FANNING THE FLAMES OF HATE: SOCIAL MEDIA AND HATE CRIME

Karsten Müller
Princeton University, Department of Economics

Carlo Schwarz
Bocconi University, Department of Economics

Abstract
This paper investigates the link between social media and hate crime. We show that antirefugee sentiment on Facebook predicts crimes against refugees in otherwise similar municipalities with higher social media usage. To establish causality, we exploit exogenous variation in the timing of major Facebook and internet outages. Consistent with a role for "echo chambers," we find that right-wing social media posts contain narrower and more loaded content than news reports. Our results suggest that social media can act as a propagation mechanism for violent crimes by enabling the spread of extreme viewpoints. (JEL: D74, J15, Z10, D72, O35)

1. Introduction

Social media has come under increasing scrutiny in recent years. In the wake of the 2016 presidential election in the United States, for example, relatively recent phenomena such as fake news, social media echo chambers, and bot farms have been subjects of widespread media coverage and public discourse (e.g., New York Times 2016, 2017a). The role of online hate speech in particular has been at the center of an

The editor in charge of this paper was Imran Rasul.

Acknowledgments: We would like to thank the editor, Imran Rasul, and four anonymous referees for their comments, which greatly improved the paper. We are also grateful to Sascha Becker, Christopher Blattman, Leonardo Bursztyn, Mirko Draca, Ruben Enikolopov, Thiemo Fetzer, Evan Fradkin, Matthew Gentzkow, Andy Guess, Vardges Levonyan, Atif Mian, Magne Mogstad, Sharun Mukand, Hans-Joachim Voth, Fabian Waldinger, Noam Yuchtman, and seminar participants at the NBER Summer Institute, University of Chicago, EEA Conference 2018, Transatlantic Doctoral Conference (LBS), Oxford Internet Institute, Geneva Academy of Humanitarian Law, Bruneck Political Economy Workshop, Leverhulme Causality Conference at the University of Warwick, Spring Meeting of Young Economists 2019, the Royal Economic Society 2019, and the UNHCR Conference on Forced Displacement for their helpful suggestions. We would also like to thank the Amadeu Antonio Stiftung for sharing their data on refugee attacks with us. Müller was supported by a Doctoral Training Centre scholarship granted by the ESRC [grant number 1500313]. Schwarz was supported by a Doctoral Scholarship from the Leverhulme Trust.

E-mail: karstenm@princeton.edu (Müller); carlo.schwarz@unibocconi.it (Schwarz)

Journal of the European Economic Association 2020 00(0):1–37 DOI: 10.1093/jeea/jvaa045

intense and polarized debate. Despite public interest and calls for policy action, there is little empirical evidence on how hateful social media content translates into real-life behavior.

In this paper, we investigate the role of social media in the propagation of hate crimes. Previous research has shown that traditional media can play a role in violent outbursts or ethnic hatred (e.g., Yanagizawa-Drott 2014; Adena et al. 2015; DellaVigna et al. 2014). In contrast to traditional media, social media platforms allow users to easily self-select into niche topics and extreme viewpoints. This preferential selection may limit the spectrum of information people absorb and create "echo chambers" (Sunstein 2009, 2017), which reinforce similar ideas (see e.g., Bessi et al. 2015; Del Vicario et al. 2016; Schmidt et al. 2017). Social media has also become a widely-consumed news source, particularly for young people: In Germany, for example, social media is among the main news sources of 18–25 year olds (Hölig and Hasebrink 2016). In the United States, around half of all adults use social media to get news and two-thirds of Facebook users use it as a news source (Pew Research Center 2018). This suggests that social media could be particularly effective in propagating hateful sentiments.

We study the link between antirefugee sentiment on Facebook and hate crimes against refugees in Germany. The German setting is motivated by the influx of around one million refugees into the country between 2015 and 2016 (BAMF 2016), which was accompanied by frequent violent crimes committed against them (see, for example, recent video coverage by New York Times 2017b). Between January 2015 and early 2017 alone, the nonprofit organization "Amadeu Antonio Stiftung" recorded around 3,300 antirefugee incidents, including over 750 cases of arson or outright assault.

We posit that social media can reinforce antirefugee sentiments, which may push some potential perpetrators over the edge to carry out violent acts. Our empirical strategy exploits differences in Facebook usage at the municipal level and weekly variation in antirefugee sentiment on social media. We create a novel measure for the salience of antirefugee hate speech on social media based on the Facebook page of the "Alternative für Deutschland" (Alternative for Germany, AfD hereafter), a relatively new right-wing party that became the third-strongest faction in the German parliament following the 2017 federal election. The AfD has positioned itself as an antirefugee and antiimmigration party. With more than 300,000 followers, 175,000 posts, 290,000 comments, and 500,000 likes (as of early 2017), their Facebook page has a broader reach than that of any other German party.[1]

This widespread reach makes the AfD's Facebook page uniquely suited to measure antirefugee sentiment on social media. In contrast to established political parties like Angela Merkel's Christian Democratic Union (CDU) or the German Social Democrats (SPD), the AfD allows users to directly post messages on its Facebook wall. The AfD is also the only party that does not explicitly outline rules of conduct, for example, by threatening to remove racist, discriminating, or otherwise hateful comments. We

1. We provide a short history of the AfD in Online Appendix A.

show that the content on the AfD page is consistently more focused on refugees than that of traditional news reports and frequently contains loaded terms that civil rights groups have identified as "hate speech." These detailed data also allow us to construct a measure of each town's exposure to Germany-wide antirefugee sentiment using the share of the population that is active on the AfD Facebook page.

Using fixed effects panel regressions, we find that—during periods of high salience of refugees on right-wing social media—antirefugee hate crimes increase in areas with higher Facebook usage. This correlation is especially pronounced for violent incidents such as assault. Controlling for a large vector of municipality characteristics, interacted with our salience measure, makes little difference for the magnitude and statistical significance of these estimates.

A concern is that our measures of exposure to right-wing social media may be correlated with unobserved municipal characteristics that explain disproportionate increases in hate crimes during times of high antirefugee sentiment. To narrow down the social media transmission channel, we provide quasi-experimental evidence using the exact timing of country-wide Facebook outages and local internet disruptions, which reduce the number of social media posts.

To begin, we study large, Germany-wide Facebook outages resulting from programming or server problems at the platform. These outages disrupt users' exposure to this particular social media platform without affecting other online channels. We find that Facebook disruptions reduce local hate crimes, particularly in areas with many AfD users. Further, during Facebook outages, higher antirefugee sentiment is not associated with a differential increase in hate crimes in areas with high Facebook usage. These results suggest that social media might play a propagating role in translating online content into offline violence.

We also exploit the precise timing of hundreds of local internet disruptions as a source of granular exogenous variation in access to social media. These local disruptions reduce a particular town's exposure to social media content although leaving Germany-wide refugee salience unaffected. Notably, the frequency of internet disruptions is geographically dispersed and largely unrelated to observable local characteristics, including AfD likes on Facebook.

We find that, although hate crimes increase in periods of higher refugee salience, this correlation disappears for municipalities experiencing an internet outage. Quantitatively, a typical internet disruption fully mediates the link between social media and hate crime. Further, once we take into account social media transmission, these internet outages themselves are no longer associated with antirefugee incidents, nor are their interactions with local internet usage or mobile internet access. These results point to social media as propagation mechanism rather than other online channels. It also makes it unlikely that we are capturing a "displacement effect" arising from potential perpetrators fixing their internet access.

We also analyze how other salient news events mediate the link of antirefugee Facebook posts with the number of violent incidents, building on Eisensee and Strömberg (2007) and Durante and Zhuravskaya (2018). Specifically, we look at the European Soccer Championship, Brexit, and Donald Trump's presidential election, all

of which crowded out the salience of refugees. Similar to our outage results, social media exposure has a significantly more muted relationship with hate crimes during these events. The link we uncover appears to be specific to antirefugee sentiment: other posts on the AfD Facebook page, for example those related to Muslims or the European Union, do not have the same predictive power for antirefugee hate crimes. Consistent with the hypothesis that social networks can act as transmission channel, the correlation with hate crime is larger in regions where AfD users show higher Facebook engagement.

When interpreting our results, we do not claim that social media itself causes crimes against refugees out of thin air. Rather, our argument is that social media can act as a propagating mechanism for hateful sentiments that likely have many fundamental sources. We find evidence for two potential channels. First, our results are driven by refugee attacks committed by groups of perpetrators. This suggests that social media may motivate collective action, consistent with existing evidence on other political outcomes such as protests (e.g., Enikolopov, Makarin, and Petrova 2016). Second, we find evidence for a spillover channel. Hate crimes are considerably more common in weeks when neighboring towns also experience them, and this is particularly true for towns with many right-wing social media users when antirefugee sentiment is elevated. In contrast, we find little evidence that social media provides useful information to perpetrators. Our results are also unlikely to be explained by persuasion effects, because we focus on high-frequency variation.

Related Literature. Our work provides evidence that social media may have effects on real-life outcomes, as measured by hate crimes. We build on existing work on media exposure and persuasion (see e.g., DellaVigna and Gentzkow 2010; DellaVigna and Ferrara 2015). In addition to the work on traditional media and violence cited previously, Dahl and DellaVigna (2009) show that—in contrast to experimental settings—violent movies decrease violent crime in the field due to displacement effects. Television has also been associated with short-lived outbursts of domestic violence (Card and Dahl 2011). In other research, Bhuller et al. (2013) demonstrate that exposure to pornographic material on the internet is linked to increased sex crime. Bursztyn et al. (2017) find that media coverage of close elections increases voter turnout, whereas Gavazza, Nardotto, and Valletti (2018) show that broadband diffusion decreased voter turnout in the United Kingdom (see also Gentzkow 2006; Manacorda and Tesei 2020). Enikolopov, Makarin, and Petrova (2016) find that social media exposure spurs protest participation in Russia by reducing coordination costs.

We contribute to this literature by investigating the role of social media in stirring up violence. Previous research has documented the prevalence of online hate speech (Oksanen et al. 2014). Other work has shown that Google search data can be used to measure racial animus (Stephens-Davidowitz 2014). In complementary work, we study the effect of Twitter usage on anti-minority sentiments in the United States (Müller and Schwarz 2018). Bursztyn et al. (2019) study the effect of social media on xenophobia in Russia. In contrast to these papers, we focus on the short-run impact of

social media posts, rather than long-run effects that may work through persuasion or changes in social norms.

Our paper also builds on research about the polarization of citizens (e.g., Fiorina and Abrams 2008). There is no consensus on whether social media increases or decreases polarization: Some authors argue that social media are divisive (Pariser 2011; Gabler 2016), whereas others find that polarization *decreases* with social media usage (Barberá 2014; Boxell, Gentzkow, and Shapiro 2017). Our work suggests that—independent of whether social media affects overall polarization or not—social media content can be associated with violent crimes.

We also build on the literature on culture and violence. Summarizing a vast body of research, Alesina and La Ferrara (2005) find that cultural and religious fragmentation predict the likelihood of civil war across countries. Voigtlander and Voth (2012) show that antisemitic violence in Germany is highly persistent: Pogroms during the era of the Black Death predict pogroms in the 1920s, Jewish deportations, and synagogue attacks during the rise of the Nazi party. Similarly, Jha (2013) shows that medieval interethnic complementarities in trade decrease the likelihood of modern Hindu–Muslim riots. These papers, however, are largely silent on the existence of volatile, short-lived bursts of sentiment leading to violent incidents. As such, our work is also related to Fouka and Voth (2013), who show that monthly variation in public acrimony between Greek and German politicians during the Greek debt crisis affected German car purchases particularly in areas of Greece where German troops committed war crimes during World War II. Our results also align with the findings of Colussi, Isphording, and Pestel (2016), who show that a higher salience of minority groups increases the likelihood of hate crimes.

While traditional media such as television are regulated in most countries, legislators are now beginning to address social media. Our work is thus particularly topical in light of the political discussions in many countries about antihate speech laws and censoring hate speech on social media. The German parliament, for example, passed an anti online hate speech law ("Netzwerkdurchsetzungsgesetz") on June 30, 2017, which threatens providers of online platforms such as Facebook with fines up to €50 million for failing to delete "criminal" content that is "obviously unlawful." The controversial law was the initiative of German Minister of Justice Heiko Maas, who lamented social media platforms' unwillingness to address "online hate crime."[2] The European Union has issued independent guidelines calling on social media companies to remove illegal hate speech as well. In the United Kingdom, the Crown Prosecution Service plans to increase prosecution of online hate crimes (The Guardian 2017; BBC 2017). Our paper serves as a first attempt to address this important topic empirically.

The paper proceeds as follows. In Section 2 we introduce the data used in our empirical analysis. Section 3 presents the results. Section 4 concludes the paper.

2. See, for example, the official statement of the German parliament on bundestag.de.

2. Data

We construct a dataset on social media activity and antirefugee hate crimes in Germany. In total, we combine data from twelve different sources which we describe in more detail in the following subsections: (1) Municipal-level data on antirefugee hate crimes; (2) Facebook data on posts, likes, and comments on the AfD page; (3) hand-collected municipal-level data on Facebook user locations; (4) municipal-level data on internet outages; (5) a hand-coded dataset on major weekly Facebook outages; (6) municipal- and county-level socioeconomic data from the German Statistical Office; (7) municipal-level voting data; (8) county-level data on broadband access; (9) municipal-level data on newspaper sales; (10) data on the content of reporting about refugees from Nexis; (11) city-level data on neo-Nazi murders and historical anti-Semitism; and (12) weekly Google search data on major news events in our sample. The final panel dataset covers 4,466 German municipalities for the 111 weeks from 1st January 2015 to 13th February 2017. Summary statistics for the main variables of interest can be found in Table 1 and Table B.3 in Online Appendix. The Online Appendix provides a comprehensive overview of the data sources and variable definitions (see Table B.4).

2.1. Antirefugee Incidents

The data on incidents targeting refugees were collected by the Amadeu Antonio Foundation and Pro Asyl (a pro asylum nongovernmental organization).[3] These data cover incidents including antirefugee graffiti, arson of refugee homes, assault, and incidents during protests in Germany between January 2015 and early 2017. This period is of particular interest because it includes the beginning and height of the refugee crisis in Germany. All 3,335 antirefugee aggressions feature a short description and are classified into four groups. The most common cases are property damage to refugee homes (2,226 incidents), followed by assault (534), incidents during antirefugee protests (339), and arson (225). Eleven events are classified as suspected cases that were still under investigation. Table B.2 in the Online Appendix lists examples for each class of antirefugee activity.

All incidents are geo-coded with an exact longitude and latitude, which we use to assign them to municipalities.[4] Figure 1 shows the location of the antirefugee incidents in our observation period for each German municipality.

3. These data are available at https://www.mut-gegen-rechte-gewalt.de/service/chronik-vorfaelle.

4. To assign coordinates to municipalities, we use the shape files provided by the GeoBasis-DE/BKG 2016 website. The shape file contains data for the 4,679 German municipalities ("Gemeindeverwaltungsverband"). A total of 213 of these municipalities do not have inhabitants (e.g., forest areas) nor antirefugee incidents. After dropping these cases, we are left with 4,466 municipalities in our estimation sample. We use the level of the "Gemeindeverwaltungsverband" because these exhibit smaller differences in their size and population than the 11,165 German "Gemeinden" and are therefore more suitable for spatial analysis according to the data provider (see link).

TABLE 1. Summary statistics for main variables.

	Level	Obs	Mean	SD	Min.	Max.
Refugee attacks						
Refugee attacks	Muni.-Week	495,726	0.007	0.099	0	8
Arson attacks	Muni.-Week	495,726	0.000	0.022	0	2
Other property damage	Muni.-Week	495,726	0.004	0.076	0	8
Assaults	Muni.-Week	495,726	0.001	0.035	0	3
Protests	Muni.-Week	495,726	0.001	0.030	0	5
Social media data						
AfD users/Pop.†	Municipality	495,726	0.301	0.286	0	8
Refugee posts	Week	495,726	84	61	2	259
Posts/AfD users	Municipality	395,493	0.554	3.882	0	118
Comments/AfD users	Municipality	395,493	1.1	7.3	0	270
Likes/AfD users	Municipality	395,493	1.8	12.3	0	370
Auxiliary ariables						
$I_{Internet\ outage}$	Muni.-Week	495,726	0.001	0.025	0.000	1.000
$I_{Facebook\ outage}$	Municipality	495,726	0.072	0.259	0.000	1.000
Baseline ontrols						
Ln(Population (2015))	Municipality	495,726	9	1	6	15
GDP/worker	County	493,617	63,095	9,846	46,835	136,763
Population density	Municipality	495,726	282	382	7	4,653
AfD vote share (2017) (in %)	Municipality	492,618	15	7	3	45
Share high school (in %)	Municipality	495,726	29	8	0	58
Share broadband access (in %)	Municipality	495,726	83	11	44	100
Share immigrants (in %)	Municipality	483,072	14	8	2	50
Asylum Seekers/Pop.	County	495,726	0.011	0.006	0.000	0.102

Notes: This table reports summary statistics for the main variables in the estimation sample. Variables tagged with a †. are scaled by population (in 1,000).

The data appear to be of high quality. Each entry has a clearly indicated source. Nearly half of the incidents in the dataset are reported by the federal government in response to inquiries by the left-wing party "Die Linke." Other sources include police reports and national or local media outlets. We hand-checked a random sample of 100 incidents and found their coding accurately reflected the information reported in the respective source.

2.2. Facebook Data on Refugee Salience

We construct a proxy for the frequency of antirefugee messaging on social media based on the Facebook page of the AfD. We chose the AfD's page because the party is by far the most popular far-right political movement in Germany. At the time of the refugee crisis, the AfD also had the highest number of Facebook followers of *any* German party. This makes their page arguably the most important platform of exchange about refugees among Germany's right-wing social media users.

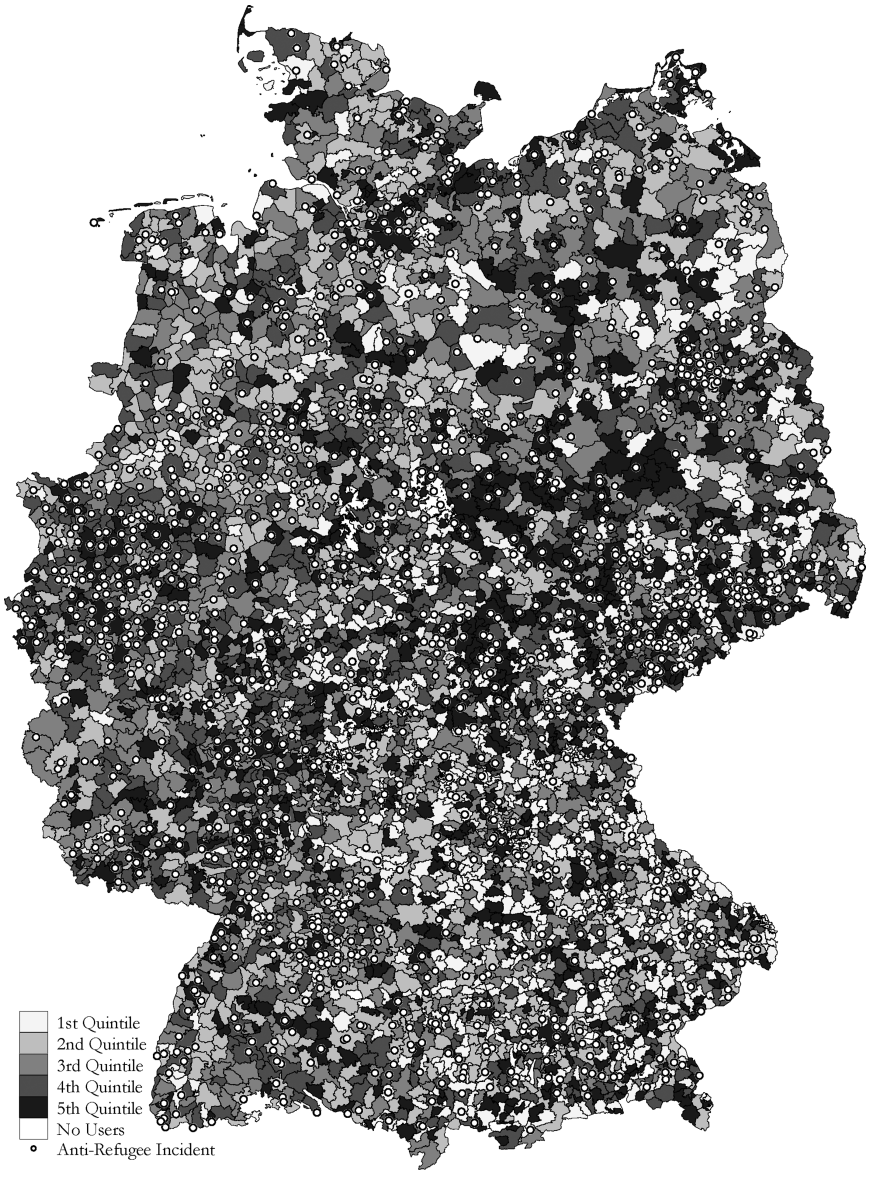

FIGURE 1. AfD Facebook usage per capita and antirefugee incidents. This map plots the number of Facebook users of the AfD page per capita for each of the 4,466 German municipalities. The dots indicate the locations of the 3,335 antirefugee incidents from the Amadeu Antonio Foundation.

We start by using the Facebook Graph API to collect all status posts, comments, and likes from the AfD Facebook page (see Online Appendix B.1 for an introduction to Facebook). The API provides a unique identifier for each post, allowing us to link posts to comments and likes, as well as the users who posted, commented, or liked anything on the page. Overall, we collected 176,153 posts, 290,854 comments, 510,268 likes, and 93,806 individual user IDs.

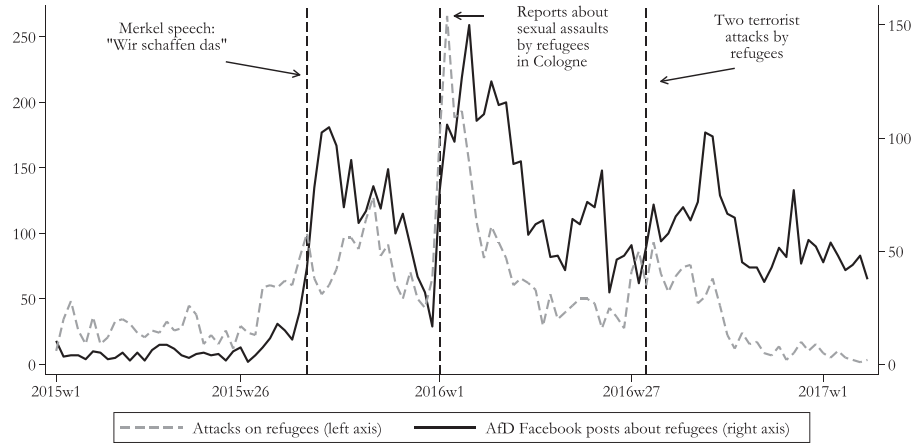

FIGURE 2. Refugee posts on social media and antirefugee incidents over time. This figure plots the number of posts about refugees on the Facebook page of the "Alternative for Germany" and the number of antirefugee incidents in Germany over time.

As our baseline measure for the salience of antirefugee hate speech on social media, we use the number of posts on the AfD Facebook page that contain the word "Flüchtling" (refugee) in any given week. The narrative in these posts centers around the idea that the "elites"—politicians and mainstream media outlets—have betrayed "the people" by allowing "streams" of illegitimate "economic refugees" to enter the country, who are described as being criminals and rapists for "cultural reasons." Table B.1 in the Online Appendix provides a few representative examples; Section 3.5 provides a more in-depth analysis. A potential downside of this approach is that we may inadvertently tag posts that do not express negative sentiments towards refugees. However, a careful content analysis of posts and comments reveals that the overwhelming majority appear to agree with the positions of the AfD. This is perhaps unsurprising given that only people who "like" the AfD Facebook page will be informed about new posts. Critics, on the other hand, have a strong incentive not to indicate publicly that they "like" the party.

We plot the total number of AfD Facebook page posts about refugees and the number of antirefugee incidents in Figure 2. Weeks with more refugee posts also tend to have more antirefugee events. Both series clearly spike during salient events related to refugees, such as Angela Merkel's widely reported statement "Wir schaffen das" ("We can do this") during a press conference on the challenges of the refugee situation. A simple time series regression of refugee attacks on AfD posts yields a R^2 of 0.34 (unreported).

2.3. Municipal-Level Facebook Measures

We construct a measure of exposure to right-wing social media at the municipal level. Because survey data about German Facebook usage are, to our knowledge, only

available at the level of the 16 federal states, we hand-collect user location data by using the unique user identifiers provided by the Facebook Graph API. Due to Facebook's privacy policy, we are only able to collect this information for people who make it publicly available.

Because we are interested in the transmission of right-wing social media sentiment, we measure exposure to it on Facebook based on users of the AfD page. In total, we can identify 93,806 users who interacted with the page at least once.[5] We were able to hand-collect and geocode a place of residence for 34,396 of these users. Overall, we were able to identify at least one AfD Facebook page user for 3,563 of the 4,466 municipalities.[6] In Figure 1, we visualize the distribution of AfD users per capita. Antirefugee incidents are concentrated in areas with more right-wing social media users. To illustrate this, Figure B.3 in the Online Appendix shows the share of municipalities with at least one refugee attack, depending on whether we can identify *at least one* AfD Facebook page user. Municipalities with AfD users are three times as likely to experience an attack during our observation period. Out of the total 3,335 attacks on refugees in our sample, 3,171 occurred in municipalities with AfD Facebook page users. A *t*-test rejects the null hypothesis of no difference between the mean of the two groups with a value of 22.95. Using the location data for AfD users, we can also assign posts, comments, and likes to municipalities. Based on these data, we construct auxiliary measures of social media interactions, e.g., the number of local posts scaled over the number of AfD users.[7]

2.4. Data on Internet and Facebook Outages

We collect data on local internet outages from Heise Online. Heise lists user reports of internet problems by telephone area codes and includes start times and duration. We use area codes to assign internet problems to municipalities; the start date and duration allow us to count the number of problems for each municipality and week.[8] The internet outage reports are geographically dispersed with no clear patterns of regional clustering (see Figure C.2(a) in the Online Appendix). The outages are also dispersed over time (see Figure C.2(b) in the Online Appendix).

5. The Facebook API does not provide data on which users "like" a page but only on users who *interact* with a page, for example by liking another user's comment. As a result, the total number of user IDs we have is smaller than the more than 300,000 people who had liked the AfD Facebook page as of 2017.

6. Note that the decision of users to disclose their location is unlikely to matter in our setting. This is because we exploit variation *within* the same location over time, which abstracts from time-invariant endogenous selection using municipality fixed effects.

7. We find that some users post and comment excessively, which leads to a few outliers in measuring how active users are in a given municipality. We therefore winsorize the number of posts, comments, and likes we can attribute to local users at the 99.9th percentile to avoid individual users driving the results.

8. If an area code spans multiple municipalities, we assign an internet outage to the municipality that overlaps most with the area code. We prefer this over to assigning the outage to all municipalities within the area code's territory because some area codes include minor overlaps with many municipalities. Assigning an internet outage to all of these municipalities would introduce substantial noise.

To validate the Heise data, we search for newspaper reports on major internet disruptions. Although the large-scale and short-lived outages discussed in the newspaper reports are not representative of the more localized and longer-lasting outages we exploit in our regressions, they do suggest that the Heise data provide a valid proxy for internet disruptions. For all major disruptions we could identify in newspapers, the Heise data suggest an increase in the number of outages specific to the internet provider experiencing the outage. Table C.1 lists several examples of newspaper reports on such outages and the respective information in our data.[9]

We focus on major outages that fulfill two criteria: (1) they have to last longer than 24 hours, and (2) they affect a significant part of the population (be in the top quartile of the reported internet problems to population ratio). This gets around the issue that some reports may reflect individual users' glitches rather than general disruptions.[10]

We also collect information on major Facebook disruptions. To identify these, we start by searching for newspaper reports of Facebook problems in our sample period. In total, we find reports on eight large outages (see Table C.2 in the Online Appendix for an overview and more details). We then validate their precise timing using the number of weekly user-reported Facebook problems on the website of "Allestörungen," a portal for aggregating user complaints on individual websites and apps. Perhaps unsurprisingly, the eight outages widely reported on in the news media are also associated with spikes in user-reported problems.

Using these data, we define a dummy variable that is 1 for weeks with Facebook outages and 0 otherwise. These outages have the advantage that they are specific to Facebook; in fact, they are uncorrelated with the total number of weekly internet outages in a given week from our Heise data. In contrast to the internet disruptions, the downside is that Facebook outages are rare, shorter, and only generate weekly variation.

2.5. Auxiliary and Control Variables

We obtain control variables from a host of sources, which are explained in more detail in the Online Appendix. Socioeconomic data on the municipality and county level are from the German Statistical Office, available via http://www.regionalstatistik.de. We include information on each municipality's population by age group, GDP per worker, population density, the share of the population with a high school degree ("Abitur"), the

9. To interpret the number of outages, note that the Heise data report an average of four reported internet outages per provider per week. That means even an increase of 15 reported outages represents a large increase.

10. In some cases, users do not seem to report the end date of the internet outage, which can lead to unlikely durations of several months. We thus winsorize the maximum duration at three weeks, but this choice is not material for our results. We scale outages over population because towns with more inhabitants mechanically also report more disruptions. As we discuss in what follows, our results are robust to using alternative definitions of this cut-off.

share of the population receiving social benefits, the share working in manufacturing, and the vote results for the 2017 German Federal Election. To control for "pull factors" of anti-minority crimes, we also obtain the share of the population that are immigrants and asylum seekers.

To measure the extent to which people use the internet, we use the share of households in a county with broadband access as well as average mobile download speeds, collected by the Federal Ministry of Transport and Digital Infrastructure (BMVI).[11] In addition, we use the number of registered *.de* internet domains per capita in a county to measure internet affinity, which has a correlation of 0.48 with broadband access.

To measure the local penetration of traditional media, we obtain data for 2016/2017 newspaper sales from the "Zeitungsmarktforschung Gesellschaft der deutschen Zeitungen (ZMG)" (Society for Market Research of German Newspapers).[12] Based on this data, we construct a measure of traditional newspaper consumption as the number of newspaper sales per capita.

For our comparison of social and more traditional media, we collected the number of total and refugee-related reports in German news media from Nexis UNI (previously LexisNexis). We use this to construct the weekly share of news reports about refugees. For further analysis, we obtained the full text of all refugee-related reports using the Lexis bulk data API, as well as all Facebook data from the pages of five major German newspapers (Welt, Frankfurter Allgemeine Zeitung (FAZ), Tageszeitung (TAZ), Süddeutsche Zeitung (SZ), and Bild).

We also include controls for the local prevalence of right-wing extremism. One such measure is the number of murders committed by neo-Nazis in each municipality from 1990 until 2016, which were collected by "Mut gegen rechte Gewalt" (Courage Against Right-Wing Violence). We complement this proxy for contemporary right-wing violence with data on the historic prevalence of anti-semitism collected by Voigtlander and Voth (2012).[13]

Finally, we obtain Google trends data on overall interest in the search terms "Brexit", "Trump", and "UEFA EM 2016" in Germany to proxy for distracting news events. Google scales the weekly number of searches for these terms on a scale from

11. Broadband access is highly correlated with publicly available survey data on individuals' internet use from Eurostat; these data are only available on the state level (see Figure B.4 in the Online Appendix).

12. These data contain the number of print newspapers sold in each municipality with more than 3,000 inhabitants. Newspapers are listed if, in any given town, they (1) sell at least 50 copies and (2) have a market share of at least 1%. To have a similar sample size across specifications, we impute values for 1,120 towns for which news paper sales data are not available, based on a municipality's population, population density, AfD vote share, and county fixed effects. However, the results are almost equivalent without imputation (available upon request).

13. From their dataset, we use the natural logarithm of one plus the number of deported Jews as well as one plus the number of letters written to "Der Stürmer," the antisemitic newspaper published by Nazi politician Julius Streicher. Towns with no information are coded as zero. We do not use scaled variables because the data from Voigtlander and Voth (2012) only cover a fraction of the municipalities in our sample.

0 to 100, where 100 marks the week with the highest search interest in the preceding 5 years. The time series plots in Figure D.1 in the Online Appendix suggest these measures are sound approximations for attention paid to Brexit, the Trump election, and the UEFA European Championship (one of the most widely followed sports events in Germany).

3. Empirical Strategy and Main Results

3.1. Empirical Strategy

We begin to investigate the link between social media and antirefugee incidents by estimating fixed effects panel regressions akin to a Bartik-type approach (Bartik 1991). In particular, we use the interaction of local right-wing Facebook usage ($AfD\ Users/Pop_i$) and weekly refugee posts on the AfD Facebook page ($Refugee\ Posts_t$) to measure the differential change of hate crimes conditional on antirefugee sentiment on social media. This empirical set-up creates variation by week and municipality, which we exploit in the following regression model:

$$\begin{aligned} Refugee\ attack_{it} = {} & \beta\ AfD\ Users/Pop_i \times Refugee\ Posts_t \\ & + \gamma\ Controls_i \times Refugee\ Posts_t \\ & + Week\ FE_t + Municipality\ FE_i + \varepsilon_{it}, \end{aligned} \tag{1}$$

The dependent variable is a dummy for the incidence of a refugee attack in municipality i in week t. β measures the differential change in antirefugee incidents conditional on Germany-wide posts about refugees on the AfD page—as a proxy of Germany-wide antirefugee sentiment on social media—and right-wing social media users per capita. We control for a host of local characteristics interacted with the refugee post measure. Because we include many fixed effects and interaction terms, we estimate 1 using ordinary least squares, which yields the linear probability model. Standard errors are clustered by municipality. We consider alternative specifications of the dependent variable and standard errors in robustness exercises.

This framework has three key features. First, it circumvents reverse causality, because refugee incidents in one town are unlikely to change antirefugee sentiment in *all other* towns. Second, our measure of social media exposure is time-invariant and thus not the result of whether a municipality experiences refugee attacks in a given week.[14] Third, a full set of fixed effects controls for unobserved heterogeneity that affects all towns at the same time (such as salient news events), as well as time-invariant differences across towns (such as a history of anti-minority violence).

14. In the robustness section in what follows, we alternatively measure local social media penetration before the start of the refugee crisis, at the cost of reducing the number of users for whom we have location data. This adjustment makes little difference for the results.

The main concern with estimating equation (1) is that *AfD Users/Pop.* may be correlated with other municipality characteristics that could explain differences in how local antirefugee attacks co-vary with the salience of refugees online. In that case, we would not be capturing a pure social media "effect." For example, the share of AfD Facebook subscribers may partially pick up general right-wing attitudes, which could lead to more antirefugee attacks in times of high refugee salience. This concern may also not be sufficiently addressed by controlling for interactions of observable municipality characteristics with the refugee salience measure.

We therefore develop an identification strategy based on Facebook and internet outages. These disruptions induce plausibly exogenous variation in people's exposure to social media although leaving other local characteristics unchanged. The first part of this empirical strategy exploits the timing of major server problems at Facebook, which disrupt access to the platform. In the second part, we build on the insight that German internet infrastructure is trailing behind that of many other European Countries (e.g., Latvia) and the Organisation for Economic Co-operation and Development average (see Financial Times 2017; OECD 2016). As a result, prolonged internet outages are relatively common. Because around 50% of worldwide Facebook users accessed the platform with their computers, many users are exposed to disruptions in internet access. In Germany, this share is likely to be even higher because of the relatively slow adoption of mobile internet.[15]

Local internet outages are widely dispersed geographically: Figure C.2(a) in the Online Appendix visualizes the distribution of disruptions per capita across Germany. The outages are also not particularly clustered in a particular time period (see Figure C.2(b) in the Online Appendix). Crucially, the frequency of internet problems is uncorrelated with the share of the population on the AfD Facebook page. As such, internet disruptions provide exogenous variation that is not already captured by our variable on local Facebook usage. The number of reported internet problems is also uncorrelated with the total number of refugee attacks in a given municipality. In fact, regressing the frequency of internet outages on a host of municipality characteristics in Figure 3 suggests that they are largely uncorrelated with observable factors: the estimated coefficients are nearly all statistically indistinguishable from zero and quantitatively small. Taken together, our interpretation is that whether an internet outage occurs in a given town and week is as good as randomly assigned with regard to unobserved other factors that might drive hate crimes.

We analyze the effect of Facebook and internet outages in a flexible empirical framework. We begin by asking whether these outages reduce antirefugee attacks, and whether they do so particularly in areas with a higher concentration of AfD Facebook users. We then study whether these disruptions also decrease our baseline correlation of local exposure to antirefugee sentiment and hate crimes. More formally, the most

15. Data on Facebook usage patterns reported on Statista.com and on mobile internet usage in Germany on (also on Statista.com) support this assessment.

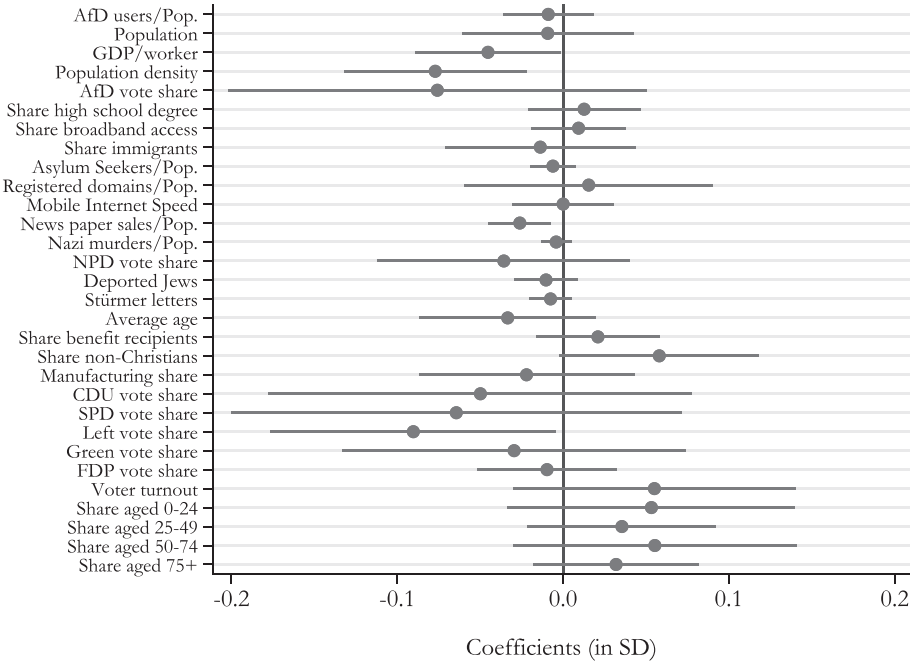

FIGURE 3. Balancedness—internet outages and local characteristics. This figure plots the coefficients of the regression $\overline{Internet\ outages}_i = \alpha + \mathbf{X}'\beta + \varepsilon_i$, where the dependent variable is the total number of internet outages in a municipality (based on our baseline definition) and $\mathbf{X}$ is a vector of local characteristics for which we plot the estimates. To make the magnitudes comparable, we standardize all variables to have a mean of zero and standard deviation of one. The 95% confidence intervals are based on standard errors clustered by municipality.

saturated regressions have the following triple difference form:

$$\begin{aligned} Refugee\ attack_{it} = {} & \beta\ AfD\ Users/Pop_i \times Refugee\ Posts_t \\ & + \lambda\ Outage_{it} \times AfD\ Users/Pop_i \times Refugee\ Posts_t \\ & + \delta_1\ Outage_{it} + \delta_2\ Outage_{it} \times Refugee\ Posts_t \\ & + \delta_3\ Outage_{it} \times AfD\ Users/Pop_i \\ & + \gamma_1\ Controls_i \times Refugee\ Posts_t \\ & + \gamma_2\ Controls_i \times Outage_{it} \\ & + Week\ FE_t + Municipality\ FE_i + \varepsilon_{it}, \end{aligned} \tag{2}$$

For the Facebook outages, which only vary by week, we replace $Outage_{it}$ with $Outage_t$.[16] For the initial tests, we focus on the estimates for δ_1 and δ_3 although

16. Note that, as a result, the estimates of δ_1 and δ_2 in equation (2) are absorbed by the week fixed effects.

excluding the coefficients β, λ, δ_2, and γ_1. That is, we ask whether outages reduce antirefugee incidents, and whether they reduce them more in areas with more AfD Facebook users. In the fully interacted regressions, the main coefficient of interest λ captures the correlation of antirefugee attacks and local exposure to antirefugee sentiment on social media, depending on whether an outage occurs. Put differently, we test whether outages break the correlation between real-life incidents and refugee salience, particularly for areas with high right-wing Facebook penetration. The vector $Controls_i \times Outage_{it}$ controls for the differential effect of outages based on observable characteristics, such as internet affinity.

The identifying assumption of this approach is that Facebook and internet outages only affect antirefugee incidents through their effect on social media exposure. This assumption is plausible for Facebook outages. In the case of internet outages, for which we have variation at the municipality-week level, one may be worried about alternative online channels. We discuss these and other potential threats to identification in the next section.

Exploiting variation in Facebook and internet outages also allow us to address the concern that towns with a stronger right-wing presence may show differential trends whenever the nationwide sentiment towards refugees changes. This is because these relatively short-lived outages are unlikely to affect the presence of deep-rooted right-wing attitudes in a municipality; absent online channels, the outages should thus not have an impact on real-life outcomes. The framework in equation (2) further addresses reverse causality concerns. If we were merely capturing that local incidents drive posts on social media, Facebook and internet outages should not reduce the number of hate crimes. Instead, they should only reduce social media activity, keeping the number of antirefugee incidents unchanged.

3.2. Panel Regression Results

We illustrate the intuition behind our regression framework in Figure 4. The figure shows a binned scatter plot of antirefugee attacks and antirefugee sentiment, split by the degree of exposure to right-wing social media. Higher refugee salience is associated with a higher probability of antirefugee attacks in both sub-samples, but the positive slope is far more pronounced for towns with an above median AfD user to population ratio (Panel (a)). Our baseline regression coefficient picks up the difference in slopes between municipalities with high and low Facebook usage.

Table 2 presents the regression results from estimating equation (1) with varying sets of control variables (interacted with refugee salience). The coefficient on the interaction of local Facebook usage and Germany-wide refugee posts is positive and highly significant in all specifications. Column (1) shows the panel regressions with the baseline control variables, which yields a coefficient 0.024 on the interaction term. This correlation does not appear to be driven by support for the AfD alone: The result holds although we control for the AfD vote share in the 2017 federal election. This highlights a distinction between our social media measure and general support for the party.

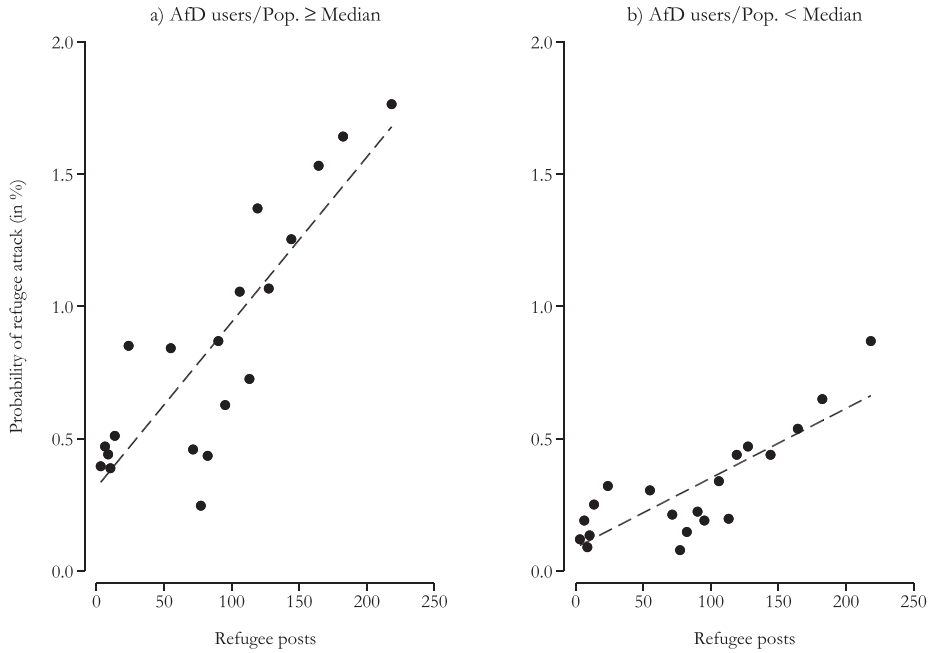

FIGURE 4. Exposure to refugee sentiment on Facebook and hate crimes. This figure plots the average number of antirefugee attacks against our measure of antirefugee sentiment for municipalities below and above the median of *AfD Users/Pop*. Refugee attacks are binned by 20 quantiles of refugee posts and residualized with respect to population.

To get a sense of the magnitudes, consider as a case study the cities of Bochum and Hannover, which are about one standard deviation apart in the ratio of AfD users to population (in 1,000s) ($\approx$0.29). Holding average antirefugee sentiment in our data constant (84 posts), this means a one standard deviation higher right-wing social media usage is associated with a 10% higher probability of an antirefugee incident relative to the mean. Table D.1 in the Online Appendix shows that this correlation is largely driven by cases of assault.

In columns (2) through (6), we introduce a richer set of controls that accounts for local right-wing attitudes, general media exposure, more socio-economic factors, and the vote shares of all major parties in the 2017 election (see Table B.3 in the Online Appendix for an overview of the control variables). In column (7), we add all interacted controls jointly. The inclusion of these covariates makes little difference to our main estimate. This is a first indication that the correlation between social media exposure and antirefugee incidents is not driven by observable municipality differences unrelated to Facebook usage.

3.3. Quasi-Experimental Evidence: Facebook and Internet Outages

To isolate the importance of social media, we next draw on internet and Facebook outages as sources of quasi-experimental variation. To count as a severe internet

TABLE 2. Baseline correlations—Facebook posts and hate crime.

		Additional interacted controls					
	Baseline controls (1)	Right Wing controls (2)	Media controls (3)	Socio-economic controls (4)	2017 vote controls (5)	Age structure controls (6)	All controls (7)
AfD users/Pop. × Refugee posts	0.024***	0.020**	0.023**	0.024**	0.021**	0.023**	0.016**
	(0.009)	(0.008)	(0.009)	(0.009)	(0.009)	(0.009)	(0.008)
Observations	479,964	479,964	479,964	474,303	479,964	476,856	474,303
R-squared	0.082	0.083	0.082	0.083	0.083	0.083	0.084
Municipalities	4,324	4,324	4,324	4,273	4,324	4,296	4,273
Municipality FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Week FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Baseline controls [8] × Posts	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Right-wing controls [4] × Posts		Yes					Yes
Media controls [4] × Posts			Yes				Yes
Socio-econ. controls [4] × Posts				Yes			Yes
Election controls [7] × Posts					Yes		Yes
Age controls [4] × Posts						Yes	Yes

Notes: This table presents the estimated coefficients from a regression of hate crimes against refugees on the interaction of local social media usage and antirefugee sentiment as in equation (1). The dependent variable is a dummy for the incidence of a refugee attack. *AfD users/Pop.* is the ratio of people with any activity on the AfD Facebook page to population. *Refugee posts* is the Germany-wide number of posts on the AfD's Facebook wall containing the word refugee ("Flüchtling"). All control variables are interacted with the *Refugee posts* measure; see text for a description of the controls. Robust standard errors in all specifications are clustered by municipality. ***, **, and * indicate statistical significance at the 0.01, 0.05, and 0.1 levels, respectively.

disruption, our baseline measure has to fulfill two criteria: (1) it has to last at least 24 hours, and (2) it has to affect a significant part of the population, that is, be in the top quartile of reported internet disruptions per capita, which vary by municipality and week (see section Section 2 for more details). This gives us 313 severe internet outages.[17]

Internet Outages. Are local internet outages severe enough to decrease a municipality's exposure to social media? We investigate this question by using a sample of posts from the AfD Facebook page for which we know the users' locations.[18] Figure 5(a) plots the local number of posts against the intensity of local internet outages. Local Facebook activity falls with outage intensity and is close to 0 as soon as we observe more than 0.25 outage reports per 10,000 inhabitants. Figure C.3 in the Online Appendix shows that we observe significantly fewer posts and comments on Facebook for municipalities that experience an internet disruption. These results lend credence to the idea that exposure to social media content is reduced in the affected municipalities and not compensated by users accessing Facebook with their mobile phones.

If internet outages indeed reduce local social media exposure, we would expect them to mediate the capacity of social media to propagate antirefugee incidents. As described in Section 3.1, we test this hypothesis by interacting the main terms of interest $AfD\ Users/Pop_i \times Refugee\ Posts_t$ with $Internet\ Problems_{it}$, our dummy for severe internet disruptions. We graphically illustrate the results in Figure 5(b). The binned scatter plot is almost identical to Figure 4, except that we plot a separate slope for municipalities that experience an internet outage. This reveals a striking pattern: although antirefugee attacks increase with antirefugee posts, this relationship disappears in municipalities that experience an internet outage. This holds true for municipalities with high and low Facebook usage.

Figure 5(b) implies that internet outages have a substantial attenuating effect. Consider the pattern in panel (a). Without outages, there is a strong correlation of refugee posts and attacks. During outages, the correlation is essentially zero. This means that the outage effect is larger than the baseline estimate for *AfD Users/Pop.* × *Refugee posts*, which is given by the slope difference of the dotted lines in panels (a) and (b). We interpret this as evidence that cutting of users from social media completely has large effects.

We next estimate versions of equation (2) and report the regression results in Table 3. Column (1) shows that internet outages reduce antirefugee violence. The coefficient of −0.003 implies that, during such outages, the probability of a refugee attack is 53% lower relative to the dependent variable mean (≈0.006). In Figure 6, we investigate the timing of this drop in incidents. Because the outages are relatively rare

17. In the Online Appendix, we show our results are robust to alternative definitions. We also exploit the eight major Facebook outages, which only vary by week. We discuss the results and their interpretation in turn.

18. These posts and comments are a sub-sample by users who publicly disclosed their location in their Facebook profiles.

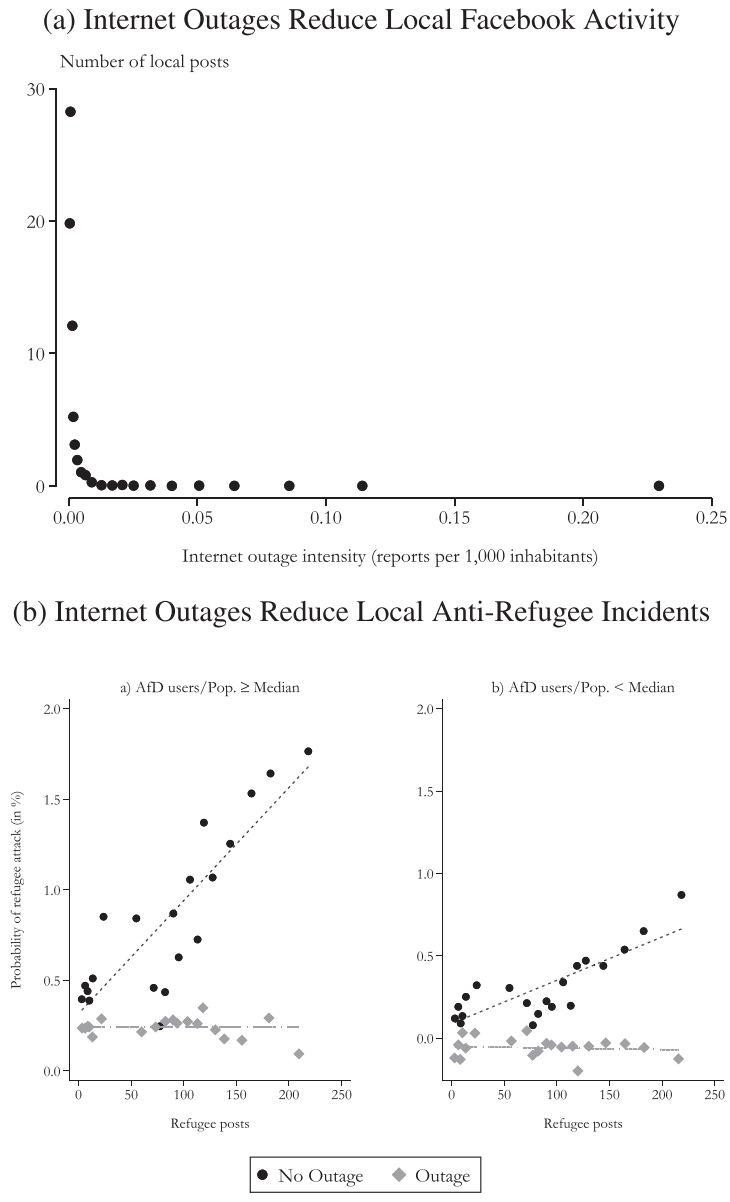

FIGURE 5. Quasi-experimental results from internet outages. Panel (a) shows a binned scatter plot of local posts on the AfD Facebook page as a function of the reports on internet outages in a given week. Panel (b) plots the average number of antirefugee attacks against our measure of antirefugee sentiment for municipalities above and below the median of *AfD Users/Pop*. Refugee attacks are binned by 20 quantiles of refugee posts. We additionally split towns by whether they experience an internet outage in a given week (gray squares). The number of antirefugee attacks is residualized with respect to population; hence, the number of attacks can be slightly below 0 in some bins.

TABLE 3. Local internet outages and social media transmission.

	(1)	(2)	(3)	(4)	(5)	(6)
Baseline interaction						
AfD users/Pop. × Refugee posts				0.024***	0.016**	0.016**
				(0.009)	(0.008)	(0.008)
AfD users/Pop. × Posts × Outage				−0.181***	−0.184***	−0.172***
				(0.058)	(0.058)	(0.057)
Outage interaction						
Outage	−0.003***	−0.000	−0.003**	−0.001	−0.002	−0.007
	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)	(0.008)
Refugee posts × Outage		−0.005***		−0.000	0.001	0.000
		(0.001)		(0.002)	(0.002)	(0.002)
AfD users/Pop. × Outage			−2.685	4.441	4.455	4.391
			(3.464)	(4.384)	(4.054)	(4.058)
Internet usage interaction						
Share broadband access × Outage						−0.000
						(0.000)
Internet domains/Pop. × Outage						0.021*
						(0.012)
Mobile broadband speed × Outage						0.000
						(0.000)
Observations	479,964	479,964	479,964	479,964	474,303	474,303
R-squared	0.082	0.082	0.082	0.082	0.084	0.084
Municipalities	4324	4324	4324	4324	4273	4273
Municipality FE	Yes	Yes	Yes	Yes	Yes	Yes
Week FE	Yes	Yes	Yes	Yes	Yes	Yes
Baseline controls [8] × Posts	Yes	Yes	Yes	Yes	Yes	Yes
All other controls [22] × Posts					Yes	Yes

Notes: This table presents the estimated coefficients from a regression of hate crimes against refugees on the interaction of local social media usage and anti-refugee sentiment as in equation (1). The dependent variable is a dummy for the incidence of a refugee attack. *AfD users/Pop.* is the ratio of people with any activity on the AfD Facebook page to population. *Refugee posts* is the Germany-wide number of posts on the AfD's Facebook wall containing the word refugee ("Flüchtling"). Internet outages are defined as municipality-weeks that are in the top quartile of the ratio of reported internet outages to population. The coefficient of "Refugee posts × Outage" is multiplied by 100 for readability. Columns (1)-(4) include the baseline controls. Columns (5) and (6) include all controls as in column (7) of table 2, interacted with *Refugee posts* (unreported). Column (6) further adds the interaction of broadband access and internet domains/pop. with local internet outages. Robust standard errors in all specifications are clustered by municipality. ***, **, and * indicate statistical significance at the 0.01, 0.05, and 0.1 levels, respectively.

in the municipality-week panel, the estimates are necessarily noisy. Nonetheless, we can see a reduction in antirefugee incidents that is sharply concentrated in the week of the internet outage.

Column (2) in Table 3 implies that this effect is driven by periods of high sentiment; it may also be driven by areas with many AfD Facebook users (column (3)) but the coefficient is not statistically significant. In columns (4) through (6), we estimate the full triple-difference model. Here, we estimate the effect of outages in areas with high social media use at times of high antirefugee sentiment. The estimates suggest that internet problems reduce social media's impact on antirefugee violence. Although the coefficient of refugee posts and social media exposure is similar to our baseline correlations, the triple interaction term with internet outages is negative

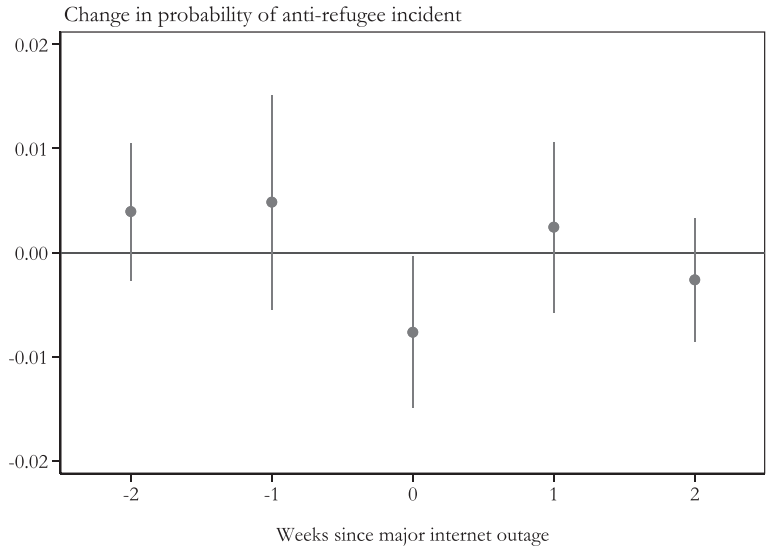

FIGURE 6. Internet outage event study. This figure plots estimates the estimates for δ from the event study regression $Attacks_{it} = \sum_{t=-2}^{2} \delta_{w=t} Outage_{it} + Fixed\ Effects + \varepsilon_{it}$, where *Outage* refers to internet outages in municipality i in week t. The 95% confidence intervals are based on standard errors clustered by municipality.

and statistically significant in all three specifications. Quantitatively, internet outages appear to mitigate the entire effect of social media. In line with the graphical evidence in Figure 5(b), we find that the triple interaction coefficient is larger than the baseline coefficient. Put differently, for a given level of antirefugee sentiment, there are fewer attacks in municipalities with high Facebook usage during an internet outage than in municipalities with low Facebook usage *without* an outage.

Could it be that the effect of internet outages is merely coincidental? As an alternative way of assessing statistical significance, we perform a randomization test. Instead of the actual internet disruptions, we randomly define 313 municipality-week pairs as placebo outages. We then estimate the same regression using 500 different sets of placebo outages. This allows us to evaluate the probability of finding a statistically significant coefficient in our dataset. Using this procedure, we find that more than 99% of the placebo triple interaction coefficients exhibit a lower t-statistic than our estimate. Our findings are thus unlikely to be purely coincidental. We show the full distribution of t-statistics from this randomization test in Figure C.6(a) in the Online Appendix.

The identifying assumption for internet outages in our framework is that they only have an effect on antirefugee hate crime through the reduced exposure to social media. Could it be that we observe reduced hate crimes because users are cut off from the internet generally, not from social media in particular? Two pieces of evidence support the idea that we capture a social media channel.

First, when we include interactions of internet disruptions with measures of internet usage (broadband access, per capita internet domains, and mobile internet access), our

main coefficient is unaffected (see column (6) in Table 3). The coefficients of the internet usage interactions are generally statistically insignificant or have the opposite of the expected sign. This is at least some indication that we are not merely capturing general internet usage. It also suggests that our findings are unlikely to capture that people are busy fixing internet access problems. If we were merely capturing such displacement effects, one would expect it to more strongly affect people in areas with high internet usage, which does not seem to be the case in the data. Second, after including the other interaction terms in columns (4) through (6), the coefficient on internet outages is no longer statistically significant. This result also supports the idea that internet outages reduce hate crime by limiting access to social media.

Another concern could be that hate crimes are less likely to be reported during internet outages. We believe this is unlikely to explain our findings because we analyze incidents that happened years in the past. Although internet outages might hamper the flow of information, it seems highly unlikely that incidents such as assault or property damages are *never* reported due to a temporary internet disruption. As further evidence, we limit our analysis to official reports by the police or the German parliament, for which social media reporting is an unlikely concern. This yields similar results (see column (1) of Table C.4 in the Online Appendix).

We also run a number of tests to rule out that our Germany-wide measure of refugee posts is affected by local internet outages. As stated previously, this appears unlikely because we focus on *local* disruptions to the internet; Table C.3 in the Online Appendix shows that the total number of internet outages in a given week is uncorrelated with the total number of Facebook posts. The outage results are also robust to using a leave-one-out measure of refugee posts (column (2)), Germany-wide posts in the previous week (column (3)), and an alternative measure based on Google search intensity for the word refugee (*Flüchtling*) in column (4). The implied magnitudes are almost equivalent.[19] This suggests that the outage effect is driven by exposure rather than the production of antirefugee content. In Table C.6 in the Online Appendix, we show additional robustness checks for alternative transformations of the dependent variable. The findings remain robust throughout. Table C.7 in the Online Appendix shows that the results also hold using alternative definitions of the outage dummy.

Facebook Outages. As further evidence for the social media transmission mechanism, we use eight major Germany-wide Facebook outages as a source of exogenous variation specific to social media access. Table C.2 in the Online Appendix outlines the details of each of the eight outages and links to relevant press reports. By definition, these outages are Facebook-specific and therefore do not affect other potential channels of online transmission.

Table C.3 in the Online Appendix shows that these outages are large enough to disrupt weekly activity on right-wing social media. Column (1) and (2) show that,

19. To see this, consider the effect implied by dividing the triple interaction coefficients by the standard deviation of these salience metrics. This suggests that internet outages have a mediating effect of 9.6, 10.5, and 11.0 for the AfD posts about refugees, the leave-one-out measure, and Google trends, respectively.

TABLE 4. Facebook outages and social media transmission.

	(1)	(2)	(3)	(4)	(5)	(6)
Baseline interaction						
AfD users/Pop. × Refugee posts			0.027***	0.027***	0.021**	0.021**
			(0.010)	(0.010)	(0.009)	(0.009)
AfD users/Pop. × Posts × Outage			−0.040*	−0.040*	−0.046**	−0.046**
			(0.021)	(0.021)	(0.022)	(0.022)
Additional outage coeffcients						
Outage	−0.001***					
	(0.000)					
AfD users/Pop. × Outage		−2.222*	1.164	1.164	1.367	3.230
		(1.273)	(1.833)	(1.833)	(1.862)	(1.969)
Observations	479,964	479,964	479,964	479,964	474,303	474,303
R-squared	0.079	0.082	0.082	0.082	0.084	0.084
Municipalities	4,324	4,324	4,324	4,324	4,273	4,273
Municipality FE	Yes	Yes	Yes	Yes	Yes	Yes
Week FE		Yes	Yes	Yes	Yes	Yes
Baseline controls [8] × Posts	Yes	Yes	Yes	Yes	Yes	Yes
All other controls [22] × Posts					Yes	Yes
All controls [30] × Outages						Yes

Notes: This table presents the estimated coefficients from a regression of hate crimes against refugees on the interaction of local social media usage and antirefugee sentiment as in equation (1). The dependent variable is a dummy for the incidence of a refugee attack. *AfD users/Pop.* is the ratio of people with any activity on the AfD Facebook page to population. *Refugee posts* is the Germany-wide number of posts on the AfD's Facebook wall containing the word refugee ("Flüchtling"). Facebook outages refer to weeks in which Facebook experienced considerable disruptions; see the Online Appendix for more details on how these are defined. Note that the other interaction terms *Outage*, *Refugee posts* and *Outage* × *Refugee posts* are absorbed by the week fixed effects in columns (3)–(5). Columns (1)–(3) include the baseline controls. Columns (4) and (5) include all controls as in column (7) of table 2, interacted with *Refugee posts*. Column (5) adds the interaction of these control variables with Facebook outages. Robust standard errors in all specifications are clustered by municipality. ***, **, and * indicate statistical significance at the 0.01, 0.05, and 0.1 levels, respectively.

during weeks with Facebook outages, there are on average 11% fewer new total posts and 24% fewer posts about refugees on the AfD page.[20] There is no evidence of such an effect in the week before. Column (5) shows that Facebook outages are also uncorrelated with the total number of weekly internet disruptions ($t = -0.41$).

We next present the results of interacting Facebook disruptions analogous to the internet outages in Table 4. The results again reveal a clear pattern. The coefficient of −0.001 in column (1) shows that the probability of an antirefugee incident is around 18% lower in weeks with major Facebook outages (relative to the unconditional probability of an attack). Figure C.4 in the Online Appendix suggests that the timing of this effect is concentrated in the week of the Facebook outage, without significant effects in the week before or after the outage. Because we solely rely on the weekly variation from the few major Facebook outages, the estimates are noisier than those

20. The average number of refugee posts in the time series is around 84. The coefficient estimate of 19.880 implies an effect of Facebook outages on posts of $-19.880/84 \approx 0.24$ relative to the mean.

for internet outages. Column (2) shows that, intuitively, this effect is also larger in areas with many users on the AfD Facebook page. The coefficient of 2.222 suggests that Facebook outages reduce the probability of a hate crime by 12% more for a one standard deviation increase in *AfD users/Pop.*[21] This is additional evidence that social media *per se* might affect hate crimes.

Next, we introduce the triple interaction of Facebook outages with social media usage and our refugee salience measure. The triple interaction is negative and statistically significant in all three specifications in columns (3) through (5). Quantitatively, we find that Facebook disruptions fully undo the baseline correlation of refugee attacks and exposure to social media sentiment. For example, consider that the coefficient of *AfD users/Pop.* and *Refugee Posts* is 0.027 in column (4) but −0.04 on the triple interaction. This implies that, in weeks of major Facebook outages, heightened refugee sentiment is not associated with a differential increase of antirefugee attacks in municipalities with higher Facebook usage.

It is worth noting that we would expect the Facebook outage coefficients to differ in magnitude from the internet outage coefficients. This is because Facebook outages eliminate the differential exposure *between* areas with high and low social media usage to antirefugee posts. In contrast, internet outages further exploit variation *within* municipalities. Because within-municipality variation induced by internet outages appears to matter more in our setting, we find smaller coefficients for Facebook outages.

We again perform a randomization test to assess the statistical significance of the Facebook outage results. We randomly assign placebo Facebook outages to eight weeks in our data, excluding the weeks in which we identified Facebook outages. We then estimate the same regression using 500 different sets of placebo outages. Using this procedure, we find that 92% of the placebo triple interaction coefficients exhibit smaller *t*-statistics. We show the full distribution of *t*-statistics from this randomization test in Figure C.6(b) in the Online Appendix. This confirms that our findings are unlikely to be a matter of coincidence.

Taken together, the evidence here suggests that the relationship of antirefugee sentiments online and hate crimes is attenuated by Facebook and internet outages. These results are most consistent with a causal propagation effect of social media.

In the Online Appendix, we conduct additional robustness exercises for our outage results. In Table C.5 in the Online Appendix, we show a range of different standard errors. We also assess our results' robustness to different transformations of the refugee attack variable and estimation methods in Table C.6 in the Online Appendix. Our results are similar when we use the number of attacks, log(1+refugee attacks) or the ratio of refugee attacks to asylum seekers as dependent variable. In all cases, the estimated coefficients are highly statistically significant.

21. In unreported results, we also find that the interaction of Facebook outages with refugee posts has a statistically significant negative coefficient.

3.4. Additional Results

Other Posts on the AfD Facebook Page. If the channel we uncover is indeed specific to refugees, we would expect a weaker correlation between refugee attacks and posts about other topics on the AfD Facebook page. We test this hypothesis in Table D.2 in the Online Appendix, where we plot the baseline estimation with refugee posts in column (1) for convenience. We also report coefficients for standardized post measures (with a mean of zero and standard deviation of one) in square brackets to compare coefficient sizes across the different posts. Next, we estimate equation (1) using all posts except those containing the word *refugee* ("Flüchtling") in column (2). The estimate is statistically indistinguishable from zero. We also repeat our baseline test using posts containing the words "Muslim," "Islam," or "EU"—the latter is motivated by the AfD's long-standing criticism of the European Union. For all these terms, we find no significant relationship between the number of posts and the number of attacks; all estimated coefficients are considerably smaller in standardized terms compared to the baseline measure. This shows the specificity of our refugee measure: the correlation we capture does not appear to be an artifact of general antiminority sentiment, but rather a predictable result of increased animosities towards refugees on social media in particular weeks.

Intensive Margin of Facebook Usage. If social media works as the propagating mechanism for hate speech, we would also expect its effect to increase with how frequently users interact with the AfD Facebook page. We explore this issue empirically in Table D.3, where we interact our main interaction term with the total number of local posts on the AfD wall and the number of comments and likes on AfD posts, all scaled over the number of AfD users in a municipality.[22] These measures of usage intensity are not systematically correlated with local Facebook penetration, city size, or population density. As such, they create additional variation in social media engagement across towns.

The results suggest that local engagement on Facebook matters: all three triple interaction terms are positive and statistically significant. Consistent with the hypothesis that social media enables hateful sentiment to spread, a higher reach per AfD user increases the correlation of social media exposure with hate crimes. These interactions work on top of our baseline interaction term, which remains similar in magnitude and highly statistically significant throughout. The smallest coefficient on the triple interaction term of 0.001 in column (3) implies that a one standard deviation increase in likes per user (around 12) increases the baseline coefficient by 25%.[23]

22. Note that we can only construct these measures on the intensive margin of municipalities where we can identify at least one AfD user. Our baseline results also hold in this sub-sample, which we show in Table E.2 in the Online Appendix.

23. To see this, consider that the total implied estimate including interaction is calculated as $0.001 \times 12 \approx 0.012$, which is about 25% than the baseline coefficient of 0.049.

TABLE 5. News shock salience and hate crime propagation.

	Brexit (1)	Trump (2)	UEFA EM 2016 (3)
AfD users/Pop. × Refugee posts	0.071***	0.096***	0.067***
	(0.018)	(0.022)	(0.017)
AfD users/Pop. × Posts × News shock	−0.019**	−0.009***	−0.002**
	(0.008)	(0.003)	(0.001)
Observations	495,726	495,726	495,726
R-squared	0.078	0.079	0.078
Municipalities	4,466	4,466	4,466
Municipality FE	Yes	Yes	Yes
Week FE	Yes	Yes	Yes

Notes: This table presents the estimated coefficients from a regression of hate crimes against refugees on the interaction of local social media usage and antirefugee sentiment as in equation (1). The dependent variable is a dummy for the incidence of a refugee attack. *AfD users/Pop.* is the ratio of people with any activity on the AfD Facebook page to population. *Refugee posts* is the Germany-wide number of posts on the AfD's Facebook wall containing the word refugee ("Flüchtling"). The news shocks refer to the Google searches as indicated in the text. Robust standard errors in all specifications are clustered by municipality. ***, **, and * indicate statistical significance at the 0.01%, 0.05%, and 0.1% levels, respectively.

Distracting News Events. As an additional piece of analysis, we investigate the role of news shocks on the transmission of online hate speech to real-world actions. We build on the evidence in Durante and Zhuravskaya (2018), who show that the Israeli army is more likely to strike against Palestinian targets when US media outlets are distracted by other news events. In our case, we hypothesize that other important news events might distract people from the topic of refugees. This is somewhat analogous to Facebook outages in that we exploit additional exogenous weekly variation: if major news events act as a distraction, they should reduce the correlation of exposure to refugee salience with hate crimes.

To measure these news shocks, we obtain Google Trends data on weekly search interest on the terms "Brexit," "Trump," and "UEFA Euro 2016." Figure D.1 in the Online Appendix shows that these spike around the respective events. In Table D.4, we show that they are indeed associated with a crowding out of refugee salience: the share of posts about refugees is markedly lower during these key events. As an example, the spike in search interest for Brexit (100 on the Google search index) is associated with an almost 30% drop in the share of refugee posts (relative to the mean).

We next investigate whether, as a result, refugee salience has a weaker link with hate crimes in the weeks these major events attracted particular news attention. If this is the case, we would expect that these events *decrease* the correlation of social media transmission with refugee attacks. As before, we implement this by including the Google trends measures as a further interaction in our panel regressions.

Table 5 plots the results. For each of the events in columns (1)–(3), we find a significant negative coefficient on the number of antirefugee incidents for the triple interaction with distracting news. The negative sign of the coefficient indicates that,

during weeks of major news events, changes in antirefugee incidents correlate less with heightened refugee salience. As the salience of other events crowds that of refugees, there are smaller increases of hate crimes in municipalities with more AfD social media users.

3.5. Differences Between Social Media And Traditional Media

How does social media differ from traditional media? And could such differences partially explain our results? Existing work has highlighted the ability of users to self-select and interact on social media (e.g., Schmidt et al. 2017). In the following, we highlight three aspects of far-right social media in Germany that may make it a particularly effective transmission mechanism for antirefugee sentiment compared to mainstream news sources.

First, Figure 7(a) shows that the share of content about refugees is consistently higher on the AfD's Facebook page compared to traditional news outlets in the Nexis data. The share of refugee mentions on Facebook is also far more volatile and spikes coincide more clearly with salient news events like Merkel's "Wir schaffen das" speech or the Cologne New Year's Eve incidents. In both of these examples, the share of refugee posts on right-wing social media is nearly 100% higher than the share of news stories on refugees, which is consistent with the idea that the topics discussed on Facebook are considerably narrower than in traditional media.

In Figure D.3.(a) in the Online Appendix, we show that this also holds true in a like-for-like comparison of the share of refugee posts on the AfD's Facebook page relative to the Facebook pages of five major German news outlets. AfD users post twice as much about refugees compared to the next-ranked newspaper. This suggests that the narrowness of content is unlikely to be explained only be the editorial constraints (e.g., space limits in newspapers) of traditional media outlets. Instead, self-selection of like-minded people into the AfD Facebook page likely also play a role. Combined with the interactive nature of social media, this result points towards an antirefugee group dynamic on the AfD's Facebook page.

Second, as argued by Sunstein (2017), self-selection of like-minded people can lead to the expression of more extreme viewpoints. To shed light on this hypothesis empirically, we compare the full text of news reports about refugees with posts on the AfD Facebook page. Existing reports on far-right hate speech on social media highlight three characteristics as typical (see for example Dinar et al. 2016; Kreißel et al. 2018; Ott and Gür-Seker 2019): (1) a belief to speak for the "true will" of the people, that is the in-group (citizens) compared to the out-group (refugees); (2) an opposition to "elites", in particular politicians and the media, who supposedly mislead or betray the people in an undemocratic way; and (3) a legitimization of discrimination against refugees by highlighting crimes by refugees, an alleged incompatibility of cultural differences, and negative repercussions for vulnerable "locals" (e.g., women, children or pensions).

We find evidence for all three of these features of right-wing hate speech on the AfD's Facebook page. Our approach is to investigate which words occur with

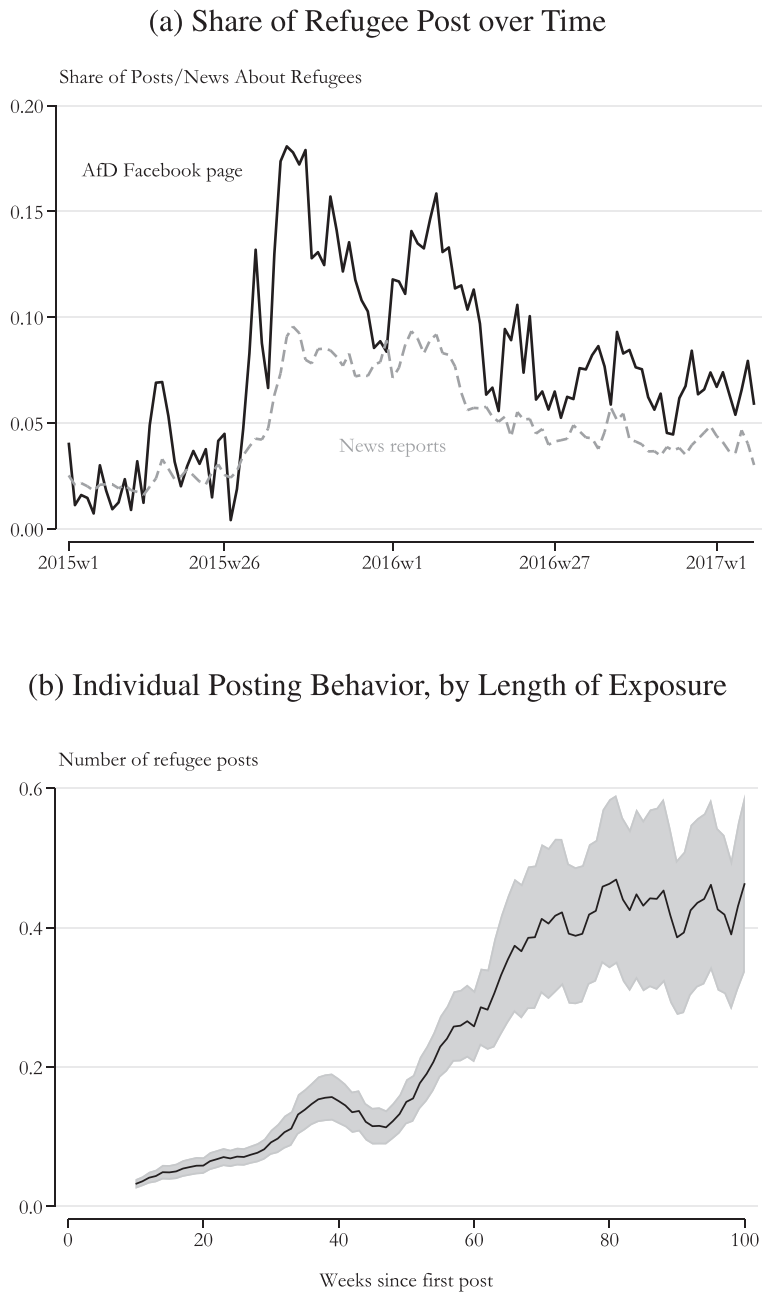

FIGURE 7. Highlighting social media echo chambers. Panel (a) plots the share of posts/reports about refugees on the AfD Facebook page and major German news outlets from Nexis. Panel (b) plots the 10-week moving average of the number of refugee posts per person as a function of a user's time spent on the AfD Facebook page, proxied by the time since the first post. The shaded area indicates 95% confidence intervals.

a higher probability in posts on the AfD page relative to news reports in the Lexis corpus.[24] We filter words using the word stems of the German terms for people, elite, democratic, press, crime, foreign, culture, refugee, betrayal, and several vulnerable groups (pensioners, children, women, homeless).

The results of this exercise in Table 6 reveal a clear pattern (see also Table D.5 in the Online Appendix). As one example, the term "Volksbetrug" (betrayal of the people) is 1,715 times more likely to appear on the AfD page than in traditional news outlets. Criticism of "elites" and the media are also far more frequent. Another main difference is how often crimes by refugees are discussed, based on the use of loaded terms like "Flüchtlingskriminalität" (refugee crime). We see expressed fears about "Fremdkulturen" (foreign cultures) and "Burkafrauen" (burka women). This analysis clearly shows that far-right ideas that have widely been interpreted as hate speech are far more pervasive on the AfD page than in traditional media reports.

We find similar results using a text analysis approach using machine learning. In particular, we train a L1 regularized logistic regression model classifier that predicts whether a text comes from the AfD Facebook page or a traditional media outlet. The classifier thereby identifies the words with the highest predictive ability for posts on the AfD Facebook page. Figure D.4 in the Online Appendix shows a word cloud of the 100 words that best separate social media from traditional media content, based on the model with the highest cross-validated out-of-sample F1 scores.[25] The size of the words represents the magnitude of the coefficients as a measure of variable importance. Consistent with the findings in Table 6, critiques of establishment parties and the economic or social costs of refugees are among the words that most uniquely identify posts on the AfD page.

Third, we investigate how individuals' posting behavior varies with the length of exposure to far-right social media content. We construct a balanced panel of users' activity on the AfD's Facebook page. In Figure 7(b), we show users' average number of posts about refugees since their first post on the page. To avoid that a changing sample composition drives our results, we restrict the analysis to the approximately 60% of users who first interacted with the AfD page before June 2015 and thus have been active on it for at least 100 weeks. The results are similar without this restriction.

The frequency of refugee posts strongly increases with users' duration on Facebook: within the first year, the average user on the AfD page goes from close to zero to posting at least once about refugees every two weeks.[26] This result suggests that the AfD page does not merely attract already active Facebook users with right-wing views, but may increase the willingness of people to express antirefugee views over time.

24. We calculate word probabilities for each corpus by dividing the number of times a word is mentioned ($Word_i$) by the total number of words in the corpus ($\sum Words_i$), for example, $P(Word_i^{News}) = Word_i^{News} / \sum Words_i^{News}$. The relative probability is the ratios between the two calculated the two probabilities, that is, $P(Word_i^{Facebook}) / P(Word_i^{News})$.

25. Note that the model was run in German and the words translated by the authors afterwards. For more details on the machine learning model, see the notes to Figure D.4 in the Online Appendix.

26. The same holds true for the total number of posts (see Figure D.3(b) in the Online Appendix).

TABLE 6. Relative word frequencies on the AfD Facebook page.

Rank	Word	Translation	Relativ prob.
Panel A: Flücht (refugee)			
1	Flüchtlingsenklaven	Refugee enclave	780
2	Flüchtlingslüge	Refugee lie	693
3	Flüchtlingsirrsinn	Refugee insanity	650
4	Flüchtlingsmafia	Refugee mafia	520
5	Flüchtlingsbefürworter	Refugee supporter	520
Panel B: Krimi (crime)			
1	Regierungskriminalität	Goverment crime	1,300
2	Diskriminierungsgesetze	Antidiscrimination laws	520
3	Schwerstkriminellen	Dangerous criminals	260
4	Fluechtlingskriminalität	Refugee crimes	260
5	Kriminalitätssteigerung	Increase in crime	260
Panel C: Presse (media)			
1	Freie Presse	Free press	390
2	Propagandapresse	Propaganda press	260
3	Presseempfang	Press meeting	260
4	Meinungspresse	Opinionated media	260
5	Nazipresse	Nazi media	260
Panel D: Volk (people)			
1	Volksbetrug	Betrayal of the people	1,715
2	volksfeindlich	Hostile to the people	780
3	volksverdummenden	Brainwashing the people	520
4	Volksverhetzungsparagraphen	Law against incitement	520
5	Volksprotesten	Protest by the people	260
Panel E: Verrat (betrayal)			
1	Volksverrats	Betrayal of the people	130
2	Vaterlandsverrat	Betrayal of the fatherland	43
3	Volksverrat	Betrayal of the people	43
4	Hochverrat	High treason	36
5	verratenen	Betrayed	32

Notes: This table plots the relative probability of words mentioned on the AfD Facebook page compared to reports by major German news outlets on Nexis. We report the results by groups of word stems identified as likely to reflecting right-wing hate speech on social media by previous work in Dinar et al. (2016).

This analysis also highlights an important distinction compared to existing research on media and violence. Yanagizawa-Drott (2014) Adena et al. (2015), and DellaVigna et al. (2014) all investigate the effect of nationalistic propaganda in settings of high ethnic tensions. In our setting, there is no nationalistic anti-minority propaganda in traditional media outlets. Rather, we find that social media provides an alternative

forum to exchange and spread extreme rhetoric and viewpoints for the fringe elements of society.

3.6. Mechanisms

In theory, multiple mechanisms could be consistent with social media playing a propagating role in real-life hate crimes. We discuss four mechanisms: information exchange, persuasion, collective action, and local spillovers. We provide suggestive evidence that collective action and local spillovers likely play a role in our setting.

First, social media might facilitate the exchange of information. In our setting, relevant information for potential perpetrators could, for example, include the locations of refugee homes and meeting points for demonstrations. We analyze the content of the refugee posts on the AfD Facebook to identify any post that might contain location information. To do so, we tag posts that either contain a zip code, mention the word "straße" (street), "weg" (path), "Flüchtlingsheim," "Asylantenheim," or "Flüchtlingsunterkunft" (all three translate to refugee home), or refer to a name of a German town or village.[27] We then manually check the content of tagged posts. This analysis suggests that although some locations like Berlin and Cologne are frequently mentioned in the posts as references to politicians and crimes committed by refugees, we find no mention about specific local information. We found no instance of zip codes or exact addresses. It hence appears unlikely that this channel is the primary driver behind our findings.

A second mechanism could be a persuasion channel, implying that social media persuades potential perpetrators that refugees may be dangerous or undeserving, which may then push some people over the edge. We believe that the timing in our setting makes this channel unlikely. In contrast to other work in Müller and Schwarz (2018) and Bursztyn et al. (2019), we focus entirely on high-frequency variation in social media posts and refugee violence. To the extent that social media changes people's attitudes, this is unlikely to happen in a single week and revert back after antirefugee salience has subsided. This is particularly true for the results on Facebook and internet outages: It seems unlikely that being cut off from social media during such disruptions reduces hate crimes because potential perpetrators become less xenophobic for a single week.

Third, social media could motivate collective action. Existing evidence in Enikolopov, Makarin, and Petrova (2016) and Manacorda and Tesei (2020) suggests that social media and mobile internet increase the incidence of protests. In our setting, users could coordinate to carry out hate crimes or learn about others' willingness to carry them out via social media. To investigate this, we rerun the panel regressions in equation (1) but limit refugee attacks to those undertaken by multiple perpetrators.[28]

27. We base the search on a comprehensive list of 2,061 German towns and 11,000 municipalities from the German statistical office, which covers villages with as little as 20 inhabitants.

28. We were able to hand-code the number of perpetrators for 28% of the hate crimes.

TABLE 7. Mechanism—antirefugee incidents, by number of perpetrators.

	Known perp. sample (1)	1 perp. (2)	<4 perp. (3)	≥4 perp. (4)
AfD users/Pop. × Refugee posts	0.010**	0.003	0.004	0.007**
	(0.005)	(0.002)	(0.003)	(0.003)
Observations	479,964	479,964	479,964	479,964
R-squared	0.081	0.037	0.046	0.055
Municipalities	4,324	4,324	4,324	4,324
Share of attacks	1	0.245	0.494	0.534
Mean of DV	0.002	0.000	0.001	0.001
Municipality FE	Yes	Yes	Yes	Yes
Week FE	Yes	Yes	Yes	Yes
Baseline controls [8] × Posts	Yes	Yes	Yes	Yes

Notes: This table presents the estimated coefficients from a regression of hate crimes against refugees on the interaction of local social media usage and antirefugee sentiment as in equation (1), where we vary the definition of the dependent variable based on the number of perpetrators. All control variables are interacted with the *Refugee posts* measure. Robust standard errors in all specifications are clustered by municipality. ***, **, and * indicate statistical significance at the 0.01, 0.05, and 0.1 levels, respectively.

In line with the collective action hypothesis, Table 7 suggests that our panel regression results are predominantly accounted for by cases with four or more perpetrators. We find no relationship for incidents with fewer than four perpetrators. Within the sub-sample where we can identify the number of perpetrators, these attacks account for a similar number of total incidents compared to the cases with more than four perpetrators. Hence, this finding is unlikely to be the result of limited statistical power.

Fourth, and somewhat relatedly, it could be that social media enables local spillovers, for example through "copy-cat" incidents. This mechanism suggests that potential perpetrators may use social media to learn about other attacks taking place, which could inspire them to carry out additional hate crimes. Because friendship networks on social media are clustered geographically (Bailey et al. 2018), this should be particularly pronounced for attacks happening nearby. We thus again rerun the panel regressions in equation (1) but now include a dummy variable if neighboring municipalities experience an attack in a given week.[29]

Table D.6 in the Online Appendix suggests that hate crimes happening in the same week nearby are associated with more antirefugee incidents. This correlation strongly interacts with the popularity of right-wing social media, particularly when antirefugee sentiment is elevated. In other words, having an attack in a neighbouring municipality

29. This is akin to the common correlated effects estimator proposed by Pesaran (2006) to hold common shocks constant.

is associated with a stronger correlation of exposure to right-wing social media and the probability of an antirefugee incident.[30]

Overall, our results appear to be most consistent with the idea that short-run bursts in antirefugee sentiment on social media can translate into real-life hate crimes by enabling coordination online, both through group actions and local spillovers.

3.7. How Many Refugee Attacks Are Caused By Online Hate Speech?

We conduct a back-of-the-envelope calculation of how many attacks against refugees would have taken place with lower antirefugee sentiment on right-wing social media. Given that we rely on high-frequency variation, this question is difficult to address. As our estimates are likely to pick up two separate facets of exposure to social media.

On one hand, it could be that exposure to antirefugee sentiment on social media merely affects the exact timing when refugee attacks occur without changing their total number. On the other hand, the time series of hate crimes and refugee posts on social media in Figure 2 exhibits prolonged overall increases in the number of antirefugee incidents with the onset of the refugee crisis. These increases are not easy to explain if antirefugee sentiment exclusively affects the timing of incidents. In our empirical setting, we cannot distinguish between these possibilities.

Despite this important caveat, we still believe it is instructive to assume social media indeed increases the number of hate crimes to illustrate the magnitudes of the results. We calculate the predicted number of attacks, based on the coefficient estimate of 0.024 from a regression with the baseline control variables (see column (1) in Table 2). Multiplying this coefficient with *AfD Users/Pop.* and *Refugee posts* gives us the estimated effect on antirefugee attacks. We sum over all observations to get the total predicted number of antirefugee attacks as a result of social media. This calculation implies that in absence of social media transmission on social media would result in 289 (10%) fewer antirefugee incidents.

4. Conclusion

Social media has become a powerful tool for sharing and disseminating information. In this paper, we investigate whether social media can play a role in propagating violent hate crimes. Our findings suggest that social media has not only become a fertile soil for the spread of hateful ideas but also motivates real-life action. By combining detailed local data on Facebook usage with user-generated content, we can shed light on the link between online posts and antirefugee incidents in Germany. Plausibly exogenous variation in disruptions to users' Facebook or internet access supports the view that some of the correlations we document reflect a causal effect.

30. Note that, although they are suggestive, we do not interpret these estimates as causal "peer effects," because we cannot distinguish them from common shocks (see Manski 1993).

Existing research shows local cultural attitudes towards foreigners are enormously persistent (e.g., Becker and Pascali 2019; Becker, Pfaff, and Rubin 2016; Voigtlander and Voth 2012, 2015). We extend this literature by showing that volatile, short-lived bursts in sentiment *within* a given location have substantial effects on people's behavior and that social media may play a role in their propagation. Our findings are particularly timely in light of recent policy debates about whether and how to "regulate" hate speech on social media. Such legislation may come at a high price: Because the lines between what constitutes free speech and hate speech can be blurred, regulation can open the door to censorship. Our work does, however, suggest that policymakers ignore online hate speech at their peril. Future research should investigate effective ways to tackle online hate speech. By quantifying the extent of the challenge, our paper takes a first step towards identifying potential harm arising from extended social media usage.

References

Adena, Maja, Ruben Enikolopov, Maria Petrova, Veronica Santarosa, and Ekaterina Zhuravskaya (2015). "Radio and the Rise of The Nazis in Prewar Germany." *Quarterly Journal of Economics*, 130, 1885–1939.

Alesina, Alberto and Eliana La Ferrara (2005). "Ethnic Diversity and Economic Performance." *Journal of Economic Literature*, 43, 721–761.

Bailey, Michael, Rachel Cao, Theresa Kuchler, Johannes Stroebel, and Arlene Wong (2018). "Social Connectedness: Measurement, Determinants, and Effects." *Journal of Economic Perspectives*, 32, 259–280.

BAMF (2016). "Aktuelle Zahlen zu Asyl." *Bundesamt für Migration und Flüchtlinge*.

Barberá, Pablo (2014). "How Social Media Reduces Mass Political Polarization: Evidence from Germany, Spain, and the US." Working paper, Job Market Paper, New York University, 46.

Bartik, Timothy J. (1991). *Who Benefits from State and Local Economic Development Policies?* Upjohn Press.

BBC (2017). "Social Media Warned to Crack Down on Hate Speech." https://www.bbc.com/news/technology-41442958

Becker, Sascha O. and Luigi Pascali (2019). "Religion, Division of Labor, and Conflict: Anti-semitism in Germany over 600 Years." *American Economic Review*, 109(5), 1764–1804.

Becker, Sascha O., Steven Pfaff, and Jared Rubin (2016). "Causes and Consequences of the Protestant Reformation." *Explorations in Economic History*, 62, 1–25.

Bessi, Alessandro, Fabiana Zollo, Michela Del Vicario, Antonio Scala, Guido Caldarelli, and Walter Quattrociocchi (2015). "Trend of Narratives in the Age of Misinformation." *PLOS ONE*, 10, 1–16.

Bhuller, Manudeep, Tarjei Havnes, Edwin Leuven, and Magne Mogstad (2013). "Broadband Internet: An Information Superhighway to Sex Crime?" *Review of Economic Studies*, 80, 1237–1266.

Boxell, Levi, Matthew Gentzkow, and Jesse M. Shapiro (2017). "Greater Internet Use Is Not Associated With Faster Growth in Political Polarization Among US Demographic Groups." *Proceedings of the National Academy of Sciences*, 114, 10612–10617.

Bursztyn, Leonardo, Davide Cantoni, Patricia Funk, and Noam Yuchtman (2017). "Polls, the Press, and Political Participation: The Effects of Anticipated Election Closeness on Voter Turnout." NBER Working Paper 23490, National Bureau of Economic Research, Inc. https://ideas.repec.org/p/nbr/nberwo/23490.html.

Bursztyn, Leonardo, Georgy Egorov, Ruben Enikolopov, and Maria Petrova (2019). "Social Media and Xenophobia: Evidence from Russia." Working Paper 26567, NBER. http://www.nber.org/papers/w26567.

Card, David and Gordon B. Dahl (2011). "Family Violence and Football: The Effect of Unexpected Emotional Cues on Violent Behavior." *Quarterly Journal of Economics*, 126, 103–143.
Colussi, Tommaso, Ingo E. Isphording, and Nico Pestel (2016). "Minority Salience and Political Extremism." IZA Discussion Papers 10417, Institute of Labor Economics (IZA). https://ideas.repec.org/p/iza/izadps/dp10417.html.
Dahl, Gordon and Stefano DellaVigna (2009). "Does Movie Violence Increase Violent Crime?" *Quarterly Journal of Economics*, 124, 677–734.
Del Vicario, Michela, Alessandro Bessi, Fabiana Zollo, Fabio Petroni, Antonio Scala, Guido Caldarelli, H. Eugene Stanley, and Walter Quattrociocchi (2016). "The Spreading of Misinformation Online." *Proceedings of the National Academy of Sciences*, 113, 554–559.
DellaVigna, Stefano, Ruben Enikolopov, Vera Mironova, Maria Petrova, and Ekaterina Zhuravskaya (2014). "Cross-Border Media and Nationalism: Evidence from Serbian Radio in Croatia." *American Economic Journal: Applied Economics*, 6, 103–132.
DellaVigna, Stefano and Eliana La Ferrara (2015). "Economic and Social Impacts of the Media." NBER Working Paper 21360, National Bureau of Economic Research, Inc. https://ideas.repec.org/p/nbr/nberwo/21360.html.
DellaVigna, Stefano and Matthew Gentzkow (2010). "Persuasion: Empirical Evidence." *Annual Review of Economics*, 2, 643–669.
Dinar, Christina, Theresa Mair, Simone Rafael, Jan Rathje, and Julia Schramm (2016). "Hetze gegen Flüchtlinge in Sozialen Medien." *Amadeu Antonio Stiftung*.
Durante, Ruben and Ekaterina Zhuravskaya (2018). "Attack When the World Is Not Watching? US News and the Israeli-Palestinian Conflict." *Journal of Political Economy*, 126, 1085–1133.
Eisensee, Thomas and David Strömberg (2007). "News Droughts, News Floods, and U. S. Disaster Relief." *Quarterly Journal of Economics*, 122, 693–728.
Enikolopov, Ruben, Alexey Makarin, and Maria Petrova (2016). "Social Media and Protest Participation: Evidence from Russia." CEPR Discussion Papers 11254, C.E.P.R. Discussion Papers. https://ideas.repec.org/p/cpr/ceprdp/11254.html.
Financial Times (2017). "Powerhouse Germany Badly Trailing Rivals in Broadband." https://www.ft.com/content/8f2e623c-7d1a-11e7-ab01-a13271d1ee9c.
Fiorina, Morris P. and Samuel J. Abrams (2008). "Political Polarization in the American Public." *Annual Review of Political Science*, 11, 563–588.
Fouka, Vasiliki and Hans-Joachim Voth (2013). "Reprisals Remembered: German-Greek Conflict and Car Sales during the Euro Crisis." CEPR Discussion Papers 9704, C.E.P.R. Discussion Papers. https://ideas.repec.org/p/cpr/ceprdp/9704.html.
Gabler, N. (2016). "The Internet and Social Media Are Increasingly Divisive and Undermining of Democracy." *Alternet*. https://www.alternet.org/2016/06/digital-divide-american-politics
Gavazza, Alessandro, Mattia Nardotto, and Tommaso Valletti (2018). "Internet and Politics: Evidence from U.K. Local Elections and Local Government Policies." *The Review of Economic Studies*, 86, 2092–2135.
Gentzkow, Matthew (2006). "Television and Voter Turnout." *Quarterly Journal of Economics*, 121, 931–972.
Hölig, Sascha and Uwe Hasebrink (2016). Reuters Institute Digital News Survey 2017: Ergebnisse für Deutschland, *Arbeitspapiere des Hans-Bredow-Instituts*, vol. Nr. 38. Verlag Hans-Bredow-Institut, Hamburg.
Jha, Saumitra (2013). "Trade, Institutions, and Ethnic Tolerance: Evidence from South Asia." *American Political Science Review*, 107, 806–832.
Kreißel, Philip, Julia Ebner, Alexander Urban, and Jakob Guhl (2018). "Hass auf Knopfdruck: Rechtsextreme Trollfabriken und das Ökosystem koordinierter Hasskampagnen im Netz." *Institute for Strategic Dialogue*.
Manacorda, Marco and Andrea Tesei (2020). "Liberation Technology: Mobile Phones and Political Mobilization in Africa." *Econometrica*, 88, 533–567.
Manski, Charles F. (1993). "Identification of Endogenous Social Effects: The Reflection Problem." *Review of Economic Studies*, 60, 531–542.
Müller, Karsten and Carlo Schwarz (2018). "From Hashtag to Hate Crime: Twitter and Anti-Minority Sentiment." Available at SSRN: https://ssrn.com/abstract=3149103.

New York Times (2016). "How Facebook Warps Our Worlds, By Frank Bruni." https://www.nytimes.com/2016/05/22/opinion/sunday/how-facebook-warps-our-worlds.html.
New York Times (2017a). "How Fiction Becomes Fact on Social Media, By Benedict Carey." https://www.nytimes.com/2017/10/20/health/social-media-fake-news.html.
New York Times (2017b). "Seeking Asylum in Germany, and Finding Hatred, By Ainara Tiefenthäler, Shane O'neill and Andrew Michael Ellis." https://www.nytimes.com/video/world/europe/100000005090433/libyan-migrant-bautzen-germany.html.
OECD (2016). "Broadband Statistics." *OECD Digital Economy Outlook*. http://dx.doi.org/10.1787/888933585229.
Oksanen, Atte, James Hawdon, Emma Holkeri, Matti Näsi, and Pekka Räsänen (2014). "Exposure to Online Hate Among Young Social Media Users." *Soul of Society: A Focus on the Lives of Children & Youth*, 18, 253–273.
Ott, Christine and Derya Gür-Seker (2019). Rechtspopulismus und Social Media: Wie Wortgebräuche in Social Media sprachkritisch betrachtet werden können, 279–318. Peter Lang AG. http://www.jstor.org/stable/j.ctvnp0hrq.12.
Pariser, Eli (2011). *The Filter Bubble: What the Internet Is Hiding From You*. Penguin UK.
Pesaran, M. Hashem (2006). "Estimation and Inference in Large Heterogeneous Panels with a Multifactor Error Structure." *Econometrica*, 74, 967–1012.
Pew Research Center (2018). "News Use Across Social Media Platforms 2018." Technical report.
Schmidt, Ana Lucía, Fabiana Zollo, Michela Del Vicario, Alessandro Bessi, Antonio Scala, Guido Caldarelli, H. Eugene Stanley, and Walter Quattrociocchi (2017). "Anatomy of News Consumption on Facebook." *Proceedings of the National Academy of Sciences*, 114, 3035–3039.
Stephens-Davidowitz, Seth (2014). "The Cost of Racial Animus on a Black Candidate: Evidence using Google Search Data." *Journal of Public Economics*, 118, 26–40.
Sunstein, Cass R. (2009). *Republic.com 2.0*. Princeton University Press.
Sunstein, Cass R. (2017). *#Republic: Divided Democracy in the Age of Social Media*. Princeton University Press.
The Guardian (2017). "CPS to Crack Down on Social Media Hate Crime, says Alison Saunders, by Vikram Dodd." https://www.theguardian.com/society/2017/aug/21/cps-to-crack-down-on-social-media-hate-says-alison-saunders.
Voigtlander, N. and H.-J. Voth (2012). "Persecution Perpetuated: The Medieval Origins of Anti-Semitic Violence in Nazi Germany." *Quarterly Journal of Economics*, 127, 1339–1392.
Voigtlander, Nico and Hans-Joachim Voth (2015). "Nazi Indoctrination and Anti-Semitic Beliefs in Germany." *Proceedings of the National Academy of Sciences of the United States of America*, 112, 7931–7936.
Yanagizawa-Drott, David (2014). "Propaganda and Conflict: Evidence from the Rwandan Genocide." *Quarterly Journal of Economics*, 129, 1947–1994.

Supplementary Data

Supplementary data are available at *JEEA* online.