Heterogeneous Recommendations: What You Might Like To Read After Watching Interstellar

Rachid Guerraoui, Anne-Marie Kermarrec, Tao Lin, Rhicheek Patra

EPFL, Switzerland

Outline

- Introduction
- X-Map
  - X-Sim
  - AlterEgo
  - Recommendation
- Experiments & Conclusion

Outline

- **Introduction**
- X-Map
  - X-Sim
  - AlterEgo
  - Recommendation
- Experiments & Conclusion

Personalization

- Personalization services, mainly recommendations, are widely employed.

Movies	Music	Books	News

Personalization

- Personalization services, mainly recommendations, are widely employed.

Movies	Music	Books	News
NETFLIX	last.fm	goodreads	Google news
IMDb	PANDORA internet radio	BookPsychic	YAHOO! NEWS

Most services are limited to personalization within a ***single*** *domain*

Heterogeneous recommendations

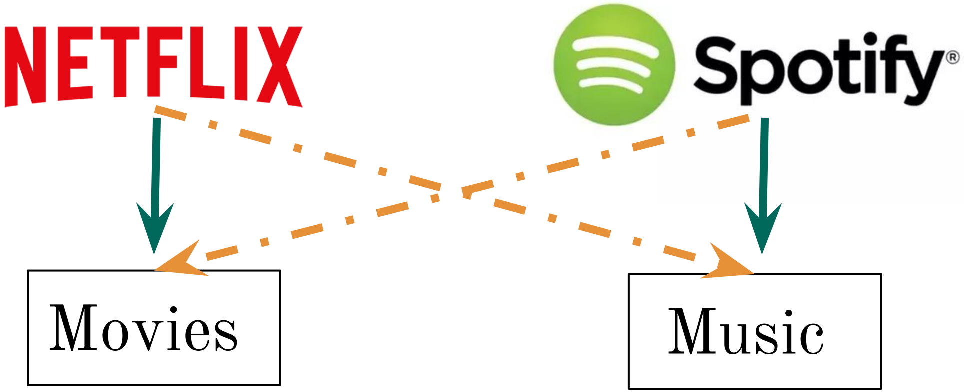
NETFLIX

Movies

Spotify

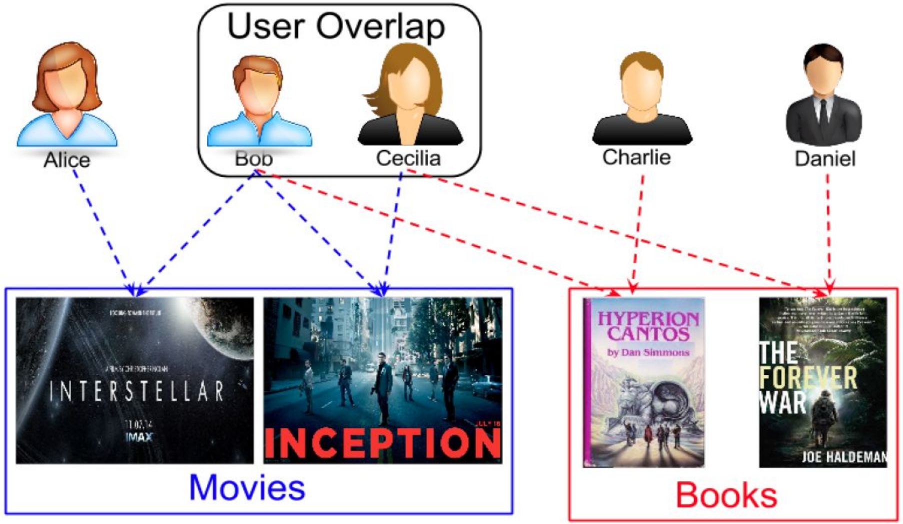
Music

Heterogeneous recommendations

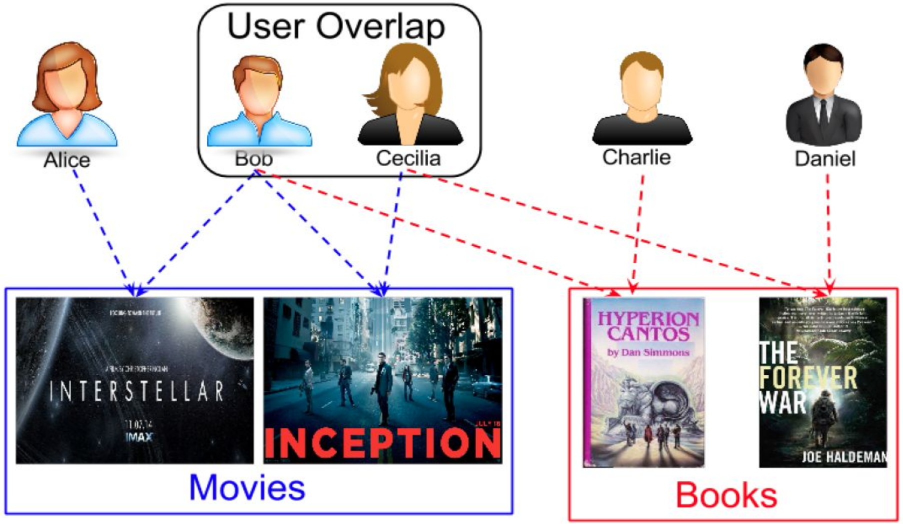


What if the companies want to venture across multiple domains?

Heterogeneous recommendations: Scenario



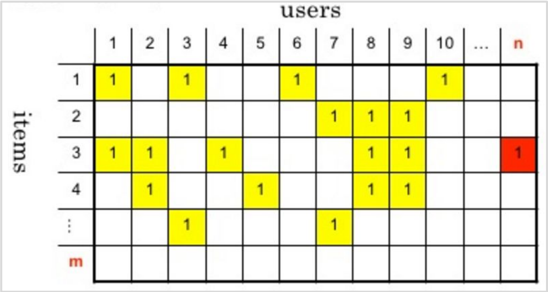
Heterogeneous recommendations: Scenario



*Given that Alice liked Interstellar, which **books** would she like to **read**?*
*Given that Daniel like The Forever War, which **movies** would he like to **watch**?*

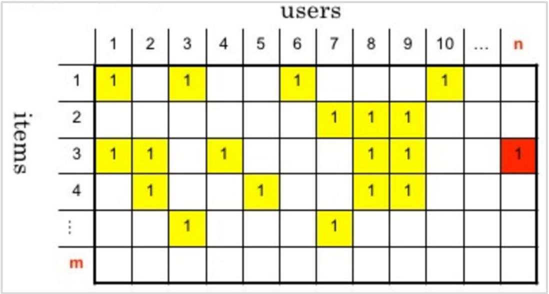
Why is heterogeneous recommendations a challenge?

- Quality
  - Standard homogenous approaches do not work
  - Preferences of users vary across domains
  - Decrease in *Density* (user-item rating matrix) affects quality
    - *Density*: Fraction of actual interactions among all the possible user-item interactions



Why is heterogeneous recommendations a challenge?

- Quality
  - Standard homogenous approaches do not work
  - Preferences of users vary across domains
  - Decrease in *Density* (user-item rating matrix) affects quality
    - *Density*: Fraction of actual interactions among all the possible user-item interactions



Density of domains in Amazon		
Books	Movies	Books + Movies
0.0204%	0.0569%	**0.0147%**

Why is heterogeneous recommendations a challenge?

- Privacy
  - General concern in homogenous recommenders trivially extends to heterogeneous ones.
  - Higher privacy concern in heterogeneous scenario due to an increase in the connections across domains.[1]
  - Straddlers, i.e., users who connect multiple domains, are at a higher privacy risk.

1. Ramakrishnan et al. "Privacy risks in recommender systems." IEEE Internet Computing 5.6 (2001): 54.

Why is heterogeneous recommendations a challenge?

- Scalability
  - Increase in information ⟶ Increased computations ⟶ Requires better scalability
  - Extend to multiple domains (movies, books, songs, electronics)
  - Additional computational overhead due to privacy preservation techniques

Challenges for heterogeneous recommendations

Quality

Scalability

Privacy

Challenges for heterogeneous recommendations

Quality

Scalability

Privacy

How to design a heterogeneous recommender to address these challenges?

Challenges for heterogeneous recommendations

<u>X-Map</u>

Quality

X-Sim

Scalability

AlterEgo

Privacy

Differential Privacy

Outline

- Introduction
- X-Map
  - **X-Sim**
  - AlterEgo
  - Recommendation
- Experiments & Conclusion

X-Sim: Baseline Similarity Graph Construction

- We use adjusted-cosine similarity to build this graph

$$s_{ac}(i,j) = \frac{\sum_{u \in Y_i \cap Y_j} (r_{u,i} - \bar{r_u})(r_{u,j} - \bar{r_u})}{\sqrt{\sum_{u \in Y_i} (r_{u,i} - \bar{r_u})^2} \sqrt{\sum_{u \in Y_j} (r_{u,j} - \bar{r_u})^2}}$$

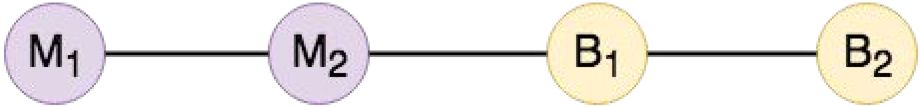
- Any two items are connected if they have common users

X-Sim: Baseline Similarity Graph Construction

- We use adjusted-cosine similarity to build this graph

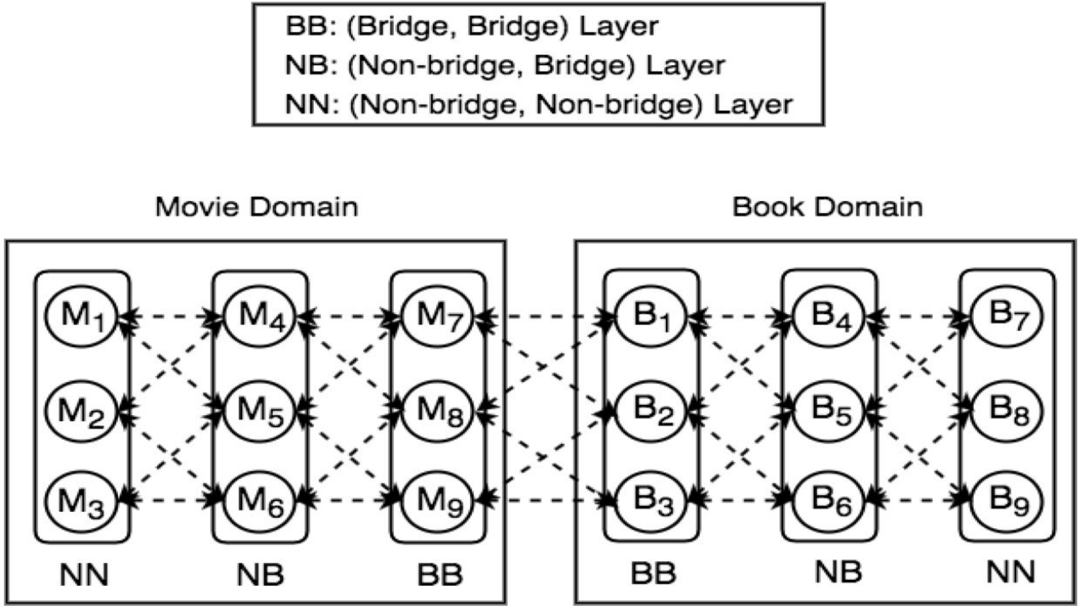
$$s_{ac}(i,j) = \frac{\sum_{u \in Y_i \cap Y_j} (r_{u,i} - \bar{r_u})(r_{u,j} - \bar{r_u})}{\sqrt{\sum_{u \in Y_i} (r_{u,i} - \bar{r_u})^2} \sqrt{\sum_{u \in Y_j} (r_{u,j} - \bar{r_u})^2}}$$

- Any two items are connected if they have common users
- We extend the current similarities using *meta-paths*
  - Meta-paths connect heterogenous items e.g., movies, books, songs
- Meta-paths captures more heterogeneous similarities



X-Sim: Layer-based K-NN

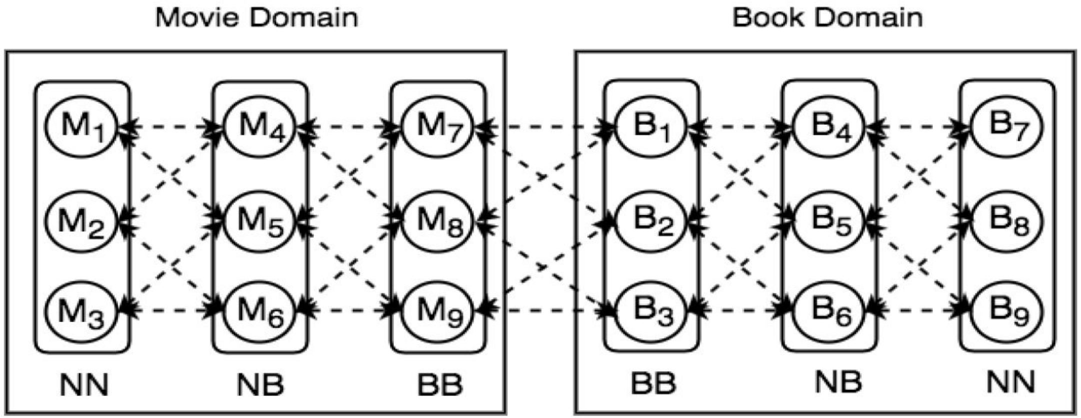
- Multiple meta-paths are possible in a heterogeneous graph
- K-NN connections across layers

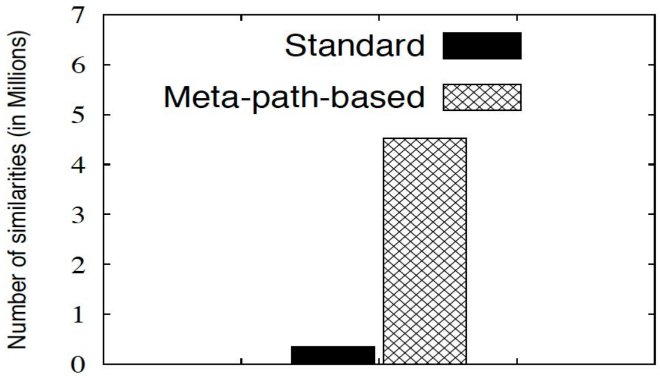


X-Sim: Layer-based K-NN

- Multiple meta-paths are possible in a heterogeneous graph
- K-NN connections across layers

BB: (Bridge, Bridge) Layer
NB: (Non-bridge, Bridge) Layer
NN: (Non-bridge, Non-bridge) Layer





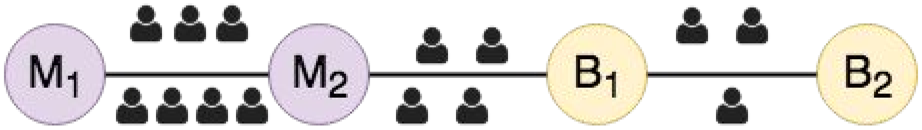
X-Sim: Meta-path based similarities

- *Weighted Significance* (adjacent items): Mutually agreeing (like/dislike) users
  - Higher number of users implies more significance

$$S_{i,j} = \underbrace{\left|Y_{i\geq\bar{i}} \cap Y_{j\geq\bar{j}}\right|}_{Mutual\ like} + \underbrace{\left|Y_{i<\bar{i}} \cap Y_{j<\bar{j}}\right|}_{Mutual\ dislike}$$

  - Normalized weighted significance

$$\widehat{S}_{i,j} = \frac{S_{i,j}}{n(Y_i \cup Y_j)}$$



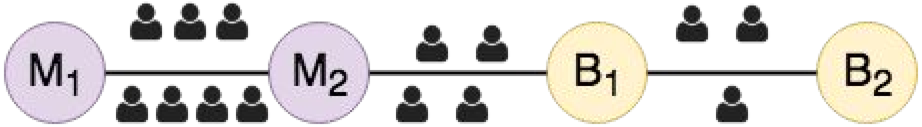
X-Sim: Meta-path based similarities

- *Weighted Significance* (adjacent items): Mutually agreeing (like/dislike) users
  - Higher number of users implies more significance

$$S_{i,j} = \underbrace{\left|Y_{i\geq\bar{i}} \cap Y_{j\geq\bar{j}}\right|}_{Mutual\ like} + \underbrace{\left|Y_{i<\bar{i}} \cap Y_{j<\bar{j}}\right|}_{Mutual\ dislike}$$

  - Normalized weighted significance

$$\widehat{S}_{i,j} = \frac{S_{i,j}}{n(Y_i \cup Y_j)}$$



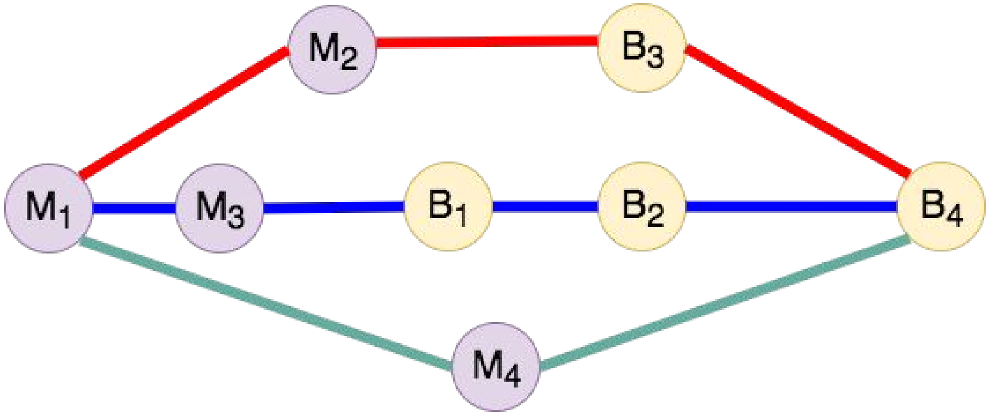
- *Meta-path-based similarity:* Baseline similarity weighted with significance

$$s_p = \frac{\sum_{t=1}^{t=k-1} S_{i_t,i_{t+1}} \cdot s_{ac}(i_t, i_{t+1})}{\sum_{t=1}^{t=k-1} S_{i_t,i_{t+1}}}$$

X-Sim: Path Certainty

- *Path Certainty:* Captures the importance of paths.
  - Longer paths are considered to be less important than shorter ones[1]

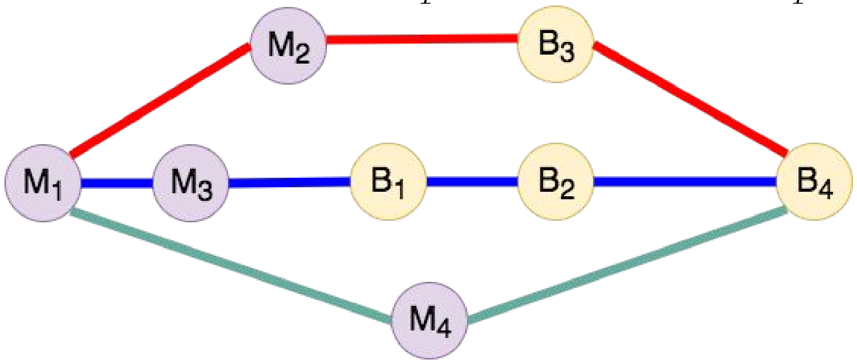
$$c_p = \prod_{t=1}^{t=k-1} \widehat{S}_{i_t, i_{t+1}}$$



[1]. Ramakrishnan et al. "Privacy risks in recommender systems." IEEE Internet Computing 5.6 (2001): 54.

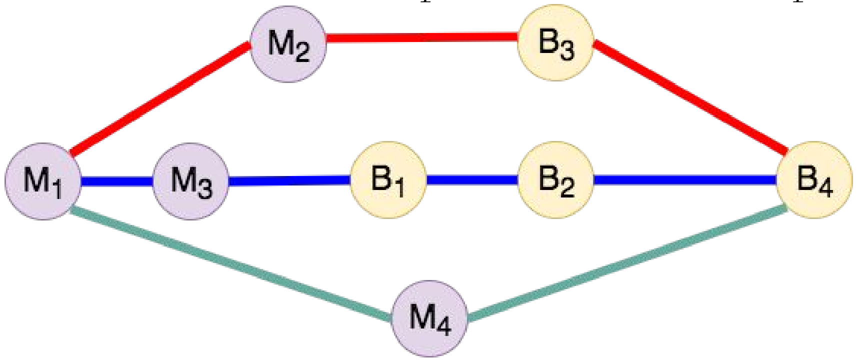
X-Sim: Cross-domain similarities

- For any two items i and j, there are multiple meta-paths between them
  - Each meta-path has similarity (s_p) and certainty (c_p) values



X-Sim: Cross-domain similarities

- For any two items i and j, there are multiple meta-paths between them
  - Each meta-path has similarity (s_p) and certainty (c_p) values



- *X-Sim:* Cross-domain similarities between two heterogeneous items i and j.

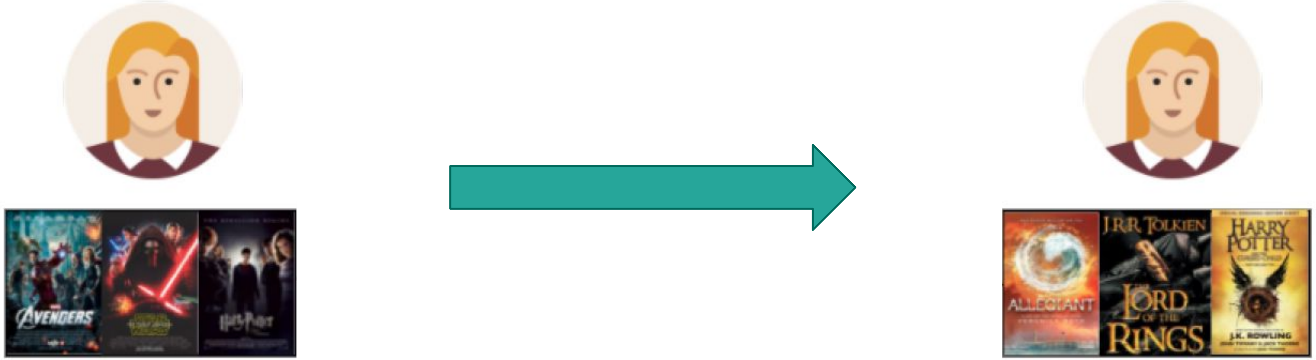
$$\text{X-SIM}(i,j) = \frac{\sum_{p \in P(i,j)} c_p \cdot s_p}{\sum_{p \in P(i,j)} c_p}$$

Outline

- Introduction
- X-Map
  - X-Sim
  - **AlterEgo**
  - Recommendation
- Experiments & Conclusion

AlterEgo generation

- Create an AlterEgo profile of the user in the target domain



Movie profile of Alice

Book AlterEgo of Alice

Alice's AlterEgo profile (in target domain) mapped from her original profile (in source domain).

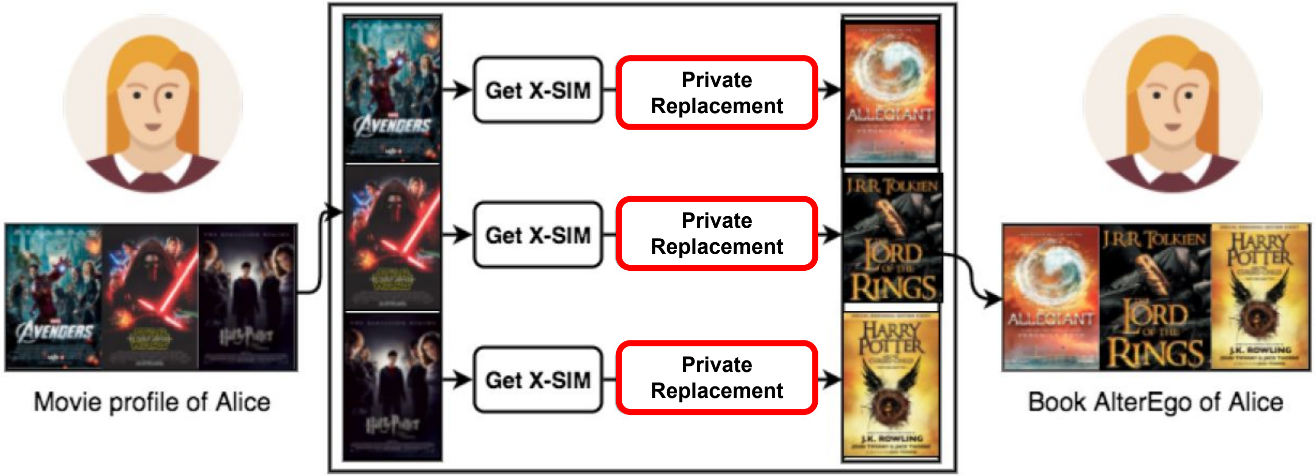
AlterEgo generation

- Create an AlterEgo profile of the user in the target domain

Alice's AlterEgo profile (in target domain) mapped from her original profile (in source domain).

AlterEgo generation (Private)

- Use probabilistic replacement (exponential mechanism for differential privacy[2])



*Alice's **<u>private</u>** AlterEgo profile (in target domain) mapped from her original profile (in source domain)*

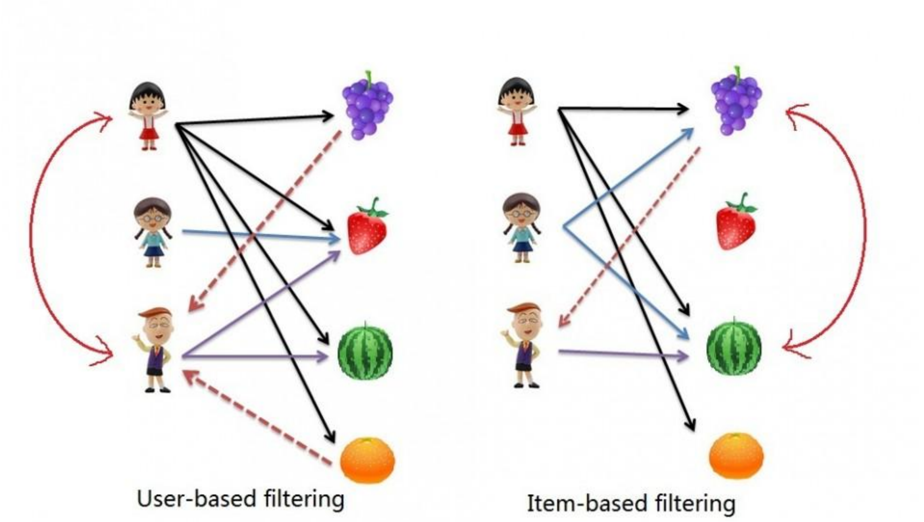
[2]. Dwork, Cynthia, and Aaron Roth. "The algorithmic foundations of differential privacy." *Foundations and Trends® in Theoretical Computer Science* 9.3–4 (2014): 211-407.

Outline

- Introduction
- X-Map
  - X-Sim
  - AlterEgo
  - **Recommendation**
- Experiments & Conclusion

Recommendation: Algorithms

- Any homogenous algorithm can be applied due to AlterEgos in X-Map.
- X-Map currently supports:
  - User-based collaborative filtering
  - Item-based collaborative filtering
- Temporal dynamics
  - AlterEgos preserve the temporal pattern
  - Capturing preference change of users
  - More accurate recommendations[3]

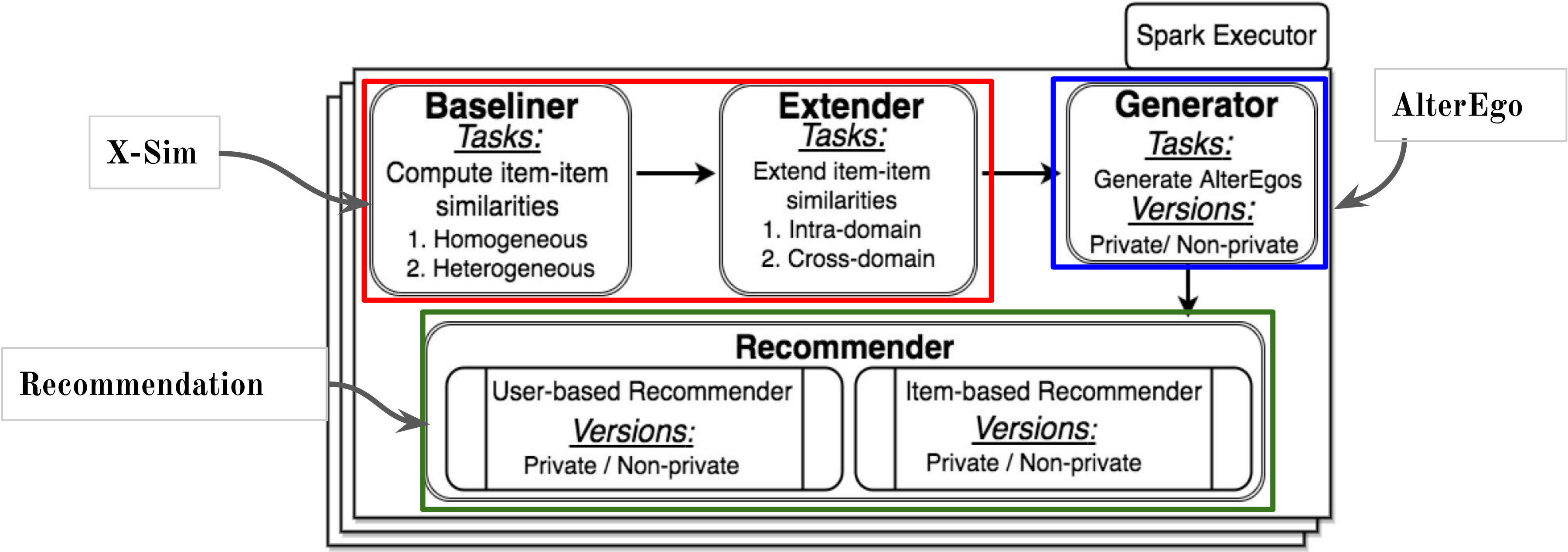


[3]. Koren, Yehuda. "Collaborative filtering with temporal dynamics." *Communications of the ACM* 53.4 (2010): 89-97.

Recommendation: Privacy

- Within-domain privacy-preserving algorithms in X-Map
    - ε-differential privacy based on recommendation-aware sensitivity.
    - Supports both user-based and item-based algorithms

Framework



Spark Executor
Baseliner
Tasks:
Compute item-item similarities
1. Homogeneous
2. Heterogeneous
Extender
Tasks:
Extend item-item similarities
1. Intra-domain
2. Cross-domain
Generator
Tasks:
Generate AlterEgos
Versions:
Private/ Non-private
Recommender
User-based Recommender
Versions:
Private / Non-private
Item-based Recommender
Versions:
Private / Non-private
X-Sim
AlterEgo
Recommendation

Outline

- Introduction
- X-Map
  - X-Sim
  - AlterEgo
  - Recommendation
- **Experiments & Conclusion**

Experimental Setup

Datasets: Amazon movies and books

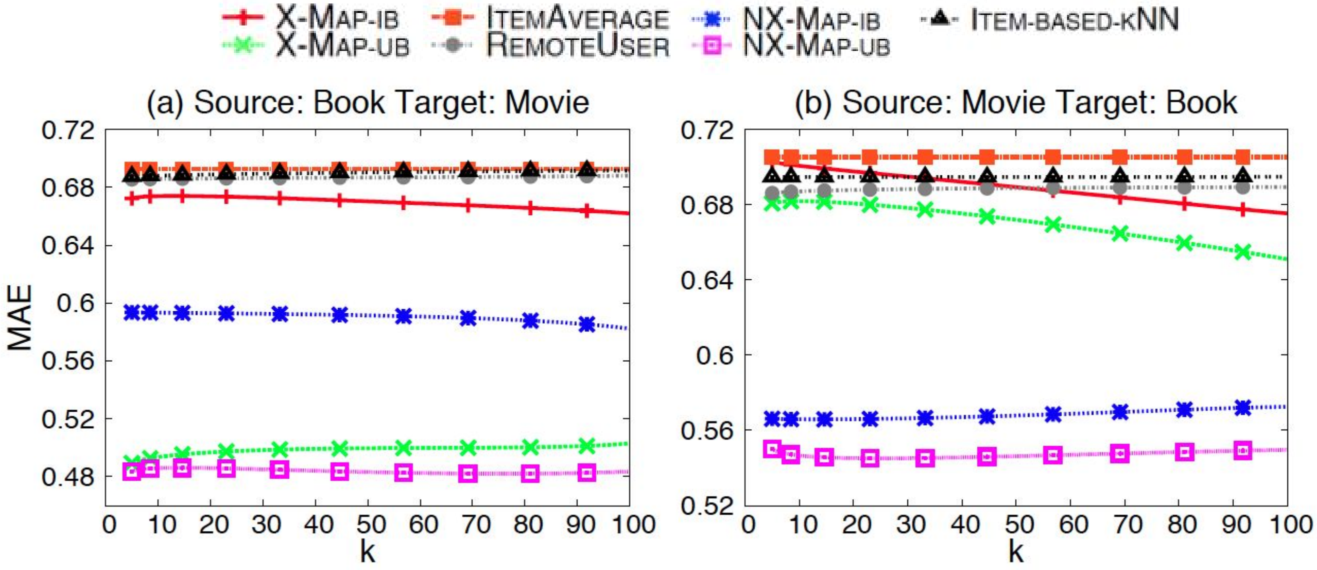
Domain	User	Items	Ratings
Movies	473,764	128,402	1,671,662
Books	725,846	403,234	2,708,839

Framework: Apache Spark

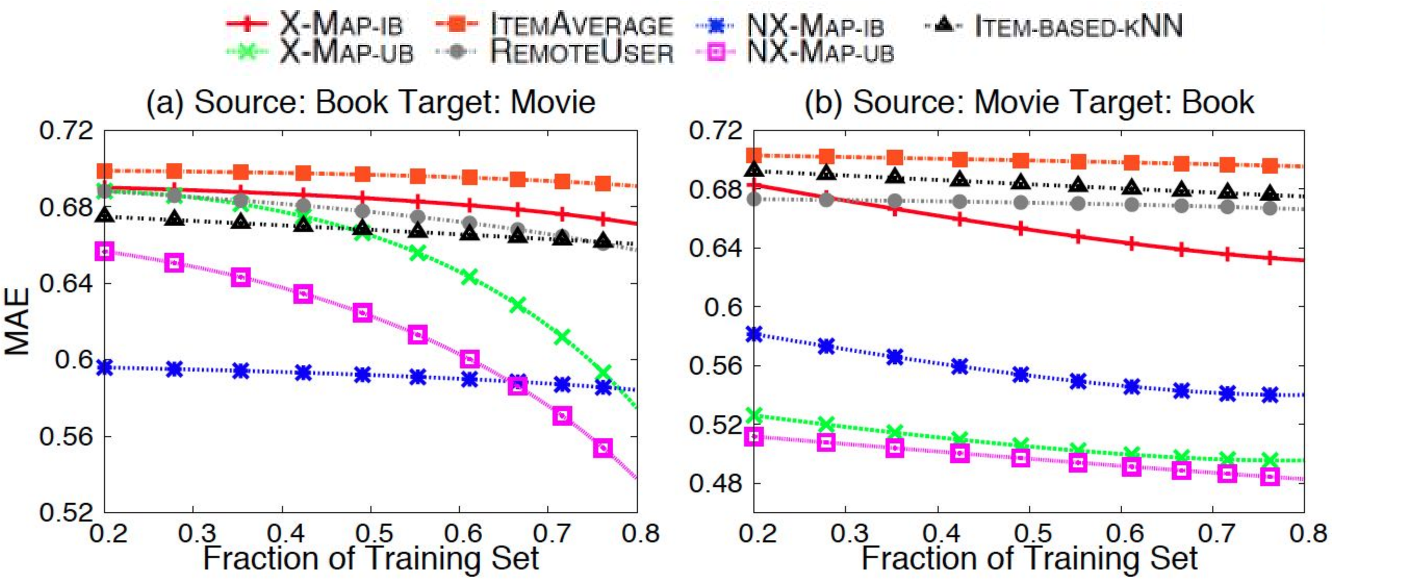
Metric: Mean Absolute Error

$$MAE = \frac{1}{n}\sum_{t=1}^{n}\left|x_t - \hat{x}_t\right|$$

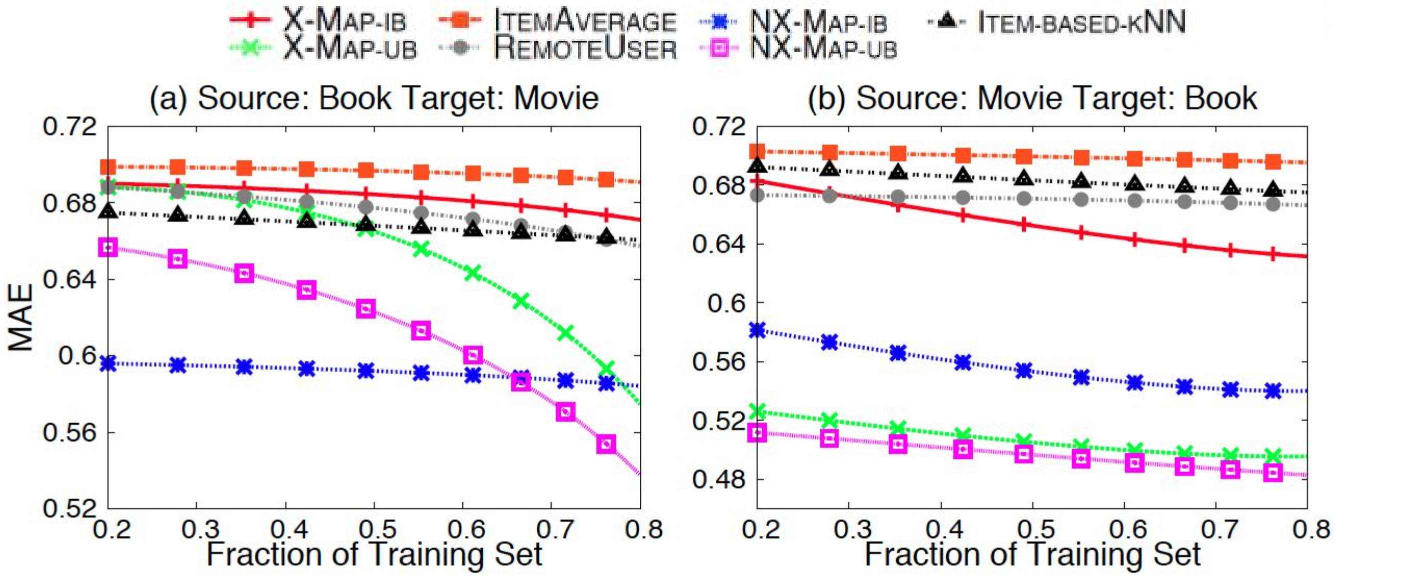
Quality: Accuracy comparison



Quality: Impact of training set

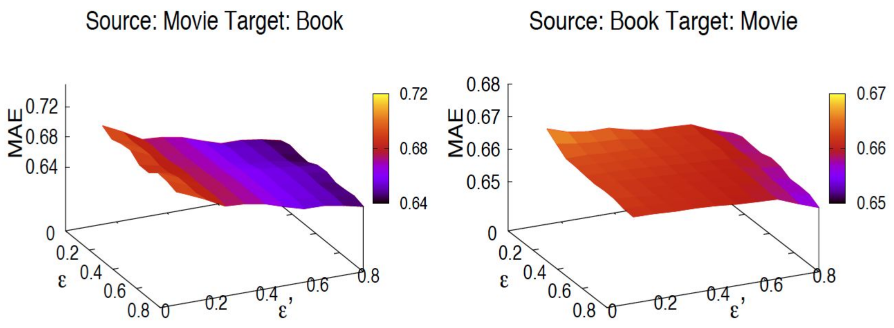


Quality: Impact of training set

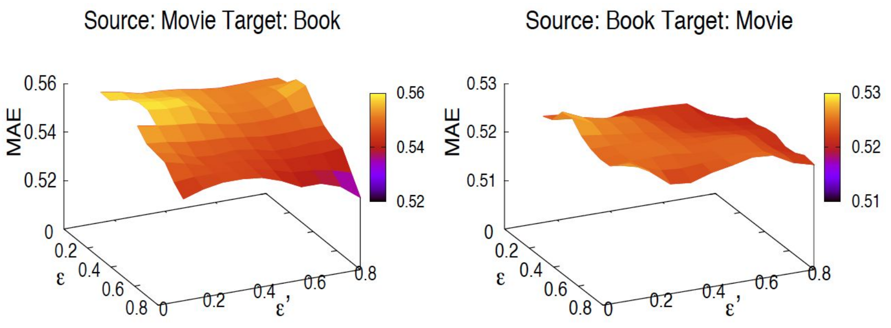


Item-to-item similarities are more static than user-to-user similarities

Privacy

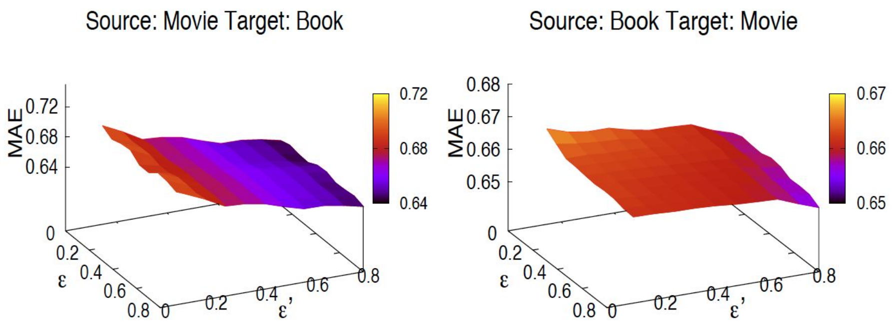


Item-based approach

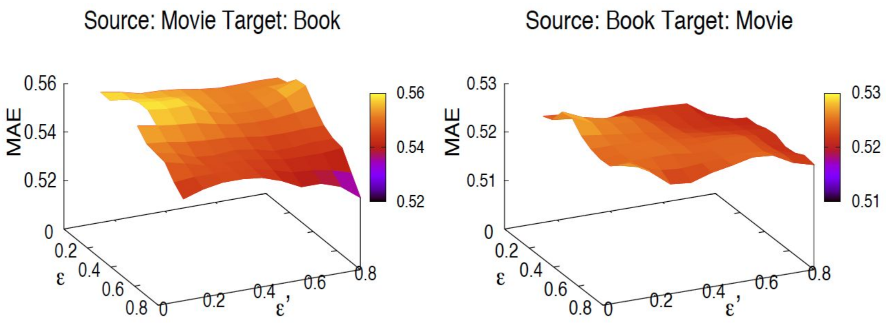


User-based approach

Privacy



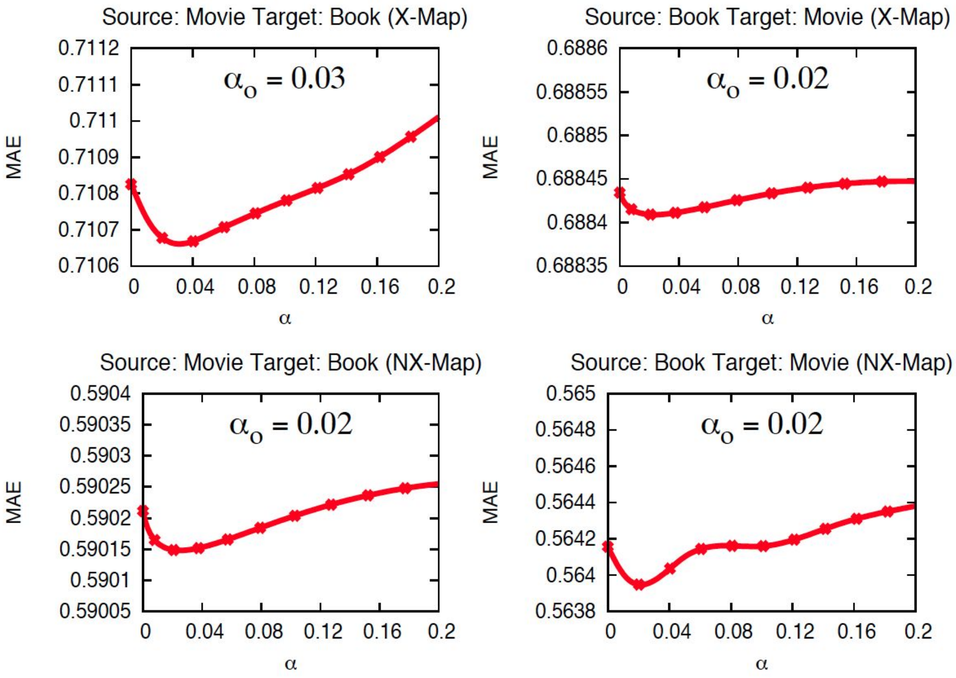
Item-based approach



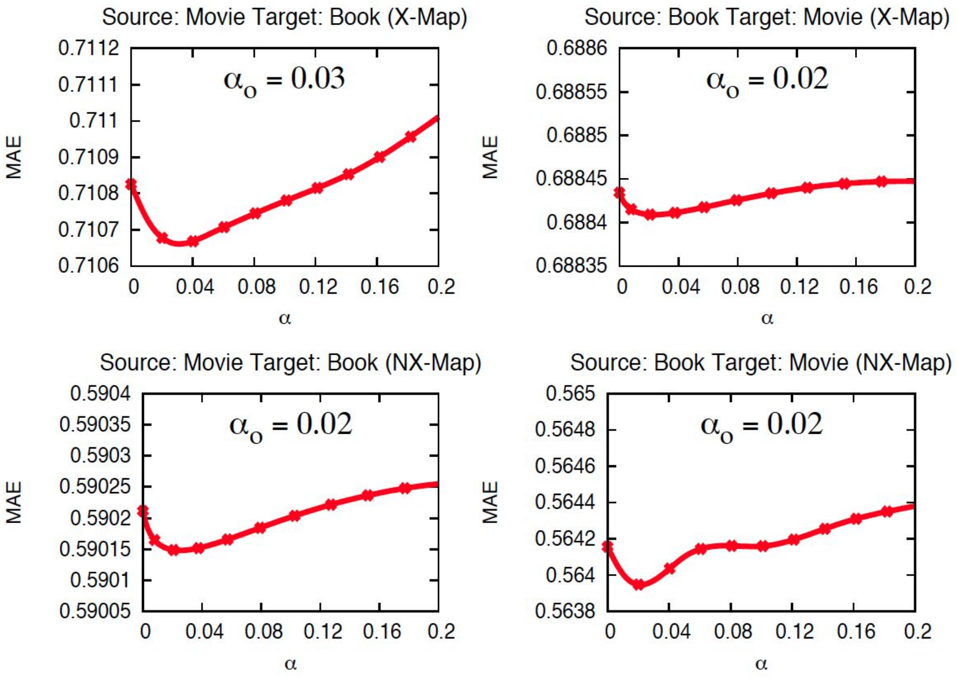
User-based approach

1. *Differential privacy: Lower ε leads to better privacy*
2. *Higher privacy leads to lower quality due to noise addition.*

Temporality



Temporality



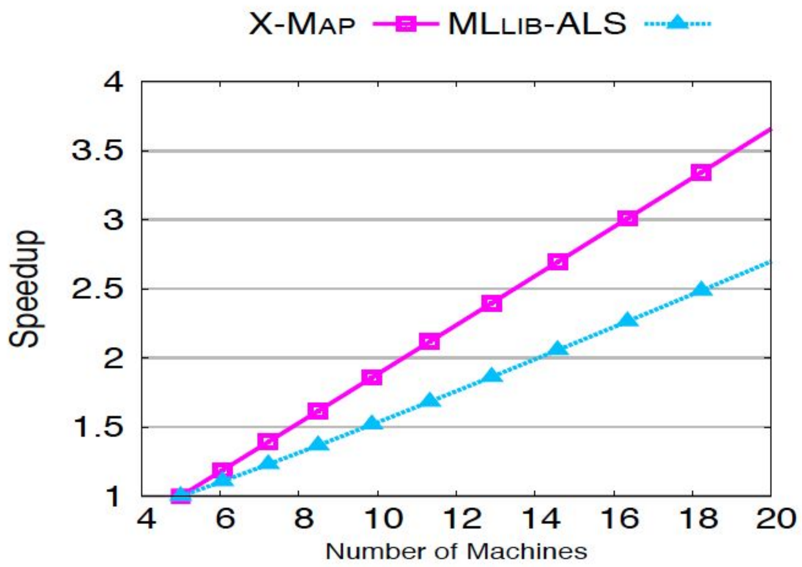
- *AlterEgos preserve temporal behavior of users.*
- *A significantly higher value of α affects negatively due to users with small profiles.*

Homogeneous scenario (Movielens)

D_1		D_2	
Genres	**Movie counts**	**Genres**	**Movie counts**
Drama	13344	Comedy	8374
Thriller	4178	Romance	4127
Action	3520	Crime	2939
Horror	2611	Documentary	2471
Adventure	2329	Sci-Fi	1743
Mystery	1514	Fantasy	1412
War	1194	Children	1139
Musical	1036	Animation	1027
Western	676	Film-Noir	330
Other	196	–	–

	NX-Map	X-Map	MLlib-ALS
MAE	**0.6027**	0.6830	0.6729

Scalability



Scales sub-linearly with multiple machines

Summary

- X-Map: Heterogenous recommender that provides
  - Quality (X-Sim, Temporality)
  - Privacy
  - Scalability
  - Prototype: http://x-map.work/

Summary

- X-Map: Heterogenous recommender that provides
    - Quality (X-Sim, Temporality)
    - Privacy
    - Scalability
    - Prototype: http://x-map.work/

Thank You

Quality: Impact of Sparsity