When Collaborative Filtering is not Collaborative: Unfairness of PCA for Recommendations

David M. Liu
Northeastern University Khoury College of Computer Sciences

Brown CNTR Tech + Society Seminar
November 7, 2024

dliu18.github.io
liu.davi@northeastern.edu
@dayvidliu

Unfairness of ML for Decision Making

Algorithms Allegedly Penalized Black Renters. The US Government Is Watching

The Department of Justice warned a provider of tenant-screening software that its technology must comply with fair housing law.

ILLUSTRATION: JACQUI VANLIEW; GETTY IMAGES

Source: Wired

Study finds gender and skin-type bias in commercial artificial-intelligence systems

Examination of facial-analysis software shows error rate of 0.8 percent for light-skinned men, 34.7 percent for dark-skinned women.

Watch Video

Larry Hardesty | MIT News Office
February 11, 2018

Source: MIT News

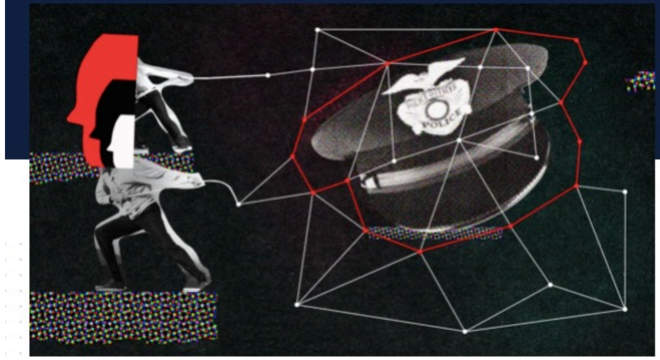
ARTIFICIAL INTELLIGENCE

Predictive policing algorithms are racist. They need to be dismantled.

Lack of transparency and biased training data mean these tools are not fit for purpose. If we can't fix them, we should ditch them.

By Will Douglas Heaven

July 17, 2020



FRANZISKA BARCZYK

Source: MIT Technology Review

Defining Group Fairness

Goal: balance classifier performance across sensitive-attribute groups

Statistical Parity – Corbett-Davies et al. (2017)

$$P(\hat{Y} = 1 \mid A = 0) = P(\hat{Y} = 1 \mid A = 1)$$

Equalized Odds – Hardt, Price, Srebro (2016)

$$P(\hat{Y} = 1 \mid A = 0, y = 1) = P(\hat{Y} = 1 \mid A = 1, y = 1)$$

Calibration – Chouldechova (2017)

$$P(Y = 1 \mid A = 0, S = s) = P(Y = 1 \mid A = 1, S = s)$$

Symbol	Meaning
Y	Ground truth label
$\hat{Y}$	Predicted label
A	Sensitive attribute
S	Risk score

Limitations of Existing Group Fairness Approaches

1. Rely on sensitive/demographic attributes
   - Are these labels available?
   - Do they characterize the appropriate/relevant groups for fairness?

2. Don't help us understand sources of *model* unfairness.
   - If a model is fair, how is it achieving fairness? And if it is not fair, why is it not?

Overview of my fairness work

Tackling the limitations:

1. Rely on demographic attributes
   - ➢ Defining group fairness with social networks [FAccT ‘23]

2. Don’t help us understand sources of unfairness
   - ➢ Identify mechanisms of unfairness in PCA collaborative filtering [In Submission]

When Collaborative Filtering is not Collaborative: Unfairness of PCA for Recommendations

In submission

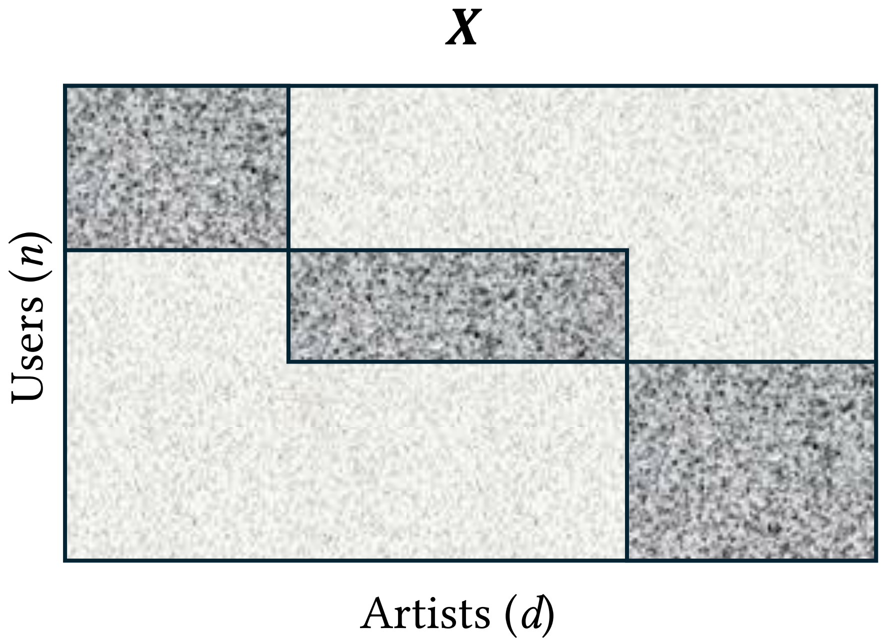
David Liu
Northeastern

Jackie Baek
NYU Stern

Tina Eliassi-Rad
Northeastern

A Motivating Example

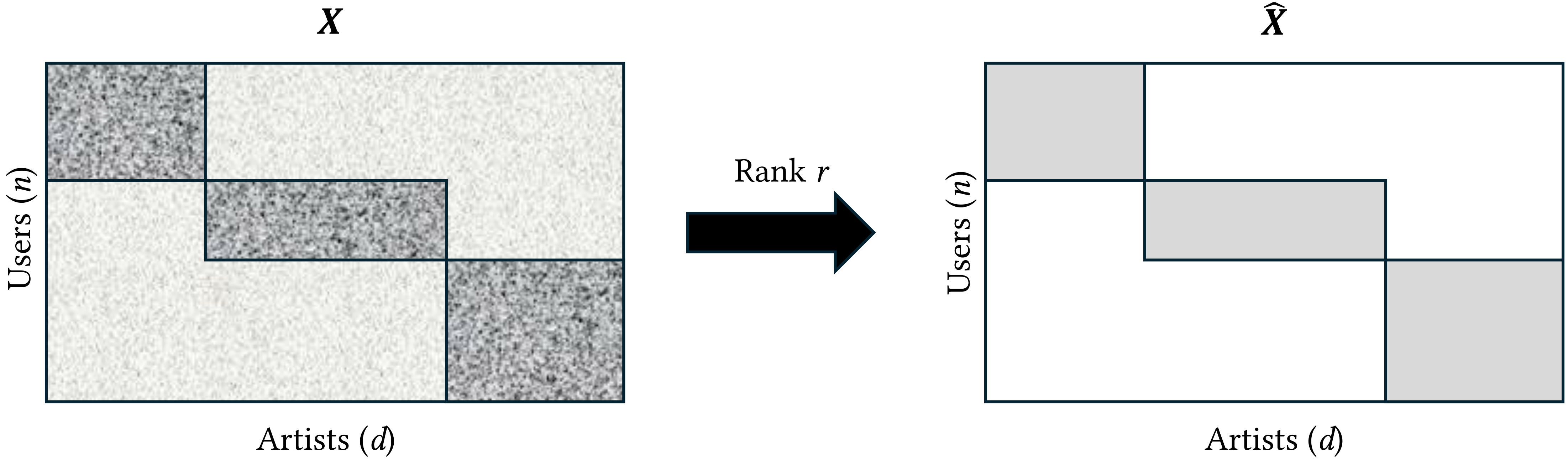
Last.fm music listening dataset of user listening counts (Cantador et al., 2011)



A Motivating Example

Last.fm music listening dataset of user listening counts (Cantador et al., 2011)

The promise of low rank: with a few latent dimensions you can well approximate a high-dimensional matrix.



A Motivating Example

Globally, as the rank budget r increases,
$|X - \hat{X}|^2$ is (exponentially) decreasing.

However, are all portions of the matrix equally well approximated?

Let *Normalized Item Error* be the artist-level error as r increases:

$$\text{Normalized Item Error} = \frac{\left|X_{.j} - \hat{X}_{.j}\right|^2}{\left|X_{.j}\right|^2}$$

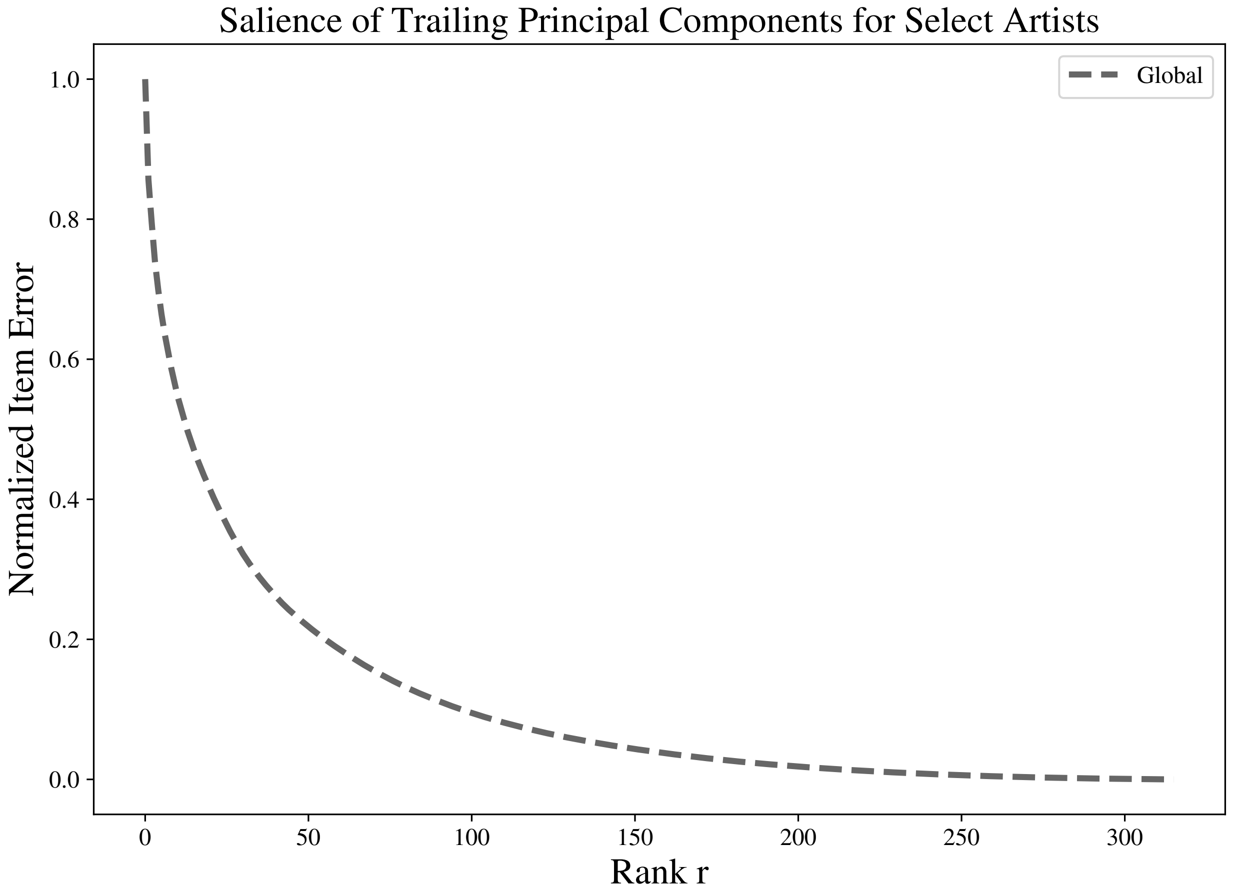
A Motivating Example

Globally, as the rank budget r increases,
$|X - \hat{X}|^2$ is (exponentially) decreasing.

However, are all portions of the matrix equally well approximated?

Let *Normalized Item Error* be the artist-level error as r increases:

$$\text{Normalized Item Error} = \frac{\left|X_{.j} - \hat{X}_{.j}\right|^2}{\left|X_{.j}\right|^2}$$



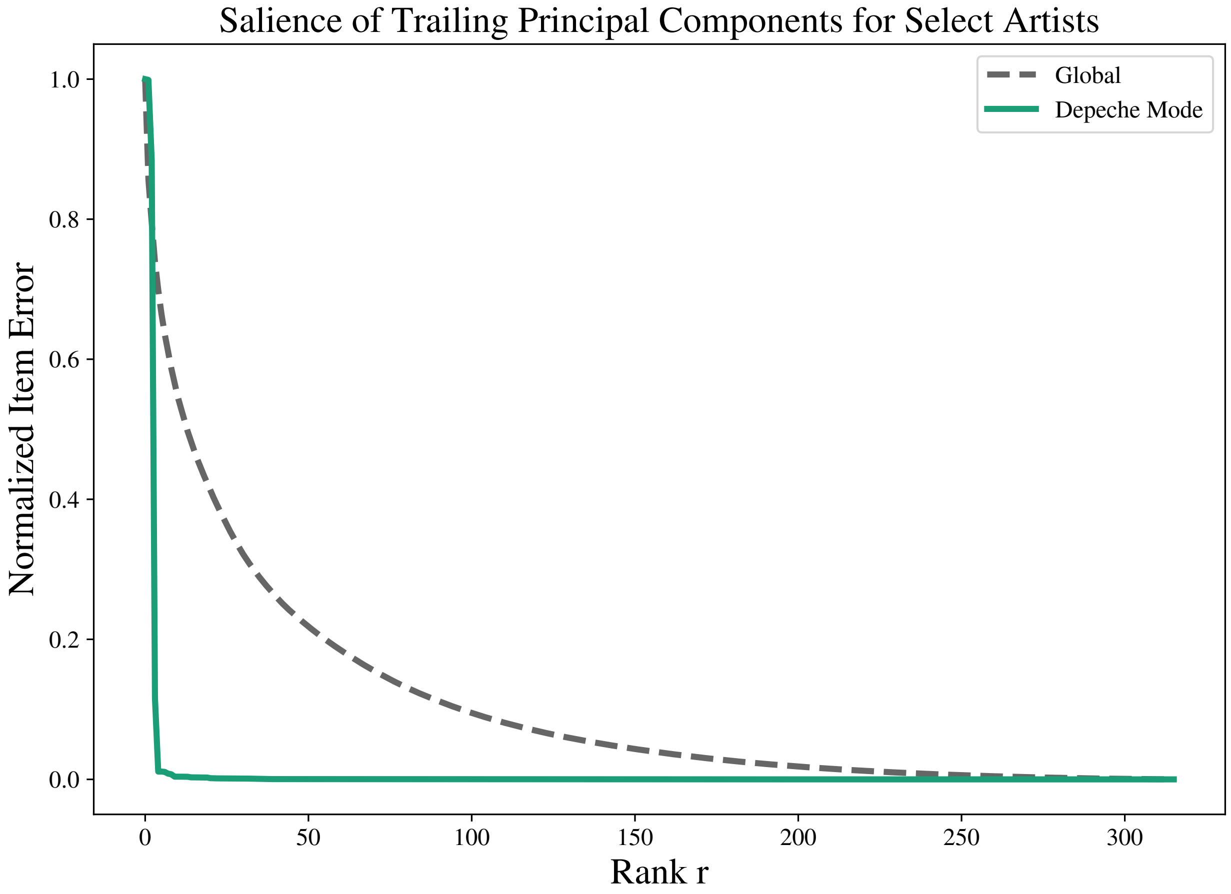
A Motivating Example

Globally, as the rank budget r increases, $|X - \hat{X}|^2$ is (exponentially) decreasing.

However, are all portions of the matrix equally well approximated?

Let *Normalized Item Error* be the artist-level error as r increases:

$$\text{Normalized Item Error} = \frac{|X_{.j} - \hat{X}_{.j}|^2}{|X_{.j}|^2}$$



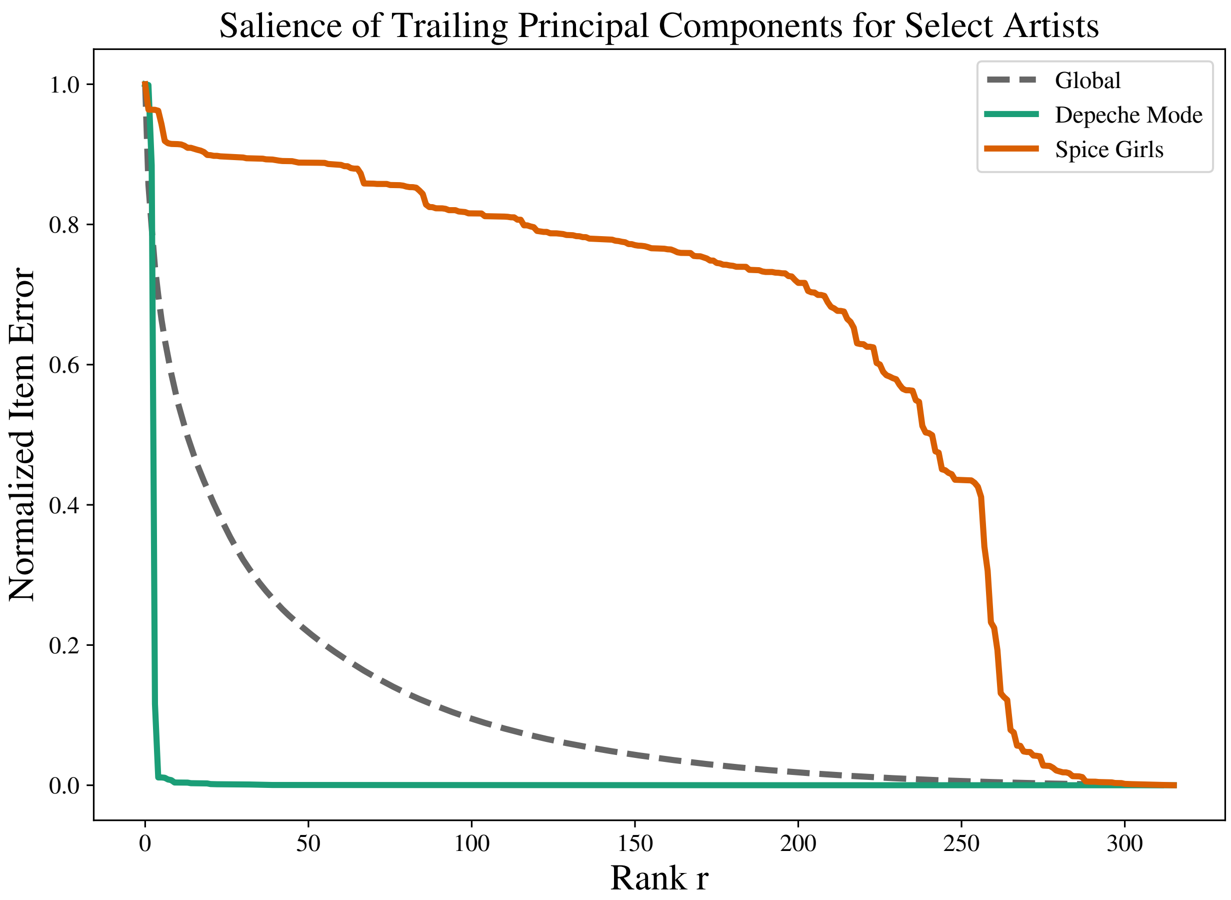
A Motivating Example

Globally, as the rank budget r increases, $|X - \hat{X}|^2$ is (exponentially) decreasing.

However, are all portions of the matrix equally well approximated?

Let *Normalized Item Error* be the artist-level error as r increases:

$$\text{Normalized Item Error} = \frac{\left|X_{.j} - \hat{X}_{.j}\right|^2}{\left|X_{.j}\right|^2}$$



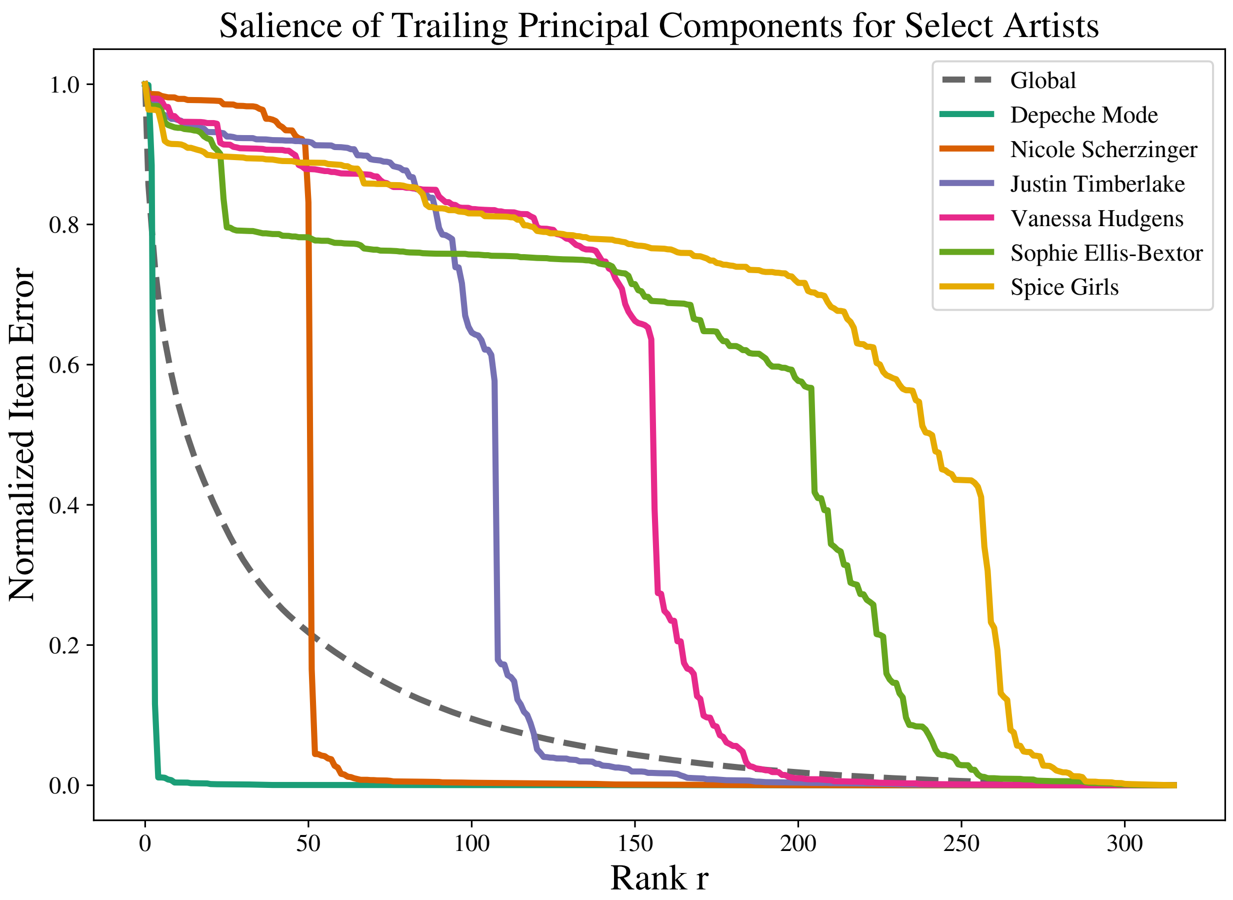
A Motivating Example

Globally, as the rank budget r increases, $|X - \hat{X}|^2$ is (exponentially) decreasing.

However, are all portions of the matrix equally well approximated?

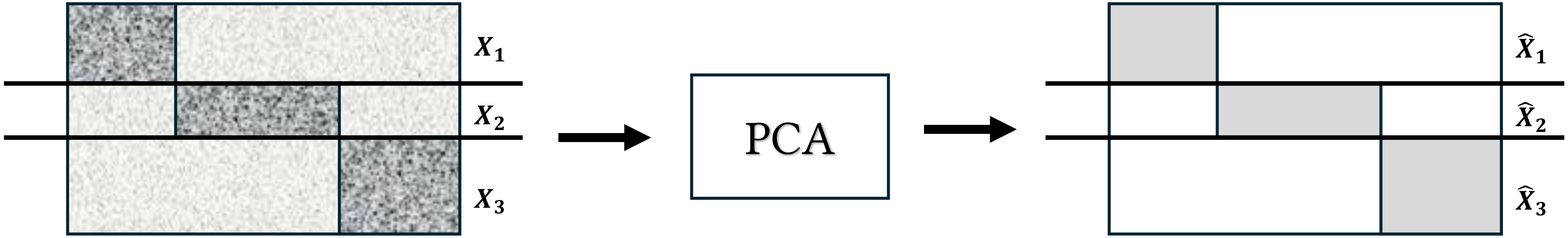
Let *Normalized Item Error* be the artist-level error as r increases:

$$\text{Normalized Item Error} = \frac{\left|X_{.j} - \hat{X}_{.j}\right|^2}{\left|X_{.j}\right|^2}$$



From Fairness Definitions to Mechanisms

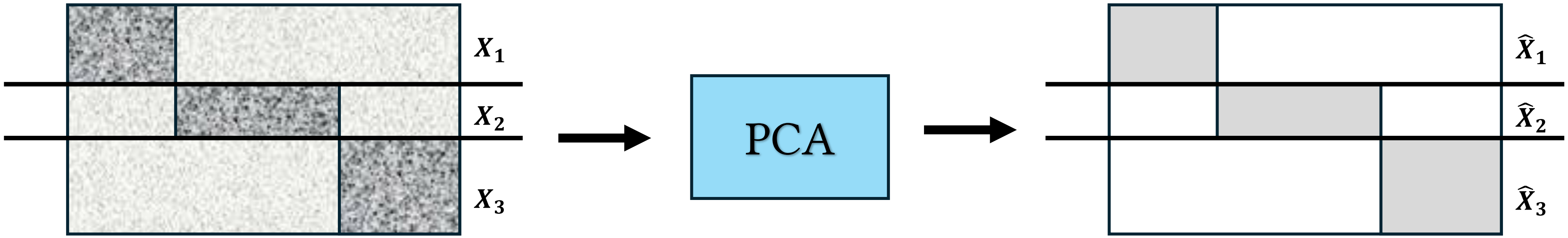
Prior work on fair PCA



$$f\left(\left|X_1 - \hat{X}_1\right|^2, \dots, \left|X_g - \hat{X}_g\right|^2\right)$$

From Fairness Definitions to Mechanisms

Prior work on fair PCA



$$f\left(\left|X_1 - \hat{X}_1\right|^2, \dots, \left|X_g - \hat{X}_g\right|^2\right)$$

RQ: What are the mechanisms of unfairness in PCA in the first place?

Outline

- **Mechanisms of unfairness in PCA**
  - Mechanism 1: unfairness for unpopular items
  - Mechanism 2: unfairness for popular items
- ***Item Weighted PCA*: an item re-weighting framework algorithm**
  - Efficient algorithm for improving representations of unpopular items
  - Optimality in stylized setting
- **Recommender system evaluation**
  - Improved recommendations for both popular and unpopular items

Mechanisms of Unfairness in PCA

Recap of PCA and Collaborative Filtering

PCA refresher: identifying r basis vectors to project data

$$\boxed{\begin{aligned}&\underset{P=UU^T}{\operatorname{argmin}} \|X - XP\|_F^2 \\ &\quad \text{s.t.} \quad U \in \mathbb{R}^{d\times r}, U^T U = I_r\end{aligned}}$$

Recap of PCA and Collaborative Filtering

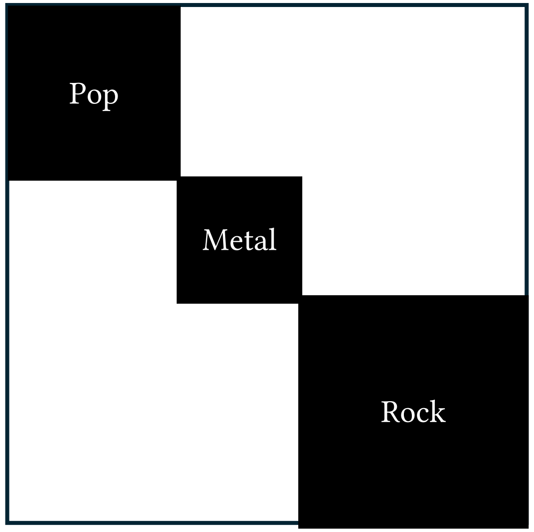
PCA refresher: identifying r basis vectors to project data

$$\underset{P=UU^T}{\operatorname{argmin}} \|X - XP\|_F^2 \quad \text{s.t.} \quad U \in \mathbb{R}^{d\times r}, U^T U = I_r$$

Collaborative filtering: leverage similarities among items to infer missing values in X.

Think of P as a d x d matrix of item similarities

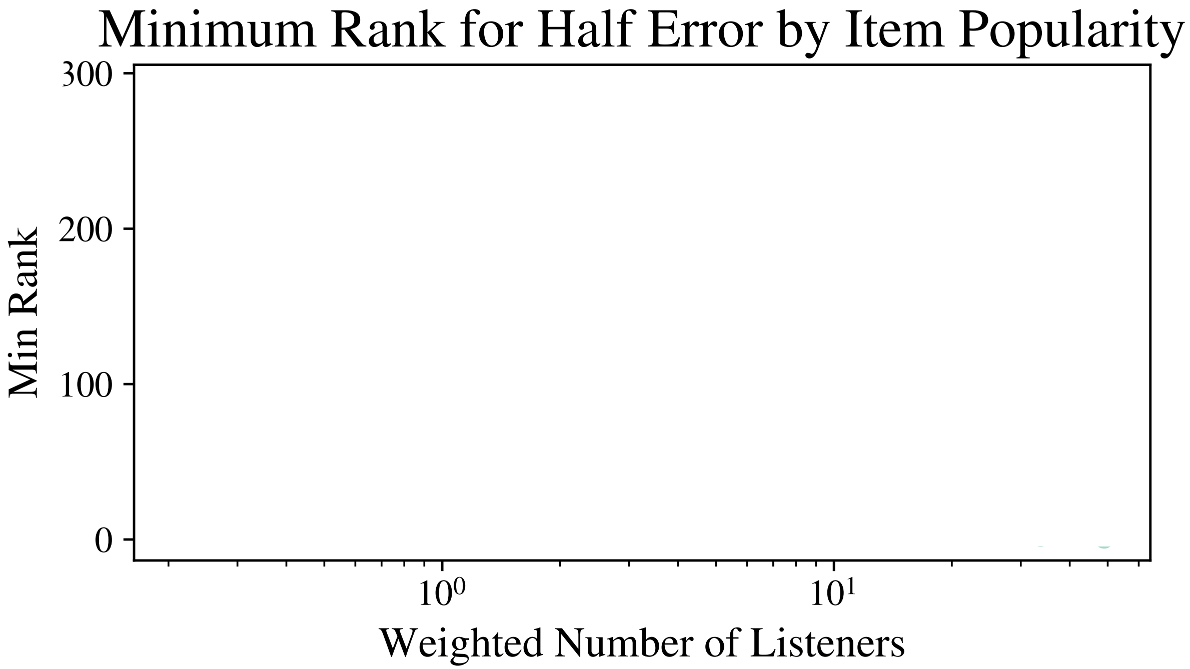
$$\hat{X}_{ij} = \sum_{j'} P_{jj'} X_{ij'}$$



Mechanism 1: Unpopular Items

The leading principal components disproportionately reconstruct entries for popular items.

$$argmin_r \frac{\left|X_{.j} - \hat{X}_{.j}\right|^2}{\left|X_{.j}\right|^2} < \frac{1}{2}$$

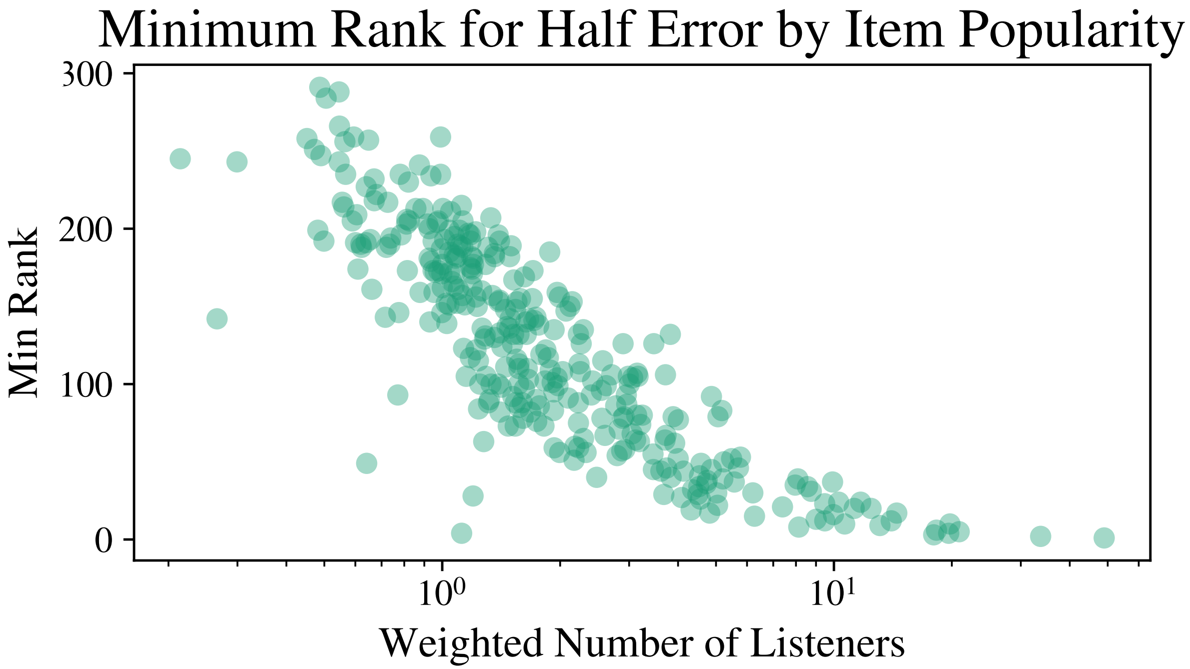


popularity

Mechanism 1: Unpopular Items

The leading principal components disproportionately reconstruct entries for popular items.

$$argmin_r \frac{\left|X_{.j} - \hat{X}_{.j}\right|^2}{\left|X_{.j}\right|^2} < \frac{1}{2}$$



popularity

Mechanism 2: Popular Items

Leading principal components specialize on individual artists as opposed to learning group information.

$$\hat{X}_{ij} = P_{jj}X_{ij} + \sum_{j' \neq j} P_{jj'}X_{ij'}$$

$$P_{jj} \approx 1 \text{ and } P_{jj'} \approx 0$$

Specialization quantified by diagonal values of P

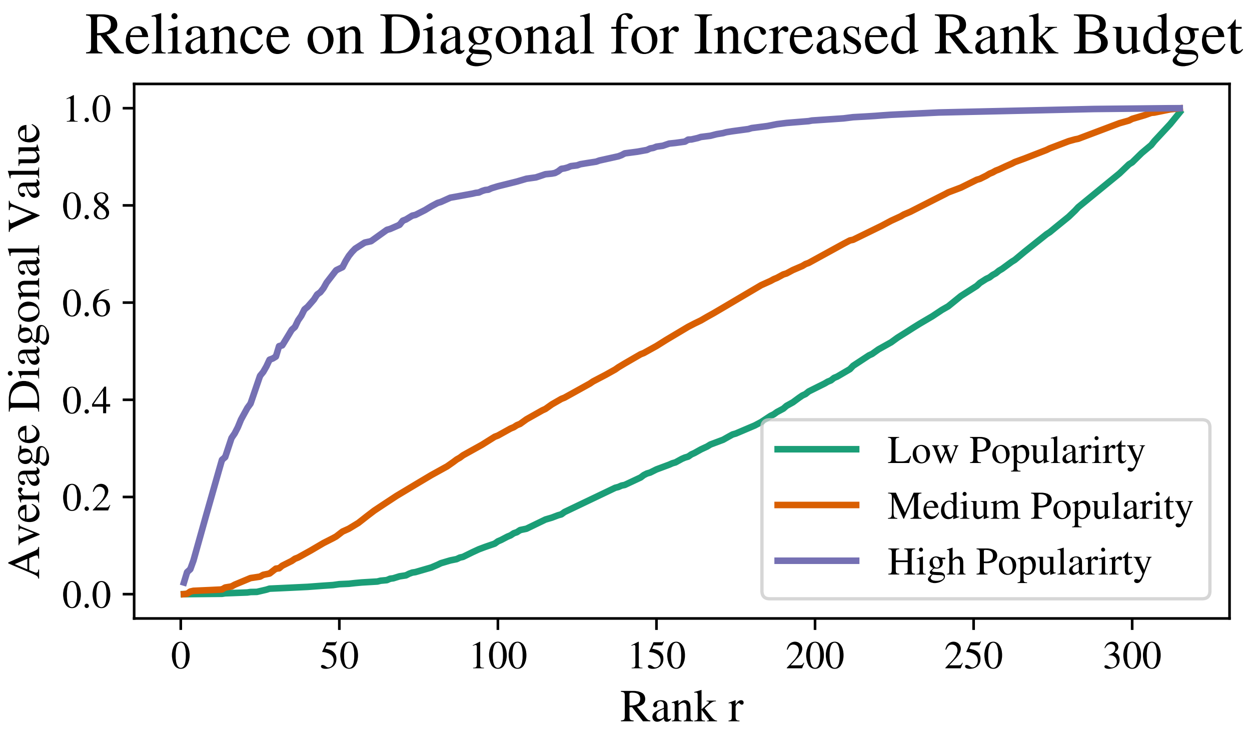
Mechanism 2: Popular Items

Leading principal components specialize on individual artists as opposed to learning group information.

$$\hat{X}_{ij} = P_{jj}X_{ij} + \sum_{j' \neq j} P_{jj'}X_{ij'}$$

$$P_{jj} \approx 1 \text{ and } P_{jj'} \approx 0$$

Specialization quantified by diagonal values of P



Item-Weighted PCA

An item reweighting algorithm for improved recommendations

Item-Weighted PCA

Main idea: ensure that $\hat{X}$ still reflects interests in unpopular items.

$$obj = \sum_{ij} w_j \left(S_{ij} * \hat{X}_{ij}\right)$$

w_j upweights less popular items

S_{ij}= sign(X_{ij}). Ensure and have the same sign.

$$\boxed{\begin{aligned} &\operatorname*{argmax}_{P} \sum_{j=1}^{d} w_j \ \langle S_{.j}, \hat{X}_{.j} \rangle \\ &\text{s.t.} \quad \operatorname{tr}(P) \le r, 0 \preceq P \preceq 1 \end{aligned}}$$

Constraints are convex relaxation of rank(P) ≤ r. In paper we show this constraint is tight.

Baselines as Instances of Item Weighted PCA

Assume: all popular items have the same popularity ($\sum_j X_{ij} = n_p$) and all unpopular items have the same popularity ($\sum_j X_{ij} = n_u$)

Two baselines:

- Vanilla PCA
- Column-normalized PCA: normalize columns of X before PCA

Both baselines are instances of Item-Weighted PCA but with differing weights.

Baselines as Instances of Item Weighted PCA

If #unpopular = $\sqrt{\frac{n_p}{n_u}}$ #popular and popularity gap is sufficiently large*

Eigenvectors of X_p: $\boldsymbol{v_1, v_2, \ldots, v_r}$ and X_u: $\boldsymbol{v_1, v_2, \ldots, v_r}$

Vanilla PCA $\quad \boldsymbol{v_1, v_2, \ldots, v_r, v_1, v_2, \ldots, v_r}$

Item-Weighted PCA $\quad \boldsymbol{v_1, v_1, v_2, v_2, v_3, v_3 \ldots}$

Normalized PCA $\quad \boldsymbol{v_1, v_2, \ldots, v_r, v_1, v_2, \ldots, v_r}$

* Details in paper

Recommender System Evaluation

Evaluation Metric and Datasets

Recommendation-based evaluation metric:

$$\frac{1}{d}\sum_{i-1}^{d} \mathsf{AUC}\left(XP'_j, y_j\right) \qquad \text{(Item AUC-ROC)}$$

P’ = P - I

Datasets

- Last.fm
  - 920 users and 316 artists.
  - Listening counts for each user and artist pair (implicit feedback)
- Movielens
  - 2,000 users and 308 movies
  - Ratings on a 5-star scale (explicit feedback)

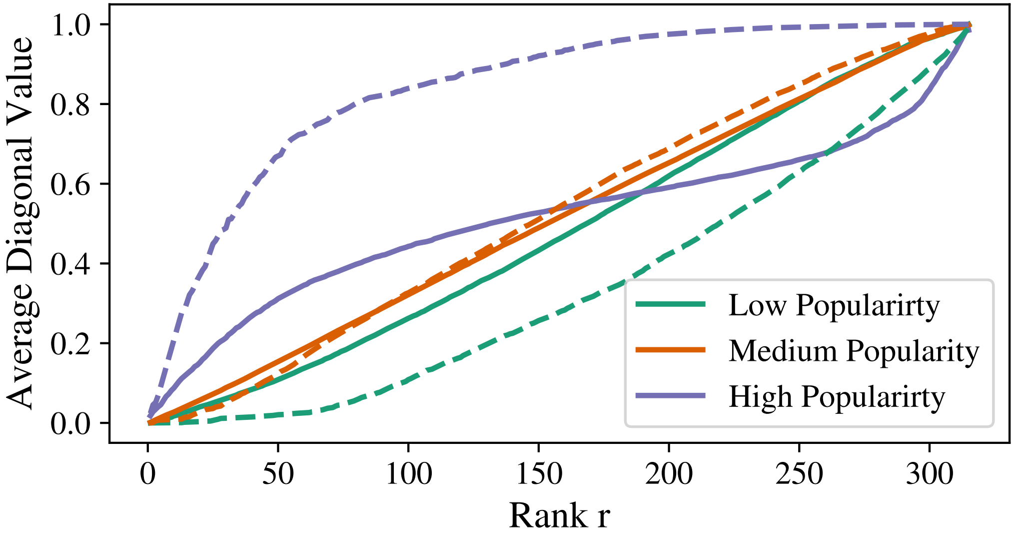
Algorithms:

- Item-weighted PCA
- Vanilla PCA
- Column-Normalized PCA

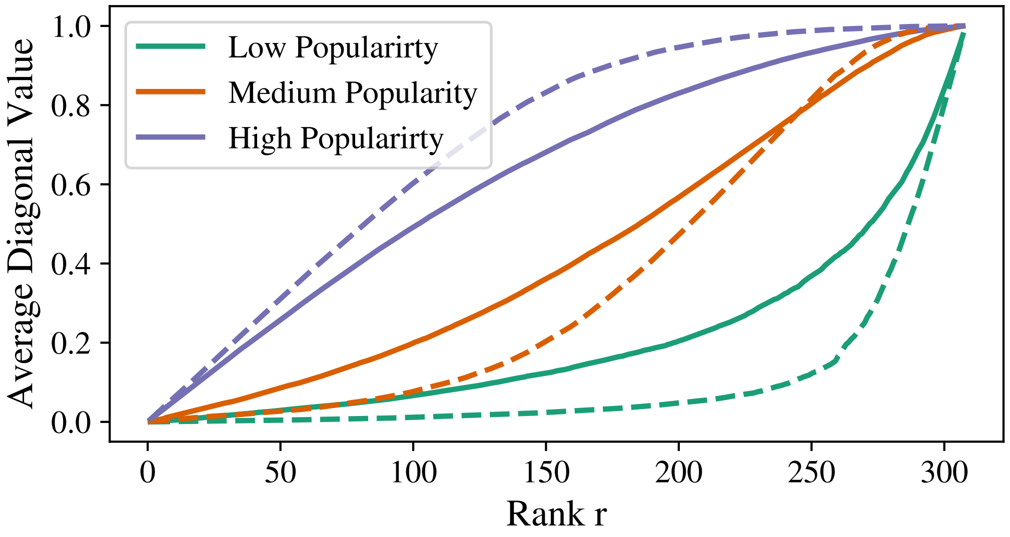
Reduced Specialization

Diagonal value as a heuristic for specialization

LastFM



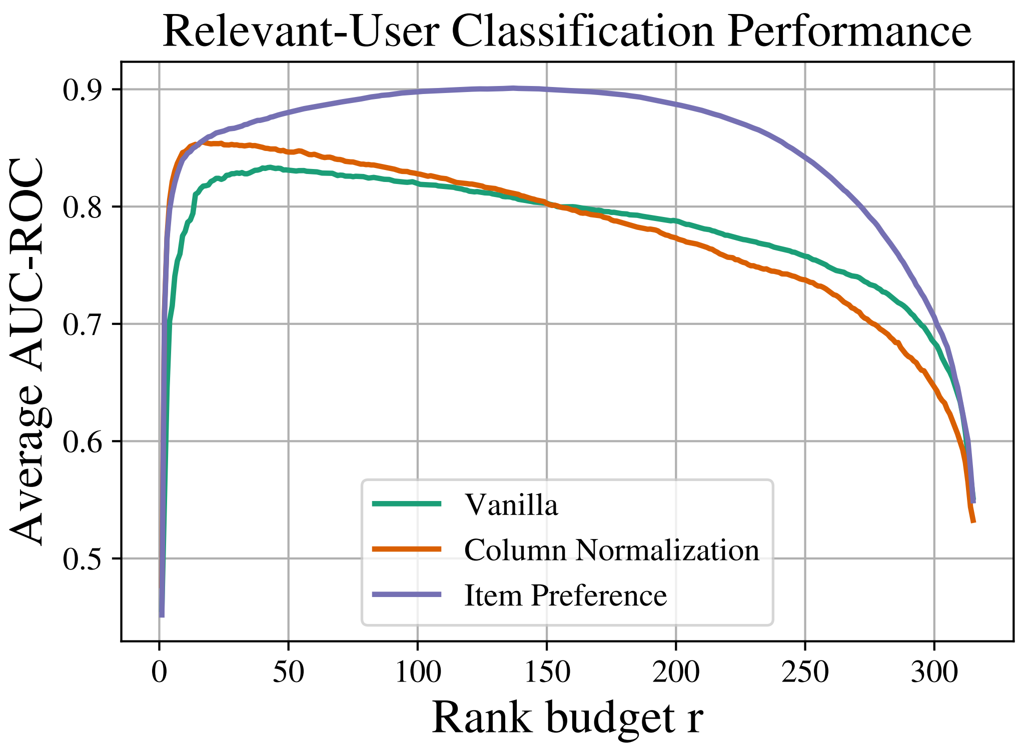
MovieLens



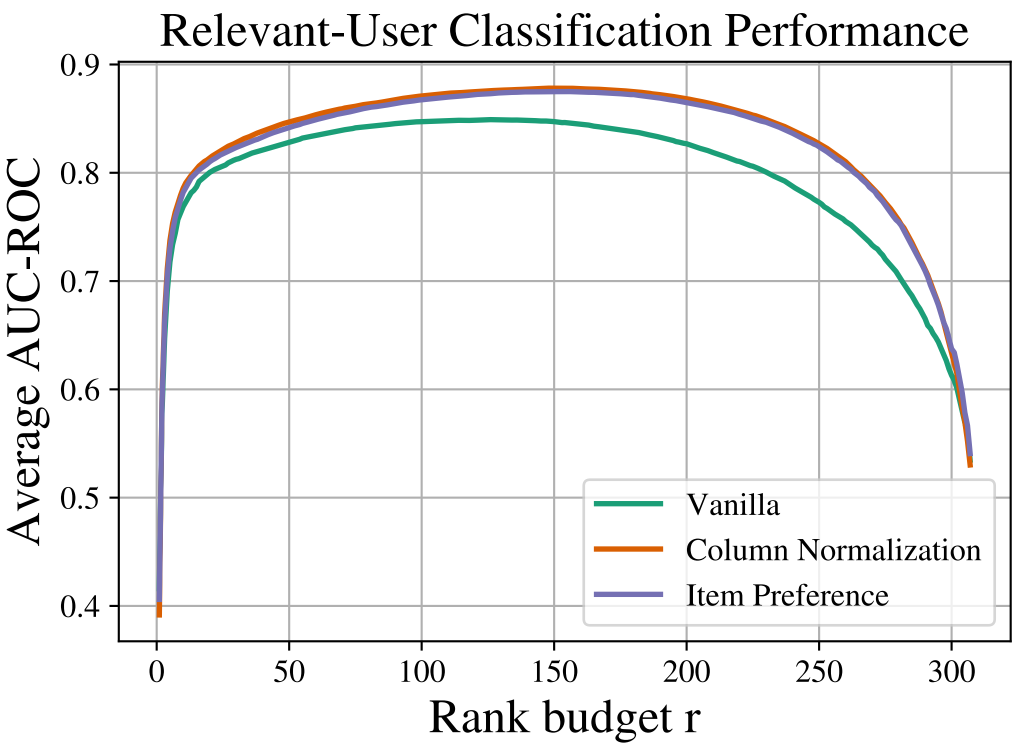
<u>Dashed</u>: Vanilla PCA
<u>Solid</u>: Item-Weighted PCA

Improved User Classification

LastFM

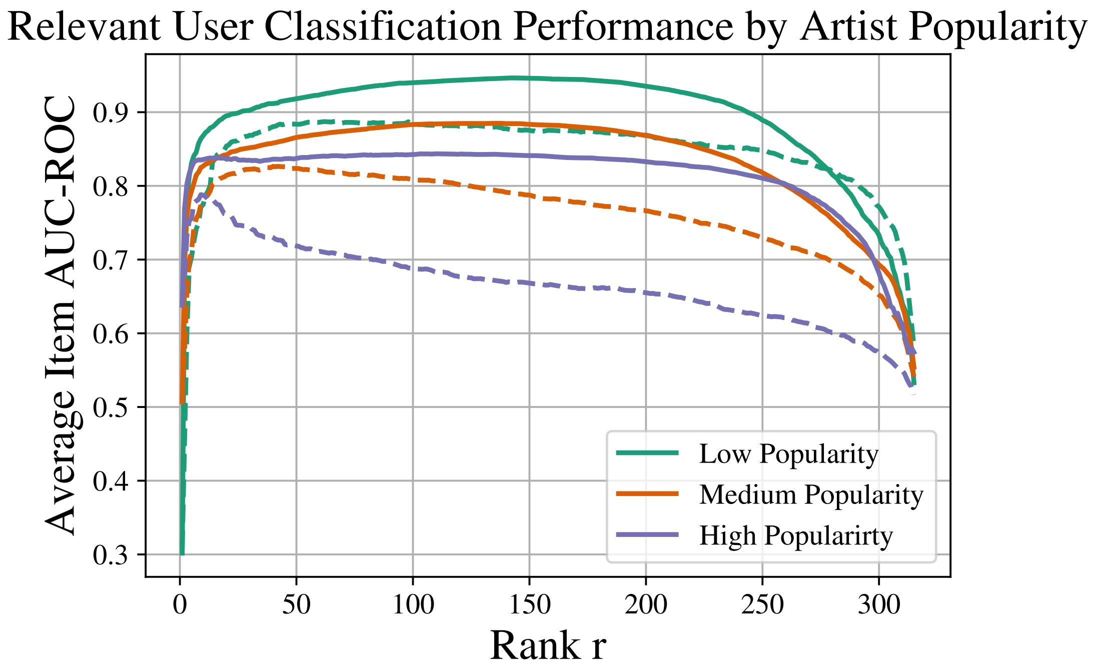


MovieLens

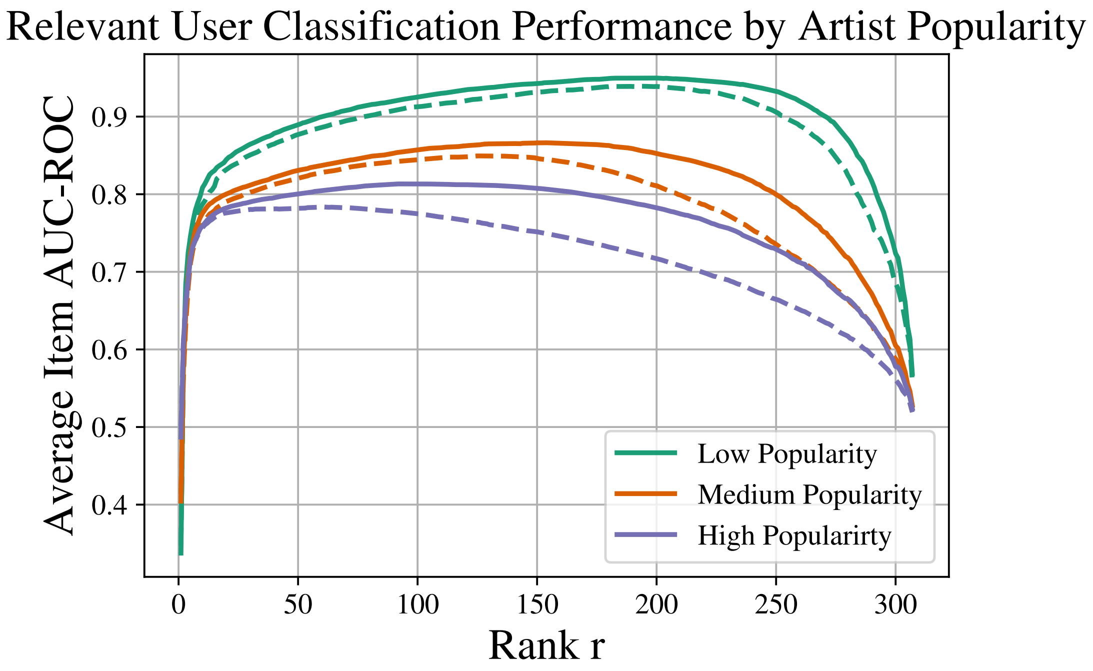


Improvement for Popular and Unpopular Artists

LastFM



MovieLens



Limitations of Item-Weighted PCA

- Solving the SDP runs in $O(d^3)$
- By upweighting unpopular items, Item-Weighted PCA may overfit to noisy data.
- The principal components are not ordered i.e. the solution for rank r+1 is not simply the solution for rank r plus an additional vector.

In-Progress Follow-up Questions

- How can we define and approach algorithmic fairness from the perspective of designing models that capture **what makes groups/individuals *unique or different*?**

- How do the preferences and behaviors of a majority group on a digital platform impact the recommendations for a minority group?

- What are the theoretical **identifiability limits** of long-tail preferences in recommender systems?

Questions?

When Collaborative Filtering is not Collaborative: Unfairness of PCA for Recommendations
David Liu, Jackie Baek, Tina Eliassi-Rad
In Submission
[arXiv 2310.09687]

On the market for postdocs beginning in Fall 2025
Please reach out if you know of any opportunities!

dliu18.github.io

liu.davi@northeastern.edu

@dayvidliu