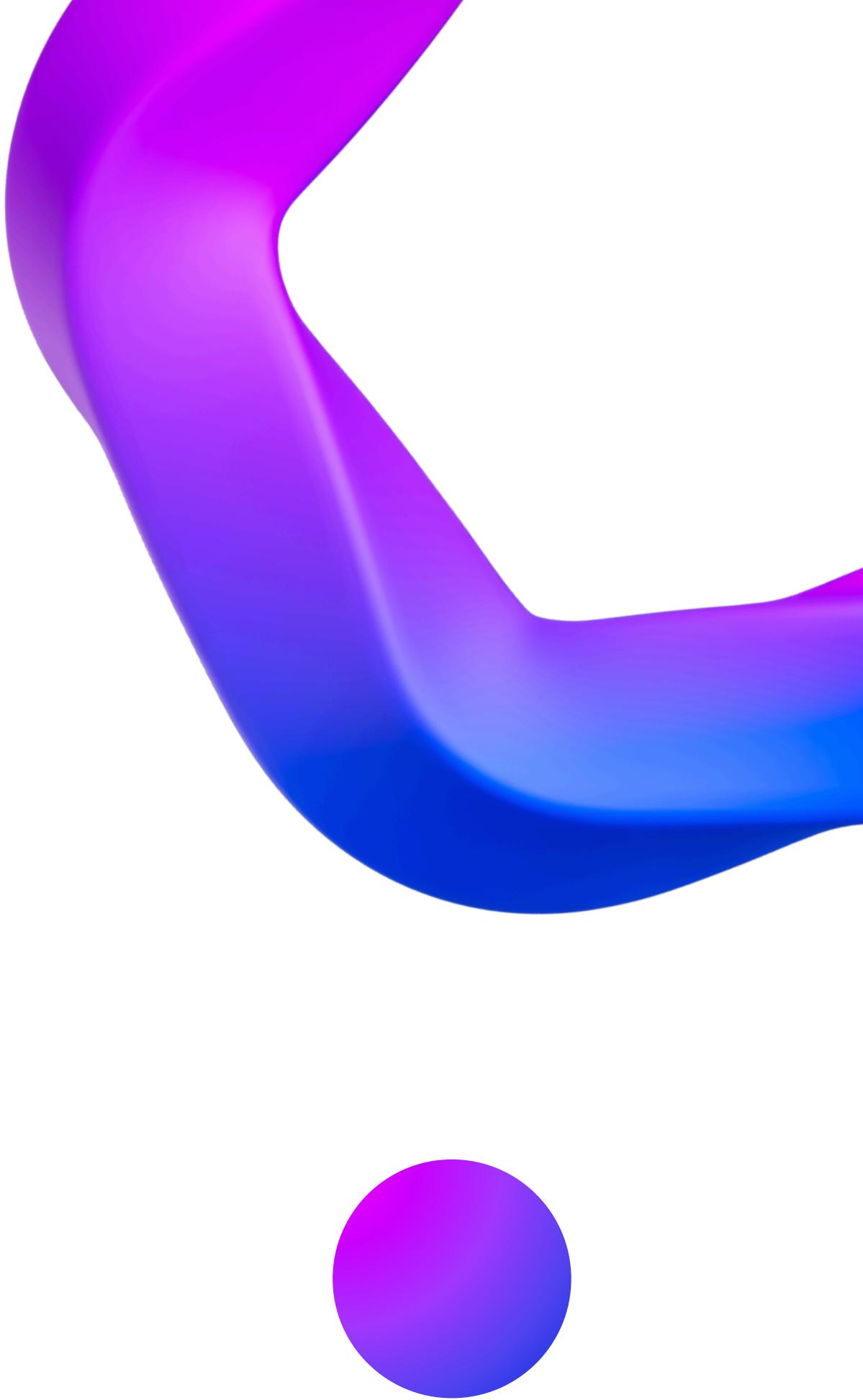




NOVA SCHOOL OF
SCIENCE & TECHNOLOGY



Knowledge graphs for social network analysis

Anna Biausque, 71935

Agnieszka Jędrzejak, 71754

Contents

- 1) Introduction & Research**
- 2) Theoretical work on graph representations**
- 3) Data Extraction & Transformation**
- 4) Modelling with knowledge graphs**
- 5) Comparative analysis**
- 6) Conclusion**

Introduction & research

A few definitions

- Social network Analysis (SNA)

Social network analysis (SNA) is the process of investigating social structures through the use of networks and graph theory.

Examples of social structures commonly visualized through social network analysis include **social media networks, meme proliferation, information circulation, ...**

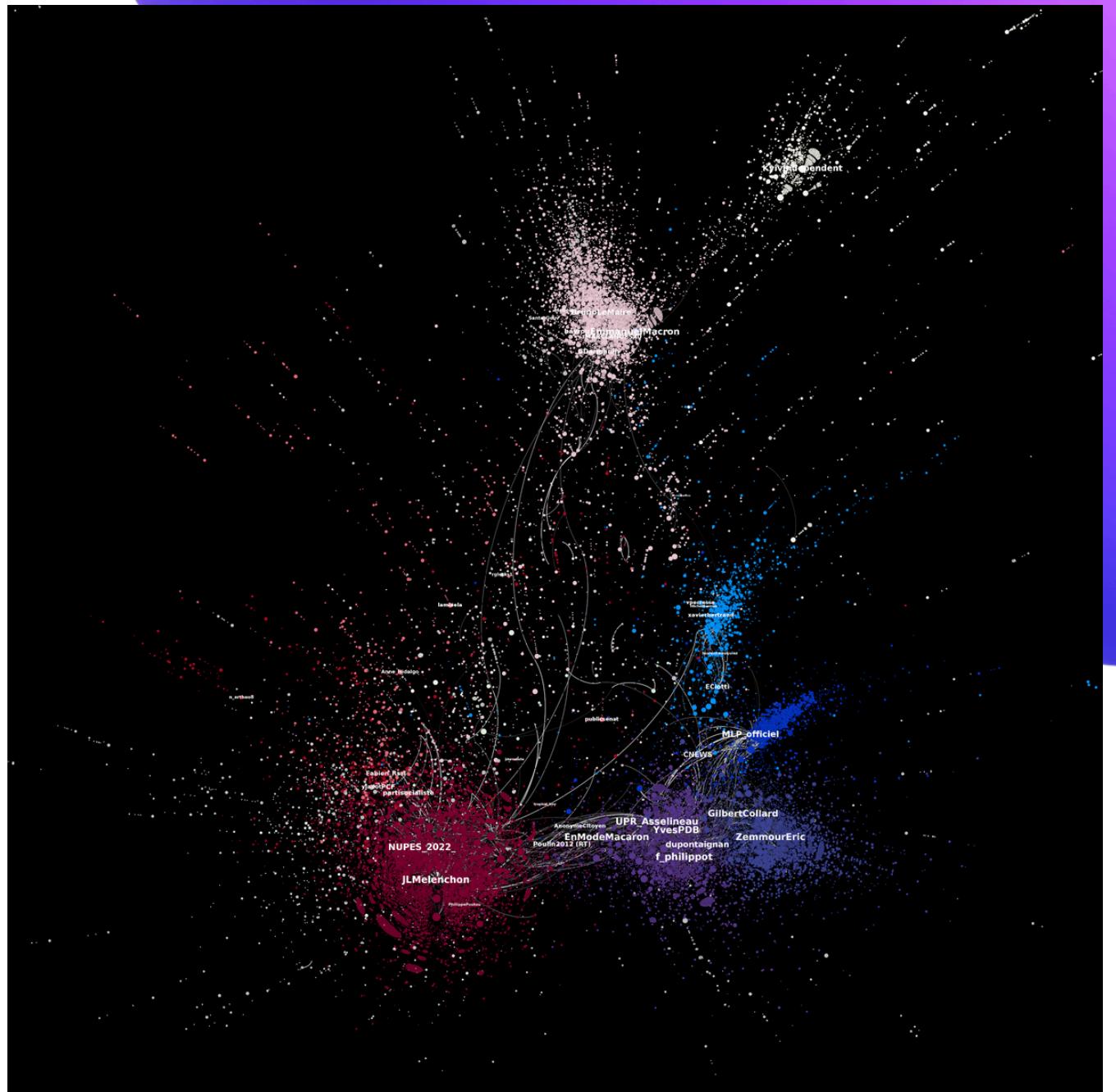


Fig. 1: Political Twittersphere from June 10 to June 16, 2022 (France).

Introduction & research

A few definitions

- **Knowledge graph**

A Knowledge Graph is a structured representation of information in the form of nodes (entities) and edges (relationships), enriched with metadata, semantic attributes, and sometimes ontologies.

It allows modeling complex knowledge in a way that is computationally exploitable.

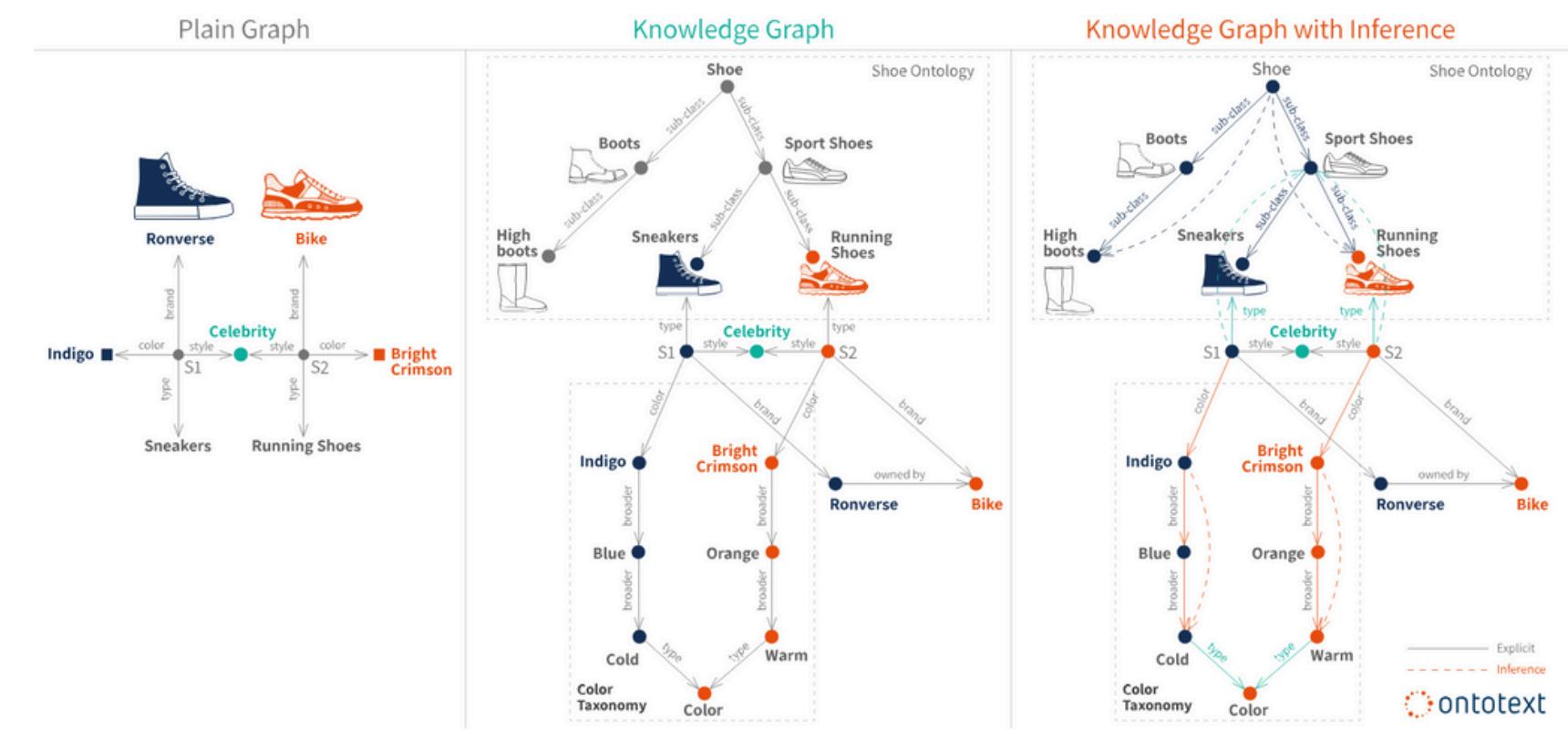


Fig. 2 : Difference between plain graph and knowledge graph

Theoretical work on graph representations

The basic SNA representation

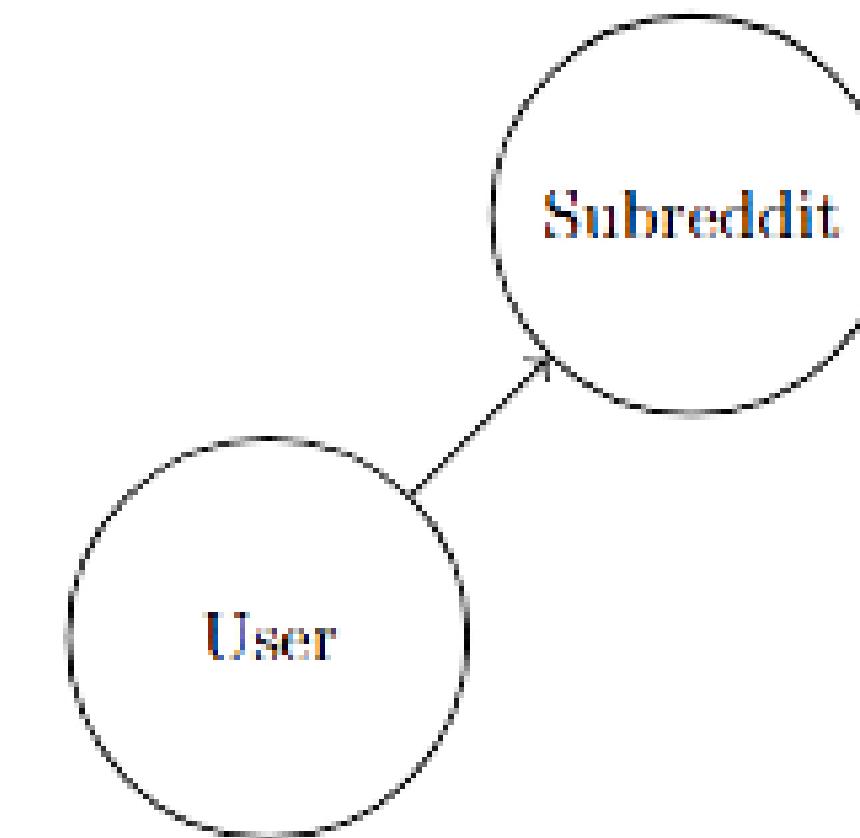


Fig. 3 : Basic model for graph representation

Theoretical work on graph representations

Beyond Subreddits : searching for topics ?

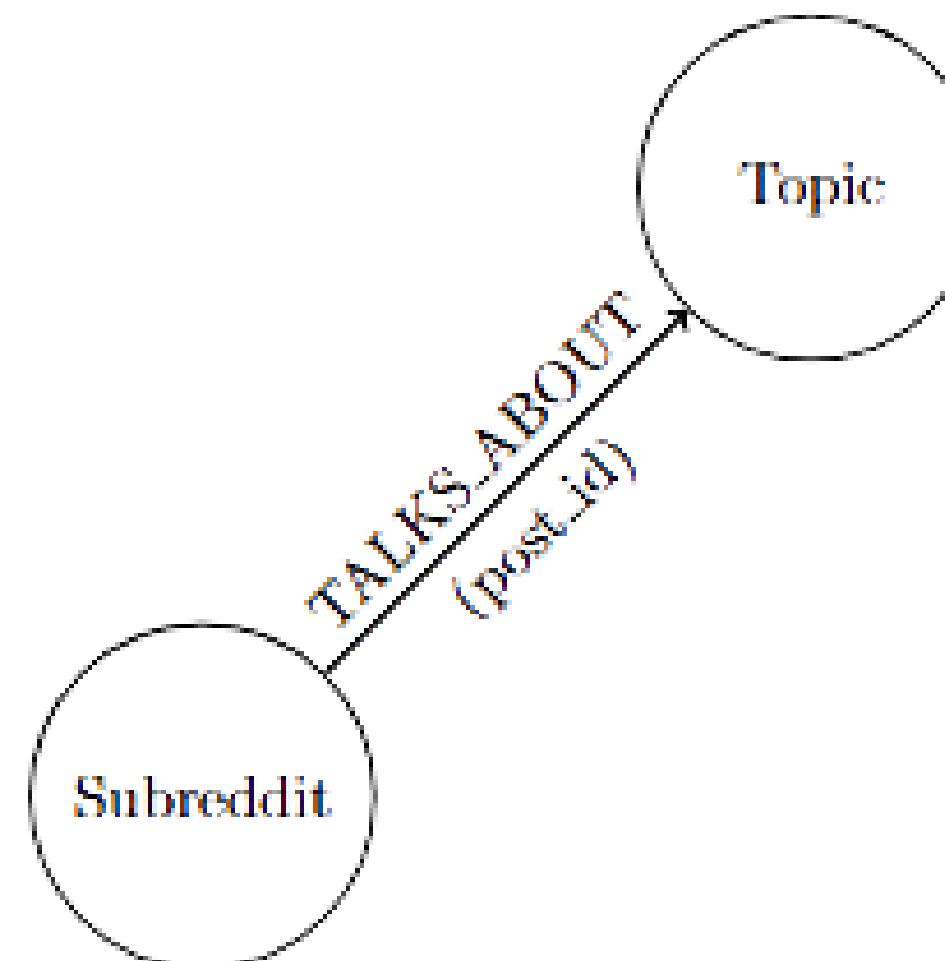


Fig. 4 : Subreddit→Topic model

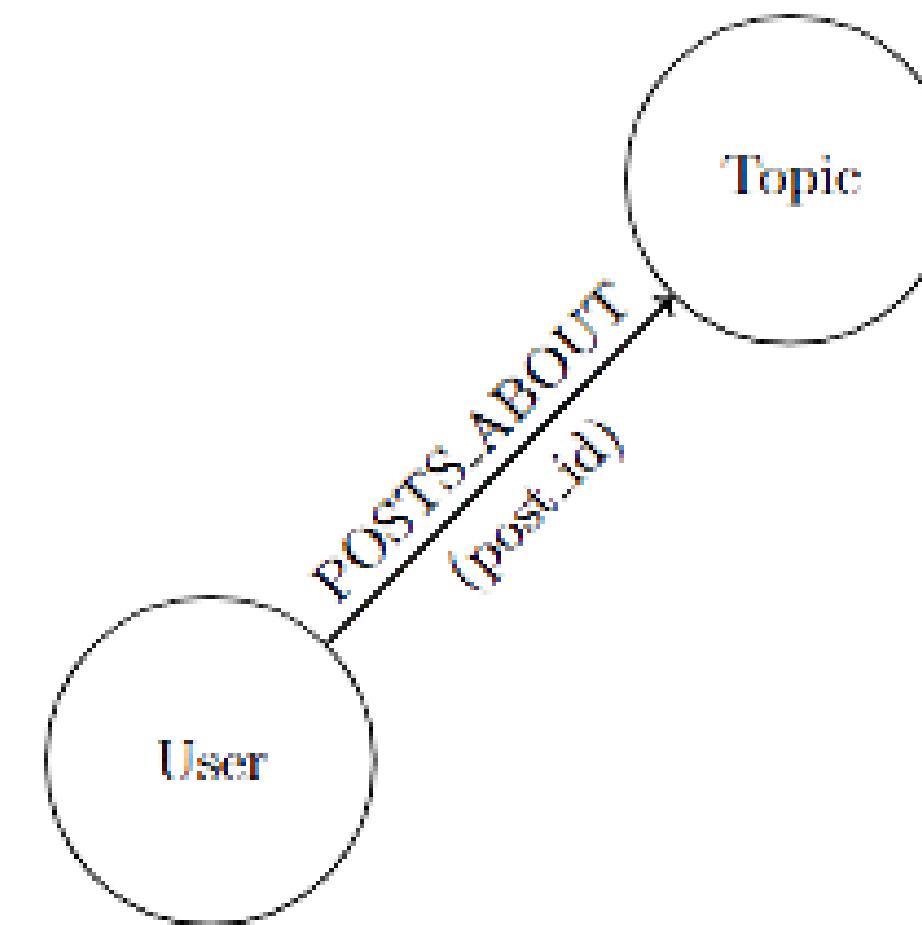


Fig. 5 : User→Topic model

Theoretical work on graph representations

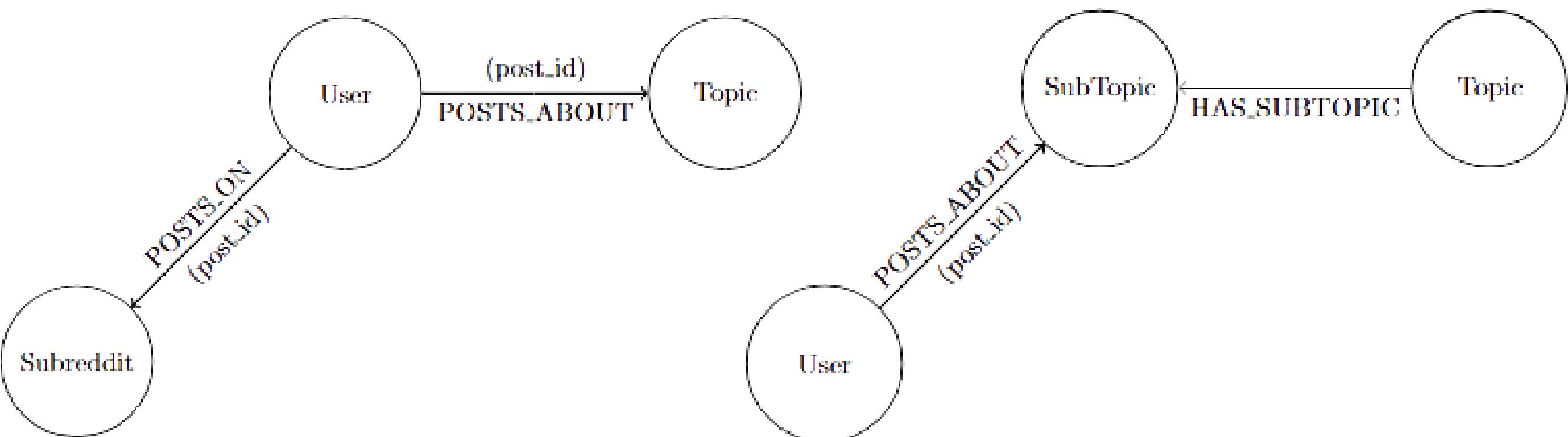


Fig. 6 : Subreddit \leftarrow User \rightarrow Topic model

Fig. 7 : User \rightarrow SubTopic \leftarrow Topic model

Theoretical work on graph representations

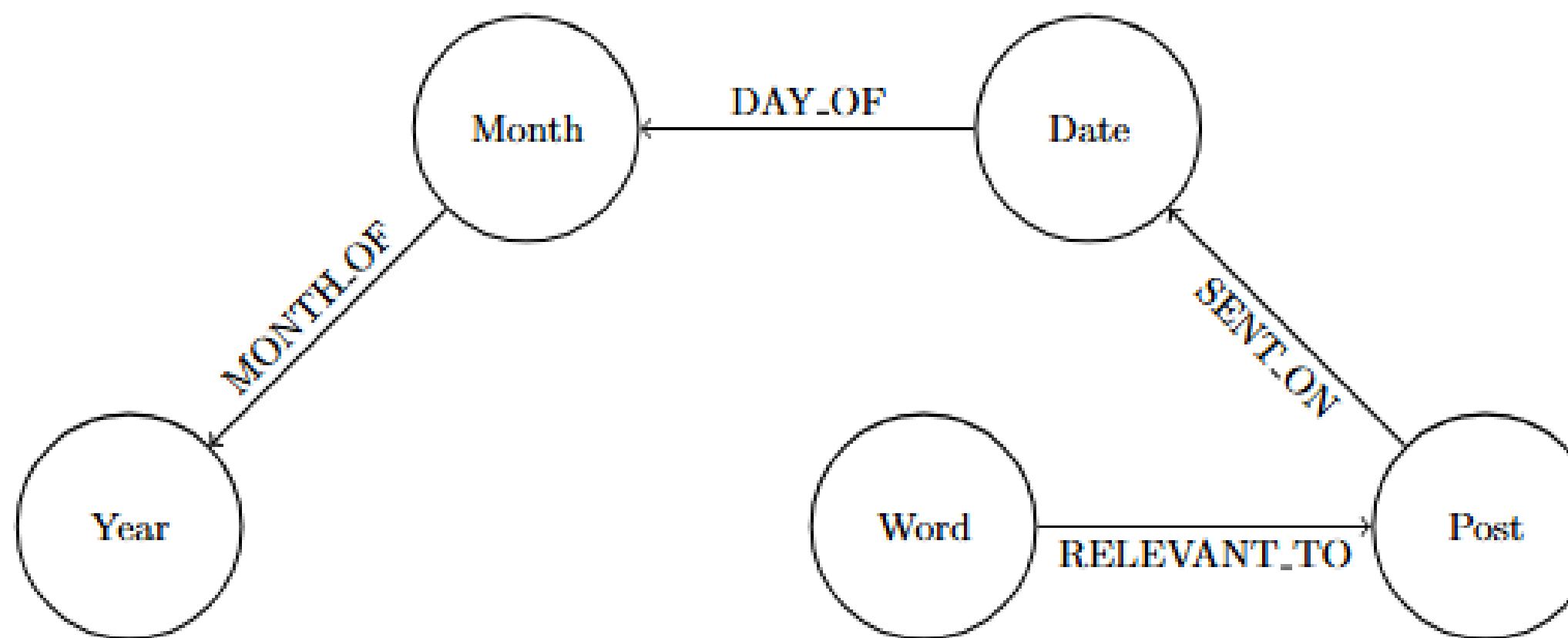


Fig. 8 : Representation with timestamps

Data extraction & transformation

- 4 million posts dataset
 - sub-dataset of 40.000 posts

| "author" | "post_id" | "subreddit" | "subreddit_id" | "content" |
|----------|-----------|-------------|----------------|-----------|
| | | | | |

- 40.300 posts dataset with timestamps

| "type" | "id" | "subreddit.id" | "created_utc" | "selftext" | "date" | "top_3_words" |
|--------|------|----------------|---------------|------------|--------|---------------|
| | | | | | | |

Data extraction & transformation

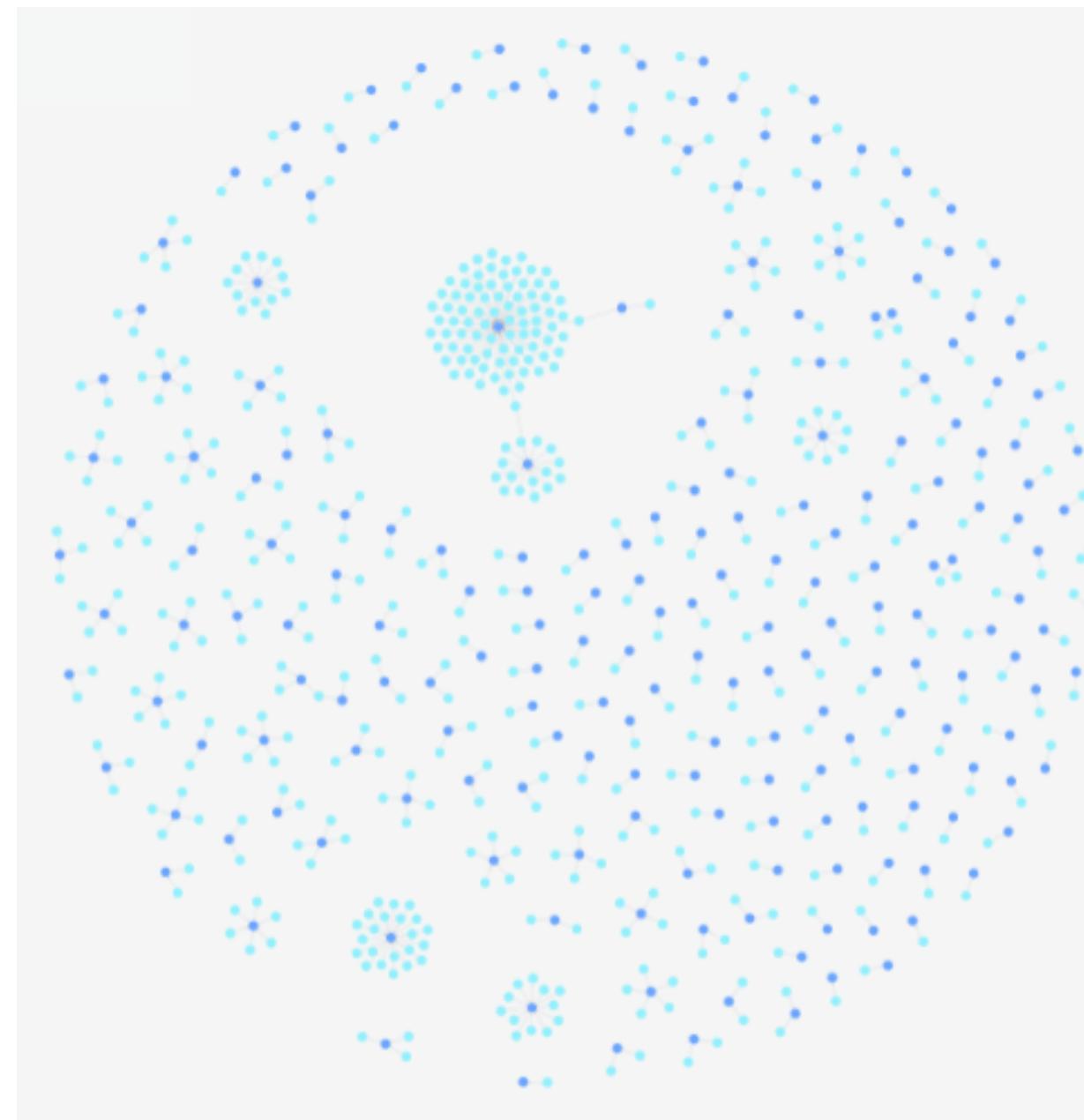


Fig. 9 : Visualization of user-subreddit relationships on smaller part of dataset

- Why not subreddits?
 - too general (e.g. AskReddit)
 - lacks interest

Data extraction & transformation

1st method: Extract topics from the posts

Goal: to generate semantically relevant words from each post --> **Clusters**

- 1) Preprocessing: clean and standardize
- 2) Topic extraction: Latent Dirichlet Allocation (LDA)
- 3) Clustering with BERT and K-means

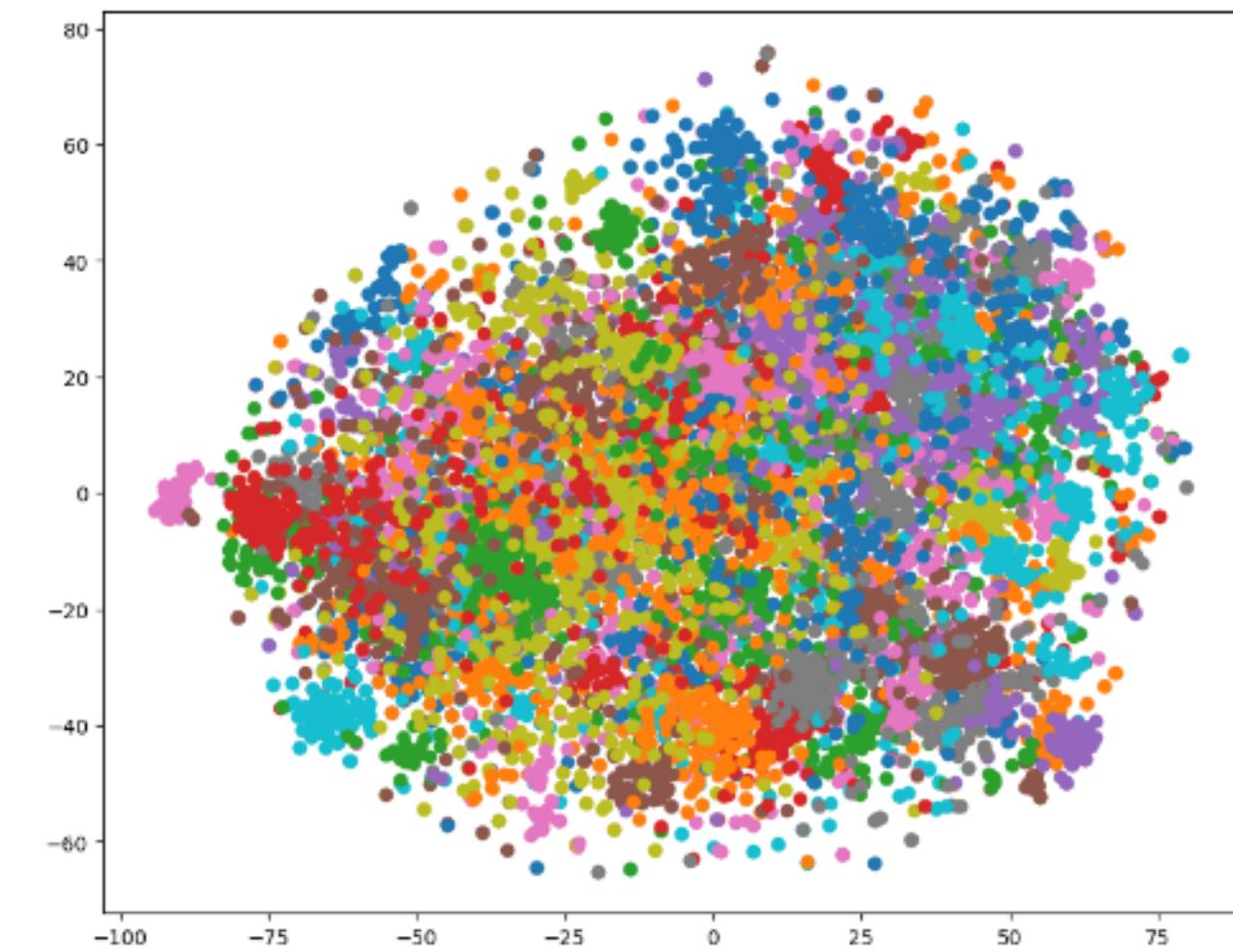


Fig. 10 : Clustering with K-means

Data extraction & transformation

2nd method: Compute similarity with a preexistent list of topics

- 1) Choice of a list of topics
- 2) Use of a Sentence Transformer to compute similarity between each post & topics
- 3) Selection of the 2 most relevant topics for each post

```
author,id,subreddit,subreddit_id,topic1,topic2
rays of darkmatter,c69al3r,math,t5_2qh0n,Environment,Nature
Stork13,c6a9nxd,funny,t5_2qh33,Art,Literature
Cloud_dreamer,c6acx4l,Borderlands,t5_2r8cd,Literature,Politics
NightlyReaper,c8onqew,gamingpc,t5_2sq2y,Game,Technology
```

Fig. 10 : Final result of topic extraction

Modeling with Neo4j

Example of basic graph model

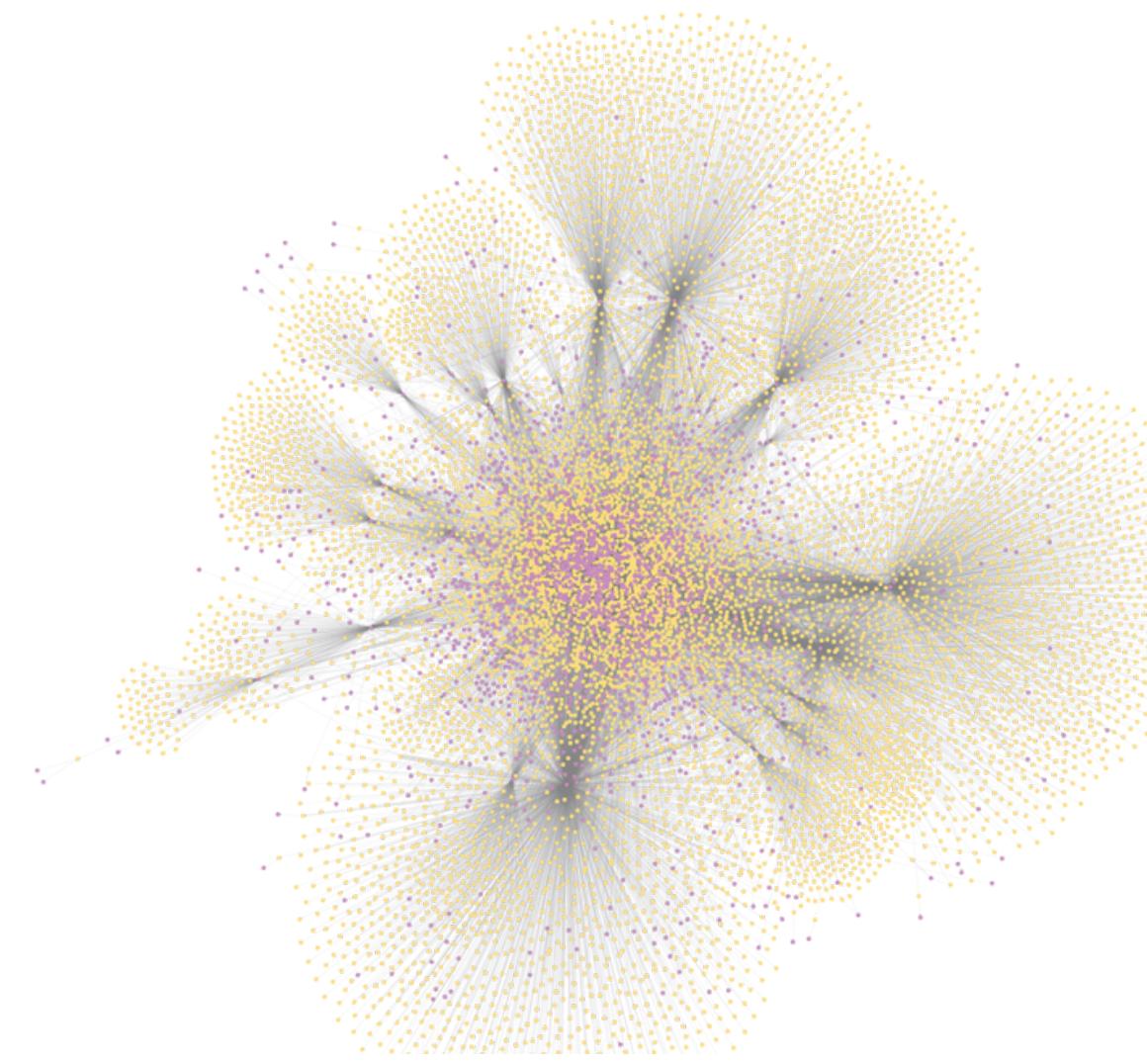


Fig. 11 : Partial visualization with Neo4j Bloom

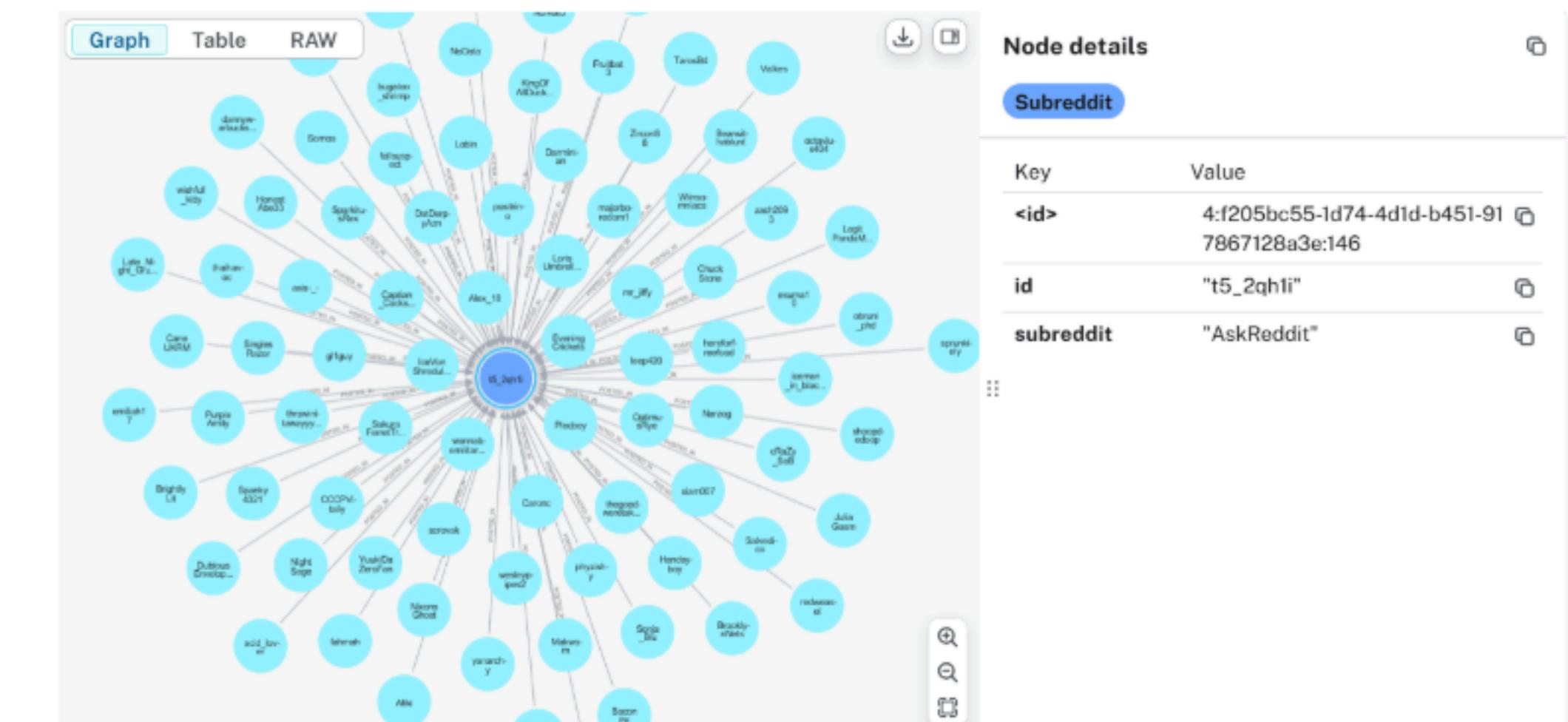


Fig. 12 : Visualization of the users who wrote on “AskReddit”

Modeling with Neo4j

Example of basic graph model

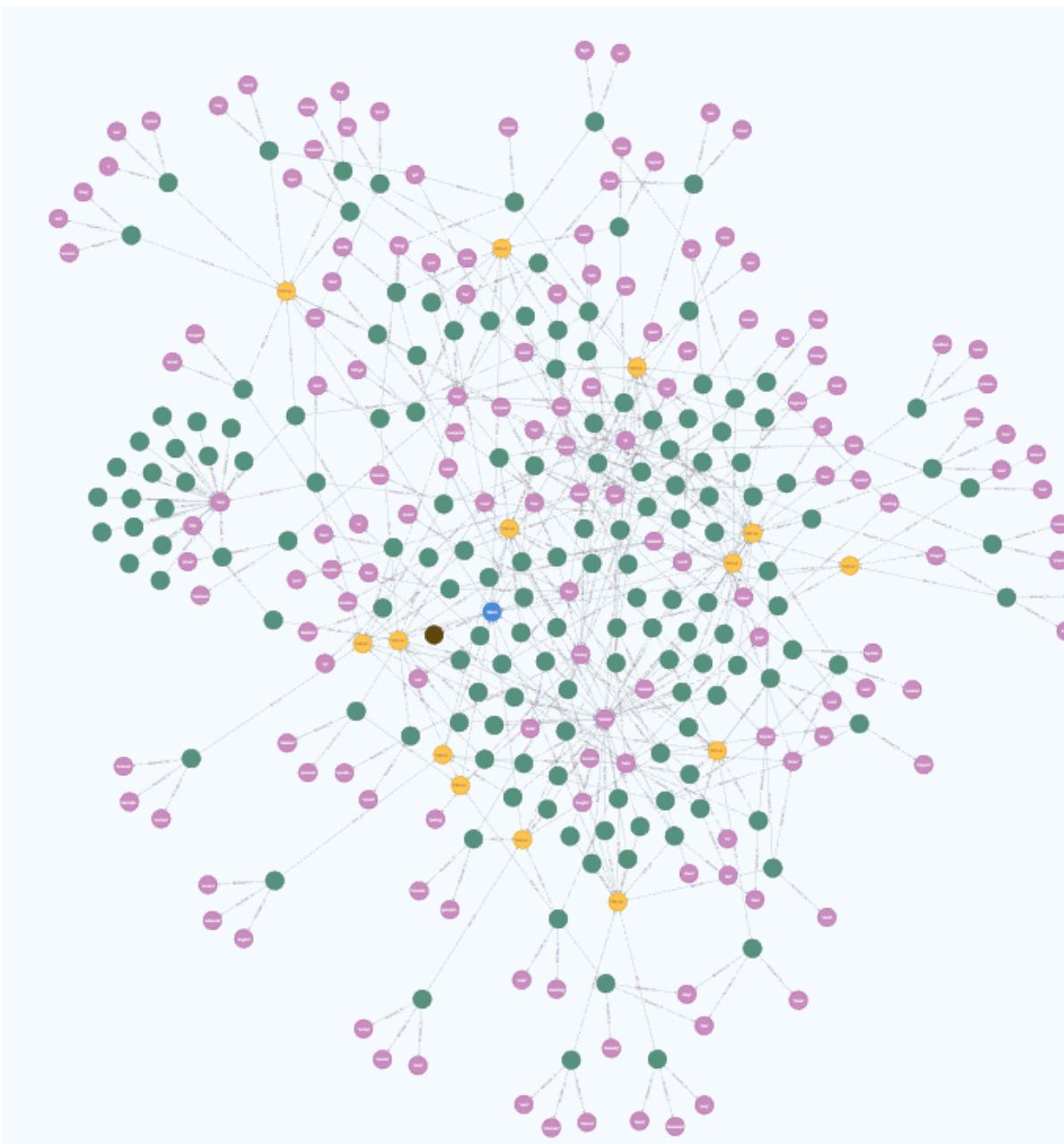


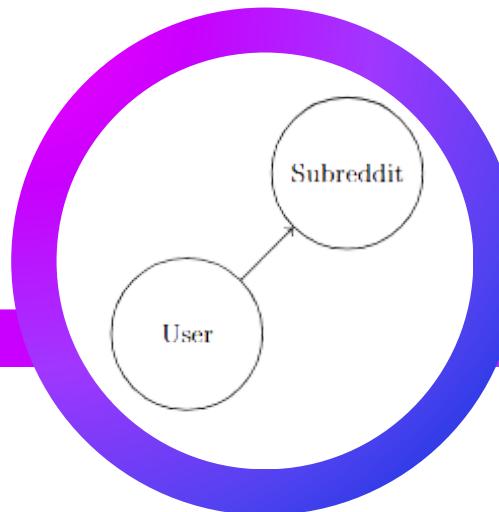
Fig.13 : Visualization in Neo4j of posts sent in March 2021

Modeling with Neo4j

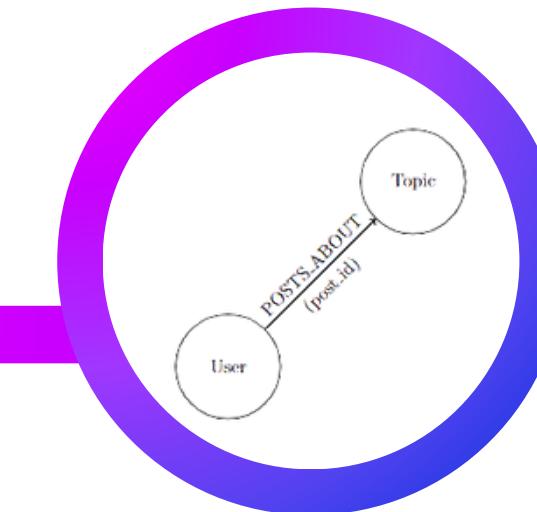
Computing metrics

| Cypher queries | GDS library |
|-------------------|------------------------|
| Degree centrality | Betweenness centrality |
| Density | Louvain's algorithm |

Comparative analysis



- Density : **5.10⁻⁶**
- Degree centrality
 - Range : **118 307**
 - Mean : **52.13**
 - Median : **2.0**
 - Standard deviation : **1242.84**



- Density : **10⁻⁵**
- Degree centrality
 - Range : **44 468**
 - Mean : **24 408.68**
 - Median : **20 822.0**
 - Standard deviation : **10 729.15**

Comparative analysis

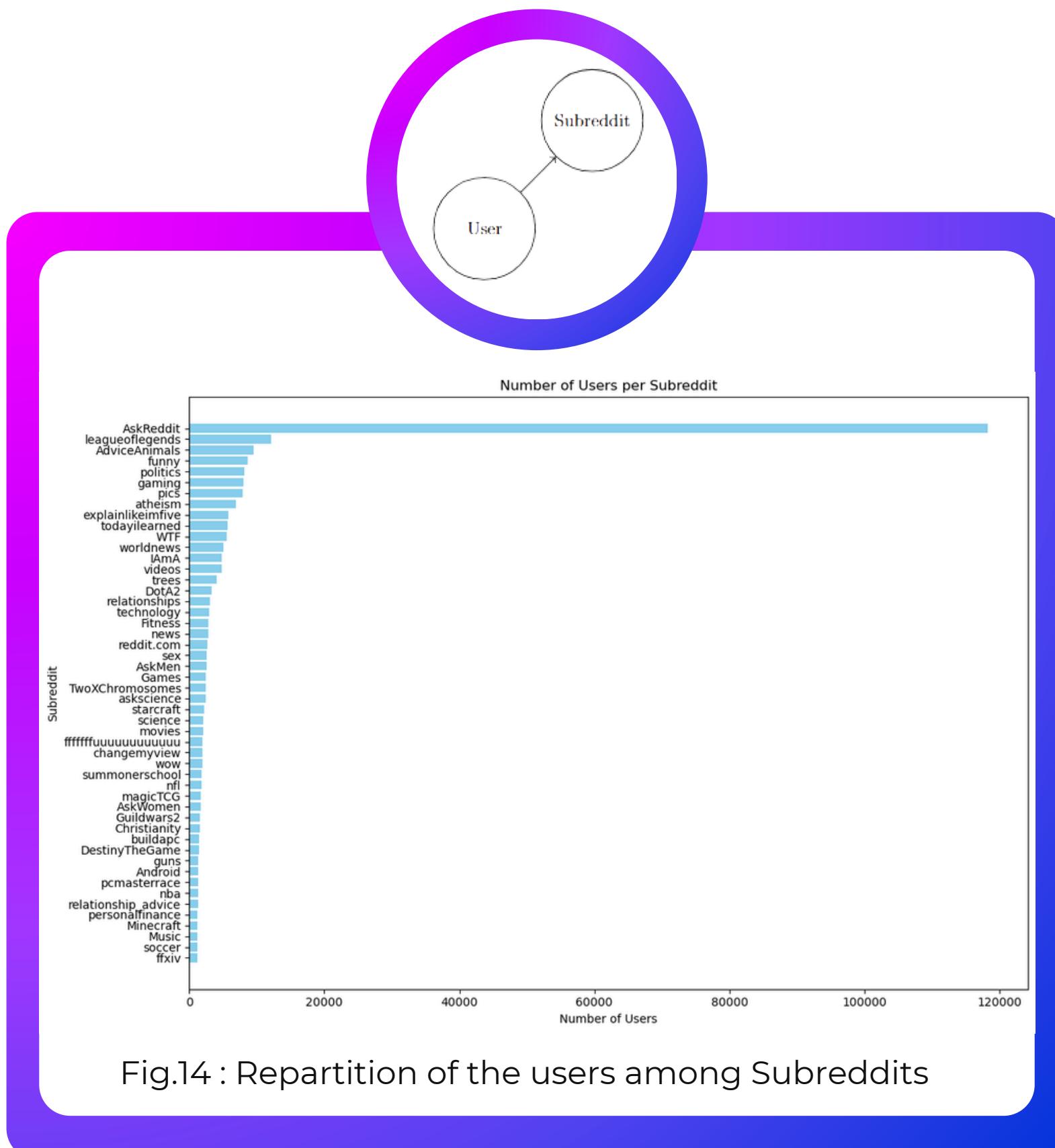


Fig.14 : Repartition of the users among Subreddits

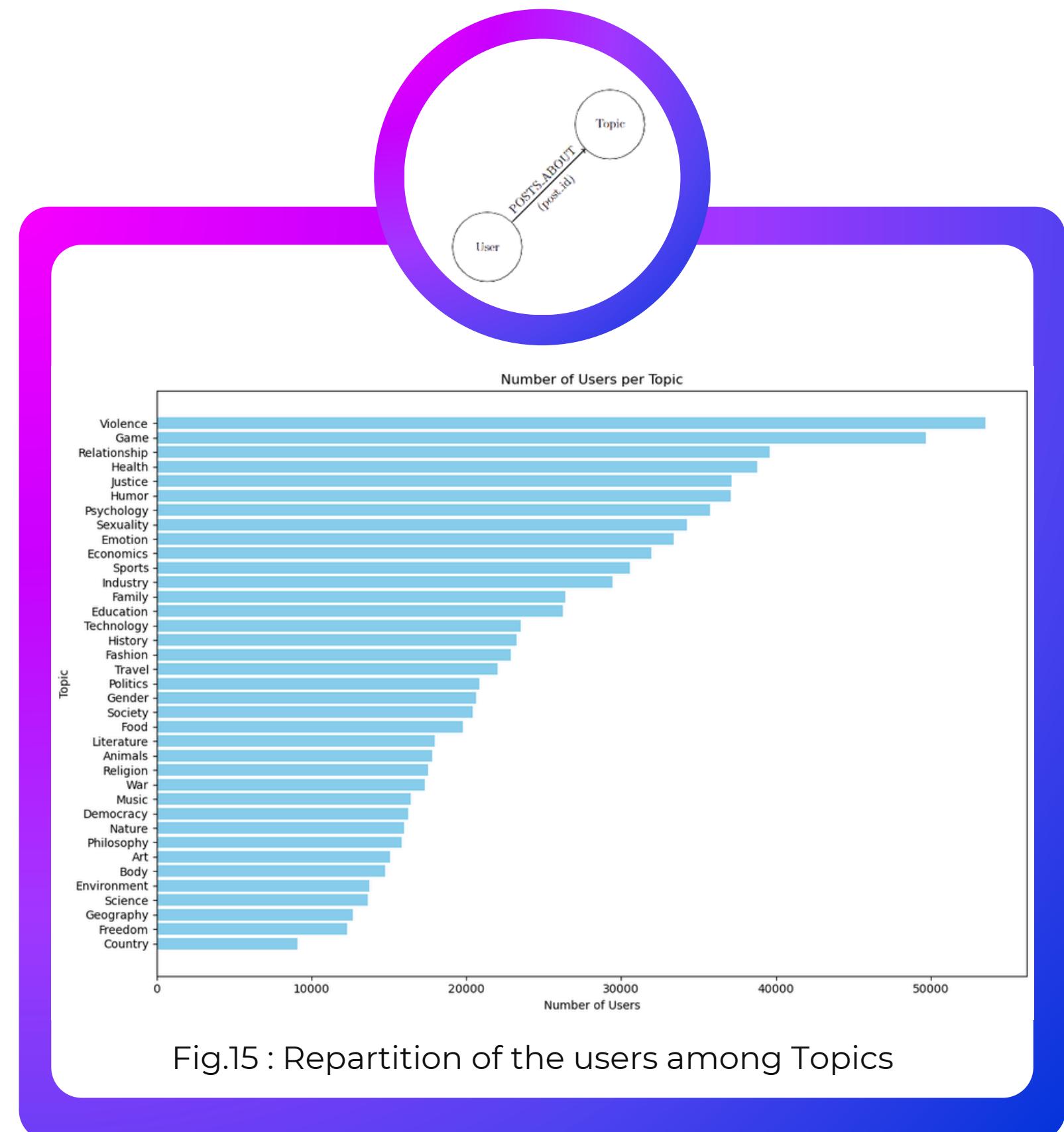


Fig.15 : Repartition of the users among Topics

Comparative analysis

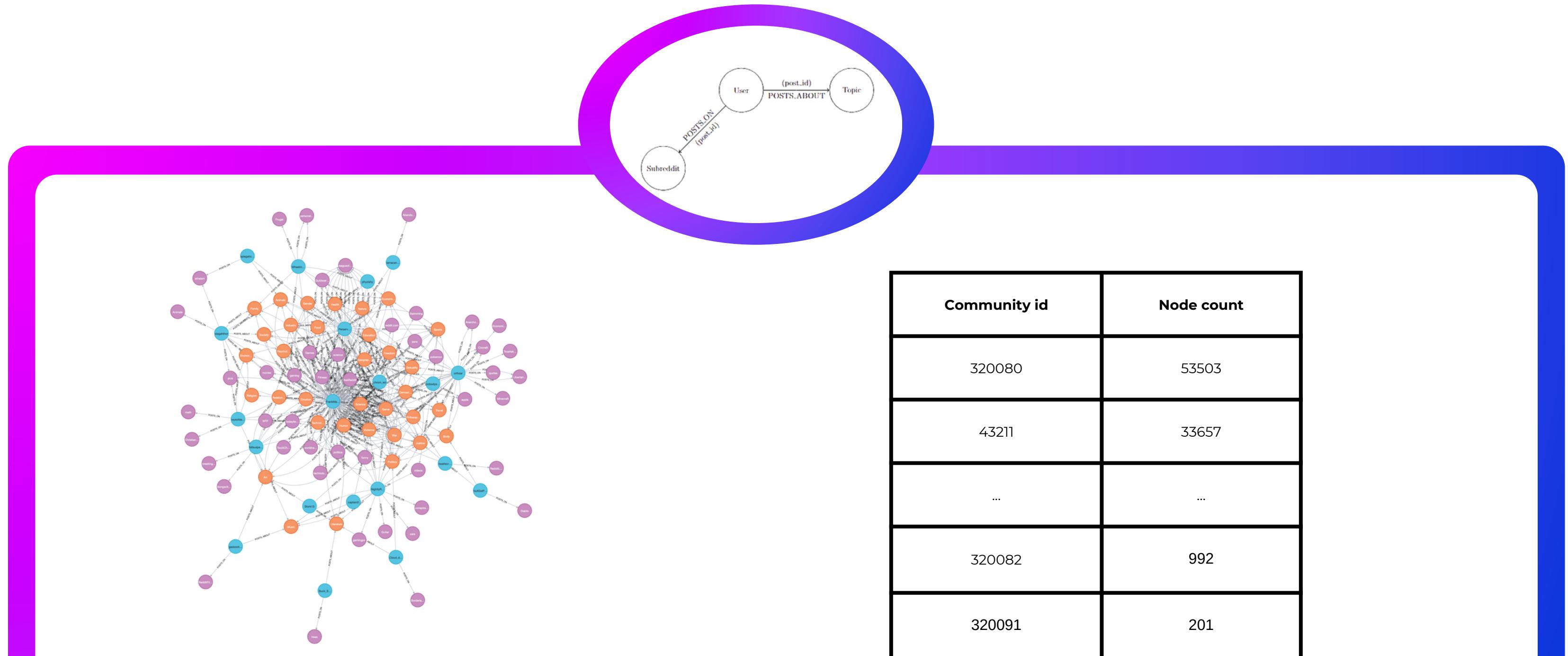


Fig.16 : Subreddit→User→Topic modeling with Neo4j

Analysis conducted with Louvain's algorithm : 16 communities

Comparative analysis

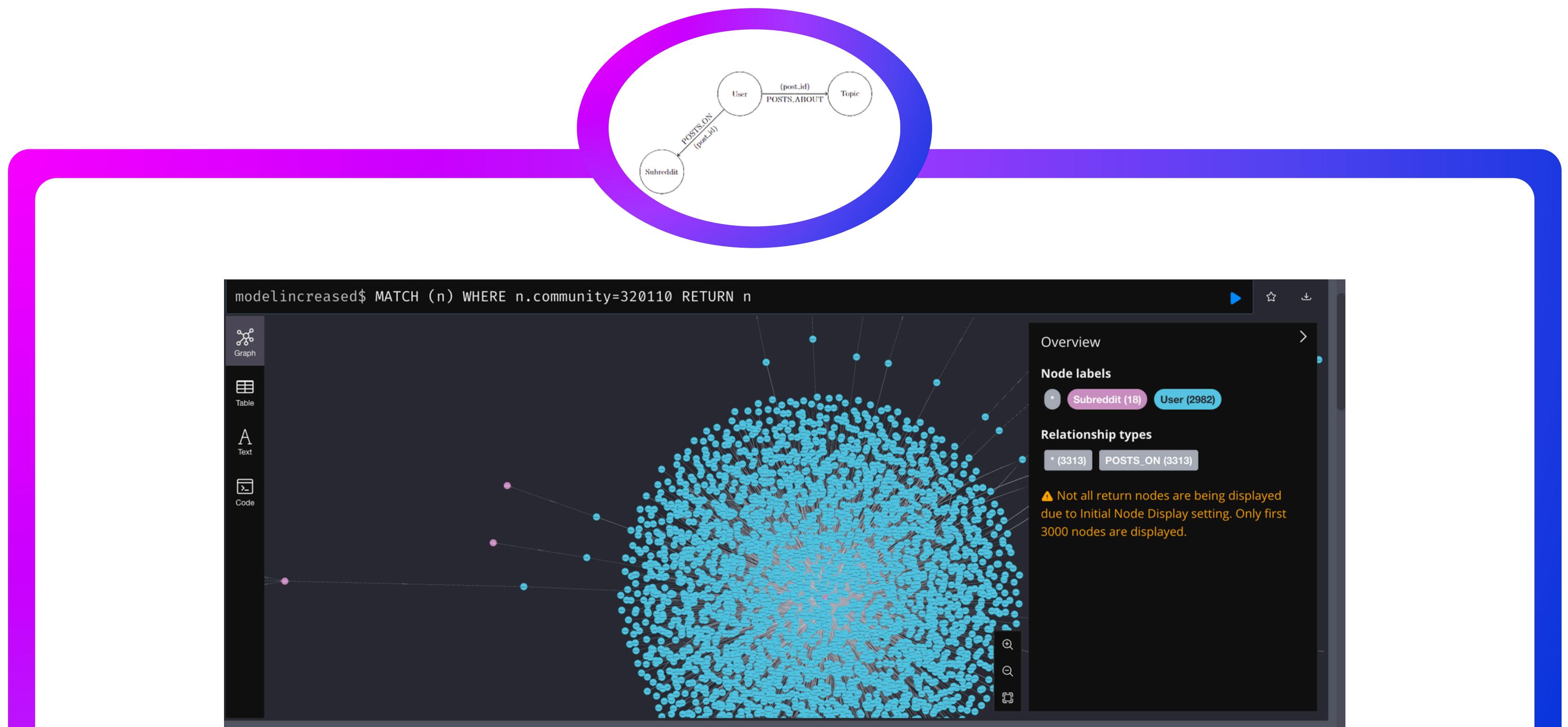
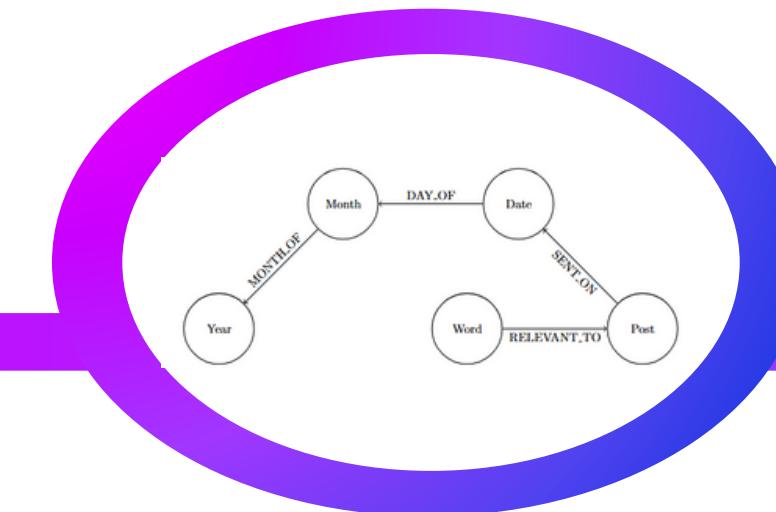


Fig.17 : Composition of a community

Timestamped analysis



- Density : **7.47×10^{-5}**
- Degree centrality
 - Range : **20292**
 - Mean : **140.91**
 - Median : **117.0**
 - Standard deviation : **115.11**

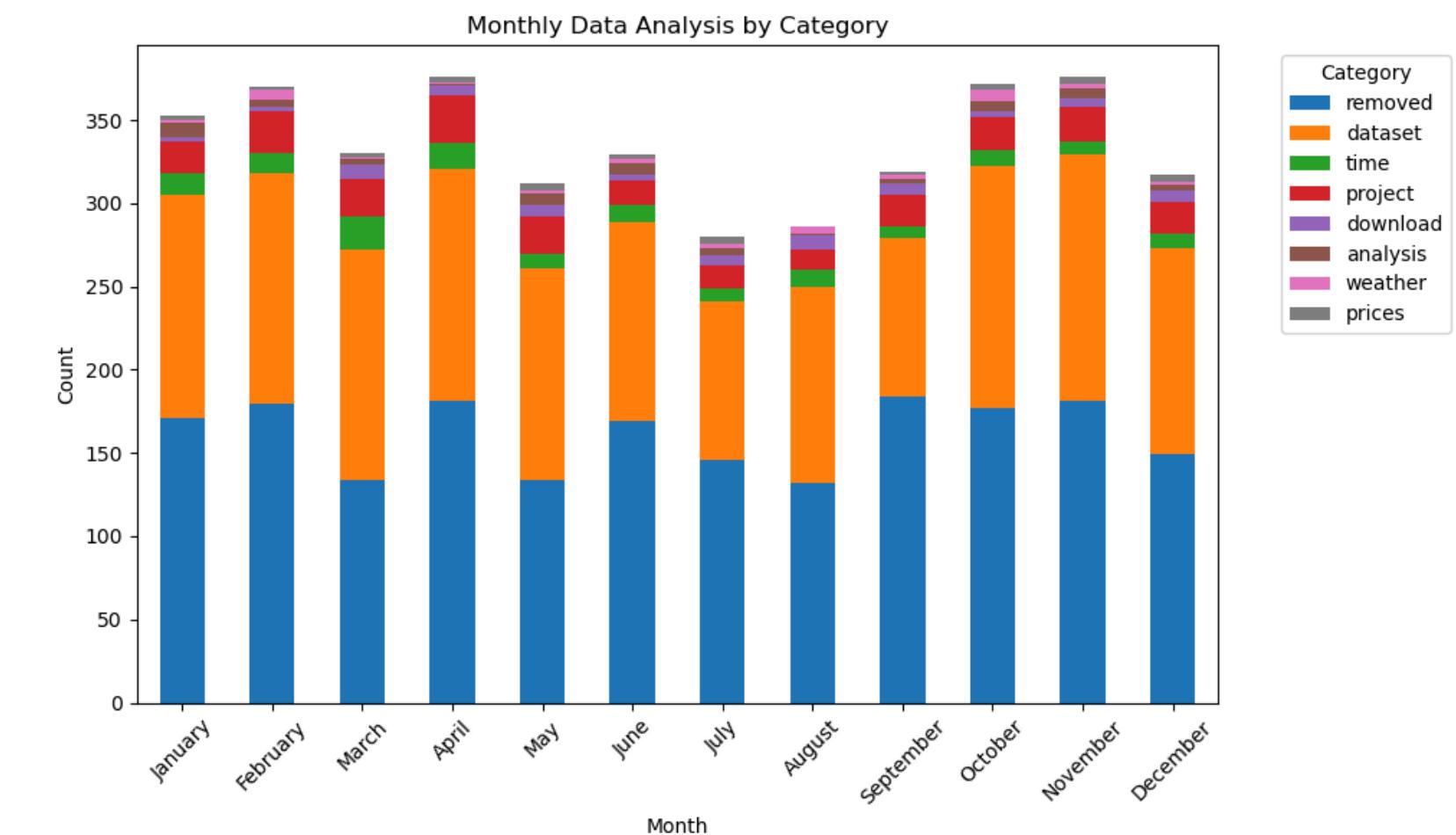


Fig.18 : Monthly analysis by category



Conclusion

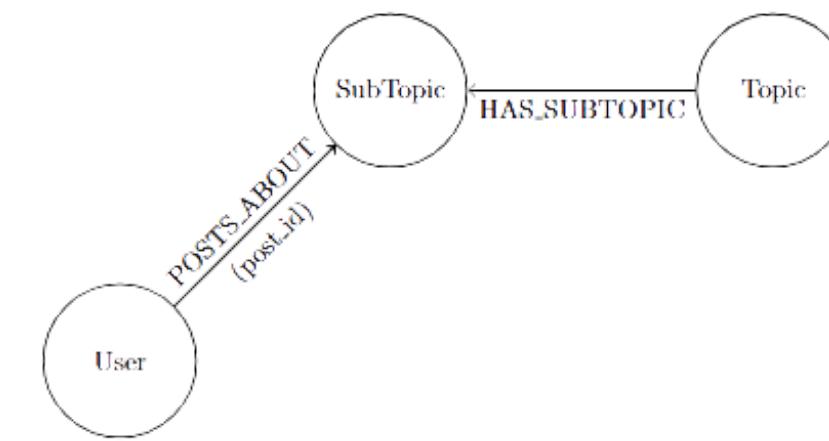
Work accomplished :

- Reflexion on theoretical knowledge graphs for SNA
- Topic extraction from the content of the posts
- Comparative analysis between model
- Comparative analysis along time

Conclusion

Areas for improvement:

- Implementation of the remaining model and comparison on communities
- Use of a more relevant dataset for timestamped analysis
- Effective visualization of the communities



Bibliography

- “**Building Knowledge Graphs: A Practitioner’s Guide.**” Graph Database & Analytics, <https://neo4j.com/knowledge-graphs-practitioners-guide/>. Accessed 14 Dec. 2024.
- **Dörpinghaus, Jens, et al. “From Social Networks to Knowledge Graphs: A Plea for Interdisciplinary Approaches.”** Social Sciences & Humanities Open, vol. 6, no. 1, Jan. 2022, p. 100337. ScienceDirect, <https://doi.org/10.1016/j.ssaho.2022.100337>.
- “**Minuit Moins Dix à l’horloge de Poutine.**” Politoscope, 1 July 2024, <https://politoscope.org/2024/07/3471/>.
- **Webis/Tldr-17** · Datasets at Hugging Face. 18 Feb. 2023, <https://huggingface.co/datasets/webis/tldr-17>.
- “**What Is a Knowledge Graph?**” Ontotext, <https://www.ontotext.com/knowledgehub/fundamentals/what-is-a-knowledge-graph/>. Accessed 14 Dec. 2024.



Thank you !