This is the accepted paper version. Full version: Papakyriakopoulos, Orestis, Morteza Shahrezaye, Juan Carlos Medina Serrano, and Simon Hegelich. "Distorting Political Communication: The Effect Of Hyperactive Users In Online Social Networks." In IEEE INFOCOM 2019-IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS), pp. 157-164. IEEE, 2019. DOI: 10.1109/INFCOMW.2019.8845094 https://ieeexplore.ieee.org/document/8845094

Distorting Political Communication: The Effect Of Hyperactive Users In Online Social Networks

Orestis Papakyriakopoulos, Morteza Shahrezaye, Juan Carlos Medina Serrano, Simon Hegelich

Bavarian School of Public Policy
Technical University of Munich
D-80333 Munich, Germany
{orestis.papakyriakopoulos,morteza.shahrezaye, juan.medina}@tum.de, simon.hegelich@hfp.tum.de

***Abstract*— Online Social Networks (OSNs) are used increasingly for political purposes. Among others, politicians externalize their views on issues, and users respond to them, initiating political discussions. Part of the discussions are shaped by hyperactive users. These are users that are over-proportionally active in relation to the mean. In this paper, we define the hyperactive user on the social media platform Facebook, both theoretically and mathematically. We apply a geometric topic modelling algorithm (GTM) on German political parties' posts and user comments to identify the topics discussed. We prove that hyperactive users have a significant role in the political discourse: They become opinion leaders, as well as set the content of discussions, thus creating an alternate picture of the public opinion. Given that, we discuss the dangers of replicating the specific bias by statistical and deep learning algorithms, which are used widely for recommendation systems and the profiling of OSN users.**

I. INTRODUCTION

Today, internet prevails as a prominent communication and information medium for citizens. Instead of watching TV or reading newspapers, increasing numbers of people get politically informed through online websites, blogs, and social media services. The latest statistics demonstrate that internet as a news source has become as important as television, with its share increasing year by year [1]. Given this shift in the means of news broadcasting, politicians have altered their tactics of communication to the society. OSNs, such as Twitter, Facebook and Instagram, have become a cornerstone of their public profiles as they use OSNs to transmit their activities and opinions on important social issues [2], [3], [4].

The growth of online communities on social media platforms have created a public amenable to political campaigning. Political parties and actors have adapted to the new digital environment [5], and besides the application of new campaigning methods as microtargeting [6], have created microblogs through which they can inform citizens of their views and activities. In addition, OSNs have enabled to users to respond to or comment on the politicians' messages, giving birth to a new type of political interaction and transforming the very nature of political communication.

On OSNs, the flow of information from politicians to citizens and back follows a different broadcasting model than the classical one [7]. Instead of journalists monitoring the political activity, political actors themselves produce messages and make them publicly available on the platforms. Each platform provides its users with access to the generated content, as well as distributes it to them through recommendation algorithms [8], [9]. The received information is then evaluated both directly or indirectly [10], [11]: The political message is interpreted immediately, or subsequently through further social interactions among citizens on the related topics. On OSNs, not only can users respond to politicians in the traditional manner -i.e. through their political activity in the society-, but also respond to or comment on the politicians' views online. This creates a new type of interactivity, as users, who actively engage in online discussions sharing their views, are able to influence the way the initial political information will be assimilated by passive users as well as directly influencing political actors.

This new form of political interactivity transforms political communication. Given the possibility of users to directly respond to the political content set by political actors, and discuss online about political issues with other citizens, OSNs emerge as a fruitful space for agonistic pluralism. They provide the necessary channels for different interests and opinions to be expressed, heard and counterposed; elements that constitute the very essence of political communication. If the discussions held are legitimized within a democratic framework, they form the basis for reaching a conflictual consensus [12], based on which societal decisions can be made. Hence, political communication on OSNs opens new possibilities for citizens to participate in the political shaping of the society, providing them with additional space to address their interests.

Problem statement

Although the above type of political communication has the potential to improve the function of democracy, OSNs possess a structural property that obstructs the unbiased constructive interaction between political actors and citizens: The activities of users on OSNs follow an extreme value distribution [13], [14], [15], [16]. Practically, this means that users are not equally active when using a specific OSN. Among others, the majority of users remain passive, or participate with a very low frequency; they either simply read the content or like, comment, tweet, etc. very rarely. On the contrary, a very small part of the users is hyperactive, as they over-proportionally interact with the platform they use. Thus, in political communication on OSNs, hyperactive users are citizens who over-proportionally externalize their political attitudes compared to the mean. This could be done by liking, commenting, tweeting or using any other interaction possibility provided by a platform to declare an attitude to a political issue.

The given activity asymmetry becomes a major issue, considering that a considerable part of the society is politically informed via OSNs. As hyperactive users externalize their political attitudes more than the others, they have the potential to distort political communication; political issues that are important to them become overrepresented on OSNs, while the views of normally active users become less visible. Hence, hyperactive users may influence the political discussions towards their ends, creating a deformed picture of the actual public opinion on OSNs. This fact violates the assumption of an equitable public political discourse as part of political communication [17], because the interests and views of normally active users appear as less important.

The above distortion of political communication is intensified by the business models of the OSN platforms. OSNs were not created to be arenas of political exchange. Their aim is to maximize the number of platform users, by keeping them satisfied [18], and to transform this social engagement to profits, i.a. through advertisement. Hence, on OSNs, users are both consumers and citizens [19]. In order to maximize their profits, OSN platforms adjust their recommendation algorithms to the content popularity, with a view to promoting information that most users will like. As hyperactive users influence asymmetrically the popularity of political content, these algorithms might replicate this asymmetry. Thus, a platform might recommend content, which is actually consistent with the political interests of hyperactive users. This phenomenon per se denotes a form of algorithmic manipulation of the political communication: The platform unintentionally magnifies hyperactive users' interests, thus posing the risk of political invisibility for the ones of normal users [20].

Last but not least, the aforementioned misrepresentation of public opinion has a direct impact on political campaigning. Contemporary political actors develop their influence strategies based on the perceived voter model [21], which presupposes the gathering of demographic and political data for the development of statistical models about the electorate's attitudes. As major part of these data is derived from social media, models that fail to take the effect of hyperactive users into account would face an important bias.

Considering the above, we want to answer following questions regarding the activity of hyperactive users:

RQ1: **How can we define hyperactive users mathematically?**

RQ2: **How can we compare and evaluate the political attitudes of hyperactive users in relation to the mean?**

Original Contribution

We mathematically define hyperactive users on OSN Facebook, and identify them on the public pages of the major German political parties. By applying a state-of-the-art topic modelling algorithm, we investigate whether they spread or like different messages on political issues other than normal users and politicians do. We prove that hyperactive users not only are responsible for a major part of online political discussions, but they also externalize different attitudes than the average user, changing the discourse taking place. We quantify their effect on content formation by measuring their popularity and showing that they adopt an opinion leader status. Finally, given the potential influence of hyperactive users on recommendation algorithms, we initiate an important discussion on OSNs as spaces of political communication.

II. Data & Method

A. Data Description

To investigate the effect of hyperactive users, we chose to analyse the public Facebook pages of the main German political parties. Our sample included CDU, CSU, SPD, FDP, Bündnis 90/Die Grünen, Die Linke, and AfD. CDU is the main conservative party of Germany, while CSU is the conservative party active in Bavaria. SPD represents the main German social-democratic party, and Die Linke the radical left. AfD has a nationalist, anti-immigrant, and neo-liberal agenda, while FDP is a conservative, neo-liberal party. Finally, Bündnis 90/Die Grünen is the German green party. We focused on Facebook, because the platform's api restrictions and its monitoring system largely prevent automated activities, as e.g. performed by social bots on other platforms [22], [23]. Therefore we could evaluate the natural behaviour of hyperactive users and not an algorithmically generated one.

We took into consideration all party posts and their reactions in the year 2016. This choice was made, because we wanted to investigate a full year of user activities. We preferred 2016 over 2017, because 2017 was an election year, with most content produced by the parties being campaign related. By contrast, 2016 was marked by the Refugee Crisis in Europe, and we were interested in evaluating the discussions on the topic. In total, by accessing the Facebook Graph API, we collected 3,261 Posts, 3,084,464 likes and 382,768 comments, made by 1,435,826 users. The sample included all posts and comments on the posts generated for the period under investigation.

B. Defining hyperactive users

We consider hyperactive users as people, whose behaviour deviates from the standard on an OSN platform. To obtain an understanding of the overall behaviour of the users, we fitted discrete power-law and extreme value distributions to mathematically describe the users' like and comment activities. Additionally, we ran bootstrapped and comparative goodness-of-fit tests based on the Kolmogorov-Smirnov [24] and the Vuong [25] statistic to evaluate the potential fits, as proposed by Clauset et al. [26]. The KS test examines the null hypothesis that the empirical sample is drawn from the reference distribution, while the Vuong test measures the log-likelihood ratio of two distributions and, based on it, investigates whether both empirical distributions are equally far from a third unidentified theoretical one.

In order to mathematically describe the activities of hyperactive users, we selected to treat them as outliers of the standard OSN population. We adopt the definitions made by Barnett and Lewis [27], Johnson and Wichern [28] and Ben-Gal [29], and see outliers not as errors, or coming from a different generative process, but as data containing important information, which is inconsistent with and deviating from the remainder of the data-set. Therefore, given the extreme skewed distribution of the activities, we followed the method proposed by Hubert and Vam der Veeken [30] and Hubert and Vandervieren [31] for outlier detection. We calculated the quartiles of our data Q_1 and Q_3, the interquartile range $IQR = Q_3 - Q_1$ and the whiskers w_1 and w_2, which extend from the Q_1 and Q_3 respectively to the following limits:

$$[Q_1 - 1.5e^{-4MC}IQR, Q_3 + 1.5e^{3MC}IQR] \quad (1)$$

where MC is the medcouple [32], a robust statistic of the distribution skewness. Data beyond the whiskers were marked as outliers.

C. Topic Modeling

After evaluating the likes and comments distributions, as well as identifying the existing hyperactive users, we prepared our data for the application of a topic modelling algorithm. As it has been shown that a noun-only topic modelling approach yields more coherent topic-bags [33], we cleaned our posts and comments from the remaining part-of-speech types. To do so, we applied the spaCy pretrained convolutional neural network (CNN) classifier [34] based on the Tiger [35] and WikiNER [36] corpuses, classified each word in our document collection, and kept only the nouns.

We wanted to investigate the various topics that users and parties discussed about but did not want to differentiate on the way they talked about them. Parties usually use a more formal language when posting on a topic than users. Therefore, there was the risk that the topic modelling algorithm would create different topics on the same issue, one for the parties and one for the users. To avoid this, we fitted our model on the user comments, and then classified the parties' posts through the trained model.

For our analysis, we applied a non-parametric Conic Scan-and-Cover (CoSAC) algorithm for geometric topic modeling [37]. Our decision was based on the fact that most topic modelling algorithms (e.g. LDA [38], NMF [39]) need a priori as input the number of topics to split the corpus. CoSAC has the advantage of electing itself the number of topics to find the most efficient topic estimates.

The algorithm presupposes that the optimal number of topics K are embedded in a V-1 dimensional probability simplex Δ^{V-1}, where V the number of words in the corpus. Each topic β_K corresponds to a set of probabilities in the word simplex. The totality of topics build hence a convex polytope $B = conv(\beta_1, ..., \beta_K)$. Each document corresponds to a point $p_m = (p_{m1}, ..., p_{mV})$ inside Polytope B, with $p_m = \sum_k \beta_k \theta_{mk}$. θ_{mk} denotes the proportion that topic k covers in document m. Finally each document is drawn from a multinomial distribution of words: $w_m \sim Multinomial(p_m, N_m)$, where N_m the number of words in document m.

The CoSAC algorithm iteratively scans the polytope B and finds the furthest point from its center Cp. It then constructs a conical region with angle ω, starting from Cp and embedding the specific point (Figure 1). All points within the cone are considered to belong in the same topic and are removed from the polytope. The procedure is iterated K-1 times, until almost no points remain in the polytope. A cone is considered sufficient if it covers at least a proportion of documents λ. After fitting the cones, CoSAC places a sphere with radius R to the polytope, to cover the remaining points. The K geometric objects and their respective points correspond to the K topics created by the algorithm. In our model, the hyperparameters were set to $\omega = 0.6, \lambda = 0.001$ and $R = 0.05$, as proposed by the authors.

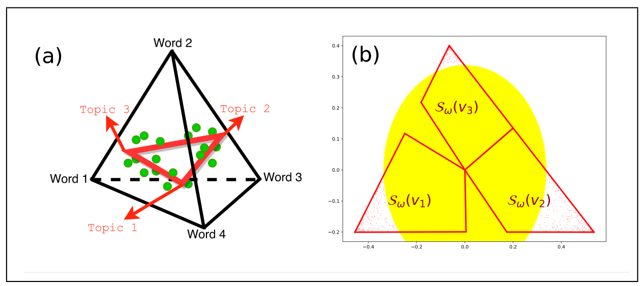

Fig. 1. (a) The topic polytope embedded in the word simplex. (b) Cones and sphere coverage of the polytope.

D. Comparison of activities

Given our topics, we wanted to evaluate the differences in the activity of normal and hyperactive users. Therefore, we calculated the empirical distributions $f(comment|topic)$ over all topics for the comments of normal and hyperactive users respectively. We pairwise compared the distributions for each topic, by applying a 2-Sample Anderson-Darling Test [40]. The test calculates the probability that the populations from which two groups of data were drawn are identical.

Besides testing the empirical comment-topic distributions, we assigned to each comment the topic with the highest probability and compared the most commented topics for normal and hyperactive users. Similarly, we assigned the classified party posts to their most probable topic and aggregated the likes of normal and hyperactive users. In this way, we were in the position to locate the concrete political interests of users.

III. Results

The results are split into three parts. First, we present our findings on the general user distribution on the investigated

TABLE I
VUONG TEST RESULTS

Log-normal vs	Likes LL-ratio (p-value)	Comments LL-ratio (p-value)
Power-law	15.1 (<0.01)	34.9 (<0.01)
Poisson	34.9 (<0.01)	12.7 (<0.01)
Exponential	12.7 (<0.01)	26.6 (<0.01)

pages. Based on that, we analyze the number and distribution of hyperactive users among the different pages. Then, we compare the behaviour between hyperactive and normal users by taking into consideration the topic modelling results and further statistical tests. Given that, we evaluate the importance and role of hyperactive users in the political discourse on OSNs.

A. *Describing user activity*

As a first result, we identified the log-normal distribution as the the best measure for describing the user activities (Figure 2). The bootstrapped KS-Tests (100 samples, two tailed) for both comments and likes failed to reject the null that our data come from a log-normal distribution (gof $<$ 0.01, p $>$ 0.05 and gof $<$ 0.01, p $>$ 0.2 respectively), while the comparative Vuong tests showed a better fit of the log-normal in comparison to the power-law, poisson and exponential distributions (Table I). Our results comply with the existing literature, which states that usually complex social network properties are log-normally distributed [15], [41], [42]. Figure 2 shows the empirical frequencies of user activities and their respective log-normal fits.

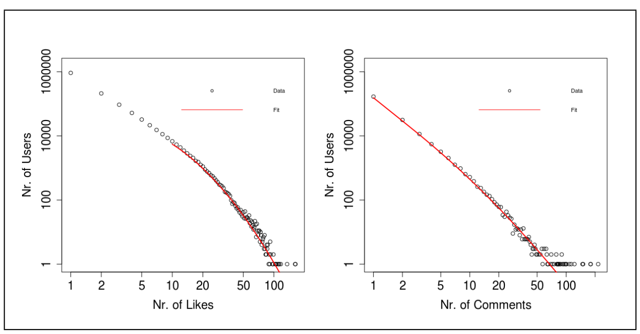

Fig. 2. (a) The topic polytope embedded in the word simplex. (b) Cones and sphere coverage of the polytope.

B. *Detecting hyperactive users*

Through our outlier detection methodology, we detected 12,295 hyperactive users on the comment section of pages, who correspond to 5.3% of the total users commenting on the pages. Due to the extreme skewness of the comments' distribution, a user was characterized as hyperactive if they made three or more comments. This is justified by the fact that actually 74% of the users under investigation made only one comment. Although hyperactive users represented 5.3% of the total commenting population, they accounted for 25.8% of the total comments generated on the parties' pages. Furthermore, 56% of these hyperactive users commented on two or more party pages, denoting that they generally interacted with users across the political spectrum. By evaluating the popularity of the users' comments, it was found that hyperactive users tend to get more support than the rest. Comments made by hyperactive users on average gained 3.52 likes, while normal users' comments only 3.07, a difference that was statistically significant (Mann-Whitney Test with continuity correction, one tailed: W = 1.4^{10}, p$<$ 0.01). This complies with previous research stating that highly active users tend to have the characteristics of opinion leaders [43].

TABLE II
HYPERACTIVE USERS PER PARTY - COMMENTS

Party	Comments by Hyperactive Users	Ratio
AfD	43,017	0.24
CDU	20,929	0.45
CSU	18,312	0.22
FDP	1,400	0.15
Die Grünen	8,946	0.36
Die Linke	2,343	0.16
SPD	3,926	0.13

Similarly, the evaluation of the pages' likes resulted in the characterization of 61,372 users as hyperactive, or 4.3% of the total users that liked the parties' posts. As before, the methodology labelled users as hyperactive if they made three or more likes, because the majority of the active Facebook population rarely interacted with the related pages. The likes of these hyperactive users accounted for 26.4% of total likes, hence having a major effect on the overall content liked. In addition, 29% of hyperactive users liked posts of more than one party, denoting again that their activities were spread over the entire parties' network. The overview of the hyperactive users' commenting and like activities for each party can be found in tables II and III. We also compared the like and comment distributions, by calculating their gini index. The measure provides a proxy for inequality, with 0 denoting perfect equality and 1 extreme inequality. In our case, we calculated a value of 0.35 and 0.45 for the comment and like distribution respectively. This denotes that like activities are more unequally distributed than the comment activities, i.e. hyperactive users play a bigger role in the formation of likes. In addition, the values denote a degree of inequality between normal and hyperactive users, but not an extreme one. Nevertheless this is misleading, because the measure does not take into consideration the inactive users. Given that information, the gini index would have been much higher in both cases.

C. *Evaluating the political attitudes of hyperactive users*

Based on the categorization of users as hyperactive or normal, we could then evaluate the results of the topic modelling algorithm. The model clustered the users' comments in 69 main topics. A major part of the topics concerned the refugee crisis of 2016 and the related discussions about Islam. A set of topics aggregated comments on German Chancellor Merkel, on the leaders of other parties, on female and male politicians and the German parties in general. There was one

TABLE III

HYPERACTIVE USERS PER PARTY - LIKES

Party	Likes by Hyperactive Users	Ratio
AfD	555,564	0.35
CDU	16,997	0.2
CSU	139,493	0.2
FDP	20,188	0.16
Die Grünen	28,777	0.19
Die Linke	24,546	0.14
SPD	29,057	0.12

topic summing up comments in English language, as well as a topic containing hyperlinks. Furthermore, the algorithm created policy related topics regarding foreign affairs, as well as the economy and labour market and the state in general. Other topics were related to the German national identity, society and solidarity, and the nature of democracy. Users also discussed about family and gender policy, homeland security, transportation and environmental policy. There were topics that included wishes, fear, ironic and aggressive speech, as well as topics aggregating user thoughts. Finally, a set of topics was about political events and communications and a number of topics included comments against mainstream media and the political system. An overview can be found in table IV. The geometric topic modelling algorithm was able to provide a broad picture of the discussion topics on the parties' pages, revealing numerous insights about the way Facebook users commented on the parties' posts. By splitting the comments into two categories, one for the ones generated by hyperactive users and one for the comments of normal users, and by assigning them to the topics to which they were mostly related, we created a stacked chart illustrating the share of hyperactive users' comments for every topic (Figure 3). It is evident that hyperactive users covered a major part of the comments, and although more active, they commented more or less similarly to the normal users among the various topics. Despite that, the Anderson-Darling tests rejected the null hypothesis that hyperactive and normal users' comments come from the same distribution for 54 out of the 69 topics. Practically, this means that the topic density distributions varied between the comments of normal and hyperactive users. This is caused when the comments contain different words in different proportions. Hence, hyperactive and normal users used different vocabularies when referring to a topic and, consequently, externalized overall different views and sentiment, or focused on different issues in each case.

Besides the fact that hyperactive users had a different behaviour on the posts' comments, our analysis showed that they also had different liking preferences. After classifying each party post to the most relevant topic, we counted the likes of the posts that belong to each topic. We took into consideration only topics that were based on either political vocabulary or politicians, and ignored topics that contained aggressive speech or sentiment, because the related vocabulary was rarely used by the parties. Figure 4 illustrates

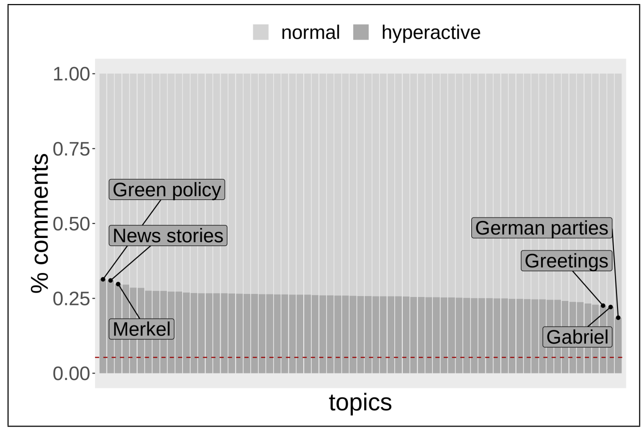

Fig. 3. Proportion of comments generated by normal and hyperactive users. The dotted red line gives the actual proportion of hyperactive users. The plot also illustrates the three most and least interesting topics for hyperactive users.

a stacked chart depicting the share of hyperactive users' likes. In contrast to the comments' chart, it is obvious that hyperactive users liked specific topics with different intensity than normal users. Even though hyperactive users performed on average 26.4% of the likes, they liked much more content related to EU politics and labour policy, while had less interest on the conservative party AfD, citizens' rights and the region of Bavaria. Therefore, it is clear that hyperactive users influence the like distribution of the public to party posts.

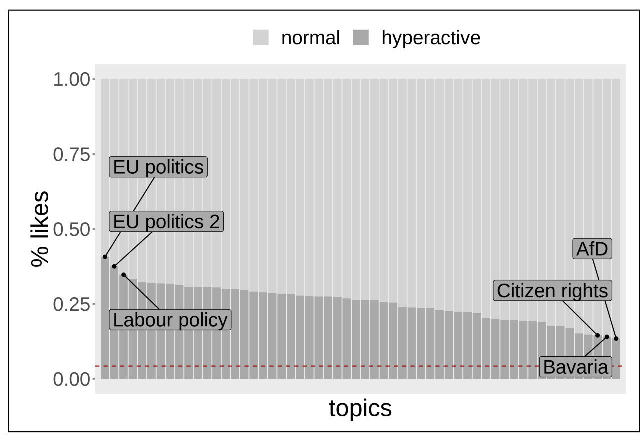

Fig. 4. Proportion of likes generated by normal and hyperactive users. The dotted red line gives the actual proportion of hyperactive users. The plot also illustrates the three most and least interesting topics for hyperactive users.

It must be noted that our analysis gives an overview of the content of posts. It cannot identify sentiment, or specific predispositions of users. For a firm understanding of the issues that were over- or under-represented by hyperactive users an additional extensive analysis is needed, which is beyond the scope of this paper. Our analysis demonstrated that, both on commenting and liking, hyperactive users have a different behaviour than the other users.

Taking the above into consideration, it was possible to show that political communication on Facebook is strongly constituted by the behaviour of hyperactive users. By describing the user like and comment activities on the platform, we managed to characterize users as hyperactive or normal through outlier detection. We proved that hyperactive users account for a significant part of the total users' activities, they participate in discussions differently from the rest, and they like different content. Moreover, they become opinion leaders, as their comments become more popular than these of the normal users. Taking Facebook as an example, we showed that user activities on OSNs are neither equally nor evenly distributed.

IV. Discussion

Given that activity asymmetries are a feature of online social networks, it is important to evaluate the consequences for science and the society. Although our analysis was concentrated on Facebook, previous research has proven that hyperactive accounts, either human or automated, have the potential to equally influence political communication in other platforms [44], [45]. The specific formation and distribution of political interactions on OSNs rises various questions regarding the role and impact of OSNs on a political level, on an algorithmic level, as well as on the intersection of both.

In the political dimension, the OSN activity asymmetries are transformed into an asymmetry of disseminated political content, as the attitudes and interests of hyperactive users appear over-proportionally in the discussions taking place. Until now, research [46], [47] has stated that OSNs suffer from a population bias: The people using OSNs are not representative of the actual society. On top of that, a content bias is now added: The content disseminated on OSNs is not even representative of the mean users' attitudes on the platform. This poses a scientific problem, as it might lead to false research results. Equally importantly, it poses a political problem, because political discussions and opinion exchange is distorted by the effect of hyperactive accounts. This happens not because the diffused information in the network is transformed or changed, but because hyperactive users strongly contribute to the type of information diffused. Their attitudes fill the communication space, leading to a bias on the political feedback to politicians, and to a shift on the issues that shape the political agenda. Although OSNs provide a more open environment to express opinions than traditional media, it ends up being partly a gathering of political echoes [48] that struggle to be imposed on each other.

In the algorithmic dimension, the extreme skewness of the activity distributions raises specific issues regarding the recommendation algorithms used by OSN platforms. The first problem is related to algorithmic accuracy: skewed data are, imbalanced data, and their raw use, either as input features or as output labels, can yield weak classification results. The imbalanced learning problem applies to both standard statistical algorithms, collaborative filtering and neural networks [49], [50], [51], with algorithms overestimating the importance of outliers and under-estimating the importance of the rest. This also happens in the case of a poor selection of a cost function [52]. Furthermore, it is proven that statistical models as Markov-chains might fail to capture the signal immanent in highly skewed data, while deep learning methods might face the same issue given power-law distributions of data [53].

The second potential problem is that an algorithm might fail, not in the sense that it might be unable to learn from the data, but rather learn the wrong signal. Hyperactive users can be seen as physical adversaries [54] of the mean user attitudes. Algorithms trained in the full data will include the bias, tracking and predicting behavioural associations that correspond to hyperactive users rather than to the population majority. It is not coincidental that the detection of adversaries in machine learning can be done by sample distribution comparison [55], in the same way as we tracked the different preferences of hyperactive users.

Solutions to the aforementioned issues exist and are usually taken into consideration by data scientists, who develop recommendation algorithms. Nevertheless, in the case of political communication, an algorithmic issue automatically becomes a political one. Recommendation systems come with a social influence bias [56], [57], i.e. have the power to change users' opinion. Hence, OSNs promoting biased political content might result in the algorithmic manipulation of political communication.

In addition, social media platforms are not designed to foster political discourses [58], but rather aim at increasing active users, in order to sell advertisement and attract funding from venture capitalists [59]. Hence, the structure and impact of recommendation algorithms distorts human behaviour [60], having transformative effects that were not foreseen a priori [61].

It is evident from the above, that each algorithm mediates and redefines the importance of political interests [62], raising further questions about the opacity of the recommendation systems [63]. In a political context, it becomes important to know as citizens, how, why and with what impact algorithms change political communication. This presupposes awareness of the data processed and, the mathematical method applied, as well as knowledge of what exactly a machine learning cost function optimizes and to what extent recommendation systems alter human behaviour. Proposals for algorithmic transparency have already been made [64], [65], [66], and wait to be applied in practice.

The above issues need to be extensively analyzed, in order to evaluate and shape the structure of political communication in the digital era. In this paper we laid the foundations for this discussion, by defining, demonstrating and quantifying the effect of hyperactive users on OSNs, through the example of Facebook. We also illustrated and defined the risks of algorithmic manipulation by the OSN recommendation systems. Future research needs to focus on the aforementioned consequences, evaluate the structure of OSNs ethically, politically and normatively as political intermediators, as well as propose and apply solutions to the newly posed problems.

TABLE IV

TOPIC MODELING, AD-TEST RESULTS AND PROPORTION OF HYPERACTIVE USERS

Nr.	Topic	AD-test gof, (p-value)	Comments	Likes
1	Immigration	3.8, (0.0)	0.27	0.30
2	Merkel	104.2, (1.0)	0.28	0.24
3	AfD	15.9, (0.0)	0.25	0.30
4	News stories	17.4, (0.0)	0.31	0.29
5	English	8.8, (0.0)	0.26	-
6	Green policy	15.1, (0.0)	0.31	0.18
7	Islam	4.8, (0.0)	0.26	0.31
8	Integration immigrants	6.7, (0.0)	0.27	0.28
9	Female politicians	9.5, (0.0)	0.26	0.22
10	Deportion immigrants	9.2, (0.0)	0.26	0.20
11	EU politics	2.5, (0.0)	0.26	0.41
12	Economic policy	6.1, (0.0)	0.28	0.31
13	Greetings	17.7, (0.0)	0.23	-
14	Polls	16.3, (0.0)	0.25	0.26
15	Union	71.2, (1.0)	0.29	0.26
16	CSU	69.2, (1.0)	0.24	0.24
17	National identity	11.5, (0.1)	0.26	0.29
18	Human rights	1.5, (0.1)	0.26	0.24
19	Security	2.6, (0.0)	0.27	0.24
20	Democracy	32.3, (0.0)	0.25	0.27
21	Citizen rights	33.9, (0.0)	0.25	0.15
22	Congratulations	26.5, (0.0)	0.24	0.26
23	Gabriel	43.2, (1.0)	0.22	0.23
24	Foreign affairs	5.0, (0.0)	0.26	0.26
25	Homeland security	17.3, (0.0)	0.25	0.25
26	Interviews	23.9, (0.0)	0.25	0.18
27	Turkey affairs	11.0, (0.0)	0.26	0.19
28	Terrorism	7.1, (0.0)	0.26	0.19
29	Fear	1.6, (0.1)	0.26	-
30	Party system	4.3, (0.0)	0.27	0.29
31	The people	3.2, (0.0)	0.27	0.27
32	News media	1.3, (0.1)	0.27	0.31
33	Erdogan	7.1, (0.0)	0.27	0.23
34	German parties	25.4, (0.0)	0.19	0.19
35	Social policy	10.9, (0.0)	0.26	0.27
36	Reflection	14.5, (0.0)	0.26	-
37	TTIP/CETA	15.7, (0.0)	0.25	0.28
38	Syria	2.4, (0.0)	0.25	0.17
39	Labour policy	20.9, (0.0)	0.24	0.30
40	Party policies	0.2, (0.3)	0.26	0.27
41	Media	32.1, (0.0)	0.25	-
42	DDR	12.9, (0.0)	0.26	0.33
43	Male politicians	2.5, (0.0)	0.25	0.28
44	East Germany	5.0, (0.0)	0.26	0.32
45	Speeches	53.6, (1.0)	0.25	-
46	Bavaria	67.1, (1.0)	0.25	0.14
47	State media	21.4, (0.0)	0.25	-
48	Female politicians 2	12.0, (0.0)	0.30	0.20
49	Bundestag	10.4, (0.0)	0.25	0.32
50	Interviews 2	16.9, (0.0)	0.25	0.28
51	Irony	42.4, (1.0)	0.26	-
52	Trump	16.2, (0.0)	0.26	0.22
53	Welfare policy	12.3, (0.0)	0.26	0.32
54	Videos	13.0, (1.0)	0.25	-
55	Government	26.1, (0.0)	0.26	0.31
56	Transportation policy	37.0, (0.0)	0.23	0.15
57	Green policy 2	3.7, (0.0)	0.27	0.20
58	Politicians	12.1, (0.0)	0.23	-
59	Public services	18.4, (0.0)	0.25	0.20
60	Gender Equality	19.7, (0.0)	0.26	0.31
61	Insults	30.5, (0.0)	0.25	-
62	Boarder security	3.4, (0.0)	0.27	0.32
63	Media 2	13.5, (0.0)	0.27	-
64	EU politics 2	2.3, (0.0)	0.25	0.38
65	Merkel 2	39.9, (0.1)	0.30	0.15
66	AfD 2	2.6, (0.0)	0.26	0.13
67	Funny	23.9, (0.0)	0.25	-
68	Germans	0.5, (0.2)	0.27	0.22
69	Labour policy 2	8.5, (0.0)	0.27	0.35

REFERENCES

[1] J. Gottfried and E. Shearer, "Americans' online news use is closing in on tv news use," Sep 2017. [Online]. Available: http://www.pewresearch.org/fact-tank/2017/09/07/americans-online-news-use-vs-tv-news-use/

[2] S. Hegelich and M. Shahrezaye, "The communication behavior of german mps on twitter: preaching to the converted and attacking opponents," *European Policy Analysis*, vol. 1, no. 2, pp. 155–174, 2015.

[3] G. S. Enli and E. Skogerbø, "Personalized campaigns in party-centred politics: Twitter and facebook as arenas for political communication," *Information, Communication & Society*, vol. 16, no. 5, pp. 757–774, 2013.

[4] V. Arnaboldi, A. Passarella, M. Conti, and R. Dunbar, "Structure of ego-alter relationships of politicians in twitter," *Journal of Computer-Mediated Communication*, vol. 22, no. 5, pp. 231–247, 2017. [Online]. Available: http://dx.doi.org/10.1111/jcc4.12193

[5] J. C. M. Serrano, S. Hegelich, M. Shahrezaye, and O. Papakyriakopoulos, *Social Media Report: The 2017 German Federal Elections*, 1st ed. TUM.University Press, 2018.

[6] O. Papakyriakopoulos, S. Hegelich, M. Shahrezaye, and J. C. M. Serrano, "Social media and microtargeting: Political data processing and the consequences for germany," *Big Data & Society*, vol. 5, no. 2, p. 2053951718811844, 2018. [Online]. Available: https://doi.org/10.1177/2053951718811844

[7] M. E. McCombs and D. L. Shaw, "The agenda-setting function of mass media," *Public opinion quarterly*, vol. 36, no. 2, pp. 176–187, 1972.

[8] E. Bakshy, S. Messing, and L. A. Adamic, "Exposure to ideologically diverse news and opinion on facebook," *Science*, vol. 348, no. 6239, pp. 1130–1132, 2015.

[9] Twitter, https://help.twitter.com/en/using-twitter/twitter-trending-faqs, 2018, online; accessed 24 August 2018.

[10] M. Hilbert, J. Vásquez, D. Halpern, S. Valenzuela, and E. Arriagada, "One step, two step, network step? complementary perspectives on communication flows in twittered citizen protests," *Social Science Computer Review*, vol. 35, no. 4, pp. 444–461, 2017.

[11] S. Choi, "The two-step flow of communication in twitter-based public forums," *Social Science Computer Review*, vol. 33, no. 6, pp. 696–711, 2015.

[12] C. Mouffe, *The democratic paradox*. Verso, 2000.

[13] N. Blenn and P. Van Mieghem, "Are human interactivity times lognormal?" *arXiv preprint*, 2016.

[14] P. Van Mieghem, N. Blenn, and C. Doerr, "Lognormal distribution in the digg online social network," *The European Physical Journal B*, vol. 83, no. 2, p. 251, 2011.

[15] K. Lerman and R. Ghosh, "Information contagion: An empirical study of the spread of news on digg and twitter social networks," in *Proceedings of the fourth International AAAI Conference on Web and Social Media*. AAAI, 2010, pp. 90–97.

[16] F. Benevenuto, T. Rodrigues, M. Cha, and V. Almeida, "Characterizing user behavior in online social networks," in *Proceedings of the 9th ACM SIGCOMM Conference on Internet Measurement*. ACM, 2009, pp. 49–62.

[17] A. Schaap, "Agonism in divided societies," *Philosophy & Social Criticism*, vol. 32, no. 2, pp. 255–277, 2006.

[18] N. Shi, M. K. Lee, C. M. Cheung, and H. Chen, "The continuance of online social networks: how to keep people using facebook?" in *Fourty-third Hawaii international conference on System sciences*. IEEE, 2010, pp. 1–10.

[19] C. R. Sunstein, *# Republic: Divided democracy in the age of social media*. Princeton University Press, 2018.

[20] T. Bucher, "Want to be on the top? algorithmic power and the threat of invisibility on facebook," *new media & society*, vol. 14, no. 7, pp. 1164–1180, 2012.

[21] E. D. Hersh, *Hacking the electorate: How campaigns perceive voters*. Cambridge University Press, 2015.

[22] Facebook, "Using the graph api - documentation." [Online]. Available: https://developers.facebook.com/docs/graph-api/using-graph-api/

[23] O. Varol, E. Ferrara, C. A. Davis, F. Menczer, and A. Flammini, "Online human-bot interactions: Detection, estimation, and characterization," in *Eleventh international AAAI conference on web and social media*, 2017.

[24] T. B. Arnold and J. W. Emerson, "Nonparametric goodness-of-fit tests for discrete null distributions." *R Journal*, vol. 3, no. 2, 2011.

[25] Q. H. Vuong, "Likelihood ratio tests for model selection and non-nested hypotheses," *Econometrica: Journal of the Econometric Society*, pp. 307–333, 1989.

[26] A. Clauset, C. R. Shalizi, and M. E. Newman, "Power-law distributions in empirical data," *SIAM review*, vol. 51, no. 4, pp. 661–703, 2009.

[27] V. Barnett and T. Lewis, *Outliers in statistical data*. Wiley, 1974.

[28] R. Johnson and D. Wichern, *Applied multivariate Statistical Analysis.* Pretice Hall, 1998.

[29] I. Ben-Gal, "Outlier detection," in *Data mining and knowledge discovery handbook*. Springer, 2005, pp. 131–146.

[30] M. Hubert and S. Van der Veeken, "Outlier detection for skewed data," *Journal of Chemometrics: A Journal of the Chemometrics Society*, vol. 22, no. 3-4, pp. 235–246, 2008.

[31] M. Hubert and E. Vandervieren, "An adjusted boxplot for skewed distributions," *Computational statistics & data analysis*, vol. 52, no. 12, pp. 5186–5201, 2008.

[32] G. Brys, M. Hubert, and A. Struyf, "A robust measure of skewness," *Journal of Computational and Graphical Statistics*, vol. 13, no. 4, pp. 996–1017, 2004.

[33] F. Martin and M. Johnson, "More efficient topic modelling through a noun only approach," in *Proceedings of the Australasian Language Technology Association Workshop 2015*, 2015, pp. 111–115.

[34] M. Honnibal and M. Johnson, "An improved non-monotonic transition system for dependency parsing," in *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, 2015, pp. 1373–1378.

[35] S. Brants, S. Dipper, P. Eisenberg, S. Hansen-Schirra, E. König, W. Lezius, C. Rohrer, G. Smith, and H. Uszkoreit, "Tiger: Linguistic interpretation of a german corpus," *Research on language and computation*, vol. 2, no. 4, pp. 597–620, 2004.

[36] J. Nothman, N. Ringland, W. Radford, T. Murphy, and J. R. Curran, "Learning multilingual named entity recognition from wikipedia," *Artificial Intelligence*, vol. 194, pp. 151–175, 2013.

[37] M. Yurochkin, A. Guha, and X. Nguyen, "Conic scan-and-cover algorithms for nonparametric topic modeling," in *Advances in Neural Information Processing Systems*, 2017, pp. 3878–3887.

[38] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent dirichlet allocation," *Journal of machine Learning research*, vol. 3, pp. 993–1022, 2003.

[39] D. D. Lee and H. S. Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature*, vol. 401, no. 6755, p. 788, 1999.

[40] F. W. Scholz and M. A. Stephens, "K-sample anderson–darling tests," *Journal of the American Statistical Association*, vol. 82, no. 399, pp. 918–924, 1987.

[41] K. Sun, "Explanation of log-normal distributions and power-law distributions in biology and social science," Department of Physics, University of Illinois at Urbana-Champaign, Tech. Rep., 2004.

[42] H. Kuninaka and M. Matsushita, "Statistical properties of complex systems-lognormal and related distributions," in *AIP Conference Proceedings*, vol. 1468. AIP, 2012, pp. 241–251.

[43] B. E. Weeks, A. Ardèvol-Abreu, and H. Gil de Zúñiga, "Online influence? social media use, opinion leadership, and political persuasion," *International Journal of Public Opinion Research*, vol. 29, no. 2, pp. 214–239, 2017.

[44] A. Thieltges, O. Papakyriakopoulos, J. C. M. Serrano, and S. Hegelich, "Effects of social bots in the iran-debate on twitter," *arXiv preprint*, 2018.

[45] C. Shao, G. L. Ciampaglia, O. Varol, K.-C. Yang, A. Flammini, and F. Menczer, "The spread of low-credibility content by social bots," *Nature communications*, vol. 9, no. 1, p. 4787, 2018.

[46] T. Correa, A. W. Hinsley, and H. G. De Zuniga, "Who interacts on the web?: The intersection of users personality and social media use," *Computers in Human Behavior*, vol. 26, no. 2, pp. 247–253, 2010.

[47] M. Duggan and J. Brenner, "The demographics of social media users - 2012," Pew Research Center's Internet & American Life Project Washington, DC, Tech. Rep., 2013.

[48] Y.-R. Lin, J. P. Bagrow, and D. Lazer, "More voices than ever? quantifying media bias in networks," in *Proceedings of the fifth International AAAI Conference on Web and Social Media*. AAAI, 2011.

[49] H. He and E. A. Garcia, "Learning from imbalanced data," *IEEE Transactions on Knowledge & Data Engineering*, vol. 9, pp. 1263–1284, 2008.

[50] Z.-H. Zhou and X.-Y. Liu, "Training cost-sensitive neural networks with methods addressing the class imbalance problem," *IEEE Transactions on Knowledge and Data Engineering*, vol. 18, no. 1, pp. 63–77, 2006.

[51] R. Pan, Y. Zhou, B. Cao, N. N. Liu, R. Lukose, M. Scholz, and Q. Yang, "One-class collaborative filtering," in *IEEE eighth International Conference on Data Mining*. IEEE, 2008, pp. 502–511.

[52] C.-C. Lee, P.-C. Chung, J.-R. Tsai, and C.-I. Chang, "Robust radial basis function neural networks," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 29, no. 6, pp. 674–685, 1999.

[53] H. W. Lin and M. Tegmark, "Criticality in formal languages and statistical physics," *arXiv preprint*, 2016.

[54] A. Kurakin, I. Goodfellow, and S. Bengio, "Adversarial examples in the physical world," *arXiv preprint*, 2016.

[55] K. Grosse, P. Manoharan, N. Papernot, M. Backes, and P. McDaniel, "On the (statistical) detection of adversarial examples," *arXiv preprint*, 2017.

[56] S. Krishnan, J. Patel, M. J. Franklin, and K. Goldberg, "A methodology for learning, analyzing, and mitigating social influence bias in recommender systems," in *Proceedings of the 8th ACM Conference on Recommender systems*. ACM, 2014, pp. 137–144.

[57] D. Cosley, S. K. Lam, I. Albert, J. A. Konstan, and J. Riedl, "Is seeing believing?: how recommender system interfaces affect users' opinions," in *Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM, 2003, pp. 585–592.

[58] S. Engelmann, J. Grossklags, and O. Papakyriakopoulos, "A democracy called facebook? participation as a privacy strategy on social media," in *Annual Privacy Forum*. Springer, 2018, pp. 91–108.

[59] M. Falch, A. Henten, R. Tadayoni, and I. Windekilde, "Business models in social networking," in *CMI International Conference on Social Networking and Communities*, 2009.

[60] D. Ruths and J. Pfeffer, "Social media for large studies of behavior," *Science*, vol. 346, no. 6213, pp. 1063–1064, 2014.

[61] B. D. Mittelstadt, P. Allo, M. Taddeo, S. Wachter, and L. Floridi, "The ethics of algorithms: Mapping the debate," *Big Data & Society*, vol. 3, no. 2, 2016.

[62] H. Nissenbaum, "From preemption to circumvention: if technology regulates, why do we need regulation (and vice versa)," *Berkeley Technology Law Journal*, vol. 26, p. 1367, 2011.

[63] J. Burrell, "How the machine thinks: Understanding opacity in machine learning algorithms," *Big Data & Society*, vol. 3, no. 1, 2016.

[64] C. Sandvig, K. Hamilton, K. Karahalios, and C. Langbort, "Auditing algorithms: Research methods for detecting discrimination on internet platforms," in *Data and discrimination: converting critical concerns into productive inquiry*, 2014, pp. 1–23.

[65] N. Diakopoulos, "Algorithmic-accountability: the investigation of black boxes," Tow Center for Digital Journalism, Tech. Rep., 2014.

[66] F. Pasquale, *The black box society: The secret algorithms that control money and information*. Harvard University Press, 2015.