

To cite: Shruti Phadke and Tanushree Mitra. 2021. Educators, Solicitors, Flamers, Motivators, Sympathizers: Characterizing Roles in Online Extremist Movements. Proc. ACM Hum.-Comput. Interact. Computer Supported Cooperative Work (CSCW' 21), (accepted May 2021).

## **Educators, Solicitors, Flamers, Motivators, Sympathizers: Characterizing Roles in Online Extremist Movements**

SHRUTI PHADKE, Information School, University of Washington, USA

TANUSHREE MITRA, Information School, University of Washington, USA

Social media provides the means by which extremist social movements, such as white supremacy and anti-LGBTQ, thrive online. Yet, we know little about the roles played by the participants of such movements. In this paper, we investigate these participants to characterize their roles, their role dynamics, and their influence in spreading online extremism. Our participants—online extremists accounts—are 4,876 public Facebook pages or groups that have shared information from the websites of 289 Southern Poverty Law Center (SPLC) designated extremist groups. Guided by theories of participatory activism, we map the information sharing features of these extremists accounts. By clustering the quantitative features followed by qualitative expert validation, we identify five roles surrounding extremist activism—*educators*, *solicitors*, *flamers*, *motivators*, *sympathizers*. For example, *solicitors* use links from extremist websites to attract donations and participation in extremist issues, whereas *flamers* share inflammatory extremist content inciting anger. We further investigate role dynamics such as, how stable these roles are over time and how likely will extremist accounts transition from one role into another. We find that roles core to the movement—*educators* and *solicitors*—are more stable, while *flamers* and *motivators* can transition to *sympathizers* with high probability. Finally, using a Hawkes process model, we test which roles are more influential in spreading various types of information. We find that *educators* and *solicitors* exert the most influence in triggering extremist link posts, whereas *flamers* are influential in triggering the spread of information from fake news sources. Our results help in situating various roles on the trajectory of deeper engagement into the extremist movements and understanding the potential effect of various counter-extremism interventions. Our findings have implications for understanding how online extremist movements flourish through participatory activism and how they gain a spectrum of allies for mobilizing extremism online.

CCS Concepts: • Human-centered computing → Empirical studies in collaborative and social computing; Social media; • Social and professional topics → Hate speech.

Additional Key Words and Phrases: online communities, extremist groups, social media, information sharing, participatory activism

### **ACM Reference Format:**

Shruti Phadke and Tanushree Mitra. 2018. Educators, Solicitors, Flamers, Motivators, Sympathizers: Characterizing Roles in Online Extremist Movements. *Proc. ACM Meas. Anal. Comput. Syst.* 37, 4, Article 111 (August 2018), 35 pages. <https://doi.org/10.1145/1122445.1122456>

---

Authors' addresses: Shruti Phadke, Information School, University of Washington, USA, Seattle, USA; Tanushree Mitra, Information School, University of Washington, USA, Seattle, USA.

---

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org).

© 2018 Association for Computing Machinery.

2476-1249/2018/8-ART111 \$15.00

<https://doi.org/10.1145/1122445.1122456>

## 1 INTRODUCTION

According to the Southern Poverty Law Center (SPLC)—a non profit organization dedicated to monitoring extremist activity in the United States—Facebook groups serve as the primary avenue for extremists to recruit new members and spread extremist propaganda [75]. Take for example, *Pissed off White Americans*—a Facebook group harboring 64K followers on their public Facebook page. It describes itself as: “*We are the sons and daughters of European Heritage, and we are tired of being treated as second class citizens. LOUD, PROUD, and very PISSED OFF!*” This page frequently posts links from a known white supremacy group, White Rabbit Radio—a Southern Poverty Law Center (SPLC) designated white nationalist group [40]. Their posts often contain calls for actions against white genocide—a white supremacy belief that there is a deliberate attempt to wipe out the white race to promote reproduction of other races. Another Facebook group, *White Lives Matter Movement* posts informational content explaining what it means to be white or “aryan” on their public Facebook page. Sharing links from the website of an SPLC designated neo-nazi group—National Vanguard [74]—they suggest norms for racial segregation and educate readers about white identity. While the *Pissed off White Americans* page uses link sharing for *soliciting* their readers’ engagement, *White Lives Matter Movement* Facebook page *educates* their readers about white identity and cultural norms. In other words, both pages play different **roles** in advancing their white supremacy ideology online. Moreover, by sharing links from the websites of known extremist organizations, both pages become participants in the extremist ecosystem on Facebook. We call them **extremist accounts**—Facebook pages and groups involved in actively sharing information from any of the 289 Southern Poverty Law Center (SPLC) designated extremist groups, and thus, operating as key players in sustaining and growing the extremist movement.

In this work, we explore the ecosystem of such online *extremist accounts* through the lens of participatory activism—the potential and magnitude of individuals and groups to engage in sociopolitical issues [57]. Previous scholars have largely studied the importance of participatory activism in positive social justice movements, such as feminist practices [33, 52] or raising awareness against police brutality [4]. On the contrary, only a handful of studies have highlighted how anti-social movements, also adopt similar participatory activism to promote a positive notion of their brand [38]. For example, terrorist organizations such as ISIS used participatory activism practices to promote the caliphate as a way of life and recruited combatants [38]. Despite their clear presence on social media, studies analyzing online anti-social movements are rare. In particular, U.S based domestic extremist movements with white supremacy, anti-LGBTQ, or anti-Immigration agendas that predominantly operate online, have not been investigated through the lens of participatory activism. In our work, we bridge this gap. Using the lens of participatory activism, we consider extremist accounts to play various social roles towards advancing extremist movements online. Specifically, we identify 4,876 extremist accounts involved in sharing links from 289 Southern Poverty Law Center (SPLC) designated extremist groups using Facebook’s CrowdTangle API. We obtain the public Facebook activity of 4,876 extremist accounts over two years (Jan’18 to Dec’19) to model their online roles in extremist movements. We further study their role dynamics and assess how influential roles are in spreading various types of information sources. We first ask:

### RQ1: What social roles are played by extremist accounts in online extremist movements?

To answer, we consider extremism as a social movement—a collective effort by a group of people aimed towards changing the society in a way that aligns with the movement’s goals [13]. Guided by theories of participation in social movements, we explore underlying behaviors of the accounts across three dimensions: drives for participation, engagement trends, and strategies of mobilization. Using features informed from these dimensions and qualitative expert validation, we identify five

roles played by extremist accounts in forwarding their social movements: *solicitors*—who solicit participation and funds for the extremist movement, *educators*—accounts that share intellectual content about extremism and prominently share and like extremist content, *flamers*—accounts that express and incite anger by posting inflammatory content, *motivators*—who are achievement oriented and go-getters of the extremist community and who post information that portrays a positive image of their extremist agenda, and *sympathizers*—accounts that are fringe supporters of the extremist movement who sparingly engage with links from the extremist websites.

These roles are based on the social media activities of the extremist accounts spanning over 6 months. How do these roles change over time? Do *sympathizers* get more involved in extremist movements and eventually become *solicitors* or *educators*? To understand, we investigate the role dynamics and ask:

**RQ2a: How long do the extremist accounts retain their original roles?**

**RQ2b: How likely are the extremist accounts in one role to transition to other roles?**

Upon measuring role retention, we find that 66% of the *solicitors* and 70% of the *educators* retain their roles throughout our analysis window. When analyzing role transition, we find that *motivators* and *flamers* transition into *sympathizers* with high probability.

As illustrated in our opening examples of *Pissed off White Americans* and *White Lives Matter Movement*, information can be valuable in putting out call for action or educating the readers about extremist ideologies. Researchers argue that information, especially disinformation spreads by flowing strategically or organically through various accounts on social media [76]. Specifically, Facebook accounts involved in extremist activities were found to use fake news, biased information and conspiracy sources to promote fundamentalist views [58]. Hence, we finally ask:

**RQ3: How influential are the roles in spreading various types of information?**

To answer, we first use Hawkes process to model the temporal and statistical characteristics of the information spread through the extremist accounts. Next, we use the parameters estimated from the Hawkes process to measure the influence of roles across four types of information sources: extremist, biased, fake news and conspiratorial sources. We consider a particular role to be *influential* in the spread of information, when a link posted by that role induces an account in another role to post the same link with high probability. We find that for information originating from extremist sources, *educators* and *solicitors* are the most influential in triggering other roles to also spread such content. Whereas, *motivators* influence other roles in spreading biased news, *flamers* are influential in the spread of fake news.

Overall our work makes the following contributions:

- We offer a framework for systematic operationalization of theoretically motivated characteristics of social movement participation.
- We present a data-driven and expert validated taxonomy of social roles in online extremist movements.
- Through a rigorous temporal statistical modeling of information flow, we reveal how different roles are influential in stimulating the spread of extremist, biased, fake and conspiratorial content on Facebook.

Our results allow us to understand how theories of social movement participation are reflected in online extremist movements and where the various roles are located on the trajectories of deeper engagement into extremism. Further, our work has implications in assessing which roles might benefit, and which might not, from interventions targeted to counter online extremism. Finally, we discuss how participatory activism might be democratizing the extremist movement and empowering the supporters with the resources and affordances to spread extremism online.

## 2 BACKGROUND

### 2.1 Participatory Activism and Extremist Movements

Activism means taking an action to effect social change [51]. Participatory activism is a kind of activism that is grounded in communities organizing to increase popular support for sociopolitical issues by strategically engaging in both, online and offline mechanisms for participation [66]. Considering the expanse of social media, researchers have connected participatory activism to the concept of “smart mobs”—people who are able to act in concert even if they don’t know each other [64, 66]. Under this new age activism, diverse set of people with overlapping interests can come together in solidarity and act without having to acknowledge who they are beyond what they support [66].

How does online participatory activism advance social causes? Previous research presents opposing perspectives on the effectiveness and the legitimacy of the online participatory activism. Researchers specifically use the terms “clicktivism” or “slacktivism” that refer to social media engagement in political micro-action [85] through low cost activities such as liking or sharing the content in order to raise awareness [67]. Writers on popular press comment that “clicktivism” or “slacktivism” are largely unproductive and ephemeral—an ideal type of activism for lazy generation [22, 47, 88]. However, communication and political science scholars argue that such low cost low risk participation is not only widespread but is also becoming a legitimate channel for political activism [24]. Specifically, Obar et. al. interviewed advocacy groups and found that all the groups viewed online participation as an effective tool for civic engagement and collective action [53]. Researchers have mostly investigated participatory activism in the context of positive social change. For example, in feminist movements, participatory activism resulted in enhanced understanding of feminism [33] and facilitated legal and public policy discourse surrounding gender based violence [52]. In other examples, participatory activism increased the visibility of police brutality against a group of environmentalists [4] and enabled successful spread of information in an authoritarian regime [62]. While these studies demonstrate how participatory activism can empower populations in social justice causes on one hand, on the other, anti-social movements, for example those advocating for terrorism and extremism, can also benefit from similar practices. One qualitative study emphasized the success of hashtag campaigns and information manifestos in ISIS’s territorial expansion in 2014 [38]. They argued that through numerous online accounts, or “media operatives,” ISIS was simultaneously able to promote its positive self-image for recruiting combatants and elicit participation from distant supporters. In other words, the “media operatives” played crucial roles in spreading diverse sets of narratives to enable recruitment into ISIS. Moreover, they did so by being a part of strategic information campaigns [38]. For example, ISIS media operatives urged their followers to download videos containing ISIS propaganda and re-upload them on various platforms so as to increase information dissemination while also evading content moderation [38].

Based on these studies, it is clear that similar networking and information sharing affordances contribute to both, positive social changes and anti-social movements. Yet, studies investigating the darker side of participatory activism are rare. In this paper, we fill this gap by asking: What are the different roles played by extremist accounts in extremist social movements? How stable or transitory are these roles? And how influential are these roles in spreading mis- and disinformation? We specifically focus on U.S domestic extremism, such as white supremacy and anti-LGBT movements and identify various roles through the lens of social movement and resource mobilization theories.

### 2.2 Extremism and Social Movements

Social movements are collective efforts to bring out the collective action fitting a specific goal or ideology [45]. Social movements emerge when constitution and function of the society is misaligned

with the movement's goals [45]. By this logic, extremist movements, such as white supremacy, become more active and aggressive when there is increased racial diversity in the society [90]. To understand the factors that may lead to the success of extremist movements, we look at theories analyzing the success of social movements. What makes social movements successful?

**2.2.1 Success of Social Movements:** Researchers attribute the success of social movements to the availability of human, material, and monetary resources [46]. Human resources include labor, experience, skills, and expertise of the members of the social movement. Material resources include property, office spaces supplies and monetary resources include funds contributed by the members. In other words, by participating in the movement, individuals make their human, material and monetary resources available to the movement [46]. These resources are then directed towards mobilizing supporters, transforming mass public into movement sympathizers and eventually bringing about the desired social change. However, the efficiency with which the resources are translated into action depends on various actors involved, such as volunteers offering human resources or supporters offering monetary resources through fundraising. Based on how various members are involved in the movement, researchers have proposed theoretical roles in social movements participation. Next, we summarize the theoretically proposed roles in social movements and discuss the challenges in adapting them to the online setting.

**2.2.2 Theoretically Identified Roles in Social Movement Participation:** Prior scholarly work has described participants in social movements from various theoretical perspectives, such as participant's role in an organizational hierarchy (e.g., leaders and followers) [48], their involvement in resource mobilization (e.g., members who distribute resources versus members who consume resources) [46] and whether they benefit from the social movement (e.g. stakeholders in the movement) [15, 46, 55, 86]. In organizational hierarchy, *leaders* are strategic decision-makers who mobilize *followers* [48]. Considering participants' involvement in resource mobilization, *constituents* provide resources to mobilize *adherents* [46] and gain sympathy from *bystanders* [80]. Moreover, scholars dichotomize the stakeholders of social movements into *potential beneficiaries*: population that directly benefit from the goals of social movement and *conscience participants*: supporters who may not directly gain from the success of the movement [15, 46, 55, 86]. How do these theoretically identified roles describe participation on online social media? Can we directly adopt the theoretical taxonomy (e.g., *leaders*, *followers*, *constituents* etc.) to describes roles in online extremist movements?

We identified two concerns with directly adopting a theoretical taxonomy, that consequently motivate the methodologies of our first research question. First, the roles derived from theories of social movements are based on physical social movement participation, such as protest events [55, 86]. Compared to online participation, physical participation requires increased commitment by members, such as on-the-ground physical presence at the protest events and significant time investment in the movement's activities [66]. On the contrary, in online activism, individuals can become a part of the movement by simply clicking, re-posting or writing short messages on relevant links [66]. Secondly, theoretically identified roles are harder to disambiguate from each other. Consider for example *leaders* defined as per the organizational hierarchy perspective and *constituents* from the resource mobilization perspective. While *leaders* lead the followers, they can also be *constituents*—the distributors of resources. Similarly, both *adherents* and *bystanders* from resource mobilization perspective can be viewed as *followers*. Considering these two points, instead of directly adopting theoretical taxonomy of roles, we consider the underlying characteristics of participation. Based on the framework of features derived from the theories of social movement participation, we develop a new taxonomy of roles for the online setting. Next, we detail the underlying characteristics that form a background for our role identification process.

### 2.3 RQ1 Background: Characteristics of Social Movement Participation

We summarize the theories in social movement participation across three dimensions: *drives for participation*, *engagement in the movement*, *strategies of mobilization*. These three dimensions, and the computational features derived from them, are at the crux of our methodology for identifying roles in the extremist movements. The first two columns in Table 2 summarize the models described in the next subsections.

**2.3.1 Drives for Social Movement Participation:** Theories of drives behind social movement participation can be broadly distilled in two groups—expectancy-value models and social-psychological models [17]. The expectancy-value models stress on the rationality of the participants whereby the individuals weigh costs and benefits of participation while decision making [36]. Particularly, they consider the risk-reward ratio of the involvement in the social movement before investing and mobilizing resources for their own benefit [44, 54]. Scholars pointed out one limitation of expectancy-value model that it underestimates the role of ideological drives and shared grievances in participants [36]. To fill this gap, social-psychological models attribute the movement participation to various psychological drives. Feelings of injustice, relative deprivation and moral outrage is at the heart of movement participation [83]. While such grievances related to social movement issues are ubiquitous, not all of them turn into protests [82]. Hence, social movement scholars also consider the sense of efficacy or achievement that drives people to participate [82]. For example, people might participate because they believe that their collective action can actually achieve the social change [21]. Further, researchers also stress the importance of group identity. The more people identify with the social groups involved in the movement, the more they are inclined towards participating in the movement [72]. Emotions also play an important role in driving people into the social movements. For example, anger is considered as *the prototypical emotion for protests* [84].

**2.3.2 Engagement Trends in Social Movements:** Various types of participants could adopt different degrees of engagement and commitment to distributing social movement related resources. Specifically based on the availability of resources, *constituents* actively distribute the resources in order to proselytize the *adherents* and *bystanders* [46]. Moreover, actors could participate in resource mobilization for a single or multiple social movements [46]. Another important aspect of the engagement is how it varies over time. People participate in the social movements with varying degrees of continuity. Corrigall-Brown characterizes such periodic engagement across four dimensions—*persistence*, *transfer*, *abeyance*, *disengagement* [10]. This characterization of participation trajectories is especially important on social media because it accords diverse group of individuals with varying degrees of interests and dedication.

**2.3.3 Strategies of Information Mobilization:** Social movements are goal-oriented. Participants in the social movements strategize the distribution and uses of resources to induce collective action and gain support [46]. Specifically, core members of the social movement mobilize resources in order to recruit volunteers, collect funds, spread their agenda and hold gatherings. Other researchers also find such solicitation strategies to be crucial for financing the social movements [6]. While solicitation may directly affect the progress of the movement, researchers also highlight resources that can indirectly create opportunities for collective action. Specifically in the social media setting, expressing opinions, thoughts and beliefs around political events [81] and reporting events to increase users' knowledge of public issues, political causes, and social movements [12, 81] are considered as key strategies of online activism.

statistics (per account)	min	max	mean	std.
posts	71	932K	7,067	30,574
link posts	23	78,571	1,614	3,915
extremist link posts	10	5,129	207	528
engagement				
page likes	106	1.8M	7,241	36,576
group members	35	2.2M	2,827	13,603

Table 1. Descriptive statistics for extremist accounts in our dataset. There are a total of 4,876 extremist accounts in our dataset.

## 2.4 Role of Information in Online Extremist Movements

Extremist movements have started leveraging social media to increase their outreach [14, 29]. According to Anti Defamation League (ADL) [2]—an anti-hate organization focusing on combating anti-semitism in USA—2018 was recorded at the most violent year in terms of deaths caused by domestic extremism since 1995 [63]. Several popular social media platforms were criticized in these attacks for enabling recruiting and spreading disinformation [19]. For example, the entire Christchurch mass shooting seemed to be orchestrated for the social media age [43]. The perpetrator of the mass shootings posted an 87 page manifesto on 8chan with anti-immigrant and anti-Muslim ideas and directed readers to a Facebook page where the shootings were live streamed. This is one of the many examples where extremists have used social media to broadcast their ideologies followed by violent events. Apart from educating the public about the extremist agenda, extremists also use online platforms to promote action [49]. Recently, the right-wing extremists have started leveraging meme and trolling culture to shape hate narratives [59]. Researchers found that extremist fringe groups collaboratively launder information even on mainstream social media platforms to appeal to the mass [58]; especially to a young audience [11].

Despite the strong visibility of extremist groups online, computational studies focusing on information operations of U.S. domestic extremist groups are rare. A recent qualitative work investigated information sharing by various extremist ideologies and find that hate groups use information sharing to recruit, radicalize and educate their follower base [58]. Specifically, they observed the use of links shared from biased news sources by white supremacy groups and links from websites hosting conspiratorial content in anti-Muslim and religious supremacy groups. In this paper we use rigorous temporal statistical models to quantify how influential various roles in extremist movements are in disseminating different types of information sources such as extremist content, fake news, biased news and conspiracies.

## 3 DATA

In this work, we focus on identifying roles in online extremist movements and assess how influential various roles are in spreading information from extremist, biased, fake news and conspiratorial information sources. Specifically, we focus on **extremist accounts**—public Facebook pages and groups that share links from extremist websites. In this section we detail our process of identifying the extremist accounts (Figure 1 (a) and (b)) and collecting their Facebook activity data. We first explain the problem of extremism on Facebook and then describe the data collection and pre-processing steps.

### 3.1 Extremism on Facebook Groups and Pages

Despite the policies against extremist content, Facebook is still crowded with groups and pages circulating and discussing violent ideas [34]. Very recently, Facebook banned a number of pages

and groups involved in the “boogaloo movement”—an anti-government movement by right wing extremists organizing for armed revolt [35]. However, this ban was not as effective as initially perceived. The pages and groups related to the boogaloo movement simply renamed and rebranded themselves with harmless-sounding names and continued their extremist activities [61]. According to the International Association of Chiefs of Police—a non-profit organization that is the world’s largest professional association for police leaders—Facebook groups and pages are at the center of extremist recruitment, radicalization, and mobilization [27]. They found that, by just re-posting and linking information from websites hosting graphic videos and other violent content, extremism related groups and pages are able to abide by Facebook’s policies against hate speech and still spread relevant information. In our first research question, we identify the roles played by various such extremist accounts in advancing the extremist movements online. In order to model these roles, we first need to identify the extremist websites, and then select the extremist accounts—Facebook pages/groups that share links from the extremist websites.

### 3.2 Identifying Extremist Websites

To identify extremist websites, we take help of the resources published by external organizations who are experts in social justice causes. Specifically, we refer to Southern Poverty Law Center’s (SPLC) website <sup>1</sup>. SPLC is a non-profit organization engaged in legal advocacy for social justice issues. Every year, SPLC releases a extremist groups’ dataset <sup>2</sup> that records the names, locations, and ideologies of extremist groups operating in the United States. We searched through SPLC’s 2018 and 2019 extremist groups datasets. For each listed extremist group, we manually searched for their official website. Note that each listed extremist group has physical headquarters across various states in the USA. By searching for their websites we identify the home for the extremist groups’ content in the online world. We identified 289 websites hosted by the listed extremist groups. For each website, we also note the website domain—hereafter referred to, as extremist domains. For example, for Virginia Dare, a white supremacy group, we record vdare.com as a website domain and for Alliance Defending Freedom—anti-LGBTQ advocacy organization—we record adflegal.org.

According to the SPLC’s policies, SPLC prioritizes identifying all U.S based hate groups regardless of the group’s “left” or “right” political leaning <sup>3</sup>. For example, the 2019 SPLC dataset contains 27 groups with Black Separatist ideology which is not on the far-right of the U.S political spectrum. However, we could identify websites for only 10 out of 27 Black Separatist groups. While we did not set out to study online extremist groups only from the far-right political spectrum, our collection of extremist websites largely belong to far-right groups representing anti-Immigration, anti-Muslim, White Supremacy and anti-LGBTQ ideologies. This skew towards far-right extremism is representative of the picture of domestic extremism in the United States. According to the Center for Strategic and International Studies (CSIS), <sup>4</sup> far-right extremism has massively outpaced far-left and other types of extremism in the United States [30]. Another independent report on Global Terrorism Index produced by the Institute for Economics and Peace reports that the far-right attacks in North America, Western Europe, and Oceania have increased by 250% since 2014, making it more lethal than the far-left extremism [18]. While we believe that our dataset is reflective of the ecosystem of domestic extremism in the United States, we discuss this skew further with respect to extremism in other countries in the Limitations and Future Directions section.

<sup>1</sup><https://www.splcenter.org/>

<sup>2</sup>see “DOWNLOAD DATA” in <https://www.splcenter.org/hate-map>

<sup>3</sup><https://www.splcenter.org/20200318/frequently-asked-questions-about-hate-groups>

<sup>4</sup>CSIS conducts policy studies and strategic analyses of political, economic and security issues throughout the world. CSIS is labeled as least biased and highly factual source of information by mediabiasfactcheck.com

### 3.3 Identifying Extremist Accounts

To identify extremist accounts—Facebook groups/pages that share links from extremist websites—we use the CrowdTangle Link Search API<sup>5</sup>. The CrowdTangle Link Search API retrieves posts by public Facebook groups and pages containing a certain link or the link domain. With the 289 identified extremist domains, we query the Link Search API separately for every domain. For extremist websites containing generic domains such as sites.google.com, we query the full link with the sub-domain, for example, sites.google.com/site/newblackliberationinstitute. For every queried domain, the API returns up to 1000 posts containing that link domain. Hence, to increase the completeness of our data, we queried every extremist domain separately for every calendar month starting from January 2018 to December 2019. For every queried domain in any calendar month, the number of returned posts was always less than 1000. This indicates that we have retrieved all public posts on Facebook available to CrowdTangle between 2018 and 2019 that share links from the identified extremist domains. Every post retrieved from CrowdTangle contains the account (page or group) name, post text, embedded links, timestamp, reactions (e.g., Like, Love, HaHa etc.), number of comments, number of public shares and 12 other fields. We aggregate all returned posts (450K posts) and find that our data contains 71,430 unique Facebook pages/groups accounts. In other words, 71,430 accounts shared at least one link from the extremist domain. User activities on social media often follow skewed, long-tailed distributions where most users contribute less frequently while fewer users are more active. We observe similar distribution with extremist links posted per account. Previously, researchers have used activity thresholds to eliminate accounts or communities that are less active [25, 39]. For example, while studying Wikipedia edits, Kumar et. al. remove the users that make less than 5 edits [39]. We decide the activity threshold by analyzing the percentile values of the extremist links per account distribution. Based on the 95<sup>th</sup> percentile cut-off, we remove all accounts that share less than 10 unique links from extremist domains. We provide additional details of the downstream analysis with various link thresholds in the Appendix. The entire data collection process happened across three weeks in May 2020. This means that the extremist accounts that were active in 2018 and 2019 but got banned before May 2020 are not included in the dataset.<sup>6</sup> Finally, we have 4,876 Facebook pages/groups remaining in our dataset. Table 1 displays the descriptive statistics for the accounts in our dataset.

### 3.4 Qualitative Validation of the Extremist Accounts

While we know that the extremist accounts posted 10 or more unique links from the extremist websites, do they promote extremist worldviews in general? In this subsection, we present our qualitative validation of the ideologies and the views promoted by the extremist accounts. To validate, we invited two experts from the Southern Poverty Law Center specializing in white supremacy and anti-LGBTQ hate groups. We requested the experts to qualitatively analyze a random sample of extremist accounts. Specifically, we randomly sampled 20 extremist accounts that share links from the white supremacy and 20 accounts that share links from anti-LGBTQ extremist websites. Next, we asked the experts to review Facebook timelines of the extremist accounts and describe their ideologies based on the content hosted and the page/group name and description. 16 of the 20 accounts posting links from white supremacy extremist websites, generally promote either far-right conspiratorial views and racists or misinformative content. Similarly, 18 of the 20 accounts posting links from anti-LGBTQ extremist websites usually peddle anti-choice, anti same sex marriage or anti-trans views. Only 2 of the 40 groups focused exclusively on memes without

<sup>5</sup><https://github.com/CrowdTangle/API/wiki>

<sup>6</sup>As of January 2021, 207 extremist accounts from our dataset (156 Facebook groups and 51 Facebook pages) have been removed from Facebook.

actively promoting any aspect of the white nationalist or anti-LGBTQ rhetoric. The complete expert analysis is available at the link in the footnote <sup>7</sup>.

While the experts validated the extremist views and the content hosted by extremist accounts, the 4,876 extremist accounts in our dataset consist of Facebook pages and groups that engage multiple Facebook users. We treat every page or a group as one extremist account that hosts content posted by its page owners or the group members. Who contributes to the content on these pages or groups? CrowdTangle, or any other Facebook API does not disclose personally identifiable information even in the public posts <sup>8 9</sup>. In other words, when requesting the content's of a post, the response will not include the name of the member who created the post. This poses a challenge in understanding the agency of posts shared on the extremist accounts. While this is a limitation of the Facebook dataset, we approximate the engagement with extremist accounts by reporting the distribution of page likes and group members in Table 1.

### 3.5 Extracting Information Shared by Extremist Accounts

We identified 4,876 extremist accounts. In our third research question, we assess how influential various accounts are in spreading links from biased, fake news and conspiracy domains in addition to the extremist domains. Towards this goal, we need all *link posts*—posts with embedded links—from the extremist accounts and not just the ones originating from extremist domains. Hence, next we extract all link posts made by the extremist accounts between 2018 and 2019 and then group the link posts by the link domain type. First, we use the CrowdTangle Post Search API to acquire all of the Facebook link posts made by the 4,876 extremist accounts. In total, we obtain 223K link posts made by extremist accounts across two years, 2018 and 2019. Next, we extract the links shared in the posts along with the timestamps and the link domains. In total there are 74,314 unique links in our dataset spanning over 1,236 link domains.

### 3.6 Data Slicing for the Downstream Analysis:

Our role identification process discussed in next sections is based on the Facebook activity of extremist accounts. Recall that we have two years (2018 and 2019) of activity for each account. Here, we determine the time unit of analysis in which we identify the roles, observe their dynamics and assess their influence in information sharing. To study the role stability over time, we need to divide the two year timespan into smaller windows and analyze how extremist accounts transition into different roles over successive time windows. We slice the data into four windows of six months— $T_1$ ,  $T_2$ ,  $T_3$ ,  $T_4$ —from January 2018 to December 2019 (Figure 1 (b)). In RQ1, we identify the roles played by extremist accounts in  $T_1$  and in RQ2, we track the role stability over  $T_2$ ,  $T_3$  and  $T_4$ . In RQ3, we model the information flow through the roles in  $T_1$ . Does the choice of six-month time window affect our observations? We experimented with different length of time windows (specifically, 2 months, 3 months and 12 months) obtaining similar results for the research questions.

### 3.7 Identifying Domain Types for Extracted Links:

In the RQ3 analysis, we model the information flow and examine the influence of various roles in spreading content from mis and disinformation based sources. Specifically, we consider 57K link posts made in  $T_1$  and identify domain types for the links. We are interested in identifying domains hosting extremist content, biased news, fake news and conspiracy content. We refer to the dataset published by OpenSources <sup>10</sup>. OpenSources maintains a professionally curated list of online sources

<sup>7</sup>[https://www.dropbox.com/s/mf3jt9xbk9z9xfw/SPLC\\_annotations.pdf?dl=0](https://www.dropbox.com/s/mf3jt9xbk9z9xfw/SPLC_annotations.pdf?dl=0)

<sup>8</sup><https://help.crowdtangle.com/en/articles/1140930-what-data-is-crowdtangle-tracking>

<sup>9</sup><https://developers.facebook.com/docs/groups-api/>

<sup>10</sup><https://github.com/BigMcLargeHuge/opensources>

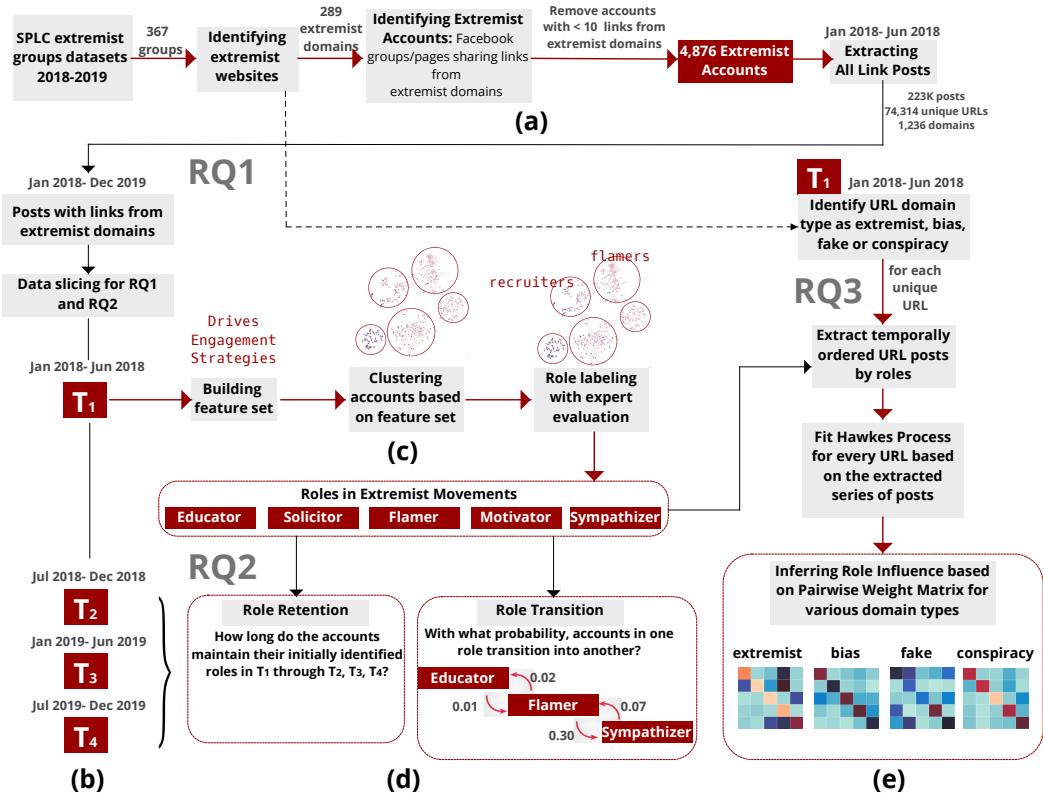


Fig. 1. Figure showing data preparation and method details for three research questions. (a) We first identify extremist websites (domains) and identify extremist accounts—Facebook pages and groups that links from extremist domains. Next, we extract posts by the extremist accounts that contain links from the extremist websites and (b) slice the dataset into four time windows of 6 months each. (c) For role identification in RQ1 we cluster extremist accounts based on the features derived from the account activity in T<sub>1</sub>. (d) We use the cluster centers from T<sub>1</sub> to re-cluster the extremist accounts and measure role retention and transition in RQ2 (also see Figure 2 for more details). (e) For RQ3, we use the link posts by the extremist accounts made in T<sub>1</sub>, group the links based on their source type (extremist, biased, fake news and conspiracy) and measure the influence of various roles in link sharing based on Hawkes process.

towards the goal of empowering people with reliable information. With the help of professionals, they annotate each source based on overall inaccuracy, extreme biases, lack of transparency, and writing styles. We use the annotations from OpenSources, and our list of originally identified extremist domains (Section 3.2), to group links by their source type as extremist, fake news source, biased source and conspiracy/pseudoscience source. For our RQ3 analysis, we remove all links that do not belong to either of these four domain types. Removing the links with domains not included in the OpenSources or the original set of extremist domains, reduces our T<sub>1</sub> link post dataset by 24%. In the Appendix, we provide the characterization of the frequent domains in the removed dataset.

#### 4 RQ1 METHOD: IDENTIFYING ROLES IN ONLINE EXTREMIST MOVEMENTS

In this research question, we identify roles played by extremist accounts in the extremist movements based on theory guided characteristics of social movement participation. Specifically, we

Characteristics of Participation	Theoretical Models	References	Behavior	Operationalization
Drives for participation	Expectancy-value models	[36, 44, 54]	Risk	Proportion of LIWC Risk words (e.g., caution, crisis, failure)
		[36, 44, 54]	Reward	Proportion of LIWC Reward words (e.g., benefit, bonus, award)
	Social-psychology models	[83]	Injustice	Proportion of MFD Fairness words (e.g., parity, fair, justice)
		[21, 82]	Achievement	Proportion of LIWC Achievement words (e.g., accomplish, ability, attain)
		[72]	Group Identity	Proportion of LIWC we words (e.g., we, ours, us)
		[84]	Anger	Proportion of LIWC anger words (e.g., resent, argue, angry)
Engagement in the movement	Degrees of participation	[46]	Proportion of links from extremist domains	Ratio of links from extremist domains to total link posts
	Degrees of participation (popularity)	[46]	Likes	Proportion of likes on extremist links to likes on the rest of the link posts
			Shares	Proportion of shares on extremist links to likes on the shares of the link posts
			Comments	Proportion of comments on extremist links to comments on the rest of the link posts
	Trends in participation	[10]	Trend	Trend line fitted on the number of extremist links posts per month
Strategies of mobilization	Opinions	[81]	Expressions of opinions	Proportion of extremist link posts containing opinion patterns (see Table 6)
	Solicitation	[6, 46]	Expressions of solicitation	Proportion of extremist link posts containing solicitation patterns (see Table 7)

Table 2. Table summarizing features used to identify roles in online extremist movements on Facebook. We build the feature set based on underlying characteristics of participation and the theoretical models describing them.

operationalize *drives for participation* (Section 2.3.1), *engagement in the movement* (Section 2.3.2) and *strategies of mobilization* (Section 2.3.3) based on the activity of extremist accounts in the six month time period  $T_1$  (Jan 2018-Jun 2018) and build a feature set (Table 2) to identify roles. Next, we represent every extremist account in our dataset with the derived features. To identify roles, we cluster the accounts based on the derived features. Finally, we qualitatively analyze and label each cluster as a role in extremist social movement with the help of experts in social psychology.

## 4.1 Operationalizing Characteristics of Social Movement Participation

4.1.1 **Drives for Social Movement Participation (6 features).** Theoretical models exploring the drives for social movement participation consider two distinct perspectives: *expectancy-value models*, whereby participation is driven by perceived risk-reward assessments and *social psychology models*, where the psychological features are the core drivers of participation. Below we detail the computational operationalizations of *drives* informed by these theoretical models explained in Section 2.3.1.

- **Expectancy-value features:** Participation in social movement could be driven by perceived risk-reward assessment related to engagement in the movement [36, 44, 54]. Analyzing the language used by the extremist accounts while sharing links from extremist websites, especially the words related to cost-benefit, could indicate whether the accounts considered the costs and benefits of participating in extremist movements. To operationalize these features, we use risk and reward lexicons from the Linguistic Inquiry and Word Count (LIWC) 2015 [79]. LIWC is designed to record words that reflect various psychological states and perceptions [79]. Specifically, we calculate the proportion of risk related words (e.g., caution, crisis, failure) and reward related words (e.g., benefit, bonus, award) used by an extremist account while sharing links from extremist websites.
- **Social Psychology features:** Guided by the social-psychology based theoretical models, here we want to measure the feelings of injustice [83], sense of achievement [21, 82], group

identity [72] and anger [84] that could serve as potential drives for participation. To identify the language related to injustice, we use Moral Foundations Dictionary that contains a systematically derived list of words pertaining to moral foundations in political ideologies [78]. Specifically, we use the “fairness” lexicon which accommodates virtue words such as rights and equality, and vice related words such as bigot, favoritism, and prejudice [20]. Next, to measure the sense of achievement, we use LIWC’s [79] achievement category which contains words such as “accomplish”, “ability”, “attain” etc. Further, language related to group identity can be reflected by the use of third person pronouns such as “we”, “us”, “ours” [28, 56, 79]. Hence we measure the third-person pronoun usage by LIWC “we” category. Finally, to measure anger related words, we use LIWC anger category containing words such as “resent”, “argue”, “angry.” For each of these lexicons, we calculate the proportion of words in each lexicon while sharing links from extremist websites.

**4.1.2 Engagement Trends in Social Movements: (5 features).** As per the details described in Section 2.3.2, here we provide our methods to characterize engagement trends in the social movements. Participants can engage with social movements in various degrees of interests and continuity [10, 46]. Hence, we calculate proportion, popularity and trends in sharing links from extremist websites.

- **Proportion of Links from Extremist Domains:** Various degrees of participation in distributing resources can reflect the involvement in social movement [46]. Hence, for every extremist account, we first calculate *proportion of links from extremist domains*—proportion of links shared from extremist domains to all links shared by that account in a given time-frame.
- **Popularity of Links from Extremist Domains:** The amount of positive reactions and interactions on the shared links could reflect how popular the posts containing extremist links are on a Facebook page/group. Hence, we calculate the average likes, shares and comments received on posts with links from extremist websites and divide it by the average likes, shares, and comments (respectively) on all links posted in a given time-frame. High values of these features indicate that the extremist content is more popular compared to the rest of the content published on that page/group.
- **Trends in Disseminating Links from Extremist Domains:** We also account for the engagement trends. For each month within the six-month period, we calculate the number of links from extremist websites posted on an extremist account and fit a line via least-square regression. Least-square regression finds optimal fit for the line by minimizing the sum of squared residuals. We calculate the trend of this fitted line to measure engagement. Specifically, positive values of trend can indicate increasing engagement and negative values can indicate disengagement in posting links from extremist websites.

**4.1.3 Strategies of Information Mobilization: (2 features):** As described in detail in Section 2.3.3, core members of the social movements may strategically solicit participation through calls for donations, volunteers and invitations for social gatherings [6, 46]. Similarly, members can also strategically create opportunities for collective action by expressing opinions, thoughts and beliefs around political events [81]. Hence, we build two features that capture the expressions of personal opinions and the language of solicitation used by extremist accounts while sharing links from extremist domains. For both features, we calculate the proportion of extremist link posts containing the expressions of opinions and solicitations respectively.

- **Expressions of Opinions in Posts with Links from Extremist Websites:** By “opinions” we refer to the expression of thoughts, beliefs and personal opinions [68]. To calculate the proportion of opinions present in posts, we extract phrases that signal expressions of

personal opinions or private states [89]. Previously, researchers studying emotional and informational support [5, 87] relied on emotional and informational support related nouns, verbs and adjectives to extract phrases related to their task. For example, Wang et. al. [87] used  $<you+MODALVERB>$  pattern to extract phrases containing suggestions (“you should” or “you must”). We use similar methodology to extract phrases related to personal opinions. How can we identify verbs, nouns and adjectives related to personal opinions? Chen et. al. argue that individuals form their opinions via cognition and internal perceptual cues [9]. Hence, we first look at LIWC 2015 cognitive processing lexicon and its subcategories that record words related to thinking, perception and expression. To construct phrase patterns, we first split all words from LIWC cognitive processing categories into their part of speech labels (verbs, nouns, adjectives). Consider the verb *prefer* and noun *preference* from LIWC cognitive processing category. Both words, when paired with different pronouns can form expressions of opinions. For example, “*I (first person subjective) prefer*” and “*My (first person possessive) preference*” both indicate personal opinions. Moreover, variations such as negations (“*I do not prefer*”) or adjectives (“*My strong preference*”) also signal opinions. Hence, we build our initial set of opinion patterns from LIWC’s cognitive processing verbs, nouns and adjectives by pairing them with appropriate pronouns and variations. We iteratively improve upon this list of phrase patterns by first, extracting sentences containing those patterns and then, manually eliminating verbs, nouns and adjectives that do not signal opinions. Table 6 in the Appendix lists the final patterns and examples. Note that our opinion extraction method is based on LIWC cognitive processing lexicon that contains limited number of words. Hence, it is possible that our opinion extraction misses out of some expressions of opinions.

- **Expressions of Solicitation in Posts with Links from Extremist Websites:** In solicitation, we want to identify expressions that demand some action on the reader’s part. Here, we are looking for calls for donations, invitations for events and protests and participation in policy advocacy (e.g., sign the petition, call your representative etc). To extract solicitation patterns, we follow similar procedure as opinion extraction but instead look at verbs, nouns and adjectives from LIWC’s social and affiliation categories. LIWC social and affiliation categories contain words such as *sign*, *call*, *contact*. We build phrase patterns and iteratively evaluate them using methods similar to opinion extraction and report final solicitation phrase patterns in Table 7 in the Appendix.

## 4.2 Clustering Extremist Accounts Based on the Derived Features

We use the features described above, to cluster the extremist accounts and label each cluster as a role in the extremist movement. We use the theory based features (described in the previous subsection) representing drives, engagement and strategies to discriminate between different roles in meaningful way. For role identification, we first need to decide on the number of roles and then use a technique that integrates structural data (feature vectors for extremist accounts) with interpretive analysis that will allow us to describe roles in a relevant way. We use K-Means clustering—an unsupervised clustering algorithm commonly used by other CSCW scholars in role identification studies [3]. For example, Arazy et. al. used K-Means to identify emergent roles in Wikipedia contributors [3]. We use the popular K-Means method with kmeans++ initialization to cluster the extremist accounts based on their activity in the six month time window  $T_1$  (Figure 1 (c)). Specifically, we represent every extremist account with a feature vector of length 13 (6 drives + 5 engagement + 2 strategies). Next, we perform a series of robustness checks to first, determine the number of clusters in order to obtain the optimal separate between different roles and then, to check the stability of clusters. All features were standardized for the downstream analysis. In K-Means algorithm, the number of clusters,  $K$  is a free parameter. We first find the best value of

Role	Frequency	Example texts used by the accounts while sharing links from extremist websites
Solicitors	5.2%	"...Sign here to demand her [Rep. Maxine Waters] immediate resignation" "...Join us in signing thank you card for President Trump" "...Stand with us to take back our streets"
Educators	10.6%	"...Escaping from motherhood: how it destroys society" "...We believe that we have the duty to instruct people in the truth of Tradition. Even if it destroys their party" "...The need for an ethnocentric society amidst "globalism" "
Flamers	18.4%	"...genuine Christians know that homosexuality is an abomination before GOD! " "...MURDERED in cold blood. Emergency: Gunfire, bodies and BLM murderers" "...house democrats vote to allow female genital mutilation..." (false information flag by Facebook pops up)
Motivators	29.4%	"...Senator Dan Halls stands with us in a passionate commitment to strengthening religious freedom" "...FREE SPEECH WINS!! Supreme court rules pregnancy centers can't be forced to advertise abortion" "...we WILL have the COURAGE to defend ourselves! borders, language and culture MATTER"
Sympathizers	36.4%	"...White South Africans petition Trump to allow them to migrate to the US" "...A jihadi cult member running for Congress as Democrat from Alaska" "...They will only see Italy on postcard: Italy turns away another migrant ship"

Table 3. Roles and the corresponding percent of extremist accounts in the dataset. The examples are of the texts written by the extremist accounts while sharing links from the extremist websites

$K$  with an elbow analysis that offers a natural trade-off between the best separation between the clusters and the number of clusters. Specifically, we train the K-Means algorithm for number of clusters ranging from 2 to 20 and plot distortions—sum of squared distances from each point to its assigned center. We observe the elbow at  $K = 5$  (see Figure 7 (a) in the Appendix). Thus we assume 5 as the optimal number of roles. We repeated the elbow experiment with other scoring parameters such as silhouette distance—a measure of how similar a data point is to its own cluster compared to other clusters—with similar results. We also check for the stability of final cluster assignments with various random seeds. Moreover, to check the robustness of our method, we perform the clustering with alternate clustering methods observing similar elbow and cluster assignments. Section A.2 contains more details about the robustness checks used for the cluster analysis. We assume that every cluster generated, represents a role in the online extremist movement. Next, we use expert guided interpretive analysis to label the roles and their descriptions.

**4.2.1 Role Labeling with Expert Evaluation.** Do our quantitatively identified clusters represent coherent roles in extremist movement participation? To evaluate, we invited a group of seven social psychology and social movement experts to first analyze and then label the clusters based on their representative characteristics. This group consisted of one senior professor, one assistant professor, one post-graduate researcher and four senior doctoral students. To reduce bias in role labeling, we selected the external evaluators who are not a part of the author group and were not involved in any of the work preceding or following this stage. We showed them mean feature values for every cluster. Additionally, we selected top 5 representative extremist accounts from each cluster—accounts with closest distance to the cluster centers. We compiled the list of top 10 most representative posts from each selected account ranked by post likes [26]. Based on this information, we asked the evaluators to come up with labels and descriptions for each cluster. In all, every evaluator looked at 250 Facebook posts (5 clusters  $\times$  5 accounts  $\times$  10 posts) and recorded the possible labels and descriptions. Finally, the first author and the evaluators together, worked and selected the best label for every cluster. We present the identified roles and comments on the expert evaluation process in the next section.

## 5 RQ1 RESULTS: ROLES IN ONLINE EXTREMIST MOVEMENTS

Here we describe the roles played by extremist accounts and their typical behaviors. Table 3 displays the frequency of the roles in the dataset alongwith the example text written by the

extremist accounts while sharing links from extremist websites. We identified five roles in the online extremist movements.

- (1) **Solicitors:** These are the accounts that solicit participation from their readers for signing petitions, attending rallies etc. On average, around 20% of their links come from extremist domains and they post extremist content with fairly consistent trend throughout the six month period (trend feature values close to 0). These accounts use high group identity language such as “we”, “our”, “us” compared to other roles. Evaluators also described them as “recruiters.” One evaluator mentioned:

*“These groups appear to be soliciting action for their hate. To some extent, they seem pretty keen on doing something about the groups they hate and are actively sharing/liking posts to promote action”*

- (2) **Educators:** Educators have distinctively high amount of extremist content in their link sharing. On an average, 50% of their links come from extremist domains. Additionally, the extremist links posts get more likes and comments compared to other material on these pages/groups. They post the extremist content with consistently high rates (trend feature > 0) throughout the six months. In qualitative evaluation, the experts pointed out that these groups share intellectual material and appear serious and sincere in propagating the fundamentals of extremist ideologies. The evaluators also suggested alternate labels such as “preachers” and “intellectuals.” According to one evaluator:

*“...they seem to take effort to make logical arguments. They are not necessarily showing anger towards other groups but are instead more focused on highlighting their own group’s worth logically/analytically”*

- (3) **Flamers:** These accounts spew toxic and inflammatory content. Around 5% of their links belong to extremist domains and the messages on the links and the link text itself often contains language suggesting anger and injustice. In other words, these pages/groups have the highest proportion of anger and injustice related words while disseminating extremist content. The extremist links posted on these accounts get higher number of shares compared to the rest of the content. The experts also described them as “fear mongerers” for attempting to cause general outrage. Immediately after looking through the posts, one evaluator commented:

*“these are clearly very strong, divisive and toxic posts”*

- (4) **Motivators:** Around 7% of the links by motivators are sourced from extremist domains. Evaluators pointed out that motivators use exceptionally positive language. While posting extremist content, they stress on the achievements and rewards associated with extremist activities. Motivators also express opinions with highest proportions compared to the other roles. Experts noted that these accounts engage in policy activism focusing on policies protecting and defending cultural and moral values. Evaluators also mentioned:

*“it almost looks like they are celebrating the in-group [people and organization involved in the extremist movement] and the sensationalized news about the in-group”*

- (5) **Sympathizers:** These accounts post extremist content links with lowest rates (2% of their Facebook link posts) and sporadically throughout the six month period. They also show low engagement in terms of likes, shares and comments on the extremist link posts. According to the experts, these groups are on the fringe of extremist ideology and might be only slightly interested in extremist causes. Experts also referred to them as “observers.” One evaluator described *sympathizers* as: *“They look more like general conservative interest groups”*

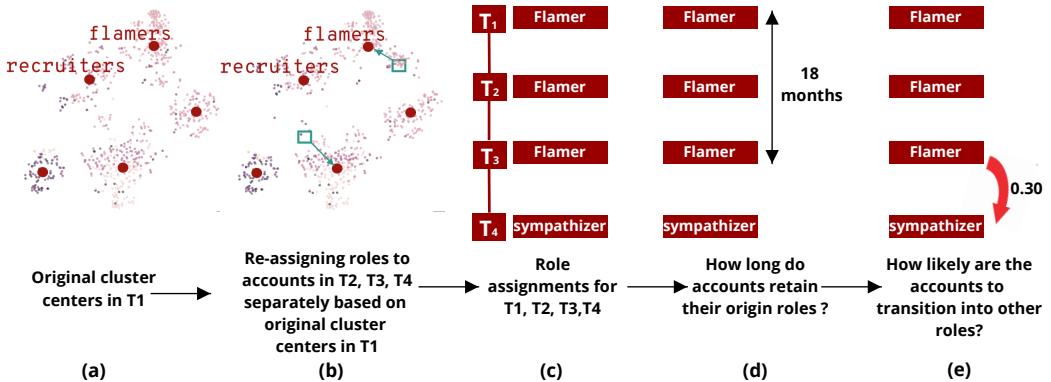


Fig. 2. Figure showing method steps for analyzing role dynamics. (a) We identified cluster centers for each role in RQ1 using the account activity in  $T_1$  (Jan 2018-Jun 2018). (b) We use those cluster centers identified in  $T_1$  and re-assign extremist accounts to roles based on their activity in  $T_2$ ,  $T_3$  and  $T_4$ . (c) For every account in the dataset, we get cluster assignments for all time periods. (d) To measure role retention, we calculate the number of time periods for which the extremist accounts maintain their original roles from  $T_1$ . (e) In role transition, we calculate the probability with which the extremist accounts may transition to different roles.

**Notes on Expert Evaluation:** Manually evaluating and labeling roles is a challenging but an insightful task. Interpreting clusters and finalizing the role labels was an iterative, discursive process. Some roles were easy to identify and label (for example, *solicitors*, *educators* and *flamers*) while others required going over additional details such as page/group names and their descriptions. For example, while the experts reached at immediate consensus for the *flamers* role, they deliberated over the *motivators* category. However, qualitative evaluation provided additional insights about the clusters that were not apparent from the original feature set. For example, we found that accounts in the *flamers* category had higher amount of posts flagged as misinformation or violent/graphic content. Similarly, *educators* shared links containing material describing extremist philosophies (for example, white identity, ethnocentrism, political philosophy) than rest of the accounts. Note that such nuances are harder to capture computationally and further extensive qualitative study of these pages might provide new theoretical insights about actors in online extremist movements.

## 6 RQ2 METHOD: MEASURING ROLE DYNAMICS

The roles we identify are data-driven. The underlying characteristics of the roles are derived mainly from the theoretical studies on physical protest events. Unlike physical social movements where members can commit to protests, meetings or other events, social media provides a dynamic, evolving space for members to engage or disengage from the social movement as they like without much accountability [10]. Hence, we also analyze the dynamics of roles based on how long the accounts retain their initial roles (role retention) and their transition probability to another role (role transition). Figure 2 displays the method steps taken to measure role retention and role transition.

**RQ2a: Measuring Role Retention:** By retention, we measure how long the extremist accounts adhere to their originally identified roles. Recall that our data spanning two years (Jan'18 to Dec'19) is sliced into four windows, each six months duration— $T_1$ ,  $T_2$ ,  $T_3$ ,  $T_4$  (Figure 1 (b)). We initially identified roles in RQ1 using the account activity in the first time period,  $T_1$  (Jan 2018 - Jun 2018). To measure role retention, we use the cluster centers (mean values for each feature for each cluster) from  $T_1$  and re-assign roles to the extremist accounts using their activity in  $T_2$ ,  $T_3$ ,  $T_4$ . For every extremist account, we now have the role assignment for  $T_1$ ,  $T_2$ ,  $T_3$ , and  $T_4$ . For example in Figure 2

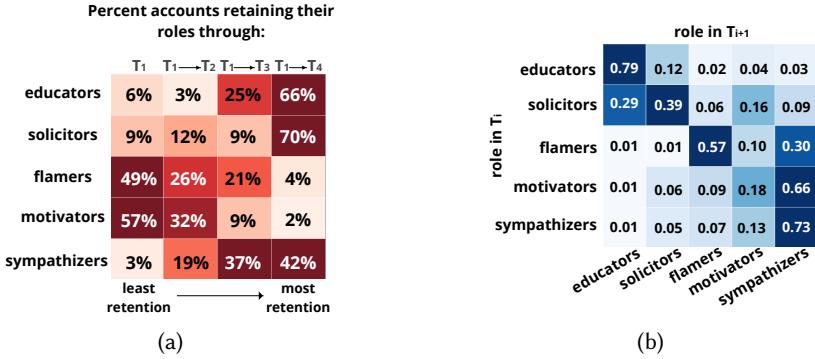


Fig. 3. Figure presenting the results of RQ2. (a) indicates role retention—proportion of extremist accounts that retain their originally identified role through subsequent time windows in the dataset. For example, 70% of the *solicitors* retain their role throughout  $T_1 \rightarrow T_4$ . Whereas 49% of *flamers* retain their roles for only the initial  $T_1$  period. (b) displays the role transition probability matrix. Rows indicate the role in  $T_i$  and columns indicate the role in  $T_{i+1}$ . The cells indicate the probability of transition from role in  $T_i$  to  $T_{i+1}$ . For example, *solicitors* may transition to *educators* in the next time period with 0.29 probability.

(c), an account stays in the *flamer* role for three consecutive time periods. That is, the role identified in  $T_1$  is retained for (3 X 6) 18 months consecutive months. Note that all time windows  $T_1$ ,  $T_2$ ,  $T_3$ , and  $T_4$  represent distinct calendar months between 2018 and 2019 as displayed in Figure 1 (b). Using the initially defined cluster centers in  $T_1$  allows us to compare the accounts' activities in  $T_2$ ,  $T_3$ ,  $T_4$  with respect to their own past states. How long do extremist accounts adhere to their initially identified role in  $T_1$  over the subsequent future time windows? To answer, we calculate the number of continuous time windows across which an account retains its initially identified role. Finally, for each role, we calculate the proportion of accounts maintain their roles across just one ( $T_1$ ), two ( $T_1 \rightarrow T_2$ ), three ( $T_1 \rightarrow T_3$ ) or all four ( $T_1 \rightarrow T_4$ ) time windows.

**RQ2b: Measuring Role Transition:** How likely are the extremist accounts to move from one role to another? For example, how likely are *sympathizers*—fringe supporters of the extremist movement—to transition to *educators*—accounts that actively distribute large proportion of extremist content? From the role retention analysis, we know which role every extremist account plays in each of the time windows— $T_1$ ,  $T_2$ ,  $T_3$ ,  $T_4$ . For every account, we consider the role played by that account in a particular time window  $T_i$  as the *state*  $S_i$  that account is in. Consequently, for every account we have sequence of four states corresponding to each of the time windows. The example account in Figure 2 (e) has states: *flamer* → *flamer* → *flamer* → *sympathizer*. Using such state sequences of all extremist accounts, we then calculate the state transition, or the role transition probability. Specifically, we calculate the pairwise transition probabilities for each pairs of roles. High probability of transition *solicitor* → *educator* will indicate that an account currently playing the role of *solicitor* is likely to transition into *educator* in the next time window with high probability.

## 7 RQ2 RESULTS: ROLE DYNAMICS

**RQ2a: Retention:** Figure 3 (a) displays the initial roles (rows) and number of time periods ( $T_1$ ,  $T_2$ ,  $T_3$ ,  $T_4$ ) for which the role was retained by an account. We find that 66% of the *educators* and 70% of the *solicitors* retain their initial role for all four time periods, that is for the entire two years. Whereas, 49% of the *flamers* and 57% of *motivators* transition to another role just after  $T_1$  (6 months). *Educators* and *solicitors* can be viewed as an elite group in extremist movements—members who

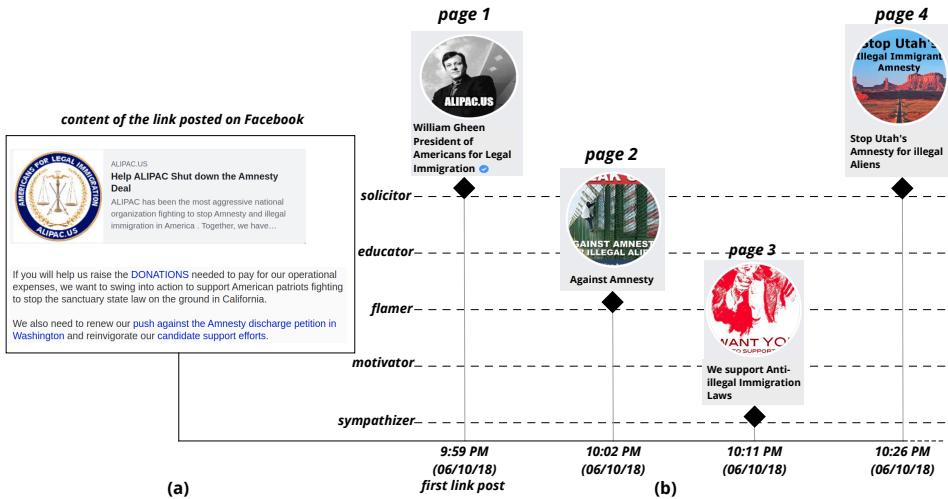


Fig. 4. An example illustrating a series of link posting events by accounts playing various roles (a) Contents of the link hosted on the website of an SPLC designated anti-Immigration group ALIPAC. This link contains call for action against stopping immigrant amnesty deal in California (b) series of link posting events on Facebook demonstrating how a link from an extremist website flows through various accounts playing roles in extremist movements. The link was first posted on a Facebook page of the president of the ALIPAC (page 1). Next, the link was posted by various accounts—Against Amnesty (page 2), We support Anti-illegal Immigration Laws (page 3) and Stop Utah's Amnesty for illegal Aliens (page 4)—playing various roles. This example demonstrates a series of 4 events—link posts by *solicitor*, *flamer*, *sympathizer* and *solicitor* accounts.

distribute (information) resources and actively recruit others [46]. On the other hand, *flamers* and *motivators* are supporters who exhibit low engagement with the links from extremist websites. Our results suggest that roles more core to the extremist movement such as *educators* and *solicitors*) are more stable. In other words, *educators* and *solicitors*) are more likely to maintain their roles and consequently their behaviour surrounding the participation in extremist movements for longer periods compared to others.

**RQ2b: Transition:** Figure 3 (b) displays a transition matrix for roles. The values in the cell indicate the probability by which an account in one role (row) would transition to another (column) in the next six months. Based on the main diagonal, accounts in most role are more likely to retain the same role in the next time window. For example, *educators* will stay *educators* in the next six months with 0.79 probability. Similarly the probability of *flamer* → *flamer* is 0.57. Notably, *solicitors* and *educators*—roles with highest engagement with links from extremist websites—have highest transition probabilities with each other compared to any other roles (*solicitor* → *educator* = 0.29 and *educator* → *solicitor* = 0.12). Moreover, *flamers* and *motivators* can transition to *sympathizers* with 0.30 and 0.66 probabilities respectively, indicating that roles with less engagement with extremist content are also less stable.

## 8 RQ3 METHOD: INFLUENTIAL ROLES IN INFORMATION SHARING

In RQ1 we identified five roles in extremist movements—*educators*, *solicitors*, *flamers*, *motivators* and *sympathizers*. In RQ2, we measured the dynamics of the roles over time. How influential are these roles in disseminating information? Previous research suggests that terrorist organizations such as ISIS are able to mobilize information through series of online accounts that knowingly or

unknowingly play a role in advancing terrorist propaganda [38]. Researchers observed a similar pattern of participatory information sharing in political disinformation campaigns [76]. In this research question, we model the information sharing by various roles to investigate how influential different roles are in spreading extremist content, fake news, biased news and conspiracy sources. We open up by a motivating example describing how information crucial to the extremist causes flows through various roles online. Next we describe the method details for modeling the information flow using Hawkes process and inferring influence.

### 8.1 Call for Shutting Down the Amnesty Deal: A Motivating Example

On June 10th 2018, ALIPAC—an anti-Immigration political action organization [69]—put out a call for action to stop amnesty deal for immigrants in California (Figure 4 (a)). This included call for donations, and participation to help support ALIPAC’s operational costs. A link containing this call for action was posted on Facebook on the same day at 9:59 PM by a verified page managed by the president of ALIPAC (page 1 in Figure 4 (b)). This Facebook verified page is also known for posting fake news from websites hosting plagiarized content [32]. According to our RQ1 analysis, this page (page 1) plays the role of *solicitor*. Three minutes after the initial post by page 1, another Facebook page—identified as *flamer*—posted the same link. Following the post by page 2, a *sympathizer* page (fringe supporter of the extremist content) as well as another *recruiter* page also posted the same link. Following these four link posts, another 32 pages posted the same link containing the call for action by an anti-Immigration group over the period of next three days. How influential was page 1 in spreading this link containing the call for action? How influential are *recruiters* or other roles in spreading the extremist content or other types of information? In the next section, we explain our information flow model that considers both temporal order of link posts and the time difference between consecutive link posts to statistically establish the influence of different roles in spreading links from various types of sources.

### 8.2 Intuitive Background for Measuring Influence in Link Sharing

Our primary goal here is to understand how influential roles are in spreading information from various types of sources. By **influence**, we measure the probability by which a link posting by one role affects link posting by other roles in future. To measure the influence of various roles, we use multivariate Hawkes process which is commonly used for modelling events on social media [23, 91]. For example, Zannettou et. al. used Hawkes process to understand the influence of various online platforms (Reddit, Twitter, 4chan) on each other in information sharing [91]. Hawkes processes are popularly used as a more sophisticated model for measuring influence compared to simple point processes [16, 65]. Specifically, Hawkes process can account for various nuances in information sharing. For example, while calculating influence, Hawkes process considers the temporal flow of link sharing, the time between consecutive link posts and the natural link posting tendencies of various roles involved. In its simplest form, modeling link sharing by different roles using Hawkes process can provide us with two parameters: (1) the natural link sharing tendency of the roles or **background rate** and (2) the pairwise influence that roles exert on each other in link sharing.

More formally, consider an **event** where a *solicitor* posts a link. This event can affect the probability by which other roles (for example, *sympathizers*) also post the same link in future. In other words, link posting by one role might increase the probability of the link posting by another role. The effect of previous events on the current event is additive and decaying. This means that, as more time passes between two events, the lesser impact the previous events will have on the current event. For each link, we extract such **series of events** (sequence of timestamps indicating the link posting by extremist accounts in various roles). Next, we provide the details for fitting Hawkes process on the extracted series of events.

### 8.3 Fitting Hawkes Process on Link Posting Events

A multivariate Hawkes model consists of  $K$  processes each with a background rate of  $\lambda_{0,k}$ . Fitting a Hawkes model on link sharing by five roles will give us the background link posting rates for each role—probability of an event caused by external factors—as well as the probabilistic estimate of the effect of the roles on each other. We use a discrete-time Hawkes model which is commonly used for modeling information flow on social media [91]. For each link, the time of posting is divided into discrete time bins of duration  $\Delta t$ , creating a series of events with time granularity of  $\Delta t$ . Fitting Hawkes model on this series of events will give us estimates for the background rate of  $\lambda_{0,k}$  and pairwise weight matrix  $W^{K \times K}$ . Appendix A.5 provides the formal mathematical representation of Hawkes process along with its parameters.

For our analysis, the **weight matrix**  $W^{K \times K}$  is most relevant. Every value  $w_{i,j}$  in the weight matrix  $W^{K \times K}$  signifies the expected number of subsequent events that will be caused in the process  $j$  after an event in the process  $i$ . In other words, value in weight matrix  $W_{i \rightarrow j}$  indicates the strength with which a link posting event by role  $i$  can affect link posting by role  $j$  [91]. We consider these weights as a proxy for **influence** of role  $i$  on role  $j$ . Note that weights from  $W$  describe the expectations of event occurrence above the background rate  $\lambda_{0,k}$  of the process where background rate  $\lambda_{0,k}$  accounts for events due to external factors.

### 8.4 Measuring Influence in Link Sharing

To understand the influence of roles, we consider the link posts made by the extremist accounts in  $T_1$  (Jan 2018-Jun 2018) and consider the role identified for every extremist account in RQ1. We assess the influence of roles by using the weight matrix  $W$  described in the previous subsection. To calculate the weight matrix, we first need to determine the number of processes ( $K$ ) and the length of discrete time windows  $\Delta t$  to bin every link posting event. Since we aim to examine how influential various roles are, we model the link postings with  $K=5$  processes—one for every role identified in RQ1. To select  $\Delta t$ , we plotted the distribution of inter arrival times—time in seconds between two consecutive link posting events—for all links. The goal for choosing  $\Delta t$  is to have a separate bin for most events. We decided  $\Delta t = 30\text{sec}$  based on the 10<sup>th</sup> percentile cut-off of the distribution. Choosing this percentile cut-off allows us to put approximately 91% of the link posting events in the bin by themselves. To eliminate links that are shared less, we select links that are shared by atleast 10 different extremist accounts spanning over atleast 3 of the 5 identified roles. Next, we fit a Hawkes model for each link using a nonparametric Expectation Maximization (EM) for parameter estimation [41]. With the EM algorithm we are able to get the background rate ( $\lambda_{0,k}$ ) as well as the weight matrix  $W$  described in the previous subsection. Both, background rates for each role and the weight matrix, contain non-negative real values. We further row normalize the weight matrix to represent the influence of a role in a row on the role in a column with values bound between 0 to 1. Note that so far we have calculated the weight matrix  $W$  for each link separately. In other words, for each unique link, we obtain pairwise influence values for roles. Remember that in Section 3.5, we recorded link domain types as extremist, biased, fake or conspiratorial based on the OpenSources annotations. Here, we calculate the overall weight matrix for each domain type by averaging the weight matrices of links based on their domain types. For example, to obtain the overall influence of roles in spreading links from fake news domains, we consider all links that are generated from fake news domains and calculate the average of their weight matrices. In the next section, we discuss our results and influential roles in spreading links from extremist, biased, fake news and conspiracy domains.

source type	#domains labeled	#labeled domains present in $T_1$	#unique links	#events (link posts)	% link posts made by				
	solicitors	educators			flamers	motivators	sympathizers		
extremist	289	231	758	16,532	12%	28%	9%	5%	44%
biased	133	94	1279	13,290	11	9%	19%	23%	36%
fake	304	107	380	3583	9%	20%	23%	13%	34%
conspiracy	154	68	936	10,001	20%	20%	10%	21%	29%

Table 4. Table describing link posting activities by various roles. We first report the number of labeled domains in each category—extremist, biased, fake and conspiracy and also specify how many of those domains are present in the link posts made in  $T_1$ . Next, we list the number of unique links and the number of link posting events for each source type. We also report what percent of such link posting events were made by various roles. For example, *solicitors* contribute to 12% of the link posts from the extremist domains.

## 9 RQ3 RESULTS: INFLUENTIAL ROLES IN INFORMATION SHARING

We summarize the number of links and number of link posting events generated for each type of information in Table 4. Looking at the number of events, *sympathizers* make up largest percent of link posting events in all source types. This is not surprising given that 36.4% of the accounts are *sympathizers*. Along the same lines, *solicitors*—who actively solicit participation by posting extremist links and *educators*—who share largest proportion of extremist links—contribute to high percent of events in posting extremist links. Interestingly, *flamers*—accounts that often post inflammatory and violent content—are also second highest in posting links from the fake news sources. Moreover, *motivators*—who focus on efficacy of policy changes and use opinionated language, make up 23% of the biased news posting events. Interestingly, *solicitors*, *educators*, *motivators* and *sympathizers*, all post links from conspiracy sources with similar rates.

Remember that we obtained weight matrix  $W$  by fitting a Hawkes model for links from every source type. We report the weight matrices for four source types in Figure 5. Value in weight matrices, for example, ( $W_{\text{solicitor} \rightarrow \text{educator}} = 0.11$  in Figure 5 (a)) indicates the strength with which *solicitors* may trigger information sharing by *educators* account. For all source types, the self weights for roles (for example,  $W_{\text{solicitor} \rightarrow \text{solicitor}} = 0.35$  in Figure 5 (a)) are highest. In other words, the main diagonal in all weight matrices in Figure 5 have the highest values. This means that extremist accounts in one role are more likely to trigger information sharing by other extremist accounts in the same role. This may be because the the accounts in the same role also share large number of same links. Note that the values for weights reported in Figure 5 are typical of influence values observed in information sharing on social media [91].

**Extremist Domains:** In all, we had listed 289 extremist website domains. Here, we analyze how important every role is in posting links from extremist domains in Facebook. Figure 5(a) displays pairwise influence of roles on each other. For example, the value of  $W_{\text{solicitors} \rightarrow \text{sympathizers}} = 0.21$  indicates how likely the link posting by *solicitors* will trigger the link posting by *sympathizers*. From RQ1 analysis, we know that *solicitors* and *educators* post large proportion of links from extremist websites. However, their higher influence on other roles indicates that they also trigger the spread of extremist content by other roles. Figure 8 in Appendix displays a post by an *educator* account citing a link from extremist domain discussing white identity. The link cited in this post is highly education in that it describes the constructs of white identity and implores the reader to not engage in interracial relationships. Surprisingly, *flamers* also have higher influence compared to *sympathizers* (0.13) and *motivators* (0.11). Overall, these results indicate that *solicitors*, *educators*, and *flamers* are most influential in spreading links from extremist sources.

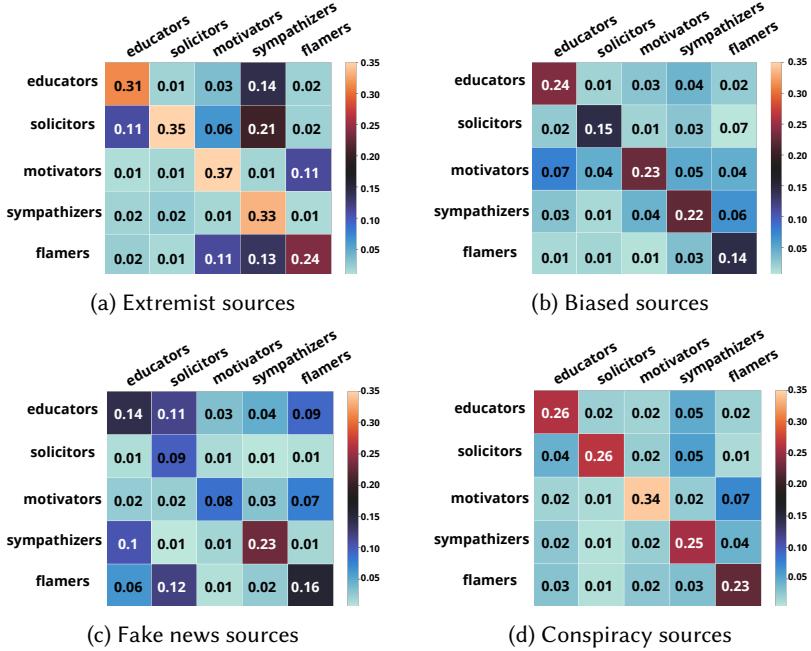


Fig. 5. Figure displaying the pairwise influence matrix for each type of source. The value of  $W_{\text{solicitors} \rightarrow \text{sympathizers}} = 0.21$  in (a) indicates how influential *solicitors* are in triggering extremist information sharing by *sympathizers*. We have used divergent color palette to distinguish between the smaller non-diagonal values that are of most relevance while discussing the results.

**Biased Domains:** As per OpenSources, biased news category contains sources that hold a specific point of view and may rely on propaganda, decontextualizing information and opinions distorted as facts<sup>11</sup>. For example, OpenSources.co lists 100percentfedup.com and dailywire.com as biased sources. While 100percentfedup.com selects stories with extreme right wing bias<sup>12</sup>, dailywire.com is strongly biased toward conservative causes and/or political affiliation<sup>13</sup>. Overall, the influence of all roles is lower while sharing links from the biased sources. However, in comparison to other roles, *motivators* are more influential in triggering biased information posting by other roles ( $W_{\text{motivators} \rightarrow \text{educators}} = 0.07$ ,  $W_{\text{motivators} \rightarrow \text{sympathizers}} = 0.05$ )

**Fake News Domains:** According to OpenSources, fake news sources entirely fabricate information or grossly distort actual news reports. In the qualitative evaluation with experts, we had observed that *flamers* post links that are flagged as fake/misinformative by Facebook fact checkers. Through weight matrix in Figure 5 (c) we quantitatively observe that *flamers* have higher influence on *solicitors* (0.12) in spreading links from fake news sources. Figure 9 in Appendix displays how *flamers* might be posting fake news to spread hate and outrage. Interestingly, *educators* are also influential in spreading fake news to *solicitors* ( $W_{\text{educators} \rightarrow \text{solicitors}} = 0.11$ ). However, it should be noted that fake news sources have lowest number of links and link posting events amongst all source types (Table 4). Overall, based on the weight matrix, *flamers* and *educators* are most influential in spreading links from fake news sources.

<sup>11</sup><https://douglasducote.com/wp-content/uploads/2019/05/OpenSources.pdf>

<sup>12</sup><https://mediabiasfactcheck.com/100-percent-fed-up/>

<sup>13</sup><https://mediabiasfactcheck.com/the-daily-wire/>

**Conspiratorial Domains:** *Sympathizers* are most susceptible to conspiracy information from *solicitors* (0.05) and *educators* (0.05). Similarly, *flamers* are susceptible to conspiratorial information sharing by *motivators*(0.07). The weights in the Figure 5 (d) indicate that *solicitors*, *educators* and *motivators* are most influential in disseminating links from conspiratorial sources.

## 10 DISCUSSION AND IMPLICATIONS

In this work we identify 5 functional roles in online extremist movements and investigate which roles are influential in spreading links from extremist, biased news, fake news and conspiracy sources. Our results can offer insights into how participatory activism advances extremist movements, how various roles are located on the pathways to deeper engagement into extremism, and what could be the possible effects of interventions in countering online extremism undertaken by these roles.

### 10.1 Online Extremist Movements and Participatory Activism

Scholars have attributed the advancement of social movements to the successful distribution of resources through its participants [46]. We observe that through participatory activism, extremist accounts, in various roles, adequately use social media to spread various types of information resources. For example, *educators* and *solicitors* dedicate a large proportion of their Facebook activity to distributing extremist content for educating and soliciting the readers into extremist movements. Moreover, the results of our third research question suggest that they also influence other roles in spreading information from extremist websites. In sum, by disseminating information through their Facebook accounts, mass educating the readers about their agenda, and soliciting funds and participation in the movements, *educators* and *solicitors* are creating human and material resources [46]. Moreover, by prominently sharing misinformative content and using toxic language, *flamers* may be raising emotional resources that create opportunities for public outrage and eventually, collective action [84] to advance the hateful agendas of their extremists movements. This distributed system of online information mobilization—distribution of various information resources through various roles online—can be compared to the democratization process in participatory activism [38]. The digital democratization process specifically consists of more equitable sharing of informational resources amongst the participants [7]. For example, by influencing other roles in sharing links from extremist domains, *solicitors* and *educators* are also *empowering* others with those extremist information. Take for example, the Facebook page of Alliance Defending Freedom (ADF)—an *educator* account in our study, representing a leading anti-LGBTQ organization [73] that the Southern Poverty Law Center has tracked for decades. This organization started with a small group of christian leaders advocating for discredited practice of conversion therapy, criminalization of LGBTQ sexual acts and opposition to the transgender rights. ADF started with 84K Facebook page likes in 2012. Today, ADF's Facebook page has over 1.7 million page likes and over 1.6 million followers. Our dataset also revealed that over 1K other Facebook groups and pages have already shared content linking to ADF's website. With 1.6 million direct followers on the ADF's official Facebook page and an additional indirect exposure to users through shares on other Facebook pages, the material created by ADF is able to reach a vast audience. This may suggest that, through participatory activism, the picture of online extremism has now shifted from a few selected radical websites and accounts to a spectrum of allies with access to extremist information and the affordances to share it with the mass.

### 10.2 Theoretical Implications: Parallels Between Theoretical and Online Roles

Some of our roles correspond to the categories of participants identified in theoretical research based on physical protest events and social movements. For example, the *educators*—accounts that primarily focus on distributing links from extremist domains—may correspond to “constituents” as

described by McCarthy and Zald [46]; “constituents” are primary distributors of resources. Similarly, our *solicitors*—who actively solicit participation via donations and gatherings—may correspond to “beneficiary constituents” [46] who stand to gain from the success, funds, and connections emerging from the movement. The *sympathizers* category may be similar to “bystanders”—a group of third-party participants, as defined by Turner et. al. [80], who might acknowledge grievances related to the issues of social movement and take a sympathetic stand. However, the *motivator* role does not resemble any of the theoretically described categories. We believe that the *motivator* role is specifically relevant in the online setting for relaying positive news and wins related to the extremist causes. Additionally, *flamers* also do not correspond to any theoretical roles. Through our data driven methods, we are able to surface these new roles that characterize online participation in extremist social movements. We believe that our framework for identifying roles based on the characteristics of participation can be extended to other social movements as well. For example, studies investigating positive social movements, such as environmentalist movement, can adopt our methodologies and identify roles and their influence in popularizing environmentalist agenda.

### 10.3 Trajectories of Extremist Movement Participation

Klandermans et. al. proposed a trajectory of social movement participation comprising four steps [36, 37]. First, people must sympathize with the ideals and goals of the movement, thus turning into potential targets for mobilization. Next, they must be targeted by core members’ mobilization attempts. Next, they must develop motivation to participate in the movement and finally overcome possible barriers and engage in collective action [77]. Through our analysis, we identify a group of accounts played the role of *sympathizers*—pages/groups expressing sympathy towards the extremist causes without getting heavily involved. 36.4% of the accounts in our dataset are *sympathizers*. Based on Klandermans’s models, *sympathizers* can also be viewed as the biggest potential group of supporters for the extremist movements. Interestingly, we also see that *educators*, *solicitors*, and *flamers* have high influence on *sympathizers* in spreading extremist content (Figure 5 (a)). This is the second step in the Klandermans’s model, whereby *sympathizers* are targeted for mobilization. In other words, *sympathizers* may lie on the first two steps of the Klandermans’s trajectories of participation. The third step consists of participants who have developed motivation for participating. The *flamer* and *motivator* roles are primarily driven by motivating factors such as anger, injustice, and the sense of achievement. Hence, *flamers* and *motivators* might be on the third step of Klandermans’s trajectory. Finally, we believe that *solicitors* and *educators* are on the last step of the participation trajectory as they actively try to educate and proselytize others through collective action. In summary, based on Klandermans’s comprising, *developing sympathy* → *getting targeted by mobilization* → *developing motivation* → *collective action*, is equivalent to the following role transitions: *sympathizers* → (*flamers* or *motivators*) → (*educators* or *solicitors*), which together represents a trajectory of deeper engagement into the extremist movements online. Can targeted interventions for counter-extremism stop users from getting induced into the deep trenches of extremist movements? Below, we discuss which roles could potentially benefit from interventions designed for countering online extremism.

### 10.4 Practical Implications: Interventions for Online Extremism Engagement

While this study focuses on identifying roles, it can also inform the design of interventions for countering extremism. Our results suggest that while accounts core to the extremist movements—*educators* and *solicitors*—tend to retain their roles, others are more likely to transition to different roles. For example, *flamers* and *motivators* become *sympathizers* with high probability. *Flamers*, *motivators* and *sympathizers* also show more sporadic engagement with sharing extremist links compared to the *educators* and *solicitors*. A study by Siegel and Badaan revealed that targeted

interventions against hate speech, such as sanctions on hateful messages, leads users to tweet less hateful content, especially if the individuals are less engaged with the hate speech in the first place [71]. On the other hand, accounts that frequently see or produce hostile language are less likely to get deterred by sanctions and may even express backlash [71]. Other researchers also report that rather than conforming to the community norms upon receiving sanctions, the producers of hostile content are more likely to move to other platforms [50] or find creative ways of continuing their hate speech [8]. For example, recall that the Facebook page, *Pissed off White Americans*, described in the Introduction, shared videos that are now banned on YouTube. However, they still made the extremist videos available to the readers by hosting them on [bitchute.com](#) which has been described as the “hotbed for violence and hate” [1]. Considering this, our results suggest that *flamers, motivators and sympathizers*—accounts infrequently exposed to extremist content—might benefit most from targeted interventions designed to counter extremism. On the contrary, *educators* and *solicitors* may retaliate or relocate to alternate platforms in response to an intervention.

## 11 LIMITATIONS AND FUTURE DIRECTIONS

Our work has some limitations which also open up promising future directions. First, our dataset contains only US-based extremist websites and most hold far-right political ideology. However, this skew towards far-right might not correctly represent the political scenarios from other countries. For example, unlike the U.S., Germany has observed increased political violence from both, far-left and far-right ideological groups [60] and is known to have a history of violence from both ends of the political spectrum [31]. Hence, while applying our study results in the context of other countries, researchers need to be cautious about the distribution of extremist ideologies in our dataset. Next, we compiled extremist accounts based on their sharing behavior, specifically the number of unique links they shared from known SPLC-designated extremist websites. While this is a common methodological choice made while choosing users/accounts for studying social media activity, a stricter selection criteria can be beneficial. For example, in addition to the frequency of the extremist link posts, extremist accounts can be selected based on the topics discussed in the posts. Moreover, while our data collection spans the activity of extremist accounts in 2018 and 2019, it was collected post-hoc, in May 2020. Once a Facebook page/group is banned, its data is no longer accessible through CrowdTangle or any official Facebook API. Thus, considering the recent Facebook bans on the white nationalist accounts, our study doesn’t contain the data for the extremist accounts that were active in 2018–2019 but got banned before May 2020. In future, researchers can collect data in real time and extend our study to understand what roles among the extremist accounts face highest moderation. Additionally, our RQ2 analysis does not account for potential cohort effect. Specifically, in RQ2, we measure role transition for extremist accounts in various time periods, across 2018 and 2019. While we normalize the features to control for overall trends happening on social media, external events happening between 2018 and 2019 (e.g., a sudden popularity of an extremist content or change in the language surrounding issues of extremism) could affect different users differently introducing cohort effects. In future, researchers can adapt our methods and extend them to controlled experiments so as to mitigate period and cohort effects. Moreover, we observe the influence of various roles in spreading information on Facebook based on the link posting activity in time. Naturally, this approach does not account for the various modes of information, such as images, memes, screenshots, videos, web articles etc. and how that might effect the role dynamics. Measuring influence for various modes of information can lead to interesting findings such as, which roles are influential in propagating memes related to extremist issues. Moreover, our influence measurement is based on the temporal characteristics of information flow. A deeper qualitative analysis of link chains and the comments on the links could help establish causal relationship of influence between various roles. Finally, our results reveal the information

ecosystem of extremist movements on just one platform—Facebook. The problem of extremism is also evident on other platforms such as Twitter and Gab. Indeed, researchers have shown that the extremist movements leverage different social media platforms towards different goals such as radicalization and mass education [58]. We encourage future researchers to extend our methods to model roles in extremist movements on other platforms or even, across the platforms.

## 12 CONCLUSION

In this work we analyze the online ecosystem of extremist movements through the lens of participatory activism. Specifically, we identify five social roles in online extremist movements: *educators*, *solicitors*, *flamers*, *motivators* and *sympathizers*. We also investigate the role dynamics and influence exerted by roles in spreading links from extremist, fake news, biased, and conspiratorial sources. We not only find that the roles core to the extremist movement (*educators* and *solicitors*) are more stable but that they also have higher influence on other roles in spreading extremist content. Our findings offer a perspective on how participatory activism might be advancing extremist movements and how various roles may be targeted for mobilization. Our results also have implications in extending theories of social movement participation and understanding the effectiveness of online counter-extremism strategies.

## 13 ACKNOWLEDGMENTS

This paper would not be possible without the expert feedback by the entire Pennebaker Language Lab directed by Dr. James Pennebaker at University of Texas Austin. We also want to acknowledge the valuable feedback from the members of Social Computing Lab at Virginia Tech and University of Washington, Seattle. This project was partially funded by the Minerva Research Initiative.

## REFERENCES

- [1] ADL. 2020. BitChute: A Hotbed of Hate | Anti-Defamation League. <https://www.adl.org/blog/bitchute-a-hotbed-of-hate>. (Accessed on 10/14/2020).
- [2] ADL. 2020. Who We Are | About Anti-Defamation League | ADL. <https://www.adl.org/who-we-are>. (Accessed on 10/15/2020).
- [3] Ofer Arazy, Hila Liifshitz-Assaf, Oded Nov, Johannes Daxenberger, Martina Balestra, and Coye Cheshire. 2017. On the "how" and "why" of emergent role behaviors in Wikipedia. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*. 2039–2051.
- [4] Balca Arda. 2015. The construction of a new sociality through social media: The case of the Gezi uprising in Turkey. *Conjunctions. Transdisciplinary Journal of Cultural Participation* 2, 1 (2015), 72–99.
- [5] Prakhar Biyani, Cornelia Caragea, Prasenjit Mitra, and John Yen. 2014. Identifying emotional and informational support in online health communities. In *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers*. 827–836.
- [6] David G Bromley and Anson D Shupe Jr. 1980. Financing the new religions: A resource mobilization approach. *Journal for the Scientific Study of Religion* (1980), 227–239.
- [7] William K Carroll and Robert A Hackett. 2006. Democratic media activism through the lens of social movement theory. *Media, culture & society* 28, 1 (2006), 83–104.
- [8] Stevie Chancellor, Jessica Annette Pater, Trustin Clear, Eric Gilbert, and Munmun De Choudhury. 2016. # thyghgapp: Instagram content moderation and lexical variation in pro-eating disorder communities. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*. 1201–1213.
- [9] Xi Chen, Shen Zhao, and Wei Li. 2019. Opinion dynamics model based on cognitive styles: field-dependence and field-independence. *Complexity* 2019 (2019).
- [10] Catherine Corrigall-Brown. 2011. *Patterns of protest: Trajectories of participation in social movements*. Stanford University Press.
- [11] Jacob Davey and Julia Ebner. 2017. The Fringe Insurgency. Connectivity, Convergence and Mainstreaming of the Extreme Right. *Institute for Strategic Dialogue* (<http://www.isdglobal.org/wp-content/uploads/2017/10/The-Fringe-Insurgency-221017.pdf>) (2017).

- [12] Claes H De Vreese and Hajo Boomgaarden. 2006. News, political knowledge and participation: The differential effects of news media exposure on political knowledge and participation. *Acta Politica* 41, 4 (2006), 317–341.
- [13] Mario Diani. 1992. The concept of social movement. *The sociological review* 40, 1 (1992), 1–25.
- [14] Joan Donovan. 2019. Extremists Understand What Tech Platforms Have Built - The Atlantic. <https://www.theatlantic.com/ideas/archive/2019/03/extremists-understand-what-tech-platforms-have-built/585136/>. (Accessed on 09/17/2020).
- [15] Bob Edwards and John D McCarthy. 2004. Resources and social movement mobilization. *The Blackwell companion to social movements* (2004), 116–152.
- [16] Paul Embrechts, Thomas Liniger, and Lu Lin. 2011. Multivariate Hawkes processes: an application to financial data. *Journal of Applied Probability* 48, A (2011), 367–378.
- [17] Myra Marx Ferree and Frederick D Miller. 1985. Mobilization and meaning: Toward an integration of social psychological and resource perspectives on social movements. *Sociological Inquiry* 55, 1 (1985), 38–61.
- [18] Institute for Economics and Peace. 2020. GTI-2020-web-1.pdf. <https://www.visionofhumanity.org/wp-content/uploads/2020/11/GTI-2020-web-1.pdf>. (Accessed on 01/11/2021).
- [19] Tony Formica. 2020. A Social (Media) Contract: Reconciling American Freedom and Security in an Age of Online Radicalization and Extremism. *Yale J. Int'l Aff.* 15 (2020), 131.
- [20] JA Frimer, R Boghrati, J Haidt, J Graham, and M Dehgani. 2019. Moral foundations dictionary for linguistic analyses 2.0. *Unpublished manuscript* (2019).
- [21] William A Gamson, William Anthony Gamson, William Anthony Gamson, and William A Gamson. 1992. *Talking politics*. Cambridge university press.
- [22] Malcolm Gladwell. 2010. Small change. *The New Yorker* 4, 2010 (2010), 42–49.
- [23] Rahul Goel, Sandeep Soni, Naman Goyal, John Paparrizos, Hanna Wallach, Fernando Diaz, and Jacob Eisenstein. 2016. The social dynamics of language change in online networks. In *International Conference on Social Informatics*. Springer, 41–57.
- [24] Max Halupka. 2018. The legitimisation of clicktivism. *Australian Journal of Political Science* 53, 1 (2018), 130–141.
- [25] William L. Hamilton, Justine Zhang, Cristian Danescu-Niculescu-Mizil, Dan Jurafsky, and Jure Leskovec. 2017. Loyalty in Online Communities. (2017). <http://arxiv.org/abs/1703.03386>
- [26] Benjamin D Horne, Sibel Adali, and Sujoy Sikdar. 2017. Identifying the social signals that drive online discussions: A case study of reddit communities. In *2017 26th International Conference on Computer Communication and Networks (ICCN)*.
- [27] IACP. 2020. Facebook and Violent Extremism Awareness Brief. <https://www.theiacp.org/resources/document/facebook-and-violent-extremism-awareness-brief>. (Accessed on 10/15/2020).
- [28] Kirk Job-Sluder and Sasha A Barab. 2004. Shared "we" and shared "they" indicators of group identity in online teacher professional development. *Designing for virtual communities in the service of learning* (2004).
- [29] NF Johnson, R Leahy, N Johnson Restrepo, N Velasquez, M Zheng, P Manrique, P Devkota, and Stefan Wuchty. 2019. Hidden resilience and adaptive dynamics of the global online hate ecology. *Nature* 573, 7773 (2019), 261–265.
- [30] Seth G Jones, Catrina Doxsee, and Nicholas Harrington. 2020. The escalating terrorism problem in the United States. (2020).
- [31] Sebastian Jungkunz. 2019. Towards a measurement of extreme left-wing attitudes. *German Politics* 28, 1 (2019), 101–122.
- [32] Alex Kaplan. 2018. The head of an anti-immigration PAC runs Facebook pages that share fake news from plagiarized sites | Media Matters for America. <https://www.mediamatters.org/facebook/head-anti-immigration-pac-runs-facebook-pages-share-fake-news-plagiarized-sites>. (Accessed on 10/15/2020).
- [33] Jessalynn Marie Keller. 2012. Virtual feminisms: Girls' blogging communities, feminist activism, and participatory politics. *Information, Communication & Society* 15, 3 (2012), 429–447.
- [34] Makenna Kelly. 2020. Facebook still hosts boogaloo extremist groups, report finds - The Verge. <https://www.theverge.com/2020/8/12/21365278/facebook-boogaloo-tech-transparency-right-wing-extremist-platform>. (Accessed on 09/13/2020).
- [35] Makenna Kelly. 2020. Facebook still hosts boogaloo extremist groups, report finds - The Verge. <https://www.theverge.com/2020/8/12/21365278/facebook-boogaloo-tech-transparency-right-wing-extremist-platform>. (Accessed on 10/15/2020).
- [36] Bert Klandermans. 1984. Mobilization and participation: Social-psychological expansions of resource mobilization theory. *American sociological review* (1984), 583–600.
- [37] Bert Klandermans and Dirk Oegema. 1987. Potentials, networks, motivations, and barriers: Steps towards participation in social movements. *American sociological review* (1987), 519–531.
- [38] Michael Krona. 2019. 5 ISIS's Media Ecology and Participatory Activism Tactics. *The Media World of ISIS* (2019), 101.
- [39] Srijan Kumar, Xikun Zhang, and Jure Leskovec. 2019. Predicting dynamic embedding trajectory in temporal interaction networks. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 1269–1278.

- [40] Ryan Lenz. 2013. Following the White Rabbit | Southern Poverty Law Center. <https://www.splcenter.org/fighting-hate/intelligence-report/2013/following-white-rabbit>. (Accessed on 10/15/2020).
- [41] Erik Lewis and George Mohler. 2011. A nonparametric EM algorithm for multiscale Hawkes processes. *Journal of Nonparametric Statistics* 1, 1 (2011), 1–20.
- [42] Scott Linderman and Ryan Adams. 2014. Discovering latent network structure in point process data. In *International Conference on Machine Learning*. 1413–1421.
- [43] Jenni Marsh and Tara Mulholland. [n.d.]. How the Christchurch terrorist attack was made for social media - CNN. <https://www.cnn.com/2019/03/15/tech/christchurch-internet-radicalization-intl/index.html>. (Accessed on 09/23/2020).
- [44] Gary T Marx and James L Wood. 1975. Strands of theory and research in collective behavior. *Annual review of sociology* 1, 1 (1975), 363–428.
- [45] John McCarthy and Mayer N Zald. 2003. Social movement organizations. *The social movements reader: Cases and concepts* (2003), 169–186.
- [46] John D McCarthy, Mayer N Zald, Gary Long, Anthony Oberschall, Anthony Orum, Kathy Pearce, Jack Seidman, and Benjamin Walter. [n.d.]. *Resource Mobilization and Social Movements: A Partial Theory*. Technical Report. <http://www.journals.uchicago.edu/t-and-c>
- [47] Evgeny Morozov. 2009. The brave new world of slacktivism. *Foreign policy* 19, 05 (2009).
- [48] Aldon Morris. 2002. Leadership in Social Movements Aldon Morris and Suzanne Staggenborg. (2002).
- [49] Karsten Müller and Carlo Schwarz. 2020. Fanning the Flames of Hate: Social Media and Hate Crime. *SSRN* (2020).
- [50] Edward Newell, David Jurgens, Haji Mohammad Saleem, Hardik Vala, Jad Sassine, Caitrin Armstrong, and Derek Ruths. 2016. User Migration in Online Social Networks: A Case Study on Reddit During a Period of Community Unrest.. In *ICWSM*. 279–288.
- [51] Permanent Culture Now. 2018. Introduction to activism. *Recuperado de < http://www permanentculturenow com/what-is-activism* (2018).
- [52] Sonia Núñez Puente, Diana Fernández Romero, and Susana Vázquez Cupeiro. 2017. Online feminist practice, participatory activism and public policies against gender-based violence in Spain. *Feminist Theory* 18, 3 (2017), 299–321.
- [53] Jonathan A Obar, Paul Zube, and Clifford Lampe. 2012. Advocacy 2.0: An analysis of how advocacy groups in the United States perceive and use social media as tools for facilitating civic engagement and collective action. *Journal of information policy* 2 (2012), 1–25.
- [54] Anthony Oberschall. 1973. *Social conflict and social movements*. Prentice hall.
- [55] Nicholas Owen. [n.d.]. The conscience constituent reconsidered. In *Other People's Struggles*. Oxford University Press.
- [56] Umashanthi Pavalanathan and Munmun De Choudhury. 2015. Identity management and mental health discourse in social media. In *WWW*. ACM.
- [57] Tsveta Petrova and Sidney Tarrow. 2007. Transactional and participatory activism in the emerging European polity: The puzzle of East-Central Europe. *Comparative political studies* 40, 1 (2007), 74–94.
- [58] Shruti Phadke and Tanushree Mitra. 2020. Many Faced Hate: A Cross Platform Study of Content Framing and Information Sharing by Online Hate Groups. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) (*CHI '20*). Association for Computing Machinery, New York, NY, USA, 1–13. <https://doi.org/10.1145/3313831.3376456>
- [59] Whitney Phillips. 2018. Data & Society – The Oxygen of Amplification. <https://datasociety.net/library/oxygen-of-amplification>. (Accessed on 09/23/2020).
- [60] Counter Extremism Project. 2020. Germany: Extremism & Counter-Extremism | Counter Extremism Project. <https://www.counterextremism.com/countries/germany>. (Accessed on 01/14/2021).
- [61] The Tech Transparency Project. 2020. Facebook's Boogaloo Problem: A Record of Failure | Tech Transparency Project. <https://www.techtransparencyproject.org/articles/facebook-boogaloo-problem-record-failure>. (Accessed on 10/15/2020).
- [62] Babak Rahimi. 2016. Vahid Online: Post-2009 Iran and the Politics of Citizen Media Convergence. *Social Sciences* 5, 4 (2016), 77.
- [63] ADL Reports. 2019. Right-Wing Extremism Linked to Every 2018 Extremist Murder in the U.S., ADL Finds | Anti-Defamation League. <https://www.adl.org/news/press-releases/right-wing-extremism-linked-to-every-2018-extremist-murder-in-the-us-adl-finds>. (Accessed on 09/23/2020).
- [64] Howard Rheingold. 2007. *Smart mobs: The next social revolution*. Basic books.
- [65] Marian-Andrei Rizoiu, Young Lee, Swapnil Mishra, and Lexing Xie. 2017. A tutorial on hawkes processes for events in social media. *arXiv preprint arXiv:1708.06401* (2017).
- [66] Debbie Rodan and Jane Mummery. 2017. *Activism and digital culture in Australia*. Rowman & Littlefield.
- [67] Dana Rotman, Sarah Vieweg, Sarita Yardi, Ed Chi, Jenny Preece, Ben Schneiderman, Peter Pirolli, and Tom Glaisyer. 2011. From slacktivism to activism: participatory culture in the age of social media. In *CHI'11 Extended Abstracts on Human Factors in Computing Systems*. 819–822.

- [68] Roser Sauri. 2008. *A factuality profiler for eventualities in text*. Ph.D. Dissertation. Ph. D. thesis, Brandeis University.
- [69] Swathi Shanmugasundaram. 2018. Anti-immigrant roundup: 7/6/18 | Southern Poverty Law Center. <https://www.splcenter.org/hatewatch/2018/07/06/anti-immigrant-roundup-7618>. (Accessed on 10/15/2020).
- [70] Simon Sheather. 2009. *A modern approach to regression with R*. Springer Science & Business Media.
- [71] Alexandra A Siegel and Vivienne Badaan. 2020. # No2Sectarianism: Experimental Approaches to Reducing Sectarian Hate Speech Online. *American Political Science Review* 114, 3 (2020), 837–855.
- [72] Bernd Simon and PG Klandermans. 2001. Toward a social psychological analysis of politicized collective identity: Conceptualization, antecedents and consequences. *American Psychologist* 56 (2001), 319–331.
- [73] SPLC. 2020. Alliance Defending Freedom | Southern Poverty Law Center. <https://www.splcenter.org/fighting-hate/extremist-files/group/alliance-defending-freedom>. (Accessed on 10/15/2020).
- [74] SPLC. 2020. National Vanguard | Southern Poverty Law Center. <https://www.splcenter.org/fighting-hate/extremist-files/group/national-vanguard>. (Accessed on 10/15/2020).
- [75] Hatewatch Staff. 2020. Facebook's Strategy for Taking Down Hate Groups is Spotty and Ineffective | Southern Poverty Law Center. <https://www.splcenter.org/hatewatch/2020/04/07/facebook-s-strategy-taking-down-hate-groups-spotty-and-ineffective>. (Accessed on 09/15/2020).
- [76] Kate Starbird, Ahmer Arif, and Tom Wilson. 2019. Disinformation as collaborative work: Surfacing the participatory nature of strategic information operations. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (2019), 1–26.
- [77] Stefan Stürmer, Bernd Simon, Michael Loewy, and Heike Jörger. 2003. The dual-pathway model of social movement participation: The case of the fat acceptance movement. *Social Psychology Quarterly* (2003), 71–82.
- [78] Hiroki Takikawa and Takuto Sakamoto. 2019. The moral-emotional foundations of political discourse: a comparative analysis of the speech records of the US and the Japanese legislatures. *Quality & Quantity* (2019), 1–20.
- [79] Yla R Tausczik and James W Pennebaker. 2010. The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of language and social psychology* 29, 1 (2010), 24–54.
- [80] Ralph H Turner. 1969. The public perception of protest. *American Sociological Review* (1969), 815–831.
- [81] Sebastián Valenzuela. 2013. Unpacking the use of social media for protest behavior: The roles of information, opinion expression, and activism. *American behavioral scientist* 57, 7 (2013), 920–942.
- [82] Jacquelien van Stekelenburg and Bert Klandermans. 2013. Social psychology of movement participation. *The Wiley-Blackwell Encyclopedia of Social and Political Movements* (2013).
- [83] Jacquelien Van Stekelenburg and Bert Klandermans. 2013. The social psychology of protest. *Current Sociology* 61, 5-6 (2013), 886–905.
- [84] Jacquelien Van Stekelenburg and Bert Klandermans. 2017. Individuals in movements: A social psychology of contention. In *Handbook of social movements across disciplines*. Springer, 103–139.
- [85] Ariadne Vromen. 2017. Digital citizenship and political engagement. In *Digital Citizenship and Political Engagement*. Springer, 9–49.
- [86] Mattias Wahlström, Abby Peterson, and Magnus Wennerhag. 2018. “CONSCIENCE ADHERENTS” REVISITED: NON-LGBT PRIDE PARADE PARTICIPANTS. *Mobilization: An International Quarterly* 23, 1 (2018), 83–100.
- [87] Yi-Chia Wang, Robert Kraut, and John M Levine. 2012. To stay or leave? The relationship of emotional and informational support to commitment in online health support groups. In *Proceedings of the ACM 2012 conference on computer supported cooperative work*. 833–842.
- [88] Micah White. 2010. Clicktivism is ruining leftist activism. *The Guardian* 12, August (2010), 2010.
- [89] Janyce Wiebe, Theresa Wilson, and Claire Cardie. 2005. Annotating expressions of opinions and emotions in language. *Language resources and evaluation* 39, 2-3 (2005), 165–210.
- [90] Mayer N Zald and Roberta Ash. 1966. Social movement organizations: Growth, decay and change. *Social forces* 44, 3 (1966), 327–341.
- [91] Savvas Zannettou, Tristan Caulfield, Emiliano De Cristofaro, Nicolas Kourteliris, Ilias Leontiadis, Michael Sirivianos, Gianluca Stringhini, and Jeremy Blackburn. 2017. The web centipede: understanding how web communities influence each other through the lens of mainstream and alternative news sources. In *Proceedings of the 2017 Internet Measurement Conference*. 405–417.

threshold for extremist links	accounts added because of the new threshold	clustering of additional accounts with original cluster centers				
		sympathizers	motivators	flamers	educators	solicitors
>8 links	536	86%	7%	4%	2%	1%
>6 links	1023	81%	5%	10%	2%	2%
>4 links	9278	75%	12%	9%	3%	1%

Table 5. Table showing the number of new accounts that could be added in the dataset with various thresholds for the number of extremist links. Our analysis is based on the threshold of >10 extremist links per account. The table also describes what percent of the new accounts fall across various roles using the cluster centers obtained in RQ1. The majority of the new accounts fall under the sympathizer role.

## A APPENDIX

### A.1 Cluster Analysis with Various Link Thresholds

In RQ1, we select extremist accounts that post atleast 10 unique extremist links in the analysis time window. While we select this threshold due to the 95th percentile of the extremist link posting distributions across all the accounts, here we present analysis for different thresholds. Specifically, we consider accounts that would have been added to the dataset if the threshold was >8, >6 and >4 extremist links posted. We find that the majority of the new accounts would have been added to the *sympathizers* category. This is not surprising as *sympathizers* have lowest proportion of extremist links compared to other roles. We also experimented with keeping all the accounts; i.e, accounts that post atleast 1 extremist link in two years. Keeping all accounts added great amount of noise in the feature calculation process. This may be because many defining clustering features (opinions, solicitations, popularity, proportion, trends) are calculated explicitly on the text and frequency of the extremist links and the accounts with only one or two extremist links may not have any words or phrases related to solicitation, opinions or any reactions on the posts. Overall, our post-hoc analysis suggests that even by lowering the threshold to >4 extremist links per account, the constitution of the educators or solicitors—roles core to the extremist movements—does not change significantly.

### A.2 Robustness Tests for Clustering

**A.2.1 Feature Correlation.** We first report the pairwise correlation values between different features used in clustering in Figure 6. Most feature pairs have low correlation. Only pairs of reward, achievement and group identity, opinions are moderately correlated. We further test for multi-collinearity using Variance Inflation Factor (VIF). VIF values of > 5 indicate high multi-collinearity [70]. All of our features have a VIF < 2 indicating low or no multi-collinearity

**A.2.2 Finding Number of Roles.** In order to determine the number of roles, we performed elbow analysis with KMeans algorithm. We used two metrics—inertia and distortions—and plotted their values across different number of clusters (or roles). Distortions are calculated as the average of the squared distances from the cluster centers and inertia is the sum of the squared distance from the cluster centers. Figure 7 (a) displays inertia and distortion plots for the KMeans algorithm where elbow can be observed at  $n = 5$  indicating that 5 is the optimal number of roles.

**A.2.3 Cluster Assignments Overlap.** We obtained role assignment for every extremist account using KMeans algorithm. As a further robustness test, we also compare the role assignments done by KMeans with a different algorithm—Agglomerative clustering. We selected Agglomerative clustering as it is hierarchical clustering algorithm operating on a different clustering mechanism compared to KMeans. Note that since KMeans and Agglomerative clustering are inherently motivated by different

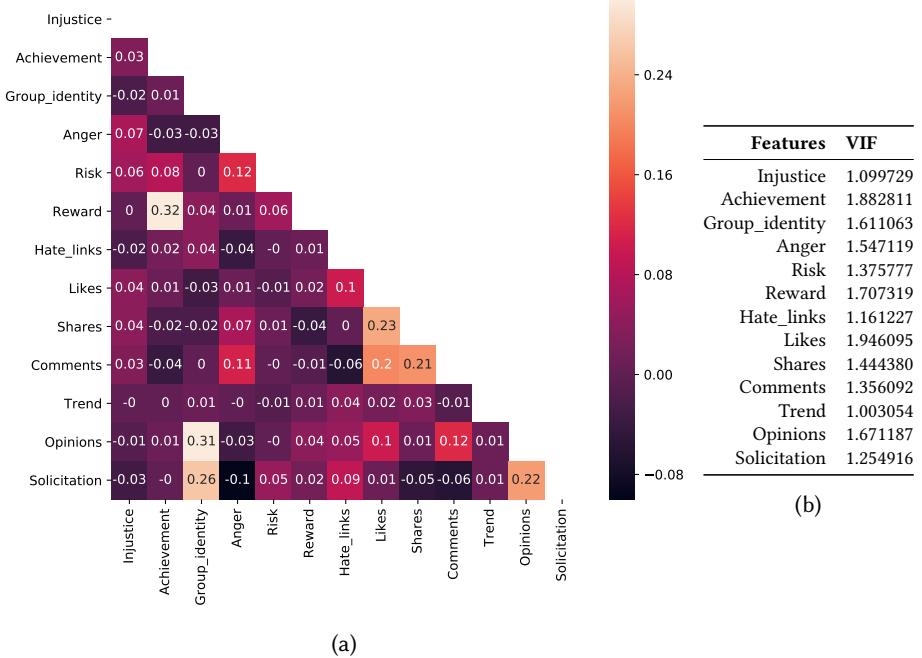


Fig. 6. (a) Figure showing pairwise Pearson correlation coefficients for all features used in the clustering. Most of the features have low correlation. Reward is moderately correlated with Achievement and Group identity is moderately correlated with Opinions. There are no high correlations. (b) Table displaying the Variance Inflation Factor (VIF) for all features. VIF is used to detect multi-collinearity. VIF exceeding 5 indicates high multi-collinearity [70] between that feature and the others. All our features have  $\text{VIF} < 2$  indicating low or no multi-collinearity.

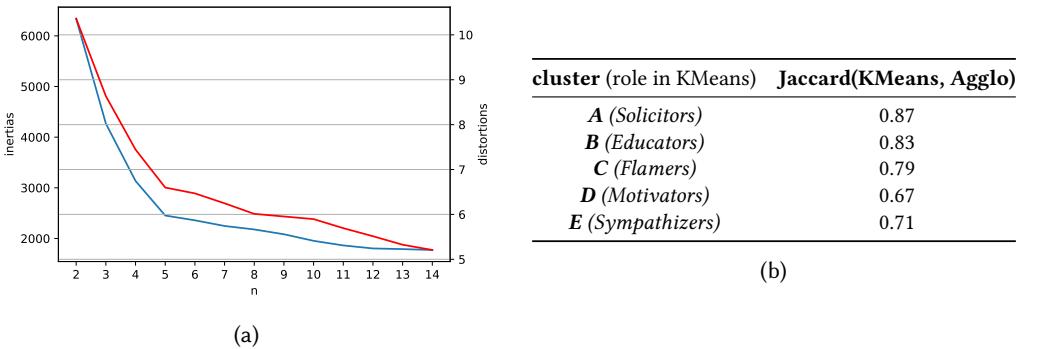


Fig. 7. Figure presenting the elbow analysis using different algorithms. (a) Plots showing distortion and inertia across different number of clusters. Optimal number of clusters can be determined by the “elbow” of the curves. (b) Jaccard similarity scores between corresponding clusters of KMeans and Agglomerative clustering algorithms.

working principles, it is harder to achieve perfect similarity between cluster assignments. However, we present Jaccard similarity scores for cluster assignments using KMeans and Agglomerative

clustering to provide an estimate of overlap between the two algorithms. We first cluster extremist accounts using KMeans algorithm and record cluster assignments. We then cluster the accounts using Agglomerative clustering and record the cluster assignments. For every cluster in the KMeans (eg., KMeans\_A), we find corresponding cluster in the Agglomerative clustering (eg., Aggro\_A) and calculate Jaccard similarity score using common extremist accounts in both clusters. Specifically, we calculate the number of common extremist accounts in Kmeans\_A and Aggro\_A and divide it by total number of unique accounts in Kmeans\_A and Aggro\_A combined. Figure 7 (b) provides the Jaccard similarity values for similar clusters obtained by two different algorithms. On an average the two algorithms have 0.77 similarity in cluster assignments.

### A.3 Tables for Opinion and Solicitation Expressions

Pronoun	POS	Variation	Examples
First person subjective (I, we)	LIWC cogproc verbs (believe, think, infer, propose)	With and without auxiliary verbs, adverbs, negations	I believe... We don't think... I really don't understand... My strong opinion... Our shared understanding of the issue...
First person possessive (my, mine, our, ours)	LIWC cogproc nouns (opinion, understanding)	With or without adjectives	I am hopeful... We are really confused... I might not be supportive...
First person subjective (I, we)	LIWC cogproc adjectives (positive, confused, unclear, hopeful)	With and without auxiliary verbs, adverbs, negations	They should... He should... They definitely must... His incomplete understanding must... Hillary must... Trump should... CAIR actually can...
Third person (he, she it, they, theirs, his, hers, its)	Modal verbs (should, must, can, could, will, might)	With or without nouns, negations, adverbs in the middle	
Proper nouns	Modal verbs (should, must, can, could, will, might)	With or without nouns, negations, adverbs in the middle	

Table 6. Table listing phrase patterns used to extract opinions. We pair various LIWC cognitive processing (cogproc) nouns, verbs and adjectives with pronouns. The last column also lists examples for each pattern.

Pronouns/POS	Pronouns/POS	Variation	Examples
please	LIWC social verbs (donate, call, register)	With and without auxiliary verbs, adverbs	Please donate... Please consider registering...
LIWC social verbs	LIWC social nouns	With or without article, adjective, preposition	Sign the petition... Register for this beautiful event...
LIWC social verbs	First person pronoun	With or without prepositions	Contact us... Register with our...
Second person (you, yours)	Modal verbs (should, must, can, could, will, might)	With or without nouns, negations, adverbs	You should... You really must... You can...
Second person (you, yours)	LIWC social noun	With or without negations, adjectives in the middle	Your wonderful donation... Your timely call... Will you...(sign this petition)? Can you ... (call your senator?)
Verbs used in requests (will, could, can, would)	Second person		

Table 7. Table listing phrase patterns used to extract expressions of solicitation. We pair various LIWC social category nouns, verbs and adjectives with pronouns. The last column also lists examples for each pattern.

### A.4 Analysis of the Domains removed in RQ3

For our RQ3 analysis, we removed all links in time period  $T_1$  sourced from domains not present in the OpenSources dataset or our list of extremist websites. Specifically, we removed 24% of the link posts. It is possible that some of the removed domains could contain extremist, fake news, biased or conspiratorial content. In order to estimate how many of the removed domains could potentially fall under extremist, fake news, biased or conspiratorial content categories, we report 100 most frequently shared domains from the removed links (See Table 8). For each of the 100 domains, we

searched for the bias and credibility ratings using mediabiasfactcheck.com and allsides.com. In Table 8 we highlight domains that have **moderate to extreme political bias (left or right)**, **low or very low factuality rating** and **strong conspiracy or pseudoscience rating**. 20 out of 100 domains are either strongly biased, have factuality or strong conspiratorial content.

foxnews.com	kirkreview.com	bible.com	cbsnews.com	jwatch.us
msn.com	newsweek.com	israelnationalnews.com	cbn.com	strangesounds.org
thehill.com	iheart.com	bloomberg.com	nationalpost.com	beecatholic.com
washingtonpost.com	<b>hannity.com</b>	themaven.net	haaretz.com	medium.com
cnn.com	<b>freebeacon.com</b>	wsj.com	<b>rushlimbaugh.com</b>	globalnews.ca
washingtontimes.com	<b>endtimeheadlines.org</b>	torontosun.com	foxbusiness.com	verelq.am
nytimes.com	<b>gatestoneinstitute.org</b>	politico.com	nydailynews.com	<b>davidharrisjr.com</b>
theguardian.com	independent.co.uk	businessinsider.com	enlaceaudio.com	snopes.com
jpost.com	atheistrepublic.com	<b>dmlnews.com</b>	eventbrite.com	realclearpolitics.com
timesofisrael.com	whitehouse.gov	forbes.com	rightspeak.net	vox.com
nypost.com	bbc.com	wikipedia.org	time.com	whoobazoo.com
elderstatement.com	patheos.com	gofundme.com	mckoysnews.com	deborahbateman.com
<b>rt.com</b>	dailymail.co.uk	<b>globalskywatch.com</b>	<b>patriotbeat.com</b>	ch7.io
<b>townhall.com</b>	fenixx.org	<b>telegraph.co.uk</b>	theresumptimes.com	observador.pt
<b>gellerreport.com</b>	<b>voiceofeurope.com</b>	apple.news	aljazeera.com	algemeiner.com
freespeechtime.net	soundcloud.com	change.org	therebel.media	godaddy.com
huffingtonpost.com	latimes.com	npr.org	chicagotribune.com	aleteia.org
reuters.com	<b>thefederalist.com</b>	ynetnews.com	mirror.co.uk	cbslocal.com
usatoday.com	spencerfernando.com	palinfo.com	<b>cnsnews.com</b>	tapwires.com
nbcnews.com	cnnbc.com	<b>bizpacreview.com</b>	msnbc.com	<b>bitchute.com</b>

Table 8. Table with 100 most frequent domains in URLs removed in the RQ3 analysis. We highlight domains that have **moderate to extreme political bias (left or right)**, **low or very low factuality rating** and **strong conspiracy or pseudoscience rating**.

## A.5 Mathematical Formulation for Hawkes Process

Consider a series of link posting events divided into the time bins  $\Delta t$ . This series is a Hawkes process if the rate of each process has the parameterized form:

$$\lambda_{t,k} = \lambda_{0,k} + \sum_{k'=1}^K \sum_{t'=1}^{t-1} s_{t',k'} \cdot h_{k' \rightarrow k}[t - t'] \quad (1)$$

where  $\lambda_{0,k}$  is the background rate,  $s$  is the matrix of events generated from common link sharing and  $h$  is an impulse response function describing the amplitude of influence that events on process (or role)  $k'$  have on the rate of process (or role)  $k$ . Guided by the descriptions in [42, 91], we can further decompose  $h$  into the weight matrix  $W$  and probability mass function  $G$ . We use the the weight matrix  $W$  of the shape  $W^{K \times K}$  to infer pairwise influence of roles.

## A.6 Example Links from Various Source Types Posted by Influential Roles



Fig. 8. Example of link posted by *educators* from an extremist domain about white identity. The content of the link is highly educational in that, it explains the constructs of the Aryan identity and emphasizes on avoiding interracial relations.

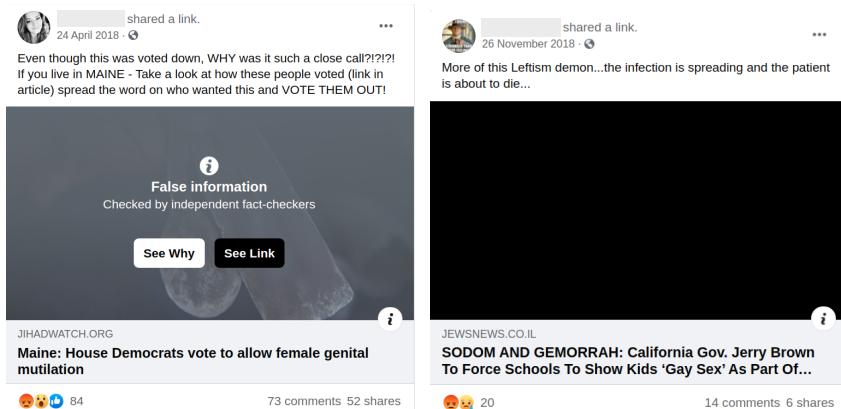


Fig. 9. Example of links posted by *flamers* from fake news domain. The first link presents false information to alarm and mobilize readers against the Democratic party. Second link uses false information alongwith fear based narratives to present anti-LGBTQ agenda.