



HAL
open science

Is Twitter just a Media Amplifier? Evidence from a data-driven, comparative Analysis of German Federal Elections 2017 and European Elections 2019

Kimmo Elo

► **To cite this version:**

Kimmo Elo. Is Twitter just a Media Amplifier? Evidence from a data-driven, comparative Analysis of German Federal Elections 2017 and European Elections 2019. 2019. halshs-02425856

HAL Id: halshs-02425856

<https://shs.hal.science/halshs-02425856>

Preprint submitted on 31 Dec 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Is Twitter just a Media Amplifier? Evidence from a data-driven, comparative Analysis of German Federal Elections 2017 and European Elections 2019

Kimmo Elo*

31st December 2019

Abstract

Despite its rather modest number of users, Twitter has gained a certain status as platform for on-time political communication and recent studies have brought evidence, that significant events tend to produce peaks in Twitter activity. This article focuses on the interaction between Twitter and “traditional” mass media channels (newspapers, magazines, TV) by analysing topics central in Twitter communication on politics. The data-driven analysis presented in the article analyses the structure and dynamics of topics discussed on Twitter during elections. The data used in this article are from two political events, the German Federal Elections 2017 and European elections 2019. Both datasets contain more than one million tweets. Empirically, the article seeks to evidence the usefulness of Twitter data when it comes to gain understanding of actual political debates in a society. Methodically, the article seeks to present how data mining, modelling and visualisation tools can be exploited in order to extract scientifically relevant, reliable and significant, and analytically interesting results from a very large dataset to be used to answer questions being relevant for political and social sciences and media studies as well.

Keywords: Data analysis; EU; Germany; Network analysis; Political communication; Topic modelling; Twitter

1 Introduction

Internet-based social media forums (Facebook, Twitter, Instagram etc.) have gained a rather established status as forums for digital politics. These forums are not just political spaces where users simply comment on political events and issues.¹ Instead, these platforms can also be seen as platforms of political action. In this article I use two other concepts to tackle political action. *Politicking* refers to actions carried out within a certain political space – in this

*Centre for Parliamentary Studies, University of Turku, FIN-20014 Turku, Finland; E-Mail: kimmo.elo@utu.fi

¹For a good overview of digital politics, see Coleman & Freelon 2015

article the Twitter – and aiming at the increase in the weight of certain political issues. *Politicization*, in turn, refers to action aiming at creating new space for politicking by marking a phenomenon as political, by detecting its political potential and thus expanding the presence of the political.² Digital politics seem to increase the speed of both politicking and politicalisation especially in issues and questions having a strong societal and political impact on a large number of people. Social media platforms play an important role in this process, since these enable events, debates and issues to spread out rapidly in a global scale. In the literature the quick-cyclic nature of social media communication is often referred to as ‘virality’ and seen as very fundamental for digital communication.³

This article focuses on Twitter communication around two major political events, the German federal elections in September 2017 and the European Parliamentary elections in May 2019. Twitter has, despite its rather limited number of users when compared to Facebook, gained a rather firm status as an important forum for real-time online commenting on current societal and political issues and discussion. From this perspective it is interesting to analyse, whether Twitter discussions actually reflect and amplify wider political debates in the society.

In order to do that the article analyses in both cases the last four (4) weeks before the official election day. In order to capture topics and themes I focus on the use of the so-called ‘hashtags’ used to connect a tweet with a certain topic. The underlying idea is rather simple and based on the assumption that hashtags serve the function to politicking and politicalisation, but also to create a certain hierarchy within a major topic, here the elections. As regards ‘politicking’ the use of the top-level hashtag – in the context of this article the ‘official’ hashtag(s) for both elections – is interpreted as a user’s action to link her tweet with the elections. As regards ‘politicalisation’ the use of other hashtags in co-occurrence with the above-mentioned top-level hashtag brings in a sub-topic to this general debate. Hence, by using other hashtags in co-occurrence with the top-level, general hashtag, the user focuses her communication by linking it to certain more specific topics. This analytical approach gains support from recent studies having evidenced that hashtag co-occurrences are a reliable base to identify and analyse thematically different sub-discussion under a certain general hashtag (like #metoo).⁴

In this article Twitter is considered as a digital extension of public space, an online, real-time platform for political communication. Hence, Twitter is not separate, nor disconnected from public discussions in polity, the political sphere for the struggle of possible futures by political activities.⁵ Against this background the political communication in Twitter is expected to reflect, even amplify wider public debates. Consequently, a thematic analysis of political communication in Twitter is expected to show a strong correlation with wider public discussion, especially with those present in other public media. However, Twitter is also expected to serve as an amplifier strengthening selected public discussions and debates in the digital realm. This article seeks to bring out evidence for both of these background assumptions.

This article understands it first and foremost as a methodological contribution

²Palonen, 2007, 42, 62-63, 65-66

³Especially Nahon & Hemsley 2013; Wang et al. 2016

⁴Feicheng & Yating 2014; Wang et al. 2016

⁵Palonen, 2007, 41-42, 59, 63

and seeks to exemplify how selected digital methods can be applied to a large dataset in order to obtain scientifically interesting and relevant results. My approach is exploratory, since the article seeks to explore patterns hidden in the data. However, the article has a strong empirical aspect seeking to improve our understanding about the role of Twitter communication as a modern form of political communication.

The structure of the article is as follows. The first section introduces the data collection procedure, the obtained datasets used in the analysis, and the methods applied. The next section presents and discusses the analytical results. The article is rounded up with a concluding section.

2 Data and Method

All results presented in this article are obtained by applying digital tools and methods to two large datasets, the first collected during the German federal elections in 2017, the second during the European Parliamentary elections in 2019. Both datasets consist solely of Twitter messages (i.e. tweets) containing at least one of the official hashtags used to tag the elections. The data collection was done via the Twitter API⁶ enabling a real-time streaming of tweets from the Twitter. The streaming itself was carried out by a small computer script written with the programming language Python and its ‘tweepy’ library⁷. The Twitter API returns a JSON formatted⁸ record for each tweet and these records were stored locally in a single file.

The first dataset consists of tweets collected during the German federal elections in September 2017. The original full dataset covers the period between 1 March 2017 and 31 December 2017. The hashtag filter used to collect tweets consisted of the following ‘official’ hashtags: #btw2017, #btw17, #bundestagswahl2017. For this paper, I broke off a datasets covering the four (4) weeks before the official election day, 24 September 2017. The most important structural characteristics of this dataset are:

- $N_{tweets} \approx 665.800$, $N_{users} \approx 171.000$
- User activity: 3.0 (mean), 1.0 (median), [1... 3590]
- Original tweets: ≈ 230.000 (35 %)

As Figure 1 shows the tweeting activity increases continuously during the last four weeks before the official election day, 24 September 2017. Thus, we can expect to be able to identify the most important topics and themes during this final stage of the election campaign.

The second dataset consist of tweets collected during the European Parliamentary elections in May 2019. The original full dataset covers the period between 1 February 2019 and 23 June 2010 (i.e. four weeks after the election day). One problem with these elections was the fact that the EU is a multi-lingual political space with 24 official languages. Since my own interest focuses on

⁶<https://developer.twitter.com/>

⁷<http://www.tweepy.org/>

⁸<https://json.org/>

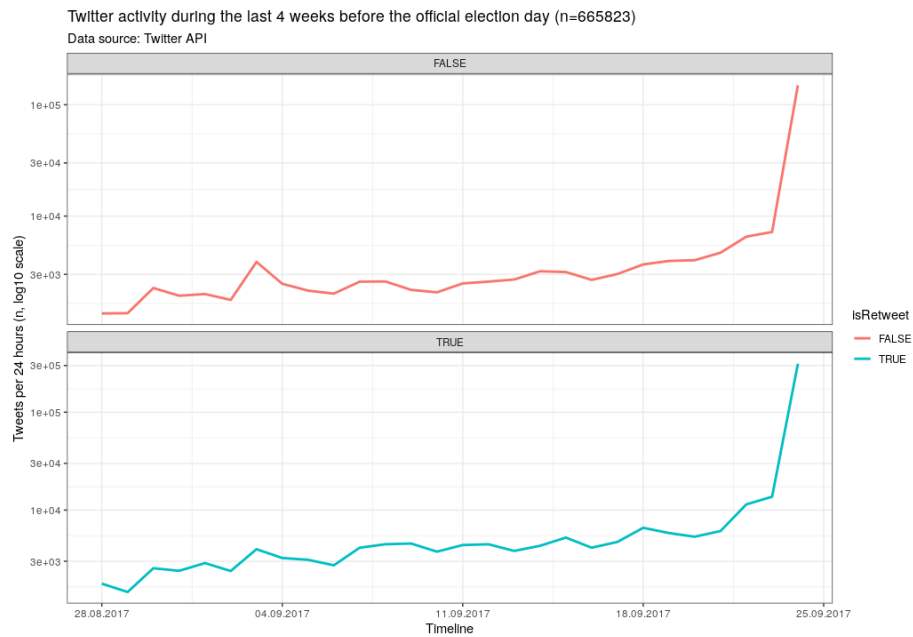


Figure 1: Twitter activity (original tweets and retweets) during the last four weeks before the German federal elections.

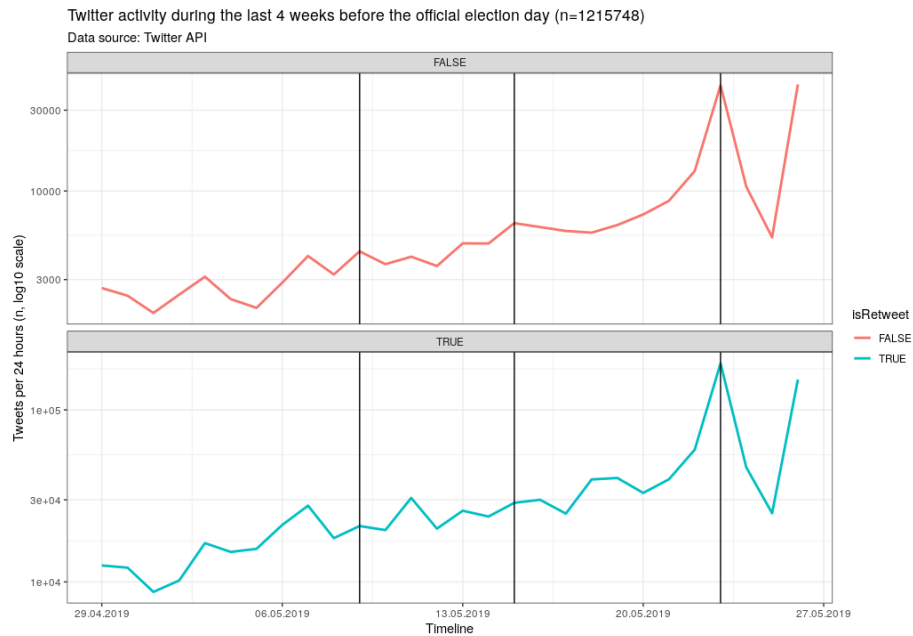


Figure 2: Twitter activity (original tweets and retweets) during the last four weeks before the European Parliamentary elections.

supranational political communication I decided to limit the data collection to Europe’s *lingua franca*, i.e. to English. This data reduction was carried out in two steps. First, the hashtag filter used to collect tweets was instructed to filter by the following hashtags: #epelections2019, #euelections2019, #ep2019, #thistimeinvoting. The last hashtag was a general hashtag used by an initiative seeking to encourage people to vote. In the second step the dataset was reduced by filtering by language, stored in the ‘language’ field in each record. The dataset was filtered to include only tweets written in English. The most important structural characteristics of this second dataset are:

- $N_{tweets} \approx 1.121.600$, $N_{users} \approx 295.000$
- User activity: 2.8 (mean), 1.0 (median), [1...683]
- Original tweets: ≈ 213.000 (18 %)

As regards the tweeting activity the European Parliamentary elections show a similar pattern like we saw in the German case, that is the tweeting activity increases during the last four weeks before the official election day (Figure 2). The peaks marked in the timeline include the European Day on 9 May 2019, the Spitzenkandidat debate on 15 May 2019, as well the brexit votings in the British parliament on 22 May 2019. Hence, also in the case of the European Parliamentary elections we can rather safely expect to be able to identify the most important topics and themes during by analysing Twitter communication during this final stage of the election campaign.

These two datasets were processed further with a specific command-line tool ‘jq’ to extract only the fields relevant for the analysis from the feature-rich JSON data. The cleaned datasets were then imported in and processed further with the statistical software R⁹. All descriptive statistical analyses have been made with R. Additionally, the network analysis and visualisation software ‘visone’¹⁰ has been used to analyse and visualise network data based on hashtag co-occurrences.

A central claim of this article is that Twitter data can be used to obtain reliable understanding of thematic patterns covered in digital debates on a certain political topic, here national and European elections. This can be achieved by dedicated exploratory digital tools, designed to mine large data in order to identify structural patterns and dynamics.

In this article I apply two different digital tools. The first tool is based on social network analysis (SNA) and focusing on hashtag co-occurrences. The main aim is to gain understanding of patterns forming distinct clusters and to interpret the clusters against current political debates. In other words, this analysis is designed to shed light on general thematic patterns having been typical for the analysed Twitter communication.

The analysis process was carried out in the following six steps. First, the dataset was filtered to include tweets containing at least two hashtags. In the second step each hashtags occurring in the same tweet were connected to all other hashtags in the same tweet. Third, hashtag pairs containing one of the top-level

⁹www.r-project.org

¹⁰www.visone.info

hashtags were removed. This step was necessary in order to shift the focus to sub-discussions. This network data preparation was completed with R. For the next steps the network data was imported in ‘visone’ for further analysis. In the fourth step the initial co-occurrence network was created and all hashtag pairs occurring only once were removed. Further, all hashtags co-occurring with less than five other hashtags were removed. This step was necessary in order to improve the reliability of cluster analysis, which was applied in the step three. This not only reduced the network in size, but also helped to focus the analysis on core patterns typical for the communication. In the fifth step the network was clustered by applying Louvain clustering method to the network. Since visone has a built-in support for Louvain method, the application was rather straightforward. Although many other clustering methods are around, Louvain has proven to be rather robust and reliable method for network clustering¹¹. The underlying assumption here was that tweets sharing a topic also share hashtags, so that an analysis of hashtag co-occurrence patterns can bring out thematic differences. In the last, six, step the clustered network was visualised as a so-called Voronoi map organising the clustered network in a visually pleasant way making the identification of clusters rather simple. Another positive thing in Voronoi maps is that it makes it rather easy to understand the cluster structure hidden in the data, plus to compare clusters in size and content. Voronoi maps were created to visualise the whole period of four weeks, since this analysis seeks to identify the main topics during the last four weeks before the election day.

The second digital tool applied in this article is *term frequency, inverse document frequency* analysis (tf-idf) widely used in text mining. This analysis focused on weekly communication and sought to identify hashtags typical for one, but not for all four weeks. Generally speaking, the very idea of this measure is to decrease the weight for commonly used words and to increase the weight for words that are not used very much in a collection of documents.¹² In this article *tf-idf* method was used to increase the weight of hashtags typical for one week and to reduce the weight of commonly used hashtags. The main aim of this approach is to gain understanding about weekly changes in themes and topics.

To sum up, the methods applied in this article offer two different views on the same communication. The first, network analysis and clustering, seeks to grasp the overall thematic structure, that is the topics having dominated the discussion during the last four weeks before the election day. The second, *tf-idf* analysis, digs deeper and sheds light on weekly discussions and changes over time.

3 Results

3.1 German Federal Elections 2017

This article seeks to find evidence for the hypothesis that in political communication Twitter is mainly used to amplify ongoing discussions. In order to test this hypothesis a comparative approach is used, in which topics and themes found in

¹¹Blondel et al. 2008

¹²A good description of the method can be found here: https://www.tidytextmining.com/tfidf.html#the-bind_tf_idf-function

Twitter discussions are reflected against topics having been in the media. Since recent studies have evidences that Twitter debates are rather ‘peaky’ and quick-living¹³, it seems rather safe to assume that - if the hypothesis is confirmed - there should be a rather strong correlation between topics present in Twitter and topics present on other media platforms.

As regards the German federal elections 2017 I decided to re-construct main public debates based on the weekly political magazine *Der Spiegel*. The magazine is a leading weekly magazine in the German-speaking Europe and covers all relevant political, societal and economic affairs. Further, *Der Spiegel* can be seen as an important influencer and opinion-builder among the public. Although the magazine is a quality magazine and, hence, more a magazine for the well-educated people, it can be considered a reliable source for the re-construction of the main political discussions in Germany during the four weeks before the election day.

Quite understandable the ongoing election campaign dominated the media landscape. *Der Spiegel* followed parties on election rallies. In the issue 35/2017 (26 August 2017) the one single TV debate between Chancellor Angela Merkel and her challenger Martin Schulz was a central topic, followed by a report on the online campaigning of the right-wing populist AfD (*Alternative für Deutschland*), which had hired consults and advisers from the USA to help to build an anti-Merkel campaign on the social media. The title story of the issue 36/2017 (2 September 2019) was dedicated to the unbalanced struggle between Merkel and Schulz, giving Merkel the position of an almost absolute favourite. The report was accompanied with a *Spiegel*-Interview with Chancellor Merkel. The issue 37/2017 (9 September 2017) tackles with its title story the political polarisation, the rise of populism and anti-Merkelism, as well the hate speech and disinformation campaigning in the Internet. The last issue considered here is the issue 38/2017 (16 September 2017), published one week before the election day. This issue is dedicated to the Federal elections and tackles questions related to possible coalitions and the high number of undecided and/or swing voters. This issue also contains the major interview with Merkel’s challenger Martin Schulz. Included in this issue is also a critical evaluation of the outgoing grand coalition, in which promises made are reflected against the actual outcomes.

In addition to these weekly topics, German public discussion were during the final stage of the federal elections overshadowed by the refugee crisis. The crisis itself started already in 2015 as a very high number of refugees came to Europe. In Germany, Chancellor Merkel’s decision to welcome refugees quickly resulted in a political split in a nationalist-conservative and a liberal-cosmopolitan blocks. The former was actively fed by the right-wing populist party AfD harshly attacking Chancellor Merkel’s refugee policy - especially in the social media and other online platforms.

In order to gain understanding about general topics during the last for weeks of the German federal election campaigns I experimented with an idea to apply network analysis to gain understanding about hashtag clusters hidden in Twitter data. My approach focuses on hashtag co-occurrences and is rooted in the idea that users participating in political debates in Twitter use hashtags to politicalise their tweets, that is to connect a tweet to a certain political topic

¹³##VIITE SEN SAKSALAISEN (KONSTANZ) TUTKIMUKSIIN, JOSSA TÄMÄ PIIKIKKYYS-IDEA!!

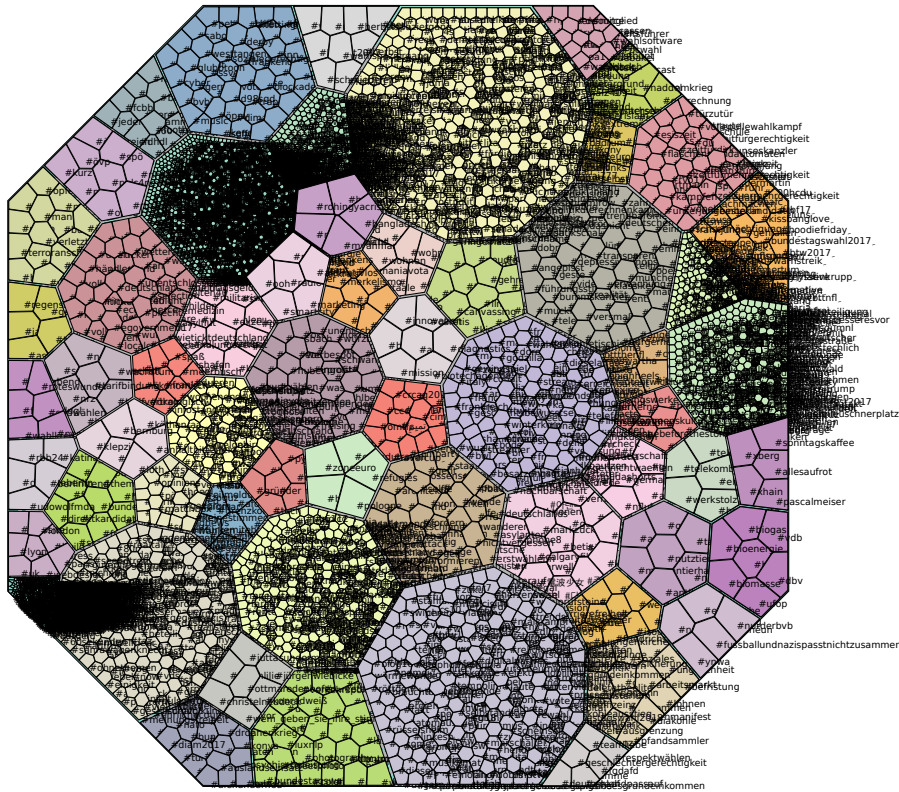


Figure 3: Thematic clusters based on hashtag co-occurrences during the last four weeks before the German federal elections (layout: Voronoi treemap). Five (5) strongest clusters are marked.

by the targeted use of hashtags. Hence, we can assume that similar topics are tagged with similar hashtags, so that they form reliable clusters and can, thus, be identified by a dedicated clustering algorithm. In other words, I focus on repetitive patterns in the data and use their connectivity and the intensity of co-occurrences as the only criteria for their belonging together and to identify salient, most influential hashtags for each topic.¹⁴ The network data used in the cluster analysis was created as described in the previous section.

The main findings of this analysis are presented in Figure 3 and Table 1. The Voronoi layout used in Figure 3 visualises each cluster as a distinct region in the graph. In order to improve the readability of the graph clusters having less than five hashtags have been removed. On the one hand, we can rather easily see how many clusters are in the data - the exact number of clusters (topics) is 58. On the other, it is rather easy to compare the sizes, that is the number of hashtags, of different clusters (topics).

The results from the cluster analysis indicate that Twitter communication during the last four weeks of the German federal elections 2017 revolved around five main topics. A closer look to the hashtag structure brings up evidence

¹⁴Paranyushkin 2011; Shim et al. 2015

Table 1: German federal elections 2017: five (5) main clusters and their most influential hashtags.

Cluster #1: AFD&Anti-Merkel ($n_{HT}=1057$)	Cluster #2: Refugee crisis ($n_{HT}=403$)	Cluster #3: Red&green ($n_{HT}=286$)	Cluster #4: Climate ($n_{HT}=230$)	Cluster #5: Democracy ($n_{HT}=180$)
#afd	#deutschland	#r2g	#deinewahl	#demokratie
#noafd	#eu	#wagenknecht	#darumgruen	#freiheit
#traudichdeutschland	#islam	#pegida	#dieselgate	#abmerkeln
#merkelmussweg	#fluechtlinge	#refugeeswelcome	#diesel	#deinestimme
#fckafd	#denkenwirnein	#gruenen	#energiewende	#gesellschaft
#afdwaehlen	#europa	#gysi	#klimaschutz	#rettetdiegrundrechte
#weidel	#asyl	#bartsch	#umwelt	#vielfalt
#gauland	#terror	#g20	#dieselskandal	#frauen
#schicksalswahl	#muslime	#oppermann	#klimawandel	#demo
#nichtmeinekanzlerin	#sicherheit	#kippling	#grün	#gerechtigkeit

that the clusters are thematically distinct (Table 1). I have analysed the main clusters and labelled each by an appropriate, descriptive label. The first and largest topic consists of 1057 hashtags and has its clear focus on the Anti-Merkel rhetoric of the right-wing populist party *Alternative für Deutschland* (AfD, Alternative for Germany). The hashtags #merkelmussweg (Merkel must go) or #nichtmeinekanzlerin (not my chancellor) are good examples of this anti-Merkel communication. The two other hashtags #weidel and #gauland refer to the AfD party leaders Alice Weidel and Alexander Gauland. It is, however, interesting to note that there is a certain counter-argumentation present in this cluster, manifested by the hashtags #noafd and #fckafd.

The second main cluster with 403 hashtags connects to discussions on the refugee crisis. Here the most influential contextual hashtags include #islam, #flüchtlinge (refugees), #asyl (asylum) and #muslime (Muslims). This very much reflects the mainstream of other public debates with a strong focus on Islam and Muslims during the German federal elections. Further, the hashtags #terror (terror/ism), #sicherheit (security) and #europa (Europe) seem to reflect the insecurity generated by different acts of terrorism in Europe after 2015.

The third topic with 286 hashtags earns the label ‘red & green’, since the main content seems to be connected to themes and questions tackled by the Left Party (*Die Linke*) or the Greens (*Die Grünen*). This interpretation is also supported by the fact, that the hashtag #r2g (red and/to green) is the most influential hashtag in this cluster. Further, the hashtag #refugeeswelcome connects to a slogan widely used by not only German, but also European left and green parties during the refugee crisis. The hashtags #wagenknecht (Sahra Wagenknecht), #gysi (Gregor Gysi), #bartsch (Dietmar Bartsch) and #kipping (Katja Kipping) refer to top figures of the Left Party. We can also see references to the G20 Hamburg Summit (#g20), as well to the Greens (#gruenen) among the most influential hashtags in this cluster.

The fourth cluster with 230 hashtags is thematically connected to climate discussions. We should, however, keep in mind, that in 2017 the climate debate had a different focus on intensity as it did in 2019 (see below). Although we can identify references to climate change (#klimawandel), this topic seem to have to main focus. First, the hashtags #dieselgate, #diesel and #dieselskandal refer to a scandal dating back to 2015 as one of Germany’s leading car manufacturer, Volkswagen, confessed to have installed illegal software to manipulate emission test results. The scandal quickly spread out to all leading German manufacturers and resulted in overheated public debates not just in Germany, but also in Europe and the USA.¹⁵

The last main topic with 180 hashtags is more general by nature. A closer analysis of the hashtags seem to allow the interpretation, that this cluster consists of value-oriented discussions on democracy, freedom and liberty. I justify this interpretation by highlighting hashtags like #demokratie (democracy), #freiheit (freedom), #rettetdiegrundrechte (save the basic rights), #vielfalt (diversity) or #gerechtigkeit (justice).

The previous analysis focused on the top discussions during the last four weeks. As the results indicated the topics having dominated Twitter discussions well

¹⁵See <https://www.cleanenergywire.org/factsheets/dieselgate-timeline-germanys-car-emissions-fraud-scandal> (last visited: 1 September, 2019)

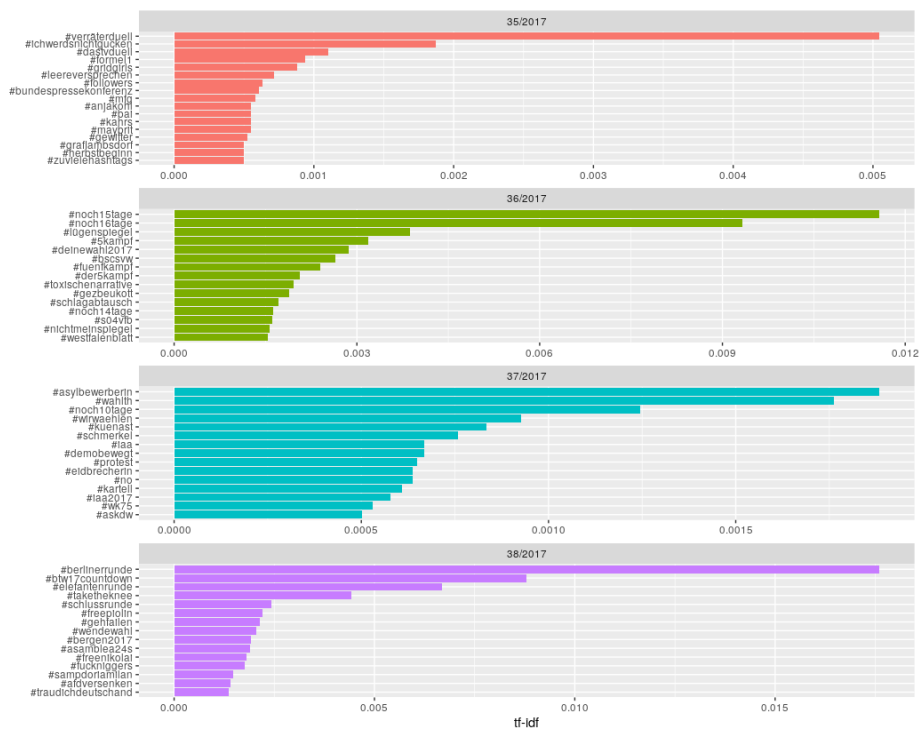


Figure 4: TOP-15 hashtags on weekly basis during the last four weeks before the German federal elections (tf-idf analysis).

reflected discussions on other media platforms. The round up the analysis I present a weekly-based analysis focusing on top hashtags per week during the last four weeks of the election campaign (Figure 4). As pointed out in the section describing data and method, the td-idf method used for this analysis seek to identify hashtags typical not for the whole period, but for each week instead.

The results from this weekly analysis indicate certain thematic shifts during the last four weeks. In the first week (week 35/2017) the TV duel on Sunday, 3 September 2017, between Angela Merkel and Martin Schulz was clearly one of the most discussed events. Once again, there were both positive/neutral - #dastvduell - and negative - #verräterduell (duel of betrayers), #ichwerdenichtgucken (I won't watch) - hashtags connected to this discussion. In the week 36/2017 the second TV, the so-called *Fünfkampf* (Battle of Five) between the Liberals, the Bavarian CSU, The Left Party, The Greens and the AfD (#5kampf, #fuenfkampf, #der5kampf) was this week's top theme¹⁶. Another theme was the approaching election day, referred to with 'countdown hashtags' #noch16tage or #noch15tage (16/15 days left). The fact that the week 37/2017 did not have any leading public event is reflected by the top hashtags. In this week the refugee debate gains a top position (#asylbewerberin - asylum seeker), followed by hashtags most probably encouraging people to vote (#noch10tage

¹⁶<https://www.daserste.de/information/politik-weltgeschehen/bundestagswahl-2017/sendung/der-fuenfkampf-nach-dem-tv-duell-100.html> (last visited: 1 September 2019)

- 10 days left, #wirwaehlen - we vote). The same applies to the election week (38/2017), of which top hashtags more or less reflect general public debates on other media platforms. The top hashtags including #berlinerrunde (Berlin round) and #elefantenrunde (Round of the Elephants) refer to the TV debate in the late evening on 24 September with high representatives of each seven parties ¹⁷.

To sum up, the results presented in this subsection bring up support for the assumption that political discussion on Twitter during the last four weeks of the German federal elections in 2017 politicised and politicalised public debates and discussions already present on different media platforms outside Twitter. Further, the results are also positive for the selection of methods used in the analysis. Both the network-based cluster analysis and tf-idf method could bring up interesting and reliable results. Used together the methods could not just shed light on general topics during the analysis period, but also focus on thematic changes in Twitter communication within the temporal framework.

3.2 European Elections 2019

Considering the European Parliamentary elections in 2019, three main topics dominated the media landscape during the spring 2019. The first topic, which was actually present also during the German federal elections in 2017, is the refugee crisis. Despite the fact that the absolutely number of refugees coming to Europe had remarkably dropped since 2015, the refugee question remained in the core of public debates in Europe. In this debate the role of European right-wing populist parties cannot be overestimated. The parties managed to keep the refugee question - very often linked to debates on and around Islam - on the top of public discussion both in traditional media, but especially in the social media and other platforms in the Internet.

The second European core discussion revolved around ‘brexit’, that the Great Britain leaving the EU. Actually, according to the original roadmap the UK should have left the EU on 29 March 2019. However, due to the chaotic political process, the UK was granted extended deadlines, finally resulting in a situation forcing the UK to participate in the European Parliamentary elections. This created very severe tensions not only between the UK and the rest of the EU, but also within the political landscape in the UK. It is worth being noted that although the ‘brexit’ referendum was held already in the summer of 2016, the ‘brexit’ question was also present during the German federal elections in 2017. However, as we have seen in the previous section, ‘brexit’ did not play any remarkable role in Twitter communication during the German federal elections.

The third main topic was climate change, a topic that hit the European publicity in October 2018 as the UN Intergovernmental Panel on Climate Change (IPCC) published its report demanding rapid and far-reaching changes in order to avoid disastrous global warming. Similar to the refugee crisis, debates on climate change quickly resulted in a strong polarisation in politics and society. Especially younger people organised themselves in order to keep this topic on top of political discussions. Maybe the most visible movement - ‘Fridays for

¹⁷<https://www.daserste.de/information/politik-weltgeschehen/bundestagswahl-2017/sendung/berliner-runde-132.html> (last visited: 1 September 2019)

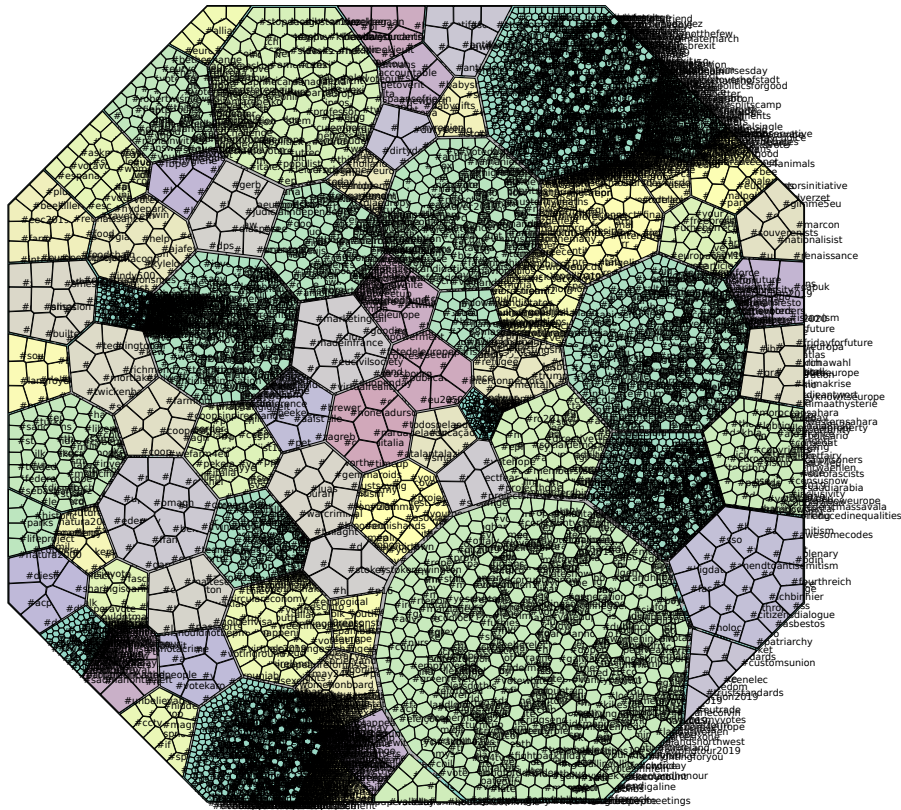


Figure 5: Thematic clusters based on hashtag co-occurrences during the last four weeks before the European Parliamentary elections (layout: Voronoi treemap).

Future' - was established by Greta Thunberg, a young Swedish environment activist born in 2003. The movement encouraged young people to get engaged on school strikes on Fridays in order to increase pressure on local (but also global) politicians in the fight against climate change. Thunberg - and more generally speaking all groups actively engaged in debates on climate change - were heavily attacked and criticised by right-wing politicians and parties seeking to question even scientific evidence of climate change.

Similar to the German federal elections 2017, the European parliamentary elections in 2019 were dominated by a rather small number of topics (Figure 5). In total, the clustering algorithm could identify 71 distinct clusters, also slightly more than in the German case. The same technical and methodological aspects described in the previous subsection apply to this subsection as well. The main findings of this analysis are presented in Figure 5 and Table 2.

The results from the cluster analysis indicate that Twitter communication during the last four weeks of the European parliamentary election was thematically dominated by three main topics: Brexit, Climate Change and Populism. 'Brexit' is not just the biggest cluster with 582 hashtags, also the clusters #2 - 'Ireland' with 442 hashtags - and cluster #5 - 'Scotland' with 175 hashtags - are clearly connected to Brexit. There is, however, an interesting dynamics between these

Table 2: European elections 2019: five (5) main clusters and their most influential hashtags.

Cluster #1: Brexit (n _{HT} =582)	Cluster #2: Ireland (n _{HT} =442)	Cluster #3: Climate change (n _{HT} =304)	Cluster #4: Populism (n _{HT} =273)	Cluster #5: Scotland (n _{HT} =175)
#brexit	#ie19	#climatechange	#europe	#votesnp
#brexitparty	#ireland	#greenwave	#italy	#itstime
#labour	#ie2019	#votegreen2019	#urban	#activesnp
#peoplesvote	#localelections2019	#votegreen	#hungary	#snp
#uk	#dublin	#fridaysforfuture	#salvini	#scotland
#ukip	#ivoted	#climateaction	#peace	#scotlandforeurope
#votelabour	#ee2019	#climate	#freedom	#independence
#changeuk	#ee2019	#climateemergency	#greennewdeal	#scotref
#farage	#divorceref	#voteclimate	#yestochange	#dissolvethunion
#deniedmyvote	#breakingbarriers	#climatestrike	#populism	#scotlandineurope

three brexit-related clusters. Whereas the ‘Brexit’-cluster itself evidence both pro - #brexitparty, #ukip, #farage (referring to the UKIP leader Nigel Farage) - and contra - #labour, #votelabour, #deniedmyvote - hashtags, the two other clusters, ‘Ireland’ and ‘Scotland’ and clearly negative to brexit. This is especially evident in the ‘Scotland’-cluster, in which the longing for independence as a part of Europe(an Union) is strongly expressed.

Hardly no surprise that climate change discussions rank to the top five clusters with 304 hashtags, especially if the high public visibility and global media coverage of the IPCC report is taken into account. Especially after the IPCC report this topic gained also in political importance, it was quickly politicalised and since then subject to controversial, partly overheated political pro- and contra discussions. As regards the European parliamentary elections, especially the Green movements in Germany and Finland, but also in France, seemed able to canalise this new environmental and climate consciousness into their electoral support. At the same time, anti-climate change politicking was one of the key domains for European right-wing populists. Consequently, this topic was highly polarised during the whole electoral campaign, which polarisation was evident on Twitter as well. But it is also worth being noted that despite its high political potential and strong public visibility, climate change was not so much politicked in the context of the European parliamentary elections as brexit.

Since both topics - brexit on the one hand, climate change on the other - have also been heavily attacked by different right-wing populist political forces, it hardly surprises that populism seems to be heavily discussed with 273 hashtags during the last four weeks of the election campaign. It is, however, worth being noted that the ‘Populism’ cluster seems to consist of non-brexit discussions revolving around Hungary - #hungary, #orban (referring to Hungarian prime minister Viktor Orbán) - or Italy - #italy and #salvini (referring to populist Northern League politician Matteo Salvini). Taking the hashtags #freedom and #peace into account, this cluster seems to consist of discussion focusing on European values, also including the hotly debated question of the rule of law.

Also in regard to the European parliamentary elections, a weekly-based tf-idf analysis brings up a more detailed view what comes to discussion dynamics and succeeds in reducing ‘noise’ causes by the permanent ‘brexit’ debate (Figure 6). In the week 18/2019 the campaign debate organised in Maastricht, the Netherlands, gains a lot of attention on Twitter (#maastrichtdebate). Interestingly, the international worker’s day on 1 May - #internationalworkersday - gains also a high frequency during the first week. The European Day celebrated on 9 May 2019 - #europeanday2019 and #9may - ranks to the top hashtags in the second week (19/2019), together with different hashtags - #registertovote, #trythegreens - encouraging voters to cast their votes.

The week 20/2019 did not have any special events, so that brexit-related hashtag gain in importance again. The election week (week 21/2019) is, rather predictable, dominated by voting-related hashtags in both positive - e.g. #ivoted - and negative - e.g. #deniedmyvote - sense. Another dominant discussion revolved around the election results, a rather expected finding.

To sum up, the results presented in this subsection also support the assumption that political discussion on Twitter merely politicked and politicalised public debates and discussions already present on different media platforms outside Twitter. As was the case in the German federal elections, also here the analytical

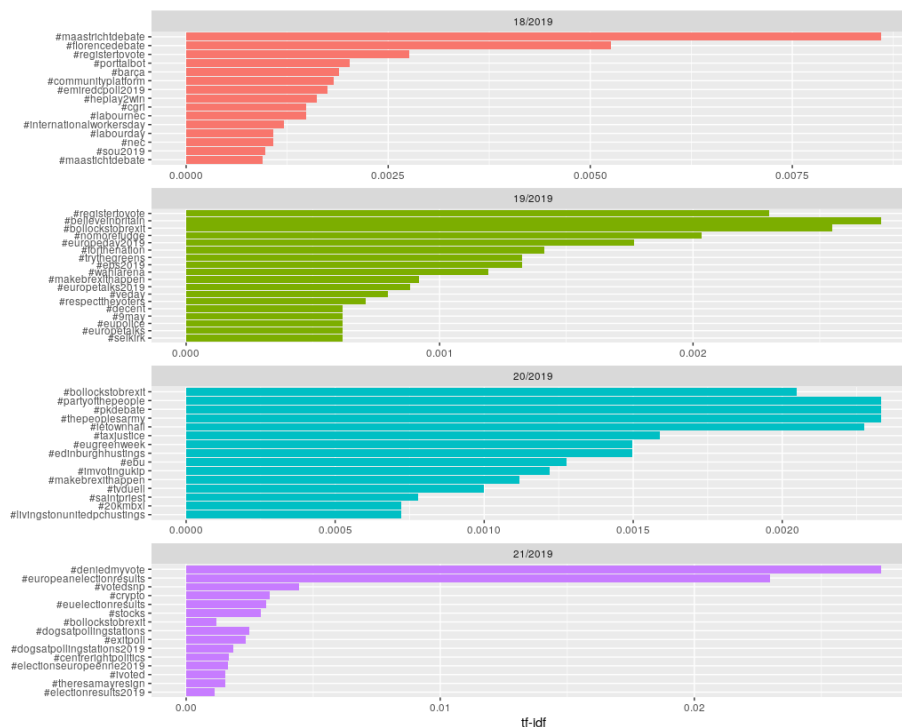


Figure 6: TOP-15 hashtags on weekly basis during the last four weeks before the European Parliamentary elections (tf-idf analysis).

and methodological tools proved their usability by bringing up reliable results.

4 Concluding remarks

This article presented a compact, data-driven comparative analysis of the German federal elections 2017 and the European parliamentary elections in 2019. From the methodological perspective the article sought to exemplify how digital tools can be applied to a large Twitter data in order to gain an improved understanding of political discussions within a selected, restricted domain.

Generally speaking the results presented in this article evidence, that a correct and appropriate selection of digital tools helps us to gain a more nuanced and truth-like understanding about the structure and dynamics of political communication on Twitter. The results presented in this article evidenced, that cluster analysis based on hashtag co-occurrence networks is an interesting alternative tool for topic modelling. Further, results from the tf-idf analysis evidenced the usefulness of this tool when it comes to understanding thematic changes and overall dynamics of Twitter discussion.

From the empirical perspective the results presented in this article seem to confirm the initial assumption, that Twitter discussions merely reflect and amplify, often also polarise, discussions already ongoing on other public and media platforms. It is, however, hard to say, whether the results presented in this paper

support or deny the idea of ‘echo chambers’ or ‘communication bubbles’¹⁸. The fact that some of the clusters analysed in this paper contained both pro and contra arguments could be interpreted as a counter-argument to the echo chamber thesis.

Despite the rather encouraging results and observations we should keep in mind that the analysis presented in this article is just *one* perspective on political communication on Twitter. Further, since Twitter data is not representative as regards the public opinion, we should be careful when drawing conclusions or generalising findings. Despite these critical points I value Twitter as an interesting source to gain understanding about how users politicalise different issues and how Twitter is used as a digital platform of politicking.

References

- Blondel, V. D., Guillaume, J.-L., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, (P10008).
- Coleman, S. & Freelon, D., Eds. (2015). *Handbook of Digital Politics*. Elgaronline.
- Feicheng, M. & Yating, L. (2014). Utilising social network analysis to study the characteristics and functions of the co-occurrence network of online tags. *Online Information Review*, 38(2), 232–247.
- Garimella, K., De Francisci Morales, G., Gionis, A., & Mathioudakis, M. (2018). Political Discourse on Social Media: Echo Chambers, Gatekeepers, and the Price of Bipartisanship. *ArXiv e-prints*.
- Herweg, S. (2013). *Politische Diskursnetzwerke und der Konflikt um das Anti-Piraterie-Abkommen ACTA*. Working Paper 15, Center for International Political Economy.
- Nahon, K. & Hemsley, J. (2013). *Going Viral*. Cambridge: Polity.
- Palonen, K. (2007). *Re-thinking Politics. Essays from a Quarter-century*. Jyväskylä: University Printing House.
- Paranyushkin, D. (2011). *Identifying the Pathways for Meaning Circulation using Text Network Analysis*. Research report, Nodus Labs.
- Shim, J., Park, C., & Wilding, M. (2015). Identifying policy frames through semantic network analysis: an examination of nuclear energy policy across six countries. *Policy Sciences*, 48(1), 51–83.
- Wang, R., Liu, W., & Gao, S. (2016). Hashtags and information virality in networked social movement: Examining hashtag co-occurrence patterns. *Online Information Review*, 40(7), 850–866.

¹⁸For example Herweg 2013; Garimella et al. 2018