Community Formation on Wikipedia

CPSC-298 Wikipedia Governance Research Project

Josiah Liao
Chapman University
Orange, California, USA
jliao@chapman.edu

Abstract

Wikipedia as a platform is built upon the principles of democracy and open contributions. In this sort of environment, communities will inevitably form around shared interests or subject matters. However, the way communities are formed around different articles and subjects, as well as the way these communities interact with each other, is a fascinating topic for study. In this paper we use Wikipedia's user databases paired with computational graphing to study how communities are formed and how users interact within different Wikipedia communities. We retrieve data on users' edits on different articles, as well as interactions on articles' talk pages and other users' talk pages. By processing this data, we can analyze the ways different users are grouped around other users and around various articles. [Leave room for summary of findings; results have not been gathered yet]

Keywords

Wikipedia, governance, community, graphs

1 Introduction

In a minimally restricted environment, it is fascinating to study how people self-govern and create communities. Wikipedia is one such minimally restricted environment. Where many moderated or otherwise controlled online environments often devolve into chaos, Wikipedia has remained relatively stable with a mostly user-managed infrastructure. Without the guidance of higher powers, communities will almost inevitably form in such a stable social system. At first glance, these communities may seem simple to categorize. However, when performing deeper analysis we may discover fascinating or shocking connections within these communities that subvert our expectations for what makes a community.

In this paper we seek to answer how communities are formed on Wikipedia and how the landscape of Wikipedia communities looks at this point in time. We look at what different communities revolve around, whether it is a subject matter, a set of articles, or a group of people. Through this we can make conclusions based on the connections between users in a free platform in this case.

By studying Wikipedia communities we can analyze what has kept Wikipedia healthy and growing in a time where online discourse is increasingly divided and self-centered. Additionally, studying the formation of online societies provides fascinating insights into the fields of sociology and psychology, revealing connections we may not expect.

This paper investigates the following research questions:

(1) What are the main large communities on Wikipedia?
(2) What draws these communities together and sustains them?
(3) How do these communities interact with each other and the platform at large?

The main contributions of this work are:

- Observation of communities on Wikipedia
- Analysis of connections between Wikipedia communities
- Use of computational graphing for online sociological research

The rest of this paper is organized as follows. Section 2 reviews related work on Wikipedia governance. Section 3 describes our data and methods. Section 4 presents our findings. Section 5 discusses implications and limitations. Section 6 concludes and suggests future work.

2 Related Work

[Brief paragraph introducing the landscape of related research]

2.1 Wikipedia Governance and Community Culture

Reagle's seminal work on Wikipedia culture [1] examines the collaborative practices and governance mechanisms that enable Wikipedia's success. [Add more discussion of related work here. Use [] to reference papers in your references.bib file]

2.2 etc

2.3 Our Work in Context

[Explain how your work differs from or builds upon existing research. What gap are you filling?]

3 Methodology

[Brief paragraph outlining your overall approach]

3.1 Data Collection

[Describe what data you collected, from where, covering what time period, etc.]

3.2 Data Processing

[Describe how you cleaned, filtered, or transformed your data]

3.3 Analysis Methods

[Describe your analytical approach. What techniques did you use? Why?]

3.4 Ethical Considerations

Optional: Include this if your project involves human subjects data, user behavior analysis, or if your institution requires ethics discussion. For basic Wikipedia article analysis, this may not be necessary - just ensure proper citation in your Data Collection section.

Example: All data consists of publicly available Wikipedia content accessed in compliance with Wikimedia's Terms of Use.

4 Results

[Brief paragraph introducing your main findings]

4.1 [Finding 1 - descriptive title]

[Present your first main finding]

4.2 [etc]

[Present your second main finding]

5 Discussion

5.1 Interpretation of Results

[What do your findings mean? How do they answer your research questions?]

5.2 Implications

[What are the broader implications of your work? For Wikipedia? For research? For practice?]

5.3 Limitations

[Be honest about limitations: data constraints, methodological issues, scope boundaries, etc.]

5.4 Future Work

[What questions remain? What should future research investigate?]

6 Conclusion

[Restate the problem you investigated]

[Summarize your approach and key findings]

[Emphasize your main contributions]

[End with a forward-looking statement about the importance of your work or future directions]

Acknowledgments

We thank . . . for . . . .

References

[1] Joseph Michael Reagle. 2010. *Good Faith Collaboration: The Culture of Wikipedia.* MIT Press.

A AI Usage Documentation

A.1 Literature Review

[Describe how you used AI agents for literature review. Reference your .prompt.md file.]

Example: We used an AI agent workflow (see the file literature-review.prompt.md) to systematically process research papers. The agent extracted summaries, methodology descriptions, and key findings from papers in our bibliography.

A.2 Data Analysis

[If you used AI for data analysis, code generation, or statistical work, document it here]

A.3 Writing Assistance

[Document any AI assistance in writing: brainstorming, editing, restructuring, etc.]

Example: We used Claude/ChatGPT to help with [specific task, e.g., "improving clarity of the abstract" or "suggesting visualizations for our data"].

A.4 Code Development

[If AI helped you write code for data collection or analysis, document it]

A.5 Verification

[How did you verify AI-generated content? What human oversight did you apply?]

All AI-generated content was reviewed, verified against primary sources, and edited by the human author(s). Factual claims were cross-checked with original papers and data.