



Twitter Is the Megaphone of Cross-platform Messaging on the White Helmets

Sameera Horawalavithana^(✉), Kin Wai Ng^{}, and Adriana Iamnitchi^{}

University of South Florida, Tampa, USA
{sameera1, kinwaing, ai}@usf.edu

Abstract. This work provides a quantitative analysis of the cross-platform disinformation campaign on Twitter against the Syrian Civil Defence group known as the White Helmets. Based on four months of Twitter messages, this article analyzes the promotion of URLs from different websites, such as alternative media, YouTube, and other social media platforms. Our study shows that alternative media URLs and YouTube videos are heavily promoted together; fact-checkers and official government sites are rarely mentioned; and there are clear signs of a coordinated campaign manifested through repeated messaging from the same user accounts.

Keywords: Cross-platform information diffusion · Disinformation campaigns

1 Introduction

Online disinformation is of critical concern for both academic research and industry. Facebook and Twitter recently detected coordinated users who were spreading rumors as part of a large network referencing a “significant state-backed information operation” [1]. Previous research in disinformation campaigns focused on single social media channels, more specifically Twitter [9]. However, several studies have shown that the interdependence between digital sources (e.g., news media outlets) and other social media platforms play a significant role in the diffusion and amplification of information [6]. For example, a coordinated group built its social presence in Google+, YouTube, Facebook, Twitter, Tumblr, Soundcloud, and Instagram around *Black Lives Matters* topics [2].

In disinformation campaigns, Twitter has played a significant role along with YouTube [12] and alternative media sites [10]. In addition to the sources of disinformation, which tend to be YouTube and alternative media sites, other parties that engage in response to disinformation campaigns are main stream media, which are typically providing factual information; government websites, which post the official positions of their countries; and fact-checking organizations that aim to reduce the spread of disinformation. Starbird et al. [12] emphasized the

importance of multiple platforms in promoting the disinformation campaign against the White Helmets.

This work studies how different categories of digital platforms are represented in Twitter conversations related to the White Helmets. The White Helmets are a search-and-rescue Syrian volunteer organization that threatened the Assad regime by its reporting and documenting of chemical attacks on civilians by the government and its allied forces. A disinformation campaign was coordinated [9] in order to discredit the White Helmets and thus delay the intervention of international organizations against the regime in power [2]. In this campaign, White Helmets are often framed as criminals and terrorists, and are responsible for staging chemical weapons attacks [9]. Based on four months of data that records two important events in the White Helmets campaign, we characterize Twitter activity in promoting content from eight categories of digital platforms: YouTube, alternative media, main stream media, social media, fact checking, government, official White Helmets websites, and others which consist of digital sources that could not be mapped to the previously mentioned categories. We capture the promotion of this content by analysing URLs shared on Twitter.

Our study shows the following. First, we confirm a recent study [12] surrounding the White Helmets that shows that YouTube is at the center of this campaign. Second, we discover that alternative media URLs are promoted via exploiting the accessibility of video content: these URLs are often co-appearing with YouTube URLs in the same tweets. Third, unlike previously believed, other social media platforms in addition to Twitter and YouTube have a significant role in this campaign. Specifically, Facebook, Gab and Steemit URLs are present along with YouTube videos on Twitter. Fourth, our data-driven investigation shows signs of coordination for promoting URLs for particular domains.

2 Data Collection and Processing

We focus on cross-platform information diffusion data specific to the White Helmets (WH) disinformation campaigns during four months starting from April 2018. The dataset was provided privately as part of DARPA SocialSim program. The data was collected from April 1, 2018 to July 31, 2018 using terms related to the WH in Twitter (listed in Table 1). This data collection period covers two critical offline events, a chemical attack in April 2018 and the intervention of the Israeli forces to save the White Helmets in July 2018. While the keywords used for the data collection were general, the resulting Twitter dataset captures a significant portion of the anti-White Helmets discussions as also seen in previous studies [12]. To understand the patterns of sharing other platforms' information in Twitter conversations, we only consider the Twitter messages that contain at least a URL, which consist of 25% from the total messages. We filtered out URLs to Twitter itself (that typically refer to other tweets) and URLs to web domains that are mentioned only once in our dataset (3,025 such URLs had no effect on information diffusion).

The main characteristics of the dataset are listed in Table 2. The resulting dataset consists of 53,297 seed messages including tweets, replies and quotes

Table 1. Keywords used for data collection.

'#cascosblancos', '#casquesblancs', '#weissehelme', '#weisshelme', '#whitehelmet', '#whitehelmets', 'caschi bianchi', 'capacetes brancos', 'cascos blancos', 'cascosblancos', 'casques blancs', 'elmetti bianchi', 'weisshelme' , , 'syria civil defence', 'syria civil defense', 'syrian civil defence', 'syrian civil defense', 'weiß helme', 'weiße helme', 'white helmet', 'WH', 'whitehelmet', 'whitehelmets'
--

(78% in English, 4% in Arabic, 3% in German, 2% in French, 1% in Russian, 12% other) and 42,230 retweets that are in response to the original tweets. While we do not analyze content other than URL mentions, including tweets in different languages captures a better picture of URL dissemination across languages.

Table 2. Dataset size. Seeds are tweets (42,087), replies (9,824) and quotes (1,386) that introduce at least one URL. Retweets (42,230) are in response to seeds. Replies are comments to tweets, quotes are comments to retweets.

	Seed messages	Retweets
# Distinct URLs	14,307	12,865
# Distinct Domains	983	957
# Users	18,162	16,578
# Records	53,297	42,230

Each message (tweet, retweet, quote, reply) in our dataset contains the following information: an assigned unique identifier, the anonymized ID of the user who posted it, the timestamp of when it was posted, and its type (whether tweet, retweet, quote, or reply). If the message is a retweet, quote, or reply, the unique identifier of the post that this message is referring to is also included. In addition, the external links (e.g., a tweet mentioning a YouTube video, or an external website domain) mentioned in the messages are pre-processed as following. The shortened URLs are expanded, and HTML parameters are removed from the URLs. The YouTube URLs are resolved to the base URL if they include a parameter referencing a specific time in the video. This pre-processing code of resolving URLs is publicly available [7].

3 Mentions of Digital Platforms in Twitter Messages

We classify the URLs mentioned in Twitter messages into eight classes: i) YouTube; ii) social media platforms (SM) (excluding YouTube); iii) mainstream media domains (MSM), iv) alternative media domains (ALT), v) government (GOV), vi) fact-checking (FACT_CHECK), vii) WH official media

outlets, and viii) other web domains (OTHER) that are not mapped to any other classes.

We used an existing classification of the mainstream and alternative media outlets as identified and publicly available [8]. The first two authors manually performed the classification for the majority of the web domains (94%) appearing in our dataset but outside of the existing identification. This approach followed an iterative and collaborative decision process to finalize the identification. However, we have to stress that the classification of alternative news media domains is considered pejorative to the classification of mainstream news media domains [10]. There are only four fact-checking web-domains (i.e., `snopes.com`, `mediabiasfactcheck.com`, `factuel.afp.com`, and `stopfake.org`) that appeared in our dataset. We identify government media outlets mainly using the regex expression `*.gov.*` The other popular government websites that have not been mapped to this regex pattern include the UK parliament (`parliament.uk`), European parliament (`europa.eu`), Ministry of Foreign Affairs France (`diplomatie.gouv.fr`), United Nations (`un.org`), and the government of Canada (`canada.ca`). We also include an intergovernmental organisation `opcw.org` (i.e., the Organisation for the Prohibition of Chemical Weapons) as a government web domain. OPCW was an official voice on the issue of chemical weapon usage in Syria during our observation period. We also identified three official websites `thesyriacampaign.org`, `syriacivildefense.org`, and `whitehelmets.org` that are directly owned or handled by WH and their subsidiaries. There are 171 (17%) web domains that are not mapped to the seven categories discussed above. This list includes web search, entertainment, university, petition web domains, etc. In the end, we identified 620 alternative media domains, 112 mainstream media domains, 50 social-media domains, and 18 government media outlets. We made this complete list of web domains and the associated media categories available online [3].

Table 3. Top-10 domains by URLs.

Domain	# URLs
facebook.com	1880
youtube.com	1811
paper.li	291
rt.com	255
mintpressnews.com	232
sputniknews.com	209
globalresearch.ca	178
southfront.org	175
theguardian.com	146
gab.ai	140

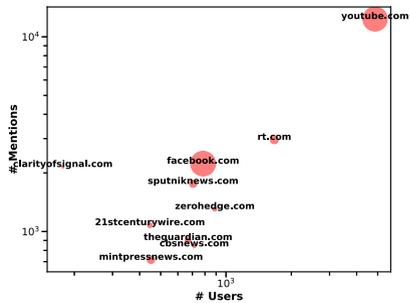


Fig. 1. Top-10 domains by mentions.

Table 3 shows the top 10 domains by the number of URLs in our dataset. Facebook and YouTube are the most popular social media platforms that

were referenced by URLs. The third most shared domain is `paper.li`, a “content curation platform that enables individuals to create newspapers based on topics they choose”. The next five domains (`rt.com`, `mintpressnews.com`, `sputniknews.com`, `globalresearch.ca`, `southfront.org`) are alternative media domains that spread master narratives in the Russia’s disinformation campaign. `Guardian.com` is the only mainstream media that breaks in to the top 10 most shared domains. This is mostly due to an active journalist, Olivia Solon, who is documenting the disinformation campaign against the WH. We also notice the URLs referenced to `gab.ai`, the notorious hate-speech platform.

Figure 1 shows the number of mentions and users involved in these URL-sharing activities. The size of the markers in this plot are proportional to the number of URLs associated with the domain. One would expect that the volume of mentions of a web domain and the number of users who share URLs from that domain are correlated with the number of distinct URLs from that domain, thus the domains that appear in Table 3 appear in Fig. 1. However, `clarityofsignal.com`, `zerohedge.com`, and `21stcenturywire.com` are the alternative media sites that have unexpected popularity for the relatively small content shared (number of URLs). Moreover, we notice that `clarityofsignal.com` acquires its many mentions from very few users: 177 users made 2,146 mentions during our observation period. Similarly, with few URLs shared (255), `rt.com` has high popularity in both users (1,670) and number of mentions (2,958). `rt.com` is directly handled by the Russian state media, while `clarityofsignal.com` promotes conspiracy theories criticizing Western governments.

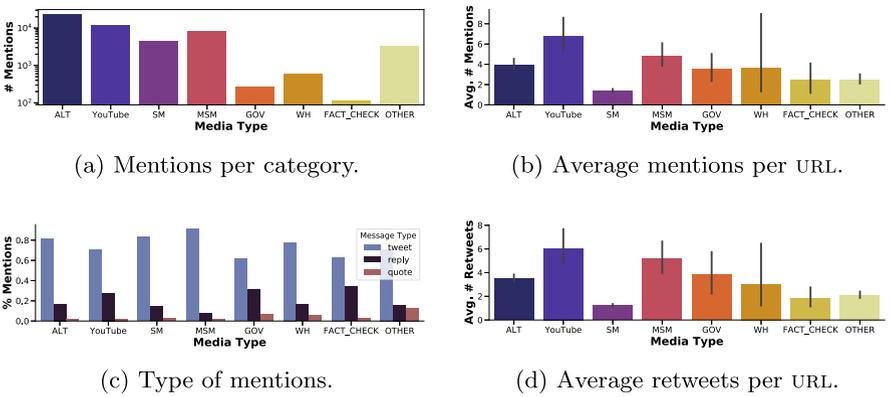


Fig. 2. URL sharing activities per media category.

Figure 2 shows the characteristics of URL mentions, the rate of mentions, and the rate of retweets received for the URL messages. The results are presented according to the eight media categories of the shared URLs. First, the majority of URL mentions are referencing alternative media domains (Fig. 2a). This is mainly due to the number of alternative media domains (63%) that participate in this

disinformation campaign. However, a YouTube URL has higher rate of mentions relative to any other media category (Fig. 2b). Mainstream media URLs also receives more mentions than alternative media. Notably, the official WH web sites received similar number of mentions compared to alternative media URLs on average (Fig. 2b). This is despite the fact that WH and alternative web domains promote opposing views on related events. The URLs referencing articles hosted at government media outlets and the fact checking websites are relatively low compared to other media domains, but such URLs have higher rate of mentions compared to social-media URLs.

Second, 25% of URLs are injected via quotes and replies, suggesting a concerted effort to promote those URLs, as shown in Fig. 2c. We note that less than 1% URLs are mentioned via quotes and replies in another Twitter dataset collected outside the topic of this study [4]. We also notice that URLs referencing government and fact-checking websites are mentioned proportionally more in replies compared to any other URLs. Unlike previously observed [11], the diffusion of URLs in this particular dataset is very limited: on average, less than one retweet is posted for every mention of a URL. A YouTube video URL and a mainstream media URL are mentioned in relatively higher number of retweets than any other URL (as shown in Fig. 2d). Compared to other media categories, YouTube videos provide audio and visuals to attract users, and the mainstream media are often used for validation and credibility.

4 Media Co-sharing Patterns

In order to understand the relationships between categories of digital platforms as expressed at tweet level, cascade level and user level, we construct three bipartite networks. The first network connects a tweet to the web categories to which the URLs that appear in that tweet belong. This network captures the co-occurrences of different web categories in the same tweet, thus capturing the potential similarity of the views expressed in the content published in those categories. We project this network in the web category space and we refer to it as the *tweet-sharing network*, where nodes are web categories and edge weights record the number of tweets that mentioned URLs from both web categories. (Note that the mentioned URLs do not have to be identical.)

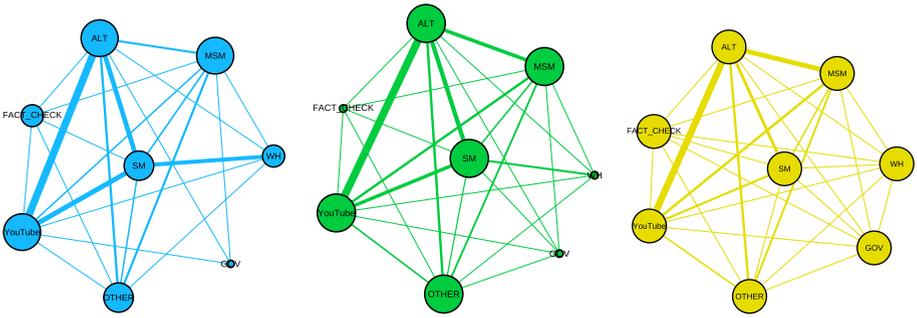
The second network connects a cascade (i.e., a conversation that starts with a tweet along with replies responding to the original tweet, or a set of endorsements via quotes and retweets) to the web categories to which the URLs that appear in that cascade belong. This network captures the co-occurrences of different web categories in the same cascade which might consist of a set of users who hold similar or opposite views. We project this network in the web category space and call it the *cascade-sharing network*; nodes are web categories and edge weights record the number of cascades that mentioned URLs from both categories.

The third network connects users with the web categories to which the URLs they share belong. This network captures the co-occurrence of different web categories in the content posted by a user and reflects the user interest over multiple

types of digital platforms. We project this network again in the category space and obtain a network that we call *user-sharing network* in which nodes are web categories and edge weights represent the number of users who posted on Twitter content that includes cascade URLs from both categories. Figure 3a and b present the tweet-sharing and cascade-sharing networks, respectively. User-sharing network is shown in Fig. 3c.

4.1 Tweet/Cascade-Sharing Network

The tweet-sharing network (Fig. 3a) is missing five edges in total. There are no tweets that contain references to government sites along with references to social media pages, fact-checkers, WH official websites or other websites. On one hand, mentioning fact-checking and government URLs could serve the same purpose of verifying facts: government sites publish the official information, fact checkers verify the information. This assumption is confirmed by the fact that the two types of information sources do not appear in the same cascade either (Fig. 3b). While mentioning fact-checking and government sites in the same tweet may be redundant or cumbersome, the lack of tweets that refer to both social media and government sites is less intuitive. However, social media URLs are mentioned in the replies of the three cascades originated from a government website, as shown in Fig. 3b. We also noticed the lack of tweets (or cascades) that mention a WH official website along with a fact-checking or a government website. We believe this fact is due to the content promoted by the WH official websites. A majority of these URLs redirects to WH donation campaign websites, mostly shared by WH sympathizers.



(a) Tweet-sharing network (b) Cascade-sharing network (c) User-sharing network

Fig. 3. Visualization of entity-sharing networks, Entity is a tweet (a, c), cascade (b), and user (d). The thickness of an edge is proportional to the number of entities that mention at least two URLs referencing both media categories.

The strongest edge connects YouTube with alternative media (via 250 tweets), while the weakest edge connects mainstream media and fact-checking

sites via one tweet. A strong relationship is also seen in the tweet-sharing network between social media and YouTube (164 tweets) and between social media and alternative media (154 tweets). We also noticed social media and WH official websites are mentioned together in 131 tweets. 98% such tweets are in Korean, and promote the WH donation campaign via posts from multiple social media platforms. Similar patterns are seen in the cascade sharing networks: YouTube and alternative media sites appear together in the largest number of conversations (577), while alternative and social media co-appear 313 times.

Because YouTube is the largest player in this ecosystem, we zoom in on the web domains that co-appear with a YouTube URL in the same tweet. The top-5 web domains are `rt.com` (78), `clarityofsignal.com` (60), `russia-insider.com` (59), `steemit.com` (56) and `facebook.com` (48) according to the number of tweets in which the website and a YouTube URL were mentioned together. For calibration, the median number of tweets is 3 (between YouTube and `21wire.tv`) and the mean is 9. Apart from `facebook.com` and `steemit.com` social-media domains, the other three domains are known to promote alternative narratives that denigrate the WH [2]. On the other hand, URLs from these three domains often co-occur in the same tweet: `russia-insider.com` and `clarityofsignal.com` co-occured with `rt.com` in 59 and 46 tweets, respectively, while `russia-insider.com` co-occured with `clarityofsignal.com` in 47 tweets. These sites are known to post the same content occasionally [10]. Moreover, `rt.com` and `clarityofsignal.com` are the two domains from the top-10 most popular in mentions that have relatively few URLs in our dataset, as previously discussed on Fig. 1.

Another observation is related to the 7th most popular web domain `21stcenturywire.com` that co-occured with a YouTube URL in 19 tweets. Videos from two YouTube channels are co-tweeted with `21stcenturywire.com` URLs: i) the official Russian Television YouTube channel, and ii) the YouTube channel of UK blogger Vanessa Beeley, a well-known anti-WH voice. Along with other user accounts, she repeatedly includes both URLs to these channels and her own articles posted on `21stcenturywire.com` in a typical card-stacking propaganda device manner [5]. This proves the cross-platform disinformation practice where a message is pushed on multiple digital platforms in their respective formats.

The top three social media sites that co-occured with a YouTube URL are `steemit.com`, `facebook.com` and `gab.ai`. We manually inspect the corresponding tweets and the content of these URLs and observe the following. First, 95% of the Twitter messages (51 out of 56) that contain both a `steemit.com` URL and a YouTube video URL have exactly the same content and are authored by the same user. The majority of such duplicated messages are in the form of replies or quotes. These messages are not available in Twitter anymore as of February 2020, which might be due to account suspension.

Second, Facebook connects to YouTube in two ways. On one hand, a pro-WH voice disseminates videos published by the official WH YouTube and Facebook accounts in the same tweet in Arabic. There are 22 tweets authored by this user with this similar pattern. On the other hand, two anti-WH users also tweet

Facebook and YouTube videos in Arabic, but this time the Facebook URLs point to the Facebook page of each user, where the same YouTube videos already shared in their tweets are also found.

Finally, `gab.ai` and YouTube are connected via 32 tweets in which they co-occur (all the articles referenced by Gab URLs in this set are not available as of February 2020). 60% of these messages are promoted by the same usernames in both Twitter and Gab. This is another example of promoting content from one platform to Twitter via exploiting the appeal of YouTube videos.

4.2 User-Sharing Network

Figure 3c that visualizes the user-sharing network suggests the following four observations. First, the network has no missing edges (full clique). But the missing edges in the tweet-sharing network appeared in the user-sharing network with very small edge weights. The weakest edges connect government and fact-checking (3), WH official web domains with fact-checking (5) or government (6) web domains. Second, the majority (1,242) of users who share alternative media URLs also share YouTube video URLs. Many alternative websites and YouTube channels that appeared in our dataset promote anti-WH content. This shows the existence of users who promote anti-WH content from different digital platforms. Third, the next largest population (859) disseminates both alternative and mainstream media. Notably, the majority of mainstream media sites that appeared in our dataset produce more pro-WH content than anti-WH content, which is analogous to the findings in [10]. Anti-WH users in Twitter often shared mainstream media and alternative media URLs in the same tweet. The alternative media articles were used as evidence to strengthen their arguments in order to debunk the narratives presented in the mainstream media articles. Fourth, very few users who tweet an alternative media or a YouTube video URL share a fact-checking URL. Out of 7,377 users who shared an alternative URL, only 32 (0.004%) users shared a fact-checking URL. One possible explanation is the disregard of fact-checking sites by anti-WH users. For example, a popular narrative among users who shared anti-WH content framed WH as a terrorist organization. `Snopes.com` verified this claim as false in an article published in December 2016, yet the alternative media articles and YouTube videos promoted the same claim even two years later.

5 Summary

This study investigates how Twitter is used to disseminate White Helmets-related information hosted on other platforms, and especially YouTube. From four months of data we identified the following patterns of URL promotion on Twitter. First, YouTube videos and alternative media URLs are often bundled in the same message. Second, main stream media and alternative media URLs are also bundled together in the same message or in messages posted by the same user. From manual inspections, the majority of such messages use alternative media content to attack the narratives and the interpretations published

by mainstream media. Third, we discovered a small number of websites with unusual popularity despite the small number of related URLs in our dataset. In some cases, these websites are aggressively and repeatedly promoted by a small group of users via mentions. And finally, very few messages include official government websites or fact-checking websites in the Twitter conversations on White Helmets, despite the active disinformation campaign going on during our observation period.

Our study demonstrates the need to develop intervention techniques to limit disinformation spread across a broader media ecology than a single platform in isolation.

Acknowledgements. This work is supported by the DARPA SocialSim Program and the Air Force Research Laboratory under contract FA8650-18-C-7825. The authors would like to thank Leidos for providing data.

References

1. Baca, M., Romm, T.: Twitter and Facebook take first actions against china for using fake accounts to sow discord in Hong Kong (2019). <https://www.washingtonpost.com/technology/2019/08/19/twitter-suspends-accounts-it-accuses-china-coordinating-against-hong-kong-protesters/>
2. DiResta, R., et al.: The Tactics & Tropes of the Internet Research Agency. New Knowledge. (2018). <https://disinformationreport.blob.core.windows.net/disinformation-report/NewKnowledge-Disinformation-Report-Whitepaper.pdf>. Accessed 28 Sept 2020
3. Horawalavithana, S.: Wh_twitter_megaphone (2018). https://github.com/SamTube405/WH_Twitter_Megaphone
4. Horawalavithana, S., Bhattacharjee, A., Liu, R., Choudhury, N.O., Hall, L., Iamnitchi, A.: Mentions of security vulnerabilities on Reddit, Twitter and GitHub. In: IEEE/WIC/ACM International Conference on Web Intelligence, pp. 200–207 (2019)
5. Lee, A., Lee, E.B.: The Fine Art of Propaganda. Harcourt, San Diego (1939)
6. Myers, S.A., Zhu, C., Leskovec, J.: Information diffusion and external influence in networks. In: 18th ACM KDD (2012)
7. PNNL, P.N.N.L.: Socialsim (2018). <https://github.com/pnnl/socialsim>
8. Starbird, K.: Examining the alternative media ecosystem through the production of alternative narratives of mass shooting events on Twitter. In: AAAI ICWSM (2017)
9. Starbird, K., Arif, A., Wilson, T.: Disinformation as collaborative work: surfacing the participatory nature of strategic information operations. Proc. ACM Hum.-Comput. Interact. **3**(CSCW), 127:1–127:26 (2019)
10. Starbird, K., Arif, A., Wilson, T., Van Koeveering, K., Yefimova, K., Scarnecchia, D.: Ecosystem or echo-system? Exploring content sharing across alternative media domains. In: AAAI ICWSM (2018)
11. Suh, B., Hong, L., Pirolli, P., Chi, E.H.: Want to be retweeted? Large scale analytics on factors impacting retweet in Twitter network. In: 2010 IEEE Second International Conference on Social Computing, pp. 177–184. IEEE (2010)
12. Wilson, T., Starbird, K.: Cross-platform disinformation campaigns: lessons learned and next steps. The Harvard Kennedy School (HKS) Misinformation Review (2020)