



Selected Papers of #AoIR2021:
The 22nd Annual Conference of the
Association of Internet Researchers
Virtual Event / 13-16 Oct 2021

“GO BACK IN YOUR FETUS CAVE”: HOW PREDATORY INFLUENCERS MANIPULATE AUDIENCES THROUGH PLATFORM RETREATS

PS Berge
University of Central Florida

Introduction

During summer of 2020, numerous high-profile figures in videogame entertainment—including Twitch and YouTube creators—were exposed for grooming minors and predatory behavior. Among them, British YouTuber Craig Thompson (‘Mini Ladd’ on YouTube), was accused of grooming and sexting underage fans. But just months after confessing to the allegations on Twitter, Thompson uploaded a new *Minecraft* video to his YouTube channel, inciting outrage online as #cancelminiladd and #getminiladdoffyoutube circulated.

YouTube’s response to survivors, however, rang hollow:

Thank you for reaching out – if you think the channel violates our Community Guidelines, you can directly report it here... (@TeamYouTube, Sept. 2020).
[Exact date and user handle redacted for survivor privacy].

YouTube’s reaction typifies the problems of cross-platform insulation—and what I call the *platform retreat*: users exposed for predatory behavior on one platform can take refuge within a different platform’s myopic moderating practices. Other scholars have noted examples of this, such as the *r/deepfakes* community enduring on GitHub after being shut down for circulating revenge porn on Reddit (Winter & Salter, 2019). Thompson’s platform retreat from Twitter to YouTube serves a dual role: he can ‘publicly confess’ on Twitter (claiming the issue is resolved), while maintaining his revenue-earning audience on YouTube. Simultaneously, YouTube is able to “seek protection for facilitating user expression, yet also seek limited liability for what those users say” (Gillespie, 2010, p. 347). This study complicates YouTube’s claim that its role in moderation ends with its channels while ignoring the networked actions of its creators.

Method

This article builds on scholarship about cross-platform digital culture, such as Burgess and Matamoros-Fernández's study that showed how #gamergate was amplified across numerous platforms (2016). Jackson et al. have likewise explored how social media networks intersect with activism and public intervention (2020). Brock's productive framework for critical technoculture discourse analysis (2018) also provides an essential model.

To explore this, I used the snsrape python scraper (JustAnotherArchivist, 2018/2020) to collect historic Twitter data, using the query "miniladd" from 6/1/2020 – 3/31/2021, resulting in a population of 34,316 tweets. I also collected comments and video network data for seven Mini Ladd videos using YouTube Data Tools, resulting in 62,911 comments. I then conducted a mixed-methods social network analysis using Orange (a data mining tool), AntConc concordance tools, and Gephi for visualization.

Analysis

The discursive patterns of the YouTube and Twitter samples were worlds apart. While the Twitter discourse was pointed, organized, and furious with Thompson, the YouTube comments featured a mixture of loyal support and toxic attacks on Thompson's fans. It became evident that it was not simply the characteristics of the platforms that insulated Thompson, but the way these platforms and their users *interacted with each other* that created opportunities for evading accountability. By comparing these two samples, I show:

1. *Thompson moderated his YouTube comments, shielding himself from coherent criticism.*

Many of the most prevalent words in the Twitter discourse (ex. "pedo" [n=1823], "pedophile" [n=1451], "minor(s)" [n=1362], "children/child" [n=885], "#getminiladdoffyoutube" [n=333]) *did not appear once* in Thompson's YouTube comments. However, misspellings, variations, and concatenations of these words *did* appear—such as "p e d o p h i l e," "pediladd," and "r/children." YouTube Studio offers "Blocked Word" filters for creators, and the data indicated that Thompson used these to censor references to his scandal in his comment sections. Thompson's critics were clearly aware of these filters and used leetspeak and concatenations to circumnavigate them. As a result, the narrative in the comments was fragmented, allusive, and enthymematic: couched in jokes and veiled references that required knowledge of the Twitter discourse to understand (Figure 1).

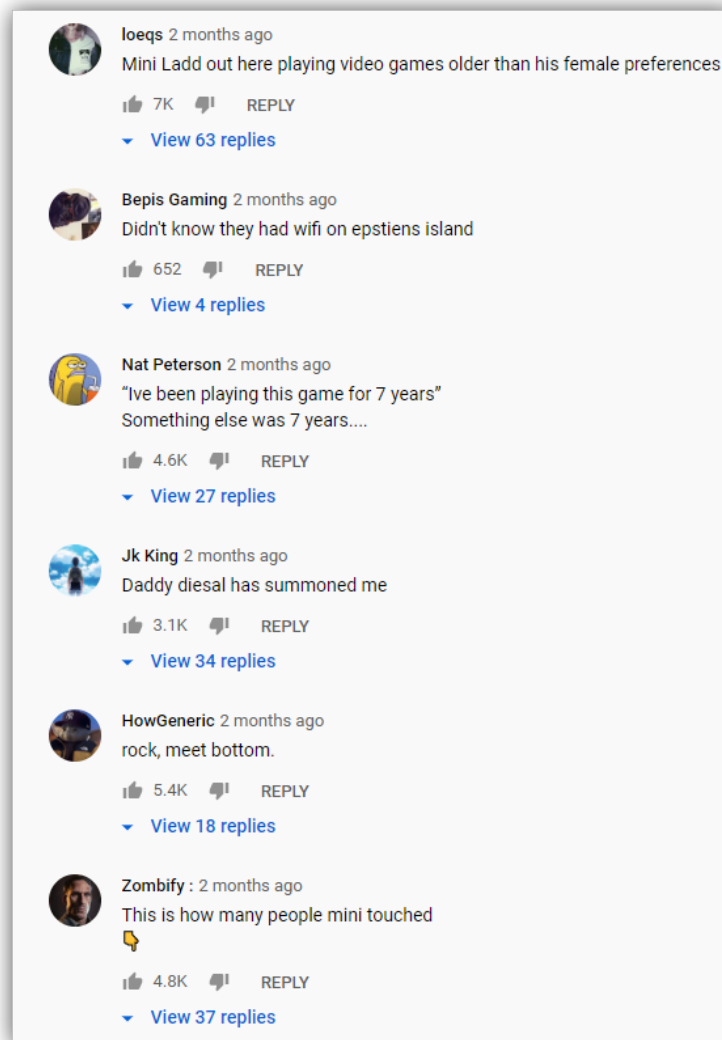


Figure 1. Allusive jokes left on Thompson's videos.

2. Thompson optimized YouTube's uploading pipeline and metadata to prevent algorithmic pairing of his content with videos referencing his behavior.

Because Thompson's primary fanbase was teenagers who were already subscribed to him, there was a direct correlation between time of upload and negative criticism on YouTube comments, as it took critics (including from Twitter) longer to reach Thompson's videos. Because of this, his fans were unlikely to be exposed to negative commentary if they arrived early. This was evidenced in a sentiment analyses of his videos, as there was a clear negative regression over time ($r = -0.19$), and a delay in negative comments.

Additionally, a network analysis of Thompson's videos showed how the algorithms shielded Thompson, while videos made by survivors were siphoned off into 'YouTube Drama' pipelines (Lewis et al., 2020). Thompson's videos were linked with other meme and gaming content, but his two videos referencing the scandal directly were separate from his regular content in the network. This was exacerbated by the way Thompson

tagged his videos. While his videos normally include over a dozen keywords, (ex. “Hilarious, Energetic, Family, Friendly, Comedy, React, gameplay”), videos referencing the scandal had only autogenerated keywords, (ex. “video, sharing, camera phone, video phone, free, upload”).

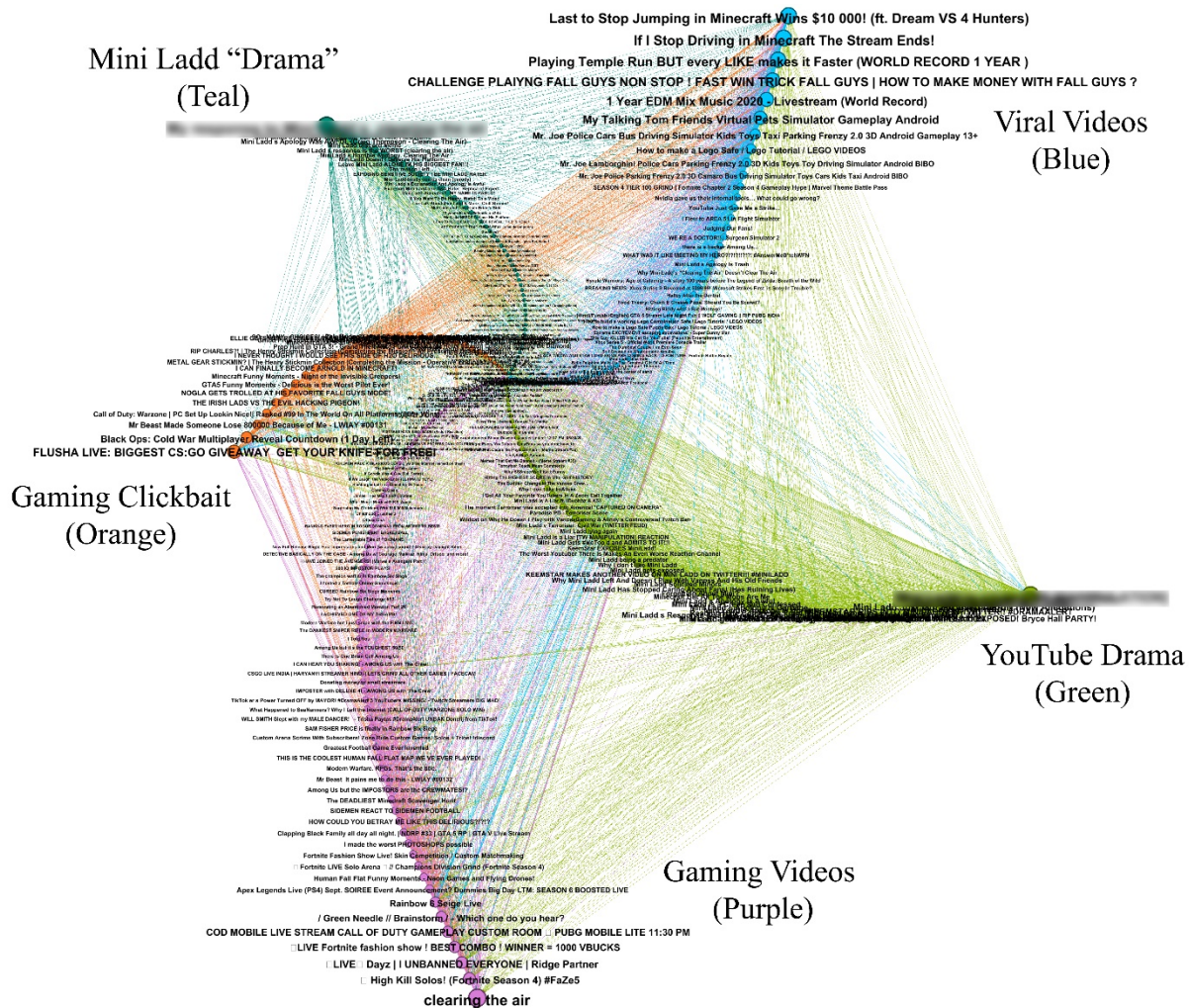


Figure 2. Visualization of video-network from “clearing the air” using Gephi with Radial Axis Layout. Survivor videos redacted for privacy.

3. Drama YouTubers, attempting to signal-boost survivors and expose Thompson, ended up counterproductively weaponizing their fanbases to create further harassment.

Animosity towards Thompson’s young fanbase was widespread in YouTube comments, with the word “fetus” appearing over 700 times (ex. “stfu fetus, before he sends you much love,” “fetus, do you want his mini ladd too?”). Drama YouTubers (Lewis et al., 2020) such as ‘diesel patches’ mobilized their fanbases through callout videos to systematically harass Thompson and his still-loyal supporters. This was exacerbated by several factors: the censorship on Thompson’s channel, where criticizing Thompson directly was difficult; the ignorance of his own fans, unlikely to be on Twitter and hence shielded from the scandal; and the toxic praxis of YouTube drama videos as a genre.

Conclusion

While scholars have excavated the role of moderation on social media platforms (Gillespie, 2010; Marwick & Caplan, 2018), this study emphasizes the need to critically examine inter-platform dynamics. Thompson's case demonstrates how cross-platform insulation can exacerbate harassment and protect predators. YouTube's policies cannot meaningfully address these issues so long as they ignore the networked structures that empower predators: ambiguous moderation policies, exploitable affordances, and cultures of harassment.

References

- Brock, A. (2018). Critical technocultural discourse analysis. *New Media & Society*, 20(3), 1012–1030. <https://doi.org/10.1177/1461444816677532>
- Burgess, J., & Matamoros-Fernández, A. (2016). Mapping sociocultural controversies across digital media platforms: One week of #gamergate on Twitter, YouTube, and Tumblr. *Communication Research and Practice*, 2(1), 79–96. <https://doi.org/10.1080/22041451.2016.1155338>
- Gillespie, T. (2010). The politics of 'platforms.' *New Media & Society*, 12(3), 347–364. <https://doi.org/10.1177/1461444809342738>
- Jackson, S. J., Bailey, M., & Foucault Welles, B. (2020). *#HashtagActivism: Networks of Race and Gender Justice*. The MIT Press. <https://doi.org/10.7551/mitpress/10858.001.0001>
- JustAnotherArchivist. (2020). *JustAnotherArchivist/snsrape* [Python]. <https://github.com/JustAnotherArchivist/snsrape> (Original work published 2018)
- Lewis, R., Marwick, A., & Partin, W. C. (2020). "We Dissect Stupidity and Respond to It": Response Videos and Networked Harassment on YouTube. SocArXiv. <https://doi.org/10.31235/osf.io/veqyj>

Marwick, A. E., & Caplan, R. (2018). Drinking male tears: Language, the manosphere, and networked harassment. *Feminist Media Studies*, 18(4), 543–559.

<https://doi.org/10.1080/14680777.2018.1450568>

Winter, R., & Salter, A. (2019). DeepFakes: Uncovering hardcore open source on GitHub. *Porn Studies*, 0(0), 1–16.

<https://doi.org/10.1080/23268743.2019.1642794>