The Anatomy of Coordination among Wikipedia Users

Grantees
Marco Minici*, Giuseppe Manco
National Research Council of Italy

Cristian Consonni†
European Commission, Joint Research Centre

Abstract

Wikipedia's large-scale, decentralized model has revolutionized knowledge creation, yet how its editors coordinate—and how this affects content quality and community health—remains insufficiently understood. This project will explore the patterns, behaviors, and networks that emerge as Wikipedia editors collectively build and refine articles. By examining various subject areas, from controversial socio-political issues to less divisive topics, we will investigate whether editorial coordination arises spontaneously or through more orchestrated efforts, and assess its impact on editor retention, diversity, and content integrity.

Our methodology integrates literature review, dataset construction, and advanced network analysis. We will identify potential subgroups of editors showing unusually high levels of coordination—either beneficial or malicious—and evaluate how these activities influence Wikipedia's mission. Findings will inform Wikimedia's strategic objectives, supporting efforts to enhance community engagement, safeguard against misinformation, and advance the Wikimedia strategic goals of "Knowledge as a Service" and "Knowledge Equity."[1]

Introduction

Wikipedia stands as one of the most prominent examples of large-scale, decentralized knowledge creation projects, yet our understanding of how contributors coordinate—and how coordination shapes article quality and community health—remains incomplete. This research project will investigate the underlying patterns, behaviors, and networks of collaboration that emerge as Wikipedia editors work together to create and curate content.

While coordinated behavior has been thoroughly investigated on social media platforms such as Facebook (Giglietto et al., 2020), and X, as well as on online forums like Reddit and 4chan (Baele et al., 2021) —both in the context of top-down manipulation (e.g., state-sponsored disinformation campaign) (Douek, 2021) or spontaneous bottom-up actions (e.g., Reddit GameStop bet or r/place collaborative art) (Mancini et al., 2022; Rappaz et al., 2018)—Wikipedia has not received similar attention, despite its fundamentally distributed

[1] https://meta.wikimedia.org/wiki/Movement_Strategy
* Primary author
† This author contributed to writing the proposal, but will not receive any funding.

operations. This gap in understanding represents a critical oversight in how Wikipedia functions and sustains its content integrity. Analyzing coordinated actions within Wikipedia, whether they arise organically (e.g., in large-scale community projects) or take on more orchestrated forms (often described as "brigading" in other contexts), can yield valuable insights. These insights, in turn, have the potential to support the Wikimedia Foundation in advancing its mission and strategic objectives.

Our proposal will address several fundamental research questions:

- What does the distribution of coordinated behavior among Wikipedia users look like?
- Do coordination patterns remain consistent across time, languages, and topics?
- How do tangible outcomes—such as editorial participation, the use of reliable sources, conflict escalation, and diversity of editorial backgrounds—relate to different degrees of coordinated activity?

We think our proposal will benefit both the multi-generational strategy and the 2030 strategic direction of the Wikimedia movement.

Multigenerational strategy

Fuel volunteer growth. We aim to examine how coordinated activity may correlate with two key factors: editorial participation and diversity of contributor backgrounds. Drawing from evidence on other platforms, we hypothesize that coordinated efforts are frequently concentrated in small, tight-knit groups, which may in turn foster stronger member retention and more persistent engagement. At the same time, these same groups could discourage newcomers or outside contributors, thereby reducing diversity. Consequently, we will investigate whether the patterns of coordinated activity remain consistent over time, across multiple languages, and in varied subject areas.

Deliver trustworthy encyclopedic content above all. On many social platforms, coordinated actions are often linked to disinformation or the manipulation of narratives. In Wikipedia's context, understanding whether unusually high levels of coordination might be tied to misleading editing is crucial. This aligns with the strategic priority of "Knowledge Integrity" identified by the Wikimedia Research Team.[2] Enhanced awareness of such behaviors can strengthen community oversight and support the auditing activity of CheckUsers by pinpointing suspicious editorial patterns. Furthermore, the growing prevalence of generative AI introduces new challenges: (1) the cost of producing misleading content is falling rapidly compared to the effort required for content verification (Tai et al., 2025); and (2) because Wikipedia serves as pre-training corpus for Large Language Models, it has become a prime target for malicious actors seeking to influence the outputs of LLMs (Floréane, 2025). By analyzing coordination within Wikipedia, we aim to propose evidence-based strategies for safeguarding the project's content and reputation.

2030 Strategic Direction

Knowledge as a service. This research project will directly support the Wikimedia Foundation's vision of delivering "Knowledge as a Service" by producing insights toward understanding whether content quality may be hindered by abnormal, highly-coordinated users. Ultimately, this strengthens Wikipedia's

[2] https://research.wikimedia.org/knowledge-integrity.html

role as a trusted source that can be used as a source of reliable information.

Knowledge equity. Our research also advances the goal of "Knowledge Equity" by investigating how different communities of editors collaborate, and whether certain patterns of coordination hinder equitable participation. Identifying potential barriers—such as closed editing circles—paves the way for more inclusive editorial processes. We will relate our findings to previous research studying community dynamics and participation in Wikipedia (Neff et al., 2013). Through a deeper understanding of coordination, we aim to help foster a Wikipedia community where all contributors, regardless of background, are empowered to create, refine, and share free knowledge.

In summary, our research proposal centers on the following questions:

- **RQ1**: What does the distribution of coordinated behavior in Wikipedia look like? Does it remain stable over time, across different languages and topics?

- **RQ2**: Is there any correlation between coordinated activity and knowledge integrity? Does this correlation change if we focus on outliers of coordination? Does coordinated activity affect the proportion of reliable sources differently than non-synergistic activity?

- **RQ3**: What is the impact of coordination on editorial activity? Does it foster user retention? Does it provide a trade-off between intra-community user retention and out-community barrier?

In the remainder of this document, we will discuss related work, outline our methods and expected outputs, address potential risks, describe community involvement, present our evaluation plan, and provide a proposed budget**.**

Date: Project kick-off: 2025-09-01 / Project end: 2026-09-01

Methods

Phase 1: Literature Review

We will begin by conducting a thorough review of existing literature on coordinated behavior in social media, online communities, and related platforms. This process will help us identify gaps in current knowledge and shape our approach by proposing novel analyses that address unresolved questions. In particular, we will investigate how coordinated behavior can lead to both positive and negative outcomes for online platforms, with a focus on core Wikipedia objectives such as user acquisition, retention, and knowledge integrity.

Phase 2: Data collection

The initial step in investigating coordinated activity on Wikipedia involves assembling datasets that will allow for meaningful analysis. To address our research questions, we plan to examine two or three major topical areas: one centered on controversial subjects (e.g., socio-political issues), and others featuring more established or less ideologically charged topics. Within each topical area, we will identify 10–30 relevant Wikipedia pages—such as the Russo-Ukrainian War or the Gaza–Israel conflict for controversial themes, and fundamental concepts like the Pythagorean Theorem or Newton's laws of motion for less contentious discussions.

We will acquire editorial activity data for each selected page through the Wikipedia APIs or

Wikipedia data dumps, covering a period of 5-10 years. This time frame includes editorial activity before the widespread adoption of generative AI tools.

For *RQ2*, we will also compile datasets capturing both malicious and organic editing, filling a notable gap in existing research. Specifically, we will (1) identify known malicious edits aimed at spreading disinformation, using open-source data from NGOs and watchdog groups, and (2) collect corresponding organic edits of the same pages during the same timeframe. Open-source datasets—such as the "Pravda" campaign (Floréane, 2025) released by CheckFirst—will serve as a starting point. We likewise plan to collaborate with the Foreign Information Manipulation Interference Institute of the European Union[3] to gather further data and share findings. These partnerships are currently under development, and could provide a valuable framework for validating datasets and methodologies. Throughout this process, we will interact with the community to be aligned with the overarching goals of Wikipedia. We will also align our research practices to the latest recommendations to protect the privacy of Wikipedia editors (Asikin-Garmager et al., 2025). For this purpose, we consider in our budget the support to participate and travel to Wikimania 2026 and the itWikiCon 2025 that will be held in Paris and Catania, respectively. We will also seek guidance and support from the Wikimedia Research team on how to handle our findings, especially for the most controversial topics.

Phase 3: Analysis

Leveraging the insights from our literature review and the data collected, we will conduct targeted analyses designed to answer our research questions.

Similarity Network Construction. A key approach to identifying coordinated activity is to build similarity networks based on predefined behavioral traces, allowing us to visualize and analyze the complexities of synergistic behavior within the platform. Since Wikipedia editors often rely on external sources, we will focus on link insertion and link removal traces to characterize coordination levels in editorial activity. To achieve this goal, we will build upon established methods on coordinated behavior analysis—e.g., Luceri et al., 2024; Pacheco et al., 2021. We provide a visual representation of the process to construct similarity networks in Figure 1, the steps are the following:

1. *Bipartite User-URL Networks.* We will construct two bipartite networks—one for link insertion and one for link removal—where users are connected to the URLs they either insert or remove. Rows in the network's adjacency matrix represent users, while columns represent URLs.

2. *TF-IDF Representation and Filtering*. Drawing on established methods , we will apply Term Frequency–Inverse Document Frequency (TF-IDF) to the user-URL matrix. To avoid skewing results toward overly frequent URLs, we will set a maximum "document" frequency at the 90th percentile, while using a minimum threshold of five occurrences per URL.

3. *User-to-User Similarity*. We will transform the bipartite user-URL network into a user-to-user similarity network by computing the cosine similarity between TF-IDF vectors for each pair of users, thereby generating an edge weight that reflects the degree of coordination.

[3] https://fimi-isac.org/

This process will yield two similarity networks—one for link insertion and one for link removal—which we may eventually merge (Luceri et al., 2024). Our approach can be flexibly applied at varying levels of granularity (e.g., pages, topics, time windows) to enhance our understanding of editorial coordination.

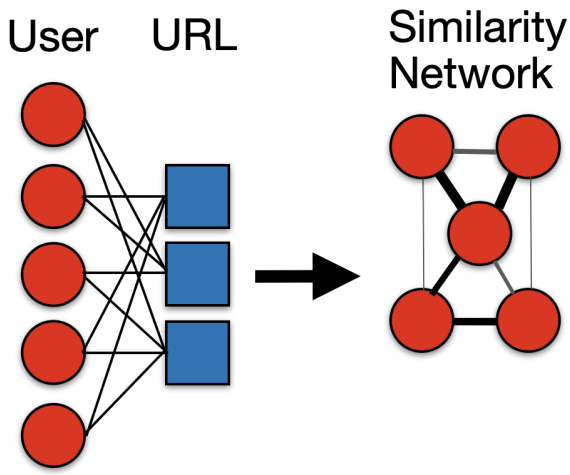

Figure 1. Creation of a similarity network from behavioral traces based on user-URL co-occurrences. We could devise two different similarity networks, one based on link insertion and the other on link removal and possibly fusing them in a unique representation following Luceri et al., 2024.

Similarity Network Analysis. After creating these similarity networks, we will examine various topological indicators to characterize coordinated activity. Metrics such as degree distribution, node centrality, and node strength (since our networks are weighted) will shed light on whether coordination is broadly dispersed among the Wikipedia user base or focused within smaller user clusters. We will also employ community detection and dense substructure mining algorithms (e.g., k-core decomposition) to identify closely knit groups of coordinated users.

By analyzing similarity networks at different levels (e.g., topic, page, time, language), we aim to pinpoint highly coordinated subgroups of users and investigate longitudinal patterns in their editorial participation.

The impact of coordinated activity. Once each user's level of coordination is established, we will assess the impact of that coordination on editorial outcomes from both individual and collective perspectives:

3. *Individual Outcomes***.** We will examine whether a user's degree of coordination at time "X" correlates positively or negatively with activity at time "X + t". We will focus our analysis over short time frames (<1 month) or we will de-correlate general trends in the case of longer periods. We will also investigate whether highly coordinated users are more likely to promote reliable sources, potentially using external validators like the Wikipedia list of Reliable/Perennial Sources[4] to gauge source quality. We are also interested to understand the impact on underrepresented groups and will analyze retention and churn effects on editors who publicly disclosed offline characteristics such as gender, native language or country of origin.
4. *Collective Outcomes.* We will investigate whether pages with substantial editorial activity from coordinated users become less accessible or welcoming to contributors outside of these groups. This is essential for understanding whether equitable access to knowledge building is maintained. We will also examine whether heightened levels of coordination lead to more frequent conflicts, such as edit wars.

[4] https://en.wikipedia.org/wiki/Wikipedia:Reliable_sources/Perennial_sources

These analyses will help us address **RQ1** and **RQ3** comprehensively.

Coordinated activity and Knowledge Integrity. Concerning RQ2, we will construct similarity networks for the specialized datasets gathered (see Phase 2). Our goal is to measure the effectiveness of unsupervised methods—such as edge filtering (e.g, removing edges with low weights) and node pruning (e.g., removing nodes with high eigenvector centrality)—in detecting malicious or suspicious behavior within these networks (Cinus et al., 2025). By comparing link insertion and link removal patterns (and their fused representation, following Luceri et al., 2024), we may identify pages at risk of being polluted by covert malicious actors, thus compromising Wikipedia's knowledge integrity.. This analysis could provide actionable insights for flagging suspicious accounts and inform programs like "CheckUsers."

Overall, our multi-phase methodology integrates literature review, data collection, and comprehensive network analyses to shed light on how coordinated user behavior in Wikipedia may influence editorial participation, content quality, and the broader goal of knowledge integrity.

Expected output

We expect our efforts to bring the following tangible results:

- 1 scientific publication in a Q1 journal or A, A* computational social science conference, e.g., ICWSM,CSCW, TheWebConf or related.
- 1 dataset paper related to coordinated activity detection.
- Relevant insights to inform the community on the benefits and risks of coordinated activity toward the strategic goal of the Wikimedia Foundation.
- As a by-product, our project will stimulate further investigations on similar topics from other scholars in the academic community.

Risks

Because our proposal is centered on computational social science rather than purely technical objectives (e.g., surpassing the current state of the art in algorithmic performance), we anticipate relatively low risk. The results of our investigation will be informative regardless of the specific outcomes we discover.

One potential concern, however, involves *RQ2*, which includes the development of an auditing methodology to detect malicious activity on Wikipedia. While this effort aligns with the Wikimedia Research Team's strategic priority of "Knowledge Integrity" and the "CheckUsers" initiative, it may be perceived by some as an attempt to suppress specific viewpoints or undermine Wikipedia's "good-faith" principle. To mitigate such concerns, we will collaborate closely with experienced Wikipedia editors, ensuring that our analyses support community values rather than conflict with them. We will create concrete opportunities to gather relevant feedback by sharing our work at public gatherings of the Wikipedia community such as Wikimania 2026 and itWikiCon 2025 (considered in our budget).

From a technical standpoint, our method is content-agnostic: we treat links as atomic entities without examining their underlying content. Our approach can also be applied to any Wikipedia language edition and across languages. This ensures that our analysis focuses solely on coordination patterns, independent of the actual viewpoints or topics involved.

Community impact plan

Building on the risk mitigation strategies noted above, we intend to work closely with experienced Wikipedia contributors to ensure our research aligns with the values of the Wikipedia community. We will maintain regular communication with editors throughout the project, using scheduled meetings and public gatherings to discuss progress, share findings, and integrate community feedback. We will place particular emphasis on understanding the needs of "CheckUsers," exploring how our analytical methods can assist them in overcoming existing challenges and achieving their objectives more effectively.

Evaluation

Because this grant is oriented toward foundational research rather than the development of specific operational tools, we propose two primary criteria for evaluating our work:

1. **Academic Contribution:** The potential to achieve high-impact publications (e.g., in Q1 journals or A/A* conferences) that will raise the profile of Wikimedia-focused research within the broader academic community.
2. **Policy and Strategy Alignment:** The extent to which our project addresses current gaps in understanding Wikipedia's internal dynamics, thereby informing policy decisions and advancing the Wikimedia Foundation's Multigenerational strategy and 2030 strategic direction.

Budget

See the report budget sheet at this link. Please note that the sheet contains the planned budget to carry out the project but this is not equal to the requested amount of money. The hosting institution will take on the compute costs and eventual costs of open access publishing on relevant journals such as Social Network Analysis and Mining or Online Social Networks and Media. The final amount of requested money will be 37,104$.

References

Asikin-Garmager, E., Zimmer, M., Ashraf, C., & Zia, L. (2025, March 13). Research and Privacy on Wikipedia. https://doi.org/10.31219/osf.io/uyxnf_v1

Baele, S. J., Brace, L., & Coan, T. G. (2021). Variations on a theme? Comparing 4chan, 8kun, and other chans' far-right "/pol" boards. Perspectives on Terrorism, 15(1), 65-80.

Cinus, Federico, et al. "Exposing Cross-Platform Coordinated Inauthentic Activity in the Run-Up to the 2024 US Election." *arXiv preprint arXiv:2410.22716* (2024).

Douek, E. (2021). Governing online speech: From" posts-as-trumps" to proportionality and probability. *Colum. L. Rev.*, *121*, 759.

Floréane M. "Pravda Network: Worldwide Expansion and LLM, Wikipedia Pollution" CheckFirst (2025) URL: https://checkfirst.network/pravda-network-worldwide-expansion-and-llm-wikipedia-pollution/

Giglietto, Fabio, et al. "It takes a village to manipulate the media: coordinated link sharing behavior during 2018 and 2019 Italian elections." Information, Communication & Society 23.6 (2020): 867-891.

Luceri, Luca, et al. "Unmasking the web of deceit: Uncovering coordinated activity to expose information operations on twitter." Proceedings of the ACM Web Conference 2024. 2024.

Mancini, A., Desiderio, A., Di Clemente, R., & Cimini, G. (2022). Self-induced consensus of Reddit users to characterise the GameStop short squeeze. *Scientific reports*, *12*(1), 13780.Pacheco, Diogo, et al. "Uncovering coordinated networks on social media: methods and case studies." Proceedings of the international AAAI conference on web and social media. Vol. 15. 2021.

Neff, Jessica J., et al. "Jointly they edit: Examining the impact of community identification on political interaction in wikipedia." PloS one 8.4 (2013): e60584.

Rappaz, J., Catasta, M., West, R., & Aberer, K. (2018, June). Latent structure in collaboration: the case of reddit r/place. In Proceedings of the International AAAI Conference on Web and Social Media (Vol. 12, No. 1).

Tai, Yuehong Cassandra, et al. "GenAI vs. Human Fact-Checkers: Accurate Ratings, Flawed Rationales." arXiv preprint arXiv:2502.14943 (2025)