

# The Role of Discussions in Collaborative Knowledge Graphs



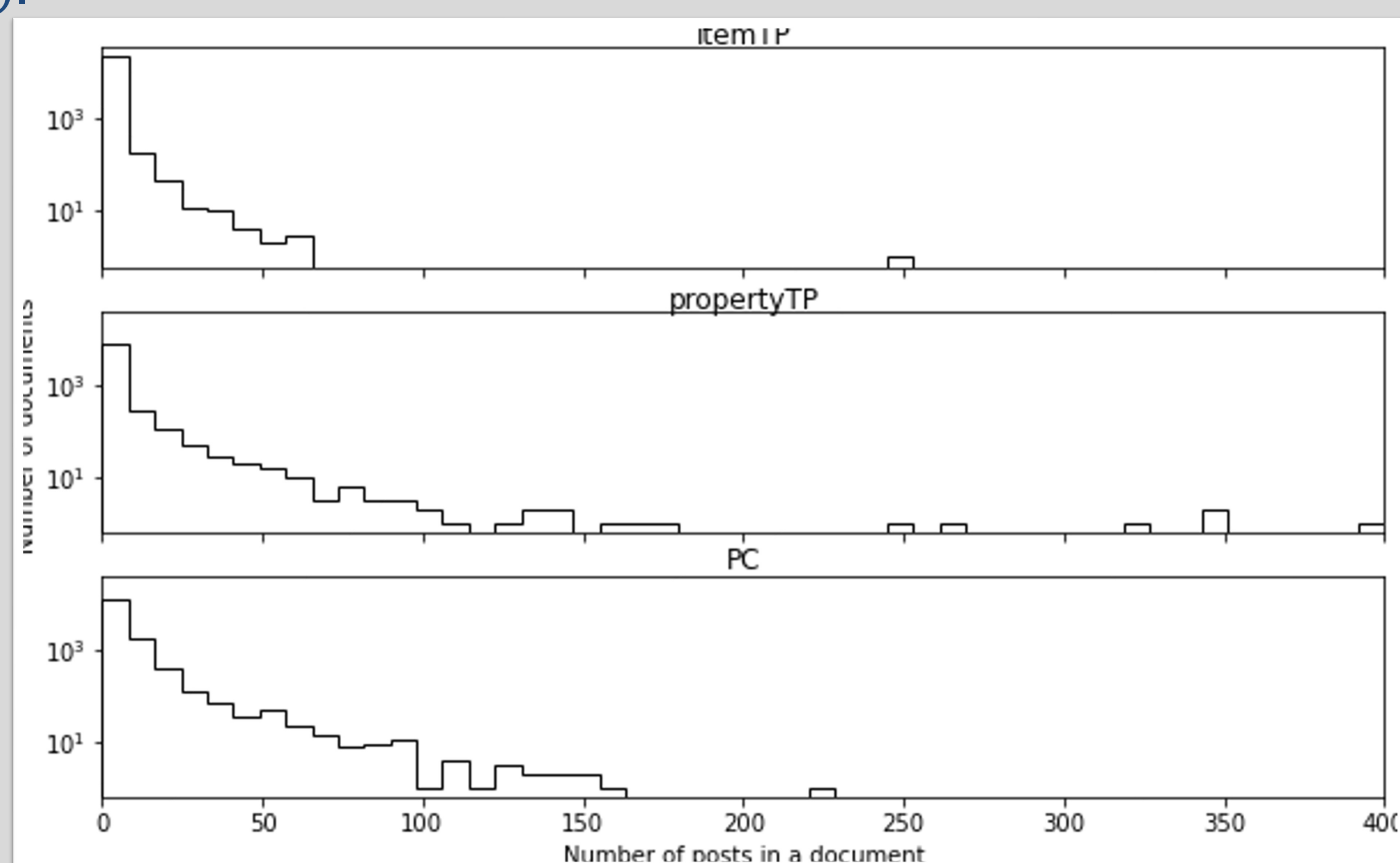
Elisavet Koutsiana, [elisavet.koutsiana@kcl.ac.uk](mailto:elisavet.koutsiana@kcl.ac.uk)

Supervised by: Prof Elena Simperl, Dr Albert Meroño-Peñuela

## Motivation

- Previous studies show that community members discuss about project coordination, guidelines, edits, and strategic planning [1] and can positively impact members' engagement and improve projects' development [2].
- Studying how the underlying community works ensures that the knowledge graph grows in the right direction and keeps high quality to support important AI applications.
- Prior studies in Wikidata have looked at various sociotechnical concerns[3-5] by interviewing editors and analysing activity logs; Wikidata discussions are missing from this line of research.

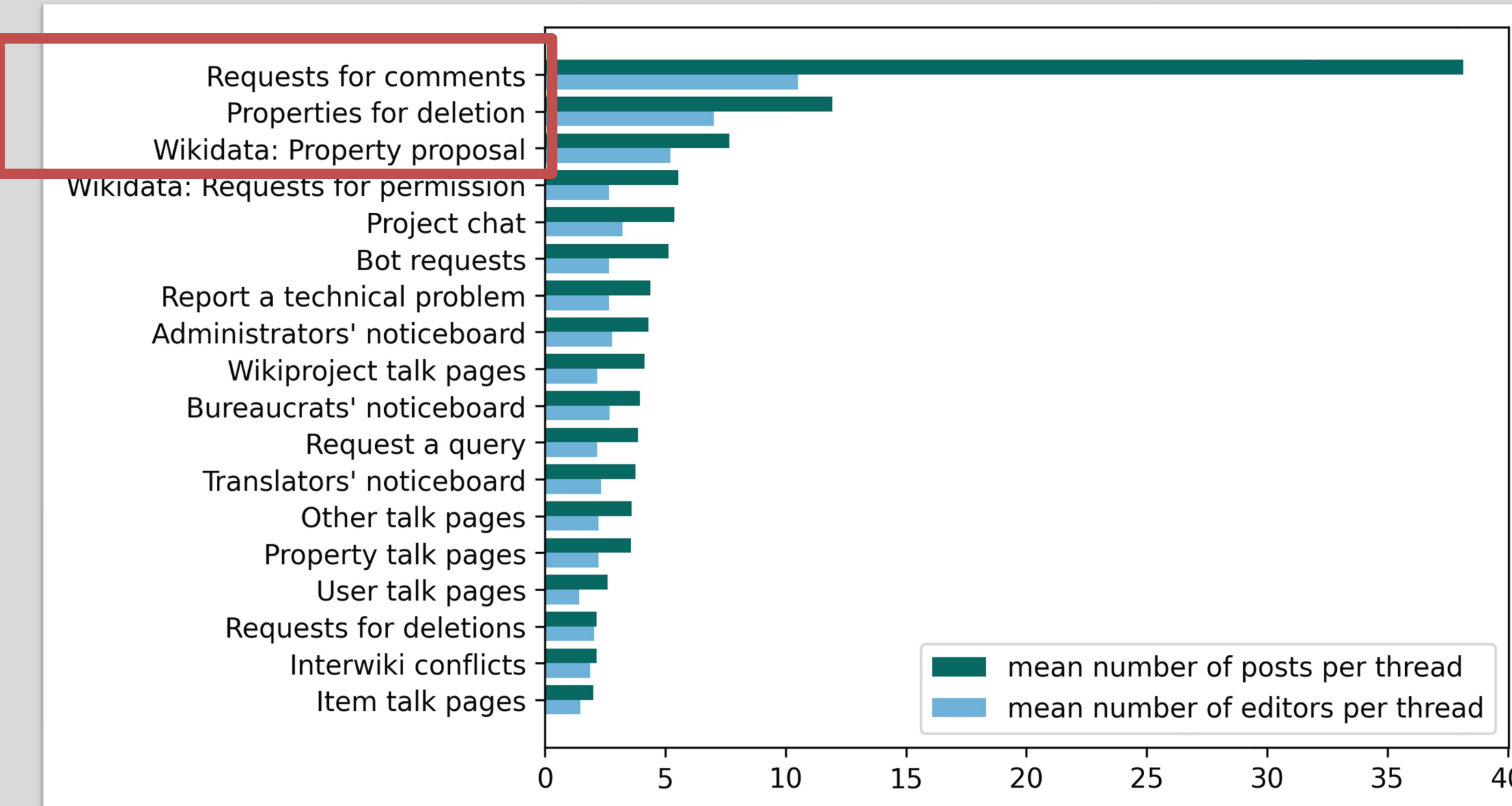
- **Methods:** descriptive statistical analysis, thematic analysis, and statistical methods
- **Dataset:** item and property talk pages, and the project-chat, a general-purpose communication page.
- **Findings:**
  - Members did not use talk pages a lot: only 0.02% of items had them.
  - Discussions followed a power-law distribution, with an overwhelming majority having one post without response (50% for items, 8% for properties, and 16% for project chat) and only a small portion presenting longer discussions (3% for items, 9% for properties, and 31% for project chat).



- The main topics of discussions revolved around knowledge engineering activities. We found activities used to classical and collaborative knowledge engineering, like conceptualisation, implementation, maintenance, role specialisation, version control, project flexibility, and tool support.

- **Methods:** descriptive statistical analysis, thematic and content analysis, measurements of radial trees, and statistical tests
- **Dataset:** a sample of discussions in Wikidata (talk pages and communication pages)
- **Findings:**

Possible controversial communication channels



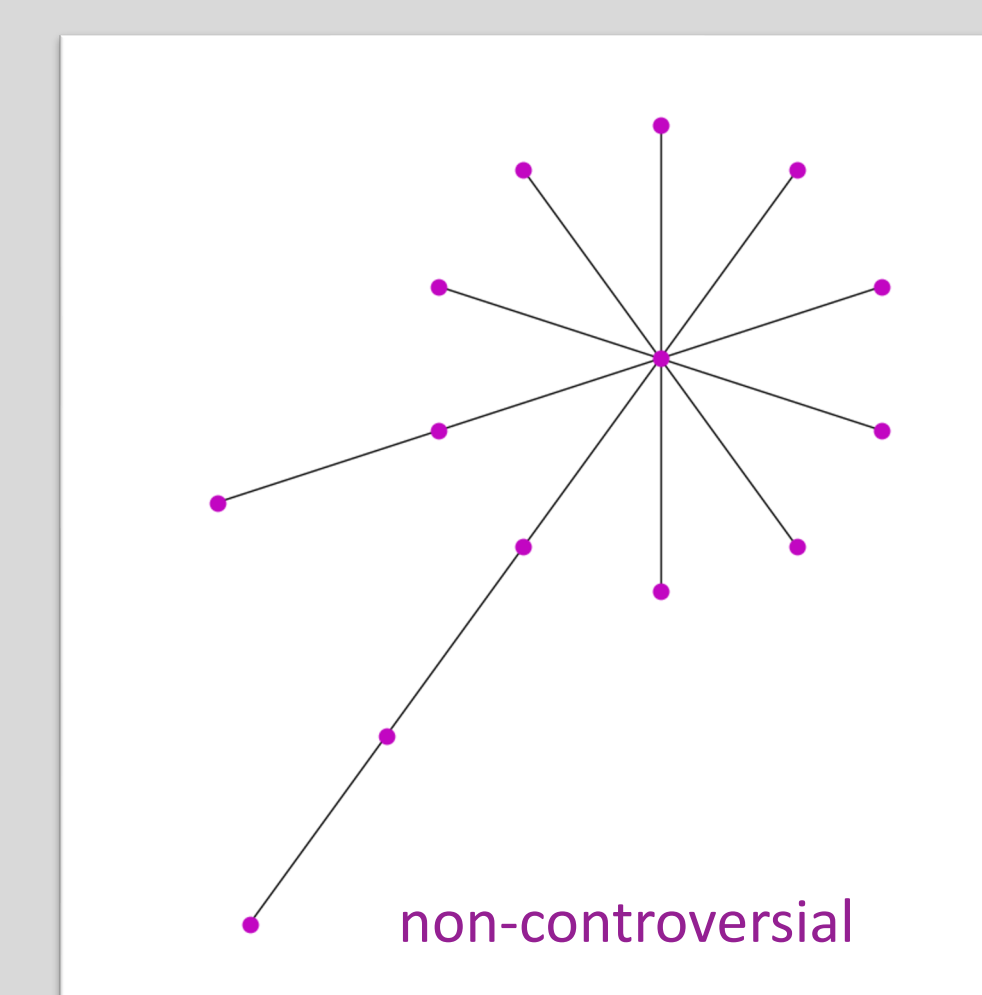
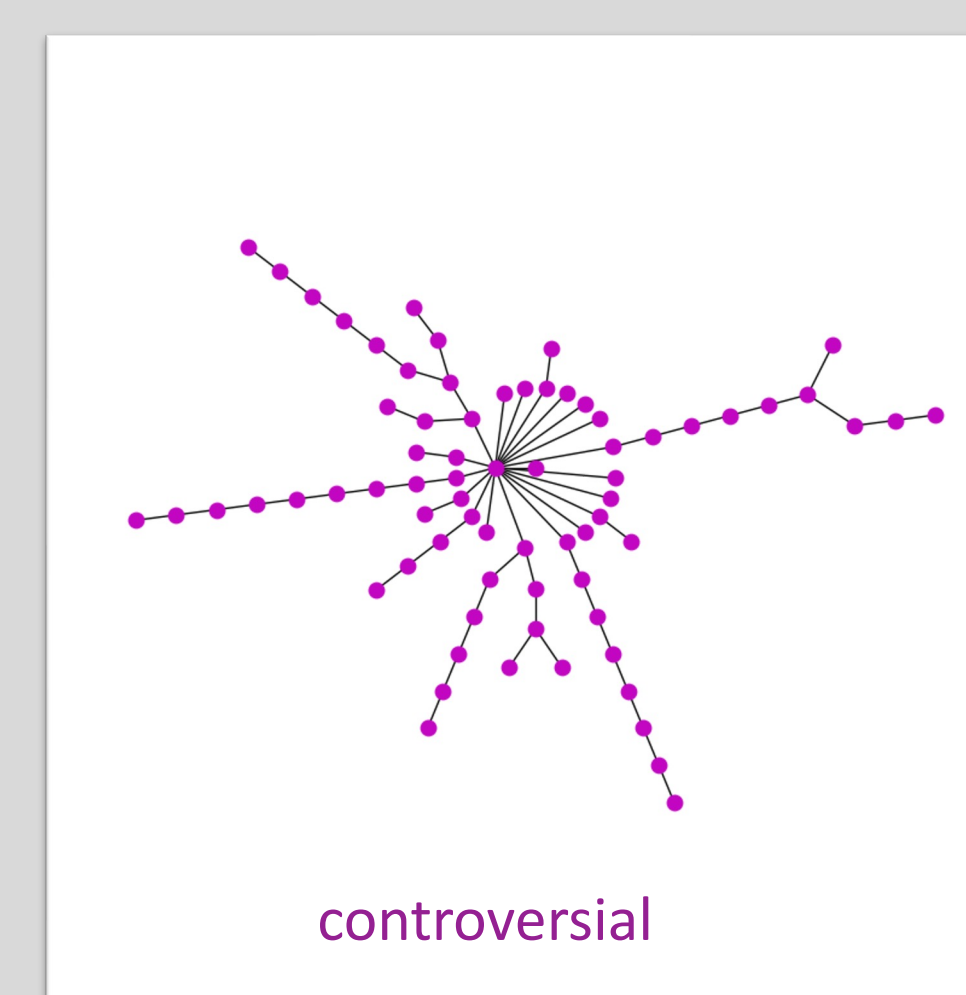
- We found little disagreement (30% of discussions identified as controversial) and very few vandalism (1%).
- The most frequently identified controversial issue was related to processes in Wikidata (52%)
- 25% of participants in controversial discussions made a few but legitimate contributions; however, they did not engage in the discussion.
- The community considered the majority of different opinions in controversial discussions (only 2% of members ignored in argumentation).
- The most used argument was disagreeing using a counter-example.
  - We found significant statistical differences between characteristics of controversial and non controversial discussions but further analysis in a bigger dataset is needed.

What discussions in Wikidata are used for?

How do Wikidata members disagree?

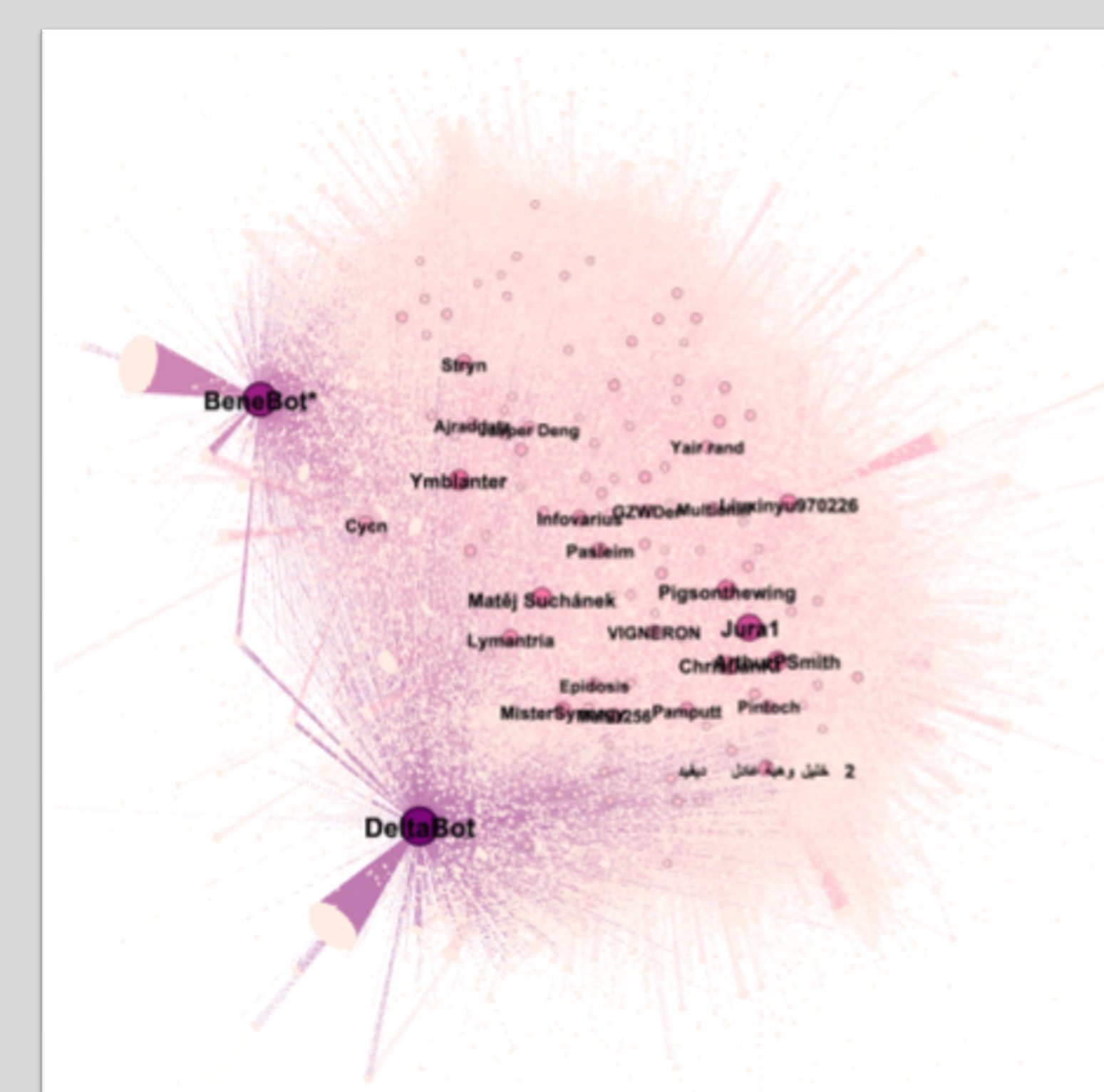
How discussions impact participation?

Who is using discussions?



- **Methods:** Text embeddings + Graph embeddings + Members' features
- **Dataset:** all discussions in Wikidata, the history of edits for members participating in discussions, and members' features like (e.g., members' age in Wikidata, their number of edits, their place in Wikidata hierarchy, if they are active or inactive).

- **Methods:** network analysis
- **Dataset:** all discussions in Wikidata
- **Findings:**
  - For degree centrality, the first two nodes are bots, and ~4k out of ~26K have 0.
  - For eigenvector centrality, first nodes are members high in the Wikidata hierarchy or work in Wikiprojects.



## References

- [1] Fernanda B Viegas et al. "Talk before you type: Coordination in Wikipedia". In: 2007 40th Annual Hawaii International Conference on System Sciences. IEEE. 2007.
- [2] Hideaki Hata et al. "GitHub Discussions: An exploratory study of early adoption". In: Empirical Software Engineering. 2022
- [3] Claudia Müller-Birn et al. "Peer-production system or collaborative ontology engineering effort: What is Wikidata?" In: Proceedings of the 11th International Symposium on Open Collaboration. 2015
- [4] Alessandro Piscopo, Christopher Phethean, and Elena Simperl. "Wikidatians Are Born: Paths to Full Participation in a Collaborative Structured Knowledge Base". In: 50th Hawaii International Conference on System Sciences. University of Hawaii. 2017
- [5] Cristina Sarasua et al. "The evolution of power and standard Wikidata editors: comparing editing behavior over time to predict lifespan and volume of edits". In: Computer Supported Cooperative Work. 2019