

Fight Fire with Fire: Hacktivists' Take on Social Media Misinformation

Filipo Sharevski
DePaul University

Benjamin Kessell
DePaul University

Abstract

In this study, we interviewed 22 prominent hacktivists to learn their take on the increased proliferation of misinformation on social media. We found that none of them welcomes the nefarious appropriation of trolling and memes for the purpose of political (counter)argumentation and dissemination of propaganda. True to the original *hacker* ethos, misinformation is seen as a threat to the democratic vision of the Internet, and as such, it must be confronted on the face with tried hacktivists' methods like deplatforming the "misinformers" and doxing or leaking data about their funding and recruitment. The majority of the hacktivists also recommended interventions for raising misinformation literacy in addition to targeted hacking campaigns. We discuss the implications of these findings relative to the emergent recasting of hacktivism in defense of a constructive and factual social media discourse.

1 Introduction

Steven Levy's portrayal of the hacker culture in his 1984 book *Hackers* largely remains the most influential reference to the public's general view of hackers [43, 65]. Recasting them Robin Hood-style activists committed to a democratic vision of the Internet [97], Levy asserts that the hacker ethos embodies several sacrosanct postulates to the public good, notably that (i) *all information should be free*, and (ii) *authority should be mistrusted and decentralization promoted* [65].

Later-day Internet hackers shifted the ideological tendency for autonomy in the cyberspace towards a vision of the Internet as a popular space for sharing any information that can nevertheless be politicized and weaponized against the neoliberal elites responsible for economic and social disarray [37]. Turning Internet activism into a form of socio-political resistance online [58], enabled a functional selection of issues that no longer necessitated a long preparation [74]. This, in turn, resulted in almost instant convergence and coordination of activities in response to the issues of interest that, over the years, became publicly visible through mass media coverage [47].

The Internet activism, expectedly, bifurcated to online campaigns concerned with the protection of the Internet as a relatively unregulated and unowned space (e.g. Anonymous, WikiLeaks, Snowden [21, 114, 116]) and online campaigns concerned with the protection of human rights and the environment (e.g. the Occupy movement, Arab Spring, Pirate Party [59, 80]). The former activism – or *hacktivism* – often is anonymous, performed in secret, and operates with a kind of impunity that the Internet technologies seem to afford so far [117]. The later activism – or *hashtag activism* – usually is public, openly used the Internet for political mobilization, operates primarily on the streets, and subjects to the dangers of crowd violence, harassment, and arbitrary arrest [100].

The hashtag activism historically utilized various Internet technologies such a petition websites (e.g. MoveOn.org for organizing political protests) or e-mail communication (e.g. Tea Party's campaign to reduce government spending and taxation) [16], but the advent of social media sites like Twitter, Facebook, and YouTube truly accelerated the self-organization and participation in the sociopolitical struggle (e.g. the #BlackLivesMatter and #SchoolStrike4Climate movements [34]). While the essential dependence on social media is apparent, both in a historical context and for the future of the hashtag activism [56], the relationship between the hacktivism and social media is a bit more complicated.

Hacktivists, in contrast, hacked various Internet technologies such as defacing websites [98], breaking into systems to "leak" and "dox" private documents [114, 118], and storm systems with traffic to cause a Denial-of-Service (DOS) [81]. Hacktivists' foray in social media mirrors these actions as campaigns were undertaken for hijacking/defacement of social media accounts (e.g., Anonymous's #OpKKK campaign [128]), doxing individuals on Twitter (e.g. the students of Covington High School [70]), and DoS Twitter topics (e.g. #IranTalks campaign [86]). But hacktivists also hacked the social media affordances for content amplification (e.g. Stay-WokeBot [36, 102]), early instances of trolling (e.g. Rick-rolls [101]), and sharing memes (e.g. Lol Cats on 4chan [21]).

Despite the intuitive versatility of social media for such

subversive operations, hacktivism became largely inactive on the mainstream platforms following some high profile run-ins with the legal authorities of the leading hacktivists [53, 124]. The apparent absence of hacktivism created a vacuum where no one actively challenged the elites, defended freedom of expression, and appended the vision of democratic social media participation. It took little time, unfortunately, for this vacuum to be appropriated by state-sponsored actors hijacking the hacking playbook for actions aimed not just against the neoliberal elites but the entire social order [32]. Bot-style amplification aided political trolling and sharing of memes in the aftermath of Brexit campaign in the UK [24] and the 2016 elections in the US [10]. The crucial difference in these instances was that the amplified memes and trolling were not pranks but damaging fake news, emotionally-charged memes, and conspiracy theories that instead of unifying the social media crowds for a cause, divided them in opposition camps pitted against each other [111].

In response to such a large-scale disruption on the social media turf, one would have plausibly expected that the hacktivists will retaliate and confront, expose, or counter hack the state-sponsored “trolls” [135]. Misinformation, back to the Levy’s depiction of hacker’s ethics [65], runs counter the (i) *all information should be free* postulate because it undermines the basic utility of information as a public good (i.e. truth and facts do not dwindle in supply as more people “consume” them and truth and facts are available to all people in a society) [31]. Misinformation also runs counter the (ii) *authority should be mistrusted and decentralization promoted* postulate because it is promulgated by a state-sponsored “shadow authority,” as evidence confirms in the aftermath of the Brexit and the 2016 US elections [48, 73, 134]. Surprisingly, the hacktivists never struck back [11], though they clearly poses the capabilities to do so, as witnessed in the Anonymous’s #OpISIS campaign, for instance, where the collective flagged about 101,000 Twitter accounts attributed to the Islamic-State [49].

The absence of response to misinformation on social media by the hacktivist community seemed quite perplexing and, in our opinion, worthy of in-depth inquiry with active “hackers” that still operate in the spirit of the Levy’s code of ethics [65]. Through personal connections and snowballing sampling, we identified 22 prominent hacker figures and set down for at least an hour-long interview with each of them to learn their take on the misinformation ecosystem, on responses to falsehoods on social media, and the way misinformation impacts and shapes the hacktivists’ agenda in the future. We found a consensus among the hacktivists against the present forms of misinformation as an ammunition for political counter(argumentation) and external propaganda. They were adamant to deplatform, dox, and expose every “misinformer” that they believe is polluting the social media discourse, and suggested ways to improve the general misinformation literacy among users in addition to these targeted operations.

To situate our study in the intersection between the hack-

tivist counter-culture and the rise of misinformation on platforms, we review the interplay between Internet activism, social media, and false information in Section 2. We look in the broader context of misinformation in Section 3 to highlight the pressing need of hacking action to reclaim the social media space true to Levy’s vision of Internet as an information exchange to the public good. In Section 4 we outline our research design and methodology. Sections 5, 6, and 7 expand on our findings and we discuss the implications of the hackers’ disposition to social media misinformation in Section 8. Finally, Section 9 concludes the paper.

2 Internet Activism and Social Media

2.1 Hashtag Activism

Online social media activism – or *slacktivism*, *clicktivism* – emerged on popular platforms as a repertoire of low-risk, low-cost expressive activities for advocacy groups’ agenda setting and political participation [99]. Social media users participated in petitions, changed personal avatars, added picture filters in support of a cause, and simply “liked” posts as an act of participation [41]. Slacktivists quickly realized they could use virality as a distinctive social media affordance to their advantage and move to use hashtags as the main drivers of mobilization, raising awareness, and demanding sociopolitical change. The practice of *hashtag activism* was instrumental for the success of social movements like #metoo, #takeaknee, and #BlackLivesMatter, allowing for mainstream visibility, expression of solidarity, and statement of victimhood [115]. This success, in turn, inspired a plethora of other movements advocating for health, human rights, social justice, and environmental issues to spur across all social media platforms and remain active within the public discourse [52].

The materialization of the hashtag activism, however noble, had to deal with the obvious threat of *hashtag hijacking* or the encroachment on viral hashtags to inject contrary perspectives into a discourse stream [126]. This “hack” against the internet activism is not just adding noise or attempting to result in a DoS, but also to disseminate hateful narratives and dilute the campaign itself (e.g. the hijacking of the #metoo hashtag [69]). Another similar threat is the *hashtag co-opting* or the contentious co-opting of the rhetoric of popular social movements (e.g. #HeterosexualPrideDay campaign co-opting the language of the mainstream LGBT movement [7]). Equally threatening is the *counter hashtagging* that concocts similar hashtags to garner opposition to well-established movements (e.g. #BlueLivesMatter countermovement to police reform in reaction to #BlackLivesMatter [61]). These antagonistic appropriations of the social media virality, consequently, enabled political extremism to creep in the public discourse and embroil users in an emotionally-charged participation [95].

In a state of emerging social media polarization, it was a question of time when fake news, offensive memes, and

conspiracy theories would be weaponized against the hashtag activism (e.g. the proliferation of fake news in the #Gunreformnow vs #NRA Twitter battle [18]). What was initially expected to remain on the fringes of the mainstream hashtag activism [33], quickly turned into an information disorder on a mass scale. Now the hashtag hijacking and co-opting developed *in parallel* with the main theme of activism, and for that, a steady and substantive feed of false and unverified information was needed. The emotionally-charged participation loomed into a global health panic (e.g. #FlattenTheCurve hashtag hijacking for spreading COVID-19 misinformation [27]) and moral panic (e.g. the QAnon's co-opting of #SaveTheChildren hashtag [83]) in addition to the already growing political panic [85].

2.2 Hacktivism

Hacktivism was a term that “Omega,” a member of the Texas-based computer-hacking group *Cult of the Dead Cow* (cDc) coined in 1996 in an email to the cDc listserv [75]. Characterized with the increasingly political ethos of hacking-for-cause, hacktivists primarily leveraged technology to advance human rights and protect the free flow of information in campaigns against the UK, US, and Chinese governments, as well as the UN [88]. In as much as hackers individually roamed the Internet, socialization was at the proper time as many of them needed establishing a strong hacktivist network. Hacktivists’ penchant for humorous memes (LOLCats) and gag hyperlinks (Rickrolls) [91] attracted an army of hackers to Christopher Poole’s 4chan.org social media website, bringing to life the notorious collective Anonymous [75].

While hacktivists never displayed a predictable trajectory of their cyberoperations and political program [21], they narrowly utilized social media for self-promotion – announcing operations with an #Op prefixed hashtags [11] – and furthering a complex relationship with other Internet activists. Anonymous cried foul on Twitter when WikiLeaks puts millions of its documents behind a pay wall [40], but also launched Operation #Ferguson of doxing the St. Louis County police chief daughter’s information in response to the shooting of the black teenager Michael Brown [9]. Hacktivists, in solidarity to the Arab spring uprisings, sent a care package composed of security tools and tactical advice though downplayed the touted “Twitter Revolution” [21].

True to their credo for utilizing Internet technologies against oppression, including social media, hacktivists launched the #OpKKK to “unhood approximately 1000 Ku Klux Klan members” hacked by gaining access to a KKK Twitter account in support of #BlackLivesMatter protesters in Ferguson, Missouri [128]. After a several years hiatus, perhaps due to arrests of some of the leading Anonymous hacktivists, the group resurfaced during the 2020 #BlackLivesMatter protests in response to the killing of George Floyd [54]. This time, in addition to the leaking of a trove of 269 giga-

bytes of confidential police data (dubbed *BlueLeaks* [64]), the hacktivists launched social bot operations to amplify a support towards #BLM and criticize police actions.

Hacktivists also utilized Internet technologies in the context of cyberwarfare. For example, the #OpIsis operation, in which lists of tens of thousands of Twitter accounts that purportedly belonged to members of ISIS or its sympathizers were leaked, was launched in response to the terrorist attacks in France in 2015 [77]. Here, in addition to the leaks, hacktivists also waged a meme war and called for a “Troll ISIS Day” to provoke and disrupt ISIS-supported social media [76]. The Anonymous group in early 2022 took on Twitter to declare a “cyber war” to Russia in response to the Ukrainian invasion, launching DoS attacks against Russian’s Federal Security Service’s website and hacking Russian streaming services to broadcast war videos from Ukraine [104].

3 Internet Activism and Misinformation

3.1 Grassroots Misinformation Operations

Hacktivists, perhaps inadvertently, authored or gave popularity to the most utilized primitives for creating, propagating, amplifying, and disseminating misinformation - *trolling* and *memes*. This negative externality is unfortunate as trolling and memes were initially used by Anonymous against what they perceived a “misinformation campaign” by the Church of Scientology [75]. The “anon” members on 4chan.org practically *hijacked* the term “troll” – initially meaning provoking others for mutual enjoyment – to abusing others for members’ own enjoyment by posting upsetting or shocking content (usually on the \b channel of 4chan.org [21]), harassing users (e.g. mocking funeral websites [12]), and spreading rumors [62]. What Anonymous did for the “lulz” (a brand of enjoyment etymologically derived from laughing-out-loud (lol)), nonetheless, showed the ease with which one could exploit the Internet technologies to be impolite, aggressive, disruptive, and manipulative to users’ emotional states [21].

Trolling initially came in textual format as comments to posts, bulletin boards, and websites “deindividualized” people’s lived experience for the “lulz” [12]. Gradually, hacktivists popularized a multimedia format of trolling or “memes,” where textual commentary is superimposed over well-known imagery, typically representing different forms of power, such as political leaders, the police, and celebrities [76]. Memes, perhaps, were the actual rite of passage to true hacktivism – moving away from the early LOLCats – as they seek to deconstruct the power represented, contest censorship, and provide political commentary [87]. Memes as content were put to hacktivist use *en masse* in operations like “Troll ISIS day,” where Anonymous proliferated memes with rubber-duck heads or rainbow stripes to ridicule ISIS propaganda imagery and disinformation narratives on Twitter [76]. Spread together with satirizing hashtags (e.g. #Daeshbags), the trolling memes