Fen Wang
PhD student, Xidian University
22, August 2019

Graph Sampling for Matrix Completion

Outlines

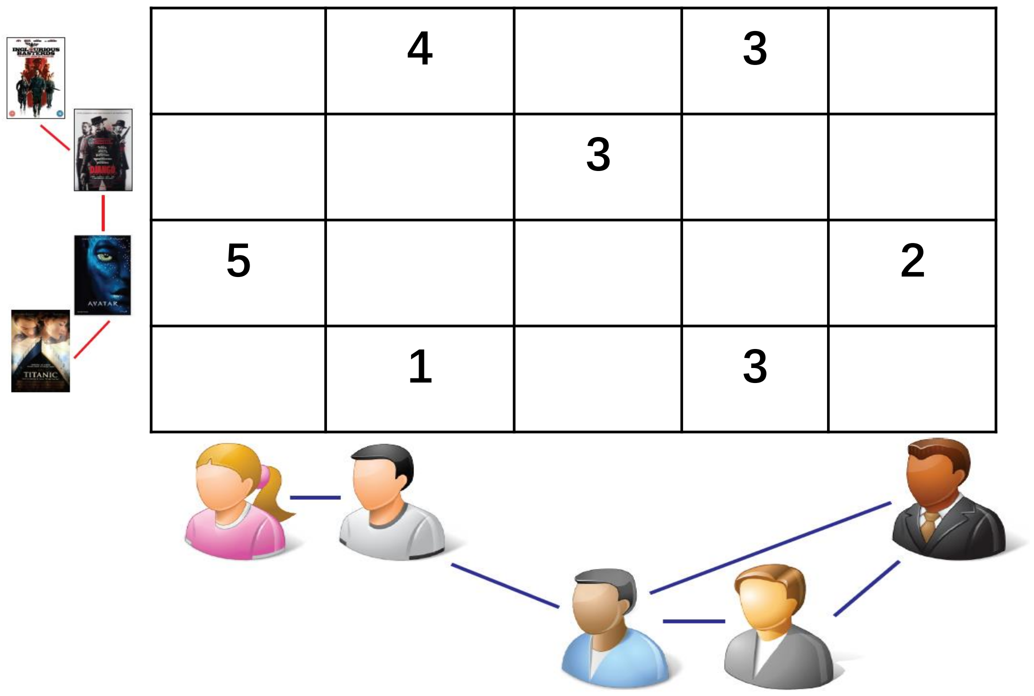
- Matrix Completion
- Sampling for Matrix Completion
- Matrix Completion with Dual Graph Smoothness Prior
- Dual Graph Smoothness Prior based Sampling
- Greedy Sampling based on Recurrent First Eigenvector Computation
- Fast Sampling via Iterative Method
- Experimental Results

Matrix Completion

- Fill in missing entries in a matrix:
  (Low-rank matrix recovery problem)

$$\min_{X\in R^{m\times n}} \operatorname{rank}(X)$$

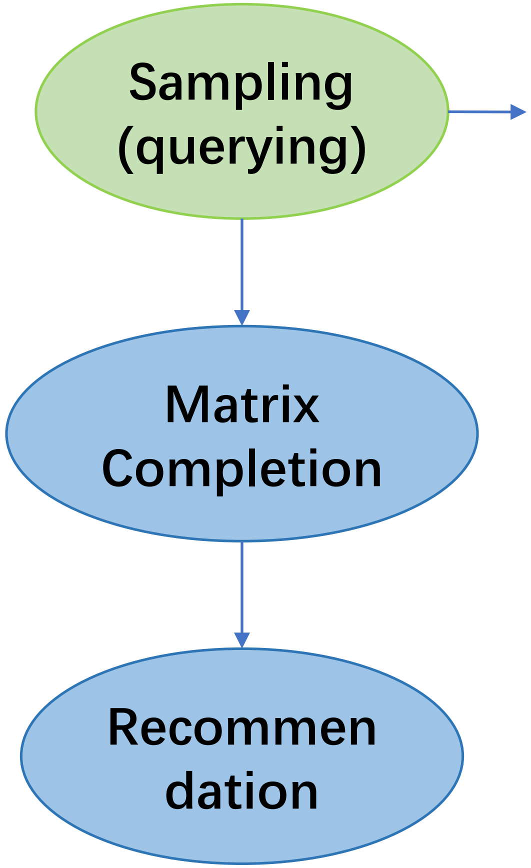
$$s.t. \quad X_{i,j} = M_{i,j}, \quad \forall i, j \in S$$



- Examples of applications:
  - **Recommendation system**—making rating prediction.
  - **Remote sensing**—infer full covariance matrix from partial correlations.
  - **Structure-from-motion** in computer vision.

[1] E. Candes and Y. Plan, "**Matrix completion with noise**," Proceedings of the IEEE, vol. 98, no.6, April 2010, pp. 925–936.

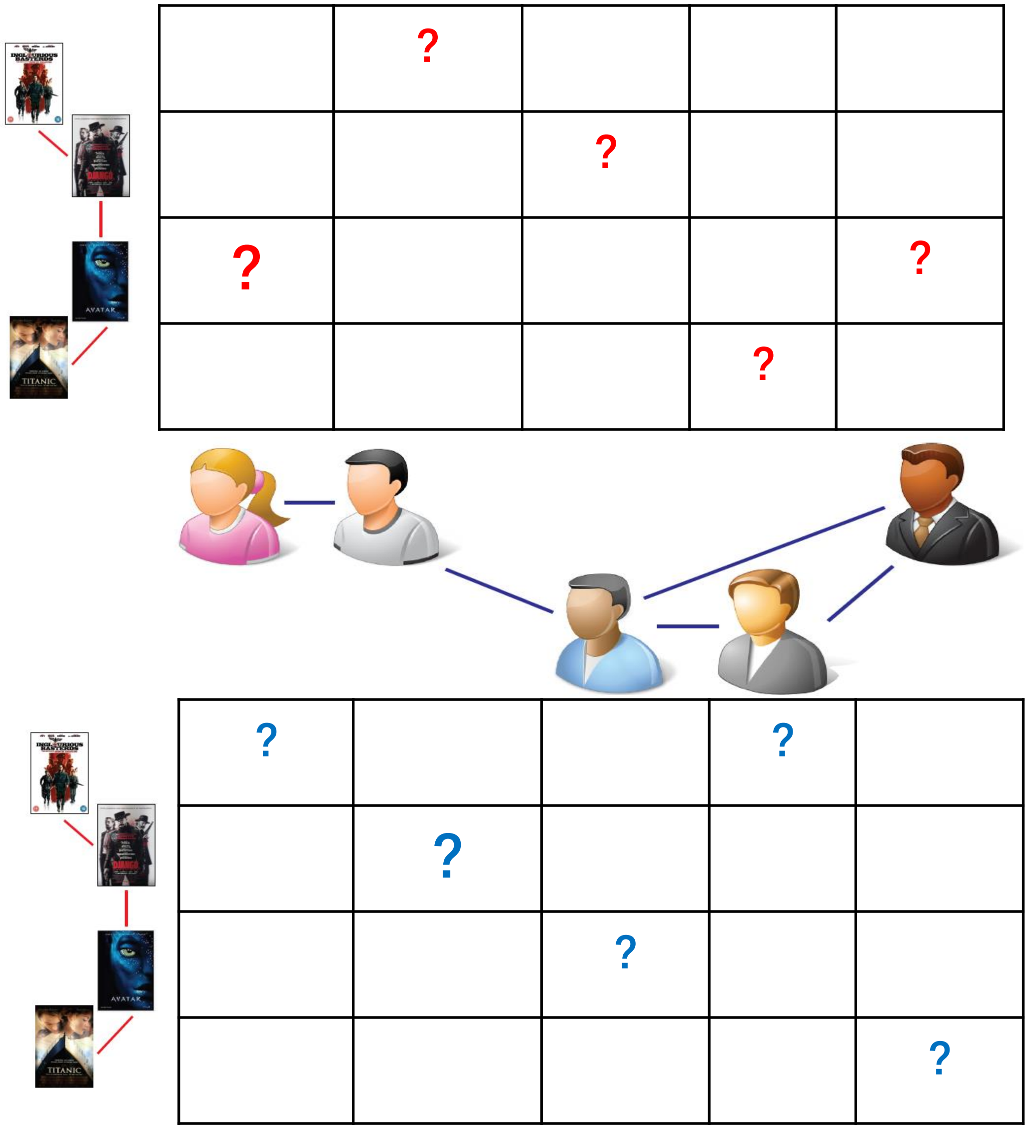
Sampling for Matrix Completion



Which entries are more **informative**, such that the unknown scores could be **recovered better** (smaller MSE w.r.t ground truth)?

Steps

1. Choose one completion method
2. Sampling formulation to minimize the completion error
3. Design sampling algorithms



Matrix Completion with Dual Graph Smoothness Prior

- Use dual graph smoothness prior to promote low rank [1]:

$$\min_{\mathbf{X}\in R^{m\times n}} f(\mathbf{X}) = \frac{1}{2}\left\|\mathbf{A}_\Omega \circ (\mathbf{X}-\mathbf{Y})\right\|_F^2 + \frac{\alpha}{2}\underline{\mathrm{Tr}\left(\mathbf{X}^{\mathrm{T}}\mathbf{L}_r\mathbf{X}\right)} + \frac{\beta}{2}\underline{\mathrm{Tr}\left(\mathbf{X}\mathbf{L}_c\mathbf{X}^{\mathrm{T}}\right)}$$

Fidelity term

$$\sum_{i=1}^{n}\mathbf{x}_i^\top \mathbf{L}_r \mathbf{x}_i \qquad \sum_{j=1}^{m}\bar{\mathbf{x}}_j^\top \mathbf{L}_c \bar{\mathbf{x}}_j$$

- Unconstrained convex objective, fast solvable via ADMM, conjugate gradient; make use of side information (better completion)

◆ Optimal solution

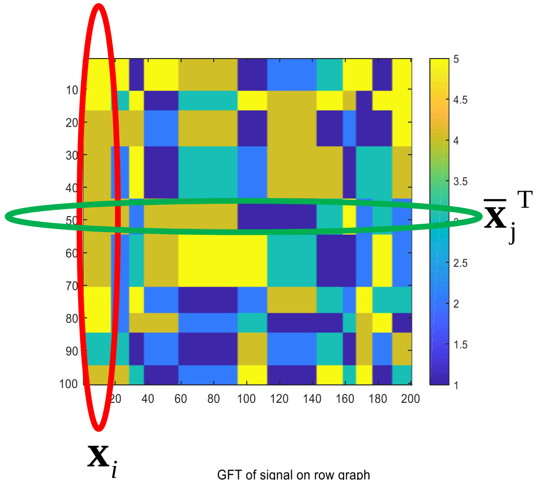
Coefficient matrix $\mathbf{Q}$

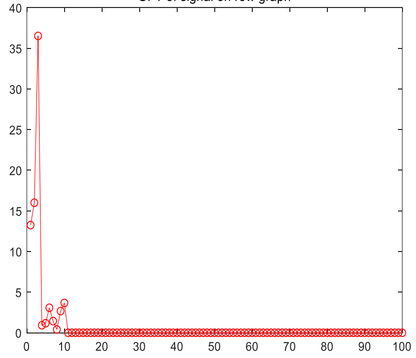
$$\left(\tilde{\mathbf{A}}_\Omega + \alpha \mathbf{I}_n \otimes \mathbf{L}_r + \beta \mathbf{L}_c \otimes \mathbf{I}_m\right)\mathrm{vec}(\mathbf{X}^*) = \mathrm{vec}(\mathbf{Y})$$

Kronecker product

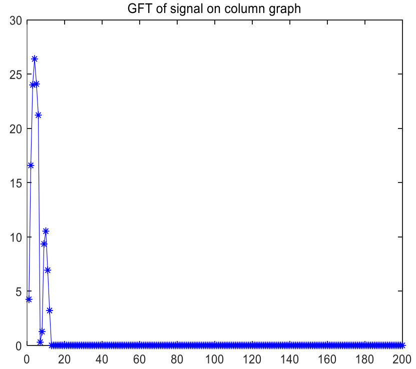
$$\tilde{\mathbf{A}}_\Omega = \mathrm{diag}\left(\mathrm{vec}\left(\mathbf{A}_\Omega\right)\right) \in R^{mn\times mn}$$

To promote **stable** computation in this linear equation, we should maximize $\lambda_{\min}(\boldsymbol{Q})$





Graph Fourier transform on $\mathbf{L}_r$



Graph Fourier transform on $\mathbf{L}_c$

[1] V. Kalofolias, X. Bresson, M. Bronstein, and P. Vandergheynst, "**Matrix completion on graphs**,"arXiv preprint arXiv:1408.1717, 2014.

Dual Graph Smoothness Prior based Sampling

- Sampling formulation

$$\max_{\Omega} \; g(\Omega) = \lambda_{\min}\left(\tilde{\mathbf{A}}_{\Omega} + \alpha \mathbf{I}_n \otimes \mathbf{L}_r + \beta \mathbf{L}_c \otimes \mathbf{I}_m\right)$$

- Lemma 1. MSE of the reconstructed signal w.r.t the original signal is upper-bounded by

$$\|vec(\mathbf{X}^*) - vec(\mathbf{X})\|_2 \leq \frac{\rho}{\lambda_{\min}(\mathbf{Q})} + vec(\mathbf{N})$$

$$where\ \rho = \|\left(\alpha \mathbf{I}_n \otimes \mathbf{L}_r + \beta \mathbf{L}_c \otimes \mathbf{I}_m\right)[vec(\mathbf{X} + \mathbf{N})]\|_2.$$

Maximizing $\lambda_{min}(\boldsymbol{Q})$ turns out minimizing the upper-bound of MSE value

Greedy Sampling based on Recurrent First Eigenvector Computation

$\Omega=\{(1,2);(2,1);(3,2)\}$; $K=3$

0	1
1	0
0	1

Matrix $\mathbf{A}_\Omega \in R^{m\times n}$

- Denote $\mathbf{L} = \alpha \mathbf{I}_n \otimes \mathbf{L}_r + \beta \mathbf{L}_c \otimes \mathbf{I}_m$

Sampling budget

$$\mathbf{Q} = \mathbf{L} + \tilde{\mathbf{A}}_\Omega = \mathbf{L} + \sum_{t=1}^{K} \mathbf{e}_{k_t} \mathbf{e}_{k_t}^\top$$

diagonal structure !!

0					
	1				
		0			
			1		
				0	
					1

Matrix $\tilde{\mathbf{A}}_\Omega = \mathrm{diag}(\mathrm{vec}(\mathbf{A}_\Omega)) \in R^{mn\times mn}$

$S=\{2,4,6\}$; $k_t \in S$

- Combinatorial → Greedy scheme

$$k_t^* = \underset{k_t \in \mathcal{S}_{t-1}^c}{\mathrm{argmax}} \ \lambda_{\min}(\mathbf{L}_{t-1} + \mathbf{e}_{k_t} \mathbf{e}_{k_t}^\top)$$

where $t \in \{1, \ldots, K\}$, $\mathcal{S}_t = \mathcal{S}_{t-1} \cup k_t^*$ with $\mathcal{S}_0 = \emptyset$, and $\mathbf{L}_t = \mathbf{L}_{t-1} + \mathbf{e}_{k_t^*} \mathbf{e}_{k_t^*}^\top$ with $\mathbf{L}_0 = \mathbf{L}$.

One needs to compute λ_{min} for all candidates for one sample. **Expensive !!!**

→ **Fast sampling method**

Greedy Sampling based on Recurrent First Eigenvector Computation

- Problem formulation $k_t^* = \underset{k_t \in \mathcal{S}_{t-1}^c}{\operatorname{argmax}} \; \lambda_{\min}(\mathbf{L}_{t-1} + \mathbf{e}_{k_t}\mathbf{e}_{k_t}^\top)$ → diag(0,0,⋯,1,0,⋯,0)

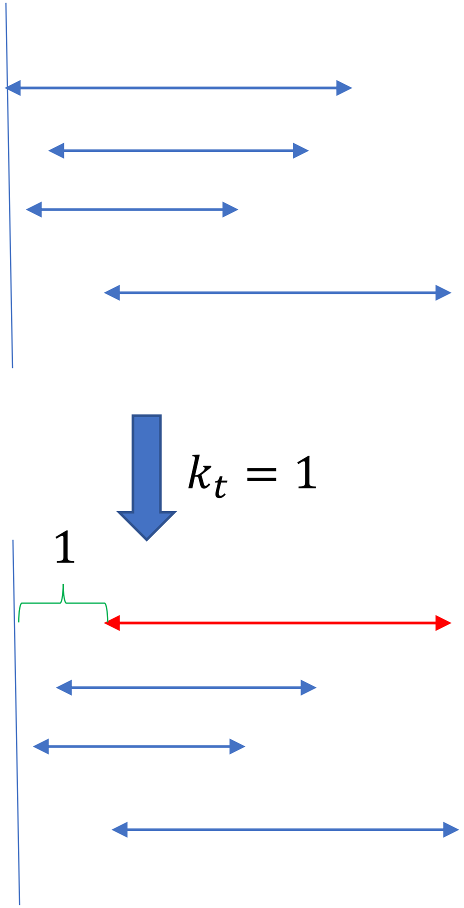
- **Gershgorin circle theorem (GCT) [1]**: $\lambda_{\min}$ Lower-bounded by left-ends of Gershgorin discs:

$$\min_i b_{i,i} - R_i \leq \lambda_{\min}$$

- Sampling node will shift one Gershgorin disc by 1 and keep the others fixed

- Corollary of GCT: $\lambda_{\min}$ must be within the disc whose node's energy is largest one in the first eigenvector

  ↳ Shift j^*-th disc with $j^* = \operatorname{argmin}_j \; |\boldsymbol{\Phi}(j)|$, where $\mathbf{L}_{t-1}\boldsymbol{\Phi} = \lambda_{\min}\,\boldsymbol{\Phi}$



To get a new sample, we just need to compute the first eigenvector once and then select the node with largest energy.

[1] R. A. Horn, R. A. Horn, and C. R. Johnson, "**Matrix analysis",** Cambridge university press, 1990

Greedy Sampling based on Recurrent First Eigenvector Computation

- **Lemma 2**. An optimal solution for the problem

$$\underset{k_t \in \mathcal{S}^c_{t-1}}{\operatorname{argmax}} \lim_{\delta \to 0} \lambda_{\min}(\mathbf{L}_{t-1} + \delta \mathbf{e}_{k_t} \mathbf{e}_{k_t}^\top)$$

is $k_t^* = \underset{k_t \in \mathcal{S}^c_{t-1}}{\operatorname{argmax}} \; |\boldsymbol{\phi}(k_t)|$

Algorithm 1 Proposed GCS sampling

Input: Sample budget K; $\mathbf{L} = \alpha \mathbf{I}_n \otimes \mathbf{L}_r + \beta \mathbf{L}_c \otimes \mathbf{I}_m$; random vector $\mathbf{v}$

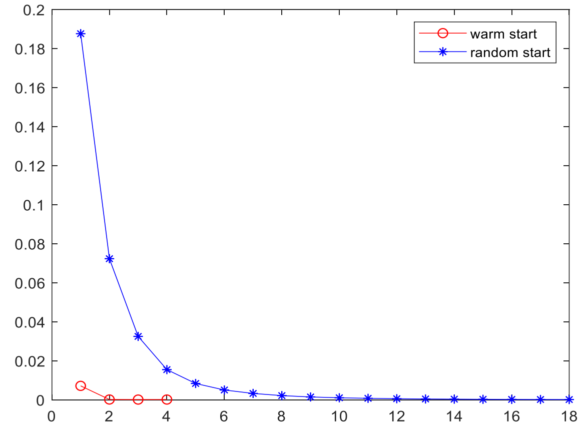
Initialization: $\mathcal{S} = \emptyset$

While $|\mathcal{S}| < K$
  compute the first eigenvector $\boldsymbol{\phi}$ of $\mathbf{L}$ with initial guess $\mathbf{v}$
  $i^* \leftarrow \max_{i \in \mathcal{S}^c} |\boldsymbol{\phi}(i)|$
  $\mathcal{S} \leftarrow \mathcal{S} \cup \{i^*\}$
  Update $\mathbf{L} = \mathbf{L} + \mathbf{e}_{i^*} \mathbf{e}_{i^*}^T$ and $\mathbf{v} = \boldsymbol{\phi}$
end While
return $\mathcal{S}$

- Fast recurrent computation with **warm start**

$$\mathbf{L}_t = \mathbf{L}_{t-1} + \mathbf{e}_{k_t^*} \mathbf{e}_{k_t^*}^\top$$ → **Small change in 1st eigenvector**

we adopt the **LOBPCG [1]** method to compute the first eigenvector, which requires an initial guess. We use $\boldsymbol{\Phi}(\mathbf{L}_{t-1})$ as input to compute $\boldsymbol{\Phi}(\mathbf{L}_t)$.



[1] Andrew V Knyazev. "**Toward the optimal preconditioned eigensolver: Locally optimal block preconditioned conjugate gradient method**" , SIAM journal on scientific computing , 23(2):517–541, 2001

Fast Sampling via Iterative Method

- For real-world datasets (movielens 100K, m=943,n=1682), computing the first eigenvector of $\mathbf{L}$ (m*n= 1586126) is too expensive even with LOBPCG.
- Decompose the matrix $\mathbf{Q}$ as follow with **split** parameter $0 < q < 1$

$$\mathbf{Q} = \left(q\tilde{\mathbf{A}} + \alpha \mathbf{I}_n \otimes \mathbf{L}_r\right) + \left((1-q)\tilde{\mathbf{A}} + \beta \mathbf{L}_c \otimes \mathbf{I}_m\right) \triangleq \mathbf{Q}_1 + \mathbf{Q}_2$$

Block diagonal

Permuted block diagonal

$$\mathbf{P}(\mathbf{L}_c \otimes \mathbf{I}_m)\mathbf{P}^\top = \mathbf{I}_m \otimes \mathbf{L}_c$$

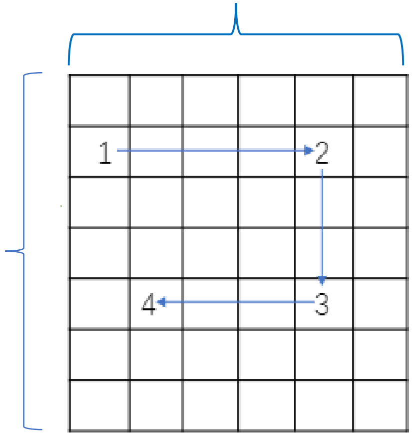
Iterative sampling design:

Sampling one node will help $\mathbf{Q_1}$ and $\mathbf{Q_2}$. We propose to iteratively sample in one cluster (block) in $\mathbf{Q_1}$ and then switch to one group (block) in $\mathbf{Q_2}$

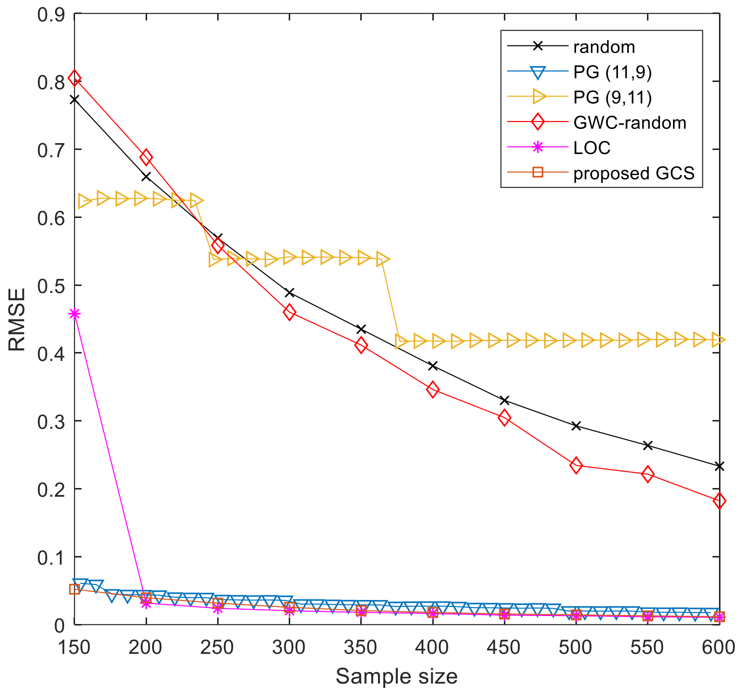
With iterative sampling, the complexity is reduced by at least a factor min{m, n}

n clusters in $\mathbf{Q_1}$

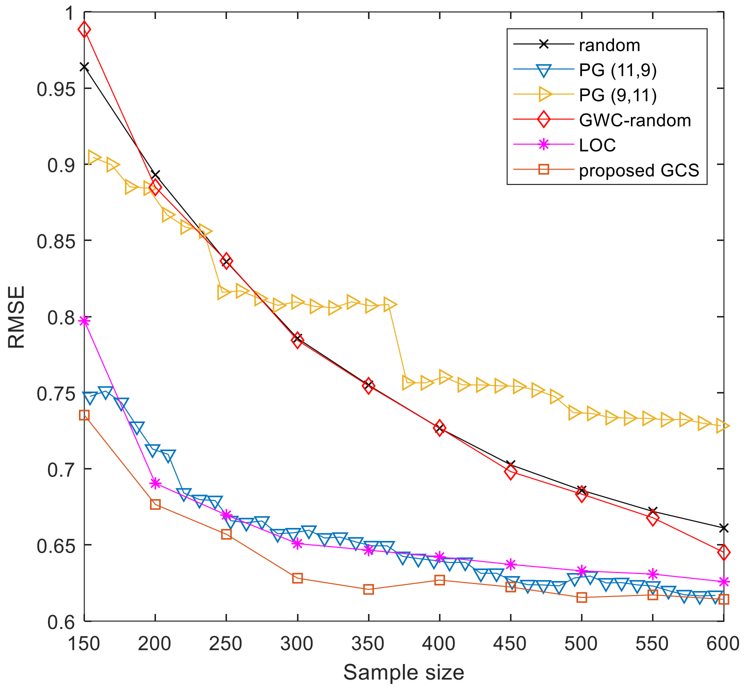
m groups in $\mathbf{Q_2}$



Experimental results

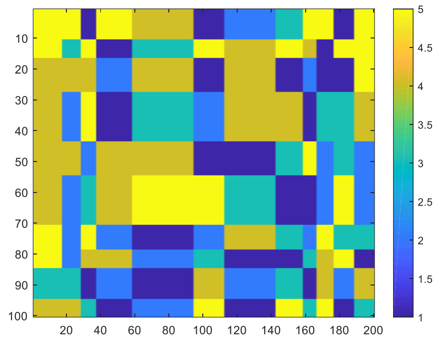


(1) Noiseless

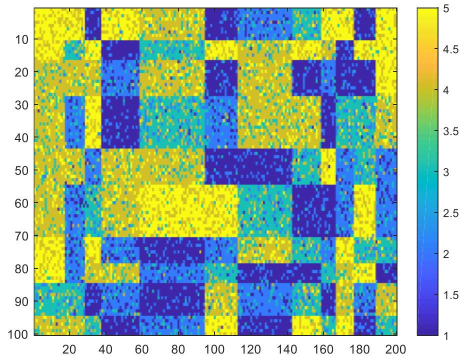


(2) Noisy

Figure: Reconstruction MSE of different sampling methods on synthetic dataset. The reconstruction method for matrix completion is dual graph smoothness based method.



Noiseless synthetic rating matrix



Noisy synthetic rating matrix

Comparison methods: PG [1]; GWC-random [2]; LOC [3]

- [1] Guillermo Ortiz-Jiménez, Mario Coutino, Sundeep Prabhakar Chepuri, and Geert Leus. "**Sampling and reconstruction of signals on product graphs**". arXiv preprint arXiv:1807.00145, 2018.
- [2] G. Puy, N. Tremblay, R. Gribonval, and P. Vandergheynst, "**Random sampling of bandlimited signals on graphs**," Applied and Computational Harmonic Analysis, vol. 44, no. 2, pp. 446–475, 2018.
- [3] A. Sakiyama, Y. Tanaka, T. Tanaka, and A. Ortega, "**Eigendecomposition-free sampling set selection for graph signals**,"IEEE Transactions on Signal Processing, 2019.

Experimental results

Table 2: RMSE for ML100K using random / IGCS sampling combined with different MC methods. Graph-based strategies are marked with ✓.

MC methods	$\mathcal{G}$?	G1	G2
IMC [36]	-	1.590 | **1.507**	1.590 | 1.600
SVT [8]	-	1.021 | 1.031	1.021 | **0.983**
GRALS [31]	✓	0.947 | 0.931	0.945 | **0.893**
GMC [17]	✓	**1.036** | 1.037	1.118 | 1.054
GC-MC [7]	✓	0.898 | 0.891	0.899 | **0.858**
NMC [24]	-	0.892 | 0.887	0.892 | **0.861**

Table 3: GRALS RMSE for different datasets using IGCS sampling with different ζ's.

dataset	random	$\zeta = 1$	$\zeta = 3$	$\zeta = 5$	$\zeta = 7$
Flixster	1.029	**0.932**	1.057	1.046	1.045
Douban	0.744	**0.715**	0.720	0.736	0.730
YahooMusic	96.987	59.172	**44.546**	52.391	47.082
ML1M	0.905	**0.829**	0.833	0.835	0.838
Book-Crossing	3.987	**3.578**	3.704	3.804	4.185
ML10M	0.706	**0.655**	0.656	0.656	0.656
Jester	0.214	**0.160**	0.162	0.162	0.165
FilmTrust	0.820	**0.668**	0.735	0.711	0.742

Intelligent sampling will improve the completion performance, in different datasets/ completions /graphs

*Some datasets information: ML100K (943*1682); ML1M (6040*3706); Douban/Yahoomusic (3000*3000)*

[31] Nikhil Rao, Hsiang-Fu Yu, Pradeep K Ravikumar, and Inderjit S Dhillon, "**Collaborative filtering with graph information: Consistency and scalable methods**", In Advances in Neural Information Processing Systems, pages 2107–2115. 2015.
[7] Rianne van den Berg, Thomas N Kipf, andMaxWelling," **Graph convolutional matrix completion**", arXiv preprint arXiv:1706.02263, 2017.
[24] D. M. Nguyen, E. Tsiligianni, and N. Deligiannis, "**Extendable neural matrix completion**" , In IEEE ICASSP, pages 6328–6332, April 2018.

Experimental results

Table 4: RMSE and IGCS sampling time on ML100K.

graph	MC methods	random	$\zeta = 1$	$\zeta = 3$	$\zeta = 5$	$\zeta = 7$
G1	GRALS	0.947	0.927	0.935	0.934	0.931
	GC-MC	0.898	0.889	0.895	0.897	0.891
	NMC	0.892	**0.880**	0.888	0.889	0.886
	Time (10^3s)	-	1.104	0.503	0.375	0.320
G2	GRALS	0.945	0.871	0.870	0.882	0.882
	GC-MC	0.899	**0.839**	0.840	0.847	0.851
	NMC	0.892	0.840	0.845	0.843	0.852
	Time (10^3s)	-	1.216	0.573	0.441	0.388

G1: side-information (feature attributes) constructed graphs

G2: collected samples constructed graphs

Execution time is decreased when the warm start parameter ζ becomes larger.

Best performance is achieved when the warm start parameter ζ is small.

[1] F. Wang, C. Yang, G. Cheung, and Y. Wang, "**Graph Sampling for Matrix Completion Using Recurrent First Eigenvector Computation**," arXiv preprint arXiv:1906.01087, 2019.

Conclusion

- Matrix completion is an overwhelming topic in data science field, but sampling for matrix completion is neglected.
- Existed sampling methods are not applicable to large real-world datasets
- Our proposed graph-based iterative sampling strategy is fast and essentially improves the completion performance
- The proposed strategy is generally useful for different completions/graphs/datasets

Future work:

- Fast sampling without explicit first eigenvector computation
- Online recommendation system

Thanks!!

Questions ?!

Contact: fenwang@stu.xidian.edu.cn