ASSESSING THE IMPACT OF GLOBAL ATTENTION ON SUBREDDIT COMMUNITY PRACTICES: THE CASE OF /R/HONGKONG

Dmitry Kuznetsov
The Chinese University of Hong Kong

Milan Ismangil
Utrecht University

Introduction

What happens to internet communities after a global media event puts them in the public spotlight? Does the flood of new members change a given community's practices, structures, and discussions? Do things go back to normal? These questions lie at the heart of our research project, which examines how a local subcommunity on the popular website Reddit changed as its matter of focus became a global discussion subject.

Specifically, this study analyses changing posting practices on /r/hongkong, a local subreddit whose popularity skyrocketed in 2019, with the Anti-Extradition Law Amendment Bill Movement garnering worldwide media coverage. Now, with global attention shifting away from the protests and the 2020 Hong Kong National Security law, the subreddit no longer exhibits the same level of activity (Figure 1). But what can be learnt beyond simply looking at the numbers? Can a qualitative change be observed on /r/hongkong?

In this extended abstract, we will briefly examine the existing research on subreddits as a community, consider the potential significance of media events and subsequent influxes of new users for community practice, outline our methodological approach, and highlight some findings.

Reddit and Community

Reddit is one of the world's most popular websites, combining features of both Web 2.0 social media platforms (networked groups, algorithmic newsfeeds) and early internet UseNet groups, involving volunteer moderation, subject specificity, and (relative)

anonymity (Squirrell, 2019). This duality of Reddit can be seen in its organisation. While much of the website (barring several private or quarantined areas) is open to all users for browsing and contribution, Reddit is structured around the establishment of interest groups focusing on particular topics congregating in specialised, moderated subreddits.
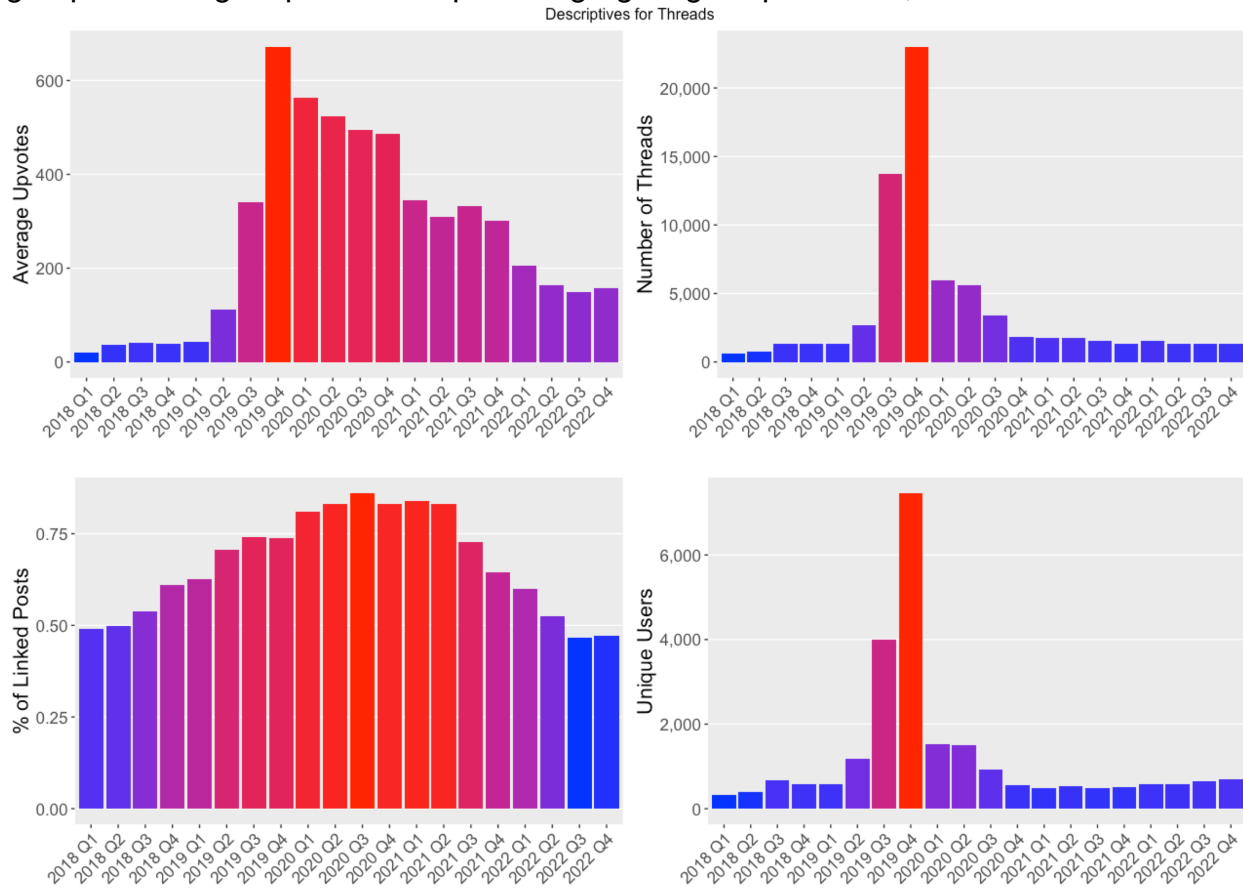

Figure 1 Threads Per Quarter on /r/hongkong; 2017-2023

This "vague" (Datta & Adar, 2019) group membership system is not based on explicit boundaries. Authors such as Squirrell (2019) highlight the importance of platform affordances (flairs; stickies; FAQ) and volunteer moderation in facilitating focused discussions in which participants exhibit member behaviour (Datta & Adar, 2019). This behaviour may be observed in their adherence to themes and rules of a given subreddit (Lê et al., 2020), in the development of collective efforts (Glassman & Kuznetcova, 2022), or traced through examination of posting behaviours (Soliman et al., 2019; Staudt Willet & Carpenter, 2020) and linguistic features (Albota, 2020; Bhandari & Armstrong, 2019).

Thus, in our discussion of community, we align closely with the work of Lyon (1987) who argues that communities are constructions featuring a trinity of common elements: common ties, shared place, and distinctive social interaction. Extending this thought further, we argue that the three elements are all mutually constitutive, and that shared place is itself a product of social interaction, rather than simply a setting for it (Massey,

1994). This is particularly important in so-called "porous" environments (Song, 2009), where it is easy for individuals to drop and drop out of conversation.

This is visible in Squirrell's (2019) investigation of reddit moderation, a task combines the complexity simultaneously enforcing community-rules and preventing spam behaviour, while also maintaining "atmosphere and epistemic integrity" (p. 1923) crucial for the sustainability of virtual communities. Thus, we argue that in virtual communities such /r/hongkong, a sudden influx of users should have a quantitative and qualitative impact, as only one aspect of Lyon's(1987) conceptualisation remains relatively constant: shared place. We are therefore particularly interested in locating changes in content and user behaviour on the subreddit when, following a media event, the population and number of posts increase exponentially.

Media Events

Our analysis of /r/hongkong aims to explore the changes in posting behaviour and content since the establishment of the subreddit. We focus on examining the impact of two years: 2014 and 2019. Both saw major social movements in Hong Kong that garnered sustained worldwide media attention. We classify these two periods as extended media events, building on the work of Dayan and Katz (1992). Since its publication, television has had to contend with the internet and social media, contributing to an age of increasingly atomized media.

Even so, media events still matter and seem to be sustained over longer periods. For example, we can consider the 2022- Ukraine War as one sustained media event, with continuous coverage conducted by both professional and amateur media organisations, leading to the formation of an almost constant viewer constituency distributed around the globe.

If a more recent term is in order, such events could be described as what Frosh and Pinchevski (2018) call media witnessing, where "every moment becomes pregnant with historical possibilities". The concept of media events, with emphasis on disruption of routine and constitution of committed audiences, guided into participation, offers a useful lens to focus our case study. With the Hong Kong protests of 2019 attracting global media attention, with online platforms helping constitute engaged discussion and activist networks (Wang & Zhou, 2021), we can gauge how this attention affects a virtual community, and what happens once it's all over.

Data & Method

As our goal is to observe the changing nature of /r/hongkong, we gathered longitudinal data to establish a baseline, to highlight how the events of 2019 impacted subreddit activity, and to evaluate whether the subreddit returned to the baseline. To collect our dataset, we used the PMAW python package (Podolak, 2021/2023) to interact with the Pushshift API (Baumgartner et al., 2020). We gathered comments posted on the subreddit from July 2010 to February 2023 (1,322,119). Data was then managed and analysed using R 4.2.2 (R Core Team, 2021) and the package Tidyverse (Wickham et al., 2019) Due to the apparent limitations of Pushshift API, a proportion of the collected

comments were not associated with a URL (217,883). Based on the data that did, we found that since October 2017, a total of 74978 unique threads were posted on the subreddit.

We used BERTopic (Grootendorst, 2022) to analyse the content of the data. We sampled some of the peak moments in user activity and look at the data qualitatively, adopting what Gerbaudo (2016) describes as "data hermeneutics" to enrich our analysis and situate the context of the study beyond quantitative abstraction. Specifically, we focused on stickied threads and subreddit rules.

Findings

First, peak activity is clearly indicated during the 2019 protests, followed by major drops in 2020 and 2021.
The subreddit saw a surge in popularity and user activity in 2019, reflected in higher average upvotes and more active users. The proportion of comments discussing linked posts also varied, suggesting changes in discussion dynamics. The top 100 users' contribution to posts varied significantly over the years, with a high of 50.3% in 2018, a low of 14.3% in 2019, and a gradual recovery to 21.5-27.7% in the subsequent years. The data on threads showed similar patterns, with a dramatic increase in activity in 2019, followed by a drop in 2022. The concentration of posting, however, increased significantly in 2020 and 2021, with the top 100 users contributing 57% and 64.1% of all threads respectively.

Topic modelling showed that discussions on the subreddit moved away from general information about Hong Kong. Starting in Quarter 2 of 2019, the most popular topics being to focus on covering the 2019 Anti-Elab protests. In 2020, discussions of protest-related issues and the National Security Law continued to dominate. In 2021, top topics focused on political issues, also covering covid-19 related concerns. In 2022, the livelihood issues seen in 2018 resurface, but popular discussions remain focused on political topics.

Qualitative analysis showed that prior to the 2019 protests, the subreddit had general rules in line with overall Reddit rules. After the surge in discussion and user base, the subreddit developed a more specific set of rules. The new rules were more prescriptive and aimed at managing the increased volume of posts and maintaining content quality. Before the 2019 protests, sticky threads were often used for community meet-ups or general announcements. The subreddit started using sticky threads for larger discussions (megathreads), and to direct users to specific content.

References

Albota, S. (2020). Linguistic and Psychological Features of the Reddit News Post. *2020 IEEE 15th International Conference on Computer Sciences and Information Technologies (CSIT)*, *1*, 295–299. https://doi.org/10.1109/CSIT49958.2020.9321991

Baumgartner, J., Zannettou, S., Keegan, B., Squire, M., & Blackburn, J. (2020). The Pushshift Reddit Dataset. *Proceedings of the International AAAI Conference on Web and Social Media*, *14*, 830–839. https://doi.org/10.1609/icwsm.v14i1.7347

Bhandari, A., & Armstrong, C. (2019). Tkol, Httt, and r/radiohead: High Affinity Terms in Reddit Communities. *Proceedings of the 5th Workshop on Noisy User-Generated Text (W-NUT 2019)*, 57–67. https://doi.org/10.18653/v1/D19-5508

Curiskis, S. A., Drake, B., Osborn, T. R., & Kennedy, P. J. (2020). An evaluation of document clustering and topic modelling in two online social networks: Twitter and Reddit. *Information Processing & Management*, *57*(2), 102034. https://doi.org/10.1016/j.ipm.2019.04.002

Datta, S., & Adar, E. (2019). Extracting Inter-Community Conflicts in Reddit. *Proceedings of the International AAAI Conference on Web and Social Media*, *13*, 146–157. https://doi.org/10.1609/icwsm.v13i01.3217

Dayan, D., & Katz, E. (1992). *Media events*. Harvard University Press.

Frosh, P., & Pinchevski, A. (2018). Media and events after *Media Events*. *Media, Culture & Society*, *40*(1), 135–138. https://doi.org/10.1177/0163443717726007

Gerbaudo, P. (2016). From Data Analytics to Data Hermeneutics. Online Political Discussions, Digital Methods and the Continuing Relevance of Interpretive Approaches. *Digital Culture & Society*, *2*(2), 95–112. https://doi.org/10.14361/dcs-2016-0207

Glassman, M., & Kuznetcova, I. (2022). The GameStop saga: Reddit communities and the emerging conflict between new and old media. *First Monday*. https://doi.org/10.5210/fm.v27i7.11766

Grootendorst, M. (2022). *BERTopic: Neural topic modeling with a class-based TF-IDF procedure* (arXiv:2203.05794). arXiv. https://doi.org/10.48550/arXiv.2203.05794

Lê, A., Farris, E. M., Sumner, J. L., & Warmka, J. M. (2020). Extra, Extra, Reddit All About it: What Reddit can Teach us About Local Politics. *The Forum*, *18*(1), 87–115. https://doi.org/10.1515/for-2020-1005

Lyon, L. (1987). *The community in urban society*. Temple University Press.

Massey, D. B. (1994). *Space, place, and gender*. University of Minnesota Press.

Podolak, M. (2023). *PMAW: Pushshift Multithread API Wrapper* [Python]. https://github.com/mattpodolak/pmaw (Original work published 2021)

R Core Team. (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Soliman, A., Hafer, J., & Lemmerich, F. (2019). A Characterization of Political Communities on Reddit. *Proceedings of the 30th ACM Conference on Hypertext and Social Media*, 259–263. https://doi.org/10.1145/3342220.3343662

Song, F. W. (2009). *Virtual communities: Bowling alone, online together*. Peter Lang.

Squirrell, T. (2019). Platform dialectics: The relationships between volunteer moderators and end users on reddit. *New Media & Society*, *21*(9), 1910–1927. https://doi.org/10.1177/1461444819834317

Staudt Willet, K. B., & Carpenter, J. P. (2020). Teachers on Reddit? Exploring contributions and interactions in four teaching-related subreddits. *Journal of Research on Technology in Education*, *52*(2), 216–233. https://doi.org/10.1080/15391523.2020.1722978

Wang, R., & Zhou, A. (2021). Hashtag activism and connective action: A case study of #HongKongPoliceBrutality. *Telematics and Informatics*, *61*, 101600. https://doi.org/10.1016/j.tele.2021.101600

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T. L., Miller, E., Bache, S. M., Müller, K., Ooms, J., Robinson, D., Seidel, D. P., Spinu, V., … Yutani, H. (2019). Welcome to the Tidyverse. *Journal of Open Source Software*, *4*(43), 1686. https://doi.org/10.21105/joss.01686