# A Game Theoretic Framework for Data Privacy Preservation in Recommender Systems

Maria Halkidi

Dept of Digital Systems
University of Piraeus

Iordanis Koutsopoulos

Dept. of Computer & Communication Engineering, Univ. of Thessaly
and CERTH

# Introduction to recommender stystems

- Recommender systems arise with various contexts
  - providing personalized search results and targeted advertising,
  - making social network related suggestions, providing personalized suggestions on various goods and services
- Users rely on recommendation systems for quick and accurate personalized expert advice and suggestions which aid them in decision making.
- The efficiency of a recommender system amounts to high-quality personalized recommendations for different users.
- The quality of recommendations for an individual user relies on
  - past experience and participation of other users in the rating process.

# Privacy preservation in recommendation systems

- Recommendation systems: data exchange between the users and third party that performs recommendations → user privacy concerns
- User optimal strategy:
  - preserve privacy by not revealing much information to third party about private personal preferences and ratings.
  - receive high-quality recommendation results
  - Fundamental tradeoff between privacy preservation and recommendation quality.
- Goal: Study interaction of multiple users, each of which selfishly aims at optimal strategy
  - Strategy of a user depends on strategies of other users !

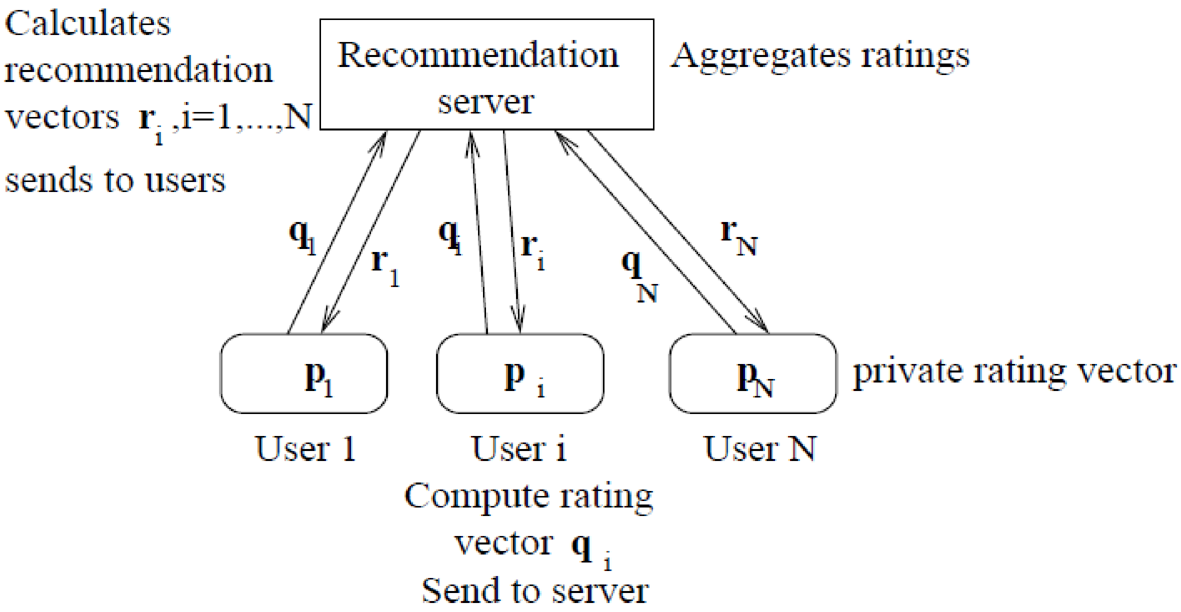
# Contributions

- We develop a mathematical framework for quantifying the goal of
  - privacy preservation and good quality recommendations
- We employ game theory to model and study the interaction of multiple users
  - we derive conditions and expressions for the Nash Equilibrium Point (NEP).
- We study a hybrid recommendation system,
  - **Collaborative filtering (CF) approaches:** a user is recommended items based on past ratings of other users.
  - **Content-based approaches** provide recommendations by comparing the content of an item to the content of items of potential interest for a user.

# Model: Ratings and recommendation

- Set $U$ of $N$ users and set of items $I$ available for recommendation.
- $S_i \subset I$: a small subset of items that a user i has already viewed, purchased (or obtained experience of)
- $p_i = (p_{ik} : k \in S_i)$: vector of ratings of user i for the items he has viewed
  - $0 \leq p_{ik} \leq P$ (continuous-valued)
  - Vector of ratings $p_i$ is private information for each user i.
- $q_i = (q_{ik} : k \in S_i)$: vector of declared ratings from user i to the server (can be different from $p_i$).
  - This is the strategy of user i

# Model: Ratings and recommendation (2)



- P = ($p_i$ : i ∈ U): ensemble of private ratings of users.
- Q = ($q_i$ : i ∈ U): ensemble of declared ratings of users to server.
- Recommendation server collects declared user profiles and issues personalized recommendations to different users.
  - Computes recommendation vector $r_i = f_i(Q) = f_i(q_1, \ldots, q_N)$ for each user i.

# Recommendation quality

- Recommendation $r_i = f_i(q_i, q_{-i})$. that each user receives depends on:
  - declared profiles of other users to the server, $q_{-i} = (q_1, \ldots, q_{i-1}, q_{i+1}, \ldots, q_N)$
  - the declared profile of this specific user, $q_i$
- $\tilde{\mathbf{r}}_i = f_i(\mathbf{p}_i, \mathbf{q}_{-i})$ : resulting recommendation vector if user i declared his true profile (with other users' profiles fixed)
- User wants to calibrate its declared profile so as to receive recommendations close to the ones he would receive if he has declared his true private profile, regardless of the declaration of others.
- This is quantified by the following constraint for user i:

$$(\mathbf{r}_i - \tilde{\mathbf{r}}_i)^2 \leq D \;\Leftrightarrow\; [f_i(\mathbf{q}_i, \mathbf{q}_{-i}) - f_i(\mathbf{p}_i, \mathbf{q}_{-i})]^2 \leq D\,,$$

  - D denotes max distortion that can be tolerated in recommendation by user i.

# Privacy metric

- Quantifies degree at which privacy is preserved for user i.
- Depends on the private profile $p_i$ and the declared profile $q_i$ of user i.
  - Actually the distance between them
  - We denote this dependence by a continuous function **$g_i(p_i, q_i)$.**

# Problem formulation (1)

- Objective from the point of view of each user i :

$$\max_{\mathbf{q}_i} g(\mathbf{p}_i, \mathbf{q}_i)$$

Privacy maximization
submit rating profile that is sufficiently far from real (private) profile

subject to:

$$[f_i(\mathbf{q}_i, \mathbf{q}_{-i}) - f_i(\mathbf{p}_i, \mathbf{q}_{-i})]^2 \leq D$$

maintain recommendation vector close enough (in distance at most D) to the one he would get if he declared true private profile.

Conflict among users, each of which tries to solve problem above from its own view (selfish / rational)

# Problem formulation (2)

- The satisfiability of the constraint of a user i depends on , $q_{-i}$ (other users' strategies)

- **Definition of NEP:** A strategy profile $\mathbf{Q}^* = (\mathbf{q}_1^*, \ldots, \mathbf{q}_N^*)$ is called Nash Equilibrium Point (NEP) if for each user i = 1, . . . ,N, the following property holds:
- In the NEP $(\mathbf{q}_1^*, \ldots, \mathbf{q}_N^*)$, no user i can further increase its privacy preservation metric by altering his declared profile to $\mathbf{q}_i \neq \mathbf{q}_i^*$ , provided that all other users stay with their NEP declared profiles.

$$g(\mathbf{p}_i, \mathbf{q}_i^*) \geq \max_{\mathbf{q}_i \in F(\mathbf{q}_{-i}^*)} g(\mathbf{p}_i, \mathbf{q}_i) \; \forall \mathbf{q}_i \neq \mathbf{q}_i^*$$

# Case Study: Hybrid Recommendation systems

- Collaborative filtering + Content-based approaches
- For each user i, the recommendation server applies the following measure to compute metrics $r_{i\ell}$ for items $\ell \notin \mathcal{S}_i,\ \ell \in \mathcal{S}_j \text{ for } j \neq i$, so as to rate them and include them in the recommendation vector for user i:

$$r_{i\ell} = \frac{1}{N-1} \underbrace{\sum_{\substack{j \neq i: \\ \ell \in \mathcal{S}_j}} q_{j\ell}}_{\text{Collaborative filtering}} \cdot \frac{1}{|\mathcal{S}_i|} \underbrace{\sum_{k \in \mathcal{S}_i} \rho_{k\ell} q_{ik}}_{\text{Content-based}}$$

$\rho_{k\ell} \in [0, 1]$ is the correlation between items k and ℓ.

The server computes the metric above for all $\ell \notin \mathcal{S}_i$ and forms vector $r_i$.

# Privacy preservation

- The function g( · ) that quantifies privacy preservation for user i is taken:

$$g(\mathbf{p}_i, \mathbf{q}_i) = \sum_{k \in \mathcal{S}_i} p_{ik}(p_{ik} - q_{ik})^2$$

- Privacy preservation increases as Euclidean distance $\sum_{k \in \mathcal{S}_i} (p_{ik} - q_{ik})^2$ increases.
- Distance is weighted by the private rating $p_{ik}$
  - among items whose private and declared rating have the same distance, it is preferable from privacy preservation perspective to change rating of items that are higher rated in reality

# Recommendation quality

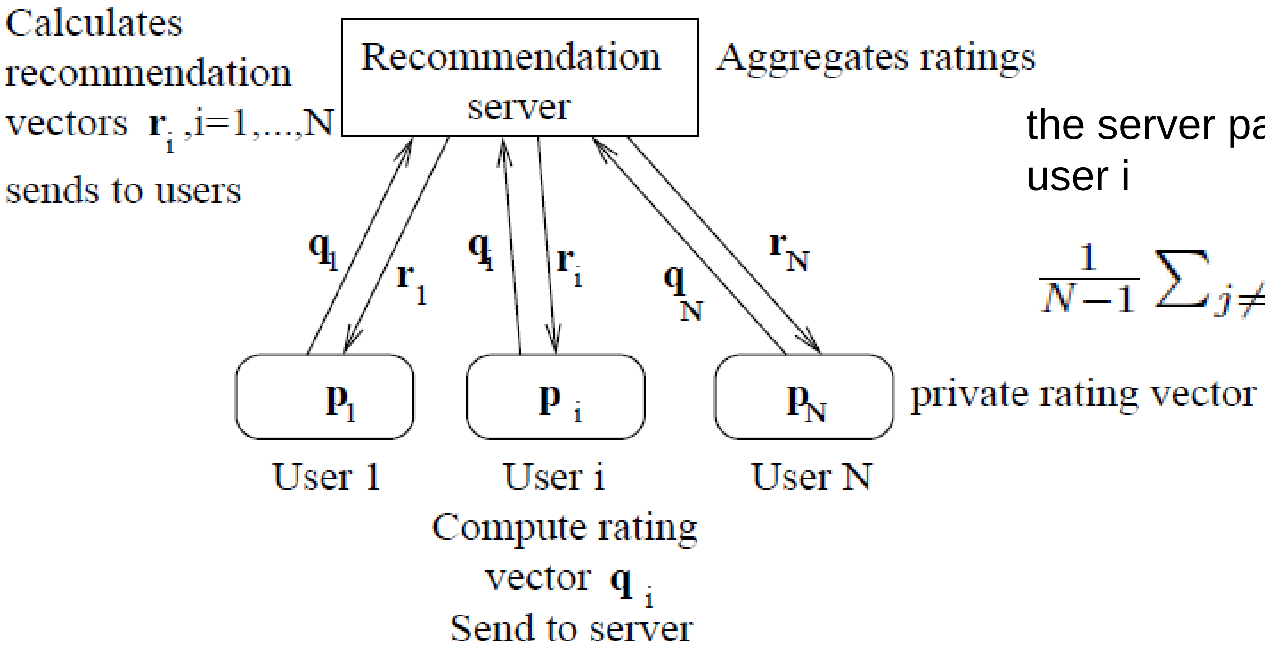
- Quality of recommendation is measured in terms of the difference between the recommendation user i gets if he declares profile $q_i$ and the one he would get if he declared the real private rating $p_i$, regardless of what other users do. Other users j ≠ i make in general declarations $q_j$.

$$(\mathbf{r}_i - \tilde{\mathbf{r}}_i)^2 \le D \Leftrightarrow [f_i(\mathbf{q}_i, \mathbf{q}_{-i}) - f_i(\mathbf{p}_i, \mathbf{q}_{-i})]^2 \le D\,,$$

- The constraint that needs to be fulfilled for acceptable recommendation quality for user i is written as

$$\frac{1}{|\mathcal{S}_i|}\sum_{k\in\mathcal{S}_i}\sum_{\ell\notin\mathcal{S}_i}\frac{1}{N-1}\sum_{\substack{j\neq i:\\ \ell\in\mathcal{S}_j}} q_{j\ell}\cdot\rho_{k\ell}(q_{ik}-p_{ik})^2 \le D$$

# System architecture



the server passes to user i

$$\frac{1}{N-1}\sum_{j\neq i} q_{j\ell}^{(t)}$$

The agent of user i solves the optimization problem

$$\max_{\mathbf{q}_i^{(t)}} g(\mathbf{p}_i, \mathbf{q}_i^{(t)}) = \sum_{k\in\mathcal{S}_i} p_{ik}\left(p_{ik} - q_{ik}^{(t)}\right)^2$$

$$\frac{1}{|\mathcal{S}_i|}\sum_{k\in\mathcal{S}_i}\sum_{\ell\notin\mathcal{S}_i}\frac{1}{N-1}\sum_{\substack{j\neq i:\\ \ell\in\mathcal{S}_j}} q_{j\ell}\cdot\rho_{k\ell}(q_{ik}-p_{ik})^2 \le D$$

# Game theoretic analysis (1)

- Set $x_{ik} = (p_{ik} - q_{ik})^2$, and $\mathbf{x}_i = (x_{ik} : k \in \mathcal{S}_i)$

- Problem (P) that is solved by each user i at each iteration is written as a Linear Programming one:

$$\max_{\mathbf{x}_i} \sum_{k \in \mathcal{S}_i} p_{ik} x_{ik}, \qquad \text{subject to:} \qquad \sum_{k \in \mathcal{S}_i} \beta_{ik} x_{ik} \leq D(N-1)$$

with

$$\beta_{ik} = \frac{1}{|\mathcal{S}_i|} \sum_{\ell \notin \mathcal{S}_i} \sum_{j \neq i : \ell \in \mathcal{S}_j} q_{j\ell} \rho_{k\ell}$$

# Game theoretic Analysis (2)

- The solution to this problem is found among the extreme points of the feasible set. Each user finds item, $k^* = \arg\min_{k \in \mathcal{S}_i} \frac{\beta_{ik}}{p_{ik}}$

and it sets

$$x_{ik^*} = D(N-1)|\mathcal{S}_i| \frac{p_{ik^*}}{\beta_{ik^*}}$$

$$q_{ik^*} = p_{ik^*} \pm \sqrt{\frac{D(N-1)|\mathcal{S}_i| p_{ik^*}}{\beta_{ik^*}}}$$

- while $q_{ik} = p_{ik}$ for other items k ≠ k*.
- Agent i maximizes its preserved privacy if it declares its true private profile for all viewed items, except one, k*, for which the quantity

$$\frac{\beta_{ik}}{p_{ik}} = \frac{\sum_{\ell \notin \mathcal{S}_i} \rho_{k\ell} \sum_{j \neq i: \ell \in \mathcal{S}_j} q_{j\ell}}{p_{ik}}$$

An item should have low correlation with items that the user has not viewed
The average declared rating of other users for this item is low.

an item has more chances to be the selected one k*, if it is highly rated in its private rating vector

is the smallest among items it has viewed.

# Cooperative user strategies

- Agents may coordinate among themselves so as to mutually benefit
- Global objective G(P,Q) needs to be defined

$G(\mathrm{P},\mathrm{Q}) = \sum_{i\in\mathcal{U}} g(\mathrm{p}_i,\mathrm{q}_i)$ or $G(\mathrm{P},\mathrm{Q}) = \min_{i\in\mathcal{U}} g(\mathrm{p}_i,\mathrm{q}_i)$

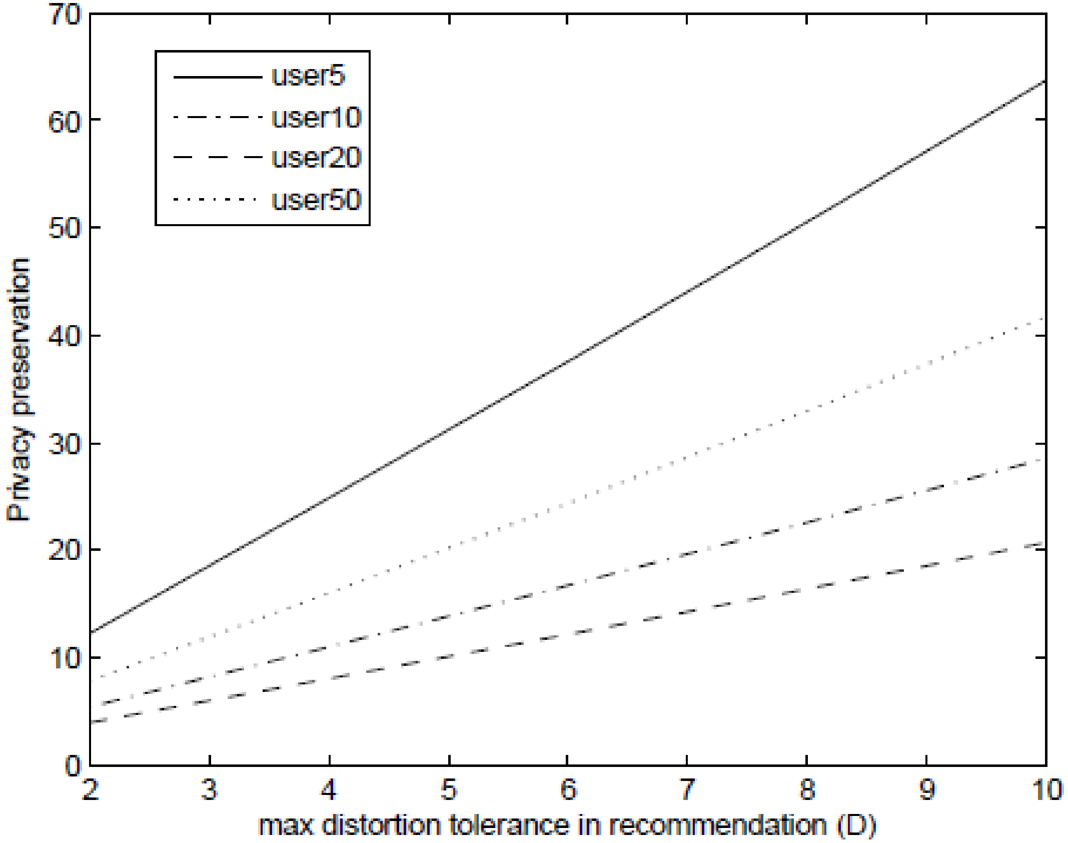
- Users act jointly in order to optimize the global constraint
- A feasible cooperation regime for N users is a joint profile declaration strategy $Q^0 = (q^0_1, \dots, q^0_N)$ such that

$$\mathrm{q}_i^0 \in F(\mathrm{q}_{-i}^0), \quad \text{and} \quad g(\mathrm{p}_i,\mathrm{q}_i^0) \geq g(\mathrm{p}_i,\mathrm{q}_i^*), \qquad \forall\, i \in \mathcal{U} \tag{1}$$

# Experimental results

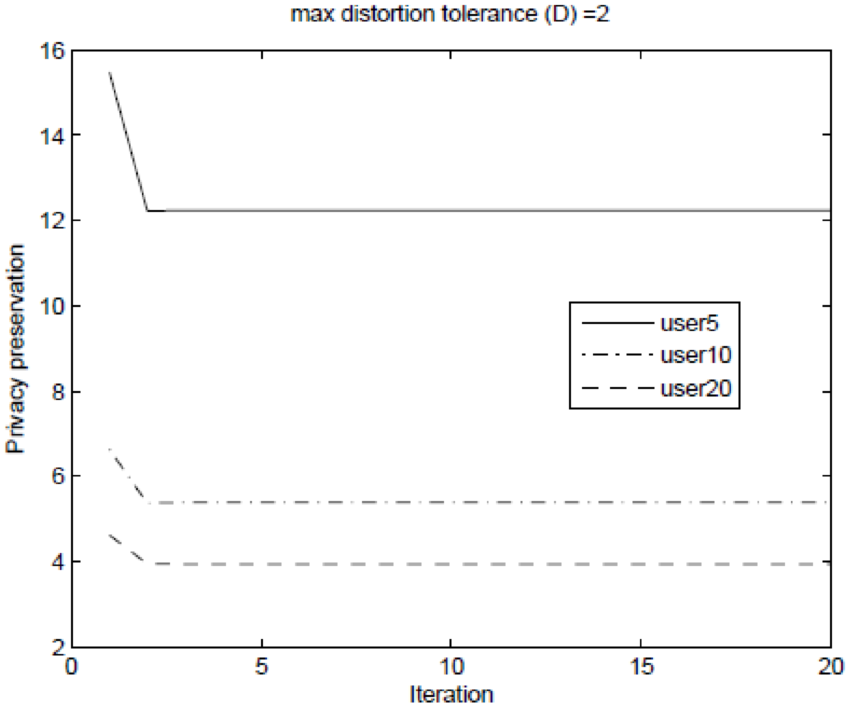
- A recommender system consisting of N = 50 users and |I| = 20 items.
- The content correlation of items is calculated a priori and announced to the agents that represent the users.
- The preferences of users for items are randomly selected for the sake of performance evaluation
- We assume that user ratings lie in the value interval [1, 5].

# Experimental results

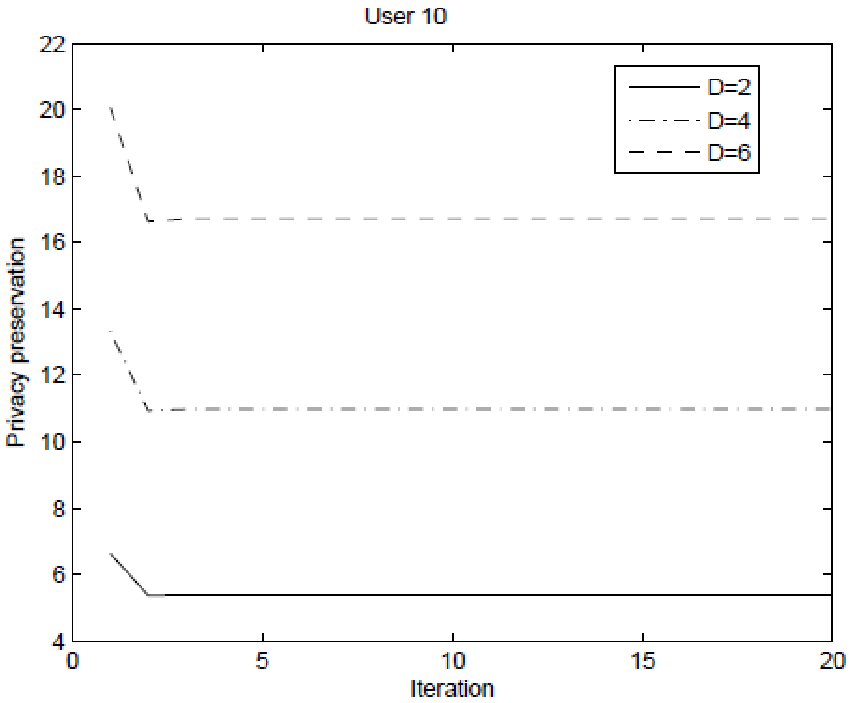


Users that are less tolerant to the error of recommendation quality, have to reveal more information about their preferences

Privacy preservation versus maximum distortion tolerance in recommendation.



Convergence of the iterative best response strategy for different users



Convergence of the iterative best response strategy for different values of D.

After a small number of iterations, usually no more than 2 − 3, the system converges and the privacy preservation metric of a user at the NEP is determined.

# Conclusions

- We took a step towards characterizing the fundamental tradeoff between privacy preservation and good quality recommendation.

- We introduced a game theoretic framework for capturing the interaction and conflicting interests of users in the context of privacy preservation in recommendation systems.

- We attempted to capture this interaction, we characterized the Nash Equilibrium Points

# Thank you!