Extended: Opportunities for Supporting Community-Scale Communication

Cristian Danescu-Niculescu-Mizil
Cornell University

Abstract

When people collaborate in large groups, they face several communication challenges when coordinating their work. Some challenges come up during the actual work—like knowing what others are doing and coordinating everyone's efforts. Other challenges happen before the work even begins, such as setting goals and deliberating on consequential choices. In this work we will explore opportunities for automated communication support by conducting a mixed-methods exploration of communication challenges arising in large-scale collaborations and by prototyping tools aimed at assisting collaborators with these challenges.

Introduction

Most instances of important infrastructure and knowledge created in our society come from people working together in a collaborative process. Take for example Wikipedia, where a large community coordinates work and plans edits and updates (Halfaker et al. 2012). Or, software development, where people come together to asynchronously collaborate on writing code.

When people collaborate in large groups, they face several challenges in coordinating their work and communicating effectively. Some challenges come up during the actual work—like knowing what others are doing and coordinating everyone's efforts. Other challenges happen before the work even begins, such as setting goals and deliberating on consequential choices. However, technological support for these collaborative processes is currently limited to content creation (e.g., generating text and tracking changes). There are no current support tools that can help manage the communication process involved.

We envision a socio-technical framework for supporting the communication process involved in community-scale collaborations. To develop this framework, in this work we propose a mixed-methods exploration of communication challenges surrounding collaboration on Wikipedia and of potential technological solutions that can help collaborators with these challenges.

In the first year of the project we will focus on understanding the opportunities for automated communication support. We will start with an analysis of existing communication data on community-wide deliberation spaces—such as those concerning Articles for Deletion (Mayfield a and Black 2019)—to identify common communication challenges. Building on these observational insights, we will design structured interviews for active Wikipedia editors, aimed at understanding what they perceive to be the most challenging aspects of communication, how they currently approach these challenges, and what they perceive as opportunities for automated communication support.

Having identified key opportunities together with the community members, in the second year of the project, we will engage in a participatory design of the prototype communication support tools. We will iterate on the design of these prototype tools together with the community members and rely on user-studies to examine how the assistance provided by these tools might fit into the editor's workflow and the extent to which they are effective in improving the collaboration process.

Impact. This project will generate both theoretical and practical contributions to the understanding and design of communication support tools for large-scale, collaborative communities. Empirically, we expect to produce a detailed characterization of communication challenges that arise in Wikipedia's collaborative workflows, especially in deliberative spaces such as Articles for Deletion. These insights will inform a broader taxonomy of breakdowns and needs that emerge in asynchronous, community-scale decision-making. Practically, the project will result in the co-designed prototypes of socio-technical tools aimed at supporting editor interactions. Through iterative design and user studies, we anticipate uncovering design principles for embedding communication assistance in editor workflows in a way that respects community norms and values. Ultimately, the project aims to lay the groundwork for scalable interventions that improve not just the efficiency of collaboration, but also its inclusiveness and deliberative quality—offering a model that can be extended beyond Wikipedia to other civic and open-source platforms.

Dates: This work proposed here will split over two years, as described above. Starting in August 2025 and continuing until August 2027.

Related work

Collaboration on Wikipedia has been widely studied as a paradigmatic example of large-scale, decentralized knowledge production. While the platform's open-editing model enables broad participation, it also presents significant communication challenges that impact the quality, efficiency, and inclusiveness of collaboration.

A foundational concern is the difficulty of coordinating among contributors with differing levels of expertise, engagement, and social capital. Kittur and Kraut (2008) noted that while Wikipedia allows for contributions from many users, only a small core of editors are deeply involved in coordinating and sustaining article development. This asymmetry can lead to misunderstandings and bottlenecks in communication, especially when newcomers lack the contextual knowledge or social signals to navigate norms effectively Halfaker et al. (2013).

Our own work introduced the largest collection of Wikipedia Talk-Space conversations (WikiConv; Hua et al 2018). This dataset was developed with the prior support from the Wikipedia Fund. We analyzed this collection to reveal communication dynamics surrounding wikipedia admin elections, including linguistic coordination (Danescu-Niculescu-Mizil et al. 2012) and politeness (Danescu-Niculescu-Mizil et al. 2013).
We have demonstrated the mixed-methods workflow in studying moderation challenges on Wikipedia and proposing potential technical support for moderation. We explored the editor's reaction to strict moderation practices (Chang et al 2019). Starting from these data-based observations we conducted a mixed method study to understand the proactive moderation practices and explore potential for algorithmic support (Schluger et al 2022). In this work plan to adopt a similar mixed-methods approach to explore communication support for community-scale collaborative practices.

Methods

Data analysis. For the data analysis section we will focus on the large-scale deliberation spaces included in WikiConv (Hua et al 2018). We will use the same methodology we used to develop this dataset to update it with recent conversations, with a focus on those involving community-wide discussions and deliberations. We will use existing methods for analyzing conversational dynamics that we have developed in previous work (such as politeness,

coordination, etc.) and introduce new tools that are specifically geared towards capturing collaboration issues (hesitance, groupthink, deliberation deadlock).

Interviews and surveys. Similar to our prior work on exploring algorithmic support for proactive moderation, we will use structured interviews with active Wikipedia editors. Initial interview questions will be developed around three main themes:

1) Types of communication breakdowns challenges;
2) Strategies currently used by editors to overcome these challenges;
3) Attitudes toward automation and tool support.

Prototypes. The exact design of the prototype communication support tools will depend on the specific support opportunities revealed by our data analysis and from the interviews. We will rely on our expertise in co-designing user-facing communication support tools together with community stakeholders (Schluger et al 2022; Chang et al. 2022). We exemplify below one potential design for a tool addressing a specific community-scale communication challenge that was identified in prior literature: overcoming deliberation deadlocks.

Overcoming deliberation deadlocks.
Deliberations on important and potentially contentious choices are an essential part of the collaboration process (Kittur et al. 2007; Billings et al. 2010). We will prototype tools to assist editors engaged in deliberation to understand and overcome deadlocks, while preserving or even improving the existing sense of community. One key challenge is identifying when progress has stalled, meaning no new points have been recently raised. To detect such moments, the tool will analyze the dynamic both at the literal level—extracting the semantics of propositions made to analyze their novelty—and at the meta-level, leveraging evidence of frustration or disengagement on the part of participants to detect a problem.

One concrete support action for helping editors overcome these deadlocks is sharing a summary of everyone's positions at that point. We envision producing new kinds of summaries that are contextual in the sense of manifesting the motivations and principles each editor has behind their position (drawing from our work on summarizing conversation dynamics; Hua et al. 2024). The tool will work with each user separately to ensure their respective summary is correct, so that inaccuracies or unhelpful connotations can be prevented, and also so that users can have a chance to see how they are perceived by others. An interesting research question that emerges is whether the contextual summary needs to be broadcast to the group, or whether providing contextual feedback just to the individual is more effective

Expected output

Scientific Publications
We will publish our results in conferences related to Human-Computer Interaction, Computer-Supported Cooperative Work, and Computational Social Science (e.g., CHI, CSCW, ICWSM).
These publications will contribute new empirical and theoretical knowledge about communication challenges in large-scale, asynchronous collaboration. Findings will help advance scholarship on socio-technical systems, coordination, and online deliberation.

Public Datasets and Annotated Corpora
We will supplement existing talk-space corpora (ConvoWizard) with community-wide discussions and deliberations (e.g., structured discussions from Wikipedia's Articles for Deletion). These will benefit social scientists, NLP researchers, and HCI researchers. These will support future research on collaborative communication, dialogue modeling, and online

governance. We have a strong track record of distributing data through our ConvoKit toolkit (Chang et al 2018).

Communication Support Tool Prototypes
Designers and developers of collaborative platforms (e.g., Wikimedia Foundation, open-source community managers), as well as HCI researchers. The prototypes will serve as proof-of-concept for tools that assist with collaborative planning, disagreement resolution, and goal-setting in online communities. These prototypes can inform future feature development or integrations within platforms like Wikipedia.
These will be open-sourced under a permissive license with technical documentation and design rationale.

Community dissemination
We will communicate our findings to the Wikipedia community through transparent, accessible, and participatory channels. This includes sharing summaries of key insights and updates on project progress via relevant community forums, such as Village Pump discussions, WikiProject pages, and the Wikimedia research newsletter. We will also host dedicated feedback sessions and participate in community events like Wikimania and WikiCon to present findings, gather input, and co-refine our interpretations. All outputs—including datasets, design frameworks, and tool prototypes—will be shared publicly with clear documentation and invitations for community feedback to ensure alignment with editor needs and values.

Community impact plan

We will engage with the community transparently from the outset, prioritizing reciprocity and mutual benefit. This includes early outreach on relevant Wikipedia forums, sharing preliminary findings, and involving editors in shaping research questions. We will adhere to established community norms, obtain feedback through iterative engagement, and ensure that any tools or findings are communicated clearly and accessibly. We will open up our code for the prototypes and engage the community in their development.

We will share our evaluation of the potential long-term impacts of the communication support tools with the community. The decision of whether to broadly deploy communication support tools that result from this work will be left to the community.

Risks

A key component of this project involves working closely with Wikipedia editors through interviews, participatory design, and tool prototyping. Gaining trust and sustained engagement from a volunteer-based community can be challenging, especially if members are skeptical of academic or technological interventions.
We will engage in community workshops (e.g., WikiCon, Wikimania) to assess community interest and recruit participants. We will focus on discussing the tradeoffs between using technology that follows a well structured framework versus non-centralized ad-hoc solutions.
Another risk is that prototypes developed in the project may not easily integrate into Wikipedia's existing technical infrastructure or may face resistance if perceived as disruptive. We will focus on lightweight, modular tools that can function independently or as browser extensions, reducing the need for platform-level integration. By involving Wikimedia developers and power users early in the design process, we can align with existing workflows and prioritize features that support, rather than replace, community-led practices.

Evaluation

The evaluation of this project will occur in two main phases: the *diagnostic phase*, focused on validating our understanding of communication challenges, and the *prototype phase*, focused on assessing the usefulness, usability, and impact of the proposed communication support tools.

Evaluation during the diagnostic phase ensures that the insights we generate are accurate, meaningful to community members, and useful for informing tool design. We will cross-reference findings from discussion data analysis with insights from structured interviews to verify consistency. We will monitor for thematic saturation in our structured interviews with Wikipedia editors—that is, the point at which additional interviews stop yielding substantially new insights. To ensure that our interpretations reflect the lived experiences of editors, we will present preliminary themes and taxonomies of communication challenges back to community members.

In the *prototype phase* we will test the usability of the support tools using think-aloud protocols and task-based evaluations with active Wikipedia editors, we will assess how intuitive and non-disruptive the tools are within real workflows. We will conduct field deployments of early prototypes (e.g., as browser extensions or third-party tools) and study how they integrate into day-to-day editing practices. We will hold iterative design and review sessions with Wikipedia editors and moderators to gather structured feedback, prioritize design revisions, and co-refine the tools. After tool usage, we will survey and interview participants to understand perceived changes in communication clarity, efficiency, inclusiveness, and civility.
To understand the broader potential for adoption and long-term impact of the communication support tools we will conduct quantitative analysis of discussion thread dynamics (e.g., thread length, conflict markers, resolution indicators) before and after tool use. We will also solicit feedback from community leaders and Wikimedia affiliates to assess the relevance and sustainability of the interventions.

Budget

The budget is included here: https://docs.google.com/spreadsheets/d/1aF72je7A3PEiORHA_3gLG0y9dtwCGIHy9V_Tk4kZyV0/edit?usp=sharing

References

Billings, Matt, and Leon A. Watts. "Understanding Dispute Resolution Online: Using Text to Reflect Personal and Substantive Issues in Conflict." In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 1447–56. Atlanta Georgia USA: ACM, 2010. https://doi.org/10.1145/1753326.1753542.

Chang, Jonathan P., and Cristian Danescu-Niculescu-Mizil. "Trajectories of Blocked Community Members: Redemption, Recidivism and Departure." In *Proceedings of WWW*, 2019. https://doi.org/10.1145/3308558.3313638.

Chang, Jonathan P., Charlotte Schluger, and Cristian Danescu-Niculescu-Mizil. "Thread With Caution: Proactively Helping Users Assess and Deescalate Tension in Their Online Discussions."

In *Proceedings of CSCW*, 545:1-545:37, 2022. https://doi.org/10.1145/3555603.

Danescu-Niculescu-Mizil, Cristian, Lillian Lee, Bo Pang, and Jon Kleinberg. "Echoes of Power: Language Effects and Power Differences in Social Interaction." In *Proceedings of WWW*, 2012. https://doi.org/10.1145/2187836.2187931.

Danescu-Niculescu-Mizil, Cristian, Moritz Sudhof, Dan Jurafsky, Jure Leskovec, and Christopher Potts. "A Computational Approach to Politeness with Application to Social Factors." In *Proceedings of ACL*, 2013. https://www.aclweb.org/anthology/P13-1025.

Halfaker, Aaron, R. Stuart Geiger, Jonathan T. Morgan, and John Riedl. "The Rise and Decline of an Open Collaboration System." *American Behavioral Scientist* 57, no. 5 (May 2013): 664–88. https://doi.org/10.1177/0002764212469365.

Halfaker, Aaron, Bryan Song, D. Alex Stuart, Aniket Kittur, and John Riedl. "NICE: Social Translucence Through UI Intervention," 101–4, 2011. https://doi.org/10.1145/2038558.2038575.

Hua, Yilun, Nicholas Chernogor, Yuzhe Gu, Seoyeon Jeong, Miranda Luo, and Cristian Danescu-Niculescu-Mizil. "How Did We Get Here? Summarizing Conversation Dynamics." In *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)*, edited by Kevin Duh, Helena Gomez, and Steven Bethard, 7452–77. Mexico City, Mexico: Association for Computational Linguistics, 2024. https://doi.org/10.18653/v1/2024.naacl-long.414.

Hua, Yiqing, Cristian Danescu-Niculescu-Mizil, Dario Taraborelli, Nithum Thain, Jeffery Sorensen, and Lucas Dixon. "WikiConv: A Corpus of the Complete Conversational History of a Large Online Collaborative Community." In *Proceedings of EMNLP*. EMNLP, 2018. https://www.aclweb.org/anthology/D18-1305.

Kittur, Aniket, and Robert E. Kraut. "Harnessing the Wisdom of Crowds in Wikipedia: Quality Through Coordination." In *Proceedings of CSCW*, 37–46, 2008. https://doi.org/10.1145/1460563.1460572.

Mayfield, Elijah, and Alan W Black. "Analyzing Wikipedia Deletion Debates with a Group Decision-Making Forecast Model" 3 (2019): 26.

Schluger, Charlotte, Jonathan P. Chang, Cristian Danescu-Niculescu-Mizil, and Karen Levy. "Proactive Moderation of Online Discussions: Existing Practices and the Potential for Algorithmic Support." *Proceedings of the ACM on Human-Computer Interaction* 6, no.

CSCW2 (November 11, 2022): 370:1-370:27. https://doi.org/10.1145/3555095.

"Understanding Dispute Resolution Online | Proceedings of the SIGCHI Conference on Human Factors in Computing Systems." Accessed April 17, 2025. https://dl.acm.org/doi/10.1145/1753326.1753542.