PLOS ONE

Ideological differences in engagement in public debate on Twitter

Felix Gaisbauer * , **Armin Pournaki‡, Sven Banisch‡, Eckehard Olbrich** ‡

Max Planck Institute for Mathematics in the Sciences, Leipzig, Germany

‡These authors also contributed equally to this work.
* felix.gaisbauer@mis.mpg.de

Abstract

This article analyses public debate on Twitter via network representations of retweets and replies. We argue that tweets observable on Twitter have both a direct and mediated effect on the perception of public opinion. Through the interplay of the two networks, it is possible to identify potentially misleading representations of public opinion on the platform. The method is employed to observe public debate about two events: The Saxon state elections and violent riots in the city of Leipzig in 2019. We show that in both cases, (i) different opinion groups exhibit different propensities to get involved in debate, and therefore have unequal impact on public opinion. Users retweeting far-right parties and politicians are significantly more active, hence their positions are disproportionately visible. (ii) Said users act significantly more confrontational in the sense that they reply mostly to users from different groups, while the contrary is not the case.

Introduction

Twitter is an immensely popular object of study for social scientists in a variety of contexts, ranging from politics [1] to crisis communication [2]. One reason for this popularity is that Twitter is open in a double sense: On the one hand, researchers can call Twitter data conveniently via an API. On the other hand—and more importantly—the content created on the platform is public by default. *In principle*, a user's activity is visible to everyone on the platform, and any user can interact with anyone else.

Due to its open platform design, user interactions on Twitter might, of all major social media platforms, come closest to what is commonly referred to as 'public debate.' While not being representative of the general public [3, 4], Twitter provides a public arena for information gathering, opinion formation and persuasion. Since journalists incorporate the platform in their daily routines [5, 6], explicitly refer to content visible on Twitter as public opinion [7], and even tend to judge tweets as newsworthy as press agency reports [8], the standpoints that are prominently featured there are reinforced in traditional media. A better understanding of how different opinion groups shape debate on the platform is therefore highly important: The image created on the platform not only affects how public opinion on certain issues is

perceived by its users, but by society more generally. Certain standpoints—advanced by committed minorities in particular—might appear more prevalent than they actually are.

This explorative study attempts to make these systematic differences in the engagement of groups with different political leaning visible. Its goal is both methodological and case-oriented: First, we propose a novel method to assess what users and lurkers perceive as public opinion on a specific issue on Twitter. Secondly, we employ this method in two case studies covering two political events in Germany.

To this end, we firstly choose a suitable theoretic underpinning for the concepts of public debate and public opinion, which have been interpreted from different angles [9, 10], and which we connect to findings on user comments and their effects on readers. Then, we describe how relevant tweets are collected, and the data transformations which yield a social-structural view on interactions between users. The proposed method relies on an interplay of network representations of two types of user interactions on Twitter: Retweets and replies. Retweet networks are used to discern opinion groups, while reply networks make it possible to assess how these groups participate in public debate. We construct said networks from a user-centered data collection for two events: A state election in the German state of Saxony and a violent clash on New Year's Eve between police and parts of the population in the city of Leipzig. We show for both cases that while retweet networks are strongly polarized, debate between users of different opinion clusters is vivid. We also show that the impression of public opinion is biased. While being a minority in number, Twitter users who mostly share content of parties to the right of the political spectrum are disproportionately active in debate and act more confrontational in the sense that they address users from different opinion groups more often.

Theory

Public debate and public opinion

A communication-based view on public debate has been provided by Vincent Price. He describes public debate as "communication processes through which publics are constituted and within which opinions on public affairs are formed" [10, p. 74]. While he invokes the analogy of a big town meeting, the technical feasibility of creating such a meeting still seemed out of reach in the early nineties: "Modern communication technologies may have enabled the enlargement of public consciousness [. . .] but they have not come close to creating any sort of town meeting at large" [10, p. 78].

With the advent of social media, the analogy appears to have become a (digital) reality. As has been stressed in the introduction, Twitter is a public medium and allows its users to interact with potentially anyone else on the platform. Users can share others' thoughts, put out their own, and organize around hashtags, thereby creating publics and attracting attention of others [11]. These processes are strongly reinforced and amplified by traditional media: Journalists incorporate social media, especially Twitter, as an established news source in their daily routine [5, 6, 8], journalistic content or events on television are discussed on the platform in parallel [12, 13], and Twitter content is often explicitly used to represent public opinion, both in qualitative (by quoting certain tweets, e.g. to underline meta-narratives) and quantitative fashion [7]. Often, social media platforms themselves provide tools or even supply journalists with data or analyses in order to get mentioned in their articles [7].

Two different, basic paradigms of public opinion can be subsumed under the terms *discursive* and *demoscopic* public opinion [10, 14]. The former refers to public opinion as a social-structural phenomenon and has a strong normative imprint. The process of arriving at public opinion—public debate—is understood as as a rational discourse between well-informed citizens [14], and should lead to the best possible decision with respect to the overall good. The

latter is related to survey research where scientists seek to aggregate the attitudes of individuals towards certain issues in a representative fashion, which yields, by majority rule or a breakdown of percentages, public opinion.

For an understanding of online interactions, both conceptions are problematic. There have been attempts to replace classical voter surveys for elections by social media observations [15, 16], but the findings have been contested by others [1] or turned out to be incorrect (Burnap et al. [16] predicted a Labour win in the 2015 UK elections). Discursive public opinion as a normative concept, on the other hand, is in general hardly accessible to empirical research. And, after all, the internet does know many compulsions besides "the unforced force of the better argument" [17, p. 306].

Elisabeth Noelle-Neumann has proposed a social-psychological approach centered around *observable* opinion expressions. ('Expressions' is interpreted very broadly: Noelle-Neumann's account includes also non-verbal modes of communication, e.g. badges that support certain political parties or even subtle facial expressions such as raised eyebrows). She conceives public opinion—or rather, what people perceive as the opinion of 'the' public—as a force of informal social pressure and control manifesting itself in "approval and disapproval of publicly observable positions and behavior" [9, p. 64]. Her operational definition of public opinion incorporates "opinions on controversial issues that one can express in public without isolating oneself" [9, p. 63]. Especially for controversial topics, individuals, being social creatures and fearing social isolation, constantly and mostly sub-consciously monitor their social environment and the mass media. They estimate the majority opinion around them, employing some "quasi-statistical sense" [18], which they then refer to as public or popular opinion. The theory hence puts strong emphasis on the role of the subjective impression of public opinion of individuals. Noelle-Neumann's spiral of silence theory states that if people realize that they hold an opinion that differs from their impression of public opinion, they tend to be less willing to express their opinion publicly any longer. This, in turn, affects the perception of public opinion of others, potentially setting off a spiralling process in which certain groups become more expressive while others fall silent. Quantified interactions in online environments (Twitter displays the number of likes, retweets, and replies below each tweet) might suggest themselves as an objective foundation for the quasi-statistical impression of the opinion climate—but as we will show, these interactions themselves can be subject to strong biases.

Comment spaces and perceived public opinion

In order to capture public debate online, comment sections, predominantly on news websites, have been the target of attention since their introduction [19–23]. Studies found that user comments on news articles affected individuals' perceptions of public opinion [20, 21]—more so than simple comparisons of likes and dislikes. In the 20th century, scholars often distinguished between representatives of interest groups that debated publicly, and a large and more spectator-like 'body' which then reacted to the debate and approved or disapproved, hence formed public opinion [10, p. 27]. This relation is reproduced by the combination of newspaper articles and user comments below, which allow directly visible engagement in larger amount and with less control than the very limited number of redacted letters to the editor.

Allowing differing points of view to reach an audience with few formal constraints, comment spaces have also been interpreted as counter-public spaces (or spheres) [23, 24], an expansion of the Habermasian concept of the public sphere. Nancy Fraser originally defined counterpublics as "parallel discursive arenas where members of subordinated social groups invent and circulate counterdiscourses to formulate oppositional interpretations of their

identities, interests, and needs" [25, p. 67], arising in response to hegemonic "publics at large" [25, p. 67]. Comment sections in general (may they be the comment sections of newspapers, or the reply thread of a tweet of a public figure on Twitter) are quite suitable for the formulation of oppositional interpretations: They are in the direct vicinity, nevertheless clearly demarcated from the interpretations and content they want to distance themselves from [23].

Lessons of these findings and interpretations are (i) that comment sections have a significant effect on how people perceive public opinion on an issue—in Noelle-Neumann's terms, the "opinion climate" [9]—and (ii) that they are hence, for all kinds of interest groups, important arenas for confrontation and contestation of certain interpretations and frames, may they be hegemonic or not. Therefore, a careful investigation of comment sections and the views expressed there is important: Which standpoints are expressed, how often, and which viewpoints (or users) remain silent? That different groups strive for the award of being called *the* public is nothing new. As Baker notes of the pre-revolutionary times in France: "Indeed, one can understand the conflicts of the Pre-Revolution as a series of struggles to fix the sociological referent of the concept in favor of one or another competing group" [26, p. 186]. Online environments, which facilitate communication and decentralize information distribution, might appear to make this competition more transparent. But they also introduce additional potential for misperceptions, not least due to differing willingness of public opinion expression of different groups [22, 27, 28]. For certain opinion groups, this can lead to a False Consensus Effect [29], according to which individuals see their own opinions as more prevalent in society than they actually are. On the other hand, groups less willing to express their opinion might underestimate their size (False Uniqueness Effect, see Mullen, Dovidio, Johnson and Copper [30]). The method in this contribution will allow an estimation of how public opinion and public debate are perceived on Twitter, and which opinion groups principally shape this impression.

Political background

The two events under consideration were the Saxon state election which took place on September 1st, 2019, and a violent clash between police and parts of the population in the city of Leipzig on New Year's Eve four months later (in the following abbreviated with NYE). The events were complementary in the sense that the election was long-anticipated, while the latter occasion was a spontaneous incident, making them suitable for comparison.

The election was of special, nation-wide interest. Since Saxony had been the birthplace of the anti-Islam movement *Pegida* in 2014, which received international attention, the election was considered a litmus test for the mobilizing potential of extreme forces. Before the election, it was not clear whether there would be the possibility of forming a majority coalition without participation of the *Alternative für Deutschland* (*AfD*), a right-wing party founded in 2013, or the democratic socialist party *Die Linke*. While the Christian democratic party *Christlich Demokratische Union Deutschlands* (*CDU*) leading the polls ruled out coalitions with both of the parties, it was publicly discussed whether parts of the *CDU* were open towards collaboration with the *AfD* [31].

The NYE incident was not anticipated, but a spontaneous event which was subsequently discussed not only in Saxony, but also in national politics. On New Year's Eve, violent riots and attacks on police officers occurred in the city of Leipzig's quarter Connewitz. The event was particularly polarizing: While some political actors framed the incident as an example of the violent potential of left-wing extremism, others accused the police of deliberate provocation [32].

Methods

Data acquisition

The data was collected in a user-centered approach. That is, all tweets that were produced by a seed set of users were gathered. Moreover, all tweets containing the Twitter handle of one or more of the seed users were collected—this included retweets, mentions, and replies to the users. With this method, not only first-order replies to a user in question could be collected, but any part of a reply tree that had been initiated by the user. In the months preceding the Saxon state elections in September 2019, a first seed set was constructed. Candidates in elections in Germany appear on electoral lists. We collected all names from these lists and checked whether the candidates had an actively maintained Twitter account. If this was the case, we included them. We also included

- state party accounts

- the leaders of the fractions in the state parliament and the local party organizations

- the Saxon members of the national parliament

- the Saxon members of the European parliament

- Saxon media accounts

- Saxony correspondents of national media

that had an active Twitter account. The seed set was expanded with a snowball-sampling method: In each sampling step, users that were not contained in the seed set but retweeted or mentioned at least once a week by users that were already in the seed set and were related to Saxony were included. The latter criterion was necessary to exclude nation-wide media accounts and national politicians. At the end of July, after seven iterations, the final seed set consisted of 270 users.

Tweets were gathered until February 2020. This allowed observation of other Saxony-related events, such as the NYE incident. Since the seed set was not perfectly tailored towards this event, we restricted the analysis to the subset of tweets containing certain event-specific keywords (*connewitz, antifa, polizei, polizist, le0101, linx, leipzig, Not-OP, notop, linke, chaoten, angriff, le3112, randal,* applied to the root tweets and incident-specific retweet network). The tweets used for the analysis of the election stem from the time period between the 25th of July and the 10th of September (364,626 tweets). For the NYE data set, tweets from December 31 until January 19 (130,685) were used. The two cases were chosen since one (NYE) represents a spontaneous event, while the other a long-anticipated election—a suitable test whether similar effects can be observed for both types of events. Moreover, the data sets were of considerable, but still maintainable size (especially regarding visualization) and therefore appropriate for testing the method.

Network representations

We used networks as a mathematical abstraction to represent two types of interaction in the data set: retweets and replies. Retweet networks were used to discern different opinion communities on Twitter, while reply networks made it possible to assess how these groups participated in public debate.

Retweet networks. Retweet interactions are represented as a directed network in which every node is a user. A link is drawn from user a to user b every time a retweets b. It has become standard practice to employ community detection algorithms to find strongly

connected clusters in a retweet network which are then interpreted as groups of users sharing an opinion or political position [33–35].

An alternative approach is the spatialisation of the retweet network via a force-directed algorithm, such as ForceAtlas2 [36], which we used in its Gephi [37] implementation. Noack [38] has shown that the energy-minimal states of force-directed layout algorithms such as ForceAtlas2 are in fact relaxations of modularity maximization. While in order to maximize modularity, nodes must be classified into discrete clusters, force-directed layouts assign continuous positions—in the case of ForceAtlas2, in a two-dimensional space. Positions in the layout are hence closely related to partition outputs of modularity maximization algorithms such as Louvain. The advantage of force-directed layout algorithms is that one can *visually* distinguish tightly-knit clusters and less dense in-between regions, which might be classified as one cluster in the Louvain algorithm. While "community detection algorithms tend to generate clear-cut and non-overlapping partitions, force-directed spatialisation reveals zones of different relational density but with blurred and uncertain borders" [39, p. 13]. Public debate and the different opinion camps cannot be clearly demarcated from one another, and regions of transition between different groups, that do not clearly belong to any one of them, are politically meaningful and should hence be discernible. As will be shown in the following section, retweet networks of both events showed polarized structure in the force-directed layout. We classified the two cohesive poles of the retweet network as two opinion clusters (the borders chosen as is visible in Fig 1A and 1B), and assigned the in-between region to a different cluster upon visual inspection.

Reply trees and reply networks. Due to the user-centered data collection, it was also possible to retrieve an exhaustive collection of all replies that were initiated by posts of the seed users. A post together with all its replies can be represented by a reply tree (see Fig 2). Only taking into account reply trees initiated by the prominent seed users corresponds quite naturally to the distinction of the previous section between representatives of interest groups debating publicly and the spectator-like body approving or disapproving subsequently [10]. These reply threads then function as spaces where different opinion groups can confront each other: They are widely visible due to the prominence of the creator of the tweet which spans up discussion, and can hence attract users of different opinion camps. Retweets, on the other hand, mainly serve to share information with one's followers. A retweet might point towards a debate, but does not imply involvement in it.

In order to gain a global view on public debate, we aggregated the combined interaction structure of all reply trees into one reply network, assigning a directed edge between two users if one had directly replied to the other in a tree (see Fig 1). (Obviously, trees are networks, too. But if we in the following speak of reply networks, we mean the bigger networks constructed in this procedure).

Some works have taken similar routes by taking into account direct user interactions in the form of mentions [34] and replies [40–42]. Sousa, Sarmento and Mendes Rodrigues [40] and Yardi and boyd [42] use a keyword-based tweet collection. This approach is useful if one is solely interested in tweets that include a certain keyword, while full conversations in a reply thread between users are not accessible with the method. Aragón, Kappler, Kaltenbrunner, Laniado and Volkovich [41] and Nuernbergk and Conrad [43] employ a user-centered collection and construct a reply network, but only between politicians on Twitter and hence do not capture debate among a more general public. Since the data sets in the present contribution include the complete reply trees below each post of one of the seed users, it was possible to gain a more general perspective on public debate that did not only include certain elites.

The classification of users from the retweet network—the information about whether they belonged to the minority or majority pole, or the in-between region—was imported into the
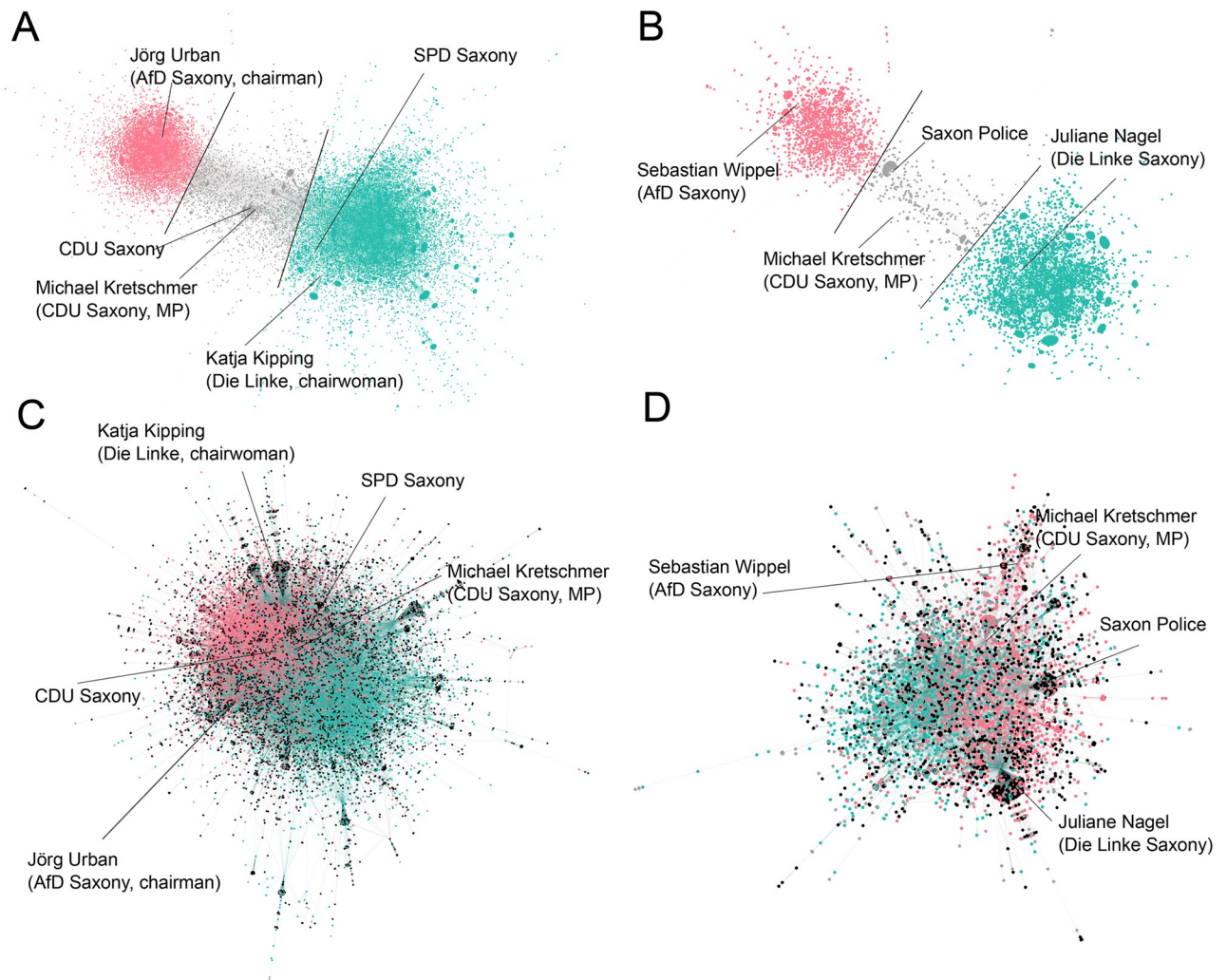
Fig 1. Retweet and reply networks. The giant component of the retweet networks (above) for the elections (A) and the NYE incident (B), and the giant component of the reply networks (C: elections, D: NYE). The retweet networks are both polarized, and users are classified according to their position in the force-directed layout. The borders for the classifications are included (black lines). The classifications are imported in the reply networks, which are not polarized. Black nodes indicate users that do not show up in the corresponding retweet network.

reply trees and networks. This made it possible to investigate how many users of the different retweet clusters were also involved in public debate, hence willing to express their opinion in discussion with others of possibly different opinions, and whether users of different opinion clusters debated mainly among each other or with others.

It must be noted here that not all users involved in debate were present in the retweet network. Hence, the classification in the reply trees and networks was not complete. Initially, around 47% of the users involved in debate in the election data set could be classified (33% for NYE). In order to include more users in the classifications, a larger retweet network was additionally constructed which included all retweets from July 2019 until the end of February 2020. The overall structure of the network was similar to the incident-specific retweet networks (see S1 File). If a user was present in the reply trees, but not present in the incident-specific retweet network, it was checked whether the user was present in the large retweet network—if so, the
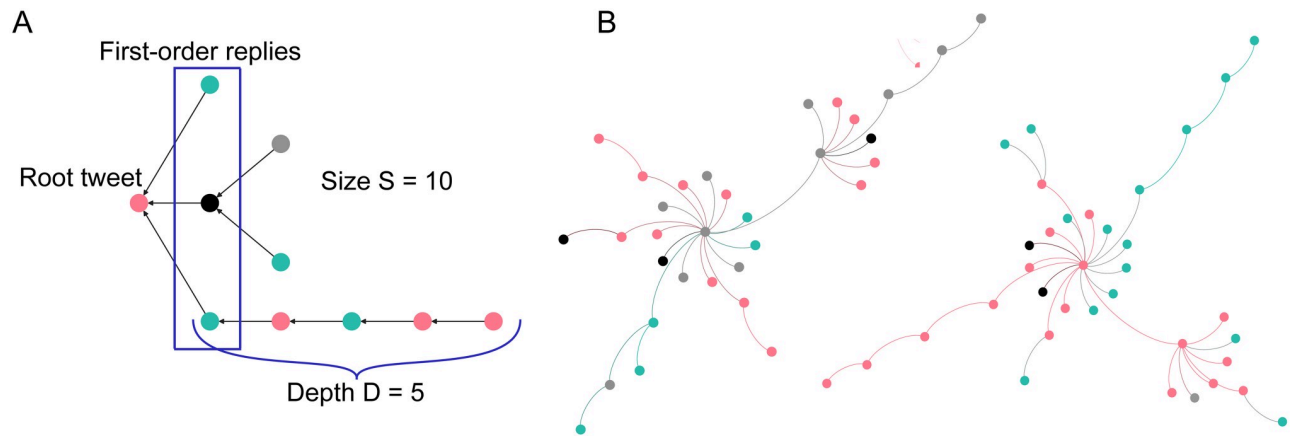
Fig 2. Reply trees. An exemplary dummy reply tree (A), along with two reply trees from the data set (B). Each node represents a tweet, and a directed edge between two nodes indicates a reply. If the users of the tweets appear in the retweet network, their replies were color-coded according to the cluster of the user (a black node indicates a reply by a user that does not appear in the retweet network). The root tweet is the original tweet by one of the seed users, while first-order replies are the direct replies to the root tweet.

user was assigned the classification from this network. With the use of the big retweet network, 63% (election) and 67% (NYE) of users present in the reply trees could be classified.

Results

Retweet networks and classification

The retweet networks for both cases are strongly polarized (see Fig 1A and 1B) in the force-directed layout.

For the election data set (31,108 users in the giant component), seed users placed in the majority pole are politicians of the parties *SPD*, *Die Linke*, and *Bündnis 90/Die Grünen* (and one politician of the *CDU*), along with media accounts (e.g. *Bild Leipzig*, *LVZ*, or *MDR Sachsen*) and left-wing activists. In the region between the two clusters, politicians of the *CDU*, *Freie Waehler* and *FDP* are placed, as well as media accounts (e.g. *MDR Aktuell*, *Bild Dresden*, or *TAG24*). The minority pole, on the other hand, includes seed users from the *AfD*, *Freie Waehler*, *Blaue Partei* and the anti-Islam movement *Pegida*. The structure of the retweet network hence quite accurately mirrors the political constellations in the run-up to the elections. Left-wing and eco-friendly parties are placed in one cluster and the right to far-right parties in another, while politicians of the market-liberal *FDP* and the center-right *CDU* are located in-between the two.

The users of the majority cluster made up 64.5% (20,052) of the retweet network, 23.1% (7,195) were part of the minority cluster, and 12.4% (3,861) of the users were in the region between the two.

A very similar structure, both in proportions and in political leanings, is given for the NYE incident. Some differences occur, however. The set of users placed in-between the two clusters (711 or 7.9%) includes the official account of the city of Leipzig and the account of the Saxon police, as well as one politician of the *AfD*, one from *Die Linke* and one *SPD* politician. The majority (6,010 users, 66.6%) and minority cluster (2,301 users, 25.5% of the giant component) show similar composition as in the election retweet network.

The classification of users on the basis of their position in the force-directed layout of the retweet network was taken as a proxy for the political position of the users in the two issues. It must be noted, however, that users of one cluster should not necessarily be interpreted as
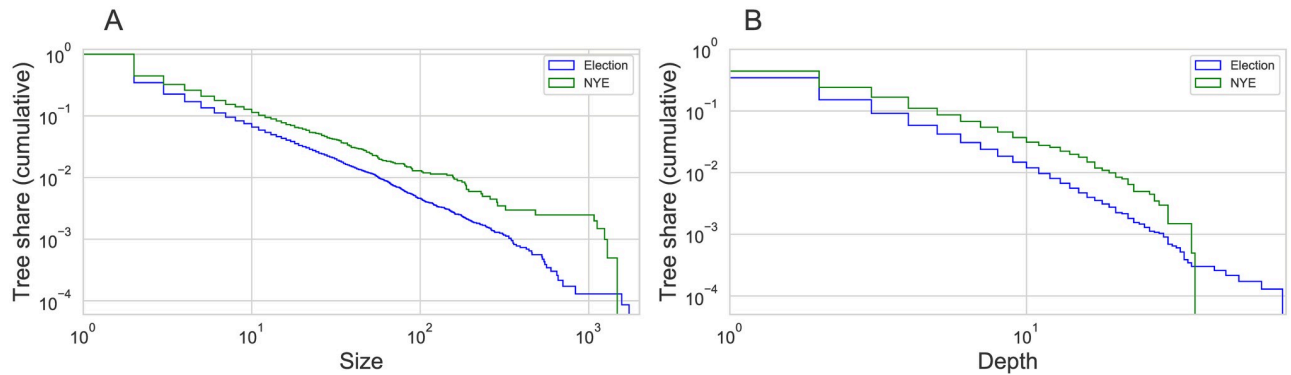
Fig 3. Reply tree size and depth. Reply tree size (A) and depth distribution (B) for the two events.

holding exactly the same opinion or political position. Rather, the clusters reflect an issue-specific fundamental political difference which is then also reflected by the classification.

A randomly selected subset of 100 users was used check whether the groups assigned to the users were plausible in the sense that the users tweeted content sympathetic to one of the parties or political figures in their assigned cluster. Out of the 100 users, 96 acted consistently with their classification, while 4 did not.

Reply trees and engagement

All reply trees that had been initiated by a root tweet of one of the seed users were taken into account in the analysis of the reply interactions. The reply trees themselves can be seen as representations of discussions triggered by single statements, and can exhibit arbitrarily complicated tree shape (see Fig 2) and have arbitrarily many participants. Not every tweet receives replies. There are 23,221 posts from the seed users in the election data set that were not replies or retweets, out of which 8,033 received at least one reply. (NYE: 2,020 posts, 897 with at least one reply).

Reply trees can be characterized by two quantities: Their size S, which is the overall number of tweets in the tree, and their depth D, which is the longest branch of the tree (Fig 2). Fig 3 shows the cumulative distribution of sizes and depths of the reply trees in the two data sets. In both data sets, around 90 percent of all reply trees have a size smaller than 10 and a depth smaller than 5. Nevertheless, reply trees can be very large: The largest tree in the election data set has 1,936 replies (NYE: 1,475), and the maximum depths are 72 and 37, respectively.

By importing the cluster classification from the respective retweet network, it was possible to compare the engagement of the different groups in the reply trees. The engagement proportions in the debate differ strongly from those in the retweet network, both in number of users and number of tweets (see Table 1).

Table 1. Users and replies from the different retweet clusters involved in the reply trees.

| | Election | | NYE | |
	Users	Replies	Users	Replies
Majority cl.	6,736 (33.3%)	30,615 (40.5%)	2,008 (32.4%)	6,403 (39.2%)
Minority cl.	4,696 (23.2%)	23,790 (31.5%)	1,727 (27.9%)	4,858 (29.8%)
Intermediate cl.	1,470 (7.0%)	6,785 (9.0%)	389 (6.3%)	1,235 (7.6%)
Not classified	7,395 (36.5%)	14,344 (19.0%)	2,072 (33.4%)	3,816 (23.4%)

Table 2. First-order replies by retweet clusters. First-order replies from users of the two poles are roughly equal in number in both data sets and make up the majority of all first-order replies. The minority cluster is even more active in replies of first order than in reply trees in general.

	Election	NYE
Majority cl.	15,347 (36.0%)	2,445 (32.7%)
Minority cl.	14,911 (35.0%)	2,451 (32.8%)
Intermediate cl.	3,710 (8.7%)	572 (7.6%)
Not classified	8,629 (20.3%)	2,014 (26.9%)

https://doi.org/10.1371/journal.pone.0249241.t002

In the election data set, participants from the majority and minority retweet clusters made up 33.3% and 23.2% of all replies, respectively. For the NYE incident, even more users from the minority retweet cluster participated in the debate—32.4% belonged the majority in the NYE retweet cluster, while 27.9% were minority users involved in the debate (see Table 1). Users placed in-between the two poles in the retweet network also participated in the debate, but less so both in number of users and in number of replies. 36.5% (election) and 33.4% (NYE) of the users that participated in the reply trees were not present in the retweet network (i.e., they did not retweet any of the seed users nor any tweets that contained the Twitter handle of one of the seed users). Interestingly, while these users make up the biggest number of users involved in the debate, they do not constitute the majority in terms of replies. In both data sets, users from the poles of the retweet network, if involved in the debate, are most active. Users from 'outside' do not tend to debate often and extensively—on average, in both data sets, they only give around two replies.

In comparison, the ratio between majority and minority pole in the corresponding retweet network was 64.5% to 23.1% (election) and 66.6% to 25.5% (NYE).

First-order replies are of special interest since they are usually directly displayed below the root tweet on Twitter. Therefore, they most probably have a stronger impact on the perception of public opinion than tweets that are at the end of a long discussion branch. The amount of first-order replies by the different clusters is displayed in Table 2. First-order replies from users of the two poles are roughly equal in number in both data sets. Minority pole users hence produced an even larger proportion of highly visible replies. Users from the intermediate region in the retweet network only account for less than 10 percent of first-order replies, while users that were not present in the retweet networks produced 20.3% (election) and 26.9% (NYE) of the replies of first order.

Comparing engagement in the form of replies and retweets makes it possible to assess whether the different opinion groups show different inclination to participate in the debate. To this end, we calculate the share of users present in the different clusters of the incident-specific retweet networks that were also present in the respective reply network (see Table 3). For both events, the groups showed significantly different behavior (election: χ^2 = 850.7, $p < 0.001$; NYE: χ^2 = 138.2, $p < 0.001$). Users of the minority cluster were roughly twice as likely to get

Table 3. Percentage of users in the incident-specific retweet networks that are active in the reply networks (seed users excluded) by cluster. The share of users from the minority pole is, in both cases, around twice as big as the share of users from the majority pole. User share from the in-between region is slightly bigger than that of the majority pole in both cases.

	Election	NYE
Majority cl.	22.4% (4,644)	16.2% (977)
Minority cl.	48.2% (3,604)	32.2% (752)
Intermediate cl.	26.4% (1,050)	22.0% (707)

https://doi.org/10.1371/journal.pone.0249241.t003

involved in the debate than users belonging to the majority cluster (election: z-score 41.0, $p < 0.001$; NYE: z-score 15.0, $p < 0.001$). Users from the in-between region were slightly more active in the debate than users from the majority pole (election: z-score 5.4, $p < 0.001$; NYE: z-score 4.0, $p < 0.001$), but still less than the minority pole (election: z-score 22.1, $p < 0.001$; NYE: z-score 46.0, $p < 0.001$).

Hence, two findings are worth stressing: (i) Minority pole users are disproportionately active in the debate compared to the majority pole, both in number of users involved and of replies written (see also the Discussion). This effect is even more pronounced in first-order replies that are, by platform design, most visible. And (ii), users from both poles of the retweet network, if they take part in the debate in the form of replies, do so more extensively than users from in-between the poles or unclassified users.

Reply networks and global interaction patterns

The reply networks give a more comprehensive structural picture of debate—it is possible to make visible patterns of discussion between different users and user groups beyond interactions in single reply trees. Each reply network was constructed by aggregating all reply interactions in the reply trees into one big network, where each node represents a user and a directed edge is created between two users if one has replied to the other.

The question of interest here is whether the groups also exhibit large-scale polarization when they discuss among each other, and whether there are differences in discussion behavior between the groups. If public debate was fragmented in the sense that discussion ties were only existent amongst a certain subset of users, this would be visible in the force-directed layout of the reply network. But, as is displayed in Fig 1C and 1D, this is the case for neither of the two events (again, spatialisation was carried out with ForceAtlas2). Users of clusters that are clearly separated in the force-directed layout of the retweet networks interact quite frequently in the form of replies.

A useful measure describing the tendency of individuals in a network to link to others with similar properties or attributes is assortativity [44], which yields one assortativity coefficient for a whole network. An assortativity coefficient of $r = 1$ means that all edges in the network only connect nodes of the same type, while for $r = -1$, the edges only connect nodes of different type (hence, the network is strongly disassortative). It has been argued that such a global view might obstruct insights into local differences between individuals or groups [45]. Using local assortativity has been proposed in order make those differences visible—each node l in a network is assigned a local assortativity score $r(l)$, such that differences in the score can be compared across all nodes. Local assortativity $r(l)$ is defined by the equation [45]

$$r(l) = \frac{1}{Q_{max}} \sum_g (e_{gg}(l) - a_g b_g), \tag{1}$$

with Q_{max} as the maximum modularity, which normalizes the assortativity coefficient. Maximum modularity is reached if all edges in the network only connect nodes of the same type. The proportion of edges in the local neighborhood of node l which connect nodes of the same type g is compared to $a_g b_g$, which is the proportion of edges between nodes of group g if one would randomly create edges between nodes, while keeping the total number of outgoing and incoming edges for each type constant. a_g here is the proportion of edges starting from nodes of group g, while b_g is the proportion of edges ending at nodes of group g. In general, e_{gh} is
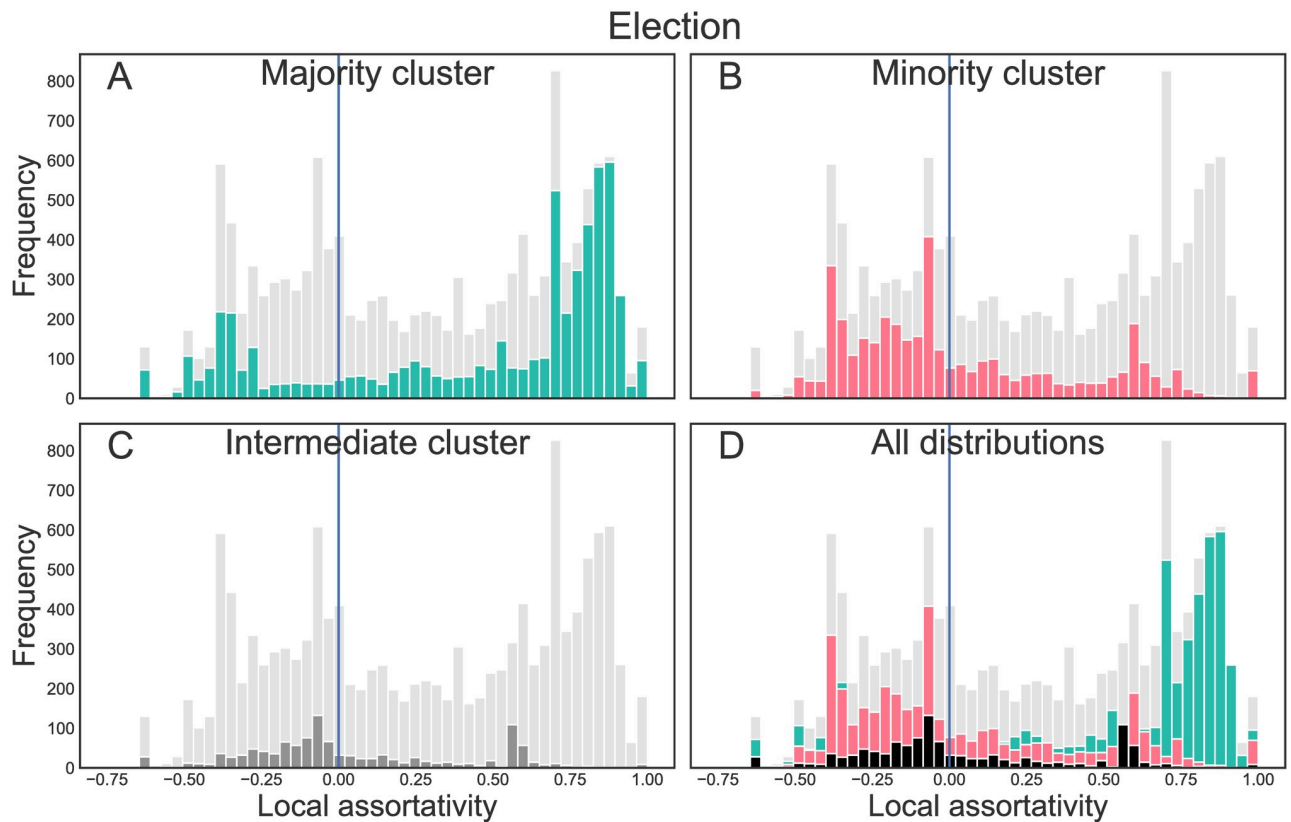
Fig 4. Local assortativity distribution for the election data. Local assortativity distribution for the reply networks of the election, split up by the users'
respective retweet clusters (A: majority cluster, B: minority cluster, C: intermediate cluster, D: comparison of all distributions). The grey distribution in
the background is the overall local assortativity distribution, the distribution of the respective subgroup(s) is plotted in the foreground and color-coded
as before.

given by

$$e_{gh} = \sum_{i:y_i=g} \sum_{j:y_j=h} w(i;l)\frac{A_{ij}}{k_i}.$$

(2)

$w(i;l)$ is a distribution over all nodes designed to capture the mixing patterns within the local
neighborhood of node l. We follow [45] in choosing the personalized PageRank vector. The
local assortativity coefficient is also capable of including incomplete metadata—since in the
current data set, not all users could be classified, this feature is beneficial. In the histograms dis-
played in Figs 4 and 5, the node contributions to the histograms were adjusted according to
the weight

$$z_l = \sum_{gh} e_{gh}(l),$$

(3)

the sum of local edge counts with *known* metadata. (Code available at [46]).

In Figs 4 and 5, the local assortativity distributions of users of the different groups are dis-
played. The distributions are multimodal, i.e. exhibit more than one peak. In the election data
set, users of the majority cluster (A) have their largest peak at a local assortativity close to 1, i.e.
they reply mainly to users of their own clusters (and, since local assortativity also takes into
account the assortativity of the broader neighborhood, also mainly interact with users who do
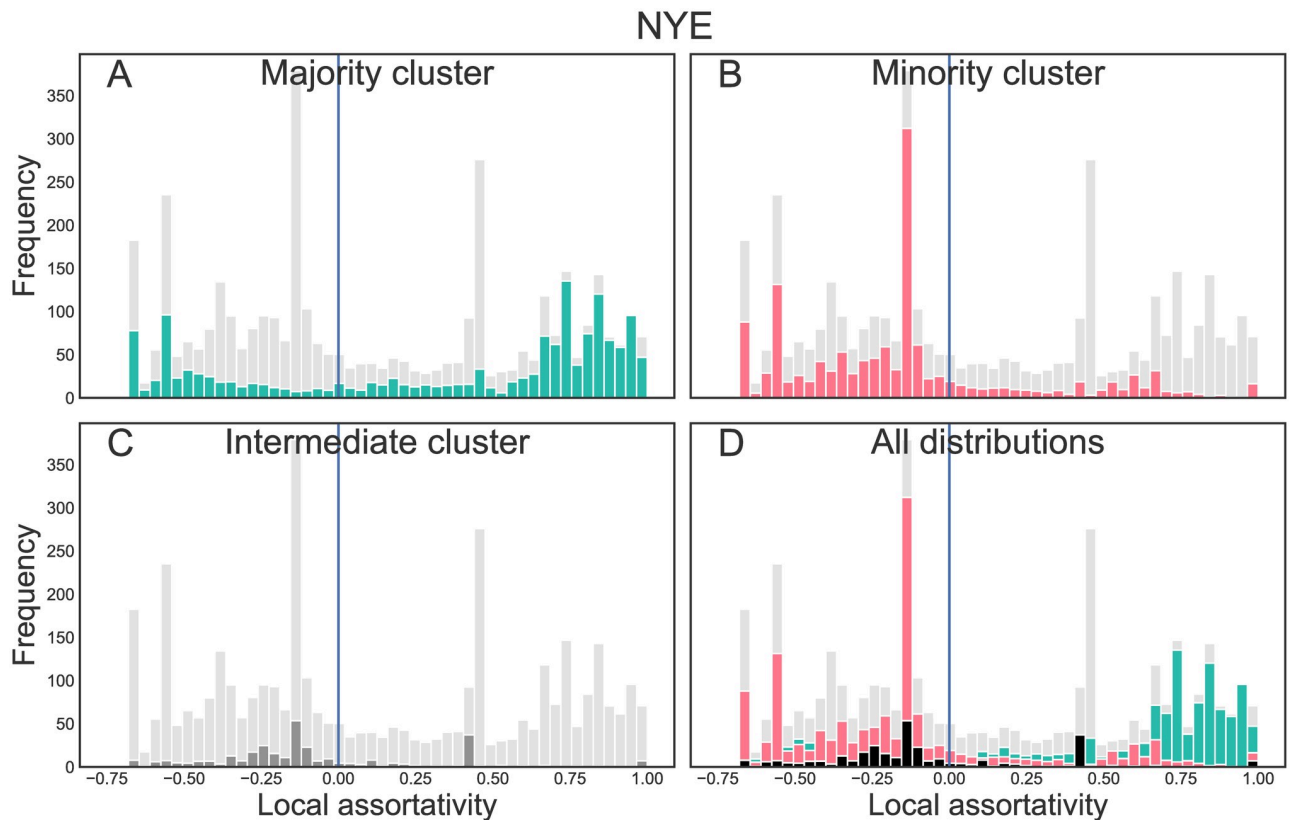
NYE

Fig 5. Local assortativity distribution for the NYE data. Local assortativity distribution for the reply networks of the NYE incident, split up by the users' respective retweet clusters (A: majority cluster, B: minority cluster, C: intermediate cluster, D: comparison of all distributions). The grey distribution in the background is the overall local assortativity distribution, the distribution of the respective subgroup(s) is plotted in the foreground and color-coded as before.

the same). A second, yet smaller peak is visible with negative local assortativity (around −0.3). Hence, some majority cluster users mainly seek debate with users from other clusters. The reverse holds for the minority cluster (B): There, most of the users reply to users from the other groups. A smaller peak is visible at positive local assortativity (around 0.6). Users from the intermediate region (C) also exhibit a multimodal distribution, with one slightly negative peak and one peak at around 0.6. The NYE data shows an even more pronounced trend: Only few users of the minority cluster get assigned a positive local assortativity. Majority group users again show a bimodal distribution with one peak close to $r_l = 1$, the other peak is now at around $r_l = -0.6$.

Further insight into the interaction patterns between the different clusters is provided by Table 4, which shows counts of replies between and within the clusters. While users of the majority cluster of the retweet network address almost 60% (election) or 50% (NYE) of their replies to other users of their own cluster, the opposite holds for the other groups: Most of their replies are directed towards users from the majority cluster. For the election, users of the intermediate cluster replied more often to others from their cluster than to the minority pole, while for the minority pole, replies to the own and replies to the intermediate cluster were almost equal in number. The number of tweets from intermediate users was considerably lower compared to tweets from the poles, hence it appears that users of the minority pole were

Table 4. Reply interactions between and within the different clusters—The columns show what cluster the replies are from, the rows who the cluster replies to. The percentages (in brackets) are given with respect to the overall number of replies a cluster has given. Note that the column entries do not sum up to the overall number of given replies since there are also replies from users that do not show up in the retweet networks, hence could not be classified.

to/from	Election			NYE		
	Majority cl.	Minority cl.	Intermediate cl.	Majority cl.	Minority cl.	Intermediate cl.
Majority cl.	18,185 (59.4%)	10,590 (44.5%)	3,596 (53.0%)	3,234 (50.5%)	2,814 (57.9%)	821 (66.5%)
Minority cl.	5,778 (18.9%)	5,964 (25.1%)	1,035 (15.3%)	1,445 (22.6%)	827 (17.0%)	162 (13.1%)
Intermediate cl.	4,770 (15.6%)	6,025 (25.3%)	1,757 (25.9%)	1,041 (16.3%)	949 (19.5%)	191 (15.5%)
Overall	30,615	23,790	6,785	6,403	4,858	1,235

https://doi.org/10.1371/journal.pone.0249241.t004

motivated to challenge both intermediate and majority pole users. A similar tendency can be found for the NYE incident.

To sum up, interaction patterns within and between the different groups are heterogeneous: Some users from every group seek debate with the differently-minded and others show a tendency to discuss amongst their own cluster. Nevertheless, the minority pole by far shows a stronger tendency to reply to users from different clusters than the majority pole.

Discussion

Vincent Price's comparison of public debate and town meetings receives a refinement in the conclusion of [10]. He states that:

> "The democratic foundations of the concept of public opinion are indisputable; far less so are the democratic foundations of day-to-day political decisions, even when they are formed out of public debate. [. . .] We may well compare public debate to a town meeting—provided we keep in mind that although some town meetings enjoy free-flowing debate, there are other meetings for which almost no one shows up, at which powerful leaders and organized coalitions dominate, and at which people with minority viewpoints are shouted down or left standing outside."

[10, p. 91]

The paragraph is quoted at length here since it illustrates aspects of public debate that might be amplified in online environments, specifically by social media. While the facilitation of communication bears the potential of enabling minorities and their concerns to gain public attention [23, 25], it can easily introduce systematic biases in the perception of public opinion. Online user comments are generally an important source of information for many and help in judging whether a product is good, a certain video is worth watching—or whether certain views should be taken into consideration in political decisions. While the experiences and opinions of others can provide a very useful basis for decision in these contexts, naive reliance on what others express online can be collectively dangerous, especially in an era in which social media shapes politics to an unprecedented extent. If groups with certain minority opinions manage to become increasingly visible, their opinions might appear more socially acceptable and accepted than they actually are. Such disparities can be problematic since perceived public opinion, as different studies have shown, can have a persuasive and/or silencing effect [9, 18, 20, 47, 48]. Under certain circumstances, committed minorities might even be able to gain public predominance, while the majority falls silent [49, 50]. In the context of Twitter, this phenomenon might be especially problematic due to its tight links with traditional media and news outlets, where Twitter content is often directly taken as representing public opinion [7]

or used as a source in routine coverage [5, 6, 8], introducing either strongly biased data or leading to a potentially unjustified alarmism in coverage.

This explorative study has two main findings:

- Disproportionately many replies come from users which constitute, if retweet networks are considered, a minority—composed of accounts by right to far-right parties, politicians, and users retweeting their content. It is hence probable that the content produced by this group influenced mere observers' perceptions of public opinion to a degree that did not reflect their real number.

- Users of different clusters also diverge in who they tend to reply to. While users from the majority cluster tend to interact mainly amongst themselves, users from the intermediate and especially from the minority cluster are more keen on confronting differently-minded others (Figs 4 and 5, Table 4).

It is hence shown that users retweeting mostly right-wing populist contents show a stronger willingness to express themselves in the form of replies in both case studies. First-order replies are nearly identical in number with those from users retweeting mostly left-leaning parties, politicians and contents. While users retweeting center to center-right parties show similar activity patterns as left-leaning ones, they are comparatively small in number, which might reflect that this voter group rarely uses Twitter. Public opinion in Noelle-Neumann's sense, assessed through reply sections, appears balanced between the two poles, while retweet networks suggest that the actual opinion proportions on the platform are very different. Noelle-Neumann also states that future expectations—i.e., that one opinion camp might gain ground in the future—might already increase the willingness of opinion expression now [18]. This is in line with the fact that right-wing populist movements have gained strength in recent years in Germany.

Previous findings parallel the results of this work: In a study from Switzerland, it has been shown that users with a right-wing political leaning engage more frequently in the comment sections of news pages [22], an effect also visible in Table 3 and in [23], where comment sections are interpreted as counter-public spaces. As an explanation for this effect in connection to the rise of right-wing populism, Schweiger [51] argues that misperceptions of the opinion climate, fuelled by news consumption through social media, lead to higher willingness of opinion expression for certain social groups—especially for those who already put less trust in established media sources, and those who lack awareness about the biases implicit in social media. The proposed method here could contribute to a substantiation of these claims and findings in subsequent studies.

With the proposed method, differences in willingness of opinion expression can be made visible. We stress that the method proposed here is not restricted to the specific cases studies. With a suitable seed set of users, any debate on Twitter can be analysed analogously. It is, nevertheless, limited in scope: It attempts to gain a comprehensive structural view on Twitter debate, but does not analyze the content of the tweets. Moreover, a proportion of users in the data sets remains unclassified by the employed method since they did not appear in the retweet networks. Complementary methods of classification should be sought.

Finally, Twitter is only one social media platform, which in addition is not representative of the general population [3]. A potentially insightful avenue for future inquiry might be the comparison of reply sections with comment sections of online newspapers [52]. Still, as we have argued, both Twitter's platform design as well as its echo in traditional media outlets at least implicitly award Twitter the role of the host of the big town meeting called public debate. We therefore deem it increasingly important to develop methods which enable a better

understanding of which viewpoints are prominently featured on the platform, and which ones remain mostly unspoken or unheard.

Supporting information

S1 Fig. Large retweet network. Figure shows the retweet network constructed out of all retweets in the data set between the 1st of July 2019 and then 24th of February 2020. If users were not present in the incident-specific retweet network, it was checked whether this large retweet network contained the users to increase the amount of classified users. Fig 5 exhibits a very similar shape as the incident-specific retweet networks. Out of 88,167 users, 71.8% belonged to the majority cluster, 18.4% to the minority cluster and 9.8% to the intermediate region.
(TIF)

S2 Fig. Time series of collection at the end of August 2019. Time series of the collected tweets. As is visible, some tweets were lost due to API problems between the 26th and the 31st of August.
(TIF)

S3 Fig. Workflow chart for analysis. Chart displaying the workflow necessary for the analysis carried out in the paper.
(TIF)

S1 File. Additional tables and explanations, showing that engagement differences cannot be caused by a potentially biased seed set. Two additional tables with engagement proportions from the large retweet network as well as a retweet network without retweeted replies.
(PDF)

Acknowledgments

Author Contributions

Conceptualization: Felix Gaisbauer.

Formal analysis: Felix Gaisbauer.

Methodology: Eckehard Olbrich.

Supervision: Sven Banisch, Eckehard Olbrich.

Validation: Felix Gaisbauer.

Visualization: Felix Gaisbauer, Armin Pournaki.

Writing – original draft: Felix Gaisbauer, Armin Pournaki, Sven Banisch, Eckehard Olbrich.

References

1. Jungherr A. Analyzing political communication with digital trace data. Cham, Switzerland: Springer. 2015;.

2. Bruns A, Burgess J. Crisis communication in natural disasters: The Queensland floods and Christchurch earthquakes. Twitter and society [Digital Formations, Volume 89]:. 2014; p. 373–384.

3. Mellon J, Prosser C. Twitter and Facebook are not representative of the general population: Political attitudes and demographics of British social media users. Research & Politics. 2017; 4(3):2053168017720008.

4. Mislove A, Lehmann S, Ahn YY, Onnela JP, Rosenquist JN. Understanding the demographics of Twitter users. In: Fifth international AAAI conference on weblogs and social media; 2011.

5. Broersma M, Graham T. Twitter as a news source: How Dutch and British newspapers used tweets in their news coverage, 2007–2011. Journalism practice. 2013; 7(4):446–464. https://doi.org/10.1080/17512786.2013.802481

6. Paulussen S, Harder RA. Social media references in newspapers: Facebook, Twitter and YouTube as sources in newspaper journalism. Journalism practice. 2014; 8(5):542–551. https://doi.org/10.1080/17512786.2014.894327

7. McGregor SC. Social media as public opinion: How journalists use social media to represent public opinion. Journalism. 2019; 20(8):1070–1086. https://doi.org/10.1177/1464884919845458

8. McGregor SC, Molyneux L. Twitter's influence on news judgment: An experiment among journalists. Journalism. 2020; 21(5):597–613. https://doi.org/10.1177/1464884918802975

9. Noelle-Neumann E. Public Opinion. Our Social Skin; 1984.

10. Price V. Public opinion. vol. 4; 1992.

11. Bruns A, Burgess J. Twitter hashtags from ad hoc to calculated publics. Hashtag publics: The power and politics of discursive networks. 2015; p. 13–28.

12. Trilling D. Two different debates? Investigating the relationship between a political debate on TV and simultaneous comments on Twitter. Social science computer review. 2015; 33(3):259–276. https://doi.org/10.1177/0894439314537886

13. Gearhart S, Kang S. Social media in television news: The effects of Twitter and Facebook comments on journalism. Electronic News. 2014; 8(4):243–259. https://doi.org/10.1177/1931243114567565

14. Schweiger W, Weihermüller M. Öffentliche Meinung als Online-Diskurs–ein neuer empirischer Zugang. Publizistik. 2008; 53(4):535–559.

15. Tumasjan A, Sprenger TO, Sandner PG, Welpe IM. Predicting elections with twitter: What 140 characters reveal about political sentiment. In: Fourth international AAAI conference on weblogs and social media; 2010.

16. Burnap P, Gibson R, Sloan L, Southern R, Williams M. 140 characters to victory?: Using Twitter to predict the UK 2015 General Election. Electoral Studies. 2016; 41:230–233. https://doi.org/10.1016/j.electstud.2015.11.017

17. Habermas J. Between facts and norms: Contributions to a discourse theory of law and democracy. John Wiley & Sons; 2015.

18. Noelle-Neumann E, Petersen T. The spiral of silence and the social nature of man. In: Handbook of political communication research. Routledge; 2004. p. 357–374.

19. Springer N, Kümpel AS. User-Generated (Dis) Content. In: Journalismus im Internet. Springer; 2018. p. 241–271.

20. Lee EJ, Jang YJ. What do others' reactions to news on Internet portal sites tell us? Effects of presentation format and readers' need for cognition on reality perception. Communication research. 2010; 37 (6):825–846. https://doi.org/10.1177/0093650210376189

21. Lee EJ. That's not the way it is: How user-generated comments on the news affect perceived media bias. Journal of Computer-Mediated Communication. 2012; 18(1):32–45. https://doi.org/10.1111/j.1083-6101.2012.01597.x

22. Friemel TN, Dötsch M. Online reader comments as indicator for perceived public opinion; 2015.

23. Toepfl F, Piwoni E. Public spheres in interaction: Comment sections of news websites as counterpublic spaces. Journal of Communication. 2015; 65(3):465–488. https://doi.org/10.1111/jcom.12156

24. Kaiser J. Public spheres of skepticism: Climate skeptics' online comments in the German networked public sphere. International Journal of Communication. 2017; 11.

25. Fraser N. Rethinking the public sphere: A contribution to the critique of actually existing democracy. Social text. 1990;(25/26):56–80. https://doi.org/10.2307/466240

26. Baker KM. In: Public opinion as political invention. Inventing the French Revolution: Essays on French Political Culture in the Eighteenth Century. Cambridge University Press; 1990. p. 167–200.

27. Kalogeropoulos A, Negredo S, Picone I, Nielsen RK. Who shares and comments on news?: A cross-national comparative analysis of online and social media participation. Social media+ society. 2017; 3(4).

28. Mustafaraj E, Finn S, Whitlock C, Metaxas PT. Vocal minority versus silent majority: Discovering the opionions of the long tail. In: 2011 IEEE Third International Conference on Privacy, Security, Risk and Trust and 2011 IEEE Third International Conference on Social Computing. IEEE; 2011. p. 103–110.

29. Ross L, Greene D, House P. The "false consensus effect": An egocentric bias in social perception and attribution processes. Journal of experimental social psychology. 1977; 13(3):279–301. https://doi.org/10.1016/0022-1031(77)90049-X

30. Mullen B, Dovidio JF, Johnson C, Copper C. In-group-out-group differences in social projection. Journal of Experimental Social Psychology. 1992; 28(5):422–440. https://doi.org/10.1016/0022-1031(92)90040-Q

31. Reissing C. Sachsens CDU-Kandidaten schließen Koalition mit AfD aus; 2019. https://www.mdr.de/nachrichten/politik/regional/cdu-schliesst-koalition-afd-aus-sachsen-100.html (last accessed: 20 June 2020).

32. Fiedler M, Jansen F. Was geschah an Silvester in Leipzig-Connewitz?; 2020. https://www.tagesspiegel.de/politik/angriff-auf-polizisten-wirft-fragen-auf-was-geschah-an-silvester-in-leipzig-connewitz/25386832.html (last accessed: 20 June 2020).

33. Conover MD, Goncalves 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; 2011. p. 192–199.

34. 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; 2011.

35. Gaumont N, Panahi M, Chavalarias D. Reconstruction of the socio-semantic dynamics of political activist Twitter networks—Method and application to the 2017 French presidential election. PLOS ONE. 2018; 13(9):1–38. https://doi.org/10.1371/journal.pone.0201879

36. 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). https://doi.org/10.1371/journal.pone.0098679 PMID: 24914678

37. Bastian M, Heymann S, Jacomy M. Gephi: an open source software for exploring and manipulating networks. In: Proceedings of the International AAAI Conference on Web and Social Media. vol. 3; 2009.

38. Noack A. Modularity clustering is force-directed layout. Physical Review E. 2009; 79(2):026102. https://doi.org/10.1103/PhysRevE.79.026102

39. Venturini T, Jacomy M, Jensen P. What do we see when we look at networks; 2019.

40. Sousa D, Sarmento L, Mendes Rodrigues E. Characterization of the twitter @ replies network: are user ties social or topical? In: Proceedings of the 2nd international workshop on Search and mining user-generated contents; 2010. p. 63–70.

41. Aragón P, Kappler KE, Kaltenbrunner A, Laniado D, Volkovich Y. Communication dynamics in twitter during political campaigns: The case of the 2011 Spanish national election. Policy & internet. 2013; 5(2):183–206. https://doi.org/10.1002/1944-2866.POI327

42. Yardi S, boyd d. Dynamic debates: An analysis of group polarization over time on twitter. Bulletin of science, technology & society. 2010; 30(5):316–327. https://doi.org/10.1177/0270467610380011

43. Nuernbergk C, Conrad J. Conversations and campaign dynamics in a hybrid media environment: Use of Twitter by members of the German Bundestag. Social Media+ Society. 2016; 2(1).

44. Newman ME. Mixing patterns in networks. Physical Review E. 2003; 67(2). https://doi.org/10.1103/PhysRevE.67.026126 PMID: 12636767

45. Peel L, Delvenne JC, Lambiotte R. Multiscale mixing patterns in networks. Proceedings of the National Academy of Sciences. 2018; 115(16):4057–4062. https://doi.org/10.1073/pnas.1713019115

46. Peel L. MultiscaleMixing; 2018. Available from: github.com/piratepeel/multiscalemixing.

47. Noelle-Neumann E. Die Schweigespirale. Öffentliche Meinung–Unsere soziale Haut. Riper [ie Piper]; 1980.

48. Neubaum G, Krämer NC. Monitoring the opinion of the crowd: Psychological mechanisms underlying public opinion perceptions on social media. Media Psychology. 2017; 20(3):502–531. https://doi.org/10.1080/15213269.2016.1211539

49. Gaisbauer F, Olbrich E, Banisch S. Dynamics of opinion expression. Physical Review E. 2020; 102(4):042303. https://doi.org/10.1103/PhysRevE.102.042303

50. Banisch S, Gaisbauer F, Olbrich E. How social feedback processing in the brain shapes collective opinion processes in the era of social media. arXiv:200308154. 2020;.

51. Schweiger W. Der (des)informierte Bürger im Netz. Springer; 2017.

52. Willaert T, Banisch S, Van Eecke P, Beuls K. Facilitating on-line opinion dynamics by mining expressions of causation. The case of climate change debates on The Guardian. arXiv:191201252. 2019;.