# The Twitter parliamentarian database: Analyzing Twitter politics across 26 countries

N.B. When citing this work, cite the original published paper.

(article starts on next page)

RESEARCH ARTICLE

# The Twitter parliamentarian database: Analyzing Twitter politics across 26 countries

Livia van Vliet[1]*, Petter Törnberg[2], Justus Uitermark[3]

**1** Department of Sociology, University of Amsterdam, Amsterdam, Netherlands, **2** Department of Space, Earth and Environment, Chalmers University of Technology, Gothenburg, Sweden, **3** Amsterdam Institute for Social Science Research, University of Amsterdam, Amsterdam, Netherlands

* l.s.vanvliet@uva.nl

## Abstract

This article introduces the Twitter Parliamentarian Database (TPD), a multi-source and manually validated database of parliamentarians on Twitter. The TPD includes parliamentarians from all European Free Trade Association countries where over 45% of parliamentarians are on Twitter as well as a selection of English-speaking countries. The database is designed to move beyond the one-off nature of most Twitter-based research and in the direction of systematic and rigorous comparative and transnational analysis. The TPD incorporates, in addition to data collected through Twitter's streaming API and governmental websites, data from the Manifesto Project Database; the Electoral System Design Database; the ParlGov database; and the Chapel Hill Expert Survey. By compiling these different data sources it becomes possible to compare different countries, political parties, political party families, and different kinds of democracies. To illustrate the opportunities for comparative and transnational analysis that the TPD opens up, we ask: What are the differences between countries in parliamentarian Twitter interactions? How do political parties differ in their use of hashtags and what is their common ground? What is the structure of interaction between parliamentarians in the transnational debate? Alongside some interesting similarities, we find striking cross-party and particularly cross-national differences in how parliamentarians engage in politics on the social media platform.

## 1 Introduction

While many authors have argued that social media data have the potential to revolutionize social science research [1, 2], scholars are just beginning to discover how they can use this new source of data to carefully design social research [3]. To make substantive and rigorous contributions to current debates in the social sciences, further steps need to be taken in terms of developing standards for data collection, preparation and analysis [3, 4]. This is particularly true for Twitter. Since this platform has afforded researchers comparatively broad access, a huge number of studies have drawn on Twitter's data to study a wide range of social processes. While the excitement about the affordances of Twitter is understandable, few studies have

begun to address the formidable challenge of systematically collecting valid and representative
data [4–6].

We contribute to this endeavor by presenting a database of parliamentarians on Twitter.
The Twitter Parliamentarian Database (TPD) contains all the tweets of members of 27 parlia-
ments (26 national parliaments and the European parliament). The database has been pains-
takingly and systematically validated to address issues of reliability and validity characteristic
of much of the existing research on Twitter politician communication. To address issues of
data availability, the TPD incorporates, in addition to data collected through Twitter's stream-
ing API and governmental websites, data from the Manifesto Project Database on the parlia-
mentarians' political parties [7]; the Electoral System Design Database on the countries'
electoral and legislative systems [8]; the Chapel Hill Expert Survey on party positions on spe-
cific issues [9]; and the ParlGov database on political parties, elections and cabinets [10].

In this paper, we carry out a tentative analysis using the data in order to demonstrate the
potential of the database. While the TPD allows for a wide range of analyses, we focus our
demonstration on comparative and transnational research, as the database fills a considerable
research gap in these fields resulting from the lack of large-scale data sources. To illustrate the
capacity of the database to answer questions pertaining to this research, we carry out three
illustrative and exploratory studies. (1) We look at parliamentarian's Twitter use across
nations, with a focus on coalitions and divisions through who retweets whom, asking: *What
are the differences between countries in parliamentarian Twitter interactions?* (2) We look at
differences between how different parties label political issues, asking: *how do political parties
differ in their use of hashtags and what is their common ground?* (3) We study the structure of
mentions between parliamentarians internationally, asking: *what is the structure of interaction
between parliamentarians in the transnational debate?* We do not here aim for definitive
answers, but rather use the analyses to illustrate the affordances of the database in relation to
research questions that were previously difficult to address. Through including a variety of
analyses we demonstrate the different functionalities and areas of research that the TPD is able
to touch upon, especially analyses of communication patterns between and within a large
number of countries, as well as the content of the communication. While the affordances of
Twitter are the same across all the countries, we find cross-party and particularly cross-
national differences in how parliamentarians engage in politics on the social media platform.

## 2 Limitations and possibilities of research on Twitter politics

### 2.1 Limitations

Twitter is not only the social media of choice for politicians, but also the social media data
source *non plus ultra* for social scientific research [11]. There has been a veritable explosion of
research on Twitter in general and Twitter politics in particular (e.g. [12–17]). The Web of
Knowledge database contained 10,653 articles with "Twitter" as a key word (as per 31st Octo-
ber 2019), with 4,112 papers produced between January 1, 2018—October 31, 2019. The com-
bination of key words "Twitter" and "politics" finds 640 articles, with 279 published in 2018-
2019—roughly one article about Twitter and politics every 3 days. Despite this large and grow-
ing body of research, there remain a number of fundamental issues with carrying out research
on Twitter data, in particular in terms of *delineation, sampling,* and *validation.*

It has long been established that the *delineation of the population* are poorly addressed in
many studies examining politics on Twitter [6, 18]. Researchers often assume that the relevant
population consists of Twitter users who index their tweets with specific hashtags. However,
there are some serious issues to consider. One is that Twitter users participating in debates do
not necessarily use hashtags associated with a particular issue and Twitter users who do use

those hashtags do not always participate in the discussion, raising difficult questions of how to decide whether a message is relevant or merely "noise" [19]. Moreover, we know that Twitter users with different political positions tend to use different hashtags [20]. While careful curation of hashtags can attenuate these problems, it cannot solve them; when tweets and users are selected through hashtags, the delineation of the population remains arbitrary to a (generally unknown) extent. When the population is arbitrarily defined, all subsequent analyses can provide evocative results at best. There are however notable attempts at tackling this issue, for instance, Bruns et al. [21] have mapped follower/followee relations to get a more relevant sample of the Australian Twittersphere.

A second issue is associated with the common use of Twitter's free Streaming API for *sampling* tweets containing keywords. While research into the Twitter Streaming API is relatively sparse and may not be up-to-date due to constant API changes, the research that has been carried out has had troubling implications for the standard approaches to gathering Twitter data, in particular the use of search word or hashtag queries. For instance, when the free streaming API was compared with paid access to the Firehose API (which reportedly gathers 100% of all tweets), the sample became less representative as the number of parameters requested increased [22–24]. The issue here is not just that samples of selected keywords are not representative but that we do not know how samples are drawn and what their biases are. Other methods of sampling have been attempted by researchers through focusing on a core group of individuals, wherein all tweets can be gathered [16]. However, no attempts to standardize sampling at the scale of the TPD have yet been made.

A third problem with the way that Twitter data is used in research is linked to the lack of contextual or background information. Twitter data tend to be difficult to connect to other data sources, meaning that little is known on the identities and political leanings of the users. Consequently, *validation* is difficult. However, contextual information is essential if we want to compare between or within different political groups. In response to this issue, researchers often attempt to infer political viewpoints from behaviors such as a follower or friend network, and hashtag use [25–28]. For instance, if people use, or are listed as, #tcot (i.e., top conservatives on Twitter), it is assumed that they are conservatives. In this case, such a strategy may result in many false positives, but will also only identify a small subset of conservative users. Also, the political viewpoints garnered through these methods may not be representative of the wide variety of political attitudes that exist, where someone may share conservative beliefs in some points, but not in others [29, 30]. The issue of misidentifying users' political leanings is aggravated by the presence of sock puppets, trolls, and bots (e.g. [19]). Moreover, inferring political views from behavior brings researchers into a legal and ethical grey zone since Twitter regulations forbid the algorithmic identification of users' political viewpoints. In response to these challenges, researchers have developed more sophisticated methods (e.g. [30–33]) to identify political viewpoints, including linking Twitter data with survey data [30, 34]. In sum, there have been some interesting ways that researchers have tried to circumvent the methodological and technical challenges posed in classifying political viewpoints. We offer a different solution by sampling parliamentarians, wherein party affiliations are known, as well as draw upon existing databases for supplementary information.

## 2.2 Possibilities

If we want to exploit the opportunities for comparative and transnational research into politics that Twitter offers, the issues of delineation, sampling and validation have to be resolved. One way forward, which is the method employed by the database at hand, is to not define populations according to the content of their tweets but to construct a panel of Twitter users whose

tweets are collected over time (e.g. [35, 36]). This approach has been taken in a number of studies that have focused on accounts of the United States' Congress, for which reasonably reliable lists are available on Twitter (see, e.g. [37–39]). Some researchers have begun to study Twitter parliamentarians outside of the United States but often consider only one or at most two countries (e.g. [16, 40, 41]), often in the specific context of election campaigns [12, 42–45]. Thus, while there is considerable research focusing on Twitter use by parliamentarians, to our knowledge, there is no research that includes a large number of countries or that connects to existing databases on parties and parliaments. Hence, the database provides new opportunities for comparative research between multiple countries by making data available that has been gathered over several years from a clearly delineated population.

Focusing on parliamentarians limits the scope of research, but the advantage is that the TPD attempts to resolve several important issues regarding the delineation of the population and sampling in a way that is straightforward, transparent, and verifiable. An additional advantage is that parliamentarians are public figures, which significantly reduces ethical issues regarding privacy protection and increases possibilities to match data obtained from Twitter with data from other sources. In an effort to capitalize on these advantages, the database presented in this article extends earlier work by: including a much greater number of countries; complementing Twitter data with data drawn from other sources; and developing elaborate procedures to maintain data accuracy. The TPD data enable a broad range of Twitter research (e.g. [12, 46–50]). While previous research focused on a limited number of cases or had to rely on convenience samples (e.g. [25, 31, 49]), the TPD allows for studying these topics at scale and in comparative and transnational perspective. As such, we present a variety of analyses to demonstrate the different functionalities of the TPD.

Several research domains can reached by the TPD through linking it to other databases, such as; the Electoral System Design Database (ESDD), the Manifesto Project, the Chapel Hill Expert Survey (CHES) and ParlGov—meaning that the TPD can extend existing comparative work to Twitter, to examine differences in for example the ways parliamentarians express themselves and engage with one another, within nations, within parties as well as transnationally. To illustrate the TPD's potential, we focus on three domains of comparative research as a demonstration of how the database could begin to answer questions in these areas.

We use several methodologies to illustrate the TPD's potential for cross-national and comparative research. First, we can use the TPD to investigate national differences in the way parliamentarians use Twitter. Political science literature has a long-standing tradition of comparing political culture and political debate among various nations [51–55], showing important differences across countries and different types of electoral and political systems. We use this literature as a way both of showing how the structure of national retweet networks can shed new light on Lijphart's classic ideas [51] on the relationship between democracy types and national political communication. Second, the TPD allows for comparison of hashtag use between different parties. Twitter provides textual data that captures the way parliamentarians express themselves and frame political issues. Hence, the TPD constitutes a powerful tool for studying subsets of political party discourse. Despite the apparent potential, there has been limited research on content of parliamentarians' tweets, and none that are on a large international scale [56]. Therefore, the TPD offers the opportunity to contribute to comparative work on discursive conflicts [57, 58] that so far had to rely on newspaper data that cover only a very small portion of political claims [59]. A third domain in which the TPD can be used is transnational communication. Since social media are often regarded as conduits for breaking geographic boundaries [60], there is a need for systematic analysis of communication flows between countries. Since the TPD not only records interactions within but also between

countries, it allows for the examination of the prevalence, nature, sources, and topics of inter-national communication networks on Twitter.

## 2.3 Limitations

The limitations of exclusively following parliamentarians on Twitter should be acknowledged. First, as mentioned, the focus on parliamentarians limits the scope of the research, since parlia-mentarian tweets only constitute a subset of political discussions on Twitter. This limits useful-ness in relation to, for instance, campaigning research, as the TPD only gathers data on parliamentarians who are already in office. It should also be noted that in some cases, parlia-mentarians may choose not to individually interact with constituents but rather present them-selves through the party account. Further, the timeliness of updating the database following elections is largely dependent on the updating of official government websites, which may not occur in some countries until coalitions are formed. In rare cases, this may be several months. While the TPD has a limited scope outside of incumbent parliamentarian communication, it may serve as a starting point for questions about other political debates, by using various tech-niques to expand from parliamentarian users to other parts of Twitter.

## 3 Data collection and database design

In selecting countries, we aim to contribute to the large and growing body of comparative and transnational analysis. We included all member states and candidate member states of the European Free Trade Association (EFTA) where over 45% of parliamentarians are on Twitter [61]. In addition, we included a number majority English speaking countries because they allow for the application of English text analysis tools and have different political systems than most EU and EFTA countries, thus contributing to variation in the dataset. The countries included are Austria, Belgium, France, Denmark, Spain, Finland, Germany, Greece, Italy, Malta, Poland, Netherlands, United Kingdom, Ireland, Sweden, New Zealand, Turkey, United States, Canada, Australia, Iceland, Norway, Switzerland, Luxembourg, Latvia and Slovenia. In addition, the database includes the European Parliament. A full list of the countries and the proportion of their parliamentarians that use Twitter can be found in the S1 Table.

To identify the *parliamentarians* in the TPD, we consulted official government websites and retrieved all the identities of incumbent members. From these websites, we also collected the parliamentarians' party affiliations and when available, data on regions and constituencies. When the official government websites did not provide the party information for parliamen-tarians (as was the case for Germany, France, Finland and Spain), we used Google searches to establish party affiliation. For those websites where English was not available, it was double checked with translation programs that the list we obtained from these websites is the most current and up-to-date list of incumbent members. Identifying parliamentarians on Twitter is occasionally challenging due to e.g. common names and parody accounts, so we followed a protocol to identify and verify Twitter accounts, which involved comparing pictures with those on the government website, examining the number of followers, scrutinizing the lists of followers, and reading the content of the tweets. If the tweets were not in English, they were translated through Google Translate.

Due to the data protection regulation, we only include parliamentarians that are in public service during the time of data collection. Hence, persons who are not currently in parliament but are campaigning to be elected are excluded, as are those who served in one legislative period but were not re-elected in the next. If an account was set to private, it was not included. Inactive accounts, defined as accounts that have not made public tweets after 2014, were also excluded. For the year of 2018, the Twitter Parliamentarian Database captured 6,281,684

tweets from 6,437 parliamentarians active on Twitter out of a possible 8,098, meaning that 79.6% of parliamentarians had an active Twitter account.

The tweets of the parliamentarians are collected using Twitter's streaming API, including their mentions, retweets, and hashtags. As mentioned with regards to prior research, the streaming API can encounter certain limitations due to its rate limit, which, when used to follow a certain query (e.g. a popular hashtag that may be used 1,000s of times per second), may be reached much faster than following a certain user, who may only tweet a couple of times per day. Hence, following certain user (in this case, a parliamentarian) results in the streaming API gathering almost all tweets of that user as the amount of tweets per minute rarely exceeds the rate limit imposed by the API.

The database is updated following elections. Therefore there may be several electoral periods for some countries, if they had an election during the time of data collection. For example, if persons A and B are elected in 2015, and there is an election in 2018, where person B is reelected but person A is not, there will be data for person B across both periods, but only for person A until the election in 2018. Data collection for a parliamentarian starts from after they are elected, hence their tweet ids before the election date are not included. The Twitter accounts of members of new parliaments are checked for several months following elections, since we discovered that it takes time for new parliamentarians to find their footing on Twitter, and that some parliamentarians may later create new accounts for their parliamentary service, separate from their personal or campaign accounts. The database includes data from as early as May 2017 for some countries, and has been continually updated since, capturing many interesting political events such as the span of Donald Trump's presidency, the Catalonia referendum, the Brexit debate. The user ids from the database can be used to retrieve older tweets from parliamentarians across the electoral periods where members were collected, or be used to gather further data. Hence, any researcher can use the Twitter ids currently provided in the TPD as a starting point for their own research to update a country of interest following an election.

We conducted cross-validation to confirm the coverage and validity of the collected users and tweets. For user accounts, only limited validation could be carried out, due to the limited availability of databases against which to cross validate. We however compared the TPD ids for members of the US congress against the *115th U.S. Congress Tweet Ids* dataset [62], finding that 93% of the ids in the member list matched. The ids that did not match in either data set were found to be missing accounts that have since been deleted. This could be due to changing accounts during time in office, which can happen as some parliamentarians have campaigning accounts. Moreover, the method of gathering accounts for the 115th U.S. Congress Tweet Ids dataset was different to the TPD, where the former mostly retrieved Twitter accounts from the congress website, which may not differentiate between campaigning and service accounts, rather than manual research which was used for the TPD. To validate the coverage of tweets, we used a random sample of 50 current parliamentarians in the TPD, and retrieved tweets from their timelines occuring between March 1, 2020 and May 1, 2020, using Twitter's REST API. We confirmed that 98% of the tweet ids that were retrieved from the timelines were found in the database. The 2% that were not found are thought to be due to server downtime.

Moreover, the database can be connected to the ESDD, the Manifesto Project, CHES and Parlgov, which enables research questions beyond countries, parliamentarians and their parties. The Manifesto Project, CHES and ParlGov can be used with the TPD in various ways. For instance, the ESDD can be used in combination with Twitter data to determine relationships between electoral systems and online politician communication, where additional variables like electoral size, number of tiers and legislative system are also available [8]. The manifesto project provides "parties' policy positions derived from a content analysis of parties' electoral

manifestos." [7] ParlGov contains information on parties, elections and cabinets for 37 countries, including all EU and most OECD democracies [10]. The CHES use expert surveys to "estimate party positioning on European integration, ideology and policy issues for national parties in a variety of European countries." [9] Thus, these databases can be used to link differences in online elite political behavior and interaction to variables like party family, political position or offline discourse. In the supplementary information of this article is the Database codebook which gives an in-depth explanation about the variables included, in which tables they can be found, and the relationships between the tables in the database. These relationships can also be seen in Fig 1. The latest data can be downloaded from the TPD website: twitterpoliticians.org.

## 4 Analyses to demonstrate the Twitter parliamentarian database as a research tool

To illustrate the TPD's capacity for comparative research, we use guiding questions to focus our demonstrations, limiting the time frame to the period from 1 January 2018 to 31 December 2018 for congruence.

### 4.1 Comparing countries: How do politicians use Twitter?

We begin by looking at the similarities and differences between countries in terms of Twitter use: *What are the differences between countries in parliamentarian Twitter interactions?* While Twitter offers the same functionalities to parliamentarians everywhere, how those functionalities are used varies significantly between countries. These differences may point to differences in political cooperation across countries. Firstly, the percentage of parliamentarians that actively use Twitter differs. Some countries may have an extremely high active Twitter base (99%, United States) where most parliamentarians tweet almost daily, whereas in other countries, the parliamentarians may have Twitter accounts, but they rarely tweet. However, on average 80% of parliamentarians per country are active on Twitter. The frequency of Twitter use varies across countries. On average, parliamentarians tweet 2.8 times per day, although there is some deviation; parliamentarians in Iceland tweet less than once per day, whereas parliamentarians in Turkey tweet 6 times per day ($min = 0.7; max = 6.6; \sigma = 1.4$).

Retweeting and mentioning exclusively between national parliamentarians make up an average of only 21% of total politician Twitter activity (min = 7%, max = 36%), with Poland having the highest proportion of Twitter activity between their parliamentarians, and Iceland with the lowest. Further, as we see in Fig 2, while mentions to other parliamentarians tend to be used much more than retweeting, there are stark differences between countries: Icelandic parliamentarians very rarely retweet, while Greek, Turkish and Canadian parliamentarians have more retweets than mentions. Conversely, we see the opposite pattern when looking at retweets and mentions to non-parliamentarians, as parliamentarians retweet more than mention when communicating with those outside of parliament.

The data on retweets and mentions can however be employed in much more powerful ways, providing new data to questions that have long been central to political science. For instance, a large literature in political science focuses on international comparison of politician cooperation, and the relation to aspects of the countries' democratic systems. This literature follows not least from Lijphart's [51] suggestion that consensus democracies, usually employing proportional representation systems, leads to "kinder, gentler" political cultures than majoritarian systems (see also e.g. [63–65]). The authors postulate that the structures of power distribution represented by the democratic system of a country (e.g. majoritarian or consensus) may encourage attitude polarization (and in turn elite conflict), or instead, foster
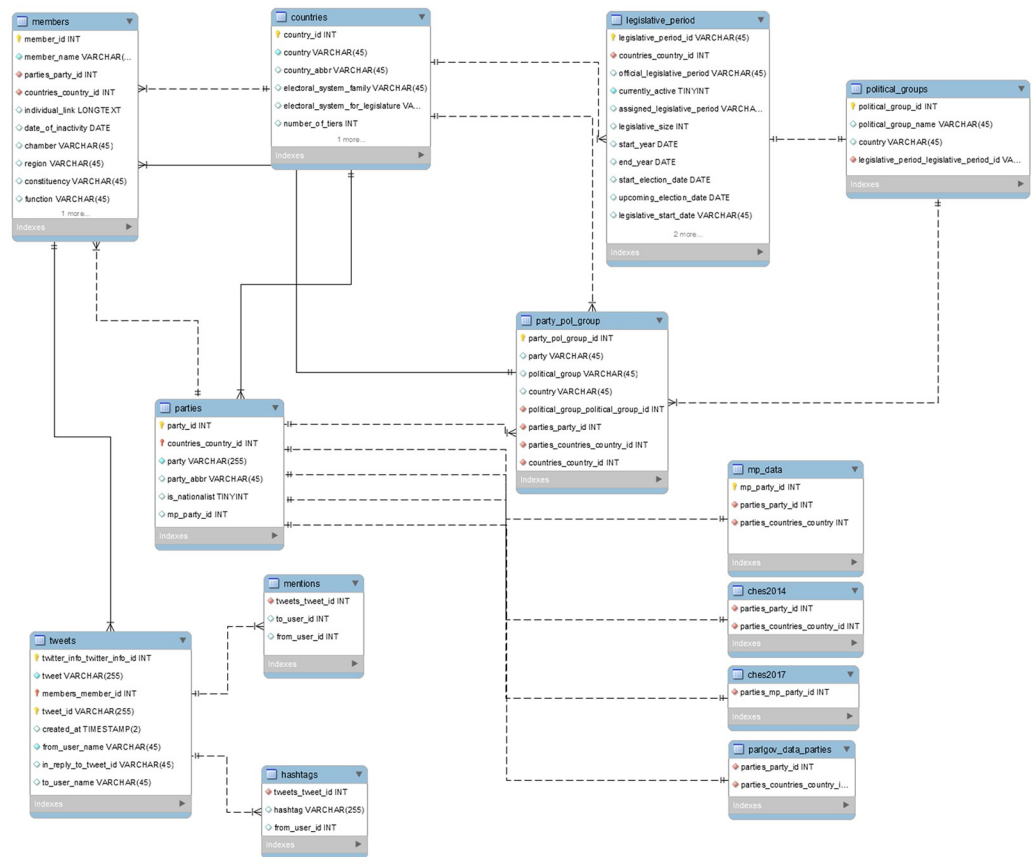
**Fig 1. Entity relationships in the TPD.** Fig 1 shows a simplified version of the entity relationships between tables in the database, the foreign keys, and their data types. (Note that not all data columns are included in the diagram for the sake of brevity).

https://doi.org/10.1371/journal.pone.0237073.g001

cooperation between politically dissimilar parties [51, 52, 66, 67]. With this reasoning, it is thought that proportional systems lead to increased cooperation [51]. Conversely, other scholars argue that due to lower barriers of entry for smaller, single-issue parties, political fragmentation is instead increased in proportional systems [66–69].

Academic work on the coalitions and divisions within parliamentarian communication networks has been limited by lack of suitable empirical data. The TPD can thus provide a useful data source for studying the coalitions and divisions among parliamentarians. By viewing retweets as endorsements (c.f. [70, 71]), the patterns of retweeting can be revealing of the political alliances within a country. Retweets can be treated as edges in a network, wherein the structure can reveal coalitions and divisions amongst parliamentarians. Such networks can be analysed in different ways, capturing different aspects of the structure of endorsements within a country. For this demonstration, we focus on the networks with 30 or more parliamentarians, and filter the networks by their giant component. We here aim only for a tentative exploration, leaving an in-depth and rigorous analysis for future research.

A simple but powerful way of operationalizing the level of cohesion within a country is to compare the fraction of retweets made to members of external parties. Countries whose parliamentarians frequently use retweets to endorse members of other parties can be assumed to have more amicable between-party relations compared to countries whose parties mostly
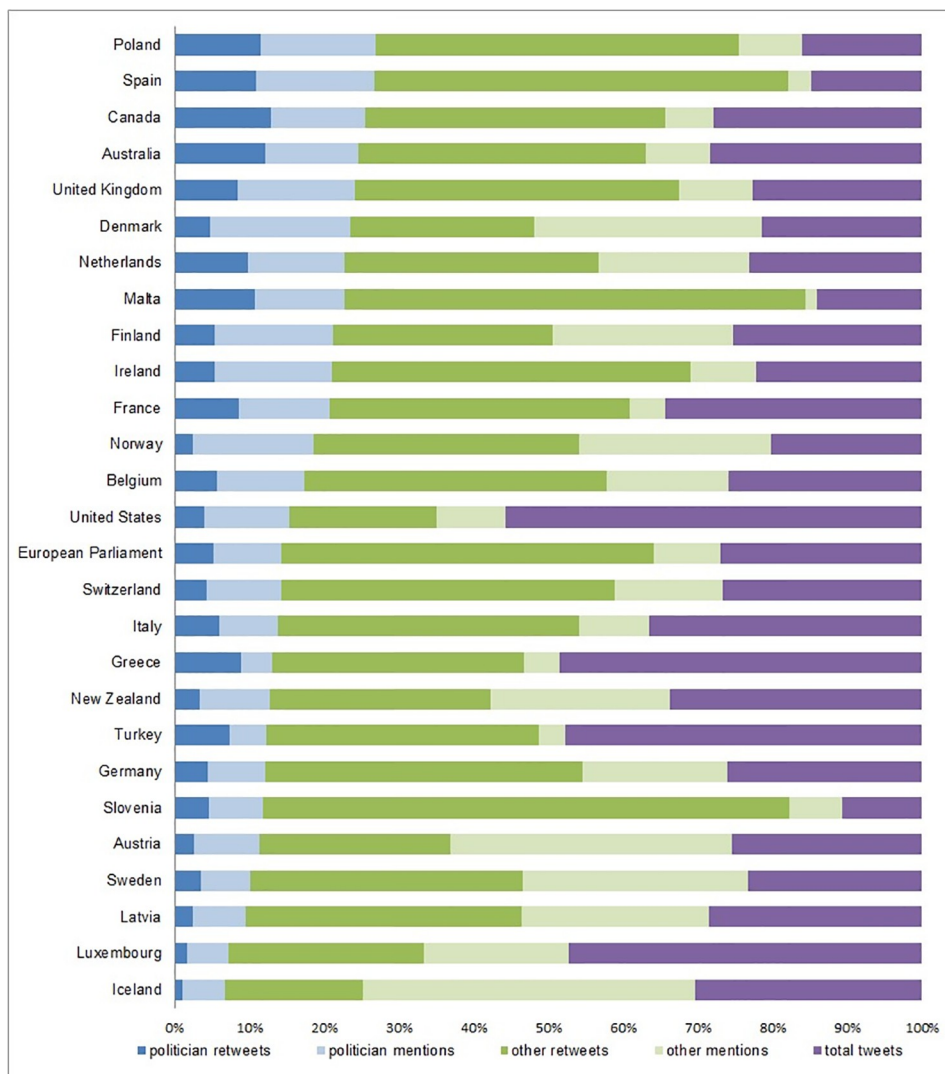
**Fig 2. Ratio of retweets & mentions.** Retweets and mentions per country as a percentage of total tweets for all countries in the database.

retweet internally. S2 Table documents the electoral system, and the average fraction of external retweets per country. It shows that there are clear differences between countries in terms of the fraction of external retweets. Although these patterns need further exploration, we see that majoritarian (M) systems tend to have a lower fraction of external retweets compared to countries with mixed and proportional representation (PR) systems. PR systems on the other hand, show wide variation in the average fraction of external retweets. Countries with a high fraction of external RTs all employ PR systems.

However, this quantitative approach has certain limitations, as it fails to capture, for instance, divisions involving multiple parties in coalitions or situations in which a fraction of a party is strongly divided from other parties. As there are many forms that divisions can take in these networks, we turn to a more qualitative approach to network analysis: Visual Network Analysis [72]. This highly flexible approach allows us to categorize the endorsement networks

according to their structure, identifying various forms of divisions and alliances within a country.

To look beyond descriptive metrics for examining the differences in parliamentarian twitter interactions, for this preliminary demonstration we focus on the networks with 30 or more nodes and filter the networks by their giant component. We plot the networks using the ForceAtlas2 visualization algorithm, which uses a number of properties to structure the networks in such a way that highly connected nodes tend to be closer to each other, and less connected nodes further away from each other [73]. In Fig 3, the nodes are colored according to their party.

Using this qualitative method identify four distinct types of political network structures (see Fig 3). Type 1 networks show a highly divided structure: there are clear divisions between clusters and very few cross-cutting ties. Type 2 networks show two large clusters that have dense connections amongst themselves and fewer external ties. Type 3 networks show a large, densely connected structure with one or two outlying parties, which are weakly connected to the others. A clear example of this is the Netherlands, where the outsider is radical-right party PVV and Germany, where the outsider is the radical-right AfD, which entered parliament in the 2017 election. Lastly, Type 4 structures exhibit one large cluster of dense connections. All the individual retweet networks and their classifications can be found in S1 Fig.

We can furthermore use additional measures to compare the categories of networks identified using VNA, to demonstrate that the networks are not only visually distinct. S2 Table also reports the modularity, number of clusters, and average clustering coefficient per country to provide measures of clustering within the networks. To confirm the stability of the modularity measure, we used the leidenalg python library to measure each network 1,000 times. We report the average modularity from these runs, as well as the most frequent number of clusters per run. Cluster assignment was based on the most frequent cluster assigned per node across the 1,000 runs. The results indicate that type 1 networks tend to have higher modularity, as well as a greater number of clusters than other networks. Type 3 networks tend to have fewer clusters
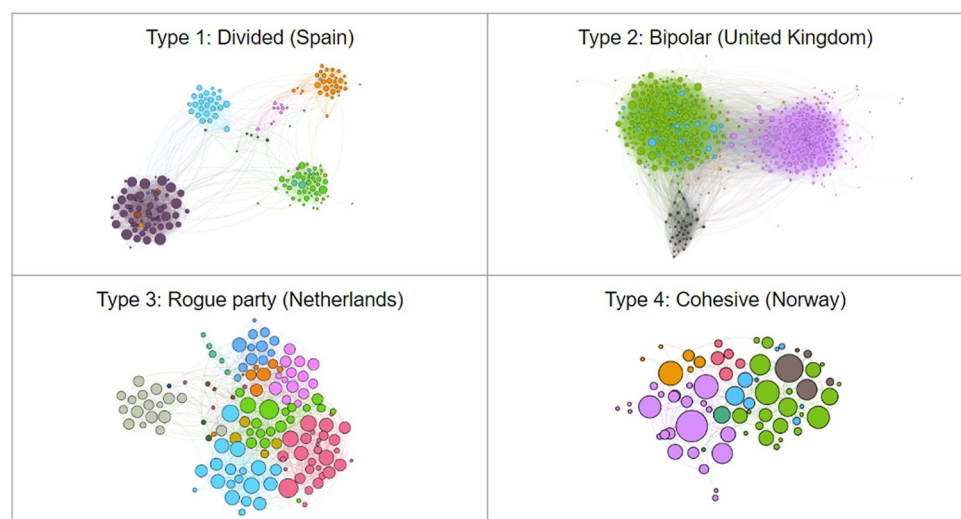


**Fig 3. Parliamentarian retweet network archetypes.** Shows the 4 distinct types of retweet networks. Each node represents an individual parliamentarian. The networks are plotted with the ForceAtlas2 visualization algorithm which plots highly connected nodes close to one another, and less connected nodes further away. The nodes are colored by party.

than the other network types. Type 1 and 2 networks tend to have higher clustering coefficients whereas type 4 networks have lower clustering coefficients. This would indicate that type 1 networks tend be more divided than types 3 and 4. The fraction of external retweets also shows that parties in type 1 and 2 networks retweet other parties much less than those in types 3 and 4. Additionally we see that most type 3 networks tend have negative kurtosis of their degree distributions, which implies that members in the network retweet each other to a similar degree, rather than rallying around a few leaders. Finally, we see that type 1 networks tend to have lower normalized eigencentrality scores, implying that nodes are less well-connected within the network.

We also look at the relationship between cluster and party membership through $\chi^2$ and Cramer's V measures, wherein Cramer's V show the strength of that relationship, thereby indicating how 'neatly' the networks cluster based on party: fragemented networks are expected to have a stronger relationship between party-cluster membership, whereas a weak relationship would indicate more overall cohesion in the network. For brevity, we only report the Cramer's V value in S2 Table, as all $\chi^2$ results indicated a significant relationship between party and cluster membership ($p < 0.00$). We see that type 1 networks have much stronger relationships between party and cluster membership than type 4 networks. Therefore, it is clear that type 4 networks generally retweet across party lines.

All in all, visual analysis and basic network measures can be combined to interpret the types of network structures that emerge from the data. The measures provided are by no means exhaustive and may differ depending on the research question. Using the information from the ESDD [8], we can explore differences in parliamentarian endorsement not only between countries, but between democratic systems. We see that Type 1 networks show a divided network with little endorsement between parties. Type 2 networks appear bipolar, with two large contending groups, and are the most common category for Majoritarian systems. Lastly, types 3 and 4 are more consensual, due to many cross-cutting endorsements beyond party lines, and are largely comprised of PR systems. These results are tentative and not exhaustive. However, they do illustrate that the database offers opportunities to compare coalitions and divisions between different countries and political systems. While our purpose here is to illustrate the potential of the dataset for studying cross-national differences through exploratory analysis, it also includes data that allows for more formal and quantitative measurement, and points to several fruitful directions for future research.

## 4.2 Comparing language between parties: Examining Hashtags

Textual data from Twitter provides an avenue into seeing how parliamentarians navigate different issues and engage in discursive conflict. Despite this potential, there has been limited research on content of parliamentarians' tweets, and none that has focused on international comparison [56]. The TPD offers the opportunity to contribute to comparative work on discursive conflicts [57, 58] that so far had to rely on newspaper data that typically do not provide a comprehensive coverage of political claims [59] and is time-consuming to collect and process [74]. Since the TPD furthermore links to the Manifesto Project, Chapel Hill, and ParlGov databases, it has the information necessary for connecting such textual analysis to a large body of work that compares different political parties and party families [7, 75–77]. In this section, we explore the possibilities for using the TPD to take a comparative approach to the study of political discourse, focusing on the differences in the use of hashtags between political parties.

The use of labels in political discourse reveal the different ways that opposing political parties discuss the same issues. This is apparent through brief exposure to political debate: while one party may speak of "tax reform," the other focuses on "tax relief." Using different labels for

the same issue indicates how central labelling is to politics [78]. The way an issue is labelled influences how we view that issue, what issues we see in the first place and enable us to make sense of what we are reading. Using certain labels over others may lead to exacerbating political divides: when every group identifies and labels its own issues in its own way, it makes conversations across partisan lines more difficult [79].

Analyzing labels used on Twitter is made easier by Twitter providing affordances for explicitly labeling tweets, in the form of hashtags. While hashtags serve many purposes, broadly speaking they can be used to index conversations [80], convey a particular point of view [48], or for issue positioning and labelling [25, 26, 48, 81]. Importantly, hashtags aid in building public perceptions of an issue by ensuring maximum visibility, and allow anyone to jump into the conversation [82]. In order to demonstrate the database as a tool for comparing issue labelling between parties, we therefore we look at hashtags, using the guiding question: *how do political parties differ in their use of hashtags and what is their common ground*?

To take a first step toward a comparison of how different parties label issues, we develop a simple computational method that captures which hashtags are partisan and which are shared. To study this, we look at the two largest parties in all countries, in terms of the number of active users on Twitter (note that these are not necessarily the parties that have the most seats in parliament). We count each hashtag used by these parliamentarians and then normalize according to the total number of hashtags used by the party. If a hashtag is used by both parties, the overlapping part of the use is seen as the intersection between the parties. As in a Venn diagram, the intersection of the sets are the common hashtags, and the non-intersecting parts are hashtags characteristic of the party. The relative size of the intersection thereby gives a measure of the similarity in hashtag use between the two parties, which corresponds to the Jaccard similarity measure [83]. Since dynamics of discursive conflicts are likely different in multiparty systems, we here focus on the majoritarian systems categorized by the ESDD: the UK, USA, Canada and Australia. The results are shown in Fig 4. While the set-based method used here is intuitive and therefore useful for exploratory analysis, a common statistical way of comparing corpora for the most overrepresented words is Log-Likelihood [84]. This analysis can be found in S3 Table, and largely matches the results in Fig 4.

A first striking takeaway from this Fig 4 is that the common ground hashtags represent only a small fraction, where the intersection is 15.7% (in the United Kingdom) or less. Most of the hashtags that are equally represented in each party (proportional to the number of total hashtags used by that party), are usually smaller, less commonly used tags. This indicates that hashtags are largely partisan. Hence, to understand what these fractions mean, we need to look closer at the way hashtags are employed by politicians.

A first thing to note is that the most common intersecting hashtags for Canada and Australia are "cdnpoli" and "auspol" respectively, although they are not used with the same relative frequency by each party. These are general country-level hashtags for marking the nation which the discussion concerns, and are broadly employed for political debate also outside of professional politicians (these hashtags have been broadly studied, e.g. [24, 85, 86]). The UK and the US do not seem to have any corresponding hashtags, which may be due to the US and UK being large enough to be the norm on the platforms, while Canadian parliamentarians need to demarcate that they are speaking of Canadian politics.

We focus here on the US for a deeper look into the ways parties make use of hashtags, and the relation to how political issues are labelled. Looking at Fig 4, we see that politicians from both parties address many of the same issues, while using very different labels. The 2018 tax reform is a major talking-point on both sides, but while Republicans refer to it using "taxreform" or "taxcutsandjobsact," Democrats instead use "goptaxscam". Interestingly, there is no common ground hashtag to denote the bill. A similar state of affairs can be identified in
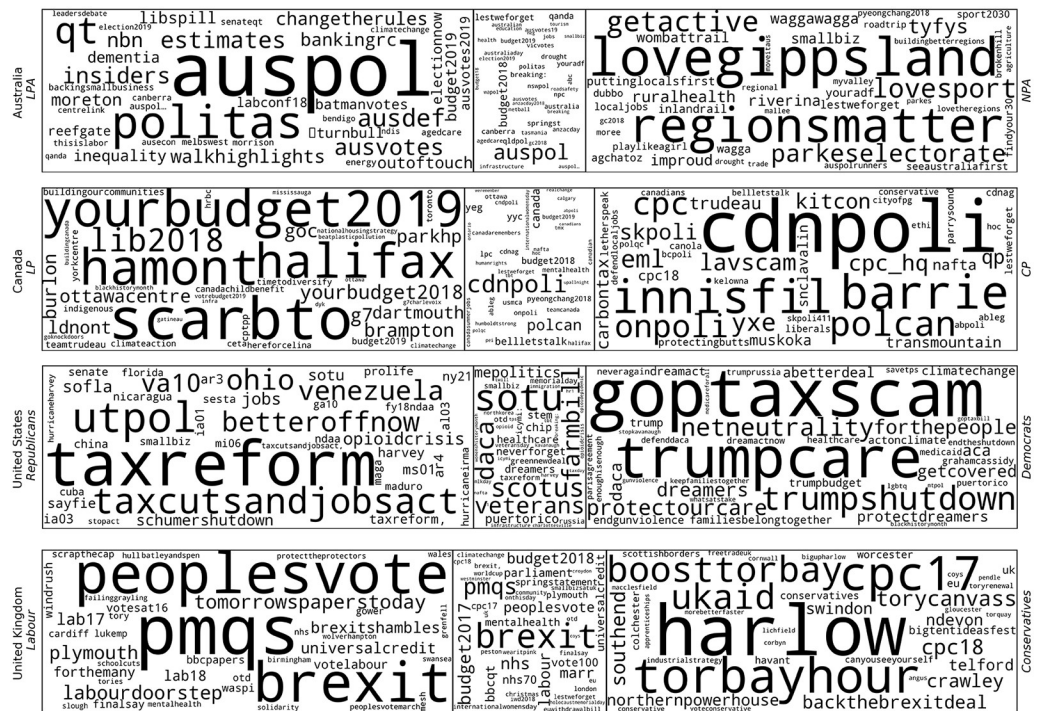
**Fig 4. Hashtag use amongst the 2 largest parties.** This figure shows partisan and common hashtags in Australia, the United States, Canada, and the United Kingdom. The middle panel shows the common ground while the side panels show the distinctive hashtags for two largest parties. The size of the middle panel is proportional to the size of the common ground relative to the parties. In the word clouds, the size of words is proportional to the frequency of use as a fraction of total hashtag use.

https://doi.org/10.1371/journal.pone.0237073.g004

relation to the 2018 government shutdown, which both parties attempt to attribute to the other party: Democrats refer to it as the "trumpshutdown," while Republican use "schumershutdown". An interesting point of note is that Democrats use the hashtag "trump," while this is not among the major hashtags for Republicans. Republicans focus on "Venezuela," as an example of the putative dangers of left-wing politics, while Democrats speak of "climate-change," an issue that is not featured among top republican talking-points.

The most important intersecting hashtags found in Fig 4 include tags which describe arenas of contention rather than specific topics of disagreement, for instance, "scotus," referring to the Supreme Court, in which highly a contested process of electing new judges played out during 2018. Another example of this is "farmbill", referring to the primary agricultural and food policy bill of the US government, which is renewed every 5 years and deals with affairs under the purview of US Department of Agriculture. While there may be disagreements about its contents, the common use hashtag suggests that representatives from both parties at least agree they are discussing the same issue.

Other important intersecting hashtags point to common ground values, to which both parties are happy to profess their support. These include "veterans," "neverforget," and "stem". Similar common ground is found in national emergencies, such as the hurricane striking Puerto Rico. However, even here the language differs somewhat between the parties: Republicans are more likely to refer to the event as "hurricaneirma", while Democrats speak of its impact on "PuertoRico". DACA (Deferred Action for Childhood Arrivals)—an Obama-era executive action, now turned Trump-vetoed bill with broad bipartisan support among voters

—is an important hashtag in the intersection, but is driven in particular by Democratic politicians, who also employ the related hashtags "dreamact", "dreamers", "protectdreamers". There is, perhaps somewhat tellingly, broad bipartisan use of the hashtag "mepolitics," denoting a criticism of the country's polarized political discourse.

As we have demonstrated, we see that politicians use hashtags to express partisan claims. While there is some common ground (for instance when it comes to the importance of caring for veterans), politicians generally use hashtags in outspokenly partisan ways: they use specific hashtags to push different kinds of issues on the agenda or to express a partisan take on the same issue. This case has served to illustrate the range of possible research opened by the TPD in examining language and discourse of politicians on Twitter.

## 4.3 Transnational communication: What is the structure of the transnational mention network?

The TPD not only allows for cross-national comparative research but also for research on international and transnational politics, by enabling analysis of how parliamentarians of different countries communicate with one another [87–89]. Through facilitating communication and mobilization of opinion across borders [90], social media platforms open up for the possibility of a cross-national dialogue. The TPD allows us to examine systematically and comprehensively whether parliamentarians are employing the affordances of the platform to engage in transnational debate. To demonstrate, we look at all mentions between politicians in the European Free Trade Association (EFTA). When a parliamentarian uses the Twitter mention functionality, it refers to another Twitter user and notifies the user that they have been tweeted about. Mentions in this case are tweets in which another parliamentarian's Twitter account is signalled with the '@' symbol. This does not include direct retweets, but it does include retweets where additional text has been added. Mentions were chosen over retweets as they are more indicative of a dialogue or debate (rather than an endorsement), as well as to provide illustration of analysis of Twitter's affordances beyond retweets in the TPD. Luxembourg and Slovenia were excluded from the analysis due to their low number of international mentions. Moreover, since we are interested in communication between national parliamentarians, we do not include the European Parliament in this analysis (see [91–93] for studies on the European Parliament). As illustrative questions, we ask; *what is the structure of interaction between parliamentarians in the transnational debate*?

We begin by examining the network of mentions between individual parliamentarians, filtered by its giant component (see Fig 5). In most countries (with the exception of Ireland and Spain), less than 55% of parliamentarians have mentioned a parliamentarian from another EFTA country. S4 Table contains the proportion of parliamentarians that made external retweets per country. The extent to which the international network is split along national lines is striking: the parliamentarians are organized in clusters that almost perfectly cut along national lines, with only a small number of parliamentarians being located outside their respective national cluster. Moreover, when there are connections between national clusters, they tend to occur between neighboring countries, with nations like Ireland and UK situated next to one other in the network.

To arrive to a more precise measure for quantifying the level of external communication between countries, we look at the fraction of external tweets, network diameter, average path length and modularity. The modularity indicates the strength of the division of the network. The network diameter measures the longest distance between any two nodes in the network, whereas average path length counts the average graph distance between all pairs of nodes. These measures help to understand the structural connectivity amongst individual national
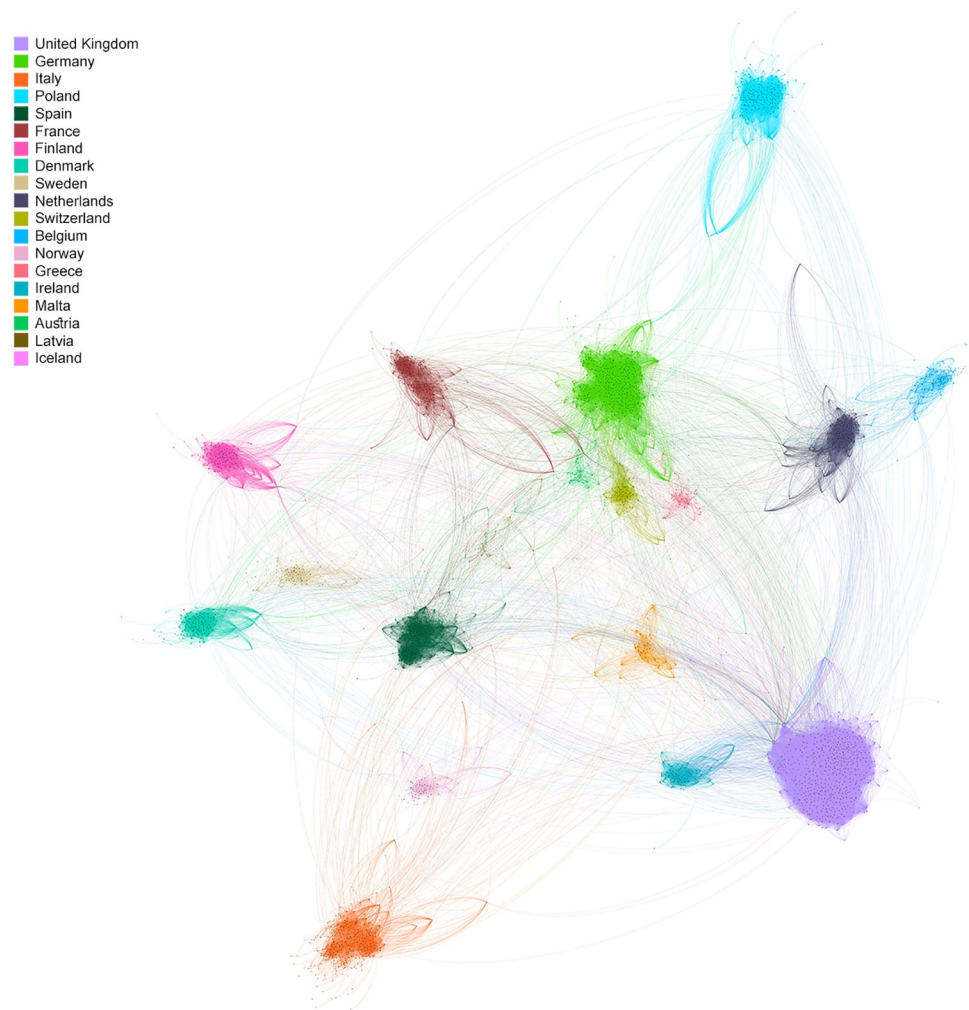
**Fig 5. International mention network.** This network shows all mentions between parliamentarians in the database. The network is made in Gephi, using the ForceAtlas2 algorithm. Node colors are set according to country.

parliamentarians on a European level. Aggregating on country level, we examine centrality with PageRank. The network diameter is 18 and the average path length is 5.8, indicating that there is quite some distance between distant nodes in this network. The network has a modularity of 0.678, suggesting that the network is relatively divisible into separate clusters. All these numbers suggest that parliamentarians communicate mostly with their compatriots, resulting in a very sparse and highly divided international communication network.

Although international communication accounts for a low proportion of the total communication (1.9% of all mentions are directed to a parliamentarian in a different country), in absolute terms it is still substantial cross-national traffic (16,955 total mentions). Looking closer at these mentions gives insight in patterns of communication across EFTA countries. We examine the fraction of external mentions per country, to determine the proportion of mentions that are dedicated to international debate [94]. When looking at fractions of external mentions, we see in Fig 6 that some countries participate minimally in international debate, through a low fraction of incoming and outgoing international mentions. Moreover, we see that the majority of all international mentions are directed towards a small number of
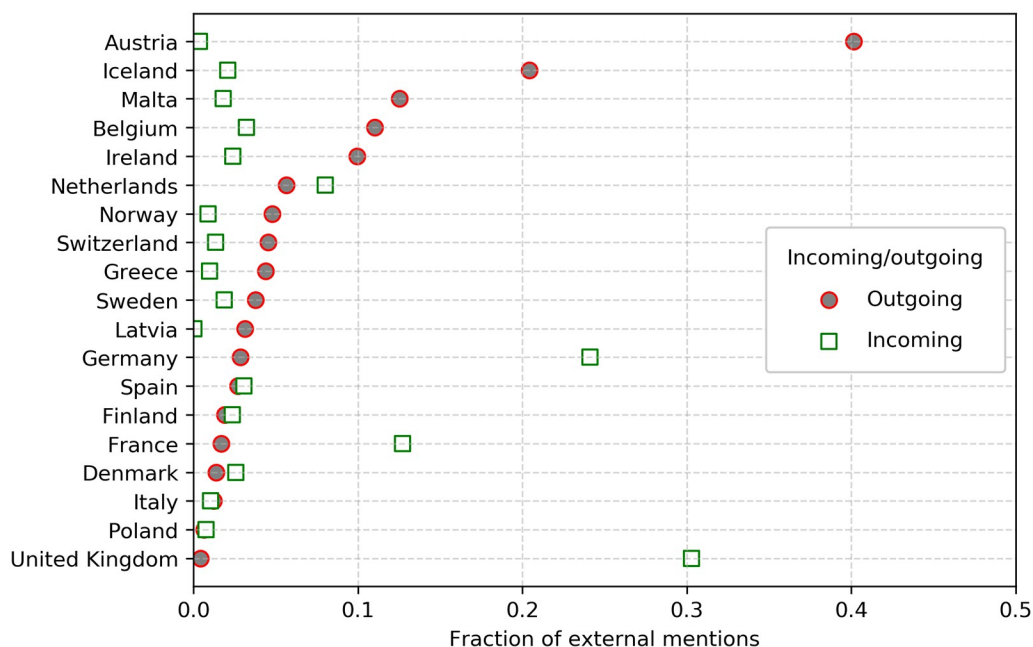
**Fig 6. Incoming and outgoing mentions.** Shows the normalized fraction of outgoing and incoming mentions. The fraction of incoming mentions shows which countries are more central to the debate whereas the outgoing mentions shows how much mention activity per country is directed to international debate.

countries, namely the United Kingdom and Germany, and to a lesser extent, France. Additionally, we see that there is a large imbalance in outgoing and incoming mentions for some smaller parliaments such as Austria and Iceland. Small countries appear relatively outward facing, as there are simply fewer parliamentarians to mention within these small countries, and more external parliamentarians, resulting in a higher fraction of external mentions. Moreover, it is noted that smaller countries tend to be more globalized than larger countries [95].

To further examine the extent of the transnational debate, Figs 7 and 8 show that the international debate is uneven, being dominated by Germany and United Kingdom. To check if countries contribute equally to the international debate, we used an independent samples Kruskall-Wallis test. We compared the relationship between normalised cross-border tweets and country. The results indicate that some countries have significantly more external ties than others (p = 0.000). This implies that some countries participate more than others in the transnational debate. For a more precise estimate of the relative importance of different countries, we calculate PageRank centrality on the adjacency matrix of the aggregated connections between countries. The results are shown in Fig 9. The figure shows that while France has a large and active parliament, they are relatively peripheral in the international debate compared to Sweden, Denmark and the Netherlands—all whom are significantly smaller in parliamentary size but higher in centrality. The sheer volume of mentions emitted by UK parliamentarians—gives it a central place in the transnational debate as shown by its PageRank. Germany, by contrast, has significantly fewer external mentions but are are just as central in terms of PageRank and relatively more connected, which is also shown by their fraction of incoming external mentions. Thus, Germany also occupies a central position in European debates.

In summary, the results of our demonstration suggest that debates among national parliamentarians remain, by and large, contained within national boundaries. The TPD not only makes it possible to study the degree of transnational communication, but also to examine the
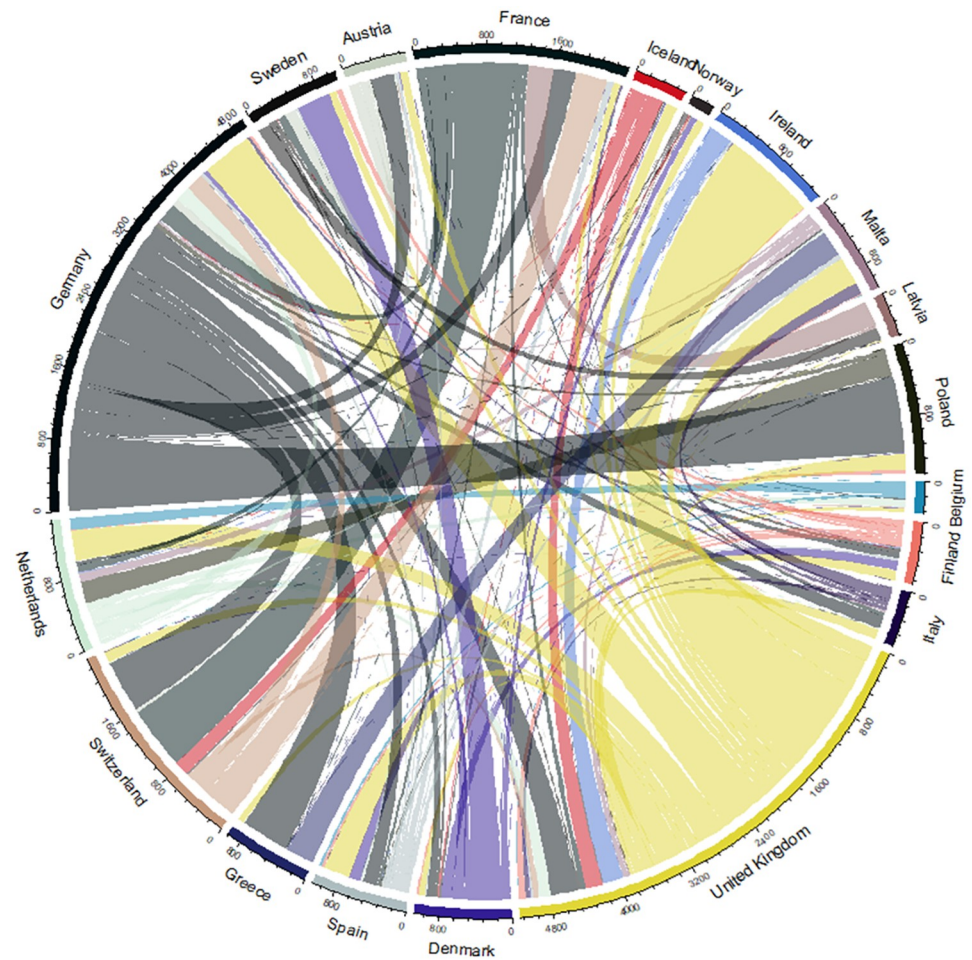
**Fig 7. Incoming and outgoing mentions.** Shows the fraction of outgoing and incoming mentions per country. It includes only international mentions, thereby showing the importance of each country in the international debate.

https://doi.org/10.1371/journal.pone.0237073.g007



**Fig 8. Absolute and normalized adjacency matrices.** Shows absolute (left) and normalized (right) adjacency matrix for the mention network. (Note that the color scheme is logarithmic).

https://doi.org/10.1371/journal.pone.0237073.g008

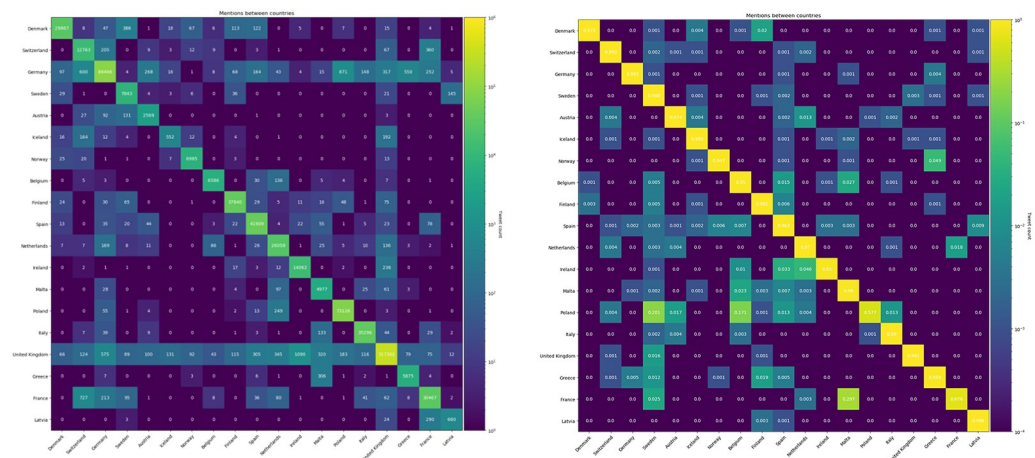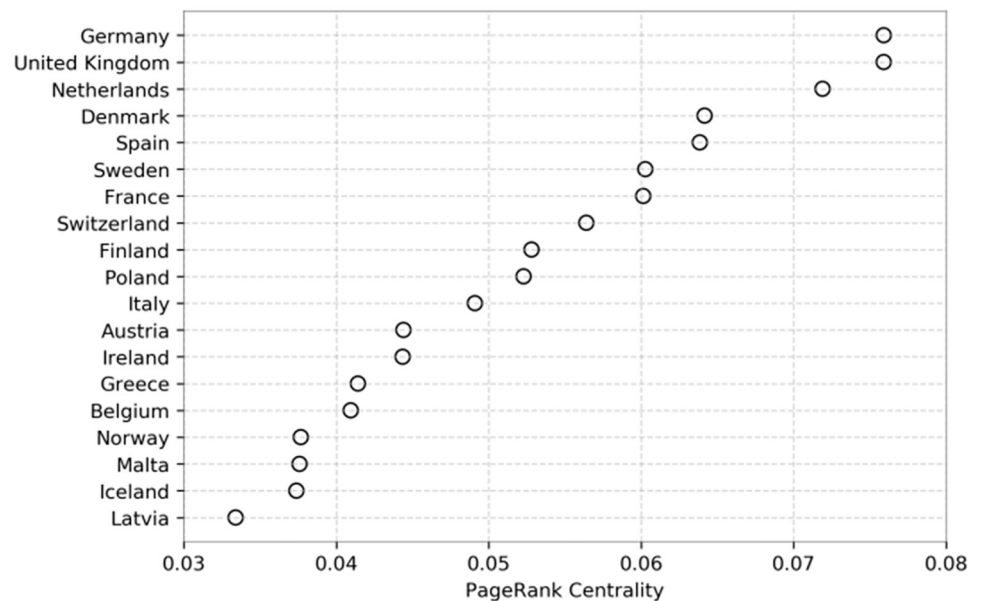**Fig 9. PageRank centrality per country.** To investigate which countries are most central in the international debate in Europe, this figure shows the PageRank centrality of the aggregated network.

position of countries, parties, or individual parliamentarians within that debate and the role of different issues. While our exploratory analysis has shown which countries are more central to transnational debates within Europe, the TPD makes it possible to study transnational politics on Twitter at a grater scale and in more detail than data sources have previously allowed for.

## 5 Conclusion

This paper presents a database that responds to the methodological issues regarding delineation, sampling and validation methods commonly used in Twitter research, using a variety of data sources along side manually validated Twitter information. The opportunities of this database for comparative and transnational research were illustrated through three tentative studies, looking at 1) national differences between parliamentarians' Twitter use, 2) differences between political parties' hashtag use and 3) the structure of transnational debates.

Through our exploratory analyses, we discover similarities in the ways which politicians use Twitter across countries. Overall when communicating with other politicians, they prefer to use mentions rather than retweets. There are also many differences: some countries have a very active, thriving Twitter culture in which all functionalities are used, others refrain from mentioning and retweeting, and yet others may not use the platform much at all. Apart from documenting differences in how Twitter is used, the database can also provide an empirical foundation for research into long-standing questions in political science. To illustrate the database as a tool for studying these questions, we used the structure of retweet networks as a proxy for coalitions and divisions in parliamentarian Twitter networks. Using Visual Network Analysis, we could distinguish four different kinds of retweet networks: bipolar, fringe party and cohesive. These structures show whether or not there is a lot of endorsement across party lines, or if the political culture seems more strictly partisan. Indeed, when looking at clustering measures, we see that networks that appear more divided have a higher clustering coefficient. We find that majoritarian systems have less external retweets and higher clustering

coefficients, and most commonly resulted in bipolar structures. On the other hand, proportional systems are the only systems that resulted in cohesive networks. This suggests that there is a correspondence between the nature of political system and patterns of political communication but further investigation is necessary to arrive at robust conclusions.

Comparisons between parties are made through using hashtags as a conduit for issue labelling. We compare countries with majoritarian systems, as they have two large opposing parties that comprise the bulk of the politicians' Twitter activity. We find that there are very few hashtags that are shared between the opposing parties, and, in line with existing literature [26, 96, 97], hashtags are used by politicians for issue positioning. The TPD is able to be used when looking at the content of parliamentarian tweets to determine which party used which hashtag, which can help provide a clear identification of partisan and shared tags within a country. This identification can then highlight issues of importance between different parties. Thus, hashtags are an interesting future avenue to study how parties label and discuss issues.

The TPD does not only allow comparisons of countries and parties but also the study of transnational communication. By way of illustration, we studied cross-European parliamentarian communication. We found that cross-national mentions constitute only a tiny portion of the total politician Twitter use, which confirmed by the fraction of external mentions per country. Therefore there is very little international interaction amongst national parliamentarians. Additionally, through PageRank measures we can see that Germany and the United Kingdom take central positions in these debates. Hence, we have demonstrated that the TPD is able to explore the degree in which cross-European communication exists for parliamentarians on Twitter, along with the position of the countries, parties or individual parliamentarians across a number of different issues with more detail than previous studies.

This paper has therefore demonstrated that the TPD is a powerful database for carrying out research on parliamentarians' use of Twitter, in particular for cross-country comparative and transnational research, which has thus far struggled with data availability. All in all, our database addresses some of the current methodological issues with Twitter research and provides a starting point for studying communication, contention and cooperation not only within countries, but also comparatively across borders. As far as we are aware, the TPD is the most comprehensive database of parliamentarians on Twitter that exists, and it is able to provide a framework for more standardized comparative methodology in politician Twitter research.

## Supporting information

**S1 File. The database codebook.**
(PDF)

**S1 Fig. Retweet networks per country, 2018.** Shows the individual country retweet networks for 2018. The nodes are colored by party and the network is visualized with the ForceAtlas2 algorithm, with each node sized by in-degree (1-10) and scaling set to 1.
(TIF)

**S1 Table. This table shows the number of MPs on Twitter per country in the database.** The legislative period id is an arbitrary id number that distinguishes each legisla-tive period.
(PDF)

**S2 Table. This table shows clustering and degree measures, as well as fractions of external mentions and the strength of relationship between party and cluster membership, applied to the individual country networks.**
(PDF)

**S3 Table. This table shows the highest Log-Likelihood values (shown in brackets) of hashtags used by the two largest parties in 2018.** High values indicate that a term occurs more frequently than chance amongst one party.
(PDF)

**S4 Table. This table shows the number and percentage of MPs who mentioned another politician in the EFTA in 2018.**
(PDF)

## Acknowledgments

We wish to acknowledge the tremendous help of Veerle Scholtes and Michele Consolini in keeping the database up-to-date and valid. We would also like to thank Jan Willem Duyvendak for his extensive feedback and suggestions, and Anna Keuchenius for her help.

## Author Contributions

**Conceptualization:** Livia van Vliet, Petter Törnberg, Justus Uitermark.

**Data curation:** Livia van Vliet.

**Formal analysis:** Livia van Vliet, Petter Törnberg, Justus Uitermark.

**Funding acquisition:** Petter Törnberg, Justus Uitermark.

**Investigation:** Livia van Vliet.

**Methodology:** Livia van Vliet, Petter Törnberg.

**Project administration:** Livia van Vliet.

**Software:** Petter Törnberg.

**Supervision:** Petter Törnberg, Justus Uitermark.

**Validation:** Livia van Vliet, Petter Törnberg.

**Visualization:** Livia van Vliet, Petter Törnberg.

**Writing – original draft:** Livia van Vliet.

**Writing – review & editing:** Livia van Vliet, Petter Törnberg, Justus Uitermark.

## References

1. Manovich L. Digital Traces in Context—100 Billion Data Rows per Second: Media Analytics in the Early 21st Century. International journal of communication. 2018; 12:16.

2. Watts DJ. A twenty-first century science. Nature. 2007; 445(7127):489. https://doi.org/10.1038/445489a PMID: 17268455

3. Salganik MJ. Bit by bit: Social research in the digital age. Princeton University Press; 2019.

4. Cihon P, Yasseri T. A biased review of biases in Twitter studies on political collective action. Frontiers in Physics. 2016; 4:34. https://doi.org/10.3389/fphy.2016.00034

5. Jungherr A, Theocharis Y. The empiricist's challenge: Asking meaningful questions in political science in the age of big data; 2017.

6. Tufekci Z. Big questions for social media big data: Representativeness, validity and other methodological pitfalls. In: Eighth International AAAI Conference on Weblogs and Social Media. AAAI Press; 2014. p. 505–514.

7. Volkens A, Lacewell O, Lehmann P, Regel S, Schultze H, Werner A. The manifesto data collection. Manifesto project (MRG/CMP/MARPOR). Berlin: Wissenschaftszentrum Berlin für Sozialforschung (WZB). 2013;.

8. IPU I. International IDEA Database on Electoral System Design at https://www.idea.int/esd/data/world. cfm. TRS = Two Round System. MMP = Mixed Member Proportional. 2019;.

9. Bakker R, Edwards E, Hooghe L, Jolly S, Marks G, Polk J, et al. Chapel Hill Expert Survey. Version 2015.1. Chapel Hill: University of North Carolina[cit 2017-07-19] Dostupné z: http://web.unc.edu/. 2014;.

10. Döring H, Manow P. Parliament and government composition database (ParlGov). An infrastructure for empirical information on parties, elections and governments in modern democracies Version. 2012; 12 (10).

11. Lazer D, Radford J. Data ex machina: introduction to big data. Annual Review of Sociology. 2017; 43:19–39. https://doi.org/10.1146/annurev-soc-060116-053457

12. Vergeer M. Twitter and political campaigning. Sociology Compass. 2015; 9(9):745–760. https://doi.org/10.1111/soc4.12294

13. Grant WJ, Moon B, Busby Grant J. Digital dialogue? Australian politicians' use of the social network tool Twitter. Australian Journal of Political Science. 2010; 45(4):579–604. https://doi.org/10.1080/10361146.2010.517176

14. Keller TR, Kleinen-von Königslöw K. Pseudo-discursive, mobilizing, emotional, and entertaining: identifying four successful communication styles of political actors on social media during the 2015 Swiss national elections. Journal of Information Technology & Politics. 2018; 15(4):358–377. https://doi.org/10.1080/19331681.2018.1510355

15. Lee EJ. Effectiveness of politicians' soft campaign on Twitter versus TV: Cognitive and experiential routes. Journal of Communication. 2013; 63(5):953–974.

16. Rauchfleisch A, Metag J. The special case of Switzerland: Swiss politicians on Twitter. New Media & Society. 2016; 18(10):2413–2431. https://doi.org/10.1177/1461444815586982

17. Enli G, Simonsen C.A. Social media logic'meets professional norms: Twitter hashtags usage by journalists and politicians Information, Communication & Society, 2018; 21(8):1081–1096

18. Ruths D, Pfeffer J. Social media for large studies of behavior. Science. 2014; 346(6213):1063–1064. https://doi.org/10.1126/science.346.6213.1063 PMID: 25430759

19. Marres N, Moats D. Mapping controversies with social media: The case for symmetry. Social Media+ Society. 2015; 1(2):2056305115604176.

20. Weber I, Garimella K. #Egypt: Visualizing Islamist vs. secular tension on Twitter. In: 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2013). IEEE; 2013. p. 1100–1101.

21. Bruns A, Moon B, Münch F, Sadkowsky T. The Australian Twittersphere in 2016: Mapping the follower/followee network. Social Media+ Society. 2017; 3(4):2056305117748162.

22. Morstatter F, Pfeffer J, Liu H, Carley KM. Is the sample good enough? comparing data from Twitter's streaming api with Twitter's firehose. In: Seventh international AAAI conference on weblogs and social media. AAAI Press; 2013. p. 400–408.

23. Joseph K, Landwehr PM, Carley KM. Two 1%s don't make a whole: Comparing simultaneous samples from Twitter's streaming API. In: International conference on social computing, behavioral-cultural modeling, and prediction. Springer; 2014. p. 75–83.

24. Bruns A, Liang YE. Tools and methods for capturing Twitter data during natural disasters. First Monday. 2012; 17(4):1–8. https://doi.org/10.5210/fm.v17i4.3937

25. Conover MD, Gonçalves B, Ratkiewicz J, Flammini A, Menczer F. Predicting the political alignment of Twitter users. In: 2011 IEEE third international conference on privacy, security, risk and trust and 2011 IEEE third international conference on social computing. IEEE; 2011. p. 192–199.

26. Hemphill L, Culotta A, Heston M. #Polar Scores: Measuring partisanship using social media content. Journal of Information Technology & Politics. 2016; 13(4):365–377. https://doi.org/10.1080/19331681.2016.1214093

27. Al Zamal F, Liu W, Ruths D. Homophily and latent attribute inference: Inferring latent attributes of twitter users from neighbors. In: Sixth International AAAI Conference on Weblogs and Social Media; 2012. p. 387–390.

28. Golbeck J, Hansen D. Computing political preference among twitter followers. In: Proceedings of the SIGCHI conference on human factors in computing systems; 2011. p. 1105–1108.

29. Kam CD, Wilking JR, Zechmeister EJ. Beyond the "narrow data base": Another convenience sample for experimental research. Political Behavior. 2007; 29(4):415–440. https://doi.org/10.1007/s11109-007-9037-6

30. Preoţiuc-Pietro D, Liu Y, Hopkins D, Ungar L. Beyond binary labels: political ideology prediction of twitter users. In: Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers); 2017. p. 729–740.

31. Barberá P. Birds of the same feather tweet together: Bayesian ideal point estimation using Twitter data. Political Analysis. 2015; 23(1):76–91. https://doi.org/10.1093/pan/mpu011

32. Pennacchiotti M, Popescu AM. A machine learning approach to twitter user classification. In: Fifth international AAAI conference on weblogs and social media; 2011. p. 281–288.

33. Sylwester K, Purver M. Twitter language use reflects psychological differences between democrats and republicans. PloS one. 2015; 10(9). https://doi.org/10.1371/journal.pone.0137422 PMID: 26375581

34. Vaccari C, Valeriani A, Barberá P, Bonneau R, Jost JT, Nagler J, et al. Social media and political communication. A survey of Twitter users during the 2013 Italian general election. Rivista italiana di scienza politica. 2013; 43(3):381–410.

35. Budak C, Watts DJ. Dissecting the spirit of Gezi: Influence vs. selection in the Occupy Gezi movement. Sociological Science. 2015; 2:370–397. https://doi.org/10.15195/v2.a18

36. Diaz F, Gamon M, Hofman JM, Kıcıman E, Rothschild D. Online and social media data as an imperfect continuous panel survey. PloS one. 2016; 11(1):e0145406. https://doi.org/10.1371/journal.pone.0145406 PMID: 26730933

37. Golbeck J, Grimes JM, Rogers A. Twitter use by the US Congress. Journal of the American Society for Information Science and Technology. 2010; 61(8):1612–1621.

38. Hemphill L, Otterbacher J, Shapiro M. What's congress doing on twitter? In: Proceedings of the 2013 conference on Computer supported cooperative work. ACM; 2013. p. 877–886.

39. Shapiro MA, Hemphill L. Politicians and the policy agenda: Does use of Twitter by the US Congress direct New York Times content? Policy & internet. 2017; 9(1):109–132. https://doi.org/10.1002/poi3.120

40. Del Valle ME, Bravo RB. Echo chambers in parliamentary Twitter networks: The catalan case. International journal of communication. 2018; 12:21.

41. Feroz Khan G, Young Yoon H, Kim J, Woo Park H. From e-government to social government: Twitter use by Korea's central government. Online Information Review. 2014; 38(1):95–113. https://doi.org/10.1108/OIR-09-2012-0162

42. Barberá P, Rivero G. Understanding the political representativeness of Twitter users. Social Science Computer Review. 2015; 33(6):712–729. https://doi.org/10.1177/0894439314558836

43. Ernst N, Engesser S, Büchel F, Blassnig S, Esser F. Extreme parties and populism: an analysis of Facebook and Twitter across six countries. Information, Communication & Society. 2017; 20(9):1347–1364. https://doi.org/10.1080/1369118X.2017.1329333

44. Graham T, Jackson D, Broersma M. New platform, old habits? Candidates' use of Twitter during the 2010 British and Dutch general election campaigns. New media & society. 2016; 18(5):765–783. https://doi.org/10.1177/1461444814546728

45. Kampf R, Manor I, Segev E. Digital diplomacy 2.0? A cross-national comparison of public engagement in Facebook and Twitter. The Hague Journal of Diplomacy. 2015; 10(4):331–362. https://doi.org/10.1163/1871191X-12341318

46. Gainous J, Wagner KM. Tweeting to power: The social media revolution in American politics. Oxford University Press; 2013.

47. Kruikemeier S. How political candidates use Twitter and the impact on votes. Computers in human behavior. 2014; 34:131–139. https://doi.org/10.1016/j.chb.2014.01.025

48. Conover MD, Ratkiewicz J, Francisco M, Gonçalves B, Menczer F, Flammini A. Political polarization on Twitter. In: Fifth international AAAI conference on weblogs and social media. AAAI Press; 2011. p. 89–96.

49. Ratkiewicz J, Conover MD, Meiss M, Gonçalves B, Flammini A, Menczer FM. Detecting and tracking political abuse in social media. In: Fifth international AAAI conference on weblogs and social media. AAAI Press; 2011. p. 297–304.

50. Hemphill L, Roback AJ. Tweet acts: how constituents lobby congress via Twitter. In: Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing. ACM; 2014. p. 1200–1210.

51. Lijphart A. Patterns of democracy: Government forms and performance in thirty-six countries. Yale University Press; 2012.

52. Norris P. Driving democracy: Do power-sharing institutions work? Cambridge University Press; 2008.

53. Kreisi H, Grande E, Lachat R, Dolezal M, Bornschier S, Frey T. West European Politics in the Age of Globalisation. Cambridge: Cambridge University Press; 2008.

**54.** Inglehart R. The silent revolution: Changing values and political styles among Western publics. Princeton University Press; 2015.

**55.** Statham P, Koopmans R, Giugni M, Passy F. Resilient or adaptable Islam? Multiculturalism, religion and migrants' claims-making for group demands in Britain, the Netherlands and France. Ethnicities. 2005; 5(4):427–459. https://doi.org/10.1177/1468796805058092

**56.** Jungherr A. Twitter use in election campaigns: A systematic literature review. Journal of information technology & politics. 2016; 13(1):72–91. https://doi.org/10.1080/19331681.2015.1132401

**57.** Koopmans R. Contested citizenship: Immigration and cultural diversity in Europe. vol. 25. U of Minnesota Press; 2005.

**58.** Ferree MM, Gamson WA, Rucht D, Gerhards J. Shaping abortion discourse: Democracy and the public sphere in Germany and the United States. Cambridge University Press; 2002.

**59.** Koopmans R. Movements and media: Selection processes and evolutionary dynamics in the public sphere. Theory and society. 2004; 33(3-4):367–391. https://doi.org/10.1023/B:RYSO.0000038603.34963.de

**60.** Hänska M, Bauchowitz S. Can social media facilitate a European public sphere? Transnational communication and the Europeanization of Twitter during the Eurozone crisis. Social Media+ Society. 2019; 5 (3):2056305119854686.

**61.** Van Vliet, L., Törnberg, P., Uitermark, J. What can we learn from how parliamentarians talk on Twitter? at http://twitterpoliticians.org/ The Twitter Parliamentarian Database. 2019.

**62.** Littman J. 115th U.S. Congress Tweet Ids; 2017. Available from: https://doi.org/10.7910/DVN/UIVHQR.

**63.** Nordinger EA. Conflict regulation in divided societies. Harvard University Press; 1972.

**64.** Lehmbruch G. A non-competitive pattern of conflict management in liberal democracies: the case of Switzerland, Austria and Lebanon. Consociational Democracy: Political Accommodation in Segmented Societies. 1974; p. 90–97.

**65.** Armingeon K. The effects of negotiation democracy: A comparative analysis. European Journal of Political Research. 2002; 41(1):81–105. https://doi.org/10.1111/1475-6765.00004

**66.** Cox GW. Centripetal and centrifugal incentives in electoral systems. American Journal of Political Science. 1990; 34(4):903–935. https://doi.org/10.2307/2111465

**67.** Horowitz DL. A democratic South Africa?: Constitutional engineering in a divided society. vol. 46. Univ of California Press; 1992.

**68.** Reilly B. Democracy in divided societies: Electoral engineering for conflict management. Cambridge University Press; 2001.

**69.** Reynolds A, Reilly B, Council NR, et al. Electoral systems and conflict in divided societies. National Academies Press; 1999.

**70.** Kim J, Yoo J. Role of sentiment in message propagation: Reply vs. retweet behavior in political communication. In: 2012 International Conference on Social Informatics. IEEE; 2012. p. 131–136.

**71.** Metaxas P, Mustafaraj E, Wong K, Zeng L, O'Keefe M, Finn S. What do retweets indicate? Results from user survey and meta-review of research. In: Ninth International AAAI Conference on Web and Social Media. AAAI Press; 2015. p. 658–661.

**72.** Decuypere M. Visual Network Analysis: a qualitative method for researching sociomaterial practice. Qualitative Research. 2019; p. 1468794118816613.

**73.** Jacomy M, Venturini T, Heymann S, Bastian M. ForceAtlas2, a continuous graph layout algorithm for handy network visualization designed for the Gephi software. PloS one. 2014; 9(6):e98679. https://doi.org/10.1371/journal.pone.0098679 PMID: 24914678

**74.** Koopmans R, Statham P. Political claims analysis: Integrating protest event and political discourse approaches. Mobilization: an international quarterly. 1999; 4(2):203–221. https://doi.org/10.17813/maiq.4.2.d7593370607l6756

**75.** Mair P, Mudde C. The party family and its study. Annual review of political science. 1998; 1(1):211–229. https://doi.org/10.1146/annurev.polisci.1.1.211

**76.** Allen TJ. All in the party family? Comparing far right voters in Western and Post-Communist Europe. Party Politics. 2017; 23(3):274–285. https://doi.org/10.1177/1354068815593457

**77.** Merz N, Regel S, Lewandowski J. The Manifesto Corpus: A new resource for research on political parties and quantitative text analysis. Research & Politics. 2016; 3(2):2053168016643346.

**78.** Lakoff G. Why it matters how we frame the environment. Environmental Communication. 2010; 4 (1):70–81. https://doi.org/10.1080/17524030903529749

**79.** Scheufele DA. Agenda-setting, priming, and framing revisited: Another look at cognitive effects of political communication. Mass Communication & Society. 2000; 3(2-3):297–316. https://doi.org/10.1207/S15327825MCS0323_07

**80.** Fitton L, Hussain A, Leaning B. Twitter for dummies. John Wiley & Sons; 2014.

**81.** Hemphill L, Culotta A, Heston M. Framing in social media: How the US Congress uses Twitter hashtags to frame political issues. Available at SSRN 2317335. 2013;.

**82.** Zappavigna M. Ambient affiliation: A linguistic perspective on Twitter. New media & society. 2011; 13 (5):788–806. https://doi.org/10.1177/1461444810385097

**83.** Niwattanakul S, Singthongchai J, Naenudorn E, Wanapu S. Using of Jaccard coefficient for keywords similarity. In: Proceedings of the international multiconference of engineers and computer scientists. vol. 1; 2013. p. 380–384.

**84.** Boukes M, Trilling D. Political relevance in the eye of the beholder: Determining the substantiveness of TV shows and political debates with Twitter data. First Monday. 2017; 22(4). https://doi.org/10.5210/fm.v22i4.7031

**85.** Bruns A, Burgess JE. The use of Twitter hashtags in the formation of ad hoc publics. In: Proceedings of the 6th European Consortium for Political Research (ECPR) General Conference 2011; 2011. p. 1–9.

**86.** Bruns A, Stieglitz S. Quantitative approaches to comparing communication patterns on Twitter. Journal of Technology in Human Services. 2012; 30(3-4):160–185. https://doi.org/10.1080/15228835.2012.744249

**87.** Dahlgren P. The Internet, public spheres, and political communication: Dispersion and deliberation. Political communication. 2005; 22(2):147–162. https://doi.org/10.1080/10584600590933160

**88.** Farrell H. The consequences of the internet for politics. Annual review of political science. 2012; 15:35–52. https://doi.org/10.1146/annurev-polisci-030810-110815

**89.** Papacharissi Z. The virtual sphere: The internet as a public sphere. New media & society. 2002; 4(1):9–27. https://doi.org/10.1177/14614440222226244

**90.** Castells M. The new public sphere: Global civil society, communication networks, and global governance. The ANNALS of the American academy of Political and Social Science. 2008; 616(1):78–93. https://doi.org/10.1177/0002716207311877

**91.** Cherepnalkoski D, Karpf A, Mozetič I, Grčar M. Cohesion and coalition formation in the European Parliament: roll-call votes and Twitter activities. PloS one. 2016; 11(11). https://doi.org/10.1371/journal.pone.0166586 PMID: 27835683

**92.** Scherpereel JA, Wohlgemuth J, Schmelzinger M. The adoption and use of Twitter as a representational tool among members of the European Parliament. European politics and society. 2017; 18(2):111–127. https://doi.org/10.1080/23745118.2016.1151125

**93.** Larsson AO. The EU Parliament on Twitter—Assessing the permanent online practices of parliamentarians. Journal of Information Technology & Politics. 2015; 12(2):149–166. https://doi.org/10.1080/19331681.2014.994158

**94.** Domínguez S, Hollstein B. Mixed methods social networks research: Design and applications. vol. 36. Cambridge University Press; 2014.

**95.** Salvatore D. Globalisation, international competitiveness and growth: advanced and emerging markets, large and small countries. Journal of International Commerce, Economics and Policy. 2010; 01(1):21–32. https://doi.org/10.1142/S179399331000007X

**96.** Jeffares S. Interpreting Hashtag politics: policy ideas in an era of social media. Springer; 2014.

**97.** Small TA. What the hashtag? A content analysis of Canadian politics on Twitter. Information, communication & society. 2011; 14(6):872–895. https://doi.org/10.1080/1369118X.2011.554572