

Evaluating The Effects of Calibrated Popularity Bias Mitigation: A Field Study

Anastasiia Klimashevskaja¹ (anastasiia.klimashevskaja@uib.no), Mehdi Elahi¹, Dietmar Jannach², Lars Skjærven³, Astrid Tessem³, Christoph Trattner¹

¹ University of Bergen, Norway
² University of Klagenfurt, Austria
³ TV 2, Norway

Media Futures



17th ACM Conference on Recommender Systems
Singapore, 18th-22nd
September 2023

Problem

- **Setting:** TV 2 – Norwegian national broadcaster, experiencing issues with **Popularity Bias** – a tendency of a Recommender System to **excessively** promote **highly popular items**
- The effect is **amplified** by the **feedback loop**
- Leads to reduced **discovery, diversity and novelty of content, lower user engagement and interest**
- **Goal:** Develop and evaluate a re-ranking method in an A/B test to mitigate bias, increasing user engagement and content diversity

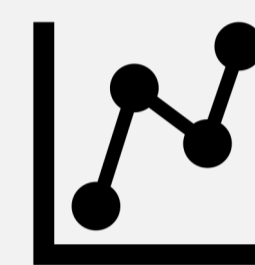
A/B Test Design



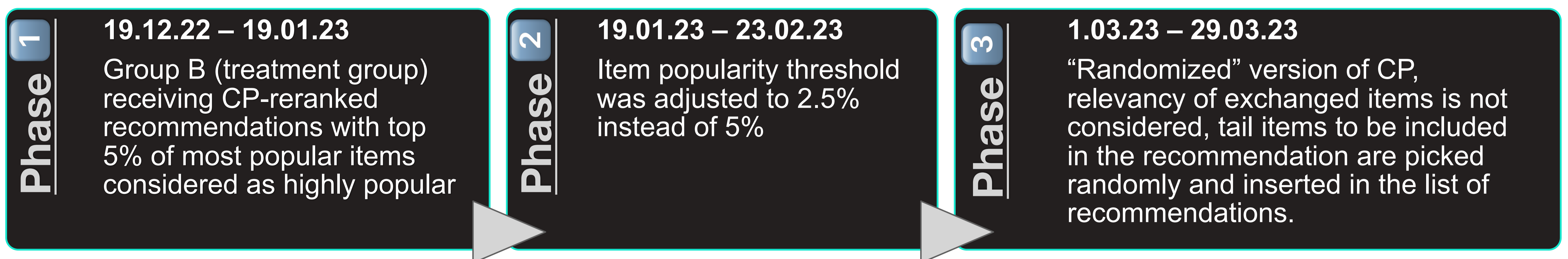
Calibrated Popularity (CP) – re-ranking that actively accounts for historical **user popularity preference**, while attempting to maintain high predicted relevance



Three consecutive A/B testing periods with varying conditions and parameters (see flowchart below)



Tracking **Click-Through Rates, Item Exposure and Watches**



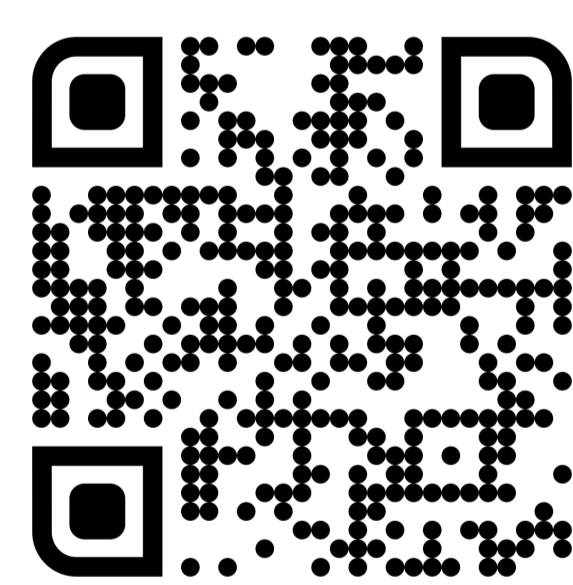
Results and Observations

We observed **CTR increase** for group B in Phases 1 and 3 (+1.25% and +2.25 respectively), **CTR drop** in Phase 2 (-4.18%)

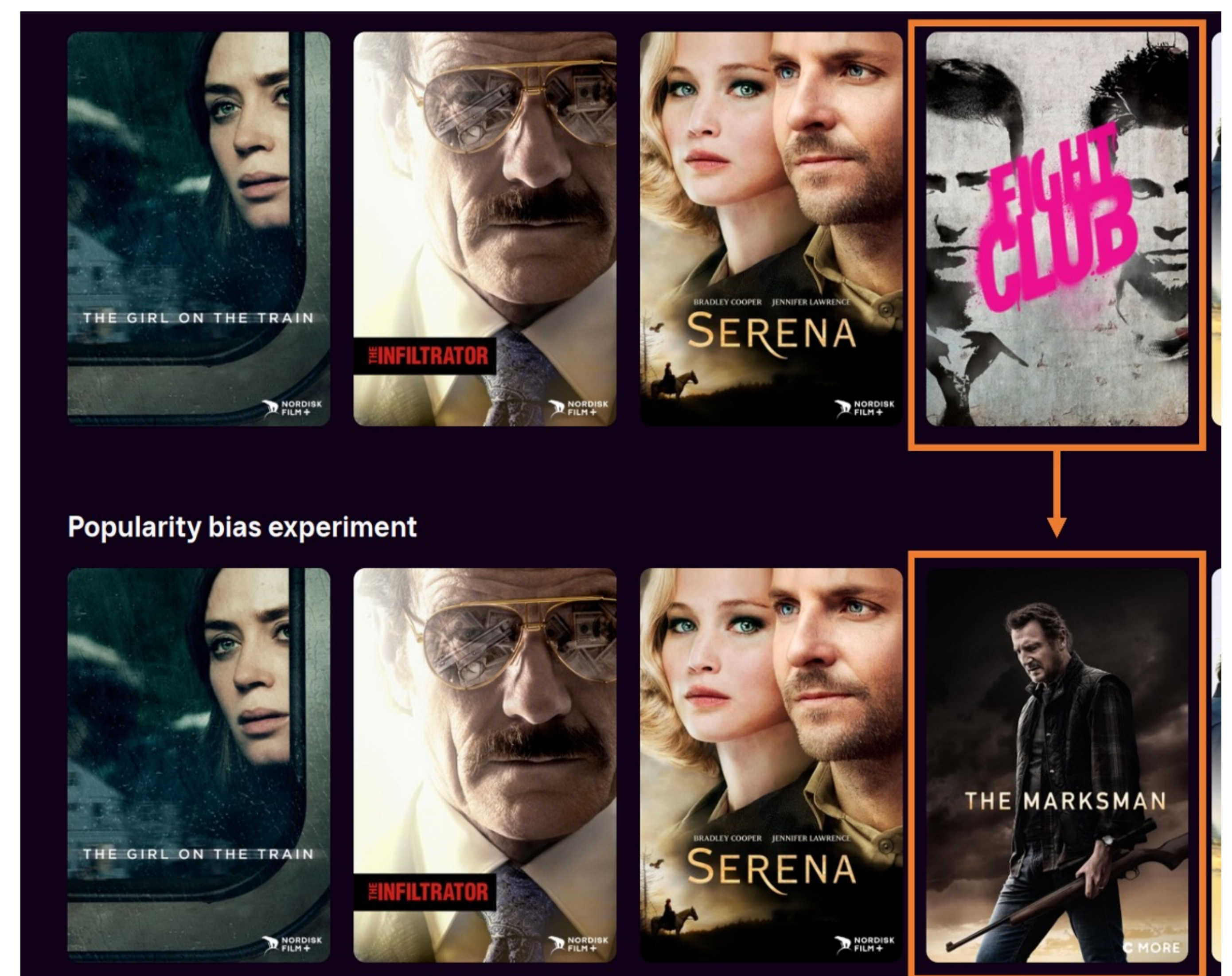
- Calibrated re-ranking can be an **effective** technique to provide less mainstream recommendations without negatively affecting user experience
- **Thresholds** and **parameters** used for the method need to be tuned and selected with **caution**
- Otherwise **undesired effects** can be observed
- Recommending a small amount of **random** items can also have **positive** effects on user exploration

More unique movie titles exposed to users in group A, while more **unique movie titles watched** by users in group B during all three phases

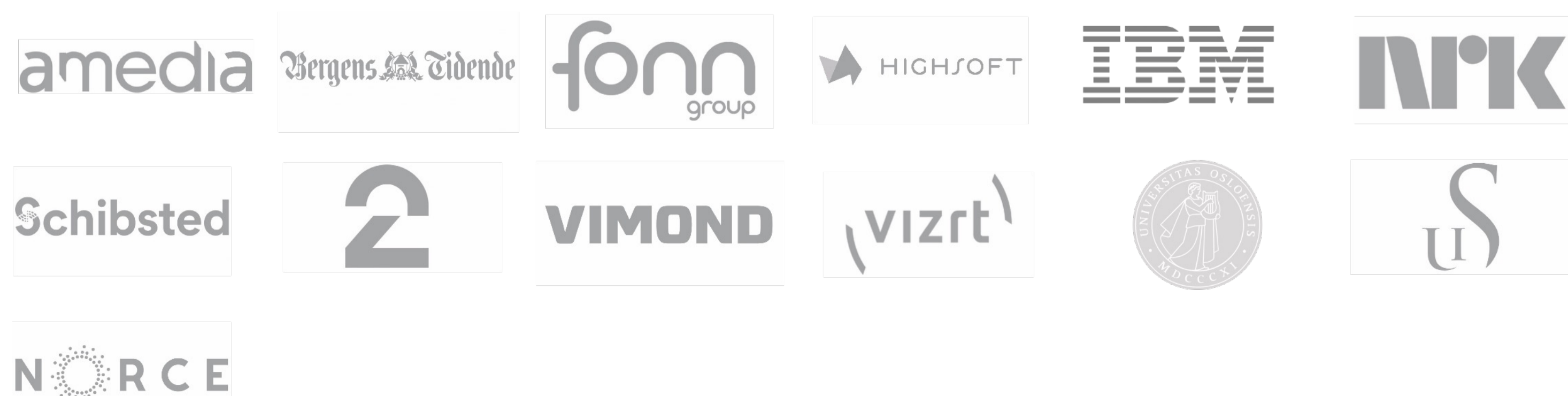
- This can be attributed to users in group A scrolling the recommendations more in search of novelty and serendipity
- At the same time group B requires less search and receives more novel recommendations



Read full paper



PARTNERS



HOST



UNIVERSITY OF BERGEN

FUNDER

This research is funded by SFI MediaFutures partners and the Research Council of Norway (grant number 309339).

