Skewness Ranking Optimization for Personalized Recommendation

Yu-Neng Chuang
CS, National Chengchi University, Taiwan

Joint work with Prof. Chuan-Ju Wang,
Chih-Ming Chen, Prof. Ming-Feng Tsai

August 5, 2020
UAI, Toronto, Canada

中央研究院
ACADEMIA SINICA

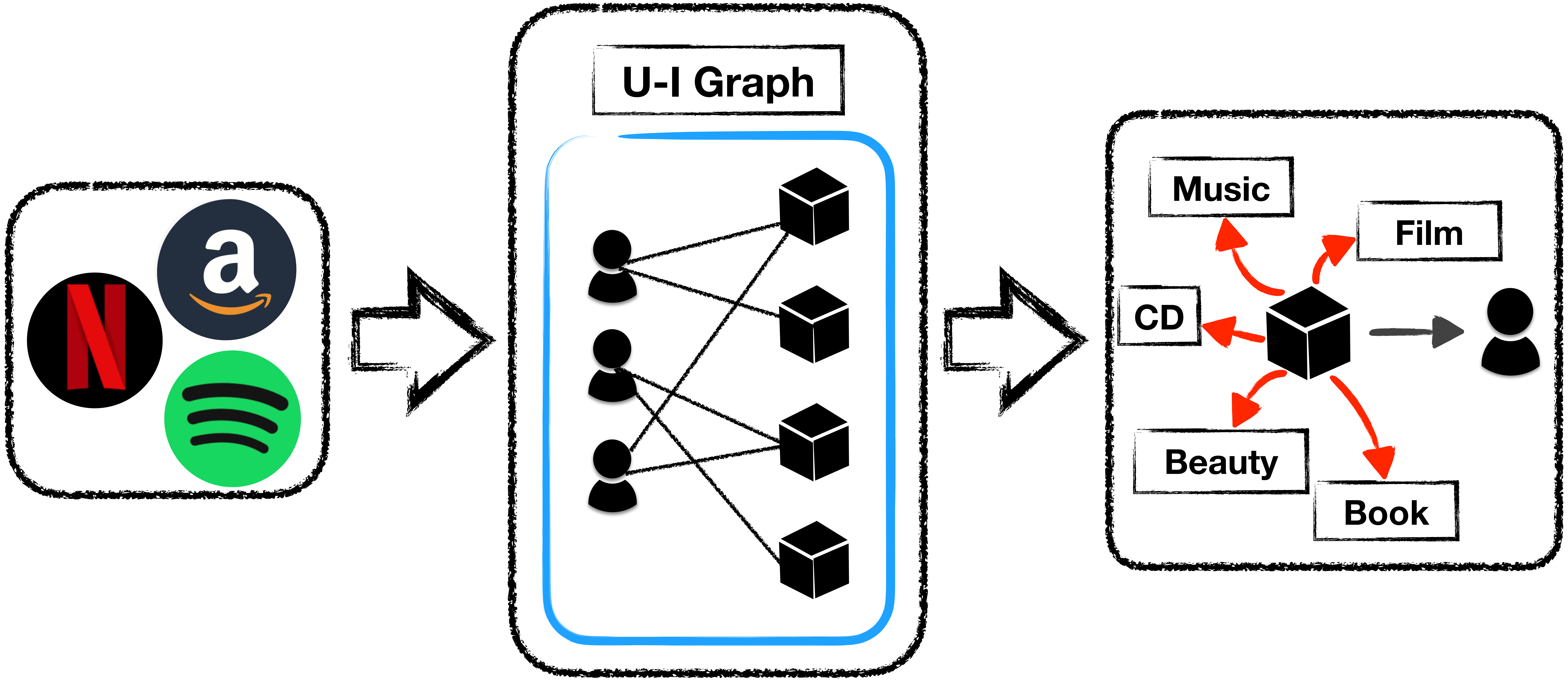
政治大学
NATIONAL CHENGCHI UNIVERSITY

Agenda

- Introduction
- Preliminaries and Observation
- Skewness Ranking Optimization (Skew-OPT)
- Experiment Results
- Conclusion

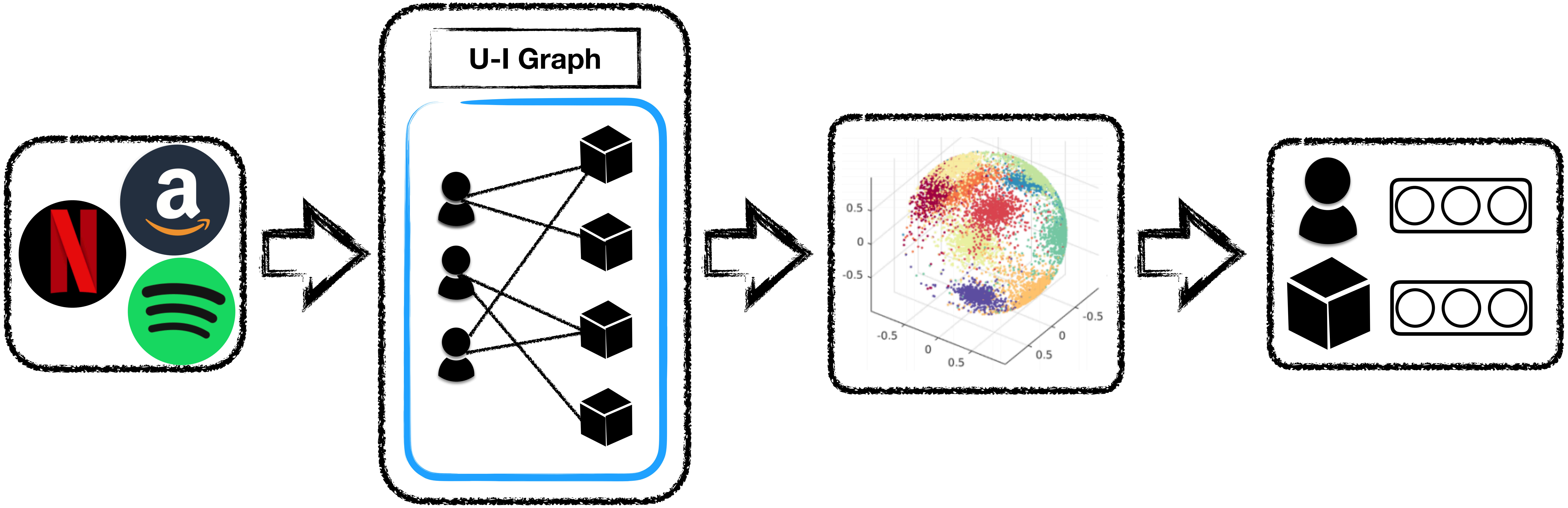
Introduction

- What is personalized recommendation?



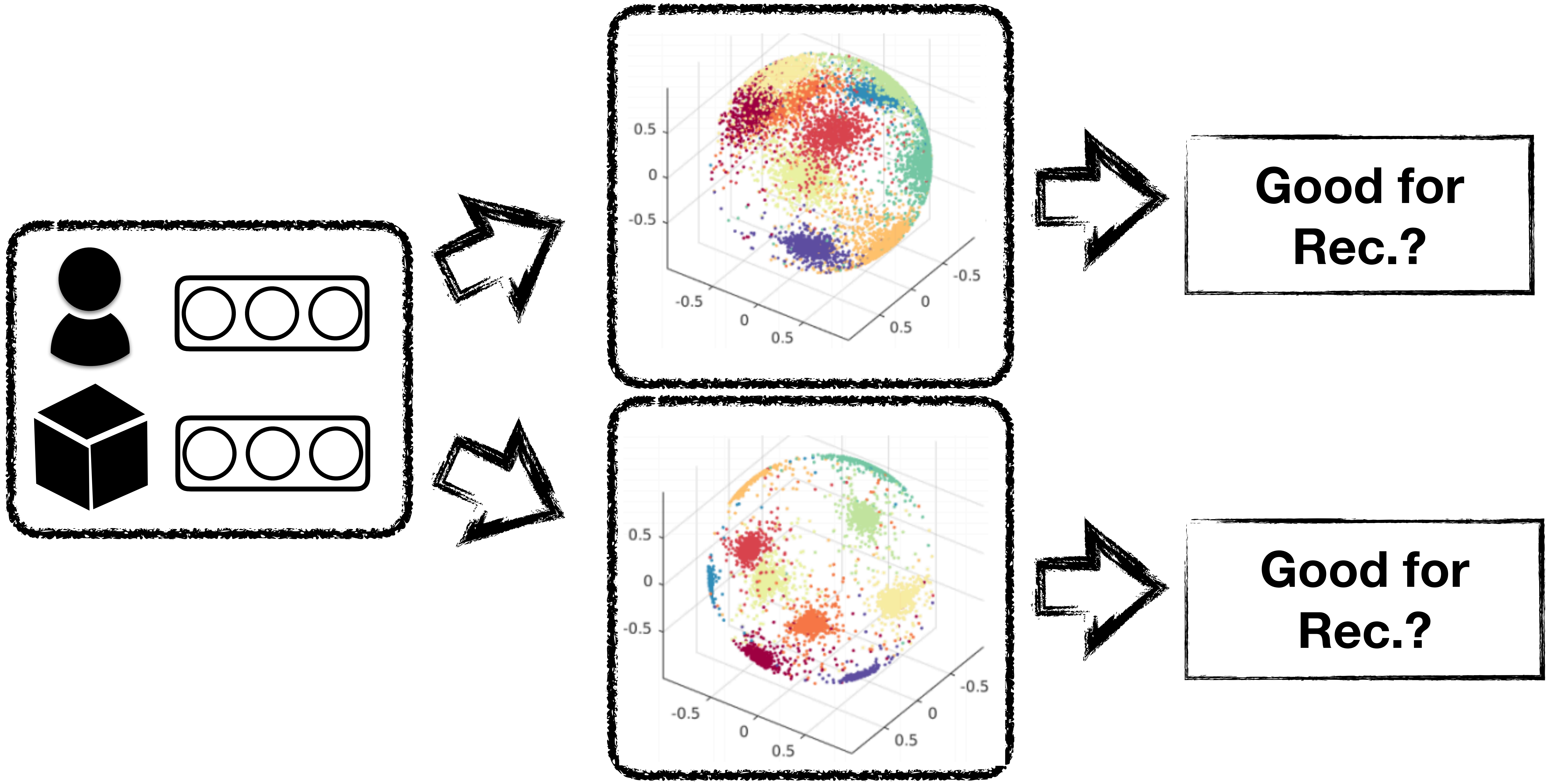
Introduction

- Each user/item can be projected into an embedding.
- All embeddings form a distribution.



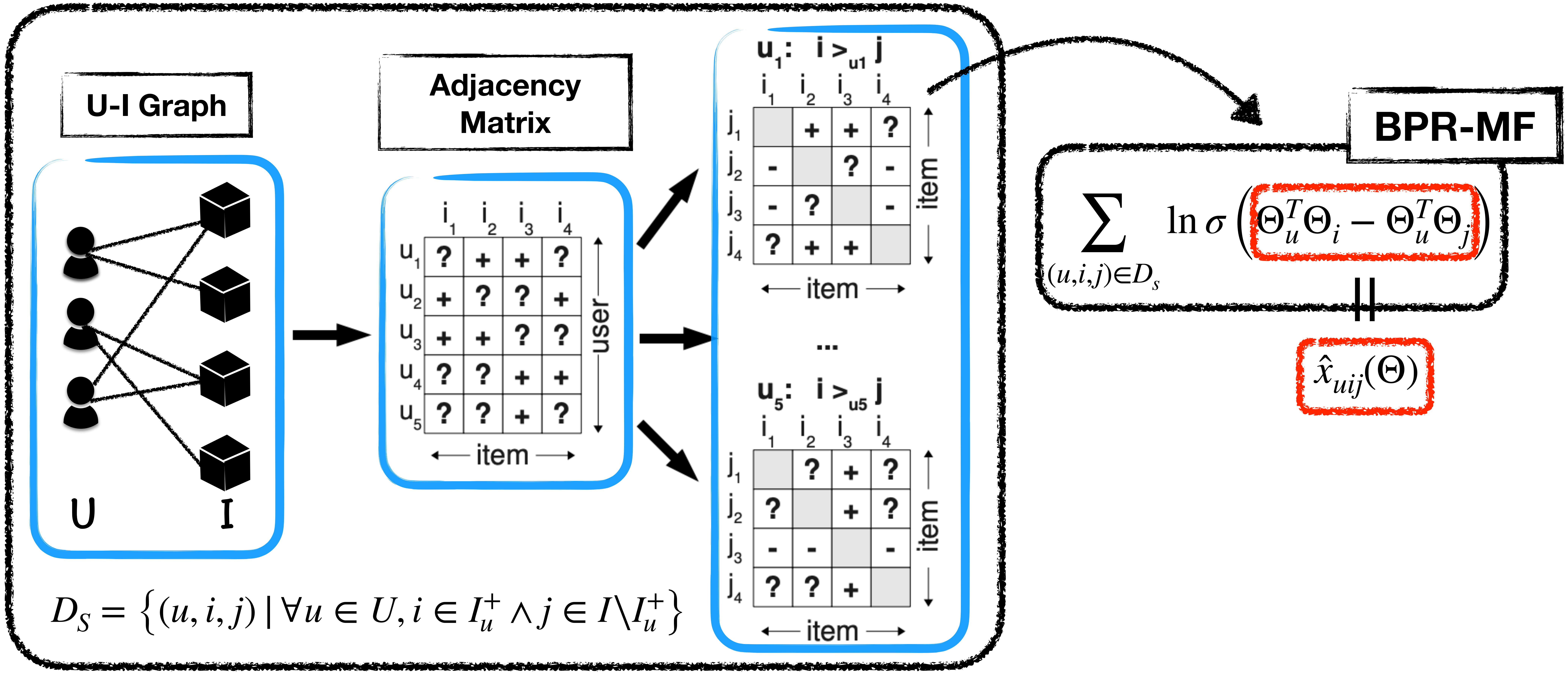
Introduction

- Goal: Find a distribution that is good for recommendation.



Preliminaries and Observation

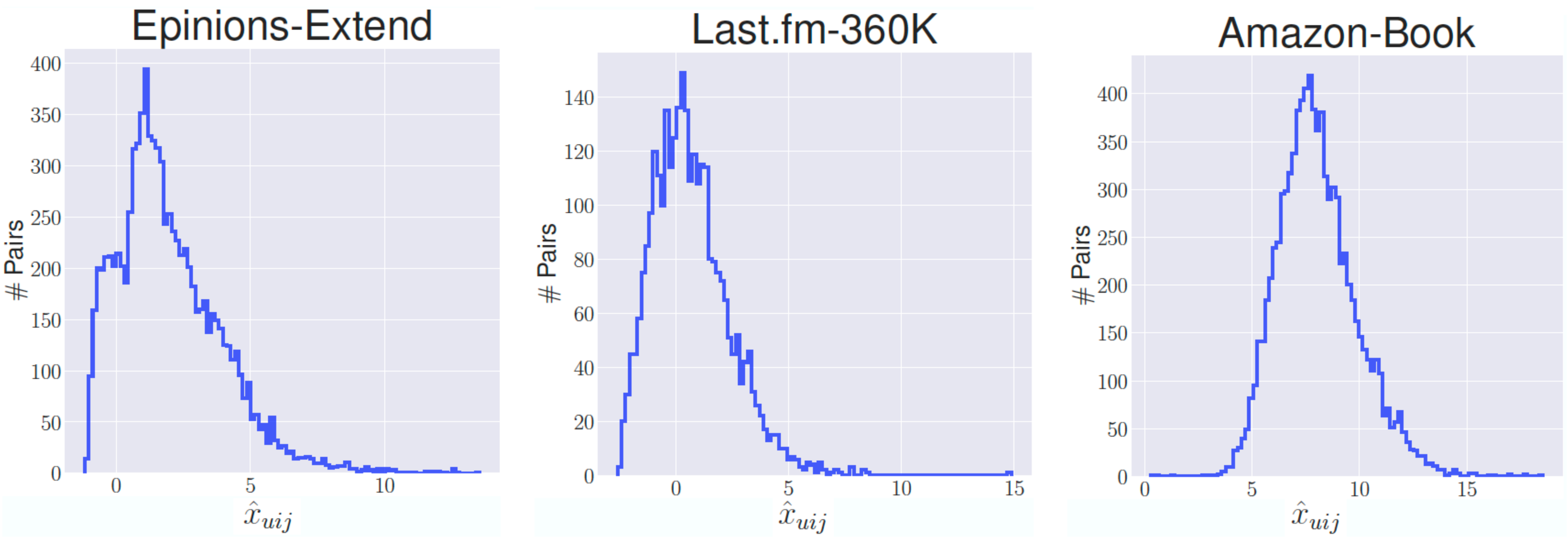
- Bayesian Personalized Ranking (BPR)



Allen Chung, National Chengchi University

Preliminaries and Observation

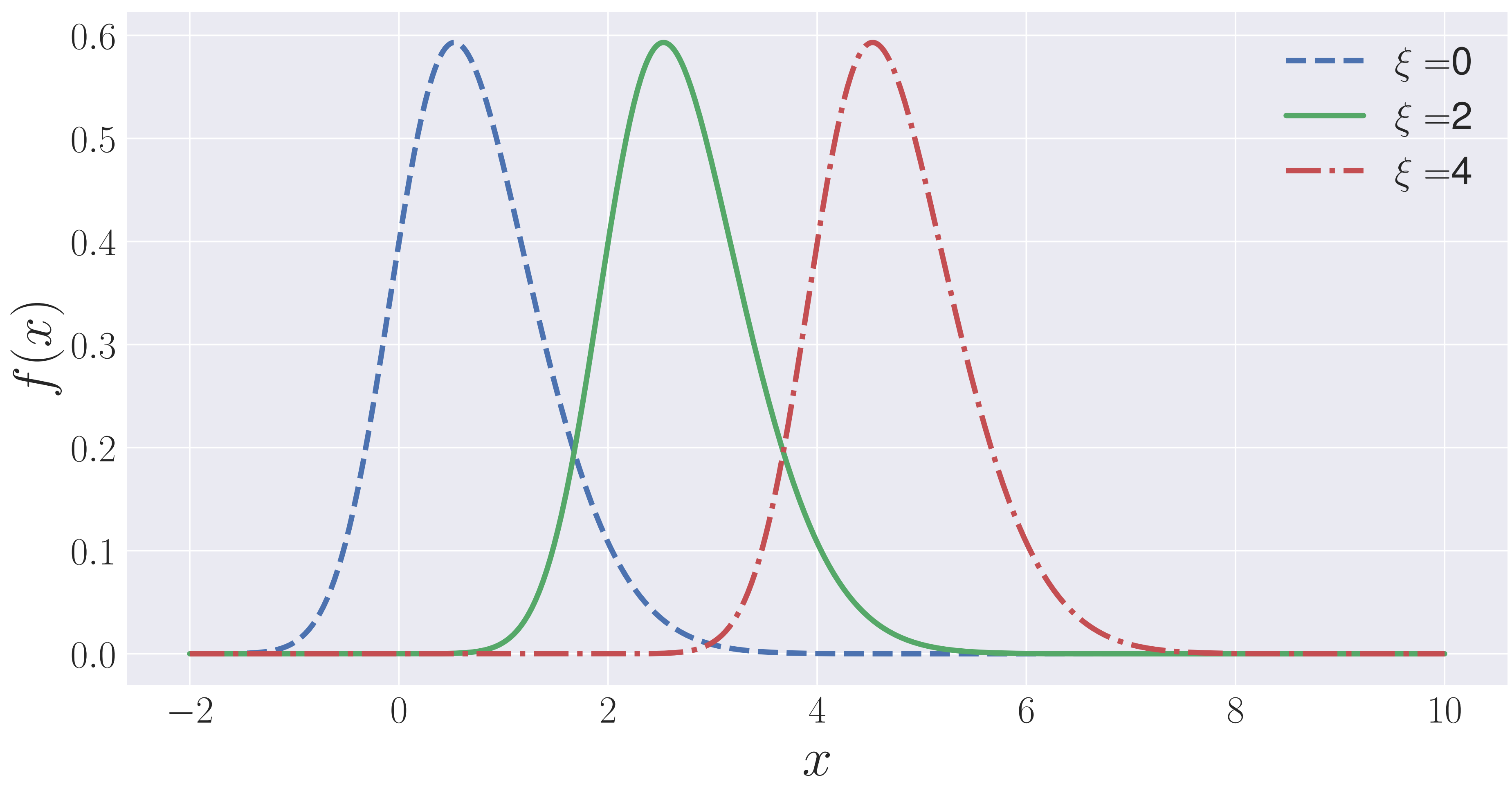
- We observed that the distributions $\hat{x}_{uij}(\Theta)$ learned from BPR are usually right-skewed—Skew normal distribution!



(a) $\hat{\gamma} = 1.09$ (b) $\hat{\gamma} = 0.373$ (c) $\hat{\gamma} = 0.08$

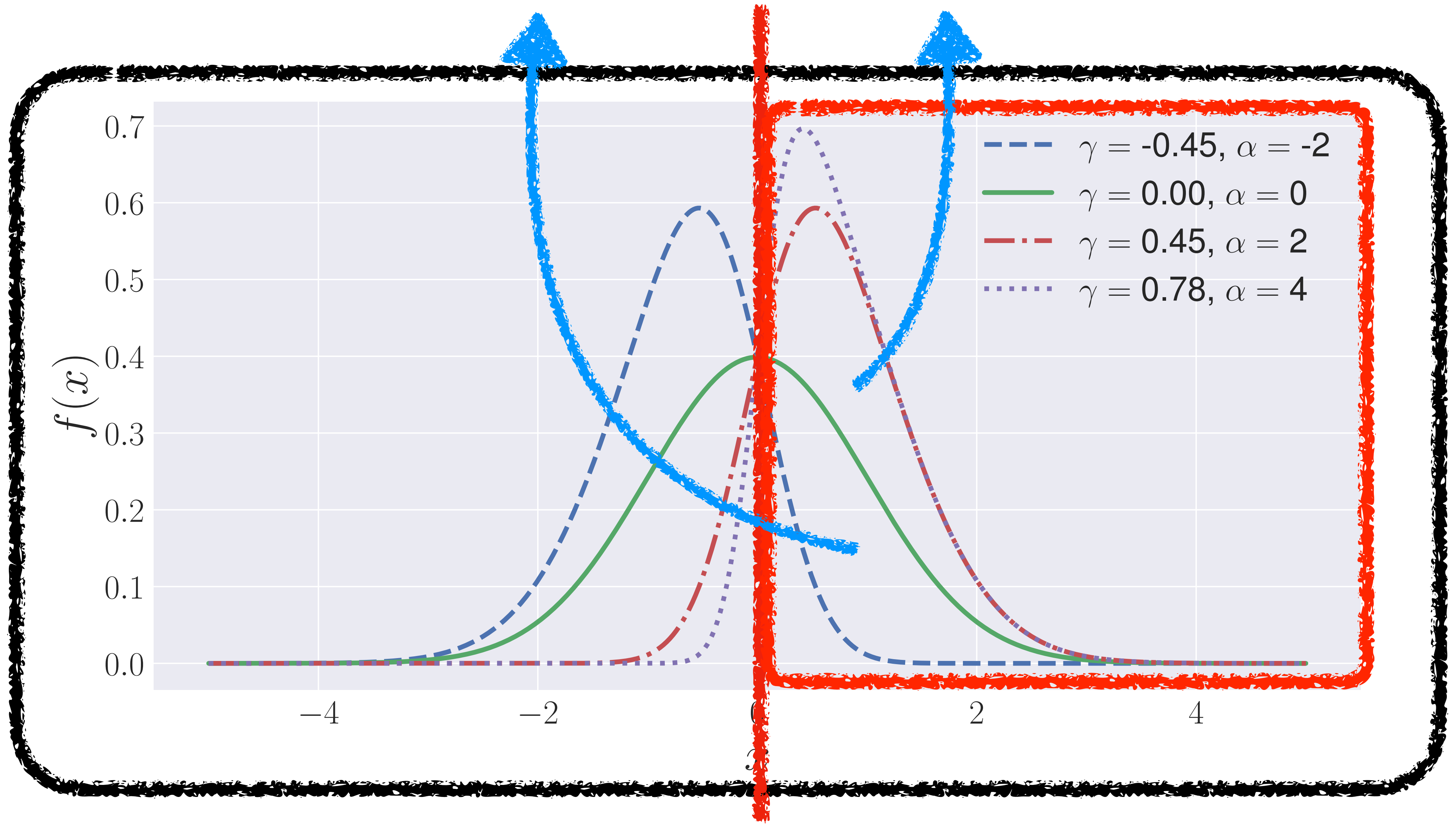
Preliminaries and Observation

- Under the assumption of skew normal distribution, there are two main ways to enlarge the $p(\hat{x}_{uij}(\Theta) > 0)$ should benefit recommendation performance:
  - Shift the distribution right-ward.
  - Maximize the shape parameter α.



right-skewed distribution

$$p(x > 0) = \int_0^\infty f(x)dx$$



Skewness Optimization Ranking

- Inspired by the observations, we manage to leverage the features of skew normal distribution to better model the personalized ranking problem.
  - NOTE: For personalized ranking, the estimator $\boxed{\hat{x}_{uij}(\Theta) = \hat{x}_{ui} - \hat{x}_{uj}}$ is to describe as the random variable X which is assumed to follow the skew normal distribution.
  - GOAL (1): To push the distribution right-ward for a larger $p(\hat{x}_{uij}(\Theta) > 0)$.
  - GOAL (2): To have a larger $p(\hat{x}_{uij}(\Theta) > 0)$ by adjusting the shape parameter.

Skewness Optimization Ranking

- Skewness optimization ranking (Skew-OPT)
  - We design the likelihood function of Skew-OPT

$$p(i >_u j \,|\, \Theta, (\xi, \omega, \eta)) = \sigma\left(\left(\frac{\hat{x}_{uij}(\Theta) - \xi}{\omega}\right)^{\eta}\right)$$

    where η is set to be odd integer.
  - The location parameter ξ allows Skew-OPT to push the distribution of the estimator $\hat{x}_{uij}$ right-ward.
  - The scale parameter ω reduces the model over-fitting for large ξ.

Allen Chung, National Chengchi University

Skewness Optimization Ranking

- Therefore, the optimization criterion of Skew-OPT becomes maximizing
  - $$\begin{aligned}\text{Skew-OPT} &:= \ln \prod_{(u,i,j)\in D_S} p\left(i >_u j \middle| \Theta, (\xi, \omega, \eta)\right) p(\Theta) \\ &= \sum_{(u,i,j)\in D_S} \ln p\left(i >_u j \middle| \Theta, (\xi, \omega, \eta)\right) + \ln p(\Theta) \\ &= \sum_{(u,i,j)\in D_S} \ln \sigma\left(\left(\frac{\hat{x}_{uij}(\Theta) - \xi}{\omega}\right)^{\eta}\right) - \lambda_{\Theta}\|\Theta\|^2. \quad (6)\end{aligned}$$
  - Skew-OPT is maximizing by utilizing the asynchronous stochastic gradient ascent for updating the learned parameters Θ.

Skewness Optimization Ranking

- Now we start to describe the relation between shape parameter α and Skew-OPT and how Skew-OPT optimize the skewness value.
  - **Lemma 1.** *Given the case that $\hat{x}_{uij}$ follows a skew normal distribution with fixed location parameter ξ and scale parameter ω, maximizing the first term of Eq. (6) for a certain η simultaneously maximizes the shape parameter α and the skewness value of the estimator, $\hat{x}_{uij}(\Theta)$.*

Max $\ln \sigma\left(\left(\frac{\hat{x}_{uij}(\Theta)-\xi}{\omega}\right)^{\eta}\right)$ $\xrightarrow{1}$ Max α $\xrightarrow{2}$ Max the skewness value

Allen Chung, National Chengchi University

Skewness Optimization Ranking

- The relation between Skew-OPT and AUC
  - We here consider micro-AUC :

$$\mathrm{AUC}^{\mathrm{micro}} := \frac{1}{|D_S|} \sum_{(u,i,j)\in D_S} \delta(\hat{x}_{uij} > 0)$$

  - Since we assume that $\hat{x}_{uij}$ follows skew normal distribution,

$$\begin{aligned}\mathrm{AUC}^{\mathrm{micro}} :=& \mathbb{E}\left[\delta(\hat{x}_{uij} > 0)\right] = \boxed{p(\hat{x}_{uij} > 0)} \\ =& 1 - F(0) \\ =& 1 - \Psi\left(\frac{0-\xi}{\omega}\right) + 2T\left(\left(\frac{0-\xi}{\omega}\right), \alpha\right)\end{aligned}$$

Skewness Optimization Ranking

- The relation between Skew-OPT and AUC
  - Skew-OPT seeks to maximize the estimator by shifting the distribution to the right, so we just discuss when $\xi > 0$.

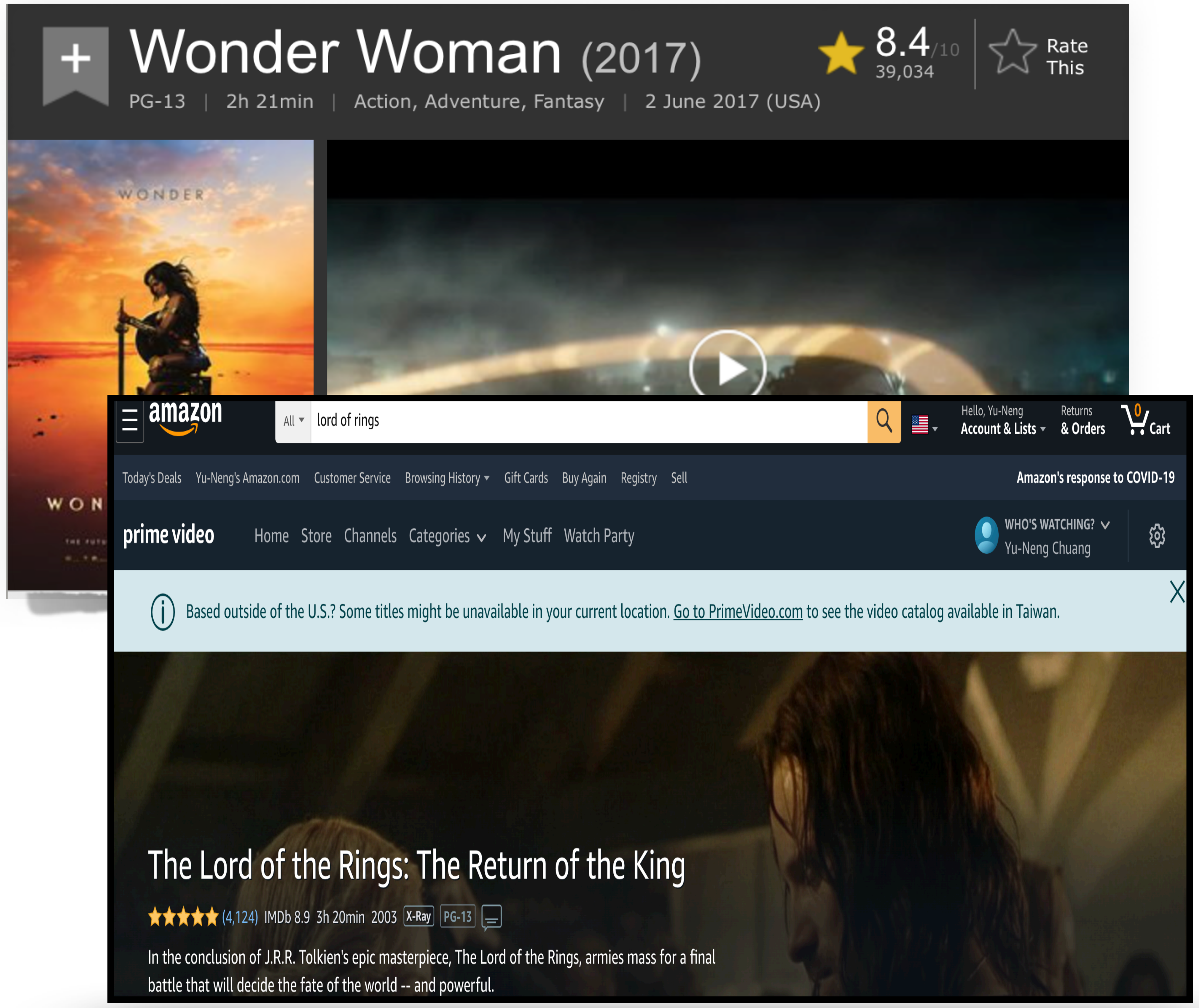
$$\begin{aligned}\lim_{\alpha\to\infty} \mathrm{AUC}^{\mathrm{micro}} :=& \mathbb{E}\left[\delta(\hat{x}_{uij} > 0)\right] = p(\hat{x}_{uij} > 0) \\ =& 1 - \Psi\left(\frac{0-\xi}{\omega}\right) + \lim_{\alpha\to\infty} 2T\left(\left(\frac{0-\xi}{\omega}\right), \alpha\right) \\ =& 1 - \Psi\left(\frac{0-\xi}{\omega}\right) + \Psi\left(\frac{0-\xi}{\omega}\right) \\ =& 1\end{aligned}$$

  - Therefore, when $\alpha \to \infty$, then $\mathrm{AUC}^{\mathrm{mirco}} \to 1$.

Experiment Results

- Datasets : Five different public real-world datasets.
  - Transfer into implicit feedback.
    - Above 3.5 points treat as preferring item.
    - Below 3.5 points treat as dislike item.

	Users	Items	Edges	Edge type
CiteULike	5,551	16,980	210,504	like/dislike
Amazon-Book	70,679	24,916	846,522	5-star
Last.fm-360K	23,566	48,123	303,4763	play count
MovieLens-Latest	259,137	40,110	24,404,096	5-star
Epinions-Extend	701,498	110,235	12,581,748	5-star



Experiment Results

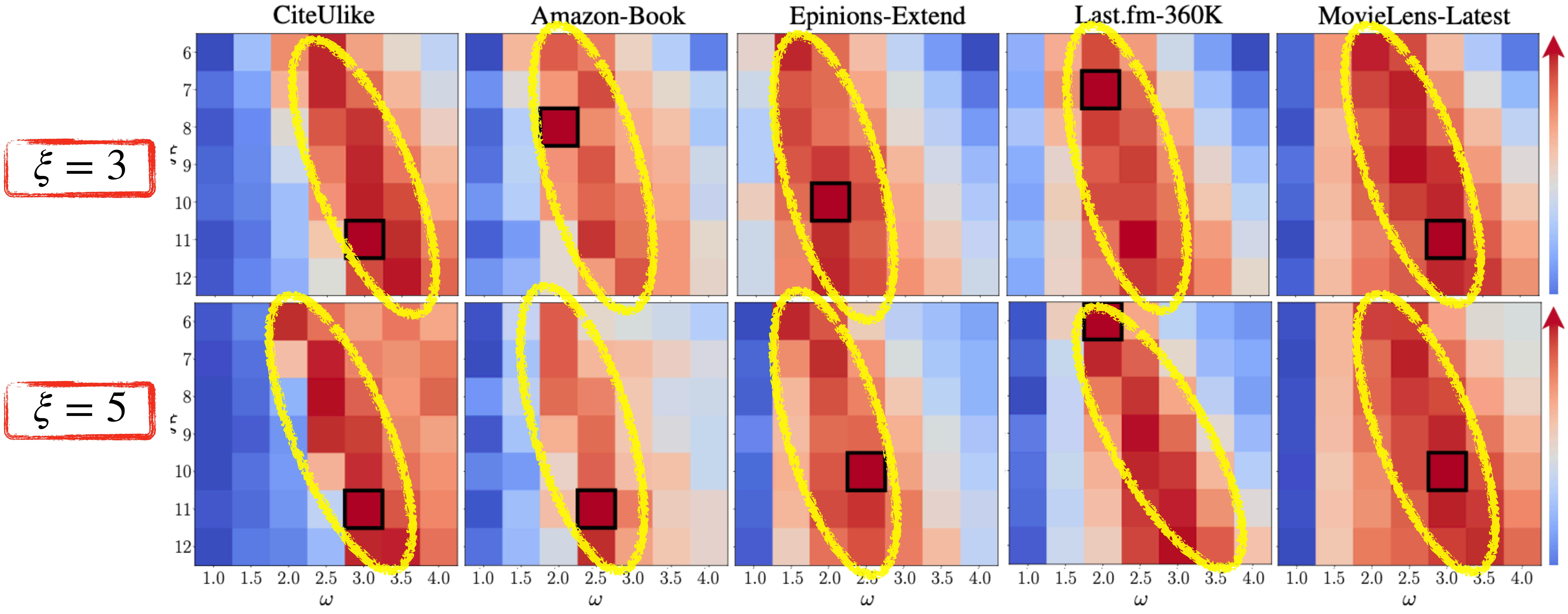
- Top-N recommendation performance

	CiteUlike		Amazon-Book		Last.fm-360K		MovieLens-Latest		Epinions-Extend	
	Recall@10	mAP@10	Recall@10	mAP@10	Recall@10	mAP@10	Recall@10	mAP@10	Recall@10	mAP@10
WRMF [10, 4]	0.2159	0.1236	0.0950	0.0374	0.1308	0.0576	0.2122	0.1061	0.1025	0.0415
BPR [11]	0.2217	0.1332	0.0972	0.0390	0.1394	0.0690	0.1952	0.1097	0.1137	0.0584
WARP [14]	0.1859	0.1033	0.0869	0.0356	† 0.1763	† 0.0937	† 0.2748	† 0.1634	0.1479	0.0711
Hop-Rec [16]	0.2232	0.1319	† 0.1072	† 0.0426	0.1701	0.0870	0.2557	0.1419	† 0.1617	† 0.0813
NGCF [13]	† 0.2321	† 0.1367	0.0818	0.0335	-	-	-	-	-	-
Skew-OPT ($\eta = 1$)	*0.2413	*0.1541	0.1069	*0.0467	*0.1976	*0.1051	0.2809	0.1636	*0.1743	*0.0914
Improv. (%)	+3.96%	+12.72%	-0.27%	+9.62%	+12.08%	+12.17%	+2.21%	+0.12%	+7.79%	+12.42%
Skew-OPT ($\eta = 3$)	*0.2481	*0.1591	***0.1173**	*0.0504	***0.2032**	***0.1103**	*0.2852	*0.1686	***0.1768**	***0.0941**
Improv. (%)	+6.89%	+16.38%	+9.42%	+18.07%	+15.25%	+17.71%	+3.78%	+3.18%	+9.33%	+15.74%
Skew-OPT ($\eta = 5$)	***0.2553**	***0.1626**	*0.1163	***0.0522**	*0.2012	*0.1083	***0.2879**	***0.1699**	*0.1758	*0.0915
Improv. (%)	+9.91%	+18.94%	+8.48%	+22.53%	+14.12%	+15.58%	+4.76%	+3.97%	+8.71%	+12.54%

- It is worthy to say that Skew-OPT win against HOP-Rec and NGCG without exploiting high-order information.

Experiment Results

- Sensitivity Analysis of the best performance

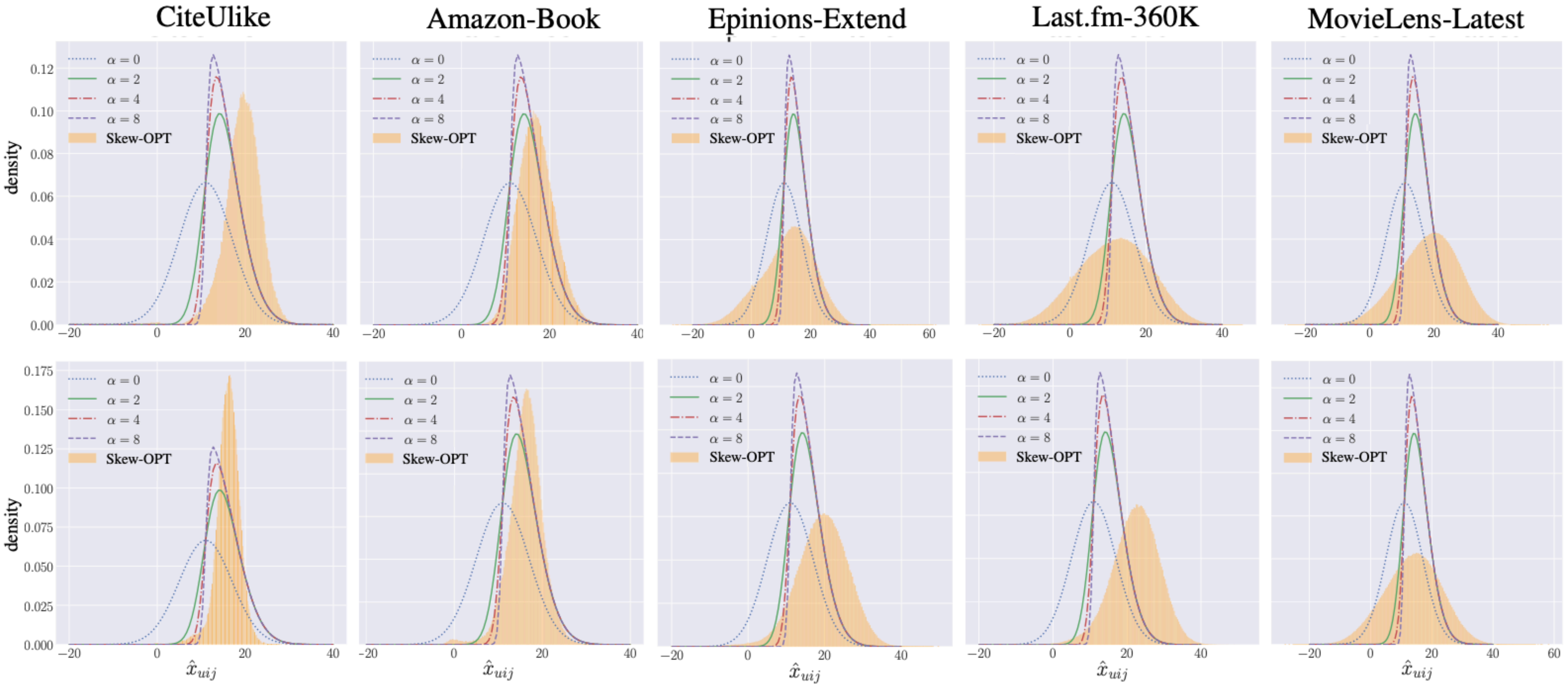


Experiment Results

- Distribution Analysis

$\xi = 3$

$\xi = 5$



Conclusion

- Skew-OPT provides probability distribution perspective to analyze the personalized recommendation problems.
- Skew-OPT leverages the feature from skew normal distribution and provides three extra degrees of freedom for ranking optimization.
- This work is first to analyze the learned embedding space for personalized recommendation task.

Skew-OPT Implementation

- Skew-OPT is now publicly available on GitHub:
  - Repo: https://github.com/cnclabs/codes.skewness.rec
- Skew-OPT is implemented on the framework of SMORe:
  - Repo: https://github.com/cnclabs/smore

Allen Chung, National Chengchi University

Thanks For Your Listening

Any Question ?