Changing the file to test the different codes should be intuitive to most people familiar with Python. The example script assumes JSON files with names in the format "username-tweets.json" are placed in a /json directory. These JSON files should be multiline, with each line containing a JSON Tweet object as defined in the Twitter API documentation. The Twitter API documentation is included in the GitHub repertoire as well.

Usage

For the most basic uses, modifying and/or appending to the included main.py should be plenty.

Requirements

These functions and scripts are based only on the core Python libraries. Only tested on python 3.6.4, but should work on any python>=3

Appendix B: Interaction rate when not taking follower count into consideration

Interaction rate for all candidates without dividing the numbers on each candidate's amount of followers. The numbers indicate average interactions per tweet.

Candidate	**Interaction rate**
Guy Verhofstadt	3289
Yanis Varoufakis	459
Frans Timmermans	249
Manfred Weber	126
Margrethe Vestager	97
Ska Keller	65
Bas Eickhout	46
Jan Zahradil	18

NTNU

Kunnskap for en bedre verden