Imputation of Single Cell RNA Sequential Data using Graph Regularized Matrix Completion Model

Harish Fulara[1], Kushagra Mahajan[1], and Gursimran Singh[1]

IIIT-Delhi, Delhi-110020, India
{harish14143,kushagra14055,gursimran14041}@iiitd.ac.in

Abstract. Matrix completion deals with the problem of finding the missing values of a matrix given some of its entries. Several important real-world problems can be projected as a matrix completion problem, including remote sensing, system identification, recommender systems, gene imputation, etc. Although the problem under the standard low rank assumption is NP-hard, Cand'es and Recht showed that it can be exactly relaxed if the number of observed entries is sufficiently large. In this work, we use graph regularized matrix completion model proposed by Kalofolias et. al [1] to impute missing values in gene-expression data. The graph regularized matrix completion model uses proximity information between the rows and columns of the matrix assuming they form communities. In this work, we have shown that graph regularized matrix completion model outperforms the standard nuclear-norm minimization based matrix completion model on most of the real world single cell RNA sequential datasets using several standard evaluation metrics such as clustering of cell populations, cell type separability, cell visualization, etc. The code for our implementation is available at: https://github.com/HarishFulara07/MCGraphs-scRNAseq

Keywords: Graph regularization · Matrix completion · Single Cell RNA

1 Introduction

1.1 Background

Single cell RNA sequence has fueled discovery and innovation in medicine over the past few years and is useful for studying cellular responses at individual cell resolution. However, insufficient quantities of starting RNA in the individual cells cause significant dropout events, introducing a large number of zero counts in the expression matrix. The measured gene expression matrix can be perceived as a partially observed version of the complete gene expression matrix. Our goal is to recover the complete gene expression matrix from the partially observed gene expression matrix using matrix completion.

Traditional strategies to solve the matrix completion problem include (1) nuclear-norm minimization, (2) matrix factorization and (3) auto-encoders. In

the standard nuclear-norm minimization based matrix completion problem, rows and columns are assumed to be completely unorganized. However, in many real-world problems like the Netflix recommendation problem, there exist relationships between users (such as their age, gender, hobbies, education, etc.) and movies (such as their genre, release year, actors, origin country, etc.). It is an intuitive observation that users form communities of people with similar taste and movies belonging to the same community have similar ratings. This information can be taken advantage of, since people sharing the same taste for a class of movies are likely to rate them similarly. The graph regularized matrix completion model proposed by Kalofolias et. al [1] makes use of graphs to encode relationships between the rows and columns of the matrix to improve upon the standard nuclear-norm minimization based matrix completion model.

In this work, we extend the algorithm of graph regularized matrix completion to sparse gene expression matrix imputation and show that it imputes the single cell RNA sequential data better than the conventional nuclear-norm minimization method.

1.2 Related Work

Mongia et. al. [4] have worked on imputing missing values in gene-expression data using deep matrix factorization techniques. Deep matrix factorization solves the standard matrix factorization problem using deep learning models. Mongia et. al. [5] have also tried to impute dropouts in single cell expression data using standard nuclear-norm based matrix completion model. To the best of our knowledge, no one has attempted to solve the problem of imputing missing gene-expression data using graph regularized matrix completion model.

2 Approach

2.1 Brief Literature Review

The problem of matrix completion is to find the values of a $m \times n$ matrix M given a sparse set Ω of observations $M_{ij} : (i, j) \in \Omega \subseteq \{1, ..., m\} \times \{1, ..., n\}$. Problems of this kind are often encountered in collaborative filtering or recommender system applications, the most famous of which is the Netflix problem, in which one tries to predict the rating that n users (columns of M) would give to m films (rows of M), given only a few ratings provided by each user. A particularly popular model is to assume that the ratings are affected by a few factors, resulting in a low-rank matrix. This leads to the rank minimization problem:

$$\min_{X \in R^{m \times n}} rank(X) \quad \text{s.t.} \quad A_\Omega(X) = A_\Omega(M) \tag{1}$$

where A_Ω is the binary observations mask (1's in place of available values and 0 for missing values) and $A_\Omega(M) = (M_{ij \in \Omega})$ denotes the observed values of M. Problem (1) is NP-hard. However, replacing rank(X) with its convex surrogate

known as the *nuclear* or *trace norm* $||X||_* = tr((XX^T)^{1/2}) = \sum_k \sigma_k$, where σ_k are the singular values of X, one can reformulate the problem (1) as

$$\min_{X \in R^{m \times n}} \gamma_n ||X||_* + ||A_\Omega(X - M)||_F^2 \tag{2}$$

2.2 Overview of Graph Regularized Matrix Completion Approach

In [1], the authors assume that rows/columns of matrix M are given on vertices of graphs. In the Netflix dataset, for e.g., the users are the vertices of a "social graph" whose edges represent, for e.g., friendship or similar tastes relations. Thus, it is reasonable to assume that connected users would give similar movie ratings, i.e. interpreting the ratings as an m-dimensional vector-valued function on the n vertices of the social graph, such a function would be smooth.

Formally, let us be given the un-directed weighted row graph $G_r = (V_r, E_r, W_r)$ with vertices $V_r = \{1, ..., m\}$ and edges $E_r \subseteq V_r \times V_r$ weighted with non-negative weights represented by the $m \times m$ matrix W_r; and respectively the column graph $G_c = (V_c, E_c, W_c)$ defined in the same way. Let $X \in R^{m \times n}$ be a matrix, which we will regard as a collection of m-dimensional column vectors denoted with subscripts $X = (x_1, ..., x_n)$, or of n-dimensional row vectors denoted with superscripts $X = ((x^1)^T, ..., (x^n)^T)^T$. Regarding the columns $x_1, ..., x_n$ as a vector-valued function defined on the vertices V_c, the smoothness assumption implies that $x_j \approx x_i$ if $(j, j') \in E_c$. We want

$$\sum_{j,j'} w^c_{jj'} ||x_j - x_{j'}||_2^2 \quad = \quad tr(XL_cX^T) \quad = \quad ||X||^2_{D,c} \tag{3}$$

to be small, where $D_c = Diag(\sum_{j'=1}^n w^c_{jj'}, L_c = D_c - W_c$ is the Laplacian of the column graph G_c, and $||.||_{D,c}$ is the graph Dirichlet semi-norm for columns. Similarly, for the rows we get a corresponding expression $tr(X^T L_c X) = ||X||^2_{D,r}$ with the Laplacian L_r of the row graph G_r. After adding these smoothness terms as regularization terms to the optimization equation, we get

$$\min_X \gamma_n ||X||_* + \frac{1}{2}||A_\Omega(X - M)||_F^2 + \frac{\gamma_r}{2}||X||^2_{D,r} + \frac{\gamma_c}{2}||X||^2_{D,c} \tag{4}$$

Alternating Direction Method of Multipliers (ADMM) is used for solving the optimization equation. ADMM consists of a splitting step followed by an augmented Lagrangian method to handle the linear equality constraint. We split (4) into two smaller optimization functions $F(X)$ and $G(Y)$, where $F(X) = \gamma_n||X||_*$ and $G(Y) = \frac{1}{2}||A_\Omega(Y-M)||_F^2 + \frac{\gamma_r}{2}||Y||^2_{D,r} + \frac{\gamma_c}{2}||Y||^2_{D,c}$. The optimization equation (4) becomes

$$\min_{X,Y \in R^{m \times n}} F(X) + G(Y) \quad s.t. \quad X = Y \tag{5}$$

The augmented Lagrangian of (5) is $\mathcal{L}(X, Y, Z) = F(X) + G(Y) + <Z, X - Y> + \frac{\rho}{2}||X - Y||_F^2$. ADMM gives the optimal solution using the following iterative scheme

$$X^{k+1} = \operatorname*{argmin}_X \mathcal{L}(X, Y^k, Z^k) \tag{6}$$

$$Y^{k+1} = \underset{Y}{\operatorname{argmin}}\, \mathcal{L}(X^{k+1}, Y, Z^k) \tag{7}$$

$$Z^{k+1} = Z^k + \rho(X^{k+1} - Y^{k+1}) \tag{8}$$

The convergence of the proposed ADMM algorithm (6) – (8) can be achieved using various mathematical approaches.

A closed-form solution exists for the sub-optimization problem (6), i.e.

$$X^{k+1} = U soft_{\frac{\gamma_n}{\rho}}(\wedge) V^T \tag{9}$$

where U, $\wedge$, V are respectively the singular value decomposition of $H = Y^k - \frac{Z^k}{\rho}$, i.e. $H = U \wedge V^T$, and $soft_{\frac{\gamma_n}{\rho}}(\wedge) = max(0, \wedge - \frac{\gamma_n}{\rho})$.

Unlike nuclear norm, there is no closed form solution for the sub-optimization problem (7). The optimal solution is found by solving the following linear equation

$$(\widetilde{A}_\Omega + \gamma_r L_r \otimes I_n + \gamma_c I_m \otimes L_c + \rho I_{mn}) vec(Y) = vec(M + \rho H) \tag{10}$$

where $vec(.)$ is the column-stack vectorization operator, and $\otimes$ is the Kronecker product, and $\widetilde{A}_\Omega = Diag(vec(A_\Omega))$, where A_Ω is observations mask matrix.

2.3 Dataset Description

We used 7 real world single cell RNA sequential datasets.

1. **Blakeley [6]:** This is a scRNA-seq dataset of the three cell lineages of the human blastocyst. It contains expression profiles of 30 cells.
2. **Quake [7]:** This is a scRNA-seq dataset of human adult cortical tissue. It contains expression profiles of 461 cells.
3. **Usoskin [8]:** This is a scRNA-seq dataset of mouse neurons obtained by performing RNA-Seq 215 on 799 dissociated single cells dissected from the mouse lumbar dorsal root 216 ganglion (DRG) distributed over a total of nine 96-well plates. It contains expression profiles of 622 cells.
4. **Kolodziejczyk [9]:** This is a scRNA-seq dataset of embryonic stem cells cultured in three different conditions: serum, 2i, and the alternative ground state a2i. It contains expression profiles of 704 cells.
5. **Zeisel [10]:** This is a scRNA-seq dataset of mouse cortex and hippocampal region. It contains expression profiles of 3005 cells.
6. **Preimplantation [11]:** This is a scRNA-seq dataset of mouse preimplantation embryos. It contains expression profiles of 317 cells.
7. **Jurkat [12]:** This is a scRNA-seq dataset of Jurkat and 293T cells, mixed in vitro at equal proportions. It contains expression profiles of 3388 cells.

2.4 Data Preprocesing

The raw single cell RNA sequential data is processed as follows:

1. **Data Filtering:** It is ensured that data has no bad cells and if a gene is detected with ≥ 3 reads in at least 3 cells, we consider it expressed. We ignore the remaining genes.
2. **Normalization:** Expression matrices are normalized by first dividing each read count by the total counts in each cell, and then by multiplying with the median of the total read counts across cells.
3. **Gene Selection:** The space complexity of the graph regularized matrix completion model is $O(m^2n^2)$ due to the intermediate matrix $(\widetilde{A}_\Omega + \gamma_r L_r \otimes I_n + \gamma_c I_m \otimes L_c + \rho I_{mn})$ required to solve the sub-optimization problem (7). This makes the graph regularized matrix completion model infeasible for large matrices. Thus, we constrain the size of our genes expression matrix by selecting only top 1000 high-dispersion (coefficient of variance) genes for imputation.
4. **Log Normalization:** A copy of the matrices are log_2 transformed following addition of 1 as pseudo-count.
5. **Imputation:** Log transformed gene-expression matrix is inputted to the graph regularized matrix completion model and the standard nuclear-norm based matrix completion model. The imputed matrices from both the models are used for further analysis.

2.5 Parameter Estimation using Grid Search

There are 6 parameters in graph regularized matrix completion model that need to be estimated.

1. $\boldsymbol{W_{r\_thresh}}$ − threshold for the weight matrix of rows graph. The values above the threshold are marked as 1 and rest of the values are marked as 0.
2. $\boldsymbol{W_{c\_thresh}}$ − threshold for the weight matrix of columns graph. The values above the threshold are marked as 1 and rest of the values are marked as 0.
3. $\boldsymbol{\gamma_n}$ − regularization constant for nuclear norm term.
4. $\boldsymbol{\gamma_r}$ − regularization constant for rows graph.
5. $\boldsymbol{\gamma_c}$ − regularization constant for columns graph.
6. $\boldsymbol{\rho}$ − step size of ADMM algorithm.

Table 1 shows the optimal parameter values for each of the real world scRNA-seq dataset obtained after an extensive grid search.

3 Results

We performed numerous experiments to evaluate the efficacy of graph regularized matrix completion approach against the nuclear-norm based matrix completion approach.

Table 1. Optimal parameter values for the real world scRNA-seq datasets obtained using grid search.

Dataset	W_{r_thresh}	W_{r_thresh}	γ_n	γ_r	γ_c	ρ
Blakeley	0.9	0.99	0.01	0.003	0.003	0.009
Quake	0.9	0.99	0.01	0.003	0.003	0.009
Usoskin	0.9	0.99	0.01	0.003	0.003	0.009
Kolodziejczyk	0.93	0.99	5	0.003	0.003	0.01
Zeisel	0.93	0.99	5	0.003	0.003	0.01
Preimplantation	0.93	0.99	5	0.003	0.003	0.01
Jurkat	0.93	0.99	5	0.003	0.003	0.01

3.1 Cell Visualization

The scRNA-seq data is reduced to two dimensions by applying Principal Component Analysis, followed by plotting the data in 2D space. Some of the plots are shown in figures 1 – 12. The plots show the clustering tendency of the classes in the imputed matrices as well as the original data.

3.2 Cell Type Separability

For any two cell groups, we first find the median of Spearman correlation values computed for each possible pair of cells within their respective groups.

Intra-cell type scatter is the average of the median correlation values, while inter-cell type scatter is the median of Spearman correlation values computed for pairs of different group cells. The difference between the intra-cell scatter and inter-cell type scatter is what is known as cell-type separability (CTS) score. Comparisons of CTS scores for the original data, data imputed using the standard nuclear-norm based matrix completion and our approach which uses graph regularization is mentioned in Table. 2. We find that our approach improves the CTS scores for the Jurkat, Zeisel, Kolodziejczyk datasets and gives marginal improvement for the Quake dataset as compared to nuclear-norm based matrix completion approach.

Table 2. Cell Type Separability Scores.

	Jurkat	Zeisel	Usoskin	Blakeley	Kolodziejczyk	Quake
Original Data	0.2284	0.1498	**0.0448**	0.1199	0.0752	**0.0682**
Nuclear Norm based Completion	0.33	0.1217	0.0364	**0.1434**	0.0558	0.0546
Graph Regularized Completion	**0.5156**	**0.2046**	0.0115	0.0461	**0.1487**	0.0605

3.3 Improvement in clustering accuracy

Clustering single cell RNA-seq data for discovering distinct cell types from a heterogeneous cell population is one of the most important applications of scRNA-

seq. But a clustering algorithm that tries to group together similar cell types can get confused by large number of non available values in single cell RNA seq data - this serves as noise. This should get fixed by a good imputation resulting in better delineation of cell types. Hence we use the K-means clustering results on all the log-transformed expression profiles for each dataset both with and without imputation. The initialization parameter K (no of clusters) in K-means algorithm has been set to the number of annotated cell types for every data. Adjusted Rand Index (ARI) was used as the primary performance metric to evaluate the correspondence between the original annotations and K-means assigned clusters. Other metrics such as AMI (Adjusted Mutual Information) and NMI (Normalized Mutual Information) have also been computed for comparison.

Table 3. K-Means performance measures.

Dataset	Metric	Original Data	Nuclear Norm	Graph Regularized
Blakeley	ARI	0.737	0.6701	0.1559
	AMI	0.6715	0.5686	0.2527
	NMI	0.7215	0.6031	0.3775
Quake	ARI	0.6476	0.49	0.3184
	AMI	0.6476	0.5519	0.3788
	NMI	0.6827	0.5742	0.5138
Usoskin	ARI	0.6996	0.2372	0.2513
	AMI	0.6943	0.27	0.2586
	NMI	0.7095	0.281	0.3706
Kolodziejczyk	ARI	0.1499	0.2427	0.4294
	AMI	0.2241	0.3155	0.4554
	NMI	0.229	0.3344	0.4621
Zeisel	ARI	0.6112	0.5794	0.6147
	AMI	0.6928	0.6572	0.6915
	NMI	0.6992	0.6632	0.6949
Preimplantation	ARI	0.5972	0.5106	0.505
	AMI	0.7426	0.704	0.65
	NMI	0.781	0.7498	0.7046
Jurkat	ARI	0.9871	0.9847	0.9847
	AMI	VSMALL	VSMALL	VSMALL
	NMI	0.9717	0.9659	0.9659

3.4 Matrix Recovery

In this experiment, we evaluate the ability of graph regularized matrix completion model to accurately recover the missing values in the gene-expression matrix at different sampling percentage levels as compared to the standard nuclear-norm based matrix completion model. Sampling at different percentage levels mimics dropout in the data that helps us evaluate the robustness of a matrix completion

model. A sampling percentage level of x% means that we sample only x% of the total observed values in the gene-expression matrix and use only those for recovering rest of the values. We use Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) metrics for quantifying the performance evaluation.

Figures 13 – 26 show the variation of MAE/RMSE with percentage sampling level for different datasets for graph regularized matrix completion approach and standard nuclear-norm based matrix completion approach. It is quite evident from these figures that graph regularized model performs better than nuclear-norm based model across all the datasets in terms of both MAE and RMSE.

4 Conclusion

In this work, we attempted to solve the problem of imputing missing gene-expression data using graph regularized matrix completion model and evaluate its performance against standard nuclear-norm based matrix completion model. After performing various evaluation experiments, we believe, in general, that graph regularized matrix completion model gives better imputations for missing values in gene-expression matrix for several real world single cell RNA sequential datasets than standard nuclear-norm based matrix completion model.

The exceeding performance of graph regularized matrix completion model can be attributed to its use of similarity information between rows and columns of a matrix, which is certainly missing in nuclear-norm based matrix completion model. The only major bottleneck of graph regularized matrix completion model is its inability to work efficiently with high dimensional matrices – though this can be improved by finding enhanced ways for solving the sub-optimization problem (7).

5 Acknowledgements

We would like to thank Dr. Angshul Majumdar and Ms. Aanchal Mongia for their guidance, help and support throughout this project.

References

1. Kalofolias, Vassilis, et al. "Matrix completion on graphs." arXiv preprint arXiv:1408.1717 (2014).
2. B. Miller, I. Albert, S. Lam, J. Konstan, and J. Riedl. MovieLens unplugged: Experiences with an occasionally connected recommender system. In Proc. Conf. Intelligent User Interfaces, 2003.
3. Talwar, Divyanshu, et al. "AutoImpute: Autoencoder based imputation of single-cell RNA-seq data." Scientific reports 8.1 (2018): 16329.
4. Mongia, Aanchal, Debarka Sengupta, and Angshul Majumdar. "deepMc: deep Matrix Completion for imputation of single cell RNA-seq data" bioRxiv (2018): 387621.
5. Mongia, Aanchal, Debarka Sengupta, and Angshul Majumdar. "McImpute: Matrix completion based imputation for single cell RNA-seq data." bioRxiv (2018): 361980.

6. Blakeley, P. et. al. "Defining the three cell lineages of the human blastocyst by single-cell rna-seq." Dev. 142. 3151–3165 (2015).
7. Darmanis, S. et al. "A survey of human brain transcriptome diversity at the single-cell level." Proc. Natl. Acad. Sci. 112, 7285–7290 (2015).
8. Usoskin, D. et. al. "Unbiased classification of sensory neuron types by large-scale single-cell rna sequencing." Nat. neuroscience 18, 145 (2015).
9. Kolodziejczyk, A. A. et. al. "Single-cell rna-sequencing of pluripotent states unlocks modular transcriptional variation." Cellstem cell17, 471–485 (2015).
10. Zeisel, A. et. al. "Cell types in the mouse cortex and hippocampus revealed by single-cell rna-seq." Sci. 347, 1138–1142 (2015).
11. Yan, L. et. al. "Single-cell rna-seq profiling of human preimplantation embryos and embryonic stem cells." Nat. structural and molecular biology 20, 1131–1139 (2013).
12. Zheng, G. X. et. al. "Massively parallel digital transcriptional profiling of single-cells." Nat. communications 8, 14049 (2017).

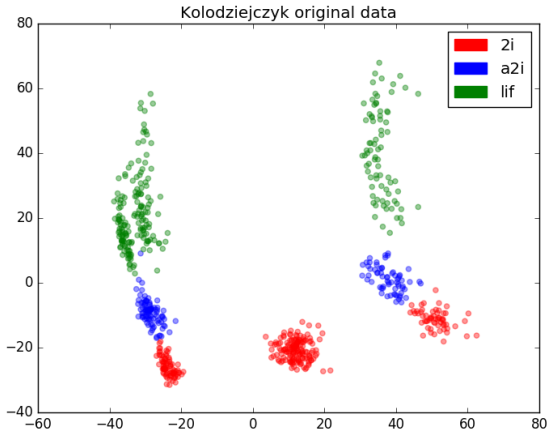

Fig. 1. Cell visualization for original Kolodziejczyk dataset.

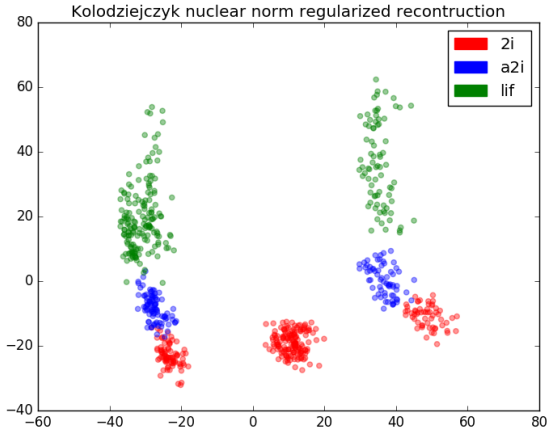

Fig. 2. Cell visualization for Kolodziejczyk data imputed using nuclear-norm based matrix completion model.

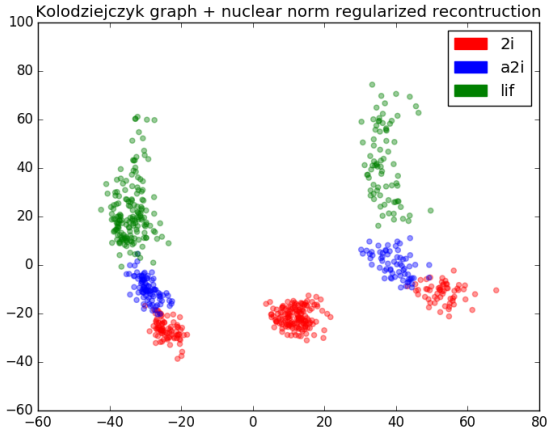

Fig. 3. Cell visualization for Kolodziejczyk data imputed using graph regularized matrix completion model.

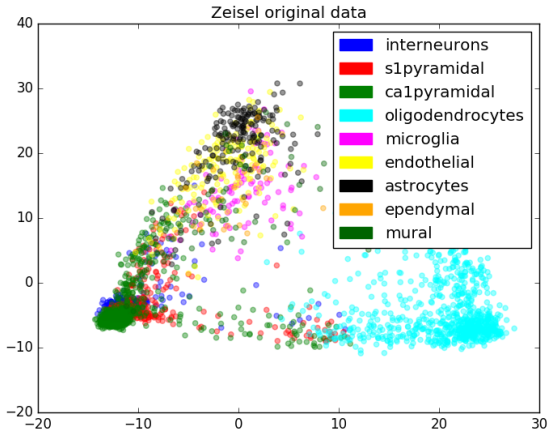

Fig. 4. Cell visualization for original Zeisel dataset.

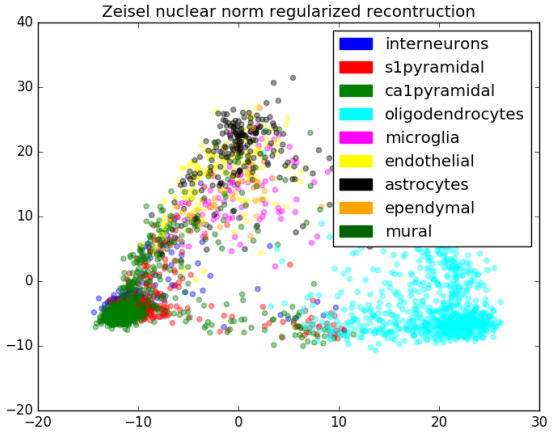

Fig. 5. Cell visualization for Zeisel data imputed using nuclear-norm based matrix completion model.

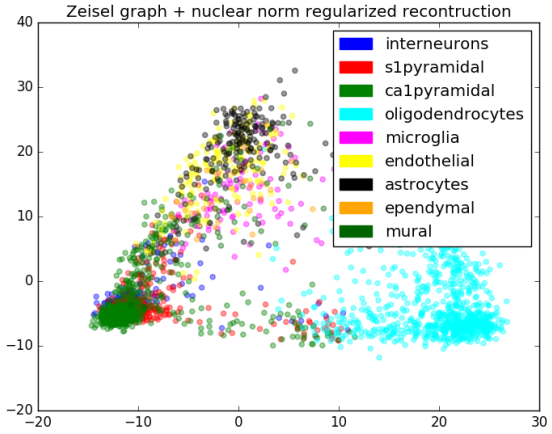

Fig. 6. Cell visualization for Zeisel data imputed using graph regularized matrix completion model.

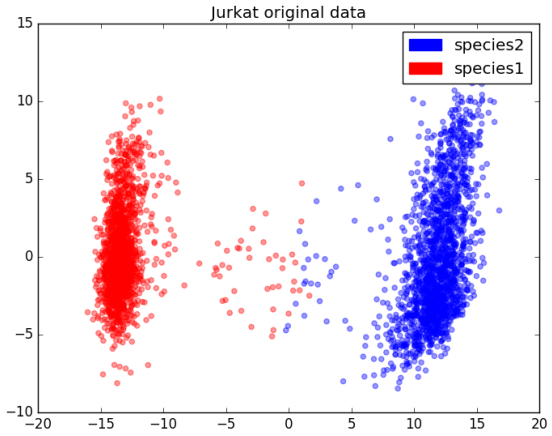

Fig. 7. Cell visualization for original Jurkat dataset.

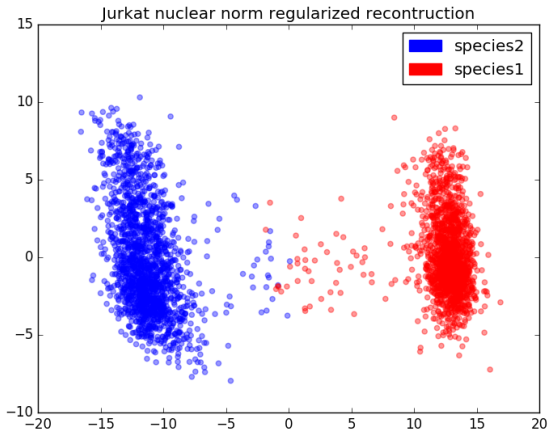

Fig. 8. Cell visualization for Jurkat data imputed using nuclear-norm based matrix completion model.

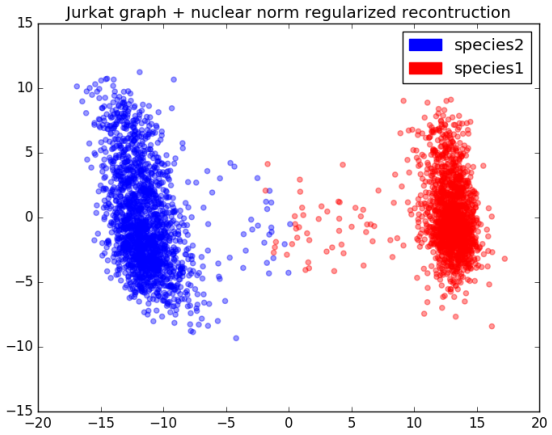

Fig. 9. Cell visualization for Jurkat data imputed using graph regularized matrix completion model.

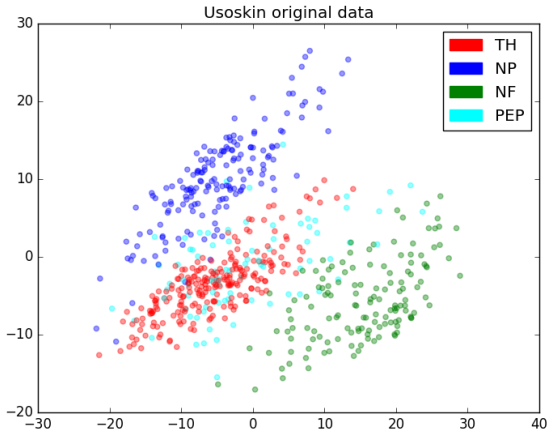

Fig. 10. Cell visualization for original Usoskin dataset.

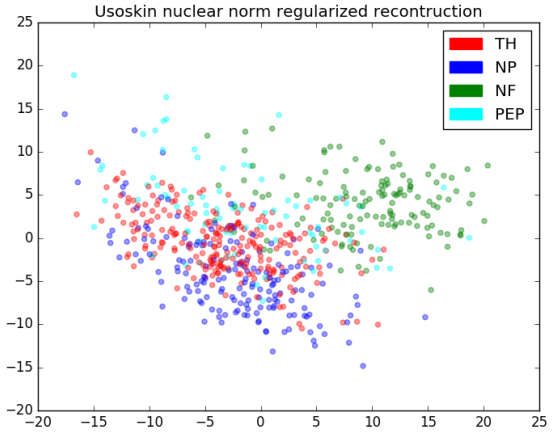

Fig. 11. Cell visualization for Usoskin data imputed using nuclear-norm based matrix completion model.

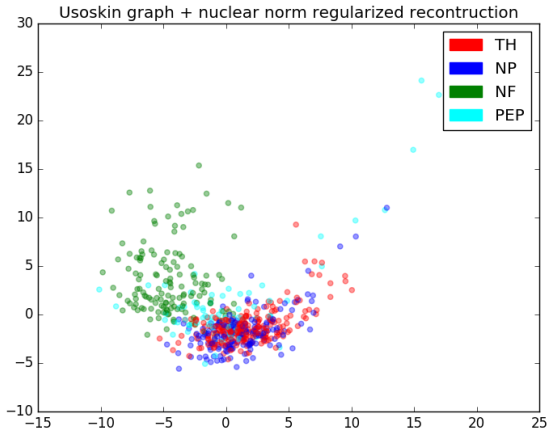

Fig. 12. Cell visualization for Usoskin data imputed using graph regularized matrix completion model.

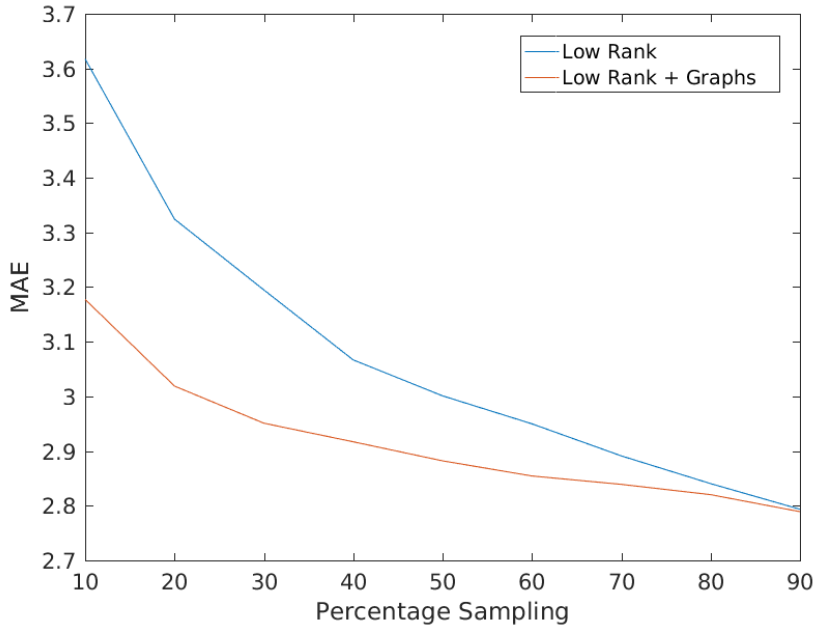

Fig. 13. Variation of MAE with percentage sampling level for Blakeley dataset for (a) Low Rank (standard nuclear-norm based matrix completion model), and (b) Low Rank + Graphs (graph regularized matrix completion model) matrix completion approaches

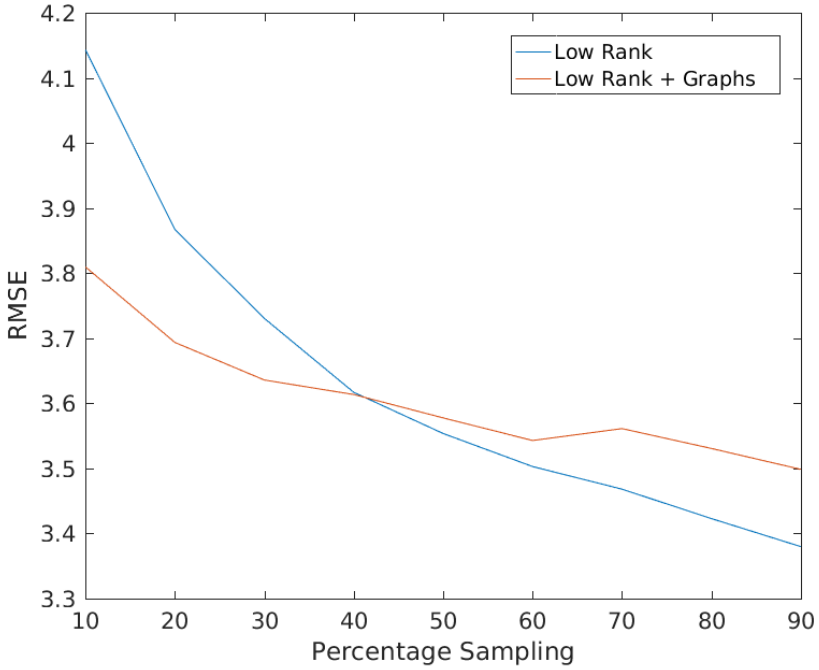

Fig. 14. Variation of RMSE with percentage sampling level for Blakeley dataset for (a) Low Rank (standard nuclear-norm based matrix completion model), and (b) Low Rank + Graphs (graph regularized matrix completion model) matrix completion approaches

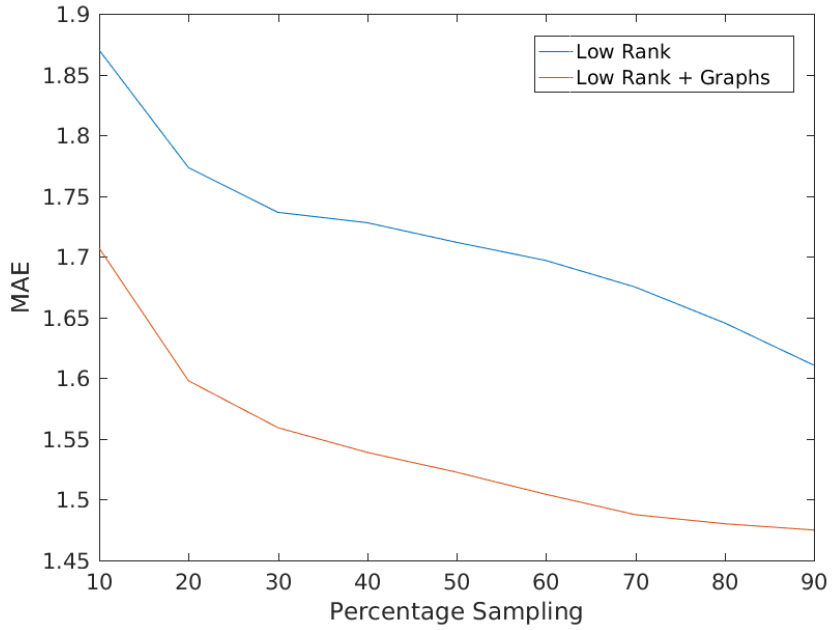

Fig. 15. Variation of MAE with percentage sampling level for Quake dataset for (a) Low Rank (standard nuclear-norm based matrix completion model), and (b) Low Rank + Graphs (graph regularized matrix completion model) matrix completion approaches

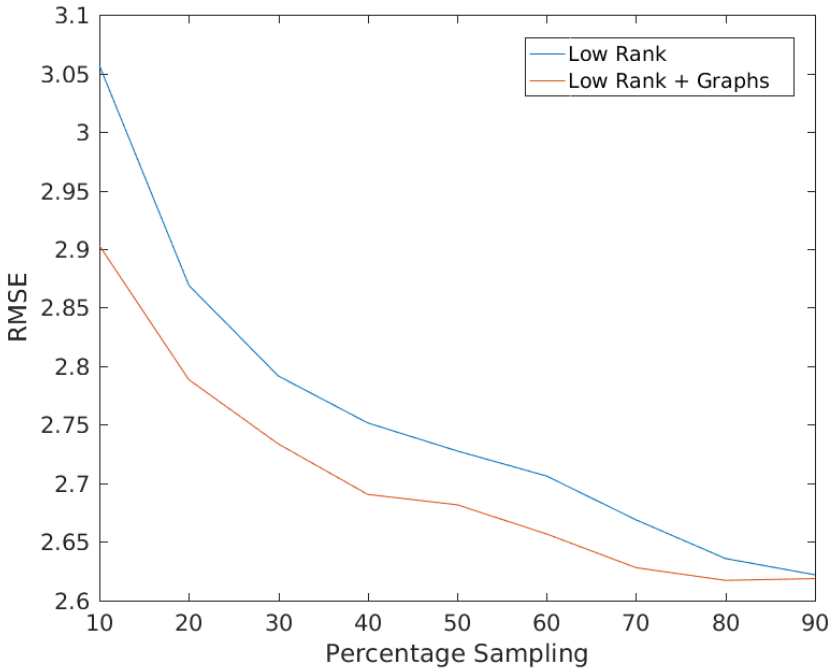

Fig. 16. Variation of RMSE with percentage sampling level for Quake dataset for (a) Low Rank (standard nuclear-norm based matrix completion model), and (b) Low Rank + Graphs (graph regularized matrix completion model) matrix completion approaches

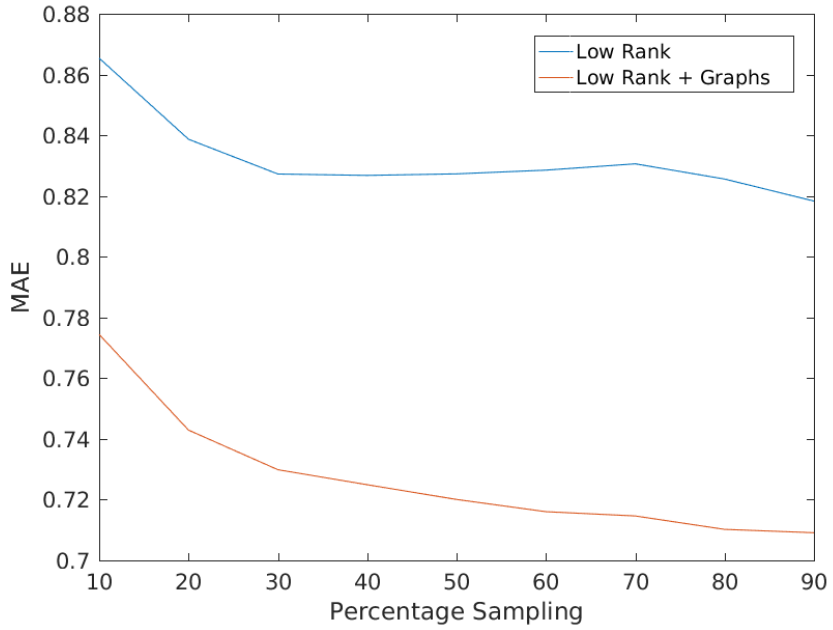

Fig. 17. Variation of MAE with percentage sampling level for Usoskin dataset for (a) Low Rank (standard nuclear-norm based matrix completion model), and (b) Low Rank + Graphs (graph regularized matrix completion model) matrix completion approaches

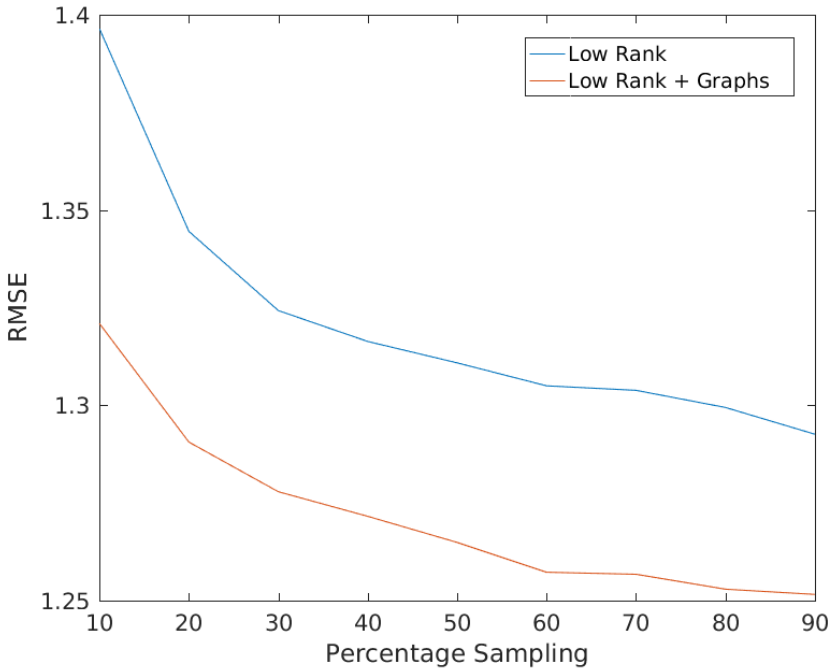

Fig. 18. Variation of RMSE with percentage sampling level for Usoskin dataset for (a) Low Rank (standard nuclear-norm based matrix completion model), and (b) Low Rank + Graphs (graph regularized matrix completion model) matrix completion approaches

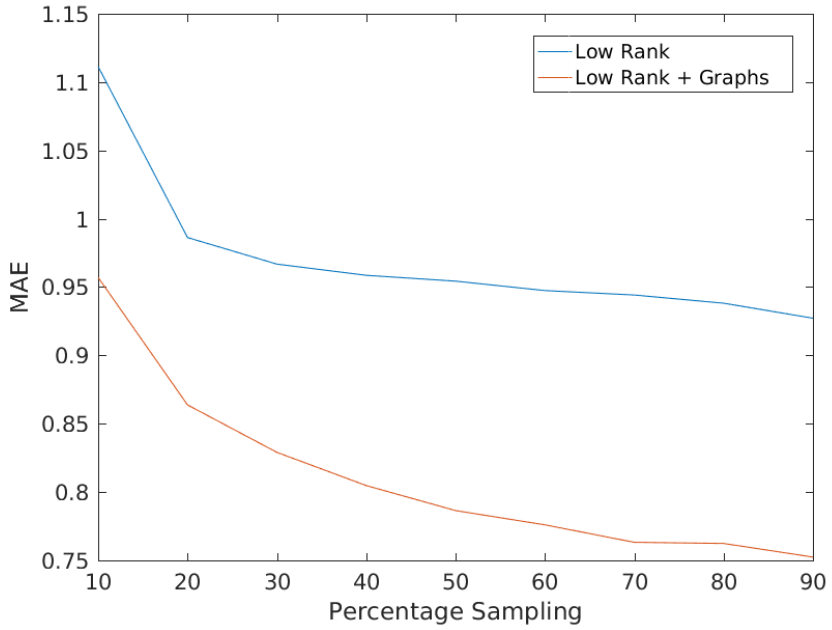

Fig. 19. Variation of MAE with percentage sampling level for Kolodziejczyk dataset for (a) Low Rank (standard nuclear-norm based matrix completion model), and (b) Low Rank + Graphs (graph regularized matrix completion model) matrix completion approaches

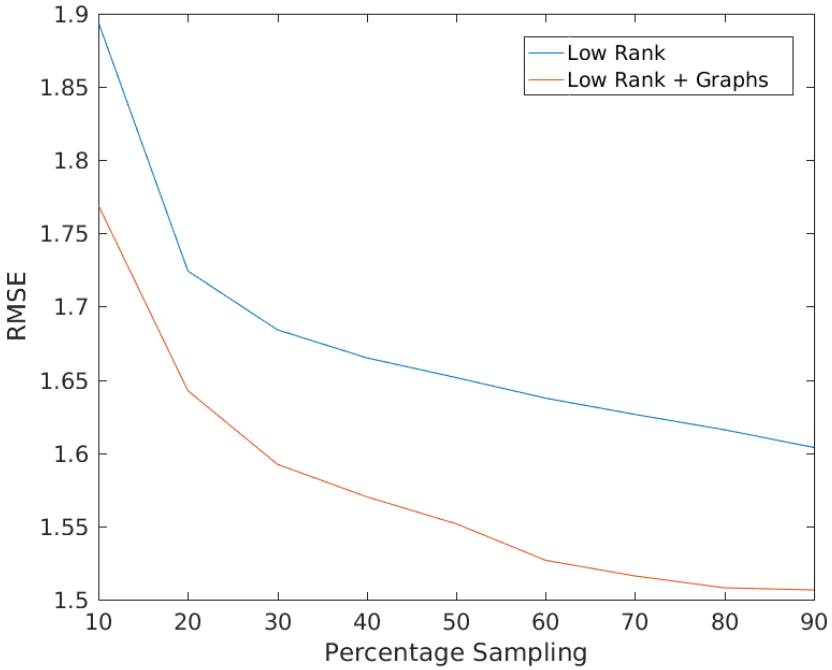

Fig. 20. Variation of RMSE with percentage sampling level for Kolodziejczyk dataset for (a) Low Rank (standard nuclear-norm based matrix completion model), and (b) Low Rank + Graphs (graph regularized matrix completion model) matrix completion approaches

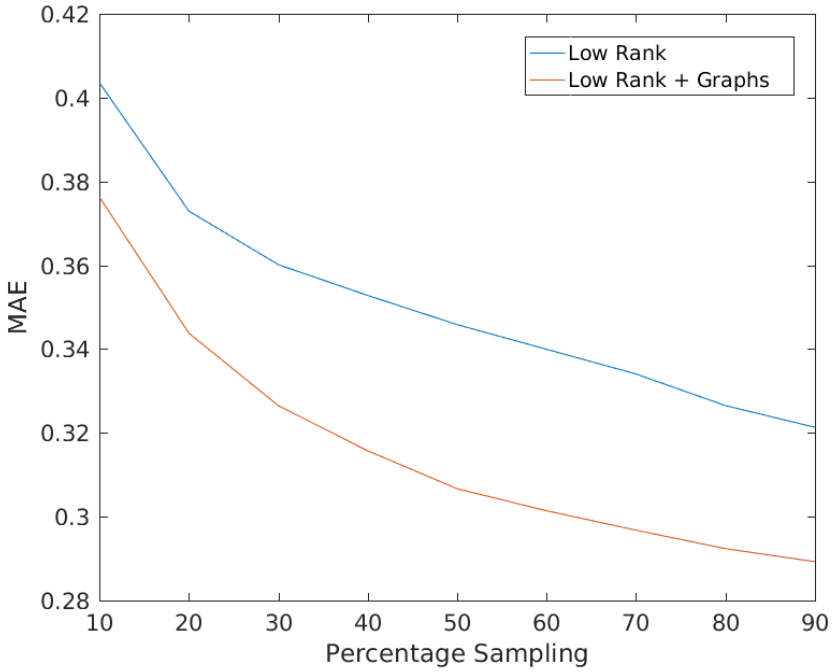

Fig. 21. Variation of MAE with percentage sampling level for Zeisel dataset for (a) Low Rank (standard nuclear-norm based matrix completion model), and (b) Low Rank + Graphs (graph regularized matrix completion model) matrix completion approaches

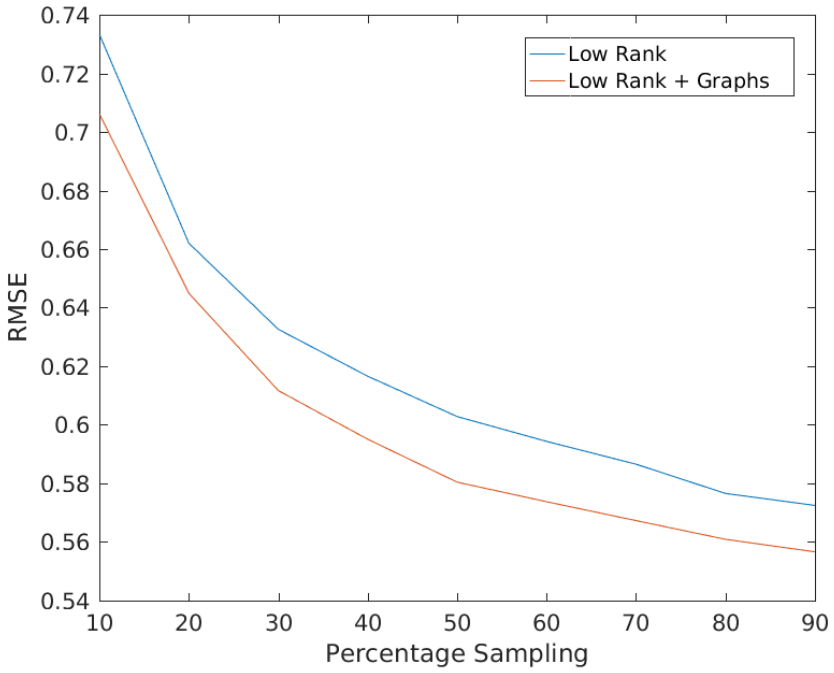

Fig. 22. Variation of RMSE with percentage sampling level for Zeisel dataset for (a) Low Rank (standard nuclear-norm based matrix completion model), and (b) Low Rank + Graphs (graph regularized matrix completion model) matrix completion approaches

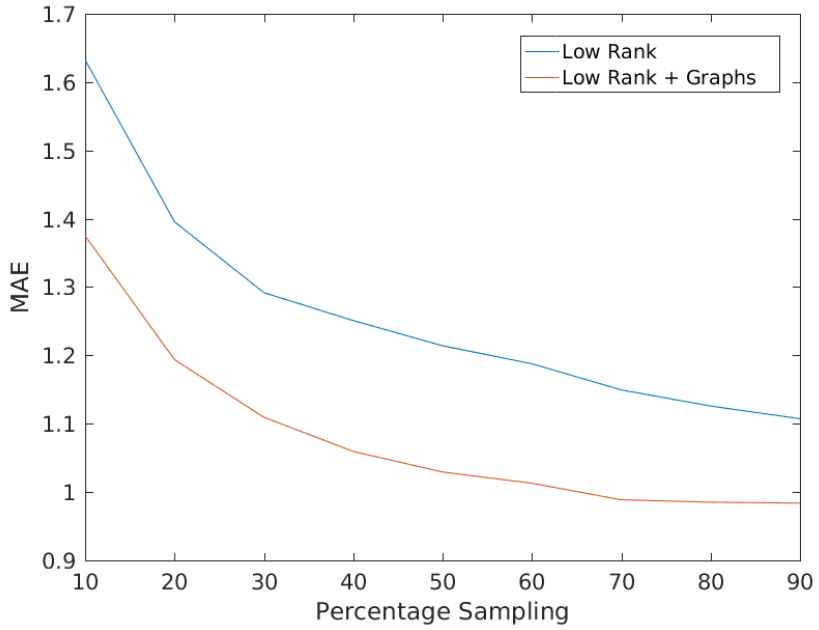

Fig. 23. Variation of MAE with percentage sampling level for Preimplantation dataset for (a) Low Rank (standard nuclear-norm based matrix completion model), and (b) Low Rank + Graphs (graph regularized matrix completion model) matrix completion approaches

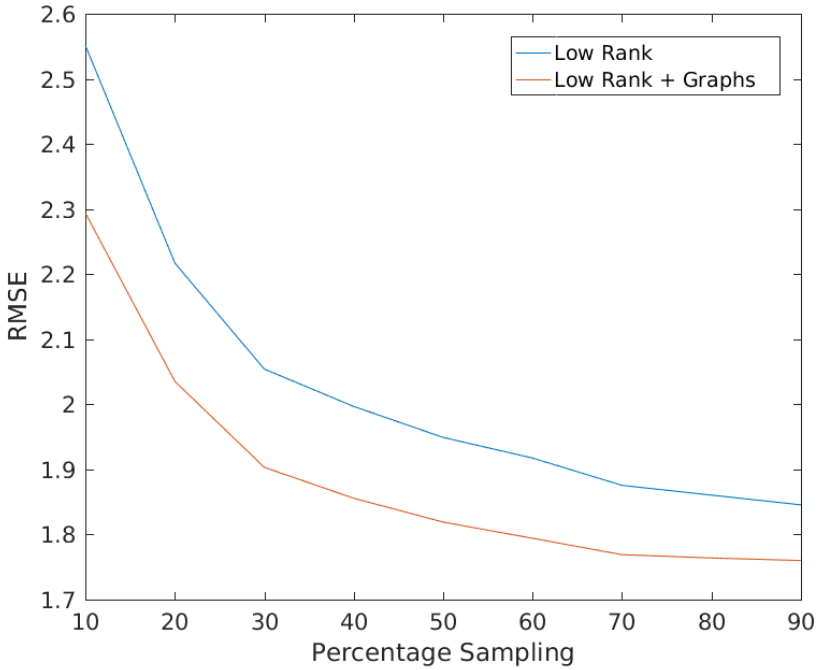

Fig. 24. Variation of RMSE with percentage sampling level for Preimplantation dataset for (a) Low Rank (standard nuclear-norm based matrix completion model), and (b) Low Rank + Graphs (graph regularized matrix completion model) matrix completion approaches

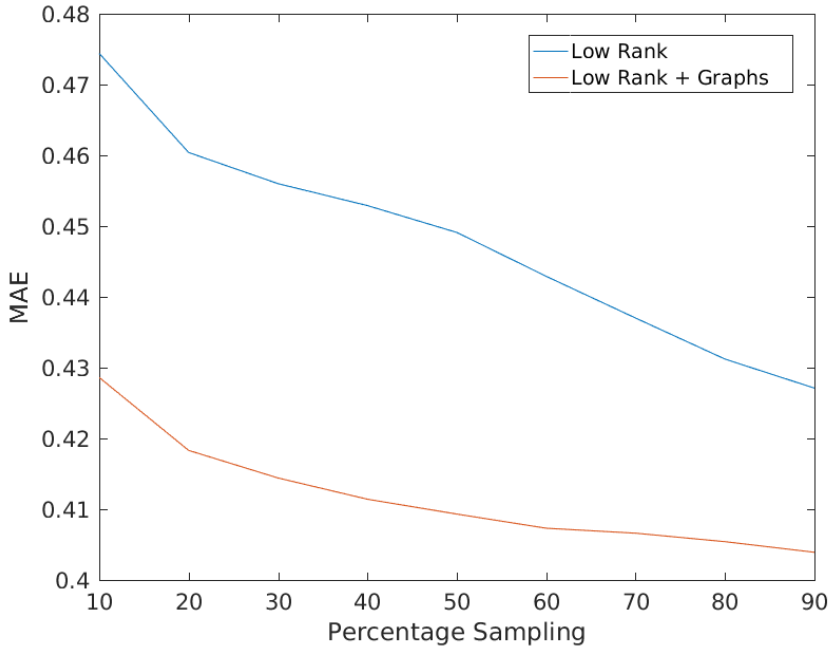

Fig. 25. Variation of MAE with percentage sampling level for Jurkat dataset for (a) Low Rank (standard nuclear-norm based matrix completion model), and (b) Low Rank + Graphs (graph regularized matrix completion model) matrix completion approaches

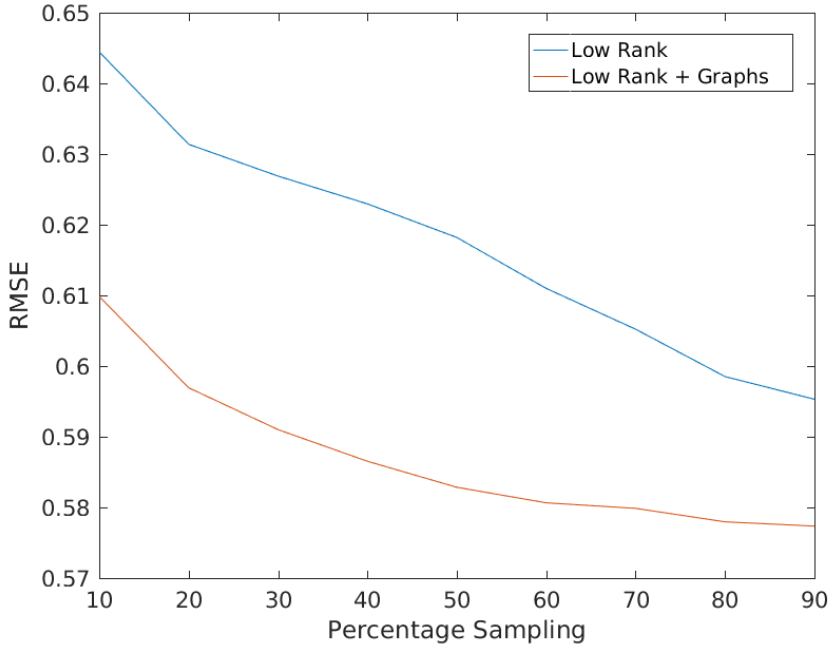

Fig. 26. Variation of RMSE with percentage sampling level for Jurkat dataset for (a) Low Rank (standard nuclear-norm based matrix completion model), and (b) Low Rank + Graphs (graph regularized matrix completion model) matrix completion approaches