Introduction

Individuals may be impacted by recommended content. We study a dynamical model of biased assimilation proposed by HJMR [2019]: preferences become more aligned with content that is enjoyed, and anti-aligned with content that is disliked.

Preference Dynamics Model

An individual's preference $p_t \in S^{d-1}$ determines their response to content $q_t$, chosen from a set $Q \subset S^{d-1}$ via the affinity $p_t^\top q_t$. The affinity affects both the rating $y_t$ and the evolution of preferences along with step-size $\eta_t$.

$$y_t = p_t^\top q_t + w_t$$

$$\tilde{p}_{t+1} = p_t + \eta_t p_t^\top q_t \cdot q_t$$

$$p_t = \tilde{p}_t / \|\tilde{p}_t\|$$

Rating Maximization with Fixed Recommendations

The dynamics make it easy to achieve high ratings; as long as $Q$ contains opposites very little needs to be known about users or items.

Informal Result: As long as $|p_t q_t| > c$ and noise is $\sigma^2$ sub-Gaussian,

$$R(T) = \sum_{t=0}^{T-1} 1 - p_t^\top p_0 \lesssim \log T$$

(See paper for more general setting)

Stationary Preferences with Randomized Recommendations

Non-manipulation [KML20] is an alternative goal. Since it may be $p_0 \notin Q$, a randomized strategy selects $q_t$ i.i.d.

Informal Result: Suppose $p_0$ is the dominant eigenvector of $E[qq^\top]$, the randomization is aligned, and $\eta_t$ decays like $1/T$. Then

$$R(T) = \sum_{t=0}^{T-1} 1 - p_t^\top p_0 \lesssim \log T$$

Implications: Mode Collapse

While non-personalized consumption leads to polarization [HJMR19, GKT21], personalized recommendations may lead initial preference (grey) to collapse to a subset of $Q$. However, randomization can keep preferences stationary.

Richness of $Q$

The item set $Q$ must be sufficiently rich for

1. Estimating initial $p_0$ from observations $y_{0:T}$ requires that $\text{span}(y_{0:T}) = \mathbb{R}^d$

2. Designing randomization for stationarity requires that $p_0 \in \text{span}(\{q \mid q \in Q, p_0^\top q > 0\})$

References

- Gaitonde, Kleinberg, Tardos, 2021. Polarization in geometric opinion dynamics. EC.