Appendix

Siwei Wang, Xinwang Liu, Li Liu, Wenxuan Tu, Xinzhong Zhu,
Jiyuan Liu, Sihang Zhou and En Zhu
{wangsiwei13,xinwangliu,twx,liujiyuan13,enzhu}@nudt.edu.cn,Li.Liu@oulu.fi

A. Details of Probability Model

The transition probability in one-step for i-th sample and j-th anchor as follows,

$$\mathbf{M}^{(1)}(x_i|c_j) = \frac{\mathbf{Z}_{ji}}{\sum_{j=1}^{m}\mathbf{Z}_{ji}}, \mathbf{M}^{(1)}(c_j|x_i) = \frac{\mathbf{Z}_{ji}}{\sum_{i=1}^{n}\mathbf{Z}_{ji}}. \quad (1)$$

Then we can easily obtain that

$$\mathbf{M}^{(2)}(x_i|x_j) = \sum_{t=1}^{k}\mathbf{M}^{(1)}(x_i|c_t)\mathbf{M}^{(1)}(c_t|x_j) = \sum_{t=1}^{k}\mathbf{Z}_{ti}\mathbf{Z}_{tj}. \quad (2)$$

Therefore $\mathbf{M}^{(2)} = \mathbf{Z}^\top\Lambda^{-1}\mathbf{Z}$ where $\Lambda_{ii} = \sum_{j=1}^{n}\mathbf{Z}_{ij}$. It is easy to prove that $\mathbf{M}^{(2)}$ is a doubly stochastic matrix where satisfies the following three properties:

- $\mathbf{1}^\top\mathbf{M}^{(2)} = \mathbf{1}$ and $\mathbf{M}^{(2)}\mathbf{1} = \mathbf{1}$;
- $\mathbf{M}^{(2)} = \mathbf{M}^{(2)\top}$.

Hence we can do spectral clustering on $\mathbf{M}^{(2)}$ to get clustering labels which can run k-means on rank-k right singular vector of $\mathbf{Z}$ [2].

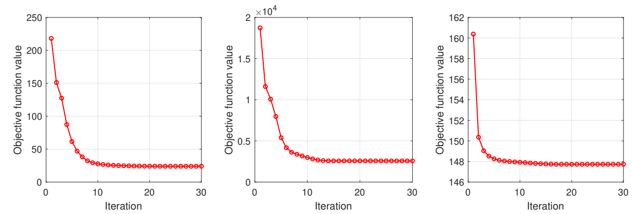
Theorem 1 *The right singular vectors of $\mathbf{Z}$ is the same as the eigenvectors of $\mathbf{Z}^\top\Delta^{-1}\mathbf{Z}$.*

Proof 1 *Suppose the singular value decomposition (SVD) of $\mathbf{Z}$ is $\mathbf{Z} = \mathbf{U}\Sigma\mathbf{V}^\top$, we can easily see that $\mathbf{Z}^\top\Lambda^{-1}\mathbf{Z} = \mathbf{V}\Sigma^\top\Lambda^{-1}\Sigma\mathbf{V}^\top$. Therefore, the right singular vector of $\mathbf{Z}$ is the same as the eigenvectors of $\mathbf{Z}^\top\Delta^{-1}\mathbf{Z}$. This completes the proof.*

According to the Theorem 1, we can conclude that the spectral embedding can be obtained by performing SVD on the consensus anchor graph $\mathbf{Z}$ which only needs $\mathcal{O}(nk^2)$ instead of existing $\mathcal{O}(n^3)$.

B. Convergence

The evolution of objective values on other five datasets are shown in Figure. 1. From these experiments, we observe that the objective values of our algorithm monotonically decrease at each iteration. These results clearly verify our algorithm's convergence.



(a) Caltech101-7 (b) Mnist (c) NGS

Figure 1. The objective of our proposed method on other benchmark datasets.

C. Technical Details

In the optimization process, some subproblems can be written as follows,

$$\max_{\mathbf{F}} \mathrm{Tr}(\mathbf{F}^\top\mathbf{U}) \;\; s.t. \;\; \mathbf{F}^\top\mathbf{F} = \mathbf{I}_k, \quad (3)$$

The optimum of $\mathbf{F}$ can be analytically obtained by Theorem 2. We offer an alternative proof for the optimum to solve the subproblem.

Theorem 2 *Suppose that the matrix $\mathbf{U}$ in Eq. (3) has the rank-k truncated singular value decomposition form as $\mathbf{U} = \mathbf{S}_k\mathbf{\Sigma}_k\mathbf{V}_k^\top$, where $\mathbf{S}_k \in \mathbb{R}^{n\times k}, \mathbf{\Sigma}_k \in \mathbb{R}^{k\times k}, \mathbf{V}_k \in \mathbb{R}^{k\times k}$. The optimization problem in Eq. (3) has a closed-form optimum as follows,*

$$\mathbf{F} = \mathbf{S}_k\mathbf{V}_k^\top, \quad (4)$$

Proof 2 *By taking the the singular value decomposition that $\mathbf{U} = \mathbf{S\Sigma V}^\top$, the Eq. (3) could be rewritten as,*

$$\mathrm{Tr}(\mathbf{F}^\top\mathbf{S\Sigma V}^\top) = \mathrm{Tr}(\mathbf{V}^\top\mathbf{F}^\top\mathbf{S\Sigma}) = \mathrm{Tr}(\mathbf{Q\Sigma}), \quad (5)$$

where $\mathbf{Q} = \mathbf{V}^\top\mathbf{F}^\top\mathbf{S}$, we have $\mathbf{QQ}^\top = \mathbf{V}^\top\mathbf{F}^\top\mathbf{SS}^\top\mathbf{FV} = \mathbf{I}_k$. We can obtain that $\mathrm{Tr}(\mathbf{V}^\top\mathbf{F}^\top\mathbf{S\Sigma}) = \mathrm{Tr}(\mathbf{Q\Sigma}) \le \sum_{i=1}^{k}\sigma_i$. Hence the optimum in Eq. (3) can be reached by the solution given as Eq. (4). In machine learning and computer vision community, Eq. (3) is called Orthogonal Procrustes Analysis which has been well studied in literature [1, 3]. Its optimum has also been provided in [1, 3]. Interested readers can read [1, 3] for other proofs.

D. Detailed Experimental Results

In this section, we provide more experimental results on benchmark datasets in the following pages.

Performance on complete NUS-wide:

Table 1. Performance on complete NUS-wide

Dataset	ACC	NMI	Purity	F-score
Nus-wide	22.46	16.42	26.62	16.48

References

[1] Trevor Hastie, Jerome H. Friedman, and Robert Tibshirani. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer Series in Statistics. Springer, 2001. 1

[2] Wei Liu, Junfeng He, and Shih-Fu Chang. Large graph construction for scalable semi-supervised learning. In *ICML*, 2010. 1

[3] Feiping Nie, Lai Tian, and Xuelong Li. Multiview clustering via adaptively weighted procrustes. In Yike Guo and Faisal Farooq, editors, *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, KDD 2018, London, UK, August 19-23, 2018*, pages 2022–2030. ACM, 2018. 1

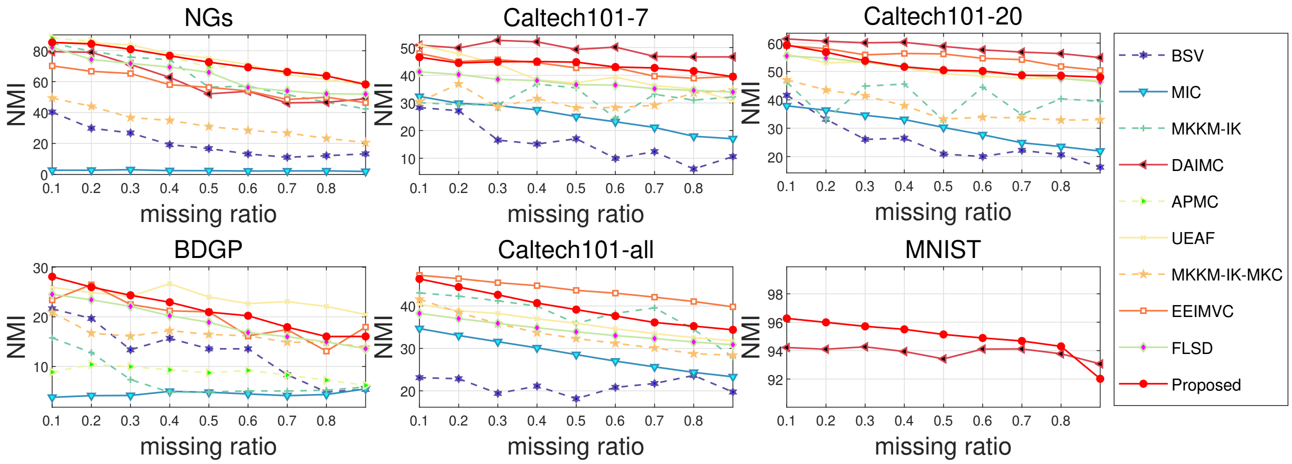


Figure 2. The clustering results of NMI metric on benchmark datasets with different incomplete ratios. Only ours can run YoutubeFace so it is omitted.

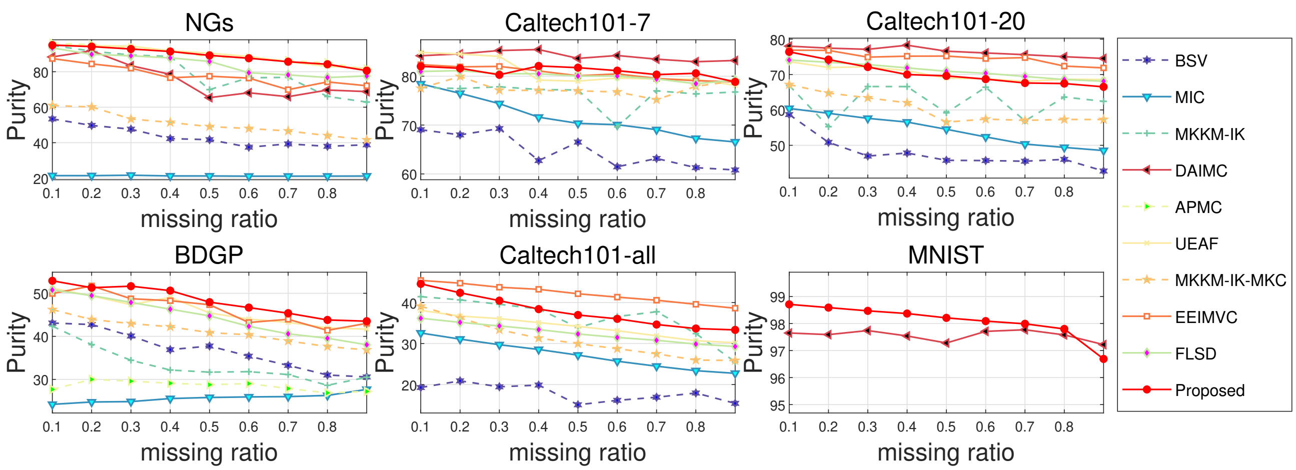


Figure 3. The clustering results of Purity metric on benchmark datasets with different incomplete ratios. Only ours can run YoutubeFace so it is omitted.

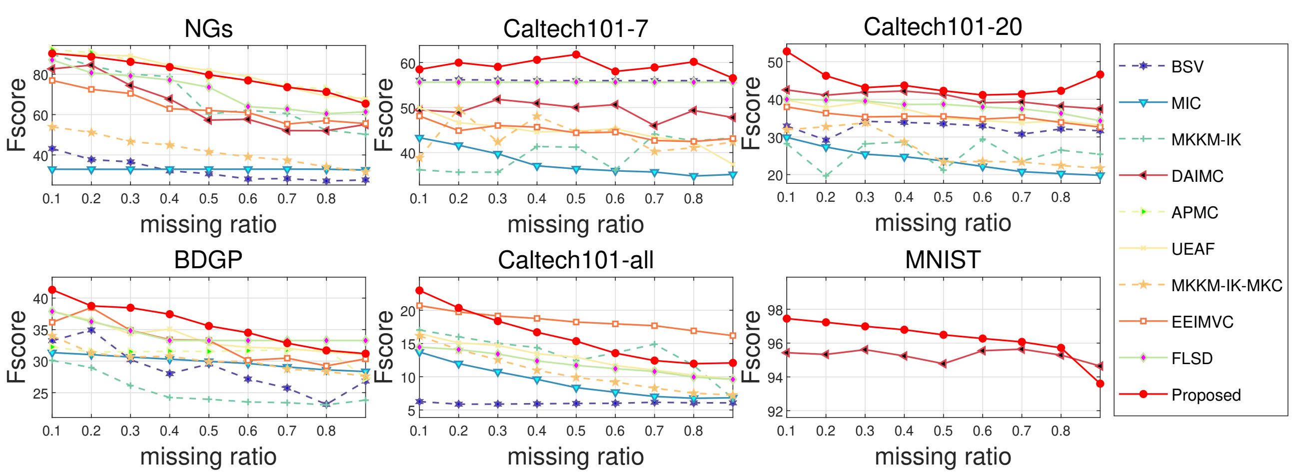


Figure 4. The clustering results of F-score metric on benchmark datasets with different incomplete ratios. Only ours can run YoutubeFace so it is omitted.