Thy Friend is My Friend: Iterative Collaborative Filtering for Sparse Matrix Estimation

Christian Borgs* Jennifer Chayes* Christina E. Lee* Devavrat Shah+

borgs@microsoft.com jchayes@microsoft.com celee@mit.edu devavrat@mit.edu

*Microsoft Research +Massachusetts Institute of Technology

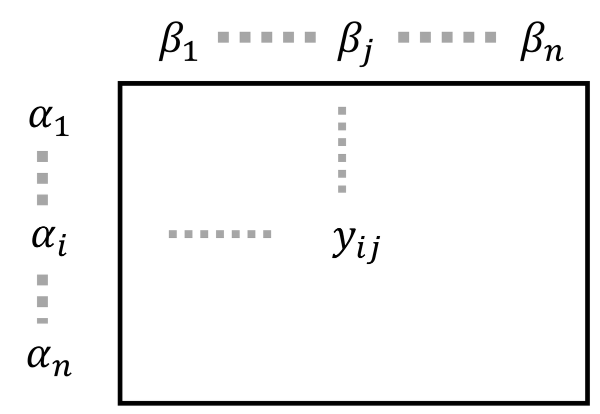
Problem Statement

Observed rating LATENT

$$\mathbb{E}[y_{ij}] = f(\alpha_i, \beta_j)$$

$$(i', j') \xrightarrow{?} \mathbb{E}[y_{i'j'}]$$

Unique IDs Predict rating



- Uniform sampling with probability p
- Row/column latent features sampled iid
- Bounded entries $y_{ui} \in [0,1]$
- Latent function has finite spectrum
  $f(\alpha_u, \beta_i) = \sum_{k=1}^{d} \lambda_k q_k(\alpha_u) q_k'(\beta_i)$

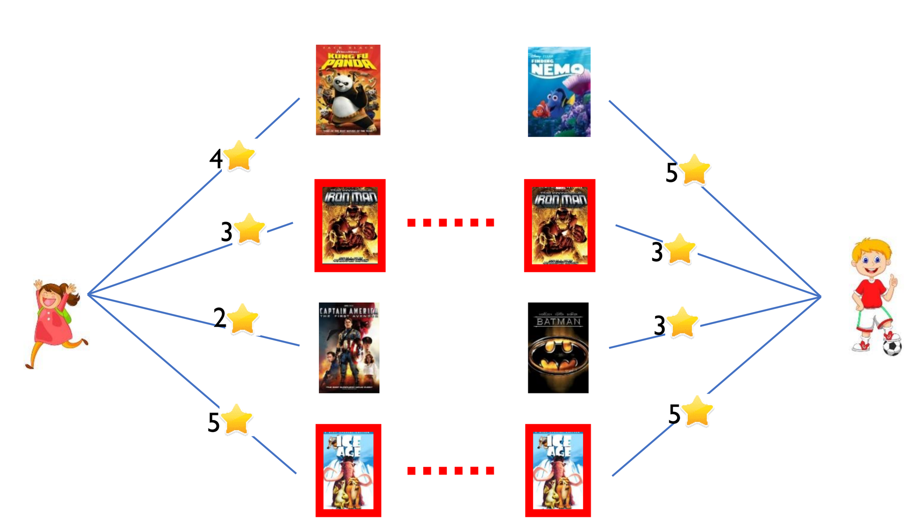
Goal: Given data $Y = \{y_{ij}\}_{ij \in \mathcal{E}}$ estimate $\mathbb{E}[Y]$

Examples include

- low rank models for matrix completion
- low rank graphon network models such as mixed membership stochastic block model

Classical Collaborative Filtering

Computing similarities relies on commonly rated movies, which requires $\Omega(n^{3/2})$ samples



Algorithm Summary

Step 1: Estimate distances by comparing product of weights along path to shared vertices in radius r neighborhoods, for $r = \Theta\left(\frac{\ln(n)}{\ln(pn)}\right)$

Step 2: Predict by averaging close neighbors

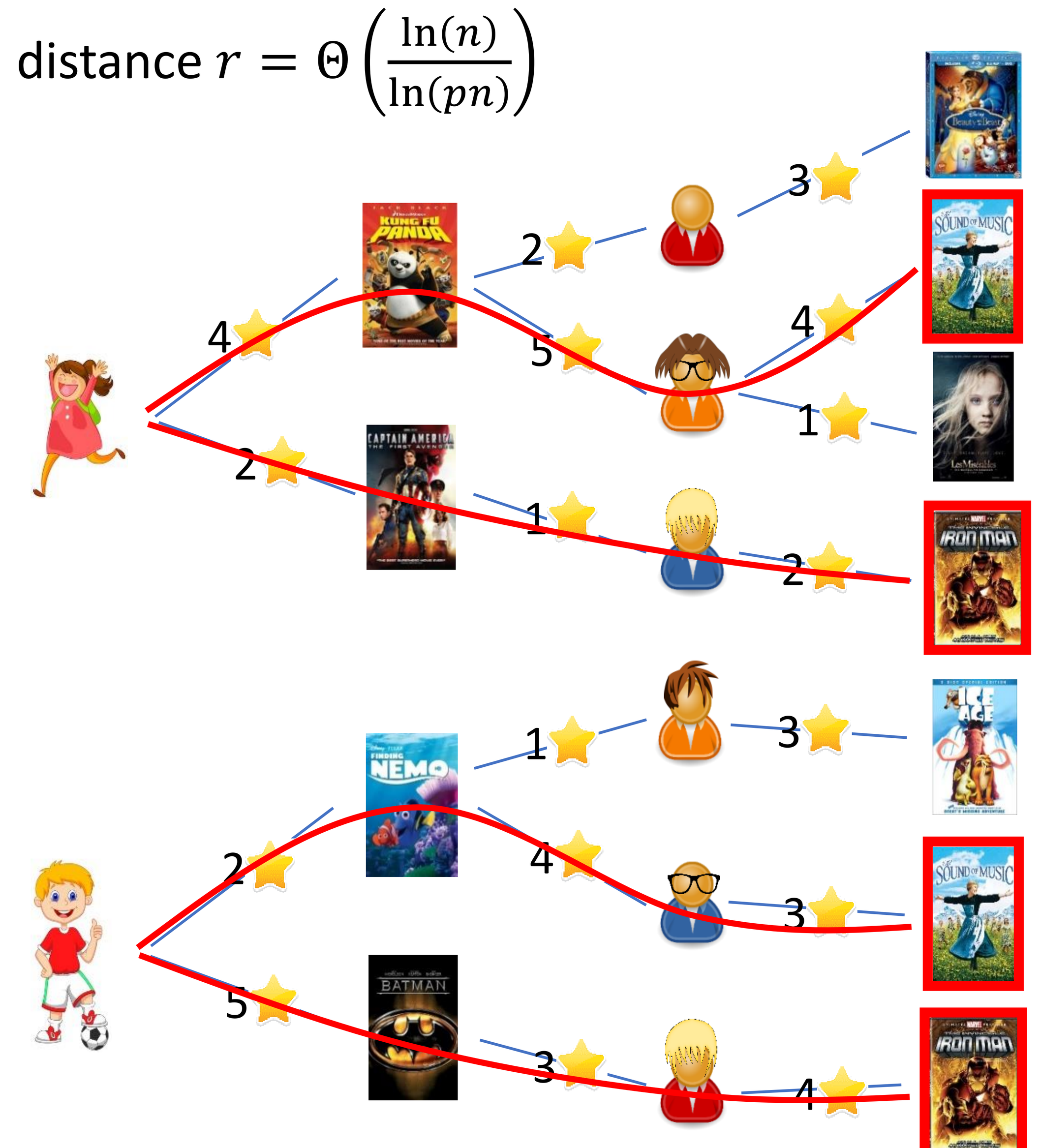
$$\hat{f}_{ui} = \frac{1}{Z} \sum_{(v,j)} y_{vj}\, \mathbb{I}(uv \text{ "close"}, ij \text{ "close"})$$

Finer details …

- Use sample splitting between algorithm steps
- Do not include loops in path (use breadth first tree)
- Reduce computation by using coarse clustering
- Multiple measurements needed when $r = \omega(1)$

Intuition: Expanding Neighborhood

We deal with sparsity by comparing product of weights along paths to common vertices at distance $r = \Theta\left(\frac{\ln(n)}{\ln(pn)}\right)$



$\mathbb{E}[Y] = U\Sigma V^T$

$$Y \cdot Y^T \cdot Y \approx U\Sigma^3 V^T$$

Compare direct neighbors $\sim \|(u - u)\Sigma\|_2^2$

Compare r distance neighbors $\sim \|(u - u)\Sigma^r\|_2^2$

*Adjust for bias by taking multiple measurements

Theorem [Borgs Chayes Lee Shah 17]

Assuming f is Lipschitz and rank d, and datapoints observed with prob $p = \omega(d^5 n^{-1})$,

$$\mathbb{E}\left[\frac{1}{n^2}\sum_{ij}\left(\hat{f}_{ij} - f(\alpha_i, \beta_j)\right)^2\right] = O(d^2 (pn)^{-2/5}).$$

If $p = \omega(d^5 n^{-1} \ln^5(n))$ with high probability,

$$\max_{ij}\left(\hat{f}_{ij} - f(\alpha_i, \beta_j)\right)^2 = O(d^2 (pn)^{-2/5}).$$

* Precise conditions of theorem have been simplified

Paper	Sample Complexity	Noise Model	Function Class
KMO10	$\Omega(dn \max(\log n, d))$	Additive, iid Gaussian	Rank d
DPBW14	$\Omega(dn \max(\log n, d))$	Binary entries	Rank d
C14	$\Omega(dn \log^6 n)$	Independent bounded	Rank d
XML14	$\Omega(n \log n)^*$	Binary entries	Rank d
AS15	$\omega(n)^*$	Binary entries	piecewise constant (d blocks)
BCLS17 (ours)	$\omega(d^5 n)$	Independent bounded	Rank d

* does not indicate dependence on d

** many results not reported here

Proof Sketch

Notation assuming symmetric model:

- $\mathbb{E}[y_{ij}] = f(\alpha_i, \alpha_j) = \sum_{k=1}^{d} \lambda_k q_k(\alpha_i) q_k(\alpha_j)$
- Let $Q_{ki} = q_k(\alpha_i)$ and diagonal $\Lambda_{kk} = \lambda_k$
- Let $\mathcal{P}_{iu}$ be shortest path from i to u
- Let $N_{iu}^{(r)} = \mathbb{I}_{\{|\mathcal{P}_{iu}|=r\}} \prod_{(a,b)\in\mathcal{P}_{iu}} y_{ab}$

Key piece of proof is to show that for all k

$$e_i^T N^{(r)} Q^T e_k \approx e_i^T Q^T \Lambda^{\mathrm{r}} e_k$$

Distance between i and j is estimated with

$$(e_i - e_j)^T N^{(r)} Y (N^{(r)})^T (e_i - e_j)$$
$$\approx (e_i - e_j)^T Q^T \Lambda^{2r+1} Q (e_i - e_j)$$
$$= \left\|\Lambda^{r+1/2} Q (e_i - e_j)\right\|_2^2$$