

Cross-platform Information Operations: Mobilizing Narratives and Building Resilience Through Both ‘Big’ and ‘Alt’ Tech

TOM WILSON, Human Centered Design & Engineering, University of Washington, USA

KATE STARBIRD, Human Centered Design & Engineering, University of Washington, USA

Despite increasing awareness and research about online strategic information operations, there remain gaps in our understanding, including how information operations leverage the wider information ecosystem and take shape on and across multiple social media platforms. In this paper we use mixed methods, including digital trace ethnography, to look beyond a single social media platform to the broader information ecosystem. We aim to understand how multiple social media platforms are used, in parallel and complementary ways, to achieve the strategic goals of online information operations. We focus on a specific case study: the contested online conversation surrounding Syria Civil Defense (the White Helmets), a group of first responders that assists civilians affected by the civil war within the country. Our findings reveal a network of social media platforms from which content is produced, stored, and integrated into the Twitter conversation. We highlight specific activities that sustain the strategic narratives and attempt to influence the media agenda. And we note that underpinning these efforts is the work of *resilience-building*: the use of alternative (non-mainstream) platforms to counter perceived threats of ‘censorship’ by large, established social media platforms. We end by discussing the implications on social media platform policy.

CCS Concepts: • **Human-centered computing** -- **Empirical studies in HCI** • **Human-centered computing** -- **Empirical studies in collaborative and social computing**

KEYWORDS: Information operations; disinformation; media manipulation; social media; Twitter;

ACM Reference format:

Tom Wilson & Kate Starbird. 2021. Cross-platform Information Operations: Mobilizing Narratives and Building Resilience Through Both ‘Big’ and ‘Alt’ Tech. In *PACM on Human Computer Interaction*, Vol. 5, CSCW2, Article 345, October 2021. ACM, New York, NY, USA. 32 pages. <https://doi.org/10.1145/3476086>

1 INTRODUCTION

While there is increasing awareness and research about online strategic information operations—broadly defined as actions by state and non-state actors to distort the information

This work is supported by the National Science Foundation, under grants 1715078 and 1749815

Authors email addresses: tomwi@uw.edu; kstarbi@uw.edu

Permission to make digital or hard copies of part or all 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 third-party components of this work must be honored. For all other uses, contact the owner/author(s).

Copyright © ACM 2021 2573-0142/2021/10 – Article 345 \$15.00
<https://doi.org/10.1145/3476086>

ecosystem to achieve strategic goals, through methods such as the dissemination of disinformation, amplification of specific accounts or content, or the manipulation of the information space [78]—we still do not fully understand the cross-platform dynamics of these operations. Because information operations attempt to leverage the wider information ecosystem [53], focusing on a single platform oversimplifies the problem, providing a limited view of the dynamics of information flow and overlooks how variations in social media platforms contribute to this flow [35].

The aim of this research is to look beyond a single social media platform to the broader information ecosystem to understand how multiple social media platforms are used, in parallel and complementary ways, to achieve the strategic goals of online information operations. In particular, we interrogate the role that different social media platforms play. In other words, how are multiple social media platforms—as sociotechnical infrastructure comprising technology, users, norms, and business practices [3]—used? We also focus on the collaborative activities of actors on and across social media platforms to unpack the kinds of collaborative “work” that users engage in while involved in multi-platform information operations.

The context of this research is the Syrian Civil War, and in particular the online conversation surrounding Syria Civil Defense (the White Helmets), a humanitarian response organization that serves as first responders, assisting civilians affected by the military actions within the country. The White Helmets (WH) predominantly operate in rebel-held (government-opposed) areas of Syria, and through their work have highlighted atrocities carried out against civilians by the Syrian government and their allies, including Russia. As a consequence, the group has been subjected to a long-term information operation that seeks to delegitimize them and weaken their support, particularly in the West [28,40,63,65,79].

Our point of departure is a Twitter dataset of English-language tweets about the WH, collected during 2017–2018 and almost four years of longitudinal observation and analysis of the online discourse surrounding the WH [64,66,79]. In this paper, we use mixed-methods research to extend this “seed” data to look beyond Twitter to other social media platforms from which content was sourced to understand how information operations function within the broader ecosystem [53]. Though this approach means we provide a “Twitter-centric” perspective of the information ecosystem, it does offer insight into how the discourse is shaped through different configurations of social media platforms, and how content flows “downstream” into the Twitter conversation—the breaking news platform of choice of major politicians or journalists [41].

The empirical findings presented in this research provide insight into the established (and largely successful) multi-platform information operation conducted against the WH. We build a picture of the surrounding ecosystem of social media platforms from which content is produced, stored, and integrated into the Twitter conversation. We extend current understanding of the collaborative aspects of online disinformation [64,80] revealing the “work” involved in conducting multi-platform information operations. One aspect of this is the work to sustain the strategic narratives and attempt to influence the agenda through *content production*, *content synthesis*, and *content mobilization* to gain the attention of specific audiences at specific times. But underpinning these efforts is the work of *resilience-building*: the use of “alt-tech”—alternative (non-mainstream) platforms—to counter perceived threat of censorship from “big-tech”, the large, established social media platforms. We end by outlining the implications of this work, particularly with regard to social media platform policy.

2 BACKGROUND

2.1 The Syrian Civil War

The context of this research is the Syrian Civil War, an ongoing conflict that began in 2012 after Arab Spring protests against the rule of Syrian President Assad turned violent. The conflict descended into a multi-sided, internationalized conflict that involves global powers such as Russia and Iran, which are allied with the Syrian government; the US, UK and other global and regional allies, which support the anti-Syrian government ‘rebels’; and ISIS, a now-weakened group that was attempting to establish a caliphate across the region [27,83]. The ongoing conflict has taken many lives and displaced millions more: accurate figures for the number of people killed are disputed, but the Syrian Observatory for Human Rights (2020) put the figure at almost 600,000. The United Nations High Commissioner for Refugees in 2018 reported that 5.6 million Syrians were refugees in other countries while a further 6.6 million were internally displaced [74].

Alongside the kinetic war ‘on the ground’, the Syrian Civil War also manifests itself online as a battle of public opinion. Supporters of the Syrian government and its allies, including Russia and Iran, take one “side”, while on the other “side” are supporters of the US and its allies, including the UK. The research contained herein examines the online manifestation of the Syrian Civil War, and how this takes shape on and across social media. A particular focus of the online discourse is the White Helmets: a volunteer rescue group which receives financial backing from countries such as the US and UK and is applauded by western governments for their volunteer response activities in the war-torn country, but is viewed by Syria and its allies as a ‘soft power’ in the region, operating on behalf of western governments, or as aligned with Islamic terrorist organizations.

2.1 Syria Civil Defense – The White Helmets

Amongst this complicated geopolitical terrain operate Syria Civil Defense, also known as the White Helmets (WH) due to their distinctive headgear. The WH first emerged as loose collaborations between local volunteers fulfilling the role of first responders, rescuing civilians and retrieving bodies in the aftermath of military attacks and airstrikes conducted by the Syrian government or their allies, Russia, in “rebel”-held areas that opposed the Syrian government. After some of the volunteers reached out for financial help from foreign humanitarian groups, James Le Mesurier, a former British army officer, established Mayday Rescue as a means to channel aid to the group, money that came from the US, UK, Canada, Denmark, Germany, and the Netherlands [10].

The group actively promoted their work to garner attention for the plight of the Syrian people caught up in the war: they captured rescues using helmet cameras [63] and were the focus of a Netflix documentary that won an Oscar in 2017 [39]. Their video footage documented atrocities committed by Syria and their allies, Russia, that the WH observed when responding to attacks. Such evidence directly challenged Russia’s claims that it respected international law and was upholding global stability through its role in the Syrian conflict [40]. In response, and in a bid to discredit the group and undermine their reporting of alleged atrocities, the WH became the targets of a sustained online disinformation campaign designed to influence public perceptions of the group [63] and in some cases justify physical attacks against them [29].

For the most part, this information operation was effective in inundating the information space with negative content about the WH and nurturing a community that was willing and able to further advance online activities against the group. As the group’s reputation became

more tarnished (e.g. by associating them with known terrorist groups), it became easier to justify physical attacks against them [29]. As of October 2018, it was reported the once 3000 strong group had lost more than 250 members to the violence [28]. In November 2019, James Le Mesurier (founder of Mayday Rescue, the organization providing financial support to the WH), was found dead in Turkey with injuries consistent with falling from a height [44,81]. The following investigation found that his death was caused by suicide and this has been attributed, at least in part, to the persistent efforts to discredit him through a disinformation campaign [10].

3. METHODS

This study is part of a broader research project encompassing almost four years of research of online activities surrounding the Syrian Civil War and the White Helmets. This broader research project has primarily relied on digital traces—metadata that is generated at the time of online action. Digital traces are rich resources of information, particularly when leveraged through methods such as digital trace ethnography [25], which places the traces in the context in which they were generated. Rather than engaging with the subjects directly or becoming an active participant in the culture (as is the case with a traditional ethnography), we engaged longitudinally with the digital traces that were left behind as artifacts of online interactions². Throughout this research project we have used a grounded, interpretive, mixed-method approach, drawing upon techniques developed within the field of crisis informatics [54,55], including the investigation of online rumors in the context of crisis events [47], and digital trace ethnography [25]. This approach is iterative and blends both quantitative and qualitative analyses—for example, by generating descriptive statistics and visualizations of the data to see from a higher level the patterns and anomalies, and then using those representations as entry points for a closer, qualitative examination of the data [79,80]. Through these methods, we generated queries, plots, and network visualizations to understand the structure and higher-level patterns of the data, plus conducted qualitative analyses of thousands of tweets, and hundreds of articles, videos, and other social media posts. By applying these mixed-methods to digital trace data, we were able to characterize communities of Twitter accounts based upon their stance toward the WH [79], and noted—among both supporters and detractors of the WH—the heterogeneous make-up of the communities as a confluence of online information activists, Western journalists, politicians, state-funded media, and agents of governments that loosely collaborate through computer-supported cooperative work [64,80]. This collaborative work takes place with little evidence of top-down orchestration [64], in contrast to the e.g. the highly-coordinated efforts of the Internet Research Agency circa 2016-17 [1]. In summary, the specific study presented in this paper is part of a larger research project that involved extensive, mixed-method analysis of digital trace data by the first author, as well as efforts of other

² At times we were brought into these discourses, such as when we shared research output, for example a network visualization, and were the recipient of tweets of support and condemnation from both “sides”. With the exception of a small number of journalists, we chose to not directly engage with accounts on either side of this discourse as we focused on our data-driven investigation.

researchers and collaborators over a period of years, in a long-form, detailed investigation of a specific case study.

This study begins with the same Twitter dataset, and the research presented herein is supported by the longitudinal observation and analysis that came before it, as described above. In an extension of our prior work, the aim of this study is to look beyond a single social media platform to the broader information ecosystem and the multi-platform dynamics of information operations.

Following the methodology applied to the broader research project outlined above, we adopt a mixed methods approach, extending upon methodologies of crisis informatics [54,55], studies of misinformation [46,68], and document-driven digital trace ethnography [25] to examine the multi-platform dynamics by following both specific content and the actors involved in its production and dissemination. We combined both an actor- and information-focused approach: identifying the individual accounts and communities that are involved and following these across the information ecosystem; and also tracking specific media and its appearance on and across multiple social media platforms³ through URL (link) sharing [22]. By applying quantitative, qualitative, and visual techniques to this data we sought to build a comprehensive understanding of this phenomenon, including revealing the activities—or ‘work’—of disparate actors that collaborate to sustain an information operation over the long term. In the following subsections we describe in more detail the specific methods we used to prepare and analyze our Twitter data, including how we extended this dataset to look beyond Twitter to other social media platforms by leveraging URLs as digital traces.

3.1 Twitter Data Collection

This research is centered around a Twitter dataset of English-language tweets referencing the “white helmets” or “whitehelmets” (case insensitive), collected in real-time using the Twitter streaming API between May 27, 2017 and June 15, 2018. The data includes tweets and retweets where the terms appear in the text field of the tweet, or in the text of the embedded quote tweet (retweet with comment). The dataset contains 913,028 English-language tweets from 218,302 distinct Twitter accounts. Unfortunately, between February 7-24 2018, our collection infrastructure suffered an outage and we are therefore unable to account for activity at that time (annotated by the grey area on the temporal plot in Figure 3).

There are two main limitations to this approach. The first is that these data—and subsequent analyses—are limited to English-language tweets, and we focus on the aspects of the WH

³ Social media are “forms of electronic communication (such as websites for social networking and microblogging) through which users create online communities to share information, ideas, personal messages, and other content” [62]. The use of the term ‘platform’ is summarized by Gillespie, 2010 [26]: “The more conceptual use of ‘platform’ leans in all of the term’s connotations: computational, something to build upon and innovate from; political, a place from which to speak and be heard;...and architectural, in that [they are] designed as an open-armed, egalitarian facilitation of expression, not an elitist gatekeeper with normative and technical restrictions” (p.352). In this way, the use of the term platform can be viewed as strategic—consistent with rhetoric surrounding the democratizing potential of the internet but at the same time serving to downplay the role of the social media companies, stressing that they are not producing information but are just hosting it. This shifts the responsibility of the content to the user and limits the liability of the company, which can deny their role as a publisher. This is supported through Section 230 of the Communications Decency Act (1997), which offers social media platforms some legal protections from what their users say and do on their infrastructure [75]

discourse that was designed for and taken up by English-speaking audiences. The aim here is to encapsulate a specific and strategic element of the WH discourse: the group has made significant and sustained efforts to communicate in English to western and global audiences, while similarly a significant portion of the anti-WH campaign has occurred within English-language discourse, attempting to counter the WHs' communications by targeting the same or similar audiences. A limitation of this approach is the acknowledgement that this represents just part of the discourse around the WH, and that the conversation may look different in other languages. The second is that this approach—beginning with 'seed' data from Twitter and building an understanding of the wider information ecosystem from there—presents a Twitter-centric view that determines the flows of information that we can see. In other words, the information ecosystem may look different if we used a collection of Facebook or YouTube data as the starting point, or collected data from all platforms concurrently, rather than collecting from Twitter first and then extending the dataset to the other platforms linked-to from Twitter.

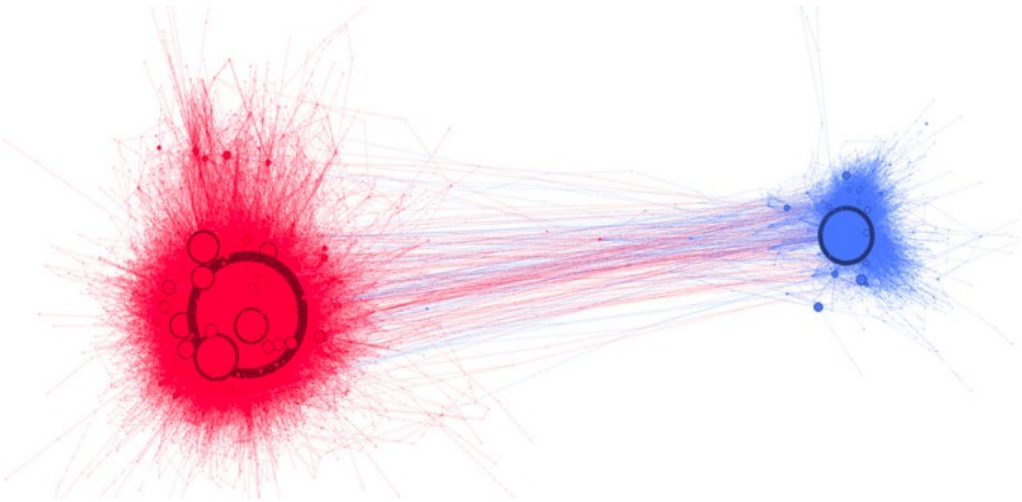


Figure 1: A retweet network visualization of the White Helmets conversation on Twitter illustrating the two distinct communities involved in the conversation: the anti-WH community (red, on the left), and the pro-WH community (blue, on the right). The nodes (circles) represent Twitter accounts which are sized relatively according to the number of times that account was retweeted. The edges (lines) between nodes represent a retweet between that pair of accounts. This is a reproduction of the network visualization published in [79].

3.2 Characterizing the communities in the WH discourse on Twitter

As detailed in [79] we generated a retweet network visualization by identifying retweet trajectories to detect communities of accounts. We used the Louvain algorithm, a greedy optimization method that optimizes the relative density of edges inside a community (modularity) compared to that outside the community [5]. Since edges represent retweets

between a pair of accounts, the algorithm effectively identified two communities⁴ of accounts that retweet one another (Figure 1). The Louvain algorithm is unsupervised, meaning the accuracy of the output is not easily estimated⁵, however this quantitative approach to community detection was cross-referenced with and supported by our qualitative analyses of hundreds of tweets, articles, and videos over the course of our broader research—in other words, the clustering was an efficient way to confirm what we had already observed through manually reading tweets and account bios and building a qualitative understanding of the conversation on Twitter. This combination of quantitative and qualitative approaches allowed us to characterize the detected communities based upon their narrative stance toward the WH. The White Helmets official account is the large central node in the blue community, with several high profile journalists, media outlets, Syrian and Russian government accounts, and political figures are prominent in the Red (these communities are further described in the methodological appendix to [79]). We describe the communities here because they are significant as we describe the methods applied to the data herein.

3.3 The pro-WH (blue) and anti-WH (red) Communities On Twitter

The first community—the pro-WH community (blue)—was almost exclusively supportive of the White Helmets. The *blue pro-WH* community contributed to and promoted the broad narratives that the White Helmets are a humanitarian response organization that provides firefighting, medical aid, rescue, response, and body recovery in rebel-held areas of Syria; that military actions taken by the Syrian/Assad regime are injuring and killing Syrian civilians, including through the use of chemical weapons; and the White Helmets are the target of a disinformation campaign.

In contrast, content from the second (*red, anti-WH*) community was almost exclusively critical of the WH. Within that network multiple, often conflicting, narratives surfaced, all of which functioned to delegitimize the White Helmets through a range of ad hominem attacks, including claims that: the WH are a propaganda construct of the West; the WH are assisted by “Western” and “mainstream” media; the WH are terrorists; the WH are “crisis actors” who stage attacks (including chemical attacks) to make the Syrian government and their Russian allies look bad; the White Helmets are agents of western governments.

3.4 Looking Beyond Twitter to Understand Cross-Platform Activities

⁴ Of the 218,302 distinct Twitter accounts in the White Helmets dataset, 168,110 (77%) were classified as members of these communities (Table 1). The remaining 50,192 accounts were not included because the accounts did not retweet nor were retweeted by another account (32,638 accounts), or; the accounts formed their own, smaller, distinct communities (17,554 accounts). Qualitative analysis of the tweets in these smaller communities revealed that the conversation was not about the White Helmets in Syria, but rather a motorcycle group in the UK and the headgear of a US football team. These unrelated accounts and their tweets were excluded from the subsequent analyses.

⁵ We thank the anonymous CSCW reviewer for their help in refining this section.

We next looked beyond Twitter to other social media platforms from which content was sourced⁶—in other words the platforms that the two communities on Twitter linked-to within their (re)tweets. This can provide insight into the structure of the wider information space [20,48,56,61] and how information operations function within this ecosystem [53]. We unshortened URLs in tweets and resolved to their destination domains. By ranking the domains by the number of (re)tweets⁷ that linked to them, we identified the five social media platforms that were linked-to in at least 1000 (re)tweets posted by accounts in pro- or anti-WH communities on Twitter: YouTube, Steemit, Facebook, Medium, and Liveleak.

As a scoping mechanism, we determined the 25 most (re)tweeted distinct URLs from each platform, and followed those URLs to gather additional data about the content being shared. We focused on the most (re)tweeted as a proxy for what was considered the most important by users [6,8,57]. While this helped us to scope the data for subsequent analysis, a limitation of this approach is that it overlooked WH content that may have been viewed as important by communities on other platforms but not by the communities on Twitter. To provide a fictional example, a heavily viewed video on YouTube that was given the ‘thumbs up’ by thousands of users would not be within our data unless it was also introduced to the Twitter conversation and retweeted enough times to make one of the top-25 most retweeted URLs.

For the 25 most (re)tweeted posts on each platform we collected the publicly available metadata (e.g. title, author/channel, published date, interactions (likes, shares), tags); and analyzed (i.e. through watching/reading) the content (video/post/article). We built a small database of the public metadata, notes that summarized the contents of the post/video, the narrative stance of the content toward the WH, and a breakdown of (re)tweets from the pro- and anti-WH communities on Twitter. Collectively, these data enabled us to build a qualitative understanding of the 125 posts and their contribution to the WH conversation on Twitter.

Examining Twitter Content Integration Through Retweet Cascades

The integration of content to the Twitter conversation was examined through retweet cascades—tweet-retweet combinations visualized temporally. We first identified “introductory” tweets—a tweet, reply, or quote tweet (non-retweet)—that contained a URL, which represented digital links to other regions of the information ecosystem and the point in time that the content was introduced into the Twitter conversation.

Introductory tweets were assigned a *retweet sequence* of 0. The second step involved finding all retweets of each introductory tweet (if they existed) and ordering these temporally. The first (temporal) retweet of an introductory tweet was assigned the *retweet sequence* 1, the second as 2, etc. In the third step, the cascades were plotted to axes, with date-time on the *x*-axis and the retweet sequence on the *y*-axis. The points on the graph were shaded by community membership (red/blue), assigned a shape (filled circle: tweet, quote tweet; filled diamond: reply;

⁶ This approach places Twitter temporally downstream insofar as content shared to Twitter has already been produced and hosted elsewhere.

⁷ We examined both tweets and retweets (herein referred to as “(re)tweets”) so we could analyze both the introduction of content to Twitter (which happens through tweets, replies, or quotes) and dissemination of content (which occurs through retweets).

empty circle: retweet), and relatively sized according to the number of followers of the (re)tweeting account (Figure 2).

The retweet cascade plots enabled interactive examination of the integration of content into the Twitter conversation, for example they reveal which community on Twitter introduced (through introductory tweets) or disseminated (through retweets) specific pieces of content. Retweet cascade plots were used as an investigative tool in the analyses presented below to help understand when and by whom content was introduced to the Twitter conversation and which accounts subsequently retweeted it. Due to space constraints we do not include these plots each time they were used in our analyses.

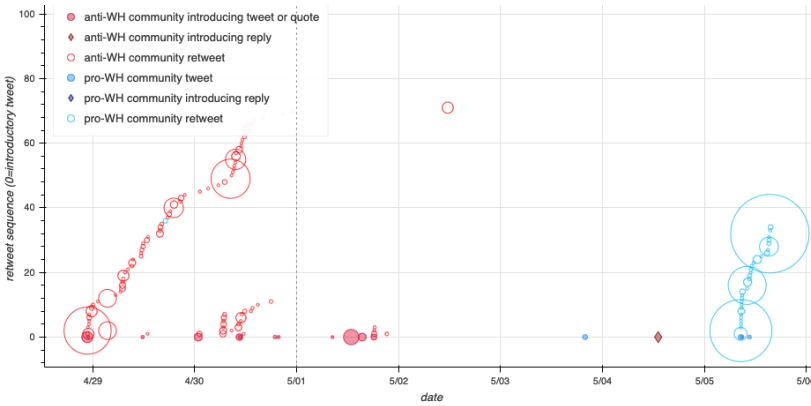


Figure 2: A retweet cascade plot used for examining content integration to Twitter. In this example, the retweet cascades show content introductions (distinct URLs) by accounts in the anti-WH red community on the left, and the pro-WH blue community to the right. The pro-WH content is introduced as a tweet (darker blue circle at $y=0$), and then retweeted by some accounts with larger followings. It receives more than 30 retweets over the course of two hours. Various anti-WH content is introduced, several of which are not retweeted, including the reply just to the left of the pro-WH cascade, which is not retweeted.

4. FINDINGS

We first describe the structure of the WH conversation on Twitter then examine each of the five social media platforms that played a role in shaping the WH discourse. We look at key platform features, provide an overview of the content, describe the introduction of content to Twitter, and summarize how the platform is ‘used’ within this information ecosystem.

4.1 The WH Conversation on Twitter

As previously reported [79], the anti-WH community (red) on Twitter is larger, containing 57.96% of all accounts in the WH discourse, and more active, generating 75.26% of (re)tweets in the conversation (Table 1). On average an account in the red community generated, over the course of our data collection 6.5 WH-related (re)tweets, versus the 3 by accounts in the blue community.

The red anti-WH community is also persistently more active: (re)tweet volumes from the blue community seldom surpass volumes of the red. There are some exceptions to this and we see increased activity from the pro-WH community in response to several key events: the murder of seven members of the WH in August 2017; an article in the Guardian (UK

newspaper) that brought attention to the disinformation campaign being conducted against the WH in December 2017; and news that the US government had ceased its funding for the group in May 2018.

We see marked increase in activity on Twitter in April 2018, which is more pronounced and sustained in the anti-WH community. This period coincides with some major events in the Syrian conflict, including the military offensive by the Syrian Army to capture for rebel-held Eastern Ghouta (Rif Dimashq offensive) that began in February 2018; heavy airstrikes of Douma that included the use of chemical munitions on April 7; missile strikes on multiple Syrian government targets by the US, UK, and France on April 14; and the May 4 announcement by the US government that it would cease funding the WH. Immediately after the Douma chemical attack on April 7 there is a sharp increase in (re)tweets, volume that is sustained through the other events. Due to the salience of YouTube URLs on Twitter during this period, we focus in more detail on this period in the YouTube section of our findings below.

community	distinct accounts	total (re)tweets	tweets	retweets	% retweets	% with url	% replies	mean (re)tweets per account
blue, pro-WH	70,670	207,938	27,708	180,230	86.67	22.72	2.42	2.94
red, anti-WH	97,440	632,638	122,966	509,672	80.56	36.96	9.20	6.49

Table 1: Overview of the blue (pro-WH) and red (anti-WH) communities on Twitter

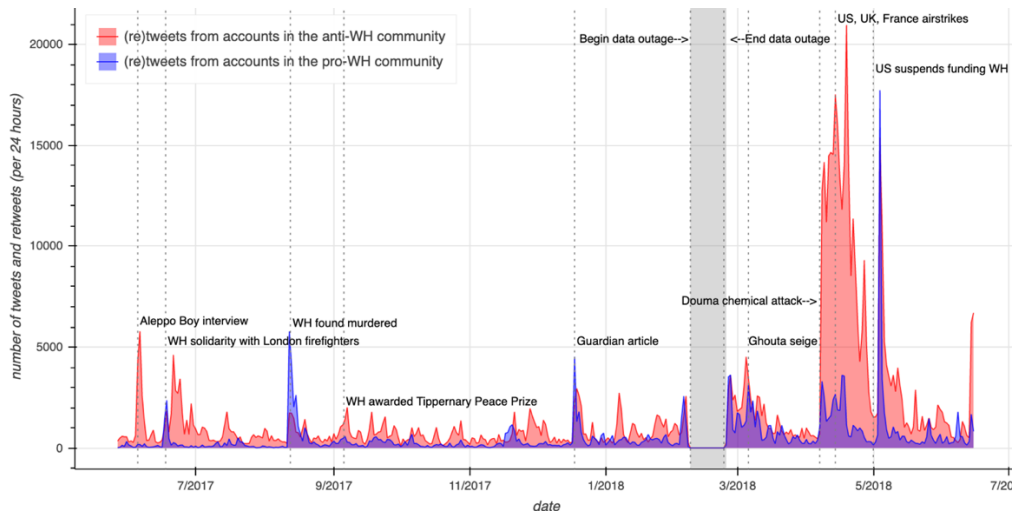


Figure 3: Temporal plot showing the number of (re)tweets per 24-hours from the red & blue communities. The plot is annotated with some key events regarding the WH & the Syrian conflict. Note the marked increase in activity—particularly from the red anti-WH community—after Mar 2018, and the data outage (grey area) in Feb 2018.

A further interesting distinction between accounts in the red and blue communities is their use of the reply function on Twitter. We observed three kinds of reply:

- *in-network replies*: an account in a community replying to an account in the same community (e.g. account in anti-WH community replies to another account in the anti-WH community);
- *cross-network replies*: an account in one community replying to account in the other community (e.g. account in anti-WH community replies to account in the pro-WH community); and
- *out-network replies*: an account in the anti- or pro-WH community replying to an account outside of these communities (i.e. an account not in our retweet network due to limited participation in the discourse on Twitter).

In general, accounts in the anti-WH community demonstrated a greater propensity to use the reply function: 9.20% of the total tweets sent by the red community were replies to other tweets (compared to 2.42% for the pro-WH community). Both the red and blue communities direct the majority of their replies in-network (3.61% and 1.22% of all tweets, respectively), but a much higher proportion of tweets from the red community are out-network replies (bottom right cell, Table 2)—with 3.25% of anti-WH community tweets replying to accounts outside of the WH discourse on Twitter. The out-network replies are generally directed at prominent politicians (e.g. *@realDonaldTrump*; *@jermeycorbyn*) and large media organizations (e.g. *@guardian*; *@cnni*; *@bbc*, *@derspiegel*). This behavior—which illustrates attempts to gain the attention of these prominent figures and organizations and their audiences as a form of content mobilization—is something we will resurface throughout findings and address in depth in our discussion.

reply from account in:	reply to account in:		
	blue (pro-WH) community	red (anti-WH) community	other (not in red/blue communities)
blue (pro-WH) community	in-network replies 1.22% (2542 replies)	cross-network replies 0.69% (1444 replies)	out-network replies 0.51% (1056 replies)
red (anti-WH) community	cross-network replies 2.33% (14729 replies)	in-network replies 3.61% (22885 replies)	out-network replies 3.25% (20592 replies)

Table 2: The percentage (count) of reply tweets that were in-network, cross-network, and out-network by each of the communities on Twitter.

social media platform	distinct urls linking to platform	num (tweets) & retweets from red community	% of red community (re)tweets containing URL	num (tweets) & retweets from blue community	% of blue community (re)tweets containing URL
youtube.com	1822	(11,170) 32,105	18.5	(630) 1,591	4.7
steemit.com	95	(693) 4,866	2.4	(1) 12	0.0
facebook.com	1660	(1,044) 2,391	1.5	(821) 1,029	3.9
medium.com	54	(121) 626	0.3	(207) 806	2.1
liveleak.com	64	(518) 1,046	0.7	(0) 2	0.0

Table 3: The number of URLs linking to content on each platform (re)tweeted by the red (anti-WH) and blue (proWH) communities on Twitter

4.2 Summary of the Top-5 Most Linked-To Social Media Platforms

Patterns in URL sharing (Table 3) demonstrate that accounts in the anti-WH community link to more externally-sourced content in their (re)tweets. Prior work has examined some elements of this wider information ecosystem, including networks of state-funded and alternative media websites that function to foment and distribute anti-WH narratives [65] and the use of YouTube by the anti-WH community on Twitter [79]. The focus herein will be on the links from Twitter to YouTube, Steemit, Facebook, Medium, and Liveleak, which were identified as the most linked-to social media platforms (Table 3). Accounts in the red community sourced content from a wider variety of platforms, including those less familiar such as Steemit and Liveleak. In the third and fifth columns from the left we include the number of tweets (in parenthesis) and retweets of each platform to illustrate both the introduction of content to Twitter (which happens through tweets, replies, or quotes) and dissemination of content (which occurs through retweets).

Table 4 provides an overview of the 25 most (re)tweeted social media posts (could be posts, articles, videos, depending on the platform) from each of these platforms, including a breakdown of pro- or anti-WH posts, the number of unique content producers, and how many posts had since been removed (as of December 2020). To the right of the table, we present the number of introductory tweets and number of retweets of the 25 posts from each platform. We also include the number of unique accounts involved in tweeting and/or retweeting as an indication of how this ‘work’ was distributed among users.

We next present our findings from Twitter: Youtube, Steemit, Facebook, Medium, and LiveLeak. For each we provide an overview of the platform and its features, a description of the types of WH content within our sample of data, and analysis of how the content was introduced into the WH conversation on Twitter.

social media platform	Breakdown of 25 most (re)tweeted tweets				Tweets (introduction to Twitter through URL)		Retweets (content dissemination)	
	pro-WH content	anti-WH content	unique content producers	content removed	num tweets	num unique accounts	num retweets	num unique accounts
youtube.com	0	25	14	1	3548	1085	13945	8004
steemit.com	0	25	5	0	589	108	4491	2887
facebook.com	5	13	9	7	190	48	2213	1693
medium.com	15	8	11	2	292	134	1405	1215
liveleak.com	0	23	2	1	470	37	959	637

Table 4: Overview of the 25 most (re)tweeted social media posts from each of the platforms, including the narrative stance toward the WH, number of unique content producers, and details of the introductions to Twitter.

4.3 YouTube

YouTube is the worlds’ largest video-sharing service and a dominant platform within the digital media environment [7]. Accounts are free and offer users unlimited space to upload and store video content to their *channel*. Social share buttons enable sharing to a wide variety of other platforms, in addition to a specific video URL that can be shared. YouTube videos can be labelled with hashtags, which group similar content and facilitate the surfacing of content through search functionality powered by Google.

YouTube has rules against sensitive content (nudity and sexual content), spam and deceptive practices, and violent or dangerous content [82]. Since YouTube was established its policies have not explicitly prohibited the publishing of videos featuring conspiracies or misleading information [50]. However, in March 2018 YouTube announced *information cues*—links to Wikipedia below content that is contested or conspiratorial in nature [50], or to denote channels that are government-funded [34].

4.3.1 WH Content On Youtube

As previously reported in [79], YouTube is the most linked-to domain in the WH discourse on Twitter, used extensively to host content containing anti-WH messaging. In this sample, 24⁸ of the most-(re)tweeted YouTube URLs include a variety of the anti-WH narratives in circulation, for example claiming terrorist affiliations; accusing the group of staging rescues; and seizing upon their funding by western governments as proof they are “agents of change” or a “soft power” in the region. YouTube content was hosted on 14 distinct channels, a heterogeneous mix of independent journalists, influential individuals in the Syrian context, government-funded media (e.g. RT), and alternative media (e.g. *One America News Network*; *UK Column*).

Several of the YouTube videos were ‘old’ i.e. they predate our Twitter data (i.e. pre-May 2017), including several compilation videos that combine “evidence” of various accusations levelled against the WH. Such compilations are flexible insofar as they can be “dug up” from the YouTube “archives” (Thorson et al. 2013) and reintegrated to the mainstream discourse to serve a variety of purposes. One example is a video from 2016 in which Boris Johnson, then Foreign Secretary of the UK Government, is watching a demonstration exercise by the WH, praising the group, and announcing additional funding for the WH. While the video demonstrates support for the WH and their activities from a western government, it was “dug up” [72] and mobilized by anti-WH voices as evidence that the WH were serving the interests of the UK government in Syria.

On the other hand, 12 out of the 24 available videos were ‘new’, uploaded to YouTube in April and May 2018 yet received enough (re)tweets to make it into the top-25 most (re)tweeted. This is due to the spike in Twitter activity—and links to YouTube—toward the end of our collection period on Twitter (Figure 3). This video content—and the subsequent efforts to introduce it to the WH conversation on Twitter—are in response to events of the Syrian conflict at this time, the most significant of which was the Douma chemical attack that took place on April 7 2018 and of which the WH provided evidence [2,45].

These newer videos reiterate many of the disruptive narratives surrounding the WH—their purported links to terrorism and their funding by the US and UK governments. For example, the most (re)tweeted YouTube video (2199 (re)tweets) is a cellphone recording of Roger Waters (musician) at a concert in early April 2018, in which Waters describes the WH as “fake...jihadists...and terrorists”. But there are also concerted efforts to propagate claims that the WH conduct ‘false flag’ events staged by ‘crisis actors’, and that they have been doing so for some time. The purpose is to discredit the WH and, by association, the evidence that the group provided of a Syrian/Russian chemical attack in Douma. To support the ‘crisis actors’ claim of false flag events, the anti-WH community appropriated the WH’s recording of themselves involved in the “*Mannequin Challenge*”, a 2016 social media trend that involved people remaining frozen in action (like mannequins) while they were filmed. This clip was reframed (and remixed as a clip in other videos) as evidence that the WH stage rescues, picked up by OANN, RT, and the Russian government to support otherwise unfounded claims that the WH

⁸ As of December 2020, 1 video is now unavailable, however During a prior round of data collection and content analysis (when the video was still available) we noted that this video was “hand held footage purportedly showing the WH cheering during an execution”. Assumedly this fell afoul of YouTube policy surrounding violent or dangerous content as was subsequently removed from the platform.

stage attacks and rescues. The tweets introducing this content to Twitter carry increasingly extreme interpretations, including claims that the WH drug children (with anesthesia) before using them to stage fake events; and that they burn barrels of plastic and wait for them to explode to create scenes (and injuries) that are consistent with a chemical attack.

4.3.2 Introducing Youtube Video Content to Twitter (April 1 -May 15 2018)

As illustrated by the temporal plot in Figure 4, YouTube is the most linked-to social media platform in the WH discourse between April 1-May 15 2018: There were 26,365 (re)tweets posted to Twitter: 8,484 (36%) of these contained a link to one of the 25 most (re)tweeted videos on YouTube. In other words, during this 6-week period more than a third of (re)tweets in the WH discourse on Twitter contained a URL to one of these 25 videos. The majority (8436) came from accounts in the red anti-WH community, versus 48 from the blue pro-WH community.

The largest peak on the graph occurs around April 15, coinciding with US, UK, and French airstrikes in Syria in response to the Douma chemical attack. Preceding this there is a substantial increase in (re)tweets containing URLs to YouTube, beginning on April 7 when the WH reported on the chemical attack in Douma. The work to undermine the credibility of the WH intensifies in the wake of the Douma chemical attack in a bid to tarnish their claims that a chemical attack took place. This sustained activity lasts almost three weeks, slowly tapering off toward the end of April, and then picking up again slightly as the US government announced it would cease funding the WH (Atwood, 2018).

Within the 8,484 (re)tweets, 1,742 are introductory tweets that function to introduce YouTube content to the Twitter conversation. More than a third of these introductory tweets (604) are posted as replies to other users. 602 replies are sent by 151 distinct users in the anti-WH community⁹. A small group of five Twitter accounts are responsible for more than half (356) of these replies, illustrating a dedication, among a small number of accounts, to repeatedly introduce this content to Twitter.

These replies are sent to 396 distinct users on Twitter. 172 in-network to accounts such as *@RT_com* (14), *@PrisonPlanet* (10), *@RussianEmbassy* (9), and *@Partisangirl* (7); 73 cross-network including to *@BBCNews* (16), *@BorisJohnson* (10), and *@guardian* (4); and 151 to out-network accounts such as *@realDonaldTrump* (14) and *@jeremycorbyn* (5).

In-network replies often brought new information or “evidence” to the attention of influential accounts in a bid to disseminate it further, as in this example which is a reply to RT, the state-funded Russian broadcaster:

Tweet by @RT_com: ‘No role for WHO’ in confirming or denying suspected chemical attack in Syria’s Douma

Reply to above tweet: @RT_com There was no chemical attack in #Douma. 13 health workers, including 8 doctors and 3 nurses, tell in detail... <<Link to YouTube video of Douma Chemical Attack False Flag Operation EXPOSED!>>

⁹ Two distinct accounts in the pro-WH community also each send a reply to themselves during this period

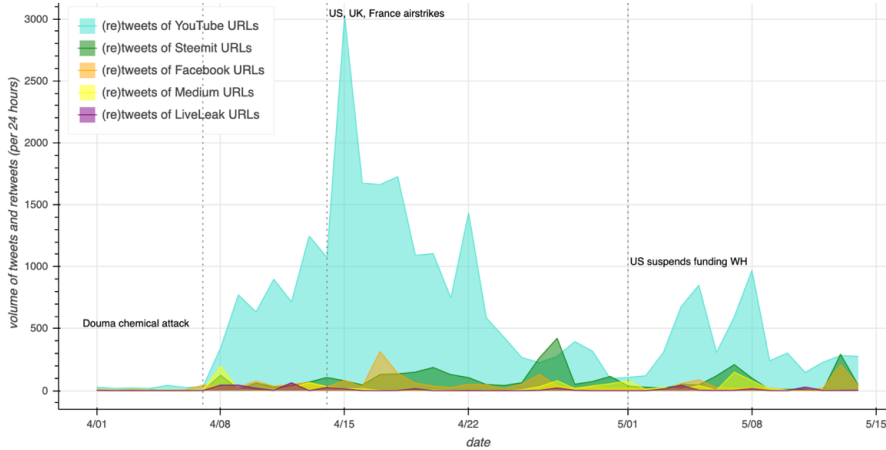


Figure 4: A temporal plot showing the number of (re)tweets, per day containing URLs to Facebook, LiveLeak, Medium, Steemit, and YouTube between April 1 2018 and May 15 2018 inclusive

In-network replies were also sent to influential accounts in the anti-WH community on Twitter. These replies attempt to garner the attention of these users to leverage their influence within the community, i.e. to retweet the content to their large number of followers who are likely to be receptive:

@Partisangirl: <<tweet is no longer available>>

@Partisangirl BREAKING!!! VITAL INFORMATION ON #SYRIA REGARDING #ChemicalAttacks #Douma <<link to YouTube video of UK Column News>>

Cross-network replies generally sought to correct or dismiss reporting, offering alternative versions of events as in the following examples to The Guardian and Boris Johnson (then UK Foreign Secretary).

@guardian: UK denounces Moscow claims country was behind 'staged' Syrian gas attack

@guardian There's a lot of info to suggest, even to a novice, that these are staged events <<link to YouTube video of Boris Johnson visiting a WH training session and pledging further monetary support for the group>

@BorisJohnson: Welcome the news of UK military strikes against major chemical weapons facilities in Syria alongside our US and French allies. The world is united in its disgust for any use of chemical weapons, but especially against civilians

@BorisJohnson ☐☐ There was no chemical attack at Douma It was a staged provocation by the White Helmets <<link to YouTube video Douma Chemical Attack False Flag Operation EXPOSED!>>

Out-network replies sought to delegitimize the WH in the eyes of prominent political figures such as Donald Trump, in effect lobbying them to not take retaliatory action against the Syrian government.

@realDonaldTrump: Many dead, including women and children, in mindless CHEMICAL attack in Syria. Area of atrocity is in lockdown and encircled by Syrian Army, making it completely inaccessible to outside world. President Putin, Russia and Iran are responsible for backing Animal Assad. Big price...

@realDonaldTrump You've been duped by the deepstate and their white helmet terrorist propagandists <<Link to CorbettReport video on YouTube titled *The White Helmets Are A Propaganda Construct*>>

A single account was responsible for the majority of replies sent during this period, resurfacing the 2016 video of Boris Johnson pledging funding for the WH (discussed above) and persistently posting it as a reply to other accounts 296 times between April 14 and May 12. The start of this effort coincided with US, UK, and French airstrikes against military targets in Syria, to highlight that the WH were funded by western governments and strategically frame this to mean that they were a military tool of these governments.

Aside from replies, other accounts also appeared ‘energized’ during this period and were actively and persistently introducing YouTube content to Twitter. Figure 5 illustrates the retweet cascades of YouTube content during this period. The vast majority of (re)tweets are from the anti-WH community, generally from accounts with smaller followings, but the content is retweeted by more influential accounts (with more followers). The largest cascade here is the Roger Waters video, receiving 1822 retweets, the majority of those over the course of a week. Overall, the vast resources of YouTube meant it was used as a source of both new and old content. While new content was being produced and introduced to Twitter in “real-time” during this period, content produced many months earlier was also being appropriated and recirculated in new context [72].

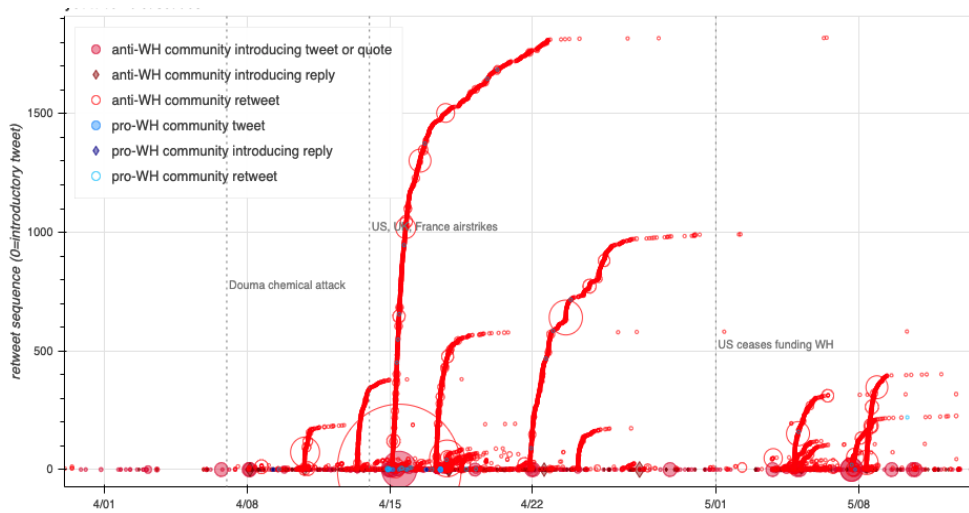


Fig 5: Retweet cascades of the top-25 most (re)tweeted YouTube videos between April 1 and May 15 2018.

4.4 Steemit

Steemit is a blockchain-based social media platform that supports community building and social interaction with cryptocurrency rewards [69]. Steemit is decentralized, existing on a

distributed peer-to-peer network, which eliminates control by a single entity (unlike e.g. Facebook). Steemit's terms of service posits "freedom of speech" as a foundational principle of civil liberties and calls out censorship as a means of limiting public discussion. It is free to join Steemit¹⁰, author posts, and upvote others' posts. Users can be rewarded in Steem (cryptocurrency) for contributions to the community, including writing an article that is upvoted (valued) by other members, or upvoting content that later becomes popular [70]. Steemit includes social share buttons to Facebook, Twitter, Reddit, and LinkedIn, plus a URL-sharing. An author must include at least one relevant tag on their post; the use of an irrelevant tag can lead to downvoting. Posts can be edited by the authoring user for 7 days, after which they are immutable. However, since all changes are logged on the public blockchain, even if deleted, traces of the content history will persist.

4.4.1 Mirroring Content on the Blockchain

Steemit is almost exclusively (re)tweeted by accounts in the red anti-WH community on Twitter, and all content within this sample promotes anti-WH narratives. There are five distinct content creators, although 19 of the articles are authored by a single user. Steemit is generally used to host compilations of content, with links to other Steemit posts, URLs to other external websites, screenshots of relevant tweets, and embedded videos. These posts offer a kind of clearing house of anti-WH content, with titles such as "massive compilation of material...", "huge cache of...", and "huge information drop".

Our analysis revealed the use of Steemit as a means of content preservation whereby a copy of content published elsewhere was also published on Steemit. This wrote the content to the blockchain and meant that in the event of its removal from another platform (e.g. due to a policy violation), it would still exist on the decentralized peer-to-peer network. Within our sample we found copies of two articles published on Medium and one video from YouTube that were 'mirrored' on Steemit. We also noted other activities on Steemit that would support efforts to preserve content: When embedding YouTube content from an alternative (non-mainstream) media source in a Steemit post, the video was often copied to D.Tube¹¹ and appeared alongside screenshots and commentary describing the video. Similarly, tweets were included as screenshot images alongside a text link to Twitter. Screenshots (as images) offer persistence, which would exist even if the source content was subsequently removed. This idea of working to preserve content — i.e. to build resilience into their information-sharing activities — is something that we will elaborate on further in our discussion.

4.4.2 Steemit-Twitter Content Integration

The majority of Steemit content is integrated to Twitter by a small number of highly active users, in particular the now-suspended *@WhiteHelmetsEXP* (382 tweets). These accounts undertake a concerted effort to attract the attention of large media organizations (e.g. *@BBCWorld* 24 replies; *@FoxNews* 24 replies) and prominent politicians (e.g. *@realDonaldTrump*

¹⁰ The Steemit FAQ reveals there is a cost to joining the Steem blockchain, but this cost is absorbed by Steemit for new signups. For this reason, users applying to join must supply their email address and phone number, which will be verified to prevent a user having multiple accounts. Some authors argue that the associated cost with participation can help with issues of authenticity online [30].

¹¹ D.Tube is short for Decentralized Tube, a decentralized open source video-sharing platform that also utilizes the Steem blockchain

19 replies; @NYCMayor 18 replies) by posting URLs to Steemit content via replies on Twitter. For example, in the reply is to Fox News (conservative cable broadcaster) the reply attempts to bring the WH “exposé” to one of their primetime presenters, Sean Hannity (also an ally of then President of the United States Donald Trump):

Tweet by @FoxNews: MONDAY: Watch @seanhannity at a new time, 9p ET, when he’ll interview Steve Bannon! #Hannityat9

Reply: @FoxNews @seanhannity White Helmets Fully Exposed as Hollywoods Favorite Terrorists-Over 200 Revealing Facebook Images <<link to Steemit post>>

We also observed accounts with large followings (re)tweeting Steemit content. For example, an anonymous account with ~100,000 followers that was—and remains in December 2020—highly active in the anti-WH discourse, posts 19 tweets that contain 10 of the 25 Steemit posts in the sample and retweets other Steemit content an additional 18 times that results in downstream retweets by other accounts in the red community.

4.5 Facebook

Facebook has grown from a social networking site connecting college friends to an influential company with billions of users [24]. Facebook offers a wide variety of functionality for its users, but important here is that it allows users to connect with others and to author content in the forms of posts (images and text), videos, and live videos. Within Facebook, this content can be shared ‘privately’ (with ‘friends’ within one’s social network) or publicly. Sharing to other platforms (e.g. Twitter) requires a user to ensure the post is shareable (i.e. the permissions are not limited to friends only) and copy the URL to a tweet. Facebook policies were somewhat in flux during 2017-2018, and content moderators were overwhelmed by the volume of posts [33]. After a Guardian investigation into the content moderation policies, in 2018, Facebook released its Community Standards [18], which outlined that although the platform values expression, they also recognize the potential for abuse. In particular, the policies focused on authenticity, safety (being a safe ‘place’), privacy, and extremist content. Although no explicit policy on mis- or dis- information was in place, Facebook said it wanted to talk about what they felt important, even if objectionable, but would consider the ‘newsworthiness’ and weigh the potential harms [18].

4.5.1 WH Content on Facebook

The sample of Facebook posts linked-to in the Twitter data is a mix of pro-WH and anti-WH content that take the form of Facebook posts (image and text) or Facebook Live (videos) produced by nine unique content producers. Seven of the 25 posts are no longer accessible, the highest proportion across the five platforms; the redirect page for these posts suggests that four of the posts were removed by the user or they changed the permissions and three are “no longer available”. Six of the seven removed posts received the majority of (re)tweets from accounts in the anti-WH community on Twitter, suggesting they contained anti-WH messaging.

Pro-WH Facebook posts were authored by the official Facebook account of the WH (4 posts of the group involved in local civil service) and the Tipperary Peace Convention (1 post announcing that the WH had been awarded the 2018 Tipperary Peace Prize¹²). Anti-WH content comes from a variety of sources including the *Russian Embassy in the USA*, Russian state-backed broadcaster RT, and a journalist with Syrian state TV al-Sama TV. The most active content producer in this sample is a Bolivian celebrity and philanthropist who posts five Facebook Live videos, on location in Syria, in which she is visiting various buildings that are purportedly occupied by the WH with claims that they are connected (physically, by underground passages) to terrorist groups affiliated with a known terrorist group Al-Nusra. In another videos, this individual interviews and discusses her findings with Pearson Sharpe, a journalist with One America News Network (OANN), a far-right, pro-Trump, conservative news network recently involved in spreading medical (covid-19) and election-related misinformation [11,21].

4.5.2 Facebook-Twitter Content Integration

Accounts in the pro-WH community on Twitter involved in content introduction and retweeting of Facebook content have, in general, relatively large followings. As illustrated below, some pro-WH content seemed to benefit—in terms of retweeting—from coordinated boosting by activists and possibly even automated accounts. Three of the pro-WH posts from the official WH Facebook account were introduced to Twitter through cross-posting by the respective *@SyriaCivilDefense* Twitter account. They were then retweeted by accounts with large followings, for example *@Free_Media_Hub*, an activist account with 225K followers (and a second account called *@Free_Media_Hub2*), and *@9b30R2gdITMbDcd* a Japanese account with (at the time) a large following of 1.18M. Interestingly, in December 2020 *@9b30R2gdITMbDcd* still exists on Twitter but the number of followers is down to 242K, suggesting that the majority of its followers were part of coordinated inauthentic behavior and were subsequently removed by Twitter. One account in the pro-WH community posts a reply to themselves, in which they translate an earlier tweet of a Facebook Live video of a WH volunteer. This is retweeted by some other accounts with modest followings, before a retweet by *@9b30R2gdITMbDcd*. This tweet ultimately receives 43 retweets. In general, there are less content introductions by the pro-WH community, and only a single reply. However, when content is introduced to Twitter, the retweet cascade quickly builds as accounts with large followings work together to promote it.

¹² The Tipperary International Peace Award was created by residents of the Irish town to counter the association between their town and war, which was created by the song “It’s a long way to Tipperary”. The annual award is given for humanitarian work.

On the anti-WH side, the timing of FB content appears to coincide with current events on the ground, for example posts in response to the reappearance of Omran Daqneesh (The ‘Aleppo Boy’) and when the WH were awarded the Tipperary International Peace Award. Despite being positive news for the WH, introduction of the Tipperary International Peace Award Facebook posts to Twitter is done by the anti-WH community who frame the award and support for the WH as a “stain on Ireland”. The Bolivian celebrity (with ~85K followers) introduces her own Facebook content to Twitter (5 introductory tweets). The anti-WH community uses a combination of in-network, cross-network, and out-network replies to draw attention to the content and mobilize it across platforms, from Facebook to Twitter. For example, one user sends 45 replies containing a URL to an anti-WH Facebook post to different Twitter posts over a 20 minute period on April 13 2018.

4.6 Medium

Medium is an online publishing platform that encourages users to express their ideas in the form of written posts. Medium posts can include embedded images, videos and content from other platforms such as Twitter. All posts include social share buttons to Facebook, Twitter, LinkedIn, or a shareable URL. Each post can have up to five tags that serve as a means of grouping content; clicking tags reveals similarly tagged posts. Medium provides users a comprehensive set of rules in its policy documents. Explicitly disallowed are hate speech, harassment, violations of privacy, deceptive or inauthentic content, graphic content, and pornography. Medium also singles out the promotion of harmful conspiracies, which includes pseudo-scientific claims, conspiracy theories with an associated history of harassment, hate, or violence, and intentional distortions, in particular systematic false claims about historic events and facts. Reported violations are investigated by the company on a case-by-case basis according to a risk analysis [52]

4.6.1 WH Content on Medium and Its Introduction to Twitter

The sample of Medium posts presents a mix of both pro-WH (15) and anti-WH (8) articles and is the only platform (in this sample) that is predominately (re)tweeted by the blue pro-WH community on Twitter. The Medium posts are generally longer articles, including a blog post of an academic paper¹³. One post has been removed by Medium for being “*in violation of the Medium rules*”, and one article has since been deleted by the author.

The pro-WH content can broadly be categorized as either research-based articles that describe the disinformation campaign against the WH or as posts that celebrate the local volunteer efforts of the WH as they organize events for children and clean schools. Two of the articles in support of the WH come in the form of obituaries to the seven WH members that were murdered in Idlib in August 2017. Several of these articles are authored by an account called *The Syria Campaign*, a group that describes itself as supporting the WH through advocacy and fundraising. On the anti-WH side, seven (out of the 8) articles are authored by a single author. These are also longer posts that call into question the role and function of the WH, both alleging ties to terrorist groups in the region and at the same time calling the group out as part of a western “war machine”. These articles are also highly critical of the western

¹³ We were also contributing authors to this published article [65]

“mainstream media”, particularly journalists, such as Olivia Solon at the Guardian who wrote an article about the disinformation campaign against the WH [63]. In general, there is a criticism of what are termed “establishment” narratives and how those that challenge these narratives are labelled as Russian trolls or “useful idiots”. In an example of cross-posting beyond Twitter—and building resilience through mirroring content on other platforms—two of this author’s posts also appear in our sample of Steemit posts. In general, due to the Medium interface and comprehensive community guidelines, Medium was generally a platform used for longer posts and opinion pieces.

Medium content introductions to Twitter are done by pro- and anti- WH communities. There are 292 introductory tweets containing a Medium URL from our sample, 201 from pro-WH accounts and 91 from anti-WH accounts. The majority of these introductions are tweets, not replies that attempt to gain the attention of large media or politicians.

4.7 LiveLeak

LiveLeak (“*redefining the media*”) is a video-sharing platform founded in 2006, emerging from a place of skepticism with the news media [12] and a desire to host real footage of real world events such as war. In 2019, the platform attracted between 16-20 million visitors per month [31]. LiveLeak describes itself as “*as free as possible*” [42], though warns users that graphic media may be removed “*on a case by case basis*”; it also prohibits use of its services by terrorist or banned organizations [43]. However, the platform has a relatively relaxed policy on what can and cannot be uploaded, and videos often contain violence [12]. LiveLeak hosted footage of Saddam Hussein’s execution and the beheading video of American journalist James Foley [71]. Its decision to ban future ISIS beheading videos and refusal to host the 2019 Christchurch shooting led to complaints from some users [31]. In 2014, LiveLeak and Ruptly (the video news agency of RT, the Russian state-funded media corporation) announced a video distribution partnership [60].

LiveLeak features a search function, and videos can be tagged to facilitate searches (although clicking tags does not initiate a search for related content but must be manually typed into the search bar). The platform does not have social share buttons: instead a “share link” must be copy-pasted elsewhere, which adds additional effort when sharing content. On Twitter LiveLeak videos appear as a text URL that when clicked redirects to the content on LiveLeak.

4.7.1 WH Content On Liveleak and Its Introduction to Twitter

23 of the LiveLeak URLs were to video content, one to the LiveLeak homepage, and one to a user’s page (the most active content creator in this sample). The 23 videos were produced by two distinct users (producing 1 and 22 videos). As of December 1 2020, only one of the videos in the sample is unavailable, due to “*a possible violation of our terms of service*”. This is consistent with a platform that exhibits relaxed policies and user base that accepts violent or gory content [31].

Aside from two notable exceptions (a video produced by a European news agency, and a compilation titled “*Tapestry of Terror*” that we discuss in detail in the next section), the production quality of videos in this sample is generally low: raw footage, captured from cell phones on the ground. All LiveLeak videos in this sample promote anti-WH messages, mainly presenting the contested narrative that the WH are aligned with terrorist groups such as Al-Nusra and Al-Qaeda in Syria. Three videos purport to show the WH involved in executions and

beheadings. In effect, LiveLeak appears to be the venue for hosting content that would otherwise fall foul of community guidelines on YouTube due to depictions of violence.

Introduction of LiveLeak content to Twitter is conducted by a small set of 37 determined users that post 470 tweets. 373 of the tweets are by a single user. In the majority of cases, the content is introduced as replies to cross-network or out-network accounts of large media organizations or politicians, including *@guardian* (77), *@cnni* (28), and *@realDonaldTrump* (8). This once again speaks to the persistent effort, by a subset of users, to lobby politicians or set the media agenda from via Twitter.

4.8 Multi-platform Concurrent Introduction of an Evidence Collage

The *Tapestry of Terror* (ToT) is a 65-minute-long video compilation that presents multiple strategic narratives about the WH, but for the most part tries to reframe the WH as aligned with regional terrorist organizations. The production is a remix of video clips, images, and screenshots of social media and websites (some of which appear in the 125 posts from across 5 social media platforms analyzed in this paper). These media resources are compiled into this production, tied together with narration that signposts the various anti-WH narratives. ToT also implicates the western mainstream media as complicit in the “conspiracy” of the WH. ToT can be described as an *evidence collage*—files that aggregate evidence [36]. In our data, ToT appears on YouTube, LiveLeak and Steemit (Table 5).

Tapestry of Terror video		Tweets (introduction to Twitter through URL)			Retweets (content dissemination)		
platform	upload/post date	num tweets	num unique accounts	avg tweet per user	num retweets	num unique accounts	avg retweet per user
YouTube	2017-03-25	207	26	8	153	141	1.1
LiveLeak	2017-03-26	374	10	37.4	442	345	1.3
Steemit	2017-12-04	41	5	8.2	26	23	1.1

Table 5: ToT appeared on 3 social media platforms in this study: YouTube, LiveLeak, and Steemit. This table breaks down the tweets and retweets of each version.

ToT was uploaded to YouTube and LiveLeak at the end of March 2017. The YouTube video description includes a call to action for others to download, reupload, and share and to prevent “*the truth be[ing] hidden and censored*”. The Steemit version of ToT was uploaded in December 2017 and appears as part of a longer post, itself an evidence collage, including screenshots from the video and embedded frames of both the YouTube and LiveLeak versions of ToT. The LiveLeak one is labelled as the “*mirrored/shared version*”. The multiple versions helped reach different communities (on LiveLeak, Youtube, and Steemit), but also served as backups, particularly in the case of the Steemit version, which would have been written to the blockchain and is therefore persistent.

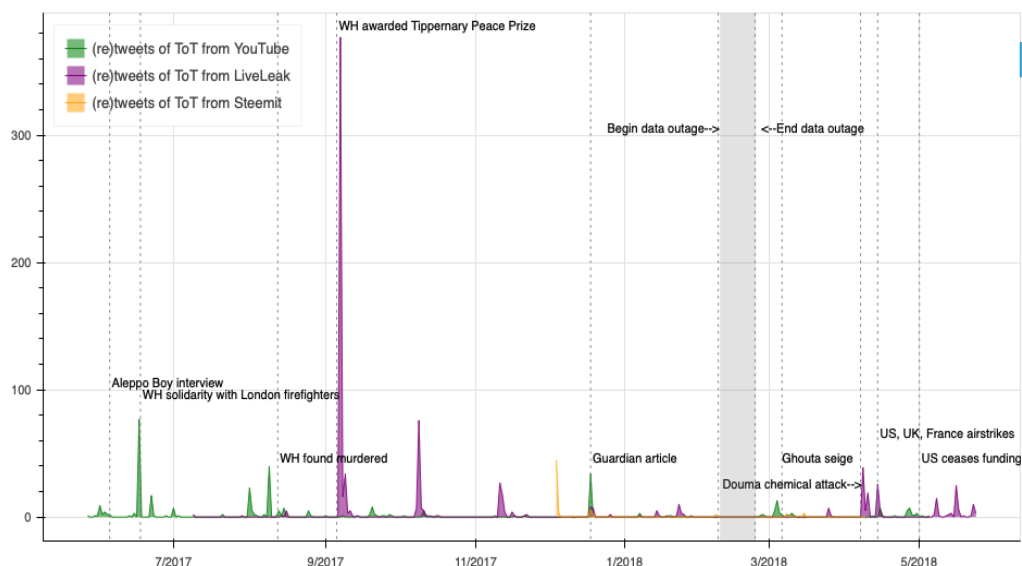


Figure 6: A temporal plot showing the number of (re)tweets, per day, for each YouTube, LiveLeak, and Steemit versions of *Tapestry of Terror*. This illustrates when the video was recirculated on Twitter.

The YouTube version of ToT is the first introduced to the Twitter conversation (Fig 6), but the LiveLeak version led to spikes in the volume of tweets. The Steemit version, uploaded several months later, only features in 67 (re)tweets, so is not visible in the plot. The LiveLeak version of ToT is introduced to the Twitter conversation 374 times by 10 distinct accounts. Concurrently, the YouTube version is introduced 205 times by 26 distinct accounts. In other words, the work of mobilizing ToT and integration into the Twitter conversation is fulfilled by a small number of dedicated users who persistently work (for between 2 and 10 months) to introduce this content to Twitter

Examining this work in more detail, the effort to introduce the LiveLeak version of ToT to Twitter was predominantly divided between two Twitter accounts: a now suspended account that was dedicated to attacking the White Helmets and another active account in this conversation that seek to bring attention to the video by replying to the tweets of major politicians (e.g. @realDonaldTrump, @jeremycorbyn) and large media organizations (e.g. @Guardian, @DerSPIEGEL, @cnni, @IrishTimes). Oftentimes, these replies were in response to tweets about unrelated matters:

@Guardian: Russian women win partial victory in Aeroflot gender discrimination case
<<link to this article on the Guardian>>

@guardian Tapestry of Terror (Highly Graphic) - White Helmets Exposed As FSA Terrorists Linked With ISIS <<shortened Link to ToT on LiveLeak>>

@RealDonaldTrump: We want our companies to hire & grow in AMERICA, to raise wages for AMERICAN workers, & to help rebuild our AMERICAN cities & towns! #USA <<USA flag emoji>>
<<image of Trump speaking outside a factory with message about raising wages for American workers>>

@realDonaldTrump Tapestry of Terror (Highly Graphic) - White Helmets Exposed As FSA Terrorists Linked With ISIS <<shortened link to ToT on LiveLeak>>

In other cases, the replies were targeted and specific, for example when CNN journalist Jake Tapper recommended people to follow the account of Bana al-Abed, a Syrian girl from Aleppo who (with the help of her mother) documented the siege of the city via Twitter. In the reply to this tweet, this is called out as “fake news” with Tapper’s credentials as a journalist challenged:

@jaketapper: Follow @AlabedBana <<link to Twitter account>>

@jaketapper @AlabedBana So disappointing if I thought u were a REAL journalist. It's #FakeNews. WATCH ↓↓↓↓↓↓↓↓↓ <<shortened link to ToT on LiveLeak>>

In total, 93% of all non-retweets containing the ToT LiveLeak URL are replies. A similar pattern of behavior exists around the YouTube version of ToT—80% of tweets are replies, predominately to major news outlets and politicians, although the replies are sent by a different group of Twitter users.

Further examination of the temporal activity reveals an interesting pattern—ToT is introduced to Twitter in response to events that present the WH in a positive light. For instance, in September 2017 the WH were awarded the Tipperary Peace Prize and in response we see the largest spike of (re)tweets of the LiveLeak version of ToT. Similar (but smaller) spikes occur in December 2017, in response to the Guardian article that reported on the disinformation campaign being waged against the WH [63]; and January 2018 when the Netflix documentary about the WH was awarded an Oscar [38]. In such cases the re-introduction of ToT was meant to distract from the positive media coverage of the WH and replace it with the contested alternative narratives presented within the video. These examples highlight YouTube’s role as a repository of content that can be referenced over time [72], and the reintroduction of ToT serves to sustain the anti-WH narratives, in particular by reintroducing them into the Twitter discourse at strategic times to distract from otherwise positive reporting of the group

5 DISCUSSION

The empirical findings presented in this research provide insight into the established (and largely successful) cross-platform information operation conducted against the WH in Syria. Our work contributes to research of multi-platform information operations by extending our initial Twitter dataset, using the URLs in tweets to build a picture of the surrounding ecosystem of social media platforms from which content is produced, stored, and integrated into the Twitter conversation. While imperfect insofar as it provides a “Twitter-centric” perspective of the wider information ecosystem, it does offer insight into how the two sides of the WH conversation make use of different configurations of social media platforms, and the ways in which the information subsequently flows into the Twitter conversation, where it can potentially be noticed by major politicians or journalists [41].

Our research leveraged the digital traces left by online interactions, following both the content and the actors involved in its production and dissemination. We adopted a CSCW perspective, meaning that we focused on the collaborative activities of these actors on and across multiple platforms, enabling us to surface the “work” involved in conducting multi-platform information operations including sustaining the narratives and attempting to set the

agenda, oftentimes to distract or counter positive reporting about the WH; and resilience-building through the use of alt-tech to counter perceived threats of censorship. We elaborate on each of these forms of work in the following sections.

5.1 Sustaining Narratives and Working to Set the agenda.

The sustaining of narratives—keeping them relevant and maintaining attention of a mainstream audience—is achieved through three forms of work: *content production*, *content synthesis*, and *content mobilization*.

On all platforms in this study we saw a small number of dedicated accounts involved in ***content production***—producing original content that supported one or more of the WH narratives in circulation by covering a new story. While the content in our sample contained both pro- and anti- WH narratives, anti-WH content was far more visible in the trace data. Content came in a variety of forms (posts, images, articles, videos), and varied in production quality. *Content production* involved authoring content; uploading it to one or more platforms (i.e. storing it); and labelling it by adding titles, tags, and descriptions (facilitating its retrieval in the future). Content production generated *communal goods* [4]—resources for the community that could be searched for, selected, and synthesized into new content.

Content synthesis is the process of synthesizing existing content resources into *evidence collages* [36]—compilations or aggregates of evidence in the form of ‘new’ (remixed) content. For example, we observed the appropriation of the WH’s “*Mannequin Challenge*” video as “evidence” that they staged rescues; the assembling of multiple URLs into Steemit posts that could serve as a clearing house of anti-WH content; and the compilation of video and image “evidence” into feature-length videos such as ToT. A particular strength of evidence collages is their flexibility—because they compile a variety of evidence that can be synthesized to support or challenge a range of narratives they are particularly useful *communal goods* that can be leveraged by other users.

Content mobilization is the activity of searching for and selecting content and trying to amplify it, or gain the attention of specific audiences, at specific times, in particular by moving content across platforms. The Twitter-centric nature of this case study generally meant that we could see content mobilization taking place by Twitter accounts, which would curate content from the other platforms and opportunistically introduce it to the Twitter conversation where it could potentially be broadcast and received by a mainstream audience, including politicians, news organizations, and journalists. A specific form of *content mobilization*, particularly salient in the anti-WH community, was to surface (search for, select) content and then publicly mobilize it *at* specific Twitter users through use of the *reply* function. This form of mobilization was often targeted at politicians and journalists to get their (and their audience’s) attention. These deliberate, and often persistent efforts to introduce content from the fringes to the mainstream are known as *trading up the chain* [36,49], which aims to build up the credibility of the message and appeal to a wider audience [14,32].

There was a distinct temporal element to this work: The anti-WH community on Twitter would work to reintroduce existing content into the Twitter discourse when the time was ‘right’, in what Chadwick [9] termed *opportunity structures* (p.64). In other words, while beneficial to publish something at the moment it occurs, there can also be an advantage in reintegrating dormant “fragments of stories” (p.64) at specific times when they will have maximized impact. In this case study, the ‘right’ time most often coincided with otherwise positive reporting of the WH. At those times (e.g. amid reports that seven WH were murdered;

when the WH were awarded the Tipperary Prize), we observed reactive and distracting dissemination practices conducted by what could be considered *episodic volunteers* [67,76]—subgroups of accounts working in alignment with the goals of the community, leveraging *opportunity structures* to take advantage of real-world events. Sometimes this was achieved through the use of influential accounts that stepped up to quickly propagate new content to shift the narrative, while other times a single or small group of accounts worked intensely—with the lightweight help of an audience who assisted in disseminating the content—over a period of hours or days to persistently reintroduce content that would challenge or distract from the positive reporting of the WH.

5.2 Resilience-building Using Alternative Infrastructure

Social media platforms are sociotechnical, comprising technological features, users, and content moderation policies, which collectively form a communication infrastructure that can support online movements [3]. Since each social media platform is a unique combination of these sociotechnical aspects, they can be particularly suitable to serve a specific role or purpose within the larger information ecosystem.

A distinct advantage of our document-driven approach (i.e. following the content through digital traces) was that it enabled us to build up a picture of the distributed ecosystem of infrastructure through which information flowed. By reconstructing the digital traces, we were able to look beyond a single platform (Twitter) to the wider information ecosystem. This allowed us to see the work of resilience-building—working toward stability against the perceived threats of “censorship” from “big tech” (a reference to the large technology companies that own the platforms such as Google, Facebook, Twitter, and Medium) [15].

The work of resilience-building was more visible on the anti-WH side of the conversation, in part due to their over-representation in our data (more active on Twitter and (re)tweeted more URLs), but also because there was explicit evidence of this strategy (and the necessity of it) in posts, articles, and videos in our sample. Resilience-building was broadly concerned with preserving content through the use of alternative platforms or *alt-tech*—a term used to describe the “alt-right”’s shift to the use of alternative (non-mainstream) platforms because of perceived threats to their stabilization as a result of efforts to deplatform and censor the movement [15]. In a similar way to several other online activist movements, including the “alt-right” [15], the anti-WH side of the discourse demonstrated a disdain for the mainstream media and the evolving content moderation policies of the established social media platforms.

In striving for resilience, the anti-WH voices sought out alternative platforms to circumnavigate policy and carve out direct distribution channels. Working toward resilience involved the strategic placement of content on a specific platform or by mirroring across multiple platforms. An artifact of this activity is visible in our data—the anti-WH community on Twitter shares more content from a greater variety of platforms that it uses as a means of storing and organizing their content. The alternative platforms of choice were LiveLeak, with its relaxed policies, editorial partnership with Russian state-funded media (Russian media were prominent members of the anti-WH discourse), and user base that has an affinity for violent content; and Steemit, the blockchain-based platform built on an immutable decentralized peer-to-peer network. These platforms provided the anti-WH voices with infrastructure that was able to protect the availability and accessibility of their content in the long-term: technologically, both platforms inherently enabled the hosting of content (with Steemit offering the additional advantage of decentralized storage not overseen by a single entity); and from a

policy perspective, both platforms claimed to oppose censorship (an implicit suggestion that other “mainstream” platforms such as Twitter, Facebook, and YouTube are engaging in censoring content). It is important to note however, that censorship describes the suppression of free speech backed up by law (i.e. conducted by a government), and here that term is being conflated with *content moderation*, which are the community guidelines set out by social media companies as private entities that lay out what is and is not appropriate [77]. In our data, LiveLeak and Steemit were used exclusively to store anti-WH content, and, in the vast majority of cases, were linked to in (re)tweets posted by accounts in the anti-WH community on Twitter. As this community worked ‘outside’ of the mainstream—and its established norms and technological solutions—preserving content was important in protecting the longevity of the operation.

But despite technologically-enabled content storage and retrieval and relatively forgiving policies, an infrastructural limitation of these alternative platforms is their lack of mainstream appeal. Steemit is among the largest blockchain-based social media platforms [30], and LiveLeak enjoyed 16-20M users per month in 2019 [31], but this is a fraction of Twitter’s 330M monthly active user base that includes major politicians and journalists. And while YouTube has 2B users, it is technologically suited to content storage and retrieval, built around recommendation algorithms that seek to keep users engaged on the platform, rather than providing a large, mainstream audience with a broadcast breaking news feed of what is *happening now*. So while these platforms may be suited to storage and retrieval, we observed the persistent work of *content mobilization* as attempts are made to integrate anti-WH content to Twitter, to reach a broader audience and sustain the anti-WH narratives in the long-term.

5.3 The Uncomfortable Truth For Social Media Platforms

Information operations and the production and dissemination of strategic narratives on and across social media platforms put the “mainstream” social media platforms (e.g. Twitter, Facebook, and YouTube, etc) in a difficult position with regards to policy and content moderation. The participatory information operation surrounding the WH brings to the forefront the fuzzy distinction between activism and active measures (Soviet parlance for “dirty tricks” and disinformation), as noted on multiple occasions by Thomas Rid, in his book charting a century of disinformation and political warfare [58]. The core social media accounts in this research are managed by genuine (as in real) people, and their activities, although persistent, do not appear automated or explicitly coordinated as was the case with other documented information operations such as the Russian government’s Internet Research Agency [13,23]. As such, the activities cannot be characterized as *coordinated inauthentic behavior*, a common signal used for remediation by the large social media platforms [19,73,79]. Further complicating matters is that, in general, the social media platforms do not have broad policies regarding mis- and dis- information: Section 230 of the Communications Decency Act (CDA) of 1996 offers them some legal protections from what their users may say and do on their infrastructure [17,75], while their leaders have previously stated they have no desire to be the “arbiters of truth” [51].

During 2020 platforms proactively introduced policies to facilitate content moderation in specific contexts, including the coronavirus pandemic and US election-related disinformation and misinformation. By design, these policies are limited in scope to these specific contexts that can be justified by the associated demonstrable offline harms (i.e. the hindering of global efforts to fight a global health crisis; voter suppression). But the development of these expanded

policies is also limited to the “mainstream” platforms: platforms such as Steemit and LiveLeak have no such policies, and the 2020 US election cycle led to the emergence of platforms such as Parler, a pro-Trump platform with minimalist community guidelines that was involved in amplifying post-election disinformation that culminated in the breaching of the US capitol in January 2021 [59].

On the one hand, without a collective effort—meaning consistent policies across platforms, such as those in place for child sexual abuse material, terrorism-related content, and more recently coordinated inauthentic behavior [16]—it remains difficult to moderate against mis- and disinformation because, as demonstrated in this research, information operations are participatory—accounts are not necessarily explicitly coordinated, and content is also produced, synthesized, and hosted on other (non-mainstream) platforms and mobilized from there. Unified agreements across platforms could make it easier to remove problematic content at its source or to prevent it moving across the information ecosystem.

On the other hand, increased cooperation between platforms to develop unified policy could lead to what have been described as *content cartels*—“*arrangements between platforms to work together to remove content or actors from their services without adequate oversight*” [16]. A unified policy is likely to favor the larger platforms that have more users, meaning that the policy for many will likely be shaped by a few. Furthermore, such efforts, particularly if enacted with a lack of transparency (in terms of development and subsequent application), will reinforce the view that platform content moderation is equal to “censorship” and drive content (and its production) elsewhere—to the “*alt-tech*” platforms where policies weaker, non-existent, or not enforced. In effect, it would just push the problematic content to the unmoderated fringes of the internet and, unless the mainstream platforms begin to deny list whole platforms, the content can and will make it back to the mainstream.

REFERENCES

- [1] Ahmer Arif, Leo G Stewart, and Kate Starbird. 2018. Acting the part: Examining information operations within #BlackLivesMatter discourse. *Proceedings of the ACM on Human-Computer Interaction* 2, November: 1:26. <http://doi.org/10.1145/3274289>
- [2] BBC News. 2018. Syria war: At least 70 killed in suspected chemical attack in Douma. *BBC News*. Retrieved January 11, 2021 from <https://www.bbc.com/news/world-middle-east-43686157>
- [3] Berkman Klein Center. 2020. Joan Donovan on domestic misinformation. *The Breakdown*. Retrieved January 12, 2021 from <https://medium.com/berkman-klein-center/the-breakdown-joan-donovan-on-domestic-misinformation-da4f60296420>
- [4] Bruce Bimber, Andrew J Flanagan, and Cynthia Stohl. 2005. Reconceptualizing collective action in the contemporary media environment. *Communication Theory* 15, 4: 365–388.
- [5] Vincent D. Blondel, Jean Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre. 2008. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*. <http://doi.org/10.1088/1742-5468/2008/10/P10008>
- [6] Danah Boyd, Scott Golder, and Gilad Lotan. 2010. Tweet, tweet, retweet: Conversational aspects of retweeting on twitter. In *Proceedings of the Annual Hawaii International Conference on System Sciences*. <http://doi.org/10.1109/HICSS.2010.412>
- [7] Jean Burgess and Joshua Green. 2018. *YouTube: Online video and participatory culture*. John Wiley & Sons.
- [8] Meeyoung Cha, Hamed Haddadi, Fabricio Benevenuto, and Krishna P. Gummadi. 2010. Measuring user influence in twitter: The million follower fallacy. In *ICWSM 2010 - Proceedings of the 4th International AAAI Conference on Weblogs and Social Media*.
- [9] Andrew Chadwick. 2017. *The hybrid media system: Politics and power*. Oxford University Press.
- [10] Martin Chulov. 2020. How Syria’s disinformation wars destroyed the co-founder of the White Helmets | News | The Guardian. *The Guardian*. Retrieved January 11, 2021 from <https://www.theguardian.com/news/2020/oct/27/syria-disinformation-war-white-helmets-mayday-rescue-james-le-mesurier>
- [11] CNN Business. 2020. Newsmax and OANN are telling lies about the election as more people tune in. *CNN Business*. Retrieved January 11, 2021 from <https://www.cnn.com/videos/business/2020/11/23/newsmax-oan-trump->

- conspiracy-theories-ratings-orig-vf.cnnbusiness
- [12] James Cook. 2014. Profile Of Hayden Hewitt, Founder Of LiveLeak - Business Insider. *Business Insider*. Retrieved January 11, 2021 from <https://www.businessinsider.com/profile-of-hayden-hewitt-founder-of-liveleak-2014-10>
 - [13] Renee DiResta, Kris Shaffer, Becky Ruppel, et al. 2019. The Tactics & Tropes of the Internet Research Agency. *U.S. Senate Documents*. Retrieved January 11, 2021 from <https://digitalcommons.unl.edu/senatedocs/2>
 - [14] Joan Donovan. 2019. How memes got weaponized: A short history | Berkman Klein Center. *MIT Technology Review*. Retrieved January 11, 2021 from <https://cyber.harvard.edu/story/2019-10/how-memes-got-weaponized-short-history>
 - [15] Joan Donovan, Becca Lewis, and Brian Friedberg. Parallel Ports: Sociotechnical Change from the Alt-Right to Alt-Tech. <http://doi.org/10.14361/9783839446706-004>
 - [16] Evelyn Douek. 2020. The Rise of Content Cartels. *Knight First Amendment Institute*. Retrieved January 11, 2021 from <https://knightcolumbia.org/content/the-rise-of-content-cartels>
 - [17] Electronic Frontier Foundation. Section 230 of the Communications Decency Act | Electronic Frontier Foundation. Retrieved January 11, 2021 from <https://www.eff.org/issues/cda230>
 - [18] Facebook. Community Standards | Introduction. Retrieved January 12, 2021 from <https://www.facebook.com/communitystandards/introduction/>
 - [19] Facebook. 2020. Community Standards | Facebook. Retrieved January 11, 2021 from https://www.facebook.com/communitystandards/inauthentic_behavior/
 - [20] Kirsten A. Foot and Steven M. Schneider. 2019. *Web Campaigning*. <http://doi.org/10.7551/mitpress/7186.001.0001>
 - [21] Jordan Freiman. 2020. YouTube suspends One America News Network channel for violating COVID-19 misinformation policy. *CBS News*. Retrieved January 11, 2021 from <https://www.cbsnews.com/news/oan-suspended-youtube-covid-19-misinformation/>
 - [22] Brian Friedberg. 2020. *Investigative Digital Ethnography: Methods for Environmental Modeling*. Retrieved January 11, 2021 from <https://mediamanipulation.org/research/investigative-digital-ethnography-methods-environmental-modeling>
 - [23] Vijaya Gadde and Yoel Roth. 2018. Enabling further research of information operations on Twitter. *Twitter Blog*. Retrieved January 11, 2021 from https://blog.twitter.com/en_us/topics/company/2018/enabling-further-research-of-information-operations-on-twitter.html
 - [24] Chaim Gartenberg. 2019. What is Facebook? Just ask Mark Zuckerberg - The Verge. *The Verge*. Retrieved January 12, 2021 from <https://www.theverge.com/2019/3/8/18255269/facebook-mark-zuckerberg-definition-social-media-network-sharing-privacy>
 - [25] Stuart Geiger and David Ribes. Trace ethnography: Following coordination through documentary practices.
 - [26] Tarleton Gillespie. 2010. The politics of 'platforms.' *New media & society* 12, 3: 347–364.
 - [27] Kathy Gilsinan. 2015. The Confused Person's Guide to the Syrian Civil War. *The Atlantic*. Retrieved April 15, 2018 from <https://www.theatlantic.com/international/archive/2015/10/syrian-civil-war-guide-isis/410746/>
 - [28] Janine di Giovanni. 2018. Why Assad and Russia Target the White Helmets. *The New York Review of Books*. Retrieved January 11, 2021 from <https://www.nybooks.com/daily/2018/10/16/why-assad-and-russia-target-the-white-helmets/>
 - [29] Emma Grey Ellis. 2017. INSIDE THE CONSPIRACY THEORY THAT TURNED SYRIA'S FIRST RESPONDERS INTO TERRORISTS. *Wired*. Retrieved from <https://www.wired.com/2017/04/white-helmets-conspiracy-theory/>
 - [30] Barbara Guidi. 2020. When Blockchain meets Online Social Networks. *Pervasive and Mobile Computing*. <http://doi.org/10.1016/j.pmcj.2020.101131>
 - [31] John Herrman. 2014. The Internet's Endless Appetite for Death Video - The New York Times. *New York Times*. Retrieved January 11, 2021 from <https://www.nytimes.com/2019/03/24/style/really-bad-stuff.html>
 - [32] Ryan Holiday. 2013. *Trust me, I'm Lying: Confessions of a Media Manipulator*. Penguin.
 - [33] Nick Hopkins. 2017. Revealed: Facebook's internal rulebook on sex, terrorism and violence | News | The Guardian. *The Guardian*. Retrieved January 12, 2021 from <https://www.theguardian.com/news/2017/may/21/revealed-facebook-internal-rulebook-sex-terrorism-violence>
 - [34] Josh Horwitz. 2018. YouTube is turning to Wikipedia to help it fight conspiracy theories — Quartz. *Quartz*. Retrieved January 11, 2021 from <https://qz.com/1228635/youtube-is-turning-to-wikipedia-to-help-it-fight-conspiracy-theories/>
 - [35] Amanda Lee Hughes, Kate Starbird, Alex Leavitt, Brian Keegan, and Bryan Semaan. 2016. Information movement across social media platforms during crisis events. *CHI'16 Extended Abstracts* 1.
 - [36] P. M. Krafft and Joan Donovan. 2020. Disinformation by Design: The Use of Evidence Collages and Platform Filtering in a Media Manipulation Campaign. *Political Communication*. <http://doi.org/10.1080/10584609.2019.1686094>
 - [38] Alexandra Larkin and Sophie Lewis. 2017. Netflix documentary "The White Helmets" wins Oscar. *CNN*. Retrieved January 11, 2021 from <https://www.cnn.com/2017/02/26/us/white-helmets-oscar/index.html>
 - [39] Alexandra Larkin and Sophie Lewis. 2017. Netflix's "The White Helmets" takes home the company's first Oscar win. *CNN*. Retrieved from <https://www.cnn.com/2017/02/26/us/white-helmets-oscar/index.html>
 - [40] Matthew Levinger. 2018. MASTER NARRATIVES OF DISINFORMATION CAMPAIGNS. *Journal of International Affairs*.
 - [41] Mark Lieberman. 2020. A growing group of journalists has cut back on Twitter, or abandoned it entirely - Poynter. *Poynter*. Retrieved January 11, 2021 from <https://www.poynter.org/reporting-editing/2020/a-growing-group-of->

- journalists-has-cut-back-on-twitter-or-abandoned-it-entirely/
- [42] LiveLeak. LiveLeak.com - Rules. Retrieved January 11, 2021 from <https://www.liveleak.com/rules>
 - [43] LiveLeak. LiveLeak.com - FAQ. Retrieved January 11, 2021 from <https://www.liveleak.com/faq>
 - [44] Louisa Loveluck. 2019. James Le Mesurier, backer of Syria's White Helmets, found dead in Istanbul. *Washington Post*. Retrieved January 11, 2021 from https://www.washingtonpost.com/world/james-le-mesurier-backer-of-syrias-white-helmets-found-dead-in-istanbul/2019/11/11/e75da222-04ab-11ea-a5e2-fccc16fa3576_story.html
 - [45] Louisa Loveluck and Eric Cunningham. 2018. Dozens killed in apparent chemical weapons attack on civilians in Syria, rescue workers say. *Washington Post*. Retrieved January 11, 2021 from https://www.washingtonpost.com/world/middle_east/dozens-killed-in-apparent-chemical-weapons-attack-on-civilians-in-eastern-ghouta--rescue-workers/2018/04/08/231bba18-3ac0-11e8-af3c-2123715f78df_story.html
 - [46] Jim Maddock. Using Historical Twitter Data for Research: Ethical Challenges of Tweet Deletions.
 - [47] Jim Maddock, Kate Starbird, Haneen Al-Hassani, Daniel E. Sandoval, Mania Orand, and Robert M. Mason. 2015. Characterizing online rumor-mongering behavior using multi-dimensional signatures. In *CSCW 2015 - Proceedings of the 2015 ACM International Conference on Computer-Supported Cooperative Work and Social Computing*. <http://doi.org/10.1145/2675133.2675280>
 - [48] Noortje Marres and Richard Rogers. 2005. Recipe for Tracing the Fate of Issues and their Publics on the Web. In *Making Things Public: Atmospheres of Democracy*.
 - [49] Alice Marwick and Rebecca Lewis. 2017. Media manipulation and disinformation online. *Data & Society Research Institute*.
 - [50] Louise Matsakis. 2018. YouTube Will Link Directly to Wikipedia to Fight Conspiracy Theories | WIRED. *Wired*. Retrieved January 11, 2021 from <https://www.wired.com/story/youtube-will-link-directly-to-wikipedia-to-fight-conspiracies/>
 - [51] Tom McCarthy. 2020. Zuckerberg says Facebook won't be "arbiters of truth" after Trump threat. *The Guardian*. Retrieved January 12, 2021 from <https://www.theguardian.com/technology/2020/may/28/zuckerberg-facebook-police-online-speech-trump>
 - [52] Medium. 2019. Medium Rules | by Medium | Medium Policy. Retrieved January 11, 2021 from <https://policy.medium.com/medium-rules-30e5502c4eb4>
 - [53] Jonathon Morgan and Renee DiResta. 2018. *Information Operations are a Cybersecurity Problem: Toward a New Strategic Paradigm to Combat Disinformation - Just Security*. Retrieved January 11, 2021 from <https://www.justsecurity.org/59152/information-operations-cybersecurity-problem-strategic-paradigm-combat-disinformation/>
 - [54] L. Palen and Kenneth M. Anderson. 2016. Crisis informatics—New data for extraordinary times. *Science* 353, June: 224–225. <http://doi.org/10.1126/science.aag2579>
 - [55] Lysia Palen, Sarah Vieweg, Sophia B. Liu, and Amanda Lee Hughes. 2009. Crisis in a Networked World. *Social Science Computer Review* 27, 4: 467–480. <http://doi.org/10.1177/0894439309332302>
 - [56] Thomas Poell. 2014. Social media and the transformation of activist communication: exploring the social media ecology of the 2010 Toronto G20 protests. *Information Communication and Society*. <http://doi.org/10.1080/1369118X.2013.812674>
 - [57] Thomas Poell and Erik Borra. 2012. Twitter, YouTube, and Flickr as platforms of alternative journalism: The social media account of the 2010 Toronto G20 protests. *Journalism*. <http://doi.org/10.1177/1464884911431533>
 - [58] Thomas Rid. 2020. *Active Measures: The Secret History of Disinformation and Political Warfare*. Farrar, Straus and Giroux.
 - [59] Tony Romm and Rachel Lerman. 2021. Amazon suspends Parler, after Google and Apple also take action. *Washington Post*. Retrieved January 12, 2021 from <https://www.washingtonpost.com/technology/2021/01/09/amazon-parler-suspension/>
 - [60] Ruptly. 2014. Ruptly Video News Agency and LiveLeak.com announce content partnership -- Ruptly | PRLog. *PR Log*. Retrieved January 11, 2021 from <https://www.prlog.org/12299623-ruptly-video-news-agency-and-liveleakcom-announce-content-partnership.html>
 - [61] Alexandra Segerberg and W. Lance Bennett. 2011. Social media and the organization of collective action: Using twitter to explore the ecologies of two climate change protests. *Communication Review*. <http://doi.org/10.1080/10714421.2011.597250>
 - [62] social media. 2020. No Title. *Merriam-Webster.com*. Retrieved from [https://www.merriam-webster.com/dictionary/social media](https://www.merriam-webster.com/dictionary/social%20media)
 - [63] Olivia Solon. 2017. How Syria's White Helmets became victims of an online propaganda machine. *The Guardian*. Retrieved January 11, 2021 from <https://www.theguardian.com/world/2017/dec/18/syria-white-helmets-conspiracy-theories>
 - [64] 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*. <http://doi.org/10.1145/3359229>
 - [65] Kate Starbird, Ahmer Arif, Tom Wilson, Katherine Van Koeveering, Katya Yefimova, and Daniel Scarnecchia. 2018. *Ecosystem or Echo-System? Exploring Content Sharing across Alternative Media Domains*. Retrieved January 11, 2021 from <http://syriacivildefense.org>
 - [66] Kate Starbird, Ahmer Arif, Tom Wilson, Katherine Van Koeveering, Katya Yefimova, and Daniel P. Scarnecchia. 2018. Ecosystem or Echo-System? Exploring Content Sharing across Alternative Media Domains. In *12th*

International AAAI Conference on Web and Social Media (ICWSM).

- [67] Kate Starbird and L. Palen. 2011. Voluntweeters: Self-organizing by digital volunteers in times of crisis. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*.
- [68] Kate Starbird, Emma S Spiro, Ahmer Arif, et al. 2015. Expressed Uncertainty and Denials as Signals of Online Rumoring. *Collective Intelligence*, December 2014: 1–4.
- [69] Steem. 2017. *An incentivized, blockchain-based, public content platform*. Retrieved January 11, 2021 from <https://steem.com/SteemWhitePaper.pdf>
- [70] Steemit. Faq – Steemit. Retrieved January 11, 2021 from https://steemit.com/faq.html#How_does_Steemit_work
- [71] Cole Stryker. 2014. Murder, Mayhem and the Evolution of Website LiveLeak. *Newsweek*. Retrieved January 11, 2021 from <https://www.newsweek.com/2014/10/10/murder-mayhem-and-evolution-website-liveleak-273963.html>
- [72] Kjerstin Thorson, Kevin Driscoll, Brian Ekdale, et al. 2013. Youtube, Twitter and the Occupy Movement. *Information, Communication & Society* 16, 3: 421–451. <http://doi.org/10.1080/1369118X.2012.756051>
- [73] Twitter. 2020. Platform manipulation and spam policy. Retrieved January 11, 2021 from <https://help.twitter.com/en/rules-and-policies/platform-manipulation>
- [74] United Nations High Commissioner for Refugees. 2018. Syria emergency. Retrieved January 12, 2021 from <https://www.unhcr.org/en-us/syria-emergency.html>
- [75] United States. 1997. 47 U.S.C. 230 - Protection for private blocking and screening of offensive material - Content Details - USCODE-2011-title47-chap5-subchapII-partI-sec230. Retrieved January 11, 2021 from <https://www.govinfo.gov/app/details/USCODE-2011-title47/USCODE-2011-title47-chap5-subchapII-partI-sec230>
- [76] Amy Volda, Ellie Harmon, Ban Al-Ani, and Donald Bren. 2012. Bridging between organizations and the public: Volunteer coordinators’ uneasy relationship with social computing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*.
- [77] Jonathan Walter. 2020. Content Moderation Is Not Synonymous With Censorship. *Public Knowledge*. Retrieved January 11, 2021 from <https://www.publicknowledge.org/blog/content-moderation-is-not-synonymous-with-censorship/>
- [78] Jen Weedon, William Nuland, and Alex Stamos. 2017. Information operations and Facebook. version 1: 27.
- [79] Tom Wilson and Kate Starbird. 2020. Cross-Platform Disinformation Campaigns: Lessons Learned and Next Steps Research question. *The Harvard Kennedy School Misinformation Review* 1, 1: 1–11. <http://doi.org/https://doi.org/10.37016/mr-2020-002>
- [80] Tom Wilson, Kaitlyn Zhou, and Kate Starbird. 2018. Assembling Strategic Narratives. *Proceedings of the ACM on Human-Computer Interaction*. <http://doi.org/10.1145/3274452>
- [81] Vivian Yee. 2019. James Le Mesurier, Backer of Syrian White Helmets, Is Found Dead in Turkey - The New York Times. *New York Times*. Retrieved January 11, 2021 from <https://www.nytimes.com/2019/11/11/world/middleeast/james-le-mesurier-white-helmets-dead.html>
- [82] YouTube. YouTube Community Guidelines & Policies - How YouTube Works. Retrieved January 11, 2021 from <https://www.youtube.com/howyoutubeworks/policies/community-guidelines/>
- [83] K. Zaram. 2013. Syria’s Civil War Plays Out On Social Media. *Associated Press*.

Received January 2021; revised April 2021; accepted May 2021