The Political Dashboard:
A Tool for Online Political Transparency

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

Technical University of Munich

{juan.medina, orestis.p, morteza.shahrezaye}@tum.de, simon.hegelich@hfp.tum.de

Abstract

Contemporary political communication is a multi- and cross-platform process. Because of its complexity, new tools are necessary to monitor and understand it. We present a system that ingests, stores, and processes political data from Twitter, Facebook, and online news articles. We visualize the data in the form of a freely accessible online dashboard. *The political dashboard* (https://political-dashboard.com/) aims to provide online political transparency and assist researchers, journalists, and the general public in understanding the German online political landscape.

Introduction

The web and the datafication of society have transformed political communication. Not only does news consumption increasingly take place online, but individuals and politicians also use online social networks as platforms for political exchange. Under this framework, political campaigns have developed new campaigning techniques, such as political microtargeting (Hersh 2015; Papakyriakopoulos et al. 2018), while traditional gatekeeping has been replaced by complex processes of news media production and consumption (King, Schneer, and White 2017). This new form of political communication occurs in a political space that spans over multiple platforms. On the one hand, it is interconnected but on the other, it is difficult to monitor and analyze. To that end, we developed *the political dashboard*, a tool that monitors digital media outlets, Facebook, and Twitter, with the aim to provide an overview of online political activities in Germany. The dashboard contributes to filtering and understanding of political information, providing multi- and cross-platform transparency.

Data Collection

To monitor politically relevant data, we continuously collect data from different online sources. Our system consists of an array of Raspberry Pi devices that either connect to application programming interfaces (APIs) or employ crawling mechanisms to retrieve data. The collection procedure differs for each data source:

- **Twitter:** We collect tweets with the help of the Twitter Streaming API[1]. It allows us to retrieve data by providing a list of hashtags and users. We carefully select 239 relevant hashtags and 13,633 users, including accounts from political parties, politicians, media portals, journalists, bloggers, and other important political actors. We collect their tweets, mentions and retweets. For the hashtag list, we selected four types: German political parties, politicians, political topics, and media sites from all political orientations. We made an effort to avoid generating bias toward a specific political ideology through the data by carefully selecting a balanced list of hashtags and users that represented the complete German political spectrum. We are further aware that in the case of hashtags, Twitter only provides a sample of the complete tweets, which can make the data biased. However, we hope that by collecting a significant number of tweets, these insights are representative of political activity on Twitter.

- **Facebook:** We include two data sources from Facebook. First, we collect the posts from 102 public political pages; these correspond to the main page of the seven German political parties in Parliament and their regional pages from the 16 German states. We use the Crowdtangle[2] service to obtain the posts. The data do not include any personal data—neither the users who interacted with the posts nor their comments. Secondly, we collect political ads that target users in Germany on this platform. To do this, we connect to the Facebook ad archive API[3]. The archive has historical ads and active ads. We constantly update our database by collecting only the current, active ads.

- **News Outlets:** To retrieve online news media articles, we use RSS feeds of the news media websites and the Python package BeautifulSoup. We select 40 online German media sources from the top sites of online traffic in Germany (Alexa). We include media from all different political orientations and only collect the news articles that appear on the political sections of each news outlet.

---

[1]https://developer.twitter.com/en

[2]https://www.crowdtangle.com
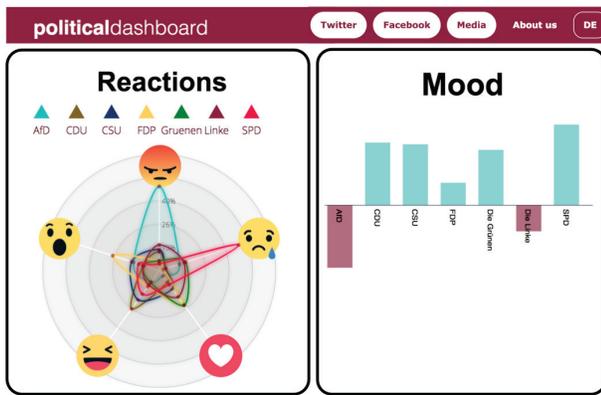
[3]https://www.facebook.com/ads/library/api

Figure 1: A screenshot of the political dashboard's Facebook page on the January 10, 2020.

After retrieval, the Raspberry Pi devices send the data to a distributed Elasticsearch database. The data is then processed with Python scripts using open source libraries. Finally, the servers send the analysis results to an application web server where live plots are created for the dashboard. The system implementation uses batch processing for the ingested data[4]. We designed the system to be able to adapt to the changing political discourse. Adding and deleting entries in MySQL tables (e.g., Twitter hashtags or users) will have a direct effect on the collection procedures.

Privacy Concerns

The constant and vast collection of data can raise privacy concerns. We neither display nor share the content of the collected tweets, Facebook posts, and news articles on the dashboard. We only provide aggregate information and analytical results. For the news articles, the full text is under strict data protection and should not be reproduced without the consent of the news media; therefore, we do not recreate the content in any way. Moreover, we do not provide individual data on the dashboard to ensure user privacy.

Description of the Dashboard

The dashboard gives a live overview of online German political trends—for example, partisanship activities, the popularity of content, and issue saliency. Although we originally conceived it as a system for internal research purposes (Serrano et al. 2018), we decided to create a front-end public tool with live analyses. Users can navigate between three web pages corresponding to each of the data sources. In the Twitter main page, we present the top hashtags and media URLs of the last 24 hours. They differ from Twitter's trending topics as we only focus on politically relevant tweets. We explicitly distinguish between biased hashtags, which are the hashtags we have pre-selected, and unbiased hashtags, which co-occur with the biased hashtags and were used by the users we follow or the users that interact with them.

---

[4]e.g., 1,834,953 tweets, 64 Facebook posts, 2,380 active ads and 214 news articles in one hour on January 10, 2020

Moreover, we show for each German political party the top hashtags used by *partisan* users. We define partisan users as those who have retweeted a political party account more than five times. A spider plot shows to which percentage the followers of each party are using the general top hashtags.

For the dashboard's Facebook page, we first show the number of posts published by the political parties in the last seven days. We then display the accumulated number of likes and shares per party. We plot the rest of the user reactions together in a spider plot as they are often in the same order of magnitude. We also apply a sentiment analysis algorithm to assign a mood score to each party (Figure 1 depicts a screenshot of the reactions and mood plot.) A second part of the Facebook page concentrates on advertising. It shows the advertisers that have more active ads and the advertisers whose ads generate the most user impressions. A map of Germany shows the percentage of the extent to which the seven political parties are targeting each state.

The third page focuses on online news. With the help of topic modeling, we process the texts and show the top seven topics, each represented by eight most important nouns per topic. The page also shows the top news articles shared on Facebook as a proxy of general online interest. We further categorize the news outlets according to their political orientation and use a spider plot to compare the proportion of articles that each media group publishes on the top topics.

The current implementation has two limitations. First, the design process focused on users interested in current online political activities; we did not design the system to allow retrieving historical data. However, we can add this functionality in the future as we store all results on our servers. Second, we are aware that the data could be biased and not replicate all online interactions. However, we made extensive efforts to minimize the bias and collect data from all political orientations. After monitoring the dashboard constantly for one year, we are confident that the results are reliable and helpful to understand the online political landscape in Germany.

Related Work and Impact

Few other websites collect online political interactions for public display. The WhatsApp monitor collects the most shared audiovisual content in WhatsApp public groups from Brazil, India, and Indonesia (Melo et al. 2019). It is part of the "Fake Elections" project, which has also developed a website that shows the number of likes and user demographics of politician's Facebook pages in Brazil (DCC 2018). Google and Facebook each provide search libraries to find political ads that were active on their platforms (Google 2019; Facebook 2019). The political dashboard stands out as it is a live monitor that shows processed analyses from three different sources. Our dashboard has already been an explicit information source to researchers, journalists, political candidates, and PR agencies. During the 2019 European elections it was used by the German Media Authorities. Since its creation, the dashboard has contributed to information extraction in multiple areas; for instance, in understanding the diffusion of right-wing and xenophobic content on social media platforms, as well as to understand partisan bias in media outlets.

References

DCC. 2018. Political Audience: Demographics of Brazilian politicians and parties on Facebook. *Eleiçoes Sem Fake*. Retrieved January 10, 2020 from http://www.audiencia-dos-politicos.dcc.ufmg.br/.

Facebook. 2019. Facebook ad libraries. Retrieved January 10, 2020 from https://www.facebook.com/ads/library/.

Google. 2019. Political advertising on google. Retrieved January 10, 2020 from https://transparencyreport.google.com/political-ads/home.

Hersh, E. D. 2015. *Hacking the electorate: How campaigns perceive voters*. Munich, Germany: Cambridge University Press.

King, G.; Schneer, B.; and White, A. 2017. How the news media activate public expression and influence national agendas. *Science* 358(6364):776–780.

Melo, P.; Messias, J.; Resende, G.; Garimella, K.; Almeida, J.; and Benevenuto, F. 2019. WhatsApp monitor: A fact-checking system for WhatsApp. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 13, 676–677.

Papakyriakopoulos, O.; Hegelich, S.; Shahrezaye, M.; and Serrano, J. C. M. 2018. Social media and microtargeting: Political data processing and the consequences for Germany. *Big Data & Society* 5(2).

Serrano, J. C. M.; Hegelich, S.; Shahrezaye, M.; and Papakyriakopoulos, O. 2018. *Social Media Report: The 2017 German Federal Elections*. Munich, Germany: TUM University Press.