2022 IEEE International Symposium on Information Theory (ISIT) | 978-1-6654-2159-1/22/$31.00 ©2022 IEEE | DOI: 10.1109/ISIT50566.2022.9834685

Graph-assisted Matrix Completion in a Multi-clustered Graph Model

Geewon Suh
EE, KAIST
Email: gwsuh91@kaist.ac.kr

Changho Suh
EE, KAIST
Email: chsuh@kaist.ac.kr

***Abstract*—We consider a matrix completion problem that exploits social graph as side information. We develop a computationally efficient algorithm that achieves the optimal sample complexity *for the entire regime of graph information* under the *multiple cluster* setting (to be detailed). The key idea is to incorporate a switching mechanism which selects the information employed in the first clustering step, between the following two types: graph & matrix ratings. Our experimental results on both synthetic and real data corroborate our theoretical result as well as demonstrate that our algorithm outperforms prior algorithms that leverage graph side information.**

A full version of this paper is accessible at: https://sites.google.com/view/gwsuh/home/full-version_isit2022

I. Introduction

Recommender systems (RSs) aim to recommend items that users may prefer or be interested in. A well-known technique for operating RSs is low-rank matrix completion, which have been extensively investigated and shown to be widely applicable [1–8]. One challenge that arises in practice is posed as the so-called *cold start problem*: new users cannot get an appropriate recommendation. A natural way to address the challenge is to enhance RSs with the aid of side information. Indeed, recent works were dedicated to improving the quality of recommendation by employing a social network graph [9–18].

There have been numerous studies providing information-theoretical insights to enhance recommender systems by exploiting graph side information [18–23]. In particular, the prior work [18] characterized the optimal sample complexity for exact matrix recovery as a function of the *quality* of graph information. Under the stochastic block model (SBM), the quality is often quantified as $I_s := (\sqrt{\alpha} - \sqrt{\beta})^2$ where α (or β) indicates the edge probability between two users in the same (or different) clusters. This work has been extended to a generalized scenario in the follow-up work [19], targeting a *multiple* cluster setting.

The prior works proposed computationally efficient algorithms but these are shown to achieve the optimal sample complexity only for a limited scenario wherein the amount of graph information is sufficiently large. Specifically, the achievable regime w.r.t. I_s reads $I_s = \omega\left(\frac{1}{n}\right)$ where n denotes the number of users in the given social graph. While an optimal and efficient algorithm had been out of reach, the very recent work [24] has provided such an algorithm in the simple two-cluster setting.

Contributions: Our contribution is to develop and analyze a computationally efficient algorithm that ensures optimality *for the entire range of* I_s in the generalized *multiple* cluster setting.

The prior algorithms [18–23] follow a well-known two-step procedure: (i) obtaining an initial estimate on user clustering via a spectral method; and (ii) recovering matrix ratings followed by iterative local refinement of clustering. Since the first clustering step is solely based on graph side information, the algorithms do not guarantee achievability for the scarce graph information regime, e.g., $I_s = O\left(\frac{1}{n}\right)$ in [18, 19]. To overcome this challenge, Suh et al. [24] proposed a switching-geared clustering strategy, which selects employed information between graph and matrix ratings depending on the quality of the two information. But the strategy and the corresponding analysis are confined to the simple two-cluster setting. Our main contribution is to generalize this mechanism to accommodate multiple clusters with theoretical guarantee; see Algorithm 1 for details on the switching strategy, as well as see Remark 4 for distinctive technical aspects of the analysis. We show that our algorithm achieves the optimal sample complexity for the entire I_s regime. The analysis is based on perturbation techniques in random matrix theory [25, 26].

Next, we conduct experiments on synthetic datasets to corroborate the theoretical guarantee of the proposed algorithm. We further demonstrate the superior performance over the other approach [18, 19], hinging solely upon graph information in the first clustering stage. We also employ real graph datasets [27, 28] to compare with other matrix completion approaches. See Fig. 4 in Section V.

Related works: In addition to [18], graph-assisted matrix completion has been explored in various multiple cluster settings [19, 20, 22, 23]. Yoon et al. [19] characterized the fundamental limit on sample probability required for matrix completion in the multiple cluster setting. Elmahdy et al. [20] considered a hierarchical cluster setting in which clusters exhibit another sub-clustering structure. Zhang et al. [22, 23] explored a richer setting which exploits as side information both social and item similarity graphs. While all of the works developed computationally efficient algorithms, the optimality of the algorithms was shown only when the amount of side information is sufficiently large, i.e., $I_s = \omega\left(\frac{1}{n}\right)$. On the other hand, our algorithm ensures the optimality for the entire I_s

regime. As mentioned earlier, Suh et al. [24] developed an efficient and optimal algorithm but the considered model is restricted to the two-cluster setting.

The initial clustering step of our algorithm relies on a line of research regarding graph-based clustering [25, 29–33] as well as other clustering methods [34, 35]. In particular, our mechanism encompasses matrix-rating-based clustering aided by singular value decomposition (SVD) [34, 35]. Several techniques [26, 36–39] employed for the analysis of SVD-based approaches form the basis of our proof. But there are some distinctive technical contributions; see Lemmas 2 and 3 for details.

Notations: For a graph $\mathcal{G}$ and two exclusive vertex sets V_1, V_2, $e(V_1, V_2)$ indicates the number of edges between V_1 and V_2.

II. Problem Formulation

Setting: We consider a rating matrix with n users (rows) and m items (columns). Each user rates items either as 1 or -1, denoting "like"/"dislike" respectively. We assign 0 for unrated items. As in [19], we focus on a simple multiple-cluster setting in which there are k equal-sized clusters of users, say $C_1, \ldots, C_k$ (for simplicity, n is assumed to be divisible by k), and $k = O(1)$. The users from the same cluster share the same rating vector over the items. For all $i \in [k]$, let $v_i \in \{-1, 1\}^{1\times m}$ be the rating vector w.r.t. cluster C_i. Assume that rating vectors $\{v_1, \ldots, v_k\}$ are linearly independent. Let $M \in \{-1, 1\}^{n\times m}$ be a rating matrix where the ith row corresponds to the rating vector of user i.

We highlight one important measure that is instrumental in describing the main result (to be stated shortly): $\delta = \frac{1}{m} \min_{i,j\in[k]} d_H(v_i, v_j)$, the normalized minimum Hamming distance w.r.t. all pairs of rating vectors $\{v_1, \ldots, v_k\}$. We partition all the possible rating matrices into subsets depending on δ. For fixed δ, $M^{(\delta)}$ be the set of rating matrices subject to δ, i.e., every pair of rating vectors has a distance less than δm.

Problem of interest: Our goal is to recover a rating matrix M given two types of information. The first is a partially observed rating matrix $Y \in \{-1, 0, 1\}^{n\times m}$. We denote by Ω the set of observed entries of Y: $\Omega = \{(i, j) \in [n] \times [m] : Y_{ij} \neq 0\}$. We assume that each entry is observed with probability $p \in [0, 1]$ independently from others. Its observation can possibly be flipped with probability $\theta \in [0, \frac{1}{2})$. In other words, the entry of Y respects $Y_{ij} \sim \mathsf{Bern}(p) \cdot (1 - 2\mathsf{Bern}(\theta)) \times M_{ij}$.

The second is an undirected social graph $\mathcal{G} = ([n], E)$, where E denotes the set of edges, capturing the connection between two users. The set $[n]$ of nodes are partitioned into k disjoint and equal-sized clusters. We assume that the graph follows the SBM with two types of edge probabilities: α (or β) for intra-cluster (or cross-cluster) users. We consider practical scenarios where users from the same cluster are more likely to be connected, i.e., $\alpha > \beta$.

Performance metric: Let ψ be the estimator of a rating matrix, taking $(Y, \mathcal{G})$ as an input. As a performance metric, we use the worst-case probability of error:

$$P_e^{(\delta)}(\psi) := \max_{M\in M^{(\delta)}} \mathbb{P}[\psi(Y, \mathcal{G}) \neq M].$$

Our goal is to develop a computationally efficient estimator ψ that drives $P_e^{(\delta)} \to 0$ as $n \to \infty$ for any p larger than the optimal sample probability p^* for a constant δ.

III. Main Results

We first state the optimal sample probability p^* characterized in [19] under the considered model. To capture the quality of social graph, we introduce $I_s := (\sqrt{\alpha} - \sqrt{\beta})^2$, a quantified measure representing the clustering capability. The higher, the easier to cluster and hence the more graph information. As in [18], we take the same assumption on n and m that yields the large deviation results in the proof: $m = \omega(\log n)$ and $\log m = o(n)$. This condition is also practically relevant, as it rules out highly asymmetric matrices.

Theorem 1 (Optimal sample probability [19]): Let

$$p^*(I_s) := \frac{1}{(\sqrt{1-\theta} - \sqrt{\theta})^2} \max\left\{\frac{k\log n - nI_s}{k\delta m}, \frac{k\log m}{n}\right\}.$$

Fix $\epsilon > 0$. If $p > (1 + \epsilon)p^*(I_s)$, $P_e^{(\delta)}(\psi) \to 0$ as $n \to \infty$ for some sequence of estimator ψ. Conversely, if $p < (1 - \epsilon)p^*(I_s)$, $P_e^{(\delta)}(\psi) \not\to 0$ as $n \to \infty$ for any ψ.

For notational simplicity, from below, we use p^* instead of $p^*(I_s)$. Yoon et al. [19] developed an efficient algorithm that achieves p^* yet only with enough graph information, formally stated below.

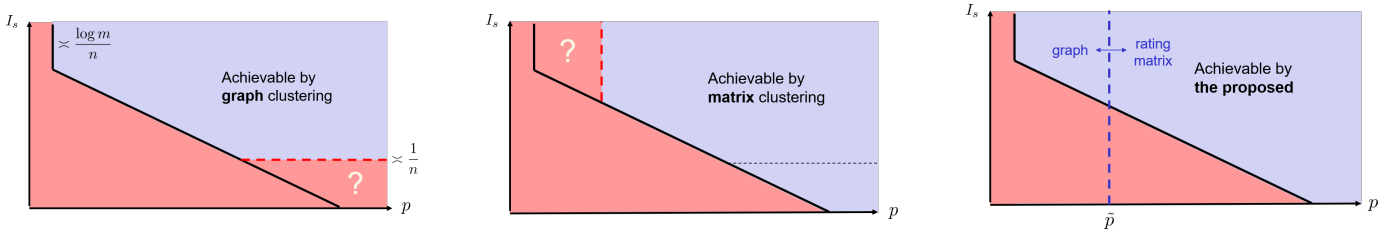
Theorem 2 (Theoretical guarantees of a prior algorithm [19]): Suppose that $I_s = \omega\left(\frac{1}{n}\right)$ and p respects the sufficient condition in Theorem 1. Then, the estimator ψ_0 in [19] exactly recovers the rating matrix M with high probability as n and m tend to infinity: $\mathbb{P}(\hat{M} = M) = 1 - o(1)$ where $\hat{M} := \psi_0(Y, \mathcal{G})$.

The optimal and computationally efficient algorithm guaranteed for the entire regime of I_s has been unknown. To answer this open question, we develop an efficient universal algorithm that promises p^* for the entire achievable regime.

Theorem 3 (Theoretical guarantees of our universal algorithm): Suppose that p respects the sufficient condition in Theorem 1. Then, our computationally-efficient algorithm exactly recovers M with high probability, approaching 1 as n and m tend to infinity.

Theorem 3 implies that there is no information-computation gap for the entire achievable regime, including the case of $I_s = O\left(\frac{1}{n}\right)$. In fact, the prior algorithm [18, 19] solely depends on the social graph information in the initial clustering step. This motivated [24] to develop a different clustering strategy which exploits both types of information.

Inspired by the proposed algorithm in [24], we incorporate a switching-geared clustering strategy that properly selects the employed information for clustering between graph and matrix ratings. While the applicability of the prior result [24] is limited to the two-cluster scenario, our algorithm ensures universal optimality for general multiple cluster settings.



(a) Graph clustering (b) Matrix clustering (c) Proposed

Fig. 1: Achievable regimes (shaded in blue color) due to: (a) graph-clustering approach (use only graph in the first clustering step); (b) matrix-rating-clustering approach (use only matrix ratings); (c) our proposed approach (use graph or matrix ratings depending on the amount of graph information). With proper switching of the employed information for clustering, our algorithm achieves the entire achievable regime promised by MLE.

Remark 1 (Switching-geared strategy): Here we briefly explain how the switching mechanism developed in [24] works. Fig. 1 depicts achievable regimes (shaded in blue color) in terms of (p, I_s). The red shaded region indicates the non-achievable regime, and the optimal sample probability $p^* = p^*(I_s)$ is indicated by the black bold line. Fig. 1(a) shows the achievable regime for the algorithm in [19] where the first clustering step employs only graph information. As implied by Theorem 2, the achievability is proved when $I_s = \omega\left(\frac{1}{n}\right)$. Fig. 1(b) refers to the case in which only matrix ratings are employed in the first clustering step. Suh et al. [24] showed the achievability when p is large: $p = \omega\left(\frac{1}{m} + \frac{1}{\sqrt{nm}}\right)$, on the other hand, the achievability is not explored in the small p regime. By switching these two clustering methods with a proper threshold, say $\tilde{p}$, it is shown in [24] that the algorithm can achieve the information-theoretical limit. See Fig. 1(c). ■

Remark 2 (Comparison to the prior work [24]): A limitation of the prior algorithm [24] comes from a restricted two-cluster scenario. Employing an additional deviation technique (see Remark 3 for details), we extend to a generalized multiple cluster setting. One interesting aspect of our algorithm is that we can make a k-unaware choice for the threshold $\tilde{p}$ so that it is independent of k. See Section IV-B for details. ■

IV. Proof of Theorem 3

A. Algorithm Description

The overall structure of our algorithm follows the prominent two-stage procedure [5, 29, 31, 40, 41]. As in [18, 19, 21, 23], we further divide the second stage into two steps, so there are three steps in total. Step 1 is the major one that invokes the key idea inspired by [24]. In Step 1, we intend to estimate all the clusters based on graph and matrix ratings. The 2nd and 3rd steps admit the standard procedures employed in [18, 19]. In Step 2, we recover the rating vectors via majority voting, given clustering information estimated in Step 1. In Step 3, we employ point-wise maximum likelihood estimation (MLE) w.r.t. each user's cluster to successively refine the cluster information. The details are described below; also see Algorithm 1.

Step 1 (Initial clustering via a switching mechanism): Notice that the regime $I_s = O\left(\frac{1}{n}\right)$ is not achievable with graph-based clustering, while being achievable with matrix-rating-based clustering. For another regime $I_s = \omega\left(\frac{1}{n}\right)$, the other way around holds. From this observation, one can suggest a switching mechanism that makes a proper selection among two clustering methods. This motivates Suh et al. [24] to propose the following rule: For a small p (a large I_s), we apply graph-based-clustering approach as in [19], and for a large p (a small I_s), we perform matrix-rating-based clustering. The authors showed that by choosing a threshold probability like $\tilde{p} = \omega\left(\frac{1}{m} + \frac{1}{\sqrt{nm}}\right)$, the proposed algorithm ensures the optimality in the two-cluster setting.

Now a natural follow-up question arises: how to develop a proper switching threshold for general k? It turns out that the same threshold employed for the two-cluster setting (chosen to be independent of k) can still be a good choice that enables the achievability for an arbitrary value of k. We show that $p > \tilde{p}$ is translated to the low $I_s = O\left(\frac{1}{n}\right)$ regime, implying that the entire I_s regime is covered by a switching strategy. See Section IV-B for proof details.

Matrix-rating-based clustering: The switching rule requires the knowledge of p which is unknown. So we employ an estimate for p, e.g., the ML estimate $\hat{p} = \frac{|\Omega|}{mn}$. If $\hat{p}$ is above the threshold $\tilde{p}$, we employ matrix-based clustering. Specifically we first apply a singular value decomposition (SVD) w.r.t. the observed matrix Y: $Y = U\Sigma V^T$. Then, from the k leading columns of U, we generate an n-by-k matrix U_Y. Next, we employ the k-means algorithm [42] w.r.t. U_Y to obtain an initial estimatie of clusters, say $\{C_l^{(0)}\}_{l\in[k]}$. See lines 3–6 in Algorithm 1. In Section IV, we will show that this clustering method guarantees weak clustering, which plays a crucial role in ensuring matrix completion together with Steps 2 and 3.

Graph-based clustering: If $\hat{p}$ is below $\tilde{p}$, we instead employ graph clustering method such as spectral clustering [31]. See lines 7 and 8 in Algorithm 1. It has been shown that graph clustering guarantees weak clustering in the considered regime [31].

Steps 2 and 3 are identical to those in [18, 19]. For completeness, we describe the procedures yet in a brief manner.

Step 2 (Exact recovery of rating vectors): Based on the initial clustering $\{C_l^{(0)}\}$, we estimate v_i via maximum likelihood estimation i.e., majority voting in the case. For item j and

cluster C_l, we collect the observed ratings of the item by the users in $C_l^{(0)}$. We set the jth entry of $\hat{v}_l$ (estimate of v_l) to be the most frequently appeared one among the collected ratings. Similarly we estimate all the entries by sweeping $l \in [k]$. See lines 10–15 for implementation details. It was shown in [18, 19] that the simple majority voting ensures exact recovery of rating vectors.

Step 3 (Exact recovery of clustering): We refine the cluster $\{C_l^{(0)}\}$ using point-wise MLE in which the likelihood is completed w.r.t. an interested user cluster while fixing the other user's clusters estimated in the prior iteration. See lines 21–28 for details. We apply $T = O(\log n)$ iterations, as it is shown that the choice of T guarantees exact clustering [18, 19]. For concreteness, as illustrated in lines 17–19, we include a procedure for estimating (α, β, θ) (via MLE) needed for the likelihood computation.

B. Proof Outline

Due to space limitation, we provide only the sketch of the proof while leaving the complete proof in [43].

The proof consists of two parts. The first is to show that the achievability proof boils down to the proof of weak recovery of matrix clustering when $p = \omega(\frac{1}{m} + \frac{1}{\sqrt{nm}})$. This will be proved in Lemma 1. The second is to prove the weak recovery for the regime of interest described in Lemma 1. See Lemma 2.

For the first part, we introduce two regimes: Regime 1 := $\{(p, I_s) : p \geq p^*, p < \tilde{p}\}$, Regime 2 := $\{(p, I_s) : p \geq p^*, p \geq \tilde{p}\}$ where $\tilde{p}$ is chosen as $\frac{\log\log n}{m}$. We claim that: (i) in Regime 1, $p < \tilde{p}$ implies $I_s = \omega\left(\frac{1}{n}\right)$; (ii) in Regime 2, $p \geq \tilde{p}$ covers $I_s = O(\frac{1}{n})$. This claim is proved in Lemma 1.

Lemma 1: If $p \geq p^*$ and $p < \tilde{p}$, then $I_s = \omega\left(\frac{1}{n}\right)$. On the other hand, if $p \geq p^*$ and $I_s = O\left(\frac{1}{n}\right)$, then $p \geq \tilde{p}$.

The second part stated in Lemma 1 implies that $p \geq \tilde{p}$ covers the remaining regime $I_s = O\left(\frac{1}{n}\right)$ as long as $p \geq p^*$. Hence, for Regime 2, it suffices to prove weak recovery of matrix clustering. The proof of this is in Lemma 2.

Lemma 2: If $p \geq p^*$ and $p = \omega\left(\frac{1}{m} + \frac{1}{\sqrt{nm}}\right)$, matrix-rating-based clustering in Step 1 guarantees weak recovery.

Proof: Let $C_l \in \mathbb{R}^{1\times k}$ indicate the centers among the points (each denoting a certain row in U_Y) that correspond to cluster C_l respectively. Let U_G be an n-by-k matrix such that the (ith row of U_G) $= C_l$ if $i \in C_l$. Here we prove that by using the approximate k-means error bound techniques being used in [25] (see Lemma 5.3 therein), the fraction of misclustered users is bounded by $\|U_Y - U_G\|_F^2$ up to a constant factor. Hence, it suffices to show $\|U_Y - U_G\|_F^2 \to 0$ for proving weak recovery. This proof is done in Lemma 3. This completes the proof of Lemma 2. ■

Lemma 3: If $p = \omega\left(\frac{1}{m} + \frac{1}{\sqrt{nm}}\right)$, $\|U_Y - U_G\|_F^2 \to 0$ with high probability as $n, m \to \infty$.

Remark 3 (Technical novelty): One major technical contribution is reflected in Lemma 3. The key step in the proof is to show that U_Y and the k leading ground-truth singular vectors are very similar. To this end, we employ perturbation bounding technique for singular subspaces in [25]. We then derive an upper bound of $\sin\Theta$ distance between U_Y and the ground-truth singular vectors as a function of the variance of Y_{ij}'s, and the k-th singular value of Y. In the case of $k = 2$, the closed form solution of the singular value can be easily computed because the ground-truth rating for each item is either same or different. But in the generalized setting, since there are 2^{k-1} cases of the ground truth for each item, direct computation is not that simple. To overcome this challenge, we employ linear algebra techniques to derive the asymptotic bound of the k-th singular value of Y as a function of n, m, k and δ. See Claim 1 in the full version [43] for details. ■

Algorithm 1: Proposed Algorithm

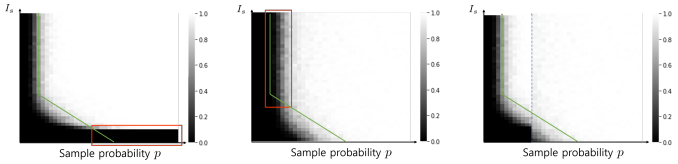
Input : Observed rating matrix $Y \in \{-1, 0, 1\}^{n\times m}$, Graph $\mathcal{G} = ([n], E)$, The number of cluster k, The number of iteration for cluster refinement T

Output: Estimate of a rating matrix $\hat{M} \in \{-1, 1\}^{n\times m}$

1 $\hat{p} \leftarrow |\Omega|/nm$;
2 **Step 1 (Initial clustering via a switching mechanism)**
3 **if** $\hat{p} > \tilde{p} := \frac{\log\log n}{m}$ **then**
4 $U\Sigma V^T \leftarrow$ singular value decomposition of Y;
5 $U_Y \leftarrow k$ leading columns of U;
6 Apply the k-means clustering w.r.t. U_Y to obtain an initial estimate for clustering: $\{C_l^{(0)}\}_{l\in[k]}$;
7 **else**
8 Apply graph-based clustering w.r.t. $\mathcal{G}$ to obtain an initial estimate for clustering: $\{C_l^{(0)}\}_{l\in[k]}$;
9 **end**
10 **Step 2 (Recovery of rating vectors)**
11 **for** *item* $j = 1$ *to* m **do**
12 **for** *cluster* $l = 1$ *to* k **do**
13 $(\hat{v}_l)_j \leftarrow \text{sign}\left(\sum_{i\in C_l^{(0)}} Y_{ij}\right)$;
14 **end**
15 **end**
16 **Step 3 (Local refinement of clustering)**
17 $\hat{\alpha} \leftarrow \frac{1}{\sum_l \binom{|C_l^{(0)}|}{2}} \left(\sum_l |\{\{i_1, i_2\} \in E : i_1, i_2 \in C_l^{(0)}\}|\right)$;
18 $\hat{\beta} \leftarrow \frac{1}{\sum_{l_1\neq l_2} |C_{l_1}^{(0)}||C_{l_2}^{(0)}|} \left(\sum_{l_1\neq l_2} e(C_{l_1}^{(0)}, C_{l_2}^{(0)})\right)$;
19 $\hat{\theta} \leftarrow |\{(i,j) \in [n]\times[m] : Y_{ij} \neq 0, Y_{ij} \neq (\hat{v}_l)_j, i \in C_l^{(0)} \text{ for any } l \in [k]\}|/|\Omega|$;
20 **for** *iteration* $t = 1$ *to* T **do**
21 $C_l^{(t)} \leftarrow \emptyset$ for all $l \in [k]$;
22 **for** *user* $i = 1$ *to* n **do**
23 **for** *cluster* $l = 1$ *to* k **do**
24 $L_l(i) \leftarrow \log\left(\frac{(1-\hat{\beta})\hat{\alpha}}{(1-\hat{\alpha})\hat{\beta}}\right) e(\{i\}, C_l^{(t-1)}) + \log\left(\frac{1-\hat{\theta}}{\hat{\theta}}\right) \sum_{j\in[m]} \mathbf{1}(Y_{ij} = (\hat{v}_l)_j)$;
25 **end**
26 $l' \leftarrow \text{argmax}_l\, L_l(i)$; $\quad L_{l'}^{(t)} \leftarrow L_{l'}^{(t)} \cup \{i\}$;
27 **end**
28 **end**
29 **for** *user* $i = 1$ *to* n **do**
30 **if** $i \in C_l^{(T)}$ **then** $\hat{M}_i \leftarrow \hat{v}_l$;
31 **end**
32 **Return** $\hat{M}$

V. Experiments

We provide Monte Carlo experiments to corroborate the main result (Theorem 3).



(a) Graph clustering (b) Matrix clustering (c) Proposed

Fig. 2: Achievable regimes of (p, I_s) due to: (a) graph-based clustering; (b) matrix-rating-based clustering; (c) our proposed algorithm. Here the brightness indicates the level of the empirical success rate; the brighter, the higher.

Synthetic data: We apply our algorithm to synthetic data generated according to the model described in Section II.

Here, we consider a four-cluster setting where $(\theta, \delta, n, m) = (0.1, 0.2, 1000, 100)$. In Fig. 2, we evaluate the performance of three algorithms via the empirical success rate for a rage of (p, I_s). Here the success rate is computed averaged over 200 random trials. The empirical success rate is visualized by the brightness level; the brighter, the higher. The green line indicates the sharp boundary indicated by the optimal sample probability $p^* = p^*(I_s)$. Fig. 2(a) shows the performance of graph-clustering-based approach. Notice that the recovery fails in the low I_s regime. See the lower right corner of Fig. 2(a). On the other hand, in Fig. 2(b) w.r.t. matrix-rating-based approach, recovery is mostly successful in the low I_s regime, however, it suffers from poor performance in the high I_s (low p) regime; see the upper left corner. By properly switching between two clustering methods depending on $\hat{p}$ (MLE of p), our algorithm shows an enhanced performance for the entire achievable regime; see Fig. 2(c).

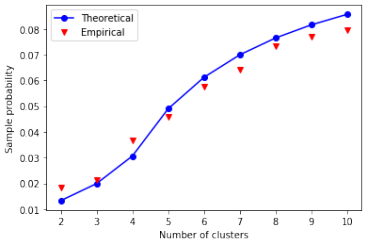
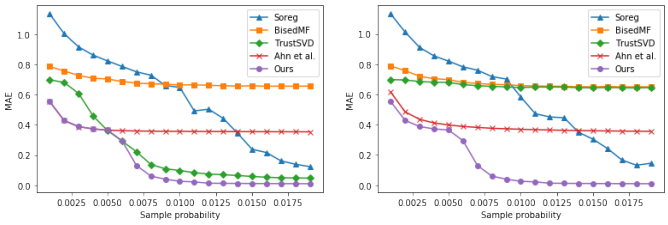


Fig. 3: Optimal (theoretical, marked in blue circle) sample probability vs. *empirical* (marked in red triangle) sample probability as a function of k. Here the empirical sample probability indicates the minimum p above which almost all matrix entries (at least 99.95% entries) are successfully recovered.

Fig. 3 demonstrates the performance for different numbers of clusters. We sweep k from 3 to 10, while keeping $\theta = 0.1$, $\delta = 0.2$, $n = 2520$, $m = 800$, $I_s = \frac{3 \log n}{n}$. For y-axis, we consider two quantities: (i) optimal sample probability; and (ii) *empirical* sample probability. We mean by the empirical sample probability the minimal p above which at least 99.95% matrix entries (e.g., $2,015,000$ out of $2,016,000$ in the considered setting) are successfully recovered. Observe that the empirical sample probability is very close to the optimal one for a range of k, corroborating our theoretical result.



(a) (b)

Fig. 4: Comparison of MAEs evaluated for various recommendation algorithms using Facebook network [28]: (a) the original network (high I_s regime); (b) the sparse network generated by subsampling (low I_s regime).

Real data: As in [18–21, 23], we consider a semi-real data setting in which a social graph is real while rating vectors are synthetically generated as per our considered model.

As a real graph, we employ Facebook network [28] having the four ground-truth clusters and $n = 4675$. We consider two network settings: (i) the original network; (ii) a sparse network generated by random subsampling (5%) of edges from the original network. We synthesize ratings to construct a rating matrix with $(n, m) = (4675, 500)$. We use mean absolute error (MAE) as a metric to compare ours with various recommendation algorithms exploiting graph side information: (i) item k-nearest neighbor (k-NN) [44]; (ii) user k-NN [44]; (iii) matrix factorization and social regularization (SoReg) [14]; (iv) biased matrix factorization (Biased MF) [45]; (v) TrustSVD [46]; and (vi) Yoon et al. [19] based solely on graph clustering. As shown in Fig. 4, our algorithm shows better performance than the other baseline algorithms, with a more pronounced gain on the sparse network scenario.

VI. Discussion

We developed a computationally efficient graph-assisted matrix completion algorithm that promises the optimal sample probability for the entire I_s regimes as well as for an arbitrary number k of clusters. The proposed algorithm generalizes the prior switching-geared strategy to multiple cluster setting with an analysis based on the derivation of asymptotic bounds of singular values. One future work of interest is to develop a *soft* version of the *hard*-decision selection mechanism proposed herein. The soft version may be practically more appearing, as it does not rely upon the estimate of p, which can be unreliable in particular for a small-sized matrix. Another future work is to go beyond the considered setting, addressing various network scenarios explored in [20, 22, 23].

Acknowledgement

This work was supported by Institute for Information & communications Technology Planning Evaluation (IITP) grant funded by the Korea government (MSIT) (No. 2019-0-01396, Development of framework for analyzing, detecting, mitigating of bias in AI model and training data).

REFERENCES

[1] M. Fazel, "Matrix rank minimization with applications," Ph.D. dissertation, PhD thesis, Stanford University, 2002.

[2] E. J. Candès and B. Recht, "Exact matrix completion via convex optimization," *Foundations of Computational mathematics*, vol. 9, no. 6, p. 717, 2009.

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

[4] E. J. Candès and T. Tao, "The power of convex relaxation: Near-optimal matrix completion," *IEEE Transactions on Information Theory*, vol. 56, no. 5, pp. 2053–2080, 2010.

[5] R. H. Keshavan, A. Montanari, and S. Oh, "Matrix completion from a few entries," *IEEE Transactions on Information Theory*, vol. 56, no. 6, pp. 2980–2998, 2010.

[6] J.-F. Cai, E. J. Candès, and Z. Shen, "A singular value thresholding algorithm for matrix completion," *SIAM Journal on optimization*, vol. 20, no. 4, pp. 1956–1982, 2010.

[7] K. Lee and Y. Bresler, "Admira: Atomic decomposition for minimum rank approximation," *IEEE Transactions on Information Theory*, vol. 56, no. 9, pp. 4402–4416, 2010.

[8] L. T. Nguyen, J. Kim, and B. Shim, "Low-rank matrix completion: A contemporary survey," *IEEE Access*, vol. 7, pp. 94 215–94 237, 2019.

[9] C. C. Aggarwal, J. L. Wolf, K.-L. Wu, and P. S. Yu, "Horting hatches an egg: A new graph-theoretic approach to collaborative filtering," in *Proceedings of the fifth ACM SIGKDD international conference on Knowledge discovery and data mining*, 1999, pp. 201–212.

[10] P. Massa and P. Avesani, "Trust-aware collaborative filtering for recommender systems," in *OTM Confederated International Conferences" On the Move to Meaningful Internet Systems"*. Springer, 2004, pp. 492–508.

[11] J. Golbeck, J. Hendler *et al.*, "Filmtrust: Movie recommendations using trust in web-based social networks," in *Proceedings of the IEEE Consumer Communications and Networking Conference*, vol. 96, no. 1. Citeseer, 2006, pp. 282–286.

[12] M. Jamali and M. Ester, "Trustwalker: a random walk model for combining trust-based and item-based recommendation," in *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, 2009, pp. 397–406.

[13] D. Cai, X. He, J. Han, and T. S. Huang, "Graph regularized nonnegative matrix factorization for data representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 8, pp. 1548–1560, 2010.

[14] H. Ma, D. Zhou, C. Liu, M. R. Lyu, and I. King, "Recommender systems with social regularization," in *Proceedings of the fourth ACM international conference on Web search and data mining*, 2011, pp. 287–296.

[15] X. Yang, Y. Guo, and Y. Liu, "Bayesian-inference-based recommendation in online social networks," *IEEE Transactions on Parallel and Distributed Systems*, vol. 24, no. 4, pp. 642–651, 2012.

[16] J. Tang, X. Hu, and H. Liu, "Social recommendation: a review," *Social Network Analysis and Mining*, vol. 3, no. 4, pp. 1113–1133, 2013.

[17] F. Monti, M. Bronstein, and X. Bresson, "Geometric matrix completion with recurrent multi-graph neural networks," in *Advances in Neural Information Processing Systems*, 2017, pp. 3697–3707.

[18] K. Ahn, K. Lee, H. Cha, and C. Suh, "Binary rating estimation with graph side information," in *Advances in Neural Information Processing Systems*, vol. 31, 2018, pp. 4272–4283.

[19] J. Yoon, K. Lee, and C. Suh, "On the joint recovery of community structure and community features," in *2018 56th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, 2018, pp. 686–694.

[20] A. Elmahdy, J. Ahn, C. Suh, and S. Mohajer, "Matrix completion with hierarchical graph side information," in *34th Conference on Neural Information Processing Systems, NeurIPS 2020*, 2020.

[21] C. Jo and K. Lee, "Discrete-valued preference estimation with graph side information," *arXiv preprint arXiv:2003.07040*, 2020.

[22] Q. Zhang, V. Y. F. Tan, and C. Suh, "Community detection and matrix completion with social and item similarity graphs," *to appear in the IEEE Transactions on Signal Processing*, 2021.

[23] Q. Zhang, G. Suh, C. Suh, and V. Y. Tan, "Mc2g: An efficient algorithm for matrix completion with social and item similarity graphs," *arXiv preprint arXiv:2006.04373*, 2020.

[24] G. Suh, S. Jeon, and C. Suh, "When to use graph side information in matrix completion," in *2021 IEEE International Symposium on Information Theory (ISIT)*. IEEE, 2021, pp. 2113–2118.

[25] J. Lei and A. Rinaldo, "Consistency of spectral clustering in stochastic block models," *The Annals of Statistics*, vol. 43, no. 1, p. 215–237, Feb 2015.

[26] T. T. Cai and A. Zhang, "Rate-optimal perturbation bounds for singular subspaces with applications to high-dimensional statistics," *The Annals of Statistics*, vol. 46, no. 1, pp. 60–89, Feb 2018.

[27] L. A. Adamic and N. Glance, "The political blogosphere and the 2004 us election: divided they blog," in *Proceedings of the 3rd international workshop on Link discovery*, 2005, pp. 36–43.

[28] A. L. Traud, P. J. Mucha, and M. A. Porter, "Social structure of facebook networks," *Physica A: Statistical Mechanics and its Applications*, vol. 391, no. 16, pp. 4165–4180, 2012.

[29] E. Abbe and C. Sandon, "Community detection in general stochastic block models: Fundamental limits and efficient algorithms for recovery," in *2015 IEEE 56th Annual Symposium on Foundations of Computer Science*. IEEE, 2015, pp. 670–688.

[30] P. Chin, A. Rao, and V. Vu, "Stochastic block model and community detection in sparse graphs: A spectral algorithm with optimal rate of recovery," in *Conference on Learning Theory*, 2015, pp. 391–423.

[31] C. Gao, Z. Ma, A. Y. Zhang, and H. H. Zhou, "Achieving optimal misclassification proportion in stochastic block models," *The Journal of Machine Learning Research*, vol. 18, no. 1, pp. 1980–2024, 2017.

[32] H. Ashtiani, S. Kushagra, and S. Ben-David, "Clustering with same-cluster queries," in *Proceedings of the 30th International Conference on Neural Information Processing Systems*, 2016, pp. 3224–3232.

[33] A. Mazumdar and B. Saha, "Query complexity of clustering with side information," in *Advances in Neural Information Processing Systems*, 2017, pp. 4682–4693.

[34] C. Boutsidis, A. Zouzias, M. W. Mahoney, and P. Drineas, "Randomized dimensionality reduction for k-means clustering," *IEEE Transactions on Information Theory*, vol. 61, no. 2, pp. 1045–1062, 2014.

[35] D. Feldman, M. Schmidt, and C. Sohler, "Turning big data into tiny data: Constant-size coresets for k-means, pca, and projective clustering," *SIAM Journal on Computing*, vol. 49, no. 3, pp. 601–657, 2020.

[36] M. Azizyan, A. Singh, and L. Wasserman, "Minimax theory for high-dimensional gaussian mixtures with sparse mean separation," *Advances in Neural Information Processing Systems*, vol. 26, pp. 2139–2147, 2013.

[37] J. Jin and W. Wang, "Influential features pca for high dimensional clustering," *The Annals of Statistics*, pp. 2323–2359, 2016.

[38] J. Jin, Z. T. Ke, W. Wang *et al.*, "Phase transitions for high dimensional clustering and related problems," *The Annals of Statistics*, vol. 45, no. 5, pp. 2151–2189, 2017.

[39] V. Vu, "A simple svd algorithm for finding hidden partitions," *Combinatorics, Probability & Computing*, vol. 27, no. 1, p. 124, 2018.

[40] Y. Chen and C. Suh, "Spectral mle: Top-k rank aggregation from pairwise comparisons," in *International Conference on Machine Learning*. PMLR, 2015, pp. 371–380.

[41] Y. Chen, G. Kamath, C. Suh, and D. Tse, "Community recovery in graphs with locality," in *International Conference on Machine Learning*. PMLR, 2016, pp. 689–698.

[42] J. MacQueen *et al.*, "Some methods for classification and analysis of multivariate observations," in *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, vol. 1, no. 14. Oakland, CA, USA, 1967, pp. 281–297.

[43] https://sites.google.com/view/gwsuh/home/full-version_isit2022.

[44] R. Pan, P. Dolog, and G. Xu, "Knn-based clustering for improving social recommender systems," in *International workshop on agents and data mining interaction*. Springer, 2012, pp. 115–125.

[45] Y. Koren, "Factorization meets the neighborhood: a multifaceted collaborative filtering model," in *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*, 2008, pp. 426–434.

[46] G. Guo, J. Zhang, and N. Yorke-Smith, "Trustsvd: Collaborative filtering with both the explicit and implicit influence of user trust and of item ratings," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 29, no. 1, 2015.