HOW DOES POLARISATION SPREAD ON TWITTER? A COMPUTATIONAL ANALYSIS OF MATTEO SALVINI'S TWITTER RHETORIC (2013-2019)

Candidate number: 1040507

Word count: 29,988

Thesis submitted in partial fulfilment of the requirements for the degree of M.Phil. in Politics (Comparative Government) in the Department of Politics and International Relations at the University of Oxford

TRINITY TERM 2021

ACKNOWLEDGEMENTS

I would like to thank my supervisor, Dr Spyros Kosmidis, for his help, guidance, and mentorship during the writing of my MPhil Thesis, as well as his overall support during my time at the University of Oxford. I wish to also thank my College Advisor, Dr Kevin Grecksch, for his assistance and availability across the two challenging years affected by the COVID-19 pandemic. Finally, I am deeply indebted to my family and friends for their unconditional love and support. This thesis is fully dedicated to my uncle Amato.

ABSTRACT

Recent studies show that polarised (or, negative emotionally-charged) social media posts lead to the public's negative attitudes towards democratic institutions and politicians. However, thus far, scholars are yet to definitively grasp how and under what conditions such polarising rhetoric spreads (i.e., it is replicated or augmented) between politicians and the public. This thesis contributes to this gap in the literature by investigating Matteo Salvini's rhetoric on Twitter, as well as his respondents' replies. In particular, this research provides a computational analysis of the levels of polarised language in both Salvini's tweets and replies to his tweets, between 2013 and 2019. With a combination of cutting-edge web-scraping methods, machine learning and time-series analyses, I study over 10 million Twitter datapoints. First, I find that polarised language spreads bi-directionally from Salvini's tweets to his respondents, and vice-versa. Second, Salvini's followers' replies not only act as an initial trigger to activate polarisation, but also sustain the use of polarised language in both Salvini and respondents across-time. Third, political content matters. Political replies are more likely to be polarised than non-political replies. Moreover, political replies trigger greater polarisation in tweets in the short-term. Instead, political tweets are more likely to display lower levels of polarisation than non-political tweets, and do not have triggering effects on replies' polarisation in the short-term. These results may shed light on how polarisation affects online behaviour in both politicians and the public. Ultimately, this thesis provides a diverse theoretical and methodological toolkit to further investigate polarisation on Twitter and beyond.

Keywords: Comparative Politics, Computational Social Science, Polarisation.

CONTENTS

ABSTRACT iii

LIST OF FIGURES vi

LIST OF TABLES vii

1 INTRODUCTION 1
1.1 The Puzzle and Research Question 3
1.2 Theoretical Contributions and Innovations 6
1.3 Chapters outline 7

2 LITERATURE REVIEW 8
2.1 Concepts and Scope conditions 8
2.2 The state of the art 12
2.2.1 Affective Polarisation (AP) 12
2.2.2 Affective Polarisation (AP)'s bi-directionality 15
2.2.3 Political Homophily (PH) 17
2.2.4 Political Content 21
2.3 Contributions 23

3 THEORY 26
3.1 Affective Polarisation (AP)'s bi-directionality 29
3.2 Affective polarisation (AP) and Political Homophily (PA) 31
3.3 AP, PH and Discussion topics 33
3.4 Triggering factors 36

4 METHODOLOGY 43
4.1 Case selection 43
4.2 Data-Collection and Coding 45
4.2.1 Data-collection 45
4.2.2 Data-coding 49
4.3 Data analysis 62
4.3.1 Variables 63
4.3.2 Estimation strategy 64

5 TIME-SERIES DIAGNOSTICS 73
5.1 Polarisation in Salvini's tweets 73
5.2 Polarisation in replies to Salvini's tweets 75

6 RESULTS . . . 78
6.1 Descriptive Statistics . . . 78
6.2 OLS models and Coefficient Plots results . . . 80
6.2.1 Tweets' polarisation . . . 81
6.2.2 Replies' polarisation . . . 86
6.3 VAR results . . . 91

7 DISCUSSIONS . . . 100
7.1 Results discussions . . . 100
7.1.1 Bi-directional relationship . . . 100
7.1.2 Followers' replies . . . 101
7.1.3 Political content . . . 102
7.1.4 Interaction between political content and followers . . . 103
7.1.5 OLS robustness checks . . . 105
7.1.6 Triggering factors . . . 106
7.1.7 Summary of the results . . . 109
7.2 Evaluation . . . 110
7.2.1 The Covid-19 impact . . . 110
7.2.2 Internal and external validity . . . 111

8 CONCLUSIONS . . . 113

9 BIBLIOGRAPHY . . . 116

10 APPENDICES . . . 135
10.1 Appendix A: Users' rights and privacy . . . 135
10.2 Appendix B: Web-scraping details . . . 136
10.3 Appendix C: Summary of hashtags details . . . 137
10.4 Appendix D: Computational details . . . 139
10.5 Appendix E: Elastic Net . . . 140
10.6 Appendix F: Tweets' polarisation time-series - stationarity tests . . . 142
10.7 Appendix G: Replies' polarisation time-series - stationarity tests . . . 143

LIST OF FIGURES

3.1 Summary of the theoretical framework . . . 27
3.2 Polarisation bi-directionality among Salvini's tweets and replies to his tweets . . 29
3.3 Relationship between tweets and replies contingently on homophilic groups . . . 31
3.4 Relationship between tweets and replies' polarisation, contingently on homophilic groups and discussion topics . . . 33
3.5 Relationship between tweets and replies' polarisation, contingently on discussion topics . . . 34
3.6 Relationship between tweets and replies' polarisation, contingent on homophilic groups and discussion topics . . . 35

4.1 An adapted version of Boehmke and Greenwell (2020)'s NN . . . 52
4.2 A random sample of replies' emotional predictions by the NN . . . 53
4.3 A random sample of replies' emotional probabilities by the NN . . . 54
4.4 Summary of the Elastic Net output . . . 58
4.5 Summary of the Elastic Net's variables' importance plot . . . 60

5.1 Time-series representing Tweets' Polarisation (2013-2019) . . . 74
5.2 Differentiated Time-series representing Tweets' Polarisation (2013-2019) . . . 75
5.3 Autocorrelation tests for Salvini's tweets' polarisation time-series . . . 75
5.4 Time-series representing Replies' Polarisation (2013-2019) . . . 76
5.5 Differentiated Time-series representing Replies' Polarisation (2013-2019) . . . 77
5.6 Autocorrelation tests for respondents' polarisation time-series . . . 77

6.1 Combined Time-series representing polarisation in Salvini's tweets and replies to his tweets . . . 80
6.2 Levels of Polarisation in Salvini's Tweets (2013-2019) . . . 81
6.3 Coefficient plot representing polarisation in Salvini's tweets, depending on sub-topics and follower repliers . . . 84
6.4 Coefficient plot representing polarisation in Salvini's tweets, depending on sub-topics and non-follower repliers . . . 85
6.5 Levels of polarisation in replies to Salvini's Tweets (2013-2019) . . . 86
6.6 Coefficient plot representing polarisation in replies to Salvini's tweets, depending on sub-topics and follower repliers. . . . 89
6.7 Coefficient plot representing polarisation in replies to Salvini's tweets, depending on sub-topics and non-follower repliers. . . . 90
6.8 Summary of cumulative IRFs output . . . 92
6.9 FEVD for each variable in the VAR model . . . 99

10.1 Stationarity tests for Salvini's tweets' polarisation time-series . . . 142
10.2 Stationarity tests for differentiated Salvini's tweets' polarisation time-series . . . 142
10.3 Stationarity tests for replies' polarisation time-series . . . 143
10.4 Stationarity tests for differentiated replies' polarisation time-series . . . 143

LIST OF TABLES

3.1 Summary 1/3 of hypotheses and sub-hypotheses 40
3.2 Summary 2/3 of hypotheses and sub-hypotheses 41
3.3 Summary 3/3 of hypotheses and sub-hypotheses 42

4.1 Summary of the collected information for each tweet and reply's ID 46
4.2 List of sampled hashtags by content type (political vs non-political) 48
4.3 Summary of the variables under analysis 63
4.4 Summary of the variables' abbreviations 71

6.1 Summary of the descriptive statistics 79

10.1 Summary 1/2 of hashtags details 137
10.2 Summary 2/2 of hashtags details 138

CHAPTER 1

INTRODUCTION

On 24 February 2017, two Roma women were bullied by a group of Italian workers, who locked them up in a cage, and shared the video of the scene online (Gasperetti, 2017; Repubblica, 2017). The Romas were accused by the workers of allegedly "foraging" the storage box of a supermarket (Gasperetti, 2017). On the same day, on Twitter, Matteo Salvini, leader of the Italian right-wing party Lega Nord, re-tweeted the video and, referring to one of the women in the cage, he tweeted,

> *"I back the workers (to whom I offer legal assistance), I don't back the "FORAGER" Romas. How miserable is this woman screaming???" (Salvini, 2017a).*

The tweet received over 2,000 interactions (i.e., retweets, favourites, and comments). C.M. (i.e., an anonymised random replier), a Salvini follower, commented on the tweet, *"Salvini, marry me!"*. R.I., a non-follower, commented, *"Salvini, you and your followers are pieces of s**t!"*. On 13 November 2017, the Italian national football team lost its World Cup qualification match against Sweden. On that day, Salvini tweeted,

> *"There are way too many foreigners, in football youth teams and in Serie A, and this is the result. #STOPTHEINVASION and give more space to Italian guys, not only on the football pitch. #italiasvezia" (Salvini, 2017b).*

This tweet received over 4,000 interactions. M.T., a Salvini follower, replied, *"lesbians, Anna Frank and even these s**ts here. We're not going to the World Cup, but at least we have some justice"*. M.B., a non-follower, replied, *"I feel ashamed for you, do**y man"*. On 30 March 2019, Salvini spoke at the World Congress of Families XIII, an annual international meeting of conservative associations dedicated to discussing Christian values. Later on, referring to a group of people, who contested his speech, he tweeted,

> *"To those feminists who are against [the concept of] family, I suggest to worry more about Islamic extremism, which wants women to be subjected, humiliated, and even beaten up" (Salvini, 2019).*

The tweet received over 3,500 interactions. P.E., a Salvini follower, replied, *"The Pope is attracted by gays and Muslims. He should be banned"*. G.F., a non-follower, commented, *"I am so glad the Pope has refused to meet you, shame on you!"*.

What do these Salvini's tweets and replies have to offer to political scientists? What kind of dynamics can we learn from them with regard to politics? Tweets and replies provide political scientists with large amounts of data to study online political discussions. Specifically, from them, one can learn more about how polarised language (or, negative emotionally-charged) may spread (i.e., being replicated or augmented) from politicians to the public, and/or vice-versa. Moreover, one can understand more about the effects of a user being a politician-follower on how such a user replies to a politician's tweet. With the expansion of social media, like Twitter, many scholars are interested in the propagation of online polarising language, and its wider effects on the quality of online democracy (Barberà et al., 2015; 2019a; 2019b; Bobba, 2019; Georgiadou et al.; Kosmidis et al., 2019).

Affective polarisation is defined as "negative animosity between parties" (Hobolt et al., 2020; Iyengar et al., 2012; 2015; 2019: 129). Such animosity refers to emotional negatively-charged language against opponents (Iyengar et al., 2012; 2015; 2019). Negatively-emotional language attachments to political messages are crucial in shaping online behaviour (Kosmidis et al., 2019; Hobolt et al., 2020). Recent scholarship on politicians' negative language on social media shows that such rhetoric has negative impacts on the public's attitudes towards democratic institutions and politicians (Kosmidis and Theocharis, 2020; Searles et al., 2018). Exposure to negative language has affected reciprocally both politicians and the

public (Abramowitz and Webster, 2016; Barberà et al., 2016; Kosmidis and Theocharis, 2020). Once experienced, users (i.e., either politicians, or the public) are likely to replicate, and/or augment, the negativity in their successive online posts. However, thus far, scholars are yet to definitively grasp whether and how such polarising rhetoric spreads from politicians to the public, and/or vice-versa.

1.1 The Puzzle and Research Question

The literature regarding the effects of polarising rhetoric on social media provides contrasting findings, which produce a theoretical and methodological puzzle. Affective polarisation theory argues that negatively-charged rhetoric affects the relationship between two opposing groups, distancing them (Iyengar et al., 2012; 2014; Inglehart and Norris, 2016). In-group members display positive emotionally-charged language when referring to allies, and exhibit hostile emotionally-charged language when referring to opponents (Iyengar et al., 2012; 2014; 2015). However, polarisation dynamics (e.g., who and how polarisation spreads) and the conditions under which it operates (e.g., users' group-belonging, discussion topics) remain elusive.

For some, polarisation spreads from politicians to the public (Druckman et al., 2019; Wells et al., 2016; Wu, 2017); for others, it is the opposite (Barberà et al., 2019a; 2019b; Theocharis et al., 2020). Other scholars focus on the conditions under which polarisation occurs. For some, it operates through political homophily, according to which political discussions on social media occur via homophilic groups, characterised by shared ideological homogeneity (Barberà et al., 2015; Bright, 2016; Garrett, 2009). Others believe that group-belonging does not affect polarisation levels as polarisation flows between users across-groups (Barberà et al., 2015; de Zúñiga and Valenzuela, 2011; Duggan and Smith, 2016). Finally, polarisation may be endogenous to conversation topics (e.g., political or non-political) beyond group-

belonging (Barberà, 2015; Adamic and Glance, 2005; Livne et al., 2011). Yet, others find that group-belonging generates ideological segregation, which is considered to be impossible to overcome and produce cross-cutting interactions (Sunstein, 2007; Conover et al., 2011; 2012; Muddiman and Stroud, 2017; Hong and Kim, 2016; Bright, 2018). These debates create a theoretical and methodological puzzle regarding the dynamics, and the conditions under which polarisation flows across users. Starting from these debates, this thesis sets out to better understand the microfoundations (i.e., dynamics, direction, and underlying conditions) of the causal link between politicians and the public's polarised rhetoric. These theoretical debates lead this dissertation to answer the following research question:

> *How and under what conditions does polarised language spread between politicians and the public?*

To answer this question, this thesis focuses on Italian politician Matteo Salvini, who is considered to be one of the most influential politicians in contemporary Italian politics (Albertazzi et al., 2018; Horowitz, 2019a; Mammone, 2019). Salvini's case provides a significant opportunity to solve this theoretical puzzle regarding polarisation for three main reasons. First, focusing on Salvini's campaign allows me to study the years in which social media has become a fundamental instrument for political campaigning in Italy (Bobba, 2019; Bobba and Quaranta, 2019; Paparo, 2018). Second, Salvini's vast number of tweets and replies to his tweets provide a consistent and large data-set to conduct cross-time research on polarisation, more than any other major Italian politician (Mazzoleni and Bracciale, 2018; Bracciale and Martella, 2017; Freschi, 2015). Third, more broadly, the Italian party system is characterised by a competitive proportional system (i.e., the difference in major parties' vote share is minimal), in which there is little policy variation across the political spectrum and, thus, the role of divisive rhetoric is accentuated to win consensus (Norris and Inglehart, 2016; Cinelli et al., 2019; Mazzoleni and Bracciale, 2019; Bobba and Roncarolo, 2018). In

other words, Italian politicians often emphasise divisive rhetoric to maintain a distinct brand and to differentiate themselves from their opponents. These three reasons will be further discussed.

Several researchers have highlighted the significance of Salvini's polarising rhetoric on social media in augmenting polarisation among the Italian public (Franzi and Madron, 2019; Horowitz, 2019b; Paparo, 2018; Martini, 2018; Sherer, 2018; Emanuele, 2018a; 2018b). However, the microfoundations of the association between his rhetoric and online behaviour remain elusive. Relatedly, previous scholarship has fallen short from understanding the conditions under which polarisation flows from Salvini to his followers and non-followers, and what discussion topics lead to polarisation (Paparo, 2018; Emanuele, 2018a). To my knowledge, previous studies have not utilised advanced data-gathering, coding, and analytical tools, which I utilise to study Salvini's rhetoric. Such tools may furnish empirical evidence to corroborate the theoretical framework that I develop. Twitter and its posting, replying and following features offer a unique environment to unveil these empirical challenges. To do so, this paper produces a computational analysis to study the use of polarised language between Salvini and his respondents. Using a combination of web-scraping (i.e., Internet data-gathering automation algorithms), machine learning (i.e., computer-aided textual classification algorithms), and time-series analyses, I study the levels of polarisation in Salvini's tweets and replies to his tweets, in turn as my dependent variables (DVs). As my independent variables, I select, respectively, tweets' polarisation (when replies' polarisation is my DV) or replies' polarisation (when tweets' polarisation is my DV), tweets and replies' content (political vs non-political), and whether replies were posted by Salvini's followers vs non-followers. The timespan of this investigation is from 15 December 2013, when Salvini became the leader of Lega Nord, to 9 August 2019, when the government he was a member of broke-down.

1.2 Theoretical Contributions and Innovations

The theoretical contribution of this thesis is five-fold in a complementary structure. First, I argue that polarised language spreads bi-directionally between Salvini's tweets and replies to his tweets. Second, within this bi-directional relationship, I argue that political content in both tweets and replies is a strong predictor of the use of polarised language. Third, I argue that Salvini's followers' replies are also positively associated with the levels of polarisation in successive tweets and replies. Fourth, after theorising this framework, I study the initial triggering factors that activated the use of polarisation. In particular, I argue that repliers have a great impact on initiating the spread of polarisation, by triggering polarised language in successive Salvini's tweets. Finally, follower repliers also trigger an initial polarised response in Salvini's tweets. These contributions are combined into a single theory of affective polarisation, its dynamics and the conditions under which it spreads. Contextually, such a theory is tested through theory-driven hypotheses to grasp how Salvini's tweets influence his respondents' polarisation levels, and/or vice-versa.

At the methodological level, this thesis sets out to contribute to the literature in various ways. First, using web-scraping tools, I develop a novel data-set on Salvini's tweets and replies to his tweets, yielding over 10 million total observations. Such data will aid me to understand more about each tweet and reply (i.e., time-stamp, text, author's being a Salvini follower or non-follower, discussion topic, etc.). Second, this thesis will code Salvini's tweets and replies to his tweets through a neural network textual classification to study the levels of polarisation in both tweets and replies. The neural network (an advanced multi-layered machine learning methods) will code each tweet and reply by emotional charges (e.g., Anger, Fear, Joy, etc.). Finally, this thesis utilises several econometric methods to study the relationship between tweets and replies' polarisation. Specifically, it conducts ordinary least squares (OLS), coefficient plots, and vector autoregression (VAR) analyses to study whether and how polarisation spreads between tweets and replies across-time. Through

these methodologies, I ultimately generate a novel data-set of Salvini's tweets and replies, which will be utilised to study theory-driven hypothesis, and can potentially be utilised by other scholars for further research purposes.

Studying affective polarisation via social media platforms is inevitably relevant to the contemporary political landscape, with various government and private sector reports seeking to better understand the evolution of online democratic processes. Noteworthy to mention are the ongoing UK and Germany's policy-making processes to tackle online polarisation and improve online political discussions (UK.gov, 2019; Die Bundesregierung, 2019). Italy, which is the country under analysis, may also benefit from such an investigation. In the private sector, my findings may support Facebook (2020)'s 'Misinformation and Polarisation' research projects on how social media companies can identify and address online polarisation.

1.3 Chapters outline

This thesis proceeds as follows. First, the literature review discusses the key concepts and scope conditions, as well as surveying the literature regarding affective polarisation, its dynamics and conditions under which it operates. Second, the theory chapter discusses an innovative theoretical framework to study affective polarisation between politicians and the public. Moreover, this chapter presents theory-driven hypotheses for the remaining chapters. Third, the methodology chapter explores data-collection, coding, analyses, as well as validation checks. Fourth, the time-series chapter conducts two main diagnostics (i.e., stationarity and autocorrelation) on the time-series under analysis. Fifth, the results chapter reports analyses from OLS and time-series models (VAR). Sixth, the discussion chapter discusses the results in the light of the theory-driven hypothesis, answering the research question, and exploring COVID-19 limitations and validity checks. Finally, the conclusion reiterates the main findings of this thesis, as well as highlighting avenues for future research.

CHAPTER 2

LITERATURE REVIEW

This chapter consists of three sections. First, it defines the key concepts and scope conditions of this thesis. Second, it provides a review of the state of the art regarding affective polarisation, and its online dynamics across group-belonging and discussion topics. Third, it discusses the main theoretical and methodological contributions of this thesis.

2.1 Concepts and Scope conditions

The definition and nature of polarisation are subject to ongoing theoretical and empirical debates to which I wish to make a theoretical and empirical contribution. Affective polarisation (AP) is a branch of the Comparative Politics literature that focuses on explaining political polarisation (Garrett and Bankert, 2018; Mason, 2013; 2015; Lelkes and Westwood, 2017). It studies the rising of the levels of animosity towards opposing groups, and favouritism towards coalition groups (Reiljan, 2016; Iyengar et al., 2012; 2019; Lelkes and Westwood, 2017; Mason, 2013; 2015). As such, the spread of polarisation can be defined as the ways in which negative language is replicated or augmented from a group (or, an individual) to another. AP stems from political psychology alongside social identity theory, according to which partisanship is the most salient social and political identity (Luttig, 2018; Mason, 2015; 2016; Iyengar, et al., 2012). Relatedly, AP argues that partisanship can be seen as a group identity, which in turn generates in-group bias, and discrimination against opposing out-groups (Iyengar et al., 2012; Miller and Conover, 2015; Iyengar and Westwood, 2015). While AP seeks to study emotional distance between partisans, it is the question of how and under what conditions this occurs that motivates identification of the dynamics of emotional group-based polarisation. On the emotional side, some scholars argue that "negative animosity between parties" fuels polarisation (Iyengar et al., 2012; 2015; 2019: 129). Such

animosity refers to emotional negatively-charged language against opponents (Iyengar et al., 2012; 2015; 2019).

Examples of emotional negatively-charged language are those sentences that display high levels of negative emotions, such as anger and disgust (Kosmidis and Theocharis, 2020). On the contrary, positive language is associated with more positive emotions, such as joy and trust (ibidem). Some researchers have considered negatively-charged language as "milder" in the cases of name-calling, mockery, or insulting others (Anderson et al., 2013; Borah, 2012; Massaro and Stryker, 2012; Kosmidis and Theocharis, 2020). Others have studied "heavier" negative language in the instances of homophobia, racism and sexism (Theocharis et al., 2016; Kosmidis and Theocharis, 2020). Therefore, polarised language is commonly perceived as a continuum in which positive language is placed at the end, while mild cases are in the centre, and heavier cases are at the extreme (Sydnor, 2019; Kosmidis and Theocharis, 2020). Contextually, Salvini's negative language has been the main driver of his success on social media (Bobba, 2019 Di Cicco and Sensales, 2019; Molinaro et al., 2018). In particular, Di Cicco and Sensales (2019) find that Salvini's angry posts result in a greater number of interactions (i.e., likes, comments, shares) by his followers, as well as greater use of strong polarised language in his followers' comments. Moreover, when focusing on positive emotions, polarisation due to positively-charged language does not succeed to fully explain why Salvini's followers may display negative language when exposed to Salvini's content (Mazzoleni and Bracciale, 2018). These findings may provide evidence for claiming that polarised negative rhetoric is a strong predictor of online mobilisation, vis-à-vis positive language.

Other researchers favour individual-level studies regarding issue positioning and AP, finding limited theoretical and empirical results. Some literature focuses on issue positioning polarisation, studying individual citizens' polarisation across ideological differences based on political and policy stands (Iyengar et al., 2012; 2019). On the other hand, AP theory is

more interested in the emotional charge of language utilised to express opinions against opponents (ibidem). I argue that these two roots of polarisation are theoretically and empirically distinct. Theoretically, scholars have shown that ideological positioning is not a sufficient condition to experience AP (Iyengar et al., 2012; 2019; Mason, 2015). Specifically, AP may be due to individual-level characteristics (Hobolt and de Vries, 2016). Such characteristics may include, above others, individuals' economic preferences (e.g., redistributive versus non-redistributive), socio-economic background (e.g., education, income), and identity (e.g., gender or age). However, evidence has shown that there is little or no correlation between issue positioning and AP (Iyengar et al., 2019; Levendusky and Malhotra 2016a; Kosmidis et al., 2019). That is, when users differ on issue positioning, AP may remain invariant, and vice-versa. Moreover, it would be theoretically unfeasible to causally infer that users' preferences and identity shape online behaviour due to the high number of theories and variables to consider. Instead, this thesis seeks to contribute to the theoretical debate regarding the variation in emotionally charged language variation among politicians and general users on Twitter. Empirically, due to Twitter's terms and conditions, researchers are not allowed to match any single user to individual-level data (Twitter, n.d.). As such, I do not have access to precise users' location, real name(s) and identity data. The lack of such data produces an ecological inference fallacy as it is empirically unfeasible to infer that aggregate-level preferences have an impact on individuals' online behaviour. Thus, also at the empirical level this thesis is not interested in making such a link between individual-level offline preferences and characteristics, and online Twitter behaviour. Doing so would be theoretically and methodologically unfeasible due to the different nature of AP (i.e., emotions) and issue positioning (i.e., individual stances on policy), and online privacy rights. For these reasons, I restrict the scope conditions of this thesis to study the spread (i.e., how negative language is replicated or augmented from an individual to another) of polarised language between politicians and the public on Twitter.

Twitter provides a strong research ground to study political communication as it offers the unique advantage of providing data from both politicians and the public (Barberà et al., 2015, 2017; 2019; Kosmidis et al., 2019; Kosmidis and Theocharis, 2020; Sydnor, 2019; Theocharis et al., 2016; 2020). Unless users set the privacy mode to restrict other users to access their Twitter data and activities (e.g., tweets, replies, favourites, retweets, etc.), posts are widely available to scholars. According to Twitter data, only a very small minority of users adopt such a private mode (Hughes and Wojcik, 2019). Moreover, posts are visible to all users by default. Therefore, users can easily interact with tweets through liking (i.e., marking posts as favourites), retweeting (i.e., re-share the tweets), or replying (i.e., commenting on the tweets). The system of hashtags, mentions, and trending topics function (i.e., a list of top-tweeted topics) can create public topic-specific discussions (Colleoni et al., 2014). In this way, Twitter overcomes offline time and space limits to access discussions as any user can interact with others regardless of when and where they tweet from (ibidem).

Others have conducted research on polarisation on Facebook given the large number of politicians and public users on this social network (Bobba, 2019; Mazzoleni and Bracciale, 2019; Sensales and Di Cicco, 2019). However, on Facebook, the public has less exposure to a politician's messages unless the public actively follows that politician. Instead, on Twitter, the public is more easily exposed to any politicians' discussion through the use of hashtags and trending topics functions. Moreover, on Twitter, access to certain activities (i.e., messaging) is not granted through "friendship requests" (i.e., asking other users to grant access to their profiles) as it happens on Facebook. Finally, contextually in Italy, conversational strategies are stronger on Twitter than on Facebook because the discussion features (e.g., replying, tweeting, hashtags) are better understood and utilised by the Italian electorate (Bracciale and Cepernich, 2018). Therefore, given the sharing tools and large availability of public data, Twitter provides a strong research ground to study the use of AP language.

In this review, I analyse three major competing, yet connected, dynamics: politicians and users' polarisation bi-directionality, group-belonging, and discussion content. Bi-directionality of polarisation refers to the bi-directional spread of polarisation from politicians to users, and vice-versa (Theocharis et al., 2016; 2020; Tromble, 2018). Group-belonging stresses the significance of online homophilic groups in determining polarisation across users (Agathangelou et al., 2017; Hong and Kim, 2016; Bright, 2018). Discussion content, instead, refers to politicians' message content, whether it is political or non-political (Barberà et al., 2015; Barberà, 2015; Theocharis et al., 2017; 2020). Therefore, this thesis is interested in understanding the dynamics of polarisation on Twitter across these competing arguments. More broadly, my research seeks to determine the conditions under which polarised language spread from politicians to users and vice-versa, such as bi-directionality of polarisation, group-belonging, and discussion topics.

2.2 The state of the art

2.2.1 Affective Polarisation (AP)

Before analysing how and under what conditions AP language spreads, this section discusses the significance of negative-emotional charge. As previously mentioned, AP theory stresses that polarised (or, highly negative) rhetoric exacerbates the relationship between two opposing groups (Iyengar et al., 2012; 2015; 2019: 129; Hobolt et al., 2020). Such a division may not stem from issue positions nor founded in policy, it may rather be the result of angry feelings towards opponents (Webster and Abramowitz, 2017; Iyengar et al., 2012; Mason, 2013; Curini and Martelli, 2015). For example, in some political contests, as AP may increase over-time, ideological divisions may shrink (Iyengar et al., 2019; Levendusky and Malhotra 2016a). Evidence from the UK and USA highlights that, when parties become less distinguishable in their policy proposals, leaders exacerbate parties' division, utilising

strong emotive language to differentiate them from their opponents (Kosmidis et al., 2019). AP language is likely to persuade undecided voters: experimental studies in the USA, UK and Colombia demonstrate that, when emotional language is utilised to derogate political opponents, it is effective in increasing exposure and popularity of the message (Suhay et al., 2018; Bail et al., 2018; Gallego et al., 2019; Munger et al., 2017). This can lead to opinion and attitudinal variation in voters who are exposed to such content, shaping groups' sense of hostility directed at their opponents (Munger, 2017; Levendusky, 2009; Iyengar and Westwood, 2014; Iyengar et al., 2012).

Others (such as Hobolt et al., 2020) have studied the impact of positively-charged language towards in-group members as a form of polarisation. Hobolt et al. (2020) find that, during the Brexit referendum campaign, positive language towards fellow remainers/leavers fortified the two ideological positions (i.e., leave vs remain). However, other research shows that what triggers polarisation in the first instance is negative language towards opponents (Bobba, 2019; Del Vigo and Ferrara, 2015; Raniolo, 2016; Prospero, 2015; Bresolin, 2014). In this vein, previous work on Salvini's rhetoric has highlighted that attacks made by Salvini managed to trigger polarised responses by both his followers and non-followers (Mazzoleni and Bracciale, 2018). Unlike in Hobolt et al. (2019)'s case, Salvini's followers (not only his non-followers) displayed negative language towards Salvini's posts. Furthermore, Salvini's use of anger has succeeded in mobilising a part of the Italian electorate who would otherwise be disconnected to politics (Cedroni, 2014; Reddy, 2001; Molinaro et al., 2018). For these reasons, what becomes important to understand is Salvini's use of negative language, as well as the negative language utilised by his respondents.

Whilst AP has sprouted in recent US politics, some scholars contend the generalisability of such polarisation outside the US. Several scholars study how the use of AP is correlated with a higher level of polarised interactions across US politicians and users (Barberà et al., 2016; Lee and Lim, 2016; Munger, 2017; Abramowitz and Webster, 2016; 2017). Often

online comments replicate the levels of polarisation, incivility and aggressive language of the original tweets (ibidem; Anderson and Huntington, 2017; Papacharissi, 2004; Muddiman and Stroud, 2017). However, the literature fails to understand whether and which types of discussion dynamics are related to greater political polarisation. In particular, it falls short from identifying whether it is due to the space of action (e.g., party system, partisanship), or actors' characteristics (e.g., leaders' charisma) (Barberà et al., 2016; Munger, 2017; Munger et al., 2017). Therefore, it becomes vital to study the effects of AP in new spaces of action, such as Italy with its highly polarised proportional system (see Cinelli et al., 2019; Norris and Inglehart, 2016), and with new actors, such as Matteo Salvini with his highly effective social media strategies (see Mazzoleni and Bracciale, 2019; Bobba and Roncarolo, 2018).

Contextual to Salvini, some scholarship focuses on cross-party research to study the spread of polarised language, however the within-case variation is yet to be understood. Mazzoleni and Bracciale (2019) study communicative strategies by Italian major leaders on Facebook. The study analyses Facebook posts and interactions by Silvio Berlusconi, Luigi Di Maio, Matteo Salvini, Giorgia Meloni and Matteo Renzi. Mazzoleni and Bracciale (2019), via linear regression models, find that Salvini is the political leader with highest rates of attacks (46%), and use of polarised language (53%, calculated as the use of negative language towards political opponents) among the sampled politicians. However, neither does the study explain the effects of such communication strategy on respondents, nor the extent to which respondents play a role in the discussion polarisation. Moreover, such research does not account for the importance of different sub-groups of repliers, different discussion topics, and triggering factors beyond solely focusing on overall effects. To overcome these limitations, a within-case analysis may aid to provide further evidence into Salvini's rhetoric, and ultimately understand how the propagation of polarised language occurs. For these reasons, I now turn to three fundamental characteristics to study the dynamics of AP: bi-directionality, homophilic groups, and discussion topics.

2.2.2 Affective Polarisation (AP)'s bi-directionality

Scholars have extensively studied the use of social media as a unidirectional channel for politicians to express their ideas to the public. However, the dynamic nature of the interactions among politicians and citizens on Twitter opens the door to novel research (Barberà et al., 2019; McLoughlin and Ward, 2020; Theocharis et al., 2020). Early supply-side theories explained that politicians prefer not to engage with online discussions due to time constraints, low levels of technological skills, or fear of losing control of the message (Stromer-Galley, 2000; Ward and Lusoli, 2005). Given the anonymity to express their ideas online, citizens often harass politicians, preventing them from engaging with the public (ibidem). However, more recent evidence shows that politicians are becoming more digitalised following Obama's 2008 and 2012 campaigns, and engage more in online discussions (Barberà et al., 2019; McLoughlin and Ward, 2020; Theocharis et al., 2020; Tromble, 2018). As such, on the supply-side, research shows that political elites widely utilise social media websites, like Twitter, to promote their personal brand, mobilise and ultimately create political support for the easy and direct access to users (Kosmidis and Theocharis, 2020; Theocharis et al., 2020). On the demand-side, public users have increased the use of Twitter for news consumption and interaction with politicians, who used to be hard-to-reach individuals (Kosmidis and Theocharis, 2020; Newman et al. 2017).

Early empirical research on online bi-directional discussions shows that politicians avoid discussing with users to avoid the risk of negative reciprocity (Tromble, 2018). Specifically, Tromble (2018) studies the role of Twitter users' tweets to and from members of the legislature in the Netherlands, the UK, and the US, in 2013. With the aid of Big-data retrieval methods and manual tweets' coding for positive, neutral or negative content, Tromble (2018) conducts regression analysis. Results show that users' demand expressed through tweets is crucial in determining whether and how politicians respond to the public. These findings become significant for the scope of this investigation, as they highlight once more the im-

portance of studying the relationship between politicians and users, and the propagation of negative language. However, as Tromble (2018) only focuses on whether the public's tweets refer to politicians, further research is needed to analyse what type of conversations occur between politicians and the public. Moreover, the paper fails to understand how the interaction between politicians and respondents may affect subsequent online behaviour. For these reasons, it is crucial to investigate both ends of online discussions, and to understand the factors that trigger polarised language.

A significant piece of research comes from Theocharis et al. (2020) who produce robust evidence on the bi-directional dynamics of emotionally charged language on Twitter. They study longitudinal data from Twitter posts from the public, which mention US Members of Congress between November 2016 and May 2017. Such research includes a yearly timespan, which allows the scholars to study electoral and non-electoral periods. Empirically, Theocharis et al. (2020) investigate the propagation of uncivil language on Twitter through machine learning textual classification, logit models for topic-prediction, and advanced Gini redistribution coefficient for negative language. The scholars find that, when tweets involve politicians, the levels of incivility are high and consistent across-time (ibidem). Furthermore, evidence shows that incivility may not be contingent on policies. Rather it is linked to broad political content (e.g., political events, such as inaugurations, scandals and others). This may highlight that one source of polarisation may stem from public users instead of politicians. Such findings are crucial in this thesis as I am interested in understanding whether polarisation has a bi-directional dynamic between politicians and the public. Moreover, one has to test whether negative language labelled as polarised may lead to similar results to uncivil language as in Theocharis et al. (2020). As such, more evidence is required to support these assumptions. Finally, further research is required to study triggering factors among politicians and respondents. Research should understand who initiates the spread of polarised language among the actors under analysis (Scholl et al., 2021). In particular,

the literature falls short of understanding the initial conditions under which either tweets or replies lead to polarised discourses in future tweets. As such, I aim to study short-term impulses that initially activate the use of polarised language.

In sum, it becomes significant to study AP as a bi-directional relationship between politicians and respondents. To contribute to this gap in the literature, I wish to investigate the type of language utilised by both actors (politicians and respondents) to study the quality of Twitter discussions.

2.2.3 Political Homophily (PH)

Studying homophily is significant to understand how group dynamics have an impact on the spread of polarisation. Homophily refers to the tendency of people to form social ties with others with whom they share similar characteristics (Ackland and Shorish, 2017; Lanzarsfeld and Merton, 1954). This principle can operate across areas of life, ranging from people's identity, such as gender and race, but also preferences, such as books and music tastes, to political attitudes (Ackland and Shorish, 2017). Group formation also occurs on the internet and, most prominently, on social media. PH theory argues that political discussions on social media occur in homophilic groups, which are groups characterised by shared ideological homogeneity (Agathangelou et al., 2017; Bright, 2018; Hong and Kim, 2016). On political terms, aggregation takes place according to users' economic and socio-cultural background (Agathangelou et al., 2017; Campbell et al. 1960; Levendusky 2009). On Twitter, users tend to be part of networks in which they resonate in terms of socio-demographic and behavioural traits; for instance, when users decide to follow politicians or other users whose ideological and political positions are similar to theirs (Conover et al., 2011; 2012; Wu et al., 2011). In this sense, following a politician is a signal of ideological and political proximity as users consent to be exposed to a greater amount of content from the politician they follow and

receive less content from opposing users (Barberà et al., 2015; Iyengar and Hahn, 2009; Oken Hodas and Lerman, 2012; Stroud, 2008;).

Twitter provides a strong habitat to study group formation because of the large group size of its users, discussions, and interactions tools (ibidem; Barberà, 2015; Barberà et al., 2015; Colleoni et al., 2014). On Twitter, users can decide to 'follow' other users (Ackland and Shorish, 2017: 28; Colleoni et al., 2014). Twitter's "follow" function is a prominent tool to study group-belonging because it allows to understand relationships among users. Homophily can occur on Twitter through three main streams (Akland and Shorish, 2017). First, reciprocity: once followed, users may reciprocate the friendship; second, transitivity: friends of friends are likely to become friends; third, sociality: when users enjoy having many social ties, they may add many users to their following lists. Under these conditions, users are likely to aggregate into homophilic groups, where an in-group is defined as the group that a person identifies with; an out-group, instead, is the group to which a person is opposed (Iyengar and Westwood, 2014; Iyengar et al., 2012; Levendusky, 2009). Moreover, thanks to numerous posts and responses, it allows researchers to study their within and cross-interactions (Barberà et al., 2015; Mencarini et al., 2019; Cresci et al., 2017; Agathangelou et al., 2017). Belonging to these online groups can lead users to ideological segregation, or cross-cutting interactions (Barberà et al., 2015; Hong and Kim, 2016; Bail et al., 2019; Bright, 2018). Starting from these accounts, one would expect Salvini's followers to resemble a homophilic group, in opposition to Salvini's non-followers group. This research seeks to study whether and how belonging to these homophilic groups affects polarisation across Salvini's tweets and replies to his tweets. I now turn to behaviour contingent on group-belonging: segregation or cross-cutting interactions.

Segregation occurs when homophilic groups favour the association of like-minded users, who share similar ideas and sources of information, reinforcing political discourses and information they already agree with (Bright, 2018; Bessi et al., 2015; Conover et al., 2011; 2012;

Weng et al., 2010; Quattrociocchi et al., 2016). Most prominently, Iyengar et al., (2012) and Garrett et al. (2014) point to the role of fragmented news and consumers in selecting 'friendly source' which provide slanted news. The role of in-group members becomes key in developing strong political identifications and biased interpretations, which generate negative evaluations of the opposing groups (Iyengar, et al., 2012; 2017; Levendusky, 2013; Mason, 2015; Miller and Conover, 2015; Westwood et al., 2018). As such, the exposition to like-minded users may lead to informational cascades, defined as the tendency of users to adopt a single line of facts' interpretation and group's views as their own (Chan et al., 2014; del Vicario et al., 2016; Lenz, 2009; Pariser, 2011). Twitter interactions, such as tweets, comments and retweets, are critical to understanding users' confirmation bias (Bright, 2018; Conover et al., 2011; 2012; Hong and Kim, 2016; Muddiman and Stroud, 2017; Sunstein, 2007). However, the persistence of such segregation is contended, as cross-cutting interactions may occur given the sharing tools (e.g, trending topic tables, discover, search) available on Twitter.

Cross-cutting interactions occur when users engage with people from different political homophilic groups (Agathangelou et al., 2017; Holbert et al., 2010). This is due to the nature of Twitter, which aims to increase the interactions between users regardless of their interests or followers (Barberà et al., 2015; de Zúñiga and Valenzuela, 2011; Duggan and Smith, 2016; Newman et al., 2017). Moreover, the need for more diverse ideas and sources of information leads users to be exposed to other homophilic groups (Knobloch-Westerwick and Kleinman, 2012; Dubois and Blank, 2018). These interactions can cause online debates and more moderate positions on political matters, which ultimately decrease the ideological distance between homophilic groups (Hong and Kim, 2016; Bright, 2018). However, long-lasting cross-cutting interactions are less common as they derive from cross-group interactions, which involve moderate users from both parties at the same time (ibidem). Moreover, some researchers predict that these interactions may only consolidate previous opinions and produce further

polarisation (Conover et al., 2011; 2012; Levendusky, 2013; Gonzalez-Bailon et al., 2010). Thus, further studies are yet to clarify when and how cross-interactions occur.

Empirically, ideological segregation and cross-cutting interactions have been widely studied, yet produced limited results. Conover et al. (2011) study political communication and group formation on Twitter during the 2010 US midterm elections. They analyse over 250,000 tweets from the six weeks prior to the elections through network clustering algorithms and manually-annotated data (ibidem). Conover et al. (2011) find that users are highly segregated on Twitter as there are limited connections across left- and right-lining groups. In particular, opposing users are more likely to retweet content from their peers rather than from their opponents, creating highly distinguishable left and right-wing chambers. However, the endurance of segregation is highly contested, as well as the actors and space in which this occurs (Barberà et al., 2015; Kosmidis et al., 2019; Hong and Kim, 2016). Indeed, empirical results challenge the segregation argument because social media, like Twitter, furnish greater space for political disagreement and large varieties of news sources (Bakshy et al. 2015; Duggan and Smith, 2016; Boxell et al., 2017). Further analyses should also be conducted on the endurance of segregation across-time, studying whether it is a short-term or a long-term status.

Similarly, Weng et al. (2010) explore segregation as the tendency of online friends (i.e., reciprocal followers) to share or post similar tweets both in terms of content and form. They analyse over one million tweets from Singapore between April 2008 and April 2009 through a probabilistic topic model. Weng et al. (2010) find that online friends tend to share similar Twitter content and behaviour. However, such research presents some limitations. First, it fails to understand whether segregation is contingent on some discussion topics, such as politics, and whether non-political discussions may lead to similar outcomes. Second, it neglects cross-time events as it mainly explores single events, such as individual election cycles, neglecting non-electoral (or, non-politically salient) periods. Finally, it does not

explain the endurance of such segregation, which may be a non-permanent users' status.

In sum, the literature does not fully explain how group-belonging affects the spread of polarisation. To contribute to this gap, I study the relationship between Salvini's tweets, and replies to his tweets, contingent on whether the respondent is a Salvini's follower.

2.2.4 Political Content

Evidence from American politics highlights that the topic and context in which tweets are shared is effective in inducing online debates between homophilic groups (Barberà et al., 2015; Adamic and Glance, 2005; Livne et al., 2011). In particular, Barberà et al. (2015) differentiate tweets between political and non-political to understand whether the effect of AP is due to a specific discussion topic. When the degree of polarisation is accounted for, they find that tweets with non-political content received, on average, more retweets (ibidem). As such, the authors (ibidem) find that non-political topics, such as the Olympics or Super Bowl, lead to greater cross-cutting interactions among groups. Thus, a promising way to promote cross-cutting interactions may be to expose people to inadvertent political contexts through non-political discussions (Brundidge, 2010; Tucker et al., 2018). Considering Barberà et al. (2015), their paper only focuses on twelve topics; more evidence may be required to further confirm, or contradict, their findings (ibidem). Moreover, research is needed to understand whether political and non-political topics can be investigated either in isolation or in combination, as non-political content can suffer from occult political propaganda. Finally, evidence is required to understand political triggering factors, which may initiate the use of polarised language among politicians and the public (Barberà et al., 2019a; 2019b). As such, further research should explain which political or non-political topics lead users to display greater polarisation in subsequent activities in the short and long period. For this reason, the thesis seeks to contribute to the understanding of tweets' topic and context regarding

AP and its impact on homophilic group formation.

Contextual to Italy, researchers study Salvini's communication on social media, focusing on his polarised rhetoric, yet with limited conclusions (Bobba, 2019; Paparo, 2018; Mazzoleni and Bracciale, 2018). In particular, Bobba (2019) investigates how Salvini's polarised language aided him to achieve greater popularity (i.e., number of likes, share, and comments) among his audience. Bobba (ibidem) finds that posts containing references to "feelings of anger, disgust, disappointment and revenge" resulted in 17% more likes than non-emotionalised messages (ibidem). Moreover, the source of the message also mattered, when Salvini posted messages on Facebook, these achieved over 40% more likes than average messages posted by his staff (ibidem).

However, while producing quantitative evidence on the effect of Salvini's posts on his audience, Bobba (2019) does not account for the type of interactions Salvini's posts receive. In this sense, evidence is required to understand whether the received comments are either negative or positive towards Salvini. This would allow researchers to have a more granular understanding of whether Salvini's polarised rhetoric flows from him to his audience, or vice-versa. Moreover, Bobba (2019) only considered 453 posts by Lega Nord and 282 posts by Salvini. Thus, further research should be conducted to analyse a greater sample of posts, and it should include reactions beyond likes, such as comments to achieve more analytical depth. Given its large data-set of tweets and replies, Twitter offers the perfect platform to investigate this. Finally, as suggested by Barberà et al. (2019a), it becomes significant to study the triggers of such a relationship between tweets and replies' polarisation. As such, Bobba (2019) only considers the overall effect of Salvini's posts on his respondents, without considering whether his polarisation comes as a result of previous comments by his respondents.

When accounting for both PH and discussion' topic, AP literature finds contrasting results.

Politicians and users that engage in political conversations tend to display higher levels of polarisation than others who engage in non-political conversations (Barberà et al., 2015; Wojcieszak and Mutz, 2009). In particular, Wojcieszak and Mutz (2009) provide research on 1000 American Internet users, who participated in online discussions, studying different discussion topics and political exchanges in 2013. They find that most cross-cutting interactions occurred in conversations in which political topics emerged secondarily, yet politics was not the central purpose of the discourse (ibidem). In other words, conversations that primarily concerned non-political topics (e.g., leisure, sports), and accidentally moved to politics, provided space for polarised discourse. For these reasons, this thesis will study not only tweets that are primarily focused on politics but also those that discuss non-political issues. Furthermore, I seek to study political and non-political replies to those tweets. Doing so would allow me to study how polarised discourse takes place within political conversations (i.e., when both tweets and replies concern political topics). Moreover, I will also study non-primarily political conversations (i.e., when only one between tweets and replies concern political matters). And finally, this thesis seeks to study non-political conversations (i.e., when neither tweets nor replies concern political matters).

In sum, more research is needed to understand how polarisation spreads across different discussion topics, namely, political versus non-political. I intend to fill this gap by studying the discussion among Salvini and his respondents, controlling for the discussion topic.

2.3 Contributions

This review has analysed the prominent literature surrounding AP on Twitter between politicians and the public. While agreeing with the existence and persistence of AP, the logical steps that remain to be studied are its dynamics and the conditions under which it operates. Departing from prominent literature, this review has highlighted theoretical and

methodological gaps to fill.

At the theoretical level, this thesis seeks to study the interactive dynamics and the underlying conditions of AP. First, it will study how AP travels from politicians to users, and vice-versa, through Salvini's tweets and replies to his tweets. This is a significant starting point to understand the portability of the theories of AP beyond the US to other spaces of interactions (such as Italy) and other political actors (such as Salvini). Second, this thesis will study the interaction between political homophily and polarised language as an underlying condition for the spread of polarisation. Specifically, it will theorise over the polarisation response of Salvini's followers versus non-followers. Third, this thesis seeks to contribute to the understanding of how political discussions lead to greater polarised language across both Salvini and his repliers. As such, it will consider how and which political topics have a positive impact on polarisation levels among the actors under analysis. Finally, this thesis considers the interaction between group-belonging and political discussions on the spread of polarisation. Having explored the potential bi-directional polarisation among politicians and the public, one can then study the most significant factors in triggering the spread of polarisation. These theoretical accounts aid to have a greater understanding of each theory individually, and in conjunction. Ultimately, the final theoretical framework seeks to propose a comprehensive understanding of AP theory.

At the methodological level, this thesis aims to provide several empirical contributions in the fields of computational social science to overcome the above-discussed limitations in the literature. First, analysing Salvini's rhetoric can provide a greater understanding of AP beyond the US. Specifically, accounting for a new place of interaction between politicians and the public, determined by little policy variation (such as the Italian party-system), as well as new actors (such as, Salvini and his respondents) furnishes novel empirical evidence to answer our theoretical questions. Second, through the use of web-scraping methods, I will create a large database of discussions through Salvini's tweets and replies to his tweets.

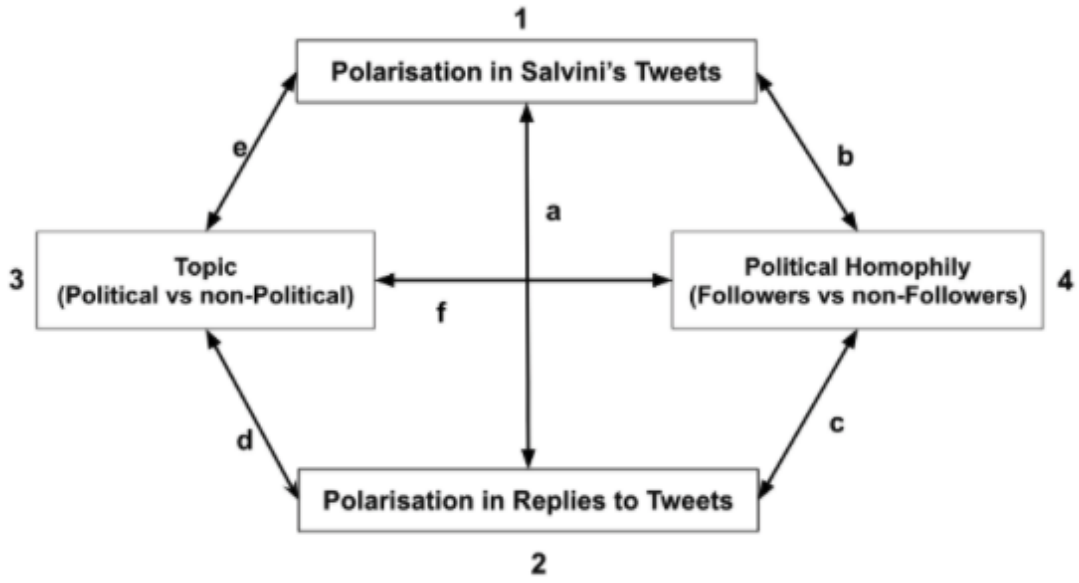
Web-scraping will also be utilised to collect users' group-belonging data to check whether a given respondent follows Salvini, and whether such group-belonging affects polarisation. Third, this paper will utilise advanced machine learning methods to code tweets and replies to measure AP as negative-charged language coefficients for each tweet and reply in the data-set. Fourth, computational methods will also be utilised to create string detection algorithms to study the specific discussion topics, and their effects on the theories under analysis. The final database will be analysed through OLS and VAR models. These methods will allow me not only to grasp the relationship between the variables across-time, but also the initial triggering factors that activated the use of polarisation. The data-gathering, coding and analysis methods will be discussed in full later.

CHAPTER 3

THEORY

This thesis is interested in understanding how and under what conditions polarised language spreads between Matteo Salvini's tweets and replies to his tweets. To solve the theoretical and methodological puzzle highlighted in the literature review, I propose a novel theoretical framework to study polarisation in Twitter discussions. First, I argue that polarised language spreads bi-directionally between Salvini and his respondents. Second, within this bi-directional relationship, I argue that political content in both tweets and replies is a strong predictor of the use of polarised language. Third, I argue that Salvini's followers' replies are also positively associated with the levels of polarisation in successive tweets and replies. Fourth, after theorising this bi-directional framework, I analyse initial triggering factors (i.e., short-term impulses that initially activate polarisation). I argue that repliers have a great impact on initiating the spread of polarisation, by triggering polarised language in successive Salvini's tweets. Finally, follower repliers also trigger an initial polarised response in Salvini's tweets. From this framework, I intend to derive hypotheses for the remaining chapters. Figure 3.1 summarises my theoretical framework.

Figure 3.1: Summary of the theoretical framework



Before engaging in my analytical work, I will first define each component of this framework.

First, at the top and bottom of the figure, there are two blocks (1, 2), such as polarisation in Salvini's tweets and replies. Each of these two blocks represents, respectively, the levels of polarised language in a given Salvini tweet (1), and a corresponding reply associated with the original tweet (2). Five arrows connect the two polarisation blocks. Four arrows (b, c, d, e) intercept discussion topics and PH (i.e., users' group-belonging) before indirectly linking tweets' polarisation and replies' polarisation. And one direct arrow (a) connects tweets to replies, and vice-versa.

Second, block 3 (on the left-hand side of the figure) represents the conversation's topic. From this block, three arrows stem (d, e, f). The first two arrows (d, e) are directly pointed towards polarisation in tweets and replies. The other arrow (f) is linked bi-directionally with PH (i.e., users' group-belonging, 4).

Third, block 4 (on the right-hand side of the figure) represents users' homophilic belonging. Such a block indicates whether a respondent to a given tweet follows Salvini, from this block three arrows stem. The first two arrows (b, c) link group-belonging to tweet's

polarisation, and replies' polarisation. The third arrow (f) links block 1 to the discussion topics.

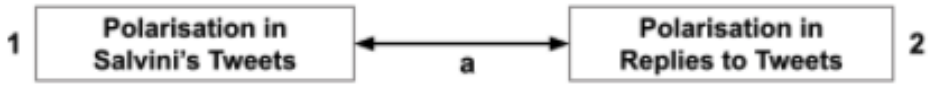
These four blocks (1, 2, 3, 4) are linked in a complementary structure to understand the spread of polarisation on Twitter. Each arrow represents an association, which will be discussed in the following sections to formulate theory-driven hypotheses. These contributions are combined into a single theory, which aims to grasp the effects of polarisation across Salvini's tweets and his respondents. Ultimately, I study who among Salvini and his respondents initiate the spread of polarised language.

Alternative explanations regarding polarisation associate AP and issue positioning, studying users' individual-level socio-economic preferences and identity. There are ongoing debates over the relationship between individual-level preferences and AP (Hobolt and De Vries, 2016). Research about such debates has found little or no correlation between issue positioning and AP (Iyengar et al., 2019; Levendusky and Malhotra 2016a). In particular, it is hard to fully comprehend the causal link between users' individual-level socio-economic preferences and identity, and their online behaviour (Iyengar et al., 2019). Making inferences on those users' characteristics would require a parsimonious selection model for economic indicators, labour market competition, fertility decisions (if accounting for users' household), and many others (Rueda and Stegmueller, 2019). Moreover, as previously discussed, Twitter's terms and conditions preclude researchers from matching Twitter data and individual-level data. As such, it would not be theoretically feasible to generalise users' preferences and online behaviour. Specifically, doing so would introduce a great level of bias due to selection bias issues. According to Pew Research, based on US evidence, not all Twitter users are eligible voters, the median Twitter is more educated than the median voter, and others (Hughes and Wojcik, 2019). Therefore, contributing to these ongoing debates is clearly beyond the scope of this investigation. Instead, the proposed theoretical framework can be utilised in future research to combine individual-level preferences and identity to online behaviour regarding

polarisation. This thesis focuses on understanding how AP (i.e., negatively-charged emotional language) flows from Salvini to public users, and vice-versa, across PH and discussion topics.

3.1 Affective Polarisation (AP)'s bi-directionality

Figure 3.2: Polarisation bi-directionality among Salvini's tweets and replies to his tweets



The first association to theorise is the one connecting tweets' polarisation and replies' polarisation. Let me consider a subsection (Figure 3.2) of the full framework, to consider the bi-directional link (a), connecting blocks 1 and 2. That is, tweets' polarisation is positively associated with replies' polarisation, and vice-versa. To explain such a link, I argue that the spread of polarisation occurs bi-directionally. The levels of polarisation in Salvini's tweet affects the levels of polarisation in replies, and vice-versa. This argument is three-fold.

As explained by supply-side-focused theories, politicians extensively utilise polarised language to attack their political opponents in their tweets (Barberá and Zeitzoff 2017; Druckman et al., 2019; Enli and Skogerbø 2013; Ott, 2017). Attacks through negative-charged language may occur for several reasons, among others gaining political advantage against an opponent, sharing political disagreement, or instigating a discussion, and others (Ott, 2017). Evidence shows that these attacks made by politicians have often been highly offensive and polarising (Ott, 2017; Kosmidis and Theocharis, 2020). As examined by Ott (2017), a recent example of such rhetoric is former US President Trump, who has utilised negative language in several tweets during his mandate. In Italy, during the 2018 elections, Italian parties have done wide use of personalisation in online communication, with an increasing use of targeting

and microtargeting techniques on social media (Bracciale and Cepernich, 2018; Bentivenga, 2015). Mazzoleni and Bracciale (2018) find that 53% of Salvini's posts include attacks on his political enemies, namely PD and migrants.

Considering demand-side-oriented theories, prominent public opinion scholarship argues that also the public has done extensive use of polarised language against politicians (Barberà et al., 2015; 2019a; McLoughlin and Ward, 2020; Tromble, 2018). Empirical evidence shows that users tend to utilise negative language often and in response to politicians' tweets (Theocharis et al., 2016; 2020). Significantly, users take advantage of their anonymity and Twitter's limited capacity to stop polarising content, to harass and troll the political discourse (Cheng et al., 2015; Theocharis et al., 2016). As such, the low cost of targeting a politician through anonymity leads to an exponential use of polarised language in users' Twitter activity (Phillips, 2016; Oz et al., 2018; Theocharis et al., 2020).

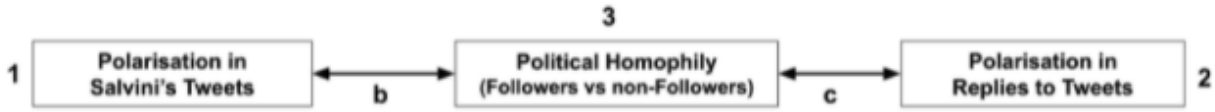
Focusing on both supply and demand sides, the spread of polarisation may occur bi-directionally (Theocharis et al., 2020; Tromble, 2018). On the one hand, one expects the repliers of politicians' tweets to either accept or reject such an attack. Those who approve such an attack may augment polarisation, further attacking the opponent. Instead, those who disapprove of such an attack may be attacking the politician's original tweet. On the other hand, when politicians are attacked by the public's replies, then they may also counteract with polarised language in successive tweets. As such, both politicians and the public play a crucial role in maintaining and spreading polarised political discourse.

Building on these theoretical accounts, I theorise that the levels of polarisation in Salvini's tweets are positively associated with those of users' replies to his tweets, and vice-versa. I expect such spread of polarisation to occur bi-directionally between the politician and his repliers. These accounts lead me to formulate the following theory-driven hypothesis, which describes the relationship between Salvini and his respondents as bi-directional.

Hypothesis 1: There is a bi-directional and positive relationship between polarised language in Salvini's tweets and replies to his tweets.

3.2 Affective polarisation (AP) and Political Homophily (PA)

Figure 3.3: Relationship between tweets and replies contingently on homophilic groups



The second and third links to discuss are those connecting tweets' polarisation and replies' polarisation through respondents' group-belonging (i.e., Salvini's followers vs non-followers). In Figure 3, such links (b, c) are both bi-directional, and link blocks 1 and 2, through 3. To explain these links, one has to consider the intersection between AP and PH. In particular, I argue that the levels of AP spread widely between Salvini and the public, across homophilic groups.

As discussed in the previous chapter, PH argues that homophilic group members tend to interact more with like-minded and less with people belonging to other groups. Limiting their exposition to only like-minded can lead users to adopt more extreme positions as they are never confronted with opponents (Colleoni et al., 2014; Stroud, 2010). In this way, homophilic groups members tend to share what their fellows and leaders write or think, creating an attitudinal or information cascade. Evidence from the USA shows that, due to Twitter homophilic groups, politicians' extreme views may be noticed, circulated, and amplified by their followers (Hong and Kim, 2016). In particular, Hong and Kim (2016) argue that social media platforms work across echo-chambers, and users hardly interact with

those who do not agree with them. As such, one would expect a politician's followers to share similar or, even, higher levels of polarisation in their replies to the original politician's tweets.

However, whilst agreeing on the homophilic group reactions argument, I contend with the echo-chamber assumption. The sharing features of Twitter (e.g., trending topics and explore function) allow users to be exposed to tweets regardless of their interests or followers (see Barberà et al., 2015; de Zúñiga and Valenzuela, 2011). Moreover, exposition of different political views may lead to greater tolerance, increasing users' understanding of the rationales supporting those views (Mutz and Mondak, 2006). As empirically demonstrated by Garrett et al. (2011), through social media people have more chances to interact and experience different viewpoints than through offline interactions. Nevertheless, in line with prominent scholarship, I argue that group-belonging is positively associated with the spread of the levels of polarisation (see Huddy et al., 2015; Mason, 2015; 2018b; Kosmidis and Theocharis, 2020). When a politician's followers agree with an attack made by the politician in the original tweet, the negative language utilised by the politician may be adopted, or augmented, in their followers' replies (Kosmidis and Theocharis, 2020). Alternatively, if a follower disagrees with the attack, they may utilise negative language to contrast the politician's point (ibidem). On the other side of the spectrum, if a non-follower is against the argument made by the politician in the original tweet, negative language may be utilised to attack the politician (ibidem) and protect their in-group member(s).

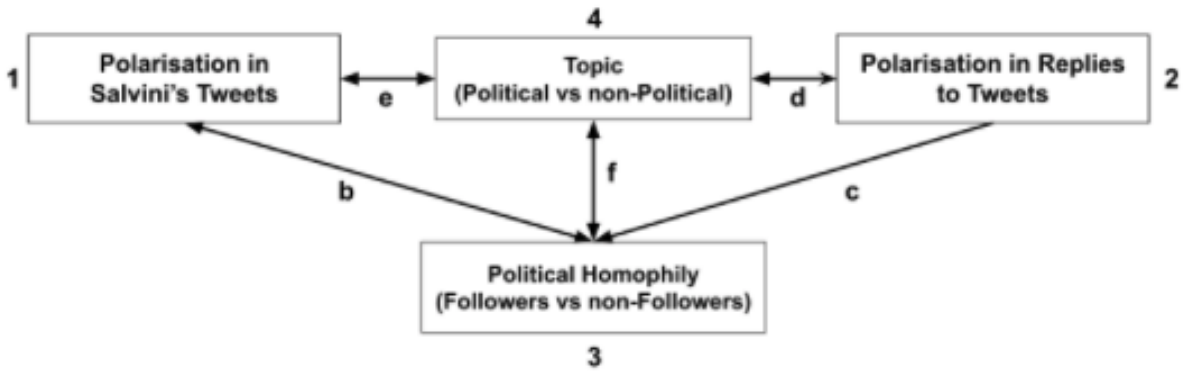
Considering Salvini's case, evidence from the 2018 elections show that, when exposed to such discussions, Salvini's followers engage with those discussion topics, and replicated polarised language in their replies to Salvini's tweets (Mazzoleni e Bracciale, 2018). Similarly, Sensales and Di Cicco (2019) find that Salvini's political posts are more likely to receive a greater number of interactions by his followers than by non-followers. Such posts are likely to initiate discussions into topics that are often neglected by more institutionalised Italian

parties, such as PD and Forza Italia (ibidem). For these reasons, I formulate the following theory-driven hypothesis on the positive association between tweets and replies' polarisation levels, contingently on the respondents being Salvini's followers or non-followers.

> **Hypothesis 2:** There is a positive relationship between Salvini's followers' replies and polarised language in successive Salvini's tweets and replies to his tweets.

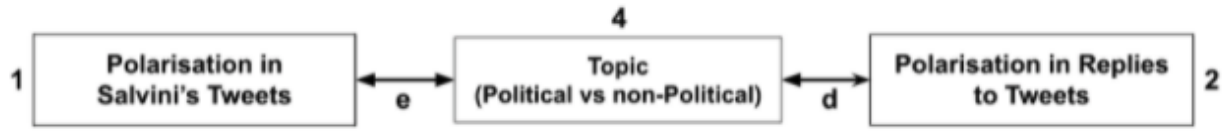
3.3 AP, PH and Discussion topics

Figure 3.4: Relationship between tweets and replies' polarisation, contingently on homophilic groups and discussion topics



The remaining three links to explore are those that stem from the discussion topics' block, linking it to tweets' polarisation, replies' polarisation, and homophilic groups' blocks. In Figure 5, two of these bi-directional arrows (d, e) represent, respectively, whether the tweets or replies' discussion topic (political or non-political) affects the spread of polarisation, and vice-versa. Moreover, the final arrow (f) refers to whether and how discussion topics are linked to homophilic group respondents, and vice-versa. Below, I discuss these two cases individually and in conjunction.

Figure 3.5: Relationship between tweets and replies' polarisation, contingently on discussion topics



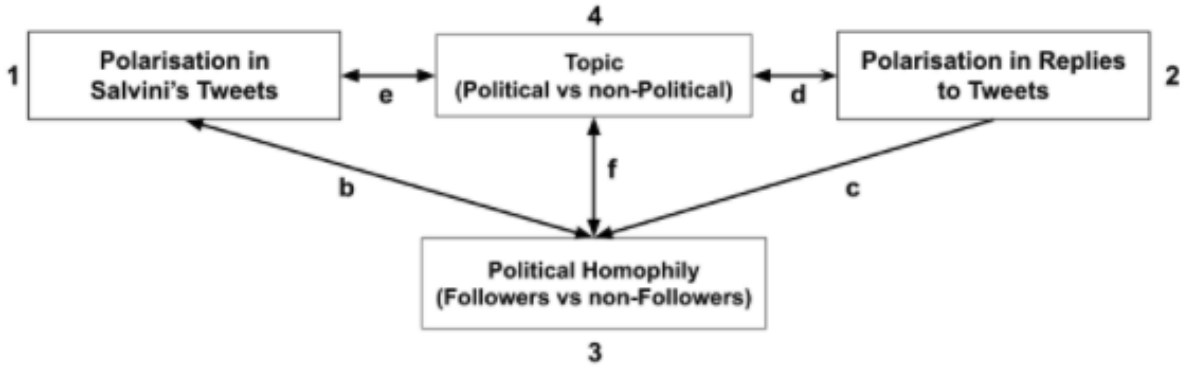
Regarding the first two arrows (d, e), I argue that there is a positive association between the polarisation in tweets and replies, contingently on political discussions. Previous research shows that the use of negative language is contingent on the discussion's topic (Barberà et al., 2015; Coe et al., 2014; Theocharis, 2020). Conversations about "hard news", such as politics, economics, immigration, are more likely to trigger politicians and citizens to utilise negative language than news about lifestyle and journalistic news (Coe et al., 2014: 669). In the same vein, Barberà et al. (2015) furnish evidence into how hard political tweets are more likely to induce both politicians and repliers to display greater levels of polarisation (ibidem). In the USA, political content is more likely than non-political content to increase polarised reactions by both politicians and public users (ibidem).

Salvini has made extensive use of political hashtags, according to Bracciale and Cepernich (2018). Salvini utilises political hashtags in 55.5% of his posts, vis-à-vis the other Italian major candidates Renzi (28%) and Di Maio (5%) (ibidem). Moreover, Di Cicco and Sensales (2019) find that Salvini's political posts resulted in a greater number of interactions by his followers, as well as greater use of strong polarised language in replies to his tweets. This may provide further evidence for assuming that political discussions lead to a greater use of polarised negative rhetoric. Therefore, I argue that, given the greater polarisation over political content than in non-political content, both Salvini and his repliers may display greater levels of polarisation. As a result, I also expect that such an increase in polarisation among Salvini and his repliers may impact the use of polarised language among them. In

other words, discussions over political content may lead both actors under analysis to increase the use of polarised language individually (i.e., when studying polarisation level in tweets and replies individually) and aggregately (i.e., when analysing both tweets and replies together). For these reasons, building on previous theories and empirical evidence, I formulate the following hypothesis.

> **Hypothesis 3.1:** There is a positive relationship between political content and the levels of polarisation in Salvini's tweets and replies to his tweets.

Figure 3.6: Relationship between tweets and replies' polarisation, contingent on homophilic groups and discussion topics



Regarding the final arrow (f) in our sampled framework, I consider the conjunction between all four blocks. In particular, if the Twitter conversation regards political content, I expect both Salvini's followers and non-followers to engage with the discussion. Going back to Barberà et al. (2015), they argue that political content increases polarisation across groups. This may be due to users consuming more information regarding non-political content, thus they are more used to engaging with different points of view on those subjects than a political matter. Instead, when dealing with political content, politicians and users are more likely to engage in animus discussions, displaying greater polarisation. Evidence from Italy

further corroborates such results. According to Paparo (2018), Salvini's rhetoric has managed to create a unified fanbase by using emotional elements against a common enemy. As such, when users, who held different political positions to Salvini, were exposed to denigratory content against who was seen as a common enemy, they reunited against the common enemy (ibidem). This unification against a common enemy was visible during the 2018 elections when often 5-Star-Movement and Salvini's supporters attacked the Democratic Party's candidates. As such, political content becomes important to understand polarisation flows between politicians and users, and vice-versa. For these reasons, I theorise that the arrow that connects polarisation in Salvini's tweets and replies through political content across groups is significant to understand how polarisation flows. In other words, when the Twitter discussions are related to political content, not only Salvini, but also both his followers and non-followers will increase their use of polarised language. For these reasons, I formulate the following hypothesis.

> **Hypothesis 3.2:** There is a positive relationship between Salvini's followers' replies and the levels of polarisation Salvini's tweets and replies to his tweets, contingent on political content.

3.4 Triggering factors

Having presented a theoretical framework to understand the dynamics of polarisation, I now turn to the triggering factors that may initiate such dynamics. In particular, I am interested in understanding under what conditions either Salvini's tweets, or replies, or both trigger initial polarised language (i.e., short-term impulses that initially activate polarisation). To do so, following Barberà et al. (2019a), I consider triggering factors, those factors that initiate polarised language in the first 15 days since the first Salvini's tweet has been posted. These will be analysed through a VAR model.

As previously discussed, often public users use strong negative language against politicians, and this precludes politicians to expose themselves in online discussions. This is due to negative replies to politicians' tweets that may undermine politicians' status (Barberà et al., 2015; 2019a; McLoughlin and Ward, 2020; Tromble, 2018). Therefore, one would not expect politicians to first expose themselves to critical issues, instead, politicians would wait for the public to lead the discussion (see Converse, 2006). In this sense, on Twitter, Barberà et al. (2019a) find empirical evidence that US politicians are more likely to follow the public, than the opposite. In particular, when online discussions are animated by polarised language, politicians may follow the public to be on the popular side, and gain consensus. Furthermore, Theocharis et al. (2020) find that politicians who display greater incivility are more likely to receive uncivil comments. Thus, politicians would not risk more attacks, and therefore decide to back the public instead of initiating negative conversations. Based on these assumptions, I develop the following hypothesis, which shows that repliers are the main triggers of polarisation vis-à-vis politicians.

> **Hypothesis 4:** Higher levels of polarisation in replies to Salvini's tweets trigger higher levels of polarisation in successive Salvini's tweets and replies.

Prominent scholars argue that strong partisans are the main driver of politicians' responsiveness on Twitter (Barberà et al.,2019a; Kosmidis and Theocharis, 2020). When active partisan users engage in political debates, they ultimately lead to issue salience because politicians allocate much more attention to their fanbase needs (Kastellec et al. 2015; Barberà et al.,2019a). On Twitter, such fanbase-attention also occurs. Politicians tend to favour accounts, who follow them back and who respond more often (vis-à-vis non-followers) to their Twitter tweets (ibidem). As such, party supporters become more influential than other public's users in increasing politicians' responsiveness. Barberà et al. (2019a) show that, as strong partisans mostly influence politicians' communication, discussions are driven by their

political interests. The influence of partisans in leading politicians' Twitter activities is a source of political polarisation (ibidem). When politicians favour their followers, they may preclude other parties' supporters to engage in political debates. As previously discussed, politicians' followers tend to display greater levels of polarisation in their language than non-followers (Huddy et al., 2015; Mason, 2015; 2018b; Kosmidis and Theocharis, 2020). Therefore, one expects that not only followers lead Twitter conversations, but they also increase polarisation.

Contextual to Salvini, Freschi (2015) qualitatively studies Salvini's communication on social media. She finds that Salvini in his communication tends to act as an amplifier of his fanbase's discussions. In particular, she finds that Salvini's posts and replies to his followers' posts support the arguments previously by his followers (ibidem). As such, Salvini replicates not only the discussion topics, but also the language utilised. Indeed, Freschi (ibidem) argues that Salvini and his fans have extensively utilised negative and offensive language against migrants and Italian southerners in their posts. This continuous support of his fanbase triggers and sustains an information and polarisation cascade across-time. Building on these previous accounts, I theorise the following hypotheses regarding the triggering effect of Salvini's followers in Salvini's polarised language:

> **Hypothesis 5.1:** Salvini's followers' replies trigger higher levels of polarisation in Salvini's tweets and replies.

> **Hypothesis 5.2:** Salvini's followers' replies trigger discussion topics in Salvini's tweets and replies.

Discussion topics are also significant triggering factors to initiate the spread of polarisation. As previously discussed, Theocharis et al. (2020) find that the use of political topics triggers

an increase in uncivil comments to politicians' tweets. Specifically, they find that public users' political replies are more likely to display polarised language than politicians' tweets on politics (ibidem). These highly polarised political replies have a subsequent impact on politicians' levels of polarisation. For this reason, one would expect that political replies may increase polarisation levels in successive tweets. Contextually in this case, empirical evidence on the language utilised by respondents to Italian politicians' posts shows a higher involvement of anger in Salvini's followers than his opponents' followers (Di Cicco and Sensales, 2019; Mazzoleni e Bracciale, 2018). Common attacks in both Salvini's followers' replies involve, among others, the left-wing Democratic Party (PD), and immigration (ibidem). Based on these accounts, I argue that polarisation may be initiated by political content in replies, which triggers a polarised reaction in successive Salvini's tweets. As previously discussed, when tweets' polarisation increases, I also expect successive replies' polarisation to increase. Therefore, as in hypothesis 3c, I expect political replies to display greater levels of polarisation as a result of political content being the main subject of Salvini's tweets. Based on these assumptions, I develop the following hypotheses.

> **Hypothesis 6.1:** Political tweets do not trigger higher levels of polarisation in Salvini's tweets and replies.

> **Hypothesis 6.2:** Political replies trigger higher levels of polarisation in Salvini's tweets and replies.

To operationalise such hypotheses empirically, I have created individual sub-hypotheses, which are summarised below.

Table 3.1: Summary 1/3 of hypotheses and sub-hypotheses

Main hypotheses	Sub-hypotheses
Hypothesis 1: There is a bi-directional and positive relationship between polarised language in Salvini's tweets and replies to his tweets.	**Hypothesis 1a (H.1a):** Higher levels of polarisation in Salvini's tweets (at t) are positively associated with lagged levels of polarisation in replies to Salvini's tweets (at *t-1*). **Hypothesis 1b (H.1b):** Higher levels of polarisation in replies to Salvini's tweets (at t) are positively associated with lagged levels of polarisation in Salvini's tweets (at *t-1*).
Hypothesis 2: There is a positive relationship between Salvini's followers' replies and polarised language in successive Salvini's tweets and replies to his tweets.	**Hypothesis 2a (H.2a):** Higher levels of polarisation in Salvini's tweets (at t) are positively associated with the lagged Salvini's followers' replies (at *t-1*). **Hypothesis 2b (H.2b):** Higher levels of polarisation in replies to Salvini's tweets (at t) are positively associated with Salvini's followers' replies (at t).
Hypothesis 3.1: There is a positive relationship between political content and the levels of polarisation in Salvini's tweets and replies to his tweets.	**Hypothesis 3a (H.3a):** Salvini's political tweets (at t) are positively associated with greater levels of polarisation in Salvini's tweets (at t). **Hypothesis 3b (H.3b):** Lagged Salvini's political tweets (at *t-1*) are positively associated with greater levels of polarisation in replies to Salvini's tweets (at t). **Hypothesis 3c (H.3c):** Lagged political replies (at *t-1*) are positively associated with higher levels of polarisation in Salvini's tweets (at t). **Hypothesis 3d (H.3d):** Political replies (at t) are positively associated with greater levels of polarisation in replies to Salvini's tweets (at t).

Note: This table outlines the main hypotheses and sub-hypotheses.

Table 3.2: Summary 2/3 of hypotheses and sub-hypotheses

Main hypotheses	Sub-hypotheses
Hypothesis 3.2: There is a positive relationship between Salvini's followers' replies and the levels of polarisation Salvini's tweets and replies to his tweets, contingent on political content.	**Hypothesis 3e (H.3e):** Higher levels of polarisation in Salvini's tweets (at t) are positively associated with lagged Salvini's followers' political replies (at t-1), contingently on political tweets (at t). **Hypothesis 3f (H.3f):** Higher levels of polarisation in replies to Salvini's tweets (at t) are positively associated with Salvini's followers' political replies (at t), contingently on lagged political tweets (at t-1).
Hypothesis 4: Higher levels of polarisation in replies to Salvini's tweets trigger higher levels of polarisation in successive Salvini's tweets and replies.	**Hypothesis 4a (H.4a):** Higher levels of polarisation in Salvini's tweets do not trigger higher levels of polarisation in successive replies to Salvini's tweets. **Hypothesis 4b (H.4b):** Higher levels of polarisation in replies to Salvini's tweets trigger higher levels of polarisation in successive Salvini's tweets.
Hypothesis 5.1: Salvini's followers' replies trigger higher levels of polarisation in Salvini's tweets and replies.	**Hypothesis 5a (H.5a):** Salvini's followers' replies to Salvini's tweets initially trigger higher levels of polarisation in Salvini's tweets. **Hypothesis 5b (H.5b):** Salvini's followers' replies to Salvini's tweets initially trigger higher levels of polarisation in replies to Salvini's tweets.

Note: This table outlines the main hypotheses and sub-hypotheses.

Table 3.3: Summary 3/3 of hypotheses and sub-hypotheses

Main hypotheses	Sub-hypotheses
Hypothesis 5.2: Salvini's followers' replies trigger discussion topics in Salvini's tweets and replies	**Hypothesis 5c (H.5c):** Salvini's followers' replies trigger the discussion topics in Salvini's tweets. **Hypothesis 5d (H.5d):** Salvini's followers' replies trigger the discussion topics in replies to Salvini's tweets.
Hypothesis 6.1: Political tweets do not trigger higher levels of polarisation in Salvini's tweets and replies.	**Hypothesis 6a (H.6a):** Political Salvini's tweets do not initially trigger higher levels of polarisation in Salvini's tweets. **Hypothesis 6b (H.6b):** Political Salvini's tweets do not initially trigger higher levels of polarisation in replies to Salvini's tweets.
Hypothesis 6.2: Political replies trigger higher levels of polarisation in Salvini's tweets and replies.	**Hypothesis 6c (H.6c):** Political replies to Salvini's tweets initially trigger higher levels of polarisation in Salvini's tweets. **Hypothesis 6d (H.6d):** Political replies to Salvini's tweets initially trigger higher levels of polarisation in replies to Salvini's tweets.

Note: This table outlines the main hypotheses and sub-hypotheses.

CHAPTER 4

METHODOLOGY

To study the dynamics of polarisation through Salvini's tweets and replies to his tweets, this paper utilises a combination of web-scraping, machine learning, and econometric methodologies in Python and R-software. This chapter discusses the case selection, procedures for data-collection, coding, and analysis, which will be implemented throughout the paper.

4.1 Case selection

The selection of Salvini's Twitter rhetoric is motivated by three main reasons. First, focusing on Salvini's Twitter campaign allows me to study the years in which social media has become a fundamental instrument for political campaigning (Bobba, 2019; Bobba and Quaranta, 2019; Paparo, 2018). Specifically, following Obama's 2008 and 2012 social media campaigns, politicians world-wide have done extensive use of such online campaigns (Barberà et al., 2019; McLoughlin and Ward, 2020; Theocharis et al., 2020; Tromble, 2018). In this thesis, I focus on the 2013-2019 period in which Twitter has becoming the major online political *agora*(Barberà et al. 2015; 2019; Tucker et al. 2018). Second, Salvini's campaign consists of a large number of tweets and interactions with the public, which provide researchers with a consistent and large data-set to conduct cross-time research on polarisation, more than any other major Italian politician (Mazzoleni and Bracciale, 2018; Bracciale and Martella, 2017; Freschi, 2015). In particular, in the last few years, Salvini has posted over 50,000 tweets, providing daily and, often, hourly Twitter-data. Salvini's posts, as well as his respondents' replies, furnish solid ground to study online political communication (Bobba, 2019; Bobba and Quaranta, 2019; Emanuele, 2018a; 2018b; Paparo, 2018). Third, and more broadly, the Italian party system is characterised by a competitive proportional system, in which there is little policy variation across the right-wing political spectrum and, thus, the role of

divisive rhetoric is accentuated to win consensus (Norris and Inglehart, 2016; Cinelli et al., 2019; Mazzoleni and Bracciale, 2019; Bobba and Roncarolo, 2018). Therefore, Italy provides a suitable natural environment to study variation in the levels of polarisation (Norris and Inglehart, 2016; Cinelli et al., 2019). Moreover, the research timespan accounts for both electoral and non-electoral periods, which allow me to study significant events in Italian politics through Salvini's Twitter activities.

While I am aware that studying a single-N case will not provide enough variation to produce generalisable results (King, Keohane, and Verba 1992), this thesis is primarily descriptive in nature to understand how, and under what conditions, polarisation spreads. As such, I rely on cross-time variation to explore the interaction dynamics within a single-N case through Salvini's tweets and replies to his tweets. Bobba and Roncarolo (2018) argue that Salvini's online rhetoric was the most influential across the Italian political spectrum during the 2018 elections, surpassing his competitors by large numbers. Thus, comparing Salvini's tweets to other Italian politicians would have introduced a weighting bias due to the disproportionate number of Salvini's tweets and replies vis-à-vis his opponents. In particular, others may argue in favour of regression weights to create frequency weight that gives more weight to other politicians and less weight to Salvini. However, using weights may result in inflated standard errors of the estimates, and thus less precise findings (Gelman, 2007, Kott, 2007).

To capture polarisation variation across-time, the timespan of this investigation goes from 15 December 2013, when Salvini became the leader of Lega Nord, to 9 August 2019, when his government broke down. This timespan allows me to focus on the most electorally successful period for Lega Nord where the party rose from 4% to 37% voting-intention (Ministero dell'Interno, n.d.). During my DPhil, I aim to expand my sample size to include new Italian parties and politicians (e.g., Brothers of Italy, Italy Forward), and a longer timespan (e.g., post-2019, including COVID-19 crisis) to further capture how polarised language flows on Twitter.

4.2 Data-Collection and Coding

The data has been collected through web-scraping, and it has been coded through neural networks textual classification. The following subsections explore the web-scraping and neural network methods utilised. In Appendix A, I discuss users' privacy and rights.

4.2.1 Data-collection

Twitter data-collection has been operationalised through an advanced web-scraping algorithm, which allowed to automate the retrieval of over 10 million data-points. This algorithm has retrieved all the website links to Salvini's root tweets (i.e., original tweets) and mention tweets (i.e. tweets posted by other users that mentioned Salvini's handle). In particular, it has retrieved the links to root tweets and mention tweets, which were posted within the research timespan (i.e., December 2013-August 2019), and included at least one of the political and/or non-political hashtag from the hashtag lists (Table 4.2). Third, having retrieved the links to the root tweets and mention tweets, I have programmed a scrutiniser to extrapolate root tweets and mentions' IDs from the links. Finally, from the retrieved tweet and mention's IDs, this paper has programmed an algorithm to retrieve information related to each ID (see Appendix B for more details). This procedure has retrieved the data presented in Table 4.1.

Table 4.1: Summary of the collected information for each tweet and reply's ID

Data	Function
Date and Time	Date and Timestamp of when the tweet/reply has been posted by Salvini, or his respondents.
Tweet ID	The unique, and randomly-generated tweet/reply's ID, which identifies the tweet/reply and its author.
Text	The text contained in a given tweet or reply.
Followers	The number of followers that a user, who posted a tweet/reply in our sample, has.
Followings	The number of other users followed by the user who posted the tweet/reply.
Number of tweets	The number of tweets posted by the user who posted the tweet/reply.
Location	The geo-location (at the national level) from which the tweet/reply has been posted.

Note: This table outlines the collected information for each Salvini's tweet and reply to his tweets.

I have set a new data-collection automation code to retrieve all the users' IDs who follow Salvini in accordance with users' privacy and rights (see Appendix A). To study homophilic group-belonging, I followed Barberà et al. (2015), and considered Salvini's followers as a homophilic group given their ideological proximity estimations (i.e., those who follow Salvini are likely to be his supporters). The automation code has been set to target Salvini's followers' list on Twitter, and retrieve all the users' ID, which appeared on the list. The retrieved users' IDs have been appended and stored in a new data-set. Subsequently, I intersected Salvini's followers' list with mentions' user IDs to understand whether those users, who mention Salvini, were Salvini's followers. From this intersection, I have created a dummy variable to identify Salvini's followers (1) and non-followers (0). This intersection allows me to study PH, and its crucial interaction with polarisation. I will be able to identify whether a given respondent belongs to a specific homophilic group, either Salvini's followers

or non-followers.

Having completed the Twitter-data, this paper has operationalised several filters to avoid ghost accounts and bots' mentions (i.e., inactive accounts and/or computer programmed accounts). To do so, this paper has followed and adapted the procedure developed by New York University's Social Media and Political Participation (SmaPP) open-source tools (Barberà et al., 2015; NYU, n.d.). In particular, I have computed five filters on the collected Salvini's mentions to only include real people. The five filters have acted on mentions' users who (1) follow 25 or more accounts, (2) have 25 or more followers, (3) have posted at least 25 tweets, (4) mention at least one of the hashtags in Table 4.2, which are discussed below, (5) tweeted from a location in Italy (this will aid me in the coding procedure to exclude non-Italian tweets).

Sampling tweets and mentions based on the hashtags mentioned, rather than on keywords, allows understanding overall debates regarding political and non-political events. For example, in Italian, the word "amici" means "friends", yet "amici" allows me to study tweets related to the Italian TV entertaining and talent-show programme called "Amici". If one were to include "amici" as a keyword, one would have retrieved tweets regarding friends/friendship, which could not tell whether the tweets are political or non-political.

To sample significant hashtags, through a scrutiniser, I have sampled 60 hashtags based on Google Trends's (n.d.) Italy's most searched political and non-political words within the 2013-2019 timespan, which have been utilised at least once by Salvini. Specifically, I have set a scrutiniser on top searched words within the political (e.g., political events, or politicians) and non-political (e.g., entertainment, music, and sport) categories. The hashtags have been evenly divided between political (30) and non-political (30) for fair representation. Political hashtags include political events (i.e., Bibbiano scandal, Elections, etc.), parties (i.e., PD, M5S, Lega, etc.) or politicians (i.e., Silvio Berlusconi, Matteo Renzi, etc.). Non-political

hashtags include events, such as sports (e.g., footballers, athletes, MotoGP, Olympic Games, etc.), and entertainment (i.e., TV shows, Italian music festivals, etc.). This thesis considers the political and non-political hashtags, which are collected in Table 4.2. Moreover, Appendix C explains each hashtag in detail.

Table 4.2: List of sampled hashtags by content type (political vs non-political)

Sampled Hashtags	Topic
'amici', 'azzurri', 'balotelli', 'bebevio', 'buongiorno', 'dovizioso', 'ferrari', 'fognini', 'gfvip', 'grandefratello', 'italiangp', 'mahmood', 'milan', 'motogp', 'maldini', 'olimpiadi', 'pellegrini', 'pizza', 'rugby', 'sanremo', 'sanremo19', 'sanremo2014', 'sanremo2015', 'sanremo2017', 'sanremo2018', 'sanremo2019', 'seriea', 'settebello', 'simoncelli', 'tramonto'	Non-Political
'berlusconi', 'bibbiano', 'boldrini', 'clandestini', 'dimaio', 'elezioni', 'europa', 'governo', 'grillo', 'immigrati', 'lega', 'm5s', 'madeinitaly', 'meloni', 'mes', 'migranti', 'pd', 'politica', 'primagliitaliani', 'renzi', 'repubblica', 'salvini', 'salvinipremier', 'senato', 'terrorismo', 'terroristi', 'ue', 'unionicivili', 'votalega', 'zaia'	Political

Note: This table outlines the sampled political and non-political hashtags. At least one of these hashtags has to be contained by a Salvini's tweet, and reply to his tweets to be sampled by my web-scraping algorithm. Appendix C explains each hashtag in detail.

For each root tweet in the data-set, which mentioned a political or non-political topic, I created a dummy variable based on a single topic. I set a new variable, 'hashtags', which collects all the hashtags in a given root or mention tweet. From the hashtag variable, I computed a dummy string detector, which separated each hashtag contained in the hashtags' column. For every detected single hashtag, a new dummy variable has been created. If a string value in an original tweet contains one of the detected hashtags, the dummy will return a 1 otherwise a 0. This allows not only to identify whether a root or mention tweet is political or non-political, but also the exact topic the tweet or mention is about.

I sampled Salvini's replies among the mention tweets. As highlighted by several scholars, Twitter's mentions can be divided into two types (McLoughlin and Ward, 2020; Theocharis et al., 2020; Tromble, 2018). First, they can be direct replies, which are tweets that mention politicians and are direct replies to a politician's tweet. Second, they can be simple mentions, which are tweets that mention a politician's username without being a response to one of the politician's tweets. To do so, I subset the mentions lists on those mention tweets which were linked to Salvini's root tweet. In sum, among all the tweets that mention Salvini, I will only consider replies to Salvini's tweet. Doing so allows me to study the effect of a given Salvini's tweet on his direct respondents, and vice-versa.

To summarise, after following the above-discussed data-gathering procedure, I have gathered 23,690 Salvini's tweets and replies to his tweets that mentioned the 60 political and non-political topics. Specifically, the data-gathering process has yielded 10,826 Salvini's unique tweets, which had at least one reply that mentioned one or more topics under analysis. Moreover, for each tweet and reply, dummy variables identify the type of content (i.e., political vs non-political), and the exact topic (see Table 4.2). Specifically for repliers, I can also identify whether respondents are Salvini's followers or not. Such a procedure allows me to study the effects of polarisation (i.e., use of negative-emotionally-charged language) in a given Salvini's tweet on polarisation in replies, and across homophilic groups (i.e., followers vs non-followers) and topic (political vs non-political).

4.2.2 Data-coding

Neural Network

This paper has programmed a pre-trained neural network textual-classification algorithm on a remote server (see Appendix D for details) to effectively perform data-coding of Salvini's tweets and replies.

Neural Network (NN) algorithms, or 'deep learning', refer to a branch of machine learning which utilises multi-layered models (Burger, 2018; Boehmke and Greenwell, 2020; Joo and Steinert-Threlkeld, 2018). Their computational form resembles one of the neurons in a biological system (Burger, 2018). For a given list of inputs, the NN performs several processing steps, and returns an output. NN works through logic gates, based on True/False boolean variables. If a condition is true, then the input moves across the NN to the next logic gate; otherwise, if the condition is false, the input follows another path to the next gate. In other words, one can imagine the NN to be composed of layers and nodes. First, the layers are vertical sections, which represent a logic gate - a condition of a textual or numerical feature. There are three types of layers: inputs (in this thesis, it is a tweet/reply's text), hidden (these are the logic gates the inputs go through), outputs (the computed results that, in this thesis, is a sentiment probability score for the given tweet/reply - ranging from 0 to 1). Second, the nodes are the computational points within each layer. For these steps to occur, a bias variable is set as a constant to the initial equation for calculations, and it is represented as a node on top of each layer of the NN. Employing a NN allows the final outputs to be the result of "a sequence of operations applied to the input data and intermediate results" (Joo and Steinert-Threlkeld, 2018). Deeper models, such as models with additional classification layers, are more likely to obtain higher coding accuracy, and capture complex data distributions than flat machine learning models (ibidem; Poggio et al., 2017).

Other scholars highlight the use of another coding approach: dictionary-coding, yet it presents methodological limitations, which can be overcome through NN. Dictionary coding refers to using pre-established psycholinguists values, accounting for affect and emotions for each word in the list, and applying those to larger populations of text (Kosmidis et al., 2019; Bradley and Lang, 1999; Dodds and Danforth, 2010; Montefinese et al., 2013; 2014). Nonetheless, this approach does not account for context and domain, and it is limited in the number of words it contains (Barberà et al., 2016). Utilising machine learning methods over-

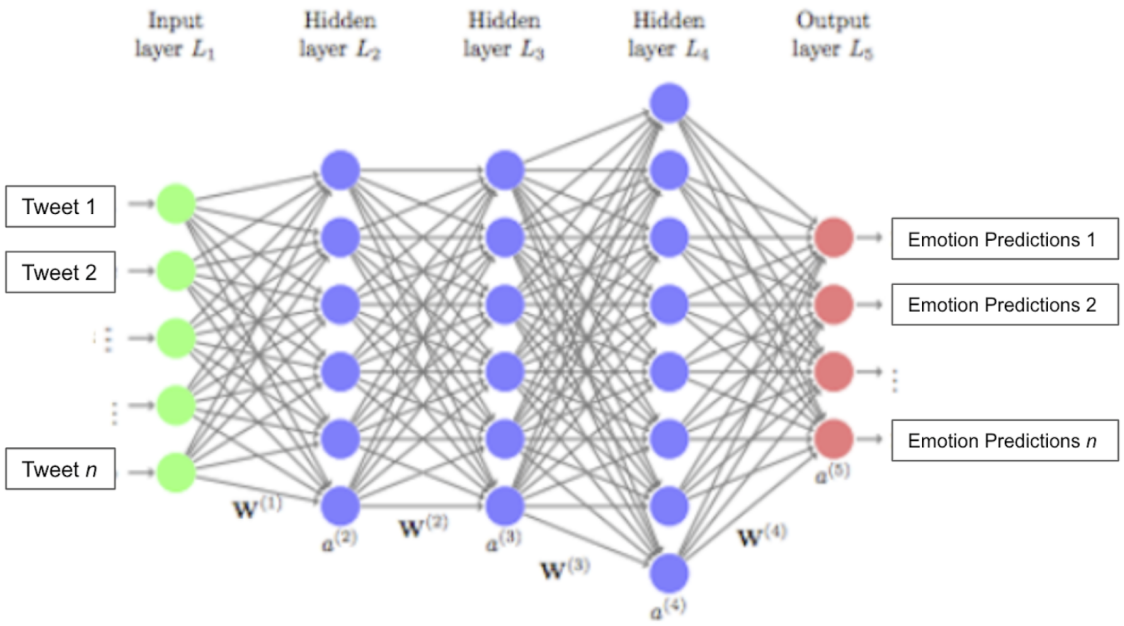
comes these issues by developing a coding algorithm through human coders and samples of text, which can be then extrapolated to the whole population of text (Barberà et al., 2019b; Barberà et al., 2016; Hastie et al., 2009). Machine learning methods allow me to code large samples of text accounting for their context, their domain and for words that may not be included in the coding dictionaries. Moreover, through machine automation, it is possible to code large sets of data in a short time, whilst maintaining high coding standards. Therefore, machine learning methods widely outperform dictionary-coding systems in coding data.

In this thesis, such algorithms are utilised to classify a given Salvini's tweet, or reply to his tweets. To measure polarised language, this paper has leveraged a pre-trained algorithm by Colneric and Demsar (2018, CD henceforth), as recently suggested by Mechkova and Wilson (2021), and Friedrichs, Stoher and Formisano (forthcoming). In particular, CD's classification produces a sentiment analysis for a given tweet or reply (see next sections for more details). Following Kosmidis and Theocharis (2020) and Marcus et al. (2020), the feelings of anger and disgust are those most correlated with uncivil/polarised reactions. These feelings are opposed to feeling of joy and trust, which may fail to explain polarisation. As such, from CD's classification, I specifically sampled anger and disgust to study AP language as both emotions are likely to carry hostile, or polarised, language. The following sections will explain in full CD's method, as well as the measurement of AP language.

CD have adapted Plutchik (1980), Ekman (1992), and Norcross et al.'s (1984) previous models for estimating several textual emotional charges into a single algorithm to identify a dominant emotion in a given tweet. Plutchik (1980) seminal work classified textual emotions across eight broad categories based on the hashtags used, such as joy and sadness, trust and disgust, fear and anger, and surprise and anticipation. Ekman (1992) classified textual emotions across six categories, also based on the hashtags used, such as anger, disgust, fear, joy, sadness, and surprise. Finally, Norcross et al.'s (1984) utilised 55 adjectives contained in the text to identify broad categories. CD (2018) departs from previous models by training

a recurrent neural network (RNN) on 73 billion tweets in the English language. This RNN classified tweets' dominant emotional charge based on their hashtags relatedly to Plutchik, Ekman and Norcross et al.'s text-labelling schemes. In other words, CD's input was a tweet's label, and trained the neural network on the rest of the text in those tweets to understand which word(s) were correlated with the tweet's labels. As a result, CD obtained a pre-trained algorithm that can take a tweet as input data, and provide the statistical likelihood of a tweet's dominant emotional charge being one of Plutchik, Ekman and Norcross et al.'s broad categories. Figure 4.1 shows an adapted version of Boehmke and Greenwell (2020)'s NN, which will be utilised in the following chapters.

Figure 4.1: An adapted version of Boehmke and Greenwell (2020)'s NN



Note: This figure shows Boehmke and Greenwell (2020)'s NN to which I have added input and output layers to visualise the functioning of the NN, which will be utilised in the following chapters.

Within CD's NN, this thesis has utilised Plutchik's 8-emotion-model because it offers a good trade-off between the less nuanced Ekman's six categories, and the complex Norcross et al.'s 55 categories. To better understand such a mechanism, two figures are provided. For demonstration purposes, the following methodology is based on a random set of replies to Salvini's tweets. Figure 4.2 shows the emotional predictions of a given reply's text (referred to 'Tweet' in Figure 4.2), within a random set of replies to Salvini's tweets, to be one of the eight macro-categories.

Figure 4.2: A random sample of replies' emotional predictions by the NN

	Tweet	Emotion
1	@matteosalvinimi "Ma io avevo detto lira, non euro" #berlusconi docet. #salvinibaciamilano	Trust
2	@matteosalvinimi @RaiPortaaPorta #abusodufficio #domenicaVotoLega ... ma sono contento che al Governo #Salvini sia... https://t.co/snSuVM53RO	Fear
3	@matteosalvinimi Ora banna anche #FreeCarolaRakete e vergognati #49milioni #Berlusconi #Siri #SalviniDimettiti	Trust
4	@matteosalvinimi Un "ministro dell'Interno" che frigna per ricevute minacce e non risponde mai su quelle che lancia... https://t.co/6GhFOBZz6k	Surprise
5	He may be no Leonardi da Vinci, but he can help you unleash innovation. Join @mtjwatts and @phoebegoh at the famous... https://t.co/9rdVps70O8	Joy
...	...	...
95	@matteosalvinimi Sempre con riferimento a "Il Giornale" di famiglia, vero Salvini? Eravate e rimanete i servitori d... https://t.co/vJCh39XBzA	Trust
96	@matteosalvinimi #Salvini Su il canale del tuo amico #Berlusconi ch'è amico del #Pd = #staseraitalia è in Diretta... https://t.co/t1BhI0q8tw	Joy
97	@matteosalvinimi @rtl1025 #sbloccacantieri #Salvini : Sei Impicciato Fine Al Collo Con I Tuoi Creditori Spiega Al P... https://t.co/MecmSxUthP	Joy
98	@matteosalvinimi Salvini volevo farle i miei auguri. Non per il fidanzamento con la leopoldina ma perchè 25 anni fa... https://t.co/Dnvxv46cVJ	Trust
99	@matteosalvinimi Era un leghista.\n Uno di loro.\n Vogliono fare solo notizia sti idioti lecchini di #Berlusconi	Trust

Note: This figure shows a random sample of replies' emotional predictions by the NN. In the figure, 'Tweets' refer to a reply's text. 'Emotion' refers to the emotional prediction made by the NN.

Secondly, the pre-trained algorithm is computed again to calculate the probability of each reply, within the random set of replies, to be in each of the eight macro-category. Figure 4.3 shows a replies' sample, where each row corresponds to a given reply's text (referred to reply 'Tweet' in Figure 4.3) with the respective calculated probabilities (ranging from 0 to 1).

Figure 4.3: A random sample of replies' emotional probabilities by the NN

	Tweet	Anger	Disgust	Fear	Joy	Sadness	Surprise	Trust	Anticipation
1	@matteosalvinimi "Ma io avevo detto lira, non euro" #berlusconi docet. #salvinibaciamilano	0.020532	0.005001	0.249487	0.084553	0.092716	0.061934	0.442439	0.043338
2	@matteosalvinimi @RaiPortaaPorta #abusodufficio #domenicaVotoLega ... ma sono contento che al Governo #Salvini sia... https://t.co/snSuVM53RO	0.026241	0.006144	0.634011	0.118627	0.043141	0.010073	0.156243	0.005520
3	@matteosalvinimi Ora banna anche #FreeCarolaRakete e vergognati #49milioni #Berlusconi #Siri #SalviniDimettiti	0.025373	0.004304	0.144700	0.065403	0.020943	0.134651	0.578813	0.025811
4	@matteosalvinimi Un "ministro dell'Interno" che frigna per ricevute minacce e non risponde mai su quelle che lancia... https://t.co/6GhFOBZz6k	0.005143	0.001999	0.016964	0.244459	0.010128	0.409836	0.296540	0.014932
5	He may be no Leonardi da Vinci, but he can help you unleash innovation. Join @mtjwatts and @phoebegoh at the famous... https://t.co/9rdVps70O8	0.002952	0.001595	0.026607	0.576968	0.017500	0.047944	0.313226	0.013209
...	...	...	...	...	...	...	...	...	...
95	@matteosalvinimi Sempre con riferimento a "Il Giornale" di famiglia, vero Salvini? Eravate e rimanete i servitori d... https://t.co/vJCh39XBzA	0.017360	0.004323	0.103049	0.270308	0.076503	0.222891	0.279072	0.026494
96	@matteosalvinimi #Salvini Su il canale del tuo amico #Berlusconi ch'è amico del #Pd = #staseraitalia è in Diretta... https://t.co/t1BhI0q8tw	0.002570	0.000214	0.030905	0.735991	0.013396	0.052627	0.146682	0.017615
97	@matteosalvinimi @rtl1025 #sbloccacantieri #Salvini : Sei Impicciato Fine Al Collo Con I Tuoi Creditori Spiega Al P... https://t.co/MecmSxUthP	0.010205	0.004857	0.129832	0.334379	0.030759	0.138668	0.301813	0.049487
98	@matteosalvinimi Salvini volevo farle i miei auguri. Non per il fidanzamento con la leopoldina ma perchè 25 anni fa... https://t.co/Dnvxv46cVJ	0.027580	0.005650	0.044493	0.272608	0.202764	0.074561	0.298883	0.073460
99	@matteosalvinimi Era un leghista.\n Uno di loro.\n Vogliono fare solo notizia sti idioti lecchini di #Berlusconi	0.008025	0.002644	0.092087	0.197680	0.200838	0.118343	0.336394	0.043987

Note: This figure shows a random sample of replies' emotional probabilities by the NN. In the figure, 'Tweets' refer to a reply's text. The other 8 columns (e.g., Anger, Disgust, Fear, etc.) refer to the emotional probabilities made by the NN.

The output nodes, or results of the textual classification of our tweets/replies, are safely stored following the procedure discussed in Appendix A. The coded-data will be utilised for the time-series analysis in later stages. I first turn to the NN validation via a machine learning model: Elastic Net.

Neural network validation

To validate the neural network for textual-classification, this paper has performed an Elastic Net (EN) regression model to predict tweets/replies' sentiment based on the CD's NN classification in R-software. In Appendix E, I discuss in detail the functioning of the EN.

Since an important concern is the predictability power of the NN classifier, the EN becomes a strong method to understand how our NN predicts the emotional component of each tweet and reply in our sample. In particular, it provides an analytical lens to study the RMSE (i.e., root-mean-square error) and R-squared (i.e., coefficient of determination) coefficients of a coding algorithm across different amounts of regularisation, to understand how well a model predicts a given outcome variable (Boehmke and Greenwell, 2020; Zou and Hastie, 2005).

Specifically, in this thesis, the EN tests how the NN predicts levels of polarisation in text, accounting for the lower error term. This validation can also be visualised graphically, having on the y-axis RMSE and R-squared coefficients, whilst having the amount of regularisation (based on the proportion of lasso penalty) on the x-axis. Finally, the EN enables me to visualise a variables' importance plot, which shows the most important words (or grams) that lead to a specific sentiment in my tweets/samples. For this thesis, I am interested in understanding which words lead to higher polarisation (i.e., the averaged coefficient of the sentiments of anger and disgust in a given tweet/reply).

To evaluate the EN, I use two metrics: the RMSE and R-squared coefficients. By definition, the RMSE represents the average magnitude of prediction error (Imai, 2017). It is calculated through the square root of the prediction error (actual outcome minus predicted outcome). In this case, I compare the test set (actual outcome) to the training set's results (predicted outcomes). The closer the RMSE coefficient is to zero, the lower the bias in our samples. An issue may arise when comparing different combinations of training and testing sets, and this is why in later stages I use 5-fold cross-validation. In other words, the RMSE coefficient allows me to better understand the bias in our models, and to account for it in our results. Moreover, I have chosen to study the R-squared because it is widely recognised as a strong determinant of how well our model fits the data. It shows the proportion of variance in the dependent variable that can be explained by the independent variable.

Prior to employing the EN validation on the NN textual-classifier, I performed basic data cleaning on the tweets/replies' text. First, I converted all @ mentions to "@" to ensure that the algorithm does not just learn which politician users are mentioned, I also removed the URLs and numbers in the tweets. Second, I tokenised the tweets' text, converting the sentences to individual words "tokens". Third, I removed stop-words, removing the most common words in the Italian language from the text sample through the R-package 'stopwords' (Benoit et al., 2021). Fourth, it has performed stemming, removing the end

of words utilising a set of heuristics, which condenses words that have different meanings depending on their form into one to allow for a smaller pool of words. Fifth, I converted tokens to a term frequency-inverse document frequency matrix (tf–idf). In other words, it has produced a numerical statistic that will enable us to understand how important a word is to a document in a corpus of texts. Finally, I normalised the tf-idf (i.e., instead of considering absolute frequencies which may lead to biased results, the paper is interested in words' relative frequency). Tweets/replies' text can now be utilised in modelling.

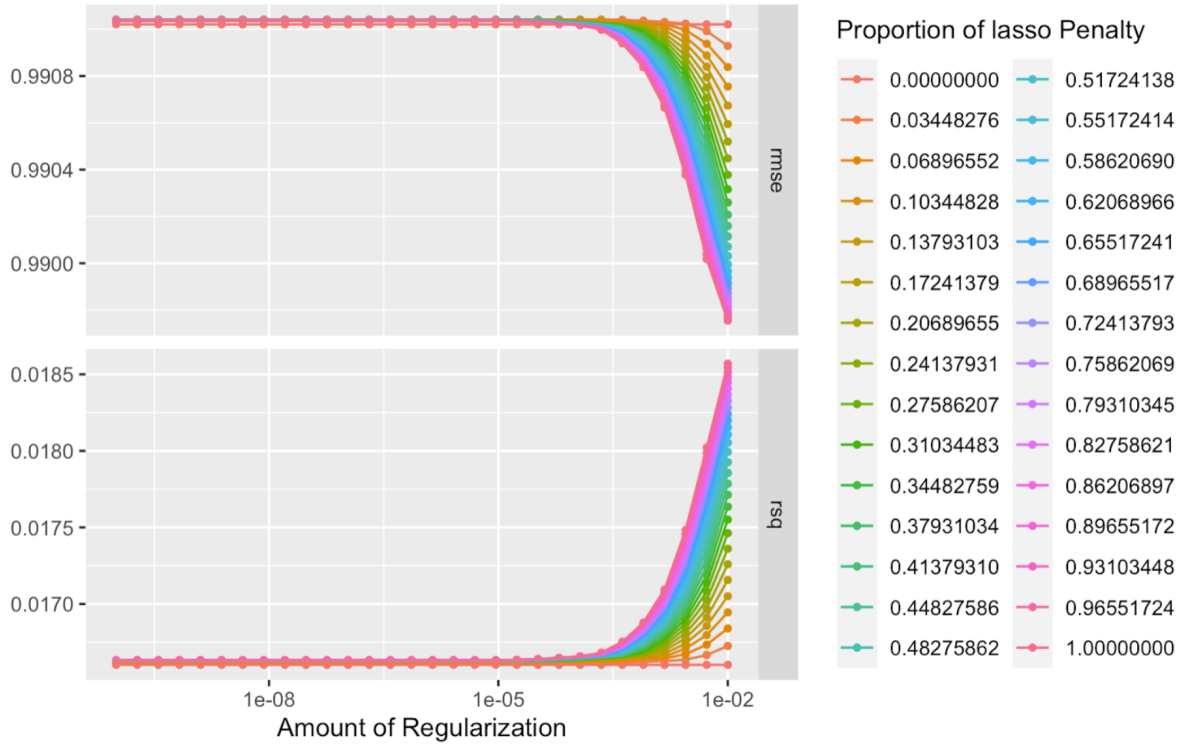
Having completed the data-cleaning procedure, I split the data into training and test sets with a ratio of 70%/30%, and I created 5-folds for cross-validation, setting the 'seed' to a random value (i.e., 5555) to study several test/training sets combinations. Moreover, I set a linear model in which the dependent variable was tweets/replies' Polarisation coefficient (i.e., the averaged coefficient of the sentiments of anger and disgust in a given tweet/reply, ranging from 0 to 1) set as numeric. The independent variable has been the Text variable (i.e., the variable containing all the tweets/replies' 'cleaned' text - see the previous paragraph for details).

The k-folds cross-validation (CV) will be significant to analyse the performance of the NN. Running machine learning models on 100% of data could lead to make false generalisations about our model, which may not fit with new data (Burger, 2018). Indeed, my model could well fit the data, however, it could not be utilised for any other data-set. Therefore, I split the tweets/replies-sample into training and testing sets to understand how well this model can predict results. Simply splitting the data into train/test sets can lead to high variance (i.e., how much prediction can change if I modelled the model using a different training set) and bias (i.e., how off the prediction is from the true relationship). To re-establish a bias-variance trade-off, I employed a k-fold CV. By definition, the k-fold CV randomly splits the data into k chunks, utilising k-1 chunks as training set and the remaining one as the testing set; it then repeats the process for all possible k-1 combinations and average across

the estimations (ibidem; Boehmke and Greenwell, 2020). This cross-validation has been shown to establish a good bias-variance trade-off because it allows me to study whether our models' results are endogenous to the combinations of the train/test set of our sample.

Following all the preparation steps, I tuned the EN model to predict tweets' sentiment based on the classification by setting a penalty range between -10 and -2. Moreover, I fitted at least 30 different tuning models (i.e., setting the 'levels' parameter to 30 when tuning the grid) to understand how predictability changed across proportions of lasso penalty. Through the 'autoplot' function, I created a plot to display the model performance for all the supplied hyperparameter values. Figure 4.4 displays a summary of the EN output.

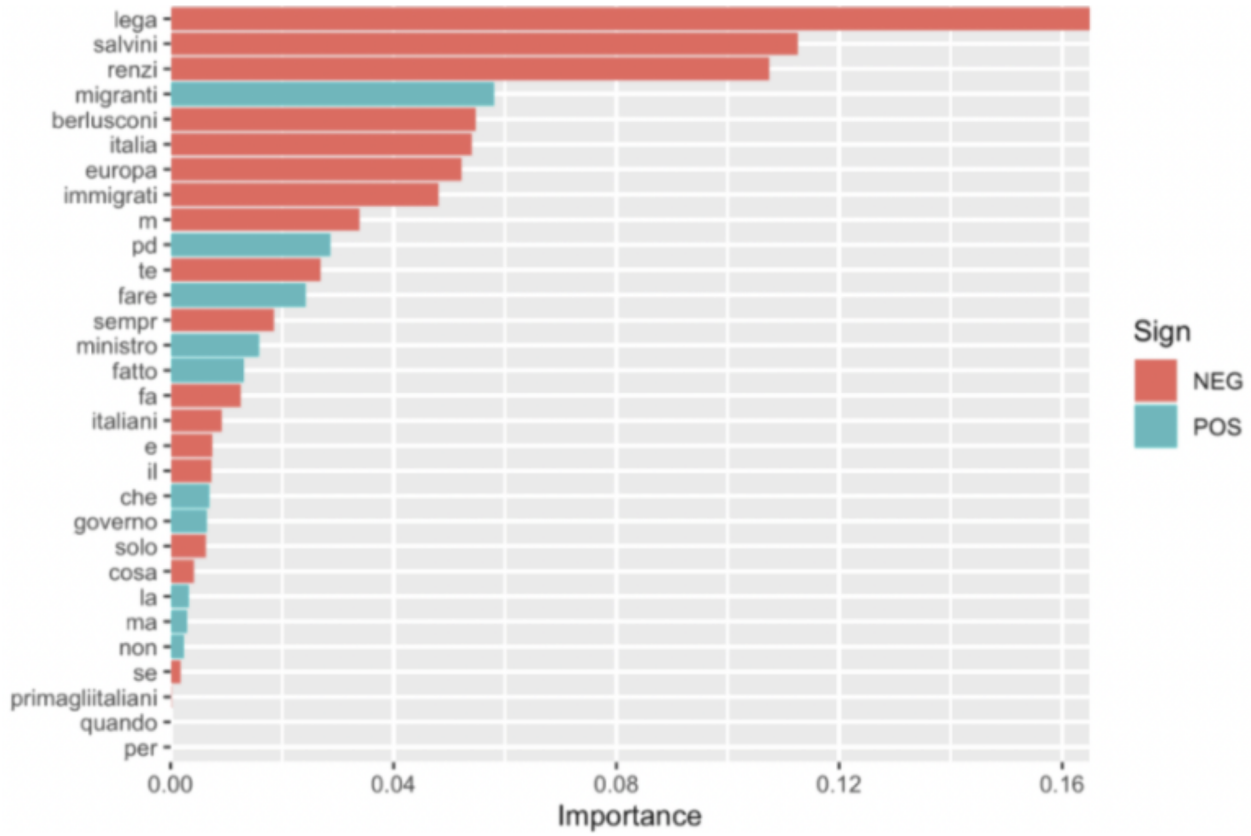
Figure 4.4: Summary of the Elastic Net output



Note: This figure represents the Elastic Net output, showing how a change in the proportion of lasso penalty (right-hand-side table) affects the model's performance. It displays the RMSE and R-squared coefficients (left-hand-side rows) across different amounts of regularisation. The key intuition to interpret the figure is to look at the trends of R-squared coefficients (which are shown to be up-ward), and RMSE coefficients (which are shown to be down-ward). As previously discussed, well performing classifications have high R-squared coefficients (or up-ward trend), and low RMSE coefficients (or down-ward trend).

Following the performance plot (Figure 4.4), as the lasso penalty increases, the model's performance increases: the R-squared coefficients increase dramatically, whilst the RMSE coefficients decrease. That is, the NN model may have well classified the tweets/replies in the sample. These graphical results can be explored at the numerical level. I selected the best lasso penalty, which allowed me to yield the highest R-squared and lowest RMSE coefficients. As such, I obtained an R-squared coefficient of 0.014, and an RMSE coefficient of 0.87. However, these coefficients show that the model has not met conventional standards. To corroborate these results, I produced a variables' importance plot, which shows the top 30 words that predict polarised tweets or replies (see Figure 4.5).

Figure 4.5: Summary of the Elastic Net's variables' importance plot



Note: This figure represents a variables' importance plot, which shows the top 30 words that predict polarised tweets or replies. The words' bars coloured in green have a positive effect on polarisation (i.e., they lead to greater levels of polarised language). The words' bars coloured in red have a negative effect on polarisation (i.e., they lead to lower levels of polarised language).

Starting from the variables' importance plot (Figure 4.5), there are two main findings. On the one hand, tweets and/or replies containing words, such as "lega", "salvini", "renzi", "berlusconi", "italia", "europa", "immigrati" and others, have a negative effect on the dependent variable (i.e., level of polarisation). On the other hand, tweets and/or replies containing words, such as "migranti", "pd", "ministro", "fatto[quotidiano]", "governo", and others, have a positive effect on the levels of polarisation. That is, when tweets and/or replies mention Salvini's persona (i.e., "salvini"), party (i.e., "lega") and allies (such as Berlusconi), the levels of polarisation are low, as expected. When tweets and/or replies mention Lega's opposition parties (such as the Italian Democratic Party, namely "pd"), minister (i.e., "ministro"), government (i.e., "governo"), and the main opposing newspaper (such as Fatto Quotidiano, namely "fatto"), the levels of polarisation rise, as expected.

However, some exceptions rise. First, according to the variables' importance plot, PD's former-leader Matteo Renzi (i.e., "renzi") is negatively associated with the polarisation variable, whilst one would expect the opposite due to the different political beliefs. Second, Europe (i.e., 'Europa') is also negatively associated with the polarisation variable, yet one would expect the opposite due to Salvini's Eurosceptic views. Nevertheless, I am interested in understanding which tweets/replies lead to polarised language despite their positive or negative magnitude. Therefore, despite having a negative association with polarisation, Renzi and Europe are still significant topics to study the spread of polarisation. Third, there is a discrepancy between the words "migrants" and "immigrants". The former is associated with greater levels of polarisation, whereas the latter is associated with lower levels of polarisation. Due to the common Salvini's anti-immigration rhetoric, one would expect that tweets/replies regarding immigration may lead to greater polarisation. This third exception may be due to a failure in the tweets/replies' text tokenisation. In fact, "immigrants" and "migrants" should have been tokenised as "-migra-". Fourth, the 'stopwords' function (i.e., removing the most common Italian words from the text sample) has not performed as wished.

In particular, several Italian articles, prepositions and conjunctions are still present in the sample, such as "the", "if", "not", and "when" (i.e., "il", "la", "se", "non", "quando"). These words may have significantly and negatively impacted the performance of the NN.

In conclusion, despite the numerical evidence, this plot shows that the NN may have performed a useful classification of Salvini's tweets and replies. At the numerical level, the RMSE and R-squared coefficients show that the NN may not be an accurate predictor of polarisation in the tweets/replies sample (i.e., not reaching conventional standards). However, this aspect has been marginally contrasted by the graphical representation of the RMSE and R-squared coefficients across higher levels of regularisation, showing that the NN performance improves with greater regularisation. Moreover, positive evidence comes from the variables' importance plot. The plot has shown that the NN has correctly (based on past events involving Salvini) coded words that referred to political events, parties or people. Furthermore, it showed that mentioning specific events (i.e., immigration) or people (i.e., Renzi) has had an impact on the levels of polarisation in the tweet/reply data-set. Some concerns regarding the NN performance may persist, and I aim to address them during my DPhil through refined textual-classification strategies.

4.3 Data analysis

Having collected and coded Salvini's tweets and replies, this section analyses the new data-set to test the theory-led hypotheses. This endeavour is divided into three complementary steps. First, this section discusses the dependent and independent variables under study. Second, it explores the time-series hypothesis-testing procedure through the estimation strategy. It also explains additional covariates and validation tests to confirm the findings.

4.3.1 Variables

The dependent variable of this thesis is, first, the levels of polarisation in Salvini's tweets, measured as the averaged coefficient of the sentiments of anger and disgust in a given tweet. And second, the levels of polarisation in replies, which are calculated analogously to tweets' polarisation. For the first DV (i.e. tweets' polarisation), the independent variables are replies' levels of polarisation, tweets' topic, replies' topic, and replies' homophilic group-belonging (i.e., Salvini's followers or non-followers). For the second DV (i.e., replies' polarisation), the independent variables are tweets' levels of polarisation, tweets' topic, replies' topic, and repliers' homophilic group-belonging. Table 4.3 summarises the variables under analysis.

Table 4.3: Summary of the variables under analysis

Variable	Description
Salvini's Tweets' polarisation	This variable represents the levels of polarisation in a given Salvini's tweet. It has been measured as the average coefficient of Anger and Fear (as a result of the NN) in a given Salvini's tweet.
Replies' polarisation	This variable represents the levels of polarisation in a given reply to a Salvini's tweet. It has been measured as the average coefficient of Anger and Fear (as a result of the NN) in a given reply to a Salvini's tweet.
Salvini's Followers' replies	This dummy variable represents whether a given reply has been posted by a Salvini's follower (1) or non-follower (0).
Political Tweets	This dummy variable represents whether a given Salvini's tweet concerns political (1) or non-political (0) matters.
Political Replies	This dummy variable represents whether a given reply to a Salvini's tweet concerns a political (1) or non-political (0) matter.

Note: This table describes the variables under investigation.

4.3.2 Estimation strategy

To test the theory-driven hypotheses, this thesis employs OLS regression and time-series analyses through a vector autoregression (VAR) model's Impulse-Repulse Functions (IRFs). This sections sets out the steps that will be followed in the next chapters.

VAR models study vectors of variables as autoregressive processes (Rachev et al., 2012; Box-Steffensmeier et al., 2014). In such models, each variable is correlated linearly with its own past lagged values, and those of the other variables in the vector model (ibidem). Future values of the process are a weighted sum of past and present values, as well as some noise and exogenous variables (ibidem). Therefore, these models allow researchers to simultaneously study different vector variables (e.g., in this case, tweets/replies time-series) across-time, accounting for potential endogeneity. Based on strong theoretical grounds, VAR models, vis-à-vis other time-series analysis (e.g., structural equation approaches), provide several methodological and theoretical advantages. Researchers can obtain granular statistical results, such as simultaneous analyses between time-series, specific lags, and controlling for endogeneity. Structural equation approaches only study individual coefficients, and deduct the model's dynamic properties (Box-Steffensmeier et al., 2014). Theoretically, the VAR models study the empirical implications of several theories through various variables, identifying the causal chain order of simultaneous effects (ibidem). Instead, structural equation approaches rely on single theories for deciding the variables to include in the models (ibidem).

Following Barberà et al. (2019)'s seminal work in one of the latest publications in the *American Political Science Review*, I utilise VAR's IRFs to study Twitter data and test my theory-driven hypotheses. Specifically, through the IRFs, one can study the impulse of a given variable A on another variable B, and the response of variable B in a multivariate model across-time (Lütkepohl, 2005; 2016). Doing so allows me to investigate the impulse response relationship between two variables in the model in a higher dimensional system.

As such, through the IRFs, one can capture how a one-time 10 percentage point increase in polarisation in a given tweet predicts the levels of polarisation in a given reply, and vice-versa (Lütkepohl, 2005; 2016; Barberà et al., 2019b). This measurement allows me to make inferences on the effect of individual shocks on each variable (Box-Steffensmeier et al., 2014). Through IRFs, I study these associations by tracing out the effect of an exogenous shock or innovation in one of the variables on some or all of the other variables (Lütkepohl, 2005; 2016). Moreover, IRFs allow me to understand how previous lags have an impact on the present variables' values (ibidem). Based on IRFs functions, I aim to study short-term impulses that initially activate the use of polarised language.

As suggested by Lütkepohl (2005), these IRFs results will be corroborated through a forecast error variance decomposition (FEVD) to check for robustness. An FEVD determines how much a forecast error variance of a given variable is explained by other variables across lags (Lütkepohl, 2005; Lee and Lee, 2020). In other words, an FEVD shows how much of each variable error variance is explained by exogenous shocks to other variables across-time. To validate my IRFs results, I expect to find low correlation between a single variable's error variance and other variables.

A central issue in VAR modelling is the choice of the number of lags. It is necessary to include enough lags to "capture the full cycle of the data", without omitting significant variations (Box-Steffensmeier et al., 2014; Brandt and Williams, 2007: 25). To do so, I will utilise three main information criteria: Akaike (AIC), Schwartz (BIC), and Hannan and Quinn (HC), which represent the measures of how well the equation system fits the data. The main intuition behind these methods relies on advanced variance-covariance matrices of the system's residuals, which are obtained by fitting the model by an OLS equation (Box-Steffensmeier et al., 2014). Each of the three criteria penalises the fit for the number of variables in the model (ibidem). The AIC is often preferred for small sample sizes because of its small forecasting features. Instead, BIC and HC are optimal for forecasting value as

samples sizes grow across-time. Therefore, for this investigation, I decided to utilise AIC because the sample size, when controlling for dates and times, is lower than 1,500 units. Moreover, this thesis is not fully interested in forecasting features, instead, I am interested in understanding how well the equation system fits the data - and this is what AIC does better than BIC and HC.

To choose the correct number of lags, I will first model the VAR, selecting the variables of interest (i.e., tweets and replies' polarisation, replies' group-belonging, and political content). Subsequently, I run a model order selection function on the VAR to find the number of lags, which is given by the minimum value of AIC. This step allows the model to select the model's lags that fits the data, while minimising the errors (Lutkepohl, 2005).Then, I run the cumulative IRFs to find to understand the intra- and inter-relationship among the variables of interest across-lags. The results of the IRFs will be also checked for robustness at the 95% confidence level.

This thesis is interested in understanding the effects of Salvini's tweets (accounting for its polarisation and discussion topic) on Twitter users and vice-versa, across respondents' homophilic groups and discussion topics. In this paper, polarisation is measured as the averaged anger and disgust coefficients calculated through the NN. First, I study the effects of polarisation in a given Salvini's tweet on replies' polarisation, and vice-versa. Second, I invert the causal arrow, studying the effects of replies' polarisation on tweets' polarisation. In both cases, I account for tweets and replies' topics (i.e., political vs non-political), as well as repliers' homophilic group-belonging (i.e., Salvini's follower vs non-follower). Ultimately, this may clarify whether and how Salvini's tweets are associated with the levels of polarisation in replies, and vice-versa. Moreover, it may aid me to understand under what conditions such cyclical-relationship occurs. Yet, before conducting causal inference analyses, I will run a series of diagnostics to understand the dynamics which underline the structure of the data-sets.

Prior to conducting such time-series diagnostics, I perform some preparation steps. I set a new variable called 'Polarisation', which represents the averaged coefficient of anger and disgust coefficients that tweets and replies display. I have created different subsets based on the author (i.e., Salvini or a respondent) and the content (i.e., political vs non-political). Finally, I have normalised the data: for each data-point, I subtracted the mean and divided by the standard error of their respective columns). This allows to plot and study different types of data (i.e., numbers of favourites, emotional coefficients, etc.) both comparatively and individually. Following the preparation steps, I have plotted the data, and graphically compare the time-series under analysis.

The first diagnostic to perform is to check for stationarity. A stationary time-series has a mean of zero, and a constant standard deviation (Box-Steffensmeier et al., 2014). Instead, a time-series is defined as non-stationary, or integrated, when it has an upward or downward trend. Integrated series are often referred to as 'unit root' as they increase and decrease over-time, but they never revert to their mean value (ibidem). These trends indicate that, within the time-series, there are shocks that accumulate over-time and have not left the system. Instead, a stationary time-series has a constant mean and variance over-time, where the effect of exogenous shocks are short-lived. In other words, stationary time series display spikes, but these revert to an equilibrium value. It is imperative to check the nature of the time-series because if a time-series is non-stationary two main issues can arise (ibidem). Non-stationary data can bias the estimates as past values of the dependent variable may influence the later values. In this way, non-stationary time-series can lead to spuriousness in the results (i.e., inflated R-squared and t-statistics). For example, when regressing one non-stationary series onto another, one may obtain statistically significant (i.e, regression's t-statistics are significant and R-squared coefficient is high) estimates, yet the relationship under analysis may be nonsensical. However, such results may be endogenous to non-consistent due to the trending nature of the non-stationary time-series.

To perform the first diagnostic (i.e., to check for stationarity), I conduct two tests: the Augmented Dickey-Fuller (ADF) and the Kwiatkowski, Perron, Schmidt, and Shin's (1992; KPSS) tests. The intuition at the basis of both the tests is that in a non-stationary series, the value of the series today does not help predict the value of the series tomorrow, whereas the opposite is true for a stationary series. Considering the equation: $Y_t = \alpha + _{t-1}$, I check whether the value of β is greater or equal to 0. If it is true, then the series Y is stationary. Otherwise, the series was non-stationary. In other words, if knowing the past value of the series helps predict the next period's change, there will be a negative coefficient in the equation. The ADF assumes the absence of a unit root as the null hypothesis (i.e. the series is non-stationary). On the contrary, the KPSS (henceforth) test adopts the absence of a unit-root as the null hypothesis (i.e., the series is stationary).

Having defined the two-stationarity test, it is important to discuss how to evaluate them. Regarding the AFD test, if the p-value is greater than 0.05, then there is weak evidence to reject the null. That is, the series is non-stationary. If the p-value is lower than 0.05, there is evidence to accept the null, meaning that the series is stationary. On the contrary, regarding the KPSS test, if the p-value is greater than 0.05, there is weak evidence to reject the null, meaning that the series is stationary. If the p-value is lower than 0.05, there is evidence to accept the null. That is, the time-series is non-Stationary. Now that I have discussed both tests individually, I discuss their interaction in four cases.

- If both tests conclude that the time-series is not stationary, then one can assume that the series is not stationary.
- If both tests conclude that the series is stationary, one can assume that the time-series is stationary.
- If the KPSS test indicates stationarity, while the ADF test indicates non-stationarity, the series is defined as trend stationary. Such a trend needs to be removed to make

series strictly stationary. Following the de-trend of the series, a re-check for stationarity is needed.

- If the KPSS test indicates non-stationarity, while the ADF indicates stationarity, the series is defined as difference stationary. In such an instance, one needs to differentiate the series to make it stationary. Following the differentiation, the new time-series is re-checked for stationarity.

As in the third and fourth cases, non-stationarity may lead to making false inferences based on spurious results (i.e., inflated R-squared and t-statistics). By removing non-stationarity, and, in later stages checking for non-autocorrelation, one can ensure that they are left with the effect driving the relationship under analysis (Box-Steffensmeier et al., 2014). To do so, I first-differentiate the variables of interest by subtracting the value of Y at time $t-1$ from its value at t. Finally, I re-check for stationarity of the new time-series via ADF and KPSS tests. If one or both time series were still non-stationary, I re-differentiate them, and re-conduct the stationary checks.

The second diagnostic to perform is an autocorrelation check (i.e., estimating whether the time series is correlated with itself at previous lags). To do so, the most prominent tool utilised to study such a property of the time-series is the correlogram. The correlogram is a graph that displays two statistics, the autocorrelation function (ACF) and the partial autocorrelation function (PACF). Such a graph shows the value of various lags of the ACF and PACF statistics, as well as the corresponding standard errors (or, error bands in the graph). This graphical representation allows us to visually inspect the series for autocorrelation, estimating whether the time-series coefficients fall within the critical region (or, estimates' standard errors). If the coefficients do fall in the error bands, meaning that earlier coefficients are not statistically significant in determining later values, one can assume that the time-series do not display autocorrelation. Thus, the time-series are not correlated

with themselves at different lags.

The ACF statistic allows to measure the correlation between Y_t and Y_{t+k}, where k represents the number of lead periods into the future. In other words, the length of the series increases when k increases, yet the confidence in ACF decreases. Despite being a powerful analytical tool, McLeary and Hay (1980) argue that it is often difficult to determine the properties of a series based only on the correlogram of an ACF when considering large orders of integration. To account for this eventuality, I will utilise the PACF(k). Such a statistic is a measure of the correlation between times-series observations that are k units apart, after controlling for (or "partialed" out) the correlation at intermediate lags. In other words, the PACF(k) measures the correlation between Y_t and Y_{t+k}, following the removal of the intermediate Y coefficient effects.

When stationarity and autocorrelation tests show that I am operating with stationary and non-autocorrelated time-series, I can perform OLS regression, coefficient plots, and VAR model analyses as previously discussed. These analyses will aid me to accept or reject the theory-driven hypotheses.

Given the limited space for the OLS and VAR equations and results output, I have adopted abbreviations for the variables' names. Such abbreviations will allowed me to display the full estimation equations and VAR's IRFs results in a single page, and study its results without overcrowding the output window. Table 4.4 summarises the abbreviations.

Table 4.4: Summary of the variables' abbreviations

Abbreviation	Variable	Description
tweets	Salvini's Tweets' polarisation	This variable represents the levels of polarisation in a given Salvini's tweet. It has been measured as the average coefficient of Anger and Fear (as a result of the NN) in a given Salvini's tweet.
replies	Replies' polarisation	This variable represents the levels of polarisation in a given reply to a Salvini's tweet. It has been measured as the average coefficient of Anger and Fear (as a result of the NN) in a given reply to a Salvini's tweet.
follow	Salvini's Followers' replies	This dummy variable represents whether a given reply has been posted by a Salvini's follower (1) or non-follower (0).
tpol	Political Tweets	This dummy variable represents whether a given Salvini's tweet concerns political (1) or non-political (0) matters.
rpol	Political Replies	This dummy variable represents whether a given reply to a Salvini's tweet concerns a political (1) or non-political (0) matter.

Note: This table summarises and discusses the variables' abbreviations which will be utilised in the VAR output.

For the OLS regression analyses, I set two multivariate linear regression equations. The first equation seeks to study the association between tweets' polarisation (DV) and lagged replies' polarisation, tweets being political, lagged replies being political, and lagged replies' homophilic-group (explanatory variables). Similarly, the second equation seeks to study the association between replies' polarisation (DV), and lagged tweets' polarisation, lagged tweets being political, replies being political, and replies' homophilic group (explanatory variables). Formally put, Equations 1 and 2:

$$Y(tweets)_t = a_0 + \beta_1(replies)_{t-1} + \beta_2(follow)_{t-1} + \beta_3(rpol)_{t-1} + \beta_4(tpol)_t + \varepsilon \quad (1)$$

$$Y\,(replies)_t = a_0 + \beta_1(tweets)_{t-1} + \beta_2(follow)_t + \beta_3(rpol)_t + \beta_4(tpol)_{t-1} + \varepsilon \quad (2)$$

To corroborate the OLS model results, following Barberà et al. (2015) and Theocharis et al. (2020), I studied individual topics within the overall political and non-political discussions. To do so, I produced coefficient plots to study the effects of the dummy variables previously created for each sub-topic within the topic samples (i.e., political vs non-political samples), on the tweets and replies' polarisation coefficients, accounting for group-belonging and discussion topics.

For the VAR model, I set a stationary time series $Y_{i,j,t}$ that represents the levels of polarisation in each Salvini's tweet (j), and reply to the tweet (i), in a given day (t) in the timespan. The VAR model accounts for previous lags of one variable in addition to the lags of other variables. Formally put, Equation 3:

$$Y_{i,j,t} = \alpha_{i,j} + \sum_{i,j} \beta_{i,j} Y_{i,j,t} + \varepsilon_{i,j,t} \quad (3)$$

Where: i encapsulate replies' polarisation coefficient, topic, and homophilic group; j represents tweets' polarisation and topic; t accounts for time variation; a represents the topic-fixed effect, assuming that the estimate is constant across topics, controlling for any time-invariant factors specific to topics that could confound unbiased estimates; $\varepsilon_{i,j,t}$ accounts for the error of the model.

CHAPTER 5

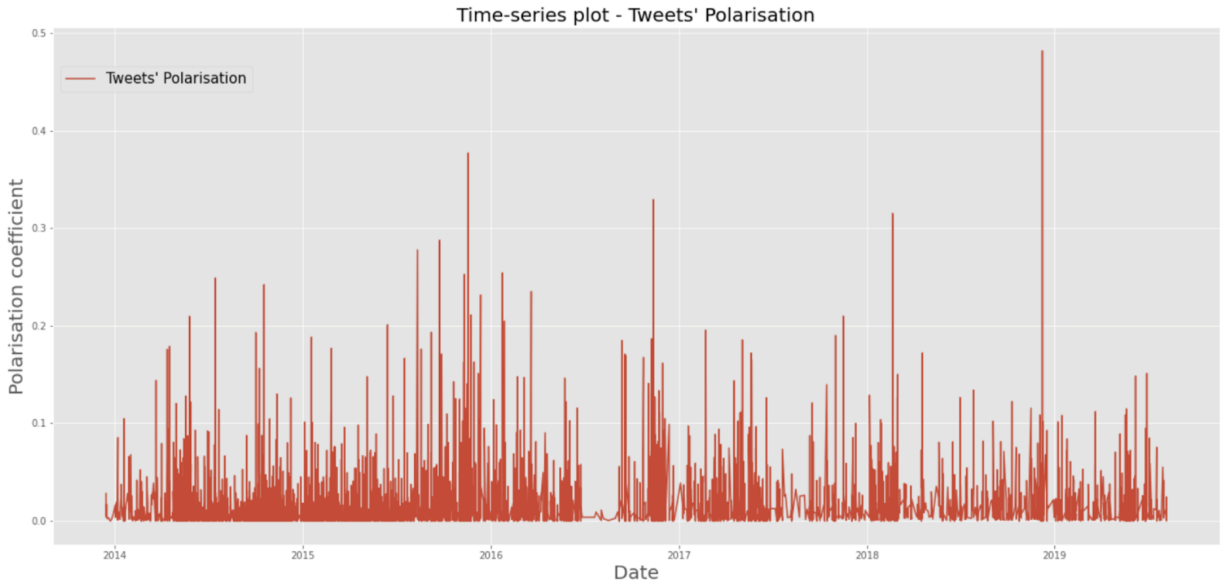
TIME-SERIES DIAGNOSTICS

Following the time-series preparation steps discussed in the methodology chapter, I conduct two diagnostics (i.e., stationary and autocorrelation tests) for the two time-series under analysis. In particular, I study the time-series, representing the levels of polarisation in Salvini's tweets, as well as the levels of polarisation in replies to Salvini's tweets. The outcomes of these two diagnostics will be utilised in the following OLS and VAR models to test the theory-driven hypotheses.

5.1 Polarisation in Salvini's tweets

Having plotted the time-series representing the polarisation of Salvini's tweets, there is no clear trend across-time (see Figure 5.1). Therefore, to check for stationarity, I utilise a first combination of ADF and KPSS stationarity tests (see Appendix F). The ADF shows that the series is stationary (p-value < 0.05), rejecting the null-hypothesis (i.e., non-stationarity). Instead, the output of the KPSS test displays that the tweets' polarisation time-series is non-stationary (p-value > 0.05), failing to reject the null (i.e., stationarity). Therefore, I can conclude that the series is difference stationary. As such, I need to differentiate the time-series to remove the unit root (or, upward trend) before moving to the next diagnostic.

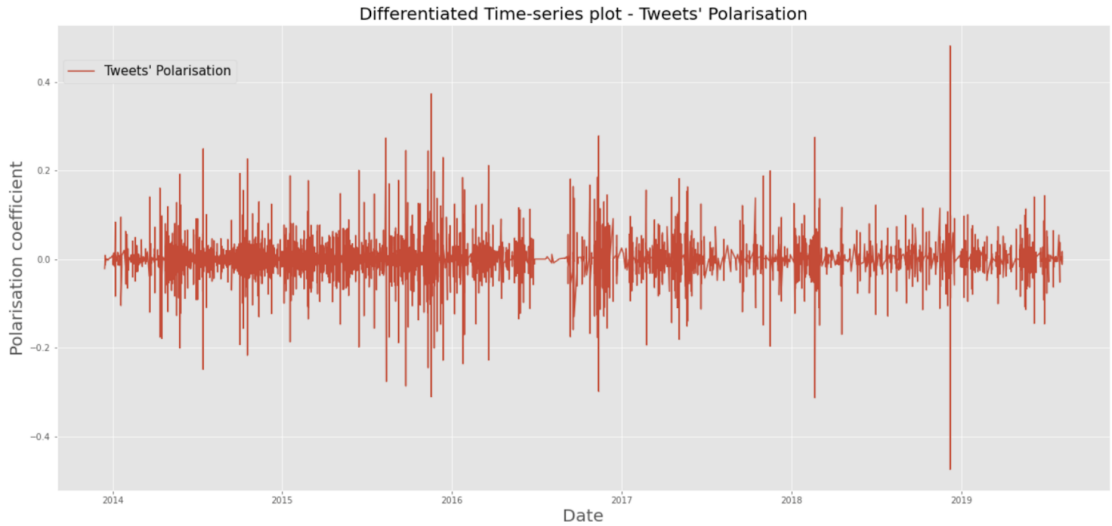
Figure 5.1: Time-series representing Tweets' Polarisation (2013-2019)



Note: This figure displays Salvini's tweets' polarisation time-series, which represents Salvini's polarised language across the timespan.

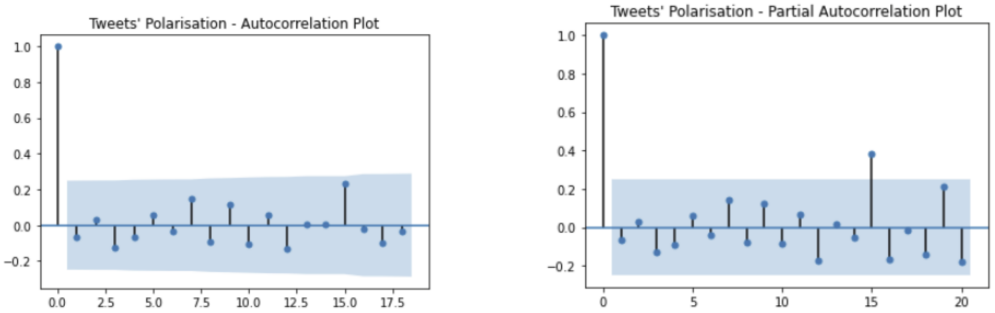
Following the differentiation, I plot the new differentiated time-series representing the tweet's polarisation. I re-conduct stationarity tests, which reveal the stationary nature of the differentiated time-series. As such, both ADF and KPSS yield statistically significant results to assume that the differentiated time-series is stationary, respectively with p-value < 0.05 (rejecting the null: non-stationarity) and p-value > 0.05 (accepting the null: stationarity). Then, I move to the next diagnostic: autocorrelation. To do so, in Figure 5.3, I run Autocorrelation and Partial-Autocorrelation tests, checking whether the time-series coefficients fall within the critical region, which represents the standard errors. If the coefficients do fall in the error bands, meaning that earlier coefficients are not statistically significant in determining later values, one can assume that the time-series do not display autocorrelation. For the differentiated tweets' polarisation time-series, both autocorrelation and partial-autocorrelation correlograms show that the time-series does not suffer from autocorrelation. Therefore, the tweets' polarisation time-series is ready for modelling.

Figure 5.2: Differentiated Time-series representing Tweets' Polarisation (2013-2019)



Note: This figure displays the differentiated Salvini's tweets' polarisation time-series, which represents Salvini's polarised language across the timespan.

Figure 5.3: Autocorrelation tests for Salvini's tweets' polarisation time-series



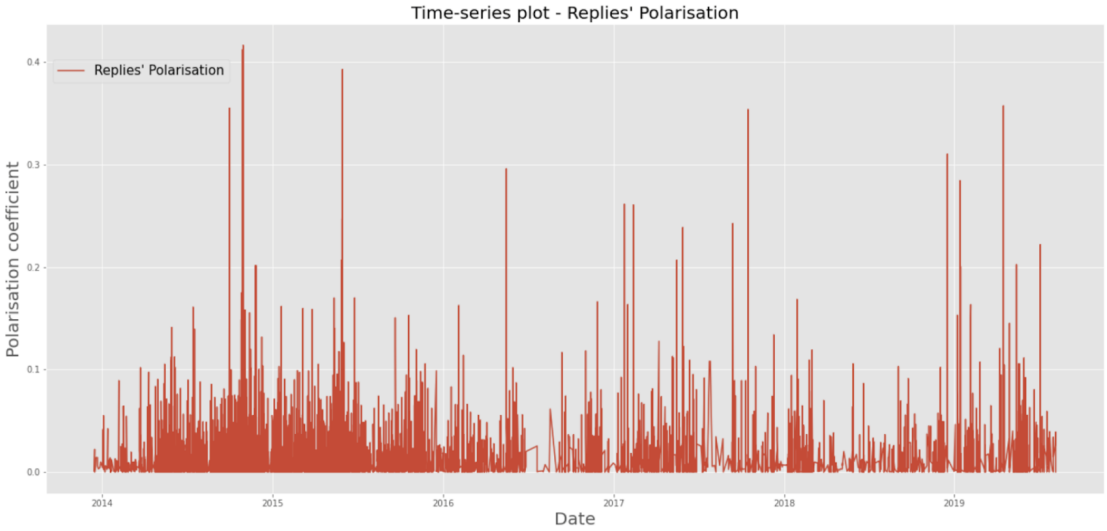
Note: These figures represent the correlograms for, respectively, Salvini's tweets' polarisation time-series autocorrelation (left figure) and partial-autocorrelation (right figure). Besides a random spike at lag 25 (in the partial-autocorrelation plot), all the estimates fall in the critical region, meaning that there is no autocorrelation.

5.2 Polarisation in replies to Salvini's tweets

As in the previous section, I now study the time-series representing the levels of polarisation in replies to Salvini's tweets. Plotting the time-series, there is no clear unit root (up/down-

trend) across-time. Thus, to check for stationarity, I conduct ADF and KPSS stationarity tests (see Appendix G). The ADF test reveals that the time-series is stationary (p-value < 0.05), rejecting the null hypothesis (i.e., non-stationarity). Instead, the KPSS test shows that the time-series is non-stationary (p-value < 0.05), rejecting the null (i.e., stationarity). Following the tests, the time-series is difference stationary. Hence, it is important to de-trend it before moving to the next diagnostic.

Figure 5.4: Time-series representing Replies' Polarisation (2013-2019)

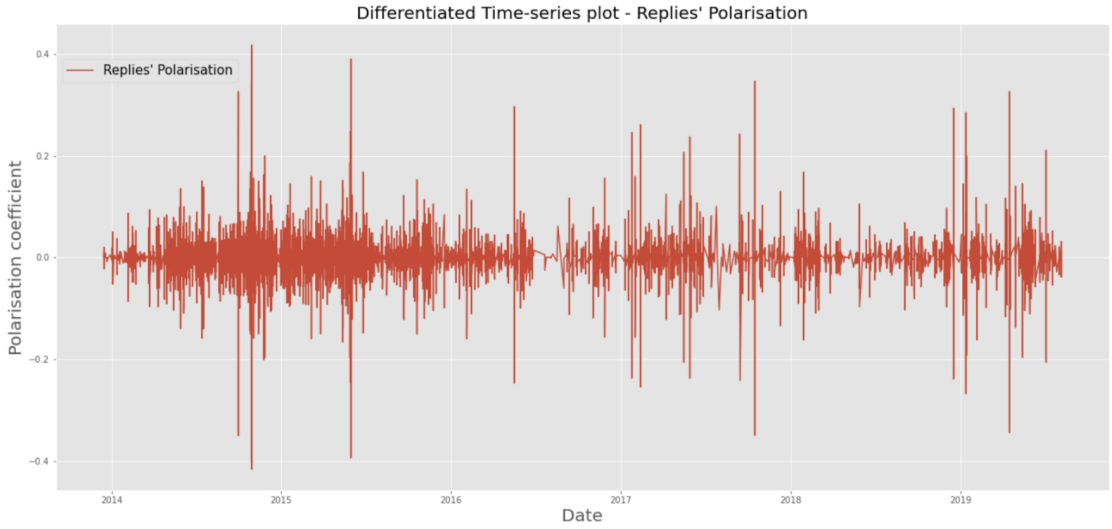


Note: This figure displays the replies' polarisation time-series, which represents Salvini's respondents' polarised language (after differentiation) across the timespan.

Following the differentiation, I plot the new differentiated time-series that represents the levels of replies' polarisation across-time. I re-conduct stationary tests, which display that the differentiated time-series is now stationary. In particular, both ADF and KPSS tests yield statistically significant results to assume that the differentiated replies' polarisation time-series is stationary, respectively with p-value < 0.05 and p-value > 0.05. Therefore, I can now move to the second diagnostic: autocorrelation. As per the tweets' polarisation time-series, I conduct both autocorrelation and partial-autocorrelation tests. Both correlograms

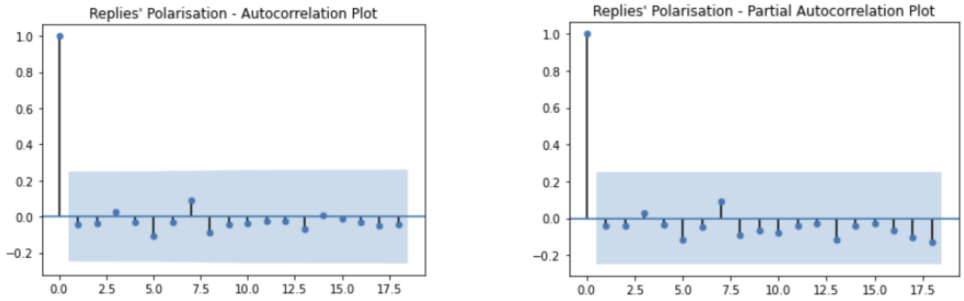
show that the differentiated time-series does not suffer from autocorrelation. For these reasons, the differentiated replies' polarisation time-series is ready for modelling.

Figure 5.5: Differentiated Time-series representing Replies' Polarisation (2013-2019)



Note: This figure displays the differentiated replies' polarisation time-series, which represents Salvini's respondents' polarised language (after differentiation) across the timespan.

Figure 5.6: Autocorrelation tests for respondents' polarisation time-series



Note: These figures represent the correlograms for, respectively, replies' polarisation time-series autocorrelation (left figure) and partial-autocorrelation (right figure). All the coefficients fall in the critical region, meaning that there is no autocorrelation.

CHAPTER 6

RESULTS

Having performed the two time-series diagnostics (stationarity and autocorrelation), this thesis can now move on to the OLS and VAR analyses. These analytical steps will allow me to study the individual theory-driven hypotheses, which will be then accepted or rejected in the discussion chapter. Before doing so, I explore the descriptive statistics of the data-set.

6.1 Descriptive Statistics

Table 6.1 provides a brief summary of the data-gathering process. I retrieved 10,826 Salvini's unique tweets, of which 7,217 (67% of the total tweets) were political. Moreover, I retrieved 12,864 unique replies to Salvini's tweets, of which 9,657 (75%) were political, and 3,176 (roughly 25%) were Salvini's followers' replies. When considering the mean of the polarisation coefficients, Salvini's tweets have the lowest average of polarisation (0.011) then, ascendantly, political replies (0.013), political tweets (0.014), overall replies (0.023), and followers' replies (0.032). Finally, the low standard deviations in tweets' polarisation (0.019), political tweets (0.017), replies' polarisation (0.015), political replies (0.017), and followers' replies (0.021) display that the average deviation from the mean in all variables is not substantial.

Table 6.1: Summary of the descriptive statistics

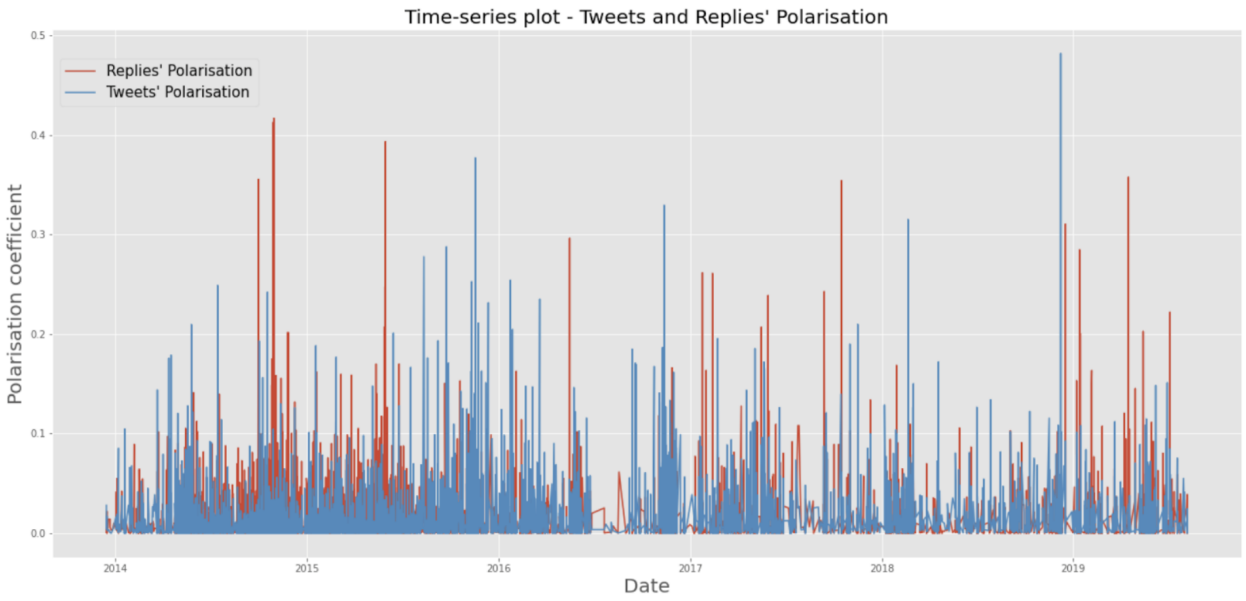
Type	N	Polarisation Mean	Polarisation Standard Deviation
Total Unique Salvini's tweets	10,826	0.011	0.019
Political tweets	7,217	0.014	0.017
Total Unique replies to Salvini's tweets	12,864	0.023	0.015
Political replies	9,657	0.013	0.017
Followers' replies	3,176	0.032	0.021

Note: This table summarises the descriptive statistics. It displays the number of retrieved tweets or replies (N), the mean of the polarisation coefficients (i.e., the averaged coefficients of the sentiments of anger and disgust in a given tweet/reply, ranging from 0 to 1) mean, the standard deviation of the polarisation coefficients.

The Figure 6.1 represents the time-series accounting for the daily polarisation in both Salvini's tweets (in blue) and replies to his tweets (in red) between December 2013 and August 2019. The graph displays that polarised language has been widely and consistently utilised by both Salvini and his respondents between December 2013 and May 2016. In this period, replies' polarisation seems to be greater than Salvini's levels of polarisation. However, between May 2016 and early-2017, there is a gap in the use of AP language. Such a gap may be due to the intensified Twitter's data-protection algorithms created before the 2016 US elections to prevent the misuse of Twitter-data (Twitter, 2016). Following 2017, there is a lower use of polarised language, which reaches its lowest intensity in late 2017 and

early 2018 prior to the Italian general elections. After the 2018 elections, when Salvini has been appointed Deputy Prime Minister, polarisation decreases dramatically in both tweets and replies. In late 2019, which corresponds to the period in which Salvini's government broke down, polarisation levels rise again substantially.

Figure 6.1: Combined Time-series representing polarisation in Salvini's tweets and replies to his tweets



Note: This figure displays the combined time-series, representing the levels of polarisation in both Salvini's tweets and replies to his tweets.

6.2 OLS models and Coefficient Plots results

I compute two equations and I evaluate them through OLS models in R-software, first for Salvini's tweets' polarisation, and second for replies' polarisation.

6.2.1 Tweets' polarisation

In this section, I test whether and how replies' polarisation has an effect on tweets' polarisation. To do so, I set the levels of polarisation in Salvini's tweets as the dependent variable, whilst setting lagged replies' polarisation, lagged respondents' group-belonging (i.e., Salvini's followers), lagged political replies, and political tweets as the independent variables.

Figure 6.2: Levels of Polarisation in Salvini's Tweets (2013-2019)

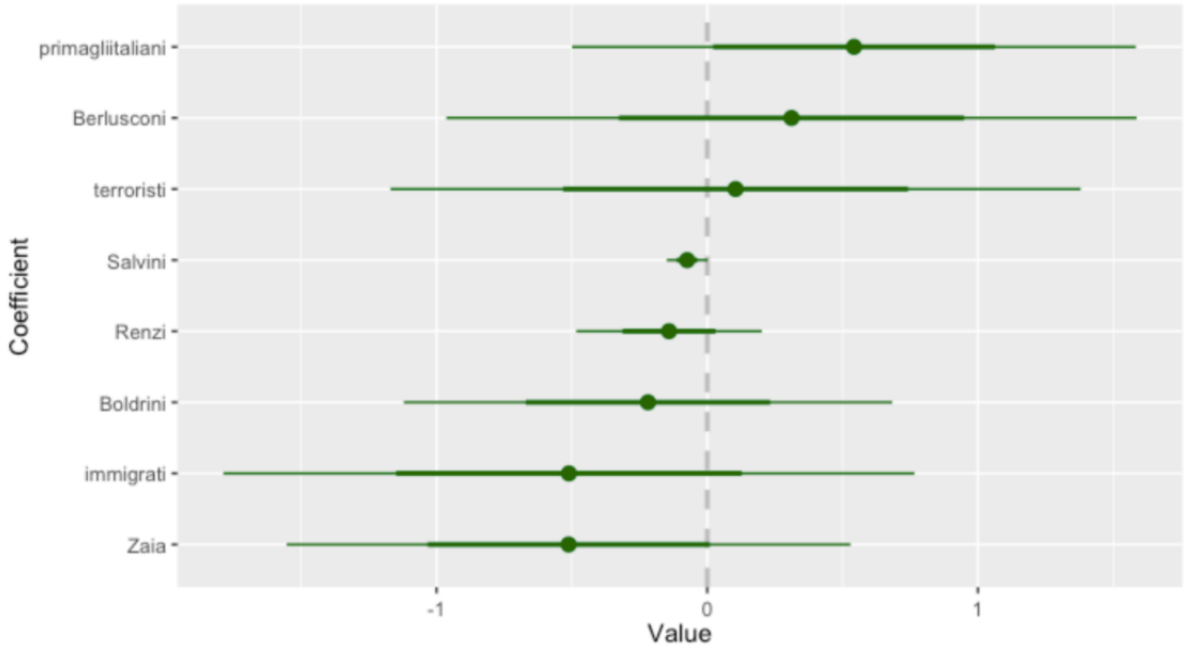
	Dependent variable:				
	Tweets' polarisation				
	(1)	(2)	(3)	(4)	(5)
Lagged Replies' Polarisation	0.021**	0.021**	0.021**	0.023**	0.023***
	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)
Lagged Followers' Replies		−0.030			−0.030
		(0.020)			(0.020)
Lagged Political Replies			0.002	0.004	0.004
			(0.041)	(0.041)	(0.041)
Political Tweets				−0.111***	−0.110***
				(0.039)	(0.039)
Constant	0.000	0.007	−0.002	0.100*	0.108**
	(0.009)	(0.010)	(0.040)	(0.054)	(0.054)
Observations	12,864	12,864	12,864	12,864	12,864
R^2	0.0004	0.001	0.0004	0.001	0.001
Note:	*p<0.1; **p<0.05; ***p<0.01				
	Standard errors in parentheses.				

Note: This figure displays the Equation 1 OLS output (computed in R-studio) to study the effect of (1) lagged replies' polarisation, (2) lagged followers' replies, (3) lagged political replies, (4) political tweets, on tweets' polarisation.

Based on Equation 1, I run five OLS models. Starting from the effect of lagged replies' polarisation on tweets' polarisation, I add an extra explanatory variable in the successive model. The results are presented in Figure 6.2. Two prominent findings emerge. On the one hand, Figure 6.2 shows that one unit increase in the variable accounting for the levels of lagged polarisation in replies is positively (yet mildly) associated with the levels of polarisation in Salvini's tweets. Such a finding is also statistically significant at the conventional levels (p-value < 0.05). When controlling for tweets and lagged replies' content, as well as lagged followers' replies, such relationship remains significant at the conventional standards. Indeed, across the five models under analysis, the effect of lagged replies' polarisation on tweets' polarisation ranges from 0.021 and 0.023. On the other hand, Figure 6.2 shows that one unit increase in the variable accounting for tweets' content being political is strongly negatively associated (-0.110) with the levels of polarisation in Salvini's tweets. In particular, such a relationship is statistically significant at the conventional levels (p-value < 0.05). This negative relationship remains robust in magnitude and statistical significance after adding lagged replies' polarisation and political topic as covariates. That is, on average, the levels of polarisation in Salvini's tweets tends to decrease when the tweets concern political content, vis-à-vis non-political content (all others being equal). For these reasons, from Equation 1, only lagged replies' polarisation and political content (in tweets) seem to be significant predictors of tweets' polarisation. Respectively, lagged replies' polarisation has a positive (yet mild) effect, and political tweets have a substantial negative effect on Salvini's tweets' polarisation.

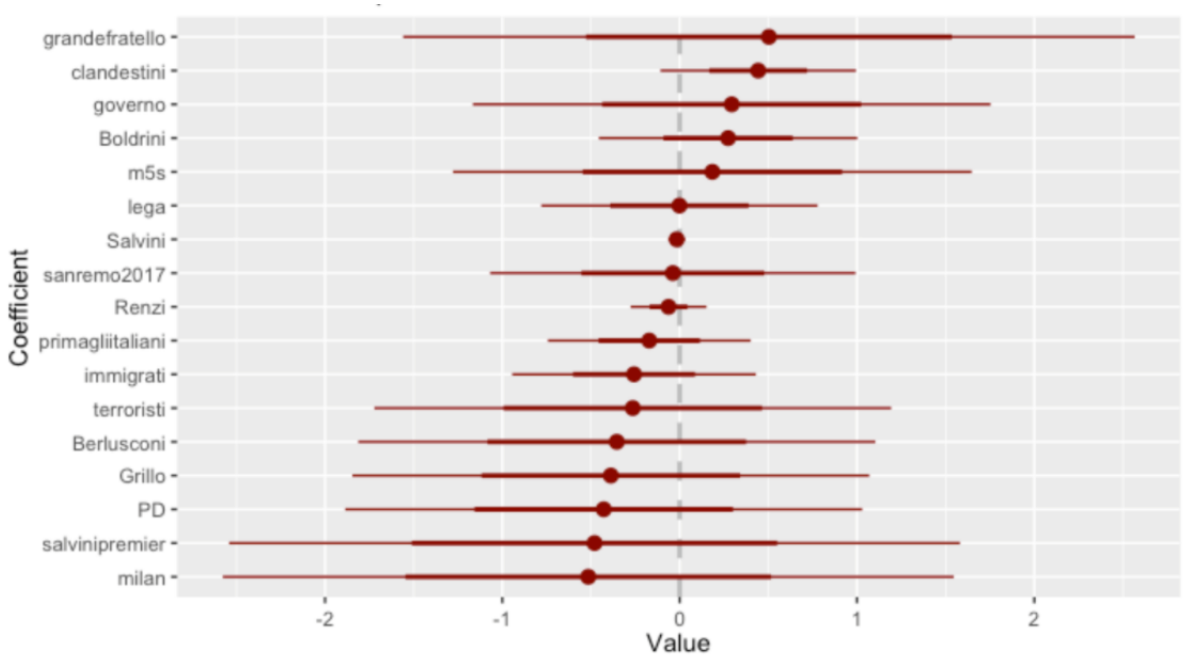
I turn to two coefficient plots to study the effect of individual topics across political and non-political content, among follower and non-follower replies, on the tweets' levels of polarisation. Coefficient plots provide more granular results into the individual topics that drive polarised language. First, Figure 6.3 shows that none of the topics, except for 'Salvini', tend to have a statistically significant effect on the DV under analysis at the 95% confidence level. Instead, when the tweets' content concerns 'Salvini', such a topic is statistically significant (p-value < 0.05) and negatively associated (yet mildly) with tweets' levels of polarisation. Second, when I turn to Figure 6.4, accounting for Salvini's lagged follower respondents, none of the topics across both political and non-political content tend to have a statistically significant effect on the levels of polarisation in Salvini's tweets, at the 95% confidence levels (p-value> 0.05). That is, on average, Salvini's tweets regarding himself tend to be less polarising vis-à-vis other topics (all others being equal). This finding is in accordance with the previous section results, which showed that the levels of polarisation in Salvini's tweets did not seem to be positively associated with political content. I now find that, when I account for the specific content of the tweets, such as 'Salvini' (seen as a political topic), tweets' polarisation slightly decreases. This means that polarisation may not flow from Salvini to his respondents when Salvini becomes the discussion topic. However, as previously seen, such a topic is only mildly associated with greater tweets' polarisation. More evidence is needed to corroborate such results.

Figure 6.3: Coefficient plot representing polarisation in Salvini's tweets, depending on sub-topics and follower repliers



Note: This figure shows a coefficient plot that displays the point estimate (for the effect of each sub-topic on Salvini's tweets' polarisation, contingent on followers' replies) and the 95% confidence intervals. The sub-topics whose confidence lines cross zero are not statistically significant in predicting polarisation in Salvini's tweets.

Figure 6.4: Coefficient plot representing polarisation in Salvini's tweets, depending on sub-topics and non-follower repliers



Note: This figure shows a coefficient plot that displays the point estimate (for the effect of each sub-topic on Salvini's tweets' polarisation, contingent on non-followers' replies) and the 95% confidence intervals. The sub-topics whose confidence lines cross zero are not statistically significant in predicting polarisation in Salvini's tweets.

6.2.2 Replies' polarisation

To test the bi-directional dynamics of polarisation, I reverse the causal arrow. Doing so I study the relationship between replies' polarisation (DV), and lagged tweets' polarisation, respondents' group-belonging (i.e., Salvini's followers), political replies, and lagged political tweets (independent variables).

Figure 6.5: Levels of polarisation in replies to Salvini's Tweets (2013-2019)

	Dependent variable:				
	Replies' polarisation				
	(1)	(2)	(3)	(4)	(5)
Lagged Tweets' Polarisation	0.021**	0.021**	0.021**	0.022**	0.023***
	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)
Followers' Replies		0.066***			0.065***
		(0.020)			(0.020)
Political Replies			0.245***	0.236***	0.237***
			(0.040)	(0.040)	(0.040)
Lagged Political Tweets				0.353***	0.351***
				(0.039)	(0.039)
Constant	0.000	−0.016	−0.233***	−0.558***	−0.573***
	(0.009)	(0.010)	(0.039)	(0.053)	(0.054)
Observations	12,864	12,864	12,864	12,864	12,864
R^2	0.0004	0.001	0.003	0.010	0.010
Note:	*p<0.1; **p<0.05; ***p<0.01 Standard errors in parentheses				

Note: This figure displays the Equation 2 OLS output (computed in R-studio) to study the effect of (1) lagged tweets' polarisation, (2) followers' replies, (3) political replies, (4) lagged political tweets, on polarisation in replies to Salvini's tweets.

Based on Equation 2, I run five OLS models. The results are presented in Figure 6.5. The

models yielded two main trends. First, one unit increase in the levels of polarisation in lagged Salvini tweets and being a Salvini's follower have a mild positive effect on replies' polarisation, which is statistically significant at the conventional standards (p-value < 0.05). In particular, when considering the variable accounting for polarisation in lagged Salvini's tweets, it has a mild positive effect on replies' polarisation, ranging from 0.021 and 0.023 across the five models. That is, on average, the levels of polarisation in Salvini's tweets have little positive effect on replies' levels of polarisation. Similarly, on average, the variable accounting for replies tweeted by Salvini's followers is positively associated with the levels of polarisation in replies. This means that replies written by Salvini's followers are likely to be 0.065 more polarised than those tweeted by non-followers.

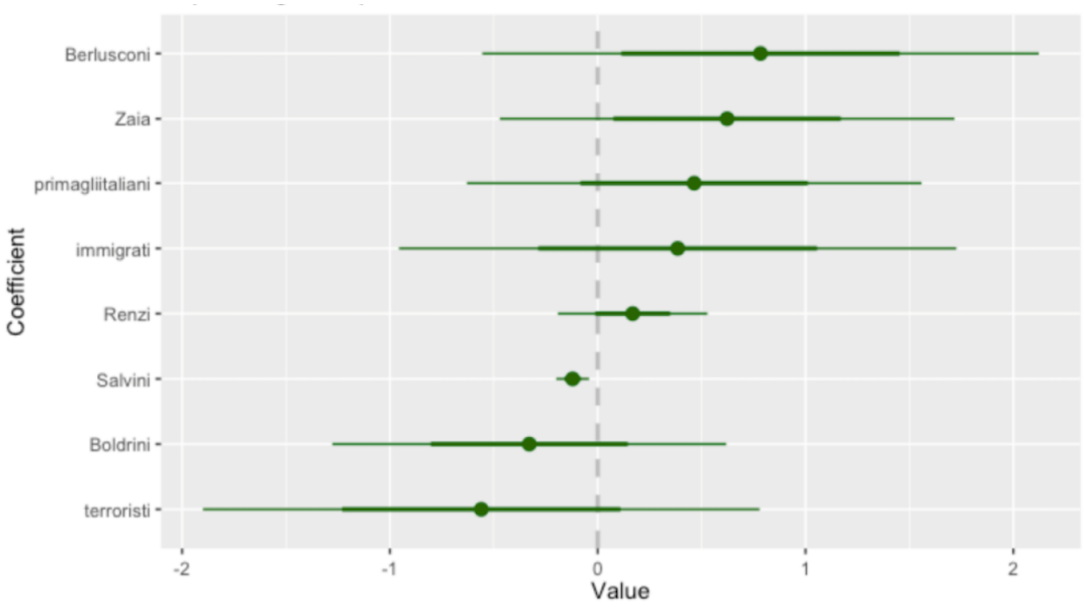
Second, when accounting for tweets and replies concerning a political content, such a content tends to increase the levels of polarisation in replies. As such, one unit increase in replies' political content is positively associated with replies' polarisation, which is statistically significant at the conventional standards (p-value < 0.05). Specifically, when replies contain political content, such replies are, on average, 0.239 more polarised vis-à-vis non-political replies (all others being equal). Moreover, when including other covariates (e.g., lagged tweets' polarisation, lagged tweets' content, and group-belonging), political replies remain a strong predictor of polarisation in replies. Finally, accounting for the political content in lagged Salvini's tweets. As such, on average, one unit increase in lagged political tweets is more strongly and positively associated (0.352) with levels of polarisation in replies than non-political tweets (all others being equal). That is, political content in Salvini's tweets tends to be the most important predictor (0.352) of polarisation in replies, when compared to political replies (0.239), being a Salvini's follower (0.065), and tweets' polarisation (0.022).

To corroborate the OLS results, I zoom into the topics within the political versus non-political content in Salvini's tweets, across followers and non-followers. First, Figure 6.6 shows that, for Salvini's followers, none of the topics is substantively associated with the dependent

variable (i.e., levels of polarisation in replies) except for 'Salvini'. In particular, the coefficient plot shows that all the confidence intervals (except for Salvini) of the magnitude for each topic's effect on polarisation contained the value 0, at the 95% confidence level. That is, there is no enough evidence against the null hypothesis (i.e., covariates do not have effect on the DV) at the conventional levels. Instead, the topic Salvini tends to be negatively associated with higher levels of polarisation in replies (p-value < 0.05).

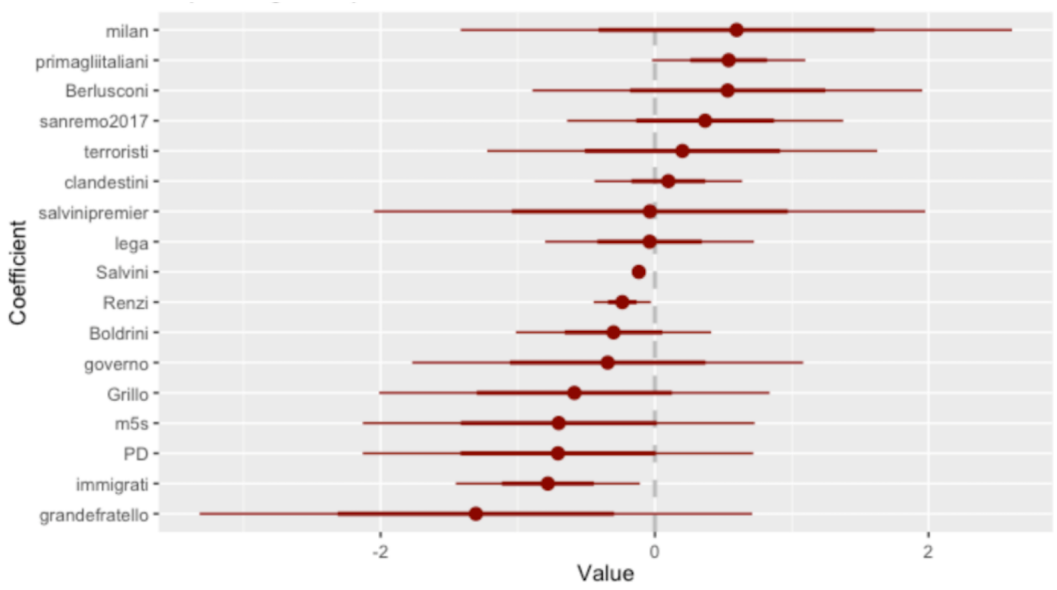
Second, Figure 6.7 shows that for Salvini's non-followers talking about 'Renzi' (i.e., Matteo Renzi - former Italian PM), 'Salvini' and 'Immigrati' (i.e., Immigrants) is negatively associated with the levels of polarisation in replies at the 95% confidence interval. Instead, all the other topics do not tend to have a statistically significant effect on the DV as all the confidence interval crosses zero at the conventional levels. Thus, from both coefficient plots, one can assume that 'Salvini', 'Renzi', and 'Immigrants', are significant topics in predicting polarisation in replies. This result may provide evidence for polarised language not travelling across political topics. Indeed, Renzi and immigrants have often been represented as political opponents in Salvini's tweets, yet there is lower polarisation in non-followers' replies. Similarly, the topic 'Salvini' is also a significant predictor of lower polarisation for both followers and non-followers.

Figure 6.6: Coefficient plot representing polarisation in replies to Salvini's tweets, depending on sub-topics and follower repliers.



Note: This figure shows a coefficient plot that displays the point estimate (for the effect of each sub-topic on respondents' polarisation, contingent on followers' replies) and the 95% confidence intervals. The sub-topics whose confidence lines cross zero are not statistically significant in predicting polarisation in replies to Salvini's tweets.

Figure 6.7: Coefficient plot representing polarisation in replies to Salvini's tweets, depending on sub-topics and non-follower repliers.



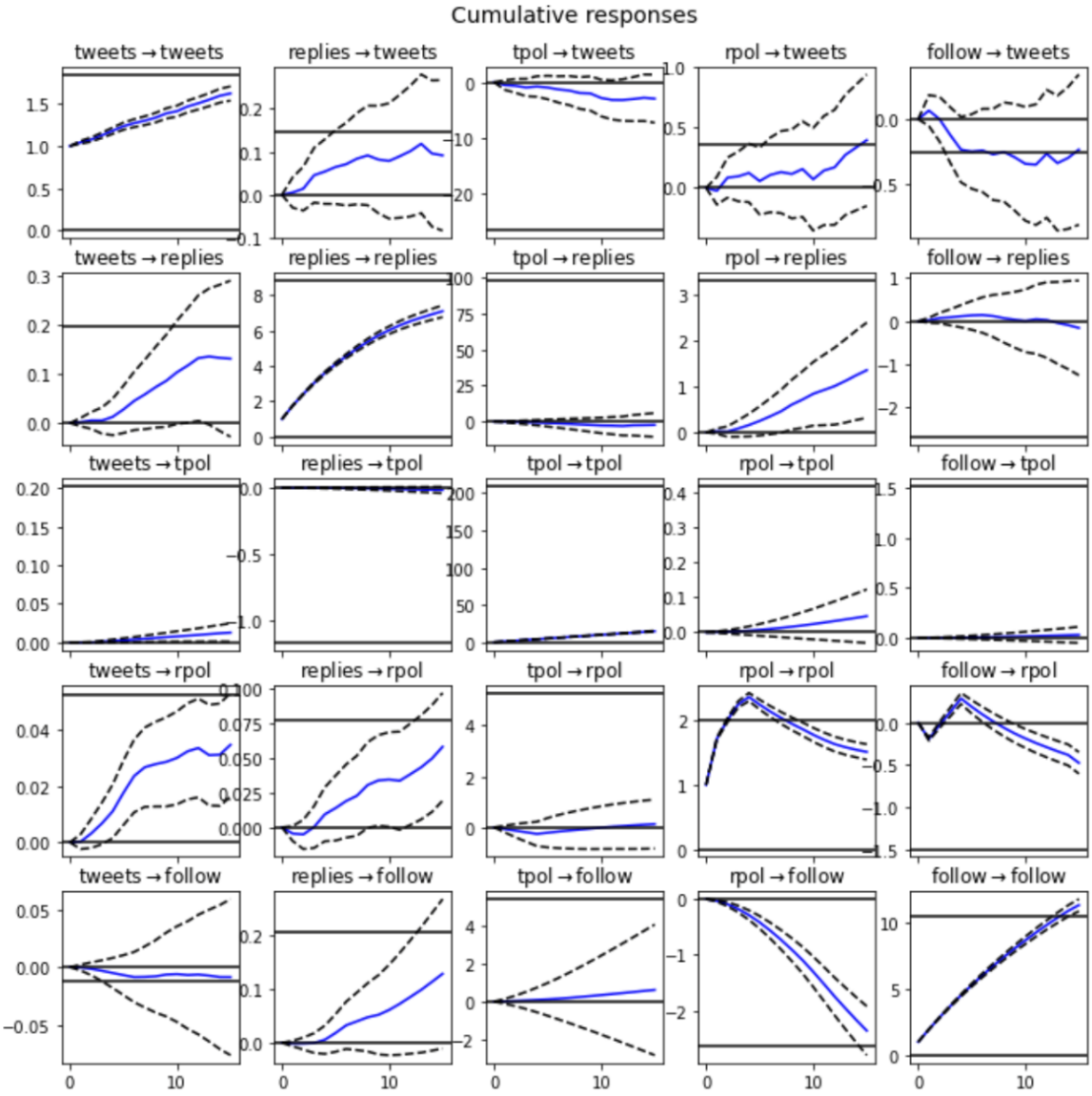
Note: This figure shows a coefficient plot that displays the point estimate (for the effect of each sub-topic on respondents' polarisation, contingent on non-followers' replies) and the 95% confidence intervals. The sub-topics whose confidence lines cross zero are not statistically significant in predicting polarisation in replies to Salvini's tweets.

6.3 VAR results

Having discussed the results of the OLS models, the following section explores the VAR model results to further understand the spread of polarised language from Salvini to his respondents, and vice-versa. As previously introduced, the VAR model allows to study the impulse of a given variable A on another variable B, and the response of variable B in a multivariate model across-time (Lütkepohl, 2005; 2016).

Equation 3 sets the VAR model. As suggested by Barberà et al. (2019), I study Salvini's tweets and replies' polarisation through 15-day IRFs. Doing so allows me to study not only how polarisation spreads, but also the triggering factors of polarisation between Salvini and his respondents. Such a 15-day window allows for the lags included in the VAR model to come into effect (ibidem). Moreover, this window enables the variables' associations to be activated and sustained through studying the Salvini's tweets and replies with the highest tweets' time exposition (ibidem). The figure below displays the IRFs output.

Figure 6.8: Summary of cumulative IRFs output



Note: This table displays the IRFs results matrix. Each panel of the matrix represents an IRF, which captures how a one-time 10 percentage point increase (y-axis) in a given variable is associated with either itself, or another variable across-time (x-axis). The main intuition regarding its interpretation is to find the statistically significant relationships in which the blue lines (i.e., the effect of one variable on either itself or another across- time) is not contained in the range drawn by the asymptotic cumulative standard error and the zero lines (i.e., the bold black lines). The dashed black lines around the blue lines mark the standard errors for the individual variables' associations.

Figure 6.8 displays IRFs output, representing a result matrix. Each panel shows the effect of each variable on either itself, or on others. Specifically, the IRFs capture how a one-time 10 percentage point increase in a given variable is associated with either itself or another variable across-time. The y-axis of each panel represents the measurement of the variables, and the x-axis shows the time variation (in days). In each panel, the blue lines represent the effect of one variable on either itself or another across-time. The dashed black lines around the blue lines mark the standard errors for the individual variables' associations. Finally, the region enclosed by the bold horizontal black line (asymptotic cumulative standard error) and zero-intercept line, indicates the 95% confidence region. This region determines the statistical significance of the IRFs. When the association between a one-time 10-percentage point increase in a given variable and either itself, or another variable, falls outside the 95% confidence region, one can assume that such a relationship is statistically significant.

Among the 25 panels in the IRFs' output matrix, only nine panels display statistically significant associations. These nine panels' associations can be summarised in three overall trends, which are then presented in a single empirical account.

The first trend regards the effects of political content in both Salvini's tweets and replies, on polarisation levels across tweets and replies.

- First, in column 4, row 1, the panel displays the association between a political reply and tweets' polarisation in the first 15 days. One 10-percentage point unit increase in political replies has a positive (yet mild) effect on tweets' polarisation on the fourteenth day at the 95% confidence level. That is, on the first days since tweets' publication, political replies are not statistically significantly associated with successive Salvini's tweets' polarisation levels. However, when reaching the fourteenth day, such association

becomes positive and significant. This means that political replies have an impact on tweets' polarisation longer after the first tweet has been posted.

- Second, in column 3, row 2, the panel displays the association between political tweets and replies' polarisation in the 15-day window. Before day 7, the association between the two variables is negligible with no statistical significance. However, after day 7, one 10-percentage point increase in political replies (i.e., when a reply concerns political matters) is negatively associated with replies' polarisation. That is, in the first few days, political replies are not likely to be associated with polarised language. However, after 7 days of publication, political replies are likely to be negatively (yet mildly) associated with polarised language. Put differently, in the first 15 days, political tweets do not seem to have a robust effect on polarisation in replies.

- Third, in column 2, row 4, the panel displays the association between replies' polarisation and political replies in the 15-day window. On the first 3 days, one 10-percentage point unit increase in replies' polarisation has a negative (yet mild) effect on political replies at the 95% confidence level. That is, in the first 3 days, greater levels of polarisation in replies is associated with political content. Such a finding confirms the previous results that the first replies and levels of polarisation are not necessarily driven by political content. Following the first 3 days, the association between replies' polarisation and political replies becomes non-statistically significant, yet it grows to become a positive association.

- Finally, in column 3, row 4, the panel displays the relationship between political tweets and political replies. From day 3 to day 5, one 10-percentage point unit increase in political tweets has a negative (yet mild) effect on political replies at the 95% confidence level. On other days in the 15-day window, such a relationship tends not to be statistically significant. That is, political tweets only lead to political replies

after a few days since publication. This finding may corroborate the hypothesis that at first political tweets do not trigger a political reply. Instead, the repliers may be more interested in non-political tweets, before engaging in political discussions.

In sum, in the first 15 days since publication, Salvini's tweets on political matters do not seem to have a positive impact on replies' polarisation and political replies. Instead, political tweets are negatively associated with both replies' polarisation and political content. In other words, political tweets tend to lead to less polarisation in repliers, and less political discussions. On the other hand, political replies have a positive effect on tweets' polarisation, and a negative effect on political replies. Therefore, while political replies increase tweets' polarisation, such political replies do not increase replies' polarisation in the first 15 days.

The second trend regards the effects of followers' replies (i.e., when a reply is posted by a Salvini's follower) on polarisation levels across tweets and replies, and discussion content.

- First, in column 5, row 1, the panel displays the association between followers' reply and tweets' polarisation in the first 15 days. On the first day, one 10-percentage point unit increase in followers' replies has a positive (yet mild) effect on tweets' polarisation at the 95% confidence level. However, following day one, such association reaches non-statistically significant levels, until the seventh day since the tweet's publication. Since day 7, one 10-percentage point unit increase in followers' replies has a negative effect on tweets' polarisation. That is, on average, in the first 15 days since publications, followers' replies have an instant positive effect on tweets' polarisation. However, following the first few days, such effects disappear before returning yet negatively after day 7. In other words, followers' replies instantly positively affect Salvini's tweets, yet in later stages such an effect becomes negative.

- Second, in column 5, row 2, the panel displays the association between followers' reply and replies' polarisation in the first 15 days. Similarly, as per Salvini's tweets, one

10-percentage point unit increase in followers' replies has a positive (yet mild) effect on replies' polarisation at the 95% confidence level. Such an association persists from day 1 to day 12, meaning that followers' replies tend to be more polarised in the first days since tweets' publication. After day 12, such association becomes non-statistically significant. Therefore, one can assume that followers' replies are strong predictors of replies' polarisation.

- Third, in column 5, row 4, the panel displays the association between followers' replies and political replies in the first 15 days. From day 3 to day 5, one 10-percentage point unit increase in followers' replies has a positive effect on political replies at the conventional levels. That is, followers' replies are likely to be political only for a short period (days 3 to 5). Prior to day 3, and after day 5, followers' replies and political content are negatively associated, yet non-statistically significantly. This means that following a short window in early days, followers' replies do not tend to discuss political matters.

In sum, in the first 15 days since tweets' publication, followers' replies tend to have a positive (yet mild) effect on polarisation in both tweets and replies. In particular, in the first 2-3 days, followers' replies trigger greater polarisation levels in tweets. Instead, followers' replies tend to be positively associated with replies' polarisation only after day 3. This means that followers' replies first affect tweets' polarisation and successively replies' polarisation. Moreover, followers' replies tend to become political only between day 3 and 5. That is, initially followers' replies are more likely to concern non-political matters before concerning politics.

The third trend regards the effects of two variables on themselves across-time. First, I discuss the effect of political replies (i.e., when a reply to Salvini's tweets concerns politics) on successive political replies. Second, I study the effect of followers' replies (i.e., when a reply is posted by a Salvini's follower) on successive followers' replies.

- First, in column 4, row 4, the panel displays the association between previous political replies on successive political replies in the first 15 days. From day 3 to day 5, one 10-percentage point unit increase in political replies has a positive effect on successive political replies at the conventional levels. That is, political replies are likely to lead to more political replies only for a short period (days 2 to 5). Prior to day 2, political replies and successive political replies are positively associated (yet without statistical significance), and after day 5, these variables become negatively associated, yet non-statistically significantly. This means that, after a short 3-day-window in early days, political replies do not tend to lead to more political replies.

- Second, in column 5, row 5, the panel displays the association between previous followers' replies on successive followers' replies in the first 15 days. After day 13, one 10-percentage point unit increase in followers' replies has a positive effect on successive followers' replies at the conventional levels. That is, followers' replies are likely to lead to more followers' replies only after 13 days. Prior to day 13, followers' replies and successive followers' replies are positively associated, yet non-statistically significantly. This means that, only after 13 days, followers' replies do tend to lead to more followers' replies.

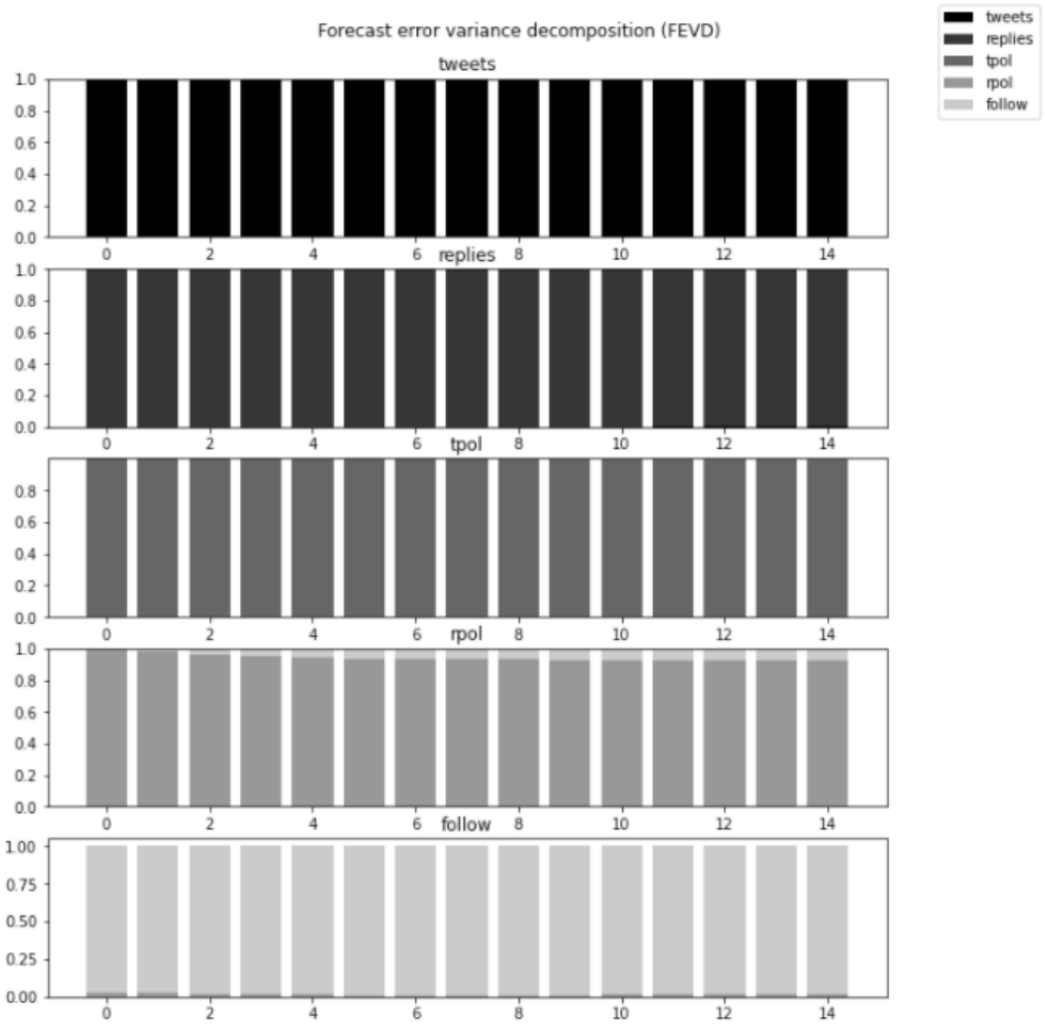
In sum, both political replies and Salvini's followers' replies are positively associated with themselves across-time in the first 15-day window. In other words, political content in replies is likely to remain salient for a short span. Similarly, Salvini's followers' replies trigger more followers' replies yes only after day 13.

To summarise, the IRFs output reveals two main trends in the spread of polarisation within the 15-day window. First, political content in Salvini's tweets seems not to trigger neither polarised replies nor political replies. Instead, political replies tend to lead to greater tweets' polarisation, yet they are not associated with greater levels of replies' polarisation nor with

political replies. This means that political replies trigger polarisation in tweets in the first 15 days since tweets' publication, yet not in replies. Second, followers' replies to Salvini's tweets play an active role in the spread of polarisation. In particular, in the first few days, following the tweets' publication, followers' replies lead to greater levels of polarisation in tweets. Instead, they trigger a polarised reaction in replies only after day 3. This means that Salvini's followers trigger a significant positive shift in polarisation in Salvini's tweets, but only later on they have an impact on replies' polarisation. Moreover, such followers' replies tend to concern political matters only between a short 2-day window. In conclusion, the main drivers of polarisation seem to be Salvini's followers. Political matter only plays a role in the spread of polarisation after a few days and for only a short range of time. Before discussing the theoretical implications of such findings, I first turn to VAR validation checks.

To validate the VAR model results, as suggested by Lütkepohl (2005), I compute a forecast error variance decomposition (FEVD) in Figure 6.9. The FEVD shows how a change in each variable's error is explained by other variables. Figure 6.9 displays that none of the variables' error terms seems to be explained by other variables. The error variance of three variables (i.e., tweets' and replies' polarisation, and political tweets) out of five is almost 100% explained by themselves. 90% of political replies' error variance is explained by the variable itself, the remaining 10% is explained by followers' replies after day-2. This means that followers' replies only explain 10% of the political replies' error variance, yet non-significantly. Similarly for the followers' replies time-series, only 5% (which decreases across-time) of its error variance is explained by political replies (the remaining 95% is explained by itself). For these reasons, the FEVD confirms the VAR results.

Figure 6.9: FEVD for each variable in the VAR model



Note: This figure displays the FEVD output, showing how a change in each variable's error is explained by other variables across the 15-day window.

CHAPTER 7

DISCUSSIONS

This chapter will first discuss the theoretical implications of the empirical results. In particular, the OLS and VAR results will aid me to accept or reject the theory-driven hypotheses. Second, it will evaluate the overall finding of this thesis through discussing the impact of the pandemic on my research, as well as discussing external and internal validity of the results.

7.1 Results discussions

The OLS models results enable to have a preliminary understanding of polarisation dynamics involving Salvini's tweets and replies, across homophilic groups and political topics. Such OLS results will be further corroborated by coefficient plots to understand the individual sub-topics' impact on polarisation. Finally, the VAR results, through the evaluation of the IRFs finding, will aid me to understand polarisation triggering factors within a 15-day window.

7.1.1 Bi-directional relationship

Hypothesis 1a: Higher levels of polarisation in Salvini's tweets (at t) are positively associated with lagged levels of polarisation in replies to Salvini's tweets (at *t-1*).

Hypothesis 1b: Higher levels of polarisation in replies to Salvini's tweets (at t) are positively associated with lagged levels of polarisation in Salvini's tweets (at *t-1*).

This section discusses the bi-directional association between tweets and replies' polarisation. First, from Figure 6.2 and 6.5, tweets and replies' levels of polarisation are bi-directionally associated across five different models. From Figure 6.2, one unit increase in replies' polarisation is positively (yet mildly) associated with tweets' polarisation, meeting the conventional statistical significance standards (p-value < 0.05). Similarly, from Figure 6.5, one unit increase in tweets' polarisation is positively (yet mildly) associated with replies polarisation, which is significant at the 95% confidence level (p-value < 0.05). Such a positive relationship is robust to the inclusion of other covariates (i.e., discussion topics and followers' replies) to the OLS models. The effects of other explanatory variables on the bi-directional relationship between tweets and replies' polarisation will be further discussed. Yet, after this preliminary analysis, I may accept hypothesis 1a and hypothesis 1b at the conventional standards (i.e., bi-directional relationship between tweets and replies' polarisation).

7.1.2 Followers' replies

Hypothesis 2a: Higher levels of polarisation in Salvini's tweets (at t) are positively associated with the lagged Salvini's followers' replies (at *t-1*).

Hypothesis 2b: Higher levels of polarisation in replies to Salvini's tweets (at t) are positively associated with Salvini's followers' replies (at *t-1*).

This section now considers the effect of accounting for followers' replies on both tweets and replies' polarisation. Figure 6.2, model 2, shows that tweets' polarisation is negatively (yet mildly) associated with followers' replies. However, such a relationship does not reach the conventional statistical significance standards (p-value > 0.05). In other words, when tweets' repliers are Salvini's followers, there is not enough evidence to assume that the levels of polarisation in tweets decrease as a result of followers replying to the tweets. This finding

may reject hypothesis 2a (i.e., follower replies having a positive effect on tweets' polarisation). On the contrary, from Figure 6.5, model 2, there is a positive (and statistically significant) association between one unit increase in followers' replies on replies' polarisation. That is, when replies are posted by Salvini's followers, these are likely to be more polarised than non-followers' replies. This finding seems to confirm hypothesis 2b (i.e., followers replies having a positive effect on replies' polarisation). As previously discussed, controlling for replies posted by Salvini's followers does not affect the persistence of a polarisation bi-directionality between Salvini's tweets and replies to his tweets.

7.1.3 Political content

Hypothesis 3a: Salvini's political tweets (at t) are positively associated with greater levels of polarisation in Salvini's tweets (at t).

Hypothesis 3b: Lagged Salvini's political tweets (at *t-1*) are positively associated with greater levels of polarisation in replies to Salvini's tweets (at t).

I discuss the effects of political content on tweets and replies' polarisation. In Figure 6.2, on the one hand, when considering political tweets, one unit increase in the variable accounting for tweets' content being political is strongly negatively associated (-0.110) with the levels of polarisation in Salvini's tweets. Such a finding may lead to failing to accept hypothesis 3a, since political tweets have a negative impact on tweets' polarisation, instead of a positive effect. Moreover, there is a positive (and statistically significant) association between one unit increase in political tweets on replies' polarisation levels. In other words, when Salvini's tweets concern political matters, these are likely to lead to polarised responses. This finding may allow me to accept hypothesis 3b (i.e., political tweets having a positive impact on replies' polarisation).

Hypothesis 3c: Lagged political replies (at *t-1*) are positively associated with higher levels of polarisation in Salvini's tweets (at t).

Hypothesis 3d: Political replies (at t) are positively associated with greater levels of polarisation in replies to Salvini's tweets (at t).

On the other hand, when considering political replies, the association between political replies and tweets' polarisation has a low magnitude (ranging from 0.02 and 0.004 across models), and it does not meet the conventional 95% statistical significance level. That is, when a reply concerns a political matter, it is not likely to lead to greater levels of polarisation in successive Salvini's tweets. This finding may reject hypothesis 3c (i.e., political replies having a positive effect on tweets' polarisation). Furthermore, there is a positive (and statistically significant) association between one unit increase in political replies on replies' polarisation levels. As such, one may expect that, on average, political replies are more likely to include polarising language than non-political replies (all being equal). This finding seems to confirm hypothesis 3d (i.e., political replies having a positive effect on replies' polarisation). Therefore, I find evidence to claim that, on average, political content either from Salvini or his repliers leads to higher levels of polarisation in replies (all being equal).

7.1.4 Interaction between political content and followers

Hypothesis 3e: Higher levels of polarisation in Salvini's tweets (at t) are positively associated with lagged Salvini's followers' political replies (at *t-1*), contingently on political tweets (at t).

I turn to evaluate the results of the interaction between political content and followers' replies on the spread of polarisation. Fist, in Figure 6.2, it is significant to mention that

when accounting for followers' political replies and political tweets, one unit increase in replies' polarisation remains positively associated with tweets' polarisation. The effect of replies' polarisation on tweets' polarisation remains and increases in magnitude from 0.021 to 0.023. Moreover, the statistical significance of such relationships increases from the 95% to the 99% confidence level. Second, when considering the effect of followers' political replies on polarisation in political tweets, I find that one unit increase in political tweets remains negatively associated with tweets' polarisation. Political tweets' effect on tweets' polarisation decreases from -0.111 (model 4) to -0.110 (model 5), thus not substantially and without affecting statistical significance, which remains at the 99% confidence level. Followers' replies' effect on tweets' polarisation does not change (-0.03, p-value > 0.05). Political replies' effect on tweets' polarisation increases, yet not substantially from 0.002 to 0.004 (p-value > 0.05). This means that tweets' polarisation is not substantially and positively associated with followers' political replies, and it is negatively associated with political tweets. In other words, these results confirm that political content does not positively affect tweets' polarisation. Moreover, there is no statistical evidence to claim that Salvini's tweets are positively associated with repliers' being Salvini's followers, contingently on political content. For these reasons, one may reject hypothesis 3e.

> **Hypothesis 3f:** Higher levels of polarisation in replies to Salvini's tweets (at t) are positively associated with Salvini's followers' political replies (at t), contingently on lagged political tweets (at *t-1*).

In Figure 6.5, accounting for followers' political replies and political tweets, one unit increase in tweets' polarisation remains positively (and statistically significantly) associated with replies' polarisation. The effect of tweets' polarisation on replies' polarisation remains increases in magnitude from 0.021 to 0.023. Moreover, the statistical significance of such relationships increases from the 95% to the 99% confidence level. Second, I consider the effect

of followers' political replies and political tweets on polarisation in replies. Political replies' effect on replies' polarisation increase, yet not substantially from 0.236 (model 4) to 0.237 (model 5), remaining statistically significant (p-value > 0.05). Similarly, followers' replies' effect on replies' polarisation decreases not substantially, remaining statistically significant (from 0.066 to 0.065, p-value < 0.05). Finally, political tweets' effect decreases from 0.353 to 0.351, thus not substantially and without affecting statistical significance, which remains at the 99% confidence level. These findings mean that followers' political replies and political tweets are positively associated with polarisation in replies. In other words, these results confirm that political content does positively affect replies' polarisation across-time. Moreover, there is statistical evidence to claim that replies to Salvini's tweets are positively associated with repliers' being Salvini's followers, contingently on political content. For these reasons, one may accept hypothesis 3f.

7.1.5 OLS robustness checks

To corroborate such results regarding tweets' polarisation, I computed two coefficient plots for political and non-political subtopics across Salvini's followers and non-followers. Figure 6.3 and 6.4 show that among both followers and non-followers, 'Salvini' as a political topic is statistically significant (yet mildly) in predicting lower levels of polarised tweets at the 95% confidence level. However, 'Salvini' as a topic only has a marginal negative effect in predicting variation in tweets' polarisation. This may partially reject hypothesis 3b as higher levels of polarisation in Salvini's tweets are negatively (yet mildly) associated with the tweets' content being political. Nevertheless, more evidence is needed to confirm such results.

Analogously, I computed two coefficient plots to study replies' polarisation across both political and non-political subtopics, while accounting for Salvini's followers and non-followers. Figures 6.6 and 6.7 show that among the sub-topics, 'Salvini', 'Renzi', and 'Immigrants'

are the most statistically significant sub-topics in predicting replies' polarisation. When accounting for political content (in both tweets and replies) and respondents' homophilic group (i.e., replies posted by Salvini's followers), 'Salvini' is the strongest predictor of lower polarisation in replies. That is, on average, 'Salvini' as a topic decreases the levels of polarisation in replies. Instead, accounting for political content and Salvini's non-followers, 'Salvini', 'Renzi', and 'Immigrants' are the strongest predictors of replies' polarisation at the conventional levels (p-value < 0.05). One unit increase in each of these sub-topics is negatively (yet mildly) associated with replies' polarisation. These results may provide further evidence to reject hypothesis 3c as higher levels of polarisation in replies to Salvini's tweets are negatively (instead of positively) associated with the replies' content being political. Nevertheless, given the low magnitude effect of such results from the coefficient plots, more evidence may be needed to confirm them.

7.1.6 Triggering factors

This section discusses the theoretical implications of the VAR results considering the theory-driven hypotheses. Through the VAR output, I study several associations among the variables within the polarisation framework in a 15-day window. This window allows me to investigate the initial triggering factors (i.e., short-term impulses that initially activate polarisation) in the spread of polarisation between Salvini's tweets and his respondents.

> **Hypothesis 4a:** Higher levels of polarisation in Salvini's tweets do not trigger higher levels of polarisation in replies to Salvini's tweets.

> **Hypothesis 4b:** Higher levels of polarisation in replies to Salvini's tweets trigger higher levels of polarisation in successive Salvini's tweets.

Following the IRFs output (Figure 6.8), I explore the overall bi-directional relationship among Salvini's tweets and replies. In the first 15 days since publication, replies' polarisation does not trigger greater tweets' polarisation. Thus, at the conventional confidence levels (95%), I may reject hypothesis 4b (i.e., replies' polarisation triggering tweets' polarisation). Similarly, reversing the causal arrow, polarisation in Salvini's does not seem to trigger replies' polarisation. Thus, I may accept hypothesis 4a (i.e., Salvini's tweets' polarisation not triggering replies' polarisation).

> **Hypothesis 5a:** Salvini's followers' replies to Salvini's tweets initially trigger higher levels of polarisation in Salvini's tweets.

> **Hypothesis 5b:** Salvini's followers' replies to Salvini's tweets initially trigger higher levels of polarisation in replies to Salvini's tweets.

This section turns to followers' replies and their effects on polarisation. The variable accounting for followers' repliers to Salvini's tweets plays a significant role in the spread of polarisation. In particular, in the first few days since tweets' publication, followers' replies lead to greater levels of polarisation in successive tweets. Such a finding may allow accepting hypothesis 5a as one 10-percentage point unit increase in Salvini's followers is positively associated with polarisation in successive Salvini's tweets. Furthermore, Salvini's followers also trigger polarised responses in overall replies. In particular, following the IRFs output, I find that Salvini's followers trigger greater polarisation in replies only after day 3. Thus, one can only partially accept hypothesis 5b as followers do lead to greater polarised responses but only after a few days since the first tweets were posted.

> **Hypothesis 5c:** Hypothesis 5c (H.5c): Salvini's followers' replies trigger the discussion topics in Salvini's tweets.

Hypothesis 5d: Salvini's followers' replies trigger the discussion topics in replies to Salvini's tweets.

Considering the discussion topics triggered by followers, according to the VAR analysis, follower's replies do not affect discussion topics in Salvini's tweets. As such, I do not find statistical evidence to claim that a past followers' reply has an effect on a successive tweet in predicting discussion topics. For these reasons, one may reject hypothesis 5c. However, VAR results suggest that followers' replies lead to political topics. I find that only, from day 3 to day 5, one 10-percentage point unit increase in followers' replies has a positive effect on political replies at the conventional levels. Therefore, one may partially accept hypothesis 5d as followers' replies seem to lead to political discussions, yet only for a small temporal window.

Hypothesis 6a: Political Salvini's tweets do not initially trigger higher levels of polarisation in Salvini's tweets.

Hypothesis 6b: Political Salvini's tweets do not initially trigger higher levels of polarisation in replies to Salvini's tweets.

Third, when considering political content, the evidence provides contrasting results among tweets and replies. According to the IRFs, Salvini's political tweets may not trigger political replies in the 15-day window under analysis. In fact, when considering the conventional 95% confidence intervals political tweets do not tend to have a statistically significant effect on the levels of polarisation neither in Salvini's tweets nor in replies. Therefore, one may reject hypotheses 6a and 6b, as there is no substantial evidence to claim that political tweets trigger polarised successive tweets and replies in the 15-day window.

Hypothesis 6c: Political replies to Salvini's tweets initially trigger higher levels of polarisation in Salvini's tweets.

Hypothesis 6d: Political replies to Salvini's tweets initially trigger higher levels of polarisation in replies to Salvini's tweets.

When studying political replies, one 10-percentage point unit increase in political replies has a positive (yet mild) effect on tweets' polarisation on the fourteenth day at the 95% confidence level. As a result, one may accept hypothesis 6c, as political replies seem to play an important role in triggering successive Salvini's tweets' polarisation. However, political replies are not associated with greater levels of successive replies' polarisation. That is, political replies trigger polarisation in tweets in the first 15 days since tweets' publication, however this does not occur in replies. Therefore, one may accept hypothesis 6c and reject hypothesis 6d for lack of statistical evidence.

7.1.7 Summary of the results

This section summarises the main results of these analyses. First, the spread of polarisation between Salvini's tweets and respondents occurs bi-directionally as tweets and replies' polarisation is positively associated across-time, and after accounting for several covariates. Second, Salvini's followers do not seem to have a positive impact on polarisation in Salvini's tweets. However, when a reply is posted by a Salvini's follower, it is likely to be associated with higher levels of polarisation in replies. Third, political content has distinguished effects on polarisation across tweets and replies. Political tweets are more likely to display a lower level of polarisation than non-political tweets. Instead, political replies are more likely to be polarised than non-political replies. When controlling for the specific topics, I find little evidence to claim that a single topic has a substantial impact on tweets and replies' polarisation levels. Yet, it is noteworthy to mention that the topic 'Salvini' is a mild predictor of polarisation in both tweets and replies. Fourth, when considering the interaction between homophilic group-belonging and political content, I also find contrasting

results among tweets and replies. On the one hand, there is no statistical evidence to claim that Salvini's tweets are positively associated with repliers' being Salvini's followers, contingently on political content. On the other hand, the levels of polarisation in replies are indeed positively associated with repliers' being Salvini's followers, contingently on political content. Finally, the most important triggering factors in tweets and replies' polarisation are followers' replies and discussion topics in the first 15 days. In particular, followers' replies are likely to trigger greater polarisation in both tweets and replies instantly. Moreover, political replies trigger greater polarisation in tweets in the first 15 days since tweets' publication. However, political tweets do not seem to have a triggering effect on replies' polarisation.

7.2 Evaluation

7.2.1 The Covid-19 impact

This paper has had several limitations due to the ongoing Covid-19 pandemic. I planned on utilising Departmental or Collegial funding for advanced Twitter data-gathering and crowd-sourced data-coding approaches. However, such funding schemes have been cancelled or postponed due to the economic consequences of the pandemic. First, with funding, I planned on obtaining a premium Twitter API to retrieve larger sets of Twitter data without size and timing limitations. To overcome these issues, as discussed in the methodology chapter, I have utilised a combination of Python libraries. As a result, the data-gathering process has taken much more time than planned as I had to study, update and synchronise different libraries and documentations. Second, for the data-coding aspect of this thesis, I aimed to conduct crowd-sourced data-coding via Qualtrics. In particular, I planned to ask online paid-participants to code a sample of Salvini's tweets and replies. Based on this coded sample, I would have trained a novel machine learning algorithm, and then coded the entire population of tweets and replies. Such a data-coding procedure would have allowed

me to produce a more accurate coding system to code tweets and replies than the pre-trained neural network utilised in this thesis.

7.2.2 Internal and external validity

The theoretical and methodological framework developed in this thesis may be further developed in future research. At the internal validity level, there are three main routes to further explore the approach. First, to enhance the findings related to the effects of AP, successive research may explore a greater variety of discussion topics on Twitter, beyond political versus non-political. In particular, one could explore subtopics, such as economic, cultural, social, and their interactions. Doing so would provide a more granular understanding of which subtopics within political and non-political discussions lead to polarised debates on Twitter. Second, future research may take further the PH findings, including the analysis of users mentioned in replies to Salvini, other than Salvini. Analysing the dynamics of mentioned users would allow me to better understand how inter- and intra-group discussions take place. For example, researchers may cross-tabulate the users mentioned in the replies to Salvini's tweets with whether those mentioned are Salvini followers or non supporters. Third, socio-economic indicators regarding people's most salient issues in Italy may be utilised to test whether the results remain significant after one controls for salient political issues. Perceived socio-economic indicators can further corroborate whether and how the results are linked to offline political and non-political activities. For example, indicators regarding the Italian north-south divide could explain whether there is intra-national differentiation among Salvini's respondents.

To test the external validity of this framework, there are also three main routes which may be undertaken. First, the timespan of the research may be expanded to include a larger number of tweets written since Salvini has been registered on Twitter, and after 2019, following the end of his mandate as Italian Vice-Prime Minister. Increasing the timespan would allow me

to better understand the effects of Salvini's tweets on Italian users across several national and European electoral cycles, terms in office, and the COVID-19 pandemic, which may furnish a great space of action to test my theory-driven hypotheses. Second, further research may explore other Italian political actors, parties, and/or movements. For example, given their online presence, Movimento 5 Stelle (i.e., 5-Star Movement), PD (i.e., Democratic Party), Forza Italia (Italy Forward) may be taken into account. In this way, one could perform similar analyses across party lines and have a fair representation of centre, right- and left-wing parties. As previously explained, I focused on Matteo Salvini because he presented the best Italian case to study the spread of polarisation. However, further scholarship may find a way to create non-biased weights and compare Salvini's results to other cases, and vice-versa. Third, another social media network may be utilised to corroborate these results. In particular, given the number of active users among the Italian population, media coverage, Facebook also provides a strong research ground. Doing so would allow further research to include another space of action for both Salvini and his respondents.

CHAPTER 8

CONCLUSIONS

The objective of this thesis was to understand how and under what conditions polarised language spreads between politicians and the public, studying Matteo Salvini's rhetoric on Twitter. This chapter reiterates the main findings of such investigation, presenting their theoretical and methodological contributions to the literature, as well as highlighting avenues for further research.

The thesis results are broken down into four complementary findings. First, it studied polarisation dynamics, analysing the bi-directional relationship between politicians' tweets and replies' polarisation. In Salvini's case, overall, tweets and replies' polarisation were positively associated, after controlling for several covariates. Second, this thesis studied how PH plays a significant role in spread polarisation in replies. Significantly, Salvini's followers' replies are likely to be more polarised than non-followers' replies. However, followers' replies do not seem to affect successive tweets' polarisation. Third, this thesis studied how political content affects the spread of polarisation. It found that, on average, political topics play a significant role in augmenting polarisation in replies, but not in tweets. In other words, when controlling for political content (in both tweets and replies), only political replies display greater levels of polarisation than non-political replies. Instead, political tweets are more likely to display lower levels of polarisation than non-political tweets. When considering both PH and discussion topics, the levels of polarisation in replies are positively associated with repliers' being Salvini's followers, contingently on political content in both tweets and replies. Instead, when reversing the causal arrow, tweets' polarisation seems not to be affected by the intersection between followers' replies and political content. Fourth, having defined the key components of such polarisation framework, I studied the most important initial triggering factors in tweets and replies' polarisation. Followers' replies are likely to instantly trigger greater polarisation in both Salvini's tweets and replies. Moreover, political replies trigger

greater polarisation in tweets in the first 15 days since tweets' publication. However, political tweets do not seem to have a triggering effect on replies' polarisation.

These findings have several theoretical and methodological implications. At the theoretical level, my contribution is three-fold. First, I found evidence to accept the intersection between polarisation supply- and demand-focused theories. As such, the dynamics of the spread of polarisation seem to show a bi-directional relationship among politicians and the public in accordance with Theocharis et al. (2020) and Kosmidis and Theocharis (2021). Second, PH and discussion topics are also likely to increase replies' polarisation, yet they lead to opposite results in tweets. In other words, I find evidence to claim that, for the public, Salvini's followers tend to post more polarised responses than non-followers, and political content (in both tweets and replies) further exacerbates polarisation. Instead, polarisation in Salvini's tweets is not positively associated with followers' replies and political content. Therefore, these results may only partially confirm PH theories (Conover et al., 2011; 2012; Hong and Kim, 2016), and discussion topics' relevance (Barberà et al., 2015; 2017; 2019a; Wojcieszak and Mutz, 2009) to increase polarisation in the public's replies. Third, in accordance to Theocharis et al. (2020), the main short-term triggers of polarisation are likely to be replies posted by politicians' followers and concerning politics. This finding may also corroborate previous research on the critical role of politicians' Twitter followers in circulating, and amplifying polarisation (Mazzoleni and Bracciale, 2018; Sensales and Di Cicco, 2019; Reddy, 2001; Molinaro et al., 2018). Finally, political content does seem to play a role in triggering subsequent polarised language as previously discovered by Theocharis et al. (2016; 2020).

At the methodological level, this thesis has offered several innovations. First, the web-scraping methods have allowed me to retrieve a large set of Salvini's tweets and replies to his tweets, as well as 10 million data-points to study polarisation. This data-gathering process has allowed me to overtake some limitations of previous literature which was based on manual sampling (see Bobba, 2019; Bracciale and Cepernich, 2018). Second, this thesis

has coded Salvini's tweets and replies through a neural network textual classification to study the levels of polarisation in both tweets and replies. This method allowed me to classify every tweet/reply in the sample by their emotional charge, studying polarisation as a continuum. Such a strategy may be further ameliorated through a new tokenisation strategy for the Italian language. Finally, it has utilised several econometric methods to study the relationship between tweets and replies' polarisation through OLS, coefficient plots, and VAR analyses. These analyses allowed me not only to understand long-term effects of tweets on replies, and vice-versa, but also triggering (or short-term) effects. Through these methodologies, I have generated a comprehensive and large data-set of Salvini's tweets and replies, which I utilised to study theory-driven hypothesis, and can potentially be utilised by other scholars for further research purposes and theories.

Due to space constraints, I had to limit the scope of my analysis, however, further questions can be answered following my theoretical and methodological framework. To enhance the internal validity of my findings, future research may consider more subtopics, such as economic, cultural, social, and their interactions. Moreover, scholars may analyse the dynamics of users mentioned by Salvini's respondents, beyond Salvini, to better understand how inter- and intra-group discussions take place. Finally, socio-economic indicators regarding people's most salient issues in Italy may be utilised to test whether these results remain significant after one controls for salient political issues. At the external validity level, future research may augment the sample size of my database by including a longer timespan, other Italian parties for comparison, and other social media as new places of interactions. Doing so would allow researchers to have more granular results into how and under what conditions polarisation spreads on Twitter and beyond.

CHAPTER 9

BIBLIOGRAPHY

Abramowitz, A.I., and Webster, S.W. (2016). 'The rise of negative partisanship and the nationalization of U.S. elections in the 21st century' *Electoral Studies*, 41, pp. 12-22.

Ackland, R., and Shorish, J. (2017). 'Political Homophily on the Web'. In Cantijoch, M., Gibson, R. and Ward, S. (eds), *Analysing Social Media Data and Web Networks*. New York: Palgrave Macmillan.

Adamic, L. A., and Glance, N. (2005). 'The political blogosphere and the 2004 U.S. election: divided they blog'. In *Proceedings of the 3rd international workshop on Link discovery*. New York, NY.

Albertazzi, B., Giovannini, A., and Seddone, A. (2018). "No regionalism please, we are Leghisti!' The transformation of the Italian Lega Nord under the leadership of Matteo Salvini', *Regional and Federal Studies*, 28(5), pp. 645-671.

Anderson, A.A., and Huntington, H.E. (2017). 'Social Media, Science, and Attack Discourse: How Twitter Discussions of Climate Change Use Sarcasm and Incivility', *Science Communication*, 39(5), pp. 598-620.

Anderson, A.A., Brossard, D., Scheufele, D.A., Xenos, M., and Ladwig, P. (2013). 'The 'Nasty Effect': Online Incivility and Risk Perceptions of Emerging Technologies', *Journal of Computer-Mediated Communication*, 19, pp. 373-87.

Arzheimer, K., and Carter, E. (2006). 'Political opportunity structures and right-wing extremist party success', *European Journal of Political Research*, 45(3), pp. 419-443.

Bail, C., Argyle, L., Taylor, B., Bumpus, J., Haohan, C., Hunzakerd, F., Leea, J., Manna, M., Merhout, F., and Volfovsky, A. (2018). 'Exposure to opposing views on social media can increase political polarization', *PNAS*, 115 (37), pp. 9216-9221.

Barberá, P, Jost, J.T., Nagler, J., Tucker, J.A., and Bonneau, R. (2015). 'Tweeting From Left to Right: Is Online Political Communication More Than an Echo Chamber?', *Psychological Science*, 26(10), pp. 1531-1542.

Barberà, P. (2014). 'Birds of the same feather tweet together: Bayesian ideal point estimation using Twitter data', *Political Analysis*, 23(1), pp. 76-91.

Barberá, P., and Zeitzoff, T. (2017). 'The New Public Address System: Why Do World Leaders Adopt Social Media?', *International Studies Quarterly*, 62, pp. 121-30.

Barberá, P., Boydstun, A., Linn, S., and McMahon, R. (2016). 'Methodological Challenges in Estimating Tone: Application to News Coverage of the U.S. Economy', *Annual meeting of the American Political Science Association, Philadelphia, Pennsylvania*.

Barberá, P., Casas, A., Nagler, J., Egan, P., Bonneau, R., Jost, J., and Tucker, J. (2019a). 'Who Leads? Who Follows? Measuring Issue Attention and Agenda Setting by Legislators and the Mass Public Using Social Media Data', *American Political Science Review*, 113(4), pp. 883-901.

Barberà, P., Gohdes, A., Zeitoff, T., and Iakhins, E. (2019b). 'From riot police to tweets: How world leaders use social media during contentious politics', *Working paper*.

Barnidge, M. (2017). 'Exposure to Political Disagreement in Social Media Versus Face-to-Face and Anonymous Online Settings', *Political Communication*, 34(2), pp. 302-321.

Benoit, K., Muhr, D., and Watanabe, K. (2021). Package 'stopwords'. Available at: https://cran.r-project.org/web/packages/stopwords/stopwords.pdf (Accessed on 30 Mar 2021).

Bentivegna, S. (2015). *A colpi di tweet. La politica in prima persona*. Bologna: Mulino.

Bessi, A., Zollo, F., Del Vicario, M., Scala, A., Caldarelli, G., and Quattrociocchi, W. (2015). 'Trend of Narratives in the Age of Misinformation', *PloS one*, 10(8).

Bobba, G. (2019). 'Social media populism - features and 'likeability' of Lega Nord communication on Facebook', *European Political Sciences*, 18, pp. 11-23.

Bobba, G., and Quaranta, M. (2019). 'Introduction to the Special Issue: A new page for Europe? Insights from the 2019 European Elections in Italy', *Italian Political Science*, 14(2).

Bobba, G., and Roncarolo, F. (2018). 'View of The likeability of populism on social media in the 2018 Italian general election', *Italian Political Science*, 13(1).

Boehmke, B., and Greenwell, B. (2020). *Hands-On Machine Learning with R*. London: Taylor Francis Group.

Borah, P. (2012). 'Does It Matter Where You Read the News Story? Interaction of Incivility and News Frames in the Political Blogosphere', *Communication Research*, 46, pp. 809-82.

Bos, L., van der Brug, W., and de Vreese, C. (2013). 'An experimental test of the impact of style and rhetoric on the perception of right-wing populist and mainstream party leaders', *Acta Politica*, 48(2), pp. 192-208.

Boxell, L., Gentzkow, M., and Shapiro, J.M. (2017). 'Greater Internet use Is Not Associated with Faster Growth in Political Polarization among US Demographic Groups', *Proceedings of the National Academy of Sciences*, 114(40), pp. 10612-10617.

Bracciale, R., and Cepernich, C. (2018). 'Hybrid 2018 campaigning: the social media habits of Italian political leaders and parties', *Italian Political Science*, 13(1), pp. 36-50.

Bradley, M.M., and Lang, P.J. (1999). 'Affective Norms for English Words (ANEW): Instruction manual and affective ratings (Technical report C-1)'. Gainesville: The Center for Research in Psychophysiology, University of Florida.

Bresolin, M. (2014). 'Sui social network Salvini batte Grillo e Renzi'. *La Stampa*, 14 December 2014.

Bright, J. (2018). 'Explaining the emergence of echo chambers on social media: the role of ideology and extremism', *Journal of Computer-Mediated Communication*, 23, pp. 17-33.

Brundidge, J. (2010). 'Encountering 'Difference' in the Contemporary Public Sphere: The Contribution of the Internet to the Heterogeneity of Political Discussion Networks', *Journal of Communication*, 60(4), pp. 680-700.

Bryden, J., and Silverman, E. (2019). 'Underlying socio-political processes behind the 2016 US election', *PLoS ONE*, 14(4), pp. 1-11.

Brzozowski, A. (2019). 'Parliament calls for EU measures to counter hostile disinformation campaigns'. [online] Euractiv. Available at: https://www.euractiv.com/section/digital/news/parliament -calls-for-eu-measures-to-counter-hostile-disinformation-campaigns/ (Accessed on 29 November 2019).

Burger, S.V. (2018). *Introduction to Machine Learning in R.* New York: O'Reilly Media.

Cedroni, L. (2014). *Politolinguistica: l'analisi del discorso politico.* Roma: Carocci editore.

Cepernich, C. (2017). *Le campagne elettorali al tempo della networked politics.* Rome: Laterza.

Chan, C.L., Lai, J., Hooi, B., and Davies, T. (2017). 'The Message or the Messenger? Inferring Virality and Diffusion Structure from Online Petition Signature Data', *Social Informatics*, pp 457-472.

Chen, T., and Guestrin, C. (2016). 'Xgboost: A scalable tree boosting system'. In *Proceedings of the 25 22nd ACM signed international conference on knowledge discovery and data mining*, pp. 785-794.

Cheng, J., Danescu-Niculescu-Mizil, C., and Leskovec, J. (2015). 'Antisocial behaviour in online discussion communities', *Social and Information Networks.*

Cinelli, M., Conti, M., Finos, L., Grisolia, F., Novak, P., Tesconi, M., Zollo, F., and Quattrociocchi, W. (2019). '(Mis)Information Operations - An Integrated Perspective', *Journal of Information Warfare*, 18.2, pp. 83-98.

Coe, K., Kenski, K., and Rains, S. A. (2014). 'Online and uncivil? Patterns and determinants of incivility in newspaper website comments', *Journal of Communication*, 64, pp. 658-679.

Colleoni, E., Rozza, A., and Arvidsson, A. (2014). 'Echo Chamber or Public Sphere? Predicting Political Orientation and Measuring Political Homophily on Twitter Using Big Data', *Journal of Communication*, 64, pp. 317-332.

Conover, M., Ratkiewicz, J., Francisco, M.R., Gonçalves, B., Menczer, F., and Flammini, A. (2011). 'Political Polarization on Twitter', *ICWSM*, 133, pp. 89-96.

Conover, M.D., Goncalves, B., Flammini, A. and Menczer, F. (2012). 'Partisan Asymmetries in Online Political Activity', *EPJ Data Science*, 1(1), pp. 1-19.

Converse, P.E. (2006). 'The Nature of Belief Systems in Mass Publics', *Critical Review*, 18(1-3), pp. 1-74.

Darmofal, D., and Strickler, R. (2019). *Demography, Politics, and Partisan Polarization in the United States, 1828-2016*. New York: Springer.

de Zúñiga, H.G., and Valenzuela, S. (2011). 'The Mediating Path to a Stronger Citizenship: Online and Offline Networks, Weak Ties, and Civic Engagement', *Communication Research*, 38(3), 397-421.

Del Vicario, M., Bessi, A., Zollo, F., Petroni, F., Scala, A., Caldarelli, G., Stanley, H.E., and Quattrociocchi, W. (2016). 'The Spreading of Misinformation Online', *Proceedings of the National Academy of Sciences*, 113(3), pp. 554- 559.

Del Vigo, F. M., and D. Ferrara. (2015). *Il metodo Salvini*. Milan: Sperling Kupfer.

Di Cicco, G., and Sensales, G. (2019). 'Psycho-social aspects of populist communication on Facebook. A first analysis of the reactions to the posts of Luigi Di Maio and Matteo Salvini', *Rassegna di Psicologia*, 36(2), pp. 61-71.

Di Cicco, G., and Sensales, G. (2019). 'Psycho-social aspects of populist communication on Facebook. A first analysis of the reactions to the posts of Luigi Di Maio and Matteo Salvini', *Rassegna di Psicologia*, 36(2), pp. 61-71.

Die Bundesregierung. (2019). 'Draft law to improve law enforcement in social networks'. [online] der Bundesregierung. Available at: https://www.bmjv.de/Shared Docs/Gesetzgebungsverfahren/Dokumente/RegE-NetzDG.pdf?–blob=publication Filev=2 (Accessed on 29 November 2019).

Dodds, P.S., and Danforth, C.M. (2010). 'Measuring the Happiness of Large-Scale Written Expression: Songs, Blogs, and Presidents', *Journal of Happiness Studies*, 11(4), pp. 441-456.

Druckman, J.N., Gubitz, S.R., Lloyd, A.M., and Levendusky, M.S. (2019). 'How Incivility on Partisan Media (De)Polarizes the Electorate', *Journal of Politics*, 81, pp. 291-95.

Dubois, E., and Blank, G. (2018). 'The echo chamber is overstated: the moderating effect of political interest and diverse media', *Information, Communication Society*, 21(5), pp. 729-745.

Duggan, M., and Smith, A. (2016). 'The Political Environment on Social Media', *Pew Research Center*.

Emanuele, V. (2018). 'Introduction to the Special Issue - Who's the winner? An analysis of the Italian general election', *Italian Political Science*, 13(1).

Emmott, B. (2019). 'Italians wait for the real Matteo Salvini to reveal himself'. [online] Financial Times. Available at: https://www.ft.com/content/c44b1f6c-b9f1-11e9-8a88-aa6628ac896c (Accessed on 29 November 2019).

Enelow, J. M., and Hinich, M.J. (1984). *The spatial theory of voting: An introduction.* New York, New York: Cambridge University Press.

Engle, R., and Yoo, S. (1987). 'Forecasting and testing in co-integrated systems', *Journal of Econometrics*, 35, pp. 143-159.

Evans-Pritchard, A. (2019). 'Italy's Matteo Salvini is a more dangerous threat to the EU outside government'. [online] The Telegraph. Available at: https://www.telegraph.co.uk/business/2019/08/28/italys-matteo-salvini-dangerous-threat-eu-outside-government/ (Accessed on 29 November 2019).

Franzi, A., and Madron, A. (2019). *Matteo Salvini. Italy, Europe and the New Right.* Rome: goWare.

Freschi, F. (2015). 'La voce del leader: strategie comunicative per la creazione del panico morale e del nemico nella pagina Facebook di Matteo Salvini'. *Masters' Thesis: University of Milan.*

Gallego, J., Martinez, J.D., Munger, K., and Vasquez-Cortes, M. (2019). 'Tweeting for peace: Experimental evidence from the 2016 Colombian Plebiscite', *Electoral Studies*, 62.

Garcia, D., Abisheva, A., Schweighofer, S., Serdult, U., and Schweitzer, F. (2017). 'Ideological and Temporal Components of Network Polarization in Online Political Participatory Media', *Policy Internet*, 7(1), pp. 46-79.

Garrett, K. N., and Bankert, A. (2018). 'The Moral Roots of Partisan Division: How Moral Conviction Heightens Affective Polarization', *British Journal of Political Science*, pp.1-20.

Garrett, R.K., Gvirsman, S.D., Johnson, B.K., Tsfati, Y., Neo, R., and Dal, A. (2014). 'Implications of pro-and counter-attitudinal information exposure for affective polarization', *Human Communication Research*, 40(3), pp.309-332.

Gasperetti, M. (2017). 'Chiusero in gabbia due donne rom, licenziati dipendenti della Lidl'. [online] Corriere della Sera. Available at: https://www.corriere.it/cronache/17-aprile-28/chiusero-gabbia-due-donne-rom-3fdb4a22-2c4a-11e7-9a20-9f76d78e4885.shtml (Accessed on 15 May 2020).

Gelman, A. (2007). 'Struggles with Survey Weights and Regression Modelling', *Statistical Science*, 22, pp. 153-164.

Germani, G. (2019). *Authoritarianism, Fascism and National Populism.* New York: Routledge.

Gittleman, J.L., and Kot, M. (1990). 'Adaptation: statistics and a null model for estimating phylogenetic effects', *Systematic Zoology*, 39, pp. 227–241.

Goldbeck, J., Robles, C., Edmondson, M., and Turner, K. (2011). 'Predicting Personality from Twitter'. In *2011 IEEE International Conference on Privacy, Security, Risk, and Trust, and IEEE International Conference on Social Computing.* Boston, MA, pp. 149-156.

Gonzalez-Bailon, S., Kaltenbrunner, A., and Banchs, R.E. (2010). 'The Structure of Political Discussion Networks: a Model for the Analysis of Online Deliberation', *Journal of Information Technology*, 25(2), pp. 230-243.

Google Trends (n.d.). 'Google Trend Search Tool'. [online] Google. Available at: https://trends.google.it/trends/?geo=IT (Accessed on 1 Dec 2020).

Gunn, E.S., and Skogerbø, S. (2013). 'Personalized Campaigns in Party-Centered Politics: Twitter and Facebook as Arenas for Political Communication', *Information, Communication Society*, 16, pp. 757-74.

Hastie, T., Tibshirani, R., and Friedman, J. (2009). 'Overview of supervised learning'. In their (eds.), *The elements of statistical learning*. Springer, pp. 9-41.

Hastie, T., Tibshirani, R., and Friedman, J. (2017). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Stanford: Springer.

Hobolt, S. B., and 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*.

Hobolt, S., Leeper, T., and Tilley, J. (2020). 'Divided by the Vote: Affective Polarization in the Wake of Brexit', *British Journal of Political Science*, pp. 1-18.

Hodas, O.N., and K. Lerman. (2012). 'How visibility and divided attention constrain social contagion'. *Proceedings of 2012 International Conference on Social Computing*. College Park, Maryland.

Holbert, R.L., Garrett, R.K., and Gleason, L.S. (2010). 'A new era of minimal effects? A response to Bennett and Iyengar', *Journal of Communication*, 60(1), pp. 15-34.

Hong, S., and Kim, S.H. (2016). 'Political polarization on Twitter: Implications for the use of social media in digital governments', *Government Information Quarterly*, 33, pp. 777-782.

Horowitz, J. (2019a). 'Italy's Government Faces a Confidence Vote Where Nothing Is Certain'. [online] The New York Times. Available at: https://www.nytimes.com/2019/08/19/world/europe/salvini-italy-confidence-vote-elections.html (Accessed on 29 November 2019).

Huddy, L., Mason, L., and Aarøe, L. (2015). 'Expressive partisanship: campaign involvement, political emotion, and partisan identity', *American Political Science Review*, 109(1), pp. 1-17.

Huddy, L., Sears, D.O., and Levy, J.S. (2013). 'Introduction: Theoretical Foundations of Political Psychology'. In their (eds.), *The Oxford Handbook of Political Psychology*. Oxford: Oxford University Press.

Hughes, A., and Wojcik, S. (2019). '10 facts about Americans and Twitter'. Pew Research Center. [online] Available at: https://www.pewresearch.org/fact-tank/2019/08/02/10-facts-about-americans-and-twitter/ (Accessed on 10 April 2021).

Imai, K. (2017). *Quantitative Social Science*. Princeton: Princeton University Press.

Iyengar, S., and Westwood, S.J. (2014). 'Fear and Loathing across Party Lines: New Evidence on Group Polarization', *American Journal of Political Science*, 59(3), pp. 690-707.

Iyengar, S., Sood, G., and Lelkes, Y. (2012). 'Affect, Not Ideology A Social Identity Perspective on Polarization', *Public Opinion Quarterly*, 76(3), pp. 405-431.

James, G., Witten, D., Hastie, T., and Tibshirani, R. (2017). *An Introduction to Statistical Learning*. New York: Springer.

Kastellec, J.P., Lax, J.R., Malecki, M., and Phillips, J.H. (2015). 'Polarizing the Electoral Connection: Partisan Representation in Supreme Court Confirmation Politics', *The Journal of Politics*, 77(3), pp. 787-804.

King, G., Keohane, R.O., and Verba, S. (1992). *Designing Social Inquiry: Scientific Inference in Qualitative Research*. Princeton: Princeton University Press.

Knobloch-Westerwick, S., and Kleinman, S. (2012). 'Pre-election selective exposure confirmation bias versus informational utility', *Communication Research*, 39(2), pp. 170-193.

Kosinski, M., Bachrach, Y., Kohli, P., Stillwell, D., and Graepel, T. (2014). 'Manifestations of user personality in website choice and behaviour on online social networks', *Machine Learning*, 95(3), pp. 357-380.

Kosmidis, S., and Theocharis, Y. (2020). 'Can Social Media Incivility Induce Enthusiasm? Evidence from Survey Experiments', *Public Opinion Quarterly*, 84, pp.284-308.

Kosmidis, S., Hobolt, S., Molloy, A., and Whitefield, S. (2019). 'Party competition and emotive rhetoric', *Comparative Political Studies*, 52(6), pp. 811-837.

Kott, P.S. (2007). 'Clarifying some Issues in the Regression Analysis of Survey Data', *Survey Research Methods*, 1, pp.11-18.

Lawrence, E., Sides, J., and Farrell, H. (2010). 'Self-Segregation or Deliberation? Blog Readership, Participation, and Polarization in American Politics', *Perspectives on Politics*, 8(1), pp.141-157.

Lazarsfeld, P. F., and Merton, R. (1954). 'Friendship as a social process: A substantive and methodological analysis'. In: Morroe, B., Abel, T. and Page, C. (Eds.), *Freedom and control in modern society*, pp. 18-66. New York, NY: Van Nostrand.

Lechner, D. (2010). 'The Estimation of Causal Effects by Difference-in-Difference Methods', *Econometrics*, 4(3), pp.165-224.

Lee, C.F., and Lee, J.C. (2020). *Handbook Of Financial Econometrics, Mathematics, Statistics, And Machine Learning*. London: World Scientific.

Lee, J., and Lim, Y. (2016). 'Gendered campaign tweets: the cases of Hillary Clinton and Donald Trump', *Public Relations Review*, 42(5), pp. 849-855.

Lelkes, Y., and Westwood, S.J. (2017). 'The Limits of Partisan Prejudice', *Journal of Politics*, 79(2), pp.485-501.

Lenz, G.S. (2009). 'Learning and Opinion Change, Not Priming: Reconsidering the Priming Hypothesis', *American Journal of Political Science*, 53(4), pp. 821-837.

Levendusky, M. (2009). 'The Microfoundations of Mass Polarization', *Political analysis*, 17, pp. 162-76.

Levendusky, M.S. (2013). 'Partisan Media Exposure and Attitudes Toward the Opposition', *Political Communication*, 30(4), pp.565-81.

Levendusky, M.S. (2013). 'Why do partisan media polarize viewers?', *American Journal of Political Science*, 57(3), pp. 611-623.

Livne, A., Simmons, M.P., Adar, E. and Adamic, L.A. (2011). 'The party is over here: Structure and content in the 2010 election'. In *Proceedings of the 5th Int'l AAAI Conference on Weblogs and Social Media (ICWSM)*. Lütkepohl, H. (2005). *New Introduction to Multiple Time Series Analysis*. New York: Springer.

Lütkepohl, H. (2008). 'Impulse Response Function'. In: *The New Palgrave Dictionary of Economics*. New York: Palgrave Macmillan.

Mammone, A. (2019). 'Salvini is turning Italian politics into a right-wing xenophobic monopoly - without action, he'll change my nation for good'. [online] Independent. Available at: https://www.independent.co.uk/voices/matteo-salvini-conte-resign-league-nationalism-anti-immigration-a9074361.html (Accessed on 29 November 2019).

Martin, L.R., and Oeppen, J.E. (1975). 'The Identification of Regional Forecasting Models Using Space: Time Correlation Functions', *Transactions of the Institute of British Geographers*, 66, pp.95-118.

Martini, F. (2019). '"The Beast," an algorithm that tells Salvini if and how "bad" to be'. [online] La Stampa. Available at: https://www.lastampa.it/esteri/la-stampa-in-english/2018/09/10/news/the-beast-an-algorithm-that-tells-salvini-if-and-how-bad-to-be-1.34044116 (Accessed 30 Nov 2019).

Mason, L. (2013). 'The rise of uncivil agreement: Issue versus behavioural polarization in the American electorate', *American Behavioural Scientist*, 57, pp.140-159.

Mason, L. (2015). 'I disrespectfully agree: the differential effects of partisan sorting on social and issue polarization', *American Journal of Political Science*, 59(1), pp. 128-45.

Mason, L. (2018a). 'Ideologues without issues: the polarizing consequences of ideological identities', *Public Opinion Quarterly*, 82(S1), pp. 280-301.

Massaro, T.M., and Stryker, R. (2012). 'Freedom of Speech, Liberal Democracy, and Emerging Evidence on Civility and Effective Democratic Engagement', *Arizona Law Review*, 54.

Mazzoleni, G. and Bracciale, R. (2018). 'Socially mediated populism - the communicative strategies of political leaders on Facebook', *Palgrave communications*, 4(50).

McLoughlin, L., and Ward, S. (2020). 'Turds, traitors and tossers: the abuse of UK MPs via Twitter', *The Journal of Legislative Studies*, 26(1), pp.47-73.

Mencarini, L., Patti, V., Sulis, E., Vignoli, D., Hernandez-Farias, D., Lai, M. (2019). 'Happy parents' tweets: An exploration of Italian Twitter data using sentiment analysis', *Demographic research*, 40(25), pp. 693-724.

Messina, A.M. (2014). 'The political and policy impacts of extreme right parties in time and context', *Ethnic and Racial Studies*, 38(8), pp. 1355-1361.

Miller, P.R., and Conover, P.J. (2015). 'Red and Blue States of Mind: Partisan Hostility and Voting in the United States', *Political Research Quarterly*, 68 (2), pp.225-39.

Ministero dell'Interno (n.d.). 'Archivio storico delle elezioni'. [online] Ministero dell'Interno. Available at: https://dait.interno.gov.it/elezioni (Accessed on 5 May 2020).

Molinaro, E., Kruglanski, A., Jasko, K., Sensales, G., Ganucci Cancellieri, U. (2018). 'Predictors of support for populism in Italy. A preliminary study', *Presented at XV Congresso dell'AIP (Sezione di Psicologia Sociale).*

Montefinese, M., Ambrosini, E., Fairfield, B. and Mammarella, N. (2013). 'Semantic memory: A feature-based analysis and new norms for Italian', *Behaviour Research Methods*, 45, pp. 440-461.

Montefinese, M., Ambrosini, E., Fairfield, B. and Mammarella, N. (2014). 'The adaptation of the Affective Norms for English Words (ANEW) for Italian', *Behaviour Research Methods*, 46, pp. 887-903.

Morales, A.J., Borondo, J., Losada, J.C. and Benito, R.M. (2015). 'Measuring Political Polarization: Twitter shows the two sides of Venezuela', *Chaos: An Interdisciplinary Journal of Non-linear Science*, 25(3), pp.1-10.

Muddiman, A. and Stroud, N.J. (2017). 'News Values, Cognitive Biases, and Partisan Incivility in Comment Sections', *Journal of Communication*, 67(4), pp. 586-609.

Munger, K. (2017). 'Experimentally Reducing Partisan Incivility on Twitter', *Working paper at the NYU Social Media and Political Participation (SMaPP).*

Munger, K., Egan, K., Nagler, J., Ronen, J., and Tucker, J.A. (2017). 'Political Knowledge and Misinformation in the Era of Social Media: Evidence from the 2015 U.K. Election', *Working paper at the NYU Social Media and Political Participation (SMaPP).*

Newman, N., Fletcher, R., Kalogeropoulos, A., Levy, D.A.L., and Nielsen, R.K. (2017). 'Reuters Institute Digital News Report 2017'. [online] Available at https:// reutersinstitute.politics.ox.ac.uk/sites/default/files/Digital%20News%20Report%202017%20 web-0.pdf (Accessed on 10 Jan 2021).

NYU (n.d.). 'The Centre for Social Media and Political Participation (SMaPP)'. [online] *University of New York*. Available at: https://csmapnyu.org/ (Accessed on 14 April 2020).

Ott, B.L. (2017). 'The Age of Twitter: Donald J. Trump and the Politics of Debasement', *Critical Studies in Media Communication*, 34, pp. 59-68.

Oz, M., Zheng, P., and Chen, G. M. (2018). 'Twitter versus Facebook: Comparing incivility, impoliteness, and deliberative attributes', *New Media Society*, 20(9), pp. 3400-3419.

Papacharissi, Z. (2004). 'Democracy online: Civility, politeness, and the democratic potential of online political discussion groups', *New Media Society*, 6(2), pp. 259-283.

Paparo, A. (2018). 'Challenger's delight- the success of M5S and Lega in the 2018 Italian general election', *Italian Political Science*, 13(1).

Pariser, E. (2011). *The filter bubble: What the Internet is hiding from you*. Penguin UK.

Phillips, W. (2016). *This is why we can't have nice things: Mapping the relationship between online trolling and mainstream culture*. Boston: MIT Press.

Prospero, M. (2015). *Il nuovismo realizzato*. Rome: Bordeaux. Quattrociocchi, W., Scala, A., and Sunstein, C. R. (2016). 'Echo chambers on Facebook'. In: Schier, S. E., and Eberly, T. E. (eds.), *Polarized: The Rise of Ideology in American Politics*. London: Rowman Littlefield.

Quercia, D., Kosinski, M., Stillwell, D., and Crowcroft, J. (2011). 'Our Twitter Profiles, Our Selves: Predicting Personality with Twitter'. In *2011 IEEE Third International Conference on Privacy, Security, Risk and Trust and 2011 IEEE Third International Conference on Social Computing*. Boston, MA, pp. 180-185.

Quercia, D., Lambiotte, R., Stillwell, D., Kosinski, M., and Crowcroft, J. (2012). 'The personality of popular Facebook users. *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work (CSCW '12)*. New York, USA, pp. 955-964.

Rachev, S.T., Mittnik, S., Fabozzi, F.J., Focardi, S.M., and Jasic, T. (2021). *Financial Econometrics: From Basics to Advanced Modelling Techniques*. New York: John Wiley Sons.

Reddy, W.M. (2001). *The navigation of feeling: A framework for the history of emotion*. Cambridge: Cambridge University Press.

Reiljan, A., (2016). *Affective Polarization in European Party Systems*. Florence: European University Institute.

Repubblica (2017). 'Licenziati alla Lidl di Follonica, Matteo Salvini si schiera contro l'azienda. [online] Repubblica. Available at: https://firenze.repubblica.it/cronaca/2017/04/29/news/licenziati-alla-lidl-di-follonica-matteo-salvini-si-schiera-con-loro-164210842/ (Accessed 15 May 2020).

Rueda, D., and Stegmueller, D. (2019). *Who Wants What? Redistribution Preferences in Comparative Perspective*. Cambridge (UK): Cambridge University Press.

Salmela, M., and Von Scheve, C. (2017). 'Emotional roots of right-wing political populism', *Social Science Information*, 56(4), pp. 567-595.

Salvini, M. (2017). Io sto con i LAVORATORI (li contatterò già oggi per offrire loro tutto il nostro sostegno, anche legale) e non con le ROM "FRUGATRICI". Ma quanto urla questa disgraziata???. [online] Twitter. Available: https://twitter.com/!/daveharte/status/150631576189931520 (Accessed 15 May 2020).

Salvini, M. (2019). Matteo Salvini. [online] Twitter. Available at: https://twitter.com/matteosalvinimi (Accessed 30 November 2019).

Schöll, N., Gaël, G., and Le Mens, G. (2021). 'Politician-Citizen Interactions and Dynamic Representation: Evidence from Twitter', *Barcelona GSE Working Paper Series*.

Sensales, G., and Di Cicco, G. (2019). 'Populism and immigration in Italian political communication. A comparison of Facebook posts of Di Maio and Salvini (2014- 2018)'. Presented at the EASP Meeting: *"Language Challenges in the 21st Century"*. Poland: Warsaw University.

Shen, C., Li, C., and Si, Y. (2016). 'Spatio-temporal autocorrelation measures for non-stationary series: A new temporally detrended spatio-temporal Moran's index', *Physics Letters*, 380, pp.106-116.

Stroud, N.J. (2008). 'Media use and political predispositions: Revisiting the concept of selective exposure', *Political Behaviour*, 30(3), pp. 341-66.

Stroud, N.J. (2010). 'Polarization and partisan selective exposure', *Journal of Communication*.

Suhay, E., Bello-Pardo, E., and Maurer, B. (2018). 'The Polarizing Effects of Online Partisan Criticism: Evidence from Two Experiments', *The International Journal of Press/Politics*, 23(1), pp. 95-115.

Sunstein, C. R. (2007). Republic.com 2.0. Princeton: Princeton University Press.

Sydnor, E. (2018). 'Platforms for Incivility: Examining Perceptions Across Different Media Formats', *Political Communication*, 35, pp. 97-116.

The Economist. (2018). 'Daily chart: Italian general election poll tracker'. [online] The Economist. Available at: https://www.economist.com/graphic-detail/2018/01/30/italian-general-election-poll-tracker (Accessed 30 Nov 2019).

Thelwall, M., Buckley, K., Paltoglou, G., Cai, D., and Kappas, A. (2010). 'Sentiment strength detection in short informal text', *Journal of the American Society for Information Science and Technology*, 61(12), 2544-2558.

Theocharis, Y., Barberá, P., Fazekas, Z., and Popa, S. (2020). 'The Dynamics of Political Incivility on Twitter', *Sage Open*, pp. 1-15.

Theocharis, Y., Barberá, P., Fazekas, Z., Popa, S., and Parnet, O. (2016). 'A Bad Workman Blames His Tweets: The Consequences of Citizens' Uncivil Twitter Use When Interacting With Party Candidates', *Journal of Communication*, 66, pp. 1007-1031.

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

Tu, Y., Yao, Q., and Zhang, R. (2015). 'Error-Correction Factor Models', *LSE Working paper.*

Tucker, J.A., Guess, A., Barberà, P., Vaccari, C., Siegel, A., Sanovich, S., Stukal, D., and Nyhan, N. (2018). 'Social Media, Political Polarization, and Political Disinformation: A Review of the Scientific Literature', *Hewlett Foundation.*

Twitter (n.d.). 'Terms and regulation'. [online] Available at: https://twitter.com/en/tos (Accessed on 10 April 2021).

Twitter (2016). 'Developer Policies to Protect People's Voices on Twitter'. [online] Available at: https://blog.twitter.com/developer/en$_us$/*topics*/*community*/2016/*developer*−*policies*−*to* − *protect* − *peoples* − *voices* − *on* − *twitter.html*(*Accessedon*10*April*2021).

UK Gov. (2019). 'Online Harms White Paper'. [online] Gov.uk. Available at: https://www.gov.uk/government/consultations/online-harms-white-paper (Accessed on 29 November 2019).

Villa, M. (2016). 'Les discours sur l'Europe dans la tweet-campagne en Italie'. In: Frame, A., Mercier, A., Brachotte, G., Thimm, C. (eds.), *Tweets from the Campaign Trail: Researching Candidates' Use of Twitter During the European Parliamentary Elections*. Paris: Peter Lang AG.

Wallis, K.F. (1987). 'Time Series Analysis of Bounded Economic Variables', *Journal of Time Series Analysis*, 8(1), pp. 115-23.

Webster, S.W., and Abramowitz, A.I. (2017). 'The Ideological Foundations of Affective Polarization in the U.S. Electorate', *American Political Research*, 45(4), pp. 621-647.

Weinberger, D. (2011). *Too Big to Know: Rethinking Knowledge Now That the Facts Aren't the Facts, Experts Are Everywhere, and the Smartest Person in the Room Is the Room*. London: Basic Books.

Wells, C., Shah, D.V., Pevehouse, J.C., Yang, J., Pelled, A., Boehm, F., Lukito, J., Ghosh, S., and Schmidt, J.L., (2016). 'How Trump Drove Coverage to the Nomination: Hybrid Media Campaigning', *Political Communication*, 33, pp. 669-76.

Wu, S., Hoffman, J. M., Mason, W.A., and Watts. D.J. (2011). 'Who says what to whom on Twitter', *Proceedings of 20th International Conference on Worldwide Web. ACM*, pp. 705-14.

Wu, T. (2017). 'How Donald Trump Wins by Losing'. [online] New York Times. Available at https://www.nytimes.com/2017/03/03/opinion/sunday/how-donald-trump-wins-by-losing.html (Accessed on 4 Jan 2021).

Zou, H., and Hastie, T. (2005). 'Regularization and Variable Selection via the Elastic Net', *Journal of the Royal Statistical Society*, 67(2), pp. 301-20.

CHAPTER 10

APPENDICES

10.1 Appendix A: Users' rights and privacy

This appendix explores Twitter users' rights and privacy guidelines the paper has strictly followed. Salvini's tweets and responses have been collected via the Twitter streaming application program interface (API). The Twitter API allows researchers to retrieve and work with data following global standards imposed by Twitter, and international governing bodies (.e.g., the European Union). Access to the Twitter API has been granted for this project, after a written interview and several checks on the objectives of this paper by Twitter and an independent panel. Given the January 2021 Twitter API regulations, this paper has achieved full access to Twitter data. The collected data has been stored in a safe and encrypted environment, Apple drive (iCloud), which has only been accessible by finger-print scan, and/or password by the hereby researcher. As users register on Twitter, they allow Twitter to access their information, and such information becomes available to developers (e.g., the hereby researcher) unless the users decide to use private accounts. Users who utilise private accounts have not been included in this study. To guarantee users' anonymisation, only user IDs (i.e., Twitter-randomly-generated users' anonymous identification numbers) have been utilised. In other words, the collected data will never point to a specific individual, screen names and other individual-specific data will not be retrieved. For these reasons, there is no risk of breaching users' privacy and rights according to Twitter's norms and regulations, and the European Union's 'General Data Protection Regulation'.

10.2 Appendix B: Web-scraping details

Twitter data have been collected through a combination of Python libraries, such as open-source snscrape (n.d.) and tweepy (n.d.). Before programming the retrieval algorithms, this paper has set several Python sleep functions, stopping the kernel's tasks every 5 minutes, and, after 5 full cycles (i.e., operation-sleep-operation), the kernel would stop for the other 25 minutes to preserve the machine's CPU before operating again. Having set the kernel's 'sleeping' cycles, this paper has programmed a tweet-link retrieval code through the snscrape library. The scrutiniser has extrapolated the tweets and replies' IDs based on the fact that every link contains repetitive patterns (i.e., "https://www.twitter.com/statuses/") followed by tweet and replies' ID (i.e., randomly assigned number to identify any Twitter activity).

10.3 Appendix C: Summary of hashtags details

Table 10.1: Summary 1/2 of hashtags details

Hashtag(s)	Details
'azzurri', 'balotelli', 'milan', 'maldini', 'seriea'	Non-political. These hashtags refer to Italian football. Respectively, they refer to the Italian national team (azzurri); footballers Mario Balottelli (balotelli) and Paolo Maldini (maldini); football team AC Milan (milan); and Italian Top football League (seriea).
'buongiorno', 'pizza', 'tramonto'	Non-political. These hashtags refer to Italian daily life and food. Respectively, they refer to 'Good Morning' wishes (buongiorno); pizza, food (pizza); sunset (tramonto).
'Olimpiadi', 'bebevio', 'pellegrini', 'fognini', 'settebello', 'rugby',	Non-political. These hashtags refer to Olympic games sports. In particular, they refer to the Olympic games (olimpiadi); Olympic Fencing Champion Bebe Vio (bebevio); Swimming Champion Federica Pellegrini (pellegrini); Tennis Champion Fabio Fognini (fognini); Italian Water polo team (settebello); Rugby sport (rugby).
'dovizioso', 'ferrari', ''motogp', 'simoncelli'	These hashtags refer to motor sports. In particular, they refer to MotoGP Driver Andrea Dovizioso (dovizioso); Formula 1 Scuderia Ferrari (ferrari); MotoGP official hashtag (motogp); MotoGP Driver Marco Simoncelli (simoncelli).
'amici', 'gfvip', 'grandefratello', 'mahmood', 'sanremo', 'sanremo19', 'sanremo2014', 'sanremo2015', 'sanremo2017', 'sanremo2018', 'sanremo2019'	Non-political. These hashtags refer to Italian entertainment and music programmes. In particular, they refer to Amici talent-show (amici); Big Brother show (gfvip, grandefratello); Italian music festivals (sanremo); winner of 2019 Sanremo Mahmood (mahmood).

Note: This tables summarises hashtags' details.

Table 10.2: Summary 2/2 of hashtags details

Hashtag(s)	Details
'berlusconi', 'meloni', 'zaia'	Political. These hashtags refer to Matteo Salvini'sallies. In particular, they refer to former Prime Minister of Italy Silvio Berlusconi (berlusconi); Leader of Brothers of Italy Giorgia Meloni (meloni); Fellow Lega Nord member Luca Zaia (zaia).
'lega', 'votalega', 'salvinipremier', 'salvini', 'primagli-italiani',	Political. These hashtags refer to Matteo Salvini's self-promotion. Specifically, they refer to Salvini's party, Lega Nord (lega); Vote for Lega (votalega); Salvini for Prime Minister (salvinipremier); Salvini's persona (salvini); Lega's slogan 'Italians First' (primagliitaliani).
'boldrini', 'dimaio', 'grillo', 'm5s', 'pd', 'renzi', 'repubblica'	Political. These hashtags refer to Matteo Salvini's major opponents. Specifically, they refer to Democratic Party member Laura Boldrini (boldrini); Leader of the Five-Star Movement Luigi Di Maio (dimaio); Italian Democratic Party (pd); former Prime Minister of Italy Matteo Renzi (renzi); top opposing newspaper Repubblica (repubblica).
'bibbiano', 'clandestini', 'immigrati', 'madeinitaly', 'mes', 'migranti', 'terrorismo', 'terroristi', 'unionicivili'	Political. These hashtags refer to Italian major scandals and policies. Notably, they refer to the Bibbiano scandal (bibbiano); Migration and Immigration issues/policies (clandestini, immigrati, migranti); Terrorism and terrorists (terrorismo, terroristi); Made in Italy (madeinitaly); Civil Unions for same-sex couples marriage (unionicivili).
'elezioni', 'europa', 'governo', 'politica', 'ue','senato'	Political. These hashtags refer to political and institutional bodies. In particular, they refer to the elections (elezioni); Europe and European Union (europa, ue); Government (governo), Politics (politica); Italian Senate (senato).

Note: This tables summarises hashtags' details.

10.4 Appendix D: Computational details

Given the computing power required to conduct the neural network textual classification, and generate a dummy for each topic, these tasks were performed on an Amazon remote server, while strictly following the privacy procedure in Appendix A. This methodology was followed for two main reasons. First, Amazon remote server service allows researchers to operate with remote machines able to perform, on average, 60,000 interactions per second on a given textual body, versus a commercial computer which can only perform between 10,000 and 20,000 interactions per second. Therefore, the Amazon server allowed this paper to achieve textual classification in a relatively short time. Second, this system allows researchers to work from their computers without sharing data with other people or entities to respect users' privacy and rights as discussed in Appendix A.

10.5 Appendix E: Elastic Net

This Appendix discusses the regression model utilised to validate the NN: Elastic Net (EN). When certain linear models' requirements are met (e.g., constant variance), we can assume that the estimated coefficients are unbiased and have the lowest variance (Boehmke and Greenwell, 2020). However, as big data contain large numbers of features, linear models' assumptions often may no longer hold (ibidem). Therefore, the models may tend to overfit the data, leading to an increase in the sampling error. To account for this occurrence, social scientists utilise regularisation techniques to improve linear models' predictions for big data analysis, namely ridge and lasso (ibidem; Zou and Hastie, 2005).

The ridge regression model minimises model complexity through a penalty parameter, which is calculated as the square of the value of the coefficient (Boehmke and Greenwell, 2020). The size of the penalty is expressed by the tuning parameter (lambda). When is equal to 0, one can assume that there is no effect, (as per a simple linear model) our model minimises the sum of squared errors (SSE). Instead, when the value of increases, the penalty becomes large and forces the coefficients toward zero (yet not completely). In this way, we shrink our estimates, increasing prediction performance. However, employing a ridge penalty presents one major disadvantage. Such a regression model includes all the predictors in the final model, making results interpretation difficult.

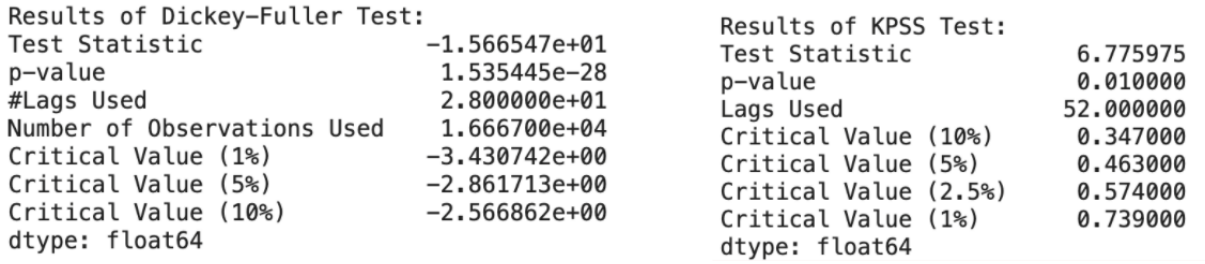
Alternatively, a lasso (least absolute shrinkage and selection operator) regression model also imposes a penalty on the coefficients (Zou and Hastie, 2005; Boehmke and Greenwell, 2020). Whilst ridge reduces variables to approximately, yet not equal to, zero; lasso forces all the coefficients to zero. Moreover, lasso performs an automated feature selection on the features with the largest and consistent signal (ibidem). In this way, we improve our model and sample the most significant variables. However, lasso presents mainly two and related issues. First, when a group of variables is highly correlated, the lasso model pushes all of them towards

zero; however, one variable may be forced fully to zero, while the others only approximately reach zero, remaining in the model. Second, the selection of the variables is not systematic, the lasso does not specify why one of the correlated variables is selected.

As we have seen, both ridge and lasso present considerable shortcomings. To take advantage of the best features of both ridge and lasso, whilst minimising their disadvantages, I will utilise an EN regression model (Zou and Hastie, 2005). By definition, an EN regularisation combines the ridge's effective regularisation with the lasso's variable selection (ibidem). Besides setting the lambda parameter in our model, we also set an alpha parameter which represents the mixing parameter between ridge ($\alpha = 0$) and lasso ($\alpha = 1$). Therefore, for the elastic model, we achieve an interval for alpha ($0 < \alpha < 1$). Doing so allows us to tune a lambda for a fixed value of alpha. The main intuition behind the elastic model is that it allows researchers to combine the benefits of ridge and lasso to improve linear models' predictions.

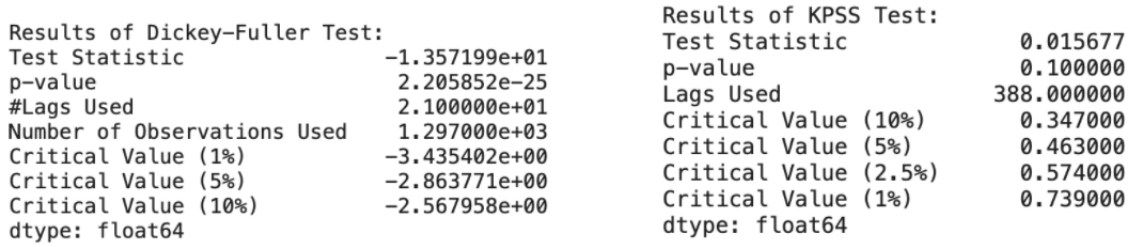
10.6 Appendix F: Tweets' polarisation time-series - stationarity tests

Figure 10.1: Stationarity tests for Salvini's tweets' polarisation time-series



Note: This figure displays stationarity tests for Salvini's tweets' polarisation time-series, respectively ADF (p-value < 0.05) and KPSS (p-value < 0.05). The tests show that the time-series needs first-differentiation because the time-series is difference stationary.

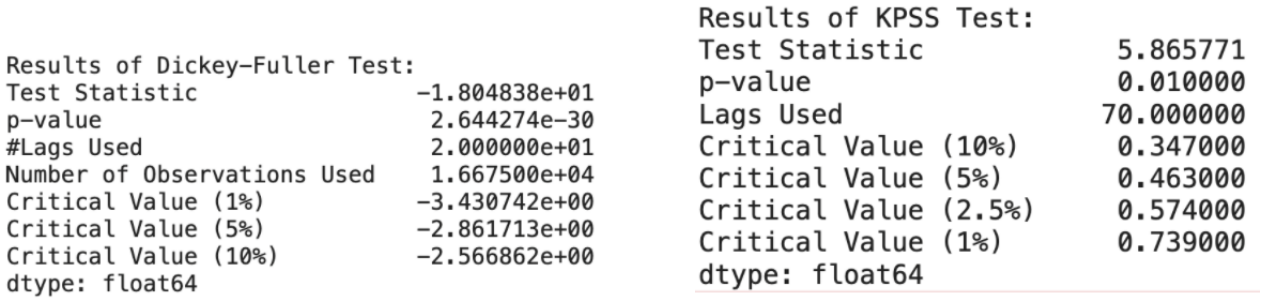
Figure 10.2: Stationarity tests for differentiated Salvini's tweets' polarisation time-series



Note: This figure displays stationarity tests for differentiated Salvini's tweets' polarisation time-series, respectively ADF (p-value < 0.05) and KPSS (p-value > 0.05). The tests show that the differentiated time-series is stationary.

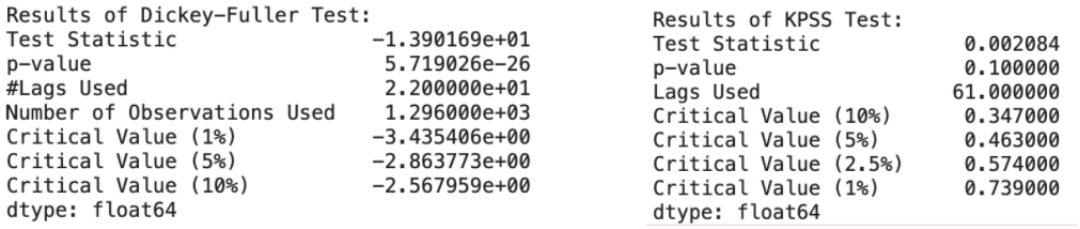
10.7 Appendix G: Replies' polarisation time-series - stationarity tests

Figure 10.3: Stationarity tests for replies' polarisation time-series



Note: This figure displays stationarity tests for replies' polarisation time-series, respectively ADF (p-value < 0.05) and KPSS (p-value < 0.05). The tests show that the time-series needs differentiation because the time-series is difference stationary.

Figure 10.4: Stationarity tests for differentiated replies' polarisation time-series



Note: This figure displays stationarity tests for differentiated replies' polarisation time-series, respectively ADF (p-value < 0.05) and KPSS (p-value > 0.05) .The tests show that the differentiated time-series is stationary.