Recommender System: An Interdisciplinary Perspective

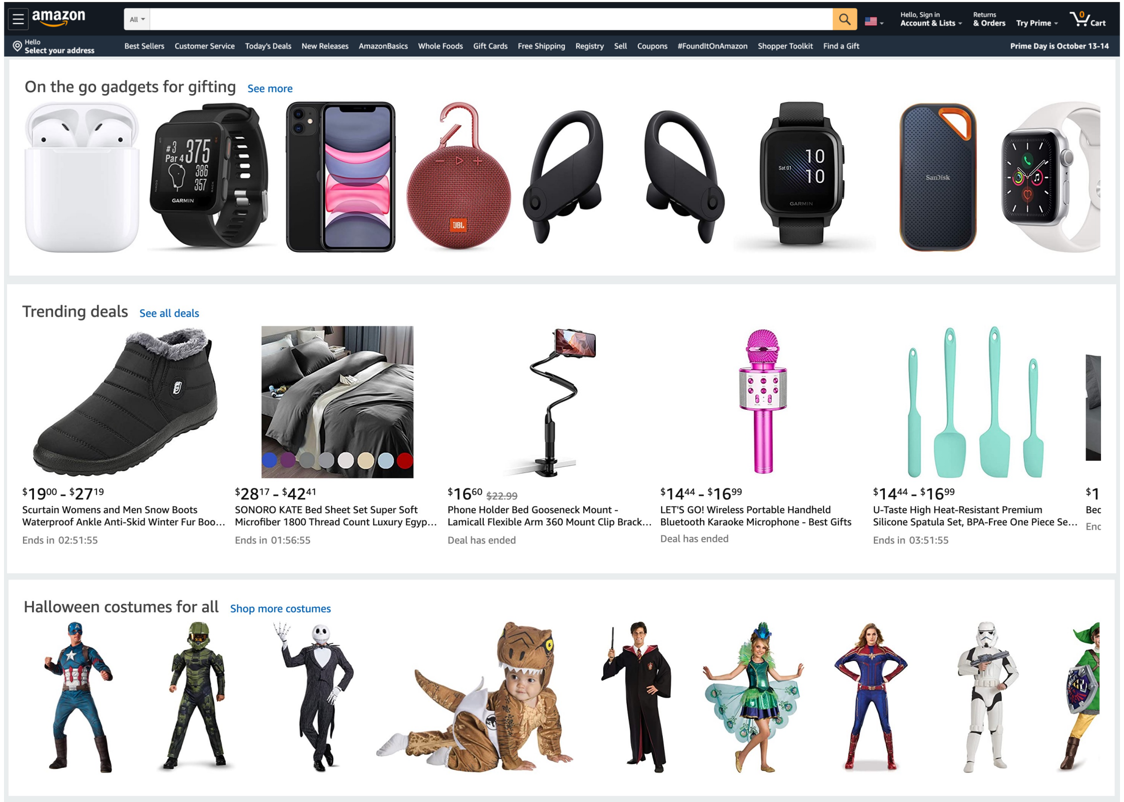
Technical, Ethical, Philosophical, Economic, and Legal Perspectives

Yongfeng Zhang
Department of Computer Science
Rutgers University
yongfeng.zhang@rutgers.edu

Outline

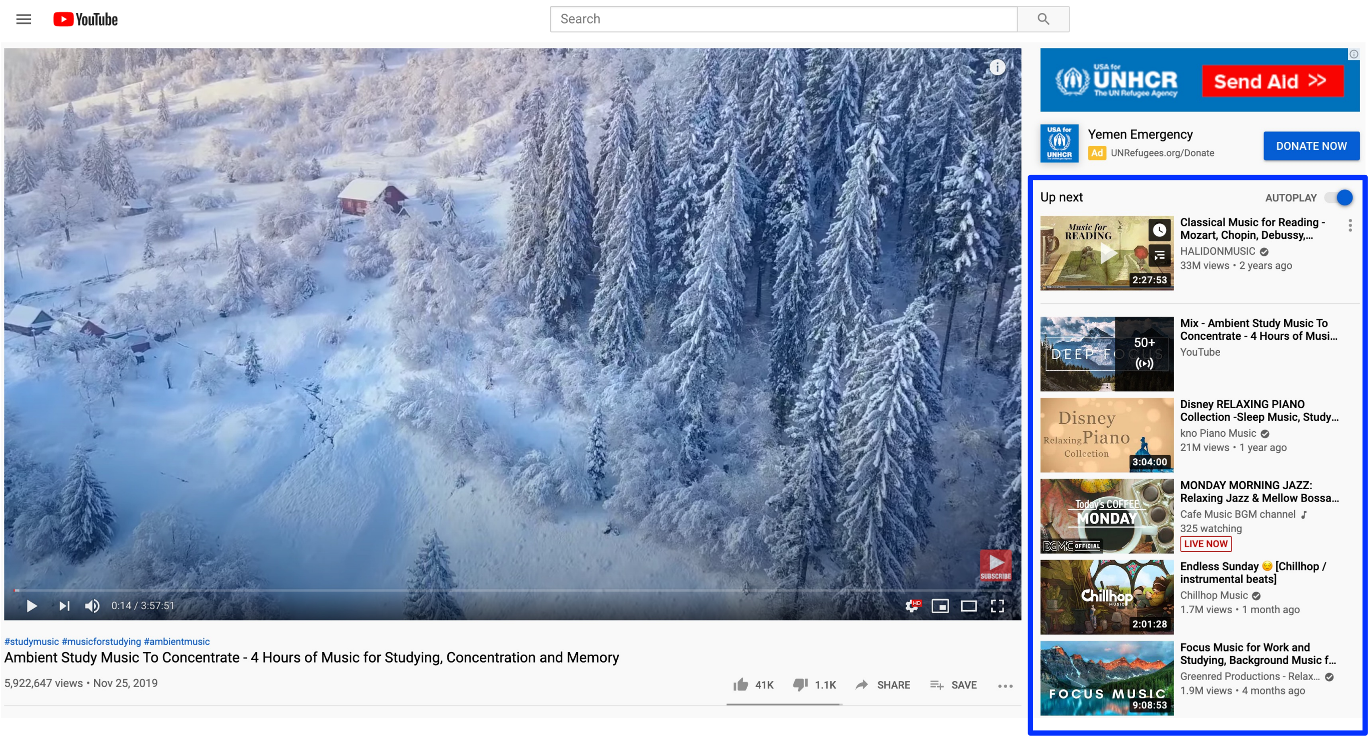
- Examples of Recommender Systems
- A Little Bit Models
- Ethics of Recommender Systems
  - Feedback Loop and Echo Chambers
  - Transparency
    - How to explain to users why certain things are recommended specifically for you
  - Bias and Fairness
    - Bias and Fairness on User-side (model bias)
    - Bias and Fairness on Provider-side (the Matthew Effect)
    - Advertising Fairness
  - Relationship between Transparency and Fairness
    - Legal Regulations on AI Ethics
- Open Discussions

Recommender Systems are Ubiquitous



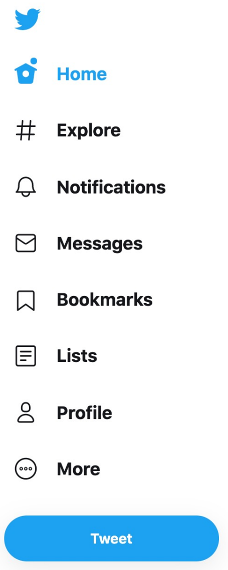
Product Recommendation
(e.g., Amazon)

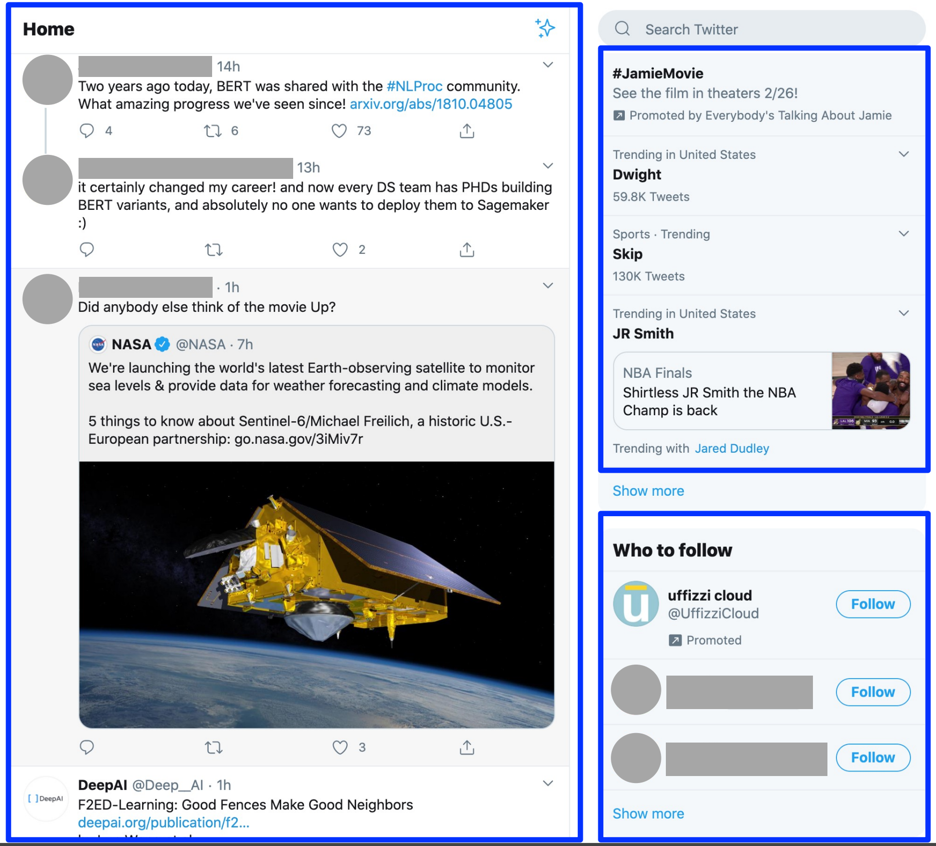
Recommender Systems are Ubiquitous



Video/Movie Recommendation (e.g., YouTube)

Recommender Systems are Ubiquitous



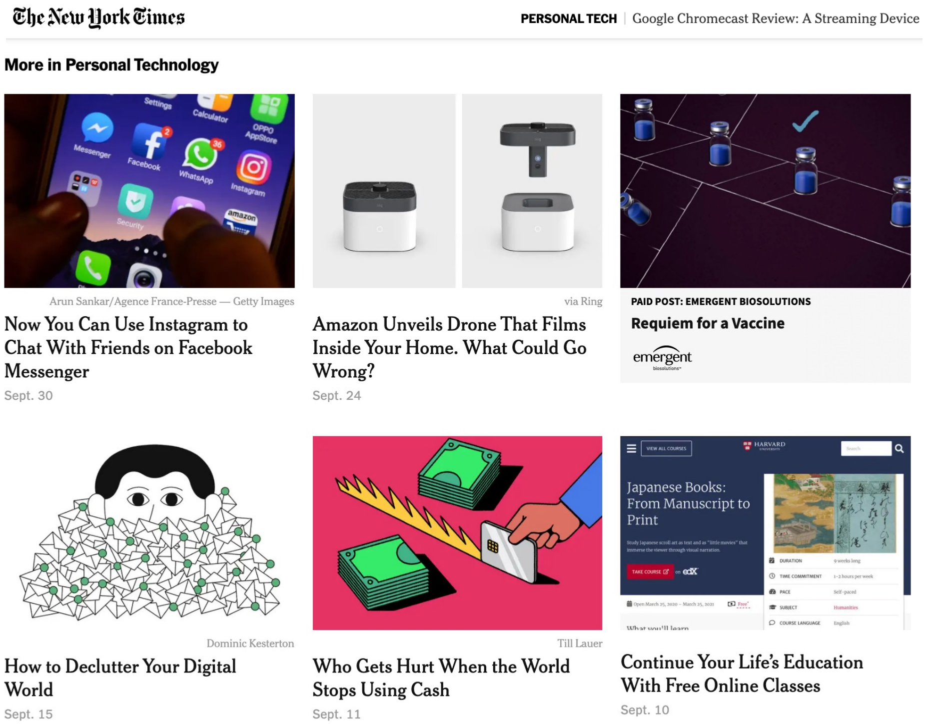


Topic Recommendation (e.g., Twitter)

Friend Recommendation (e.g., Twitter)

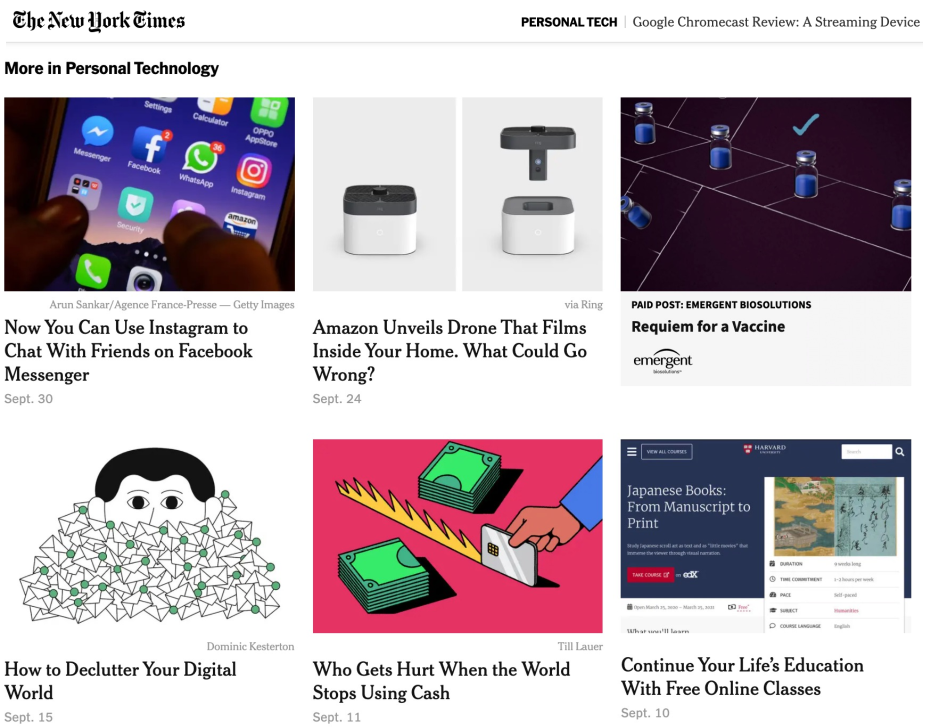
News Feed Recommendation (e.g., Twitter)

Recommender Systems are Ubiquitous



News Recommendation
(e.g., The New York Times)

Recommender Systems are Ubiquitous



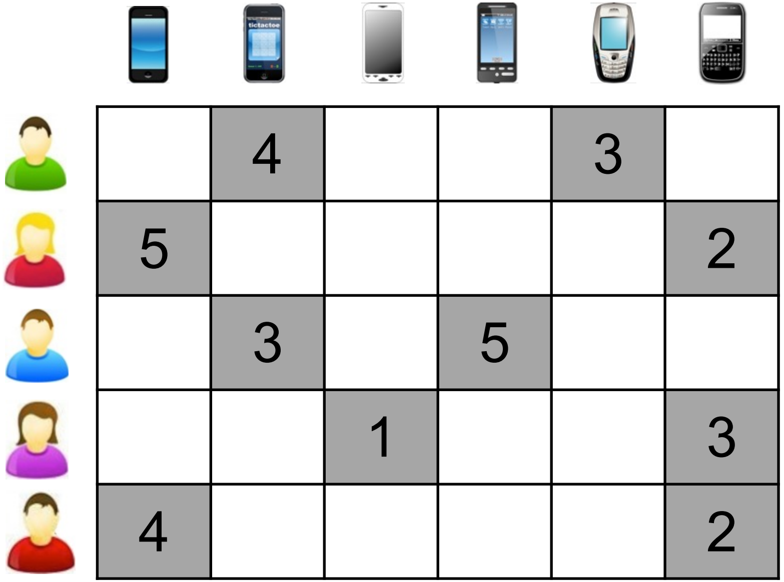
News Recommendation
(e.g., The New York Times)

Key of RS: Personalization

- The key feature of model recommender system is “Personalization”
  - Provide different and personalized items for different users

(Just a little bit) Models

- Input



User-Item Rating Matrix

Example Rating on Amazon

(Just a little bit) Models

- The Key Problem

?	4	?	?	3	?
5	?	?	?	?	2
?	3	?	5	?	?
?	?	1	?	?	3
4	?	?	?	?	2

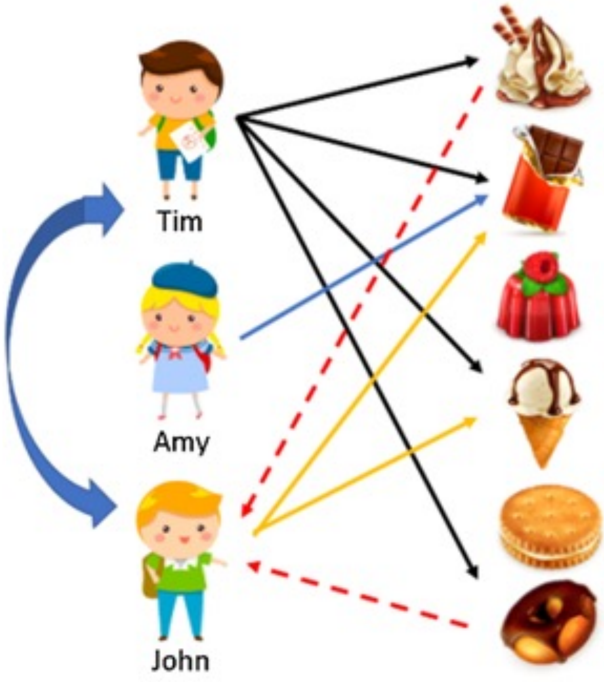
A key task: Predict the missing ratings

Predict the Missing Ratings

User-based Collaborative Filtering

- Consider user ***x***
- Find set ***N*** of other users whose ratings are "**similar**" to ***x***'s ratings
  - sim($\boldsymbol{x}$, $\boldsymbol{y}$) = cos($\boldsymbol{r_x}$, $\boldsymbol{r_y}$) = $\frac{r_x \cdot r_y}{||r_x|| \cdot ||r_y||}$
- Estimate ***x***'s ratings based on ratings of users in ***N***

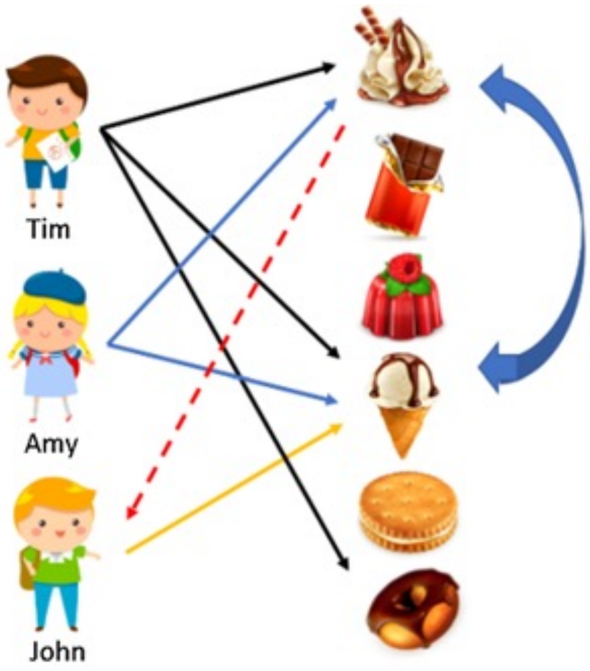
Key Idea: Recommend those items that other similar users liked



(a) User-based filtering

Item-based Collaborative Filtering

- Consider user ***x***, consider an item ***i*** the user liked
- Find set ***N*** of other items whose ratings are "**similar**" to *i*'s ratings
  - sim(i, j) = cos(r_i, r_j) = $\frac{r_i \cdot r_j}{||r_i|| \cdot ||r_j||}$
- Recommend these items



(b) Item-based filtering

Key Idea: Recommend those items that are similar to what the user have already liked

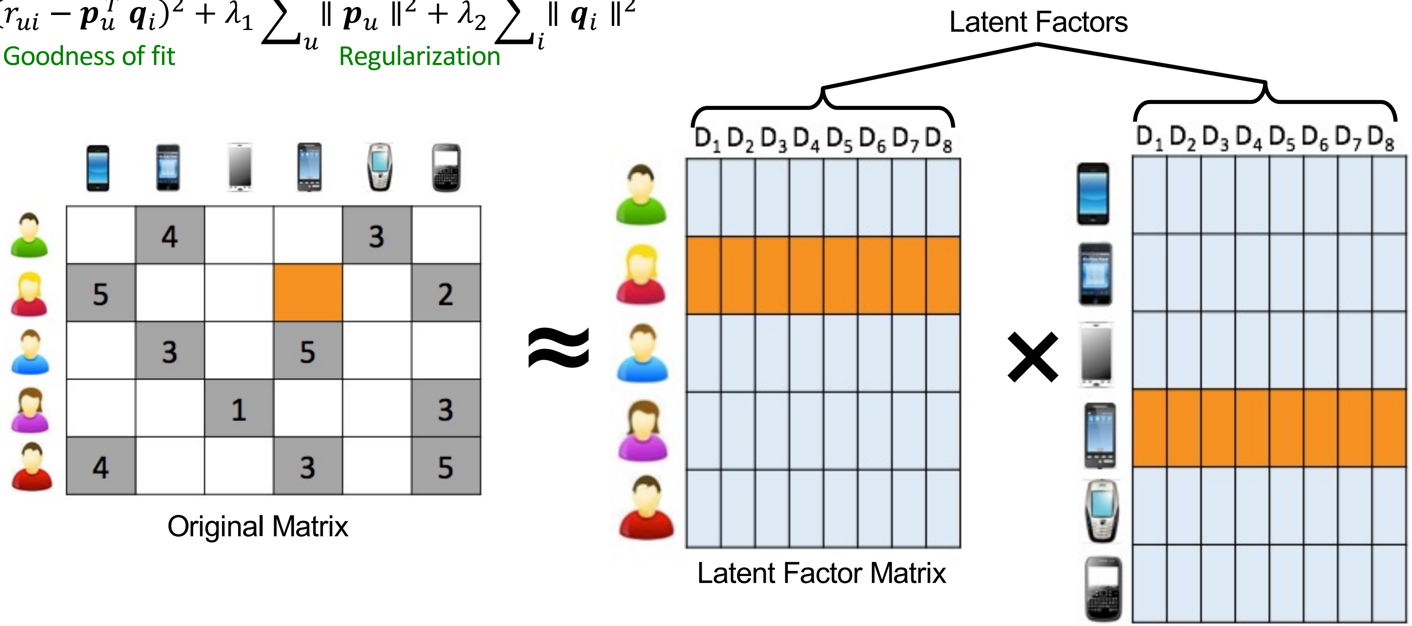
A Machine Learning-based Model

- Latent Factor Models for Matrix Completion

$\hat{r}_{ui} = \boldsymbol{p}_u^T \, \boldsymbol{q}_i$ $\quad$ $\boldsymbol{p}_u$: The "learned" feature vector for user *u*
$\boldsymbol{q}_i$: The "learned" feature vector for item *i*

$$\min_{\boldsymbol{p},\boldsymbol{q}} \underbrace{\sum_{(u,i)\in R} (r_{ui} - \boldsymbol{p}_u^T \, \boldsymbol{q}_i)^2}_{\text{Goodness of fit}} + \underbrace{\lambda_1 \sum_u \| \boldsymbol{p}_u \|^2 + \lambda_2 \sum_i \| \boldsymbol{q}_i \|^2}_{\text{Regularization}}$$

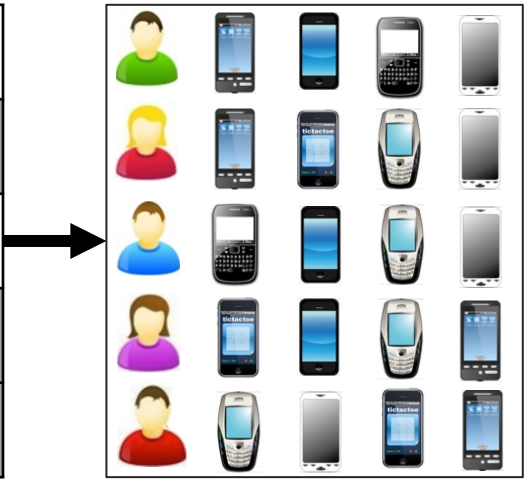
User/Item Personalization by Vector Representations



(Just a little bit) Models

- Providing Recommendations by Ranking

4.5	4	2.7	4.7	3	3.6
5	3.8	1.2	4.2	3.5	2
3.3	3	2.6	5	2.8	3.7
4.2	4.5	1	2.4	3.2	3
4	3.7	3.9	3.2	4.6	2



Ethics of Recommender Systems

- Feedback Loop and Echo Chambers
- Transparency
  - How to explain to users why certain things are recommended specifically for you
- Bias and Fairness
  - Bias and Fairness on User-side (model bias)
  - Bias and Fairness on Provider-side (the Matthew Effect)
- Relationship between Transparency and Fairness
- Open Discussions

Ethics of Recommender Systems

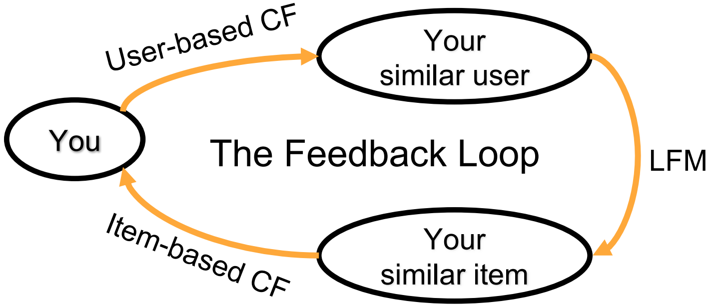
- Technology is neither good nor bad; nor is it neutral
  - It all depends on how we use it, and we should use technology in a responsible way
  - E.g., Atomic theory: nuclear power station and clean energy vs. nuclear bombs
- Recommender Systems
  - Helps users to find good items in a sea of items
  - May also bring counter-effects

Feedback Loops and Echo Chambers

- What is Echo Chamber

  An echo chamber is "an environment where a person only encounters information or opinions that reflect and reinforce their own.

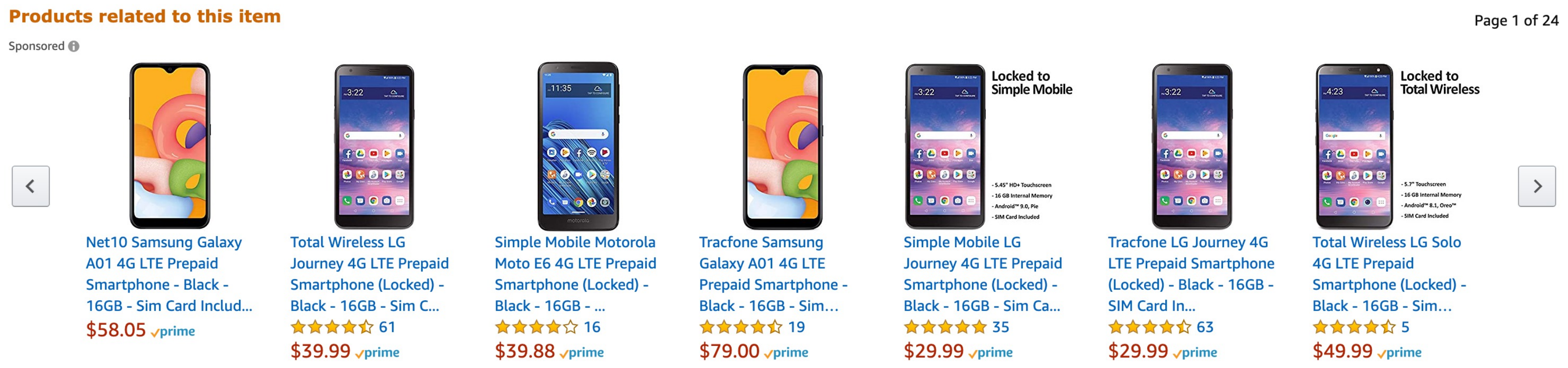
- Why RS creates Echo Chambers



The more you like something, the more RS will recommend similar things, and thus you like them even more.

Echo Chambers in E-commerce

- System always recommend similar products
- Even always recommend products that you already bought (e.g., phones)

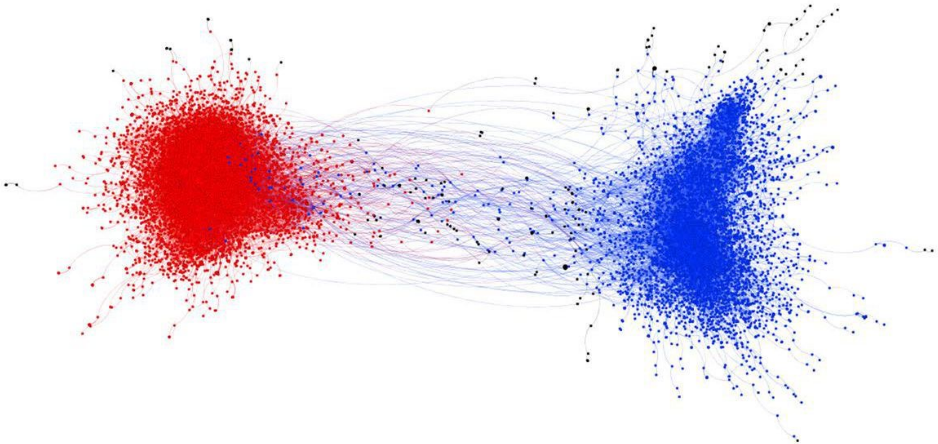


Echo Chambers in Video/Movie/Book Recommendation

- Always recommend similar videos/movies/books
- Prevents you from and exploring a much richer and diversified world
- Prevents you from thinking outside of the box

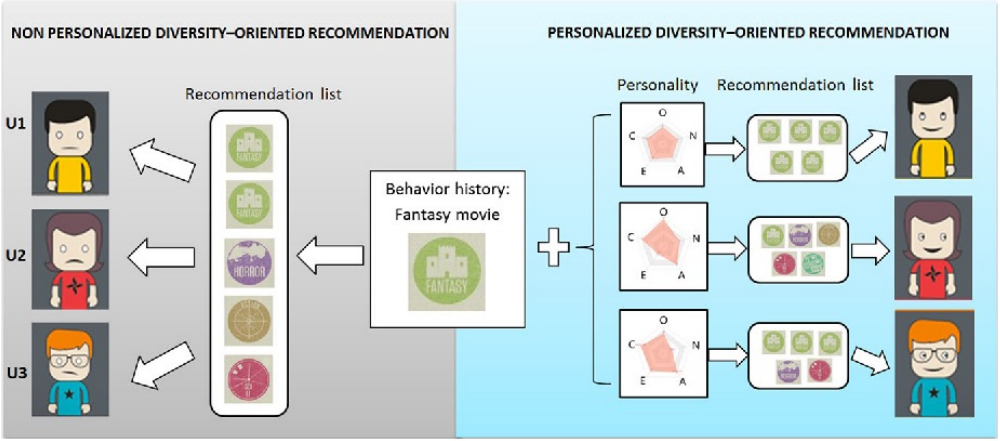
Echo Chambers in Social Networks (e.g., Twitter)

- The Social Echo Chamber
  - Makes all your connections like-minded persons as you
  - Makes all your news feed recommendation similar to what you have liked
  - Makes it difficult to explore new ideas and opinions different from yours
  - Makes everyone feel the whole world thinks the same way as you think
  - May even reinforce someone’s extremist ideas



How to Avoid Echo Chambers

- An Active Research Area in RS, ML, AI
- The key is to take care of the diversity in recommendation
  - Provide similar item recommendations, and meanwhile some dis-similar recommendations
- A trade-off between utility and diversity
  - Best if the dis-similar recommendations are also what the user likes, e.g., personalized diversity

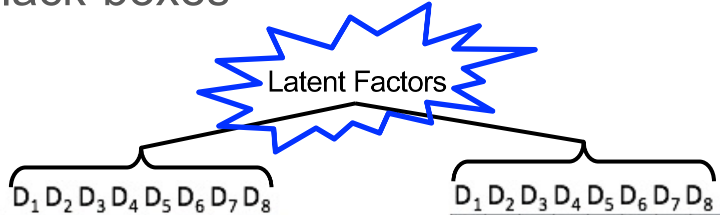


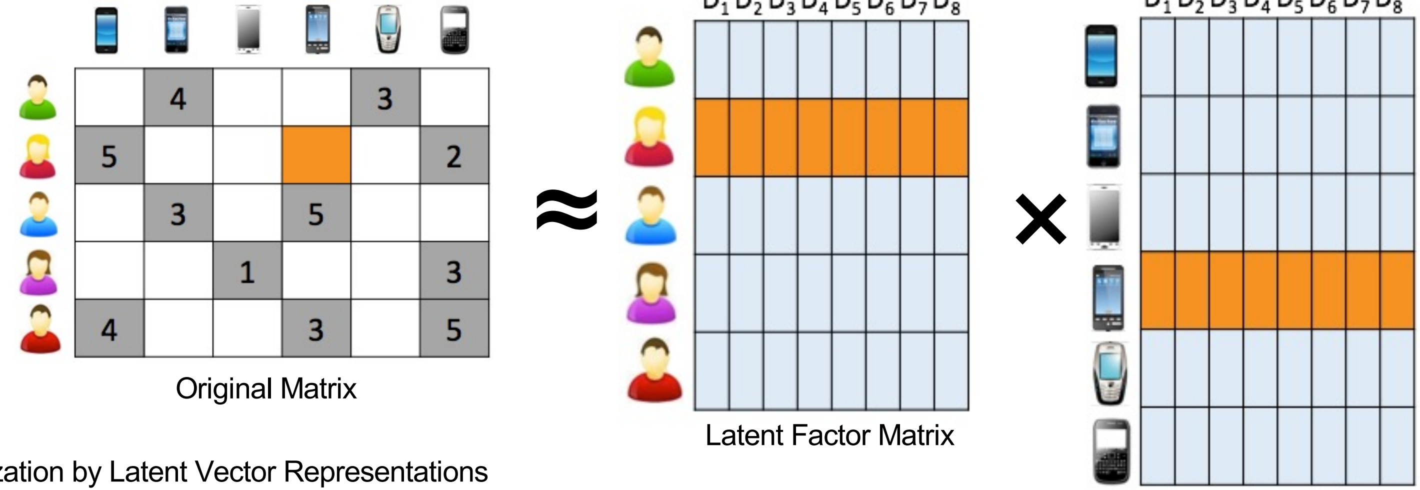
Transparency of Recommendation

- Tell users why something is recommended: Explainable Recommendation
- Many machine learning models are black-boxes

$\hat{r}_{ui} = \boldsymbol{p}_u^T \boldsymbol{q}_i$ $\boldsymbol{p}_u$: The "learned" feature vector for user *u*
$\boldsymbol{q}_i$: The "learned" feature vector for item *i*

$$\min_{\boldsymbol{p},\boldsymbol{q}} \underbrace{\sum_{(u,i)\in R} (r_{ui} - \boldsymbol{p}_u^T \boldsymbol{q}_i)^2}_{\text{Goodness of fit}} + \underbrace{\lambda_1 \sum_u \| \boldsymbol{p}_u \|^2 + \lambda_2 \sum_i \| \boldsymbol{q}_i \|^2}_{\text{Regularization}}$$





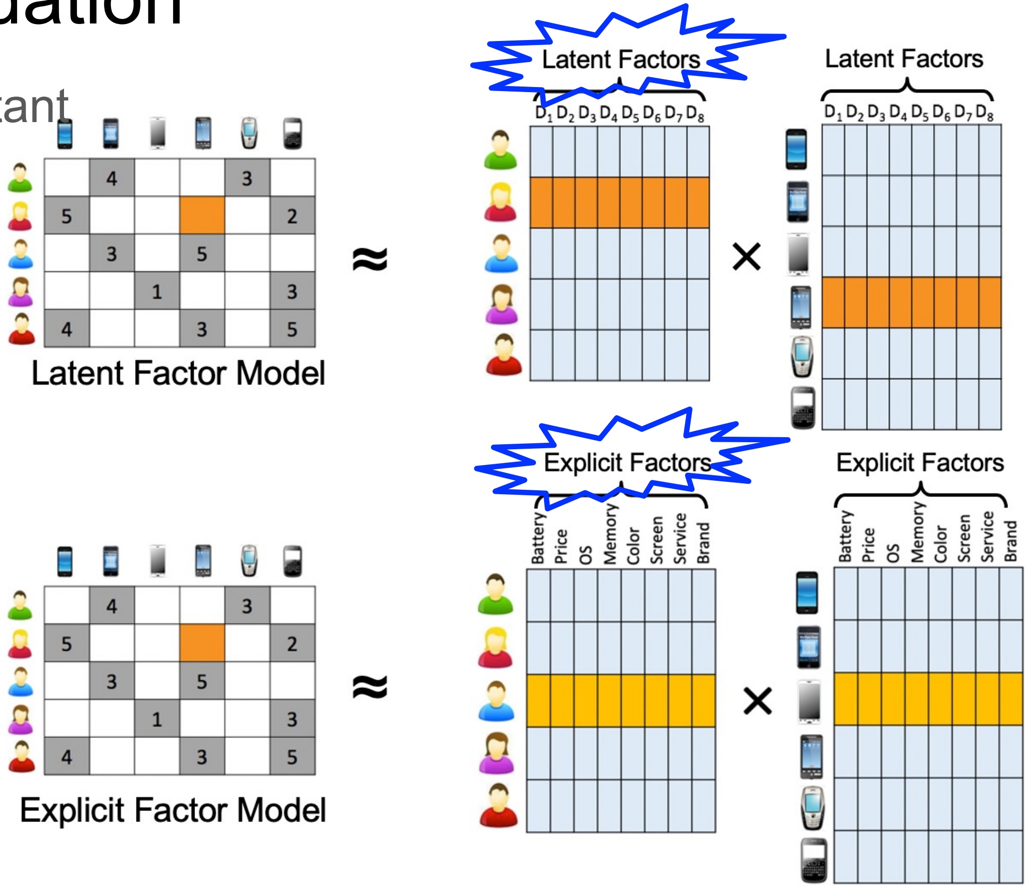
Original Matrix

Latent Factor Matrix

User/Item Personalization by Latent Vector Representations

Explainable Recommendation

- Letting the users know **why** is important
  - Help users to make the **right decisions**
  - Otherwise human might be “**controlled**” by algorithms
- Explainable Recommendation
  - More generally: Explainable AI (XAI)
  - An active Research area in AI
  - An example: Explicit Factor Model
  - Explanation: We recommend because you may be interested in [feature], and this item performs well on that feature.



Bias and Fairness in Recommendation

- A broad problem, broadly classified into two problems
  - User-side fairness, Provider-side fairness
- Why? Usually RS works in two-sided markets/environments
  - RS is actually a resource allocation problem

The *Prosumer* Paradigm:

Consumers – *items* – *Producers*

Buyers – Goods – Sellers

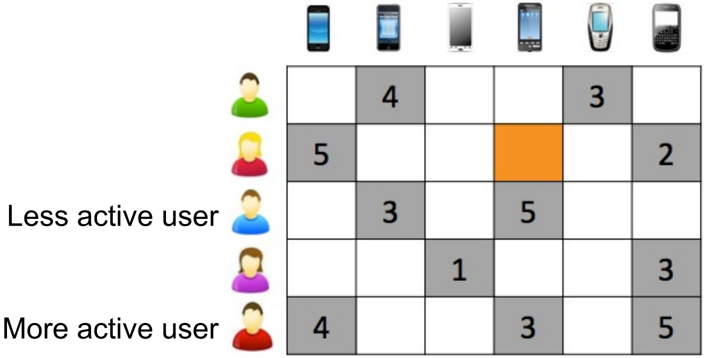
Freelancer – Jobs – Employers

Borrowers – Money – Lenders

Passengers – Services – Drivers

User-side Fairness in RS

- Where does unfairness come from?
  - Some users are **more active**
    - e.g., more time to explore items, more money to buy items
  - They contribute **more training data** to the ML model
  - The model behavior may be **dominated** by active users.
    - e.g., tend to recommend items that the active use likes to everyone



Original Matrix

$$\hat{r}_{ui} = \boldsymbol{p}_u^T \boldsymbol{q}_i$$

$\boldsymbol{p}_u$: The “learned” feature vector for user u

$\boldsymbol{q}_i$: The “learned” feature vector for item i

$$\min_{\boldsymbol{p},\boldsymbol{q}} \underbrace{\sum_{(u,i)\in R} (r_{ui} - \boldsymbol{p}_u^T \boldsymbol{q}_i)^2}_{\text{Goodness of fit}} + \underbrace{\lambda_1 \sum_u \| \boldsymbol{p}_u \|^2 + \lambda_2 \sum_i \| \boldsymbol{q}_i \|^2}_{\text{Regularization}}$$

User-side Fairness in RS

Observation: Top 5% active users' data may dominate a ML algorithm

Dataset	CDs & Vinyl								Clothing							
Measures (%)	Overall		Inactive Users		Active Users		GRU		Overall		Inactive Users		Active Users		GRU	
	NDCG	F_1	NDCG	F_1	NDCG	F_1	NDCG	F_1	NDCG	F_1	NDCG	F_1	NDCG	F_1	NDCG	F_1
HeteroEmbed	6.992	3.576	6.526	3.373	15.843	7.429	9.317	4.056	3.221	1.404	3.121	1.348	**5.130**	2.461	2.009	1.113
Fair HeteroEmbed	**8.094**	**4.019**	**7.674**	**3.820**	**16.074**	**7.801**	**8.400**	**3.981**	**3.494**	**1.536**	**3.484**	**1.482**	3.691	**2.556**	**0.207**	**1.074**
PGPR	6.947	3.571	6.526	3.373	14.943	7.324	8.417	3.951	2.856	1.240	2.787	1.198	**4.197**	2.036	1.410	0.833
Fair PGPR	**8.045**	**4.019**	**7.675**	**3.820**	**15.074**	**7.801**	**7.399**	**3.261**	**3.101**	**1.314**	**3.089**	**1.274**	3.322	**2.078**	**0.233**	**0.804**
KGAT	5.411	3.357	5.038	3.162	12.498	7.046	7.460	3.884	3.021	1.305	2.931	1.254	4.741	2.259	1.810	1.005
Fair KGAT	**5.640**	**3.492**	**5.295**	**3.318**	12.366	6.791	**7.081**	**3.473**	**3.206**	**1.393**	**3.119**	**1.347**	**4.843**	**2.262**	**1.724**	**0.915**

$$\max_{Q_{ij}} \quad \mathcal{R} = \sum_{i=1}^{m} \mathcal{R}_{\text{rec}}(\mathbf{Q_i})$$

$$\text{s.t.} \quad \sum_{j=1}^{N} Q_{ij} = K, \; Q_{ij} \in \{0, 1\}$$

$$GRU(G_1, G_2, \mathbf{Q}) < \varepsilon_1$$

$$GREU(G_1, G_2, \mathbf{Q}) < \varepsilon_2$$

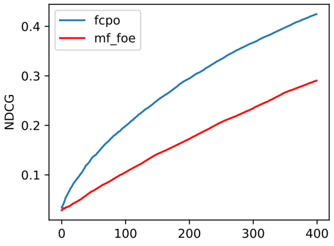
$$GRU(G_1, G_2, \mathbf{Q}) = \left| \frac{1}{|G_1|} \sum_{i \in G_1} \mathcal{F}(\mathbf{Q_i}) - \frac{1}{|G_2|} \sum_{i \in G_2} \mathcal{F}(\mathbf{Q_i}) \right|$$

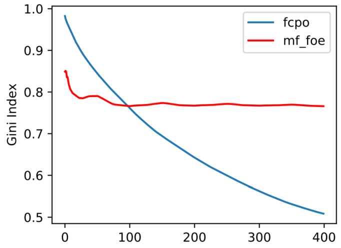
$$GEDU(G_1, G_2, \mathbf{Q}) = \left| \frac{1}{|G_1|} \sum_{i \in G_1} f(\mathbf{Q_i}) - \frac{1}{|G_2|} \sum_{i \in G_2} f(\mathbf{Q_i}) \right|$$

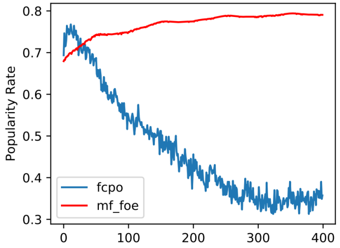
Item-side Fairness in RS

- Where does unfairness come from?
  - Some providers are big, some are small
    - Big retailors like Walmart vs Family-owned small-business retailors
  - Big providers have more budget for advertising and marketing
    - Thus their items get more exposure in E-commerce
  - The more exposure they get, the more users buy their items
  - RS thinks these items are more liked by users, and thus recommend these items even more
  - It becomes even more difficult for small-businesses to survive in the environment
  - The **Matthew Effect**: The rich get even richer and the poor get even poorer
  - This is unhealthy to the national economy!

$$Gini\ Index(\mathcal{G}) = \frac{1}{2|\mathcal{I}|^2\bar{g}} \sum_{i=1}^{|\mathcal{I}|} \sum_{j=1}^{|\mathcal{I}|} |g_i - g_j|$$

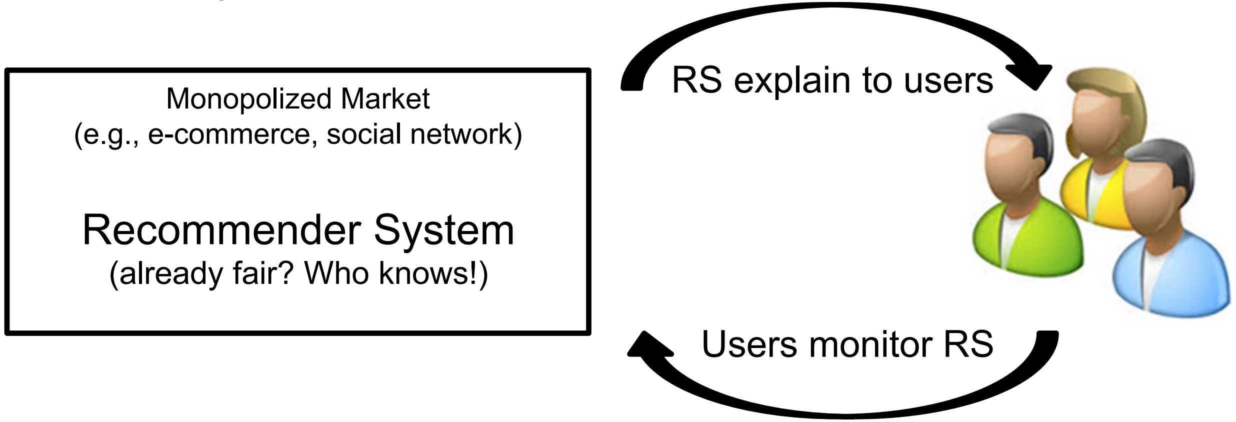






Relationship between Transparency and Fairness

- Transparency and Fairness benefit each other: Explainable Fairness
- Legal Regulatory Approaches to AI Ethics
    - E.g., EU General Data Protection Regulation (GDPR), The California Privacy Act of 2018
    - Emphasize the trustworthiness, robustness, transparency, and fairness of algorithmic decisions in AI systems.



A healthy virtuous cycle between user and system, and thus a healthy online economy.

Wrap up

- Examples of Recommender Systems
- A Little Bit Models
- Ethics of Recommender Systems
    - Feedback Loop and Echo Chambers
    - Transparency
        - How to explain to users why certain things are recommended specifically for you
    - Bias and Fairness
        - Bias and Fairness on User-side (model bias)
        - Bias and Fairness on Provider-side (the Matthew Effect)
        - Advertising Fairness
    - Relationship between Transparency and Fairness
        - Legal Regulations on AI Ethics
- Open Discussions


THANKS