

Social media and microtargeting: Political data processing and the consequences for Germany

Big Data & Society
July–December 2018: 1–15
© The Author(s) 2018
DOI: 10.1177/2053951718811844
journals.sagepub.com/home/bds



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

Abstract

Amongst other methods, political campaigns employ microtargeting, a specific technique used to address the individual voter. In the US, microtargeting relies on a broad set of collected data about the individual. However, due to the unavailability of comparable data in Germany, the practice of microtargeting is far more challenging. Citizens in Germany widely treat social media platforms as a means for political debate. The digital traces they leave through their interactions provide a rich information pool, which can create the necessary conditions for political microtargeting following appropriate algorithmic processing. More specifically, data mining techniques enable information gathering about a people's general opinion, party preferences and other non-political characteristics. Through the application of data-intensive algorithms, it is possible to cluster users in respect of common attributes, and through profiling identify whom and how to influence. Applying machine learning algorithms, this paper explores the possibility to identify micro groups of users, which can potentially be targeted with special campaign messages, and how this approach can be expanded to large parts of the electorate. Lastly, based on these technical capabilities, we discuss the ethical and political implications for the German political system.

Keywords

Microtargeting, social media, Germany, influence, datafication, electorate

Introduction

The contemporary digital revolution is constantly transforming the political world. Datafication (Mayer-Schönberger and Cukier, 2013), i.e. the categorization, quantification and aggregation of phenomena into databases, and their further algorithmic processing, have opened new opportunities in understanding and evaluating complex social phenomena. More specifically the use of social media and the internet has resulted in the creation of enormous databases that contain information about citizens' personal and political preferences. Based on these *Big Political Data* a new type of data-driven interaction between politics and citizens emerges through social media. In its core lies the application of advanced statistical and machine learning algorithms, the possibilities of which enable the development of new political strategies. Consequently,

political actors have started using newly developed tools in order to analyse citizens' behaviour and to influence the electoral body. One of these methods is microtargeting, which allows the formulation of personalized messages and their direct delivery to groups and individuals (Agan, 2007), hence creating a promising tool for electoral campaigning and opinion formation.

In this paper, we demonstrate a *proof of concept* regarding the ways political actors could establish the conditions for political microtargeting in Germany, through the utilization of social media platforms.

Bavarian School of Public Policy, Technical University Munich, Germany

Corresponding author:

Orestis Papakyriakopoulos, Bavarian School of Public Policy, Technical University Munich, Richard-Wagner-Strasse 1, 80333 Munich, Germany.
Email: orestis.papakyriakopoulos@tum.de



The scope of our analysis is to identify the possibilities and dangers of microtargeting in electoral campaigning, taking into consideration ‘state of the art’ technology. Therefore, we apply our method to Facebook data that could actually be used in political campaigning. Initially, we explain the theory behind microtargeting and discuss existing obstacles that prevent its application. Second, we illustrate our methodology and present our results. Lastly, we evaluate data-driven microtargeting ethically and comment on its political consequences.

Microtargeting in theory

Microtargeting is a strategic process intended to influence voters through the direct transmission of stimuli, which are formed based on the preferences and characteristics of an individual. First of all, microtargeting presupposes the collection of large amounts of data able to depict the political preferences and other non-political characteristics of voters. This data can be either manually collected or acquired through data-mining and can include information ranging from a person’s name, address, and voting history to more abstract properties such as a person’s opinion about political and non-political topics, their social activity and cultural background. The gathered data are then processed with the aid of appropriate *machine learning* algorithms, while the acquired results depend on the type of algorithm used. It is then possible to make predictions about specific variables, for example, the outcome of a political decision (supervised learning) or identification of patterns in the data through clustering (unsupervised learning). Implementing the latter, political actors are in a position to detect sub-groups of voters that share common demographic and attitudinal traits (Barbu, 2014). Based on the algorithmic results, they can then generate messages or plan actions aimed at influencing each specific sub-group or person (often called nanotargeting (Edsall, 2012)), leading to their potential mobilization or de-mobilization.

Microtargeting was first applied to a limited extent in the US 2000 Federal Elections by the Republican Party (Panagopoulos, 2015). Since then, the increasing datafication of societies has provided fertile ground for its expansion as a political strategy. A milestone for its application was the 2008 Federal Elections (Franz and Ridout, 2010), when the Democratic Party campaign applied the strategy at full scale. Today, microtargeting is a standard online and offline (Panagopoulos, 2015) campaigning method in the US as it overcomes problems of classical political campaigning. First of all, it

has the potential to partly track the predispositions or general interests of a voter (Ellul, 1966), and based on them, to modify the candidates’ public images in a way that complies with the voters’ opinions (Bond and Messing, 2015; Capara et al., 1999). Furthermore, by directly communicating individual- or group-specific messages, candidates are able to reduce the risk of alienating other voters that might disagree on a topic (Woo, 2015). Another advantage is that microtargeting allows political actors to target voters from the entire political spectrum, rather than exclusively developing their campaign on the characteristics of the *median voter* (Downs, 1957), as was the case in the past. Finally, given that opinion polls in the 2016 US and 2017 German elections failed to make plausible forecasts of election results, microtargeting provides a methodology to overcome political decisions based solely on survey polls. Despite the above advantages, it is important to note that there is no comprehensive study that proves the effectiveness of microtargeting (Jungheer, 2017; Karpf, 2016); to date it remains a promise emerging from the technological state of the art.

One of the main reasons behind the success of microtargeting in the US is the loose legal framework, which allows political actors to almost freely create, acquire and use databases that contain personal information. It is characteristic that there is no dedicated data protection law or a concept of ‘sensitive’ personal data in the US legislation. Hence, there is no general legislative framework exclusively dealing with the protection of a person’s privacy rights (Sotto and Simpson, 2015). Although legal frameworks, as the FTC, ECPA, HIPAA, etc., indeed aim to regulate the monitoring of personal data and their protection in their respective fields, the administration of data policies takes place usually only indirectly, by laws that might impose purpose limitations or time limits on the data retention (Boehm, 2015). Furthermore, the US law presents significant gaps concerning the protection of individual privacy (Ohm, 2014): e.g. the datafication or reuse of information acquired as a by-product of providing services is largely unregulated (Strandburg, 2014: 22). Consequently, such legal inconsistencies facilitate the development of huge political databases, which can then be used for political campaigning (Bennett, 2016).

Contrary to the US, the legal framework applicable in Germany significantly limits the potential of microtargeting. Germany’s privacy law complies with the EU-directive on the processing of personal data. The General Data Protection Regulation (EU-Directive, 2016) provides an extensive regulatory framework for

the protection of privacy and personal data, their acquisition, use and exchange. The GDPR thoroughly describes the limits and responsibilities of data controllers and processors, supports the subjects' rights to privacy and consent, and stipulates the exact regulating role of public authorities. Furthermore, the German data protection law explicitly defines the conditions and cases in which someone is able to access and use personal data (Däubler et al., 2016) and lays down the rights of persons affected (Broy, 2017), strongly limiting data exploitation.

Barrier 1: Privacy and data protection policy

Some authors¹ have argued therefore that microtargeting cannot be applied in German politics. However, despite the legal restrictions, there is ample leeway for it on social media platforms (Papakyriakopoulos et al., 2017). The reason is that the German privacy law permits the collection and processing of public personal data stemming from social media, as long as the individuals' interests are not challenged (Dorschel, 2015). The GDPR clearly states that given the appropriate safeguards, personal data on political opinion can be used for electoral activities (EU-Directive, 2016: 11). In addition, users on social media services consent to companies using their personal data for commercial and other activities, by opting in. Hence, the legal requirements for using social media data as basis for political microtargeting are met. Given the fact that users agree to publish on social media a huge amount of data about their political and non-political preferences and behaviour, these platforms are an ideal source for political knowledge extraction. Social media have become a key environment for political campaigns, as the majority of politicians can use them to communicate directly with the electoral body (Barberá and Zeitzoff, 2017; Hegelich and Shahrezaye, 2015; Medina Serrano et al., 2019; Nulty et al., 2016; Stier et al., 2017). That aside, political actors often perform organized influencing strategies on social media, frequently trespassing the legal limits set (Weedon et al., 2017).

Barrier 2: Data bias

The legal framework is not the only obstacle for successful microtargeting. The type of data subjected to algorithmic process and their entailed results can sometimes lead to spurious political action. In our case, the world of social media is not identical to the offline world. Hence, political preferences appearing on

social media platforms cannot be assumed to be the same for the actual electorate. The politically active user population on Facebook is in no way representative of the whole population of a country (Ruths and Pfeffer, 2014), while the expression of an opinion online does not fully correspond to a coherent political statement (a like is not a vote; Hegelich and Shahrezaye, 2015). Furthermore, the evaluation of social media data is bound with multiple methodological issues (Hegelich, 2017). Still, the case of the United States has shown that political campaigning is more than ever based on data, from which an electorate's image is derived, also known as *perceived voter model* (Hersh, 2015). This model may be misleading but nevertheless used, as it reduces the complexity in campaign decision-making. Due to the fact that it is almost impossible to causally link a campaigning tool to election results, microtargeting is used as long as it is assumed to have a successful influence – even if in reality it might not. The difficulties in causal inference arise – amongst others – from potential self-fulfilling prophecies: should a campaigning tool identify a target group, the campaign will increase interaction with this group. This special attention might yield positive results; but these results could have also been the same for a totally different group, as well. Despite the above, microtargeting is applied, even if it might be epistemologically impossible to evaluate its exact impact.

Data and method

In this paper, we demonstrate how politicians in Germany can create the conditions for microtargeting based on data from the social media platform *Facebook* and we evaluate its ethical and political consequences. Facebook was chosen as a data source for three reasons: (1) the German Facebook population is larger and less selective than that of Twitter. (2) It is part of the company's business model to offer targeted advertisement services for political campaigning, the possibilities of which we are exploring. (3) Contrary to the US, where there are extensive political databases with personal identifiers (Bennett, 2016), in Germany this is not the case. Hence, social media provide a straightforward way to acquire knowledge for microtargeting.

For our *proof of concept*, we analysed the public Facebook pages of the German political parties and their supporters: Our sample includes the following parties: Christlich Demokratische Union (CDU), Christlich Soziale Union (CSU), Sozialdemokratische Partei Deutschlands (SPD), Bündnis 89/ Die Grünen, Die Linke, and Alternative für Deutschland (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.

For each political page, we evaluated user “Likes” on political posts and assigned a partisanship to each user according to their preferences (Figure 1). Following a standard microtargeting technique, we focused our study on users who have liked content on pages of more than one political party. The reason behind this decision is that the specific group of voters, also named as cross-pressured partisans, has the highest likelihood to be influenced, as they are both undecided and engaged in politics (Ellul, 1966; Hersh, 2015). After identifying the relevant groups, we applied machine learning algorithms to cluster the various pages’ posts and created a mapping of 55 different topics, to which each of the posts might be assigned. To achieve this, we performed topic modelling analysis by applying a *Latent Dirichlet Allocation* algorithm (Blei et al., 2003). In this way, we demonstrate how someone can detect individual political topics of interest and how these can be later used to shape targeted messages for each micro group of users.

Prerequisite for the application of microtargeting is the existence of a rich database containing voters’ characteristics and preferences. Therefore, we mined data from 570 public pages related to the major political parties in Germany through the Facebook Graph API, and analysed posts and Likes. We selected the pages by searching the respective party names in the name field of the Facebook pages. We then classified manually our results, and removed irrelevant pages.² We mined every post generated by the administrators of the pages since their creation, the Likes each post got, and the unique IDs and profile names of the users liking them. Usually, the profile name of a Facebook account tends to be the same with the real name of the account holder, as Facebook maintains a real name policy.³ In total, we collected 251,947 posts with 6,347,448 Likes related to them and identified the activity of 1,208,740 unique users. This is only data related to the pages mined, hence the actual size of trackable users is even larger. We define a user who has liked at least one post of a party as partisan, and a user who has liked posts on pages of two or more parties as a cross-pressured partisan. Of course, the act of liking per se does not make someone a party partisan, but in this

case it provides a plausible classification method for the users. Furthermore, it does not distort the microtargeting process, as microtargeting targets the identification of voter’s predispositions and not to definitely certify someone’s exclusive support to a party. As shown in Figure 1, around 50% of the active users per party have made only one Like. This is typical of Big Data applications on social media phenomena, where the information for the majority of users is low.

Along with the identification of potential cross-pressured partisans, we wanted to identify the specific content that they find interesting. Therefore, we applied the LDA topic modelling algorithm (Blei et al., 2003) to classify 251,947 posts. LDA has many advantages over other standard text-mining algorithms (Grimmer and Stewart, 2013), as it can recognize complex relations in text-datasets. The algorithm has the ability to cluster posts in a certain number of topics, where each topic is a set of words that characterize different contents. Hence, someone can evaluate all the posts without having to investigate them one by one. LDA assigns a probability for each post belonging to a specific topic. Then, by ascribing to each post the topic with the highest probability and by detecting the users who liked it, we can explicitly track the topics that each user is interested in.

The LDA algorithm is a three-level hierarchical Bayesian model that predicts the probabilities of words and documents belonging to a number of topics K given the empirical distribution of words (or n-grams) in a corpus (Blei et al., 2003, 2002). In our case, the corpus consists of the total number of posts M under investigation, while each post corresponds to a document d , which is a sequence of N_d words. LDA is a generative model, i.e. it assumes the probability distributions of topics over words β_k , of documents over topics θ_d and predicts the probability that a specific word in a specific document will belong to a specific topic. This Bayesian admixture can be described by the following probability distributions

$$\theta_d \sim \text{Dir}_K(\alpha)$$

$$\beta_k \sim \text{Dir}_V(\eta)$$

$$z_w \sim \text{Multinom}_K(\theta_d)$$

$$w | z_w \sim \text{Multinom}_V(\beta_k)$$

where V is the number of unique words existing in the corpus, and α and η are Dirichlet parameters. Multinomial distribution z_w gives the probability that

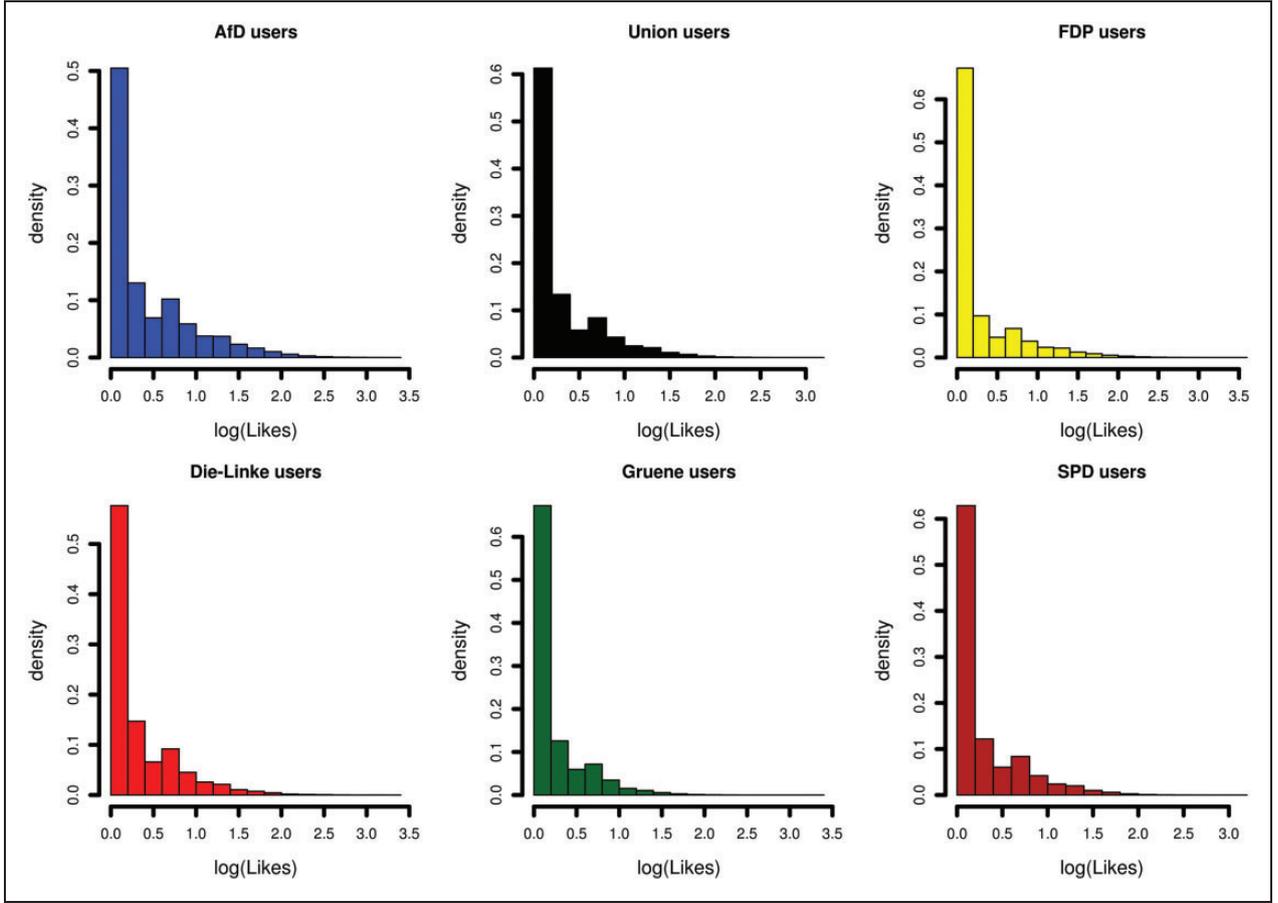


Figure 1. Likes distribution for the users on parties' pages.

a topic will be assigned to a word, given the distribution of topics over documents. Finally, multinomial distribution $w | z_w$ gives the probability that the model will generate a specific word in a specific document given a topic (Figure 2).

In our case, we want to create topics about the content of our corpus based on the empirical distribution of words over documents. Given the complexity of the model and the fact that the initial distributions are assumed and not empirically provided, we randomly assign topics to words and documents and we follow a Markov chain Monte Carlo procedure to update their values (Griffiths, 2002). By iteratively applying a Markov chain, we can converge to the assumed distributions and hence sample from them (Gilks et al., 1995; Roberts and Smith, 1994) the probability $P(z_w | w)$ that a word in a document belongs to a specific topic. More specifically, we used a collapsed Gibbs sampling Markov chain Monte Carlo (MCMC) (Geman and Geman, 1984) method to identify the relevant topics.

The specific algorithm comes with the advantage of integrating out the probability distributions β_k, θ_d (Darling, 2011). Thus as part of the iterative Markov chain, one can calculate the targeted probabilities through the process

for each document : $d_i = (1 \dots M)$

for each term in a document $i = (1 \dots N_{d_i})$

$$P(z_i = j | z_{-i}, V) = \frac{v_{-i,j}^{w=i} + \eta}{\sum_{w=1}^V v_{-i,j}^w + V\eta} \frac{n_{-i,j}^{d_i} + \alpha}{\sum_{k=1}^K n_{-i,k}^{d_i} + K\alpha}$$

where i is the concrete appearance of a word, $-i$ denotes its exclusion and j is a topic. $v_{-i,j}^{w=i}$ corresponds to the number of times word i is assigned to topic j , without its current appearance and index $\sum_{w=1}^V v_{-i,j}^w$ gives the total number of words in the corpus assigned to topic j excluding i . Furthermore, $n_{-i,j}^{d_i}$ contains the total number of words in document d_i that are assigned to topic j without i . Finally, $n_{-i}^{d_i}$ corresponds to the

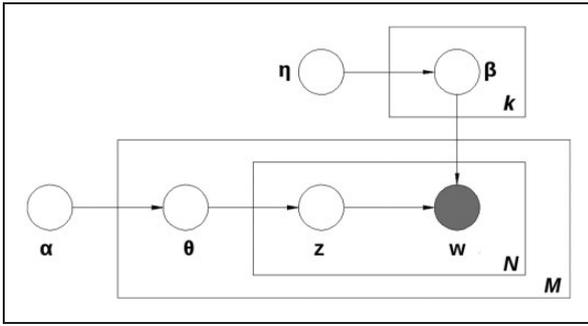


Figure 2. Plate notation for the Latent Dirichlet Allocation algorithm.

total number of words in the document, again not including i .

Necessary for the creation of a useful LDA model is the election of an appropriate number of topics, in order to split the content into interpretable subgroups. Electing a small number of topics results in a clustering of posts, from which one cannot identify concrete political topics of interest. On the contrary, if the number of topics is too large, the algorithm selects many words as topic-important that actually have no political value. To overcome this issue, we applied a topic optimization algorithm proposed by Deveaud et al. (2014). More specifically, we calculated the Jensen–Shannon divergence between topics for multiple LDA models through the equation

$$D(k_i, k_j) = \frac{1}{2} \sum_{w=1}^V \beta_{i,w} \log\left(\frac{\beta_{i,w}}{\beta_{j,w}}\right) + \frac{1}{2} \sum_{w=1}^V \beta_{j,w} \log\left(\frac{\beta_{j,w}}{\beta_{i,w}}\right)$$

where i, j are two different topics in a model and $\beta_{i,w}, \beta_{j,w}$ the probability density values of the distribution β_k for a word w in the corpus V and each topic, respectively, then selected the model that maximizes the sum of the Jensen–Shannon divergence for all topic combinations given the expression

$$K_{opt} = \operatorname{argmax} \frac{1}{K(K-1)} \sum_{k_i, k_j=1}^K D(k_i, k_j)$$

Based on the optimization process (Figure 3), we concluded on an LDA model with 55 topics. In order to sort and visualize topics according to their similarity, we used the method proposed by Sievert and Shirley (2014). We used the already calculated Jensen–Shannon

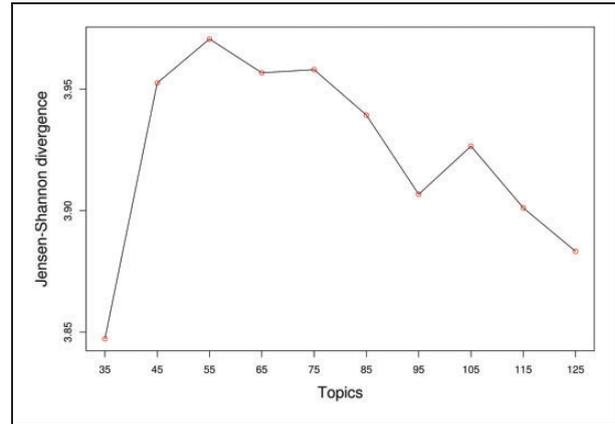


Figure 3. Topic optimization process. The model with the highest Jensen–Shannon convergence contained 55 topics.

divergences for the unique 1485 topic combinations and created a distance matrix. On it, we applied a principle component analysis algorithm (Hotelling, 1933) and we plotted the first three components.

Results

The first result of our analysis was the specification of the political content of the investigated posts. The LDA algorithm clustered the posts in 55 topics that can be split into three main categories. These categories were chosen manually, and do not denote that they are the optimal ones; still their election makes the results much more interpretable.⁴ The first category includes topics related to general political issues, such as social involvement (topic 1), education (topics 2, 15), national economy (topic 4) and homeland security (topic 32). Some topics do not only illustrate the relevance of posts to a political issue, but also the exact opinion underlying them. For example, topics 10 and 12 are both migration related, but topic 10 includes posts that are refugee-friendly, while topic 12 contains posts that demand a stricter migration policy. In addition, there are topics that analyse political parties (topic 39) or persons (topic 38). In the same category, also exists a set of topics (9, 27, 14) that contain posts that do not make concrete political statements, but declare uncertainty and reflection.⁵ The second category includes topics that are related to political actors and candidates, but not as part of a political discussion. They summarize posts about political events, media appearances and electoral campaigning. Finally, the third category contains topics that are location related and discuss political problems about regions. For example, topic 54 includes

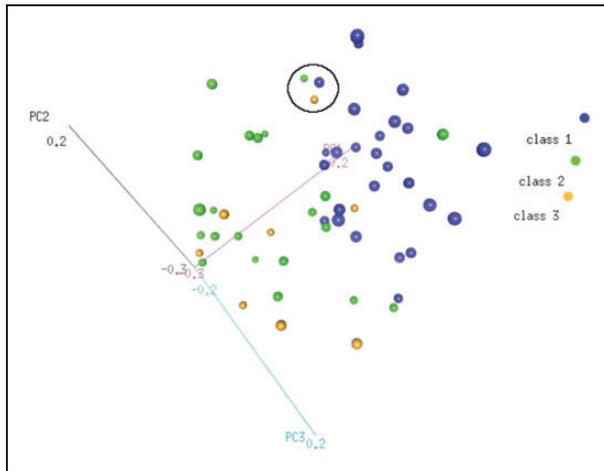


Figure 4. Topic distance visualization with the help of PCA. Circled are topics 21, 43, 38.

posts about Berlin, topic 31 about Hamburg and topic 43 about Bavaria.

In order to evaluate and verify our topic classification, we visualized the relationship between the developed topics in a three-dimensional space with the help of PCA (Figure 4). Each sphere corresponds to a different topic, while their size is proportional to the number of posts they contain. Their distance in 3D-space functions as a measure of their content similarity. It is visible that three categories classify topics into unique clusters. As expected though, there is some overlapping between categories, as a topic might contain keywords belonging to more than one categories. For example topics 21, 43, 38 appear very close, even though we classified them differently (Table 1). This occurs because they all include a combination of posts of all classes. Topic 21 is about AfD, including both posts about its political background and the elections. Topic 38 is about Angela Merkel and her political activity, as well as her party structure. Finally, topic 43 is about Bavaria, including a number of posts about the regional CSU party and its candidates.

In our analysis, we identified a total of 58,532 cross-pressured users. Figure 5 shows that cross-pressured users tend to like more frequently than the average Facebook partisan. This however does not mean that cross-pressured partisans tend to be more active; on the contrary, it denotes that we can only trace cross-pressured partisans, when the users are more active online. This has an important implication for the perceived voter's model: The selection of cross-pressured partisans as targeted population comes with the advantage that they behave as multipliers, and thus their

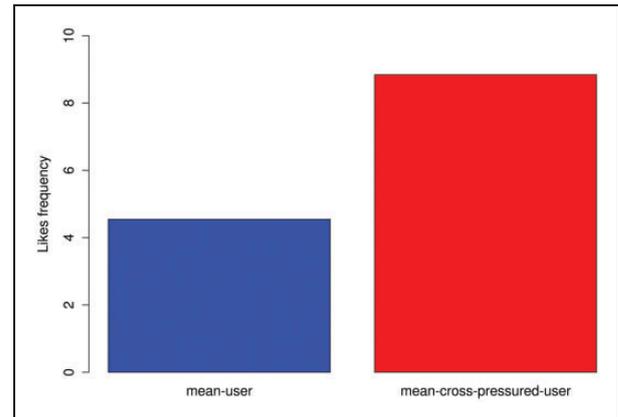


Figure 5. Average Likes frequency for the mean and the cross-pressured user.

potential influence will contribute to the motivation of other users as well.

Figure 6 shows the ratio of cross-pressured partisans between parties. In the given dataset, more or less 10% of the page users for each party are cross-pressured. This does not mean though, that this number corresponds to the actual electorate, as the descriptive results are biased through our statistical sample and the structure of the social media platform. Nevertheless, it is possible to recognize certain predispositions of the electorate, as for example an increased interaction of Union and FDP users and the almost non-existent overlap of users that are interested in both Die Grünen and AfD.

After the concretization of the topics of interest, microtargeting can be performed in two ways: one can either initially focus on single users and then track afterwards the topics they are interested in, or select specific topics and then identify users interested in them. To demonstrate how further steps of the microtargeting process could be realised, we choose randomly topic 4 as an example. Topic 4 includes, amongst others, the words: Euro, Steuergeld, Milliarde, Zuschuss, Kosten, i.e. it is linked to German economic policy. It is possible to analyse the relevance of this topic for each party, as well as to identify users who like the topic. In this case, we find Union coalition posts that talk about the German economy and identify the relevant cross-pressured partisans. Then, we randomly pick one of the users to investigate all the other topics that are of interest to her. Our random cross-pressured user has also liked FDP posts, and as Table 2 shows, she has also expressed interest in political issues of Schleswig Holstein and homeland security. Hence, we can identify significant political topics of

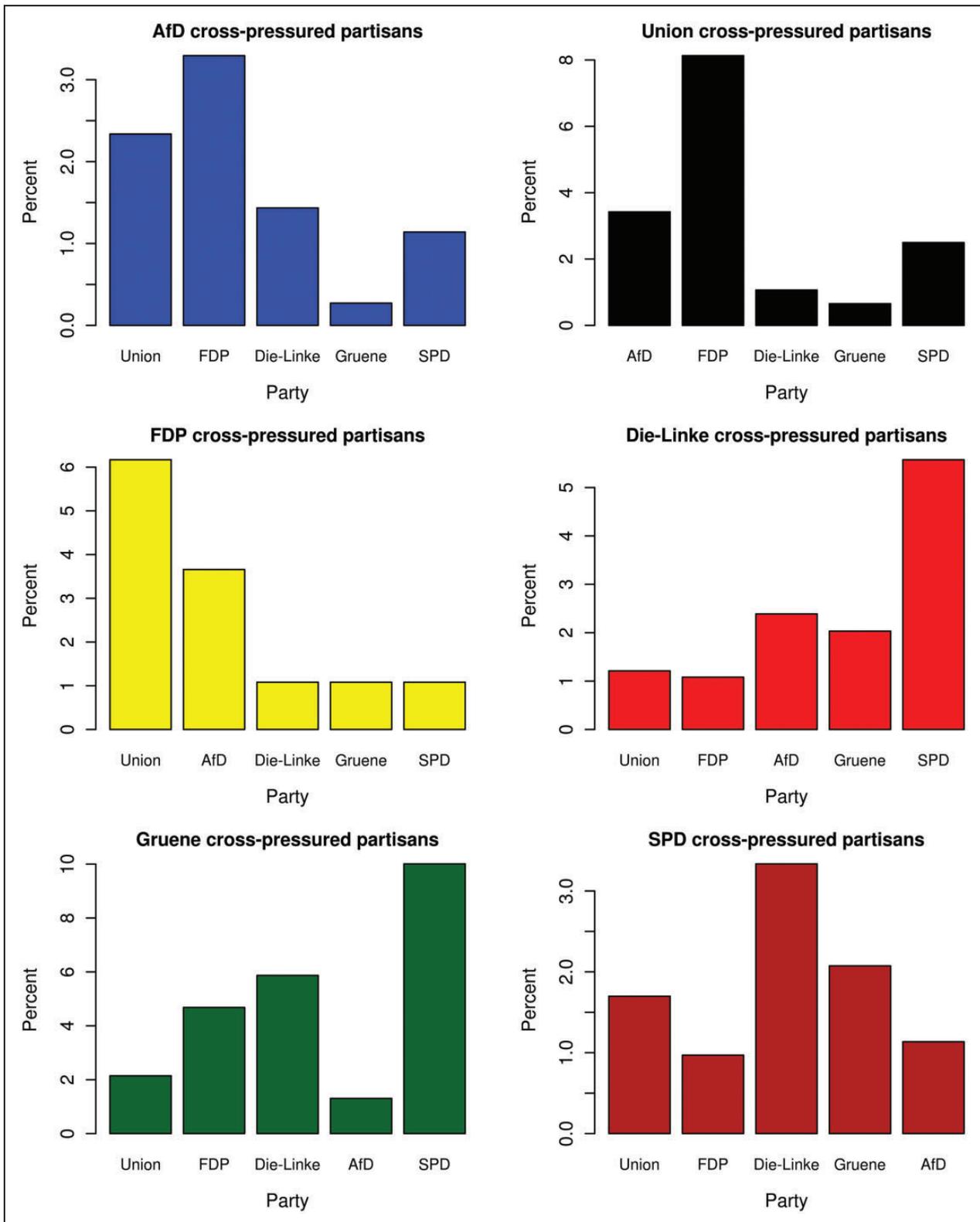


Figure 6. Percentage of cross-pressured partisans per party.

Table 1. Extended keywords for topics 21, 38, 43.

Topics	Extended keywords
21	AfD, Partei, rechtspopulist, Position, Altparteien, Wahlen, Argument, Stimmen, vertreten, Gegner
38	Merkel, Angela, Kanzlerin, CDU, Union, CSU, Seehofer, Volk, Flüchtlingspolitik, Terroranschlag
43	Bayern, München, Freistaat, Wahlprogramm, CSU, muss, Regierung, Generalsekretarin, Schalzwedel, Grüne

Table 2. Topics of interest for an example-user and for Union-SPD cross-pressured users.

Target	Topic	keywords				
Example user	4	Euro	Steuergerl	Milliar	Zuschuss	kosten
	32	Innenminister	Polizei	ermittelt	Justitz	Kriminalität
	51	Schleswig	Holstein	Kiel	Rostock	Schwerin
Union-SPD cross-pressured users	8	Islam	Muslim	Christlich	Religion	Kirche

interest for the user, as well as political parties to which, the user is positively inclined.

The topic modelling algorithm, however, does not illustrate if the user thinks positively or negatively of a political topic, i.e. it does not trace their exact political attitudes to the issues. To do this it would be necessary to apply a sentiment analysis algorithm to the parties' posts, or a qualitative analysis thereof. In the current research, we did not perform a sentiment analysis. Given the results of the sentiment analysis, the person's political evaluations of political topics and party sympathy, a campaign-maker has adequate information to create personalized messages and communicate them through micro-targeted advertisement.

Similarly, it is possible to identify topics that are important for groups of strategic importance. For example, partisans that are cross-pressured by the Union and SPD are highly interested in topic 8, which is related to Islam and Christianity. Thus, after the combination with a sentiment analysis, the creation of an advertisement specifically related to this topic can provide additional advantage to a political party, as it might mobilize an important part of the electorate towards its ends. Of course, the content of a personal message can be further specialised, as it is always possible to access recursively the full post that a user liked, and locate exactly its content in relation to the topic it belongs to.

Given the mined Facebook data, we proved that there is an extensive dataset for potential microtargeting in German politics available in social media services. Although national privacy regulations usually forbid the direct acquirement and use of personal

data, data existing on social media platforms provide a fruitful source for microtargeting. By mining and structuring the content of 570 German political pages, we managed to detect over 58,000 cross-pressured users through their Likes. The selection of this sub-population was based on the idea that they are people both active in politics and potentially undecided on their exact party preference. Hence, communicating a message to them is of greater value than to people who are strict supporters of one party or are not interested in politics at all. In order to track topics of interest of cross-pressured users, we applied simple machine learning algorithms on the pages' content and found the most common issues discussed. Finally, we connected the topics with the users through their posts' Likes, finding out valuable political information about them. Accompanied with a sentiment analysis algorithm, the necessary knowledge can be gathered for the creation of personalized messages. Last step is to contact the users, a process that should be adapted to and compliant with the legal frameworks.

The communication of the message could theoretically be performed in two ways: One could cluster users sharing common characteristics and directly target them through the platform's advertisement service, which allows campaigners to define custom target audiences. This comes with the advantage that there is no need for manual matching of users to their real world identities, as it suffices to communicate the message to them through the platform. The second way is to manually look at a person's further public activity on Facebook, and given additional sociodemographic data available, try to find another communication path (e.g. email, mail, phone number, etc.). Although the second way is time-

consuming, complicated, and sometimes inadequate, gathering socio-demographic data about individuals and then targeting them offline is actually what is intensively done in US campaigns (Hersh, 2015: 77). Still, in EU the feasibility of the strategy is much lower, due to the existing privacy laws. For the second way to be applicable, political actors should develop platforms, applications, or services, through which they would get the person's consent to target them with the related messages.

The processing of the social media political dataset also comes with specific limitations. The inferences drawn reveal only part of a person's political characteristics, and only if indeed someone's online behaviour matches their actual political preferences. Furthermore, the users detected online might not have a voting right in Germany, making the sampling process biased and distorting the advertisement process.

The presented results serve as a proof of concept. We have thoroughly described how microtargeting based on social media data could be performed. The analysis was focused on Germany, where the acquisition of relevant data is usually problematic. The described method can be extended through further actions in both online and offline campaigning. For example, parties have already started promoting apps to connect the digital and analogue campaigning.⁶ These apps help to analyse the reactions of people, giving feedback to the campaign-managers about their campaigning tactics. Furthermore, the combination of the app data with data coming from social media can provide even more insights on the relevant issues. The processed social media data can also be used to complement standard opinion prediction techniques. Existing census data about demographic characteristics and public record data about past voting behaviour can be combined with results from the topic modelling and sentiment analysis algorithms and hence explain the features of political behaviour.

In our study, we focused only on the detection of voters' political topics of interest, however part of the microtargeting process is also the evaluation of the personalized advertisement's success. This can be done after the first application of microtargeting, through analysis of click-statistics, performance of surveys and the actual election results. Furthermore, after the calibration of the process, the generation of microtargeting data can be highly automated. This of course raises the question of whether politicians' positions would still be a result of their actual opinions or just an algorithmic creation for attracting voters. Finally, machine learning algorithms can predict the users' interest in further topics or parties, even if they have not liked them on the platform. Further data would be required for this,

which in this case were not taken into consideration, but are still publicly available online (Kosinski et al., 2013). By collecting data from other social media interactions, e.g. likes on news media or other non-political pages, one can train models and assign probabilities of someone being interested in a political issue or party. In this way, political knowledge can be extracted about users that actually did not actually interact with any party-related content on the platform and hence be included as audience of political microtargeting.

Discussion

The penetration of datafication into people's privacy is once more proven through our investigation, as we were able to gather and process a large amount of user data from the social media platform Facebook. Hence, from our perspective, it is important to evaluate the impact of the latest technological advances on the ethical and political life of our society. The discussion that has already started regarding the application of data-intensive algorithms to social networks (e.g. social bots (Thieltges et al., 2016), using algorithms for social engineering (Strohmaier and Wagner, 2014)), must now be also extended to the effect of microtargeting as a technology driven campaigning method. As the new technological capabilities raise questions regarding the limits of ethical political influence and the potential transformation of political behaviour in contemporary society, our task is to identify and reflect on the newly emerged issues.

The study showed, that through machine learning, it is possible to track someone's interests and subsequently develop personalized political advertisement that can be used to influence social media users. Hence, the first question emerging is whether microtargeting might lead to the manipulation of voters. The transmission of a personalized message does not per se signify the manipulation of a person, as each individual possesses the freedom to decide whom to vote for. As the public is offering more and more voluntarily their information in exchange for online or offline services (Barbu, 2014) though, algorithms tend to become more precise in evaluating personal preferences and attitudes. As microtargeting could potentially contact the person directly with a very well adapted message, it might achieve what is called instant influence: trigger the person's mind to develop a conditioned response the way the political actors desire (Cialdini, 2007). This happens, because in cases of fast incoming information stimuli, the individual does not process them rationally (Simon, 1996). On the contrary, the

information is assimilated intuitively, creating a phenomenalist connection between the message and the political party (Piaget, 1947). Of course, framing a party successfully also presupposes other psychological, social and political preconditions to be present (Domke et al., 1998; Schmitt-Beck, 2003), which cannot be formed by simply sending well-adapted personal messages. But given these conditions, a systematic application of microtargeting might lead to a ‘progression from thought to action artificially’ (Ellul, 1966). A reaction to this issue is the conscious understanding of the person that they are being microtargeted. In this way, they would be in position to evaluate a message totally differently, knowing that the incoming stimuli are already adapted to their own attitudes. The rule of the conscious over the unconscious is a precondition for the society to remain autonomous (Castoriadis, 1997).

This type of consciousness is not only needed at the moment of evaluating a political message, but must also exist at the level of privacy. It is common that through the use of apps and online platforms, people voluntarily provide their personal data and allow their further usage as a by-product of the service. It is important for users to become aware of what they are agreeing on, and what consequences their actions have. In this direction, certain normative and legal imperatives have already been formulated: Transparency of data collection, processing and application (Barocas et al., 2017), autonomy of the subject on having control of their own personal data (McDermott, 2017), and (in)visibility: the right of the subject to choose if and to know how personal data might be collected and used (Taylor, 2017), are stated as necessary for supporting someone’s privacy. The EU General Data Protection Regulation makes also steps towards this direction, by explicitly incorporating transparency and consent in its regulatory claims.

Despite the regulatory efforts, the act of a user opting in, given a very long document of terms and conditions, where how personal data might be used is outlined in a short and general manner does not signify transparency, or actual consent (Strandburg, 2014). Especially regarding personal data for microtargeting, the information that should be presented to the subject in order to give their consent should clarify exactly what information is going to be collected, how, by whom and for what purpose. This is a prerequisite for the subjects’ expectations about the collected data to coincide with the actual data usage (Barocas and Nissenbaum, 2014). At the same time, the individuals should be emancipated, by both getting to know through access to the history of their personal data used by services (Kennedy and Moss,

2015), and realizing how datafication has pragmatically altered the contemporary social structure.

Important for the ethical evaluation of microtargeting, as well as for data privacy, is also who acquired the related data, not only how. For us being able to gain access to the aforementioned dataset poses a dilemma: Should public data, for which users have provided their consent to be used and further processed, become openly available, or should they remain only under the control of the initial gatherer? The question is relevant more than ever to the present discussion, given the contemporary Facebook data scandal (Facebook, 2018a, 2018b), as well as the platform’s decision to significantly limit the data available through its application programming interface (API). On the one hand, making data broadly open might result to an uncontrolled data mining phenomenon (Pasquale, 2015), with private data becoming a part of the public sphere. On the other hand, the possession of these public data only by the original gatherer might result in the problem of a knowledge monopoly, making the data holder much more powerful in economic and political terms than other social actors.

The specific case study would have a different form, if the data were collected under the new API rules of the platform. Important public data for microtargeting, as user likes, cannot be downloaded in an automated way. If public online data are accessible only to the extent platforms decide, and political actors can target users exclusively through the targeting services provided, then the political system itself becomes contingent to technological companies. Electing microtargeting as a political campaigning strategy thus presupposes the constant compliance of political actors with the existing political and legal conditions (Kruschinski and Haller, 2017), as well as with the market structures and the dominant online platform decisions.

Another issue regarding microtargeting is related to the perceived voter model. Given that the majority of users in social networks are relatively inactive, the danger exists that politicians will concentrate on the analysis of data provided by the more active users, even if that sample is not representative of the population (Barberá and Rivero, 2015). The less data one can gather about a person, the more inexact can their attitude-prediction be. Thus, a campaign might be developed based on falsely assessed voters’ attitudes. If political campaigns are highly or exclusively data-driven, it leads to the perceived voter phenomenon (Hersh, 2015): All campaigning decisions are based to an algorithmically calculated electorate and thus, any forecasts are dependent on the nature of the collected data. Given that social media data always possess a

certain rate of bias (Ruths and Pfeffer, 2014), it is possible that political actors might perform a campaigning on a ‘constructed’ reality and not on an actual one. Of course, gathering of even more data is not a solution. If someone observes campaigning in the US, they might question the independency of the electorate: US parties’ campaigns aim for the mobilization or de-mobilization of specific social groups, demographic layers and geographic populations in order to strategically achieve their goals (Hersh, 2015; Kreiss, 2016; Persily, 2017). Furthermore huge public databases contain extensive data about the majority of the electorate and their voting history. The discussion about microtargeting and data privacy is already under way in Europe and the newly emerged issues should be assessed.

This study demonstrates through its ‘proof of concept’ certain possibilities and dangers of microtargeting, in order to initiate an important debate for the political system. To expand this discussion, further qualitative and quantitative research is needed, in order to uncover: (1) How political communication on social media influences the formation of political attitudes in terms of polarization, political mobilization and opinion formation? (2) What is the effect of political campaigning services offered by social media and other internet platforms? (3) At which level current privacy policies protect individuals and what else could be done? The answers to the aforementioned questions, if given, can redefine how the political discourse should be performed in the digital age.

Declaration of conflicting interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: the German Research Foundation (DFG) and the Technical University of Munich within the funding programme Open Access Publishing.

ORCID iD

Orestis Papakyriakopoulos  <http://orcid.org/0000-0003-4680-0022>

Notes

1. See e.g. Christl (2016) and Thiele (2017).
2. The pages of CDU and CSU were classified together under the term *Union*.

3. <https://www.facebook.com/help/112146705538576> (accessed 21 March 2018).
4. Appendix 1 contains the full description of the topics created, as well as their important keywords.
5. The topics contain keywords as e.g. Vielleicht, aber, glaube, nachdenken.
6. E.g. CDU’s app ‘connect17’.

References

- Agan T (2007) Silent marketing: Micro-targeting. Available at: <http://gaia.adage.com/images/random/microtarget031207.pdf> (accessed 27 March 2018).
- Barberá P and Rivero G (2015) Understanding the political representativeness of twitter users. *Social Science Computer Review* 33(6): 712–729.
- Barberá P and Zeitoff T (2017) The new public address system: Why do world leaders adopt social media? *International Studies Quarterly* 62(1): 121–130.
- Barbu O (2014) Advertising, microtargeting and social media. *Procedia – Social and Behavioral Sciences* 163: 44–49.
- Barocas S, Bradley E, Honavar V, et al. (2017) Big data, data science, and civil rights. *ArXiv*. Available at: <https://arxiv.org/abs/1706.03102> (accessed 5 November 2018).
- Barocas S and Nissenbaum H (2014) Big data’s end run around anonymity and consent. *Privacy, Big Data, and the Public Good: Frameworks for Engagement* 1: 44–75.
- Bennett CJ (2016) Voter databases, micro-targeting, and data protection law: Can political parties campaign in Europe as they do in North America? *International Data Privacy Law* 6(4): 261–275.
- Blei DM, Ng AY and Jordan MI (2002) Latent Dirichlet allocation. In: Dietterich TG, Becker S and Ghahramani Z (eds) *Advances in Neural Information Processing Systems 14*. Cambridge, MA: MIT Press, pp. 601–608.
- Blei DM, Ng AY and Jordan MI (2003) Latent Dirichlet allocation. *Journal of Machine Learning Research* 3: 993–1022.
- Boehm F (2015) A comparison between us and EU data protection legislation for law enforcement purposes. Available at: www.europarl.europa.eu/RegData/etudes/2015/536459/IPOL_STU%282015%29536459_EN.pdf (accessed 28 March 2018).
- Bond R and Messing S (2015) Quantifying social media’s political space: Estimating ideology from publicly revealed preferences on Facebook. *American Political Science Review* 109(1): 62–78.
- Broy D (2017) Germany: Starting implementation of the GDPR-brief overview of the government bill for a new Federal Data Protection Act. *European Data Protection Law Review* 3: 93.
- Capara GV, Barbaranelli C and Zimbardo PG (1999) Personality profiles and political parties. *Political Psychology* 20(1): 175–197.
- Castoriadis C (1997) *The Imaginary Institution of Society*. Cambridge, MA: MIT Press.

- Christl W (2016) Big data im wahlkampf: An ihren daten sollt ihr sie erkennen. Available at: <http://www.faz.net/aktuell/feuilleton/medien/big-data-im-wahlkampf-ist-microtargeting-entscheidend-14582735.html> (accessed 28 March 2018).
- Cialdini RB (2007) *Influence: The Psychology of Persuasion*. New York, NY: Harper Collins.
- Darling WM (2011) A theoretical and practical implementation tutorial on topic modeling and Gibbs sampling. In: *Proceedings of the 49th annual meeting of the association for computational linguistics: Human language technologies*, Portland, Oregon, 19 – 24 June 2011. pp.642–647. Madison, WI: Omnipress Inc.
- Däubler W, Klebe T, Wedde P, et al. (2016) *Bundesdatenschutzgesetz*. Frankfurt: Bund-Verlag.
- Deveaud R, SanJuan E and Bellot P (2014) Accurate and effective latent concept modeling for ad hoc information retrieval. *Document numérique* 17(1): 61–84.
- Domke D, Shah DV and Wackman DB (1998) Media priming effects: Accessibility, association, and activation. *International Journal of Public Opinion Research* 10(1): 51–74.
- Dorschel J (2015) *Praxishandbuch Big Data: Wirtschaft–Recht–Technik*. Wiesbaden: Springer-Verlag.
- Downs A (1957) An economic theory of political action in a democracy. *Journal of Political Economy* 65(2): 135–150.
- Edsall TB (2012) Let the nanotargeting begin. Available at: <https://campaignstops.blogs.nytimes.com/2012/04/15/let-the-nanotargeting-begin/> (accessed 28 March 2018).
- Ellul J (1966) *Propaganda*. New York, NY: Knopf.
- EU-Directive (2016) Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). *Official Journal of the European Union* L119: 1–88.
- Facebook (2018a) Hard questions: Update on Cambridge Analytica. Available at: <https://newsroom.fb.com/news/2018/03/hard-questions-cambridge-analytica/> (accessed 20 September 2018).
- Facebook (2018b) Suspending Cambridge analytica and SCL group from Facebook. Available at: <https://newsroom.fb.com/news/2018/03/suspending-cambridge-analytica/> (accessed 20 September 2018).
- Franz MM and Ridout TN (2010) Political advertising and persuasion in the 2004 and 2008 presidential elections. *American Politics Research* 38(2): 303–329.
- Geman S and Geman D (1984) Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence*. Vol. 6. IEEE, pp.721–741.
- Gilks W, Richardson S and Spiegelhalter D (1995) *Markov Chain Monte Carlo in Practice*. London: Chapman and Hall.
- Griffiths T (2002) Gibbs sampling in the generative model of latent Dirichlet allocation. Available at: <https://people.cs.umass.edu/~wallach/courses/s11/cmpsci791s/s/readings/griffiths02gibbs.pdf> (accessed 28 March 2018).
- Grimmer J and Stewart BM (2013) Text as data: The promise and pitfalls of automatic content analysis methods for political texts. *Political Analysis* 21(3): 267–297.
- Hegelich S (2017) R for social media analysis. In: Luke Sloan AQH (ed.) *The SAGE Handbook of Social Media Research Methods*. London: SAGE Publications, Chapter 28.
- Hegelich S and Shahrezaye M (2015) The communication behavior of German MPS on twitter: Preaching to the converted and attacking opponents. *European Policy Analysis* 1(2): 155–174.
- Hersh ED (2015) *Hacking the Electorate: How Campaigns Perceive Voters*. New York, NY: Cambridge University Press.
- Hotelling H (1933) Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology* 24(6): 417.
- Jungherr A (2017) Einsatz Digitaler Technologie im Wahlkampf. *Schriftreihe Medienkompetenz* 10111: 92–101.
- Karpp D (2016) The partisan technology gap. In: Gordon E and Mihailidis P (eds) *Civic Media: Technology, Design, Practice*. Cambridge, MA: MIT Press, pp. 199–216.
- Kennedy H and Moss G (2015) Known or knowing publics? Social media data mining and the question of public agency. *Big Data & Society* 2(2): 2053951715611145.
- Kosinski M, Stillwell D and Graepel T (2013) Private traits and attributes are predictable from digital records of human behavior. *Proceedings of the National Academy of Sciences* 110(15): 5802–5805.
- Kreiss D (2016) *Prototype Politics: Technology-Intensive Campaigning and the Data of Democracy*. New York, NY: Oxford University Press.
- Kruschinski S and Haller A (2017) Restrictions on data-driven political micro-targeting in Germany. *Internet Policy Review* 6(4): 1–23.
- McDermott Y (2017) Conceptualising the right to data protection in an era of big data. *Big Data & Society* 4(1): 2053951716686994.
- Mayer-Schönberger V and Cukier K (2013) *Big Data – A Revolution that Will Transform how We Live, Work, and Think*. Orlando, FL: Houghton Mifflin Harcourt.
- Medina Serrano JC, Hegelich S, Shahrezaye M, et al. (2019) *Social Media Report: The 2017 German Federal Elections*. Munich: TUM University Press.
- Nulty P, Theocharis Y, Popa SA, et al. (2016) Social media and political communication in the 2014 elections to the European parliament. *Electoral Studies* 44: 429–444.
- Ohm P (2014) Changing the rules: General principles for data use and analysis. *Privacy, Big Data, and the Public Good: Frameworks for Engagement* 1: 96–111.
- Panagopoulos C (2015) All about that base. *Party Politics* 22(2): 22–190.
- Papakyriakopoulos O, Shahrezaye M, Thielges A, et al. (2017) Social media und microtargeting in Deutschland. *Informatik-Spektrum* 40(4): 327–335.
- Pasquale F (2015) *The Black Box Society: The Secret Algorithms that Control Money and Information*. Cambridge, MA: Harvard University Press.

- Persily N (2017) Can democracy survive the internet? *Journal of Democracy* 28(2): 63–76.
- Piaget J (1947) *The Psychology of Intelligence*. London, New York: Routledge.
- Roberts G and Smith A (1994) Simple conditions for the convergence of the Gibbs sampler and Metropolis-Hastings algorithms. *Stochastic Processes and their Applications* 49(2): 207–216.
- Ruths D and Pfeffer J (2014) Social media for large studies of behavior. *Science* 346(6213): 1063–1064.
- Schmitt-Beck R (2003) Mass communication, personal communication and vote choice: The filter hypothesis of media influence in comparative perspective. *British Journal of Political Science* 33(2): 233–259.
- Sievert C and Shirley KE (2014) Ldavis: A method for visualizing and interpreting topics. In: *Proceedings of the workshop on interactive language learning, visualization, and interfaces*, Baltimore, Maryland, 27 June 2014, pp. 63–70.
- Simon HA (1996) *The Sciences of the Artificial*. Cambridge, MA: MIT Press.
- Sotto LJ and Simpson AP (2015) Data protection & privacy: United States. Available at: https://www.huntonprivacyblog.com/wp-content/uploads/sites/18/2011/04/DDP2015_United_States.pdf (accessed 28 March 2018).
- Stier S, Posch L, Bleier A, et al. (2017) When populists become popular: Comparing Facebook use by the right-wing movement Pegida and German political parties. *Information, Communication & Society* 20(9): 1–24.
- Strandburg KJ (2014) Monitoring, datafication and consent: Legal approaches to privacy in the big data context. In: Lane J, Stodden V, Bender S, et al. (eds) *Privacy, Big Data and the Public Good*. New York, NY: Cambridge University Press, pp. 5–43.
- Strohmaier M and Wagner C (2014) Computational social science for the world wide web. *IEEE Intelligent Systems* 29(5): 84–88.
- Taylor L (2017) What is data justice? The case for connecting digital rights and freedoms globally. *Big Data & Society* 4(2): 2053951717736335.
- Thiele M (2017) Die wahlschlacht der datenbanken. Available at: <http://www.tagesspiegel.de/themen/freie-universitaet-berlin/wahlkampf-mit-big-data-die-wahlschlacht-%der-datenbanken/19938576.html> (accessed 28 March 2018).
- Thieltges A, Schmidt F and Hegelich S (2016) The devils triangle: Ethical considerations on developing bot detection methods. In: *Proceedings of the 2016 AAAI spring symposium, Stanford University, Palo Alto, California, 21 – 23 March 2016*, Vol. 2123, pp.253–257.
- Weedon J, Nuland W and Stamos A (2017) Information operations and Facebook. Technical report, Facebook. Available at: <https://fbnewsroomus.files.wordpress.com/2017/04/facebook-and-information-operations-v1.pdf> (accessed 28 March 2018).
- Woo HY (2015) *Strategic communication with verifiable messages*. PhD Thesis, University of California, USA.

Appendix I

Table 3. Topics overview for category I: general political issues. Topic content keywords.

1	State / Citizens / Social involvement	Bürgerinnen	Gestalten	Zusammenhalten	Engagement	Landkreis
2	Education / Thoughts	Gymnasium	Lösung	Lernen	Klasse	Anforderung
3	Law	Bundesverfassungsgericht	Verfassung	Urteil	Grundgesetz	Bundesrepublik
4	Economy	Euro	Steuergerld	Milliard	Zuschuss	kosten
5	Transportation policy	Flughafen	Nahverkehr	Bahn	Mitarbeiter	Verkehrspolitik
6	Democracy / People / Germany	Demokratie	Volk	Elite	Freiheit	Bürger
7	Against left-wing radicalism	Linksextremisten	Antifa	Gewalttat	Straftat	Polizei
8	Religion / Islam / Christianity	Islam	Muslim	Christlich	Religion	Kirche
9	Thoughts	Vielleicht	Aber	Glaube	Eigentlich	Ich
10	Refugee policy / for	Unterkunft	Fluchtling	Asybewerber	Aufnahme	Geflüchtet
11	Energy policy	Energie	Umwelt	Klimaschutz	Landwirtschaft	Energiepolitik
12	Refugee policy / Against	Fluchtling	Asyl	Abschiebung	illegal	Asylverfahren
14	Austerity / Unemployment / Thoughts	Jobcenter	eigentlich	soll	Sparen	Rettung
15	Education	Schule	Kinder	Eltern	Bildung	Lehre
16	Foreign policy	EU	Russland	Ukrain	USA	Turkei
18	Social policy / Hartz 4 / Poverty	Hartz IV	Armut	Sozial	Gerecht	Rente
19	Greek crisis	Griechenland	Bank	Finanz	Steuerzahl	Schuld
20	Housing policy	Wohnung	Wohnraum	Miete	Verwaltung	Wohnungsbau

(continued)

Table 3. Continued

21	AfD (political discussion)	AfD	Partei	rechtspopulist	Position	Altparteien
23	Against Pegida	Demonstration	Pegida	Nazis	Rassismus	gegen
24	Against right-wing radicals, racism	Diskriminierung	Homophobie	Rechtsextremismus	Freiheit	Rassisten
25	Income / Workers unions	Mindestlohn	Arbeitsgeber	Arbeitsnehmer	Gewerkschaft	Arbeitsbedingung
26	German left-wing history	DDR	Rosa Luxemburg	NATO	Geschichte	Revolution
27	Thoughts	Nachdenken	Denkst	Wahrheit	Du	Einfach
28	Family	Frau	Mann	Mutter	Familie	Kinder
32	Homeland security	Innenminister	Polizei	Ermittelt	Justitz	Kriminalität
37	Against TTIP/CETA	TTIP	CETA	Stopp	unterschreiben	Aktion

Table 4. Topics overview for category 2: political actors' activity.

Topic content keywords						
13	Political events	Eingeladen	Veranstaltung	Lädt	Vortrag	Diskussion
22	Greetings	Gruss	Liebe	Freunde	Melden	Spenden
29	Die Grünen	Grünen	Bündnis	Landtag	Grün-linke	Sachsen
30	After political events	Danke	Besuch	toll	Fotos	Impression
33	Congratulations	Glückwunsch	Herzlich	Gratulieren	Wahlgang	Wiedergewählt
34	Schwesig (Politician)	Schwesig	Manuela	Andrea	Nahles	Frau
35	Political Coalitions	rot	grün	Schwarz	Gelb	Koalition
38	Merkel	Merkel	Angela	Kanzlerin	CDU	Union
39	Petry	Petry	Lucke	Alternative	Deutschland	AfD
40	Election campaign	Daum	Druck	Wahlkampf	Stimmen	Sonntag
41	Candidates	Wahlkreis	Kandidat	Landesliste	Nominiert	Listenplatz
42	Wagenknecht	Mannheim	Wagenknecht	sahra	Linksjugend	Freiburg
45	Twitter	Twitter	Schaut	Teilen	Mitmachen	Abstimmen
46	Various politicians	Gabriel	Schulz	Gauck	Bundespräsident	Steinmeier
48	Debates/ TV	live	Aktuell	TV	gleich	Fernsehen
49	FDP/ Rheinland	Pfalz	Rheinland	Liberal	FDP	Liberte
50	Greetings/ Thank you	Wünsche	Spass	Gut	frohe	Feiertag
52	Lindner/ FDP	Lindner	Christian	NRW	Bundesvorsitzender	Kubicki
53	Die LINKE	die Linke	linksfraktion	Riexinger	Kipping	themen
55	German news media	Focus	Welt	Spiegel	Interview	Zeitung

Table 5. Topics overview for category 3: regional topics.

Topic content keywords						
117	NRW politicians	Münster	Bochum	Bezirksvertreter	Essen	Ruhr
31	Hamburg	Altona	Hamburg	Landesparteitag	Bezirkversammlung	Bürgerschaft
36	Leipzig AfD	Leipzig	Kreisvorsitzende	Kreisverband	Vorstand	Mitglied
43	Bayern	Bayern	München	Freistaat	Wahlprogramm	CSU
44	Baden-Württemberg	Baden	Württemberg	Ministerpräsident	Stuttgart	bw
47	Bielefeld/ Koblenz	Bielefeld	Koblenz	Mainz	Rülke	Theurer
51	Hamburg/ Schleswig-Holstein	Schleswig	Holstein	Kiel	Rostock	Schwerin
54	Berlin	Berlin	Tempelhof	Lichtenberg	Schöneberg	Bezirk