Automated election audits - Analyzing Exposure to Political Content on Social Media with a Case Study of TikTok in Germany's 2024 Regional Elections

Tjaden, Jasper[1 2]; Wolfgram, Johannes[1]; Philipp, Aaron[1]; Weissmann, Sarah[1]; Bobzien, Licia[1]; Kohler, Ulrich[1]; Verwiebe, Roland[1]

[1] *Faculty of Economics and Social Sciences, Potsdam Social Media Monitor, University of Potsdam, Potsdam, Brandenburg, Germany*

[2] *corresponding author: jasper.tjaden@uni-potsdam.de*

Abstract

A main obstacle to the study of social media's impact on democracy is limited access to platform data. We present sock-puppet audits as a method for political scientists to collect social media data building on the algorithm audit literature in computer science and experimental approaches in the social sciences. Data is collected by scraping feeds of automated users (bots) of varied profile and behavioral characteristics. Our framework produces real-time data suitable for both quantitative and qualitative analysis without relying on platform-controlled APIs. First, we demonstrate the step-by-step execution of sock-puppet audits using pseudocode. Second, we apply the approach to a case study on political content exposure on TikTok during Germany's 2024 regional elections (N=230.000 videos). Within just three weeks on the platform, we find that new users without any political interest have a 3-4 times higher likelihood of being exposed to the right-wing populist party (AFD) relative to mainstream parties. We further discuss cross-platform opportunities for application of this approach as well as legal, ethical and technical challenges.

Keywords: social media, algorithm, audit, TikTok, scraping, election, communication, sock puppet

1. Introduction

One central emerging discussion among political scientists is the role of social media in the functioning of modern democracies (Howard and Kollanyi 2016; Metaxas and Mustafaraj 2012; Guess et al. 2023; Ferrara 2020; Stier et al. 2020; Jungherr, Rivero, and Gayo-Avello 2020). Some scholars are concerned that social media undermines the democratic process by spreading disinformation, fake news, and increasing polarization and mistrust (e.g.Ferrara 2020; Freelon and Wells 2020; Garimella et al. 2018). Others argue that the potential negative impact of social media is exaggerated, and that social media can even lead to an increase in access to useful political information (e.g. Nyhan et al. 2023; Scharkow et al. 2020).

One major obstacle for political scientists aiming to advance our understanding of this important question is limited access to social media data (Pasquale 2015; Freelon 2018). Despite the growing influence of digital platforms in political life, the private companies operating these platforms restrict data access, often for proprietary reasons, which limits meaningful scrutiny. When access is granted, it remains partial, heavily curated, and may not accurately reflect the platforms' influence on public opinion. This lack of transparency not only hinders scientific inquiry but also leaves the digital sphere largely unregulated and unaccountable to the public, raising urgent questions about corporate power and democratic accountability for the digital societies of the 21st century (McKay and Tenove 2021; Bennett and Livingston 2018).

In this paper, we present a framework for political scientists to gather comprehensive, user-centric, longitudinal data from social media platforms. This interdisciplinary framework integrates experimental research design with algorithmic audits, in particular, “sock

puppet audits" which have been applied in the field of computer science (Asplund et al. 2020; Bandy 2021; Lam et al. 2023; Srba et al. 2023). In essence, researchers create experimentally varied user accounts on social media platforms and monitor the content these accounts encounter. Accounts are automated to simulate real user behavior. Data is systematically extracted via web scraping.

We argue that sock-puppet audits have several advantages contributing to the study of political communication on social media: First, it shifts focus from studying inputs – i.e. what political actors do on social media (e.g. Beltran et al. 2020) – to studying outputs, i.e. what content reaches users (Chadwick 2017). Second, experimental variation of user behaviour allows further theory-driven hypothesis testing flexible to a myriad of research questions. Third, beyond quantitative analysis, sock-puppet audits can be used to collect text and video content for further qualitative (content/discourse) analysis across different platforms.

After briefly reviewing sock puppet audits in Section 2, Section 3 illustrates the application of the approach using pseudocode. We then apply sock-puppet audits in Section 4 to monitor political content exposure on TikTok during Germany's 2024 regional elections. Finally, Section 5 discusses the broader potential of this approach, particularly its ability to further unpack the inner workings of platform algorithms in determining exposure to political content, and close with a discussion on implementations challenges and overall limitations.

2. Sock puppet audits' potential for political science research

One dominant practice in the study of political communication on social media is focusing on the input dimension via hand-picking individual posts, topics (via hashtags) or accounts based on a priori selection (Cervi, Tejedor Calvo, and Blesa 2023; Grantham 2024; Moir 2023). This approach is well-suited to better understand the strategy, content and discourse produced by specific (political) actors. However, the approach is not able to measure the output dimension, i.e. the type and degree to which users are eventually exposed to this content. Data collection approaches able to capture output data include a) data donations and tracking, b) data sharing agreements, and c) the use of APIs. Data donations or tracking is costly because of incentivizing users to share their personal feeds. It also is subject to selection bias regarding (unobserved) characteristics of users who are willing to participate. Data sharing agreements are rare, and just like APIs, they are controlled by the platform companies. We provide a more detailed review of these approaches in Annex I, including reference examples from the field of political sciences.

Our sock-puppet audits are designed to measure the exposure to political content independent of platform control. The appraoch builds on the algorithm audit literature in the field of computer science. *Algorithm auditing*, or, more specifically *sock puppet auditing* systematically uses web scraping as a technique to assess algorithms for violations of laws (most famously, gender and race discrimination in job advertisements and recruitment) (Asplund et al. 2020; Bandy 2021; Hussein, Juneja, and Mitra 2020; Sandvig et al. 2014; Wachter, Mittelstadt, and Floridi 2017). Sock puppets refer to software that imper-

sonates addressees of an algorithm by creating user accounts or programmatically-constructed traffic (Sandvig et al. 2014, 13). Web scraping is then used to extract information which the fabricated accounts are exposed to. *Web scraping* pulls information from any element in the source text of a web page or social media feed (Luscombe, Dick, and Walby 2022; Speckmann 2021).[1] Scraping is flexible and customizable to any platform and research question. However, scraping is technically challenging, requires consistent maintenance and involves legal and ethical considerations (see last section for discussion).

A great advantage of sock puppet audits is their ability to implement experimental research designs by varying profile characteristics analogous to (field) experiments in social science. All elements of the "puppet" are held constant except for information in the user profile. To the best of our knowledge, sock puppet audits have not been applied to the study of political communication on social media.

We extend the "sock puppet" approach by also manipulating the bots' behavior (i.e. political interest) in addition to profile characteristics (age, sex). As the user variation is in the hands of the researcher, this feature opens the door to testing a wide range of theories for various research questions and presents a major advantage over observational data collected through APIs. It widens the application of audits from the field of (legal) discrimination to questions relevant to political scientists.

Bots are frequently linked to malicious activities such as artificially inflating traffic, spreading misinformation or extreme content (Srba et al. 2023; Haroon et al. 2023; Keller

[1] For example, Osmundsen et al. (2021) scraped the Twitter feeds of 2,300 users to analyze the determinants of fake news sharing (Osmundsen et al. 2021).

and Klinger 2019). Our approach, in contrast, demonstrates an ethical application of this technology for research purposes - without inflicting harm on real users.

3. Using sock puppet audits for measuring political exposure on social media

3.1. Framework

The framework for the data collection pipeline follows a series of steps, which we describe in detail below: 1) Defining the exposure(s) 2) Creation of accounts 3) Automating user behavior, 4) Extracting contents, 5) Enhancing data, 6) Analysis. Figure A1, A2 and A3 in the Supplementary Materials provide a graphical illustration of the overall bot deployment framework, logic and navigation procedure in addition to pseudocode included below.

1) Defining the exposure(s)

The key advantage of sock-puppet audits, compared to other approaches for collecting data from social media platforms, especially APIs, is that they enable experimental research designs (Haroon et al. 2023; Chen et al. 2021). Researchers may vary the characteristics of their users (bots) in terms of self-reported profile information (gender, age, profile picture, name, interests etc.) and user behavior (what they do when using the platform). The design of the bots depends on the research question and should be informed by relevant theories and literature.

Bot characteristics are analogous to the “treatment” or “exposure” variable in an experimental setting as researchers are interested in the (causal) effect of, for example, the gender of the user on content exposure. In our application, we also vary the specific behavior of users while holding constant other user characteristics such as gender or age. Examples of user behavior may include scrolling through the feed, engaging with content, creating content, using search features, and adjusting the duration and frequency of platform use. For example, varying user behavior allows political scientists to mimic different types of voters with varying interests and ideological leanings.

Location of the user is another important feature influencing content exposure which can also be varied by letting bots log on to the platform from different locations.

2) Creation of accounts

Some social media platforms (e.g., Facebook, LinkedIn) require a user account to access content, while others may not (e.g., TikTok, YouTube). Additionally, platforms tailor content for users based on their behavior, location, and profile information. Setting up user accounts allows researchers to control these variables guided by theory. Most platforms require an email address and a phone number.

3) Automating user behavior at scale

After bot design and creation, bot behavior is automated. Manual operation of the bot is time-consuming, resource-intensive and a potential bias for the research design given that it is difficult to keep behavior constant across bots when operated by humans. The automation process requires knowledge of programming languages, automation tools, and system deployment strategies. Researchers may utilize automation frameworks like the Selenium WebDriver for web applications, or Appium for mobile applications, which

allow for simulating user interactions within browsers or mobile interfaces. This enables bots to navigate complex websites and apps that rely heavily on JavaScript, dynamic content loading, or proprietary mobile interfaces where third-party services (e.g. Octoparse) often fail. Figure 1 shows pseudocode to illustrate how bots can be set up to simulate user behavior. By modularizing these functions, researchers can easily adjust the behavior to match different experimental conditions.

Figure 1: Pseudocode for user behavior simulation

User Behavior Simulation

Pseudocode

```
def simulate_user_behavior():
    initialize_user_profile()
    while session_active():
        content = fetch_current_content()
        if
        ↪ should_interact_with_content(content):
            interact_with_content(content)
        perform_navigation()
        wait_random_interval()
    end_session()

def initialize_user_profile():
    set_profile_attributes()
    set_behavioral_parameters()

def should_interact_with_content(content):
    return (content_matches_preferences(content)
    ↪ and
            within_interaction_limits())

def interact_with_content(content):
    action = choose_interaction_type(content)
    perform_action(action)

def perform_navigation():
    action = choose_navigation_action()
    execute_navigation(action)

def wait_random_interval():
    time.sleep(random.uniform(min_wait,
    ↪ max_wait))
```

Annotations:

`simulate_user_behavior()`:
Represents the main loop of the bots browsing session. It initializes the user profile and then continuously fetches content, decides whether to interact, performs interactions, navigates through the platform, and waits random intervals to mimic human behavior.

`initialize_user_profile()`:
Sets up the bots profile based on the experimental design, including attributes like age, gender, location, and behavioral parameters such as interaction frequencies and content preferences. This could be achieved via login into an existing account or creation of a new one.

`should_interact_with_content(content)`:
Determines whether the bot should interact with the current content. This decision is based on whether the content matches the bots predefined preferences and whether interaction limits (e.g., maximum minutes per session) have not been exceeded.

`interact_with_content(content)`:
Simulates user interactions such as liking, sharing, or other. The type of interaction is chosen based on the content's characteristics and the bots behavior profile.

`perform_navigation()`:
Simulates navigation actions, such as scrolling to the next piece of content, switching sections (e.g., from the home feed to the search page), or opening menus.

`wait_random_interval()`:
Introduces random delays between actions to mimic natural human browsing patterns, which helps avoid detection by bot detection algorithms.

Bot operation will likely need constant maintenance as platforms frequently implement changes. Maintaining the bots involves continuous monitoring of the platform for any changes in the site/app's structure, layout, or underlying code. Automated tests and alerting systems can be implemented to detect when the bots are corrupted (see section 5.2). For maintenance, containerization technologies are recommended (e.g. Docker). This allows researchers to package the bots into standardized units that run consistently across various computing environments. Containerization simplifies deployment, ensures consistency, and facilitates version control, making it easier to roll out updates during the data collection phase. It also allows for capturing location-specific content variations by deploying containers on servers in different geographical locations. Figure 2 shows pseudocode for managing the deployment of bots at scale.

Figure 2: Pseudocode for bot deployment at scale

Bot Deployment

Pseudocode

```
def deploy_bot_instances():
    deployment_schedule =
    ↪ create_deployment_schedule()
    for scheduled_time, bot_config in
    ↪ deployment_schedule:
        wait_until(scheduled_time)
        start_bot_instance(bot_config)


def create_deployment_schedule():
    schedule = []
    for bot in bot_list:
        scheduled_time = assign_start_time(bot)
        bot_config = configure_bot(bot)
        schedule.append((scheduled_time,
        ↪ bot_config))
    return schedule


def start_bot_instance(bot_config):
    container = launch_container(bot_config)
    monitor_bot(container)


def monitor_bot(container):
    while container.is_running():
        if has_exceeded_runtime(container):
            stop_container(container)
        else:
            check_container_health(container)
        time.sleep(monitoring_interval)
    collect_results(container)
    remove_container(container)
```

Annotations:

`deploy_bot_instances()`:
Iterates over a deployment schedule, launching each bot at its assigned time with specific configurations.

`create_deployment_schedule()`:
Assigns start times and configurations based on the experimental design. It ensures that bots start at appropriate times to simulate different user behaviors and usage patterns.

`start_bot_instance(bot_config)`:
Launches a bot instance within a container using the specified configuration.

`monitor_bot(container)`:
Continuously checks the bot during its session. If the bot exceeds its runtime limit or encounters issues, it is stopped. Monitoring ensures that resources are used efficiently and that bots operate within defined parameters.

Content on social media can be strongly tailored to the location from where users log in (Casero-Ripollés, Micó-Sanz, and Díez-Bosch 2020). Therefore, if the research question requires accounts to appear as if they're in specific locations, researchers have several options. They can achieve geographic diversity by utilizing virtual private networks (VPNs) or proxies to simulate different geographical locations without physically relocating resources. Alternatively, researchers can deploy actual computational resources in target regions using distributed servers or deploying mini-servers (e.g. Raspberry Pi units).

4) Extracting information

While the bots are navigating the social media platform, researchers employ a scraper to extract the desired information from the website. Building a scraper varies significantly in terms of difficulty depending on the platform's complexity, and the level of bot detection employed by the site/app. Figure 3 shows pseudocode illustrating the data extraction process.

Figure 3: Pseudocode for data extraction

Data Extraction

Pseudocode

```
def extract_content_data():
    while session_active():
        content = get_current_content()
        data = extract_relevant_data(content)
        store_data(data)
        perform_navigation()
        wait_random_interval()
    end_session()


def extract_relevant_data(content):
    data = {
        'content_id': content.id,
        'timestamp': get_current_timestamp(),
        'content_type': content.type,
        'author': content.author,
        'engagement_metrics': {
            'likes': content.likes,
            'shares': content.shares,
            'comments': content.comments
        },
        'hashtags': content.hashtags,
        'metadata': content.metadata

        # Add more variables if applicable
    }
    return data


def store_data(data):
    database.insert(data)
```

Annotations:

`extract_content_data()`:
Represents the data extraction process during the bots active session. As the bot navigates the platform, it extracts data from each piece of content encountered.

`extract_relevant_data(content)`:
Specifies the data fields to be extracted from the content. This may include identifiers, timestamps, content types (e.g., video, image, text), author information, engagement metrics (likes, shares, comments), hashtags, and any other metadata available.

`store_data(data)`:
Handles the storage of the extracted data.

5) Data analysis

Numerical information (views, shares, comments etc.) can be used to assess the performance of video content on the platform. Text data can be used for automated text analysis (for example, sentiment analysis) (Liu 2012) or classification algorithms (for example, detecting specific types of content). Text and audio-visual data can also be used for qualitative content, narrative or discourse analysis (Rahmat et al. 2024; Pérez-Escoda et al. 2021; Yadlin-Segal and Oppenheim 2021). It is important that researchers clarify their research question in detail prior to data collection. Once sock-puppet audits are set up, data can only be generated prospectively, and it is difficult to recuperate missing

variables after launch. We offer further reflection on analytical potential of this approach in section 5.2.

4. Application: Exposure to political content on TikTok in the context of German regional elections in 2024

4.1. Context

In the following, we demonstrate the steps outlined in the previous section using a case study about political content on TikTok ahead of three regional elections in Germany in 2024. This research is motivated by the significant rise in support for the right-wing populist party, Alternative for Germany (AfD), among young voters in various elections in 2024. Notably, in the 2024 European elections in Germany, the AfD's vote share among 16- to 24-year-olds tripled compared to the 2019 election, a result that surprised many political observers.

The research question is how much new, politically neutral, young voters are exposed to political content of right-wing populist parties without actively searching for it. Following the idea that digital platforms tend to reinforce existing social inequalities, we hypothesize that exposure to digital content is influenced by socio-structural variables affecting voting behavior (Kulachai, Lerdtomornsakul, and Homyamyen 2023). We thus expect: (1) Higher exposure to non-traditional party content among young users due to weaker traditional party loyalties, (2) greater exposure to right-wing populist content among male users, reflecting offline voting patterns, and (3) increased political content exposure for politically active users through algorithm recognition of engagement patterns. Taking a

dynamic perspective and following the theoretical idea that early engagement with political content can reinforce exposure to (similar) political content (Bakshy, Messing, and Adamic 2015; Garrett 2009), we further hypothesize that exposure to political content, and therefore being politically active, will increase general political exposure over time (4).

Social media has become one of the primary sources through which young people stay informed about current events. TikTok, in particular, is highly popular among younger audiences, with 64% of Germans under the age of 24 reporting usage of the platform within the last four weeks, spending an average of 43 minutes per day on it (Bobzien et al. 2023). Additionally, 40% of young adults in the U.S. and 30% in Germany cite accessing news as one of their reasons for using TikTok (mpfs 2023; Shearer et al. 2024). Given these trends, TikTok may play a significant role in shaping the political views and opinions of young people.

4.2. Data collection

Between 13/08/2024 and 06/10/2024, we deployed 34 bots on TikTok – 7 in Saxony, 8 in Thuringia and 19 in Brandenburg. Bots were online for, on average, 49 days (min= 22, max=55). We varied the accounts regarding age (17-23 years old vs. 44 years old), gender (male, female, diverse, not specified), residence (Saxony, Thuringia, Brandenburg) and user behavior (political interest vs. no active political interest). The implementation of different age groups and gender was achieved through profile information (self-reported pronouns, name of person, and date of birth). The residence was varied by using servers in different geo-locations in Saxony, Thuringia, and Brandenburg through which the bots log into their accounts.

Bots were programmed to scroll through their "For You page" until a video with hashtags relating to their interest (see Table 1) appeared. When it did, the user watches the video for its full length, but no longer than 2 minutes and liked it. To ensure balance in the interest of the bots over time, we "nudged" the bots once per session, after 35–45 videos, towards one of their assigned interests at random. We refer to nudging when bots searched for a specific hashtag in the top search bar rather than simply scrolling through their feed. The combination of feed with occasional nudges avoids pushing bots into a specific direction too early while ensuring variation in exposure across user behavior types. In the case of users with an interest in elections, videos with political content are watched in full but not liked. By this, we stipulate that politically interested users may watch videos with opposing ideological viewpoints merely out of interest, but not because they support the contents.[2]

Table 1: Hashtags to control search behavior of bots

Non-political hash-tags		Political hashtags	
German	English	German	English
#freunde	#friends	#landagswahlen2024	#stateelections20204
#fürdich	#foryou	#wahlen2024	#elections2024
#hunde	#dogs		
#ideen	#ideas		
#kochen	#cooking		
#lustig	#comic		
#reisen	#travelling		
#witzig	#funny		

Note: All bots were programmed to occasionally actively search for content with non-political hashtags such as videos relating to funny videos, travel and cooking. 18 of the 34 accounts, *additionally,* actively

[2] The bots did, however, like content including political parties when they found it indirectly and through other interests, for example when a political hashtag was included with other hashtags of interest. This emulated, to a degree, the natural user behavior and gave us a rough measurement of the increase in party content a bot gets when it starts to care about a party.

searched for content related to the elections without looking for any partisan or ideological content revealing a political opinion; see Table A1 and A2 in the Annex for the complete list of hashtags.

Bots scrolled through the suggested videos for approximately an hour each day without engaging with other users. The accounts did not leave comments, did not follow other accounts or publish their own content. Hence, the accounts were not supposed to exert any influence on the TikTok ecosystem and only act as passive observers.

During the study period, a total of 229,807 videos were collected – 44,349 in Saxony, 58,469 in Thuringia, and 126,989 in Brandenburg. The data collection included all videos that were shown to the bots by the platform, including engagement metrics (number of likes and comments) video description, video creator, used hashtags and music. Additionally, metadata of political videos was collected using an open-source solution including more engagement metrics (views, shares, saves) and time and location of the upload (Freelon et al. 2024).

Table 2 illustrates the structure of the data created from the bot. The data is best described as a hierarchical data set with video events nested in sessions and bots. Or, stated differently, for each bot we have observations on each video presented to them in their daily sessions.

Table 2: Structure of the dataset

Bot-ID	Date	Time	VideoID	Variables
1	20240901	10:21	A1b2c3	…
1	20240901	10:22	Q6r7s8	…
:	:	:	:	:
1	20240902	12:53	G2h314	:
:	:	:	:	:
2	20240901	16:23		:
:	:	:	:	:

The variables obtained to describe each video are shown in Table 3, including an anonymized real-world example for one specific unit. The entry in the variable “description” was translated into English.

Table 3 Variables obtained

Variable	Example
Bot ID	wm_mp
Author	creator_lifestyle_XX
Description	Today, creator_lifestyle_XX shows you 5 ideas for your kitchen that make sense! Which is your favorite? #küchenideen
Hashtags	küchenideen
Likes	not shown for data protection reasons
Comments	not shown for data protection reasons
Saves	1850
Views	425000
Watched	True
Liked	False
Saved	False
Is nudged	False
Duration (s)	180
Create Time	not shown for data protection reasons
Location	AT
Verified	False
Advertisement	False
Timestamp	not shown for data protection reasons

The data is used to generate two central variables for our application: Exposure to Official Party Accounts (OPAs) and exposure to Party-Affiliated Content (PAC). OPA is defined by the number of times a video appears in the user feed which was posted by an official party representative (including official party accounts and the accounts of official candidates of that party). PAC is defined by the number of times a video appears which was linked to a party via party-specific hashtags (i.e. "#AFD") (see Annex II, Table A1 and Table A2 for a full list of OPAs and PAC hashtags).

Table 4 shows a summary of the bot activity within our observation window (55 days). Bots watched approximately the same number of videos, although the total number of videos varied quite substantially.

Table 4: Summary of bot activities

characteristic		B. Total number of exposed videos	C. Total number of watched videos	D. Total number of liked videos	E. Total number of exposed videos from OPAs	F. Total number of exposed videos containing PAC
Age	young	178,795	51,714	39,463	330	2,679
	old	51,012	14,776	11,259	74	689
Gender	divers	54,129	15,186	11,502	94	769
	female	57,416	16,793	12,802	86	806
	male	61,985	18,280	14,115	123	1,106
	none	56,277	16,231	12,303	101	687
Political Interest	yes	121,548	34,793	26,443	327	2,571
	no	108,259	31,697	24,279	77	797
State	Saxony	44,349	12,468	9,561	85	632
	Thuringia	58,469	16,478	12,522	111	910
	Brandenburg	126,989	37,544	28,639	208	1,826
total		229,807	66,490	50,722	404	3368

Note: Data collected by the authors Sep/Oct 2024. Table reports aggregate statistics of all created bots by bot characteristic, behavior and state. Column E: OPAs stands for Official Party Accounts and includes accounts linked to parties at the federal level and the regional level, the respective youth organizations of the parties and the top 10 election candidates of each party in each state, given they were TikTok members. Column F: PAC includes videos involving party-specific hashtags regardless of who uploaded the video (official and unofficial accounts).

4.3. Results

In this section, we report results from the data analysis. The purpose is to demonstrate examples of how this data can be used in various ways.

4.3.1. Exposure to Official Party Accounts (OPA)

On average, across all parties and all users (bots), the OPA exposure rate was 0.18% (0.0018), equivalent to 1.8 videos by political parties for every 1.000 videos which appeared in the feed. Standardized by the time spent online, the rate indicates that the average user in our study saw a video from a political party every 516 minutes of scrolling through their feed.

There is large heterogeneity in the exposure rates across political parties with higher exposure rates for the right-wing *Alternative fuer Deutschland* (AFD) [5.46 videos per 10.000] and the left-wing party *Buendnis Sahra Wagenknecht* (BSW) [5.30] than for the center left party (*Sozialdemokratische Partei Deutschlands* (SPD) [1.32] and center-right party *Christlich Demokratische Union* (CDU) [1.57], the far-left party *Die Linke [0.79]*, the greens *Buendnis 90 - Die Gruenen* [3.51] and liberal party *Freiheitlich-Demokratische Partei* (FDP) [0]. The overall exposure risk is low given that a) our users are new to the platform, b) have no clear political leaning, and c) only consume videos for one hour a day. However, the relative differences in exposure are striking: The average, new, politically neutral, user in our study is 3 to 4 times more likely to be exposed to right-wing AFD content compared to centrist parties like the SPD and CDU.

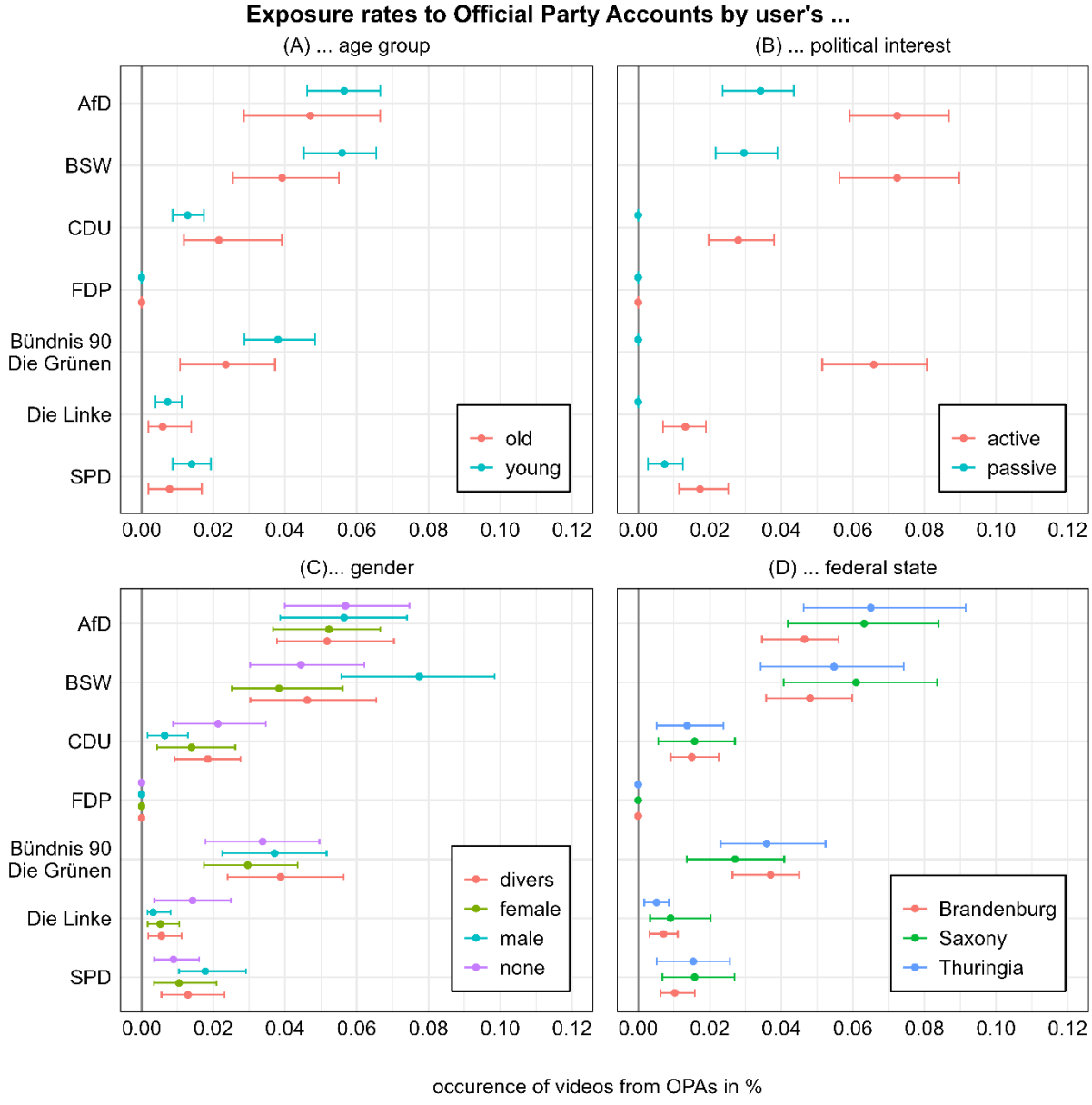
In 39% of cases, the first video by a political party was by the AFD followed by 24% by BSW, 21% by Greens and only 9% by CDU and 6% by Die Linke. SPD and FDP never appeared first. The average user was more than 5 times more likely to encounter AFD in their feeds before CDU, die Linke, SPD and FDP.

Figure 4 shows the breakdown of the exposure rate by the treated sock puppets' characteristics age, behavior (political interest or not), gender and state. These effects can be interpreted as causal given that assignment of the characteristics was exogenous, and all features of the bots and the platform environment remained constant.

Exposure varies substantially by political interest. Active bots are more exposed to OPA compared to passive bots, with exposure rates to OPA for AFD party at 0.072% for active bots versus 0.029% for passive bots. Passive bots are more likely to be exposed to AFD than active bots are to other parties (with the exception of BSW). Passive bots without any interest in politics only receive content from far-right or far-left parties.

We also see a tendency that OPA exposure is higher among younger users albeit heterogenous across parties: differences between young and old users in OPA exposure are higher for AfD and BSW than for other parties. Our analysis also reveals limited gender differences and regional variation across the three federal states of Brandenburg, Sachsen, and Thuringia.

Figure 4: Exposure rates for Official Party Accounts by user characteristics



Note: Data collected by the authors Sep/Oct 2024 (N=229,807). 95%-confidence intervals were calculated using bootstrapping.

4.3.2. Exposure to Party-affiliated content

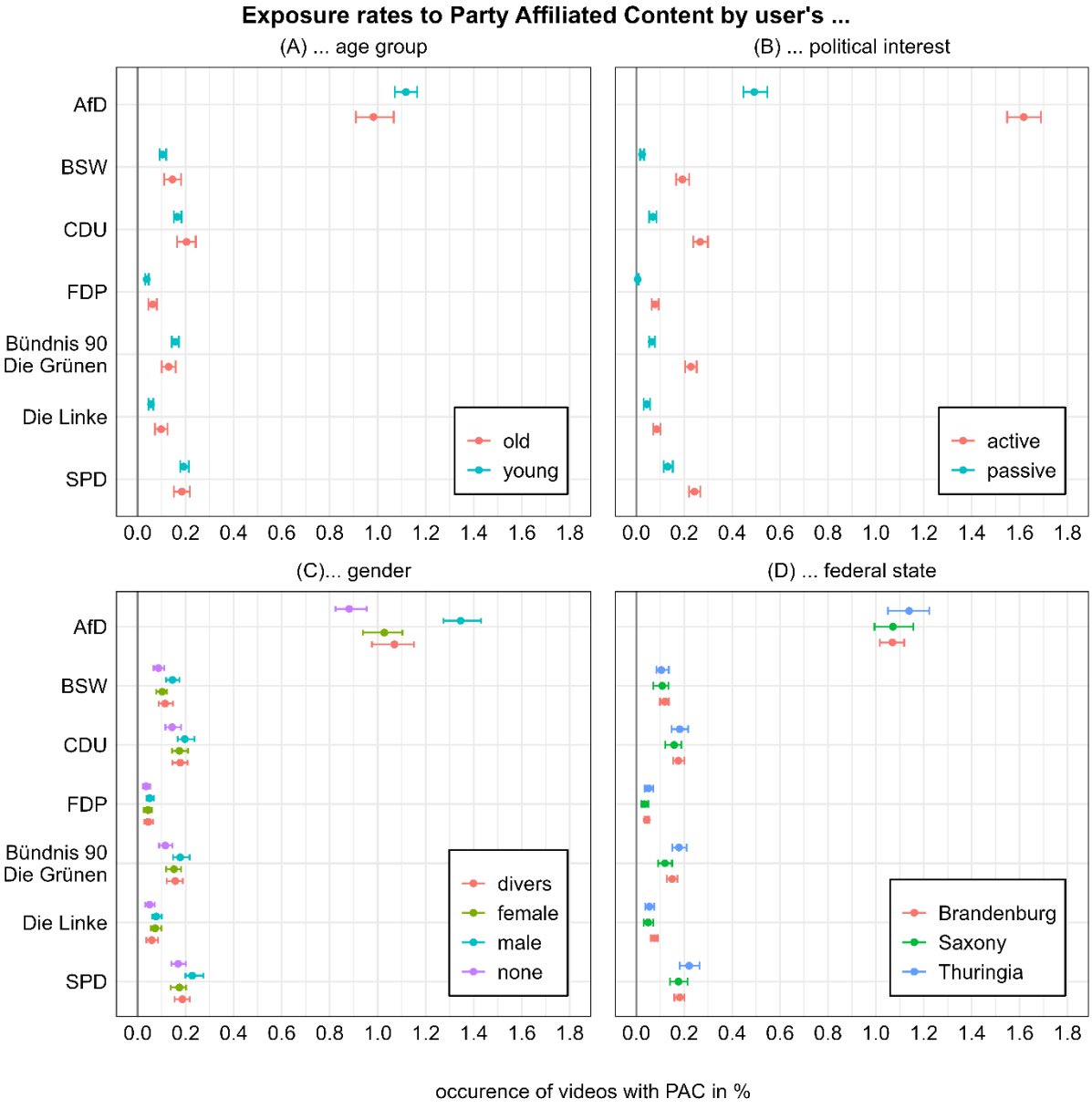
The overall rate of exposure to party-affiliated content, i.e. any videos with a party-affiliated hashtag, is 0.014 (14 videos per 1000). The risk of being exposed to political content which is not spread by official political parties is 8 times higher than content produced by official party accounts. Standardized by the time spent online, the rate indicates that the average user in our study saw a video with political content affiliated to a party every 41 minutes of scrolling through their feed.

Exposure rates vary substantially across parties. The AFD has by far the highest PAC exposure rate [110 videos per 10.000]. All other parties have rates below 20 per 10.000 videos. Our new, politically neutral users in the study are exposed to 6 times more content by AFD than CDU or SPD. In 39% of cases, the first political video which appeared in the feed was affiliated with the AFD followed by 27% for die Linke, 21% for SPD, 9% for Greens, and 3% for BSW.

Figure 5 shows the exposure rates for party-affiliated content by age group, political interest, gender and federal state. We find little variation between age groups, gender, and regions for all parties except the AFD. For the AFD, we find a modest, positive effect for male (vs. female) and younger (vs. older) users on PAC exposure. For all parties, we find significant differences between politically interested (active) and politically not interested (passive) users. For example, the exposure rates for active users are 2-8 times higher compared to passive user (e.g. SPD = 0.13% vs. 0.24%; CDU = 0.07% vs. 0.27%; Greens =0.06% vs. 0.23%, AFD=0.49% vs 1.6%). Strikingly, passive users, who are not searching for any political information, received, on average, more AFD affiliated content than users did who actively search information from any other party.

This large gap in the PAC exposure rate between AFD and other parties suggests that right-wing political content is more popular and more widely available on the platform compared to content affiliated with moderate or even far-left parties. While far-left party BSW was almost as successful as the far-right party AFD in reaching users with their official content, AFD reaches far more users through unofficial sources.

Figure 5: Exposure rates to Party Affiliated Content by user characteristics



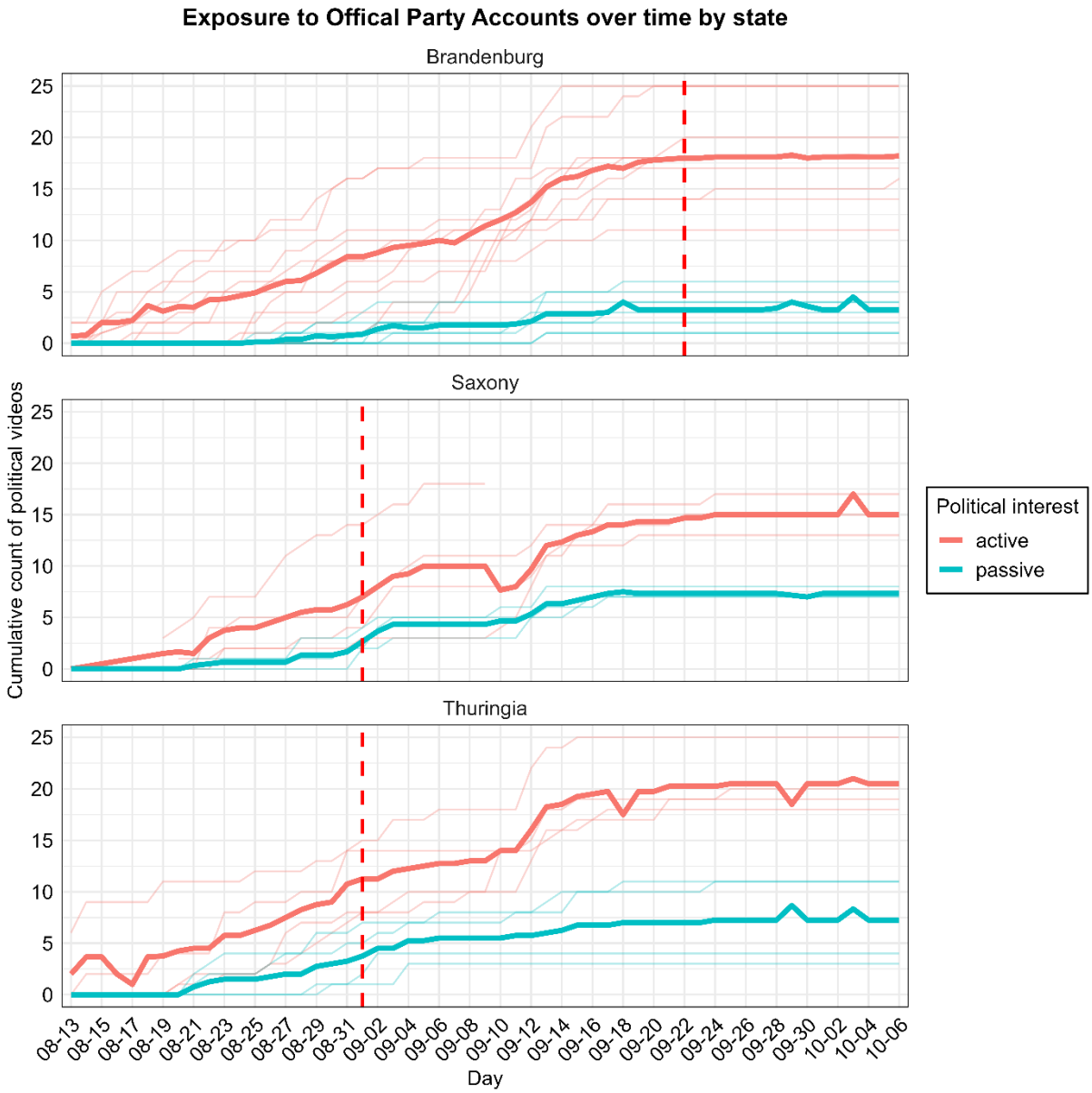
Note: Data collected by the authors Sep/Oct 2024 (N=229,807). 95%-confidence intervals were calculated using bootstrapping.

4.3.3. Exposure to political content over time by user characteristics

The exposure to OPA increases over time while differences in exposure between active and passive users as well as different parties increasingly diverge over time. In line with our hypothesis, the algorithm responds faster to users who reveal political interest. This suggests that early engagement with political content - which has a higher likelihood to be right-wing content – reinforces exposure to right-wing party content despite initial neutral stance.

Figure 6 also shows the day of the respective election in each state. No clear differences in terms of exposure trends depending on the timing of the election can be observed.

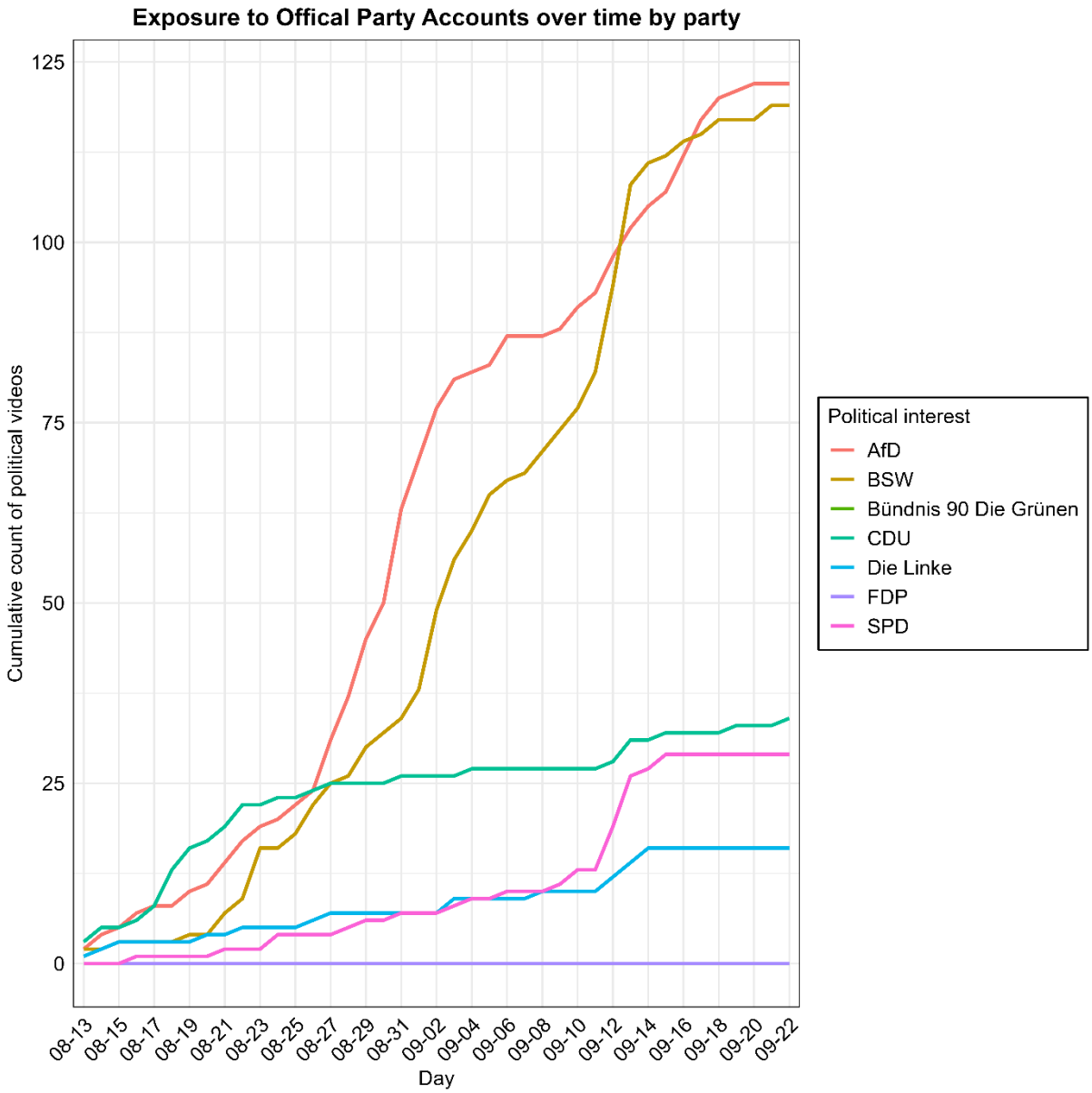
Figure 6: Cumulate exposure to Official Party Accounts by day



Note: Data collected by the authors Sep/Oct 2024 (N=229,807). Drops in the average cumulative exposure (think lines) are due to temporary drop-out of individual bots (thin lines) as a result of platform changes. After adjustments, bots resumed after 1-2 working days. The dotted line represents the date of the regional election. Bots with "active" political interests searched for information about the regional election. Bots with "passive" political interest did not search information about the election. Political search terms (see Table 1) did not reveal any ideological or political preference.

Figure 7 shows the difference in cumulative exposure to official party accounts by the respective political party. The graph shows that the AFD and BSW quickly gain momentum and follow a steeper trajectory compared to other parties.

Figure 7: Cumulate exposure to Official Party Accounts by party



Note: Data collected by the authors Sep/Oct 2024 (N=229,807).

5. Discussion

5.1. Research potential

We presented an application of sock-puppet audits to study exposure to (far-right) political content in the context of elections. We showed how this approach can enable researchers to study causal effects of user characteristics such as age, gender, and

political interest on social media exposure. In this section, we argue that the general framework provides the opportunity to be applied to many research questions due to its flexibility. Relying exclusively on official (often paid) APIs provided by companies limits the research potential as companies severely limit the information available through APIs or even decide to close it (see Freelon 2018). Data donation and data tracking require expensive incentives for users and suffer from sample selection (Boeschoten et al. 2022; Breuer, Bishop, and Kinder-Kurlanda 2020). Sock-puppet audits allow to collects data that are entirely customizable to the need of researchers in terms of scope, observation periods, location, measurements etc. This method is, in principle, applicable to any social media platform and adaptable to different research questions.

We see three general areas of research where sock-puppet audits can yield valuable insights:

5.1.1. User behavior

In our application, we varied whether users were "politically interested" (i.e. actively searching for information about the election) or not. User behavior can be manipulated in any way. Users could vary according to different markers revealing their political leanings. Users could follow different news outlets. Users could also vary by the time they are online. Depending on the research question, the number of variations in the bot behavior is only limited by the programming skills of the researcher and the theory that informs them.

5.1.2. Longitudinal analysis

The longitudinal data (illustrated in Figure 6 and Figure 7) enables unique ways to study how and how quickly prior user behavior affects later content curation. This angle is

particularly interesting when comparing exposure across various platforms as “filter bubbles” may form at different speeds. The longitudinal dimension also offers the potential to understand how users who start out with no ideological leanings (in terms of their revealed interest on the platform) can enter ecosystems that are dominated by either far-left or far-right content, and presumably add to polarization.

5.1.3. Mechanisms and algorithm bias

Platform algorithms are well-kept business secrets. As a result, the public and researchers do not fully understand why certain content is displayed. Commonly proposed factors determining content exposure include the number of interactions (clicks, likes, shares, comments), the level of activity on the platform, third-party accounts and multipliers, specific content strategies (emotional, funny, short, simple). Given the number and complexity of potential influencing factors, researchers struggle to isolate what drives exposure to social media content – especially when platforms tailor content exposure to each individual user depending on their platform behavior or account characteristics. The presented sock-puppet audits are helpful in this context because, by design, they eliminate a host of potential “confounders” relating to user behavior, user characteristics, other user input to the platform, and current events, as these factors are constant across bots or, if intended, vary across bots in ways designed by researchers. This approach allows to focus attention on the residual variation in exposure which is unrelated to individual user behavior and, as a result, likely due to the popularity of the content on the platform (observable), the behavior of other users (partially observable), off-platform behavior (largely unobserved) and other remaining unknown features of the algorithm (unobserved).

5.2. Implementation challenges

While appreciating its potential (see 5.1.), it must be acknowledged that implementing the sock-puppet audits presents several challenges involving mainly bot detection, platform changes and legal and ethical issues.

5.2.1. Bot creation and detection

Creating accounts at scale is a significant hurdle. Platforms often have measures in place to detect and prevent the creation of inauthentic or suspicious accounts. Automated account creation may rely on tools from unreliable sources, which can pose legal and ethical concerns. Manual account creation, while more reliable, is time-consuming and resource-intensive.

Once created, social media platforms employ bot detection algorithms to detect automated users to varying degrees of intensity. The use of bots violates some of the terms of references of certain platforms (e.g. Facebook, Instagram), while others do not prohibit bot activity in principle (Twitter, TikTok) (Stricklin and McBride 2020).

Three common ways in which platforms try to identify non-human users include Completely Automated Public Turing test to tell Computers and Humans Apart (CAPTCHA), bot detection algorithms; and rate limits.[3]

[3] CAPTCHA is an effective tool for preventing bots from accessing web services and ensuring that web services are accessed by humans. A common application of CAPTCHAs is selecting images showing small parts of a larger object from a larger set of images. This is an easy task for humans but difficult for bots to perform. Yet, available Software can be included in Selenium to solve CAPTCHA's with high success rates. Bot detection uses statistical models to predict whether a user is a real human or a bot. The models are

In addition to behavioral patterns, platforms may also use IP address analysis as part of their bot detection strategies. Datacenter IP addresses, often associated with cloud services (i.e. AWS, Google Cloud, Azure) or VPNs, may be flagged as suspicious. To mitigate this, researchers can use residential or mobile IP addresses, which appear more authentic to platforms. However, this approach comes with additional costs and ethical considerations, particularly regarding the sourcing of these IPs. While proxies and VPNs offer versatility in IP management, they may be impractical for high data volume tasks (e.g. video streaming on TikTok) due to bandwidth limitations and associated costs.

5.2.2. Dynamic platform changes

As mentioned above, social media platforms frequently update their interfaces and functionalities, which can render existing automation scripts ineffective. Researchers must be prepared to adjust their bots promptly in response to such changes to ensure continuity in data collection.

By utilizing containerization and flexible configuration management, updates to the scraper can be rolled out efficiently across all bots. Since the bots retrieve their configurations and instructions from a centralized database, adjustments to scraping strategies or data extraction targets can be implemented without redeploying the entire bot infrastructure. This ensures that data collection remains consistent and adapts swiftly to platform changes.

---

applied to analyze patterns of interaction of users with the website. Repetitive patterns or few pauses between actions, for example, are more characteristic of bots compared to humans. Rate limits regulate the number of requests that can be sent to a server in a particular time window. Websites can easily identify bots when the request coming from one IP address (i.e. one computer) exceeds the rate limit. To minimize the risk of bot detection, researchers can design bots to mimic human behavior as closely as possible. This includes incorporating random delays between actions, varying the sequence of interactions, and simulating realistic session durations and activity patterns. Avoiding excessive or unnatural interaction rates can help prevent triggering automated detection systems.

5.2.3. Legal and ethical considerations

Sock-puppet audits pose questions regarding legal and ethical standards. The main legal concerns are 1) violating platform's terms of service, and 2) infringing upon data protection of other users on the platform.

The current thinking on both issues may vary substantially depending on the country, the purpose of data collection, the platform, and is fast evolving (DeVito, Richards, and Inglesby 2020).

In the United States, a few court cases have ruled that massive data extraction causing commercial damage is illegal (Dilmegani 2024). Scraping for research purposes, however, rather than for commercial gain, is generally seen as less problematic. Some countries explicitly allow the scraping of data for research purposes, although companies prohibit the scraping of their own data in their terms of service (Sebastian Golla and Müller 2020). Publicly available data are generally not a concern as long as they are not used for commercial gain. For example, in a famous recent example, a US judge sided with a research team which scraped Twitter data for the purpose of identifying hate speech (Fung 2024).

Data protection issues arise when scraping sensitive personally identifying information (PII) such as name, email, postal address, etc., especially when information can be used to harm individuals or vulnerable groups (Salah, Canca, and Bariş 2022). In the European Union, for example, scraping PII without "legitimate interest" can potentially raise concerns related to compliance with the General Data Protection Regulation (GDPR) (Szwed 2021). Scrapers typically collect information without the knowledge and consent of indi-

viduals including their account names, posts and comments. Standard research practices, such as anonymization and aggregation, should be applied in these cases. One option to mitigate ethical concerns is debriefing after the execution of an experiment, i.e. revealing which bots did not represent real humans. In our application, we only further processed posts by official political actors whose communication is of public interest. In any case, we recommend consulting formal ethics review committee before launching a study involving social media data collection.[4]

6. Conclusion

This paper introduced sock-puppet audits as a method for political scientists to measure social media exposure to political content. The approach automates users on social media and systematically extracts the content visible to them. Users (bots) vary experimentally by profile characteristics and behavior. We presented all steps involved in executing the sock-puppet audits and showcased an application of this approach on regional elections in Germany in 2024. We showed that new, neutral users have a 3-4 times higher likelihood of being exposed to right-wing content compared to "mainstream" parties and that these gaps widen for politically interested users.

Despite technical demands and a fast-evolving legal context, we argue that this approach offers large potential for political science scholars to study political communication on social media. Sock-puppet audits are generally scalable to any platform and different research questions. In a post-API age (Freelon 2018), where large companies further reduce access to social media data, this approach provides researchers a tool to assess

[4] We obtained ethics approval by the University of Potsdam ethics board (decision 180/2023).

the potentially harmful impact of platforms on voters. We encourage others to apply this approach to help build a more solid evidence base for studying the role of social media in increasingly digitalized democracies.

Data Availability Statement

Replication code for the analysis is provided as part of the submission and will be uploaded online upon acceptance of the paper.

Competing Interests

The authors do not report any competing interests

Funding

The project was funded by the University of Potsdam without relying on third-party funding.

References

Asplund, Joshua, Motahhare Eslami, Hari Sundaram, Christian Sandvig, and Karrie Karahalios. 2020. "Auditing Race and Gender Discrimination in Online Housing Markets." *Proceedings of the International AAAI Conference on Web and Social Media* 14 (1):24-35. doi: 10.1609/icwsm.v14i1.7276.

Bakshy, Eytan, Solomon Messing, and Lada A. Adamic. 2015. "Exposure to ideologically diverse news and opinion on Facebook." *Science* 348 (6239):1130-1132. doi: doi:10.1126/science.aaa1160.

Bandy, Jack. 2021. "Problematic Machine Behavior: A Systematic Literature Review of Algorithm Audits." *Proc. ACM Hum.-Comput. Interact.* 5 (CSCW1):Article 74. doi: 10.1145/3449148.

Beltran, Javier, Aina Gallego, Alba Huidobro, Enrique Romero, and Lluís Padró. 2020. "Male and female politicians on Twitter: A machine learning approach." *European Journal of Political Research* 60. doi: 10.1111/1475-6765.12392.

Bennett, W. Lance, and Steven Livingston. 2018. "The disinformation order: Disruptive communication and the decline of democratic institutions." *European Journal of Communication* 33 (2):122-139. doi: 10.1177/0267323118760317.

Bobzien, Licia, Fabian Kalleitner, Ulrich Kohler, and Roland Verwiebe. 2023. "Upcoming Modules: Digital platforms and life satisfaction." https://companion-is.soep.de/Innovative%20Modules/2023/Digital%20platforms%20and%20life%20satisfaction.html.

Boeschoten, Laura, Jef Ausloos, Judith E. Möller, Theo Araujo, and Daniel L. Oberski. 2022. "A framework for privacy preserving digital trace data collection through data donation." *Computational Communication Research* 4 (2):388-423. doi: 10.5117/CCR2022.2.002.BOES.

Breuer, Johannes, Libby Bishop, and Katharina Kinder-Kurlanda. 2020. "The practical and ethical challenges in acquiring and sharing digital trace data: Negotiating public-private partnerships." *New Media & Society* 22 (11):2058-2080. doi: 10.1177/1461444820924622.

Casero-Ripollés, Andreu, Josep-Lluís Micó-Sanz, and Míriam Díez-Bosch. 2020. "Digital Public Sphere and Geography: The Influence of Physical Location on Twitter's Political Conversation." *2020* 8 (4):11. doi: 10.17645/mac.v8i4.3145.

Cervi, Laura, Santiago Tejedor Calvo, and Fernando Blesa. 2023. "TikTok and Political Communication: The Latest Frontier of Politainment? A Case Study." *Media and Communication* 11. doi: 10.17645/mac.v11i2.6390.

Chadwick, Andrew. 2017. *The Hybrid Media System: Politics and Power*: Oxford University Press.

Chen, Wen, Diogo Pacheco, Kai-Cheng Yang, and Filippo Menczer. 2021. "Neutral bots probe political bias on social media." *Nature Communications* 12 (1):5580. doi: 10.1038/s41467-021-25738-6.

DeVito, N. J., G. C. Richards, and P. Inglesby. 2020. "How we learnt to stop worrying and love web scraping." *Nature* 585 (7826):621-622. doi: 10.1038/d41586-020-02558-0.

Dilmegani, Cem. 2024. "Is Web Scraping Legal? Ethical Web Scraping Guide." AI Multiple Research. https://research.aimultiple.com/web-scraping-ethics/.

Ferrara, Emilio. 2020. "Bots, Elections, and Social Media: A Brief Overview." In *Disinformation, Misinformation, and Fake News in Social Media: Emerging Research Challenges and Opportunities*, edited by Kai Shu, Suhang Wang, Dongwon Lee and Huan Liu, 95-114. Cham: Springer International Publishing.

Freelon, Deen. 2018. "Computational Research in the Post-API Age." *Political Communication* 35 (4):665-668. doi: 10.1080/10584609.2018.1477506.

Freelon, Deen, Philip Kreißel, Parker Bach, Timo Bäuerle, Christina Walker, Temesgen Mesfin Tewolde, Dan Phiffer, Billy Beniar, and Thomas Ruizt. 2024. Pyktok. GitHub.

Freelon, Deen, and Chris Wells. 2020. "Disinformation as Political Communication." *Political Communication* 37 (2):145-156. doi: 10.1080/10584609.2020.1723755.

Fung, Brian. 2024. "Judge tosses Elon Musk's case against hate speech watchdog in excoriating rebuke." *CNN* https://edition.cnn.com/2024/03/25/tech/judge-tosses-elon-musks-case-against-hate-speech-watchdog/index.html.

Garimella, Kiran, Gianmarco De Francisci Morales, Aristides Gionis, and Michael Mathioudakis. 2018. "Political Discourse on Social Media: Echo Chambers, Gatekeepers, and the Price of Bipartisanship." Proceedings of the 2018 World Wide Web Conference, Lyon, France.

Garrett, R. Kelly. 2009. "Politically Motivated Reinforcement Seeking: Reframing the Selective Exposure Debate." *Journal of Communication* 59 (4):676-699. doi: 10.1111/j.1460-2466.2009.01452.x.

Grantham, Susan. 2024. "The rise of TikTok elections: the Australian Labor Party's use of TikTok in the 2022 federal election campaigning." *Communication Research and Practice* 10 (2):181-199. doi: 10.1080/22041451.2024.2349451.

Guess, Andrew M., Neil Malhotra, Jennifer Pan, Pablo Barberá, Hunt Allcott, Taylor Brown, Adriana Crespo-Tenorio, Drew Dimmery, Deen Freelon, Matthew Gentzkow, Sandra González-Bailón, Edward Kennedy, Young Mie Kim, David Lazer, Devra Moehler, Brendan Nyhan, Carlos Velasco Rivera, Jaime Settle, Daniel Robert Thomas, Emily Thorson, Rebekah Tromble, Arjun Wilkins, Magdalena Wojcieszak, Beixian Xiong, Chad Kiewiet de Jonge, Annie Franco, Winter Mason, Natalie Jomini Stroud, and Joshua A. Tucker. 2023. "How do social media feed algorithms affect attitudes and behavior in an election campaign?" *Science* 381 (6656):398-404. doi: 10.1126/science.abp9364.

Haroon, Muhammad, Magdalena Wojcieszak, Anshuman Chhabra, Xin Liu, Prasant Mohapatra, and Zubair Shafiq. 2023. "Auditing YouTube's recommendation system for ideologically congenial, extreme, and problematic recommendations." *Proceedings of the National Academy of Sciences* 120 (50):e2213020120. doi: 10.1073/pnas.2213020120.

Howard, Philip N., and Bence Kollanyi. 2016. "Bots, #Strongerin, and #Brexit: Computational Propaganda During the UK-EU Referendum." *arXiv* (2016-1):1606.06356 doi: 10.48550/arXiv.1606.06356.

Hussein, Eslam, Prerna Juneja, and Tanushree Mitra. 2020. "Measuring Misinformation in Video Search Platforms: An Audit Study on YouTube." *Proc. ACM Hum.-Comput. Interact.* 4 (CSCW1):Article 48. doi: 10.1145/3392854.

Jungherr, Andreas, Gonzalo Rivero, and Daniel Gayo-Avello. 2020. *Retooling Politics: How Digital Media Are Shaping Democracy*.

Keller, Tobias R., and Ulrike Klinger. 2019. "Social Bots in Election Campaigns: Theoretical, Empirical, and Methodological Implications." *Political Communication* 36 (1):171-189. doi: 10.1080/10584609.2018.1526238.

Kulachai, Waiphot, Unisa Lerdtomornsakul, and Patipol Homyamyen. 2023. "Factors Influencing Voting Decision: A Comprehensive Literature Review." *Social Sciences* 12 (9):469. doi: 10.3390/socsci12090469.

Lam, Michelle S., Ayush Pandit, Colin H. Kalicki, Rachit Gupta, Poonam Sahoo, and Danaë Metaxa. 2023. "Sociotechnical Audits: Broadening the Algorithm Auditing Lens to Investigate Targeted Advertising." *Proc. ACM Hum.-Comput. Interact.* 7 (CSCW2):Article 360. doi: 10.1145/3610209.

Liu, Bing. 2012. *Sentiment Analysis and Opinion Mining*, *Synthesis Lectures on Human Language Technologies*.

Luscombe, Alex, Kevin Dick, and Kevin Walby. 2022. "Algorithmic thinking in the public interest: navigating technical, legal, and ethical hurdles to web scraping in the social sciences." *Quality & Quantity* 56:1023-1044. doi: 10.1007/s11135-021-01164-0.

McKay, Spencer, and Chris Tenove. 2021. "Disinformation as a Threat to Deliberative Democracy." *Political Research Quarterly* 74 (3):703-717. doi: 10.1177/1065912920938143.

Metaxas, Panagiotis T., and Eni Mustafaraj. 2012. "Social Media and the Elections." *Science* 338 (6106):472-473. doi: 10.1126/science.1230456.

Moir, Aidan. 2023. "The Use of TikTok for Political Campaigning in Canada: The Case of Jagmeet Singh." *Social Media + Society* 9 (1):20563051231157604. doi: 10.1177/20563051231157604.

mpfs. 2023. JIM-Studie 2023 Jugend, Information, Medien. In *Jugend, Information, Medien*: Medienpädagogischer Forschungsverbund Südwest.

Nyhan, Brendan, Jaime Settle, Emily Thorson, Magdalena Wojcieszak, Pablo Barberá, Annie Y. Chen, Hunt Allcott, Taylor Brown, Adriana Crespo-Tenorio, Drew Dimmery, Deen Freelon, Matthew Gentzkow, Sandra González-Bailón, Andrew M. Guess, Edward Kennedy, Young Mie Kim, David Lazer, Neil Malhotra, Devra Moehler, Jennifer Pan, Daniel Robert Thomas, Rebekah Tromble, Carlos Velasco Rivera, Arjun Wilkins, Beixian Xiong, Chad Kiewiet de Jonge, Annie Franco, Winter Mason, Natalie Jomini Stroud, and Joshua A. Tucker. 2023. "Like-minded sources on Facebook are prevalent but not polarizing." *Nature* 620 (7972):137-144. doi: 10.1038/s41586-023-06297-w.

Osmundsen, Mathias, Alexander Bor, Peter Bjerregaard Vahlstrup, Anja Bechmann, and Michael Bang Petersen. 2021. "Partisan Polarization Is the Primary Psychological Motivation behind Political Fake News Sharing on Twitter." *American Political Science Review* 115 (3):999-1015. doi: 10.1017/S0003055421000290.

Pasquale, Frank. 2015. *The Black Box Society*. Cambridge, MA and London, England: Harvard University Press.

Pérez-Escoda, Ana, Luis Miguel Pedrero-Esteban, Juana Rubio-Romero, and Carlos Jiménez-Narros. 2021. "Fake News Reaching Young People on Social Networks: Distrust Challenging Media Literacy." *Publications* 9 (2):24. doi: 10.3390/publications9020024.

Rahmat, Wahyudi, Refa Lina Tiawati, Edwar Kemal, Ricci Gemarni Tatalia, Harizqi Azri, and Yosi Wulandari. 2024. "How Do the Young People Picture Out Their Use, Activeness, and Connectivity on Social Media? A Discourse Analysis Approach." *Journal of Communication Inquiry* 0 (0):01968599231174848. doi: 10.1177/01968599231174848.

Salah, Albert Ali, Cansu Canca, and Erman Bariş. 2022. "Ethical and Legal Concerns on Data Science for Large-Scale Human Mobility." In *Data Science for Migration and Mobility*, edited by Albert Ali Salah, Emre Eren Korkmaz and Tuba Bircan, 24-48. Oxford: British Academy.

Sandvig, Christian, Kevin Hamilton, Karrie Karahalios, and Cédric Langbort. 2014. "Auditing Algorithms: Research Methods for Detecting Discrimination on Internet Platforms." Data and Discrimination, a Pre-Conference at the 64th Annual Meeting of the International Communication Association, Seattle, USA.

Scharkow, Michael, Frank Mangold, Sebastian Stier, and Johannes Breuer. 2020. "How social network sites and other online intermediaries increase exposure to news." *Proceedings of the National Academy of Sciences* 117 (6):2761-2763. doi: doi:10.1073/pnas.1918279117.

Sebastian Golla, and Denise Müller. 2020. "Web Scraping Social Media: Legitimate Research or a Breach of Contract?" *PRIF Blog*.

Shearer, Elisa, Sarah Naseer, Jacob Liedke, and Katerina Eva Matsa. 2024. TikTok users' experiences with news. In *How Americans Get News on TikTok, X, Facebook and Instagram*: PEW Research Center.

Speckmann, Felix. 2021. "Web scraping: A useful tool to broaden and extend psychological research." *Zeitschrift für Psychologie* 229 (4):241-244. doi: 10.1027/2151-2604/a000470.

Srba, Ivan, Robert Moro, Matus Tomlein, Branislav Pecher, Jakub Simko, Elena Stefancova, Michal Kompan, Andrea Hrckova, Juraj Podrouzek, Adrian Gavornik, and Maria Bielikova. 2023. "Auditing YouTube's Recommendation Algorithm for Misinformation Filter Bubbles." *ACM Trans. Recomm. Syst.* 1 (1):Article 6. doi: 10.1145/3568392.

Stier, Sebastian, Arnim Bleier, Haiko Lietz, and Markus Strohmaier. 2020. "Election Campaigning on Social Media: Politicians, Audiences, and the Mediation of Political Communication on Facebook and Twitter." In *Studying Politics Across Media*, 50-74.

Stricklin, Kasey , and Megan K. McBride. 2020. Social Media Bots: Laws, Regulations, and Platform Policies CNA.

Szwed, Patrycja. 2021. "Is web scraping legal? A short guide on scraping under EU law." accessed 09/09/2024. https://discoverdigitallaw.com/is-web-scraping-legal-short-guide-on-scraping-under-the-eu-jurisdiction/.

Wachter, Sandra, Brent Mittelstadt, and Luciano Floridi. 2017. "Transparent, explainable, and accountable AI for robotics." *Science Robotics* 2 (6):eaan6080. doi: 10.1126/scirobotics.aan6080.

Yadlin-Segal, Aya, and Yael Oppenheim. 2021. "Whose dystopia is it anyway? Deepfakes and social media regulation." *Convergence* 27 (1):36-51. doi: 10.1177/1354856520923963.