

Solving the long tail issue in recommender systems

By optimising tripartite graphs

Thomas Crul

t.c.crul@student.tudelft.nl

Supervisors:

Frans Oliehoek

Aleksander Czechowski

Oussama Azizi

Davide Mambelli

1 The long tail issue

While most recommender systems show great performance and accuracy when popular items are concerned, the error rate tends to increase towards the low-ranked items that reside in the long tail of the itemset.

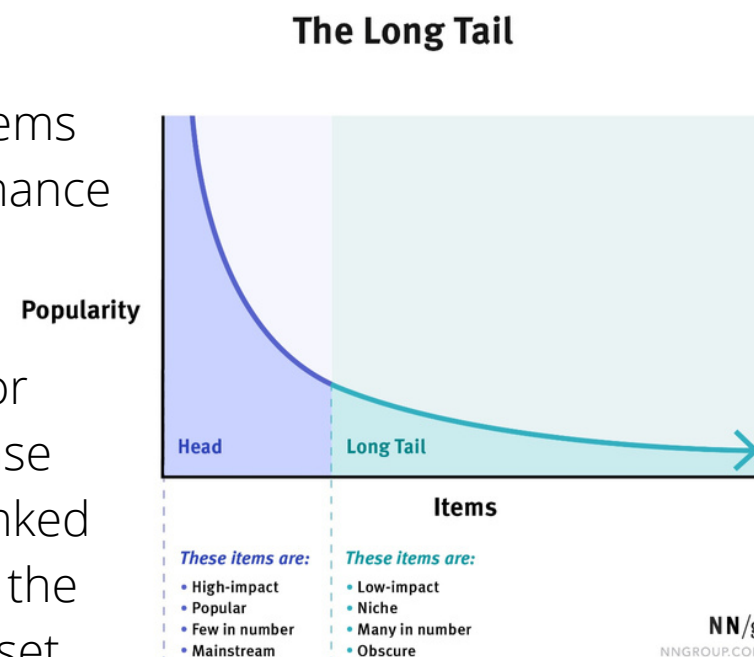


Figure 1: An illustration of the long tail

Aim of the study

Investigate the impact of the design of the **additional layer** on the performance of the tripartite model in solving the long tail issue.

2

3 The graphs

The **tripartite graphs** contain three disjoint groups of nodes: users (U), items (I), and the main layer of interest: **the additional layer** (C). This final layer aims to effectively characterise either U or I. All options present in the MovieLens dataset are covered:

- Basic genre (18) - Release year (81) - Full genre (301)
- Gender (2) - Age (6) - Occupation (21) - Zip code (3,439)

The numbers indicate the number of nodes in the intermediate layer.

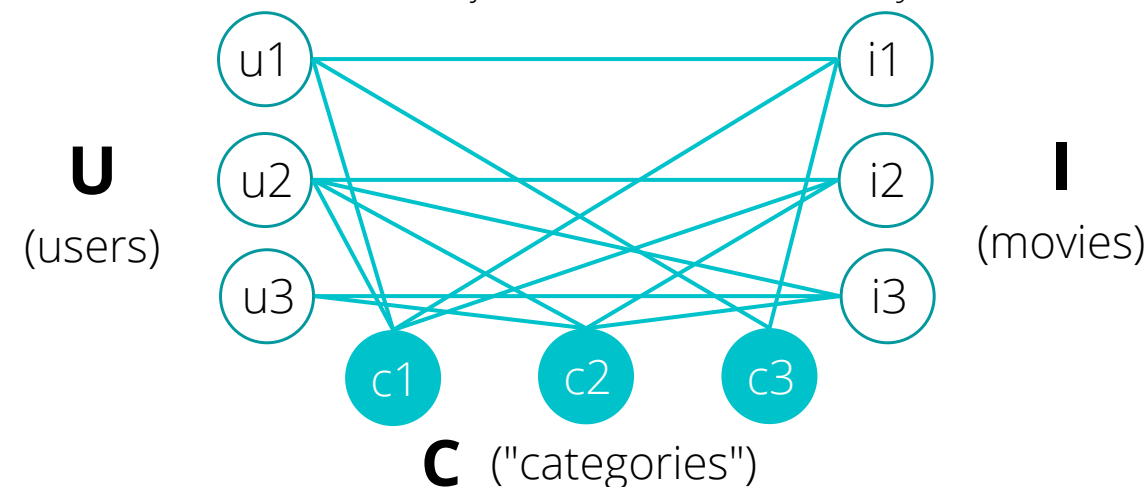


Figure 2: An example of a tripartite graph

4 Results

Recommendations for users are generated by traversing the tripartite graphs through a **Markov process**. This process effectively finds nodes 'closest' to a query user of interest. The recommendations generated were evaluated on **Recall@N** and **diversity** and results are shown in figure 3.

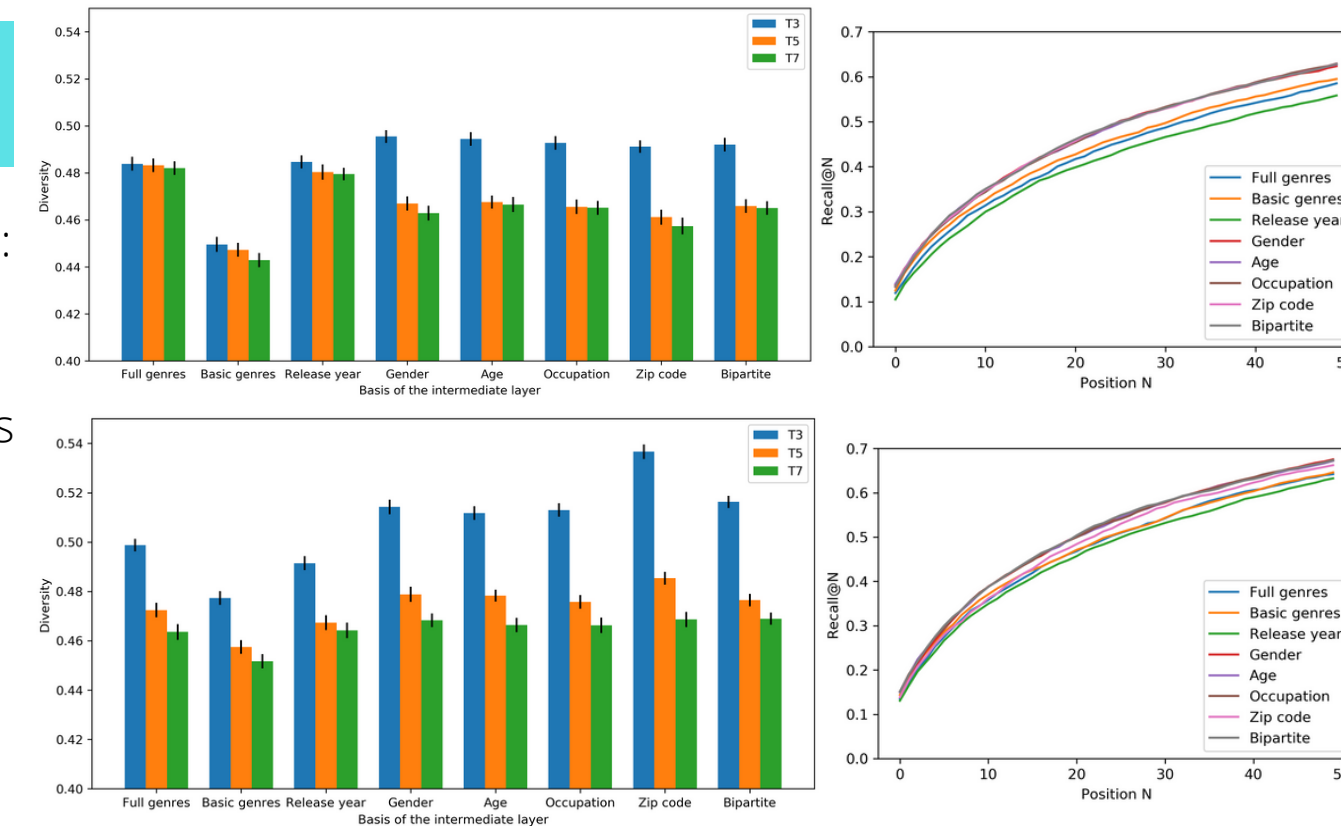


Figure 3: The results of the experiments conducted during the research

5 Conclusions

- **User-focused** graphs outperform item-focused graphs, yet the **bipartite graph** shows equal performance
- **Normalisation** of the transition matrices significantly improves the performance of all graph options
- Previous papers describing the graph-based recommendation models show **poor reproducibility**

Limitations

Only the **MovieLens dataset** and the categories it provides are considered.

6