
Volunteer Work: Mapping the Future of Moderation Research

Charles Kiene

University of Washington
ckiene@uw.edu

Eshwar Chandrasekharan

Shagun Jhaver

Georgia Institute of Technology
eshwar3@gatech.edu
sjhaver3@gatech.edu

Joseph Seering

Carnegie Mellon University
jseering@andrew.cmu.edu

Kat Lo

University of California, Irvine
kmlo@uci.edu

Bryan Dosono

Airbnb, Inc.
bryan.dosono@airbnb.com

Kate Grandprey-Shores

Riot Games, Inc.
shor0090@umn.edu

Jialun “Aaron” Jiang

Brianna Dym

University of Colorado, Boulder
aaron.jiang@colorado.edu
brianna.dym@colorado.edu

Sarah Gilbert

University of Maryland
sagilber@umd.edu

Donghee Yvette Wohn

New Jersey Institute of Technology
yvettewohn@gmail.com

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.

CSCW '19 Companion, November 9–13, 2019, Austin, TX, USA

© 2019 Copyright is held by the owner/author(s).

ACM ISBN 978-1-4503-6692-2/19/11.

<https://doi.org/10.1145/3311957.3359443>

KEYWORDS

Content moderation; online communities; governance; volunteer work; platforms; social media.

ABSTRACT

Research on the governance of online communities often requires exchanges and interactions between researchers and moderators. While a growing body of work has studied commercial content moderation in the context of platform governance and policy enforcement, only a small number of studies have begun to explore the work of unpaid, volunteer community moderators who manage the millions of different subcommunities that exist on platforms. This workshop will create a pathway for future scholars to tackle the challenges and opportunities of research on volunteer community moderators and establish best practices for engaging with volunteer moderators without disrupting their work. Through lightning talks, collaborative brainstorming exercises, and small-group activities applying principles to research practice, workshop participants will bring together their diverse experiences and perspectives to map the future of moderation research. Both industry and academic researchers as well as experienced moderators will lead this one-day workshop that may accommodate up to 20 participants.

INTRODUCTION

Volunteer online community moderators play a vital role in maintaining millions of subcommunities on the Internet. These moderators keep their communities operating by managing disputes between community members, moderating inappropriate content, and making decisions that shape the development of their communities. Despite their importance in contributing to the success and safety of online communities, only a handful of studies have examined volunteer community moderator practices, the rules that they make, the resources they draw on to moderate, and the challenges they encounter from social or technological change [3, 6, 7, 9, 12, 13]. More research on volunteer community moderators is needed to understand and support them and their communities, but this growing field must be guided by a set of ethical principles and strategies to inform better research practices while minimizing disruptions to community moderator work and their communities.

BACKGROUND

Moderators face numerous challenges, which can both inform focus for research and guide methods for engaging with moderators. For example, Matias [7] and Seering et al. [10], note both personal and interpersonal struggles moderators face from the long, unpaid hours they frequently work in maintaining their communities to the challenges in maintaining relationships with each other, their communities, and sometimes even platform administrators. Moderator work is mediated by the technologies made available through the platforms, which change frequently and without warning. While some of this work involves moderating inappropriate or off-topic content that's against community rules [10], volunteer community moderators must also manage the emotional toll that

accompanies these tasks, like the exposure to abusive or negative content or dealing with harassment for enforcing community rules [13]. Leaders in volunteer community moderation teams have even more responsibilities, like delegating tasks to other moderators, recruiting and training new moderators, and managing major disruptions to the community [6, 10]. Understanding the nature of this work and respectfully navigating the challenges moderators face is important for the integrity of social computing research. Researchers must navigate the ethical tensions in respect to volunteer moderation work while recognizing moderators' expertise.

With this workshop, we aim to give new researchers in this domain a strong foundation and to build from the expertise of established researchers in generating a shared understanding of the challenges, opportunities, and best practices for engagement in future work in community moderation. This proposed full day workshop will begin with lightning talks and small group exercises, and it will conclude with full-group discussions of future research proposals based on principles established earlier in the workshop. Development of a more nuanced collective understanding of the challenges moderators face will help researchers both to identify the right questions to ask and to engage with moderators respectfully in the pursuit of answers.

WORKSHOP STRUCTURE

Intros & Lightning Talks: Participants will introduce themselves in lightning talks summarizing their proposed future work and relevant experience in the space.

Large Group Discussions: Participants will engage in an open discussion in relation to common themes and challenges highlighted from the lightning talks.

Principle Generation Break-out Groups: Participants will group around similar questions and methods and generate a best-practices-for-researchers-draft and/or generate questions that probe at how to better understand best practices.

Presentations of Results: Groups present proposed study design in a short presentation. Looking Forward: Participants can remotely collaborate on writing best practices and research agenda document.

WORKSHOP ACTIVITIES AND GOALS

This workshop will establish a set of principles for researchers looking to explore moderation and online community practices around safety. We will generate a list of best practices for working with online communities and their volunteer moderators. Second, we will generate a guiding document outlining the opportunities for needed research on volunteer moderators. We also aim to use this workshop to help welcome new researchers to study community moderation. This research space can benefit from the addition of a wider variety of perspectives, both from established faculty with a growing interest in the area and from newer researchers who are looking to find their footing in a supportive community. These themes will guide discussion: *current focuses of volunteer moderation research, recruitment and community access, privacy and data collection, and research in marginalized or vulnerable communities.*

Current focuses of volunteer moderation research. Recently published work has focused most notably on Facebook, Reddit, and Twitch [5–7, 10]. A strong majority of this work has used interview methods [6–8, 10, 11, 13], with a small number using other ethnographic methods [12], quantitative analyses [1, 9], or experiments [8]. This imbalance suggests several questions for exploration – what types of questions in this space are best answered with each methodological approach? What are the challenges of using each method in this particular research space? Has the high volume of qualitative work in this space resulted from the appropriateness of these methods in answering early questions, or is it a result of the backgrounds of researchers who have thus far been involved?

Recruitment and community access. Whether or not a social computing researcher is directly recruiting volunteer moderators, they are likely to encounter them or pass labor onto them during research in this domain. Researchers' methods often take volunteer community moderators away from managing their communities. Additionally, these potential informants may not know how to respond to recruitment messages or what they should consider in determining whether to grant researchers access, as they may not be fully knowledgeable on the practices and ethics of human subjects research [4]. In this workshop we will imagine strategies for future research where moderators are informed and/or recruited in a way that is easy to understand but contains the most important information that volunteer community moderators need to approve community access.

Design interventions and moderation tools. Moderators and community managers are often constrained in the actions they can take to govern their communities as a result of strict technological affordances of the platform. Without effective tools for scaling content moderation, moderators may be limited in their ability to curb norm and rule violating behavior and risk their communities collapsing into disarray. How can we design sociotechnical systems and tools to better support moderation teams in the future? Are automated (e.g., machine learning) tools to handle the challenges of scale on large communities a solution? How do we design these new technologies with respect to fairness, accountability, transparency, and ethics in socio-algorithmic online governance? How should research involving sociotechnical interventions and testing new technologies be carried out in this space?

Privacy and data collection. Following calls for better strategies of informed consent in recruitment messages, future social computing researchers must also pay careful attention to how they collect, store, and share data and identifiers obtained from volunteer community moderators. While potential informants may be comfortable with their usernames being published alongside interview quotes or their communities identified in papers, we have a responsibility as social computing researchers to minimize the risk that may accompany exposure in a future publication. On the other hand, collecting trace data on online communities through web scraping tools may be technically acceptable within the policies of some online platforms that host these communities. How can we be more intentional about gaining consent from volunteer community moderators to collect data? What should volunteer community moderators know about how we record, store, and share their data and identifiers?

Marginalized or vulnerable communities. Moderators of online communities with marginalized or vulnerable groups may feel a sense of responsibility to provide a safe space and to protect members from harassment, hate speech, and other harmful acts more so than in other kinds of online communities. In some cases, this may lead to increased skepticism about the intentions of researchers and the impact that research may have on their communities [2]. What does this mean for studying moderation in

marginalized and vulnerable communities? What experience and/or background should researchers focus on developing prior to beginning work in these spaces?

WORKSHOP DURATION AND PARTICIPANTS

This will be a one-day workshop on November 10, 2019. Participants will submit a position paper proposing a future study investigating online communities managed by volunteer community moderators. These papers are not intended to be fully-developed research plans but rather talking points for the workshop. Participants should be prepared to revise these plans during the workshop based on group discussion. Organizers will review position paper submissions with a focus on identifying a breadth of perspectives and levels of experience in the field. Examples may be research proposals aimed at understanding the challenges of volunteer community moderators of vulnerable online communities, the systems of organization and work of volunteer community moderators across different contexts or sociotechnical systems, or the negotiations between community moderators and other sources of governance such as platform policies or the law. Please email your position paper to communitymoderationworkshop@gmail.com.

Website

More details about this workshop, including deadlines, workshop schedule, and organizer backgrounds, can be found at: <https://sites.google.com/view/cscw2019modworkshop>.

REFERENCES

- [1] Eshwar Chandrasekharan, Mattia Samory, Shagun Jhaver, Hunter Charvat, Amy Bruckman, Cliff Lampe, Jacob Eisenstein, and Eric Gilbert. 2018. The internet's hidden rules: an empirical study of reddit norm violations at micro, meso, and macro scales. *Proc. ACM Hum.-Comput. Interact.* 2, CSCW (Nov. 2018), 32:1–32:25. <https://doi.org/10.1145/3274301>
- [2] Brianna Dym and Casey Fiesler. 2018. Vulnerable and online: fandom's case for stronger privacy norms and tools. In *Companion of the 2018 ACM Conference on Computer Supported Cooperative Work and Social Computing (CSCW '18)*. ACM, New York, NY, USA, 329–332. <https://doi.org/10.1145/3272973.3274089> event-place: Jersey City, NJ, USA.
- [3] Casey Fiesler, Jialun" Aaron" Jiang, Joshua McCann, Kyle Frye, and Jed R. Brubaker. 2018. Reddit rules! Characterizing an ecosystem of governance.. In *ICWSM*. 72–81. <http://cmci.colorado.edu/~cafi5706/icwsm18-redditrules.pdf>
- [4] Casey Fiesler and Nicholas Proferes. 2018. "Participant" perceptions of twitter research ethics. *Social Media + Society* 4, 1 (Jan. 2018), 2056305118763366. <https://doi.org/10.1177/2056305118763366>
- [5] Shagun Jhaver, Sucheta Ghoshal, Amy Bruckman, and Eric Gilbert. 2018. Online harassment and content moderation: the case of blocklists. *ACM Trans. Comput.-Hum. Interact.* 25, 2 (March 2018), 12:1–12:33. <https://doi.org/10.1145/3185593>
- [6] Charles Kiene, Andrés Monroy-Hernández, and Benjamin Mako Hill. 2016. Surviving an "Eternal September": How an online community managed a surge of newcomers. In *Proceedings of the 2016 ACM Conference on Human Factors in Computing Systems (CHI '16)*. ACM, New York, NY, 1152–1156. <https://doi.org/10.1145/2858036.2858356>
- [7] J. Nathan Matias. 2016. Going dark: Social factors in collective action against platform operators in the Reddit blackout. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*. ACM, New York, NY, 1138–1151. <https://doi.org/10.1145/2858036.2858391>

- [8] J. Nathan Matias. 2019. The civic labor of volunteer moderators online. *Social Media + Society* 5, 2 (April 2019), 1–12. <https://doi.org/10.1177/2056305119836778>
- [9] Joseph Seering, Robert Kraut, and Laura Dabbish. 2017. Shaping pro and anti-social behavior on twitch through moderation and example-setting. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing (CSCW '17)*. ACM, New York, NY, USA, 111–125. <https://doi.org/10.1145/2998181.2998277> event-place: Portland, Oregon, USA.
- [10] Joseph Seering, Tony Wang, Jina Yoon, and Geoff Kaufman. 2019. Moderator engagement and community development in the age of algorithms. *New Media & Society* (Jan. 2019), 1461444818821316. <https://doi.org/10.1177/1461444818821316>
- [11] Jhaver Shagun, Iris Birman, Eric Gilbert, and Amy Bruckman. 2019. Human-Machine Collaboration for Content Regulation: The Case of Reddit Automoderator. In *ACM Transactions on Computer-Human Interaction (TOCHI 2019)*. <https://doi.org/10.1145/3338243>
- [12] Janet Sternberg and Yao Sun. 2014. Misbehavior in Cyber Places: The Regulation of Online Conduct in Virtual Communities on the Internet. *International Journal of Communication* 8, 0 (June 2014), 3. <https://ijoc.org/index.php/ijoc/article/view/2906>
- [13] Yvette Wohn. 2019. Volunteer moderators in twitch micro communities: how they get involved, the roles they play, and the emotional labor they experience. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI'19)*. ACM, Glasgow, UK. <https://doi.org/10.1145/3290605.3300390>