# On the Fundamental Limits of Matrix Completion: Leveraging Hierarchical Similarity Graphs

Junhyung Ahn, *Member, IEEE*, Adel Elmahdy, *Graduate Student Member, IEEE*, Soheil Mohajer, *Member, IEEE*, and Changho Suh, *Fellow, IEEE*

***Abstract*— We study a matrix completion problem which leverages a hierarchical structure of social similarity graphs as side information in the context of recommender systems. We assume that users are categorized into clusters, each of which comprises sub-clusters (or what we call "groups"). We consider a hierarchical stochastic block model that well respects practically-relevant social graphs and follows a low-rank rating matrix model. Under this setting, we characterize the information-theoretic limit on the number of observed matrix entries (i.e., optimal sample complexity) as a function of the quality of graph side information (to be detailed) by proving sharp upper and lower bounds on the sample complexity. One important consequence of this result is that leveraging the hierarchical structure of similarity graphs yields a substantial gain in sample complexity relative to the one that simply identifies different groups without resorting to the relational structure across them. Another implication of the result is when the graph information is rich, the optimal sample complexity is proportional to the number of clusters, while it nearly stays constant as the number of groups in a cluster increases. We empirically demonstrate through extensive experiments that the proposed algorithm achieves the optimal sample complexity.**

***Index Terms*— Recommender systems, matrix completion problem, graph side information.**

## I. Introduction

IN RECENT years, personalized recommender systems have emerged in an extensive range of Web applications to predict the preferences of its users and provide them with new and relevant items based on scarce data about the users and/or items [1]. There are two major paradigms of recommender systems: (i) content-based filtering systems; (ii) collaborative filtering systems. Content-based filtering approach exploits a profile of users' preferences and/or properties of the items to carry out the recommendation task. On the other hand, collaborative filtering approach recommends new items to users based on similarity measures between users and items. The main advantage of collaborative filtering over content-based filtering is that they do not require domain knowledge since the embeddings are automatically learned, and more interactions between the users and the items lead to a more accurate and relevant new recommendations. Inspired by the Netflix challenge, a well-known technique for predicting missing ratings in collaborative filtering frameworks is low-rank matrix completion, which is the main interest of this paper. Given partial observation of a matrix of users by items, the goal is to develop an algorithm to accurately predict the values of the missing ratings. One of the prime challenges of collaborative filtering systems that rely on user-item interactions is the "cold start problem" in which high-quality recommendations are not feasible for new users/items that bear little or no information. One prominent technique to overcome the problem with cold start users is to incorporate the community information into the framework of recommender systems in order to enhance the recommendation quality. Motivated by the social homophily theory [2] that users within the same community are more likely to share similar preferences, socially-aware collaborative filtering approach exploits the social network among users and provides recommendations based on the similarity measures of users who have direct or indirect social relationships with a given user.

A plethora of research works has explored the idea of exploiting the information inferred by social graphs to enhance the performance of recommender systems from an algorithmic perspective [1], [3], [4], [5], [6], [7], [8], [9], [10], [11], [12], [13], [14], [15], [16], [17], [18], [19], [20], [21]. However, few works were dedicated to developing theoretical insights on the usefulness of graph side information on the quality of recommendation, and characterizing the maximal achievable gain due to side information, e.g., [22] and [23]. Recently, a number of works [24], [25], [26], [27] have investigated the problem of interest from an information-theoretic perspective. Ahn et al. [24] considered a matrix completion problem with $n$ users and $m$ items, and studied a simplified model where

Manuscript received 1 January 2023; revised 17 July 2023; accepted 22 October 2023. Date of publication 28 December 2023; date of current version 16 February 2024. The work of Junhyung Ahn and Changho Suh was supported in part by the Korea Advanced Institute of Science and Technology (KAIST) Grand Challenge 30 Program funded by KAIST under Grant N11230067; and in part by the Institute of Information and Communications Technology Planning and Evaluation (IITP) Grant funded by the Korean Government [Ministry of Science and ICT (MSIT)], Development and Study of AI Technologies to Inexpensively Conform to Evolving Policy on Ethics, under Grant 2022-0-00184. The work of Adel Elmahdy and Soheil Mohajer was supported in part by the National Science Foundation under Grant CCF-1749981. An earlier version of this paper was presented in part at the 2020 Advances in Neural Information Processing Systems Conference and in part at the 2022 IEEE International Symposium on Information Theory [DOI: 10.1109/ISIT50566.2022.9834449]. *(Junhyung Ahn and Adel Elmahdy contributed equally to this work.) (Corresponding author: Adel Elmahdy.)*

Junhyung Ahn and Changho Suh are with the School of Electrical Engineering, Korea Advanced Institute of Science and Technology (KAIST), Daejeon 34141, South Korea (e-mail: tonyahn96@kaist.ac.kr; chsuh@kaist.ac.kr).

Adel Elmahdy and Soheil Mohajer are with the Department of Electrical and Computer Engineering, University of Minnesota, Minneapolis, MN 55455 USA (e-mail: adel@umn.edu; soheil@umn.edu).

Communicated by S. Jalali, Associate Editor for Machine Learning and Statistics.

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TIT.2023.3345902.

Digital Object Identifier 10.1109/TIT.2023.3345902

there are two clusters of users, and the users of each cluster share the same rating over the items. A sharp threshold on the sample complexity is derived as a function of the quality of the social graph information, and the gain due to the information provided by the social graph is theoretically quantified. Furthermore, the authors proposed an efficient rating estimation algorithm that provably achieves the minimum sample complexity for reliable recovery of the ground truth rating matrix. Follow-up works have investigated different models of the matrix completion problem proposed in [24]. Yoon et al. [25] considered a general setting with $K$ hidden communities of possibly different sizes, where each community is associated with only one feature, and hence the users of each community provide the same binary rating over the items. Unlike [24] and [25] where one-sided graph side information (i.e., user-to-user similarity graph) is considered, Zhang et al. [26] studied the benefits of two-sided graph side information depicted by user-to-user and item-to-item similarity graphs. Interestingly, the theoretical analysis demonstrates that there is a synergistic effect, under some scenarios, stemming from considering two pieces of graph side information. This implies that observing both graphs is necessary to reduce the sample complexity under those scenarios. Jo and Lee [27] relaxed the assumption in [24] on the preference matrix whose element at row $i$ and column $j$ denotes the probability that user $i$ likes item $j$, and proposed a new model in which the unknown entries of the preference matrix can take discrete values drawn from a finite set of probabilities. While the works of [24], [25], [26], and [27] lay out the theoretical foundation for the problem, they impose a number of strict assumptions on the system model. In particular, users of the same cluster are assumed to have the same ratings for all items, which limits the practicality of the proposed models for real-world data.

A natural hypothesis in the theory of recommender systems is that the unknown rating matrix has an intrinsic structure of being low-rank. This hypothesis is sensible because it is generally believed that only a few factors contribute to one's preference. Prior works [24], [25], [26], [27] assume that each cluster is represented by a rank-one matrix, and users within a cluster share the same rating vector over items. In this work, we relax this assumption and study a more generalized framework where each cluster is represented by a rank-$r$ matrix. More specifically, we consider a matrix completion problem where the users are categorized into $c$ clusters, each of which comprises $g$ sub-clusters, or what we call "groups", producing a hierarchical structure in which the features of different groups within a cluster are broadly similar to each other; however, they are different from the features of the groups in other clusters. The goal is to reliably retrieve the rating matrix under the proposed generalized model, utilizing the information provided by the noisy partial observation of the rating matrix, as well as the hierarchical social graph.

The main contributions of this paper are summarized as follows. First, we characterize an information-theoretic threshold for reliable matrix recovery as a function of the quantified quality of the considered hierarchical graph side information (to be detailed) by establishing matching upper and lower bounds on the sample complexity. An implication of the result is that our algorithm, which leverages the hierarchical graph structure, yields a substantial gain in sample complexity, compared to a simple variant of [24] and [25] that does not exploit the relational structure across rating vectors of groups. We also reveal that when graph information is rich enough to perfectly retrieve the structures of clusters and groups, the optimal sample complexity increases linearly as the number of clusters increases. Otherwise, the optimal sample complexity remains almost constant even though the number of groups in a cluster increases. Next, we develop a matrix completion algorithm that starts with hierarchical graph clustering, which produces an exact recovery of clusters yet almost exact recovery of groups. Then, rating vectors are estimated followed by iterative local refinement of groups. We conduct extensive experiments to demonstrate that the optimal sample complexity is achieved by the proposed algorithm, which is a practically-appealing contribution.

A preliminary version of the main results of this paper has been reported in [28] for $(c, g, r, q) = (2, 3, 2, 2)$ and [29] for any $(c, g, r, q)$, where $q$ denotes the order of the finite field from which the rating matrix entries are selected. In this paper, we characterize the optimal sample complexity result for any $(c, g, r, q)$, and present the complete achievability and converse proofs. Furthermore, we propose an algorithm that achieves the optimal sample complexity for all $(c, g, r, q)$. While numerous low-complexity matrix completion algorithms have been proposed, it remains an open problem to develop optimization algorithms with provable performance guarantees for a generic class of matrices [30]. This work makes substantial progress on this long-standing open problem. We also emphasize on the fact that this work is a non-trivial extension of [24] and [25], as will be delineated in the following sections.

### A. Related Works

*1) Connection to Low-Rank Matrix Completion Problems:* The objective of low-rank matrix completion, a recurring problem in collaborative filtering [1], is to recover an unknown low-rank matrix from partial, and possibly noisy, sampling of its entries [30]. Since the rank minimization problem is NP-hard, accurate reconstruction is generally ill-posed and computationally intractable. However, exploiting the fact that the structure of the matrix is of low-rank makes the exploration for a solution worthwhile. One direction of research is geared towards studying low-rank matrix completion where the observed subset of matrix entries is exactly known. Under certain conditions, upper bounds on the number of observed entries, which are uniformly drawn at random, are developed to ensure successful reconstruction with high probability [31], [32], [33]. A fundamental open question in the literature of low-rank matrix completion with exact observation is how to find a low-rank matrix that is consistent with the partial observation of its entries. This question stems from the fact that the sparse basis of the low-rank matrix is unknown and that the basis is drawn from a continuous space. The performance guarantees provided by existing algorithms only hold when certain incoherence assumptions on the singular vectors of the matrix are satisfied. By and large, theoretical guarantees on the reconstruction performance are not established even for the rank-one case, and hence, our understanding of the problem is far from complete. Numerous algorithms for low-rank matrix completion have been proposed over the years. If the rank information of the original low-rank matrix is unknown, various techniques based on nuclear norm minimization are proposed [31], [32], [33], [34], [35], [36], [37]. On the other hand, if the rank is known in advance, techniques based on Frobenius norm minimization are proposed [38], [39], [40], [41], [42], [43], [44], [45]. Another interesting and practical research direction is investigating low-rank matrix completion when the observed entries are contaminated by noise. The objective is to seek a low-rank matrix that best approximates the original matrix, and find an upper bound on the root-mean squared error [32], [46].

In this work, we also consider a low-rank matrix completion problem which has been an open problem for decades. Even for simple settings, such as rank-1 or rank-2 matrices, the optimal sample complexity has been open for decades, although some upper and lower bounds are derived. The matrix of our consideration in this work is of rank $r \geq 2$. Thus, we make progress on this long-standing open problem by exploiting the structural property posed by our considered application.

*2) Algorithms for Recommender Systems With Graph Side Information:* The idea of exploring the value of incorporating graph side information into collaborative filtering approaches has gained a lot of attention from the research community [3]. There are two primary approaches of collaborative learning [1]: (i) latent factor approach, and (ii) neighborhood approach. Latent factor approach learns latent features for users and items from the observed ratings. Most successful realizations of this approach hinge on matrix factorization which characterizes the latent characteristics of users and items by two low-rank user and item-feature matrices inferred from the rating patterns. One direction to integrate graph side information in this approach is by adding some regularization terms to the loss function of the matrix factorization model [4], [5], [6], [7], [8]. Another direction is to develop matrix factorization frameworks that fuse the user-item rating matrix with the social network of the users [9], [10], [11], [12]. Moreover, a robust online matrix completion on graphs is designed and analyzed in [13] that exploits the graph information to recover the incomplete rating matrix entries in the presence of outlier noise. On the other hand, for the neighborhood approach, the prediction of rating information is based on computing the relationships among items or users. The recommendation accuracy in this approach can be enhanced by incorporating the information provided by the social graphs into the neighborhood definition [14], [15], [16], [17], [18], [19]. Lately, recent works have proposed novel architectures for graph convolutional neural networks that fully exploit the structure of item/user graphs [20], [21].

Few works in the literature have provided theoretical insights on the usefulness of side information for the matrix completion problem, e.g., [22] and [23]. Chiang et al. [22] proposed a dirty statistical model to exploit the feature-based side information, yet to be robust to feature noise, in matrix completion applications. They provided theoretical guarantees that the proposed model achieves lower sample complexity than the standard matrix completion (with no graph information) under the condition that the features are not too noisy. Rao et al. [23] proposed a scalable graph regularized matrix completion, and derived consistency guarantees to demonstrate the gain due to the graph side information. It is worth mentioning that the maximal achievable gain due to graph side information is not characterized in these works. Prior works [24], [25] revealed the information-theoretic limit, though the assumption that users in a cluster share the same rating vector limits the practicality of the considered model. We relax the assumption on the rank of the rating matrix by considering a more generalized framework, which is a rank-r matrix.

*3) Connection to Community Detection in Stochastic Block Model:* The statistical model that we consider for the theoretical analysis of our proposed algorithm relies on the hierarchical SBM, which has been shown to well respect many practically-relevant scenarios [47], [48]. The proposed algorithm builds in part upon prominent clustering [49], [50] and hierarchical clustering [51], [52] algorithms, although it exhibits a notable distinction in other matrix-completion-related procedures.

*4) Connection to Clustering Problems With Side Information:* Recently, some works explored a dual problem where clustering is performed with a partially observed matrix as side information. Ashtiani et al. [53] proved that having few pairwise queries leads to more efficient k-means clustering, which is NP-hard in general. Mazumdar and Saha [54] explored the benefits of the similarity matrix, which is used to cluster similar points together, to reduce the adaptive query complexity. In both works, information-theoretic lower bounds are proved, and efficient clustering algorithms are designed. The matrix completion problem with graph side information can be seen as a natural extension to the clustering problem if we shift our focus to recovering the structure of (hierarchical) clusters instead of reconstructing the rating matrix. While the focus of [54] is on finding the clusters, we are interested in revealing the structures of groups and matrix completion.

### *B. Notation*

Row vectors and matrices are denoted by lowercase letters (e.g., $v$) and uppercase letters (e.g., $X$), respectively. Random matrices are denoted by boldface uppercase letters (e.g., $\mathbf{X}$), while their realizations are denoted by uppercase letters (e.g., $X$). Sets are denoted by calligraphic letters (e.g., $\mathcal{Z}$). Let $\mathbb{F}_q$ be a finite field of order $q$ for some prime number $q$. Let $\mathbf{0}_{n\times m}$ and $\mathbf{1}_{n\times m}$ be all-zero and all-one matrices of dimension $n \times m$, respectively. For a matrix $X \in \mathbb{F}_q^{n\times m}$, let $X^\intercal$ denote the transpose of $X$. Let $X(r,t)$ denote the matrix entry at row $r$ and column $t$. Furthermore, let $X(i,:)$ and $X(:,j)$ denote the $i$th row and $j$th column of matrix $X$, respectively. Furthermore, for sets $\mathcal{I}$ and $\mathcal{J}$, the submatrix of $X$, that is obtained by rows $i \in \mathcal{I}$ and columns $j \in \mathcal{J}$, is denoted by $X(\mathcal{I},\mathcal{J})$. Let $\Lambda(X,Y)$ denote the number of different entries between matrices $X$ and $Y$ for $X, Y \in \mathbb{F}_q^{n\times m}$. For $u, v \in \mathbb{F}_q^{1\times m}$, we use $[u;\ v]$ to denote $[u^\intercal\ v^\intercal]^\intercal$. Let $v(t)$

denote the $t$th entry of $v$. Moreover, the Hamming distance between two vectors $u$ and $v$ is denoted by $d_{\mathrm{H}}(u, v)$. It is defined as the number of entries in which $u$ and $v$ differ. Let $\|u\|_1$ denote the $\ell_1$ norm of vector $u$. For an integer $n \geq 1$, let $[n]$ denote the set of integers $\{1, 2, \ldots, n\}$. For integers $n$ and $m$ where $n \leq m$, define $[n : m]$ as $\{n, n+1, \ldots, m\}$. Let $\{0, 1, \ldots, q-1\}^n$ be the set of all $n$-digit sequences whose digits are drawn from $\mathbb{F}_q$. Moreover, we use $\mathbb{1}[\cdot]$ to denote the indicator function. Finally, $\mathcal{G} = ([n], \mathcal{E})$ denotes an undirected graph $\mathcal{G}$ where $[n]$ is the set of $n$ vertices labeled by integers in $[n]$; and $\mathcal{E}$ is the set of undirected edges. The elements of $\mathcal{E}$ are given by pairs $(a, b)$ for $a, b \in [n]$.

### C. Paper Outline

The remainder of this paper is organized as follows. We first present the problem formulation of the rating matrix and hierarchical similarity graph in Section II. Then, the main results and implications of the results are presented in Section III. Next, the achievability proof is presented in Section IV, while the converse proof is presented in Section V. In Section VI, we show that the proposed algorithm achieves the optimal sample complexity via experimental results. Finally, the paper is concluded and directions for future research are discussed in Section VII.

## II. Problem Formulation

Consider a rating matrix $X \in \mathbb{F}_q^{n \times m}$ where $n$ denotes the number of users and $m$ denotes the number of items. The ratings of the $r$th user over $m$ items are given by the rating vector that corresponds to the $r$th row of $X$ for $r \in [n]$. The user similarity graphs (e.g., social graphs) are leveraged as side information during the matrix completion procedure to enhance the item recommendation quality. More specifically, we consider a hierarchical similarity graph that consists of $c$ disjoint clusters, and each cluster comprises $g$ disjoint sub-clusters (or that we call "groups"). For the sake of tractable mathematical analysis, we assume equal-sized clusters and groups. The theoretical guarantees, however, hold as long as the group sizes are order-wise same (see Section III). According to social homophily theory [2], users within the same community (i.e., who are more likely to be connected in a social graph) are more likely to share similar preferences of items. This results in a low-rank structure of the rating matrix since the rows of the rating matrix associated with such users are highly likely to be similar [30]. To capture this crucial fact in our model, we make the following assumptions: (i) the rating vectors of the users who belong to the same group are equal, and hence there are $gc$ distinct rating vectors in total; (ii) the rating vectors of the groups of a given cluster are different, yet intimately-related to each other through a linear subspace of $r$ basis vectors for some integer $r \leq g$. Let $v_i^{(x)}$ denote the rating vector of the users in cluster $x$ and group $i$ for $x \in [c]$ and $i \in [g]$. Let $R^{(x)} \in \mathbb{F}_q^{g \times m}$ denote a matrix whose rows are the rating vectors of the groups in cluster $x$ for $x \in [c]$. The set of $g$ rows of $R^{(x)}$ (i.e., set of $g$ rating vectors of the groups in cluster $x$) is spanned by any subset of $r$ rows of $R^{(x)}$.

Let $X_0$ denote the ground truth rating matrix. Each instance of the problem corresponds to a rating matrix $X_0$. Equivalently, $X_0$ can be represented by a set of rating vectors $\mathcal{V}_0 = \{u_i^{(x)} : x \in [c], i \in [g]\}$ and a user partitioning $\mathcal{Z}_0$. For instance, consider a problem with $n = 12$ users in $c = 2$ clusters and $g = 3$ groups. If the rating matrix is given by

$$X_0 = \Big[u_1^{(1)};\ u_2^{(1)};\ u_1^{(2)};\ u_3^{(2)};\ u_3^{(1)};\ u_2^{(1)};\ u_2^{(2)};\ u_1^{(2)};\ u_3^{(1)}; \quad u_3^{(2)};\ u_2^{(2)};\ u_1^{(1)}\Big], \tag{1}$$

then we have the set of rating vectors as

$$\mathcal{V}_0 = \left\{u_1^{(1)},\ u_2^{(1)},\ u_3^{(1)},\ u_1^{(2)},\ u_2^{(2)},\ u_3^{(2)}\right\}, \tag{2}$$

and the user partitioning as

$$\mathcal{Z}_0 = \{Z_0(1,1)=\{1,12\},\ Z_0(1,2)=\{2,6\},\ Z_0(1,3)=\{5,9\}, \quad Z_0(2,1)=\{3,8\},\ Z_0(2,2)=\{7,11\},\ Z_0(2,3)=\{4,10\}\}. \tag{3}$$

Formally, $\mathcal{Z}_0$ is a family of subsets of $[n]$ that partitions the set of all users $[n]$ into $c$ clusters and $g$ groups (per cluster). That is,

$$\mathcal{Z}_0 = \Big\{ \{Z_0(x,i)\}_{x \in [c],\, i \in [g]} : Z_0(x,i) \subseteq [n], \quad Z_0(x,i) \cap Z_0(y,j) = \varnothing \text{ for } (x,i) \neq (y,j), \quad \bigcup_{x \in [c]} \bigcup_{i \in [g]} Z_0(x,i) = [n] \Big\}, \tag{4}$$

where $Z_0(x,i)$ denotes the set of users who belong to cluster $x$ and group $i$ for $x \in [c]$ and $i \in [g]$.

The main goal of the problem of interest is to find the best estimate of $X_0$ with the knowledge of two types of observations:

1) partial and noisy observation $Y$ of $X_0$. For every $r \in [n]$ and $t \in [m]$, let $Y(r,t) \in \mathbb{F}_q \cup \{*\}$, where $*$ denotes no observation. Let the set of observed entries of $X_0$ be denoted by $\Omega = \{(r,t) \in [n] \times [m] : Y(r,t) \neq *\}$. The partial observation is modeled by assuming that each entry of $X_0$ is observed with probability $p \in [0,1]$, independently from others. Moreover, the potential noise in the observation is modeled by considering a random uniform noise distribution, that is, the noise is not adversarial (i.e., not deterministic). In particular, we assume that each observed entry $X_0(r,t)$, for $(r,t) \in \Omega$, can be possibly flipped to any element of the set $\{0, 1, \ldots, q-1\} \setminus X_0(r,t)$ with a uniform probability of $\theta/(q-1)$ for $\theta \in [0, (q-1)/q)$. The reasons behind choosing this noise model are: (i) the uniform noise distribution is the worst-case distribution in discrete channels; and (ii) this model captures the fact that there may exist a fraction of group members whose ratings are close to the majority ratings, yet they are not exactly identical. Hence, the majority ratings can be considered as the ground truth, while the ratings of this fraction of users can be seen as noisy versions of the ground truth. Furthermore, since this fraction of users have some ratings that are different from the

majority, each of such ratings can take a value that is randomly and uniformly selected from the set of all possible ratings different from that of the majority;

2) user similarity graph $\mathcal{G} = ([n], \mathcal{E})$. A vertex represents a user, and an edge captures a social connection between two users. The set $[n]$ of vertices is partitioned into $c$ disjoint clusters, each of which is of size $n/c$ users. Each cluster is further partitioned into $g$ disjoint groups, each of which is of size $n/(cg)$ users. The user similarity graph is generated as per the hierarchical stochastic block model (HSBM) [51], [55], which is a generative model for random graphs exhibiting hierarchical cluster behavior. In this model, every two nodes in the graph are connected by an edge, independent of all other nodes, with probabilities given by

$$\begin{aligned}
\widetilde{\alpha} &:= \mathbb{P}\left[(a,b) \in \mathcal{E} : a, b \in Z_0(x,i)\right], \text{ for } x \in [c],\ i \in [g], \\
\widetilde{\beta} &:= \mathbb{P}\left[(a,b) \in \mathcal{E} : a \in Z_0(x,i), b \in Z_0(x,j)\right], \\
&\quad \text{for } x \in [c],\ i,j \in [g],\ i \neq j, \\
\widetilde{\gamma} &:= \mathbb{P}\left[(a,b) \in \mathcal{E} : a \in Z_0(x,i), b \in Z_0(y,j)\right], \\
&\quad \text{for } x,y \in [c],\ x \neq y,\ i,j \in [g].
\end{aligned} \tag{5}$$

Here, we assume that edge probabilities are scaling with the size of the problem. In particular, we assume

$$\widetilde{\alpha} = \alpha \frac{\log n}{n}, \quad \widetilde{\beta} = \beta \frac{\log n}{n}, \quad \widetilde{\gamma} = \gamma \frac{\log n}{n}, \tag{6}$$

where $\alpha$, $\beta$ and $\gamma$ are positive real numbers such that $\alpha \geq \beta \geq \gamma$. In other words, there is an edge between two users in the *same group* within a cluster with probability $\alpha\frac{\log n}{n}$; there is an edge between two users in *different groups* but within the *same cluster* with probability $\beta\frac{\log n}{n}$; and there is an edge between two users in *different clusters* with probability $\gamma\frac{\log n}{n}$. Here, $\frac{\log n}{n}$ refers to a scaling factor implying that the term becomes relatively smaller as the number of nodes $n$ increases, signifying that the growth rate of edges in the graph is slower compared to the logarithm of the number of nodes [56]. Note that the considered edge probabilities guarantee the disappearance of isolated vertices (i.e., vertices of degree zero) in the user similarity graph, which is a necessary property for exact recovery in the stochastic block model (SBM). Furthermore, motivated by the social homophily theory [2], we study the problem of interest when users within the same group (or cluster) are more likely to be connected than those in different groups (or clusters). That is why we assume that $\alpha \geq \beta \geq \gamma$.

Let $\psi$ denote an estimator (decoder) that takes as input a pair $(Y, \mathcal{G})$ where $Y$ is the incomplete and noisy rating matrix and $\mathcal{G}$ is the user similarity graph and outputs a completed rating matrix $X \in \mathbb{F}_q^{n \times m}$. Note that both the set of rating vectors $\mathcal{V}$ and the user partitioning[1] $\mathcal{Z}$ can be recovered from the completed rating matrix $X$ and vice versa. Hence, with a slight abuse of notation, we may interchangeably use $X$ or $(\mathcal{V}, \mathcal{Z})$ as the output of the estimator. The former notation is adopted when we are interested in the entries of the rating matrix, while the latter notation is used when we are interested in either the set of rating vectors or the user partitioning.

[1] To be more precise, $\mathcal{Z}$ can be recovered from $X$ up to a relabeling of the clusters and groups, i.e., we can only identify which users belong to the same group/cluster, but the label associated to a group/cluster cannot be identified.

One key parameter that is instrumental in both expressing the main result (see Section III) as well as proving the main theorem is the discrepancy between the rating vectors. Let $\delta_g$ be the minimum normalized Hamming distance among distinct pairs of rating vectors of groups *within the same cluster*. Let $\delta_c$ be the counterpart with respect to distinct pairs of rating vectors *across different clusters*. More formally, $\delta_g$ and $\delta_c$ are given by

$$\begin{aligned}
\delta_g &= \frac{1}{m} \min_{x \in [c]} \min_{\substack{i,j \in [g] \\ i \neq j}} d_{\mathrm{H}}\left(u_i^{(x)}, u_j^{(x)}\right), \\
\delta_c &= \frac{1}{m} \min_{\substack{i,j \in [g] \\ x,y \in [c], x \neq y}} d_{\mathrm{H}}\left(u_i^{(x)}, u_j^{(y)}\right).
\end{aligned} \tag{7}$$

As will be elaborated on in the next section, our result hinges on $\delta := (\delta_g, \delta_c)$. We provide theoretical guarantees for the recovery of all rating matrices $M$ in which the rating vectors maintain a minimum level of dissimilarity. Formally, we define $\mathcal{M}^{(\delta)}$ to be the set of rating matrices $M = (\mathcal{V}, \mathcal{Z})$ such that the following properties are satisfied:

- the set of rating vectors $\mathcal{V}$ must satisfy the property that the minimum normalized Hamming distance among the rating vectors in different groups within the same cluster and those in different clusters are not smaller than $\delta_g$ and $\delta_c$, respectively;
- the user partitioning $\mathcal{Z}$ must satisfy the property that the size of clusters is $n/c$ users, while the size of the groups is $n/(gc)$ users.

The performance metric we consider to provide theoretical guarantees on the quality of recommendation is the worst-case probability of error $P_e$. In other words, the quality of the estimator is defined by its accuracy of estimation of the *most difficult* ground truth matrix $M = (\mathcal{V}, \mathcal{Z}) \in \mathcal{M}^{(\delta)}$. Therefore, we apply a minimax optimization approach wherein the objective is to find the estimator that minimizes the maximum risk (i.e., minimizes the worst-case probability of error). This can be expressed as

$$\inf_{\psi} P_e^{(\delta)}(\psi) = \inf_{\psi} \max_{M \in \mathcal{M}^{(\delta)}} \mathbb{P}\left[\psi(Y, \mathcal{G}) \neq M\right]. \tag{8}$$

Based on the proposed problem formulation and performance metric, we aim at characterizing the optimal sample complexity (i.e., the minimum number of entries of the rating matrix that is required to be observed). For $X_0 \in \mathbb{R}_q^{n \times m}$, this number is concentrated around $nmp^\star$ in the limit of $n$ and $m$, for exact rating matrix recovery. Here, $p^\star$ denotes a sharp threshold on the observation probability such that the following conditions, in the limit of $n$ and $m$, are satisfied:

- when $p > p^\star$, there exists an estimator such that the error probability can be made arbitrarily close to 0;
- when $p < p^\star$, the error probability does not converge to zero no matter what and whatsoever.

$$p^{\star}=\frac{1}{\left(\sqrt{1-\theta}-\sqrt{\frac{\theta}{q-1}}\right)^{2}}\max\left\{\frac{gc}{g-r+1}\frac{\log m}{n},\ \frac{\log n}{\delta_g m}\left(1-\frac{\left(\sqrt{\alpha}-\sqrt{\beta}\right)^{2}}{gc}\right),\ \frac{\log n}{\delta_c m}\left(1-\frac{\left(\sqrt{\alpha}-\sqrt{\gamma}\right)^{2}+(g-1)\left(\sqrt{\beta}-\sqrt{\gamma}\right)^{2}}{gc}\right)\right\}. \tag{9}$$

$$p^{\star}=\frac{1}{\left(\sqrt{1-\theta}-\sqrt{\theta}\right)^{2}}\max\left\{3\,\frac{\log m}{n},\ \frac{\log n}{\delta_g m}\left(1-\frac{\left(\sqrt{\alpha}-\sqrt{\beta}\right)^{2}}{6}\right),\ \frac{\log n}{\delta_c m}\left(1-\frac{\left(\sqrt{\alpha}-\sqrt{\gamma}\right)^{2}}{6}-\frac{\left(\sqrt{\beta}-\sqrt{\gamma}\right)^{2}}{3}\right)\right\}. \tag{10}$$

## III. Main Results

### A. Information-Theoretic Limits

As in [28], we assume that $m = \omega(\log n)$ and $\log m = o(n)$ hold in order to ease the proof via large deviation theories. These assumptions are also practically relevant as they rule out the possibility of having highly asymmetric matrices (i.e., extremely tall and wide matrices).

*Theorem 1 (Optimal Sample Complexity):* Let $m=\omega(\log n)$ and $\log m = o(n)$. Also, consider $(q, \theta, c, g, r)$ to be constants such that $q$ is prime, $\theta \in [0, (q-1)/q)$, and $r \leq g$. Let $p^{\star}$ be given by (9), shown at the top of the page.

For any constant $\epsilon > 0$, if $p \geq (1+\epsilon)p^{\star}$, there exists an estimator $\psi$ that outputs a rating matrix $X \in \mathcal{M}^{(\delta)}$ given $Y$ and $\mathcal{G}$ such that $\lim_{n\to\infty} P_e^{(\delta)}(\psi) = 0$. Conversely, if $p \leq (1-\epsilon)p^{\star}$, then $\lim_{n\to\infty} P_e^{(\delta)}(\psi) \neq 0$ for any estimator $\psi$.

*Remark 1 (Technical Novelty):* The technical distinctions with respect to the prior works [24], [25], [28] are four-folded. First, the likelihood computation requires more involved combinatorial arguments due to the hierarchical structure of the similarity graph; see Lemma 1 in Section IV. Second, more sophisticated upper and lower bounding techniques that exploit the relational structure across different groups are developed in Lemmas 3 and 4. Next, novel typical and atypical error analyses are proposed in Lemmas 5 and 6. Finally, novel failure proof techniques are presented in Section V.

The following remark demonstrates that the problem setting considered in [28] is a special case of the general setting considered in this paper, and their result is subsumed by our generalized result presented in Theorem 1.

*Remark 2:* Setting $(c, g, r, q) = (2, 3, 2, 2)$, the optimal observation probability $p^{\star}$ reduces to (10), shown at the top of the page, which is equal to the shape threshold on $p$ characterized[2] by [28] under the setting of $(c, g, r, q) = (2, 3, 2, 2)$.

### B. Implications of Theorem 1

We investigate the relationship between the optimal sample complexity $nmp^{\star}$, where $p^{\star}$ is characterized by (9) in Theorem 1, and different parameters related to the rating matrix as well as the hierarchical user similarity graph.

*Remark 3:* The optimal sample complexity increases as $\delta_g$ (or $\delta_c$) decreases. This is due to the fact that as the Hamming distance between rating vectors of two users in different groups within the same cluster (or in different clusters) decreases, it becomes harder to distinguish the rating vectors, and hence it leads to imperfect user grouping (or clustering). Thus, one has to sample more entries of the rating matrix in order to exactly recover the rating matrix.

*Remark 4:* It is evident from (9) that the optimal sample complexity increases as $\theta$ grows. Furthermore, as $\theta$ approaches $(q-1)/q$, each sampled entry of the rating matrix can take any of the $q$ possible values with a uniform probability of $1/q$, and hence an infinite sample complexity is theoretically required to exactly recover the entries of the rating matrix.

*1) Quality of the Hierarchical Similarity Graph:* In order to better illustrate the relationship between the optimal sample complexity and the quality of the hierarchical graph, we define the following quality parameters:

$$I_{\alpha,\beta} := \left(\sqrt{\alpha}-\sqrt{\beta}\right)^{2}, \quad I_{\alpha,\gamma} := \left(\sqrt{\alpha}-\sqrt{\gamma}\right)^{2}, \quad I_{\beta,\gamma} := \left(\sqrt{\beta}-\sqrt{\gamma}\right)^{2}. \tag{11}$$

Intuitively, as $I_{\alpha,\beta}$ increases, it becomes easier to distinguish users in different groups within the same cluster. On the other hand, larger values of $I_{\alpha,\gamma}$ and $I_{\beta,\gamma}$ lead to better user clustering. The optimal sample complexity reads different values depending on the quality parameters of the hierarchical graph. More specifically, we define three regimes as follows:

1) the first term in the right-hand-side (RHS) of (9) is activated when $I_{\alpha,\beta}, I_{\alpha,\gamma}$ and $I_{\beta,\gamma}$ are large enough so that the grouping and clustering information is reliable. Hence, this regime is coined as *"perfect clustering/grouping regime"*;
2) the second term in the RHS of (9) is activated when $I_{\alpha,\beta}$ is small such that the grouping information is not reliable. Therefore, this regime is coined as *"grouping-limited regime"*;
3) the third term in the RHS of (9) is activated when $I_{\alpha,\gamma}$ and $I_{\beta,\gamma}$ are small such that the clustering information is not reliable, and[3] $\delta_g > \delta_c$. Thus, this regime is coined as *"clustering-limited regime"*.

In what follows, we analyze the optimal sample complexity under each regime and highlight the novel technical contributions in the achievability proof (see Section IV) as well as the converse proof (see Section V). For illustrative simplicity, we focus on the noiseless case where $\theta = 0$.

*Remark 5:* (Perfect Clustering/Grouping Regime) The optimal sample complexity reads $(gc/(g-r+1))m\log m$. Since

[2]The optimal sample complexity for general $(c, g, r, q)$, given by (9), is conjectured by [28]. However, the achievability and converse proofs are provided only for $(c, g, r, q) = (2, 3, 2, 2)$ in [28]. In this paper, we present complete achievability and converse proofs for any $(c, g, r, q)$.

[3]It is evident from (9) and $\alpha \geq \beta \geq \gamma$ that the third term in the RHS of (9) is inactive whenever $\delta_g \leq \delta_c$.

the grouping and clustering information is reliable, groups and clusters can be recovered from the similarity graph. However, further increments of the values of these quality parameters do not yield further improvement in the sample complexity, and hence, the sample complexity gain from the similarity graph is saturated in this regime. Moreover, it should be noted that a naive generalization of [24], [25] requires $crm \log m$ observations since there are $r$ independent rating vectors to be estimated for each of the $c$ clusters, and each rating vector requires $m \log m$ observations under the considered random sampling due to the coupon-collecting effect. On the other hand, we leverage the relational structure (i.e., linear dependency) across rating vectors of different groups, reflected by the underlying linear MDS code structure (to be detailed in Section IV), and hence this serves to estimate the $rc$ rating vectors more efficiently, precisely by a factor of $r(g-r+1)/g$ improvement, thus yielding $(gc/(g-r+1))m \log m$.

*Remark 6:* (Grouping-Limited Regime) The optimal sample complexity reads

$$\frac{1}{\delta_g}\left(1-\frac{\left(\sqrt{\alpha}-\sqrt{\beta}\right)^2}{gc}\right) n \log n \;=\; \frac{1}{\delta_g}\left(1-\frac{I_{\alpha,\beta}}{gc}\right) n \log n,$$

which is a decreasing function of $I_{\alpha,\beta}$. This sample complexity coincides with that of [25] in which the considered similarity graph consists of only $gc$ clusters. This implies that exploiting the relational structure across different groups does not help improve sample complexity when grouping information is not reliable. Furthermore, since the clustering information is reliable, clusters can be recovered from the similarity graph. However, further increments of $I_{\alpha,\gamma}$ and $I_{\beta,\gamma}$ do not yield further reduction in the sample complexity, and hence the sample complexity gain from these two quality parameters is saturated in this regime.

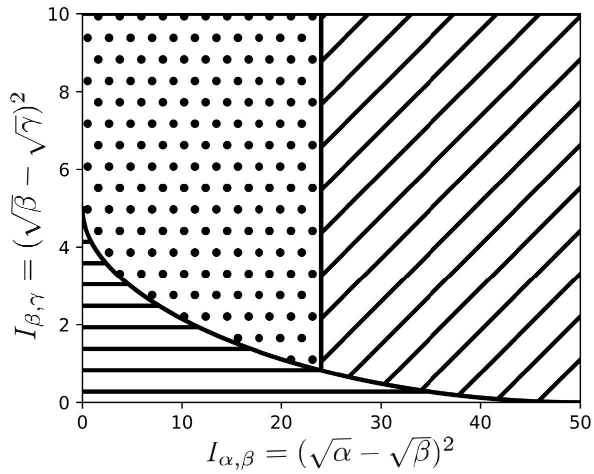
*Remark 7:* (Clustering-Limited Regime) The optimal sample complexity reads

$$\frac{1}{\delta_c}\left(1-\frac{\left(\sqrt{\alpha}-\sqrt{\gamma}\right)^2+(g-1)\left(\sqrt{\beta}-\sqrt{\gamma}\right)^2}{gc}\right) n \log n$$
$$=\frac{1}{\delta_c}\left(1-\frac{I_{\alpha,\gamma}+(g-1)I_{\beta,\gamma}}{gc}\right) n \log n,$$

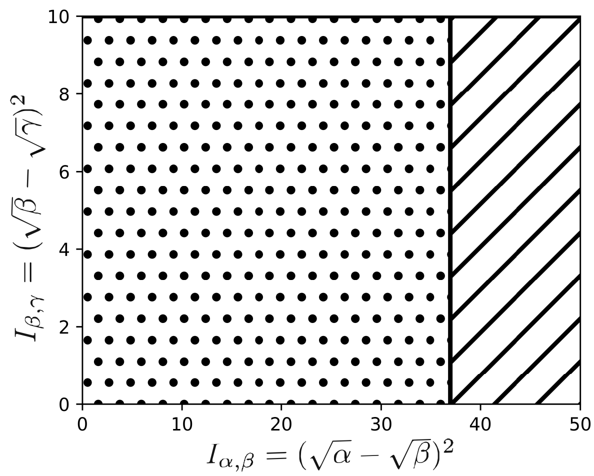
which is a decreasing function of $I_{\alpha,\gamma}$ and $I_{\beta,\gamma}$. This is the most challenging scenario which has not been explored by any prior works. Since the clustering information is not reliable, it is not possible to recover the groups and clusters from the similarity graph. Moreover, it should be noted that when $\beta=\gamma$, i.e., groups and clusters are indistinguishable, we have $I_{\alpha,\beta}=I_{\alpha,\gamma}$ and $I_{\beta,\gamma}=0$. As a result, it boils down to a problem setting of $gc$ clusters, and hence the optimal sample complexity reads

$$\frac{1}{\delta_c}\left(1-\frac{\left(\sqrt{\alpha}-\sqrt{\gamma}\right)^2}{gc}\right) n \log n = \frac{1}{\delta_c}\left(1-\frac{I_{\alpha,\beta}}{gc}\right) n \log n,$$

Compared to the optimal sample complexity expression for the grouping-limited regime, the only distinction appears in the denominator, in which $\delta_g$ is replaced with $\delta_c$ due to the fact that $\delta_c < \delta_g$.




(a) $\delta_g = \frac{1}{3}$ and $\delta_c = \frac{1}{6}$.




(b) $\delta_g = \frac{1}{6}$ and $\delta_c = \frac{1}{3}$.

Fig. 1. Let $(n, m, \theta, c, g, r, q) = (4000, 500, 0, 10, 5, 3, 5)$. (a) The different regimes of the optimal sample complexity reported in (9) for $\delta_g > \delta_c$. (b) The different regimes of the optimal sample complexity reported in (9) for $\delta_g < \delta_c$. For both sub-figures, diagonal stripes, dots, and horizontal stripes refer to the perfect clustering/grouping, grouping-limited, and clustering-limited regime, respectively.

Consider a problem setting where $n=4000$, $m=500$, $\theta=0$, $c=10$, $g=5$, $r=3$ and $q=5$. Fig. 1a and Fig. 1b depict the different regimes of the optimal sample complexity as a function of $(I_{\alpha,\beta}, I_{\beta,\gamma})$. In Fig. 1a, where $\delta_g=1/3$ and $\delta_c=1/6$, the region depicted by diagonal stripes corresponds to the perfect clustering/grouping regime and the first term in the RHS of (9) is active. The graph quality parameters $I_{\alpha,\beta}$, $I_{\beta,\gamma}$, and consequently $I_{\alpha,\gamma}$ are large, and graph information is rich enough to perfectly retrieve the clusters and groups. The region depicted by dots corresponds to the grouping-limited regime, where the second term in the RHS of (9) is active. In this regime, graph information suffices to exactly recover the clusters, but we need to rely on rating observation to exactly recover the groups. Finally, the third term in the RHS of (9) is active in the region captured by horizontal stripes. This indicates the clustering-limited regime, where neither clustering nor grouping is exact without the side information of the rating vectors. On the other hand, Fig. 1b depicts the case where $\delta_g=1/6$ and $\delta_c=1/3$. It is worth noting that

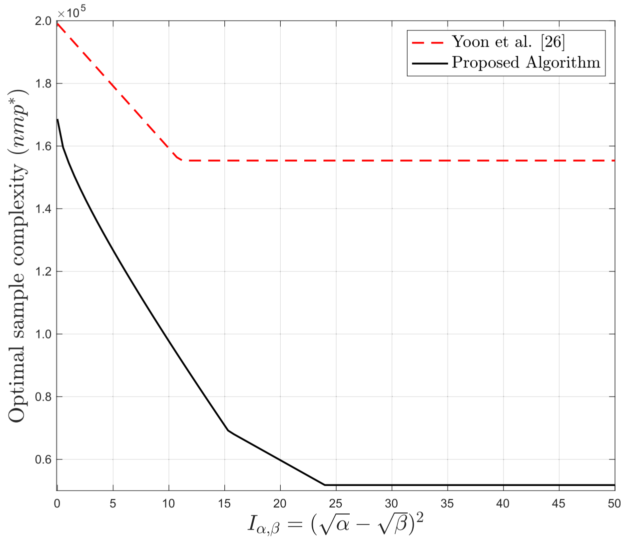



Fig. 2. Let $(n, m, \theta, c, g, r, q) = (4000, 500, 0, 10, 5, 3, 5)$. Comparison between the sample complexity reported in (9) and that of [25] for $\beta = 5$, $\gamma = 1$, $\delta_g = 1/3$ and $\delta_c = 1/6$.

in practically-relevant systems, we have $\delta_g < \delta_c$, i.e., rating vectors of users in the same cluster are expected to be more similar than those in a different cluster. Therefore, the third regime (i.e., clustering-limited regime) vanishes in Fig. 1b.

*2) Benefit of Hierarchical Graph Structure:* Consider a problem setting where $n = 4000$, $m = 500$, $\theta = 0$, $c = 10$, $g = 5$, $r = 3$ and $q = 5$. Fig. 2 compares the optimal sample complexity, as a function of $I_{\alpha,\beta}$, between the one reported in (9) and that of [25] for $\delta_g = 1/3$, $\delta_c = 1/6$, $\beta = 5$ and $\gamma = 1$. It should be noted that [25] leverages neither the hierarchical structure of the graph nor the linear dependency among the rating vectors. Thus, the problem formulated in Section II will be translated to a graph that consists of $gc$ clusters whose rating vectors are linearly independent in the setting of [25]. Also, note that the minimum Hamming distance for [25] is $\delta_c$. In Fig. 2, we can see that the noticeable gain in the sample complexity of our result in the diagonal parts of the plot (i.e., clustering-limited and grouping-limited regimes on the left side) is due to leveraging the hierarchical graph structure, while the improvement in sample complexity in the flat part of the plot (i.e., perfect clustering/grouping regime) is a consequence of exploiting the relational structure (i.e., linear dependency) among the rating vectors within each cluster.

## IV. The Achievability Proof

In this section, we prove the achievability part of Theorem 1, that is if the condition on $p$ in (9) holds, then there exists an estimator $\psi$ such that $\lim_{n\to\infty} P_e^{(\delta)}(\psi) = 0$. To this end, we prove that $\lim_{n\to\infty} P_e^{(\delta)}(\psi_{\text{ML}}) = 0$ where $\psi_{\text{ML}}$ is the maximum likelihood (ML) estimator if all the following inequalities hold:

**Perfect Clustering/Grouping Regime:**

$$\frac{g-r+1}{gc} nI_r \ \geq\ (1+\epsilon)\log m, \tag{12}$$

**Grouping-Limited Regime:**

$$\delta_g m I_r + \frac{I_{\alpha,\beta}}{gc}\log n \ \geq\ (1+\epsilon)\log n, \tag{13}$$

**Clustering-Limited Regime:**

$$\delta_c m I_r + \frac{I_{\alpha,\gamma}}{gc}\log n + \frac{(g-1)I_{\beta,\gamma}}{gc}\log n \ \geq\ (1+\epsilon)\log n, \tag{14}$$

where

$$I_r := p\left(\sqrt{1-\theta} - \sqrt{\frac{\theta}{q-1}}\right)^2, \quad I_{\alpha,\beta} := (\sqrt{\alpha}-\sqrt{\beta})^2,$$
$$I_{\alpha,\gamma} := (\sqrt{\alpha}-\sqrt{\gamma})^2, \quad I_{\beta,\gamma} := (\sqrt{\beta}-\sqrt{\gamma})^2. \tag{15}$$

Throughout the proof, let $p = \Theta((\log n)/n)$, and let $q$ and $\theta$ be constants such that $q$ is prime and $\theta \in [0, 1]$. We first present the structure of the ground truth rating matrix and the underlying linear maximum distance separable code (MDS) code structure in Section IV-A. Next, we introduce a number of auxiliary lemmas in Section IV-B. Finally, we present the achievability proof of Theorem 1 in Section IV-C.

The achievability proof is based on maximum likelihood estimation (MLE). We first evaluate the likelihood for a given clustering/grouping of users and the corresponding rating matrix. Next, we provide an upper bound on the worst-case probability of error, which is given by the probability that the likelihood of the ground truth rating matrix is less than that of a candidate rating matrix. Then, we partition the candidate rating matrices into two sets, typical and atypical sets. A typical (or atypical) set denotes the set of rating matrices that have a relatively small (or large) number of error entries compared to the ground truth matrix. Finally, we conduct typical and atypical error analyses as follows. In the typical error analysis, we provide a tight upper bound on the cardinality of the typical set and a loose upper bound on the error probability of a candidate matrix. On the other hand, in the atypical error analysis, we provide a loose upper bound on the cardinality of the typical set and a tight upper bound on the error probability of a candidate matrix. These analyses are based on the fact that the size of the set candidate matrices with a small number of error entries is relatively larger than the one with a large number of error entries. Based on these bounds, we show that the probability of error for any candidate matrix in the typical set is negligibly smaller than the cardinality of the typical set of matrices, and hence, this leads to convergence of the overall worst-case probability of error to zero as $n$ and $m$ goes to infinity. Hence, the worst-case probability of error vanishes in the limit of $n$ and $m$. This completes the achievability proof.

### A. The Structure of Ground Truth Rating Matrix

For $X_0 \in \mathcal{M}^{(\delta)}$, let $X_0 = (\mathcal{V}_0, \mathcal{Z}_0)$ denote the ground truth rating matrix, where

$$\mathcal{V}_0 = \left\{u_i^{(x)} : x \in [c], i \in [g]\right\}, \ \mathcal{Z}_0 = \left\{\{Z_0(x,i)\}_{x\in[c],\, i\in[g]}\right\}, \tag{16}$$

and $\mathcal{Z}_0$ follows the conditions given in (4), but omitted for the sake of brevity. For $x \in [c]$, let $R_0^{(x)} \in \mathbb{F}_q^{g\times m}$ be a matrix obtained by stacking all the rating vectors of cluster $x$, i.e., $\{u_i^{(x)} : i \in [g]\}$. Consequently, $X_0$ is an $n \times m$ matrix where its $r$th row is equal to $u_i^{(x)}$ if and only if $r \in Z_0(x,i)$. Furthermore, let the output of an estimator $\psi$ (i.e., the completed rating matrix) be denoted by $X = (\mathcal{V}, \mathcal{Z})$, where $X \in \mathbb{F}_q^{n\times m}$,

$$\mathcal{V} = \left\{ v_i^{(x)} : x \in [c], i \in [g] \right\},\ \ \mathcal{Z} = \left\{ \{Z(x,i)\}_{x\in[c],\, i\in[g]} \right\}, \quad (17)$$

and $\mathcal{Z}$ follows the conditions listed in (4).

We construct a ground truth rating matrix $X_0$ that we are supposed to recover using the maximum likelihood estimator $\psi_{\mathrm{ML}}$. Recall from Section II that the hierarchical graph consists of $c$ clusters, and each cluster comprises $g$ equal-sized groups. The set of $g$ rating vectors of cluster $x$ is spanned by any subset of $r$ rating vectors for $x \in [c]$. Without loss of generality, assume that the set of users who belong to cluster $x$ and group $i$ is given by $\{k{+}1, k{+}2, \ldots, k{+}\frac{n}{cg}\}$ for $x \in [c]$, $i \in [g]$ and $k = (x{-}1)\frac{n}{c} + (i{-}1)\frac{n}{cg}$. Let us consider a $(g, r)$ linear MDS code over the finite field $\mathbb{F}_q$, where $g$ is the length of the code and $r$ is its dimension. This code can be defined as a linear subspace of the vector space $\mathbb{F}_q^g$ with dimension $r$. From the literature of error-correcting codes, a $(g, r)$ linear MDS code in $\mathbb{F}_q$ exhibits a minimum distance of $g - r + 1$, and hence reaches the Singleton bound [57]. Let the set of $g$ ground truth rating vectors of the groups in cluster $x$ be a $(g, r)$ MDS code. Hence, the set of $g$ rows of $R_0^{(x)}$ spanned by any subset of $r$ rows of $R_0^{(x)}$. Let $\Phi^{(x)} \in \mathbb{F}_q^{g\times r}$ be a generator matrix of the $(g, r)$ MDS code, and $W^{(x)} \in \mathbb{F}_q^{r\times m}$ be the basis matrix (with rank $r$), such that

$$R_0^{(x)} = \Phi^{(x)} W^{(x)}, \quad \text{for } x \in [c]. \quad (18)$$

In the remaining of this paper, without loss of generality, we make the following assumptions:

- the first row of $R_0^{(1)}$ be given by $R_0^{(1)}(1,:) = \mathbf{1}_{1\times n}$;
- for $x \in [c]$, let $\Phi^{(x)} = \Phi$ where $\Phi$ is a systematic generator matrix such that $\Phi = [I_{r\times r}\ A^\mathsf{T}]^\mathsf{T}$ and $A \in \mathbb{F}_q^{(g-r)\times r}$. This ensures that the first $r$ rows of $\Phi^{(x)}$ are linearly independent, and hence the first $r$ rows of $R_0^{(x)}$ are linearly independent by (18).

Based on the aforementioned assumptions[4], the entries of each column of $X_0$ can take values from a set of $q^{cr-1}$ possible column vectors. This is due to the fact that the first row of $X_0$ is fixed to all-one vector, and the last $g - r$ rows of $R_0^{(x)}$, for $x \in [c]$, can be constructed by linear combinations of its first $r$ rows. Hence, we have a total of $q^{cg-(1+c(g-r))} = q^{cr-1}$ different choices. Let the set of columns of $X_0$ be partitioned into $q^{cr-1}$ sections, where the columns of each section correspond to one choice of the possible $q^{cr-1}$ vectors. Let the number of columns of each section be $s_\ell m$, where $0 \le s_\ell \le 1$ for $\ell \in \{0, 1, \ldots, q{-}1\}^{cr-1}$ and $\sum_{\ell\in\{0,1,\ldots,q-1\}^{cr-1}} s_\ell = 1$. Let $\mathcal{S}_\ell$ denote the $\ell$th column section of $X_0$, and hence $s_\ell = |\mathcal{S}_\ell|/m$. Accordingly, let each row $u_i^{(x)}$ of $X_0$ be partitioned into $q^{cr-1}$ sections, denoted[5] by $\{u_i^{(x)}(\ell) : \ell \in \{0, 1, \ldots, q{-}1\}^{cr-1}\}$, for $x \in [c]$ and $i \in [g]$. We assume that the MDS code structure is known a priori, and hence, the output matrix follows the MDS code structure imposed on the construction of $X_0$. In the following example, we give an illustrative description of the proposed construction of the ground truth rating matrix $X_0$.

*1) Illustrative Example:* Consider the setting with parameters $(c, g, r, q) = (2, 3, 2, 2)$. Under this setting, the generator matrix and the basis matrix of each cluster are given by (22) and (23), shown at the top of the next page, respectively, in which $0 \le s_\ell \le 1$ for $\ell \in \{0,1\}^3$, and $\sum_{\ell\in\{0,1\}^3} s_\ell = 1$. Therefore, from (18), we obtain (24), shown at the top of the next page, Consequently, $X_0$ is given by (25), shown at the top of the next page, which is the same construction of $X_0$ provided in [28] for the special case of $(c, g, r, q) = (2, 3, 2, 2)$. ♦

### *B. The Auxiliary Lemmas*

We present six auxiliary lemmas used to prove the achievability part of Theorem 1. Before each lemma, we introduce the necessary terminologies and notations.

Let $\mathsf{L}(X)$ denotes[6] the negative log-likelihood of a candidate rating matrix $X = (\mathcal{V}, \mathcal{Z})$ given a fixed input pair $(Y, \mathcal{G})$. More formally, we have

$$\mathsf{L}(X) = \begin{cases} -\log \mathbb{P}\left[(Y, \mathcal{G}) \mid \mathbf{X} = X\right] & \text{if } X \in \mathcal{M}^{(\delta)}, \\ \infty & \text{otherwise.} \end{cases} \quad (19)$$

We show that the likelihood expression hinges on two factors: (i) the difference between the estimated ratings (entries of $X$) and the observed ratings (elements of $Y$); and (ii) the dissimilarity between the graph induced by the partitioning in $\mathcal{Z}$ and the observed graph $\mathcal{G}$. As defined in Section I-B, we denote by $\Lambda(X, Y)$ the number of mismatched entries between $X$ and $Y$. For a user partitioning $\mathcal{Z}$, let $\mathcal{P}_\alpha(\mathcal{Z})$ denote the set of pairs of users within any group; $\mathcal{P}_\beta(\mathcal{Z})$ denote the set of pairs of users in different groups within any cluster; and $\mathcal{P}_\gamma(\mathcal{Z})$ denote the set of pairs of users in different clusters. Formally, we have

$$\begin{aligned} \mathcal{P}_\alpha(\mathcal{Z}) &= \{(a,b) : a \in Z(x,i),\ b \in Z(x,I), \\ &\qquad \text{for } x \in [c],\ i \in [g]\}, \\ \mathcal{P}_\beta(\mathcal{Z}) &= \{(a,b) : a \in Z(x,i),\ b \in Z(x,j), \\ &\qquad \text{for } x \in [c],\ i,j \in [g],\ i \ne j\}, \\ \mathcal{P}_\gamma(\mathcal{Z}) &= \{(a,b) : a \in Z(x,i),\ b \in Z(y,j), \\ &\qquad \text{for } x,y \in [c],\ x \ne y,\ i,j \in [g]\}. \end{aligned} \quad (20)$$

Recall from Section II that the user partitioning induced by any rating matrix in $\mathcal{M}^{(\delta)}$ should satisfy the property that all groups have equal sizes of $n/(cg)$ users. This implies that the sizes of $\mathcal{P}_\alpha(\mathcal{Z})$, $\mathcal{P}_\beta(\mathcal{Z})$ and $\mathcal{P}_\gamma(\mathcal{Z})$ are constants and are given by

$$\begin{aligned} |\mathcal{P}_\alpha(\mathcal{Z})| &= gc\binom{n/(gc)}{2}, \quad |\mathcal{P}_\beta(\mathcal{Z})| = c\binom{g}{2}(n/(gc))^2, \\ |\mathcal{P}_\gamma(\mathcal{Z})| &= \binom{c}{2}(n/c)^2, \end{aligned} \quad (21)$$

[4] The purpose of introducing the first assumption about $R_0^{(1)}$ is to simplify the achievability proof. Nevertheless, it is important to mention that the same proof can be used for scenarios where the first row of $R_0^{(1)}$ is not exactly an all-one vector. As long as the rating vectors of $R_0^{(1)}$ follow the MDS code structure, the proof remains valid.

[5] A similar interpretation goes for $\{v_i^{(x)}(\ell) : \ell \in \{0, 1, \ldots, q-1\}^{cr-1}\}$.

[6] With a slight abuse of notation, we omit the dependence on $(Y, G)$ in the likelihood function for notational compactness.

$$\Phi^{(1)} = \Phi^{(2)} = \Phi = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \end{bmatrix}. \tag{22}$$

$$W^{(1)} = \left[\begin{array}{c|c|c|c|c|c|c|c} \mathbf{1}_{1\times s_{000}m} & \mathbf{1}_{1\times s_{001}m} & \mathbf{1}_{1\times s_{010}m} & \mathbf{1}_{1\times s_{011}m} & \mathbf{1}_{1\times s_{100}m} & \mathbf{1}_{1\times s_{101}m} & \mathbf{1}_{1\times s_{110}m} & \mathbf{1}_{1\times s_{111}m} \\ \hline \mathbf{0}_{1\times s_{000}m} & \mathbf{0}_{1\times s_{001}m} & \mathbf{0}_{1\times s_{010}m} & \mathbf{0}_{1\times s_{011}m} & \mathbf{1}_{1\times s_{100}m} & \mathbf{1}_{1\times s_{101}m} & \mathbf{1}_{1\times s_{110}m} & \mathbf{1}_{1\times s_{111}m} \end{array}\right].$$
$$W^{(2)} = \left[\begin{array}{c|c|c|c|c|c|c|c} \mathbf{0}_{1\times s_{000}m} & \mathbf{0}_{1\times s_{001}m} & \mathbf{1}_{1\times s_{010}m} & \mathbf{1}_{1\times s_{011}m} & \mathbf{0}_{1\times s_{100}m} & \mathbf{0}_{1\times s_{101}m} & \mathbf{1}_{1\times s_{110}m} & \mathbf{1}_{1\times s_{111}m} \\ \hline \mathbf{0}_{1\times s_{000}m} & \mathbf{1}_{1\times s_{001}m} & \mathbf{0}_{1\times s_{010}m} & \mathbf{1}_{1\times s_{011}m} & \mathbf{0}_{1\times s_{100}m} & \mathbf{1}_{1\times s_{101}m} & \mathbf{0}_{1\times s_{110}m} & \mathbf{1}_{1\times s_{111}m} \end{array}\right]. \tag{23}$$

$$R_0^{(1)} = \Phi^{(1)}W^{(1)} = \left[\begin{array}{c|c|c|c|c|c|c|c} \mathbf{1}_{1\times s_{000}m} & \mathbf{1}_{1\times s_{001}m} & \mathbf{1}_{1\times s_{010}m} & \mathbf{1}_{1\times s_{011}m} & \mathbf{1}_{1\times s_{100}m} & \mathbf{1}_{1\times s_{101}m} & \mathbf{1}_{1\times s_{110}m} & \mathbf{1}_{1\times s_{111}m} \\ \hline \mathbf{0}_{1\times s_{000}m} & \mathbf{0}_{1\times s_{001}m} & \mathbf{0}_{1\times s_{010}m} & \mathbf{0}_{1\times s_{011}m} & \mathbf{1}_{1\times s_{100}m} & \mathbf{1}_{1\times s_{101}m} & \mathbf{1}_{1\times s_{110}m} & \mathbf{1}_{1\times s_{111}m} \\ \hline \mathbf{1}_{1\times s_{000}m} & \mathbf{1}_{1\times s_{001}m} & \mathbf{1}_{1\times s_{010}m} & \mathbf{1}_{1\times s_{011}m} & \mathbf{0}_{1\times s_{100}m} & \mathbf{0}_{1\times s_{101}m} & \mathbf{0}_{1\times s_{110}m} & \mathbf{0}_{1\times s_{111}m} \end{array}\right].$$
$$R_0^{(2)} = \Phi^{(2)}W^{(2)} = \left[\begin{array}{c|c|c|c|c|c|c|c} \mathbf{0}_{1\times s_{000}m} & \mathbf{0}_{1\times s_{001}m} & \mathbf{1}_{1\times s_{010}m} & \mathbf{1}_{1\times s_{011}m} & \mathbf{0}_{1\times s_{100}m} & \mathbf{0}_{1\times s_{101}m} & \mathbf{1}_{1\times s_{110}m} & \mathbf{1}_{1\times s_{111}m} \\ \hline \mathbf{0}_{1\times s_{000}m} & \mathbf{1}_{1\times s_{001}m} & \mathbf{0}_{1\times s_{010}m} & \mathbf{1}_{1\times s_{011}m} & \mathbf{0}_{1\times s_{100}m} & \mathbf{1}_{1\times s_{101}m} & \mathbf{0}_{1\times s_{110}m} & \mathbf{1}_{1\times s_{111}m} \\ \hline \mathbf{0}_{1\times s_{000}m} & \mathbf{1}_{1\times s_{001}m} & \mathbf{1}_{1\times s_{010}m} & \mathbf{0}_{1\times s_{011}m} & \mathbf{0}_{1\times s_{100}m} & \mathbf{1}_{1\times s_{101}m} & \mathbf{1}_{1\times s_{110}m} & \mathbf{0}_{1\times s_{111}m} \end{array}\right]. \tag{24}$$

$$X_0 = \left[\begin{array}{c|c|c|c|c|c|c|c} \mathbf{1}_{\frac{n}{6}\times s_{000}m} & \mathbf{1}_{\frac{n}{6}\times s_{001}m} & \mathbf{1}_{\frac{n}{6}\times s_{010}m} & \mathbf{1}_{\frac{n}{6}\times s_{011}m} & \mathbf{1}_{\frac{n}{6}\times s_{100}m} & \mathbf{1}_{\frac{n}{6}\times s_{101}m} & \mathbf{1}_{\frac{n}{6}\times s_{110}m} & \mathbf{1}_{\frac{n}{6}\times s_{111}m} \\ \hline \mathbf{0}_{\frac{n}{6}\times s_{000}m} & \mathbf{0}_{\frac{n}{6}\times s_{001}m} & \mathbf{0}_{\frac{n}{6}\times s_{010}m} & \mathbf{0}_{\frac{n}{6}\times s_{011}m} & \mathbf{1}_{\frac{n}{6}\times s_{100}m} & \mathbf{1}_{\frac{n}{6}\times s_{101}m} & \mathbf{1}_{\frac{n}{6}\times s_{110}m} & \mathbf{1}_{\frac{n}{6}\times s_{111}m} \\ \hline \mathbf{1}_{\frac{n}{6}\times s_{000}m} & \mathbf{1}_{\frac{n}{6}\times s_{001}m} & \mathbf{1}_{\frac{n}{6}\times s_{010}m} & \mathbf{1}_{\frac{n}{6}\times s_{011}m} & \mathbf{0}_{\frac{n}{6}\times s_{100}m} & \mathbf{0}_{\frac{n}{6}\times s_{101}m} & \mathbf{0}_{\frac{n}{6}\times s_{110}m} & \mathbf{0}_{\frac{n}{6}\times s_{111}m} \\ \hline \mathbf{0}_{\frac{n}{6}\times s_{000}m} & \mathbf{0}_{\frac{n}{6}\times s_{001}m} & \mathbf{1}_{\frac{n}{6}\times s_{010}m} & \mathbf{1}_{\frac{n}{6}\times s_{011}m} & \mathbf{0}_{\frac{n}{6}\times s_{100}m} & \mathbf{0}_{\frac{n}{6}\times s_{101}m} & \mathbf{1}_{\frac{n}{6}\times s_{110}m} & \mathbf{1}_{\frac{n}{6}\times s_{111}m} \\ \hline \mathbf{0}_{\frac{n}{6}\times s_{000}m} & \mathbf{1}_{\frac{n}{6}\times s_{001}m} & \mathbf{0}_{\frac{n}{6}\times s_{010}m} & \mathbf{1}_{\frac{n}{6}\times s_{011}m} & \mathbf{0}_{\frac{n}{6}\times s_{100}m} & \mathbf{1}_{\frac{n}{6}\times s_{101}m} & \mathbf{0}_{\frac{n}{6}\times s_{110}m} & \mathbf{1}_{\frac{n}{6}\times s_{111}m} \\ \hline \mathbf{0}_{\frac{n}{6}\times s_{000}m} & \mathbf{1}_{\frac{n}{6}\times s_{001}m} & \mathbf{1}_{\frac{n}{6}\times s_{010}m} & \mathbf{0}_{\frac{n}{6}\times s_{011}m} & \mathbf{0}_{\frac{n}{6}\times s_{100}m} & \mathbf{1}_{\frac{n}{6}\times s_{101}m} & \mathbf{1}_{\frac{n}{6}\times s_{110}m} & \mathbf{0}_{\frac{n}{6}\times s_{111}m} \end{array}\right]. \tag{25}$$

for any user partitioning. Furthermore, for a graph $\mathcal{G}$ and a user partitioning $\mathcal{Z}$, define $e_\alpha(\mathcal{G}, \mathcal{Z})$ as the number of edges within any group; $e_\beta(\mathcal{G}, \mathcal{Z})$ as the number of edges across groups within any cluster; and $e_\gamma(\mathcal{G}, \mathcal{Z})$ as the number of edges across clusters. More formally, we have

$$e_\mu(\mathcal{G}, \mathcal{Z}) = \sum_{(a,b)\in\mathcal{P}_\mu(\mathcal{Z})} \mathbb{1}\left[(a,b)\in\mathcal{E}\right], \tag{26}$$

for $\mu \in \{\widetilde{\alpha}, \widetilde{\beta}, \widetilde{\gamma}\}$. The following lemma provides a precise expression of $\mathsf{L}(X)$.

*Lemma 1:* For a given (and fixed) input pair $(Y, \mathcal{G})$ and any $X \in \mathcal{M}^{(\delta)}$, we have

$$\begin{aligned}\mathsf{L}(X) &= \log\left((q-1)\frac{1-\theta}{\theta}\right)\Lambda(Y, X) \\ &\quad + \sum_{\mu\in\{\widetilde{\alpha},\widetilde{\beta},\widetilde{\gamma}\}}\left[\log\left(\frac{1-\mu}{\mu}\right)e_\mu(\mathcal{G}, \mathcal{Z}) - \log(1-\mu)|\mathcal{P}_\mu(\mathcal{Z})|\right],\end{aligned} \tag{27}$$

where $\widetilde{\alpha}$, $\widetilde{\beta}$ and $\widetilde{\gamma}$ are the edge probabilities defined in (5).

*Proof:* We refer to Appendix A for the proof of Lemma 1. ∎

The following lemma provides an upper bound on the worst-case probability of error $P_e^{(\delta)}(\psi_{\text{ML}})$.

*Lemma 2:* For the maximum likelihood estimator $\psi_{\text{ML}}$, we have

$$P_e^{(\delta)}(\psi_{\text{ML}}) \leq \sum_{X\neq X_0} \mathbb{P}\left[\mathsf{L}(X_0) \geq \mathsf{L}(X)\right]. \tag{28}$$

*Proof:* We refer to Appendix B for the proof of Lemma 2. ∎

For a ground truth rating matrix $X_0 = (\mathcal{V}_0, \mathcal{Z}_0)$; a candidate rating matrix $X = (\mathcal{V}, \mathcal{Z})$; and a tuple $T \in \mathcal{T}^{(\delta)}$, define the following disjoint sets:

- define $\mathcal{P}_{\text{d}} = \mathcal{P}_{\text{d}}(X_0, X)$ as the set of matrix entries where $X \neq X_0$. Formally, we have

$$\mathcal{P}_{\text{d}} = \{(r,t) \in [n]\times[m] : X(r,t) \neq X_0(r,t)\}; \tag{29}$$

- define $\mathcal{P}_{\widetilde{\beta}\to\widetilde{\alpha}} = \mathcal{P}_{\widetilde{\beta}\to\widetilde{\alpha}}(\mathcal{Z}_0, \mathcal{Z})$ as the set of pairs of users where the two users of each pair belong to different groups of the same cluster in $X_0$ (and therefore they are connected with probability $\widetilde{\beta}$), but they are estimated to be in the same group in $X$ (and hence, given the estimator output, the belief for the existence of an edge between these two users is $\widetilde{\alpha}$). Formally, we have

$$\begin{aligned}\mathcal{P}_{\widetilde{\beta}\to\widetilde{\alpha}} = \{&(a,b) : a \in Z_0(x, i_1) \cap Z(y, j), \\ &b \in Z_0(x, i_2) \cap Z(y, j), \\ &\text{for } x, y \in [c],\ i_1, i_2, j \in [g],\ i_1 \neq i_2\}.\end{aligned} \tag{30}$$

On the other hand, define $\mathcal{P}_{\widetilde{\alpha}\to\widetilde{\beta}} = \mathcal{P}_{\widetilde{\alpha}\to\widetilde{\beta}}(\mathcal{Z}_0, \mathcal{Z})$ as

$$\begin{aligned}\mathcal{P}_{\widetilde{\alpha}\to\widetilde{\beta}} = \{&(a,b) : a \in Z_0(x, i) \cap Z(y, j_1), \\ &b \in Z_0(x, i) \cap Z(y, j_2), \\ &\text{for } x, y \in [c],\ i, j_1, j_2 \in [g],\ j_1 \neq j_2\};\end{aligned} \tag{31}$$

- define $\mathcal{P}_{\widetilde{\gamma}\to\widetilde{\alpha}} = \mathcal{P}_{\widetilde{\gamma}\to\widetilde{\alpha}}(\mathcal{Z}_0, \mathcal{Z})$ as the set of pairs of users where the two users of each pair belong to different clusters in $X_0$ (and therefore they are connected with probability $\widetilde{\gamma}$), but they are estimated to be in the same group in $X$ (and hence, given the estimator output, the belief for the existence of an edge between these two users is $\widetilde{\alpha}$). Formally, we have

$$\mathcal{P}_{\widetilde{\gamma}\to\widetilde{\alpha}} = \{(a,b) : a\in Z_0(x_1,i_1)\cap Z(y,j),\ b\in Z_0(x_2,i_2)\cap Z(y,j),\ \text{for } x_1,x_2,y\in[c],\ x_1\neq x_2,\ i_1,i_2,j\in[g]\}. \tag{32}$$

On the other hand, define $\mathcal{P}_{\widetilde{\alpha}\to\widetilde{\gamma}} = \mathcal{P}_{\widetilde{\alpha}\to\widetilde{\gamma}}(\mathcal{Z}_0,\mathcal{Z})$ as

$$\mathcal{P}_{\widetilde{\alpha}\to\widetilde{\gamma}} = \{(a,b) : a\in Z_0(x,i)\cap Z(y_1,j_1),\ b\in Z_0(x,i)\cap Z(y_2,j_2),\ \text{for } x,y_1,y_2\in[c],\ y_1\neq y_2,\ i,j_1,j_2\in[g]\}; \tag{33}$$

- define $\mathcal{P}_{\widetilde{\gamma}\to\widetilde{\beta}} = \mathcal{P}_{\widetilde{\gamma}\to\widetilde{\beta}}(\mathcal{Z}_0,\mathcal{Z})$ as the set of pairs of users where the two users of each pair belong to different clusters in $X_0$ (and therefore they are connected with probability $\widetilde{\gamma}$), but they are estimated to be in different groups of the same cluster in $X$ (and hence, given the estimator output, the belief for the existence of an edge between these two users is $\widetilde{\beta}$). Formally, we have

$$\mathcal{P}_{\widetilde{\gamma}\to\widetilde{\beta}} = \{(a,b) : a\in Z_0(x_1,i_1)\cap Z(y,j_1),\ b\in Z_0(x_2,i_2)\cap Z(y,j_2),\ \text{for } x_1,x_2,y\in[c],\ x_1\neq x_2,\ i_1,i_2,j_1,j_2\in[g],\ j_1\neq j_2\}. \tag{34}$$

On the other hand, define $\mathcal{P}_{\widetilde{\beta}\to\widetilde{\gamma}} = \mathcal{P}_{\widetilde{\beta}\to\widetilde{\gamma}}(\mathcal{Z}_0,\mathcal{Z})$ as

$$\mathcal{P}_{\widetilde{\beta}\to\widetilde{\gamma}} = \{(a,b) : a\in Z_0(x,i_1)\cap Z(y_1,j_1),\ b\in Z_0(x,i_2)\cap Z(y_2,j_2),\ \text{for } x,y_1,y_2\in[c],\ y_1\neq y_2,\ i_1,i_2,j_1,j_2\in[g],\ i_1\neq i_2\}. \tag{35}$$

Let $\mathsf{B}_i^{(\sigma)}$ denote the $i$th Bernoulli random variable with parameter $\sigma\in\{p,\theta,\frac{1}{q-1},\widetilde{\alpha},\widetilde{\beta},\widetilde{\gamma}\}$. Define the following sets of independent Bernoulli random variables:

$$\left\{\mathsf{B}_i^{(p)} : i\in\mathcal{P}_{\rm d}\right\},\ \left\{\mathsf{B}_i^{(\theta)} : i\in\mathcal{P}_{\rm d}\right\},\ \left\{\mathsf{B}_i^{\left(\frac{1}{q-1}\right)} : i\in\mathcal{P}_{\rm d}\right\},\ \left\{\mathsf{B}_i^{(\mu)} : i\in\mathcal{P}_{\mu\to\nu},\ \mu,\nu\in\left\{\widetilde{\alpha},\widetilde{\beta},\widetilde{\gamma}\right\},\ \mu\neq\nu\right\}. \tag{36}$$

Now, define $\mathbf{B} = \mathbf{B}\left(\mathcal{P}_{\rm d}, \{\mathcal{P}_{\mu\to\nu} : \mu,\nu\in\{\widetilde{\alpha},\widetilde{\beta},\widetilde{\gamma}\},\ \mu\neq\nu\}\right)$ as given in (42), shown at the bottom of the page. In the following lemma, we write each summand in (28) in terms of (42).

*Lemma 3:* For any $X\in\mathcal{X}(T)$ and $T\in\mathcal{T}^{(\delta)}$, we have

$$\mathbb{P}\left[\mathsf{L}(X_0)\geq\mathsf{L}(X)\right] = \mathbb{P}\left[\mathbf{B}\geq 0\right]. \tag{37}$$

*Proof:* We refer to Appendix C for the proof of Lemma 3. ■

The following lemma provides an upper bound of the RHS of (37).

*Lemma 4:* For any $\{\mathcal{P}_{\mu\to\nu} : \mu,\nu\in\{\widetilde{\alpha},\widetilde{\beta},\widetilde{\gamma}\},\ \mu\neq\nu\}$, we have

$$\mathbb{P}\left[\mathbf{B}\geq 0\right]\leq\exp\left[-(1+o(1))\left(|\mathcal{P}_{\rm d}|I_r + P_{\widetilde{\alpha}\leftrightarrow\widetilde{\beta}}\,I_{\alpha,\beta}\frac{\log n}{n} + P_{\widetilde{\alpha}\leftrightarrow\widetilde{\gamma}}\,I_{\alpha,\gamma}\frac{\log n}{n} + P_{\widetilde{\beta}\leftrightarrow\widetilde{\gamma}}\,I_{\beta,\gamma}\frac{\log n}{n}\right)\right], \tag{38}$$

where

$$P_{\widetilde{\alpha}\leftrightarrow\widetilde{\beta}} = \frac{\left|\mathcal{P}_{\widetilde{\beta}\to\widetilde{\alpha}}\right| + \left|\mathcal{P}_{\widetilde{\alpha}\to\widetilde{\beta}}\right|}{2},\quad P_{\widetilde{\alpha}\leftrightarrow\widetilde{\gamma}} = \frac{|\mathcal{P}_{\widetilde{\gamma}\to\widetilde{\alpha}}| + |\mathcal{P}_{\widetilde{\alpha}\to\widetilde{\gamma}}|}{2},\quad P_{\widetilde{\beta}\leftrightarrow\widetilde{\gamma}} = \frac{\left|\mathcal{P}_{\widetilde{\gamma}\to\widetilde{\beta}}\right| + \left|\mathcal{P}_{\widetilde{\beta}\to\widetilde{\gamma}}\right|}{2}. \tag{39}$$

*Proof:* We refer to Appendix D for the proof of Lemma 4. ■

In the following, we show that the error event in (28), i.e., $\{\mathsf{L}(X_0)\geq\mathsf{L}(X) : X\neq X_0\}$, depends solely on two sets of key parameters which dictate the relationship between $X$ and $X_0$:

1) the first set includes counters to identify the number of users in cluster $x$ and group $i$ whose rating vector $u_i^{(x)}$ in $X_0$ is changed to the rating vector $v_j^{(y)}$ of users in cluster $y$ and group $j$ in $X$, for $x,y\in[c]$ and $i,j\in[g]$. Formally, we define

$$n_{i,j}^{(x,y)} = \left|\{r : r\in Z_0(x,i)\cap Z(y,j)\}\right|, \tag{40}$$

for $0\leq n_{i,j}^{(x,y)}\leq\frac{n}{gc}$;

2) the second set provides the Hamming distance between vectors $u_i^{(x)}$ and $v_i^{(x)}$, for $x\in[c]$ and $i\in[g]$. Formally, we define

$$d_{i,j}^{(x,y)} = d_{\rm H}\left(u_i^{(x)}, v_j^{(y)}\right), \tag{41}$$

where $0\leq d_{i,j}^{(x,y)}\leq m$.

$$\begin{aligned}
\mathbf{B} := {} & \log\left((q-1)\frac{1-\theta}{\theta}\right)\sum_{i\in\mathcal{P}_{\rm d}}\mathsf{B}_i^{(p)}\left[\left(1+\mathsf{B}_i^{\left(\frac{1}{q-1}\right)}\right)\mathsf{B}_i^{(\theta)}-1\right] \\
& + \left(\log\frac{(1-\widetilde{\beta})\widetilde{\alpha}}{(1-\widetilde{\alpha})\widetilde{\beta}}\right)\left(\sum_{i\in\mathcal{P}_{\widetilde{\beta}\to\widetilde{\alpha}}}\mathsf{B}_i^{(\widetilde{\beta})} - \sum_{i\in\mathcal{P}_{\widetilde{\alpha}\to\widetilde{\beta}}}\mathsf{B}_i^{(\widetilde{\alpha})}\right) + \left(\log\frac{1-\widetilde{\alpha}}{1-\widetilde{\beta}}\right)\left(\left|\mathcal{P}_{\widetilde{\beta}\to\widetilde{\alpha}}\right| - \left|\mathcal{P}_{\widetilde{\alpha}\to\widetilde{\beta}}\right|\right) \\
& + \left(\log\frac{(1-\widetilde{\gamma})\widetilde{\alpha}}{(1-\widetilde{\alpha})\widetilde{\gamma}}\right)\left(\sum_{i\in\mathcal{P}_{\widetilde{\gamma}\to\widetilde{\alpha}}}\mathsf{B}_i^{(\widetilde{\gamma})} - \sum_{i\in\mathcal{P}_{\widetilde{\alpha}\to\widetilde{\gamma}}}\mathsf{B}_i^{(\widetilde{\alpha})}\right) + \left(\log\frac{1-\widetilde{\alpha}}{1-\widetilde{\gamma}}\right)\left(|\mathcal{P}_{\widetilde{\gamma}\to\widetilde{\alpha}}| - |\mathcal{P}_{\widetilde{\alpha}\to\widetilde{\gamma}}|\right) \\
& + \left(\log\frac{(1-\widetilde{\gamma})\widetilde{\beta}}{(1-\widetilde{\beta})\widetilde{\gamma}}\right)\left(\sum_{i\in\mathcal{P}_{\widetilde{\gamma}\to\widetilde{\beta}}}\mathsf{B}_i^{(\widetilde{\gamma})} - \sum_{i\in\mathcal{P}_{\widetilde{\beta}\to\widetilde{\gamma}}}\mathsf{B}_i^{(\widetilde{\beta})}\right) + \left(\log\frac{1-\widetilde{\beta}}{1-\widetilde{\gamma}}\right)\left(\left|\mathcal{P}_{\widetilde{\gamma}\to\widetilde{\beta}}\right| - \left|\mathcal{P}_{\widetilde{\beta}\to\widetilde{\gamma}}\right|\right).
\end{aligned} \tag{42}$$

Based on these two parameters, the set of rating matrices $\mathcal{M}^{(\delta)}$ is partitioned into a number of classes of matrices $\mathcal{X}(T)$. Here, each matrix class $\mathcal{X}(T)$ is defined as the set of rating matrices that is characterized by a tuple $T$ where

$$T = \left( \left\{ n_{i,j}^{(x,y)} \right\}_{x,y \in [c],\, i,j \in [g]}, \left\{ d_{i,j}^{(x,y)} \right\}_{x,y \in [c],\, i,j \in [g]} \right). \quad (43)$$

Define $\mathcal{T}^{(\delta)}$ as the set of all non-all-zero tuples $T$. Therefore, we can write $\mathcal{M}^{(\delta)} = \bigcup_{T \in \mathcal{T}^{(\delta)}} \mathcal{X}(T)$.

Next, we analyze the performance of the ML decoder by comparing the ground truth user partitioning with that of the decoder. For a non-negative constant

$$\tau \in (0,\ (\epsilon \log m - (2+\epsilon)\log(2q))/(2(1+\epsilon)\log m)),$$

with

$$\epsilon > \max\left\{ \frac{2\log 2}{\log n}, \frac{2(g-r+1)\log 2}{\log(2qm)}, \frac{2\log(2q)}{\log(m/2q)} \right\},$$

define $\sigma(x,i)$ as the set of pairs of cluster and group in $\mathcal{Z}$ whose number of overlapped users with $\mathcal{Z}_0(x,i)$ exceeds a $(1-\tau)$-fraction of the group size. Formally, we have

$$\sigma(x,i) = \left\{ (y,j) \in [c] \times [g] : |Z_0(x,i) \cap Z(y,j)| \geq (1-\tau)\frac{n}{gc} \right\}. \quad (44)$$

Note that $\tau < 0.5$, which implies that $|\sigma(x,i)| \leq 1$ since the size of any group is $n/(gc)$ users. For $|\sigma(x,i)| = 1$, let $\sigma(x,i) = \{(\sigma(x), \sigma(i|x))\}$. Accordingly, partition the set $\mathcal{T}^{(\delta)}$ into two subsets $\mathcal{T}^{(\delta)}_{\text{small}}$ and $\mathcal{T}^{(\delta)}_{\text{large}}$ that are defined as follows:

$$\mathcal{T}^{(\delta)}_{\text{small}} = \Big\{ T \in \mathcal{T}^{(\delta)} : \forall (x,i) \in [c] \times [g] \text{ s.t. } |\sigma(x,i)| = 1, \\ d_{i,\,\sigma(i|x)}^{(x,\,\sigma(x))} \leq \tau m \min\{\delta_c, \delta_g\} \Big\}, \quad (45)$$

$$\mathcal{T}^{(\delta)}_{\text{large}} = \Big\{ T \in \mathcal{T}^{(\delta)} : \exists (x,i) \in [c] \times [g] \text{ s.t. } (|\sigma(x,i)| = 0) \Big\} \\ \cup \Big\{ T \in \mathcal{T}^{(\delta)} : \forall (x,i) \in [c] \times [g] \text{ s.t. } |\sigma(x,i)| = 1, \\ \exists (x,i) \in [c] \times [g] \text{ s.t. } d_{i,\,\sigma(i|x)}^{(x,\,\sigma(x))} > \tau m \min\{\delta_c, \delta_g\} \Big\}. \quad (46)$$

Intuitively, when $T \in \mathcal{T}^{(\delta)}_{\text{small}}$, the class of matrices $\mathcal{X}(T)$ corresponds to the typical (i.e., small) error set. On the other hand, when $T \in \mathcal{T}^{(\delta)}_{\text{large}}$, the class of matrices $\mathcal{X}(T)$ corresponds to the atypical (i.e., large) error set that has negligible probability mass.

The following two lemmas provide an upper bound on the RHS of (38) under different classes of candidate rating matrices, and evaluating the limits as $n$ and $m$ tend to infinity.

*Lemma 5:* For any $\{\mathcal{P}_{\mu \to \nu} : \mu, \nu \in \{\widetilde{\alpha}, \widetilde{\beta}, \widetilde{\gamma}\},\ \mu \neq \nu\}$, we have[7]

$$\lim_{n,m \to \infty} \sum_{T \in \mathcal{T}^{(\delta)}_{\text{small}}} \sum_{X \in \mathcal{X}(T)} \exp\Big[ -(1+o(1)) \Big( |\mathcal{P}_{\text{d}}| I_r \\ + \frac{\log n}{n} \left( P_{\widetilde{\alpha} \leftrightarrow \widetilde{\beta}} I_{\alpha,\beta} + P_{\widetilde{\alpha} \leftrightarrow \widetilde{\gamma}} I_{\alpha,\gamma} + P_{\widetilde{\beta} \leftrightarrow \widetilde{\gamma}} I_{\beta,\gamma} \right) \Big) \Big] = 0. \quad (47)$$

[7]As $n$ tends to infinity, $m$ also tends to infinity since $m = \omega(\log n)$.

*Proof:* We refer to Appendix E for the proof of Lemma 5. ∎

*Lemma 6:* For any $\{\mathcal{P}_{\mu \to \nu} : \mu, \nu \in \{\widetilde{\alpha}, \widetilde{\beta}, \widetilde{\gamma}\},\ \mu \neq \nu\}$, we have

$$\lim_{n,m \to \infty} \sum_{T \in \mathcal{T}^{(\delta)}_{\text{large}}} \sum_{X \in \mathcal{X}(T)} \exp\Big[ -(1+o(1)) \Big( |\mathcal{P}_{\text{d}}| I_r \\ + \frac{\log n}{n} \left( P_{\widetilde{\alpha} \leftrightarrow \widetilde{\beta}} I_{\alpha,\beta} + P_{\widetilde{\alpha} \leftrightarrow \widetilde{\gamma}} I_{\alpha,\gamma} + P_{\widetilde{\beta} \leftrightarrow \widetilde{\gamma}} I_{\beta,\gamma} \right) \Big) \Big] = 0. \quad (48)$$

*Proof:* We refer to Appendix F for the proof of Lemma 6. ∎

### C. The Achievability Proof of Theorem 1

The worst-case probability of error $P_e^{(\delta)}(\psi_{\text{ML}})$ is upper bounded by

$$\begin{aligned} & P_e^{(\delta)}(\psi_{\text{ML}}) \\ & \leq \sum_{X \neq X_0} \mathbb{P}\left[ \mathsf{L}(X_0) \geq \mathsf{L}(X) \right] \quad (49) \\ & = \sum_{\substack{X \neq X_0, \\ X \in \mathcal{M}^{(\delta)}}} \mathbb{P}\left[ \mathsf{L}(X_0) \geq \mathsf{L}(X) \right] \quad (50) \\ & = \sum_{T \in \mathcal{T}^{(\delta)}} \sum_{X \in \mathcal{X}(T)} \mathbb{P}\left[ \mathsf{L}(X_0) \geq \mathsf{L}(X) \right] \quad (51) \\ & \leq \sum_{T \in \mathcal{T}^{(\delta)}} \sum_{X \in \mathcal{X}(T)} \exp\Big[ -(1+o(1)) \Big( |\mathcal{P}_{\text{d}}| I_r \\ & \quad + \frac{\log n}{n} \left( P_{\widetilde{\alpha} \leftrightarrow \widetilde{\beta}} I_{\alpha,\beta} + P_{\widetilde{\alpha} \leftrightarrow \widetilde{\gamma}} I_{\alpha,\gamma} + P_{\widetilde{\beta} \leftrightarrow \widetilde{\gamma}} I_{\beta,\gamma} \right) \Big) \Big] \quad (52) \\ & = \sum_{T \in \mathcal{T}^{(\delta)}_{\text{small}}} \sum_{X \in \mathcal{X}(T)} \exp\Big[ -(1+o(1)) \Big( |\mathcal{P}_{\text{d}}| I_r \\ & \quad + \frac{\log n}{n} \left( P_{\widetilde{\alpha} \leftrightarrow \widetilde{\beta}} I_{\alpha,\beta} + P_{\widetilde{\alpha} \leftrightarrow \widetilde{\gamma}} I_{\alpha,\gamma} + P_{\widetilde{\beta} \leftrightarrow \widetilde{\gamma}} I_{\beta,\gamma} \right) \Big) \Big] \\ & \quad + \sum_{T \in \mathcal{T}^{(\delta)}_{\text{large}}} \sum_{X \in \mathcal{X}(T)} \exp\Big[ -(1+o(1)) \Big( |\mathcal{P}_{\text{d}}| I_r \\ & \quad + \frac{\log n}{n} \left( P_{\widetilde{\alpha} \leftrightarrow \widetilde{\beta}} I_{\alpha,\beta} + P_{\widetilde{\alpha} \leftrightarrow \widetilde{\gamma}} I_{\alpha,\gamma} + P_{\widetilde{\beta} \leftrightarrow \widetilde{\gamma}} I_{\beta,\gamma} \right) \Big) \Big], \quad (53) \end{aligned}$$

where (49) follows from Lemma 2; (50) follows from the definition of negative log-likelihood in (19); (51) follows from the definition of the tuples characterizing matrix classes in (43); (52) follows from Lemma 3 and Lemma 4; and finally (53) follows from the definitions of $\mathcal{T}^{(\delta)}_{\text{small}}$ and $\mathcal{T}^{(\delta)}_{\text{large}}$ in (45) and (46), respectively.

Finally, following Lemma 5 and Lemma 6, as $n$ and $m$ tend to infinity, the limit of the worst-case probability of error $P_e^{(\delta)}(\psi_{\text{ML}})$ in (53) can be evaluated as (54), shown at the top of the next page. This concludes the achievability proof of Theorem 1. ∎

## V. The Converse Proof

In this section, we prove the converse part of Theorem 1. More precisely, we show that if the condition on $p$ in (9) holds,

$$
\begin{aligned}
&\lim_{n,m\to\infty} P_e^{(\delta)}(\psi_{\text{ML}}) \\
&\le \lim_{n,m\to\infty} \Bigg( \sum_{T\in\mathcal{T}_{\text{small}}^{(\delta)}} \sum_{X\in\mathcal{X}(T)} \exp\left[-(1+o(1))\left(|\mathcal{P}_{\text{d}}|I_r + P_{\widetilde{\alpha}\leftrightarrow\widetilde{\beta}}\, I_{\alpha,\beta}\frac{\log n}{n} + P_{\widetilde{\alpha}\leftrightarrow\widetilde{\gamma}}\, I_{\alpha,\gamma}\frac{\log n}{n} + P_{\widetilde{\beta}\leftrightarrow\widetilde{\gamma}}\, I_{\beta,\gamma}\frac{\log n}{n}\right)\right] \\
&\quad + \sum_{T\in\mathcal{T}_{\text{large}}^{(\delta)}} \sum_{X\in\mathcal{X}(T)} \exp\left[-(1+o(1))\left(|\mathcal{P}_{\text{d}}|I_r + P_{\widetilde{\alpha}\leftrightarrow\widetilde{\beta}}\, I_{\alpha,\beta}\frac{\log n}{n} + P_{\widetilde{\alpha}\leftrightarrow\widetilde{\gamma}}\, I_{\alpha,\gamma}\frac{\log n}{n} + P_{\widetilde{\beta}\leftrightarrow\widetilde{\gamma}}\, I_{\beta,\gamma}\frac{\log n}{n}\right)\right]\Bigg) \\
&= 0. \qquad (54)
\end{aligned}
$$

then $\lim_{n\to\infty} P_e^{(\delta)}(\psi) \neq 0$ for any estimator $\psi$. To this end, we prove that $\lim_{n\to\infty} P_e^{(\delta)}(\psi) \neq 0$ for any estimator $\psi$ and any ground truth rating matrix $X_0 \in \mathcal{M}^{(\delta)}$, if either of the following conditions holds:

**Perfect Clustering/Grouping Regime:**

$$\frac{g-r+1}{gc} nI_r \;\le\; (1-\epsilon)\log m, \tag{55}$$

**Grouping-Limited Regime:**

$$\delta_g m I_r + \frac{I_{\alpha,\beta}}{gc}\log n \;\le\; (1-\epsilon)\log n, \tag{56}$$

**Clustering-Limited Regime:**

$$\delta_c m I_r + \frac{I_{\alpha,\gamma}}{gc}\log n + \frac{(g-1)I_{\beta,\gamma}}{gc}\log n \;\le\; (1-\epsilon)\log n, \tag{57}$$

where $I_r$, $I_{\alpha,\beta}$, $I_{\alpha,\gamma}$ and $I_{\beta,\gamma}$ are defined in (15). Throughout the proof, let $p = \Theta((\log n)/n)$, and let $q$ and $\theta$ be constants such that $q$ is prime and $\theta \in [0,1]$. We first present a number of auxiliary lemmas in Section V-A. Then, we present the converse proof of Theorem 1 in Section V-B.

In the converse proof, we establish a lower bound on the error probability and show that it is minimized when employing the maximum likelihood estimator. Next, we prove that if $p$ is smaller than any of the three terms in the RHS of (9), then there exists another solution that yields a larger likelihood, compared to the ground truth matrix. More precisely, for any estimator and any ground truth rating matrix, we have the following three cases:

- if
$$p \le \frac{(1-\epsilon)gc\log m}{\left(\sqrt{1-\theta}-\sqrt{\frac{\theta}{(q-1)}}\right)^2 (g-r+1)n},$$
there exists a class of matrices that is obtained by replacing one column of the ground truth rating matrix with a carefully chosen sequence and yields a larger likelihood than the one of the ground truth rating matrix;
- if
$$p \le \frac{\log n}{\left(\sqrt{1-\theta}-\sqrt{\frac{\theta}{(q-1)}}\right)^2 \delta_g m}\left((1-\epsilon) - \frac{(\sqrt{\alpha}-\sqrt{\gamma})^2}{gc}\right),$$
there exists a class of rating matrices that is obtained by swapping the rating vectors of two users in the same cluster yet from distinct groups such that the Hamming distance between their rating vectors is $m\delta_g$. We show that the likelihood of any rating matrix from this class is greater than the one of the ground truth rating matrix;
- and finally, if
$$p \le \frac{\log n}{\left(\sqrt{1-\theta}-\sqrt{\frac{\theta}{(q-1)}}\right)^2 \delta_c m}\left((1-\epsilon) - \frac{\left(\sqrt{\alpha}-\sqrt{\gamma}\right)^2 + (g-1)\left(\sqrt{\beta}-\sqrt{\gamma}\right)^2}{gc}\right),$$
there exists a class of rating matrices, which is obtained by swapping the rating vectors of two users in distinct clusters such that the Hamming distance between their rating vectors is $m\delta_c$. We demonstrate that any rating matrix from this class yields a larger likelihood than the one of the ground truth rating matrix.

For each case, we show that the maximum likelihood estimator will fail in the limit of $n$ and $m$ by selecting one of the rating matrices from the respective class instead of the ground truth rating matrix.

### *A. The Auxiliary Lemmas*

We present three auxiliary lemmas that are used to prove the converse part of Theorem 1. Before each lemma, we introduce terminologies and notations needed for the statement of the lemma.

First, let $S$ denote the success event that a rating matrix is correctly estimated (i.e., exactly recovered). It is defined as

$$S := \bigcap_{X\neq X_0} \left[\mathsf{L}(X) > \mathsf{L}(X_0)\right], \tag{58}$$

where $\mathsf{L}(X)$ is the negative log-likelihood of a candidate rating matrix $X$, defined in (19). The following lemma introduces a lower bound on the infimum of the worst-case probability of error.

*Lemma 7:* For any estimator $\psi$, we have

$$\inf_{\psi} P_e^{(\delta)}(\psi) \ge \mathbb{P}\left[S^c\right]. \tag{59}$$

*Proof:* We refer to Appendix G for the proof of Lemma 7. ■

Next, we will use the following lemma together with Lemma 3, to provide a lower bound on the probability that $\mathsf{L}(X_0)$ is greater than or equal to $\mathsf{L}(X)$.

*Lemma 8:* For any

$$\left\{\mathcal{P}_{\mu\to\nu} : |\mathcal{P}_{\mu\to\nu}| = |\mathcal{P}_{\nu\to\mu}|\,,\ \mu,\nu \in \left\{\widetilde{\alpha},\widetilde{\beta},\widetilde{\gamma}\right\},\ \mu \neq \nu\right\},$$

we have

$$\begin{aligned}&\mathbb{P}\left[\mathbf{B} \geq 0\right]\\ &\geq \frac{1}{4}\exp\Bigg[-(1+o(1))\Bigg(|\mathcal{P}_{\mathrm{d}}|I_r + \left|\mathcal{P}_{\widetilde{\beta}\to\widetilde{\alpha}}\right| I_{\alpha,\beta}\frac{\log n}{n}\\ &\quad + |\mathcal{P}_{\widetilde{\gamma}\to\widetilde{\alpha}}|\, I_{\alpha,\gamma}\frac{\log n}{n} + \left|\mathcal{P}_{\widetilde{\gamma}\to\widetilde{\beta}}\right| I_{\beta,\gamma}\frac{\log n}{n}\Bigg)\Bigg],\end{aligned} \tag{60}$$

where the random variable $\mathbf{B}$ is defined in (42).

*Proof:* We refer to Appendix H for the proof of Lemma 8. ∎

Finally, we present a lemma that guarantees the existence of two subsets of users with specific properties.

*Lemma 9:* Consider sets $Z_0(x,i)$ and $Z_0(y,j)$ for distinct pairs $(x,i),(y,j) \in [c]\times[g]$. As $n\to\infty$, with probability approaching 1, there exist two subsets $\widetilde{Z}_0(x,i) \subset Z_0(x,i)$ and $\widetilde{Z}_0(y,j) \subset Z_0(y,j)$ with cardinalities $|\widetilde{Z}_0(x,i)| \geq \frac{n}{\log^3 n}$ and $|\widetilde{Z}_0(y,j)| \geq \frac{n}{\log^3 n}$ such that there are no edges between the vertices in $\widetilde{Z}_0(x,i)\cup\widetilde{Z}_0(y,j)$. That is,

$$\mathcal{E}\cap\left(\left(\widetilde{Z}_0(x,i)\cup\widetilde{Z}_0(y,j)\right)\times\left(\widetilde{Z}_0(x,i)\cup\widetilde{Z}_0(y,j)\right)\right)=\varnothing. \tag{61}$$

*Proof:* We refer to Appendix I for the proof of Lemma 9. ∎

### B. The Converse Proof of Theorem 1

In order to prove the converse part of Theorem 1, we demonstrate that $\lim_{n,m\to\infty}\mathbb{P}[S] = 0$ if any of the conditions given by (55), (56) or (57) holds. In the following, we show the claim for each condition in (55), (56) or (57) separately.

*1) Failure in the Perfect Clustering/Grouping Regime:* In this proof, we introduce a class of rating matrices, where each matrix in this class is obtained by replacing one column of $X_0$ with a carefully chosen sequence. Then, we prove that if (55) holds, then, with high probability, the ML estimator will fail by selecting one of the rating matrices from this class instead of $X_0$.

Recall the partitioning of the columns of $X_0$ defined in Section IV, and note that there exists (at least) one section $\mathcal{S}_\ell$ such that $s_\ell = |\mathcal{S}_\ell|/m$ is bounded away from zero (i.e., not vanishing with $m$ and $n$). For each $k\in\mathcal{S}_\ell$, define $X_{\langle k\rangle}\in\mathbb{F}_q^{n\times m}$ as a rating matrix that is identical to $X_0$ except for its $k$th, which will be determined below. Recall from Section IV that $R_0^{(1)}\in\mathbb{F}_q^{g\times m}$, the submatrix of $X_0$ associated with the first cluster, is obtained by stacking some codeword vectors from a $(g,r)$ MDS code with generator matrix $\Phi^{(1)}$. Let $w\in\mathbb{F}_q^{g\times 1}$ be another codeword from this MDS code such that

$$d_{\mathrm{H}}\left(w,\ R_0^{(1)}(:,k)\right) = g-r+1. \tag{62}$$

The existence of such a column vector $w$ is guaranteed due to the fact that the $(g,r)$ MDS code in $\mathbb{F}_q$ has a minimum distance of $g-r+1$. Consequently, the entries of $X_{\langle k\rangle}$ are given by

$$X_{\langle k\rangle}(r,t) = \begin{cases} w(1) & \text{if } r\in Z(1,1) \text{ and } t=k,\\ w(2) & \text{if } r\in Z(1,2) \text{ and } t=k,\\ \vdots & \vdots\\ w(g) & \text{if } r\in Z(1,g) \text{ and } t=k,\\ X_0(r,t) & \text{otherwise.}\end{cases} \tag{63}$$

Furthermore, given $X_0$ and $X_{\langle k\rangle}$, we have

$$\begin{aligned}&\mathcal{P}_{\mathrm{d}} = \left\{(r,k): r\in Z(1,i) \text{ for } i\in[g], X_{\langle k\rangle}(r,k)\neq X_0(r,k)\right\},\\ &\mathcal{P}_{\widetilde{\beta}\to\widetilde{\alpha}} = \mathcal{P}_{\widetilde{\alpha}\to\widetilde{\beta}} = \varnothing,\\ &\mathcal{P}_{\widetilde{\gamma}\to\widetilde{\alpha}} = \mathcal{P}_{\widetilde{\alpha}\to\widetilde{\gamma}} = \varnothing,\\ &\mathcal{P}_{\widetilde{\gamma}\to\widetilde{\beta}} = \mathcal{P}_{\widetilde{\beta}\to\widetilde{\gamma}} = \varnothing,\end{aligned} \tag{64}$$

according to their definitions in (29)–(34). Therefore, the cardinalities of the sets in (64) are given by

$$\begin{aligned}|\mathcal{P}_{\mathrm{d}}| &= \frac{n}{gc}(g-r+1),\\ \left|\mathcal{P}_{\widetilde{\beta}\to\widetilde{\alpha}}\right| &= |\mathcal{P}_{\widetilde{\alpha}\to\widetilde{\gamma}}| = 0,\\ |\mathcal{P}_{\widetilde{\gamma}\to\widetilde{\alpha}}| &= |\mathcal{P}_{\widetilde{\alpha}\to\widetilde{\gamma}}| = 0,\\ \left|\mathcal{P}_{\widetilde{\gamma}\to\widetilde{\beta}}\right| &= \left|\mathcal{P}_{\widetilde{\beta}\to\widetilde{\gamma}}\right| = 0.\end{aligned} \tag{65}$$

For each $X_{\langle k\rangle}$ where $k\in\mathcal{S}_\ell$, the probability that the negative log-likelihood of $X_{\langle k\rangle}$ is greater than that of $X_0$ is upper bounded by

$$\begin{aligned}&\mathbb{P}\left[\mathsf{L}(X_{\langle k\rangle}) > \mathsf{L}(X_0)\right]\\ &= 1-\mathbb{P}\left[\mathsf{L}(X_{\langle k\rangle}) \leq \mathsf{L}(X_0)\right]\\ &= 1-\mathbb{P}\Bigg[\log\left((q-1)\frac{1-\theta}{\theta}\right)\\ &\qquad \times\sum_{i\in\mathcal{P}_{\mathrm{d}}}\mathsf{B}_i^{(p)}\left(\left(1+\mathsf{B}_i^{\left(\frac{1}{q-1}\right)}\right)\mathsf{B}_i^{(\theta)}-1\right) \geq 0\Bigg] &(66)\\ &= 1-\mathbb{P}\Bigg[\log\left((q-1)\frac{1-\theta}{\theta}\right)\\ &\qquad \times\sum_{i=1}^{\frac{n}{gc}(g-r+1)}\mathsf{B}_i^{(p)}\left(\left(1+\mathsf{B}_i^{\left(\frac{1}{q-1}\right)}\right)\mathsf{B}_i^{(\theta)}-1\right) \geq 0\Bigg] &(67)\\ &\leq 1-\frac{1}{4}\exp\left(-(1+o(1))\frac{g-r+1}{gc}nI_r\right) &(68)\\ &\leq \exp\left[-\frac{1}{4}\exp\left(-(1+o(1))\frac{g-r+1}{gc}nI_r\right)\right], &(69)\end{aligned}$$

where (66) follows from Lemma 3 and (64); (67) follows from (65); and (68) is an immediate consequence of Lemma 8.

Finally, since $s_\ell$ is bounded away from zero, the probability of exact rating matrix recovery is upper bounded by

$$\begin{aligned}\mathbb{P}[S] &\leq \mathbb{P}\left[\bigcap_{k\in\mathcal{S}_\ell}\left(\mathsf{L}(X_{\langle k\rangle}) > \mathsf{L}(X_0)\right)\right] &(70)\\ &= \prod_{k\in\mathcal{S}_\ell}\mathbb{P}\left[\mathsf{L}(X_{\langle k\rangle}) > \mathsf{L}(X_0)\right] &(71)\end{aligned}$$

$$\leq \left(\exp\left[-\frac{1}{4}\exp\left(-(1+o(1))\frac{g-r+1}{gc}nI_r\right)\right]\right)^{s_\ell m} \quad (72)$$

$$= \exp\left[-\frac{1}{4}s_\ell \exp\left(-(1+o(1))\frac{g-r+1}{gc}nI_r + \log m\right)\right]$$

$$\leq \exp\left[-\frac{1}{4}s_\ell \exp\Big(-\big((1+o(1))(1-\epsilon)-1\big)\log m\Big)\right] \quad (73)$$

$$\leq \exp\left[-\frac{1}{4}s_\ell \exp\Big(\big(\epsilon - o(1)(1-\epsilon)\big)\log m\Big)\right], \quad (74)$$

where (70) follows from the definition in (58); (71) holds since the events $\{\mathsf{L}(X_{\langle k\rangle}) > \mathsf{L}(X_0) : k \in \mathcal{S}_\ell\}$ are mutually independent due to the fact that each event corresponds to a different column $k$ within $\mathcal{S}_\ell$; (72) follows from (69); and (73) follows from the condition in (55). Therefore, we obtain

$$\lim_{n,m\to\infty} \mathbb{P}[S] \leq \lim_{n,m\to\infty} \exp\left[-\frac{1}{4}s_\ell \exp\Big(\big(\epsilon - o(1)(1-\epsilon)\big)\log m\Big)\right] = 0, \quad (75)$$

which shows that if the condition in (55) holds, then the ML estimator will fail to find $X_0$ with high probability.

*2) Failure in the Grouping-Limited Regime:* Without loss of generality, assume $\delta_g m = d_{\mathrm{H}}\left(u_1^{(1)}, u_2^{(1)}\right)$, i.e., the rating vectors of groups 1 and 2 in cluster 1 have the minimum Hamming distance among distinct pairs of rating vectors of groups within the same cluster. In this proof, we introduce a class of rating matrices, which are obtained by switching two users between groups 1 and 2 in cluster 1. Then, we prove that if (56) holds, then, with high probability, the ML estimator will fail by selecting one of the rating matrices from this class instead of $X_0$.

Applying Lemma 9 to $(x,i) = (1,1)$ and $(y,j) = (1,2)$, we conclude that there exist some subsets $\widetilde{Z}_0(1,1) \subset Z_0(1,1)$ and $\widetilde{Z}_0(1,2) \subset Z_0(1,2)$ with $|\widetilde{Z}_0(1,1)| = |\widetilde{Z}_0(1,2)| = \frac{n}{\log^3 n}$, such that the subgraph induced by all the vertices in the set $\widetilde{Z}_0(1,1) \cup \widetilde{Z}_0(1,2)$ is edge-free. Define $X_{\langle a,b\rangle} \in \mathbb{F}_q^{n\times m}$, for $a \in \widetilde{Z}_0(1,1)$ and $b \in \widetilde{Z}_0(1,2)$, as a rating matrix that is identical to $X_0$ except for its $a$th and $b$th rows, which are swapped. More formally, the entries of $X_{\langle a,b\rangle}$ are given by

$$X_{\langle a,b\rangle}(r,:) = \begin{cases} X_0(b,:) = u_2^{(1)} & \text{if } r = a, \\ X_0(a,:) = u_1^{(1)} & \text{if } r = b, \\ X_0(r,:) & \text{otherwise.} \end{cases} \quad (76)$$

The user partitioning $\mathcal{Z}_{\langle a,b\rangle}$ induced by $X_{\langle a,b\rangle}$ is given by

$$Z_{\langle a,b\rangle}(x,i) = \begin{cases} Z_0(1,1) \cup \{b\} \setminus \{a\} & \text{if } (x,i) = (1,1), \\ Z_0(1,2) \cup \{a\} \setminus \{b\} & \text{if } (x,i) = (1,2), \\ Z_0(x,i) & \text{otherwise.} \end{cases} \quad (77)$$

Furthermore, given $X_0$ and $X_{\langle a,b\rangle}$, we have

$$\mathcal{P}_{\mathrm{d}} = \left\{(r,t) : r \in \{a,b\}, t \in [m],\ X_{\langle a,b\rangle}(r,t) \neq X_0(r,t)\right\},$$
$$\mathcal{P}_{\widetilde{\beta}\to\widetilde{\alpha}} = \{(a,h) : h \in Z_0(2,1)\setminus\{b\}\} \cup \{(b,h) : h \in Z_0(1,1)\setminus\{a\}\},$$
$$\mathcal{P}_{\widetilde{\alpha}\to\widetilde{\beta}} = \{(a,h) : h \in Z_0(1,1)\setminus\{a\}\} \cup \{(b,h) : h \in Z_0(1,2)\setminus\{b\}\},$$
$$\mathcal{P}_{\widetilde{\gamma}\to\widetilde{\alpha}} = \mathcal{P}_{\widetilde{\alpha}\to\widetilde{\gamma}} = \varnothing,$$
$$\mathcal{P}_{\widetilde{\gamma}\to\widetilde{\beta}} = \mathcal{P}_{\widetilde{\beta}\to\widetilde{\gamma}} = \varnothing, \quad (78)$$

according to their definitions in (29)–(34). Therefore, the cardinalities of the sets in (78) are given by

$$\begin{aligned} |\mathcal{P}_{\mathrm{d}}| &= d_{\mathrm{H}}\left(X_{\langle a,b\rangle}(a,:), X_0(a,:)\right) + d_{\mathrm{H}}\left(X_{\langle a,b\rangle}(b,:), X_0(b,:)\right) \\ &= d_{\mathrm{H}}\left(X_0(b,:), X_0(a,:)\right) + d_{\mathrm{H}}\left(X_0(a,:), X_0(b,:)\right) \\ &= 2\delta_g m, \\ \left|\mathcal{P}_{\widetilde{\beta}\to\widetilde{\alpha}}\right| &= |\mathcal{P}_{\widetilde{\alpha}\to\widetilde{\gamma}}| = 2\left(\frac{n}{cg} - 1\right), \\ |\mathcal{P}_{\widetilde{\gamma}\to\widetilde{\alpha}}| &= |\mathcal{P}_{\widetilde{\alpha}\to\widetilde{\gamma}}| = 0, \\ \left|\mathcal{P}_{\widetilde{\gamma}\to\widetilde{\beta}}\right| &= \left|\mathcal{P}_{\widetilde{\beta}\to\widetilde{\gamma}}\right| = 0. \end{aligned} \quad (79)$$

For each $X_{\langle a,b\rangle}$ where $a \in \widetilde{Z}_0(1,1)$ and $b \in \widetilde{Z}_0(1,2)$, we have

$$\begin{aligned} &\mathsf{L}(X_0) - \mathsf{L}\left(X_{\langle a,b\rangle}\right) \\ &= \log\left((q-1)\frac{1-\theta}{\theta}\right) \sum_{i\in\mathcal{P}_{\mathrm{d}}} \mathsf{B}_i^{(p)} \left[\left(1 + \mathsf{B}_i^{\left(\frac{1}{q-1}\right)}\right)\mathsf{B}_i^{(\theta)} - 1\right] \\ &\quad + \log\left(\frac{(1-\widetilde{\beta})\widetilde{\alpha}}{(1-\widetilde{\alpha})\widetilde{\beta}}\right) \sum_{j\in\mathcal{P}_{\widetilde{\beta}\to\widetilde{\alpha}}} \left(\mathsf{B}_j^{(\widetilde{\beta})} - \mathsf{B}_j^{(\widetilde{\alpha})}\right) \quad (80) \\ &= \log\left((q-1)\frac{1-\theta}{\theta}\right) \sum_{i=1}^{2\delta_g m} \mathsf{B}_i^{(p)} \left[\left(1 + \mathsf{B}_i^{\left(\frac{1}{q-1}\right)}\right)\mathsf{B}_i^{(\theta)} - 1\right] \\ &\quad + \log\left(\frac{(1-\widetilde{\beta})\widetilde{\alpha}}{(1-\widetilde{\alpha})\widetilde{\beta}}\right) \sum_{j=1}^{2\left(\frac{n}{cg}-1\right)} \left(\mathsf{B}_j^{(\widetilde{\beta})} - \mathsf{B}_j^{(\widetilde{\alpha})}\right), \quad (81) \end{aligned}$$

where (80) follows from Lemma 3 and (78); and (81) follows from (79). Therefore, the probability that the negative log-likelihood of $X_{\langle a,b\rangle}$ is greater than that of $X_0$ is upper bounded by

$$\begin{aligned} &\mathbb{P}\left[\mathsf{L}\left(X_{\langle a,b\rangle}\right) > \mathsf{L}(X_0)\right] \\ &= 1 - \mathbb{P}\left[\mathsf{L}(X_0) - \mathsf{L}\left(X_{\langle a,b\rangle}\right) \geq 0\right] \\ &\leq 1 - \frac{1}{4}\exp\left[-(1+o(1))\left(2\delta_g m I_r + 2\left(\frac{n}{cg} - 1\right) I_{\alpha,\beta}\frac{\log n}{n}\right)\right] \quad (82) \\ &\leq \exp\left[-\frac{1}{4}\exp\left[-(1+o(1))\left(2\delta_g m I_r + 2\left(\frac{n}{cg} - 1\right) I_{\alpha,\beta}\frac{\log n}{n}\right)\right]\right], \quad (83) \end{aligned}$$

where (82) follows from (81) and Lemma 8.

Finally, the probability of exact rating matrix recovery is upper bounded by (89), shown at the top of the next page, where (84), shown at the top of the next page, follows from the definition in (58); (85), shown at the top of the next page, holds since the events $\{\mathsf{L}\left(X_{\langle a,b\rangle}\right) > \mathsf{L}(X_0) : a \in \widetilde{Z}_0(1,1), b \in \widetilde{Z}_0(1,2)\}$ are mutually independent, which is a consequence of the fact that there are no edges among the vertices in $\widetilde{Z}_0(1,1) \cup \widetilde{Z}_0(1,2)$, as per Lemma 9; (86), shown at the top of the next page, follows from (83); in (87), shown at the top of the next page, we used the fact that $|\widetilde{Z}_0(1,1)| = |\widetilde{Z}_0(1,2)| = \frac{n}{\log^3 n}$; and (88), shown at the top of the next page, follows from the

$$\mathbb{P}[S] \leq \mathbb{P}\left[\bigcap_{\substack{a\in\widetilde{Z}_0(1,1)\\ b\in\widetilde{Z}_0(1,2)}} \left(\mathsf{L}\left(X_{\langle a,b\rangle}\right) > \mathsf{L}(X_0)\right)\right] \tag{84}$$

$$= \prod_{\substack{a\in\widetilde{Z}_0(1,1)\\ b\in\widetilde{Z}_0(1,2)}} \mathbb{P}\left[\mathsf{L}\left(X_{\langle a,b\rangle}\right) > \mathsf{L}(X_0)\right] \tag{85}$$

$$\leq \left(\exp\left[-\frac{1}{4}\exp\left[-(1+o(1))\left(2\delta_g m I_r + 2\left(\frac{n}{cg}-1\right)I_{\alpha,\beta}\frac{\log n}{n}\right)\right]\right]\right)^{\left|\widetilde{Z}_0(1,1)\right|\cdot\left|\widetilde{Z}_0(1,2)\right|} \tag{86}$$

$$= \exp\left[-\frac{n^2}{4\log^6 n}\exp\left[-(1+o(1))\left(2\delta_g m I_r + 2\left(\frac{n}{cg}-1\right)I_{\alpha,\beta}\frac{\log n}{n}\right)\right]\right] \tag{87}$$

$$\leq \exp\left(-\frac{n^2}{4\log^6 n}\exp\left(-2(1+o(1))(1-\epsilon)\log n\right)\right) \tag{88}$$

$$\leq \exp\left(-\frac{n^{2(\epsilon-o(1)(1-\epsilon))}}{4\log^6 n}\right). \tag{89}$$

condition in (56). Therefore, we obtain

$$\lim_{n,m\to\infty} \mathbb{P}[S] \leq \lim_{n,m\to\infty} \exp\left(-\frac{n^{2(\epsilon-o(1)(1-\epsilon))}}{4\log^6 n}\right) = 0,$$

which shows that if the condition in (56) holds, then the ML estimator will fail to find $X_0$ with high probability.

*3) Failure in the Clustering-Limited Regime:* The proof follows the same structure as that presented in Section V-B.2 where the condition in (56) holds. Without loss of generality, assume that the rating vectors of group 1 in cluster 1 and group 2 in cluster 2 have the minimum Hamming distance among distinct pairs of rating vectors across different clusters, i.e., $d_{\mathrm{H}}\left(u_1^{(1)}, u_2^{(2)}\right) = \delta_c m$. Note that the corresponding groups defined by such rating vectors belong to different clusters, as opposed to the same cluster in Section V-B.2. In this proof, we introduce a class of rating matrices, which are obtained by switching two users between group 1 in cluster 1 and group 2 in cluster 2. Then, we prove that if (57) holds, then, with high probability, the ML estimator will fail by selecting one of the rating matrices from this class, instead of $X_0$.

We use Lemma 9 for $(x,i) = (1,1)$ and $(y,j) = (2,2)$. This implies that there exist subsets $\widetilde{Z}_0(1,1) \subset Z_0(1,1)$ and $\widetilde{Z}_0(2,2) \subset Z_0(2,2)$ with $|\widetilde{Z}_0(1,1)| = |\widetilde{Z}_0(2,2)| = \frac{n}{\log^3 n}$, such that the subgraph induced by all the vertices in the set $\widetilde{Z}_0(1,1) \cup \widetilde{Z}_0(2,2)$ is edge-free. Similar to (76) and (77) in Section V-B.2, define $X_{\langle a,b\rangle} \in \mathbb{F}_q^{n\times m}$, for $a \in \widetilde{Z}_0(1,1)$ and $b \in \widetilde{Z}_0(2,2)$, as

$$X_{\langle a,b\rangle}(r,:) = \begin{cases} X_0(b,:) = u_2^{(2)} & \text{if } r = a, \\ X_0(a,:) = u_1^{(1)} & \text{if } r = b, \\ X_0(r,:) & \text{otherwise.} \end{cases} \tag{90}$$

The corresponding user partitioning $\mathcal{Z}_{\langle a,b\rangle}$ for the rating matrix in (90) is given by

$$Z_{\langle a,b\rangle}(x,i) = \begin{cases} Z_0(1,1) \cup \{b\} \setminus \{a\} & \text{if } (x,i) = (1,1), \\ Z_0(2,2) \cup \{a\} \setminus \{b\} & \text{if } (x,i) = (2,2), \\ Z_0(x,i) & \text{otherwise.} \end{cases} \tag{91}$$

Furthermore, given $X_0$ and $X_{\langle a,b\rangle}$, we can identify the following sets

$$\begin{aligned}
&\mathcal{P}_{\mathrm{d}} = \left\{(r,t) : r \in \{a,b\}, t \in [m], X_{\langle a,b\rangle}(r,t) \neq X_0(r,t)\right\}, \\
&\mathcal{P}_{\widetilde{\beta}\to\widetilde{\alpha}} = \mathcal{P}_{\widetilde{\alpha}\to\widetilde{\beta}} = \varnothing, \\
&\mathcal{P}_{\widetilde{\gamma}\to\widetilde{\alpha}} = \{(a,h) : h \in Z_0(2,2)\setminus\{b\}\} \cup \{(b,h) : h \in Z_0(1,1)\setminus\{a\}\}, \\
&\mathcal{P}_{\widetilde{\alpha}\to\widetilde{\gamma}} = \{(a,h) : h \in Z_0(1,1)\setminus\{a\}\} \cup \{(b,h) : h \in Z_0(2,2)\setminus\{b\}\}, \\
&\mathcal{P}_{\widetilde{\gamma}\to\widetilde{\beta}} = \left\{(a,h) : h \in \bigcup_{i\in[g]\setminus\{2\}} Z_0(2,i)\right\} \cup \left\{(b,h) : h \in \bigcup_{i\in[g]\setminus\{1\}} Z_0(1,i)\right\}, \\
&\mathcal{P}_{\widetilde{\beta}\to\widetilde{\gamma}} = \left\{(a,h) : h \in \bigcup_{i\in[g]\setminus\{1\}} Z_0(1,i)\right\} \cup \left\{(b,h) : h \in \bigcup_{i\in[g]\setminus\{2\}} Z_0(2,i)\right\},
\end{aligned} \tag{92}$$

according to their definitions in (29)–(34). Thus, the size of the sets in (92) are given by

$$\begin{aligned}
|\mathcal{P}_{\mathrm{d}}| &= d_{\mathrm{H}}\left(X_{\langle a,b\rangle}(a,:), X_0(a,:)\right) + d_{\mathrm{H}}\left(X_{\langle a,b\rangle}(b,:), X_0(b,:)\right) \\
&= d_{\mathrm{H}}\left(X_0(b,:), X_0(a,:)\right) + d_{\mathrm{H}}\left(X_0(a,:), X_0(b,:)\right) \\
&= 2\delta_c m, \\
\left|\mathcal{P}_{\widetilde{\beta}\to\widetilde{\alpha}}\right| &= |\mathcal{P}_{\widetilde{\alpha}\to\widetilde{\gamma}}| = 0, \\
|\mathcal{P}_{\widetilde{\gamma}\to\widetilde{\alpha}}| &= |\mathcal{P}_{\widetilde{\alpha}\to\widetilde{\gamma}}| = 2\left(\frac{n}{cg}-1\right), \\
\left|\mathcal{P}_{\widetilde{\gamma}\to\widetilde{\beta}}\right| &= \left|\mathcal{P}_{\widetilde{\beta}\to\widetilde{\gamma}}\right| = 2\left(\frac{g-1}{gc}\right)n.
\end{aligned} \tag{93}$$

For each $X_{\langle a,b\rangle}$ with $a \in \widetilde{Z}_0(1,1)$ and $b \in \widetilde{Z}_0(2,2)$, we have

$$\begin{aligned}
&\mathsf{L}(X_0) - \mathsf{L}\left(X_{\langle a,b\rangle}\right) \\
&= \log\left((q-1)\frac{1-\theta}{\theta}\right)\sum_{i\in\mathcal{P}_{\mathrm{d}}} \mathsf{B}_i^{(p)}\left[\left(1+\mathsf{B}_i^{\left(\frac{1}{q-1}\right)}\right)\mathsf{B}_i^{(\theta)} - 1\right] \\
&\quad + \left(\log\frac{(1-\widetilde{\gamma})\widetilde{\alpha}}{(1-\widetilde{\alpha})\widetilde{\gamma}}\right)\sum_{i\in\mathcal{P}_{\widetilde{\gamma}\to\widetilde{\alpha}}}\left(\mathsf{B}_i^{(\widetilde{\gamma})} - \mathsf{B}_i^{(\widetilde{\alpha})}\right)
\end{aligned}$$

$$+\left(\log\frac{(1-\widetilde{\gamma})\widetilde{\beta}}{(1-\widetilde{\beta})\widetilde{\gamma}}\right)\sum_{i\in\mathcal{P}_{\widetilde{\gamma}\to\widetilde{\beta}}}\left(\mathsf{B}_i^{(\widetilde{\gamma})}-\mathsf{B}_i^{(\widetilde{\beta})}\right) \tag{94}$$

$$=\log\left((q-1)\frac{1-\theta}{\theta}\right)\sum_{i=1}^{2\delta_c m}\mathsf{B}_i^{(p)}\left[\left(1+\mathsf{B}_i^{\left(\frac{1}{q-1}\right)}\right)\mathsf{B}_i^{(\theta)}-1\right]$$
$$+\left(\log\frac{(1-\widetilde{\gamma})\widetilde{\alpha}}{(1-\widetilde{\alpha})\widetilde{\gamma}}\right)\sum_{i=1}^{2\left(\frac{n}{cg}-1\right)}\left(\mathsf{B}_i^{(\widetilde{\gamma})}-\mathsf{B}_i^{(\widetilde{\alpha})}\right)$$
$$+\left(\log\frac{(1-\widetilde{\gamma})\widetilde{\beta}}{(1-\widetilde{\beta})\widetilde{\gamma}}\right)\sum_{i=1}^{2\left(\frac{g-1}{gc}\right)n}\left(\mathsf{B}_i^{(\widetilde{\gamma})}-\mathsf{B}_i^{(\widetilde{\beta})}\right), \tag{95}$$

where (94) follows from Lemma 3 and (92); and (95) follows from (93). Thus, the probability that the negative log-likelihood of $X_{\langle a,b\rangle}$ is greater than that of $X_0$ is upper bounded by

$$\mathbb{P}\left[\mathsf{L}\left(X_{\langle a,b\rangle}\right)>\mathsf{L}(X_0)\right]$$
$$=1-\mathbb{P}\left[B\left(2\delta_c m,\,0,\,2\left(\frac{n}{cg}-1\right),\,2\left(\frac{g-1}{gc}\right)n\right)\geq 0\right]$$
$$\leq 1-\frac{1}{4}\exp\left[-(1+o(1))\left(2\delta_c mI_r+2\left(\frac{n}{cg}-1\right)I_{\alpha,\gamma}\frac{\log n}{n}\right.\right.$$
$$\left.\left.+2\left(\frac{g-1}{gc}\right)nI_{\beta,\gamma}\frac{\log n}{n}\right)\right] \tag{96}$$
$$\leq\exp\left[-\frac{1}{4}\exp\left[-(1+o(1))\left(2\delta_c mI_r+2\left(\frac{n}{cg}-1\right)I_{\alpha,\gamma}\frac{\log n}{n}\right.\right.\right.$$
$$\left.\left.\left.+2\left(\frac{g-1}{gc}\right)nI_{\beta,\gamma}\frac{\log n}{n}\right)\right]\right], \tag{97}$$

where (96) follows from (95) and Lemma 8.

Finally, the probability of exact matrix recovery is upper bounded by (102) presented at the bottom of this page. Note that (98), shown at the bottom of the page, follows from (58); (99), shown at the bottom of the page, holds since the events $\{\mathsf{L}\left(X_{\langle a,b\rangle}\right)>\mathsf{L}(X_0): a\in\widetilde{Z}_0(1,1), b\in\widetilde{Z}_0(2,2)\}$ are mutually independent (as there are no edges between the vertices in $\widetilde{Z}_0(1,1)\cup\widetilde{Z}_0(2,2)$), as per Lemma 9; (100), shown at the bottom of the page, follows from (96); and (101), shown at the bottom of the page, follows from the condition in (57), and $|\widetilde{Z}_0(1,1)|=|\widetilde{Z}_0(1,2)|=\frac{n}{\log^3 n}$. Thus, we obtain

$$\lim_{n,m\to\infty}\mathbb{P}[S]=\lim_{n,m\to\infty}\exp\left(-\frac{n^{2(\epsilon-o(1)(1-\epsilon))}}{4\log^6 n}\right)=0,$$

which shows that if the condition in (57) holds, then the ML estimator will fail to find $X_0$ with high probability.

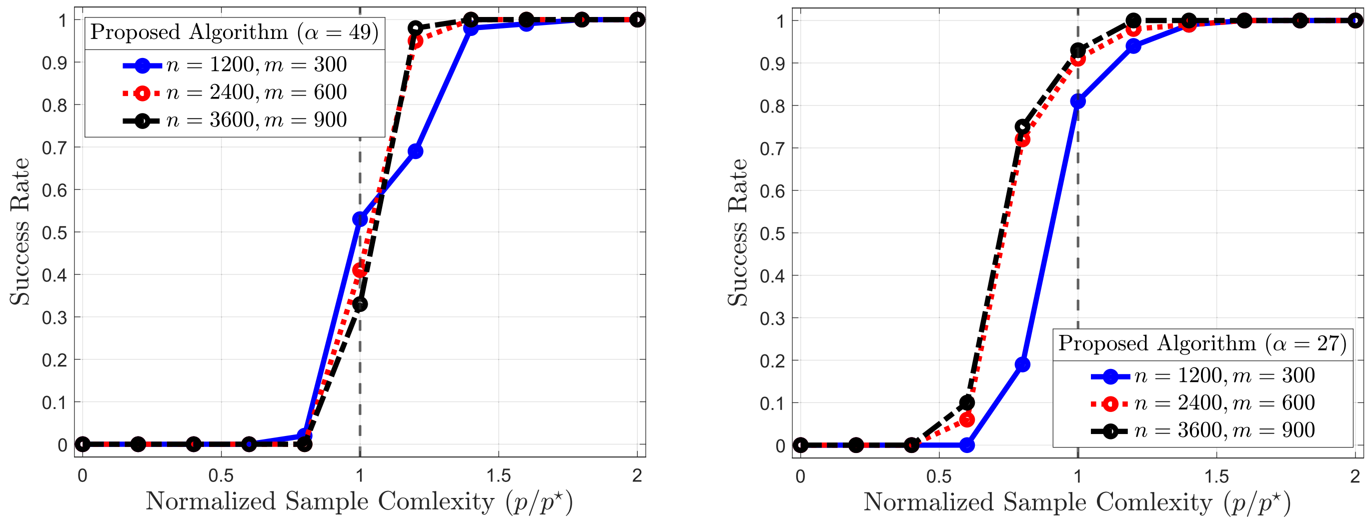
Since $\lim_{n,m\to\infty}\mathbb{P}[S]=0$ is proved under each of the three conditions stated in (55), (56) and (57), the converse proof of Theorem 1 is concluded. ■

## VI. SIMULATION RESULTS

We conduct several Monte Carlo experiments to show that our proposed algorithm achieves $p^*$ characterized by Theorem 1. The proposed algorithm is built in part upon the computationally efficient matrix completion algorithm, proposed in [28]. The idea is to first find a good initial estimate of clusters, groups, and ratings and then successively refine this estimate until the optimal solution is reached. The distinction of our proposed algorithm compared to [28] is the stage of exact recovery of rating vectors, which is based on maximum likelihood (ML) decoding of users' ratings based on the partial and noisy observations. The overview of this phase is as follows: for $x\in[c]$, and $t\in[m]$, we count the number of corresponding ratings between $t$th column of the observation matrix of cluster $x$ and candidate rating vectors that follow the structure of $(g,r)$ MDS code. Then, we set the one rating vector that corresponds the most as an estimated vector.

To formalize the given description in a mathematical framework, let $Z(x,i)$ denote the initial estimation users of cluster $x$ and group $i$, and let $Z(x,:)=\cup_{i\in[g]}Z(x,i)$. The maximum likelihood (ML) decoder is denoted by $\Pi(v)$, which performs the aforementioned comparisons and counting on $v$, and then outputs a column vector, which is denoted by $\widehat{u}_t^{(x)}\in\mathbb{F}_q^{(n/c)\times 1}$. Let the $j$th element of the rating vector $\widehat{u}_t^{(x)}$ be denoted by $\widehat{u}_t^x(j)$, for $x\in[c]$, and $t\in[m]$. The pseudocode of the phase is given by Algorithm 1. The term $\widehat{u}_t^{(x)}\left(\frac{n}{gc}(i-1)+1\right)$ in line 4 of Algorithm 1 refers to

$$\mathbb{P}[S]\leq\mathbb{P}\left[\bigcap_{\substack{a\in\widetilde{Z}_0(1,1)\\ b\in\widetilde{Z}_0(2,2)}}\left(\mathsf{L}\left(X_{\langle a,b\rangle}\right)>\mathsf{L}(X_0)\right)\right] \tag{98}$$
$$=\prod_{\substack{a\in\widetilde{Z}_0(1,1)\\ b\in\widetilde{Z}_0(2,2)}}\mathbb{P}\left[\mathsf{L}\left(X_{\langle a,b\rangle}\right)>\mathsf{L}(X_0)\right] \tag{99}$$
$$\leq\left(\exp\left[-\frac{1}{4}\exp\left[-(1+o(1))\left(2\delta_c mI_r+2\left(\frac{n}{cg}-1\right)I_{\alpha,\gamma}\frac{\log n}{n}+2\left(\frac{g-1}{gc}\right)nI_{\beta,\gamma}\frac{\log n}{n}\right)\right]\right]\right)^{\left|\widetilde{Z}_0(1,1)\right|\cdot\left|\widetilde{Z}_0(2,2)\right|} \tag{100}$$
$$\leq\exp\left(-\frac{n^2}{4\log^6 n}\exp\left(-2(1+o(1))(1-\epsilon)\log n\right)\right) \tag{101}$$
$$\leq\exp\left(-\frac{n^{2(\epsilon-o(1)(1-\epsilon))}}{4\log^6 n}\right). \tag{102}$$




(a) (b)

Fig. 3. The success rate of the proposed algorithm as a function of $p/p^\star$ for different values of $n$, $m$, and $\alpha$. The problem setting is characterized by $(c, g, q, r) = (3, 4, 5, 3)$, $\theta = 0.01$, $(\beta, \gamma) = (9, 0.5)$, and $(\delta_g, \delta_c) = (1/3, 1/3)$. The MDS code structure is given by $u_4^{(x)} = u_1^{(x)} + u_2^{(x)} + u_3^{(x)}$ for $x \in [3]$. We study the two cases: (a) perfect clustering/grouping regime ($\alpha = 49$); and (b) grouping-limited regime where ($\alpha = 27$).

**Algorithm 1** Exact Recovery of Rating Vectors

**function** VECRCV $(n, m, Y)$
  **for** $x \in [c]$ and $t \in [m]$ **do**
    $\widehat{u}_t^{(x)} \leftarrow \Pi(Y(Z(x, :), t))$
    $\widehat{v}_i^{(x)}(t) \leftarrow \widehat{u}_t^{(x)}\left(\frac{n}{gc}(i-1)+1\right)$
  **end for**
  **return** $\{\widehat{v}_i^{(x)} : x \in [c], i \in [g]\}$
**end function**

the rating of group $i$. Note that pseudocodes of other phases can be obtained just by replacing the number of clusters and groups with $c$ and $g$ in the pseudocodes in [28], respectively. Thus, we omit the pseudocode of other phases in this paper.

The synthetic data is generated as per the model in Section II. We consider a problem setting in which we have $c = 3$ clusters, $g = 4$ groups per cluster, finite field of order $q = 5$, and $r = 3$ basis vectors per group. The MDS code structure is given by $u_4^{(x)} = u_1^{(x)} + u_2^{(x)} + u_3^{(x)}$ for $x \in [3]$. Furthermore, the parameters of observation noise, graph, and rating vectors are set to $\theta = 0.01$, $(\beta, \gamma) = (9, 0.5)$ and $(\delta_g, \delta_c) = (1/3, 1/3)$, respectively.

In Figs. 3a and 3b, we evaluate the performance of the proposed algorithm and quantify the empirical success rate as a function of the normalized sample complexity $p/p^\star$ over 100 randomly drawn realizations of rating vectors and hierarchical graphs. The results are reported for various values of $n$ and $m$, while the ratio $n/m = 4$ is preserved. Fig. 3a depicts the case of $\alpha = 49$, which corresponds to the perfect clustering/grouping regime, while Fig. 3b illustrates the case of $\alpha = 27$, which corresponds to the grouping-limited regime. In both figures, we observe a phase transition[8] in the success rate at $p = p^\star$, and the phase transition gets sharper as $n$ and $m$ increase. Figs. 3a and 3b imply that our proposed algorithm achieves $p^*$ characterized by Theorem 1 in different regimes when the graph side information is not scarce.

Next, we highlight the sample complexity gain from leveraging the relational structure among the rating vectors. We compare the performance of the proposed algorithm against that of [24], which does not consider the relational structure among the rating vectors under the two different settings. In Figs. 4a and 4b, we consider a problem setting in which we have $n = 1200$ users, $m = 300$ items, $c = 2$ clusters, $g = 3$ groups per cluster, $r = 2$ basis vectors per group, and finite field of size $q = 2$. The MDS code structure is given by $u_3^{(x)} = u_1^{(x)} + u_2^{(x)}$ for $x \in [2]$. In Figs. 4c and 4d, we set $(n, m, \theta, \gamma, c, g, r, q) = (2400, 600, 0, 0.5, 3, 4, 3, 5)$, and the MDS code structure is given by $u_4^{(x)} = u_1^{(x)} + u_2^{(x)} + u_3^{(x)}$ for $x \in [3]$. In both cases, the parameters of observation noise and rating vectors are set to $\theta = 0$ and $(\delta_g, \delta_c) = (1/2, 1/2)$, respectively. Figs. 4a and 4c depict the success rates of the proposed algorithm under various values of $p$ and $I_{\alpha,\beta}$, while Figs. 4b and 4d depict those of the algorithm presented in [25].

We set the range of values of $I_{\alpha,\beta}$ to span the grouping-limited and perfect clustering/grouping regimes, where we set $\gamma = 0.5$. The empirical success rate is depicted by the grayscale heat map, averaged over 100 randomly drawn realizations of rating vectors and hierarchical graphs. The orange line reflects the optimal sample complexity characterized by Theorem 1, where the vertical line implies the sample complexity in the perfect clustering/grouping regime, while the diagonal line means the sample complexity in the grouping-limited regime. In Fig. 4a and 4c, the phase transition occurs near the vertical line in the perfect clustering/grouping regime. However, the transition in Fig. 4b and 4d does not occur at the optimal observation probability $p^*$ given in Theorem 1, which demonstrates the sample complexity gain resulting from leveraging the relational structure among the rating vectors.

[8]The transition is ideally a step function at $p = p^\star$ as $n, m \to \infty$.

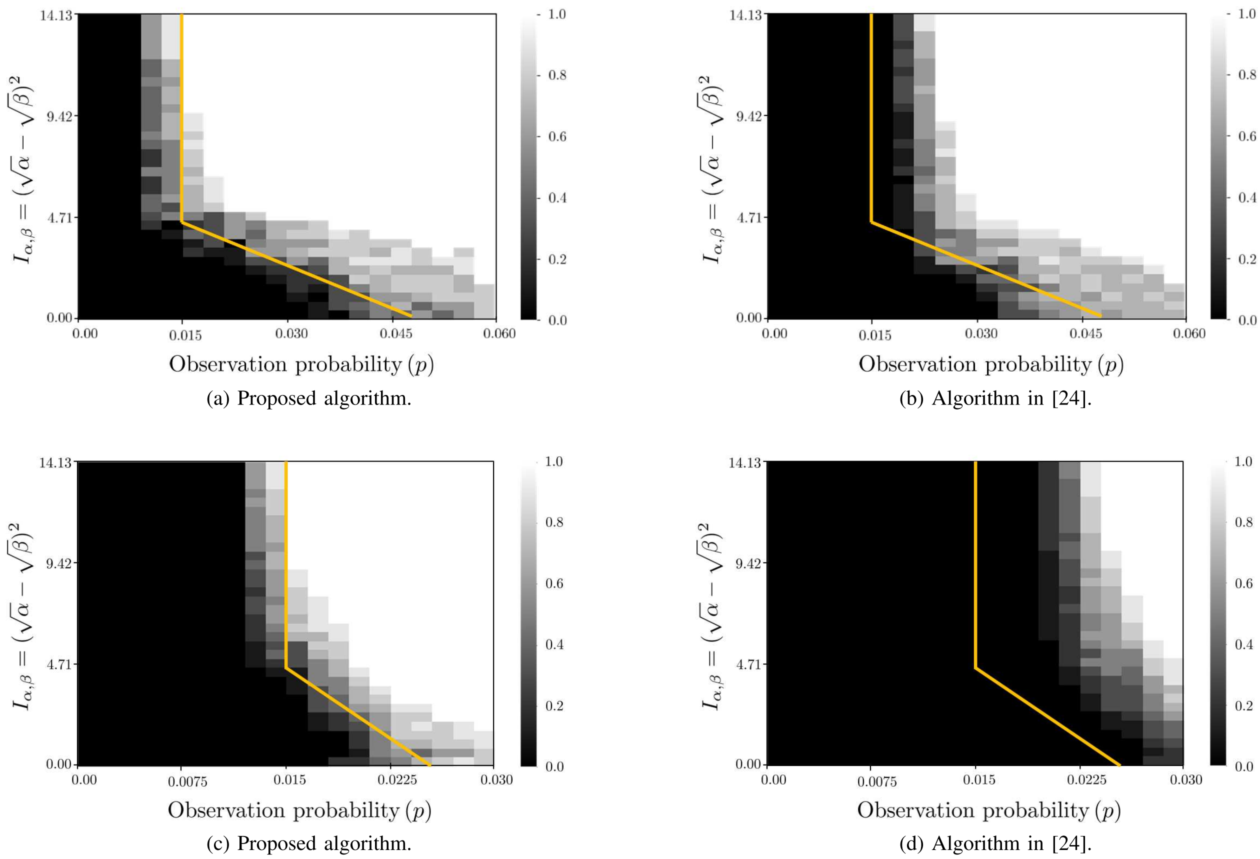



Fig. 4. A comparison between the success rates of the proposed algorithm, denoted as (a) and (c), in contrast to those presented in [24], labeled as (b) and (d). The problem setting for cases (a) and (b) is characterized by $(n, m, \theta, \gamma, c, g, r, q) = (1200, 300, 0, 0.5, 2, 3, 2, 2)$, and $(\delta_g, \delta_c) = (1/2, 1/2)$. On the other hand, for cases (c) and (d), the problem is defined by $(n, m, \theta, \gamma, c, g, r, q) = (2400, 600, 0, 0.5, 3, 4, 3, 5)$, and $(\delta_g, \delta_c) = (1/2, 1/2)$.

## VII. Conclusion

In this paper, we consider a rating matrix that consists of $n$ users and $m$ items, and a hierarchical similarity graph that consists of $c$ disjoint clusters, and each cluster comprises $g$ disjoint groups. The rating vectors of the groups in a given cluster are different but related to each other through a linear subspace of $r$ basis vectors. We characterize the optimal sample complexity to jointly recover the hierarchical structure of the similarity graph as well as the rating matrix entries. We propose a matrix completion algorithm that is based on the maximum likelihood estimation and achieves the characterized sample complexity. The optimality of the proposed achievable scheme was demonstrated through a matching converse proof. We demonstrate that the optimal sample complexity hinges on the quality of side information of the hierarchical similarity graph. We also highlight the fact that leveraging the graph side information enables us to achieve a significant gain in sample complexity, compared to existing schemes that identify different groups without taking into consideration the hierarchical structure across them.

An important research follow-up direction is to develop a computationally efficient algorithm to achieve the sharp threshold on the optimal sample complexity characterized in this paper. Another research direction is to characterize the optimal sample complexity for a more general case of $c$ clusters, each of which comprises an arbitrary number of groups of possibly different numbers of users.

## Appendix A
## Proof of Lemma 1

From the definition in (19), the negative log-likelihood of a candidate rating matrix $X = (\mathcal{V}, \mathcal{Z})$, for $X \in \mathcal{M}^{(\delta)}$, given a fixed input pair $(Y, \mathcal{G})$ can be written as

$$\begin{aligned}\mathsf{L}(X) &= -\log \mathbb{P}\left[(Y, \mathcal{G}) \mid \mathbf{X} = X\right] \\ &= -\log\left(\mathbb{P}\left[Y \mid \mathbf{X} = X\right] \mathbb{P}\left[\mathcal{G} \mid \mathbf{X} = X\right]\right) \\ &= -\log \mathbb{P}\left[Y \mid \mathbf{X} = X\right] - \log \mathbb{P}\left[\mathcal{G} \mid \mathbf{X} = X\right], \end{aligned} \tag{103}$$

where

$$\begin{aligned}\mathbb{P}\left[Y|\mathbf{X} = X\right] &= p^{|\Omega|}\left(1-p\right)^{nm-|\Omega|}\left(\frac{\theta}{q-1}\right)^{\Lambda(Y,X)} \\ &\quad \times (1-\theta)^{|\Omega|-\Lambda(Y,X)}, \end{aligned} \tag{104}$$

$$\begin{aligned}\mathbb{P}\left[\mathcal{G} \mid \mathbf{X} = X\right] &= \widetilde{\alpha}^{e_\alpha(\mathcal{G},Z)}(1-\widetilde{\alpha})^{|\mathcal{P}_\alpha(\mathcal{Z})|-e_\alpha(\mathcal{G},Z)} \\ &\quad \times \widetilde{\beta}^{e_\beta(\mathcal{G},Z)}(1-\widetilde{\beta})^{|\mathcal{P}_\beta(\mathcal{Z})|-e_\beta(\mathcal{G},Z)} \\ &\quad \times \widetilde{\gamma}^{e_\gamma(\mathcal{G},Z)}(1-\widetilde{\gamma})^{|\mathcal{P}_\gamma(\mathcal{Z})|-e_\gamma(\mathcal{G},Z)}. \end{aligned} \tag{105}$$

Consequently, $\mathsf{L}(X)$ is given by

$$\mathsf{L}(X)=\log\left((q-1)\frac{1-\theta}{\theta}\right)\Lambda\left(Y,X\right)+\sum_{\mu\in\{\widetilde{\alpha},\widetilde{\beta},\widetilde{\gamma}\}}\left[\log\left(\frac{1-\mu}{\mu}\right)e_{\mu}\left(\mathcal{G},\mathcal{Z}\right)-\log(1-\mu)|\mathcal{P}_{\mu}\left(\mathcal{Z}\right)|\right]. \tag{106}$$

This completes the proof of Lemma 1. ∎

## Appendix B
## Proof of Lemma 2

The worst-case probability of error $P_e^{(\delta)}(\psi_{\text{ML}})$ for the maximum likelihood estimator $\psi_{\text{ML}}$ is upper bounded by

$$\begin{aligned}P_e^{(\delta)}(\psi_{\text{ML}}) &= \max_{M\in\mathcal{M}^{(\delta)}}\mathbb{P}\left[\psi_{\text{ML}}(Y,\mathcal{G})\neq M\right]\\ &=\mathbb{P}\left[\psi_{\text{ML}}(Y,\mathcal{G})\neq X_0\mid \mathbf{M}=X_0\right] &(107)\\ &=\mathbb{P}\left[\bigcup_{X\neq X_0}\mathsf{L}(X)\leq\mathsf{L}(X_0)\right] &(108)\\ &\leq\sum_{X\neq X_0}\mathbb{P}\left[\mathsf{L}(X)\leq\mathsf{L}(X_0)\right], &(109)\end{aligned}$$

where (107) holds since $X_0\in\mathcal{M}^{(\delta)}$ by the construction of $X_0$ presented in Section IV, and the error event $\{\psi_{\text{ML}}(Y,\mathcal{G})\neq M\}$ is statistically identical over all $M\in\mathcal{M}^{(\delta)}$; (108) follows from the fact that the output of the maximum likelihood estimator is different from the ground truth rating matrix $X_0$ only if there exists a candidate rating matrix $X$ whose negative log-likelihood is less than or equal to that of $X_0$; and (109) follows from the union bound. This completes the proof of Lemma 2. ∎

## Appendix C
## Proof of Lemma 3

By Lemma 1, the LHS of (37) can be written as

$$\begin{aligned}&\mathsf{L}\left(X_0\right)-\mathsf{L}(X)\\ &=\log\left((q-1)\frac{1-\theta}{\theta}\right)\underbrace{\left(\Lambda\left(Y,X_0\right)-\Lambda\left(Y,X\right)\right)}_{\mathsf{Term}_1}\\ &\quad+\underbrace{\sum_{\mu\in\{\widetilde{\alpha},\widetilde{\beta},\widetilde{\gamma}\}}\left[\log\left(\frac{1-\mu}{\mu}\right)\left(e_{\mu}\left(\mathcal{G},\mathcal{Z}_0\right)-e_{\mu}\left(\mathcal{G},\mathcal{Z}\right)\right)\right]}_{\mathsf{Term}_2}\\ &\quad+\underbrace{\sum_{\mu\in\{\widetilde{\alpha},\widetilde{\beta},\widetilde{\gamma}\}}\left[\log(1-\mu)\left(|\mathcal{P}_{\mu}\left(\mathcal{Z}\right)|-|\mathcal{P}_{\mu}\left(\mathcal{Z}_0\right)|\right)\right]}_{\mathsf{Term}_3}. &(110)\end{aligned}$$

In what follows, we evaluate each of the three terms in (110).

Recall from Section I-B that $\Lambda\left(A,B\right)$ denotes the number of different entries between the matrices $A_{n\times m}$ and $B_{n\times m}$. Therefore, $\mathsf{Term}_1$ can be expanded as

$$\begin{aligned}&\mathsf{Term}_1\\ &=\Lambda\left(Y,X_0\right)-\Lambda\left(Y,X\right)\\ &=\sum_{(r,t)\in\Omega}\left(\mathbb{1}\left[Y(r,t)\neq X_0(r,t)\right]\right)-\sum_{(r,t)\in\Omega}\left(\mathbb{1}\left[Y(r,t)\neq X(r,t)\right]\right)\\ &=\sum_{(r,t)\in\Omega}(nm-\mathbb{1}\left[Y(r,t)=X_0(r,t)\right])-(nm-\mathbb{1}\left[Y(r,t)=X(r,t)\right])\\ &=\sum_{\substack{(r,t)\in\Omega:\\ X(r,t)\neq X_0(r,t)}}\mathbb{1}\left[Y(r,t)=X(r,t)\right]-\mathbb{1}\left[Y(r,t)=X_0(r,t)\right] &(111)\\ &=\sum_{\substack{i\in|\{(r,t)\in[n]\times[m]:\\ X(r,t)\neq X_0(r,t)\}|}}\mathsf{B}_i^{(p)}\mathsf{B}_i^{\left(\frac{\theta}{q-1}\right)}-\mathsf{B}_i^{(p)}\left(1-\mathsf{B}_i^{(\theta)}\right) &(112)\\ &=\sum_{\substack{i\in|\{(r,t)\in[n]\times[m]:\\ X(r,t)\neq X_0(r,t)\}|}}\mathsf{B}_i^{(p)}\left[\mathsf{B}_i^{(\theta)}\mathsf{B}_i^{\left(\frac{1}{q-1}\right)}-\left(1-\mathsf{B}_i^{(\theta)}\right)\right] &(113)\\ &=\sum_{i\in\mathcal{P}_{\text{d}}}\mathsf{B}_i^{(p)}\left[\left(1+\mathsf{B}_i^{\left(\frac{1}{q-1}\right)}\right)\mathsf{B}_i^{(\theta)}-1\right], &(114)\end{aligned}$$

where (111) follows since

$$\mathbb{1}\left[Y(i,j)=X(i,j)\right]=\mathbb{1}\left[Y(i,j)=X_0(i,j)\right]$$

if $X(i,j)=X_0(i,j)$, the first term of each summand in (112) follows since the probability that the observed rating matrix entry is $X(i,j)$, which is not equal to $X_0(i,j)$, is $p(\theta/(q-1))$, while the second term of each summand in (112) follows since the probability that the observed rating matrix entry is $X_0(i,j)$ is $p(1-\theta)$, for every $(i,j)\in[n]\times[m]$; (113) follows since $\mathsf{B}_l^{(\theta/(q-1))}=\mathsf{B}_l^{(\theta)}\mathsf{B}_l^{(1/(q-1))}$; and finally (114) follows from (29).

Next, we expand $\mathsf{Term}_2$ in (110). We first evaluate the quantity $e_{\alpha}\left(\mathcal{G},\mathcal{Z}_0\right)-e_{\alpha}\left(\mathcal{G},Z\right)$ as (118), shown at the top of the next page, where (117), shown at the top of the next page, holds since the edges that remain after the subtraction are: (i) edges that exist in the same group in $\mathcal{Z}_0$, but are estimated to be in different groups within the same cluster in $\mathcal{Z}$; (ii) edges that exist in the same group in $\mathcal{Z}_0$, but are estimated to be in different clusters in $\mathcal{Z}$; (iii) edges that exist in different groups within the same cluster in $\mathcal{Z}_0$, but are estimated to be in the same group in $\mathcal{Z}$; (iv) edges that exist in different clusters in $\mathcal{Z}_0$, but are estimated to be in the same group in $\mathcal{Z}$; and finally, (118) follows from (31)–(32). In a similar way, one can evaluate the following quantities:

$$\begin{aligned}&e_{\widetilde{\beta}}\left(\mathcal{G},\mathcal{Z}_0\right)-e_{\widetilde{\beta}}\left(\mathcal{G},\mathcal{Z}\right)\\ &=\sum_{i=1}^{\mathcal{P}_{\widetilde{\beta}\to\widetilde{\alpha}}}\mathsf{B}_i^{(\widetilde{\beta})}+\sum_{i=1}^{\mathcal{P}_{\widetilde{\beta}\to\widetilde{\gamma}}}\mathsf{B}_i^{(\widetilde{\beta})}-\sum_{i=1}^{\mathcal{P}_{\widetilde{\alpha}\to\widetilde{\beta}}}\mathsf{B}_i^{(\widetilde{\alpha})}-\sum_{i=1}^{\mathcal{P}_{\widetilde{\gamma}\to\widetilde{\beta}}}\mathsf{B}_i^{(\widetilde{\gamma})}, &(115)\\ &e_{\widetilde{\gamma}}\left(\mathcal{G},\mathcal{Z}_0\right)-e_{\widetilde{\gamma}}\left(\mathcal{G},\mathcal{Z}\right)\\ &=\sum_{i=1}^{\mathcal{P}_{\widetilde{\gamma}\to\widetilde{\alpha}}}\mathsf{B}_i^{(\widetilde{\gamma})}+\sum_{i=1}^{\mathcal{P}_{\widetilde{\gamma}\to\widetilde{\beta}}}\mathsf{B}_i^{(\widetilde{\gamma})}-\sum_{i=1}^{\mathcal{P}_{\widetilde{\alpha}\to\widetilde{\gamma}}}\mathsf{B}_i^{(\widetilde{\alpha})}-\sum_{i=1}^{\mathcal{P}_{\widetilde{\beta}\to\widetilde{\gamma}}}\mathsf{B}_i^{(\widetilde{\beta})}. &(116)\end{aligned}$$

Consequently, $\mathsf{Term}_2$ can be written as (120), shown at the top of the next page, where (119), shown at the top of the next page, follows from (116)–(118).

$$\begin{aligned}
&e_\alpha(\mathcal{G},\mathcal{Z}_0)-e_\alpha(\mathcal{G},Z)\\
&=|\{(a,b)\in\mathcal{E}: a\in Z_0(x,i)\cap Z(y,j_1), b\in Z_0(x,i)\cap Z(y,j_2),\ \text{for } x,y\in[c], i,j_1,j_2\in[g], j_1\neq j_2\}|\\
&\quad+|\{(a,b)\in\mathcal{E}: a\in Z_0(x,i)\cap Z(y_1,j_1), b\in Z_0(x,i)\cap Z(y_2,j_2),\ \text{for } x,y\in[c],\ y_1\neq y_2,\ i,j_1,j_2\in[g]\}|\\
&\quad-|\{(a,b)\in\mathcal{E}: a\in Z_0(x,i_1)\cap Z(y,j), b\in Z_0(x,i_2)\cap Z(y,j),\ \text{for } x,y\in[c],\ i_1,i_2,j\in[g],\ i_1\neq i_2\}|\\
&\quad-|\{(a,b)\in\mathcal{E}: a\in Z_0(x_1,i_1)\cap Z(y,j), b\in Z_0(x_2,i_2)\cap Z(y,j),\ \text{for } x_1,x_2,y\in[c],\ x_1\neq x_2,\ i_1,i_2,j\in[g]\}| \quad (117)\\
&=\sum_{i=1}^{\mathcal{P}_{\widetilde{\alpha}\to\widetilde{\beta}}}\mathsf{B}_i^{(\widetilde{\alpha})}+\sum_{i=1}^{\mathcal{P}_{\widetilde{\alpha}\to\widetilde{\gamma}}}\mathsf{B}_i^{(\widetilde{\alpha})}-\sum_{i=1}^{\mathcal{P}_{\widetilde{\beta}\to\widetilde{\alpha}}}\mathsf{B}_i^{(\widetilde{\beta})}-\sum_{i=1}^{\mathcal{P}_{\widetilde{\gamma}\to\widetilde{\alpha}}}\mathsf{B}_i^{(\widetilde{\gamma})}. \quad (118)
\end{aligned}$$

---

$$\begin{aligned}
\mathsf{Term}_2&=\sum_{\mu\in\widetilde{\alpha},\widetilde{\beta},\widetilde{\gamma}}\left[\log\left(\frac{1-\mu}{\mu}\right)(e_\mu(\mathcal{G},\mathcal{Z}_0)-e_\mu(\mathcal{G},\mathcal{Z}))\right]\\
&=\log\left(\frac{1-\widetilde{\alpha}}{\widetilde{\alpha}}\right)\left(\sum_{i\in\mathcal{P}_{\widetilde{\alpha}\to\widetilde{\beta}}}\mathsf{B}_i^{(\widetilde{\alpha})}+\sum_{i\in\mathcal{P}_{\widetilde{\alpha}\to\widetilde{\gamma}}}\mathsf{B}_i^{(\widetilde{\alpha})}-\sum_{i\in\mathcal{P}_{\widetilde{\beta}\to\widetilde{\alpha}}}\mathsf{B}_i^{(\widetilde{\beta})}-\sum_{i\in\mathcal{P}_{\widetilde{\gamma}\to\widetilde{\alpha}}}\mathsf{B}_i^{(\widetilde{\gamma})}\right)\\
&\quad+\log\left(\frac{1-\widetilde{\beta}}{\widetilde{\beta}}\right)\left(\sum_{i\in\mathcal{P}_{\widetilde{\beta}\to\widetilde{\alpha}}}\mathsf{B}_i^{(\widetilde{\beta})}+\sum_{i\in\mathcal{P}_{\widetilde{\beta}\to\widetilde{\gamma}}}\mathsf{B}_i^{(\widetilde{\beta})}-\sum_{i\in\mathcal{P}_{\widetilde{\alpha}\to\widetilde{\beta}}}\mathsf{B}_i^{(\widetilde{\alpha})}-\sum_{i\in\mathcal{P}_{\widetilde{\gamma}\to\widetilde{\beta}}}\mathsf{B}_i^{(\widetilde{\gamma})}\right)\\
&\quad+\log\left(\frac{1-\widetilde{\gamma}}{\widetilde{\gamma}}\right)\left(\sum_{i\in\mathcal{P}_{\widetilde{\gamma}\to\widetilde{\alpha}}}\mathsf{B}_i^{(\widetilde{\gamma})}+\sum_{i\in\mathcal{P}_{\widetilde{\gamma}\to\widetilde{\beta}}}\mathsf{B}_i^{(\widetilde{\gamma})}-\sum_{i\in\mathcal{P}_{\widetilde{\alpha}\to\widetilde{\gamma}}}\mathsf{B}_i^{(\widetilde{\alpha})}-\sum_{i\in\mathcal{P}_{\widetilde{\beta}\to\widetilde{\gamma}}}\mathsf{B}_i^{(\widetilde{\beta})}\right) \quad (119)\\
&=\left(\log\frac{(1-\widetilde{\beta})\widetilde{\alpha}}{(1-\widetilde{\alpha})\widetilde{\beta}}\right)\left(\sum_{i\in\mathcal{P}_{\widetilde{\beta}\to\widetilde{\alpha}}}\mathsf{B}_i^{(\widetilde{\beta})}-\sum_{i\in\mathcal{P}_{\widetilde{\alpha}\to\widetilde{\beta}}}\mathsf{B}_i^{(\widetilde{\alpha})}\right)+\left(\log\frac{(1-\widetilde{\gamma})\widetilde{\alpha}}{(1-\widetilde{\alpha})\widetilde{\gamma}}\right)\left(\sum_{i\in\mathcal{P}_{\widetilde{\gamma}\to\widetilde{\alpha}}}\mathsf{B}_i^{(\widetilde{\gamma})}-\sum_{i\in\mathcal{P}_{\widetilde{\alpha}\to\widetilde{\gamma}}}\mathsf{B}_i^{(\widetilde{\alpha})}\right)\\
&\quad+\left(\log\frac{(1-\widetilde{\gamma})\widetilde{\beta}}{(1-\widetilde{\beta})\widetilde{\gamma}}\right)\left(\sum_{i\in\mathcal{P}_{\widetilde{\gamma}\to\widetilde{\beta}}}\mathsf{B}_i^{(\widetilde{\gamma})}-\sum_{i\in\mathcal{P}_{\widetilde{\beta}\to\widetilde{\gamma}}}\mathsf{B}_i^{(\widetilde{\beta})}\right). \quad (120)
\end{aligned}$$

---

Finally, $\mathsf{Term}_3$ is evaluated as (126), shown at the top of the next page, where (124), shown at the top of the next page follows from

$$\mathcal{P}_\mu(\mathcal{Z}_0)=\bigcup_{\nu\in\{\widetilde{\alpha},\widetilde{\beta},\widetilde{\gamma}\}}\mathcal{P}_{\mu\to\nu},\quad \mathcal{P}_\mu(\mathcal{Z})=\bigcup_{\nu\in\{\widetilde{\alpha},\widetilde{\beta},\widetilde{\gamma}\}}\mathcal{P}_{\nu\to\mu}; \quad (121)$$

and (125), shown at the top of next page holds since $\{\mathcal{P}_{\mu\to\nu}:\mu,\nu\in\{\widetilde{\alpha},\widetilde{\beta},\widetilde{\gamma}\},\ \mu\neq\nu\}$ is a collection of disjoint sets. Plugging (114), (120), and (126), shown at the top of next page into (110), we arrive at (127), shown at the top of next page, which implies (37). This completes the proof of Lemma 3. ∎

## Appendix D
## Proof of Lemma 4

We start by defining three groups of random variables. Recall from Section IV that $\mathsf{B}_i^{(\sigma)}$ with $\sigma\in\{p,\theta,\frac{1}{q-1},\widetilde{\alpha},\widetilde{\beta},\widetilde{\gamma}\}$ denotes a Bernoulli random variable with parameter $\sigma$, that is, $\mathbb{P}[\mathsf{B}_i^{(\sigma)}=1]=1-\mathbb{P}[\mathsf{B}_i^{(\sigma)}=0]=\sigma$. For $p=\Theta\left(\frac{\log n}{n}\right)$, a constant $\theta\in[0,1]$ and $i\in\mathcal{P}_{\mathrm{d}}$, we define the first (group of) random variable $\mathbf{U}_i=\mathbf{U}_i(p,\theta,q)$ as

$$\begin{aligned}
&\mathbf{U}_i(p,\theta,q)\\
&=\log\left((q-1)\frac{1-\theta}{\theta}\right)\mathsf{B}_i^{(p)}\left[\left(1+\mathsf{B}_i^{(\frac{1}{q-1})}\right)\mathsf{B}_i^{(\theta)}-1\right]\\
&=\begin{cases}-\log\left((q-1)\frac{1-\theta}{\theta}\right) & \text{w.p. } p(1-\theta),\\ 0 & \text{w.p. } (1-p)+p\theta\left(1-\frac{1}{q-1}\right),\\ \log\left((q-1)\frac{1-\theta}{\theta}\right) & \text{w.p. } p\theta\frac{1}{q-1}.\end{cases} \quad (122)
\end{aligned}$$

The moment generating function (MGF) $M_{\mathbf{U}_i(p,\theta,q)}(t)$ of $\mathbf{U}_i(p,\theta,q)$ at $t=1/2$ is evaluated as

$$\begin{aligned}
&M_{\mathbf{U}_i(p,\theta,q)}\left(\frac{1}{2}\right)\\
&=\mathbb{E}\left[\exp\left(\frac{1}{2}\mathbf{U}_i(p,\theta,q)\right)\right]\\
&=\left[p(1-\theta)\exp\left(-\frac{1}{2}\log\left((q-1)\frac{1-\theta}{\theta}\right)\right)\right]\\
&\quad+\left[1-p+p\theta\left(1-\frac{1}{q-1}\right)\right]\\
&\quad+\left[\frac{p\theta}{q-1}\exp\left(\frac{1}{2}\log\left((q-1)\frac{1-\theta}{\theta}\right)\right)\right]\\
&=p\sqrt{\frac{\theta(1-\theta)}{q-1}}+1-p+p\theta\left(1-\frac{1}{q-1}\right)+p\sqrt{\frac{\theta(1-\theta)}{q-1}}\\
&=1-p\left(1-\theta-2\sqrt{\frac{\theta(1-\theta)}{q-1}}+\frac{\theta}{q-1}\right)\\
&=1-p\left(\sqrt{1-\theta}-\sqrt{\theta/(q-1)}\right)^2, \quad (123)
\end{aligned}$$

$$\begin{aligned}\mathsf{Term}_3 &= \sum_{\mu\in\{\widetilde{\alpha},\widetilde{\beta},\widetilde{\gamma}\}} \left[\log(1-\mu)\left(|\mathcal{P}_\mu(\mathcal{Z})|-|\mathcal{P}_\mu(\mathcal{Z}_0)|\right)\right]\\
&= \sum_{\mu\in\{\widetilde{\alpha},\widetilde{\beta},\widetilde{\gamma}\}} \left(\log(1-\mu)\left|\bigcup_{\nu\in\{\widetilde{\alpha},\widetilde{\beta},\widetilde{\gamma}\}}\mathcal{P}_{\nu\to\mu}\right|\right) - \sum_{\mu\in\{\widetilde{\alpha},\widetilde{\beta},\widetilde{\gamma}\}} \left(\log(1-\mu)\left|\bigcup_{\nu\in\{\widetilde{\alpha},\widetilde{\beta},\widetilde{\gamma}\}}\mathcal{P}_{\mu\to\nu}\right|\right) \qquad (124)\\
&= \sum_{\mu\in\{\widetilde{\alpha},\widetilde{\beta},\widetilde{\gamma}\}} \left[\log(1-\mu)\left(|\mathcal{P}_{\widetilde{\alpha}\to\mu}|+\left|\mathcal{P}_{\widetilde{\beta}\to\mu}\right|+|\mathcal{P}_{\widetilde{\gamma}\to\mu}|\right)\right] - \sum_{\mu\in\{\widetilde{\alpha},\widetilde{\beta},\widetilde{\gamma}\}} \left[\log(1-\mu)\left(|\mathcal{P}_{\mu\to\widetilde{\alpha}}|+\left|\mathcal{P}_{\mu\to\widetilde{\beta}}\right|+|\mathcal{P}_{\mu\to\widetilde{\gamma}}|\right)\right] \qquad (125)\\
&= \log(1-\widetilde{\alpha})\left(\left|\mathcal{P}_{\widetilde{\beta}\to\widetilde{\alpha}}\right|+|\mathcal{P}_{\widetilde{\gamma}\to\widetilde{\alpha}}|\right) - \log(1-\widetilde{\alpha})\left(\left|\mathcal{P}_{\widetilde{\alpha}\to\widetilde{\beta}}\right|+|\mathcal{P}_{\widetilde{\alpha}\to\widetilde{\gamma}}|\right) + \log(1-\widetilde{\beta})\left(\left|\mathcal{P}_{\widetilde{\alpha}\to\widetilde{\beta}}\right|+\left|\mathcal{P}_{\widetilde{\gamma}\to\widetilde{\beta}}\right|\right)\\
&\quad - \log(1-\widetilde{\beta})\left(\left|\mathcal{P}_{\widetilde{\beta}\to\widetilde{\alpha}}\right|+\left|\mathcal{P}_{\widetilde{\beta}\to\widetilde{\gamma}}\right|\right) + \log(1-\widetilde{\gamma})\left(|\mathcal{P}_{\widetilde{\alpha}\to\widetilde{\gamma}}|+\left|\mathcal{P}_{\widetilde{\beta}\to\widetilde{\gamma}}\right|\right) - \log(1-\widetilde{\gamma})\left(|\mathcal{P}_{\widetilde{\gamma}\to\widetilde{\alpha}}|+\left|\mathcal{P}_{\widetilde{\gamma}\to\widetilde{\beta}}\right|\right)\\
&= \left(\log\frac{1-\widetilde{\alpha}}{1-\widetilde{\beta}}\right)\left(\left|\mathcal{P}_{\widetilde{\beta}\to\widetilde{\alpha}}\right|-\left|\mathcal{P}_{\widetilde{\alpha}\to\widetilde{\beta}}\right|\right) + \left(\log\frac{1-\widetilde{\alpha}}{1-\widetilde{\gamma}}\right)\left(|\mathcal{P}_{\widetilde{\gamma}\to\widetilde{\alpha}}|-|\mathcal{P}_{\widetilde{\alpha}\to\widetilde{\gamma}}|\right) + \left(\log\frac{1-\widetilde{\beta}}{1-\widetilde{\gamma}}\right)\left(\left|\mathcal{P}_{\widetilde{\gamma}\to\widetilde{\beta}}\right|-\left|\mathcal{P}_{\widetilde{\beta}\to\widetilde{\gamma}}\right|\right). \qquad (126)\end{aligned}$$

$$\begin{aligned}\mathsf{L}(X_0)-\mathsf{L}(X) &= \log\left((q-1)\frac{1-\theta}{\theta}\right)\sum_{i\in\mathcal{P}_{\rm d}}\mathsf{B}_i^{(p)}\left[\left(1+\mathsf{B}_i^{\left(\frac{1}{q-1}\right)}\right)\mathsf{B}_i^{(\theta)}-1\right]\\
&\quad+\left(\log\frac{(1-\widetilde{\beta})\widetilde{\alpha}}{(1-\widetilde{\alpha})\widetilde{\beta}}\right)\left(\sum_{i\in\mathcal{P}_{\widetilde{\beta}\to\widetilde{\alpha}}}\mathsf{B}_i^{(\widetilde{\beta})}-\sum_{i\in\mathcal{P}_{\widetilde{\alpha}\to\widetilde{\beta}}}\mathsf{B}_i^{(\widetilde{\alpha})}\right)+\left(\log\frac{1-\widetilde{\alpha}}{1-\widetilde{\beta}}\right)\left(\left|\mathcal{P}_{\widetilde{\beta}\to\widetilde{\alpha}}\right|-\left|\mathcal{P}_{\widetilde{\alpha}\to\widetilde{\beta}}\right|\right)\\
&\quad+\left(\log\frac{(1-\widetilde{\gamma})\widetilde{\alpha}}{(1-\widetilde{\alpha})\widetilde{\gamma}}\right)\left(\sum_{i\in\mathcal{P}_{\widetilde{\gamma}\to\widetilde{\alpha}}}\mathsf{B}_i^{(\widetilde{\gamma})}-\sum_{i\in\mathcal{P}_{\widetilde{\alpha}\to\widetilde{\gamma}}}\mathsf{B}_i^{(\widetilde{\alpha})}\right)+\left(\log\frac{1-\widetilde{\alpha}}{1-\widetilde{\gamma}}\right)\left(|\mathcal{P}_{\widetilde{\gamma}\to\widetilde{\alpha}}|-|\mathcal{P}_{\widetilde{\alpha}\to\widetilde{\gamma}}|\right)\\
&\quad+\left(\log\frac{(1-\widetilde{\gamma})\widetilde{\beta}}{(1-\widetilde{\beta})\widetilde{\gamma}}\right)\left(\sum_{i\in\mathcal{P}_{\widetilde{\gamma}\to\widetilde{\beta}}}\mathsf{B}_i^{(\widetilde{\gamma})}-\sum_{i\in\mathcal{P}_{\widetilde{\beta}\to\widetilde{\gamma}}}\mathsf{B}_i^{(\widetilde{\beta})}\right)+\left(\log\frac{1-\widetilde{\beta}}{1-\widetilde{\gamma}}\right)\left(\left|\mathcal{P}_{\widetilde{\gamma}\to\widetilde{\beta}}\right|-\left|\mathcal{P}_{\widetilde{\beta}\to\widetilde{\gamma}}\right|\right). \qquad (127)\end{aligned}$$

and hence, we have

$$\begin{aligned}-\log M_{\mathbf{U}_i(p,\theta,q)}\left(\frac{1}{2}\right) &= -\log\left[1-p\left(\sqrt{1-\theta}-\sqrt{\frac{\theta}{q-1}}\right)^2\right]\\
&= p\left(\sqrt{1-\theta}-\sqrt{\frac{\theta}{q-1}}\right)^2+O\left(p^2\right) \qquad (128)\\
&= (1+o(1))\left(\sqrt{1-\theta}-\sqrt{\frac{\theta}{q-1}}\right)^2 p\\
&= (1+o(1))I_r, \qquad (129)\end{aligned}$$

where (128) follows from the Taylor expansion of the function $\log(1-x)$ at $x=p\left(\sqrt{1-\theta}-\sqrt{\frac{\theta}{q-1}}\right)^2$, which converges for $p=\Theta\left(\frac{\log n}{n}\right)$. Next, for $\mu,\nu=\Theta\left(\frac{\log n}{n}\right)$ and $i\in\mathcal{P}_{\mu\to\nu}$, define the second (set of) random variables $\mathbf{V}_i=\mathbf{V}_i(\mu,\nu)$ as

$$\begin{aligned}\mathbf{V}_i(\mu,\nu) &= \left(\log\frac{(1-\mu)\nu}{(1-\nu)\mu}\right)\left(\mathsf{B}_i^{(\mu)}-\mathsf{B}_i^{(\nu)}\right)\\
&= \begin{cases}-\log\frac{(1-\mu)\nu}{(1-\nu)\mu} & \text{w.p. } (1-\mu)\nu,\\ 0 & \text{w.p. } (1-\mu)(1-\nu)+\mu\nu,\\ \log\frac{(1-\mu)\nu}{(1-\nu)\mu} & \text{w.p. } \mu(1-\nu).\end{cases} \qquad (130)\end{aligned}$$

The MGF of $\mathbf{V}_i(\mu,\nu)$ at $t=1/2$ is evaluated as

$$\begin{aligned}&M_{\mathbf{V}_i(\mu,\nu)}\left(\frac{1}{2}\right)=\mathbb{E}\left[\exp\left(\frac{1}{2}\mathbf{V}_i(\mu,\nu)\right)\right]\\
&=\left[(1-\mu)\nu\,\exp\left(-\frac{1}{2}\log\frac{(1-\mu)\nu}{(1-\nu)\mu}\right)\right]\\
&\quad+[(1-\mu)(1-\nu)+\mu\nu]+\left[\mu(1-\nu)\,\exp\left(\frac{1}{2}\log\frac{(1-\mu)\nu}{(1-\nu)\mu}\right)\right]\\
&=(1-\mu)\nu\sqrt{\frac{(1-\nu)\mu}{(1-\mu)\nu}}+(1-\mu)(1-\nu)+\mu\nu\\
&\quad+(1-\nu)\mu\sqrt{\frac{(1-\mu)\nu}{(1-\nu)\mu}}\\
&=\mu\nu+2\sqrt{(1-\mu)(1-\nu)\mu\nu}+(1-\mu)(1-\nu)\\
&=\left(\sqrt{\mu\nu}+\sqrt{(1-\mu)(1-\nu)}\right)^2. \qquad (131)\end{aligned}$$

Therefore, we have

$$\begin{aligned}&-\log M_{\mathbf{V}_i(\mu,\nu)}\left(\frac{1}{2}\right)\\
&=-2\log\left(\sqrt{\mu\nu}+\sqrt{(1-\mu)}\sqrt{(1-\nu)}\right)\\
&=-2\log\left[\sqrt{\mu\nu}+\left(1-\frac{1}{2}\mu+O\left(\mu^2\right)\right)\left(1-\frac{1}{2}\nu+O\left(\nu^2\right)\right)\right] \qquad (132)\end{aligned}$$

$$= -2\log\left[\sqrt{\mu\nu} + \left(1 - \frac{1}{2}\mu - \frac{1}{2}\nu + O\left(\mu^2+\nu^2\right)\right)\right]$$
$$= -2\log\left[1 - \left(\frac{1}{2}\mu + \frac{1}{2}\nu - \sqrt{\mu\nu} + O\left(\mu^2+\nu^2\right)\right)\right]$$
$$= \left(\sqrt{\mu}-\sqrt{\nu}\right)^2 + O\left(\mu^2+\nu^2\right) \quad (133)$$
$$= (1+o(1))\left(\sqrt{\mu}-\sqrt{\nu}\right)^2$$
$$= \begin{cases} (1+o(1))I_{\alpha,\beta}\frac{\log n}{n} & \text{if } \mu=\widetilde{\beta},\ \nu=\widetilde{\alpha}, \\ (1+o(1))I_{\alpha,\gamma}\frac{\log n}{n} & \text{if } \mu=\widetilde{\gamma},\ \nu=\widetilde{\alpha}, \\ (1+o(1))I_{\beta,\gamma}\frac{\log n}{n} & \text{if } \mu=\widetilde{\gamma},\ \nu=\widetilde{\beta}, \end{cases} \quad (134)$$

where (132) follows form the Taylor expansion of the functions $\sqrt{1-\mu}$ and $\sqrt{1-\nu}$, which both converge since $\mu,\nu = \Theta\left(\frac{\log n}{n}\right)$; and (133) follows from the Taylor expansion of the function $\log(1-x)$, for $x = \frac{1}{2}\mu + \frac{1}{2}\nu - \sqrt{\mu\nu} + O\left(\mu^2+\nu^2\right)$, which also converges for $\mu,\nu = \Theta\left(\frac{\log n}{n}\right)$.

Finally, for $\mu,\nu = \Theta\left(\frac{\log n}{n}\right)$ and $i \in \mathcal{P}_{\mu\to\nu}$, define the third (group of) random variable $\mathbf{W}_i = \mathbf{W}_i(\mu,\nu)$ as

$$\mathbf{W}_i(\mu,\nu) = \left(\log\frac{1-\nu}{1-\mu}\right) + \left(\log\frac{(1-\mu)\nu}{(1-\nu)\mu}\right)\mathsf{B}_i^{(\mu)}$$
$$= \begin{cases} \log\frac{\nu}{\mu} & \text{w.p. } \mu, \\ \log\frac{1-\nu}{1-\mu} & \text{w.p. } (1-\mu). \end{cases} \quad (135)$$

The moment generating function $M_{\mathbf{W}_i(\mu,\nu)}(t)$ of $\mathbf{W}_i(\mu,\nu)$ at $t=1/2$ is evaluated as

$$M_{\mathbf{W}_i(\mu,\nu)}\left(\frac{1}{2}\right)$$
$$= \mathbb{E}\left[\exp\left(\frac{1}{2}\mathbf{W}_i(\mu,\nu)\right)\right]$$
$$= \left[\mu\exp\left(\frac{1}{2}\log\frac{\nu}{\mu}\right)\right] + \left[(1-\mu)\exp\left(\frac{1}{2}\log\frac{1-\nu}{1-\mu}\right)\right]$$
$$= \sqrt{\mu\nu} + \sqrt{(1-\mu)(1-\nu)}. \quad (136)$$

Thus, we can write

$$-\log M_{\mathbf{W}_i(\mu,\nu)}\left(\frac{1}{2}\right)$$
$$= -\log\left(\sqrt{\mu\nu} + \sqrt{(1-\mu)}\sqrt{(1-\nu)}\right)$$
$$= \frac{1}{2}(1+o(1))\left(\sqrt{\mu}-\sqrt{\nu}\right)^2$$
$$= \begin{cases} \frac{1}{2}(1+o(1))I_{\alpha,\beta}\frac{\log n}{n} & \text{if } \mu=\widetilde{\beta},\ \nu=\widetilde{\alpha}, \\ \frac{1}{2}(1+o(1))I_{\alpha,\gamma}\frac{\log n}{n} & \text{if } \mu=\widetilde{\gamma},\ \nu=\widetilde{\alpha}, \\ \frac{1}{2}(1+o(1))I_{\beta,\gamma}\frac{\log n}{n} & \text{if } \mu=\widetilde{\gamma},\ \nu=\widetilde{\beta}, \end{cases} \quad (137)$$

where (137) follows from (134). Next, we present the following proposition that is used in the proof of Lemma 4. The proof of the proposition is presented at the end of this appendix.

*Proposition 1:* For $\mu,\nu = \Theta\left(\frac{\log n}{n}\right)$, let $\mathbf{A} = \mathbf{A}(\mu,\nu)$ be a random variable that is defined as

$$\mathbf{A}(\mu,\nu) = \left(\log\frac{(1-\mu)\nu}{(1-\nu)\mu}\right)\left(\sum_{i\in\mathcal{P}_{\mu\to\nu}}\mathsf{B}_i^{(\mu)} - \sum_{i\in\mathcal{P}_{\nu\to\mu}}\mathsf{B}_i^{(\nu)}\right)$$
$$+ \left(\log\frac{1-\nu}{1-\mu}\right)\left(|\mathcal{P}_{\mu\to\nu}| - |\mathcal{P}_{\nu\to\mu}|\right), \quad (138)$$

where $\{\mathsf{B}_i^{(\mu)} : i \in \mathcal{P}_{\nu\to\mu}\}$ and $\{\mathsf{B}_i^{(\nu)} : i \in \mathcal{P}_{\nu\to\mu}\}$ are sets of independent and identically distributed Bernoulli random variables. The moment generating function $M_{\mathbf{A}(\mu,\nu)}(t)$ of $\mathbf{A}(\mu,\nu)$ at $t=1/2$ is given by

$$M_{\mathbf{A}(\mu,\nu)}(t)$$
$$= \exp\left(-(1+o(1))\frac{|\mathcal{P}_{\mu\to\nu}| + |\mathcal{P}_{\nu\to\mu}|}{2}\left(\sqrt{\mu}-\sqrt{\nu}\right)^2\right)$$
$$= \begin{cases} \exp\left(-(1+o(1))P_{\widetilde{\alpha}\leftrightarrow\widetilde{\beta}}\, I_{\alpha,\beta}\frac{\log n}{n}\right) & \text{if } \mu=\widetilde{\beta},\ \nu=\widetilde{\alpha}, \\ \exp\left(-(1+o(1))P_{\widetilde{\alpha}\leftrightarrow\widetilde{\gamma}}\, I_{\alpha,\gamma}\frac{\log n}{n}\right) & \text{if } \mu=\widetilde{\gamma},\ \nu=\widetilde{\alpha}, \\ \exp\left(-(1+o(1))P_{\widetilde{\beta}\leftrightarrow\widetilde{\gamma}}\, I_{\beta,\gamma}\frac{\log n}{n}\right) & \text{if } \mu=\widetilde{\gamma},\ \nu=\widetilde{\beta}. \end{cases} \quad (139)$$

Let $\{\mathbf{U}_i(p,\theta,q) : i \in \mathcal{P}_{\rm d}\}$, and $\{\mathbf{A}(\widetilde{\beta},\widetilde{\alpha}), \mathbf{A}(\widetilde{\gamma},\widetilde{\alpha}), \mathbf{A}(\widetilde{\gamma},\widetilde{\beta})\}$ be sets of independent and identically distributed random variables defined as per (122), and (138) in Proposition 1. Note that the sets $\{\mathcal{P}_{\mu\to\nu} : \mu,\nu \in \{\widetilde{\alpha},\widetilde{\beta},\widetilde{\gamma}\},\ \mu\neq\nu\}$ are disjoint as per their definitions given by (29)–(34). Consequently, the LHS of (38) is upper bounded by

$$\mathbb{P}\left[\mathbf{B}\geq 0\right]$$
$$= \mathbb{P}\left[\left(\sum_{i\in\mathcal{P}_{\rm d}}\mathbf{U}_i(p,\theta,q)\right) + \mathbf{A}(\widetilde{\beta},\widetilde{\alpha}) + \mathbf{A}(\widetilde{\gamma},\widetilde{\alpha}) + \mathbf{A}(\widetilde{\gamma},\widetilde{\beta}) \geq 0\right]$$
$$\leq \left(M_{\mathbf{U}_i(p,\theta,q)}\left(\frac{1}{2}\right)\right)^{|\mathcal{P}_{\rm d}|}\left(M_{\mathbf{A}(\widetilde{\beta},\widetilde{\alpha})}\left(\frac{1}{2}\right)\right)$$
$$\times\left(M_{\mathbf{A}(\widetilde{\gamma},\widetilde{\alpha})}\left(\frac{1}{2}\right)\right)\left(M_{\mathbf{A}(\widetilde{\gamma},\widetilde{\beta})}\left(\frac{1}{2}\right)\right) \quad (140)$$
$$= \exp\Bigg(-(1+o(1))\Bigg(|\mathcal{P}_{\rm d}|I_r + P_{\widetilde{\alpha}\leftrightarrow\widetilde{\beta}}\, I_{\alpha,\beta}\frac{\log n}{n}$$
$$+ P_{\widetilde{\alpha}\leftrightarrow\widetilde{\gamma}}\, I_{\alpha,\gamma}\frac{\log n}{n} + P_{\widetilde{\beta}\leftrightarrow\widetilde{\gamma}}\, I_{\beta,\gamma}\frac{\log n}{n}\Bigg)\Bigg), \quad (141)$$

where in (140) we used the Chernoff bound at $t=\frac{1}{2}$ for mutually independent random variables $\{\mathbf{U}_i(p,\theta,q) : i\in\mathcal{P}_{\rm d}\}$, and $\{\mathbf{A}(\widetilde{\beta},\widetilde{\alpha}), \mathbf{A}(\widetilde{\gamma},\widetilde{\alpha}), \mathbf{A}(\widetilde{\gamma},\widetilde{\beta})\}$; and finally (141) follows from (129), and (139) in Proposition 1. This completes the proof of Lemma 4. ∎

It remains to prove Proposition 1. In the following, we present the proof of the proposition. *Proof:* [Proof of Proposition 1] We distinguish two cases based on the sizes of the sets $\mathcal{P}_{\mu\to\nu}$ and $\mathcal{P}_{\nu\to\mu}$. First, assume $|\mathcal{P}_{\mu\to\nu}| \geq |\mathcal{P}_{\nu\to\mu}|$. In this case, the random variable $\mathbf{A}(\mu,\nu)$ can be expressed as

$$\mathbf{A}(\mu,\nu) = \sum_{i\in\mathcal{P}_{\nu\to\mu}}\left(\left(\log\frac{(1-\mu)\nu}{(1-\nu)\mu}\right)\left(\mathsf{B}_i^{(\mu)} - \mathsf{B}_i^{(\nu)}\right)\right)$$
$$+ \sum_{i\in\mathcal{P}_{\mu\to\nu}\setminus\mathcal{P}_{\nu\to\mu}}\left(\left(\log\frac{1-\nu}{1-\mu}\right) + \left(\log\frac{(1-\mu)\nu}{(1-\nu)\mu}\right)\mathsf{B}_i^{(\mu)}\right) \quad (142)$$
$$= \sum_{i\in\mathcal{P}_{\nu\to\mu}}\mathbf{V}_i + \sum_{i\in\mathcal{P}_{\mu\to\nu}\setminus\mathcal{P}_{\nu\to\mu}}\mathbf{W}_i, \quad (143)$$

where (142) holds since the sets $\mathcal{P}_{\nu\to\mu}$ and $\mathcal{P}_{\mu\to\nu}$ are disjoint; and (143) follows from (130) and (135).

Then, we have

$$M_{\mathbf{A}(\mu,\nu)}\left(\frac{1}{2}\right) = \mathbb{E}\left[\exp\left(\frac{1}{2}\mathbf{A}(\mu,\nu)\right)\right]$$
$$= \mathbb{E}\left[\left(\prod_{i\in\mathcal{P}_{\nu\to\mu}} \exp\left(\frac{1}{2}\mathbf{V}_i(\mu,\nu)\right)\right) \times \left(\prod_{i\in\mathcal{P}_{\mu\to\nu}\setminus\mathcal{P}_{\nu\to\mu}} \exp\left(\frac{1}{2}\mathbf{W}_i(\mu,\nu)\right)\right)\right]$$
$$= \left(\prod_{i\in\mathcal{P}_{\nu\to\mu}} \mathbb{E}\left[\exp\left(\frac{1}{2}\mathbf{V}_i(\mu,\nu)\right)\right]\right) \times \left(\prod_{i\in\mathcal{P}_{\mu\to\nu}\setminus\mathcal{P}_{\nu\to\mu}} \mathbb{E}\left[\exp\left(\frac{1}{2}\mathbf{W}_i(\mu,\nu)\right)\right]\right) \tag{144}$$
$$= \left(\prod_{i\in\mathcal{P}_{\nu\to\mu}} M_{\mathbf{V}_i(\mu,\nu)}\left(\frac{1}{2}\right)\right)\left(\prod_{i\in\mathcal{P}_{\mu\to\nu}\setminus\mathcal{P}_{\nu\to\mu}} M_{\mathbf{W}_i(\mu,\nu)}\left(\frac{1}{2}\right)\right)$$
$$= \left[\exp\left(-(1+o(1))\left(\sqrt{\mu}-\sqrt{\nu}\right)^2\right)\right]^{|\mathcal{P}_{\nu\to\mu}|} \times \left[\exp\left(-\frac{1}{2}(1+o(1))\left(\sqrt{\mu}-\sqrt{\nu}\right)^2\right)\right]^{|\mathcal{P}_{\mu\to\nu}|-|\mathcal{P}_{\nu\to\mu}|} \tag{145}$$
$$= \exp\left(-(1+o(1))\frac{|\mathcal{P}_{\mu\to\nu}|+|\mathcal{P}_{\nu\to\mu}|}{2}\left(\sqrt{\mu}-\sqrt{\nu}\right)^2\right), \tag{146}$$

in which, (144) follows from the fact that the random variables $\{\mathbf{V}_i : i\in\mathcal{P}_{\nu\to\mu}\}$ and $\{\mathbf{W}_i : i\in\mathcal{P}_{\mu\to\nu}\setminus\mathcal{P}_{\nu\to\mu}\}$ are independent; and (145) is a consequence of (134) and (137). This shows the claim of the proposition for the first case.

Next, consider the second case, where $|\mathcal{P}_{\mu\to\nu}| \le |\mathcal{P}_{\nu\to\mu}|$. In a similar way, the random variable $\mathbf{A}$ can be written as

$$\mathbf{A}(\mu,\nu) = \sum_{i\in\mathcal{P}_{\mu\to\nu}}\left[\left(\log\frac{(1-\mu)\nu}{(1-\nu)\mu}\right)\left(\mathsf{B}_i^{(\mu)}-\mathsf{B}_i^{(\nu)}\right)\right] + \sum_{i\in\mathcal{P}_{\nu\to\mu}\setminus\mathcal{P}_{\mu\to\nu}}\left[\left(\log\frac{1-\mu}{1-\nu}\right)+\left(\log\frac{(1-\nu)\mu}{(1-\mu)\nu}\right)\mathsf{B}_i^{(\nu)}\right]$$
$$= \sum_{i\in\mathcal{P}_{\mu\to\nu}}\mathbf{V}_i + \sum_{i\in\mathcal{P}_{\nu\to\mu}\setminus\mathcal{P}_{\mu\to\nu}}\mathbf{W}_i. \tag{147}$$

Following the same procedure presented above, one can show that $M_{\mathbf{A}(\mu,\nu)}\left(\frac{1}{2}\right)$ for the second case can also be simplified to the expression given in (146). This completes the proof of Proposition 1. ∎

## Appendix E
## Proof of Lemma 5

The LHS of (47) is given by (152), shown at the bottom of the page. In what follows, we derive upper bounds on $\mathsf{Term}_1$ and $\mathsf{Term}_2$ of (152) for a fixed non-all-zero tuple $T\in\mathcal{T}^{(\delta)}_{\text{small}}$, given by

$$T = \left(\left\{n_{i,j}^{(x,y)}\right\}_{x,y\in[c],\, i,j\in[g]}, \left\{d_{i,j}^{(x,y)}\right\}_{x,y\in[c],\, i,j\in[g]}\right), \tag{148}$$

according to (43).

(1) Upper Bound on $\mathsf{Term}_1$: The size of the set $\mathcal{X}(T)$ is given by (153), shown at the top of the next page, which follows from the fact that the number of ways to count the rating matrices subject to $\{n_{i,j}^{(x,y)} : x,y\in[c], i,j\in[g]\}$, and subject to $\{d_{i,j}^{(x,y)} : x,y\in[c], i,j\in[g]\}$ are independent. We denote the first and second term of (153) by $\mathsf{Term}_{1,1}$ and $\mathsf{Term}_{1,2}$, respectively. Next, we provide upper bounds on $\mathsf{Term}_{1,1}$ and $\mathsf{Term}_{1,2}$.

(1-1) Upper Bound on $\mathsf{Term}_{1,1}$: We can write

$$\mathsf{Term}_{1,1} = \left|\mathcal{X}\left(\left\{n_{i,j}^{(x,y)} : x,y\in[c], i,j\in[g]\right\}, \left\{\widehat{d}_{i,j}^{(x,y)} = 0 : x,y\in[c], i,j\in[g]\right\}\right)\right|$$
$$= \prod_{x\in[c]}\prod_{i\in[g]}\binom{n/(gc)}{n_{i,1}^{(x,1)},\cdots,n_{i,g}^{(x,1)},n_{i,1}^{(x,2)},\cdots,n_{i,g}^{(x,c)}} \tag{149}$$
$$\le \prod_{x\in[c]}\prod_{i\in[g]}\left(\frac{n}{gc}\right)^{\sum_{(y,j)\neq(\sigma(x),\sigma(i|x))} n_{i,j}^{(x,y)}} \tag{150}$$
$$\le \prod_{x\in[c]}\prod_{i\in[g]}\exp\left[\left(\sum_{(y,j)\neq(\sigma(x),\sigma(i|x))} n_{i,j}^{(x,y)}\right)\log n\right]$$
$$= \exp\left[\log n\left(\sum_{x\in[c]}\sum_{i\in[g]}\sum_{(y,j)\neq(\sigma(x),\sigma(i|x))} n_{i,j}^{(x,y)}\right)\right]. \tag{151}$$

Note that the equality in (149) follows from the definitions in (40) and (43); (150) follows from the definition of a multinomial coefficient, and the fact that $\binom{n}{k}\le n^k$.

(1-2) Upper Bound on $\mathsf{Term}_{1,2}$: An upper bound on $\mathsf{Term}_{1,2}$ is given by (154), shown at the top of the next page.

Recall from Section II that $R_0^{(x)}\in\mathbb{F}_q^{g\times m}$ denotes a matrix that is obtained by stacking all the rating vectors of cluster $x$ given by $\{u_i^{(x)} : i\in[g]\}$ for $x\in[c]$, and whose columns are elements of $(g,r)$ MDS code. Similarly, define $R^{(x)}\in\mathbb{F}_q^{g\times m}$

$$\lim_{n,m\to\infty}\sum_{T\in\mathcal{T}^{(\delta)}_{\text{small}}}\sum_{X\in\mathcal{X}(T)}\exp\left[-(1+o(1))\left(|\mathcal{P}_{\mathsf{d}}|I_r + P_{\widetilde{\alpha}\leftrightarrow\widetilde{\beta}}\, I_{\alpha,\beta}\frac{\log n}{n} + P_{\widetilde{\alpha}\leftrightarrow\widetilde{\gamma}}\, I_{\alpha,\gamma}\frac{\log n}{n} + P_{\widetilde{\beta}\leftrightarrow\widetilde{\gamma}}\, I_{\beta,\gamma}\frac{\log n}{n}\right)\right]$$
$$= \lim_{n,m\to\infty}\sum_{T\in\mathcal{T}^{(\delta)}_{\text{small}}}\underbrace{|\mathcal{X}(T)|}_{\mathsf{Term}_1}\underbrace{\exp\left[-(1+o(1))\left(|\mathcal{P}_{\mathsf{d}}|I_r + P_{\widetilde{\alpha}\leftrightarrow\widetilde{\beta}}\, I_{\alpha,\beta}\frac{\log n}{n} + P_{\widetilde{\alpha}\leftrightarrow\widetilde{\gamma}}\, I_{\alpha,\gamma}\frac{\log n}{n} + P_{\widetilde{\beta}\leftrightarrow\widetilde{\gamma}}\, I_{\beta,\gamma}\frac{\log n}{n}\right)\right]}_{\mathsf{Term}_2}. \tag{152}$$

$$\mathsf{Term}_1 = \underbrace{\left|\mathcal{X}\left(\left\{n_{i,j}^{(x,y)} : x,y\in[c], i,j\in[g]\right\}, \left\{\widehat{d}_{i,j}^{(x,y)} = 0 : x,y\in[c], i,j\in[g]\right\}\right)\right|}_{\mathsf{Term}_{1,1}} \times \underbrace{\left|\mathcal{X}\left(\left\{\widehat{n}_{i,j}^{(x,y)} = 0 : x,y\in[c], i,j\in[g]\right\}, \left\{d_{i,j}^{(x,y)} : x,y\in[c], i,j\in[g]\right\}\right)\right|}_{\mathsf{Term}_{1,2}}, \quad (153)$$

$$\begin{aligned}\mathsf{Term}_{1,2} &= \left|\mathcal{X}\left(\left\{\widehat{n}_{i,j}^{(x,y)} = 0 : x,y\in[c], i,j\in[g]\right\}, \left\{d_{i,j}^{(x,y)} : x,y\in[c], i,j\in[g]\right\}\right)\right| \\ &\le \left|\mathcal{X}\left(\left\{\widehat{n}_{i,j}^{(x,y)} = 0 : x,y\in[c], i,j\in[g]\right\}, \left\{d_{i,\sigma(i|x)}^{(x,\sigma(x))} : x\in[c], i\in[g]\right\},\right.\right. \\ &\qquad \left.\left.\left\{\widehat{d}_{i,j}^{(x,y)} = t : 0\le t\le m, x,y\in[c], i,j\in[g], (y,j)\ne(\sigma(x),\sigma(i|x))\right\}\right)\right| \\ &\le \prod_{z\in[c]}\left|\mathcal{X}\left(\left\{\widehat{n}_{i,j}^{(x,y)} = 0 : x,y\in[c], i,j\in[g]\right\}, \left\{d_{i,\sigma(i|z)}^{(z,\sigma(z))} : i\in[g]\right\},\right.\right. \\ &\qquad \left\{\widehat{d}_{i,\sigma(i|x)}^{(x,\sigma(x))} = t : 0\le t\le m,\ x\in[c]\setminus\{z\},\ i\in[g]\right\}, \\ &\qquad \left.\left.\left\{\widehat{d}_{i,j}^{(x,y)} = t : 0\le t\le m, x,y\in[c], i,j\in[g],\ (y,j)\ne(\sigma(x),\sigma(i|x))\right\}\right)\right|. \quad (154)\end{aligned}$$

as a matrix that is obtained by stacking all the rating vectors of cluster $x$ given by $\{u_i^{(x)} : i\in[g]\}$ for $x\in[c]$, and whose columns are also elements of $(g,r)$ MDS code. Furthermore, define the binary matrix $\widehat{R}^{(x)}\in\mathbb{F}_q^{g\times m}$ as follows:

$$\widehat{R}^{(x)}(i,t) = \mathbb{1}\left[R_0^{(x)}(i,t)\ne R^{(x)}(i,t)\right],$$

for $x\in[c]$, $i\in[g]$, and $t\in[m]$. Note that $\widehat{R}^{(x)}(i,t)=1$ when there is an error in estimating the rating of the users in cluster $x$ and group $i$ for item $t$ for $x\in[c]$, $i\in[g]$ and $t\in[m]$. Let any non-zero column of $\widehat{R}^{(x)}$ be denoted as an "error column".

Then, for a given cluster $x\in[c]$, we enumerate all possible matrices $\widehat{R}^{(x)}$ subject to a given number of error columns. To this end, define $f^{(x)}$ as the total number of error columns of $\widehat{R}^{(x)}$. Moreover, define $\kappa$ as the number of possible configurations of an error column. Let $\{w_k : k\in[\kappa]\}$ be the set of all possible error columns. Note that this set only depends on the problem setting and the MDS code structure. For instance, for $(c,g,q,r)=(2,3,2,2)$ and $u_3^{(x)} = u_1^{(x)} + u_2^{(x)}$ for $x\in[2]$, we have $\kappa=3$ since the possible configurations of an error column are given by

$$w_1 = \begin{bmatrix}1\\0\\1\end{bmatrix},\ w_2 = \begin{bmatrix}1\\1\\0\end{bmatrix},\ w_\kappa = w_3 = \begin{bmatrix}0\\1\\1\end{bmatrix}.$$

For $x\in[c]$ and $k\in[\kappa]$, let $f_k^{(x)}$ denote the number of columns in the matrix $\widehat{R}^{(x)}$ that are equal to $w_k$. Note that $0\le f_k^{(x)}\le f^{(x)}$ and $\sum_{k=1}^{\kappa} f_k^{(x)} = f^{(x)}$. For cluster $x\in[c]$, let $\mathcal{R}^{(x)}(f^{(x)},\{w_k : k\in[\kappa]\})$ denote the set of matrices $R^{(x)}\in\mathbb{F}_q^{g\times m}$ characterized by $f^{(x)}$ and $\{w_k : k\in[\kappa]\}$. The size of $\mathcal{R}^{(x)}(f^{(x)},\{w_k : k\in[\kappa]\})$ can be bounded by

$$\begin{aligned}&\left|\mathcal{R}^{(x)}\left(f^{(x)},\{w_k : k\in[\kappa]\}\right)\right| \\ &\le \binom{m}{f^{(x)}}\binom{f^{(x)}+\kappa-1}{\kappa-1}(q-1)^{gf^{(x)}} \quad (155)\\ &\le m^{f^{(x)}}\, 2^{f^{(x)}+\kappa-1}\, q^{gf^{(x)}} \quad (156)\\ &\le 2^{q^g-1} q^g (2qm)^{f^{(x)}}, \quad (157)\end{aligned}$$

where

- (155) follows by first choosing $f^{(x)}$ columns from $m$ columns to be error columns, then counting the number of integer solutions of $\sum_{k=1}^{\kappa} f_k^{(x)} = f^{(x)}$, and lastly counting the number of estimation error combination within the $g$ entries of each of the $f^{(x)}$ error columns;
- (156) follows from bounding the first binomial coefficient by $\binom{a}{b}\le a^b$, and the second binomial coefficient by $\binom{a}{b}\le\sum_{i=1}^{a}\binom{a}{i} = 2^a$, for $a\ge b$;
- and finally (157) follows from $\kappa\le q^g$, that is due to the fact that each entry of a rating matrix column can take one of $q$ values.

Next, for a given cluster $x\in[c]$, we evaluate the maximum number of error columns among all candidate matrices $R^{(x)}$. On one hand, row-wise counting of the error entries in $R^{(x)}$, compared to $R_0^{(x)}$, yields

$$\sum_{i\in[g]} d_{i,\sigma(i|x)}^{(x,\sigma(x))}. \quad (158)$$

On the other hand, column-wise counting of the error entries in $\widehat{R}^{(x)}$ (i.e., number of ones) yields

$$\sum_{k\in[\kappa]} \|w_k\|_1\, f_k^{(x)}. \quad (159)$$

From (154), we are interested in the class of candidate rating matrices where the clustering and grouping are done correctly without any errors in user associations to their respective clusters and groups. Therefore, the expressions given by (158) and (159) count the elements of the same set, and hence we obtain

$$\begin{aligned}\sum_{i\in[g]} d_{i,\sigma(i|x)}^{(x,\sigma(x))} &= \sum_{k\in[\kappa]} \|w_k\|_1\, f_k^{(x)} \\ &\ge (g-r+1)\sum_{k\in[\kappa]} f_k^{(x)} \quad (160)\\ &= (g-r+1) f^{(x)}, \quad (161)\end{aligned}$$

where (160) follows since the MDS code structure is known at the decoder side, and the fact that minimum distance between any two codewords in a $(g, r)$ linear MDS code is $g - r + 1$. Therefore, by (161), we get

$$\max f^{(x)} = \frac{1}{g-r+1} \sum_{i\in[g]} d_{i,\sigma(i|x)}^{(x,\sigma(x))}. \tag{162}$$

Finally, by (154) and (162), $\mathsf{Term}_{1,2}$ can be further upper bounded by

$$\begin{aligned}
\mathsf{Term}_{1,2} &\le \prod_{z\in[c]} \sum_{\ell=1}^{\max f^{(z)}} \left|\mathcal{R}^{(z)}\left(f^{(z)} = \ell, \{w_k : k \in [\kappa]\}\right)\right| \\
&\le \prod_{z\in[c]} \sum_{\ell=1}^{\max f^{(z)}} 2^{q^g-1}\, q^g\, (2qm)^\ell \qquad (163) \\
&\le \prod_{z\in[c]} 2^{q^g-1}\, q^g\, m^{\max f^{(z)}} \sum_{\ell=1}^{\max f^{(z)}} (2q)^\ell \\
&\le \prod_{z\in[c]} 2^{q^g-1}\, q^g\, m^{\max f^{(z)}}\, (2q)^{\max f^{(z)}+1} \qquad (164) \\
&= \prod_{z\in[c]} 2^{q^g}\, q^{g+1}\, (2qm)^{\max f^{(z)}} \\
&= \left(2^{q^g} q^{g+1}\right)^c (2qm)^{\sum_{x\in[c]}\max f^{(x)}} \\
&= c_0 \exp\left(\frac{\log(c_1 m)}{g-r+1} \sum_{x\in[c]}\sum_{i\in[g]} d_{i,\sigma(i|x)}^{(x,\sigma(x))}\right), \qquad (165)
\end{aligned}$$

where in (163) we used the bound in (157); (164) follows from $\sum_{\ell=1}^{L} x^\ell \le \sum_{\ell=0}^{L} x^\ell = (x^{L+1}-1)/(x-1) \le x^{L+1}$ for $x > 2$; and (165) follows by setting $c_0 = \left(2^{q^g} q^{g+1}\right)^c \ge 1$ and $c_1 = 2q \ge 1$.

Substituting (151) and (165) into (153), an upper bound on $\mathsf{Term}_1$ is thus given by

$$\begin{aligned}
\mathsf{Term}_1 \le c_0 \exp\Bigg[&\log n \left(\sum_{x\in[c]}\sum_{i\in[g]}\sum_{(y,j)\neq(\sigma(x),\sigma(i|x))} n_{i,j}^{(x,y)}\right) \\
&+ \frac{\log(c_1 m)}{g-r+1}\left(\sum_{x\in[c]}\sum_{i\in[g]} d_{i,\sigma(i|x)}^{(x,\sigma(x))}\right)\Bigg]. \qquad (166)
\end{aligned}$$

(2) Upper Bound on $\mathsf{Term}_2$: To this end, we derive lower bounds on the cardinalities of different sets in the exponent of $\mathsf{Term}_2$. Recall from (45) that

$$\begin{aligned}
\mathcal{T}_{\text{small}}^{(\delta)} &= \Big\{T \in \mathcal{T}^{(\delta)} : \forall (x,i) \in [c]\times[g] \text{ s.t. } |\sigma(x)| = 1, \\
&\qquad |\sigma(i|x)| = 1, d_{i,\,\sigma(i|x)}^{(x,\,\sigma(x))} \le \tau m \min\{\delta_g, \delta_c\}\Big\}, \\
&= \Big\{T \in \mathcal{T}^{(\delta)} : \forall (x,i) \in [c]\times[g] \text{ s.t. } \\
&\qquad n_{i,\sigma(i|x)}^{(x,\sigma(x))} \ge (1-\tau)\frac{n}{gc},\ d_{i,\sigma(i|x)}^{(x,\sigma(x))} \le \tau m \min\{\delta_g, \delta_c\}\Big\},
\end{aligned} \tag{167}$$

where (167) follows from (44).

(2-1) Lower Bound on $|\mathcal{P}_{\text{d}}|$: For $T \in \mathcal{T}_{\text{small}}^{(\delta)}$, a lower bound on $|\mathcal{P}_{\text{d}}|$ is given in (171), shown at the bottom of the page, where (168), shown at the bottom of the page, follows from the definitions in (29), (40) and (41); (169), shown at the bottom of the page, follows from (167) and the triangle inequality; and (170), shown at the bottom of the page, follows from (167) and the fact that the minimum Hamming distance between any two different rating vectors in $\mathcal{V}$ is $\min\{\delta_g, \delta_c\}m$.

$$\begin{aligned}
|\mathcal{P}_{\text{d}}| &= \sum_{x\in[c]}\sum_{i\in[g]}\sum_{y\in[c]}\sum_{j\in[g]} n_{i,j}^{(x,y)} d_{i,j}^{(x,y)} \qquad (168) \\
&= \left[\sum_{x\in[c]}\sum_{i\in[g]} n_{i,\sigma(i|x)}^{(x,\sigma(x))} d_{i,\sigma(i|x)}^{(x,\sigma(x))}\right] + \left[\sum_{x\in[c]}\sum_{i\in[g]}\sum_{j\in[g]\setminus\sigma(i|x)} n_{i,j}^{(x,\sigma(x))} d_{i,j}^{(x,\sigma(x))}\right] + \left[\sum_{x\in[c]}\sum_{y\in[c]\setminus\sigma(x)}\sum_{i\in[g]}\sum_{j\in[g]} n_{i,j}^{(x,y)} d_{i,j}^{(x,y)}\right] \\
&\ge \left(\sum_{x\in[c]}\sum_{i\in[g]} d_{i,\sigma(i|x)}^{(x,\sigma(x))}\left((1-\tau)\frac{n}{gc}\right)\right) + \left(\sum_{x\in[c]}\sum_{i\in[g]}\sum_{j\in[g]\setminus\sigma(i|x)} n_{i,j}^{(x,\sigma(x))}\left(d_{\text{H}}\left(v_{\sigma(i|x)}^{(\sigma(x))}, v_j^{(\sigma(x))}\right) - d_{i,\sigma(i|x)}^{(x,\sigma(x))}\right)\right) \\
&\quad + \left(\sum_{x\in[c]}\sum_{y\in[c]\setminus\sigma(x)}\sum_{i\in[g]}\sum_{j\in[g]} n_{i,j}^{(x,y)}\left(d_{\text{H}}\left(v_{\sigma(i|x)}^{(\sigma(x))}, v_j^{(y)}\right) - d_{i,\sigma(i|x)}^{(x,\sigma(x))}\right)\right) \qquad (169) \\
&\ge (1-\tau)\frac{n}{gc}\left(\sum_{x\in[c]}\sum_{i\in[g]} d_{i,\sigma(i|x)}^{(x,\sigma(x))}\right) + (\delta_g m - \delta_g \tau m)\left(\sum_{x\in[c]}\sum_{i\in[g]}\sum_{j\in[g]\setminus\sigma(i|x)} n_{i,j}^{(x,\sigma(x))}\right) \\
&\quad + (\delta_c m - \delta_c \tau m)\left(\sum_{x\in[c]}\sum_{y\in[c]\setminus\sigma(x)}\sum_{i\in[g]}\sum_{j\in[g]} n_{i,j}^{(x,y)}\right) \qquad (170) \\
&= (1-\tau)\left[\frac{n}{gc}\left(\sum_{x\in[c]}\sum_{i\in[g]} d_{i,\sigma(i|x)}^{(x,\sigma(x))}\right) + \delta_g m\left(\sum_{x\in[c]}\sum_{i\in[g]}\sum_{j\in[g]\setminus\sigma(i|x)} n_{i,j}^{(x,\sigma(x))}\right) + \delta_c m\left(\sum_{x\in[c]}\sum_{y\in[c]\setminus\sigma(x)}\sum_{i\in[g]}\sum_{j\in[g]} n_{i,j}^{(x,y)}\right)\right]. \qquad (171)
\end{aligned}$$

(2-2) Lower Bound on $|\mathcal{P}_{\widetilde{\beta}\to\widetilde{\alpha}}|$ and $|\mathcal{P}_{\widetilde{\alpha}\to\widetilde{\beta}}|$: For $T\in\mathcal{T}^{(\delta)}_{\text{small}}$, a lower bound on $|\mathcal{P}_{\widetilde{\beta}\to\widetilde{\alpha}}|$ is given by

$$\begin{aligned}|\mathcal{P}_{\widetilde{\beta}\to\widetilde{\alpha}}| &= \sum_{x\in[c]}\sum_{y\in[c]}\sum_{i\in[g]}\sum_{j\in[g]\setminus i}\sum_{k\in[g]} n_{i,k}^{(x,y)}n_{j,k}^{(x,y)} \quad (172)\\ &\geq \sum_{x\in[c]}\sum_{i\in[g]}\sum_{j\in[g]\setminus i}\sum_{k\in[g]} n_{i,k}^{(x,\sigma(x))}n_{j,k}^{(x,\sigma(x))}\\ &\geq \sum_{x\in[c]}\sum_{i\in[g]}\sum_{j\in[g]\setminus i}\sum_{k\in[g]\setminus\sigma(i|x)} n_{i,k}^{(x,\sigma(x))}n_{j,k}^{(x,\sigma(x))}\\ &= \sum_{x\in[c]}\sum_{i\in[g]}\sum_{k\in[g]\setminus\sigma(i|x)} n_{i,k}^{(x,\sigma(x))}\left(\sum_{j\in[g]\setminus i} n_{j,k}^{(x,\sigma(x))}\right)\\ &\geq \sum_{x\in[c]}\sum_{i\in[g]}\sum_{k\in[g]\setminus\sigma(i|x)} n_{i,k}^{(x,\sigma(x))}\left((1-\tau)\frac{n}{gc}\right) \quad (173)\\ &= (1-\tau)\frac{n}{gc}\sum_{x\in[c]}\sum_{i\in[g]}\sum_{j\in[g]\setminus\sigma(i|x)} n_{i,j}^{(x,\sigma(x))}, \quad (174)\end{aligned}$$

where (172) follows from (40) and (30), and (173) follows from (167). Similarly, $|\mathcal{P}_{\widetilde{\alpha}\to\widetilde{\beta}}|$ can be bounded by

$$\begin{aligned}|\mathcal{P}_{\widetilde{\alpha}\to\widetilde{\beta}}| &= \sum_{x\in[c]}\sum_{y\in[c]}\sum_{i\in[g]}\sum_{j\in[g]}\sum_{k\in[g]\setminus j} n_{i,j}^{(x,y)}n_{i,k}^{(x,y)} \quad (175)\\ &\geq \sum_{x\in[c]}\sum_{i\in[g]}\sum_{j\in[g]}\sum_{k\in[g]\setminus j} n_{i,j}^{(x,\sigma(x))}n_{i,k}^{(x,\sigma(x))}\\ &\geq \sum_{x\in[c]}\sum_{i\in[g]}\sum_{k\in[g]\setminus\sigma(i|x)} n_{i,\sigma(i|x)}^{(x,\sigma(x))}n_{i,k}^{(x,\sigma(x))}\\ &\geq \sum_{x\in[c]}\sum_{i\in[g]}\sum_{k\in[g]\setminus\sigma(i|x)} (1-\tau)\frac{n}{gc}n_{i,k}^{(x,\sigma(x))}\\ &= (1-\tau)\frac{n}{gc}\sum_{x\in[c]}\sum_{i\in[g]}\sum_{j\in[g]\setminus\sigma(i|x)} n_{i,j}^{(x,\sigma(x))}, \quad (176)\end{aligned}$$

where (175) follows from the definitions in (40) and (31). Therefore, by (174) and (176), we obtain

$$\frac{|\mathcal{P}_{\widetilde{\beta}\to\widetilde{\alpha}}|+|\mathcal{P}_{\widetilde{\alpha}\to\widetilde{\beta}}|}{2} \geq (1-\tau)\frac{n}{gc}\sum_{x\in[c]}\sum_{i\in[g]}\sum_{j\in[g]\setminus\sigma(i|x)} n_{i,j}^{(x,\sigma(x))}. \quad (177)$$

(2-3) Lower Bound on $|\mathcal{P}_{\widetilde{\gamma}\to\widetilde{\alpha}}|$ and $|\mathcal{P}_{\widetilde{\alpha}\to\widetilde{\gamma}}|$: For $T\in\mathcal{T}^{(\delta)}_{\text{small}}$, a lower bound on $|\mathcal{P}_{\widetilde{\gamma}\to\widetilde{\alpha}}|$ is given by

$$\begin{aligned}|\mathcal{P}_{\widetilde{\gamma}\to\widetilde{\alpha}}| &= \sum_{x\in[c]}\sum_{y\in[c]}\sum_{z\in[c]\setminus x}\sum_{i\in[g]}\sum_{j\in[g]}\sum_{k\in[g]} n_{i,j}^{(x,y)}n_{k,j}^{(z,y)} \quad (178)\\ &\geq \sum_{x\in[c]}\sum_{y\in[c]\setminus\sigma(x)}\sum_{z\in[c]\setminus x}\sum_{i\in[g]}\sum_{j\in[g]}\sum_{k\in[g]} n_{i,j}^{(x,y)}n_{k,j}^{(z,y)}\\ &= \sum_{x\in[c]}\sum_{y\in[c]\setminus\sigma(x)}\sum_{i\in[g]}\sum_{j\in[g]} n_{i,j}^{(x,y)}\left(\sum_{z\in[c]\setminus x}\sum_{k\in[g]} n_{k,j}^{(z,y)}\right)\\ &\geq \sum_{x\in[c]}\sum_{y\in[c]\setminus\sigma(x)}\sum_{i\in[g]}\sum_{j\in[g]} n_{i,j}^{(x,y)}\left((1-\tau)\frac{n}{gc}\right) \quad (179)\\ &= (1-\tau)\frac{n}{gc}\sum_{x\in[c]}\sum_{y\in[c]\setminus\sigma(x)}\sum_{i\in[g]}\sum_{j\in[g]} n_{i,j}^{(x,y)}, \quad (180)\end{aligned}$$

where (178) follows from (40) and (32), and (179) follows from (167). Similarly, we can bound $|\mathcal{P}_{\widetilde{\alpha}\to\widetilde{\gamma}}|$ as

$$\begin{aligned}|\mathcal{P}_{\widetilde{\alpha}\to\widetilde{\gamma}}| &= \sum_{x\in[c]}\sum_{y\in[c]}\sum_{z\in[c]\setminus y}\sum_{i\in[g]}\sum_{j\in[g]}\sum_{k\in[g]} n_{i,j}^{(x,y)}n_{i,k}^{(x,z)} \quad (181)\\ &\geq \sum_{x\in[c]}\sum_{z\in[c]\setminus\sigma(x)}\sum_{i\in[g]}\sum_{k\in[g]}\sum_{j\in[g]} n_{i,j}^{(x,\sigma(x))}n_{i,k}^{(x,z)}\\ &\geq \sum_{x\in[c]}\sum_{z\in[c]\setminus\sigma(x)}\sum_{i\in[g]}\sum_{k\in[g]} n_{i,\sigma(i|x)}^{(x,\sigma(x))}n_{i,k}^{(x,z)}\\ &= \sum_{x\in[c]}\sum_{z\in[c]\setminus\sigma(x)}\sum_{i\in[g]}\sum_{k\in[g]} (1-\tau)\frac{n}{gc}n_{i,k}^{(x,z)}\\ &= (1-\tau)\frac{n}{gc}\sum_{x\in[c]}\sum_{y\in[c]\setminus\sigma(x)}\sum_{i\in[g]}\sum_{j\in[g]} n_{i,j}^{(x,y)}, \quad (182)\end{aligned}$$

where (181) follows from the definitions in (40) and (33). Therefore, by (180) and (182), we obtain

$$\frac{|\mathcal{P}_{\widetilde{\gamma}\to\widetilde{\alpha}}|+|\mathcal{P}_{\widetilde{\alpha}\to\widetilde{\gamma}}|}{2} \geq (1-\tau)\frac{n}{gc}\sum_{x\in[c]}\sum_{y\in[c]\setminus x}\sum_{i\in[g]}\sum_{j\in[g]} n_{i,j}^{(x,y)}. \quad (183)$$

(2-4) Lower Bound on $|\mathcal{P}_{\widetilde{\gamma}\to\widetilde{\beta}}|$ and $|\mathcal{P}_{\widetilde{\beta}\to\widetilde{\gamma}}|$: For $T\in\mathcal{T}^{(\delta)}_{\text{small}}$, a lower bound on $|\mathcal{P}_{\widetilde{\gamma}\to\widetilde{\beta}}|$ is given by

$$\begin{aligned}|\mathcal{P}_{\widetilde{\gamma}\to\widetilde{\beta}}| &= \sum_{x\in[c]}\sum_{y\in[c]}\sum_{z\in[c]\setminus x}\sum_{i\in[g]}\sum_{k\in[g]}\sum_{j\in[g]}\sum_{\ell\in[g]\setminus j} n_{i,j}^{(x,y)}n_{k,\ell}^{(z,y)} \quad (184)\\ &\geq \sum_{x\in[c]}\sum_{y\in[c]\setminus\sigma(x)}\sum_{z\in[c]\setminus x}\sum_{i\in[g]}\sum_{k\in[g]}\sum_{j\in[g]}\sum_{\ell\in[g]\setminus j} n_{i,j}^{(x,y)}n_{k,\ell}^{(z,y)}\\ &= \sum_{x\in[c]}\sum_{y\in[c]\setminus\sigma(x)}\sum_{i\in[g]}\sum_{j\in[g]} n_{i,j}^{(x,y)}\left(\sum_{\ell\in[g]\setminus j}\sum_{z\in[c]\setminus x}\sum_{k\in[g]} n_{k,\ell}^{(z,y)}\right)\\ &\geq \sum_{x\in[c]}\sum_{y\in[c]\setminus\sigma(x)}\sum_{i\in[g]}\sum_{j\in[g]} n_{i,j}^{(x,y)}\left(\sum_{\ell\in[g]\setminus j}(1-\tau)\frac{n}{gc}\right) \quad (185)\\ &= (g-1)(1-\tau)\frac{n}{gc}\sum_{x\in[c]}\sum_{y\in[c]\setminus x}\sum_{i\in[g]}\sum_{j\in[g]} n_{i,j}^{(x,y)}, \quad (186)\end{aligned}$$

where (184) follows from (40) and (34), and (185) follows from (167). Similarly, $|\mathcal{P}_{\widetilde{\beta}\to\widetilde{\gamma}}|$ can be bounded by

$$\begin{aligned}|\mathcal{P}_{\widetilde{\beta}\to\widetilde{\gamma}}| &= \sum_{x\in[c]}\sum_{y\in[c]}\sum_{z\in[c]\setminus y}\sum_{i\in[g]}\sum_{k\in[g]\setminus i}\sum_{j\in[g]}\sum_{\ell\in[g]} n_{i,j}^{(x,y)}n_{k,\ell}^{(x,z)} \quad (187)\\ &\geq \sum_{x\in[c]}\sum_{z\in[c]\setminus\sigma(x)}\sum_{i\in[g]}\sum_{k\in[g]\setminus i}\sum_{j\in[g]}\sum_{\ell\in[g]} n_{i,j}^{(x,\sigma(x))}n_{k,\ell}^{(x,z)}\\ &= \sum_{x\in[c]}\sum_{z\in[c]\setminus\sigma(x)}\sum_{k\in[g]}\sum_{i\in[g]\setminus k} n_{i,\sigma(x|i)}^{(x,\sigma(x))}\sum_{\ell\in[g]} n_{k,\ell}^{(x,z)}\\ &\geq \sum_{x\in[c]}\sum_{z\in[c]\setminus\sigma(x)}\sum_{k\in[g]}\sum_{i\in[g]\setminus k} (1-\tau)\frac{n}{gc}\sum_{\ell\in[g]} n_{k,\ell}^{(x,z)}\\ &= \sum_{x\in[c]}\sum_{z\in[c]\setminus\sigma(x)}\sum_{k\in[g]} (g-1)(1-\tau)\frac{n}{gc}\sum_{\ell\in[g]} n_{k,\ell}^{(x,z)}\\ &= (g-1)(1-\tau)\frac{n}{gc}\sum_{x\in[c]}\sum_{y\in[c]\setminus\sigma(x)}\sum_{i\in[g]}\sum_{j\in[g]} n_{i,j}^{(x,y)}, \quad (188)\end{aligned}$$

$$\begin{aligned}
\mathsf{Term}_2 &= \exp\left[-(1+o(1))\left(|\mathcal{P}_{\mathrm{d}}|I_r + P_{\widetilde{\alpha}\leftrightarrow\widetilde{\beta}}\, I_{\alpha,\beta}\frac{\log n}{n} + P_{\widetilde{\alpha}\leftrightarrow\widetilde{\gamma}}\, I_{\alpha,\gamma}\frac{\log n}{n} + P_{\widetilde{\beta}\leftrightarrow\widetilde{\gamma}}\, I_{\beta,\gamma}\frac{\log n}{n}\right)\right] \\
&\le \exp\left[-(1-\tau)\left(\frac{nI_r}{gc}\left(\sum_{x\in[c]}\sum_{i\in[g]} d_{i,\sigma(i|x)}^{(x,\sigma(x))}\right) + \left(\delta_g m I_r + \frac{nI_{\alpha,\beta}}{gc}\frac{\log n}{n}\right)\left(\sum_{x\in[c]}\sum_{i\in[g]}\sum_{j\in[g]\setminus\sigma(i|x)} n_{i,j}^{(x,\sigma(x))}\right)\right.\right. \\
&\quad \left.\left. + \left(\delta_c m I_r + \frac{nI_{\alpha,\gamma}}{gc}\frac{\log n}{n} + \frac{(g-1)nI_{\beta,\gamma}}{gc}\frac{\log n}{n}\right)\left(\sum_{x\in[c]}\sum_{y\in[c]\setminus\sigma(x)}\sum_{i\in[g]}\sum_{j\in[g]} n_{i,j}^{(x,y)}\right)\right)\right] \\
&\le \exp\left[-(1-\tau)(1+\epsilon)\left(\frac{\log m}{g-r+1}\left(\sum_{x\in[c]}\sum_{i\in[g]} d_{i,\sigma(i|x)}^{(x,\sigma(x))}\right) + \log n\left(\sum_{x\in[c]}\sum_{i\in[g]}\sum_{j\in[g]\setminus\sigma(i|x)} n_{i,j}^{(x,\sigma(x))}\right)\right.\right. \\
&\quad \left.\left. + \log n\left(\sum_{x\in[c]}\sum_{y\in[c]\setminus\sigma(x)}\sum_{i\in[g]}\sum_{j\in[g]} n_{i,j}^{(x,y)}\right)\right)\right] \qquad (189) \\
&\le \exp\left[-\left(1+\frac{\epsilon}{2}\right)\left(\frac{\log(c_1 m)}{g-r+1}\left(\sum_{x\in[c]}\sum_{i\in[g]} d_{i,\sigma(i|x)}^{(x,\sigma(x))}\right) + \log n\left(\sum_{x\in[c]}\sum_{i\in[g]}\sum_{j\in[g]\setminus\sigma(i|x)} n_{i,j}^{(x,\sigma(x))}\right)\right.\right. \\
&\quad \left.\left. + \log n\left(\sum_{x\in[c]}\sum_{y\in[c]\setminus\sigma(x)}\sum_{i\in[g]}\sum_{j\in[g]} n_{i,j}^{(x,y)}\right)\right)\right] \qquad (190) \\
&= \exp\left[-\left(1+\frac{\epsilon}{2}\right)\left(\frac{\log(c_1 m)}{g-r+1}\left(\sum_{x\in[c]}\sum_{i\in[g]} d_{i,\sigma(i|x)}^{(x,\sigma(x))}\right) + \log n\left(\sum_{x\in[c]}\sum_{i\in[g]}\sum_{(y,j)\neq(\sigma(x),\sigma(i|x))} n_{i,j}^{(x,y)}\right)\right)\right]. \qquad (191)
\end{aligned}$$

where (187) follows from the definitions in (40) and (35). Therefore, from (186) and (188), we obtain

$$\frac{|\mathcal{P}_{\widetilde{\gamma}\to\widetilde{\beta}}| + |\mathcal{P}_{\widetilde{\beta}\to\widetilde{\gamma}}|}{2} \ge (1-\tau)\,(g-1)\frac{n}{gc}\sum_{x\in[c]}\sum_{y\in[c]\setminus x}\sum_{i\in[g]}\sum_{j\in[g]} n_{i,j}^{(x,y)}. \qquad (192)$$

Plugging (171), (177), (183) and (192) into definition of $\mathsf{Term}_2$ in (152), we can upper bound $\mathsf{Term}_2$ as given in (191), shown at the top of the page. Here, (189), shown at the top of the page, follows from the sufficient conditions in (12), (13) and (14); and (190), shown at the top of the page, holds since

$$\tau \le \frac{\epsilon\log m - (2+\epsilon)\log(2q)}{2(1+\epsilon)\log m},$$

which implies

$$(1-\tau)(1+\epsilon)\log m \ge \left(1+\frac{\epsilon}{2}\right)\log(c_1 m),$$

and

$$(1-\tau)(1+\epsilon) \ge 1 + \frac{\epsilon}{2}.$$

Finally, using (166) and (191), we can upper bound the function in the RHS of (152), as given in (200), shown at the bottom of the next page. Here,

- in (196), shown at the top of the next page, readily follows from (167);
- in (197), shown at the top of the next page, we break the summation into three summations, and use the fact that the enumeration of the first element of the set is independent of the enumeration of the second element;
- in (198), shown at the top of the next page, we use the fact that the number of integer solutions of $\sum_{i=1}^{n} x_i = s$ is equal to $\binom{s+n-1}{n-1}$;
- in (199), shown at the top of the next page, for $a \ge b$, we bound the binomial coefficient by $\binom{a}{b} \le \sum_{i=0}^{a}\binom{a}{i} \le 2^a$;
- and finally in (200), we evaluate the infinite geometric series, where
$$\epsilon > \max\left\{\frac{2\log 2}{\log n}, \frac{2(g-r+1)\log 2}{\log m}\right\}.$$

Therefore, from (200), the RHS of (152) can be simplified as given in (201), shown at the bottom of the next page. Note that as $n$ tends to infinity, the condition on $\epsilon$ becomes

$$\epsilon > \lim_{n,m\to\infty}\max\left\{\frac{2\log 2}{\log n}, \frac{2(g-r+1)\log 2}{\log(c_1 m)}, \frac{2\log c_1}{\log(m/c_1)}\right\} = 0. \qquad (193)$$

This completes the proof of Lemma 5. ■

## Appendix F
## Proof of Lemma 6

In order to prove the lemma, we first partition the set $\mathcal{T}_{\text{large}}^{(\delta)}$ into two disjoint subsets (regimes), denoted by $\mathcal{R}_1$ and $\mathcal{R}_2$. They are defined as

$$\mathcal{R}_1 = \left\{T \in \mathcal{T}_{\text{large}}^{(\delta)} : \exists (x,i) \in [c]\times[g] \text{ s.t. } |\sigma(x,i)| = 0\right\}, \qquad (194)$$

$$\mathcal{R}_2 = \left\{T \in \mathcal{T}_{\text{large}}^{(\delta)} : \forall (x,i) \in [c]\times[g] \text{ s.t. } |\sigma(x,i)| = 1, \text{ and } \exists (x,i) \in [c]\times[g] \text{ s.t. } d_{i,\sigma(i|x)}^{(x,\sigma(x))} > \tau m\min\{\delta_c,\delta_g\}\right\}. \qquad (195)$$

$$\sum_{T\in\mathcal{T}^{(\delta)}_{\text{small}}} |\mathcal{X}(T)| \ \exp\left[-(1+o(1))\left(|\mathcal{P}_{\text{d}}|I_r + P_{\widetilde{\alpha}\leftrightarrow\widetilde{\beta}}\, I_{\alpha,\beta}\frac{\log n}{n} + P_{\widetilde{\alpha}\leftrightarrow\widetilde{\gamma}}\, I_{\alpha,\gamma}\frac{\log n}{n} + P_{\widetilde{\beta}\leftrightarrow\widetilde{\gamma}}\, I_{\beta,\gamma}\frac{\log n}{n}\right)\right].$$

$$\leq \sum_{T\in\mathcal{T}^{(\delta)}_{\text{small}}} c_0 \exp\left[-\frac{\epsilon}{2}\left(\frac{\log(c_1 m)}{g-r+1}\left(\sum_{x\in[c]}\sum_{i\in[g]} d^{(x,\sigma(x))}_{i,\sigma(i|x)}\right) + \log n\left(\sum_{x\in[c]}\sum_{i\in[g]}\sum_{(y,j)\neq(\sigma(x),\sigma(i|x))} n^{(x,y)}_{i,j}\right)\right)\right].$$

$$= \sum_{\ell_1=0}^{gc\tau\min\{\delta_g,\delta_c\}m}\sum_{\ell_2=0}^{\tau n}\left|\left\{\sum_{x\in[c]}\sum_{i\in[g]} d^{(x,\sigma(x))}_{i,\sigma(i|x)}=\ell_1,\ \sum_{x\in[c]}\sum_{i\in[g]}\sum_{(y,j)\neq(\sigma(x),\sigma(i|x))} n^{(x,y)}_{i,j}=\ell_2\right\}\right|$$

$$\times\exp\left(-\frac{\epsilon\log(c_1 m)}{2(g-r+1)}\ell_1 - \frac{\epsilon\log n}{2}\ell_2\right) \tag{196}$$

$$= \sum_{\ell_1=1}^{gc\tau\min\{\delta_g,\delta_c\}m}\left|\left\{\sum_{x\in[c]}\sum_{i\in[g]} d^{(x,\sigma(x))}_{i,\sigma(i|x)}=\ell_1\right\}\right|\left|\left\{\sum_{x\in[c]}\sum_{i\in[g]}\sum_{(y,j)\neq(\sigma(x),\sigma(i|x))} n^{(x,y)}_{i,j}=0\right\}\right|\exp\left(-\frac{\epsilon\log(c_1 m)}{2(g-r+1)}\ell_1\right)$$

$$+\sum_{\ell_2=1}^{\tau n}\left|\left\{\sum_{x\in[c]}\sum_{i\in[g]} d^{(x,\sigma(x))}_{i,\sigma(i|x)}=0\right\}\right|\left|\left\{\sum_{x\in[c]}\sum_{i\in[g]}\sum_{(y,j)\neq(\sigma(x),\sigma(i|x))} n^{(x,y)}_{i,j}=\ell_2\right\}\right|\exp\left(-\frac{\epsilon\log n}{2}\ell_2\right)$$

$$+\sum_{\ell_1=1}^{gc\tau\min\{\delta_g,\delta_c\}m}\sum_{\ell_2=1}^{\tau n}\left|\left\{\sum_{x\in[c]}\sum_{i\in[g]} d^{(x,\sigma(x))}_{i,\sigma(i|x)}=\ell_1\right\}\right|\left|\left\{\sum_{x\in[c]}\sum_{i\in[g]}\sum_{(y,j)\neq(\sigma(x),\sigma(i|x))} n^{(x,y)}_{i,j}=\ell_2\right\}\right|$$

$$\times\exp\left(-\frac{\epsilon\log(c_1 m)}{2(g-r+1)}\ell_1 - \frac{\epsilon\log n}{2}\ell_2\right) \tag{197}$$

$$= \sum_{\ell_1=1}^{gc\tau\min\{\delta_g,\delta_c\}m}\binom{\ell_1+gc}{gc}\exp\left(-\frac{\epsilon\log(c_1 m)}{2(g-r+1)}\ell_1\right)+\sum_{\ell_2=1}^{\tau n}\binom{\ell_2+gc}{gc}\exp\left(-\frac{\epsilon\log n}{2}\ell_2\right)$$

$$+\sum_{\ell_1=1}^{gc\tau\min\{\delta_g,\delta_c\}m}\sum_{\ell_2=1}^{\tau n}\binom{\ell_1+gc-1}{gc-1}\binom{\ell_2+gc-1}{gc-1}\exp\left(-\frac{\epsilon\log(c_1 m)}{2(g-r+1)}\ell_1 - \frac{\epsilon\log n}{2}\ell_2\right) \tag{198}$$

$$\leq \sum_{\ell_1=1}^{gc\tau\min\{\delta_g,\delta_c\}m} 2^{(\ell_1+gc)}\ (c_1 m)^{\left(-\frac{\epsilon}{2(g-r+1)}\ell_1\right)} + \sum_{\ell_2=1}^{\tau n} 2^{(\ell_2+gc)}\ n^{\left(-\frac{\epsilon}{2}\ell_2\right)}$$

$$+\sum_{\ell_1=1}^{gc\tau\min\{\delta_g,\delta_c\}m} 2^{(\ell_1+gc)}\ (c_1 m)^{\left(-\frac{\epsilon}{2(g-r+1)}\ell_1\right)}\left(\sum_{\ell_2=1}^{\tau n} 2^{(\ell_2+gc)}\ n^{\left(-\frac{\epsilon}{2}\ell_2\right)}\right) \tag{199}$$

$$\leq 2^{gc}\sum_{\ell_1=1}^{\infty}\left(2\,(c_1 m)^{\left(-\frac{\epsilon}{2(g-r+1)}\right)}\right)^{\ell_1} + 2^{gc}\sum_{\ell_2=1}^{\infty}\left(2n^{\left(-\frac{\epsilon}{2}\right)}\right)^{\ell_2}$$

$$+2^{2gc}\sum_{\ell_1=1}^{\infty}\left(2\,(c_1 m)^{\left(-\frac{\epsilon}{2(g-r+1)}\right)}\right)^{\ell_1}\left[\sum_{\ell_2=1}^{\infty}\left(2n^{\left(-\frac{\epsilon}{2}\right)}\right)^{\ell_2}\right]$$

$$= 2^{gc}\frac{2\,(c_1 m)^{\left(-\frac{\epsilon}{2(g-r+1)}\right)}}{1-2\,(c_1 m)^{\left(-\frac{\epsilon}{2(g-r+1)}\right)}} + 2^{gc}\frac{2n^{\left(-\frac{\epsilon}{2}\right)}}{1-2\,n^{\left(-\frac{\epsilon}{2}\right)}} + 2^{2gc}\frac{2\,(c_1 m)^{\left(-\frac{\epsilon}{2(g-r+1)}\right)}}{1-2\,(c_1 m)^{\left(-\frac{\epsilon}{2(g-r+1)}\right)}}\frac{2n^{\left(-\frac{\epsilon}{2}\right)}}{1-2\,n^{\left(-\frac{\epsilon}{2}\right)}}. \tag{200}$$

$$\lim_{n,m\to\infty}\sum_{T\in\mathcal{T}^{(\delta)}_{\text{small}}}\sum_{X\in\mathcal{X}(T)}\exp\left[-(1+o(1))\left(|\mathcal{P}_{\text{d}}|I_r + P_{\widetilde{\alpha}\leftrightarrow\widetilde{\beta}}\, I_{\alpha,\beta}\frac{\log n}{n} + P_{\widetilde{\alpha}\leftrightarrow\widetilde{\gamma}}\, I_{\alpha,\gamma}\frac{\log n}{n} + P_{\widetilde{\beta}\leftrightarrow\widetilde{\gamma}}\, I_{\beta,\gamma}\frac{\log n}{n}\right)\right]=0. \tag{201}$$

$$\lim_{n,m\to\infty}\sum_{T\in\mathcal{T}^{(\delta)}_{\text{large}}}\sum_{X\in\mathcal{X}(T)}\exp\left[-(1+o(1))\left(|\mathcal{P}_{\text{d}}|I_r+P_{\widetilde{\alpha}\leftrightarrow\widetilde{\beta}}\,I_{\alpha,\beta}\frac{\log n}{n}+P_{\widetilde{\alpha}\leftrightarrow\widetilde{\gamma}}\,I_{\alpha,\gamma}\frac{\log n}{n}+P_{\widetilde{\beta}\leftrightarrow\widetilde{\gamma}}\,I_{\beta,\gamma}\frac{\log n}{n}\right)\right]$$

$$\leq\lim_{n,m\to\infty}\sum_{T\in\mathcal{T}^{(\delta)}_{\text{large}}}|\mathcal{X}(T)|\exp\left[-\left(|\mathcal{P}_{\text{d}}|I_r+P_{\widetilde{\alpha}\leftrightarrow\widetilde{\beta}}\,I_{\alpha,\beta}\frac{\log n}{n}+P_{\widetilde{\alpha}\leftrightarrow\widetilde{\gamma}}\,I_{\alpha,\gamma}\frac{\log n}{n}+P_{\widetilde{\beta}\leftrightarrow\widetilde{\gamma}}\,I_{\beta,\gamma}\frac{\log n}{n}\right)\right] \quad (202)$$

$$=\lim_{n,m\to\infty}\left(\sum_{T\in\mathcal{R}_1}|\mathcal{X}(T)|\exp\left[-\left(|\mathcal{P}_{\text{d}}|I_r+P_{\widetilde{\alpha}\leftrightarrow\widetilde{\beta}}\,I_{\alpha,\beta}\frac{\log n}{n}+P_{\widetilde{\alpha}\leftrightarrow\widetilde{\gamma}}\,I_{\alpha,\gamma}\frac{\log n}{n}+P_{\widetilde{\beta}\leftrightarrow\widetilde{\gamma}}\,I_{\beta,\gamma}\frac{\log n}{n}\right)\right]\right.$$

$$\left.+\sum_{T\in\mathcal{R}_2}|\mathcal{X}(T)|\exp\left[-\left(|\mathcal{P}_{\text{d}}|I_r+P_{\widetilde{\alpha}\leftrightarrow\widetilde{\beta}}\,I_{\alpha,\beta}\frac{\log n}{n}+P_{\widetilde{\alpha}\leftrightarrow\widetilde{\gamma}}\,I_{\alpha,\gamma}\frac{\log n}{n}+P_{\widetilde{\beta}\leftrightarrow\widetilde{\gamma}}\,I_{\beta,\gamma}\frac{\log n}{n}\right)\right]\right). \quad (203)$$

Now, we can start from the LHS of (48), and simplify it as given in (203), shown at the top of the page. Note that the summand in (48) does not depend on the actual rating matrix $X$, and hence the effect of the inner summation is only the number of such matrices, i.e., $|\mathcal{X}(T)|$, as given in (202), shown at the top of the page. In what follows, we derive upper bounds on each summation term corresponding to regimes $\mathcal{R}_1$ and $\mathcal{R}_2$.

### *A. Large Grouping Error Regime*

This regime corresponds to $\mathcal{R}_1$ characterized by (194). Suppose that there exist a cluster $x_0\in[c]$ and a group $i_0\in[g]$ such that $|\sigma(x_0,i_0)|=0$. From (44), we get

$$n^{(x_0,y)}_{i_0,j}=|Z_0(x_0,i_0)\cap Z(y,j)|\leq(1-\tau)\frac{n}{gc}, \quad (204)$$

for every $\forall(y,j)\in[c]\times[g]$. We further partition the set $\mathcal{R}_1$ into three sub-regimes, namely $\mathcal{R}_{1,1}$, $\mathcal{R}_{1,2}$, and $\mathcal{R}_{1,3}$, that are defined as

$$\mathcal{R}_{1,1}=\Big\{T\in\mathcal{R}_1\colon \exists\mu>0,\ \exists(y_1,j_1)\in[c]\times[g],\ \exists(y_2,j_2)\in[c]\times[g]\ \text{s.t. } n^{(x_0,y_1)}_{i_0,j_1}\geq\mu n,\ n^{(x_0,y_2)}_{i_0,j_2}\geq\mu n\Big\}, \quad (205)$$

$$\mathcal{R}_{1,2}=\Big\{T\in\mathcal{R}_1\colon \exists\mu>0,\ \exists(y_1,j_1)\in[c]\times[g]\ \text{s.t. } n^{(x_0,y_1)}_{i_0,j_1}\geq\mu n\Big\}, \quad (206)$$

$$\mathcal{R}_{1,3}=\Big\{T\in\mathcal{R}_1\colon \forall\mu>0,\ \forall(y,j)\in[c]\times[g]\ \text{s.t. } n^{(x_0,y)}_{i_0,j}<\mu n\Big\}. \quad (207)$$

(1) Sub-regime 1-1: Consider the sub-regime $\mathcal{R}_{1,1}$ as given in (205). Suppose that there exist a constant $\mu>0$, and two distinct pairs $(y_1,j_1),(y_2,j_2)\in[c]\times[g]$ such that

$$n^{(x_0,y_1)}_{i_0,j_1}\geq\mu n,\ \text{ and } n^{(x_0,y_2)}_{i_0,j_2}\geq\mu n. \quad (208)$$

Recall that there are a total of $n$ users, each of which belongs to one of the $gc$ groups. Hence, the number of user-to-group associations can be (loosely) bounded by $(gc)^n$. On the other hand, there are $m$ items, each with a rating in $\mathbb{F}_q$ from each of the $gc$ groups of users. Hence, each item rating vector can be one of $q^{gc}$ possible vectors across all users. Therefore, a loose upper bound on the number of matrices that belong to matrix class $\mathcal{X}(T)$ is given by

$$|\mathcal{X}(T)|\leq(gc)^n\left(q^{gc}\right)^m,\quad \forall T\in\mathcal{T}^{(\delta)}. \quad (209)$$

Next, we can lower bound the cardinality of the set $\mathcal{P}_{\text{d}}$ as

$$\begin{aligned}|\mathcal{P}_{\text{d}}|&=\sum_{x\in[c]}\sum_{i\in[g]}\sum_{y\in[c]}\sum_{j\in[g]}n^{(x,y)}_{i,j}d^{(x,y)}_{i,j} &(210)\\ &\geq n^{(x_0,y_1)}_{i_0,j_1}d^{(x_0,y_1)}_{i_0,j_1}+n^{(x_0,y_2)}_{i_0,j_2}d^{(x_0,y_2)}_{i_0,j_2}\\ &>\mu n\left(d^{(x_0,y_1)}_{i_0,j_1}+d^{(x_0,y_2)}_{i_0,j_2}\right) &(211)\\ &\geq\mu n\, d_{\text{H}}\left(v^{(y_1)}_{j_1},v^{(y_2)}_{j_2}\right) &(212)\\ &\geq\mu\min\{\delta_g,\delta_c\}\,nm, &(213)\end{aligned}$$

where (210) follows from the definitions of $\mathcal{P}_{\text{d}}$, $n^{(x,y)}_{i,j}$, and $d^{(x,y)}_{i,j}$ in (29), (40) and (41), respectively; (211) follows from (208); (212) follows from the triangle inequality; and (213) holds since the minimum Hamming distance between any two different rating vectors in $\mathcal{V}$ is $\min\{\delta_g,\delta_c\}m$. Furthermore, if $y_1=y_2$, then $|\mathcal{P}_{\widetilde{\alpha}\to\widetilde{\beta}}|$ is lower bounded by

$$\begin{aligned}|\mathcal{P}_{\widetilde{\alpha}\to\widetilde{\beta}}|&=\sum_{x\in[c]}\sum_{y\in[c]}\sum_{i\in[g]}\sum_{k\in[g]}\sum_{j\in[g]\setminus k}n^{(x,y)}_{i,j}n^{(x,y)}_{i,k} &(214)\\ &\geq n^{(x_0,y_1)}_{i_0,j_1}n^{(x_0,y_2)}_{i_0,j_2}\geq(\mu n)^2, &(215)\end{aligned}$$

where (214) follows from the definitions in (40) and (31). On the other hand, if $y_1\neq y_2$, then $|\mathcal{P}_{\widetilde{\alpha}\to\widetilde{\gamma}}|$ is lower bounded by

$$\begin{aligned}|\mathcal{P}_{\widetilde{\alpha}\to\widetilde{\gamma}}|&=\sum_{x\in[c]}\sum_{z\in[c]}\sum_{y\in[c]\setminus z}\sum_{i\in[g]}\sum_{k\in[g]}\sum_{j\in[g]}n^{(x,y)}_{i,j}n^{(x,z)}_{i,k} &(216)\\ &\geq n^{(x_0,y_1)}_{i_0,j_1}n^{(x_0,y_2)}_{i_0,j_2}\geq(\mu n)^2, &(217)\end{aligned}$$

where (216) follows from (40) and (33).

Finally, the first summation term in the RHS of (202) is upper bounded by

$$\begin{aligned}&\sum_{T\in\mathcal{R}_{1,1}}|\mathcal{X}(T)|\exp\Bigg[-\Bigg(|\mathcal{P}_{\text{d}}|I_r+P_{\widetilde{\alpha}\leftrightarrow\widetilde{\beta}}\,I_{\alpha,\beta}\frac{\log n}{n}\\ &\qquad+P_{\widetilde{\alpha}\leftrightarrow\widetilde{\gamma}}\,I_{\alpha,\gamma}\frac{\log n}{n}+P_{\widetilde{\beta}\leftrightarrow\widetilde{\gamma}}\,I_{\beta,\gamma}\frac{\log n}{n}\Bigg)\Bigg]\\ &\leq\sum_{T\in\mathcal{R}_{1,1}}|\mathcal{X}(T)|\exp\Bigg[-\Bigg(|\mathcal{P}_{\text{d}}|I_r+\frac{I_{\alpha,\beta}}{2}\frac{\log n}{n}|\mathcal{P}_{\widetilde{\alpha}\to\widetilde{\beta}}|\\ &\qquad+\frac{I_{\alpha,\gamma}}{2}\frac{\log n}{n}|\mathcal{P}_{\widetilde{\alpha}\to\widetilde{\gamma}}|\Bigg)\Bigg]\\ &=\exp\left[-\left(c_2nm\frac{\log m}{n}+c_3\frac{\log n}{n}n^2\right)\right]\sum_{T\in\mathcal{R}_{1,1}}|\mathcal{X}(T)|\end{aligned} \quad (218)$$

$$|\mathcal{P}_{\rm d}| = \sum_{x\in[c]}\sum_{i\in[g]}\sum_{y\in[c]}\sum_{j\in[g]} n_{i,j}^{(x,y)} d_{i,j}^{(x,y)} \geq n_{i_0,j_0}^{(x_0,y_0)} d_{i_0,j_0}^{(x_0,y_0)}$$

$$> \left((1-\tau)\frac{n}{gc}\right)(\tau m \min\{\delta_c,\delta_g\}) \tag{219}$$

$$= \left(\frac{(1-\tau)\tau\min\{\delta_c,\delta_g\}}{2gc}\right)mn + \left(\frac{(1-\tau)\tau\min\{\delta_c,\delta_g\}}{2gc}\right)mn$$

$$\geq \left(\frac{(1-\tau)\tau\min\{\delta_c,\delta_g\}}{2gc}\right)mn + ((1-\tau)m)\left(\sum_{x\in[c]}\sum_{i\in[g]}\sum_{(y,j)\neq\sigma(x,i)} n_{i,j}^{(x,y)}\right) \tag{220}$$

$$= \left(\frac{(1-\tau)\tau\min\{\delta_c,\delta_g\}}{2gc}\right)mn + ((1-\tau)m)\left(\sum_{x\in[c]}\sum_{i\in[g]}\sum_{j\in[g]\backslash\sigma(i|x)} n_{i,j}^{(x,\sigma(x))} + \sum_{x\in[c]}\sum_{y\in[c]\backslash\sigma(x)}\sum_{i\in[g]}\sum_{j\in[g]} n_{i,j}^{(x,y)}\right)$$

$$\geq c_4 mn + ((1-\tau)m)\left[\delta_g\left(\sum_{x\in[c]}\sum_{i\in[g]}\sum_{j\in[g]\backslash\sigma(i|x)} n_{i,j}^{(x,\sigma(x))}\right) + \delta_c\left(\sum_{x\in[c]}\sum_{y\in[c]\backslash\sigma(x)}\sum_{i\in[g]}\sum_{j\in[g]} n_{i,j}^{(x,y)}\right)\right]. \tag{221}$$

---

$$\leq \exp\left[-(c_2 m\log m + c_3 n\log n) + n\log(gc) + mgc\log q\right] \tag{222}$$

$$= \exp\left[-\left(m\left(c_2\log m - gc\log q\right) + n\left(c_3\log n - \log(gc)\right)\right)\right], \tag{223}$$

where $c_2$ and $c_3$ in (218) are some positive constants; (218) follows from (12), (213), (215) and (217); and (222) follows from (209).

(2) Sub-regime 1-2: This sub-regime corresponds to $\mathcal{R}_{1,2}$ characterized by (206). Suppose that there exists only one pair $(y_1,j_1)\in[c]\times[g]$, and a constant $\mu>0$ such that $n_{i_0,j_1}^{(x_0,y_1)}\geq\mu n$. This implies that

$$n_{i_0,j}^{(x_0,y)} < \frac{\tau}{(gc-1)gc}n, \quad \text{for } (y,j)\neq(y_0,j_0). \tag{224}$$

Therefore, by (224), we have

$$n_{i_0,j_0}^{(x_0,y_0)} = \frac{n}{gc} - \sum_{(y,j)\neq(y_0,j_0)} n_{i_0,j}^{(x_0,y)} > \frac{n}{gc} - (gc-1)\frac{\tau}{(gc-1)}\frac{n}{gc} = (1-\tau)\frac{n}{gc}. \tag{225}$$

However, this is in contradiction with (204). Hence, we conclude that sub-regime $\mathcal{R}_{1,2}$ is impossible to exist.

(3) Sub-regime 1-3: This sub-regime corresponds to $\mathcal{R}_{1,3}$ characterized by (207). Due to the fact that

$$\sum_{y\in[c]}\sum_{j\in[g]} n_{i_0,j}^{(x_0,y)} = \frac{n}{gc}, \tag{226}$$

there should be at least one pair $(y_1,j_1)$ with $n_{i_0,j_1}^{(x_0,y_1)}\geq\mu n$, for some $\mu>0$. However, this is in contradiction with (207). Thus, we conclude that sub-regime $\mathcal{R}_{1,3}$ is impossible to exist.

As a result, we conclude that

$$\sum_{T\in\mathcal{R}_1}|\mathcal{X}(T)|\exp\left[-\left(|\mathcal{P}_{\rm d}|I_r + P_{\widetilde{\alpha}\leftrightarrow\widetilde{\beta}}\, I_{\alpha,\beta}\frac{\log n}{n} + P_{\widetilde{\alpha}\leftrightarrow\widetilde{\gamma}}\, I_{\alpha,\gamma}\frac{\log n}{n} + P_{\widetilde{\beta}\leftrightarrow\widetilde{\gamma}}\, I_{\beta,\gamma}\frac{\log n}{n}\right)\right]$$
$$\leq \exp\left[-\left(m\left(c_2\log m - gc\log q\right) + n\left(c_3\log n - \log(gc)\right)\right)\right]. \tag{227}$$

### B. Large Rating Estimation Error Regime

This regime corresponds to $\mathcal{R}_2$ characterized by (195). Suppose that the following conditions hold:

- For every $(x,i)\in[c]\times[g]$, there is a pair $(y,j)\in[c]\times[g]$ such that $|\sigma(x,i)|=1$. More precisely, (from (44)) for every $(x,i)\in[c]\times[g]$ we have
$$\exists(y,j)=(\sigma(x),\sigma(i|x))\in[c]\times[g]:\; n_{i,j}^{(x,y)}=|Z_0(x,i)\cap Z(y,j)|\geq(1-\tau)\frac{n}{gc}; \tag{228}$$
- There exists $(x_0,i_0)\in[c]\times[g]$ with $|\sigma(x_0,i_0)|=1$, and
$$d_{i_0,j_0}^{(x_0,y_0)} = d_{i_0,\sigma(i_0|x_0)}^{(x_0,\sigma(x_0))} > \tau m\min\{\delta_c,\delta_g\}. \tag{229}$$

We first provide an upper bound on $|\mathcal{X}(T)|$. By (153), (151) and (209), an upper bound on $|\mathcal{X}(T)|$ is given by

$$|\mathcal{X}(T)| \leq (q^{gc})^m \exp\left[\left(\sum_{x\in[c]}\sum_{i\in[g]}\sum_{\substack{(y,j)\neq\\(\sigma(x),\sigma(i|x))}} n_{i,j}^{(x,y)}\right)\log n\right]. \tag{230}$$

Next, we provide a lower bound on $|\mathcal{P}_{\rm d}|$. Based on (228), if there exists at least one other pair $(\widehat{y},\widehat{j})\neq(\sigma(x),\sigma(i|x))$ for some $(x,i)\in[c]\times[g]$ such that

$$n_{i,\widehat{j}}^{(x,\widehat{y})} = \left|Z_0(x,i)\cap Z(\widehat{y},\widehat{j})\right| \geq \mu n, \tag{231}$$

for some constant $\mu>0$, then the analysis of this case boils down to Sub-regime 1-1. Therefore, we assume that for every $(x,i)\in[c]\times[g]$, we have

$$n_{i,j}^{(x,y)} < \mu n,\ \forall(y,j)\neq(\sigma(x),\sigma(i|x)),\ \forall\mu>0. \tag{232}$$

Consequently, the size of the set $\mathcal{P}_{\rm d}$ can be lower bounded as in (221), shown at the top of the page. Here, (219), shown at the top of the page, follows from (228) and (229); (220), shown at the top of the page, follows from (232) for $\mu=(\tau\min\{\delta_g,\delta_c\})/(2(gc-1)(gc)^2)$; and (221) follows by setting $c_4=((1-\tau)\tau\min\{\delta_c,\delta_g\})/(2gc)$, where $0\leq c_4<1$.

On the other hand, recall from (177), (183) and (192) that

$$\frac{|\mathcal{P}_{\widetilde{\beta}\to\widetilde{\alpha}}|+|\mathcal{P}_{\widetilde{\alpha}\to\widetilde{\beta}}|}{2} \geq \frac{(1-\tau)n}{gc}\sum_{x\in[c]}\sum_{i\in[g]}\sum_{j\in[g]\setminus\sigma(i|x)} n_{i,j}^{(x,\sigma(x))}, \tag{233}$$

$$\frac{|\mathcal{P}_{\widetilde{\gamma}\to\widetilde{\alpha}}|+|\mathcal{P}_{\widetilde{\alpha}\to\widetilde{\gamma}}|}{2} \geq \frac{(1-\tau)n}{gc}\sum_{x\in[c]}\sum_{y\in[c]\setminus\sigma(x)}\sum_{i\in[g]}\sum_{j\in[g]} n_{i,j}^{(x,y)}, \tag{234}$$

and

$$\frac{|\mathcal{P}_{\widetilde{\gamma}\to\widetilde{\beta}}|+|\mathcal{P}_{\widetilde{\beta}\to\widetilde{\gamma}}|}{2} \geq (g-1)\frac{(1-\tau)n}{gc}\sum_{x\in[c]}\sum_{y\in[c]\setminus\sigma(x)}\sum_{i\in[g]}\sum_{j\in[g]} n_{i,j}^{(x,y)}. \tag{235}$$

Finally, the second term in the RHS of (202) is upper bounded as in (243), given at the bottom of this page. Here,

$$\begin{aligned}
&\sum_{T\in\mathcal{R}_2}|\mathcal{X}(T)|\exp\left[-\left(|\mathcal{P}_{\rm d}|I_r + P_{\widetilde{\alpha}\leftrightarrow\widetilde{\beta}}\, I_{\alpha,\beta}\frac{\log n}{n} + P_{\widetilde{\alpha}\leftrightarrow\widetilde{\gamma}}\, I_{\alpha,\gamma}\frac{\log n}{n} + P_{\widetilde{\beta}\leftrightarrow\widetilde{\gamma}}\, I_{\beta,\gamma}\frac{\log n}{n}\right)\right] \\
&\leq \sum_{T\in\mathcal{R}_2}|\mathcal{X}(T)|\exp\left(-c_4 mn\frac{\log m}{n}\right)\exp\left[-(1-\tau)\left(\left(\delta_g m I_r + \frac{n}{gc}I_{\alpha,\beta}\frac{\log n}{n}\right)\sum_{x\in[c]}\sum_{i\in[g]}\sum_{j\in[g]\setminus\sigma(i|x)} n_{i,j}^{(x,\sigma(x))}\right.\right. \\
&\quad \left.\left.+\left(\delta_c m I_r + \frac{nI_{\alpha,\gamma}}{gc}\frac{\log n}{n} + \frac{(g-1)nI_{\beta,\gamma}}{gc}\frac{\log n}{n}\right)\sum_{x\in[c]}\sum_{y\in[c]\setminus x}\sum_{i\in[g]}\sum_{j\in[g]} n_{i,j}^{(x,y)}\right)\right] \qquad (236) \\
&\leq \sum_{T\in\mathcal{R}_2}\exp\left[\log n\left(\sum_{x\in[c]}\sum_{i\in[g]}\sum_{(y,j)\neq\sigma(x,i)} n_{i,j}^{(x,y)}\right)\right]\times (q^{gc})^m\exp\left(-c_4 m\log m\right) \\
&\quad\times\exp\left[-(1-\tau)(1+\epsilon)\log n\left(\sum_{x\in[c]}\sum_{i\in[g]}\sum_{j\in[g]\setminus\sigma(i|x)} n_{i,j}^{(x,\sigma(x))} + \sum_{x\in[c]}\sum_{y\in[c]\setminus x}\sum_{i\in[g]}\sum_{j\in[g]} n_{i,j}^{(x,y)}\right)\right] \qquad (237) \\
&\leq \sum_{T\in\mathcal{R}_2}(q^{gc})^m\exp\left(-c_4 m\log m\right)\exp\left[\log n\left(\sum_{x\in[c]}\sum_{i\in[g]}\sum_{(y,j)\neq\sigma(x,i)} n_{i,j}^{(x,y)}\right)\right] \\
&\quad\times\exp\left[-\left(1+\frac{\epsilon}{2}\right)\log n\left(\sum_{x\in[c]}\sum_{i\in[g]}\sum_{(y,j)\neq\sigma(x,i)} n_{i,j}^{(x,y)}\right)\right] \qquad (238) \\
&= \exp\left[-m\left(c_4\log m - gc\log q\right)\right]\sum_{T\in\mathcal{R}_2}\exp\left[-\frac{\epsilon}{2}\log n\left(\sum_{x\in[c]}\sum_{i\in[g]}\sum_{(y,j)\neq\sigma(x,i)} n_{i,j}^{(x,y)}\right)\right] \\
&= \exp\left[-m\left(c_4\log m - gc\log q\right)\right]\sum_{\ell=0}^{\tau n}\left|\left\{\sum_{x\in[c]}\sum_{i\in[g]}\sum_{(y,j)\neq(\sigma(x),\sigma(i|x))} n_{i,j}^{(x,y)} = \ell\right\}\right|\exp\left(-\frac{\epsilon\log n}{2}\ell\right) \qquad (239) \\
&= \exp\left[-m\left(c_4\log m - gc\log q\right)\right]\left[\left|\left\{\sum_{x\in[c]}\sum_{i\in[g]}\sum_{(y,j)\neq(\sigma(x),\sigma(i|x))} n_{i,j}^{(x,y)} = 0\right\}\right|\right. \\
&\quad\left.+\sum_{\ell=1}^{\tau n}\left|\left\{\sum_{x\in[c]}\sum_{i\in[g]}\sum_{(y,j)\neq(\sigma(x),\sigma(i|x))} n_{i,j}^{(x,y)} = \ell\right\}\right|\exp\left(-\frac{\epsilon\log n}{2}\ell\right)\right] \qquad (240) \\
&= \exp\left[-m\left(c_4\log m - gc\log q\right)\right]\left[1+\sum_{\ell=1}^{\tau n}\binom{\ell+gc}{gc}\exp\left(-\frac{\epsilon\log n}{2}\ell\right)\right] \qquad (241) \\
&\leq \exp\left[-m\left(c_4\log m - gc\log q\right)\right]\left[1+\sum_{\ell_2=1}^{\tau n}2^{(\ell_2+gc)}\; n^{\left(-\frac{\epsilon}{2}\ell_2\right)}\right] \qquad (242) \\
&\leq \exp\left[-m\left(c_4\log m - gc\log q\right)\right]\left[1+2^{gc}\sum_{\ell_2=1}^{\infty}\left(2n^{-\epsilon/2}\right)^{\ell_2}\right] \\
&= \exp\left[-m\left(c_4\log m - gc\log q\right)\right]\left[1+2^{gc}\frac{2n^{-\epsilon/2}}{1-2\;n^{-\epsilon/2}}\right]. \qquad (243)
\end{aligned}$$

- in (236), shown at the bottom of the previous page, follows from (221) and (233)-(235);
- in (237), shown at the bottom of the previous page, follows from the sufficient conditions in (12), (13) and (14);
- in (238), shown at the bottom of the previous page, follows from

$$\tau \leq [\epsilon \log m - (2+\epsilon)\log(2q)]/[2(1+\epsilon)\log m)],$$

which implies that $(1-\tau)(1+\epsilon) \geq (1+(\epsilon/2))$;
- in (239), shown at the bottom of the previous page, readily follows from (228);
- in (240), shown at the bottom of the previous page, we break the summation into two summations and use the fact that the enumeration of the first element of the set is independent of the enumeration of the second element;
- in (241), shown at the bottom of the previous page, we use the fact that the number of integer solutions of $\sum_{i=1}^{n} x_i = s$ is equal to $\binom{s+n-1}{n-1}$;
- in (242), shown at the bottom of the previous page, for $a \geq b$, we bound each binomial coefficient by $\binom{a}{b} \leq \sum_{i=0}^{a}\binom{a}{i} \leq 2^a$;
- and finally in (243),we evaluate the infinite geometric series, where $\epsilon > (2\log 2)/\log n$.

By (227) and (243), the RHS of (203) is upper bounded by

$$\begin{aligned}
&\lim_{n,m\to\infty} \sum_{T\in\mathcal{T}^{(\delta)}_{\text{large}}} \sum_{X\in\mathcal{X}(T)} \exp\Bigg[-(1+o(1)) \\
&\left(|\mathcal{P}_{\text{d}}|I_r + \frac{\log n}{n}\left(P_{\widetilde{\alpha}\leftrightarrow\widetilde{\beta}}I_{\alpha,\beta} + P_{\widetilde{\alpha}\leftrightarrow\widetilde{\gamma}}I_{\alpha,\gamma} + P_{\widetilde{\beta}\leftrightarrow\widetilde{\gamma}}I_{\beta,\gamma}\right)\right)\Bigg] \\
&\leq \lim_{n,m\to\infty}\Bigg(\exp\left[-\left(m\left(c_2\log m - gc\log q\right)\right.\right. \\
&\qquad\qquad \left.\left.+n\left(c_3\log n - \log(gc)\right)\right)\right] \\
&\quad + \exp\left[-m\left(c_4\log m - gc\log q\right)\right]\left[1 + 2^{gc}\frac{2n^{-\frac{\epsilon}{2}}}{1-2\,n^{-\frac{\epsilon}{2}}}\right]\Bigg) \\
&= 0. \qquad (244)
\end{aligned}$$

Note that as $n$ tends to infinity, the condition on $\epsilon$ becomes

$$\epsilon > \lim_{n,m\to\infty} \max\left\{\frac{2\log 2}{\log n}, \frac{2(g-r+1)\log 2}{\log(c_1 m)}, \frac{2\log c_1}{\log(m/c_1)}\right\} = 0. \qquad (245)$$

This completes the proof of Lemma 6. ∎

## APPENDIX G
## PROOF OF LEMMA 7

We start with the proof with the minimax optimization approach in (8) to minimize the maximum worst-case probability of error as follows:

$$\begin{aligned}
&\inf_{\psi} P_e^{(\delta)}(\psi) \\
&= \inf_{\psi}\max_{M\in\mathcal{M}^{(\delta)}} \mathbb{P}\left[\psi(Y^\Omega, G) \neq M\right] \\
&\geq \inf_{\psi}\max_{M\in\mathcal{M}^{(\delta)}} \mathbb{P}\left[\psi(Y^\Omega, G) \neq M,\ \mathbf{M} = M\right] \\
&= \inf_{\psi}\max_{M\in\mathcal{M}^{(\delta)}} \mathbb{P}\left[\psi(Y^\Omega, G) \neq M \mid \mathbf{M} = M\right] \qquad (246) \\
&= \inf_{\psi}\max_{M\in\mathcal{M}^{(\delta)}} \sum_{X\neq M} \mathbb{P}\left[\psi(Y^\Omega, G) = X \mid \mathbf{M} = M\right] \\
&= \max_{M\in\mathcal{M}^{(\delta)}} \sum_{X\neq M} \mathbb{P}\left[\psi_{\text{ML}}(Y^\Omega, G) = X \mid \mathbf{M} = M\right] \qquad (247) \\
&\geq \sum_{X\neq X_0} \mathbb{P}\left[\psi_{\text{ML}}(Y^\Omega, G) = X \mid \mathbf{M} = X_0\right] \qquad (248) \\
&= \sum_{X\neq X_0} \mathbb{P}\left[\mathsf{L}(X) \leq \mathsf{L}(X_0)\right] \qquad (249) \\
&\geq \mathbb{P}\left[\bigcup_{X\neq X_0} \left(\mathsf{L}(X) \leq \mathsf{L}(X_0)\right)\right] \qquad (250) \\
&= \mathbb{P}\left[S^c\right], \qquad (251)
\end{aligned}$$

where (246) holds for $\mathbf{M}$ with uniform distribution; (247) follows from the fact that the maximum likelihood estimator is optimal under a uniform prior; (248) follows since $X_0 \in \mathcal{M}^{(\delta)}$ whose construction is given in Section IV; (249) follows by the definition of maximum likelihood estimation; (250) follows from the union bound; and finally (251) follows from (58). This completes the proof of Lemma 7. ∎

## APPENDIX H
## PROOF OF LEMMA 8

Recall from Appendix D the definition of $\mathbf{U}_i = \mathbf{U}_i(p,\theta,q)$ in (122), and the expression of $-\log M_{\mathbf{U}_i(p,\theta,q)}\left(\frac{1}{2}\right)$ in (129). Define a related random variable $\widehat{\mathbf{U}}_i = \widehat{\mathbf{U}}_i(p,\theta,q)$ that has the same sample space as $\mathbf{U}_i(p,\theta,q)$, but its probability mass function is given by

$$f_{\widehat{\mathbf{U}}_i(p,\theta,q)}(u) = \frac{\exp\left(\frac{1}{2}u\right) f_{\mathbf{U}_i(p,\theta,q)}(u)}{M_{\mathbf{U}_i(p,\theta,q)}\left(\frac{1}{2}\right)}. \qquad (252)$$

More formally, $\widehat{\mathbf{U}}_i(p,\theta,q)$ is defined as

$$\widehat{\mathbf{U}}_i(p,\theta,q) = \begin{cases} -\log\left((q-1)\frac{1-\theta}{\theta}\right) & \text{w.p. } \frac{1}{M_{\mathbf{U}_i}\left(\frac{1}{2}\right)}\sqrt{\frac{\theta(1-\theta)}{q-1}}p, \\ 0 & \text{w.p. } \frac{(1-p)+p\theta\left(1-\frac{1}{q-1}\right)}{M_{\mathbf{U}_i}\left(\frac{1}{2}\right)}, \\ \log\left((q-1)\frac{1-\theta}{\theta}\right) & \text{w.p. } \frac{1}{M_{\mathbf{U}_i}\left(\frac{1}{2}\right)}\sqrt{\frac{\theta(1-\theta)}{q-1}}p, \end{cases} \qquad (253)$$

from which one can readily show that

$$\mathbb{E}\left[\widehat{\mathbf{U}}_i(p,\theta,q)\right] = 0, \qquad (254)$$

$$\begin{aligned}
\mathsf{Var}\left[\widehat{\mathbf{U}}_i(p,\theta,q)\right] &= \frac{2\left(\log\left((q-1)\frac{1-\theta}{\theta}\right)\right)^2\sqrt{\frac{\theta(1-\theta)}{q-1}}p}{1-\left(\sqrt{1-\theta}-\sqrt{\frac{\theta}{q-1}}\right)^2 p} \\
&= O(p). \qquad (255)
\end{aligned}$$

Similarly, we can use the definition of $\mathbf{V}_j = \mathbf{V}_j(\mu,\nu)$ in (130) in Appendix D, and the expression of $-\log M_{\mathbf{V}_j(\mu,\nu)}\left(\frac{1}{2}\right)$ in (134). Define a related random variable $\widehat{\mathbf{V}}_j = \widehat{\mathbf{V}}_j(\mu,\nu)$ that has the same sample space as $\mathbf{V}_j(\mu,\nu)$, but its probability mass function is given by

$$f_{\widehat{\mathbf{V}}_j(\mu,\nu)}(v) = \frac{\exp(\frac{1}{2}v) f_{\mathbf{V}_j(\mu,\nu)}(v)}{M_{\mathbf{V}_j(\mu,\nu)}(\frac{1}{2})}. \qquad (256)$$

More formally, $\widehat{\mathbf{V}}_j(\mu,\nu)$ is defined as

$$\widehat{\mathbf{V}}_j(\mu,\nu)=\begin{cases}-\log\frac{(1-\mu)\nu}{(1-\nu)\mu} & \text{w.p. } \frac{\sqrt{(1-\mu)(1-\nu)\mu\nu}}{M_{\mathbf{V}_j}\left(\frac{1}{2}\right)},\\ 0 & \text{w.p. } \frac{((1-\mu)(1-\nu)+\mu\nu)}{M_{\mathbf{V}_j}\left(\frac{1}{2}\right)},\\ \log\frac{(1-\mu)\nu}{(1-\nu)\mu} & \text{w.p } \frac{\sqrt{(1-\mu)(1-\nu)\mu\nu}}{M_{\mathbf{V}_j}\left(\frac{1}{2}\right)}.\end{cases} \tag{257}$$

Note that $\widehat{\mathbf{V}}_j(\mu,\nu)$ is a random variable that takes only three values, and hence its mean and variance can be easily evaluated as

$$\mathbb{E}\left[\widehat{\mathbf{V}}_j(\mu,\nu)\right]=0, \tag{258}$$

$$\begin{aligned}\mathsf{Var}\left[\widehat{\mathbf{V}}_j(\mu,\nu)\right]&=\frac{2\left(\log\frac{(1-\mu)\nu}{(1-\nu)\mu}\right)^2\sqrt{(1-\mu)(1-\nu)\mu\nu}}{\left(\sqrt{\mu\nu}+\sqrt{(1-\mu)(1-\nu)}\right)^2}\\&=O\left(\sqrt{\mu\nu}\right).\end{aligned} \tag{259}$$

Let

$$\begin{aligned}&\{\mathbf{U}_i(p,\theta,q): i\in\mathcal{P}_{\rm d}\},\\&\{\mathbf{V}_j(\widetilde{\beta},\widetilde{\alpha}): j\in\mathcal{P}_{\widetilde{\beta}\to\widetilde{\alpha}}\},\\&\{\mathbf{V}_k(\widetilde{\gamma},\widetilde{\alpha}): k\in\mathcal{P}_{\widetilde{\gamma}\to\widetilde{\alpha}}\},\\&\{\mathbf{V}_\ell(\widetilde{\gamma},\widetilde{\beta}): \ell\in\mathcal{P}_{\widetilde{\gamma}\to\widetilde{\beta}}\}\end{aligned}$$

be sets of independent and identically distributed random variables defined as per (122) and (130). Note that the sets $\mathcal{P}_{\rm d}$, $\mathcal{P}_{\widetilde{\beta}\to\widetilde{\alpha}}$, $\mathcal{P}_{\widetilde{\gamma}\to\widetilde{\alpha}}$ and $\mathcal{P}_{\widetilde{\gamma}\to\widetilde{\beta}}$ (defined by (29), (30), (32) and (34), respectively) are disjoint sets. Similarly, let

$$\begin{aligned}&\{\widehat{\mathbf{U}}_i(p,\theta,q): i\in\mathcal{P}_{\rm d}\},\\&\{\widehat{\mathbf{V}}_j(\widetilde{\beta},\widetilde{\alpha}): j\in\mathcal{P}_{\widetilde{\beta}\to\widetilde{\alpha}}\},\\&\{\widehat{\mathbf{V}}_k(\widetilde{\gamma},\widetilde{\alpha}): k\in\mathcal{P}_{\widetilde{\gamma}\to\widetilde{\alpha}}\},\\&\{\widehat{\mathbf{V}}_\ell(\widetilde{\gamma},\widetilde{\beta}): \ell\in\mathcal{P}_{\widetilde{\gamma}\to\widetilde{\beta}}\}\end{aligned}$$

be sets of independent and identically distributed random variables defined as per (252) and (256).

For $\{\mathcal{P}_{\mu\to\nu}:|\mathcal{P}_{\mu\to\nu}|=|\mathcal{P}_{\nu\to\mu}|,\ \mu,\nu\in\{\widetilde{\alpha},\widetilde{\beta},\widetilde{\gamma}\},\ \mu\neq\nu\}$, we express the random variable of interest $\mathbf{B}$ from (42) as

$$\begin{aligned}\mathbf{B}=&\sum_{i\in\mathcal{P}_{\rm d}}\log\left((q-1)\frac{1-\theta}{\theta}\right)\mathsf{B}_i^{(p)}\left[\left(1+\mathsf{B}_i^{\left(\frac{1}{q-1}\right)}\right)\mathsf{B}_i^{(\theta)}-1\right]\\&+\sum_{j\in\mathcal{P}_{\widetilde{\beta}\to\widetilde{\alpha}}}\log\left(\frac{(1-\widetilde{\beta})\widetilde{\alpha}}{(1-\widetilde{\alpha})\widetilde{\beta}}\right)\left(\mathsf{B}_j^{(\widetilde{\beta})}-\mathsf{B}_j^{(\widetilde{\alpha})}\right)\\&+\sum_{k\in\mathcal{P}_{\widetilde{\gamma}\to\widetilde{\alpha}}}\log\left(\frac{(1-\widetilde{\gamma})\widetilde{\alpha}}{(1-\widetilde{\alpha})\widetilde{\gamma}}\right)\left(\mathsf{B}_k^{(\widetilde{\gamma})}-\mathsf{B}_k^{(\widetilde{\alpha})}\right)\\&+\sum_{\ell\in\mathcal{P}_{\widetilde{\gamma}\to\widetilde{\beta}}}\log\left(\frac{(1-\widetilde{\gamma})\widetilde{\beta}}{(1-\widetilde{\beta})\widetilde{\gamma}}\right)\left(\mathsf{B}_\ell^{(\widetilde{\gamma})}-\mathsf{B}_\ell^{(\widetilde{\beta})}\right),\\=&\sum_{i\in\mathcal{P}_{\rm d}}\mathbf{U}_i(p,\theta,q)+\sum_{j\in\mathcal{P}_{\widetilde{\beta}\to\widetilde{\alpha}}}\mathbf{V}_j(\widetilde{\beta},\widetilde{\alpha})\\&+\sum_{k\in\mathcal{P}_{\widetilde{\gamma}\to\widetilde{\alpha}}}\mathbf{V}_k(\widetilde{\gamma},\widetilde{\alpha})+\sum_{\ell\in\mathcal{P}_{\widetilde{\gamma}\to\widetilde{\beta}}}\mathbf{V}_\ell(\widetilde{\gamma},\widetilde{\beta}).\end{aligned} \tag{260}$$

Following a similar proof technique used for the proof of [58, Lemma 5.2], the probability that $\mathbf{B}$ is non-negative can be lower bounded by (268), shown at the top of the next page, where

- the summation in (265), shown at the top of the next page, is over

$$\mathcal{R}(\xi)=\Bigg\{\{u_i\}_{i\in\mathcal{P}_{\rm d}},\{v_j\}_{j\in\mathcal{P}_{\widetilde{\beta}\to\widetilde{\alpha}}},\{v_k\}_{j\in\mathcal{P}_{\widetilde{\gamma}\to\widetilde{\alpha}}},\{v_\ell\}_{j\in\mathcal{P}_{\widetilde{\gamma}\to\widetilde{\beta}}}:\\ 0\leq\left(\sum_{i\in\mathcal{P}_{\rm d}}u_i+\sum_{j\in\mathcal{P}_{\widetilde{\beta}\to\widetilde{\alpha}}}v_j+\sum_{k\in\mathcal{P}_{\widetilde{\gamma}\to\widetilde{\alpha}}}v_k+\sum_{\ell\in\mathcal{P}_{\widetilde{\gamma}\to\widetilde{\beta}}}v_\ell\right)<\xi\Bigg\};$$

  moreover, (265) follows from the independence of the random variables $\{\mathbf{U}_i: i\in\mathcal{P}_{\rm d}\}$, $\{\mathbf{V}_j: j\in\mathcal{P}_{\widetilde{\beta}\to\widetilde{\alpha}}\}$, $\{\mathbf{V}_k: k\in\mathcal{P}_{\widetilde{\gamma}\to\widetilde{\alpha}}\}$ and $\{\mathbf{V}_\ell:\ell\in\mathcal{P}_{\widetilde{\gamma}\to\widetilde{\beta}}\}$.
- in (266), shown at the top of the next page, holds since

$$\exp\left[\frac{1}{2}\left(\sum_{i\in\mathcal{P}_{\rm d}}u_i+\sum_{j\in\mathcal{P}_{\widetilde{\beta}\to\widetilde{\alpha}}}v_j+\sum_{k\in\mathcal{P}_{\widetilde{\gamma}\to\widetilde{\alpha}}}v_k+\sum_{\ell\in\mathcal{P}_{\widetilde{\gamma}\to\widetilde{\beta}}}v_\ell\right)\right]<\exp\left(\frac{\xi}{2}\right);$$

- in (267), shown at the top of the next page, follows from (252) and (256); and
- in (268), is a consequence of (129) and (134), and the fact that random variables $\{\widehat{\mathbf{U}}_i: i\in\mathcal{P}_{\rm d}\}$, $\{\widehat{\mathbf{V}}_j: j\in\mathcal{P}_{\widetilde{\beta}\to\widetilde{\alpha}}\}$, $\{\widehat{\mathbf{V}}_k: k\in\mathcal{P}_{\widetilde{\gamma}\to\widetilde{\alpha}}\}$ and $\{\widehat{\mathbf{V}}_\ell:\ell\in\mathcal{P}_{\widetilde{\gamma}\to\widetilde{\beta}}\}$ are independent.

It should be noted that (268) holds for any value of $\xi$. In particular, we choose $\xi_n$ that satisfies the following two conditions:

$$\lim_{n\to\infty}\frac{\xi_n}{|\mathcal{P}_{\rm d}|I_r+\frac{\left(|\mathcal{P}_{\widetilde{\beta}\to\widetilde{\alpha}}|I_{\alpha,\beta}+|\mathcal{P}_{\widetilde{\gamma}\to\widetilde{\alpha}}|I_{\alpha,\gamma}+|\mathcal{P}_{\widetilde{\gamma}\to\widetilde{\beta}}|I_{\beta,\gamma}\right)\log n}{n}}=0, \tag{261}$$

$$\lim_{n\to\infty}\frac{|\mathcal{P}_{\rm d}|p+|\mathcal{P}_{\widetilde{\beta}\to\widetilde{\alpha}}|\sqrt{\widetilde{\alpha}\widetilde{\beta}}+|\mathcal{P}_{\widetilde{\gamma}\to\widetilde{\alpha}}|\sqrt{\widetilde{\alpha}\widetilde{\gamma}}+|\mathcal{P}_{\widetilde{\gamma}\to\widetilde{\beta}}|\sqrt{\widetilde{\beta}\widetilde{\gamma}}}{\xi_n^2}=0, \tag{262}$$

for any $\mathcal{P}_{\rm d}$, $\mathcal{P}_{\widetilde{\beta}\to\widetilde{\alpha}}$, $\mathcal{P}_{\widetilde{\gamma}\to\widetilde{\alpha}}$ and $\mathcal{P}_{\widetilde{\gamma}\to\widetilde{\beta}}$ such that at least one of these sets is non-empty.[9] One instance of $\xi_n$ that satisfies both (261) and (262) is given by

$$\xi_n=\left(\max\left\{|\mathcal{P}_{\rm d}|,\left|\mathcal{P}_{\widetilde{\beta}\to\widetilde{\alpha}}\right|,|\mathcal{P}_{\widetilde{\gamma}\to\widetilde{\alpha}}|,\left|\mathcal{P}_{\widetilde{\gamma}\to\widetilde{\beta}}\right|\right\}\frac{\log n}{n}\right)^{\frac{2}{3}}. \tag{263}$$

Consequently, (261) implies that the exponent of the exponential term in (268) can be asymptotically approximated as

$$\begin{aligned}&-(1+o(1))\Bigg[|\mathcal{P}_{\rm d}|I_r+\frac{\log n}{n}\Big(|\mathcal{P}_{\widetilde{\beta}\to\widetilde{\alpha}}|I_{\alpha,\beta}\\&\qquad+|\mathcal{P}_{\widetilde{\gamma}\to\widetilde{\alpha}}|I_{\alpha,\gamma}+|\mathcal{P}_{\widetilde{\gamma}\to\widetilde{\beta}}|I_{\beta,\gamma}\Big)\Bigg]-\frac{1}{2}\xi_n\\&\simeq-(1+o(1))\Bigg[|\mathcal{P}_{\rm d}|I_r+|\mathcal{P}_{\widetilde{\beta}\to\widetilde{\alpha}}|I_{\alpha,\beta}\frac{\log n}{n}\\&\qquad+|\mathcal{P}_{\widetilde{\gamma}\to\widetilde{\alpha}}|I_{\alpha,\gamma}\frac{\log n}{n}+|\mathcal{P}_{\widetilde{\gamma}\to\widetilde{\beta}}|I_{\beta,\gamma}\frac{\log n}{n}\Bigg].\end{aligned} \tag{264}$$

[9] If the sets $\mathcal{P}_{\rm d}$, $\mathcal{P}_{\widetilde{\beta}\to\widetilde{\alpha}}$, $\mathcal{P}_{\widetilde{\gamma}\to\widetilde{\alpha}}$ and $\mathcal{P}_{\widetilde{\gamma}\to\widetilde{\beta}}$ are all empty, then (60) is trivially true.

$$
\begin{aligned}
&\mathbb{P}\left[\mathbf{B} \geq 0\right] \\
&= \mathbb{P}\left[\sum_{i\in\mathcal{P}_{\mathrm{d}}} \mathbf{U}_i(p,\theta,q) + \sum_{j\in\mathcal{P}_{\widetilde{\beta}\to\widetilde{\alpha}}} \mathbf{V}_j(\widetilde{\beta},\widetilde{\alpha}) + \sum_{k\in\mathcal{P}_{\widetilde{\gamma}\to\widetilde{\alpha}}} \mathbf{V}_k(\widetilde{\gamma},\widetilde{\alpha}) + \sum_{\ell\in\mathcal{P}_{\widetilde{\gamma}\to\widetilde{\beta}}} \mathbf{V}_\ell(\widetilde{\gamma},\widetilde{\beta}) \geq 0\right] \\
&\geq \mathbb{P}\left[0 \leq \sum_{i\in\mathcal{P}_{\mathrm{d}}} \mathbf{U}_i(p,\theta,q) + \sum_{j\in\mathcal{P}_{\widetilde{\beta}\to\widetilde{\alpha}}} \mathbf{V}_j(\widetilde{\beta},\widetilde{\alpha}) + \sum_{k\in\mathcal{P}_{\widetilde{\gamma}\to\widetilde{\alpha}}} \mathbf{V}_k(\widetilde{\gamma},\widetilde{\alpha}) + \sum_{\ell\in\mathcal{P}_{\widetilde{\gamma}\to\widetilde{\beta}}} \mathbf{V}_\ell(\widetilde{\gamma},\widetilde{\beta}) < \xi\right] \\
&= \sum_{\mathcal{R}(\xi)} \left[\prod_{i\in\mathcal{P}_{\mathrm{d}}} f_{\mathbf{U}_i(p,\theta,q)}(u_i) \prod_{j\in\mathcal{P}_{\widetilde{\beta}\to\widetilde{\alpha}}} f_{\mathbf{V}_j(\widetilde{\beta},\widetilde{\alpha})}(v_j) \prod_{k\in\mathcal{P}_{\widetilde{\gamma}\to\widetilde{\alpha}}} f_{\mathbf{V}_k(\widetilde{\gamma},\widetilde{\alpha})}(v_k) \prod_{\ell\in\mathcal{P}_{\widetilde{\gamma}\to\widetilde{\beta}}} f_{\mathbf{V}_\ell(\widetilde{\gamma},\widetilde{\beta})}(v_\ell)\right] \qquad (265) \\
&\geq \frac{\left(M_{\mathbf{U}_i(p,\theta,q)}\left(\frac{1}{2}\right)\right)^{|\mathcal{P}_{\mathrm{d}}|}\left(M_{\mathbf{V}_j(\widetilde{\beta},\widetilde{\alpha})}\left(\frac{1}{2}\right)\right)^{|\mathcal{P}_{\widetilde{\beta}\to\widetilde{\alpha}}|}\left(M_{\mathbf{V}_k(\widetilde{\gamma},\widetilde{\alpha})}\left(\frac{1}{2}\right)\right)^{|\mathcal{P}_{\widetilde{\gamma}\to\widetilde{\alpha}}|}\left(M_{\mathbf{V}_\ell(\widetilde{\gamma},\widetilde{\beta})}\left(\frac{1}{2}\right)\right)^{|\mathcal{P}_{\widetilde{\gamma}\to\widetilde{\beta}}|}}{\exp\left(\frac{1}{2}\xi\right)} \\
&\quad\times \sum_{\mathcal{R}(\xi)} \left[\prod_{i\in\mathcal{P}_{\mathrm{d}}} \frac{\exp\left(\frac{1}{2}u_i\right) f_{\mathbf{U}_i(p,\theta,q)}(u_i)}{M_{\mathbf{U}_i(p,\theta)}\left(\frac{1}{2}\right)} \prod_{j\in\mathcal{P}_{\widetilde{\beta}\to\widetilde{\alpha}}} \frac{\exp\left(\frac{1}{2}v_j\right) f_{\mathbf{V}_j(\widetilde{\beta},\widetilde{\alpha})}(v_j)}{M_{\mathbf{V}_j(\widetilde{\beta},\widetilde{\alpha})}\left(\frac{1}{2}\right)}\right. \\
&\quad\left.\times \prod_{k\in\mathcal{P}_{\widetilde{\gamma}\to\widetilde{\alpha}}} \frac{\exp\left(\frac{1}{2}v_k\right) f_{\mathbf{V}_k(\widetilde{\gamma},\widetilde{\alpha})}(v_k)}{M_{\mathbf{V}_k(\widetilde{\gamma},\widetilde{\alpha})}\left(\frac{1}{2}\right)} \prod_{\ell\in\mathcal{P}_{\widetilde{\gamma}\to\widetilde{\beta}}} \frac{\exp\left(\frac{1}{2}v_\ell\right) f_{\mathbf{V}_\ell(\widetilde{\gamma},\widetilde{\beta})}(v_\ell)}{M_{\mathbf{V}_\ell(\widetilde{\gamma},\widetilde{\beta})}\left(\frac{1}{2}\right)}\right] \qquad (266) \\
&= \exp\left[|\mathcal{P}_{\mathrm{d}}| \log M_{\mathbf{U}_i(p,\theta,q)}\left(\frac{1}{2}\right) + |\mathcal{P}_{\widetilde{\beta}\to\widetilde{\alpha}}| \log M_{\mathbf{V}_j(\widetilde{\beta},\widetilde{\alpha})}\left(\frac{1}{2}\right)\right. \\
&\quad\left. + |\mathcal{P}_{\widetilde{\gamma}\to\widetilde{\alpha}}| \log M_{\mathbf{V}_k(\widetilde{\gamma},\widetilde{\alpha})}\left(\frac{1}{2}\right) + |\mathcal{P}_{\widetilde{\gamma}\to\widetilde{\beta}}| \log M_{\mathbf{V}_\ell(\widetilde{\gamma},\widetilde{\beta})}\left(\frac{1}{2}\right) - \frac{1}{2}\xi\right] \\
&\quad\times \sum_{\mathcal{R}(\xi)} \left[\prod_{i\in\mathcal{P}_{\mathrm{d}}} f_{\widehat{\mathbf{U}}_i(p,\theta,q)}(u_i) \prod_{j\in\mathcal{P}_{\widetilde{\beta}\to\widetilde{\alpha}}} f_{\widehat{\mathbf{V}}_j(\widetilde{\beta},\widetilde{\alpha})}(v_j) \prod_{k\in\mathcal{P}_{\widetilde{\gamma}\to\widetilde{\alpha}}} f_{\widehat{\mathbf{V}}_k(\widetilde{\gamma},\widetilde{\alpha})}(v_k) \prod_{\ell\in\mathcal{P}_{\widetilde{\gamma}\to\widetilde{\beta}}} f_{\widehat{\mathbf{V}}_\ell(\widetilde{\gamma},\widetilde{\beta})}(v_\ell)\right] \qquad (267) \\
&= \exp\left[-(1+o(1))\left(|\mathcal{P}_{\mathrm{d}}| I_r + |\mathcal{P}_{\widetilde{\beta}\to\widetilde{\alpha}}| I_{\alpha,\beta}\frac{\log n}{n} + |\mathcal{P}_{\widetilde{\gamma}\to\widetilde{\alpha}}| I_{\alpha,\gamma}\frac{\log n}{n} + |\mathcal{P}_{\widetilde{\gamma}\to\widetilde{\beta}}| I_{\beta,\gamma}\frac{\log n}{n}\right) - \frac{1}{2}\xi\right] \\
&\quad\times \mathbb{P}\left[0 \leq \sum_{i\in\mathcal{P}_{\mathrm{d}}} \widehat{\mathbf{U}}_i(p,\theta,q) + \sum_{j\in\mathcal{P}_{\widetilde{\beta}\to\widetilde{\alpha}}} \widehat{\mathbf{V}}_j(\widetilde{\beta},\widetilde{\alpha}) + \sum_{k\in\mathcal{P}_{\widetilde{\gamma}\to\widetilde{\alpha}}} \widehat{\mathbf{V}}_k(\widetilde{\gamma},\widetilde{\alpha}) + \sum_{\ell\in\mathcal{P}_{\widetilde{\gamma}\to\widetilde{\beta}}} \widehat{\mathbf{V}}_\ell(\widetilde{\gamma},\widetilde{\beta}) < \xi\right]. \qquad (268)
\end{aligned}
$$

Furthermore, the probability term in (268) can be lower bounded as

$$
\begin{aligned}
&\mathbb{P}\left[0 \leq \sum_{i\in\mathcal{P}_{\mathrm{d}}} \widehat{\mathbf{U}}_i(p,\theta,q) + \sum_{j\in\mathcal{P}_{\widetilde{\beta}\to\widetilde{\alpha}}} \widehat{\mathbf{V}}_j(\widetilde{\beta},\widetilde{\alpha})\right. \\
&\qquad\left. + \sum_{k\in\mathcal{P}_{\widetilde{\gamma}\to\widetilde{\alpha}}} \widehat{\mathbf{V}}_k(\widetilde{\gamma},\widetilde{\alpha}) + \sum_{\ell\in\mathcal{P}_{\widetilde{\gamma}\to\widetilde{\beta}}} \widehat{\mathbf{V}}_\ell(\widetilde{\gamma},\widetilde{\beta}) < \xi_n\right] \\
&\geq \frac{1}{2} - \mathbb{P}\left[\sum_{i\in\mathcal{P}_{\mathrm{d}}} \widehat{\mathbf{U}}_i(p,\theta,q) + \sum_{j\in\mathcal{P}_{\widetilde{\beta}\to\widetilde{\alpha}}} \widehat{\mathbf{V}}_j(\widetilde{\beta},\widetilde{\alpha})\right. \\
&\qquad\left. + \sum_{k\in\mathcal{P}_{\widetilde{\gamma}\to\widetilde{\alpha}}} \widehat{\mathbf{V}}_k(\widetilde{\gamma},\widetilde{\alpha}) + \sum_{\ell\in\mathcal{P}_{\widetilde{\gamma}\to\widetilde{\beta}}} \widehat{\mathbf{V}}_\ell(\widetilde{\gamma},\widetilde{\beta}) \geq \xi_n\right] \qquad (269) \\
&\geq \frac{1}{2} - \frac{1}{\xi_n^2}\left(|\mathcal{P}_{\mathrm{d}}| \mathsf{Var}\left[\widehat{\mathbf{U}}_i(p,\theta)\right] + |\mathcal{P}_{\widetilde{\beta}\to\widetilde{\alpha}}| \mathsf{Var}\left[\widehat{\mathbf{V}}_j(\widetilde{\beta},\widetilde{\alpha})\right]\right. \\
&\qquad\left. + |\mathcal{P}_{\widetilde{\gamma}\to\widetilde{\alpha}}| \mathsf{Var}\left[\widehat{\mathbf{V}}_k(\widetilde{\gamma},\widetilde{\alpha})\right] + |\mathcal{P}_{\widetilde{\gamma}\to\widetilde{\beta}}| \mathsf{Var}\left[\widehat{\mathbf{V}}_\ell(\widetilde{\gamma},\widetilde{\beta})\right]\right) \qquad (270) \\
&= \frac{1}{2} - \frac{1}{\xi_n^2}\left(|\mathcal{P}_{\mathrm{d}}| O(p) + |\mathcal{P}_{\widetilde{\beta}\to\widetilde{\alpha}}| O(\sqrt{\widetilde{\alpha}\widetilde{\beta}})\right. \\
&\qquad\left. + |\mathcal{P}_{\widetilde{\gamma}\to\widetilde{\alpha}}| O(\sqrt{\widetilde{\alpha}\widetilde{\gamma}}) + |\mathcal{P}_{\widetilde{\gamma}\to\widetilde{\beta}}| O(\sqrt{\widetilde{\beta}\widetilde{\gamma}})\right) \qquad (271) \\
&= \frac{1}{2} - o(1) > \frac{1}{4}, \qquad (272)
\end{aligned}
$$

where (269) is due to the symmetry of the random variables $\widehat{\mathbf{U}}_i$, $\widehat{\mathbf{V}}_j$, $\widehat{\mathbf{V}}_k$ and $\widehat{\mathbf{V}}_\ell$; (270) follows from Chebyshev's inequality; (271) holds by substituting the variances with (255) and (259); and finally (272) is a consequence of (262). Plugging (264) and (272) into (268), we obtain

$$
\begin{aligned}
&\mathbb{P}\left[B\left(\mathcal{P}_{\mathrm{d}}, \mathcal{P}_{\widetilde{\beta}\to\widetilde{\alpha}}, \mathcal{P}_{\widetilde{\gamma}\to\widetilde{\alpha}}, \mathcal{P}_{\widetilde{\gamma}\to\widetilde{\beta}}\right) \geq 0\right] \\
&\geq \frac{1}{4}\exp\left[-(1+o(1))\left(|\mathcal{P}_{\mathrm{d}}| I_r\right.\right. \\
&\qquad\left.\left. + \frac{\log n}{n}\left(|\mathcal{P}_{\widetilde{\beta}\to\widetilde{\alpha}}| I_{\alpha,\beta} + |\mathcal{P}_{\widetilde{\gamma}\to\widetilde{\alpha}}| I_{\alpha,\gamma} + |\mathcal{P}_{\widetilde{\gamma}\to\widetilde{\beta}}| I_{\beta,\gamma}\right)\right)\right],
\end{aligned}
$$

which is the desired bound in (60). This concludes the proof of Lemma 8. ■

## APPENDIX I
## PROOF OF LEMMA 9

The proof hinges on the alteration method [59]. We present a construction to show the existence of subsets $\widetilde{Z}_0(x,i)$ and $\widetilde{Z}_0(y,j)$ with the desired property. Define $r := \frac{n}{\log^3 n}$. We start by sampling two random subsets $\overline{Z}_0(x,i) \subset Z_0(x,i)$ and $\overline{Z}_0(y,j) \subset Z_0(y,j)$ with cardinalities $|\overline{Z}_0(x,i)| = 2r$ and $|\overline{Z}_0(y,j)| = 2r$, respectively. Then, we prune these sets to obtain the desired edge-free subsets. To this end, for any pair of nodes $a, b \in \overline{Z}_0(x,i) \cup \overline{Z}_0(y,j)$, we remove both $a$ and $b$ from $\overline{Z}_0(x,i) \cup \overline{Z}_0(y,j)$ if $(a,b) \in \mathcal{E}$. We continue this process until the remaining set of nodes is edge-free. Let $\mathcal{P}$ be the set of nodes that we remove from $\overline{Z}_0(x,i) \cup \overline{Z}_0(y,j)$ throughout the pruning process. The expected value of $|\mathcal{P}|$ is upper bounded by

$$\begin{aligned}
\mathbb{E}[|\mathcal{P}|] &\leq 2\mathbb{E}\left[\sum_{a,b\,\in\,\overline{Z}_0(x,i)\cup\overline{Z}_0(y,j)} \mathbb{1}\left[(a,b)\in\mathcal{E}\right]\right] \\
&= 2\left(\sum_{a,b\in\overline{Z}_0(x,i)} \mathbb{E}[\mathbb{1}\left[(a,b)\in\mathcal{E}\right]] + \sum_{a,b\in\overline{Z}_0(y,j)} \mathbb{E}[\mathbb{1}\left[(a,b)\in\mathcal{E}\right]]\right. \\
&\qquad \left. + \sum_{a\in\overline{Z}_0(x,i)} \sum_{b\in\overline{Z}_0(y,j)} \mathbb{E}[\mathbb{1}\left[(a,b)\in\mathcal{E}\right]]\right) \\
&= 2\left(\sum_{a,b\in\overline{Z}_0(x,i)} \widetilde{\alpha} + \sum_{a,b\in\overline{Z}_0(y,j)} \widetilde{\alpha} + \sum_{a\in\overline{Z}_0(x,i)} \sum_{b\in\overline{Z}_0(y,j)} \widetilde{\beta}\right) \\
&= 2\left[\binom{2r}{2}\widetilde{\alpha} + \binom{2r}{2}\widetilde{\alpha} + (2r)^2\widetilde{\beta}\right] \leq 16\, r^2\widetilde{\alpha}, \qquad (273)
\end{aligned}$$

where the last inequality holds since $\widetilde{\beta} \leq \widetilde{\alpha}$. Using Markov's inequality for the non-negative random variable $|\mathcal{P}|$, we obtain

$$\mathbb{P}\left[|\mathcal{P}| \geq r\right] \leq \frac{\mathbb{E}\left[N\right]}{r} \leq \frac{16n}{\log^3 n}\widetilde{\alpha} = \Theta\left(\frac{n}{\log^3 n} \times \frac{\log n}{n}\right) = o(1). \qquad (274)$$

Therefore, the number of remaining nodes (after pruning) satisfies

$$\begin{aligned}
&\mathbb{P}\left[|\overline{Z}_0(x,i) \cup \overline{Z}_0(y,j) \setminus \mathcal{P}| > 3r\right] \\
&\qquad = \mathbb{P}\left[|\mathcal{P}| < r\right] = 1 - \mathbb{P}\left[|\mathcal{P}| \geq r\right] = 1 - o(1).
\end{aligned}$$

Hence, $\overline{Z}_0(x,i) \setminus \mathcal{P}$ and $\overline{Z}_0(y,j) \setminus \mathcal{P}$ both have at least $3r$ elements. This, together with $|\overline{Z}_0(x,i)| = |\overline{Z}_0(y,j)| = 2r$, implies that $\overline{Z}_0(x,i) \setminus \mathcal{P}$ and $\overline{Z}_0(y,j) \setminus \mathcal{P}$ each have at least $r$ elements. Therefore, we can choose $r$ elements from each of $\overline{Z}_0(x,i) \setminus \mathcal{P}$ and $\overline{Z}_0(y,j) \setminus \mathcal{P}$ to form the desired sets $\widetilde{Z}_0(x,i)$ and $\widetilde{Z}_0(y,j)$, respectively. This completes the proof of Lemma 9. ■

## REFERENCES


[1] Y. Koren, R. Bell, and C. Volinsky, "Matrix factorization techniques for recommender systems," *Computer*, vol. 42, no. 8, pp. 30–37, Aug. 2009.

[2] M. McPherson, L. Smith-Lovin, and J. M. Cook, "Birds of a feather: Homophily in social networks," *Annu. Rev. Sociol.*, vol. 27, no. 1, pp. 415–444, Aug. 2001.

[3] J. Tang, X. Hu, and H. Liu, "Social recommendation: A review," *Social Netw. Anal. Mining*, vol. 3, no. 4, pp. 1113–1133, 2013.

[4] D. Cai, X. He, J. Han, and T. S. Huang, "Graph regularized nonnegative matrix factorization for data representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 8, pp. 1548–1560, Aug. 2011.

[5] M. Jamali and M. Ester, "A matrix factorization technique with trust propagation for recommendation in social networks," in *Proc. 4th ACM Conf. Recommender Syst.*, 2010, pp. 135–142.

[6] W.-J. Li and D.-Y. Yeung, "Relation regularized matrix factorization," in *Proc. Int. Joint Conf. Artif. Intell. (IJCAI)*, 2009, pp. 1126–1131.

[7] H. Ma, D. Zhou, C. Liu, M. R. Lyu, and I. King, "Recommender systems with social regularization," in *Proc. 4th ACM Int. Conf. Web Search Data Mining*, Feb. 2011, pp. 287–296.

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

[9] H. Ma, H. Yang, M. R. Lyu, and I. King, "SoRec: Social recommendation using probabilistic matrix factorization," in *Proc. 17th ACM Conf. Inf. Knowl. Manage.*, 2008, pp. 931–940.

[10] H. Ma, I. King, and M. R. Lyu, "Learning to recommend with social trust ensemble," in *Proc. 32nd Int. ACM SIGIR Conf. Res. Develop. Inf. Retr.*, Jul. 2009, pp. 203–210.

[11] 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 *Proc. AAAI Conf. Artif. Intell.*, 2015, pp. 123–129.

[12] H. Zhao, Q. Yao, J. T. Kwok, and D. L. Lee, "Collaborative filtering with social local models," in *Proc. IEEE Int. Conf. Data Mining (ICDM)*, Nov. 2017, pp. 645–654.

[13] S. Chouvardas, M. A. Abdullah, L. Claude, and M. Draief, "Robust online matrix completion on graphs," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Mar. 2017, pp. 4019–4023.

[14] P. Massa and P. Avesani, "Controversial users demand local trust metrics: An experimental study on $Epinions.com$ community," in *Proc. AAAI*, 2005, pp. 121–126.

[15] J. Golbeck and J. Hendler, "FilmTrust: Movie recommendations using trust in web-based social networks," in *Proc. IEEE Consum. Commun. Netw. Conf.*, vol. 1, Feb. 2006, pp. 282–286.

[16] M. Jamali and M. Ester, "TrustWalker: A random walk model for combining trust-based and item-based recommendation," in *Proc. 15th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Jun. 2009, pp. 397–406.

[17] M. Jamali and M. Ester, "Using a trust network to improve top-N recommendation," in *Proc. 3rd ACM Conf. Recommender Syst.*, Oct. 2009, pp. 181–188.

[18] X. Yang, Y. Guo, and Y. Liu, "Bayesian-inference-based recommendation in online social networks," *IEEE Trans. Parallel Distrib. Syst.*, vol. 24, no. 4, pp. 642–651, Apr. 2013.

[19] X. Yang, H. Steck, Y. Guo, and Y. Liu, "On top-k recommendation using social networks," in *Proc. 6th ACM Conf. Recommender Syst.*, Sep. 2012, pp. 67–74.

[20] F. Monti, M. M. Bronstein, and X. Bresson, "Geometric matrix completion with recurrent multi-graph neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 30, 2017, pp. 3697–3707.

[21] R. van den Berg, T. N. Kipf, and M. Welling, "Graph convolutional matrix completion," 2017, *arXiv:1706.02263*.

[22] K.-Y. Chiang, C.-J. Hsieh, and I. S. Dhillon, "Matrix completion with noisy side information," in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, 2015, pp. 3447–3455.

[23] N. Rao, H F. Yu, P K. Ravikumar, and I. S. Dhillon, "Collaborative filtering with graph information: Consistency and scalable methods," in *Proc. Adv. Neural Inf. Process. Syst.*, 2015, pp. 2107–2115.

[24] K. Ahn, K. Lee, H. Cha, and C. Suh, "Binary rating estimation with graph side information," in *Proc. Adv. Neural Inf. Process. Syst. (NeurIPS)*, 2018, pp. 4272–4283.

[25] J. Yoon, K. Lee, and C. Suh, "On the joint recovery of community structure and community features," in *Proc. 56th Annu. Allerton Conf. Commun., Control, Comput. (Allerton)*, Oct. 2018, pp. 686–694.

[26] Q. Zhang, V. Y. F. Tan, and C. Suh, "Community detection and matrix completion with social and item similarity graphs," *IEEE Trans. Signal Process.*, vol. 69, pp. 917–931, 2021.

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

[28] A. Elmahdy, J. Ahn, C. Suh, and S. Mohajer, "Matrix completion with hierarchical graph side information," in *Proc. Adv. Neural Inf. Process. Syst. (NeurIPS)*, vol. 33, 2020, pp. 9061–9074.

[29] A. Elmahdy, J. Ahn, S. Mohajer, and C. Suh, "The optimal sample complexity of matrix completion with hierarchical similarity graphs," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jun. 2022, pp. 2409–2414.

[30] L. T. Nguyen, J. Kim, and B. Shim, "Low-rank matrix completion: A contemporary survey," *IEEE Access*, vol. 7, pp. 94215–94237, 2019.
[31] E. J. Candès and B. Recht, "Exact matrix completion via convex optimization," *Found. Comput. Math.*, vol. 9, no. 6, pp. 717–772, Dec. 2009.
[32] R. H. Keshavan, A. Montanari, and S. Oh, "Matrix completion from a few entries," *IEEE Trans. Inf. Theory*, vol. 56, no. 6, pp. 2980–2998, Jun. 2010.
[33] E. J. Candes and T. Tao, "The power of convex relaxation: Near-optimal matrix completion," *IEEE Trans. Inf. Theory*, vol. 56, no. 5, pp. 2053–2080, May 2010.
[34] M. Fazel, "Matrix rank minimization with applications," Ph.D. thesis, Dept. Elect. Eng., Stanford Univ., Stanford, CA, USA, 2002.
[35] J.-F. Cai, E. J. Candès, and Z. Shen, "A singular value thresholding algorithm for matrix completion," *SIAM J. Optim.*, vol. 20, no. 4, pp. 1956–1982, 2010.
[36] M. Fornasier, H. Rauhut, and R. Ward, "Low-rank matrix recovery via iteratively reweighted least squares minimization," *SIAM J. Optim.*, vol. 21, no. 4, pp. 1614–1640, 2011.
[37] K. Mohan and M. Fazel, "Iterative reweighted algorithms for matrix rank minimization," *J. Mach. Learn. Res.*, vol. 11, pp. 3441–3473, Jan. 2012.
[38] K. Lee and Y. Bresler, "ADMiRA: Atomic decomposition for minimum rank approximation," *IEEE Trans. Inf. Theory*, vol. 56, no. 9, pp. 4402–4416, Sep. 2010.
[39] Z. Wang, M.-J. Lai, Z. Lu, W. Fan, H. Davulcu, and J. Ye, "Rank-one matrix pursuit for matrix completion," in *Proc. Int. Conf. Mach. Learn. (ICML)*, 2014, pp. 91–99.
[40] J. Tanner and K. Wei, "Low rank matrix completion by alternating steepest descent methods," *Appl. Comput. Harmon. Anal.*, vol. 40, no. 2, pp. 417–429, Mar. 2016.
[41] Z. Wen, W. Yin, and Y. Zhang, "Solving a low-rank factorization model for matrix completion by a nonlinear successive over-relaxation algorithm," *Math. Program. Comput.*, vol. 4, no. 4, pp. 333–361, Dec. 2012.
[42] W. Dai and O. Milenkovic, "SET: An algorithm for consistent matrix completion," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, Mar. 2010, pp. 3646–3649.
[43] B. Vandereycken, "Low-rank matrix completion by Riemannian optimization," *SIAM J. Optim.*, vol. 23, no. 2, pp. 1214–1236, Jan. 2013.
[44] Y. Hu, D. Zhang, J. Ye, X. Li, and X. He, "Fast and accurate matrix completion via truncated nuclear norm regularization," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 9, pp. 2117–2130, Sep. 2013.
[45] J.-Y. Gotoh, A. Takeda, and K. Tono, "DC formulations and algorithms for sparse optimization problems," *Math. Program.*, vol. 169, no. 1, pp. 141–176, May 2018.
[46] E. J. Candes and Y. Plan, "Matrix completion with noise," *Proc. IEEE*, vol. 98, no. 6, pp. 925–936, Jun. 2010.
[47] A. L. Traud, P. J. Mucha, and M. A. Porter, "Social structure of Facebook networks," *Phys. A*, vol. 391, no. 16, pp. 4165–4180, Aug. 2012.
[48] M. E. J. Newman, "The structure of scientific collaboration networks," *Proc. Nat. Acad. Sci. USA*, vol. 98, no. 2, pp. 404–409, Jan. 2001.
[49] E. Abbe and C. Sandon, "Community detection in general stochastic block models: Fundamental limits and efficient algorithms for recovery," in *Proc. IEEE 56th Annu. Symp. Found. Comput. Sci.*, Oct. 2015, pp. 670–688.
[50] C. Gao, Z. Ma, A. Y. Zhang, and H. H. Zhou, "Achieving optimal misclassification proportion in stochastic block models," *J. Mach. Learn. Res.*, vol. 18, pp. 1980–2024, Jan. 2017.
[51] V. Lyzinski, M. Tang, A. Athreya, Y. Park, and C. E. Priebe, "Community detection and classification in hierarchical stochastic blockmodels," *IEEE Trans. Netw. Sci. Eng.*, vol. 4, no. 1, pp. 13–26, Jan. 2017.
[52] V. Cohen-Addad, V. Kanade, F. Mallmann-Trenn, and C. Mathieu, "Hierarchical clustering: Objective functions and algorithms," *J. ACM*, vol. 66, no. 4, pp. 1–42, 2019.
[53] H. Ashtiani, S. Kushagra, and S. Ben-David, "Clustering with same-cluster queries," in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, 2016, pp. 3216–3224.
[54] A. Mazumdar and B. Saha, "Query complexity of clustering with side information," in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, 2017, pp. 4682–4693.
[55] E. Abbe, "Community detection and stochastic block models: Recent developments," *The J. Mach. Learn. Res. (JMLR)*, vol. 18, no. 1, pp. 6446–6531, 2017.
[56] E. Abbe, A. S. Bandeira, and G. Hall, "Exact recovery in the stochastic block model," *IEEE Trans. Inf. Theory*, vol. 62, no. 1, pp. 471–487, Jan. 2016.
[57] F. J. MacWilliams and N. J. A. Sloane, *The Theory Error Correcting Codes*. Amsterdam, The Netherlands: Elsevier, 1977, vol. 16.
[58] A. Y. Zhang and H. H. Zhou, "Minimax rates of community detection in stochastic block models," *Ann. Statist.*, vol. 44, no. 5, pp. 2252–2280, Oct. 2016.
[59] N. Alon and J. H. Spencer, *The Probabilistic Method*. Hoboken, NJ, USA: Wiley, 2004.

**Junhyung Ahn** (Member, IEEE) received the B.S. and M.S. degrees in electrical engineering from the Korea Advanced Institute of Science and Technology (KAIST) in 2019 and 2022, respectively. He is currently a Machine Learning Engineer at Dable, serving as a Technical Research Personnel for military service. His research interests include information theory and its applications in machine learning.

**Adel Elmahdy** (Graduate Student Member, IEEE) received the B.Sc. degree in electrical engineering from Alexandria University, Egypt, and the M.Sc. degree in wireless communications from Nile University, Egypt. He is currently pursuing the Ph.D. degree with the Department of Electrical and Computer Engineering, University of Minnesota, Minneapolis, MN, USA. He was a Research Intern at Microsoft Research, Amplitude Analytics, and Vectara. His research interests span a broad range of topics focusing largely on information theory and coding, and its applications in machine learning. He was a recipient of the Graduate Fellowship from Nile University and the ADC Fellowship from the University of Minnesota.

**Soheil Mohajer** (Member, IEEE) received the B.Sc. degree in electrical engineering from the Sharif University of Technology, Tehran, Iran, in 2004, and the M.Sc. and Ph.D. degrees in communication systems from École Polytechnique Fédérale de Lausanne (EPFL), Switzerland, in 2005 and 2010, respectively. He was affiliated as a Post-Doctoral Researcher with Princeton University from 2010 to 2011, and the University of California at Berkeley from 2011 to 2013. He is currently an Associate Professor of electrical and computer engineering with the University of Minnesota. His research interests include information theory and its applications in distributed storage systems, distributed optimization, and statistical machine learning. He received the NSF CAREER Award in 2018. He has served as an Editor for IEEE Transactions on Communications from 2019 to 2021.

**Changho Suh** (Fellow, IEEE) received the B.S. and M.S. degrees in electrical engineering from the Korea Advanced Institute of Science and Technology (KAIST), Daejeon, South Korea, in 2000 and 2002, respectively, and the Ph.D. degree in electrical engineering and computer sciences from UC Berkeley in 2011. From 2002 to 2006, he was with Samsung Electronics. From 2011 to 2012, he was a Post-Doctoral Associate with the Research Laboratory of Electronics, Massachusetts Institute of Technology (MIT). He is currently a Professor of electrical engineering at KAIST. He was a recipient of numerous awards in research and teaching, such as the 2022 Google Research Award, the 2021 James L. Massey Research and Teaching Award for Young Scholars from the IEEE Information Theory Society, the 2020 LINKGENESIS Best Teacher Award, the 2019 AFOSR Grant, the 2019 Google Education Grant, the 2018 IEIE/IEEE Joint Award, the 2015 IEIE Haedong Young Engineer Award, the 2015 Bell Labs Prize Finalist, the 2013 IEEE Communications Society Stephen O. Rice Prize, the 2011 David J. Sakrison Memorial Prize, the 2009 IEEE ISIT Best Student Paper Award, and the five Department Teaching Awards in 2013, 2019, 2020, 2021, and 2022. He is a Distinguished Lecturer of the IEEE Information Theory Society from 2020 to 2022, the General Chair of the Inaugural East Asian School of Information Theory 2021, and a member of the Young Korean Academy of Science and Technology. He is an Associate Editor of Machine Learning for IEEE Transactions on Information Theory, a Guest Editor for the IEEE Journal on Selected Areas in Information Theory, an Editor for IEEE Information Theory Newsletter, an Area Editor for *IEEE BITS the Information Theory Magazine*, the Area Chair of NeurIPS from 2021 to 2022, and a Senior Programme Committee of IJCAI from 2019 to 2021.