Community Detection and Matrix Completion with Two-Sided Graph Side-Information

Qiaosheng (Eric) Zhang†, Vincent Y. F. Tan†, and Changho Suh§

Abstract. We consider the problem of recovering communities of users and communities of items (such as movies) based on a partially observed rating matrix as well as side-information in the form of similarity graphs of the users and items. The user-to-user and item-to-item similarity graphs are generated according to the celebrated stochastic block model (SBM). We develop lower and upper bounds on the minimum expected number of observed ratings (also known as the sample complexity) needed for this recovery task. These bounds are functions of various parameters including the quality of the graph side-information which is manifested in the intra- and inter-cluster probabilities of the SBMs. We show that these bounds match for a wide range of parameters of interest, and match up to a constant factor of two for the remaining parameter regime. Our information-theoretic results quantify the benefits of the two-sided graph side-information for recovery, and further analysis reveals that the two pieces of graph side-information produce an interesting synergistic effect under certain scenarios. This means that if one observes only one of the two graphs, then the required sample complexity worsens to the case in which none of the graphs is observed. Thus both graphs are strictly needed to reduce the sample complexity.

Key words. Community detection, matrix completion, stochastic block model, graph side-information.

AMS subject classifications. 68R10, 68P30, 68Q87.

1. Introduction. Recommender systems aim to accurately predict users’ preferences and recommend appropriate items for users based on available data that is usually scant and/or of low quality. For example, Netflix’s movie recommender system relies heavily on the rating matrix that comprises users’ evaluations of movies, and various recommendation algorithms (such as collaborative filtering [7]) have been developed. However, merely adopting or exploiting the available ratings may not be sufficient for high-quality recommendations, since (i) the rating matrices are usually highly incomplete (e.g., about 99% ratings in Netflix’s rating dataset are missing [24, 13]), and (ii) the preferences of new users are always unavailable (i.e., the cold start problem). Meanwhile, it has been noticed that community information—either the communities of users (e.g., the friendships in Facebook) or the communities of movies (e.g., the categories/genres of movies in the Netflix database)—may effectively improve the quality of recommendations [15, 12, 4] and tackle the cold start problem [10], as users in the same community are more likely to share similar preferences, and movies in the same community are more likely to attract similar users.

While most of the attention has focused on the algorithmic developments of the graph-
aided recommender systems (see [23] for a review of social recommender systems) as well as their accompanying analyses, the fundamental limits of such problems are also worth exploring. Ahn et al. [4] considered the problem of recovering the binary rating matrix (which comprises users’ ratings to movies) based on a partially observed matrix and a user-to-user similarity graph. They characterized a sharp threshold on the minimum expected number of observed ratings needed for recovery as a function of the “quality” of the user-to-user graph and the amount of noise in the measurements, and also quantified the gains due to the graph side-information. In practice, the item-to-item similarity graph is sometimes also available; hence one may ask whether the additional item-to-item graph provides strictly more benefits for recovery, and whether observing two pieces of graph side-information simultaneously has synergistic effects. This work precisely addresses the aforementioned questions by investigating, from an information-theoretic perspective, the roles and benefits of the two-sided graph side-information for this recovery problem.

We consider a concrete example of movie recommender systems with \( n \) users and \( m \) movies. For simplicity, we consider a simple setting wherein users’ ratings to movies are either 0 (dislike) or 1 (like). Users are partitioned into communities of men and women (of equal size), while movies are partitioned into communities of action movies and romance movies (of equal size). The assumptions on binary ratings and two equal-sized communities are mainly for ease of presentation, and extensions to general settings are certainly also possible. Typically, action movies attract more men and romance movies attract more women, but we also allow the existence of atypical action movies and romance movies. The nominal ratings from a certain community of users to a certain type of movies are pre-specified in Table 1.1 The \( n \times m \) binary rating matrix comprises \( n \) users’ ratings to all the \( m \) movies. A personalized rating of each individual rating is a perturbed version of the corresponding nominal rating (being flipped with probability less than \( 1/2 \)), modeling the different preferences of users in the same community to a certain movie.

Under this setting, three pieces of information are observed by the learner: (i) Entries in the rating matrix that are observed/sampled (independently) with a fixed sample probability, (ii) The user-to-user similarity graph that is generated according to the widely-adopted symmetric stochastic block model (SBM) [2, 9] with two equal-sized communities (which is also known as the planted bisection model), and (iii) The item-to-item similarity graph that is generated according to another independent symmetric SBM with two equal-sized communities. The task here is to exactly recover the communities of men and women, the communities of action movies and romance movies, as well as to uncover the atypical action and romance movies. It is worth highlighting that the existence of atypical movies makes our task strictly more difficult compared to that of recovering only the binary rating matrix, since movies that attract more men than women (resp. attract more women than men) may be considered as either typical action movies or atypical romance movies (resp. either typical romance movies or atypical action movies). Essentially, recovering only the binary rating matrix, to be detailed

---

1As an initial effort, we assume there is only one type of atypical movies that is completely different from typical ones (e.g., typical action movies attract more men while atypical action movies attract more women). For future work, one may extend to a more general setting wherein three types of atypical movies with different likabilities are considered (e.g., atypical action movies may attract more women than men, both men and women, or neither).
Table 1: Nominal ratings from a certain community of users to a certain type of movies

|                | Action movies | Romance movies |
|----------------|---------------|----------------|
|                | Typical | Atypical | Typical | Atypical |
| Men            | 1       | 0        | 0       | 1        |
| Women          | 0       | 1        | 1       | 0        |

in Remark 2.2 in Subsection 2.4, can be viewed as a by-product of our task.

**Main Contributions.** The main contributions of this paper are summarized as follows. We develop lower and upper bounds on the minimum sample complexity needed for recovery as a function of the “qualities” of the user-to-user similarity graph (i.e., the row graph) and the item-to-item similarity graph (i.e., the column graph). The “qualities” of the graphs can be quantified by the difference between the intra- and inter-cluster connection probabilities of the SBMs that govern them. These bounds match for a wide range of parameters of interest, and match up to a constant factor of two for the remaining parameter regime in which the bounds are loose. Our theoretical studies show that, from the viewpoint of the sample complexity, gains due to the two-side graph side-information appear for a wide range of parameters. More interestingly, we show that there exists a certain regime in which there is a synergistic effect of the two-sided graph side-information—simultaneously observing both graphs is helpful for recovery, while observing only one graph is equivalent to observing neither.

**Related Works.** This work is closely related to the community detection problem and the matrix completion problem. While there is a vast literature on these two topics (especially from the algorithmic and experimental perspectives), in the following discussions we mainly focus on theoretical works that provide provable guarantees, especially those that focus on fundamental limits (in the absence of complexity considerations).

The community detection problem, which aims to partition vertices into different communities (or clusters) based on the density of connections, has been well-studied from an information-theoretic perspective [2, 17, 3, 11, 8]. In particular, the sharp threshold has been established for exact recovery of communities. We refer readers to [1] for an excellent survey. Moreover, it has been shown that side-information (e.g., edge weights [6], node values [19, 21, 20, 25], similarity information between data points [16], etc.) is also helpful in recovering hidden communities. Roughly speaking, in our problem setting, we are required to recover the communities of users and movies as well as the rating information (in the form of a complete binary matrix) given realizations from two symmetric SBMs (the item-to-item and the user-to-user similarity graphs) together with a partially observed rating matrix. Also, we note that the task in [24] (joint recovery of rows and columns communities) is similar to ours, but therein, the graph information is not available.

The matrix completion problem focuses on the recovery of low-rank matrices from sparse observations, and has wide applications in recommender systems [22]. Unlike the standard setting in which the linear dependence of rows and columns of the low-rank matrix is unstructured, the graph side-information in recommender systems also imposes additional structures on the low-rank matrix to be completed. Such models and the usefulness of graph side-
information have also been investigated in the literature. For instance, the works of [4, 26] considered a specific binary matrix completion problem with the aid of one-sided graph side-information (the user-to-user similarity graph), while [12] and [18] studied matrix completion models whereby additional proximity information about both rows and columns is available. The task in this paper, as mentioned earlier and to be detailed in Subsection 2.4, is strictly more challenging than merely recovering a low-rank matrix; nonetheless, one can view the problem of recovering a low-rank matrix with two-sided graph side-information as a by-product of our task.

**Outline.** The rest of this paper is organized as follows. We formally describe our model in Section 2. In Section 3, we present the main results of this paper, and reveal the benefits of the two-sided graph side-information by thoroughly analyzing two examples. Section 4 and Section 5 respectively provide proofs of the lower and upper bounds presented in Section 3. Section 6 concludes this work and proposes several directions that are fertile avenues for future research.

2. Problem statement.

2.1. Notation. Random variables are denoted by upper-case letters (e.g., \(X\)), while their realizations are denoted by lower-case letters (e.g., \(x\)). Vectors are denoted by boldface letters (e.g., \(x\)), and sets are denoted by calligraphy letters (e.g., \(X\)). The Hamming distance between two vectors \(x\) and \(x'\) are denoted by \(d_H(x, x')\). For any integer \(a \geq 1\), \([a]\) represents the set of integers \(\{1, \ldots, a\}\). For any integers \(a, b\) such that \(a < b\), \([a : b]\) represents the set of integers \(\{a, a+1, \ldots, b\}\). For any event \(\mathcal{E}\), the conditional probability \(P(\cdot | \mathcal{E})\) is abbreviated as \(P_{\mathcal{E}}(\cdot)\).

Throughout this paper we use standard asymptotic notations [14, Ch. 3.1] to describe the limiting behaviour of functions/sequences.

2.2. Model. Consider \(n\) users and \(m\) movies, and we require\(^2\) \(m = \omega(\log n)\) and \(n = \omega(\log m)\) for technical reasons. The sets of men and women are respectively denoted by \(M\) and \(W\), where \(M, W \subset [n]\), \(|M| = |W| = n/2\), and \(M \cap W = \emptyset\). The sets of action and romance movies are respectively denoted by \(A\) and \(R\), where \(A, R \subset [m]\), \(|A| = |R| = m/2\), and \(A \cap R = \emptyset\). Meanwhile, there may also exist an unknown-sized subset of atypical action movies \(A_0 \subseteq A\) and an unknown-sized subset of atypical romance movies \(R_0 \subseteq R\). Recall that the nominal ratings from users to movies are stated in Table 1. This reflects our assumption that typical action movies and atypical romance movies attract more men than women, while typical romance movies and atypical action movies attract more women than men. We define \(v_{MA} \in \{0, 1\}^{m/2}\) as the length-\(m/2\) nominal rating vector from men to action movies, where the 1’s in \(v_{MA}\) correspond to men’s nominal ratings to typical action movies, and 0’s in \(v_{MA}\) correspond to men’s nominal ratings to atypical action movies. The other three nominal rating vectors \(v_{MR}\) (men to romance movies), \(v_{WA}\) (women to action movies), and \(v_{WR}\) (women to romance movies) can be defined and interpreted similarly. By assumption, \(d_H(v_{MA}, v_{WA}) = d_H(v_{MR}, v_{WR}) = m/2\).

Let \(A(i)\) and \(R(i)\) be the \(i\)-th smallest elements of \(A\) and \(R\), respectively. For each \(j \in A\) (resp. \(j \in R\)), we define \(i_j (i_j \in [m/2])\) as the index such that \(A(i_j) = j\) (resp. \(R(i_j) = j\)).

\(^2\)This requirement essentially excludes very “fat” and very “thin” matrices.
The length-\(m\) vectors \(v_M\) and \(v_W\), which denote men’s and women’s preferences to all the \(m\) movies, are respectively defined as

\[
\begin{align*}
  v_M(j) = \begin{cases} 
    v_{MA}(i_j), & \text{if } j \in A, \\
    v_{MR}(i_j), & \text{if } j \in R,
  \end{cases} \\
  v_W(j) = \begin{cases} 
    v_{WA}(i_j), & \text{if } j \in A, \\
    v_{WR}(i_j), & \text{if } j \in R.
  \end{cases}
\end{align*}
\]

Note that \(d_H(v_M, v_W) = m\). For instance, if \(m = 6, A = \{1, 3, 5\}, v_{MA} = [1, 1, 1, 0, 0, 0], v_{MR} = [0, 0, 0, \ldots, 0]\), the length-\(m\) vector \(v_M\) then equals \([1, 0, 1, 0, 1, 0]\).

The taste of each individual man or woman also differs from the nominal taste of the communities. For each individual man or woman, the rating vector to action movies is obtained by passing the nominal rating vectors \(v_{MA}\) or \(v_{WA}\) through a binary symmetric channel with crossover probability \(\theta_{A}\) (denoted by \(\text{BSC}(\theta_{A})\)), where \(\theta_{A} \in (0, 1]\) is the personalization parameter for action movies. Similarly, the rating vector to romance movies is obtained by passing the nominal rating vectors \(v_{MR}\) or \(v_{WR}\) through a \(\text{BSC}(\theta_{R})\), where \(\theta_{R} \in (0, 1]\) is the personalization parameter for romance movies. Note that the difference between \(\theta_{A}\) and \(\theta_{R}\) is an important statistic for distinguishing action and romance movies.

Let \(\xi_{M, W, A, R, v_{M}, v_{W}}\) be an aggregation of the parameters of interest, and we sometimes abbreviate \(\xi_{M, W, A, R, v_{M}, v_{W}}\) as \(\xi\) for notational convenience. The sets of men, women, action movies, and romance movies (associated with \(\xi\)) are respectively denoted by \(\xi_M, \xi_W, \xi_A, \) and \(\xi_R\), while the associated nominal vectors are denoted by \(\xi_{v_M}\) and \(\xi_{v_W}\). In order to avoid any indeterminacies, without loss of generality,\(^3\) we also assume that the majority of the first \(n/2\) users are men (i.e., \(|\xi_M \cap [n/2]| > n/4\)), and the majority of the first \(m/2\) movies are action movies (i.e., \(|\xi_A \cap [m/2]| > m/4\)). The parameter space \(\Xi\) is the collection of valid parameters \(\xi_{M, W, A, R, v_{M}, v_{W}}\), i.e.,

\[
\Xi \triangleq \left\{ \xi_{M, W, A, R, v_{M}, v_{W}} \mid \begin{array}{l}
\xi_M, \xi_W \subseteq [n], \ |\xi_M| = |\xi_W| = \frac{n}{2}, \ |\xi_M \cap [n/2]| > \frac{n}{4} \\
\xi_A, \xi_R \subseteq [m], \ |\xi_A| = |\xi_R| = \frac{m}{2}, \ |\xi_A \cap [m/2]| > \frac{m}{4} \\
|\xi_{v_M}, \xi_{v_W}| \in \{0,1\}^m, \ d_H(\xi_{v_M}, \xi_{v_W}) = m
\end{array} \right\}.
\]

For any \(\xi \in \Xi\), the corresponding \(n \times m\) non-personalized binary rating matrix \(B_{\xi}\) denotes users’ nominal ratings to all the movies, wherein the \(i\)-th row of \(B_{\xi}\) equals \(v_M\) if \(i \in \xi_M\), and equals \(v_W\) if \(i \in \xi_W\). The \(n \times m\) personalized binary rating matrix \(V_{\xi}\) denotes all the users’ actual ratings to all the movies. The \(i\)-th row of \(V_{\xi}\) represents the \(i\)-th user’s ratings to all the movies, whereas the \(j\)-th column of \(V_{\xi}\) represents all the users’ rating to the \(j\)-th movie. Effectively, the \(j\)-th column of \(V_{\xi}\) is obtained by passing the \(j\)-th column of \(B_{\xi}\) through a \(\text{BSC}(\theta_{A})\) if \(j \in \xi_A\), and a \(\text{BSC}(\theta_{R})\) if \(j \in \xi_R\).

2.3. Observations. For any \(\xi \in \Xi\), the learner observes three pieces of information.

1. The partially observed matrix \(V^{\Omega}\). For each \((i, j) \in [n] \times [m]\), the \((i, j)\)-th entry of \(V^{\Omega}\) is given by

\[
(V^{\Omega})_{ij} = \begin{cases} 
  (V_{\xi})_{ij}, & \text{with probability } p, \\
  \perp, & \text{with probability } 1 - p.
\end{cases}
\]

\(^3\)Without this assumption, for any \(\xi\), one can always find a \(\xi’\) with \(\xi’_{v_M} = \xi_M, \xi’_{v_W} = \xi_W, \xi’_{v_R} = \xi_A,\) and \(\xi’_{v_A} = \xi_R\) (i.e., simultaneously flipping the communities of users and movies) such that \(\xi\) and \(\xi’\) are statistically indistinguishable.
Figure 1: An illustration of $V^Ω$, $G_1$, and $G_2$ that are generated according to the model parameterized by $ξ$, where $ξ_M = \{1, 2, 3, 4\}$ (gray), $ξ_W = \{5, 6, 7, 8\}$ (orange), $ξ_A = \{1, 2, 3, 4, 5, 6\}$ (blue), $ξ_{A_0} = \{6\}$ (light blue), $ξ_R = \{7, 8, 9, 10, 11, 12\}$ (red), and $ξ_{R_0} = \{12\}$ (pink).

where $⊥$ denotes the erasure symbol, and $p$ is called the sample probability. The sample complexity then equals $nmp$, which corresponds to the expected number of observed entries.

2. The row graph $G_1$ with $n$ nodes corresponding to the $n$ users. For any pairs of nodes $i \neq j$, it is connected with probability $α_1 = \frac{a_1 \log n}{n}$ if both $i$ and $j$ are in the same community (either in $ξ_M$ or $ξ_W$), and is connected with probability $β_1 = \frac{b_1 \log n}{n}$ otherwise. The total number of edges in $G_1$ is denoted by $|E_1|$.

3. The column graph $G_2$ with $m$ nodes corresponding to the $m$ movies. For any pairs of nodes $i \neq j$, it is connected with probability $α_2 = \frac{a_2 \log m}{m}$ if both $i$ and $j$ are in the same community (either in $ξ_A$ or $ξ_R$), and is connected with probability $β_2 = \frac{b_2 \log m}{m}$ otherwise. The total number of edges in $G_2$ is denoted by $|E_2|$.

An example of the three pieces of information $V^Ω, G_1, G_2$ is illustrated in Figure 1. Note that $a_1, a_2, b_1, b_2$ are constants, while $α_1, β_1 = Θ(\frac{\log n}{n})$ and $α_2, β_2 = Θ(\frac{\log m}{m})$. We define $I_1 = (\sqrt{a_1} - \sqrt{b_1})^2$ as the “quality” of the row graph $G_1$, since a larger difference between $a_1$ and $b_1$ makes recovery easier. A well-known result [2] shows that exact recovery is possible if $I_1 > 2$ and impossible if $I_1 < 2$. Similarly, we define $I_2 = (\sqrt{a_2} - \sqrt{b_2})^2$ as the “quality” of the column graph $G_2$.

2.4. Objective. Based on the observations $V^Ω, G_1$, and $G_2$, the learner aims to use an estimator $φ = φ(V^Ω, G_1, G_2)$ to recover $ξ$, i.e., the communities of men and women ($ξ_M$ and $ξ_W$), the communities of action movies and romance movies ($ξ_A$ and $ξ_R$), and the corresponding nominal vectors $ξ_{VM}$ and $ξ_{VW}$. Note that recovering the nominal vectors $ξ_{VM}$ and $ξ_{VW}$ is equivalent to recovering the sets of atypical action movies $ξ_{A_0}$ and atypical romance movies $ξ_{R_0}$.
Definition 2.1 (Exact recovery). For any estimator $\phi$, the maximum probability of error is defined as
\[
P_{\text{err}}(\phi) \triangleq \max_{\xi \in \Xi} P_{\xi}(\phi(V, G_1, G_2) \neq \xi),
\]
where $P_{\xi}(\cdot)$ represents the probability of error when $V, G_1,$ and $G_2$ are generated according to the model parameterized by $\xi$. An estimator $\phi$ satisfies the exact recovery property if $P_{\text{err}}(\phi)$ goes to zero as $n$ tends to infinity.

As a by-product, an estimator $\phi$ with a vanishing $P_{\text{err}}(\phi)$ is also able to reliably recover the non-personalized binary rating matrix $B_\xi$ with high probability. However, as stated in Remark 2.2 below, the ability of merely recovering the binary rating matrix $B_\xi$ does not suffice for our task.

Remark 2.2. For two different instances $\xi \neq \xi'$, their corresponding non-personalized binary rating matrices $B_\xi$ and $B_{\xi'}$ may be the same. This can be shown via the following example with $m = 6, n = 2$:

- $\xi_M = \{1\}$, $\xi_W = \{2\}$, $\xi_A = \{1, 2, 3\}$, $\xi_R = \{4, 5, 6\}$, $\xi_{vM} = [1, 0, 0, 1, 0, 0, 0]$, and $\xi_{vW} = [0, 0, 1, 1, 1, 1, 1]$. In particular, the 3-rd movie is the atypical action movie, and all the other movies are typical;
- $\xi'_M = \{1\}$, $\xi'_W = \{2\}$, $\xi'_A = \{1, 2, 4\}$, $\xi'_R = \{3, 5, 6\}$, $\xi'_{vM} = [1, 1, 0, 0, 0, 0]$, and $\xi'_{vW} = [0, 0, 1, 1, 1, 1, 1]$. In particular, the 4-th movie is the atypical action movie, and all the other movies are typical.

In both cases, their corresponding non-personalized binary rating matrices are given by

\[
B_\xi = B_{\xi'} = \begin{pmatrix}
1 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 1 & 1 & 1
\end{pmatrix}.
\]

Essentially, if a movie attracts on average more men than women, it may be considered as either a typical action movie or an atypical romance movie. This observation also applies to movies that attract more women than men. Because of this flexibility in the classification of movies, a single non-personalized binary rating matrix may correspond to multiple distinct partitions of the set of $m$ movies into $m/2$ action movies and $m/2$ romance movies. Therefore, the ability of recovering $B_\xi$ does not necessarily guarantee the ability of recovering $\xi$.\footnote{However, one may also note that reliable recovery of the non-personalized binary rating matrix $B_\xi$ does imply the reliable recovery of the communities of men and women.}

This is why we emphasize in Section 1 that our task is strictly more difficult compared to that of recovering only the binary rating matrix.

3. Main results. For ease of presentation, the following results are stated in terms of the sample probability $p$ (which is proportional to the sample complexity $nmp$), and we sometimes use these two notions interchangeably. Theorem 3.1 and Theorem 3.3 below respectively provide a lower bound and an upper bound on the sample probability $p$, as a function of $n, m$, personalization parameters $\theta_A, \theta_R$, and the “qualities” of the row and column graphs $I_1$ and
symbol | description
--- | ---
m | number of movies
n | number of users
p | sample probability
\(nmp\) | sample complexity
\(\theta_A, \theta_R\) | personalization probabilities for action and romance movies
\(\tau_{UV}, U, V \in \{A, R\}\) | functions of \(\theta_A\) and \(\theta_R\)
\(\nu_{UV}, U, V \in \{A, R\}\) | functions of \(\theta_A\) and \(\theta_R\)
\(\alpha_1, \beta_2\) | intra- and inter-cluster probabilities of \(G_1\)
\(a_1, b_1\) | normalized intra- and inter-cluster probabilities of \(G_1\)
\(\alpha_2, \beta_2\) | intra- and inter-cluster probabilities of \(G_2\)
\(a_2, b_2\) | normalized intra- and inter-cluster probabilities of \(G_2\)
\(I_1\) and \(I_2\) | “qualities” of \(G_1\) and \(G_2\)
\(B_{UV}, U, V \in \{A, R\}\) | sets of movies satisfying (4.11)
\(t_{UV}, U, V \in \{A, R\}\) | number of movies satisfying (4.11)

Table 2: Table of parameters

\(I_2\). Before stating the main results, we first define two functions of \(\theta_A\) and \(\theta_R\) as follows:

\[
\begin{align*}
\tau_{UV} & \triangleq 1 - \sqrt{\theta_U \theta_V} - \sqrt{(1 - \theta_U)(1 - \theta_V)}, \quad \text{for } U, V \in \{A, R\}, \\
\nu_{UV} & \triangleq 1 - \sqrt{\theta_U(1 - \theta_V)} - \sqrt{\theta_V(1 - \theta_U)}, \quad \text{for } U, V \in \{A, R\}.
\end{align*}
\]

**Theorem 3.1.** (a) Consider the regime in which \(\theta_A \neq \theta_R\). For any \(\epsilon > 0\), if

\[
p \geq \max \left\{ \frac{(2(1 + \epsilon) - I_1) \log n}{(\nu_{AA} + \nu_{RR})m}, \frac{(1 + \epsilon) \log m \cdot (2(1 + \epsilon) - I_2) \log m}{\min\{\nu_{AA}, \nu_{RR}\} \cdot n}, \frac{(2(1 + \epsilon) - I_1) \log n}{2\tau_{AR}n} \right\},
\]

then there exists an estimator \(\phi\) satisfying \(\lim_{n \to \infty} P_{err}(\phi) = 0\).

(b) Consider the regime in which \(\theta_A = \theta_R\). For any \(\epsilon > 0\), if \(I_2 \geq 2(1 + \epsilon)\) and

\[
p \geq \max \left\{ \frac{(2(1 + \epsilon) - I_1) \log n}{(\nu_{AA} + \nu_{RR})m}, \frac{(1 + \epsilon) \log m \cdot (2(1 + \epsilon) - I_2) \log m}{\min\{\nu_{AA}, \nu_{RR}\} \cdot n} \right\},
\]

then there exists an estimator \(\phi\) satisfying \(\lim_{n \to \infty} P_{err}(\phi) = 0\).

In particular, the estimator \(\phi\) in **Theorem 3.1** can be chosen as the maximum likelihood estimator \(\phi_{ML}\).

**Remark 3.2.** By noting that \(\theta_A = \theta_R\) implies \(\tau_{AR} = 0\), the achievability result in (3.2) can be interpreted as a limiting consequence of (3.1) as \(\theta_A \to \theta_R\). When \(I_2 \geq 2(1 + \epsilon)\), the third term of (3.1) is non-positive and thus plays no role in the overall expression. When \(I_2 < 2(1 + \epsilon)\), no achievability result is provided since the third term of (3.1) becomes infinity (indeed, **Theorem 3.3(b)** below shows that exact recovery is impossible).
Theorem 3.3. (a) Consider the regime in which \( \theta_A \neq \theta_R \). For any \( \epsilon > 0 \), if

\[
(3.3) \quad p < \max \left\{ \frac{(2(1 - \epsilon) - I_1) \log n}{(\nu_{AA} + \nu_{RR}) m}, \frac{(1 - \epsilon) \log m}{\min\{\nu_{AA}, \nu_{RR}\} \cdot n}, \frac{(1 - \epsilon) - I_2) \log m}{2\tau_{AR} n} \right\},
\]

then \( \lim_{n \to \infty} P_{err}(\phi) = 1 \) for any estimator \( \phi \).

(b) Consider the regime in which \( \theta_A = \theta_R \). For any \( \epsilon > 0 \), if \( I_2 < 2(1 - \epsilon) \) or

\[
(3.4) \quad p < \max \left\{ \frac{(2(1 - \epsilon) - I_1) \log n}{(\nu_{AA} + \nu_{RR}) m}, \frac{(1 - \epsilon) \log m}{\min\{\nu_{AA}, \nu_{RR}\} \cdot n} \right\},
\]

then \( \lim_{n \to \infty} P_{err}(\phi) = 1 \) for any estimator \( \phi \).

Remark 3.4. Theorem 3.3 presents a strong converse result in the sense that the probability of error of any estimator asymptotically goes to one as long as \( p \) is smaller than the given upper bound.

A sharp threshold of the sample probability \( p \) is established when \( \theta_A = \theta_R \), as we have matching upper and lower bounds for this regime. When \( \theta_A \neq \theta_R \), the characterization of the sample probability \( p \) is order optimal—in particular, the upper and lower bounds match exactly for a wide range of parameter space, and match up to a constant factor of two for the remaining parameter space. We now provide two examples to illustrate the bounds in Theorem 3.1 and Theorem 3.3. These examples quantify the benefit of the graph side-information.

Example 3.5 \((n = m)\). Note that regardless of the values of \( I_1 \) (the quality of the row graph) and \((\theta_A, \theta_R)\), in both (3.1) and (3.2), the first term \( \frac{(2(1 + \epsilon) - I_1) \log n}{(\nu_{AA} + \nu_{RR}) m} \) is always upper bounded by the second term \( \frac{(1 + \epsilon) \log m}{\min\{\nu_{AA}, \nu_{RR}\} \cdot n} \) since

\[
\frac{(2(1 + \epsilon) - I_1) \log n}{(\nu_{AA} + \nu_{RR}) m} \leq \frac{2(1 + \epsilon) \log m}{(\nu_{AA} + \nu_{RR}) n} \leq \frac{(1 + \epsilon) \log m}{\min\{\nu_{AA}, \nu_{RR}\}}.
\]

This implies that when \( n = m \), the first term is inactive, and observing the row graph \( G_1 \) (with \( I_1 > 0 \)) does not help to reduce the sample complexity compared to the scenario in which \( G_1 \) is not observed. In fact, the above inequality and conclusions can be generalized to every \((m, n)\)-pair such that \( n \leq m \).

Another natural question to ask is that from the achievability’s perspective, whether observing the column graph \( G_2 \) (with \( I_2 > 0 \)) helps to reduce the sample complexity compared to the scenario in which \( G_2 \) is not observed. It turns out that the answer depends on the values of \((\theta_A, \theta_R)\), but does not depend on whether or not the row graph \( G_1 \) is observed. We assume the slackness parameter \( \epsilon = 0 \), and analyze three different instances in the following.

(i) When \((\theta_A, \theta_R)\) falls into the red region (including the boundary between the red and green regions) in Figure 2a, the sample probability in (3.1) is dominated by the second term when \( G_2 \) is not observed (i.e., \( I_2 = 0 \)), hence observing the column graph \( G_2 \) does not reduce the sample probability. Figure 2c plots the sample probability \( p \) as a function of \( I_2 \) for \((\theta_A, \theta_R) = (0.4, 0.1)\), and we note that \( p \) keeps unchanged as \( I_2 \) increases. This conclusion intuitively makes sense since the “big difference” between \( \theta_A \)
(a) Partitions of the \((\theta_A, \theta_R)\)-plane for the achievability part.

(b) Partitions of the \((\theta_A, \theta_R)\)-plane for the converse part.

(c) \(\theta_A = 0.4\) and \(\theta_R = 0.1\).

(d) \(\theta_A = 0.3\) and \(\theta_R = 0.15\)

Figure 2: Figure 2a shows the dominant term of (3.1) in Theorem 3.1(a) for \(I_1 = I_2 = \epsilon = 0\) and different values of \((\theta_A, \theta_R)\). Figure 2b shows the dominant term of (3.3) in Theorem 3.3(a) for \(I_1 = I_2 = \epsilon = 0\) and different values of \((\theta_A, \theta_R)\). In both (3.1) and (3.3), the second term dominates when \((\theta_A, \theta_R)\) falls into the red region, while the third term dominates when \((\theta_A, \theta_R)\) falls into the green region. Figure 2c and Figure 2d plot the sample probability \(p\) as a function of \(I_2\) (for \(n = m = 10,000\) and arbitrary \(I_1\)). In Figure 2d, as \(I_2\) increases, the sample probability first decreases before approaching the threshold (i.e., the vertical dotted line), and then stays constant after exceeding the threshold.

and \(\theta_R\) makes it easy to distinguish the action and romance movies from the partially observed matrix \(V^\Omega\), and the column graph then becomes useless. Note that in this regime, the converse also matches the achievability (as illustrated in Figure 2c), since the second terms in (3.1) and (3.3) are the same.

(ii) When \((\theta_A, \theta_R)\) falls into the green region in Figure 2a and satisfies \(\theta_A \neq \theta_R\), observing
the column graph \( G_2 \) does help to reduce the sample probability. This is because the sample probability is dominated by the third term of (3.1) when \( G_2 \) is not observed (i.e., \( I_2 = 0 \)), and increasing \( I_2 \) effectively decreases the third term. As illustrated in Figure 2d for \((\theta_A, \theta_R) = (0.3, 0.15)\), the sample probability with any positive \( I_2 \) is strictly smaller than that with \( I_2 = 0 \). Another interesting phenomenon is that once \( I_2 \) exceeds the “threshold” \( 2 - \frac{2\tau_{AR}}{\min\{\nu_{AA}, \nu_{RR}\}} \) (which is strictly positive for \((\theta_A, \theta_R)\)-pairs falling into the green region), the second term of (3.1) then becomes active, and the gain of increasing \( I_2 \) saturates. The vertical dotted line in Figure 2d denotes the threshold for \((\theta_A, \theta_R) = (0.3, 0.15)\), and as \( I_2 \) increases, the sample probability first decreases and then stays constant. Further, we note that the achievability and converse match if and only if \( I_2 \) exceeds the threshold (as illustrated in Figure 2d), since the second terms in both (3.1) and (3.3) become the dominant terms when \( I_2 \) exceeds the threshold.

(iii) When \( \theta_A = \theta_R \), according to Theorem 3.1(b) and Theorem 3.3(b), exact recovery is impossible if \( I_2 < 2 \), and is possible otherwise. Thus, observing the column graph \( G_2 \) is helpful only when the quality of \( G_2 \) is sufficiently high (i.e., \( I_2 > 2 \)).

Generally speaking, the achievability and converse match if the first term or the second term in (3.1) is the dominant term, and do not match if the third term in (3.1) dominates. This argument is also applicable to the following (more complicated) example.

**Example 3.6** \((n = 5m)\). Again, we assume the slackness parameter \( \epsilon = 0 \). Figure 3a partitions the \((\theta_A, \theta_R)\)-plane into three different regions—the yellow, red, and green regions respectively denote the collections of \((\theta_A, \theta_R)\)-pairs in which the first, the second, and the third term in (3.1) is the dominant term (when \( I_1 = I_2 = 0 \)). Figure 3b is similar to Figure 3a except that it considers the three terms in (3.3) for the converse part. Comparing Figure 3a and Figure 3b, we note that the boundary between the red and yellow regions does not change since the first and the second terms in both (3.1) and (3.3) are the same, while the boundary between the yellow and green regions move inwards (closer to the 45° line) since the third term in (3.3) is smaller than that in (3.1).

Considering the achievability part, both the row graph \( G_1 \) and the column graph \( G_2 \) may or may not be helpful for exact recovery. We analyze several different instances in the following.

(i) When \((\theta_A, \theta_R)\) falls into the red region (including the boundary between the red and the yellow regions) in Figure 3a, the sample probability in (3.1) is dominated by the second term when no graph is observed (i.e., \( I_1 = I_2 = 0 \)), hence neither the row graph \( G_1 \) nor the column graph \( G_2 \) helps to reduce the sample probability. The phenomenon is similar to that in Example 3.5(i).

(ii) When \((\theta_A, \theta_R)\) falls into the yellow region in Figure 3a, the sample probability in (3.1) is dominated by the first term, hence observing the row graph \( G_1 \) (with \( I_1 > 0 \)) reduces the sample probability (compared to the scenario in which \( G_1 \) is not observed). Figure 3c plots the sample probability \( p \) as a function of \( I_1 \) for \((\theta_A, \theta_R) = (0.3, 0.03)\) and three different values of \( I_2 \), and we note that

- regardless of the value of \( I_2 \), the sample probability with any positive \( I_1 \) is strictly smaller than that with \( I_1 = 0 \);
Figure 3: Figure 3a shows the dominant term of (3.1) in Theorem 3.1(a) for $I_1 = I_2 = \epsilon = 0$ and different values of $(\theta_A, \theta_R)$. Figure 3b shows the dominant term of (3.3) in Theorem 3.3(a) for $I_1 = I_2 = \epsilon = 0$ and different values of $(\theta_A, \theta_R)$. In both (3.1) and (3.3), the first, the second, and the third terms are respectively the dominant term when $(\theta_A, \theta_R)$ falls into the yellow region, the red region, and the green region. Figure 3c and Figure 3e plot the sample probability $p$ as a function of $I_1$ for $n = 5m = 10,000$ and two different values of $(\theta_A, \theta_R)$. Figure 3d and Figure 3f plot the sample probability $p$ as a function of $I_2$ for $n = 5m = 10,000$ and two different values of $(\theta_A, \theta_R)$.

- the column graph $G_2$ (with $I_2 > 0$) is also helpful when $I_1$ exceeds the red point, and is useless otherwise.

Further, the dotted black curve in Figure 3c indicates that the converse matches the achievability when $I_1$ does not exceed the red point (i.e., when the first term in (3.1) dominates), or $I_1$ and $I_2$ are both large enough (i.e., when the second term in (3.1) dominates).

(iii) When $(\theta_A, \theta_R)$ falls into the green region in Figure 3a and satisfies $\theta_A \neq \theta_R$, the sample probability in (3.1) is dominated by the third term, hence observing the column graph $G_2$ (with $I_2 > 0$) reduces the sample probability. Figure 3d plots the sample probability $p$ as a function of $I_2$ for $(\theta_A, \theta_R) = (0.3, 0.15)$ and three different values of $I_1$, and we note that

- regardless of the value of $I_1$, the sample probability with any positive $I_2$ is
strictly smaller than that with \( I_2 = 0 \);

- the row graph \( G_1 \) (with \( I_1 > 0 \)) is also helpful when \( I_2 \) exceeds the red point, and is useless otherwise.

Comparing the colored curves (for achievability) and black curves (for converse) in Figure 3d, we note that the achievability and converse match when \( I_2 \) is large enough (i.e., when the third term in (3.1) is no longer dominant).

(iv) When \( (\theta_A, \theta_R) \) is on the boundary between the yellow and green regions in Figure 3a, the first and the third terms in (3.1) are equal and the sample probability is dominated by both of the two terms. In this regime, observing both the row and column graphs \( G_1 \) and \( G_2 \) (with \( I_1 > 0, I_2 > 0 \)) reduces the sample probability compared to the scenario in which neither is observed. More interestingly, observing only one of the two graphs is equivalent to observing neither. Thus, there is a synergistic effect when both pieces of side-information (i.e., both graphs) are observed. The above argument is also illustrated in Figure 3c and Figure 3f, which respectively plot the sample probability \( p \) as functions of \( I_1 \) and \( I_2 \) for the boundary point \((\theta_A, \theta_R) = (0.35, 0.1156)\).

(v) When \( \theta_A = \theta_R \), observing the column graph \( G_2 \) is helpful only when \( I_2 > 2 \) (similar to Example 3.5(iii)). Conditioned on the observation of \( G_2 \) with \( I_2 > 2 \), observing the row graph \( G_1 \) (with \( I_1 > 0 \)) also reduces the sample probability. This is because the sample probability in (3.2) is dominated by the first term when \( G_1 \) is not observed, and increasing \( I_1 \) effectively decreases the first term.

4. Proof of Achievability.

4.1. Maximum likelihood estimator \( \phi_{\text{ML}} \). We use the maximum likelihood estimator \( \phi_{\text{ML}} \) to reconstruct \( \hat{\xi}_{\mathcal{M}, \mathcal{W}, A, A', R, v_M, v_W} \), which includes the communities \( \hat{\xi}_\mathcal{M}, \hat{\xi}_\mathcal{W}, \hat{\xi}_A, \hat{\xi}_R \), and the nominal rating vectors \( \hat{\xi}_{v_M}, \hat{\xi}_{v_W} \). For any \( \xi \in \Xi \), the negative log-likelihood of \( \xi \) is defined as

\[
L(\xi) \triangleq -\log \mathbb{P}_\xi(V^\Omega, G_1, G_2).
\]

Then, the estimation rule of \( \phi_{\text{ML}} \) is given by

\[
\hat{\xi} = \phi_{\text{ML}}(V^\Omega, G_1, G_2) = \arg\min_{\xi \in \Xi} L(\xi).
\]

Without loss of generality, we assume the ground truth is \( \xi^*_{\mathcal{M}, \mathcal{W}, A, A', R, v_M, v_W} \), where \( \xi^*_\mathcal{M} = [1 : \frac{n}{2}], \xi^*_{v_M} = [\frac{n}{2} + 1 : n], \xi^*_A = [1 : \frac{m}{2}], \xi^*_R = [\frac{m}{2} + 1 : m] \), and the nominal vectors \( \xi^*_{v_M}, \xi^*_{v_W} \) can be chosen arbitrarily as long as they satisfy \( d_H(\xi^*_{v_M}, \xi^*_{v_W}) = m \). Even though it may not be a priori clear, as shown in equation (4.9) to follow, the probabilities of error for different ground truths with different sizes of atypical action and romance movies are exactly the same. Hence,

\[
P_{\text{err}}(\phi_{\text{ML}}) = \max_{\xi \in \Xi} \mathbb{P}_\xi(\phi_{\text{ML}}(V^\Omega, G_1, G_2) \neq \xi) = \mathbb{P}_{\xi^*}(\phi_{\text{ML}}(V^\Omega, G_1, G_2) \neq \xi^*).
\]

4.2. Notations and Definitions. For any \( \xi \in \Xi \), we define the number of edges crossing the two communities \( \xi_{\mathcal{M}} \) and \( \xi_{\mathcal{W}} \) as \( e(\xi_{\mathcal{M}}, \xi_{\mathcal{W}}) \), and the number of edges crossing the two communities \( \xi_A \) and \( \xi_R \) as \( e(\xi_A, \xi_R) \), i.e.,

\[
e(\xi_{\mathcal{M}}, \xi_{\mathcal{W}}) \triangleq \sum_{i \in \xi_{\mathcal{M}}} \sum_{j \in \xi_{\mathcal{W}}} Y_{ij}, \quad e(\xi_A, \xi_R) \triangleq \sum_{i \in \xi_A} \sum_{j \in \xi_R} Y'_{ij},
\]
Note that \( \Pi \) where \( U, V \in \{ A^* \) and similar to (4.2), we respectively define \( \Pi \) as the set of index pairs \((i, j)\) such that the \((i, j)\)-th entry in the non-personalized binary rating matrix \( B_\xi \) corresponds to the action movies (resp. the romance movies), i.e.,

\[
I_{\xi_U} = \{ (i, j) : (i, j) \in [n] \times \xi_U, \ U \in \{ A, R \}. 
\]

The set of index pairs \((i, j)\) such that the \((i, j)\)-th entry in the partially observed matrix \( V^\Omega \) is not erased is denoted by

\[
\Omega \界定 \{ (i, j) \in [n] \times [m] : (V^\Omega)_{ij} \neq \perp \} ,
\]

and the set of index pairs \((i, j)\) that belong to both \( I_{\xi_U} \) and \( \Omega \) is denoted by

\[
(4.1) \quad \Omega_{\xi_U} \界定 \{ (i, j) \in I_{\xi_U} \cap \Omega = \{ (i, j) \in I_{\xi_U} : (V^\Omega)_{ij} \neq \perp \}, \ \text{for} \ U \in \{ A, R \}.
\]

Further, we define \( \Pi_{\xi_U} \) as the set of index pairs \((i, j)\) such that \((i, j) \in \Omega_{\xi_U}\) and the \((i, j)\)-th entry in \( B_\xi \) equals the \((i, j)\)-th entry in \( V^\Omega \), i.e.,

\[
(4.2) \quad \Pi_{\xi_U} \界定 \{ (i, j) \in \Omega_{\xi_U} : (B_\xi)_{ij} = (V^\Omega)_{ij} \}, \ \text{for} \ U \in \{ A, R \}.
\]

Now we consider the ground truth \( \xi^* \) together with another instance \( \xi \in \Xi \) such that \( \xi \neq \xi^* \).

For \( U, V \in \{ A, R \} \), let

\[
S_{\xi_U, \xi_V} \界定 \{ (i, j) : (i, j) \in [n] \times (\xi_U^* \cap \xi_V) \}, \ \text{for} \ U, V \in \{ A, R \},
\]

be the set of index pairs \((i, j)\) that belong to both \( I_{\xi_U} \) and \( I_{\xi_V} \), and we further partition \( S_{\xi_U, \xi_V} \) into two subsets

\[
S_{\xi_U, \xi_V}^e \界定 \{ (i, j) \in S_{\xi_U, \xi_V} : (B_\xi^*)_{ij} = (B_\xi)_{ij} \}, \quad S_{\xi_U, \xi_V}^{ue} \界定 \{ (i, j) \in S_{\xi_U, \xi_V} : (B_\xi^*)_{ij} \neq (B_\xi)_{ij} \} = S_{\xi_U, \xi_V} \setminus S_{\xi_U, \xi_V}^e ,
\]

where \( S_{\xi_U, \xi_V}^e \) (resp. \( S_{\xi_U, \xi_V}^{ue} \)) contains all the pairs \((i, j)\) \( \in S_{\xi_U, \xi_V} \) such that the \((i, j)\)-th entry in \( B_\xi^* \) and the \((i, j)\)-th entry in \( B_\xi \) are coincident (resp. different). Similar to (4.1), we define the set of index pairs \((i, j)\) that belong to both \( S_{\xi_U, \xi_V} \) and \( \Omega \) as

\[
\Omega_{\xi_U, \xi_V} \界定 \{ (i, j) \in S_{\xi_U, \xi_V} : (V^\Omega)_{ij} \neq \perp \}, \ \text{for} \ U, V \in \{ A, R \},
\]

and similar to (4.2), we respectively define \( \Pi_{\xi_U, \xi_V}^* \) and \( \Pi_{\xi_U, \xi_V} \) as

\[
\Pi_{\xi_U, \xi_V}^* \界定 \{ (i, j) \in \Omega_{\xi_U, \xi_V} : (B_\xi^*)_{ij} = (V^\Omega)_{ij} \}, \quad \Pi_{\xi_U, \xi_V} \界定 \{ (i, j) \in \Omega_{\xi_U, \xi_V} : (B_\xi)_{ij} = (V^\Omega)_{ij} \}.
\]

Note that \( \Pi_{\xi_U} \subseteq \Omega_{\xi_U} \subseteq I_{\xi_U} \) for \( U \in \{ A, R \}, \) and \( \Pi_{\xi_U, \xi_V}^*, \Pi_{\xi_U, \xi_V} \subseteq \Omega_{\xi_U, \xi_V} \subseteq S_{\xi_U, \xi_V} \) for \( U, V \in \{ A, R \}. \)
4.3. Analysis. For any instance $\xi \in \Xi$, the independence of the observations $V^\Omega, G_1, G_2$ yields that

$$P_\xi(V^\Omega, G_1, G_2) = P_\xi(V^\Omega)P_\xi(G_1)P_\xi(G_2).$$

By noting that

$$P_\xi(V^\Omega) = p^{|\Omega|}(1 - p)^{mn - |\Omega|} \prod_{U \in \{A, R\}} (1 - \theta_U)^{|\Pi_{\xi U}|} \theta_U^{|\Omega_{\xi U}| - |\Pi_{\xi U}|},$$

$$P_\xi(G_1) = \beta_1^e(\xi_M, \xi_W)(1 - \beta_1)^{(\frac{2}{3})^2 - e(\xi_M, \xi_W)} \alpha_1^{|E_1| - e(\xi_M, \xi_W)}(1 - \alpha_1)^{2(n^2/2 - |E_1| - e(\xi_M, \xi_W))},$$

$$P_\xi(G_2) = \beta_2^e(\xi_A, \xi_R)(1 - \beta_2)^{(\frac{2}{3})^2 - e(\xi_A, \xi_R)} \alpha_2^{|E_2| - e(\xi_A, \xi_R)}(1 - \alpha_2)^{2(n^2/2 - |E_2| - e(\xi_A, \xi_R))},$$

one can show that the negative log-likelihood of $\xi$ is given by

$$L(\xi) = -\log P_\xi(V^\Omega, G_1, G_2) = -\log P_\xi(V^\Omega) + \log P_\xi(G_1) + \log P_\xi(G_2)$$

$$= e(\xi_M, \xi_W) \log \left(\frac{\alpha_1(1 - \beta_1)}{\beta_1(1 - \alpha_1)}\right) + e(\xi_A, \xi_R) \log \left(\frac{\alpha_2(1 - \beta_2)}{\beta_2(1 - \alpha_2)}\right)$$

$$- \sum_{U \in \{A, R\}} |\Pi_{\xi U}| \log \left(\frac{1 - \theta_U}{\theta_U}\right) + |\Omega_{\xi U}| \log(\theta_U) + C_0,$$

(4.3)

where $C_0$ is a constant that is independent of $\xi$.

To bound the probability of error of the maximum likelihood estimator $\phi_{ML}$, we first apply the union bound to obtain that

$$P_{err}(\phi_{ML}) = P_{\xi^*}(\phi_{ML}(V^\Omega, G_1, G_2) \neq \xi^*) \leq \sum_{\xi \in \Xi, \xi \neq \xi^*} P_{\xi^*}(L(\xi) \leq L(\xi^*),$$

and then analyze the term $P_{\xi^*}(L(\xi) \leq L(\xi^*))$ for different $\xi \neq \xi^*$. By noting that $\Pi_{\xi U}^* = \bigcup_{V \in \{A, R\}} \Pi_{\xi U}^{\xi_U, \xi_V}$ and $\Pi_{\xi V} = \bigcup_{U \in \{A, R\}} \Pi_{\xi U}^{\xi_U, \xi_V}$, we have

$$\log P_{\xi^*}(V^\Omega) = \sum_{U \in \{A, R\}} |\Pi_{\xi U}^*| \log \left(\frac{1 - \theta_U}{\theta_U}\right) + |\Omega_{\xi U}| \log(\theta_U) + C_0$$

(4.4)

$$= \sum_{U \in \{A, R\}} \sum_{V \in \{A, R\}} |\Pi_{\xi U}^{\xi_U, \xi_V}| \log \left(\frac{1 - \theta_U}{\theta_U}\right) + |\Omega_{\xi U}^{\xi_U, \xi_V}| \log(\theta_U) + C_0,$$

(4.5)

$$\log P_{\xi}(V^\Omega) = \sum_{U \in \{A, R\}} \sum_{V \in \{A, R\}} |\Pi_{\xi U}^{\xi_U, \xi_V}| \log \left(\frac{1 - \theta_V}{\theta_V}\right) + |\Omega_{\xi U}^{\xi_U, \xi_V}| \log(\theta_V) + C_0,$$
and romance movies in parameter that quantifies the amount of overlap between the communities of action movies \(4.7\).

\[
\xi^* = \sum_{(i,j) \in S^w_{\xi^*}} \mathbb{1}\{ (B_{\xi^*})_{ij} = (V^\Omega)_{ij} \} + \sum_{(i,j) \in S^w_{\xi^*}} \mathbb{1}\{ (B_{\xi^*})_{ij} \neq (V^\Omega)_{ij} \},
\]

\[
\Omega_{\xi^*} = \sum_{(i,j) \in S^w_{\xi^*}} \mathbb{1}\{ (V^\Omega)_{ij} \neq \perp \} + \sum_{(i,j) \in S^w_{\xi^*}} \mathbb{1}\{ (V^\Omega)_{ij} \neq \perp \}.
\]

Subtracting (4.5) from (4.4), we have

\[
\log \mathbb{P}_{\xi^*}(V^\Omega) - \log \mathbb{P}_\xi(V^\Omega)
= \sum_{U,V \in \{A,R\}} |\Pi_{\xi^*}^u| \log \left( \frac{1 - \theta_U}{\theta_U} \right) - |\Pi_{\xi^*}^e| \log \left( \frac{1 - \theta_V}{\theta_V} \right) + |\Omega_{\xi^*}^e| \log \left( \frac{\theta_U}{\theta_V} \right)
= \sum_{U,V \in \{A,R\}} \left[ \sum_{(i,j) \in S^w_{\xi^*}} \log \left( \frac{1 - \theta_U}{\theta_U(1 - \theta_V)} \right) \mathbb{1}\{ (B_{\xi^*})_{ij} = (V^\Omega)_{ij} \} + \log \left( \frac{1 - \theta_U}{\theta_U} \right) \mathbb{1}\{ (B_{\xi^*})_{ij} \neq (V^\Omega)_{ij} \} \right]
+ \sum_{(i,j) \in S^w_{\xi^*}} \mathbb{1}\{ (V^\Omega)_{ij} \neq \perp \}
= \sum_{U,V \in \{A,R\}} \left[ \sum_{(i,j) \in S^w_{\xi^*}} \log \left( \frac{1 - \theta_U}{\theta_U(1 - \theta_V)} \right) Z_{ij} (1 - \Theta_{ij}^U) + \log \left( \frac{\theta_U}{\theta_V} \right) Z_{ij} \right]
+ \sum_{(i,j) \in S^w_{\xi^*}} \log \left( \frac{\theta_V}{1 - \theta_V} \right) Z_{ij} \Theta_{ij}^U + \log \left( \frac{1 - \theta_U}{\theta_U} \right) Z_{ij} (1 - \Theta_{ij}^U) + \log \left( \frac{\theta_U}{\theta_V} \right) Z_{ij},
\]

where \( \{ Z_{ij} \} \overset{1.i.d.}{=} \text{Bern}(p), \{ \Theta_{ij}^U \} \overset{1.i.d.}{=} \text{Bern}(\theta_U) \) for \( U \in \{A,R\} \).

Let \( k_1 \overset{\Delta}{=} |\xi_M \setminus \xi^*_M| = |\xi_W \setminus \xi^*_W| \) be the parameter that quantifies the amount of overlap between the communities of men and women in \( \xi^* \) and \( \xi \), and \( k_2 \overset{\Delta}{=} |\xi_R \setminus \xi^*_R| = |\xi_K \setminus \xi^*_K| \) be the parameter that quantifies the amount of overlap between the communities of action movies and romance movies in \( \xi^* \) and \( \xi \). One can show that

\[
\log \mathbb{P}_{\xi^*}(G_1) - \log \mathbb{P}_\xi(G_1) = \log \left( \frac{(1 - \beta_1)\alpha_1}{(1 - \alpha_1)\beta_1} \right) \sum_{i=1}^{nk_1-2k_2^2} (Y_i - X_i),
\]

\[
\log \mathbb{P}_{\xi^*}(G_2) - \log \mathbb{P}_\xi(G_2) = \log \left( \frac{(1 - \beta_2)\alpha_2}{(1 - \alpha_2)\beta_2} \right) \sum_{i=1}^{nk_2-2k_2^2} (Y'_i - X'_i),
\]
where \( \{Y_i\} \sim \text{i.i.d.} \text{Bern}(\beta_1), \{X_i\} \sim \text{i.i.d.} \text{Bern}(\alpha_1), \{Y'_i\} \sim \text{i.i.d.} \text{Bern}(\beta_2), \) and \( \{X'_i\} \sim \text{i.i.d.} \text{Bern}(\alpha_2) \).

For notational convenience we set \( c = \log \left( \frac{1-\beta_1\alpha_1}{1-\alpha_1}\right), d = \log \left( \frac{1-\beta_2\alpha_2}{1-\alpha_2}\right) \), \( C = nk_1 - 2k_1^2, \) \( D = mk_2 - 2k_2^2 \). By applying the Chernoff bound \( \mathbb{P}(X > \xi) \leq \min_{t>0} e^{-t\xi} \cdot \mathbb{E}(e^{tX}) \) with \( t = 1/2 \), and combining (4.6), (4.7), and (4.8), we have

\[
\mathbb{P}_{\xi}(L(\xi) \leq L(\xi^*)) = \mathbb{P}\left(c \sum_{i=1}^{C} (Y_i - X_i) + d \sum_{i=1}^{D} (Y'_i - X'_i)
+ \sum_{U,V \in \{A,R\}} \sum_{(i,j) \in S_{\xi_U,\xi_V}^e} \log \left( \frac{1 + \frac{1-\theta_U}{\theta_U}(1-\theta_V)}{\frac{1-\theta_U}{\theta_U}(1-\theta_V)} \right) Z_{ij}(\Theta(U)_j - 1) - \log \left( \frac{\theta_U}{\theta_V} \right) Z_{ij}\right)
+ \left[ \sum_{(i,j) \in S_{\xi_U,\xi_V}^e} \log \left( \frac{1 - \theta_U}{\theta_U} \right) Z_{ij}(\Theta(U)_j - 1) + \log \left( \frac{1 - \theta_V}{\theta_V} \right) Z_{ij}\Theta(U)_j - \log \left( \frac{\theta_U}{\theta_V} \right) Z_{ij}\right] \geq 0
\leq \exp\left\{ -C I_1 \frac{\log n}{n} - D I_2 \frac{\log m}{m} \right\} \prod_{U,V \in \{A,R\}} \left(1 - p_{UV}\right)^{S_{\xi_U,\xi_V}^e} \left(1 - p_{UV}\right)^{S_{\xi_U,\xi_V}^w}
\leq \exp\left\{ -C I_1 \frac{\log n}{n} - D I_2 \frac{\log m}{m} - p \cdot \left[ \sum_{U,V \in \{A,R\}} S_{\xi_U,\xi_V}^e \cdot \tau_{UV} + S_{\xi_U,\xi_V}^w \cdot \nu_{UV} \right]\right\}.\]

For \( U, V \in \{A, R\} \), let \( B_{UV} \) be the set of movies belonging to both \( \xi_U^* \) and \( \xi_V \) and the ratings in \( \xi^e_{U} \) and \( \xi^e_{V} \) are different (as illustrated in Figure 4), and let \( t_{UV} \) be the cardinality of \( B_{UV} \), i.e.,

\[
B_{UV} \triangleq \{|i \in [m] : i \in \xi^*_U \cap \xi_V, \xi^e_{U}(i) \neq \xi^e_{V}(i)\}, \quad t_{UV} \triangleq |B_{UV}|.
\]

We then provide expressions for \( |S_{\xi^e_U,\xi_V}^e|, |S_{\xi^e_U,\xi_V}^w| \) and \( |S_{\xi^e_U,\xi_V}^{uw}| \) as follows:

\[
|S_{\xi^e_U,\xi_V}^e| = \begin{cases} n \left( \frac{m}{2} - k_2 \right), & \text{if } (U, V) = (A, A), (R, R), \\
nk_2, & \text{if } (U, V) = (A, R), (R, A). \end{cases}
\]

\[
|S_{\xi^e_U,\xi_V}^{uw}| = 2 \left( \frac{n}{2} - k_1 \right) t_{UV} \left| \sum_{U,V \in \{A,R\}} S_{\xi_U,\xi_V}^e \cdot \tau_{UV} + S_{\xi_U,\xi_V}^w \cdot \nu_{UV} \right|.
\]

\[
|S_{\xi^e_U,\xi_V}^{we}| = 2 \left( \frac{n}{2} - k_1 \right) \left( |S_{\xi^e_U,\xi_V}^e| - t_{UV} \right) + 2k_1 t_{UV},
\]

\[
|S_{\xi^e_U,\xi_V}^{we}| = 2 \left( \frac{n}{2} - k_1 \right) \left( |S_{\xi^e_U,\xi_V}^e| - t_{UV} \right) + 2k_1 t_{UV},
\]
Then we have

$$
\sum_{U,V \in \{A,R\}} |S^u_{\xi_{UV}}| \cdot \tau_{UV} + |S^w_{\xi_{UV}}| \cdot \nu_{UV}
$$

$$
= \tau_{AR} \left( 2 \left( \frac{n}{2} - k_1 \right) (k_2 - t_{AR}) + 2k_1 t_{AR} \right) + \tau_{RA} \left( 2 \left( \frac{n}{2} - k_1 \right) (k_2 - t_{RA}) + 2k_1 t_{RA} \right)
$$

$$
+ \nu_{AR} \left( 2 \left( \frac{n}{2} - k_1 \right) t_{AR} + 2k_1 (k_2 - t_{AR}) \right) + \nu_{RA} \left( 2 \left( \frac{n}{2} - k_1 \right) t_{RA} + 2k_1 (k_2 - t_{RA}) \right)
$$

$$
+ \nu_{AA} \left( 2 \left( \frac{n}{2} - k_1 \right) t_{AA} + 2k_1 \left( \frac{m}{2} - k_2 - t_{AA} \right) \right)
$$

$$
+ \nu_{RR} \left( 2 \left( \frac{n}{2} - k_1 \right) t_{RR} + 2k_1 \left( \frac{m}{2} - k_2 - t_{RR} \right) \right)
$$

$$
= 2\tau_{AR} nk_2 + (2\nu_{AR} - 2\tau_{AR}) \left( 2k_1 k_2 + \frac{n}{2}(t_{AR} + t_{RA}) - 2k_1 (t_{AR} + t_{RA}) \right)
$$

$$
+ 2k_1 \left( \frac{m}{2} - k_2 \right) (\nu_{AA} + \nu_{RR}) + \nu_{AA} t_{AA}(n - 4k_1) + \nu_{RR} t_{RR}(n - 4k_1)
$$

$$
\geq 2\tau_{AR} nk_2 + 2k_1 \left( \frac{m}{2} - k_2 \right) (\nu_{AA} + \nu_{RR}) + (t_{AA} + t_{RR})(n - 4k_1) \min\{\nu_{AA}, \nu_{RR}\}.
$$

(4.12)

where the last step holds since \( \nu_{AR} \geq \tau_{AR} \) for any \( \theta_A, \theta_R \in \left( 0, \frac{1}{2} \right) \). Substituting (4.12) into (4.10), we have

$$
\mathbb{P}_{\xi^*}(L(\xi) \leq L(\xi^*))
$$

$$
\leq \exp \left\{ - (nk_1 - 2k_1^2) I_1 \frac{\log n}{n} - (mk_2 - 2k_2^2) I_2 \frac{\log m}{m}
$$

$$
- p \cdot \left[ 2\tau_{AR} nk_2 + 2k_1 \left( \frac{m}{2} - k_2 \right) (\nu_{AA} + \nu_{RR}) + (t_{AA} + t_{RR})(n - 4k_1) \min\{\nu_{AA}, \nu_{RR}\} \right] \right\}.
$$

Note that the expression above only depends on \( k_1, k_2, t_{AA}, t_{RR} \). Let \( \Xi_{\xi^*}(k_1, k_2, t_{AA}, t_{RR}) \) be the set of \( \xi \) with the same error probability \( \mathbb{P}_{\xi^*}(L(\xi) \leq L(\xi^*)) \), and \( T \) be the set of tuples

Figure 4: An illustration of \( B_{AA}, B_{AR}, B_{RA}, \) and \( B_{RR} \) corresponding to \( \xi^* \) and \( \xi \). Action movies are labelled in blue while romance movies are labelled in red—in particular, light blue and pink respectively denote atypical action and romance movies.
of interest, i.e.,

\[
\mathcal{T} \triangleq \left\{ (k_1, k_2, t_{AA}, t_{RR}) \neq (0, 0, 0, 0) : k_1 \in \left[ 0 : \frac{n}{4} \right], k_2 \in \left[ 0 : \frac{m}{4} \right], t_{AA} \in \left[ 0 : \frac{m}{2} - k_2 \right], t_{RR} \in \left[ 0 : \frac{m}{2} - k_2 \right] \right\}.
\]

For any \( \epsilon > 0 \), we define an auxiliary parameter \( \delta_\epsilon \), which is independent of both \( m \) and \( n \), as \( \delta_\epsilon \triangleq \min \left\{ \epsilon, \frac{\sqrt{47}}{8(1 + \epsilon)} \right\} \). The following analysis decomposes the parameter space \( \Xi \) into different “types” \( \Xi_\epsilon^* (k_1, k_2, t_{AA}, t_{RR}) \), i.e.,

\[
P_{err}(\phi_{ML}) \leq \sum_{\xi \in \Xi_\Xi \neq \xi^*} P_{\xi^*}(L(\xi) \leq L(\xi^*))
= \sum_{(k_1, k_2, t_{AA}, t_{RR}) \in \mathcal{T}} |\Xi_{\xi^*}(k_1, k_2, t_{AA}, t_{RR})| \cdot P_{\xi^*}(L(\xi) \leq L(\xi^*))
\leq \sum_{i=1}^{4} \sum_{(k_1, k_2, t_{AA}, t_{RR}) \in \mathcal{T}_i} |\Xi_{\xi^*}(k_1, k_2, t_{AA}, t_{RR})| \cdot P_{\xi^*}(L(\xi) \leq L(\xi^*))
\]

(4.13)

where \( \mathcal{T} \subset (\cup_{i=1}^{4} \mathcal{T}_i) \) and the four sets \( \mathcal{T}_1, \mathcal{T}_2, \mathcal{T}_3, \) and \( \mathcal{T}_4 \) are respectively defined as

\[
\mathcal{T}_1 \triangleq \left\{ (k_1, k_2, t_{AA}, t_{RR}) : k_1 \in \left[ 1 : \delta_\epsilon n \right], k_2 \in \left[ 0 : \delta_\epsilon m \right], t_{AA} \in \left[ 0 : \frac{m}{2} - k_2 \right], t_{RR} \in \left[ 0 : \frac{m}{2} - k_2 \right] \right\},
\]

\[
\mathcal{T}_2 \triangleq \left\{ (k_1, k_2, t_{AA}, t_{RR}) : k_1 = 0, k_2 \in \left[ 1 : \delta_\epsilon m \right], t_{AA} \in \left[ 0 : \frac{m}{2} - k_2 \right], t_{RR} \in \left[ 0 : \frac{m}{2} - k_2 \right] \right\},
\]

\[
\mathcal{T}_3 \triangleq \left\{ (k_1, k_2, t_{AA}, t_{RR}) : k_1 > \delta_\epsilon n, k_2 \in \left[ 0 : \frac{m}{4} \right], t_{AA} \in \left[ 0 : \frac{m}{2} - k_2 \right], t_{RR} \in \left[ 0 : \frac{m}{2} - k_2 \right] \right\},
\]

\[
\mathcal{T}_4 \triangleq \left\{ (k_1, k_2, t_{AA}, t_{RR}) : k_1 \in \left[ 0 : \frac{n}{4} \right], k_2 > \delta_\epsilon m, t_{AA} \in \left[ 0 : \frac{m}{2} - k_2 \right], t_{RR} \in \left[ 0 : \frac{m}{2} - k_2 \right] \right\}.
\]

We now show that the probabilities of error induced by each of the four terms in (4.13) are vanishing as \( n \) tends to infinity.

**4.3.1. Case 1: \( k_1 \leq \delta_\epsilon n, k_2 \leq \delta_\epsilon m \).** When \( k_1 \leq \delta_\epsilon n \) and \( k_2 \leq \delta_\epsilon m \), we have

\[
(nk_1 - 2k_1^2) = 2k_1 \left( \frac{n}{2} - k_1 \right) \geq 2k_1 n \left( \frac{1}{2} - \delta_\epsilon \right),
\]

\[
(mk_2 - 2k_2^2) = 2k_2 \left( \frac{m}{2} - k_2 \right) \geq 2k_2 m \left( \frac{1}{2} - \delta_\epsilon \right).
\]
Therefore,
\[ \Pr_{\xi^*}(L(\xi) \leq L(\xi^*)) \]
\[ \leq \exp \left\{ -2k_1 n \left( \frac{1}{2} - \delta_e \right) I_1 \log \frac{n}{n} - 2k_2 m \left( \frac{1}{2} - \delta_e \right) I_2 \log \frac{m}{m} \right. \]
\[ - p \left[ 2\tau_{AR} n k_2 + 2k_1 m \left( \frac{1}{2} - \delta_e \right) (\nu_{AA} + \nu_{RR}) + (t_{AA} + t_{RR}) n (1 - 4\delta_e) \min \{\nu_{AA}, \nu_{RR}\} \right] \}
\leq \exp \left\{ -2k_1 \left[ \left( \frac{1}{2} - \delta_e \right) I_1 (\log n) + pm \left( \frac{1}{2} - \delta_e \right) (\nu_{AA} + \nu_{RR}) \right] \right. \]
\[ (4.14) \]
\[ - 2k_2 \left[ \left( \frac{1}{2} - \delta_e \right) I_2 (\log m) + pm \tau_{AR} \right] - pm(t_{AA} + t_{RR})(1 - 4\delta_e) \min \{\nu_{AA}, \nu_{RR}\} \right\} \]

Since \( p \geq \frac{(2(1+\epsilon)-I_1) \log n}{(\nu_{AA} + \nu_{RR}) m} \), one may check that the first part of (4.14) satisfies
\[ 2k_1 \left[ \left( \frac{1}{2} - \delta_e \right) I_1 (\log n) + pm \left( \frac{1}{2} - \delta_e \right) (\nu_{AA} + \nu_{RR}) \right] \geq \left(1 + \frac{\epsilon}{2}\right) 2k_1 \log n. \]

We now consider the second part of (4.14). Recall from the statement of Theorem 3.1 that
- when \( \theta_A \neq \theta_R \), we set \( p \geq \frac{(2(1+\epsilon)-I_2) \log m}{2\tau_{AR} n} \) and \( p \geq \frac{(1+\epsilon) \log m}{\min \{\nu_{AA}, \nu_{RR}\} n} \);
- when \( \theta_A = \theta_R \) and \( I_2 \geq 2(1+\epsilon) \), we set \( p \geq \frac{(1+\epsilon) \log m}{\min \{\nu_{AA}, \nu_{RR}\} n} \).

Then, for both of the regimes, we have\(^5\)
\[ 2k_2 \left[ \left( \frac{1}{2} - \delta_e \right) I_2 (\log m) + pm \tau_{AR} \right] + pm(t_{AA} + t_{RR})(1 - 4\delta_e) \min \{\nu_{AA}, \nu_{RR}\} \]
\[ \geq \left(1 + \frac{\epsilon}{2}\right) (2k_2 + t_{AA} + t_{RR}) \log m. \]

Moreover, the size of \( \Xi_{\xi^*}(k_1, k_2, t_{AA}, t_{RR}) \) can be bounded from above as
\[
|\Xi_{\xi^*}(k_1, k_2, t_{AA}, t_{RR})| = \left( \frac{n}{k_1} \right)^2 \left( \frac{m}{k_2} \right)^2 \left( \frac{m}{2} - k_2 \right) \left( \frac{m}{2} - k_2 \right) \sum_{t_{AA} = 0}^{k_2} \sum_{t_{RR} = 0}^{k_2} \left( \frac{k_2}{t_{AA}} \right) \left( \frac{k_2}{t_{RR}} \right) \]
\[ \leq n^{2k_1 m^{2k_2 m^{-k_2}} m^{-t_{AA} + t_{RR}}} \cdot g^{k_2} \]
\[ \leq \exp (2k_1 \log n) \exp ((2k_2 + t_{AA} + t_{RR}) \log m + 2k_2). \]

Therefore,
\[
\sum_{(k_1, k_2, t_{AA}, t_{RR}) \in T_1} |\Xi_{\xi^*}(k_1, k_2, t_{AA}, t_{RR})| \cdot \Pr_{\xi^*}(L(\xi) \leq L(\xi^*)) \]
\[ \leq e^{-\epsilon k_1 \log n} \sum_{k_2 \in [0, \delta_m]} \sum_{t_{AA} \in [0, \frac{m}{2} - k_2]} m^{-\frac{k_2}{t_{AA}}} \sum_{t_{RR} \in [0, \frac{m}{2} - k_2]} m^{-\frac{k_2}{t_{RR}}} \]
\[ \leq 16 m^{-\epsilon}. \]

\(^5\)It is worth noting that \( \tau_{AR} = 0 \) when \( \theta_A = \theta_R \).
and similarly,
\[
\sum_{(k_1, k_2, t_{AA}, t_{RR}) \in T_2} |\mathbb{E}_\xi^* (k_1, k_2, t_{AA}, t_{RR})| \cdot \mathbb{P}_\xi^* (L(\xi) \leq L(\xi^*)) \leq 32n^\epsilon.
\]

### 4.3.2. Case 2: \( k_1 > \delta n \) or \( k_2 > \delta m \).
In this case, one may check that
\[
\mathbb{P}_\xi^* (L(\xi) \leq L(\xi^*)) \leq \exp \{-\Omega(pm)\} = \exp \{-\Omega(m \log m + n \log n)\}.
\]
Since the number of partitions of the set of \( n \) users into \( n^2 \) men and \( n^2 \) women is upper bounded by \( 2^n \), the number of partitions of the set of \( m \) movies into \( m^2 \) action movies and \( m^2 \) romance movies is upper bounded by \( 2^m \), and the number of possible nominal rating vectors is upper bounded by \( 2^{2m} \), we have
\[
|\Xi| \leq 2^n + 3m.
\]
Therefore, one can show that the error events corresponding to \( k_1 > \delta n \) or \( k_2 > \delta m \) occur with vanishing probabilities, i.e.,
\[
\sum_{(k_1, k_2, t_{AA}, t_{RR}) \in T_3 \cup T_4} |\mathbb{E}_\xi^* (k_1, k_2, t_{AA}, t_{RR})| \cdot \mathbb{P}_\xi^* (L(\xi) \leq L(\xi^*))
\leq |\Xi| \cdot \exp \{-\Omega(m \log m + n \log n)\} = o(1).
\]

### 5. Proof of Converse.
In the following, we show that for any \( \epsilon > 0 \), the probability of error \( P_{\text{err}} \) goes to one as \( n \) tends to infinity for the following two settings:
- when \( \theta_A \neq \theta_R \), the sample probability
  \[
  p < \max \left\{ \left( \frac{2(1 - \epsilon) - I_1}{\nu_{AA} + \nu_{RR}} \right) \log n, \left( (1 - \epsilon) - I_2 \right) \log m, \frac{(1 - \epsilon) \log m}{\min\{\nu_{AA}, \nu_{RR}\} \cdot n} \right\};
  \]
- when \( \theta_A = \theta_R \), \( I_2 < 2(1 - \epsilon) \) or the sample probability
  \[
  p < \max \left\{ \left( \frac{2(1 - \epsilon) - I_1}{\nu_{AA} + \nu_{RR}} \right) \log n, \left( (1 - \epsilon) - I_2 \right) \log m \right\}.
  \]
As argued in [5], the maximum likelihood estimator \( \phi_{\text{ML}} \) minimizes the probability of error \( P_{\text{err}} \) defined in (2.1), and one can further assume that the ground truth is \( \xi^* \) without loss of generality, i.e.,
\[
\inf_{\phi} P_{\text{err}}(\phi) \geq \mathbb{P}_{\xi^*} \left( \phi_{\text{ML}}(V, G_1, G_2) \neq \xi^* \right).
\]
Hence, it suffices to analyze the probability of error with respect to \( \phi_{\text{ML}} \) and \( \xi^* \). Before stating the detailed proofs, we first provide two technical lemmas showing the tightness of the Chernoff bound.
Lemma 5.1 (Adapted from Lemma 4 of [5]). For integers $K, L_1, L_2 > 0$, let \( \{Y_i\}_{i=1}^K \) i.i.d. Bern(\( \beta_1 \)), \( \{X_i\}_{i=1}^K \) i.i.d. Bern(\( \alpha_1 \)), \( \{Z_i\}_{i=1}^{L_1} \) i.i.d. Bern(\( p \)), \( \{\Theta^{A'}_i\}_{i=1}^{L_1} \) i.i.d. Bern(\( \theta_A \)), \( \{\Theta^{R'}_i\}_{i=1}^{L_2} \) i.i.d. Bern(\( \theta_R \)), and assume that \( \alpha_1, \beta_1, p = o(1) \) and \( \max\{\sqrt{\alpha_1 \beta_1} K, pL\} = \omega(1) \). Then,

\[
P\left( \sum_{i=1}^K (Y_i - X_i) + \log \left( \frac{1 - \theta_A}{\theta_A} \right) \sum_{i=1}^{L_1} Z_i (2\Theta^{A'}_i - 1) + \log \left( \frac{1 - \theta_R}{\theta_R} \right) \sum_{i=1}^{L_2} Z'_i (2\Theta^{R'}_i - 1) \right) \geq \frac{1}{4} \exp \left\{ -(1 + o(1)) K I_1 \log \frac{n}{m} - (1 + o(1)) L_1 p \nu_{AA} - (1 + o(1)) L_2 p \nu_{RR} \right\}.
\]

Proof. See [5, Appendix A-D].

Lemma 5.2. For integers $K, L_1, L_2 > 0$, let \( \{Y'_i\}_{i=1}^K \) i.i.d. Bern(\( \beta_2 \)), \( \{X'_i\}_{i=1}^K \) i.i.d. Bern(\( \alpha_2 \)), \( \{Z_i\}_{i=1}^{L_1} \) i.i.d. Bern(\( p \)), \( \{\Theta^{A}_i\}_{i=1}^{L_1} \) i.i.d. Bern(\( \theta_A \)), \( \{\Theta^{R}_i\}_{i=1}^{L_2} \) i.i.d. Bern(\( \theta_R \)), and assume that \( \alpha_2, \beta_2, p = o(1) \) and \( \max\{\sqrt{\alpha_2 \beta_2} K, pL\} = \omega(1) \). Then,

\[
P\left( \sum_{i=1}^K (Y'_i - X'_i) + \sum_{i=1}^{L_1} \log \left( \frac{1 - \theta_A}{\theta_A} \right) Z_i (\Theta^{A}_i - 1) - \log \left( \frac{\theta_A}{\theta_R} \right) Z'_i \right) \geq \frac{1}{4} \exp \left\{ -(1 + o(1)) K I_2 \log \frac{m}{n} - (1 + o(1)) (L_1 + L_2) p \tau_{AR} \right\}.
\]

Proof. See Appendix A.

In Subsection 5.1, 5.2, and 5.3 below, we show that if the sample probability \( p \) satisfies (5.1) or (5.2), the probability of error goes to one even if we restrict our analysis to a particular subset of the parameter space \( \Xi \) (essentially, we focus on a subset of events that are likeliest to cause errors).

5.1. Case 1: restricted to \( \Xi^{*}(0, 0, 1, 0) \) or \( \Xi^{*}(0, 0, 0, 1) \). Suppose \( p < \frac{(1 - \epsilon) \log m}{L_1 \nu_{AA}} \). Note that the probability of success \( P_{\text{suc}} \), which equals \( 1 - P_{\text{err}} \), can be bounded from above as

\[ P_{\text{suc}} = \mathbb{P}_{\Theta^{*}} \left( \bigcap_{\xi \in \Xi : \xi \neq \xi^*} \{ L(\xi) > L(\xi^*) \} \right) \]
\[ = \mathbb{P}_{\Theta^{*}} \left( \bigcap_{(k_1, k_2, \ell_{AA}, \ell_{RR}) \in T} \bigcap_{\xi \in \Xi^{*}(k_1, k_2, \ell_{AA}, \ell_{RR})} \{ L(\xi) > L(\xi^*) \} \right) \]
\[ \leq \mathbb{P}_{\Theta^{*}} \left( \bigcap_{\xi \in \Xi^{*}(0, 0, 0, 1)} \{ L(\xi) > L(\xi^*) \} \right). \]
Substituting $k_1 = k_2 = t_{RR} = 0$ and $t_{AA} = 1$ into (4.9), we have that for any $\xi \in \Xi_{\xi^*}(0,0,1,0)$,

$$P_{\xi^*}(L(\xi) \leq L(\xi^*)) = \mathbb{P} \left( \sum_{(i,j) \in S_{\xi^*}t_{A}} \log \left( \frac{1 - \theta_A}{\theta_A} \right) Z_{ij}(2\Theta_A^A - 1) \geq 0 \right) \geq \frac{1}{4} \exp \left\{ -(1 + o(1))np_{t_{AA}} \right\},$$  

(5.3)

where (5.3) follows from Lemma 5.1 by substituting $L_1 = n$ and $K = L_2 = 0$. Hence, for sufficiently large $n$,

$$P_{\xi^*}(L(\xi) > L(\xi^*)) = 1 - P_{\xi^*}(L(\xi) \leq L(\xi^*)) \leq 1 - \frac{1}{4} \exp \left\{ -(1 + o(1))np_{t_{AA}} \right\} = \exp \left\{ -\frac{1}{4} e^{-(1+o(1))np_{t_{AA}}} \right\}.$$

One may check that conditioned on the ground truth $\xi^*$, the events $\{L(\xi) > L(\xi^*)\}$ for different $\xi \in \Xi_{\xi^*}(0,0,1,0)$ are mutually independent. By noting that $p < \frac{(1 - \epsilon) \log m}{\nu_{t_{AA}} n}$ and there are $\frac{m}{2}$ mutually independent events, we have

$$P_{\xi^*} \left( \bigcap_{\xi \in \Xi_{\xi^*}(0,0,1,0)} \{L(\xi) > L(\xi^*)\} \right) = \prod_{\xi \in \Xi_{\xi^*}(0,0,1,0)} P_{\xi^*} \left( L(\xi) > L(\xi^*) \right)$$

$$= \exp \left\{ -\frac{m}{8} e^{-(1+o(1))np_{t_{AA}}} \right\} = \exp \left\{ -\frac{1}{8} e^{\log m - (1 + o(1))np_{t_{AA}}} \right\} \leq \exp \left\{ -\frac{1}{8} m^{\frac{\epsilon}{8}} \right\},$$

which implies that the probability of success $P_{\text{suc}}$ is upper bounded by $\exp \left\{ -\frac{1}{8} m^{\frac{\epsilon}{8}} \right\}$.

Similarly, when $p < \frac{(1 - \epsilon) \log m}{\nu_{t_{RR}} n}$, one can show that the probability of success $P_{\text{suc}}$ is upper bounded by $\exp \left\{ -\frac{1}{8} m^{\frac{\epsilon}{8}} \right\}$ by replacing $\Xi_{\xi^*}(0,0,1,0)$ with $\Xi_{\xi^*}(0,0,0,1)$ in the above analysis (i.e., set $k_1 = 0, k_2 = 0, t_{AA} = 0, t_{RR} = 1$). Therefore, as long as

$$p < \min \left\{ \frac{(1 - \epsilon) \log m}{\nu_{t_{AA}} n}, \frac{(1 - \epsilon) \log m}{\nu_{t_{RR}} n} \right\} = \frac{(1 - \epsilon) \log m}{\min\{\nu_{t_{AA}}, \nu_{t_{RR}}\} \cdot n},$$

the probability of success converges to zero.

**5.2. Case 2: restricted to $\Xi_{\xi^*}(1,0,0,0)$**. Suppose $p < \frac{2(1 - \epsilon) - L_1 \log n}{(\nu_{t_{AA}} + \nu_{t_{RR}}) m}$. Consider the ground truth tuple $\xi^*$. For each $i \in \xi_M$ and $j \in \xi_W$, we define the tuple $\xi^*_{\text{row}}$ to be identical to $\xi^*$ except that $(\xi^*_{\text{row}})^M = \xi^*_M \setminus \{i\}$ and $(\xi^*_{\text{row}})^W = \xi^*_W \cup \{i\}$, the tuple $\xi^*_{\text{col}}(j)$ to be identical to $\xi^*$ except that $(\xi^*_{\text{col}})^M = \xi^*_M \cup \{j\}$ and $(\xi^*_{\text{col}})^W = \xi^*_W \setminus \{j\}$, and the tuple $\xi^*_{\text{row}}(i,j)$ to be
identical to $\xi^*$ except that the $i$-th user and the $j$-th user are swapped. The probability of success is bounded from above as

$$P_{\text{succ}} \leq \mathbb{P}_{\xi^*} \left( \bigcap_{\xi \in \xi^*(1,0,0)} \{ L(\xi) > L(\xi^*) \} \right) = \mathbb{P}_{\xi^*} \left( \bigcap_{i \in \xi_M^*, j \in \xi_W^*} \{ L(\xi_{\text{row}}^{(i,j)}) > L(\xi^*) \} \right).$$

Recall that $\xi_M^* = [1 : \frac{n}{2}]$ and $\xi_W^* = [\frac{n}{2} + 1 : n]$. Let $r_1 = \frac{n}{\log^2 n}$, and $G_1 = [1 : 2r_1] \cup [\frac{n}{2} + 1 : \frac{n}{2} + 2r_1]$. Consider the random row graph $G_1$ that comprises $n$ user nodes. The following lemma shows that with high probability, the number of isolated user nodes in the set $G_1$ (i.e., the nodes that are not connected to any other nodes in $G_1$) is at least $3r_1$.

**Lemma 5.3.** With probability at least $1 - \exp\left(-\frac{\eta^2(a_1 + b_1)n}{\log^3 n}\right)$ for any $\eta \in (0, 1)$, the number of isolated nodes in $G_1$ is at least $3r_1$.

**Proof.** Let $N \triangleq 2\left(\frac{2r_1}{2}\right) = 4r_1^2 - 2r_1$ and $N' \triangleq 4r_1^2$, and

$$X \triangleq \sum_{i=1}^{N} X_i + \sum_{i=1}^{N'} Y_i$$

be the number of edges in $G_1$, where $\{X_i\}_{i=1}^{N}$ i.i.d. $\text{Bern}(\alpha_1)$, and $\{Y_i\}_{i=1}^{N'}$ i.i.d. $\text{Bern}(\beta_1)$. The number of non-isolated nodes is at most $2X$. Note that $\mathbb{E}(X) = Na_1 + N'\beta_1$, which lies in the interval $[3r_1^2(a_1 + \beta_1), 4r_1^2(a_1 + \beta_1)]$ for sufficiently large $n$. For any $\eta \in (0, 1)$, by the Chernoff bound, we have

$$\mathbb{P}\left(X \geq (1 + \eta) 4r_1^2(a_1 + \beta_1)\right) \leq \mathbb{P}\left(X \geq (1 + \eta) \mathbb{E}(X)\right) \leq \exp\left(-\frac{\eta^2 \mathbb{E}(X)}{3}\right) \leq \exp\left(-\frac{\eta^2(a_1 + b_1)n}{\log^3 n}\right).$$

Therefore, with probability at least $1 - \exp\left(-\frac{\eta^2(a_1 + b_1)n}{\log^3 n}\right)$, $X \leq (1 + \eta) 4r_1^2(a_1 + \beta_1) < r_1/2$, and the number of non-isolated nodes is at most $r_1$. \qed

Let $\Delta_1$ be the event that the number of isolated nodes in $G_1$ is at least $3r_1$. Conditioned on $\Delta_1$, Lemma 5.3 implies that one can find a subset $\xi_M^* \subset \xi_M^*$ and a subset $\xi_W^* \subset \xi_W^*$ such that $|\xi_M^*| = |\xi_W^*| = r_1$ and all the nodes in $\xi_M^* \cup \xi_W^*$ are not connected to one another. Then,
the term in (5.4) can be further bounded from above as
\[
P_{\xi^*} \left( \bigcap_{i \in \xi^*_1, j \in \xi^*_w} \left\{ L \left( \xi^*_w(i, j) > L(\xi^*) \right) \right\} \right)
\leq P_{\xi^*} \left( \bigcap_{i \in \xi^*_1, j \in \xi^*_Q_1} \left\{ L \left( \xi^*_w(i, j) > L(\xi^*) \right) \right\} \right)
= P_{\xi^*}(\Delta_1) P \left( \bigcap_{i \in \xi^*_1, j \in \xi^*_Q_1} \left\{ L \left( \xi^*_w(i, j) > L(\xi^*) \right) \bigg| \Delta_1 \right\} \right)
+ P_{\xi^*}(\Delta_1^c) P \left( \bigcap_{i \in \xi^*_1, j \in \xi^*_Q_1} \left\{ L \left( \xi^*_w(i, j) > L(\xi^*) \right) \bigg| \Delta_1^c \right\} \right),
\]
(5.5)

Lemma 5.4 (Adapted from Lemma 6, [4]). Conditioned on \(\xi^*\) and \(\Delta_1\), if \(L(\xi^*_w(i)) \leq L(\xi^*)\) and \(L(\xi^*_w(j)) \leq L(\xi^*)\) for some \(i \in \xi^*_1\) and \(j \in \xi^*_Q_1\), then \(L(\xi^*_w(i, j)) \leq L(\xi^*)\).

\textbf{Proof.} See [5, Appendix A-F].

Based on Lemma 5.4, one can bound the first term of (5.5) from above as
\[
P_{\xi^*} \left( \bigcap_{i \in \xi^*_1, j \in \xi^*_Q_1} \left\{ L \left( \xi^*_w(i, j) > L(\xi^*) \right) \bigg| \Delta_1 \right\} \right)
\]
(5.6)

where the decomposition in (5.6) is inspired by the proof technique in [5]. Without loss of generality, we assume \(1 \in \xi^*_1\) and \(\frac{3}{2} + 1 \in \xi^*_Q_1\). It is worth noting that conditioned on \(\Delta_1\) and \(\xi^*\), the events \(\left\{ L(\xi^*_w(i)) > L(\xi^*) \right\}\) for different \(i \in \xi^*_1\) are mutually independent, thus
\[
P_{\xi^*} \left( \bigcap_{i \in \xi^*_1} \left\{ L(\xi^*_w(i)) > L(\xi^*) \bigg| \Delta_1 \right\} \right) = P_{\xi^*} \left( L(\xi^*_w(1)) > L(\xi^*) \bigg| \Delta_1 \right)_{|\xi^*_1|}.
\]
Similarly, the events \(\left\{ L(\xi^*_w(j)) > L(\xi^*) \right\}\) for different \(j \in \xi^*_Q_1\) are also mutually independent, and
\[
P_{\xi^*} \left( \bigcap_{j \in \xi^*_Q_1} \left\{ L(\xi^*_w(j)) > L(\xi^*) \bigg| \Delta_1 \right\} \right) = P_{\xi^*} \left( L(\xi^*_w(\frac{3}{2} + 1)) > L(\xi^*) \bigg| \Delta_1 \right)_{|\xi^*_Q_1|}.
\]
Remark 5.5. The main purpose of introducing $\xi_{P_1}$ and $\xi_{Q_1}$ is to ensure that the events 
\(\{L(\xi_{\text{row}}^{(i)}) > L(\xi^{*})\}_{i \in \xi_{P_1}}\) are mutually independent and the events 
\(\{L(\xi_{\text{row}}^{(j)}) > L(\xi^{*})\}_{j \in \xi_{Q_1}}\) are mutually independent.

Moreover, by noting that $P_{\xi^{*}}(\Delta_1) \geq 1 - \exp\left(-\frac{n^2(a_1 + b_1)n}{\log^4 n}\right)$, one can establish the inequality below:

\[
P_{\xi^{*}}\left(L(\xi_{\text{row}}^{(1)}) > L(\xi^{*})\right) \\
= P_{\xi^{*}}\left(L(\xi_{\text{row}}^{(1)}) > L(\xi^{*})|\Delta_1\right) P_{\xi^{*}}(\Delta_1) + P_{\xi^{*}}\left(L(\xi_{\text{row}}^{(1)}) > L(\xi^{*})|\Delta_1^c\right) P_{\xi^{*}}(\Delta_1^c) \\
\geq \left(1 - \exp\left(-\frac{n^2(a_1 + b_1)n}{\log^4 n}\right)\right) \cdot P_{\xi^{*}}\left(L(\xi_{\text{row}}^{(1)}) > L(\xi^{*})|\Delta_1\right).
\]

In the following, it remains to provide a lower bound on $P_{\xi^{*}}\left(L(\xi_{\text{row}}^{(1)}) > L(\xi^{*})\right)$. Note that

\[
P_{\xi^{*}}\left(L(\xi_{\text{row}}^{(1)}) \leq L(\xi^{*})\right) \\
= P\left(\sum_{i=1}^{\frac{n}{2}-1} c(Y_i - X_i) + CY_2 + \sum_{U \in \{A,R\}} \sum_{(i,j) \in S_{U,L}} \log\left(1 - \frac{\theta_U}{\theta_U}\right) Z_{ij}(2\Theta_U^U - 1) \geq 0\right) \\
\geq P\left(\sum_{i=1}^{\frac{n}{2}-1} c(Y_i - X_i) + \sum_{U \in \{A,R\}} \sum_{i=1}^{\frac{m}{2}} \log\left(1 - \frac{\theta_U}{\theta_U}\right) Z_i(2\Theta_i^U - 1) \geq 0\right) \\
\geq \frac{1}{4} \exp\left\{-\left(\frac{n}{2} - 1\right)(1 + o(1))I_1 \log \frac{n}{I_1} - \sum_{U \in \{A,R\}} \frac{m}{2}(1 + o(1))p_U\right\},
\]

where inequality (5.7) follows from Lemma 5.1 by substituting $K = \frac{n}{2} - 1$ and $L_1 = L_2 = \frac{m}{2}$. Hence,

\[
P_{\xi^{*}}\left(L(\xi_{\text{row}}^{(1)}) > L(\xi^{*})\right) |_{\xi_{P_1}} \\
= \left(1 - P_{\xi^{*}}\left(L(\xi_{\text{row}}^{(1)}) \leq L(\xi^{*})\right)\right) |_{\xi_{P_1}} \\
\leq \exp\left\{-\frac{1}{4} e^{-\left(\frac{n}{2} - 1\right)(1 + o(1))I_1 \log \frac{n}{I_1} - \sum_{U \in \{A,R\}} \frac{m}{2}(1 + o(1))p_U}\right\} |_{\xi_{P_1}} \\
= \exp\left\{-\frac{1}{4} e^{-\left(1 - o(1)) \log n - \frac{1}{2}(1 + o(1))I_1(\log n) - \sum_{U \in \{A,R\}} \frac{m}{2}(1 + o(1))p_U}\right\} |_{\xi_{P_1}} \\
\leq \exp\left(-\frac{1}{4} n^{\frac{1}{2}}\right),
\]
where the last step follows from the facts that \( p < \frac{(2(1-\epsilon)-I_1)\log n}{(v, u, A + v, u, R) m} \) and \( |\xi_{P_1}| = r_1 \). Therefore,

\[
P_{\xi^*} \left( L(\xi_{row}^{(1)}) > L(\xi^*) | \Delta_1 \right) ) |^{1}_{\xi_{P_1}} \right) \leq \left( \frac{1}{1 - \exp \left(-\eta^2(a_1 + b_1)n \right)} \right) |^{1}_{\xi_{P_1}} \right) \leq 2 \exp \left(-\frac{1}{4} n^2 \right).
\]

Similarly, one can show that

\[
P_{\xi^*} \left( L(\xi_{row}^{(2)} > L(\xi^*) | \Delta_1 \right) ) |^{1}_{\xi_{P_1}} \right) \leq 2 \exp \left(-\frac{1}{4} n^2 \right).
\]

Combining the inequalities above, we conclude that when \( p < \frac{(2(1-\epsilon)-I_1)\log n}{(v, u, A + v, u, R) m} \) and \( n \) grows without bound,

\[
P_{\text{suc}} \leq P_{\xi^*} \left( L(\xi_{row}^{(1)}) > L(\xi^*) | \Delta_1 \right) ) |^{1}_{\xi_{P_1}} \right) + P_{\xi^*} \left( L(\xi_{row}^{(2)} > L(\xi^*) | \Delta_1 \right) ) |^{1}_{\xi_{P_1}} \right) + P_{\xi^*}(\Delta_1^c)
\]

which implies that the probability of success converges to zero.

### 5.3. Case 3: restricted to \( \Xi_{\xi^*}(0, 1, 0, 0) \)

Consider the ground truth tuple \( \xi^* \). For each \( i \in \xi^*_A \) and \( j \in \xi^*_R \), we define the tuple \( \xi^*_\text{col}(i) \) to be identical to \( \xi^* \) except that \( (\xi^*_\text{col}(i))_{A} = \xi^*_A \setminus \{i\} \) and \( (\xi^*_\text{col}(i))_{R} = \xi^*_R \cup \{i\} \), the tuple \( \xi^*_\text{col}(j) \) to be identical to \( \xi^* \) except that \( (\xi^*_\text{col}(j))_{A} = \xi^*_A \cup \{j\} \) and \( (\xi^*_\text{col}(j))_{R} = \xi^*_R \setminus \{j\} \), and the tuple \( \xi^*_\text{col}(i,j) \) to be identical to \( \xi^* \) except that the \( i \)-th movie and the \( j \)-th movie are swapped. The probability of success is bounded from above as

\[
(5.8) \quad P_{\text{suc}} \leq \mathbb{P}_{\xi^*} \left( \bigcap_{i \in \xi^*_A} \left\{ L(\xi^*) > L(\xi^{(i)}) \right\} \bigcap_{j \in \xi^*_R} \left\{ L(\xi^{(i,j)}) > L(\xi^*) \right\} \right).
\]

Recall that the ground truth \( \xi^* \) satisfies \( \xi^*_A = [1 : \frac{m}{2}] \) and \( \xi^*_R = [\frac{m}{2} + 1 : m] \). Let \( r_2 = \frac{m}{\log^2 m} \) and \( G_2 = [1 : 2r_2] \cup [\frac{m}{2} + 1 : \frac{m}{2} + 2r_2] \). Consider the random column graph \( G_2 \) that comprises \( m \) movie nodes. Similar to Lemma 5.3, one can show that with probability at least \( 1 - \exp \left(-\eta^2(a_2 + b_2)m \right) \) for any \( \eta \in (0, 1) \), the event \( \Delta_2 \) — the number of isolated movie nodes in the set \( G_2 \) (i.e., the nodes that are not connected to any other nodes in \( G_2 \)) is at least \( 3r_2 \) occurs. Conditioned on \( \Delta_2 \) and \( \xi^* \), one can find a subset \( \xi^*_{P_2} \subset \xi^*_A \) and a subset \( \xi^*_{Q_2} \subset \xi^*_R \) such that \( |\xi^*_{P_2}| = |\xi^*_{Q_2}| = r_2 \) and all the nodes in \( \xi^*_{P_2} \cup \xi^*_{Q_2} \) are not connected to one another.

#### 5.3.1. When \( \theta_A \neq \theta_R \)

Suppose \( p < \frac{(1(1-\epsilon)-I_2)\log n}{2(2, 4, A^2) m} \). Let \([\xi^*_{P_2}]_i \) and \([\xi^*_{Q_2}]_i \) respectively be the \( i \)-th elements of \( \xi^*_{P_2} \) and \( \xi^*_{Q_2} \), where \( i \in [1 : r_2] \). Let

\[
[\xi^*_{P_2,Q_2}] = \{(\xi^*_{P_2})_i, (\xi^*_{Q_2})_i\}_{i=1}^{r_2}
\]
and note that $|\xi_{P_2, Q_2}| = r_2$. Then, the term in (5.8) can be further bounded from above as

$$
P_{\xi^*} \left( \bigcap_{(i,j) \in \xi_{P_2, Q_2}} \left\{ L \left( \xi_{col}^{(i,j)} \right) > L(\xi^*) \right\} \right)$$

$$\leq P_{\xi^*} \left( \bigcap_{(i,j) \in \xi_{P_2, Q_2}} \left\{ L \left( \xi_{col}^{(i,j)} \right) > L(\xi^*) \right\} \right)$$

$$\leq P_{\xi^*} \left( \bigcap_{(i,j) \in \xi_{P_2, Q_2}} \left\{ L \left( \xi_{col}^{(i,j)} \right) > L(\xi^*) \right\} \right) = P_{\xi^*} \left( \left\{ L \left( \xi_{col}^{(i, \frac{m}{2}+1)} \right) > L(\xi^*) \right\} \right)_{\Delta_2}^{\xi_{P_2, Q_2}}.$$  

(5.9)

Without loss of generality, we assume $(1, \frac{m}{2}+1) \in \xi_{P_2, Q_2}$. The key observation is that conditioned on $\Delta_2$ and $\xi^*$, the events $\left\{ L(\xi_{col}^{(i,j)}) > L(\xi^*) \right\}$ for different $(i, j) \in \xi_{P_2, Q_2}$ are mutually independent, thus

$$P_{\xi^*} \left( \bigcap_{(i,j) \in \xi_{P_2, Q_2}} \left\{ L \left( \xi_{col}^{(i,j)} \right) > L(\xi^*) \right\} \right) = P_{\xi^*} \left( \left\{ L \left( \xi_{col}^{(i, \frac{m}{2}+1)} \right) > L(\xi^*) \right\} \right)_{\Delta_2}^{\xi_{P_2, Q_2}}.$$  

Since $k_2 = 1$, the parameters $t_{AR}$ and $t_{R,A}$ can be chosen to equal either 0 or 1, but in the following we consider the scenario in which both $t_{AR}$ and $t_{R,A}$ equal zero. Note that

$$P_{\xi^*} \left( L(\xi_{col}^{(1, \frac{m}{2}+1)}) \leq L(\xi^*) \right)$$

$$= \mathbb{P} \left( d \sum_{i=1}^{m-2} (Y_i' - X_i') + \sum_{(i,j) \in S_{\xi_{P_2, Q_2}}} \log \left( \frac{1 - \theta_A}{1 - \theta_R} \right) \frac{\theta_R}{\theta_A} Z_{ij} (\Theta_i A - 1) - \log \left( \frac{\theta_A}{\theta_R} \right) Z_{ij} \right)$$

$$+ \sum_{(i,j) \in S_{\xi_{P_2, Q_2}}} \log \left( \frac{1 - \theta_R}{1 - \theta_A} \right) \frac{\theta_A}{\theta_R} Z_{ij} (\Theta_i R - 1) - \log \left( \frac{\theta_R}{\theta_A} \right) Z_{ij} \geq 0$$

$$\geq \mathbb{P} \left( d \sum_{i=1}^{m-2} (Y_i' - X_i') + \sum_{i=1}^{n} \log \left( \frac{1 - \theta_A}{1 - \theta_R} \right) \frac{\theta_R}{\theta_A} Z_i (\Theta_i A - 1) - \log \left( \frac{\theta_A}{\theta_R} \right) Z_i \right)$$

$$+ \sum_{i=1}^{n} \log \left( \frac{1 - \theta_R}{1 - \theta_A} \right) \frac{\theta_A}{\theta_R} Z_i (\Theta_i R - 1) - \log \left( \frac{\theta_R}{\theta_A} \right) Z_i \geq 0$$

(5.10)$$\geq \frac{1}{4} \exp \left\{ -(1 + o(1)) I_2 (\log m) - (1 + o(1)) 2 np t_{AR} \right\},$$

where (5.10) follows from Lemma 5.2 by substituting $K = (m - 2)$ and $L_1 = L_2 = n$. The
Hence, for sufficiently large $n$, the probability of success is bounded from above as

$$P_{\text{suc}} \leq \mathbb{P}_{\xi^*} \left( L\left(\xi^{\ast}_{\text{col}}(\tfrac{m}{2} + 1) > L(\xi^*)\right) \right)^{\mathbb{E}}\left(\xi_{P_2},Q_2 \right) + \mathbb{P}_{\xi^*}(\Delta_2')$$

$$\leq \left( \frac{1}{1 - \exp\left(-\frac{n^2(a_0 + b_0)^2}{\log^* m}\right)} \right) \mathbb{P}_{\xi^*} \left( L\left(\xi^{\ast}_{\text{col}}(\tfrac{m}{2} + 1) > L(\xi^*)\right) \right)^{\mathbb{E}}\left(\xi_{P_2},Q_2 \right) + \mathbb{P}_{\xi^*}(\Delta_2')$$

$$\leq 3 \exp\left(-\frac{1}{4} m^2\right).$$

### 5.3.2. When $\theta_A = \theta_R$. Suppose $I_2 \leq 2(1 - \epsilon)$. Without loss of generality, we assume that $1 \in \xi_{P_2}$ and $\tfrac{m}{2} + 1 \in \xi_{Q_2}$. Similar to Lemma 5.4, we have the following inequalities:

$$\mathbb{P}_{\xi^*} \left( \bigcap_{i \in \xi_A, j \in \xi_R} \left\{ L\left(\xi^{(i,j)}_{\text{col}}\right) > L(\xi^*) \right\} \right)$$

$$= \mathbb{P}_{\xi^*} \left( \bigcap_{i \in \xi_{P_2}, j \in \xi_{Q_2}} \left\{ L\left(\xi^{(i,j)}_{\text{col}}\right) > L(\xi^*) \right\} \bigg| \Delta_2 \right)$$

$$\leq \mathbb{P}_{\xi^*} \left( \bigcap_{i \in \xi_{P_2}} \left\{ L\left(\xi^{(i)}_{\text{col}}\right) > L(\xi^*) \right\} \bigg| \Delta_2 \right) + \mathbb{P}_{\xi^*} \left( \bigcap_{j \in \xi_{Q_2}} \left\{ L\left(\xi^{(j)}_{\text{col}}\right) > L(\xi^*) \right\} \bigg| \Delta_2 \right)$$

$$\leq \mathbb{P}_{\xi^*} \left( L\left(\xi^{(1)}_{\text{col}}\right) > L(\xi^*) \bigg| \Delta_2 \right) + \mathbb{P}_{\xi^*} \left( L\left(\xi^{(\tfrac{m}{2} + 1)}_{\text{col}}\right) > L(\xi^*) \bigg| \Delta_2 \right),$$

where (5.13) holds since conditioned on $\Delta_2$ and $\xi^*$, the events $\{ L\left(\xi^{(i)}_{\text{col}}\right) > L(\xi^*) \}$ for different $i \in \xi_{P_2}$ are mutually independent, and the events $\{ L\left(\xi^{(j)}_{\text{col}}\right) > L(\xi^*) \}$ for different $j \in \xi_{Q_2}$ are also mutually independent. By applying Lemma 5.1 and noting that $I_2 < 2(1 - \epsilon)$, we have

$$\mathbb{P}_{\xi^*} \left( L\left(\xi^{(1)}_{\text{col}}\right) \leq L(\xi^*) \right) = \mathbb{P} \left( \sum_{i=1}^{\frac{m}{2} - 1} (Y_i' - X_i') + dY_{\frac{m}{2}}' \geq 0 \right)$$

$$\geq \mathbb{P} \left( \sum_{i=1}^{\frac{m}{2} - 1} (Y_i' - X_i') \geq 0 \right) \geq \frac{1}{4} e^{-\frac{1}{2}(1+o(1))I_2(\log m)},$$
and

\begin{equation}
\mathbb{P}_{\xi^*} \left( L(\xi_{\text{col}}^{(1)}(\xi^*) > L(\xi^*)) \right) = 1 - \mathbb{P}_{\xi^*} \left( L(\xi_{\text{col}}^{(1)}(\xi^*) \leq L(\xi^*)) \right) \leq \exp \left( -\frac{1}{4} m \right).
\end{equation}

Similarly,

\begin{equation}
\mathbb{P}_{\xi^*} \left( L(\xi_{\text{col}}^{(m+1)}(\xi^*) > L(\xi^*)) \right) \leq \exp \left( -\frac{1}{4} m \right).
\end{equation}

By combining (5.13), (5.14), (5.15), and the fact that \( \Delta_2 \) occurs with probability at least \( 1 - \exp \left( -\frac{\eta^2(a^2+b^2)m}{\log^3 m} \right) \) for any \( \eta \in (0,1) \), one can eventually conclude that the probability of success is at most \( 5 \exp \left( -\frac{1}{4} m \right) \), which completes the proof.

Remark 5.6. Note that the above analysis for \( \theta_A \neq \theta_R \) is sub-optimal—the number of events that are likeliest to cause errors is \( |\xi^*_P| \times |\xi^*_Q| = O \left( \frac{m^2}{(\log m)^4} \right) \); however, among them only \( O(m/(\log m)^2) \) independent events are extracted to \( \xi^*_P,Q \), as shown in equation (5.9). Hence, a factor of two is lost in the converse part. Furthermore, due to the fact that \( \theta_A \) and \( \theta_R \) are unequal, the approach used for \( \theta_A = \theta_R \) and Subsection 5.2 (i.e., split \( \mathbb{P}_{\xi^*} \left( \bigcap_{i \in \xi^*_P,j \in \xi^*_Q} \{ L(\xi_{\text{col}}^{(i,j)}(\xi^*) > L(\xi^*) \} | \Delta_2 \right) \) into two individual terms as per (5.13)) does not yield a tight converse as well.

6. Conclusion and future directions. This paper investigates a novel community recovery problem based on a partially observed rating matrix and two-sided graph side-information. Our inner and outer bounds on the optimal sample probability quantifies the gains due to graph side-information; in particular, there exists a certain regime in which simultaneously observing two pieces of graph side-information is critical to reduce the optimal sample probability.

Finally, we put forth three promising directions for future work.

1. While the information-theoretic characterization in this work is optimal in a certain parameter regime and order-optimal in the remaining parameter regime, one would also expect that overcoming the challenge discussed in Remark 5.6 and establishing a sharp threshold by filling the small gap for the regime in which our bounds do not match would be a fruitful endeavour.

2. In addition to the fundamental limits, another direction that is worth exploring is the algorithmic developments and analyses for the problem as described in Section 2.

3. It would also be natural and interesting to investigate a more general setting. For instance, users’ ratings to movies may not necessarily be binary, and one may also assume that both users and movies form multiple unequal-sized communities, modelling different attributes of users (e.g., age, nationality) and movies (e.g., genre, language, duration).
Appendix A. Proof of Lemma 5.2.

Let

\[ A_i \triangleq d(Y'_i - X'_i), \text{ for } i \in [1 : K], \]
\[ B_j \triangleq \log \left( \frac{(1 - \theta_A) \theta_R}{(1 - \theta_R) \theta_A} \right) Z_i(\Theta'_j - 1) - \log \left( \frac{\theta_A}{\theta_R} \right) Z_j, \text{ for } j \in [1 : L_1], \]
\[ C_k \triangleq \log \left( \frac{(1 - \theta_R) \theta_A}{(1 - \theta_A) \theta_R} \right) Z'_k(\Theta'_k - 1) - \log \left( \frac{\theta_R}{\theta_A} \right) Z'_k, \text{ for } k \in [1 : L_2], \]

One can check that

\[
E(e^{\frac{1}{2}A_i}) = \left( \sqrt{\alpha_2 \beta_2} + \sqrt{(1 - \alpha_2)(1 - \beta_2)} \right)^2,
\]
\[
E(e^{\frac{1}{2}B_j}) = E(e^{\frac{1}{2}C_k}) = 1 - p \left( 1 - \sqrt{\theta_A \theta_R} - \sqrt{(1 - \theta_A)(1 - \theta_R)} \right),
\]

and

\[
- \log E(e^{\frac{1}{2}A_i}) = (1 + o(1))I_2 \frac{\log m}{m},
\]
\[
- \log E(e^{\frac{1}{2}B_j}) = - \log E(e^{\frac{1}{2}C_k}) = p \tau_{AR} + O(p^2).
\]

We further denote \( P_A(a) \triangleq P(A_i = a), P_B(b) \triangleq P(B_i = b), P_C(c) \triangleq P(C_i = c), \) and

\[ Z \triangleq \sum_{i=1}^K A_i + \sum_{j=1}^{L_1} B_j + \sum_{k=1}^{L_2} C_k, \]

then

\[ P(Z > 0) = \frac{\prod_{i=1}^K P_A(a_i) \prod_{j=1}^{L_1} P_B(b_j) \prod_{k=1}^{L_2} P_C(c_k)}{\sum_{a_i + \sum_j b_j + \sum_k c_k > 0} \frac{E(e^{\frac{1}{2}A_i})}{e^{\frac{1}{2}A_i}} \frac{E(e^{\frac{1}{2}B_j})}{e^{\frac{1}{2}B_j}} \frac{E(e^{\frac{1}{2}C_k})}{e^{\frac{1}{2}C_k}}}
\]

\[ \geq \sum_{\{a_i\}, \{b_j\}, \{c_k\}: \sum_i a_i + \sum_j b_j + \sum_k c_k \in (0, \nu)} \prod_{i=1}^K e^{\frac{1}{2}A_i} P_A(a_i) \prod_{j=1}^{L_1} e^{\frac{1}{2}B_j} P_B(b_j) \prod_{k=1}^{L_2} e^{\frac{1}{2}C_k} P_C(c_k) \frac{E(e^{\frac{1}{2}A_i})}{E(e^{\frac{1}{2}A_i})} \frac{E(e^{\frac{1}{2}B_j})}{E(e^{\frac{1}{2}B_j})} \frac{E(e^{\frac{1}{2}C_k})}{E(e^{\frac{1}{2}C_k})}
\]

\[ = \exp \left( -(1 + o(1))K I_2 \frac{\log m}{m} - (1 + o(1))(L_1 + L_2)p \tau_{AR} - \frac{1}{2} \nu \right).
\]

\[ \tag{A.1}
\]
By choosing \( \nu = \max\{K \frac{\log m}{m}, ((L_1 + L_2)p)^{\frac{1}{4}}\} \), we guarantee that the first part of (A.1) equals
\[
\exp\left( -(1 + o(1))KL_2 \frac{\log m}{m} - (1 + o(1))(L_1 + L_2)p \tau_{AR} \right).
\]

Let \( \{U_i\}_{i=1}^K \) be i.i.d. random variables distributed according to \( P_U(u) \triangleq e^{\frac{1}{2}u} P_A(u)/\mathbb{E}(e^{\frac{1}{2}A_i}) \), \( \{V_j\}_{j=1}^{L_1} \) be i.i.d. random variables distributed according to \( P_V(v) \triangleq e^{\frac{1}{2}v} P_B(v)/\mathbb{E}(e^{\frac{1}{2}B_i}) \), and \( \{W_k\}_{k=1}^{L_2} \) be i.i.d. random variables distributed according to \( P_W(w) \triangleq e^{\frac{3}{2}w} P_C(w)/\mathbb{E}(e^{\frac{3}{2}C_i}) \). Then, it remains to show that the second part of (A.1), which exactly equals \( \mathbb{P}(0 < \sum_{i=1}^K U_i + \sum_{j=1}^{L_1} V_j + \sum_{k=1}^{L_2} W_k < \nu) \), is at least \( \frac{1}{2} \). One may check that \( \sum_{i=1}^K U_i + \sum_{j=1}^{L_1} V_j + \sum_{k=1}^{L_2} W_k \) is a symmetric random variable with zero mean and variance \( \mathcal{O}(K \frac{\log m}{m} + (L_1 + L_2)p) \). By symmetry and the Chebyshev’s inequality, we have that \( \mathbb{P}(\sum_{i=1}^K U_i + \sum_{j=1}^{L_1} V_j + \sum_{k=1}^{L_2} W_k > 0) \) approaches \( \frac{1}{2} \) and \( \mathbb{P}(\sum_{i=1}^K U_i + \sum_{j=1}^{L_1} V_j + \sum_{k=1}^{L_2} W_k < \nu) \) approaches one asymptotically. This completes the proof.

REFERENCES

[1] E. Abbe, Community detection and stochastic block models: recent developments, The Journal of Machine Learning Research, 18 (2017), pp. 6446–6531.
[2] E. Abbe, A. S. Bandeira, and G. Hall, Exact recovery in the stochastic block model, IEEE Transactions on Information Theory, 62 (2015), pp. 471–487.
[3] E. Abbe and C. Sandon, Community detection in general stochastic block models: Fundamental limits and efficient algorithms for recovery, in IEEE 56th Annual Symposium on Foundations of Computer Science, 2015, pp. 670–688.
[4] K. Ahn, K. Lee, H. Cha, and C. Suh, Binary rating estimation with graph side information, in Advances in Neural Information Processing Systems, 2018, pp. 4272–4283.
[5] K. Ahn, K. Lee, H. Cha, and C. Suh, Binary rating estimation with graph side information: Supplementary material, 2018. Available at https://papers.nips.cc/paper/7681-binary-rating-estimation-with-graph-side-information.
[6] C. Aicher, A. Z. Jacobs, and A. Clauset, Adapting the stochastic block model to edge-weighted networks, arXiv preprint arXiv:1305.5782, (2013).
[7] D. Goldberg, D. Nichols, B. M. Oki, and D. Terry, Using collaborative filtering to weave an information tapestry, Communications of the ACM, 35 (1992), pp. 61–71.
[8] B. Hajek, Y. Wu, and J. Xu, Information limits for recovering a hidden community, IEEE Transactions on Information Theory, 63 (2017), pp. 4729–4745.
[9] P. W. Holland, K. B. Laskey, and S. Leinhardt, Stochastic blockmodels: First steps, Social networks, 5 (1983), pp. 109–137.
[10] M. Jamali and M. Ester, A matrix factorization technique with trust propagation for recommendation in social networks, in Proceedings of the fourth ACM conference on Recommender systems, 2010, pp. 135–142.
[11] V. Jog and P.-L. Loh, Information-theoretic bounds for exact recovery in weighted stochastic block models using the Renyi divergence, arXiv preprint arXiv:1509.06418, (2015).
[12] V. Kalofolias, X. Bresson, M. Bronstein, and P. Vandergheynst, Matrix completion on graphs, arXiv preprint arXiv:1408.1717, (2014).
[13] Y. Koren, R. Bell, and C. Volinsky, Matrix factorization techniques for recommender systems, Computer, (2009), pp. 30–37.
[14] C. E. Leiserson, R. L. Rivest, T. H. Cormen, and C. Stein, Introduction to algorithms, vol. 6, MIT press Cambridge, MA, 2001.
[15] H. Ma, D. Zhou, C. Liu, M. R. Lyu, and I. King, *Recommender systems with social regularization*, in Proceedings of the fourth ACM international conference on Web search and data mining, 2011, pp. 287–296.

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

[17] E. Mossel, J. Neeman, and A. Sly, *Reconstruction and estimation in the planted partition model*, Probability Theory and Related Fields, 162 (2015), pp. 431–461.

[18] N. Rao, H.-F. Yu, P. K. Ravikumar, and I. S. Dhillon, *Collaborative filtering with graph information: Consistency and scalable methods*, in Advances in Neural Information Processing Systems, 2015, pp. 2107–2115.

[19] H. Saad and A. Nosratinia, *Community detection with side information: Exact recovery under the stochastic block model*, IEEE Journal of Selected Topics in Signal Processing, 12 (2018), pp. 944–958.

[20] H. Saad and A. Nosratinia, *Exact recovery in community detection with continuous-valued side information*, IEEE Signal Processing Letters, 26 (2018), pp. 332–336.

[21] H. Saad and A. Nosratinia, *Recovering a single community with side information*, arXiv preprint arXiv:1809.01738, (2018).

[22] N. Srebro, J. Rennie, and T. S. Jaakkola, *Maximum-margin matrix factorization*, in Advances in Neural Information Processing Systems, 2005, pp. 1329–1336.

[23] J. Tang, X. Hu, and H. Liu, *Social recommendation: A review*, Social Network Analysis and Mining, 3 (2013), pp. 1113–1133.

[24] J. Xu, R. Wu, K. Zhu, B. Hajek, R. Srikant, and L. Ying, *Jointly clustering rows and columns of binary matrices: Algorithms and trade-offs*, in ACM SIGMETRICS Performance Evaluation Review, vol. 42, 2014, pp. 29–41.

[25] J. Yang, J. McAuley, and J. Leskovec, *Community detection in networks with node attributes*, in IEEE 13th International Conference on Data Mining, 2013, pp. 1151–1156.

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