Testing High-dimensional Multinomials with Applications to Text Analysis

T. Tony Cai[*], Zheng Tracy Ke[†], and Paxton Turner[†]

Abstract

Motivated by applications in text mining and discrete distribution inference, we investigate the testing for equality of probability mass functions of K groups of high-dimensional multinomial distributions. A test statistic, which is shown to have an asymptotic standard normal distribution under the null, is proposed. The optimal detection boundary is established, and the proposed test is shown to achieve this optimal detection boundary across the entire parameter space of interest. The proposed method is demonstrated in simulation studies and applied to analyze two real-world datasets to examine variation among consumer reviews of Amazon movies and diversity of statistical paper abstracts.

Keywords: authorship attribution, closeness testing, consumer reviews, martingale central limit theorem, minimax optimality, topic model

1 Introduction

Statistical inference for multinomial data has garnered considerable recent interest [Diakonikolas and Kane, 2016, Balakrishnan and Wasserman, 2018]. One important application is in text mining, as it is common to model the word counts in a text document by a multinomial distribution [Blei et al., 2003]. We consider a specific example in marketing, where the study of online customer ratings and reviews has become a trending topic [Chevalier and Mayzlin, 2006, Zhu and Zhang, 2010, Leung and Yang, 2020]. Customer reviews are a good proxy to the overall word of mouth (WOM) and can significantly influence customers' decisions [Zhu and Zhang, 2010]. Many research works aim to understand the patterns in online reviews and their impacts on sales. Classical studies only use the numerical ratings but ignore the rich text reviews because of their unstructured nature. More recent works have revealed the importance of analyzing text reviews [Chevalier and Mayzlin, 2006], especially for hedonic products such as books, movies, and hotels. A question of great interest is to detect the heterogeneity in reviewers' response styles. For example, Leung and Yang [2020] discovered that younger travelers, women, and travelers with less review expertise tend to give more positive reviews and that guests staying in

[*]University of Pennsylvania
[†]Harvard University

high-class hotels tend to have more extreme response styles than those staying in low-class hotels. Knowing such differences will offer valuable insights for hotel managers and online rating/review sites.

The aforementioned heterogeneity detection can be cast as a hypothesis test on multinomial data. Suppose reviews are written on a vocabulary of p distinct words. Let $X_i \in \mathbb{R}^p$ denote the word counts in review i. We model that

$$X_i \sim \text{Multinomial}(N_i, \Omega_i), \qquad 1 \leq i \leq n, \tag{1.1}$$

where N_i is the total length of review i and $\Omega_i \in \mathbb{R}^p$ is a probability mass function (PMF) containing the population word frequencies. These reviews are divided into K groups by reviewer characteristics (e.g., age, gender, new/returning customer), product characteristics (e.g., high-class versus low-class hotels), and numeric ratings (e.g., from 1 star to 5 stars), where K can be presumably large. We view Ω_i as representing the 'true response' of review i. The "average response" of a group k is defined by a weighted average of the PMFs:

$$\mu_k = (n_k \bar{N}_k)^{-1} \sum_{i \in S_k} N_i \Omega_i, \qquad 1 \leq k \leq K. \tag{1.2}$$

Here $S_k \subset \{1, 2, \ldots, n\}$ is the index set of group k, $n_k = |S_k|$ is the total number of reviews in group k, and $\bar{N}_k = n_k^{-1} \sum_{i \in S_k} N_i$ is the average length of reviews in group k. We would like to test

$$H_0: \quad \mu_1 = \mu_2 = \ldots = \mu_K. \tag{1.3}$$

When the null hypothesis is rejected, it means there exist statistically significant differences among the group-wise "average responses".

We call (1.1)-(1.3) the "K-sample testing for equality of average PMFs in multinomials" or "K-sample testing for multinomials" for short. Interestingly, as K varies, this problem includes several well-defined problems in text mining and discrete distribution inference as special cases.

1. *Global testing for topic models.* Topic modeling [Blei et al., 2003] is a popular text mining tool. In a topic model, each Ω_i in (1.1) is a convex combination of M topic vectors. Before fitting a topic model to a corpus, it is often desirable to determine if the corpus indeed contains multiple topics. This boils down to the global testing problem, which tests $M = 1$ versus $M > 1$. Under the null hypothesis, Ω_i's are equal to each other, and in the alternative hypothesis, Ω_i's can take continuous values in a high-dimensional simplex. This is a special case of our problem with $K = n$ and $n_k = 1$.

2. *Authorship attribution* [Mosteller and Wallace, 1963, Kipnis, 2022]. In these applications, the goal is to determine the unknown authorship of an article from other articles with known authors. A famous example [Mosteller and Wallace, 2012] is to determine the actual authors of a few Federalist Papers written by three authors but published under a single pseudonym. It can be formulated [Mosteller and Wallace, 1963, Kipnis, 2022] as testing the equality of population word frequencies between the

article of interest and the corpus from a known author, a special case of our problem with $K = 2$.

3. *Closeness between discrete distributions* [Chan et al., 2014, Bhattacharya and Valiant, 2015, Balakrishnan and Wasserman, 2019]. There has been a surge of interest in discrete distribution inference. Closeness testing is one of most studied problems. The data from two discrete distributions are summarized in two multinomial vectors Multinomial(N_1, μ) and Multinomial(N_2, θ). The goal is to test $\mu = \theta$. It is a special case of our testing problem with $K = 2$ and $n_1 = n_2 = 1$.

In this paper, we provide a unified solution to all the aforementioned problems. The key to our methodology is a flexible statistic called DELVE (DE-biased and Length-assisted Variability Estimator). It provides a general similarity measure for comparing groups of discrete distributions such as count vectors associated with text corpora. Similarity measures (such as the classical cosine similarity, log-likelihood ratio statistic, and others) are fundamental in text mining and have been applied to problems in distribution testing [Kim et al., 2022], computational linguistics [Gomaa et al., 2013], econometrics [Hansen et al., 2018], and computational biology [Kolodziejczyk et al., 2015]. Our method is a new and flexible similarity measure that is potentially useful in these areas.

We emphasize that our setting does not require that the X_i's in the same group are drawn from the same distribution. Under the null hypothesis (1.3), the group-wise means are equal, but the Ω_i's within each group can still be different from each other. As a result, the null hypothesis is composite and designing a proper test statistic is non-trivial.

1.1 Our results and contributions

The dimensionality of the testing problem is captured by (n, p, K) and $\bar{N} := n^{-1} \sum_{i=1}^{n} N_i$. We are interested in a high-dimensional setting where

$$n\bar{N} \to \infty, \quad p \to \infty, \quad \text{and} \quad n^2\bar{N}^2/(Kp) \to \infty. \tag{1.4}$$

In most places of this paper, we use a subscript n to indicate asymptotics, but our method and theory do apply to the case where n is finite and $\bar{N} \to \infty$. In text applications, $n\bar{N}$ is the total count of words in the corpus, and a large $n\bar{N}$ means either there are sufficiently many documents, or the documents are sufficiently long. Given that $n\bar{N} \to \infty$, we further allow (p, K) to grow with n at a speed such that $Kp \ll n^2\bar{N}^2$. In particular, our settings allow K to range from 2 to n, so as to cover all the application examples.

We propose a test that enjoys the following properties:

(a) *Parameter-free null distribution*: We show that the test statistic $\psi \to N(0, 1)$ under H_0. Even under the null hypothesis (1.3), the model contains a large number of free parameters because the null hypothesis is only about the equality of "average" PMFs but still allows (N_i, Ω_i) to differ within each group. As an appealing property, the null distribution of ψ does not depend on these individual multinomial parameters; hence, we can always conveniently obtain the asymptotic p-value for our proposed test.

(b) *Minimax optimal detection boundary*: We define a quantity $\omega_n := \omega_n(\mu_1, \mu_2, \ldots, \mu_K)$ in (3.5) that measures the difference among the K group-wise mean PMF's. It satisfies that $\omega_n = 0$ if and only if the null hypothesis holds, and it has been properly normalized so that ω_n is bounded under the alternative hypothesis (provided some mild regularity conditions hold). We show that the proposed test has an asymptotic full power if $\omega_n^4 n^2 \bar{N}^2/(Kp) \to \infty$. We also provide a matching lower bound by showing that the null hypothesis and the alternative hypothesis are asymptotically indistinguishable if $\omega_n^4 n^2 \bar{N}^2/(Kp) \to 0$. Therefore, the proposed test is minimax optimal. Furthermore, in the boundary case where $\omega_n^4 n^2 \bar{N}^2/(Kp) \to c_0$ for a constant $c_0 > 0$, for some special settings, we show that $\psi \to N(0,1)$ under H_0, and $\psi \to N(c_1, 1)$, under H_1, with the constant c_1 being an explicit function of c_0.

To the best of our knowledge, this testing problem for a general K has not been studied before. The existing works primarily focused on closeness testing and authorship attribution (see Section 1.2), which are special cases with $K = 2$. In comparison, our test is applicable to any value of K, offering a unified solution to multiple applications. Even for $K = 2$, the existing works do not provide a test statistic that has a tractable null distribution. They determined the rejection region and calculated p-values using either a (conservative) large-deviation bound or a permutation procedure. Our test is the first one equipped with a tractable null distribution. Our results about the optimal detection boundary for a general K are also new to the literature. By varying K in our theory, we obtain the optimal detection boundary for different sub-problems. For some of them (e.g., global testing for topic models, authorship attribution with moderate sparsity), the optimal detection boundary was not known before; hence, our results help advance the understanding of the statistical limits of these problems.

1.2 Related literature

First, we make a connection to discrete distribution inference. Let $X \sim \text{Multinomial}(N, \Omega)$ represent a size-N sample from a discrete distribution with p categories. The one-sample closeness testing aims to test $H_0 : \Omega = \mu$, for a given PMF μ. Existing works focus on finding the minimum separation condition in terms of the ℓ^1-norm or ℓ^2-norm of $\Omega - \mu$. Balakrishnan and Wasserman [2019] derived the minimum ℓ^1-separation condition and proposed a truncated chi-square test to achieve it. Valiant and Valiant [2017] studied the "local critical radius", a local separation condition that depends on the "effective sparsity" of μ, and they proposed a "2/3rd + tail" test to achieve it. In the two-sample closeness testing problem, given $X_1 \sim \text{Multinomial}(N_1, \Omega_1)$ and $X_2 \sim \text{Multinomial}(N_2, \Omega_2)$, it aims to test $H_0 : \Omega_1 = \Omega_2$. Again, this literature focuses on finding the minimum separation condition in terms of the ℓ^1-norm or ℓ^2-norm of $\Omega_1 - \Omega_2$. When $N_1 = N_2$, Chan et al. [2014] derived the minimum ℓ^1-separation condition and proposed a weighted chi-square test to attain it. Bhattacharya and Valiant [2015] extended their results to the unbalanced case where $N_1 \neq N_2$, assuming $\|\Omega_1 - \Omega_2\|_1 \geq p^{-1/12}$. This assumption was later removed by Diakonikolas and Kane [2016], who established the minimum ℓ^1-separation condition in full generality. Kim et al. [2022] proposed a two-sample kernel U-statistic and showed that it attains the minimum ℓ^2-separation condition.

Since the two-sample closeness testing is a special case of our problem with $K = 2$ and $n_1 = n_2 = 1$, our test is directly applicable. An appealing property of our test is its tractable asymptotic null distribution of $N(0, 1)$. In contrast, for the chi-square statistic in Chan et al. [2014] or the U-statistic in [Kim et al., 2022], the rejection region is determined by either an upper bound from concentration inequalities or a permutation procedure, which may lead to a conservative threshold or need additional computational costs. Regarding the testing power, we show in Section 4.3 that our test achieves the minimum ℓ^2-separation condition, i.e., our method is an optimal "ℓ^2 testor." Our test can also be turned into an optimal "ℓ^1 testor" (a test that achieves the minimum ℓ^1-separation condition) by re-weighting terms in the test statistic (see Section 4.3).

Next, we make a connection to text mining. In this literature, a multinomial vector $X \sim \text{Multinomial}(N, \Omega)$ represents the word counts for a document of length N written with a dictionary containing p words. In a topic model, each Ω_i is a convex combination of M "topic vectors": $\Omega_i = \sum_{k=1}^{M} w_i(k) A_k$, where each $A_k \in \mathbb{R}^p$ is a PMF and the combination coefficient vector $w_i \in \mathbb{R}^K$ is called the "topic weight" vector for document i. Given a collection of documents $X_1, X_2, \ldots, X_n$, the global testing problem aims to test $M = 1$ versus $M > 1$. Interestingly, the optimal detection boundary for this problem has never been rigorously studied. As we have explained, this problem a special case of our testing problem with $K = n$. Our results (a) provide a test statistic that has a tractable null distribution and (b) reveal that the optimal detection boundary is $\omega_n^2 \asymp (\sqrt{n}\bar{N})^{-1}\sqrt{p}$. Both (a) and (b) are new results. When comparing our results with those about estimation of A_k's [Ke and Wang, 2022], it suggests that global testing requires a strictly lower signal strength than topic estimation.

For authorship attribution, Kipnis [2022] treats the corpus from a known author as a single document and tests the null hypothesis that this combined document and a new document have the same population word frequencies. It is a two-sample closeness testing problem, except that sparsity is imposed on the difference of two PMFs. Kipnis [2022] proposed a test which applies an "exact binomial test" to obtain a p-value for each word and combines these p-values using Higher Criticism [Donoho and Jin, 2004]. Donoho and Kipnis [2022] analyzed this test when the number of "useful words" is $o(\sqrt{p})$, and they derived a sharp phase diagram (a related one-sample setting was studied in Arias-Castro and Wang [2015]). In Section 4.2, we show that our test is applicable to this problem and has some nice properties: (a) tractable null distribution; (b) allows for $s \geq c\sqrt{p}$, where s is the number of useful words; and (c) does not require documents from the known author to have identical population word frequencies, making the setting more realistic. On the other hand, when $s = o(\sqrt{p})$, our test is less powerful than the one in Kipnis [2022], Donoho and Kipnis [2022], as our test does not utilize sparsity explicitly. We can further improve our test in this regime by modifying the DELVE statistic to incorporate sparsity (see the remark in Section 4.2).

1.3 Organization

The rest of this paper is arranged as follows. In Section 2, we introduce the test statistic and explain the rationale behind it. We then present in Section 3 the main theoretical results, including the asymptotic null distribution, power analysis, a matching lower bound, the study of two special cases ($K = n$ and $K = 2$), and a discussion of the contiguity regime. Section 4 applies our results to text mining and discrete distribution testing. Simulations are in Section 5 and real data analysis is in Section 6. The paper is concluded with a discussion in Section 7. All proofs are in the appendix.

2 The DELVE Test

Recall that $X_i \sim \text{Multinomial}(N_i, \Omega_i)$ for $1 \leq i \leq n$. There is a known partition $\{1, 2, \ldots, n\} = \cup_{k=1}^K S_k$. Write $n_k = |S_k|$, $\bar{N}_k = n_k^{-1} \sum_{i \in S_k} N_i$, and $\bar{N} = n^{-1} \sum_{i=1}^n N_i$. In (1.2), we have defined the group-wise mean PMF $\mu_k = (n_k \bar{N}_k)^{-1} \sum_{i \in S_k} N_i \Omega_i$. We further define the overall mean PMF $\mu \in \mathbb{R}^p$ by

$$\mu := \frac{1}{n\bar{N}} \sum_{k=1}^K n_k \bar{N}_k \mu_k = \frac{1}{n\bar{N}} \sum_{i=1}^n N_i \Omega_i. \tag{2.1}$$

We introduce a quantity $\rho^2 = \rho^2(\mu_1, \ldots, \mu_K)$ by

$$\rho^2 := \sum_{k=1}^K n_k \bar{N}_k \|\mu_k - \mu\|^2. \tag{2.2}$$

This quantity measures the variations across K group-wise mean PMFs. It is true that the null hypothesis (1.3) holds if and only if $\rho^2 = 0$. Inspired by this observation, we hope to construct an unbiased estimator of ρ^2 and develop it to a test statistic.

We can easily obtain the minimum variance unbiased estimators of μ_k and μ:

$$\hat{\mu}_k = \frac{1}{n_k \bar{N}_k} \sum_{i \in S_k} X_i, \qquad \text{and} \qquad \hat{\mu} = \frac{1}{n\bar{N}} \sum_{k=1}^K n_k \bar{N}_k \hat{\mu}_k = \frac{1}{n\bar{N}} \sum_{i=1}^n X_i. \tag{2.3}$$

For each $1 \leq j \leq p$, let μ_{kj}, μ_j, $\hat{\mu}_{kj}$ and $\hat{\mu}_j$ represent the jth entry of μ_k, μ, $\hat{\mu}_k$ and $\hat{\mu}$, respectively. A naive estimator of ρ^2 is

$$\widetilde{T} = \sum_{j=1}^p \widetilde{T}_j, \qquad \text{where} \quad \widetilde{T}_j = \sum_{k=1}^K n_k \bar{N}_k (\hat{\mu}_{kj} - \hat{\mu}_j)^2. \tag{2.4}$$

This estimator is biased. In Section C.1 of the appendix , we show that $\mathbb{E}[\widetilde{T}_j] = \sum_{k=1}^K \big[n_k \bar{N}_k (\mu_{kj} - \mu_j)^2 + \big(\frac{1}{n_k \bar{N}_k} - \frac{1}{n\bar{N}} \big) \sum_{i \in S_k} N_i \Omega_{ij} (1 - \Omega_{ij}) \big]$. It motivates us to debias $\widetilde{T}_j$ by using an unbiased estimate of $\Omega_{ij}(1 - \Omega_{ij})$. By elementary properties of the multinomial distribution, $\mathbb{E}[X_{ij}(N_i - X_{ij})] = N_i(N_i - 1)\Omega_{ij}(1 - \Omega_{ij})$. We thereby use $\frac{1}{N_i(N_i-1)} X_{ij}(N_i - X_{ij})$ to estimate $\Omega_{ij}(1 - \Omega_{ij})$. This gives rise to an unbiased estimator of ρ^2 as

$$T = \sum_{j=1}^p T_j, \quad T_j = \sum_{k=1}^K \Big[n_k \bar{N}_k (\hat{\mu}_{kj} - \hat{\mu}_j)^2 - \Big(\frac{1}{n_k \bar{N}_k} - \frac{1}{n\bar{N}} \Big) \sum_{i \in S_k} \frac{X_{ij}(N_i - X_{ij})}{N_i - 1} \Big]. \tag{2.5}$$

Lemma 2.1. *Under Models* (1.1)-(1.2)*, the estimator in* (2.5) *satisfies that* $\mathbb{E}[T] = \rho^2$*.*

To use T for hypothesis testing, we need a proper standardization of this statistic. In Sections A.1-A.2 of the appendix , we study $\mathbb{V}(T)$, the variance of T. Under mild regularity conditions, it can be shown that $\mathbb{V}(T) = \Theta_n \cdot [1 + o(1)]$, where

$$\Theta_n := 4\sum_{k=1}^{K}\sum_{j=1}^{p} n_k\bar{N}_k(\mu_{kj}-\mu_j)^2\mu_{kj} + 2\sum_{k=1}^{K}\sum_{i\in S_k}\sum_{j=1}^{p}\Big(\frac{1}{n_k\bar{N}_k}-\frac{1}{n\bar{N}}\Big)^2\frac{N_i^3}{N_i-1}\Omega_{ij}^2 \tag{2.6}$$
$$+\frac{2}{n^2\bar{N}^2}\sum_{1\leq k\neq \ell\leq K}\sum_{i\in S_k}\sum_{m\in S_\ell}\sum_{j=1}^{p} N_iN_m\Omega_{ij}\Omega_{mj} + 2\sum_{k=1}^{K}\sum_{\substack{i\in S_k, m\in S_k,\\ i\neq m}}\sum_{j=1}^{p}\Big(\frac{1}{n_k\bar{N}_k}-\frac{1}{n\bar{N}}\Big)^2 N_iN_m\Omega_{ij}\Omega_{mj}.$$

In Θ_n, the first term vanishes under the null, so it suffices to estimate the other three terms in Θ_n. By properties of multinomial distributions, $\mathbb{E}[X_{ij}X_{mj}] = N_iN_m\Omega_{ij}\Omega_{mj}$, $\mathbb{E}[X_{ij}^2] = N_i^2\Omega_{ij}^2 + N_i\Omega_{ij}(1-\Omega_{ij})$, and $\mathbb{E}[X_{ij}(N_i - X_{ij})] = N_i(N_i-1)\Omega_{ij}(1-\Omega_{ij})$. It inspires us to estimate $\Omega_{ij}\Omega_{mj}$ by $\frac{X_{ij}X_{mj}}{N_iN_m}$ and estimate Ω_{ij}^2 by $\frac{X_{ij}^2}{N_i^2} - \frac{X_{ij}(N_i-X_{ij})}{N_i^2(N_i-1)} = \frac{X_{ij}^2 - X_{ij}}{N_i(N_i-1)}$. Define

$$V = 2\sum_{k=1}^{K}\sum_{i\in S_k}\sum_{j=1}^{p}\Big(\frac{1}{n_k\bar{N}_k}-\frac{1}{n\bar{N}}\Big)^2\frac{X_{ij}^2-X_{ij}}{N_i(N_i-1)} + \frac{2}{n^2\bar{N}^2}\sum_{k\neq\ell}\sum_{i\in S_k}\sum_{m\in S_\ell}\sum_{j=1}^{p}X_{ij}X_{mj}$$
$$+2\sum_{k=1}^{K}\sum_{\substack{i\in S_k, m\in S_k,\\ i\neq m}}\sum_{j=1}^{p}\Big(\frac{1}{n_k\bar{N}_k}-\frac{1}{n\bar{N}}\Big)^2 X_{ij}X_{mj}. \tag{2.7}$$

The test statistic we propose is as follows (in the rate event $V < 0$, we simply set $\psi = 0$):

$$\psi = T/\sqrt{V}. \tag{2.8}$$

We call ψ the *DEbiased and Length-adjusted Variability Estimator (DELVE)*. In Section 3.1, we show that under mild regularity conditions, $\psi \to N(0,1)$ under the null hypothesis. For any fixed $\alpha \in (0,1)$, the asymptotic level-α DELVE test rejects H_0 if

$$\psi > z_\alpha, \qquad \text{where } z_\alpha \text{ is the } (1-\alpha)\text{-quantile of } N(0,1). \tag{2.9}$$

2.1 The special cases of $K = n$ and $K = 2$

As seen in Section 1, the application examples of $K = n$ and $K = 2$ are particularly intriguing. In these cases, we give more explicit expressions of our test statistic.

When $K = n$, we have $S_k = \{i\}$ and $\hat{\mu}_{kj} = N_i^{-1}X_{ij}$. The null hypothesis becomes $H_0 : \Omega_1 = \Omega_2 = \ldots = \Omega_n$. The statistic in (2.5) reduces to

$$T = \sum_{j=1}^{p}\sum_{i=1}^{n}\Big[\frac{(X_{ij}-N_i\hat{\mu}_j)^2}{N_i} - \Big(1-\frac{N_i}{n\bar{N}}\Big)\frac{X_{ij}(N_i-X_{ij})}{N_i(N_i-1)}\Big]. \tag{2.10}$$

Moreover, in the variance estimate (2.7), the last term is exactly zero, and it can be shown that the third term is negligible compared to the first term. We thereby consider a simpler variance estimator by only retaining the first term in (2.7):

$$V^* = 2\sum_{i=1}^n \sum_{j=1}^p \Big(\frac{1}{N_i} - \frac{1}{n\bar{N}}\Big)^2 \frac{X_{ij}^2 - X_{ij}}{N_i(N_i-1)}. \tag{2.11}$$

The simplified DELVE test statistic is $\psi^* = T/\sqrt{V^*}$.

When $K=2$, we observe two collections of multinomial vectors, denoted by $\{X_i\}_{1\leq i\leq n}$ and $\{G_i\}_{1\leq i\leq m}$. We assume for $1\leq i\leq n$ and $1\leq j\leq m$,

$$X_i \sim \text{Multinomial}(N_i, \Omega_i), \qquad G_j \sim \text{Multinomial}(M_j, \Gamma_j). \tag{2.12}$$

Write $\bar{N} = n^{-1}\sum_{i=1}^n N_i$ and $\bar{M} = m^{-1}\sum_{i=1}^m M_i$. The null hypothesis becomes

$$H_0: \quad \eta = \theta, \qquad \text{where } \eta = \frac{1}{n\bar{N}}\sum_{i=1}^n N_i\Omega_i, \text{ and } \theta = \frac{1}{m\bar{M}}\sum_{i=1}^m M_i\Gamma_i, \tag{2.13}$$

where θ and η are the two group-wise mean PMFs. We estimate them by $\hat{\eta} = (n\bar{N})^{-1}\sum_{i=1}^n X_i$ and $\hat{\theta} = (m\bar{M})^{-1}\sum_{i=1}^m G_i$. The statistic in (2.5) has an equivalent form as follows:

$$T = \frac{n\bar{N}m\bar{M}}{n\bar{N}+m\bar{M}}\Big[\|\hat{\eta}-\hat{\theta}\|^2 - \sum_{i=1}^n\sum_{j=1}^p \frac{X_{ij}(N_i - X_{ij})}{n^2\bar{N}^2(N_i-1)} - \sum_{i=1}^m\sum_{j=1}^p \frac{G_{ij}(M_i - G_{ij})}{m^2\bar{M}^2(M_i-1)}\Big]. \tag{2.14}$$

The variance estimate (2.7) has an equivalent form as follows:

$$\begin{aligned} V = {} & \frac{4\sum_{i=1}^n \sum_{i'=1}^m \sum_{j=1}^p X_{ij}G_{i'j}}{(n\bar{N}+m\bar{M})^2} + \frac{2m^2\bar{M}^2\big[\sum_{i=1}^n \frac{X_{ij}^2 - X_{ij}}{N_i(N_i-1)} + \sum_{1\leq i\neq i'\leq n} X_{ij}X_{i'j}\big]}{n^2\bar{N}^2(n\bar{N}+m\bar{M})^2} \\ & + \frac{2n^2\bar{N}^2\big[\sum_{i=1}^m \frac{G_{ij}^2 - G_{ij}}{M_i(M_i-1)} + \sum_{1\leq i\neq i'\leq m} G_{ij}G_{i'j}\big]}{m^2\bar{M}^2(n\bar{N}+m\bar{M})^2}. \end{aligned} \tag{2.15}$$

The DELVE test statistic is $\psi = T/\sqrt{V}$.

2.2 A variant: DELVE+

We introduce a variant of the DELVE test statistic to better suit real data. Let $\hat{\mu}$, T and V be as in (2.3), (2.5) and (2.7). Define

$$\psi^+ = T/\sqrt{V^+}, \qquad \text{where} \quad V^+ = V\cdot\big(1 + \|\hat{\mu}\|_2 T/\sqrt{V}\big). \tag{2.16}$$

We call (2.16) the DELVE+ test statistic. In theory, this modification has little effect on the key properties of the test. To see this, we note that $\|\hat{\mu}\|_2 = o_{\mathbb{P}}(1)$ in high-dimensional settings. Suppose $T/\sqrt{V} \to N(0,1)$ under H_0. Since $\|\hat{\mu}\|_2 \to 0$, it is seen immediately that $V^+/V \to 1$; hence, the asymptotic normality also holds for ψ^+. Suppose $T/\sqrt{V}\to\infty$ under the alternative hypothesis. It follows that $V^+ \leq 2\max\{V, \|\hat{\mu}\|_2\cdot T\sqrt{V}\}$ and $\psi^+ \geq \frac{1}{\sqrt{2}}\min\{T/\sqrt{V},\, \|\hat{\mu}\|_2^{-1}(T/\sqrt{V})^{1/2}\}\to\infty$. We have proved the following lemma:

Lemma 2.2. *As $n\bar{N}\to\infty$, suppose $\|\hat{\mu}\|_2 \to 0$ in probability. Under H_0, if $T/\sqrt{V} \to N(0,1)$, then $T/\sqrt{V^+}\to N(0,1)$. Under H_1, if $T/\sqrt{V}\to\infty$, then $T/\sqrt{V^+}\to\infty$.*

In practice, this modification avoids extremely small p-values. In some real datasets, V is very small and leads to an extremely small p-value in the original DELVE test. In DELVE+, as long as T is positive, ψ^+ is smaller than ψ, so that the p-value is adjusted.

In the numerical experiments, we consider both DELVE and DELVE+. For theoretical analysis, since these two versions have almost identical theoretical properties, we only focus on the original DELVE test statistic.

3 Theoretical Properties

We first present the regularity conditions. For a constant $c_0 \in (0,1)$, we assume

$$\min_{1\leq i\leq n} N_i \geq 2, \qquad \max_{1\leq i\leq n} \|\Omega_i\|_\infty \leq 1-c_0, \qquad \max_{1\leq k\leq K} \frac{n_k\bar{N}_k}{n\bar{N}} \leq 1-c_0. \tag{3.1}$$

In (3.1), the first condition is mild. The second condition is also mild: note that $\|\Omega_i\|_1 = 1$ for each i; this condition excludes those cases where one of the p categories has an extremely dominating probability in the PMF Ω_i. In the third condition, $n_k\bar{N}_k$ is the total number of counts in all multinomials of group k, and this condition excludes the extremely unbalanced case where one group occupies the majority of counts. Note that in the special case of $K=2$, we relax this condition to allow for severely unbalanced groups (see Section 3.4).

Recall that $\mu_k = \frac{1}{n_k\bar{N}_k}\sum_{i\in S_k} N_i\Omega_i$ is the mean PMF within group k. We also define a 'covariance' matrix of PMF's for group k by $\Sigma_k = \frac{1}{n_k\bar{N}_k}\sum_{i\in S_k} N_i\Omega_i\Omega_i'$. Let

$$\alpha_n := \max\left\{\sum_{k=1}^K \frac{\|\mu_k\|_3^3}{n_k\bar{N}_k}, \quad \sum_{k=1}^K \frac{\|\mu_k\|^2}{n_k^2\bar{N}_k^2}\right\} \Big/ \left(\sum_{k=1}^K \|\mu_k\|^2\right)^2, \tag{3.2}$$

and

$$\beta_n := \max\left\{\sum_{k=1}^K\sum_{i\in S_k} \frac{N_i^2}{n_k^2\bar{N}_k^2}\|\Omega_i\|_3^3, \quad \sum_{k=1}^K \|\Sigma_k\|_F^2\right\} \Big/ (K\|\mu\|^2). \tag{3.3}$$

We assume that as $n\bar{N}\to\infty$,

$$\alpha_n = o(1), \qquad \beta_n = o(1), \qquad \text{and} \quad \frac{\|\mu\|_4^4}{K\|\mu\|^4} = o(1). \tag{3.4}$$

Here α_n and β_n only depend on group-wise quantities, such as μ_k, Σ_k and $\sum_{i\in S_k} N_i^2\|\Omega_i\|_3^3$; hence, a small number of 'outliers' (i.e., extremely large entries) in Ω has little effect on α_n and β_n. Furthermore, in a simple case where $\max_k n_k \leq C\min_k n_k$, $\max_k \bar{N}_k \leq C\min_k \bar{N}_k$ and $\|\Omega\|_{\max} = O(1/p)$, it holds that $\alpha_n = O(\max\{\frac{1}{n\bar{N}}, \frac{Kp}{n^2\bar{N}^2}\})$, $\beta_n = O(\max\{\frac{K^2}{n^2p}, \frac{1}{p}\})$ and $\frac{\|\mu\|_4^4}{K\|\mu\|^4} = O(\frac{1}{Kp})$. When $n\bar{N}\to\infty$ and $p\to\infty$, (3.4) reduces to $n^2\bar{N}^2/(Kp)\to\infty$. This condition is necessary for successful testing, because our lower bound in Section 3.3 implies that the two hypotheses are asymptotically indistinguishable if $n^2\bar{N}^2/(Kp)\to 0$.

3.1 The asymptotic null distribution

Under the null hypothesis, the K group-wise mean PMF's $\mu_1, \mu_2, \ldots, \mu_K$, are equal to each other, but this hypothesis is still highly composite, as (N_i, Ω_i) are not necessarily the same within each group. We show that the DELVE test statistic always enjoys a parameter-free asymptotic null distribution. Let T, Θ_n and V be as in (2.5)-(2.7). The next two theorems are proved in the appendix.

Theorem 3.1. *Consider Models* (1.1)-(1.2)*, where the null hypothesis* (1.3) *holds. Suppose* (3.1) *and* (3.4) *are satisfied. As* $n\bar{N} \to \infty$, $T/\sqrt{\Theta_n} \to N(0,1)$ *in distribution.*

Theorem 3.2. *Under the conditions of Theorem 3.1, as* $n\bar{N} \to \infty$, $V/\Theta_n \to 1$ *in probability, and* $\psi := T/\sqrt{V} \to N(0,1)$ *in distribution.*

By Theorem 3.2, the asymptotic p-value is computed via $1 - \Phi(\psi)$, where $\Phi(\cdot)$ is the cumulative distribution function of the standard normal. Moreover, for any fixed $\alpha \in (0,1)$, the rejection region of the asymptotic level-α test is as given in (2.9).

The proofs of Theorems 3.1-3.2 contain two key steps: in the first step, we decompose T into the sum of mutually uncorrelated terms. We introduce a set of independent, mean-zero random vectors $\{Z_{ir}\}_{1\le i\le n, 1\le r\le N_i}$, where $Z_{ir} \sim \text{Multinomial}(1, \Omega_i) - \Omega_i$. By properties of multinomial distributions, $X_i = N_i\Omega_i + \sum_{r=1}^{N_i} Z_{ir}$ in distribution. We plug it into (2.5) to obtain $T = T_1 + T_2 + T_3 + T_4$, where T_1 is a linear form of $\{Z_{ir}\}$, T_2, T_3 and T_4 are quadratic forms of $\{Z_{ir}\}$, and the four terms are uncorrelated with each other (details are contained in Section A of the appendix). In the second step, we construct a martingale for each term T_j. This is accomplished by rearranging the double-index sequence Z_{ir} to a single-index sequence and then successively adding terms in this sequence to T_j. We then apply the martingale central limit theorem (CLT) [Hall and Heyde, 2014] to prove the asymptotic normality of each T_j. The asymptotic normality of T follows by identifying the dominating terms in T_1-T_4 (as model parameters change, the dominating terms can be different) and studying their joint distribution. This step involves extensive calculations to bound the conditional variance and to verify the Lindeberg conditions of the martingale CLT, as well as numerous subtle uses of the Cauchy-Schwarz inequality to simplify the moment bounds.

3.2 Power analysis

Under the alternative hypothesis, the PMF's $\mu_1, \mu_2, \ldots, \mu_K$ are not the same. In Section 2, we introduce a quantity ρ^2 (see (2.2)) to capture the total variation in μ_k's, but this quantity is not scale-free. We define a scaled version of ρ^2 as

$$\omega_n = \omega_n(\mu_1, \mu_2, \ldots, \mu_K) := \frac{1}{n\bar{N}\|\mu\|^2} \sum_{k=1}^{K} n_k \bar{N}_k \|\mu_k - \mu\|^2. \tag{3.5}$$

It is seen that $\omega_n \le \max_k\{\frac{\|\mu_k-\mu\|^2}{\|\mu\|^2}\}$, which is properly scaled.

Theorem 3.3. *Consider Models* (1.1)-(1.2)*, where* (3.1) *and* (3.4) *are satisfied. Then,* $\mathbb{E}[T] = n\bar{N}\|\mu\|^2\omega_n^2$, *and* $\mathbb{V}(T) = O\big(\sum_{k=1}^{K} \|\mu_k\|^2\big) + \mathbb{E}[T] \cdot O\big(\max_{1\le k\le K} \|\mu_k\|_\infty\big)$.

For the DELVE test to have an asymptotically full power, we need $\mathbb{E}[T] \gg \sqrt{\mathbb{V}(T)}$. By Theorem 3.3, this is satisfied if $\mathbb{E}[T] \gg \sqrt{\sum_k \|\mu_k\|^2}$ and $\mathbb{E}[T] \gg \max_k \|\mu_k\|_\infty$. Between these two requirements, the latter one is weaker; hence, we only need $\mathbb{E}[T] \gg \sqrt{\sum_{k=1}^K \|\mu_k\|^2}$. It gives rise to the following theorem:

Theorem 3.4. *Under the conditions of Theorem 3.3, we further assume that under the alternative hypothesis, as $n\bar{N} \to \infty$,*

$$\mathrm{SNR}_n := \frac{n\bar{N}\|\mu\|^2\omega_n^2}{\sqrt{\sum_{k=1}^K \|\mu_k\|^2}} \quad \to \quad \infty. \tag{3.6}$$

The following statements are true. Under the alternative hypothesis, $\psi \to \infty$ in probability. For any fixed $\alpha \in (0,1)$, the level-α DELVE test has an asymptotic level of α and an asymptotic power of 1. If we choose $\alpha = \alpha_n$ such that $\alpha_n \to 0$ and $1 - \Phi(\mathrm{SNR}_n) = o(\alpha_n)$, where Φ is the CDF of $N(0,1)$, then the sum of type I and type II errors of the DELVE test converges to 0.

The detection boundary in (3.6) has simpler forms in some special cases. For example, if $\|\mu_k\| \asymp \|\mu\|$ for $1 \le k \le K$, then $\mathrm{SRN}_n \asymp n\bar{N}\omega_n^2\|\mu\|/\sqrt{K}$. If, furthermore, all entries of μ are at the same order, which implies $\|\mu\| \asymp p^{-1/2}$, then $\mathrm{SRN}_n \asymp n^2\bar{N}^2\omega_n^2/\sqrt{Kp}$. In this case, the detection boundary simplifies to $\omega_n^4 n^2\bar{N}^2/(Kp) \to \infty$.

Remark 1 (The low-dimensional case $p = O(1)$)**.** Although we are primarily interested in the high-dimensional setting $p \to \infty$, it is also worth investigating the case $p = O(1)$. We can show the same detection boundary for our test, but the asymptotic normality may not hold, because the variance estimator V in (2.7) is not guaranteed to be consistent. To fix this issue, we propose a variant of our test by replacing V with a refined variance estimator $\widetilde{V}$, which is consistent for a finite p. The expression of $\widetilde{V}$ is a little complicated. Due to space limits, we relegate it to Section E of the appendix.

3.3 A matching lower bound

We have seen that the DELVE test successfully separates two hypotheses if $\mathrm{SNR}_n \to \infty$, where SNR_n is as defined in (3.6). We now present a lower bound to show that the two hypotheses are asymptotically indistinguishable if $\mathrm{SNR}_n \to 0$.

Let $\ell_i \in \{1, 2, \dots, K\}$ denote the group label of X_i. Write $\xi = \{(N_i, \Omega_i, \ell_i)\}_{1 \le i \le n}$. Let μ_k, α_n, β_n, and ω_n be the same as defined in (1.2), (3.2), (3.3), and (3.5), respectively. For each given $(n, p, K, \bar{N})$, we write $\mu_k = \mu_k(\xi)$ to emphasize its dependence on parameters, and similarly for $\alpha_n, \beta_n, \omega_n$. For any $c_0 \in (0,1)$ and sequence ϵ_n, define

$$\mathcal{Q}_n(c_0, \epsilon_n) := \Big\{\xi = \{(N_i, \Omega_i, \ell_i)\}_{i=1}^n : \text{(3.1) holds for } c_0,\ \max(\alpha_n(\xi), \beta_n(\xi)) \le \epsilon_n\Big\} \tag{3.7}$$

Furthermore, for any sequence δ_n, we define a parameter class for the null hypothesis and a parameter class for the alternative hypothesis:

$$\mathcal{Q}_{0n}^*(c_0, \epsilon_n) = \mathcal{Q}_n(c_0, \epsilon_n) \cap \{\xi : \omega_n(\xi) = 0\},$$

$$\mathcal{Q}_{1n}^*(\delta_n; c_0, \epsilon_n) = \mathcal{Q}_n(c_0, \epsilon_n) \cap \left\{ \xi : \frac{n\bar{N}\|\mu(\xi)\|^2\omega_n^2(\xi)}{\sqrt{\sum_{k=1}^K \|\mu_k(\xi)\|^2}} \geq \delta_n \right\}. \tag{3.8}$$

Theorem 3.5. *Fix a constant* $c_0 \in (0,1)$ *and two positive sequences* ϵ_n *and* δ_n *such that* $\epsilon_n \to 0$ *as* $n \to \infty$. *For any sequence of* $(n, p, K, \bar{N})$ *indexed by* n, *we consider Models* (1.1)-(1.2) *for* $\Omega \in \mathcal{Q}_n(c_0, \epsilon_n)$. *Let* $\mathcal{Q}_{0n}^*(c_0, \epsilon_n)$ *and* $\mathcal{Q}_{1n}^*(\delta_n; c_0, \epsilon_n)$ *be as in* (3.8). *If* $\delta_n \to 0$, *then* $\limsup_{n\to\infty} \inf_{\Psi\in\{0,1\}} \big\{\sup_{\xi\in\mathcal{Q}_{0n}^*(c_0,\epsilon_n)} \mathbb{P}_\xi(\Psi = 1) + \sup_{\xi\in\mathcal{Q}_{1n}^*(\delta_n;c_0,\epsilon_n)} \mathbb{P}_\xi(\Psi = 0)\big\} = 1$.

By Theorem 3.5, the null and alternative hypotheses are asymptotically indistinguishable if $\mathrm{SRN}_n \to 0$. Combining it with Theorem 3.4, the DELVE test achieves the minimax optimal detection boundary.

3.4 The special case of $K = 2$

The special case of $K = 2$ is found in applications such as closeness testing and authorship attribution. We study this case more carefully. Given $\{X_i\}_{1\leq i\leq n}$ and $\{G_i\}_{1\leq i\leq m}$, we assume

$$X_i \sim \text{Multinomial}(N_i, \Omega_i), \qquad G_j \sim \text{Multinomial}(M_j, \Gamma_j). \tag{3.9}$$

Write $\bar{N} = n^{-1}\sum_{i=1}^n N_i$ and $\bar{M} = m^{-1}\sum_{i=1}^m M_i$. The null hypothesis becomes

$$H_0: \quad \eta = \theta, \qquad \text{where } \eta = \frac{1}{n\bar{N}}\sum_{i=1}^n N_i\Omega_i, \text{ and } \theta = \frac{1}{m\bar{M}}\sum_{i=1}^m M_i\Gamma_i, \tag{3.10}$$

where θ and η are the two group-wise mean PMFs. In this case, the test statistic ψ has a more explicit form as in (2.14)-(2.15).

In our previous results for a general K, the regularity conditions (e.g., (3.1)) impose restrictions on the balance of sample sizes among groups. For $K = 2$, the severely unbalanced setting is interesting (e.g., in authorship attribution, $n = 1$ and m can be large). We relax the regularity conditions to the following ones:

Condition 3.1. *Let* θ *and* η *be as in* (3.10) *and define two matrices* $\Sigma_1 = \frac{1}{n\bar{N}}\sum_{i=1}^n N_i\Omega_i\Omega_i'$ *and* $\Sigma_2 = \frac{1}{m\bar{M}}\sum_{i=1}^m M_i\Gamma_i\Gamma_i'$. *We assume that the following statements are true (a) For* $1 \leq i \leq n$ *and* $1 \leq j \leq m$, $N_i \geq 2$, $\|\Omega_i\|_\infty \leq 1 - c_0$, $M_j \geq 2$, *and* $\|\Gamma_j\|_\infty \leq 1 - c_0$, *where* $c_0 \in (0,1)$ *is a contant, (b)* $\max\big\{\big(\frac{\|\eta\|_3^3}{n\bar{N}} + \frac{\|\theta\|_3^3}{m\bar{M}}\big), \big(\frac{\|\eta\|_2^2}{n^2\bar{N}^2} + \frac{\|\theta\|_2^2}{m^2\bar{M}_2^2}\big)\big\}/\big\|\frac{m\bar{M}}{n\bar{N}+m\bar{M}}\eta + \frac{n\bar{N}}{n\bar{N}+m\bar{M}}\theta\big\|^4 = o(1)$, *(c)* $\max\big\{\sum_i \frac{N_i^2}{n^2\bar{N}^2}\|\Omega_i\|_3^3, \sum_i \frac{M_i^2}{m^2\bar{M}^2}\|\Gamma_i\|_3^3, \|\Sigma_1\|_F^2 + \|\Sigma_2\|_F^2\big\}/\|\mu\|^2 = o(1)$, *and (d)* $\|\mu\|_4^4/\|\mu\|^4 = o(1)$.

Condition (a) is similar to (3.1), except that we drop the sample size balance requriement. Conditions (b)-(d) are equivalent to (3.4) but have more explicit expressions for $K = 2$.

Theorem 3.6. *In Model* (3.9), *we test the null hypothesis* H_0: $\theta = \mu$. *As* $\min\{n\bar{N}, m\bar{M}\} \to \infty$, *suppose Condition 3.1 is satisfied. Under the alternative hypothesis, we further assume*

$$\frac{\|\eta - \theta\|^2}{\big(\frac{1}{n\bar{N}} + \frac{1}{m\bar{M}}\big)\max\{\|\eta\|, \|\theta\|\}} \to \infty. \tag{3.11}$$

Consider the DELVE test statistic $\psi = T/\sqrt{V}$. The following statements are true. Under the null hypothesis, $\psi \to N(0,1)$ in distribution. Under the alternative hypothesis, $\psi \to \infty$ in probability. Moreover for any fixed $\alpha \in (0,1)$, the level-α DELVE test has an asymptotic level of α and an asymptotic power of 1.

Compared with the theorems for a general K, first, Theorem 3.6 allows the two groups to be severely unbalanced and reveals that the detection boundary depends on the harmonic mean of $n\bar{N}$ and $m\bar{M}$. Second, the detection boundary is expressed using $\|\eta - \theta\|$, which is easier to interpret.

3.5 The special case of $K = n$

The special case of $K = n$ is interesting for two reasons. First, the application example of global testing in topic models corresponds to $K = n$. Second, for any K, when Ω_i's within each group are assumed to be the same (e.g., this is the case in closeness testing of discrete distributions), it suffices to aggregate the counts in each group, i.e., let $Y_k = \sum_{i \in S_k} X_i$ and operate on $Y_1, \ldots, Y_K$ instead of the original X_i's; this reduces to the case of $K = n$.

When $K = n$, the null hypothesis has a simpler form:

$$H_0: \quad \Omega_i = \mu, \qquad 1 \leq i \leq n. \tag{3.12}$$

Moreover, under the alternative hypothesis, the quantity ω_n^2 in (3.5) simplifies to

$$\omega_n = \omega_n(\Omega_1, \Omega_2, \ldots, \Omega_n) = \frac{1}{n\bar{N}\|\mu\|^2} \sum_{i=1}^{n} N_i \|\Omega_i - \mu\|^2. \tag{3.13}$$

The DELVE test statistic also has a simplified form as in (2.10)-(2.11). We can prove the same theoretical results under *weaker conditions*:

Condition 3.2. *We assume that the following statements are true: (a) For a constant $c_0 \in (0,1)$, $2 \leq N_i \leq (1-c_0)n\bar{N}$ and $\|\Omega_i\|_\infty \leq 1 - c_0$, $1 \leq i \leq n$, and (b) $\max\{\sum_i \frac{\|\Omega_i\|_3^3}{N_i}, \sum_i \frac{\|\Omega_i\|^2}{N_i^2}\}/(\sum_i \|\Omega_i\|^2)^2 = o(1)$, and $(\sum_i \|\Omega_i\|_3^3)/(n\|\mu\|^2) = o(1)$*

When $K = n$, Condition (a) is equivalent to (3.1); and Condition (b) is weaker than (3.4), as we have dropped the requirement $\frac{\|\mu\|_4^4}{K\|\mu\|^4} = o(1)$. We obtain weaker conditions for $K = n$ because the dominant terms in T differ from those for $K < n$.

Theorem 3.7. *In Model (1.1), we test the null hypothesis (3.12). As $n \to \infty$, we assume that Condition 3.2 is satisfied. Under the alternative, we further assume that*

$$\frac{n\bar{N}\|\mu\|^2\omega_n^2}{\sqrt{\sum_{i=1}^n \|\Omega_i\|^2}} \to \infty. \tag{3.14}$$

Let T and V^* be the same as in (2.10)-(2.11). Consider the simplified DELVE test statistic $\psi^* = T/\sqrt{V^*}$. The following statements are true. Under the null hypothesis, $\psi^* \to N(0,1)$ in distribution. Under the alternative hypothesis, $\psi^* \to \infty$ in probability. Moreover for any fixed $\alpha \in (0,1)$, the level-α DELVE test has an asymptotic level of α and an asymptotic power of 1.

The detection boundary in (3.14) has a simpler form if $\sum_i \|\Omega_i\|^2 \asymp n\|\mu\|^2$. In this case, (3.14) is equivalent to $\sqrt{n}\bar{N}\|\mu\|\omega_n^2 \to \infty$. Additionally, if all entries of μ are at the same order, then $\|\mu\| \asymp 1/\sqrt{p}$, and (3.14) further reduces to $\sqrt{n\bar{N}^2/p} \cdot \omega_n^2 \to \infty$.

3.6 A discussion of the contiguity regime

Our power analysis in Section 3.2 concerns $\mathrm{SNR}_n \to \infty$, and our lower bound in Section 3.3 concerns $\mathrm{SNR}_n \to 0$. We now study the contiguity regime where SNR_n tends to a constant. For illustration, we consider a special choice of parameters, which allows us to obtain a simple expression of the testing risk.

Suppose $K = n$ and $N_i = N$ for all $1 \leq i \leq n$. Consider the pair of hypotheses:

$$H_0: \ \Omega_{ij} = p^{-1}, \qquad \text{v.s.} \qquad H_1: \ \Omega_{ij} = p^{-1}(1 + \beta_n \delta_{ij}), \tag{3.15}$$

where $\{\delta_{ij}\}_{1\leq i\leq n, 1\leq j\leq p}$ satisfy that $|\delta_{ij}| = 1$, $\sum_{j=1}^p \delta_{ij} = 0$ and $\sum_{i=1}^n \delta_{ij} = 0$. Such δ_{ij} always exist.[1] The SNR_n in (3.6) satisfies that $\mathrm{SNR}_n \asymp (N\sqrt{n}/\sqrt{p})\beta_n^2$. We thereby set

$$\beta_n^2 = \frac{\sqrt{2p}}{N\sqrt{n}} \cdot a, \qquad \text{for a constant } a > 0. \tag{3.16}$$

Since $K = n$ here, we consider the simplified DELVE test statistic ψ^* as in Section 3.5.

Theorem 3.8. *Consider Model* (1.1) *with* $N_i = N$*. For a constant* $a > 0$*, let the null and alternative hypotheses be specified as in* (3.15)-(3.16)*. As* $n \to \infty$*, if* $p = o(N^2 n)$*, then* $\psi^* \to N(0, 1)$ *under* H_0 *and* $\psi^* \to N(a, 1)$ *under* H_1*.*

Let Φ be the cumulative distribution function of the standard normal. By Theorem 3.8, for any fixed constant $t \in (0, a)$, if we reject the null hypothesis when $\psi^* > t$, then the sum of type I and type II errors converges to $[1 - \Phi(t)] + [1 - \Phi(a - t)]$.

4 Applications

As mentioned in Section 1, our testing problem includes global testing for topic models, authorship attribution, and closeness testing for discrete distributions as special examples. In this section, the DELVE test is applied separately to these three problems.

4.1 Global testing for topic models

Topic modeling [Blei et al., 2003] is a popular tool in text mining. It aims to learn a small number of "topics" from a large corpus. Given n documents written using a dictionary of p words, let $X_i \sim \mathrm{Multinomial}(N_i, \Omega_i)$ denote the word counts of document i, where N_i is the length of this document and $\Omega_i \in \mathbb{R}^p$ contains the population word frequencies. In a topic model, there exist M topic vectors $A_1, A_2, \ldots, A_M \in \mathbb{R}^p$, where each A_k is a PMF. Let

[1] For example, we can first partition the dictionary into two halves and then partition all the documents into two halves; this divides $\{1, 2, \ldots, p\} \times \{1, 2, \ldots, n\}$ into four subsets; we construct δ_{ij}'s freely on one subset and then specify the δ_{ij}'s on the other three subsets by symmetry.

$w_i \in \mathbb{R}^M$ be a nonnegative vector whose entries sum up to 1, where $w_i(k)$ is the "weight" document i puts on topic k. It assumes

$$\Omega_i = \sum_{k=1}^{M} w_i(k) A_k, \qquad 1 \leq i \leq n. \tag{4.1}$$

Under (4.1), the matrix $\Omega = [\Omega_1, \Omega_2, \ldots, \Omega_n]$ admits a low-rank nonnegative factorization.

Before fitting a topic model, we would like to know whether the corpus indeed involves multiple topics. This is the global testing problem: $H_0 : M = 1$ v.s. $H_1 : M > 1$. When $M = 1$, by writing $A_1 = \mu$, the topic model reduces to the null hypothesis in (3.12). We can apply the DELVE test by treating each X_i as a separate group (i.e., $K = n$).

Corollary 4.1. *Consider Model* (1.1) *and define a vector $\xi \in \mathbb{R}^n$ by $\xi_i = \bar{N}^{-1} N_i$. Suppose that $\Omega = \mu \mathbf{1}_n'$ under the null hypothesis, with $\mu = n^{-1}\Omega\xi$, and that Ω satisfies* (4.1) *under the alternative hypothesis, with $r := \mathrm{rank}(\Omega) \geq 2$. Suppose $\bar{N}/(\min_i N_i) = O(1)$. Denote by $\lambda_1, \lambda_2, \ldots, \lambda_r > 0$ the singular values of $\Omega[\mathrm{diag}(\xi)]^{1/2}$, arranged in the descending order. We further assume that under the alternative hypothesis,*

$$\bar{N} \cdot \frac{\sum_{k=2}^{r} \lambda_k^2}{\sqrt{\sum_{k=1}^{r} \lambda_k^2}} \to \infty. \tag{4.2}$$

For any fixed $\alpha \in (0, 1)$, the level-α DELVE test has an asymptotic level α and an asymptotic power 1*.*

The least-favorable configuration in the proof of Theorem 3.5 is in fact a topic model that follows (4.1) with $M = 2$. Transferring the argument yields the following lower bound that confirms the optimality of DELVE for the global testing of topic models.

Corollary 4.2. *Let $\mathcal{R}_{n,M}(\epsilon_n, \delta_n)$ be the collection of $\{(N_i, \Omega_i)\}_{i=1}^n$ satisfying the following conditions: 1) Ω follows the topic model* (4.1) *with M topics; 2) Condition 3.2 holds with $o(1)$ replaced by $\leq \epsilon_n$; 3) $\bar{N}(\sum_{k=2}^{r} \lambda_k^2)/(\sum_{k=1}^{r} \lambda_k^2)^{1/2} \geq \delta_n$. If $\epsilon_n \to 0$ and $\delta_n \to 0$, then $\limsup_{n\to\infty} \inf_{\Psi\in\{0,1\}} \Big\{ \sup_{\mathcal{R}_{n,1}(\epsilon_n,0)} \mathbb{P}(\Psi = 1) + \sup_{\cup_{M\geq 2}\mathcal{R}_{n,M}(\delta_n,\delta_n)} \mathbb{P}(\Psi = 0) \Big\} = 1$.*

The detection boundary (4.2) can be simplified when $M = O(1)$. Following Ke and Wang [2022], we define $\Sigma_A = A' H^{-1} A$ and $\Sigma_W = n^{-1} W W'$, where $A = [A_1, A_2, \ldots, A_M]$, $W = [w_1, w_2, \ldots, w_n]$ and $H = \mathrm{diag}(A\mathbf{1}_M)$. Ke and Wang [2022] argued that it is reasonable to assume that eigenvalues of these two matrices are at the constant order. If this is true, with some mild additional regularity conditions, each λ_k is at the order of $\sqrt{n/p}$. Hence, (4.2) reduces to $\sqrt{n}\bar{N}/\sqrt{p} \to \infty$. In comparison, Ke and Wang [2022] showed that a necessary condition for any estimator $\hat{A} = [\hat{A}_1, \hat{A}_2, \ldots, \hat{A}_M]$ to achieve $\frac{1}{M}\sum_{k=1}^{M} \|\hat{A}_k - A_k\|_1 = o(1)$ is $\sqrt{n\bar{N}/p} \to \infty$. We conclude that consistent estimation of topic vectors requires strictly stronger conditions than successful testing.

4.2 Authorship attribution

In authorship attribution, given a corpus from a known author, we want to test whether a new document is from the same author. It is a special case of our testing problem with $K = 2$. We can directly apply the results in Section 3.4. However, the setting in Section 3.4 has no sparsity. Kipnis [2022], Donoho and Kipnis [2022] point out that the number of words with discriminating power is often much smaller than p. To see how our test performs under sparsity, we consider a sparse model. As in Section 3.4, let

$$X_i \sim \text{Multinomial}(N_i, \Omega_i),\ 1 \le i \le n, \quad \text{and} \quad G_i \sim \text{Multinomial}(M_i, \Gamma_i),\ 1 \le i \le m. \tag{4.3}$$

Let $\bar{N}$ and $\bar{M}$ be the average of N_i's and M_i's, respectively. Write $\eta = \frac{1}{n\bar{N}} \sum_{i=1}^{n} N_i \Omega_i$ and $\theta = \frac{1}{m\bar{M}} \sum_{i=1}^{m} M_i \Gamma_i$. We assume for some $\beta_n > 0$,

$$\eta_j = \theta_j, \text{ for } j \notin S, \qquad \text{and} \qquad \left|\sqrt{\eta_j} - \sqrt{\theta_j}\right| \ge \beta_n, \text{ for } j \in S. \tag{4.4}$$

Corollary 4.3. *Under the model* (4.3)-(4.4)*, consider testing $H_0 : S = \emptyset$ v.s. $H_1 : S \neq \emptyset$, where Condition 3.1 is satisfied. Let η_S and θ_S be the sub-vectors of η and θ restricted to the coordinates in S. Suppose that under the alternative hypothesis,*

$$\frac{\beta_n^2 \cdot (\|\eta_S\|_1 + \|\theta_S\|_1)}{\left(\frac{1}{n\bar{N}} + \frac{1}{m\bar{M}}\right) \max\{\|\eta\|, \|\theta\|\}} \to \infty. \tag{4.5}$$

As $\min\{n\bar{N}, m\bar{M}\} \to \infty$, the level-$\alpha$ DELVE test has an asymptotic level α and an asymptotic power 1*. Furthermore, if $n\bar{N} \asymp m\bar{M}$ and $\min_{j \in S}(\eta_j + \theta_j) \ge cp^{-1}$ for a constant $c > 0$, then* (4.5) *reduces to $n\bar{N}\beta_n^2 |S|/\sqrt{p} \to \infty$.*

Donoho and Kipnis [2022] studied a case where $N = M$, $n = m = 1$, $p \to \infty$,

$$|S| = p^{1-\vartheta}, \qquad \text{and} \qquad \beta_n = c \cdot N^{-1/2}\sqrt{\log(p)}. \tag{4.6}$$

When $\vartheta > 1/2$ (i.e., $|S| = o(\sqrt{p})$), they derived a phase diagram for the aforementioned testing problem (under a slightly different setting where the data distributions are Poisson instead of multinomial). They showed that when $\vartheta > 1/2$ and c is a properly large constant, a Higher-Criticism-based test has an asymptotically full power. Donoho and Kipnis [2022] did not study the case of $\vartheta \le 1/2$. By Corollary 4.3, when $\vartheta \le 1/2$ (i.e., $|S| \ge C\sqrt{p}$), the DELVE test has asymptotically full power.

Remark 2. When $\vartheta > 1/2$ in (4.6), the DELVE test is powerless. However, this issue can be resolved by borrowing the idea of maximum test or Higher Criticism test [Donoho and Jin, 2004] from the classical multiple testing. For example, recalling T_j in (2.5), we can use $\max_{1 \le j \le p}\{T_j/\sqrt{V_j}\}$ as the test statistic, where V_j is a proper estimator of the variance of T_j. We leave a careful study of this idea to future work.

4.3 Closeness testing between discrete distributions

Two-sample closeness testing is a subject of intensive study in discrete distribution inference [Bhattacharya and Valiant, 2015, Chan et al., 2014, Diakonikolas and Kane, 2016, Kim et al., 2022]. It is a special case of our problem with $K = 2$ and $n_1 = n_2 = 1$. We thereby apply both Theorem 3.6 and Theorem 3.7.

Corollary 4.4. *Let Y_1 and Y_2 be two discrete variables taking values on the same p outcomes. Let $\Omega_1 \in \mathbb{R}^p$ and $\Omega_2 \in \mathbb{R}^p$ be their corresponding PMFs. Suppose we have N_1 samples of Y_1 and N_2 samples of Y_2. The data are summarized in two multinomial vectors: $X_1 \sim \text{Multinomial}(N_1, \Omega_1), X_2 \sim \text{Multinomial}(N_2, \Omega_2)$. We test $H_0 : \Omega_1 = \Omega_2$. Write $\mu = \frac{1}{N_1+N_2}(N_1\Omega_1 + N_2\Omega_2)$. Suppose $\min\{N_1, N_2\} \geq 2$, $\max\{\|\Omega_1\|_\infty, \|\Omega_2\|_\infty\} \leq 1 - c_0$, for a constant $c_0 \in (0, 1)$. Suppose $\frac{1}{(\sum_{k=1}^2 \|\Omega_k\|^2)^2} \max\big\{\sum_{k=1}^2 \frac{\|\Omega_k\|_3^3}{N_k}, \sum_{k=1}^2 \frac{\|\Omega_k\|^2}{N_k^2}\big\} = o(1)$, and $\frac{1}{n\|\mu\|^2}\sum_{k=1}^2 \|\Omega_k\|_3^3 = o(1)$. We assume that under the alternative hypothesis,*

$$\frac{\|\Omega_1 - \Omega_2\|^2}{\left(N_1^{-1} + N_2^{-1}\right)\max\{\|\Omega_1\|,\, \|\Omega_2\|\}} \to \infty. \tag{4.7}$$

As $\min\{N_1, N_2\} \to \infty$, the level-$\alpha$ DELVE test has level α and power 1, asymptotically.

We notice that (4.7) matches with the minimum ℓ^2-separation condition for two-sample closeness testing [Kim et al., 2022, Proposition 4.4]. Therefore, our test is an optimal ℓ^2-testor. Although other optimal ℓ^2-testors have been proposed [Chan et al., 2014, Bhattacharya and Valiant, 2015, Diakonikolas and Kane, 2016], they are not equipped with tractable null distributions.

Remark 3. We can modify the DELVE test to incorporate frequency-dependent weights. Given any nonnegative vector $w = (w_1, w_2, \ldots, w_p)'$, define $T(w) := \sum_{j=1}^p w_j T_j$ where T_j is the same as in (2.5). These weights w_j serve to adjust the contributions of different words. For example, consider $w_j = \left(\max\{1/p,\ \hat{\mu}_j\}\right)^{-1}$. This kind of weights have been used in discrete distribution inference [Balakrishnan and Wasserman, 2019, Chan et al., 2014] to turn an optimal ℓ^2 testor to an optimal ℓ^1 testor. We can similarly study the power of this modified test, except that we need an additional assumption $n\bar{N} \gg p$ to guarantee that $\hat{\mu}_j$ is a sufficiently accurate estimator of μ_j.

5 Simulations

The proposed DELVE test is computationally efficient and easy to implement. In this section, we investigate its numerical performance in simulation studies. Real data analysis will be carried out in Section 6.

Experiment 1 (Asymptotic normality) . Given $(n, p, K, N_{\min}, N_{\max}, \alpha)$, we generate data as follows: first, we divide $\{1, \ldots, n\}$ into K equal-size groups. Next, we draw $\Omega_1^{alt}, \ldots, \Omega_n^{alt}$ i.i.d. from Dirichlet$(p, \alpha\mathbf{1}_p)$. Third, we draw $N_i \overset{iid}{\sim} \text{Uniform}[N_{\min}, N_{\max}]$ and set $\Omega_i^{null} = \mu$, where $\mu := \frac{1}{nN}\sum_i N_i\Omega_i^{alt}$. Last, we generate $X_1, \ldots, X_n$ using Model (1.1). We consider three sub-experiments. In Experiment 1.1, $(n, p, K, N_{\min}, N_{\max}, \alpha) = (50, 100, 5, 10, 20, 0.3)$. In Experiment 1.2, α is changed to 1, and the other parameters are the same. When $\alpha = 1$, Ω_i^{alt} are drawn from the uniform distribution of the standard probability simplex; in comparison, $\alpha = 0.3$ puts more mass near the boundary of the standard probability simplex. In Experiment 1.3, we keep all parameters the same as in Experiment 1.1, except that (p, K) are changed to $(300, 50)$. For each sub-experiment, we generate 2000 data sets under the null hypothesis and plot the histogram of the DELVE test statistic ψ (in

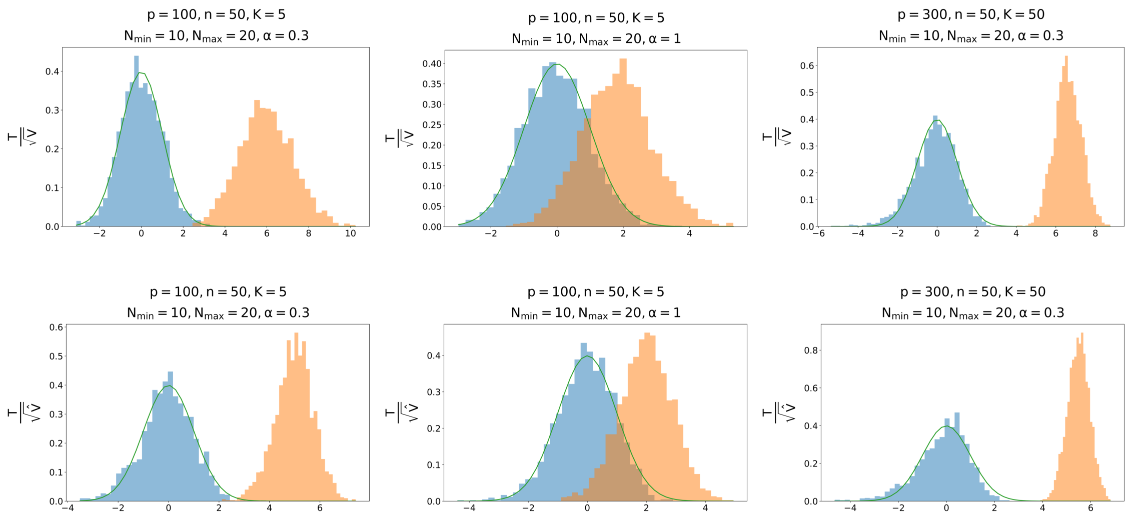


Figure 1: Histograms of the DELVE statistic (top three panels) and the DELVE+ statistic (bottom three panels) in Experiments 1.1-1.3. In each plot, the blue and orange histograms correspond to the null and alternative hypotheses, respectively; and the green curve is the density of $N(0,1)$.

blue); similarly, we generate 2000 data sets under the alternative hypothesis and plot the histogram of ψ (in orange). The results are contained on the top three panels of Figure 1. In Section 2.2, we introduced a variant of DELVE, called DELVE+, in which the variance estimator V is replaced by an adjusted one. DELVE+ has similar theoretical properties as DELVE but is more suitable for real data. We plot the histograms of the DELVE+ test statistics on the bottom three panels of Figure 1.

We have several observations. In all sub-experiments, when the null hypothesis holds, the histograms of both DELVE and DELVE+ fit the standard normal density reasonably well. This supports our theory in Section 3.1. Second, when (p, K) increase, the finite sample effect becomes slightly more pronounced (c.f., Experiment 1.3 versus Experiment 1.1). Third, the tests have power in differentiating two hypotheses. As α decreases or K increases, the power increases, and the histograms corresponding to two hypotheses become further apart. Last, in the alternative hypothesis, DELVE+ has smaller mean and variance than DELVE. By Lemma 2.2, these two have similar asymptotic behaviors. The simulation results suggest that they have noticeable finite-sample differences.

Experiment 2 (Power curve). Similarly as before, we divide $\{1, 2, \ldots, n\}$ into K equal-size groups and draw $N_i \sim \mathrm{Uniform}[N_{\min}, N_{\max}]$. In this experiment, the PMF's Ω_i are generated in a different way. Under the null hypothesis, we generate $\mu \sim \mathrm{Dirichlet}(p/2, \alpha \mathbf{1}_{p/2})$ and set $\Omega_i^{null} = \tilde{\mu}$, where $\tilde{\mu}_j = \frac{1}{2}\mu_j$ for $1 \leq j \leq p/2$ and $\tilde{\mu}_j = \frac{1}{2}\mu_{p+1-j}$ for $p/2 + 1 \leq j \leq p$. Under the alternative hypothesis, we draw $z_1, \ldots, z_K$, $b_1, \ldots, b_{p/2} \overset{iid}{\sim} \mathrm{Rademacher}(1/2)$ and then let $\Omega_{ij}^{alt} = \mu(1 + \tau_n z_k b_j)$, for all i in group k and $1 \leq j \leq p/2$, and $\Omega_{ij}^{alt} = \mu(1 + \tau_n z_k b_j)$ for $p/2 + 1 \leq j \leq p$. By applying our theory in Section 3.2 together with some calculations, we obtain that the signal-to-noise ratio is captured by $\lambda := K^{-1/2} n \bar{N} \|\mu\| \tau_n$. We consider three sub-experiments, Experiment 2.1-2.3, in which the parameter values of

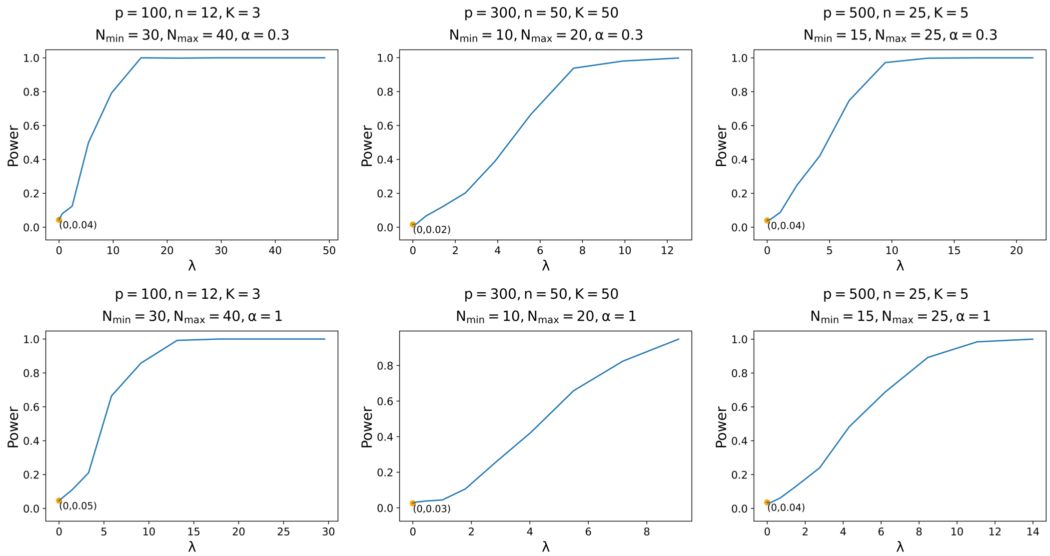


Figure 2: Power diagrams (based on 500 repetitions) at level 5%. The x-axis plots the SNR $\lambda(\omega_n) = K^{-1/2} n\bar{N}\|\mu\| \cdot \omega_n$.

$(n, p, K, N_{\min}, N_{\max}, \alpha)$ are the same as in Experiments 1.1-1.3. For each sub-experiment, we consider a grid of 10 equally-spaced values of λ. When $\lambda = 0$, it corresponds to the null hypothesis; when $\lambda > 0$, it corresponds to the alternative hypothesis. For each λ, we generate 500 data sets and compute the fraction of rejections of the level-5% DELVE test. This gives a power curve for the level-5% DELVE test, in which the first point corresponding to $\lambda = 0$ is the actual level of the test. The results are contained on the top three panels of Figure 2. We repeat the same experiments for the DELVE+ test, which results are on the bottom three panels of Figure 2. In all three experiments, the actual level of our proposed tests is $\leq 5\%$, suggesting that our tests perform well at controlling the type-I error. As λ increases, the power gradually increased to 1, suggesting that λ is a good metric of the signal-to-noise ratio. This supports our theory in Section 3.2.

6 Real Data Analysis

We apply our proposed methods on two real corpora: one consists of abstracts of research papers in four statistics journals, and the other consists of movie reviews on Amazon. For the analysis of real data, we use DELVE+, which modifies the variance estimator in DELVE and reduces the occurrence of extremely small p-values.

6.1 Abstracts of statisticians

We use the data set from Ji and Jin [2016]. It contains the bibtex information of all published papers in four top-tier statistics journals, *Annals of Statistics*, *Biometrika*, *Journal of the American Statistical Association*, and *Journal of the Royal Statistical Society - Series B*, from 2003 to the first half of 2012. We pre-process the abstracts of papers by tokenization

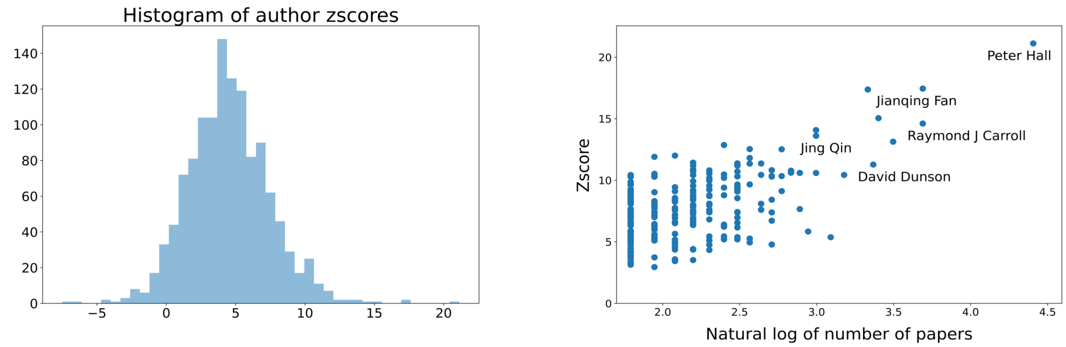


Figure 3: (**Left**) Histogram of nonzero DELVE Z-scores for all authors in the dataset. The mean is 4.52 and the standard deviation is 2.94. (**Right**) Scatter plot of author DELVE scores versus the natural log of the number of papers with five statisticians identified.

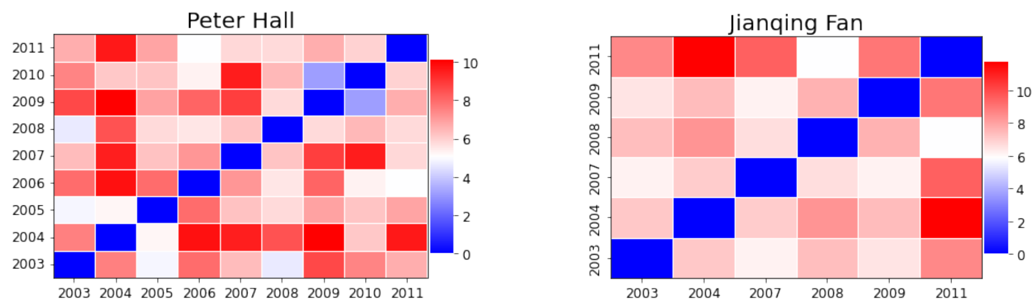


Figure 4: Pairwise Z-score plots for Peter Hall (left) and Jianqing Fan (right). In the cell (x, y), we compare the corpus of an author's abstracts from time x with the corpus of that author's abstracts from time y. The heatmap shows the value of DELVE+ with $K = 2$ for each cell.

and stemming and turn each abstract to a word count vector.

We conduct two experiments. In the first one, we fix an author and treat the collection of his/her co-authored abstracts as a corpus. We apply DELVE+ with $K = n$, where n is the total number of abstracts written by this author. The Z-score measures the "diversity" or "variability" of this authors' abstracts. An author with a high Z-score possesses either diverse research interests or a variable writing style. A number of authors have only 1–2 papers in this data set, and the variance estimator V is often negative; we remove all those authors. In Figure 3 (left panel), we plot the histogram of Z-scores of all retained authors. The mean is 4.52 and the standard deviation is 2.94. In Figure 3 (right panel), we show the scatter plot of Z-score versus logarithm of the number of abstracts written by this author, and a few prolific authors who have many papers and a large Z-score are labeled. For example, Peter Hall has the most papers in this dataset (82 papers in total). Hall's Z-score is larger than 20, implying a huge diversity in his abstracts. There is also a positive association between Z-score and total papers. It suggests that senior authors have more diversity in their abstracts, which is as expected.

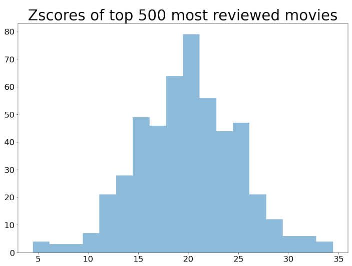
Year	Title	Journal
2011	Nonparametric independence screening in sparse ultra-high-dimensional additive models	JASA
2011	Penalized composite quasi-likelihood for ultrahigh dimensional variable selection	JRSS-B
2011	Multiple testing via FDR_L for large-scale imaging data	Ann. Stat.
2012	Vast volatility matrix estimation using high-frequency data for portfolio selection	JASA
2012	A road to classification in high dimensional space: the regularized optimal affine discriminant	JRSS-B
2012	Variance estimation using refitted cross-validation in ultrahigh dimensional regression	JRSS-B

Year	Title	Journal
2004	Low order approximations in deconvolution and regression with errors in variables	JRSS-B
2004	Nonparametric inference about service time distribution from indirect measurements	JRSS-B
2004	Cross-validation and the estimation of conditional probability densities	JASA
2004	Nonparametric confidence intervals for receiver operating characteristic curves	Biometrika
2004	Bump hunting with non-Gaussian kernels	Ann. Stat.
2004	Attributing a probability to the shape of a probability density	Ann. Stat.

Figure 5: (**Left**) Jianqing Fan's papers in the dataset of Ji and Jin [2016] from 2011 to 2012. (**Right**) Peter Hall's papers in the dataset of Ji and Jin [2016] from 2004.

In the second experiment, we divide the abstracts of each author into groups by publication year. We divide Peter Hall's abstracts into 9 groups, and each group corresponds to one year. We divide Jianqing Fan's abstracts into 6 groups, with unequal window sizes to make all groups have roughly equal numbers of abstracts. Our test can be used to detect differences between all groups, but to see more informative results, we do a pairwise comparison: for each pair of groups, we apply DELVE+ with $K = 2$. This yields a pairwise plot of Z-scores. The plot reveals the temporal patterns of this author in abstract writing. Figure 4 shows the results for Peter Hall and Jianqing Fan.

There are interesting temporal patterns. For Jianqing Fan (right panel of Figure 4), the group consisting of his 2011-2012 abstracts has comparably large Z-scores in the pairwise comparison with other groups. To interpret this , we gathered the titles and abstracts of all his papers in the dataset and compared the ones before/after 2011. He published six papers in these journals during 2011-2012, whose titles are listed on the left of Figure 5. We see that his papers in this period had a strong emphasis on screening and variable selection: four out of the six papers mention this subject in their titles and/or abstracts. This shows a departure from his previously published topics such as covariance estimation (a focus from 2007–2009) and semiparametric estimation (a focus before 2010). Though Jianqing Fan had previously published papers on variable selection and screening in these journals, he had never published so many in such a short time period. For Peter Hall (left panel of Figure 4), the group of 2004 abstracts have comparably large Z-scores in the pairwise comparison with other groups. We examined the titles and abstracts of his 6 papers published in 2004 in this data set. All of his 2004 papers, except the first one, mention *bandwidth selection* or *smoothing parameters*, and in the last 4 papers, *bandwidth selection* plays a central role. For instance, *Bump hunting with non-Gaussian kernels, (Ann. Stat., 2004)* studies the relationship between the number of modes of a kernel density estimator and its bandwidth parameter. Though Peter Hall's 2014 papers concern many nonparametric statistics topics, we find that bandwidth selection is a theme underlying his research in these journals in 2004.



Rank	Title	Z-Score	Total reviews
1	Prometheus	34.44	813
2	Expelled: No Intelligence Allowed	34.17	830
3	V for Vendetta	32.24	815
4	Sin City	31.72	828
5	No Country for Old Men	30.57	819
⋮	⋮	⋮	⋮
16	John Adams	20.78	857
17	Cars	19.98	902
18	Food, Inc.	17.81	876
19	Jeff Dunham: Arguing with Myself	4.96	860
20	Jeff Dunham: Spark of Insanity	4.46	877

Figure 6: (**Left**) Histogram of Z-scores for the 500 most-reviewed movies. The mean is 19.97 and the standard deviation is 5.07. (**Right**) Z-scores for the top 20 most reviewed movies.

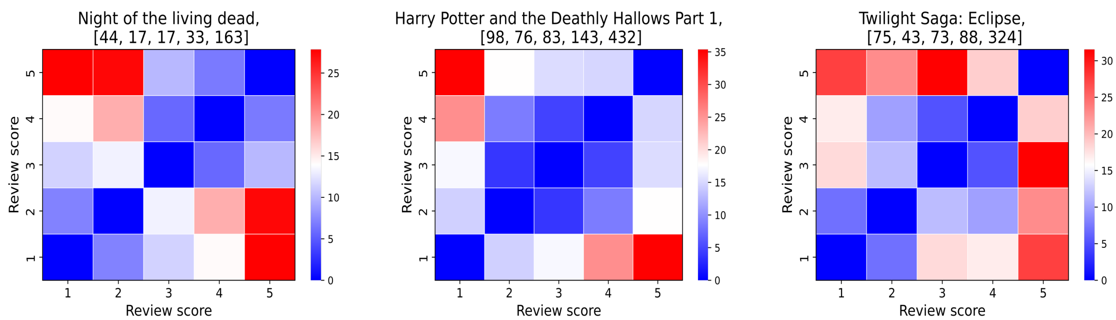


Figure 7: Pairwise Z-scores for 3 movies. In each cell, we use DELVE+ to compare reviews associated to a pair of star ratings. For each movie, the title list the number of reviews of each rating from 1–5.

6.2 Amazon movie reviews

We analyze Amazon reviews from the dataset Maurya [2018] that consists of 1,924,471 reviews of 143,007 visual media products (ie, DVDs, Bluray, or streams). We examine products with the largest number of reviews. Each product's review corpus is cleaned and stemming is used to group together words with the same root. We obtain word counts for each review and a term-document matrix of a product's review corpus. In the first experiment, we fix a movie and apply DELVE+ with $K = n$ to the corpus consisting of all reviews of this movie. In Figure 6 (left panel), we plot the histogram of Z-scores for the top 500 most reviewed movies. The mean is 19.97 and the standard deviation is 5.07. Compared with the histogram of Z-scores for statistics paper abstracts, there is much larger diversity in movie reviews. In Figure 6 (right panel), we list the 5 movies with the highest Z-scores and lowest Z-scores out of the 20 most reviewed movies. Each movie has more than 800 reviews, but some have surprisingly low Z-scores. The works by comedian Jeff Dunham have the lowest Z-scores, suggesting strong homogeneity among the reviews. The 2012 horror film *Prometheus* has the highest degree of review diversity among the 20 most reviewed movies. In the second experiment, we divide the reviews of each movie into

5 groups by star rating. We compare each pair of groups using DELVE+ with $K = 2$, resulting in a pairwise Z-score plot. In Figure 7, we plot this for 3 popular movies. We see a variety of polarization patterns among the scores. In *Harry Potter and the Deathly Hallows Part I*, DELVE+ signifies that the reviews with ratings in the range 2–4 stars are all similar. We see a smooth gradation in how the 1-star reviews differ from those from 2–4 stars, and similarly for 5-star reviews versus those from 2–4 stars. *Twilight Saga: Eclipse* shows three clusters: 1–2 stars, 3–4 stars, and 5 star, while *Night of the living dead* shows two clusters: 1–2 stars and 3–5 stars.

7 Discussions

We examine the testing for equality of PMFs of K groups of high-dimensional multinomial distributions. The proposed DELVE statistic has a parameter-free limiting null that allows for computation of Z-scores and p-values on real data. DELVE achieves the optimal detection boundary over the whole range of parameters $(n, p, K, \bar{N})$, including the high-dimensional case $p \to \infty$, which is very relevant to applications in text mining.

This work leads to interesting questions for future study. So far the focus is on testing, but one can also consider inference for ρ^2 from (2.2), which measures the heterogeneity among the group-wise means. Consistent variance estimation under the alternative uses a similar strategy, though we omit the calculations in this paper. Establishing asymptotic normality of DELVE under the alternative would then lead to asymptotic confidence intervals for our heterogeneity metric ρ^2. Based on the plots in Section 5, it is possible that stronger regularity conditions are needed to obtain a pivotal distribution under the alternative. As in the two-sample multinomial testing problems described in Kipnis and Donoho [2021], Kipnis [2022], such as author attribution, we may also consider an alternative where all the group means are the same except for a small set of "giveaway words". It is interesting to develop a procedure for identifying these useful words. As discussed in Section 4.2, we may modify DELVE by using a version based on the maximum test or higher criticism. Another extension is to go beyond 'bag-of-words' style analysis and use different types of counts besides raw word frequencies. One option is to apply a suitably modified DELVE to the counts of multi-grams in the corpus and another is to combine words with similar meanings into a 'superword' and use superword counts as the basis for DELVE. To do this, we can combine words that are close together in some word embedding. We leave these interesting tasks for future work.

Acknowledgments The research of T. Tony Cai was supported in part by NSF Grant DMS-2015259 and NIH grant R01-GM129781. The research of Zheng Tracy Ke was supported in part by NSF CAREER Grant DMS-1943902.

Notational conventions for the appendix: We write $A \lesssim B$ (respectively, $A \gtrsim B$) if there exists an absolute constant $C > 0$ such that $A \le C \cdot B$ (respectively $A \ge C \cdot B$). If both $A \lesssim B$ and $B \lesssim A$, we write $A \asymp B$. The implicit constant C may vary from line to line. For sequences a_t, b_t indexed by an integer $t \in \mathbb{N}$, we write $a_t \ll b_t$ if $b_t/a_t \to \infty$ as $t \to \infty$, and we write $a_t \gg b_t$ if $a_t/b_t \to \infty$ as $t \to \infty$. We also may write $a_t = o(b_t)$ to denote $a_t \ll b_t$. In particular, we write $a_t = (1 + o(1))b_t$ if $a_t/b_t \to 1$ as $t \to \infty$.

A Properties of T and V

We recall that

$$X_i \sim \text{Multinomial}(N_i, \Omega_i), \qquad 1 \le i \le n. \tag{A.1}$$

For each $1 \le k \le K$, define

$$\mu_k = \frac{1}{n_k \bar{N}_k} \sum_{i \in S_k} N_i \Omega_i \ \in \ \mathbb{R}^p, \qquad \Sigma_k = \frac{1}{n_k \bar{N}_k} \sum_{i \in S_k} N_i \Omega_i \Omega_i' \ \in \ \mathbb{R}^{p \times p}. \tag{A.2}$$

Moreover, let

$$\mu = \frac{1}{n\bar{N}} \sum_{k=1}^{K} n_k \bar{N}_k \mu_k = \frac{1}{n\bar{N}} \sum_{i=1}^{n} N_i \Omega_i\,, \quad \Sigma = \frac{1}{n\bar{N}} \sum_{k=1}^{K} n_k \bar{N}_k \Sigma_k = \frac{1}{n\bar{N}} \sum_{i} N_i \Omega_i \Omega_i' \tag{A.3}$$

The DELVE test statistic is $\psi = T/\sqrt{V}$, where T is as in (2.5) and V is as in (2.7). As a preparation for the main proofs, in this section, we study T and V separately.

A.1 The decomposition of T

It is well-known that a multinomial with the number of trials equal to N can be equivalently written as the sum of N independent multinomials each with the number of trials equal to 1. This inspires us to introduce a set of independent, mean-zero random vectors:

$$\{Z_{ir}\}_{1 \le i \le n, 1 \le r \le N_i}, \qquad \text{with } Z_{ir} = B_{ir} - \mathbb{E}B_{ir}, \ \text{ and } \ B_{ir} \sim \text{Multinomial}(1, \Omega_i). \tag{A.4}$$

We use them to get a decomposition of T into mutually uncorrelated terms:

Lemma A.1. *Let $\{Z_{ir}\}_{1 \le i \le n, 1 \le r \le N_i}$ be as in (A.4). For each $Z_{ir} \in \mathbb{R}^p$, let $\{Z_{ijr}\}_{1 \le j \le p}$ denote its p coordinates. Recall that $\rho^2 = \sum_{k=1}^{K} n_k \bar{N}_k \|\mu_k - \mu\|^2$. For $1 \le j \le p$, define*

$$\begin{aligned}
U_{1j} &= 2\sum_{k=1}^{K} \sum_{i \in S_k} \sum_{r=1}^{N_i} (\mu_{kj} - \mu_j) Z_{ijr}, \\
U_{2j} &= \sum_{k=1}^{K} \sum_{i \in S_k} \sum_{1 \le r \ne s \le N_i} \Big(\frac{1}{n_k \bar{N}_k} - \frac{1}{n\bar{N}} \Big) \frac{N_i}{N_i - 1} Z_{ijr} Z_{ijs}, \\
U_{3j} &= -\frac{1}{n\bar{N}} \sum_{1 \le k \ne \ell \le K} \sum_{i \in S_k} \sum_{m \in S_\ell} \sum_{r=1}^{N_i} \sum_{s=1}^{N_m} Z_{ijr} Z_{mjs},
\end{aligned}$$

$$U_{4j} \;=\; \sum_{k=1}^{K} \sum_{\substack{i\in S_k, m\in S_k\\ i\neq m}} \sum_{r=1}^{N_i}\sum_{s=1}^{N_m}\Big(\frac{1}{n_k\bar{N}_k}-\frac{1}{n\bar{N}}\Big) Z_{ijr}Z_{mjs}.$$

Then, $T = \rho^2 + \sum_{\kappa=1}^4 \mathbf{1}_p' U_\kappa$. *Moreover,* $\mathbb{E}[U_\kappa] = \mathbf{0}_p$ *and* $\mathbb{E}[U_\kappa U_\zeta'] = \mathbf{0}_{p\times p}$ *for* $1 \leq \kappa \neq \zeta \leq 4$.

A.2 The variance of T

By Lemma A.1, the four terms $\{\mathbf{1}_p' U_\kappa\}_{1\leq\kappa\leq 4}$ are uncorrelated with each other. Therefore,

$$\mathrm{Var}(T) = \mathrm{Var}(\mathbf{1}_p'U_1) + \mathrm{Var}(\mathbf{1}_p'U_2) + \mathrm{Var}(\mathbf{1}_p'U_3) + \mathrm{Var}(\mathbf{1}_p'U_4).$$

It suffices to study the variance of each of these four terms.

Lemma A.2. *Let* U_1 *be the same as in Lemma A.1. Define*

$$\Theta_{n1} = 4\sum_{k=1}^{K} n_k\bar{N}_k \big\|\mathrm{diag}(\mu_k)^{1/2}(\mu_k-\mu)\big\|^2 \tag{A.5}$$

$$L_n = 4\sum_{k=1}^{K} n_k\bar{N}_k \big\|\Sigma_k^{1/2}(\mu_k-\mu)\big\|^2 \tag{A.6}$$

Then $\mathrm{Var}(\mathbf{1}_p'U_1) = \Theta_{n1} - L_n$. *Furthermore, if* $\max_{1\leq k\leq K}\|\mu_k\|_\infty = o(1)$, *then* $\mathrm{Var}(\mathbf{1}_p'U_1) = o(\rho^2)$.

Lemma A.3. *Let* U_2 *be the same as in Lemma A.1. Define*

$$\Theta_{n2} = 2\sum_{k=1}^{K}\Big(\frac{1}{n_k\bar{N}_k}-\frac{1}{n\bar{N}}\Big)^2 \sum_{i\in S_k}\frac{N_i^3}{N_i-1}\|\Omega_i\|^2 \tag{A.7}$$

$$A_n = 2\sum_{k=1}^{K}\Big(\frac{1}{n_k\bar{N}_k}-\frac{1}{n\bar{N}}\Big)^2 \sum_{i\in S_k}\frac{N_i^3}{N_i-1}\|\Omega_i\|_3^3 \tag{A.8}$$

Then

$$\Theta_{n2} - A_n \leq \mathrm{Var}(\mathbf{1}_p'U_2) \leq \Theta_{n2}.$$

Furthermore, if

$$\max_{1\leq k\leq K}\Big\{\frac{\sum_{i\in S_k} N_i^2\|\Omega_i\|_3^3}{\sum_{i\in S_k} N_i^2\|\Omega_i\|^2}\Big\} = o(1), \tag{A.9}$$

then $\mathrm{Var}(\mathbf{1}_p'U_2) = [1+o(1)]\cdot\Theta_{n2}$.

Lemma A.4. *Let* U_3 *be the same as in Lemma A.1. Define*

$$\Theta_{n3} = \frac{2}{n^2\bar{N}^2}\sum_{k\neq\ell}\sum_{i\in S_k}\sum_{m\in S_\ell}\sum_{j} N_iN_m\Omega_{ij}\Omega_{mj} \tag{A.10}$$

$$B_n = 2\sum_{k\neq\ell}\frac{n_kn_\ell\bar{N}_k\bar{N}_\ell}{n^2\bar{N}^2}\mathbf{1}_p'(\Sigma_k\circ\Sigma_\ell)\mathbf{1}_p \tag{A.11}$$

Then

$$\Theta_{n3} - B_n \leq \mathrm{Var}(\mathbf{1}_p'U_3) \leq \Theta_{n3} + B_n.$$

Lemma A.5. *Let U_4 be the same as in Lemma A.1. Define*

$$\Theta_{n4} = 2\sum_{k=1}^{K} \sum_{\substack{i\in S_k, m\in S_k\\ i\neq m}} \sum_{j} \Big(\frac{1}{n_k \bar{N}_k} - \frac{1}{n\bar{N}}\Big)^2 N_i N_m \Omega_{ij}\Omega_{mj}. \tag{A.12}$$

$$E_n = 2\sum_{k} \sum_{\substack{i\in S_k, m\in S_k,\\ i\neq m}} \sum_{1\leq j,j'\leq p} \Big(\frac{1}{n_k \bar{N}_k} - \frac{1}{n\bar{N}}\Big)^2 N_i N_m \Omega_{ij}\Omega_{ij'}\Omega_{mj}\Omega_{mj'} \tag{A.13}$$

Then

$$\Theta_{n4} - E_n \leq \mathrm{Var}(\mathbf{1}_p' U_4) \leq \Theta_{n4} + E_n$$

.

Using Lemmas A.2-A.5, we derive regularity conditions such that the first term in $\mathrm{Var}(\mathbf{1}_p' U_\kappa)$ is the dominating term. Observe that $\Theta_n = \Theta_{n1} + \Theta_{n2} + \Theta_{n3} + \Theta_{n4}$, where the quantity Θ_n is defined in (2.6). The following intermediate result is useful.

Lemma A.6. *Suppose that* (3.1) *holds. Then*

$$\Theta_{n2} + \Theta_{n3} + \Theta_{n4} \asymp \sum_k \|\mu_k\|^2. \tag{A.14}$$

Moreover, under the null hypothesis, $\Theta_n \asymp K\|\mu\|^2$.

The next result is useful in proving that our variance estimator V is asymptotically unbiased.

Lemma A.7. *Suppose that* (3.1) *holds, and recall the definition of Θ_n in* (2.6)*. Define*

$$\beta_n = \frac{\max\Big\{ \sum_k \sum_{i\in S_k} \frac{N_i^2}{n_k^2 \bar{N}_k^2} \|\Omega_i\|_3^3\,,\ \sum_k \|\Sigma_k\|_F^2 \Big\}}{K\|\mu\|^2}. \tag{A.15}$$

If $\beta_n = o(1)$, then under the null hypothesis, $\mathrm{Var}(T) = [1 + o(1)] \cdot \Theta_n$.

We also study the case of $K = 2$ more explicitly. In the lemmas below we use the notation from Section 3.4. First we have an intermediate result analogous to Lemma A.6 that holds under weaker conditions.

Lemma A.8. *Consider $K = 2$ and suppose that* $\min N_i \geq 2$, $\min M_i \geq 2$ *Then*

$$\Theta_{n2} + \Theta_{n3} + \Theta_{n4} \asymp \Big\| \frac{m\bar{M}}{n\bar{N} + m\bar{M}}\eta + \frac{n\bar{N}}{n\bar{N} + m\bar{M}}\theta \Big\|^2.$$

Moreover, under the null hypothesis, $\Theta_n \asymp \|\mu\|^2$.

The next result is a version of Lemma A.7 for the case $K = 2$ that holds under weaker conditions.

Lemma A.9. *Suppose that* $\min_i N_i \geq 2$ *and* $\min_i M_i \geq 2$*. Define*

$$\beta_n^{(2)} = \frac{\max\Big\{ \sum_i N_i^2 \|\Omega_i\|^3,\ \sum_i M_i^2 \|\Gamma_i\|^3\,,\ \|\Sigma_1\|_F^2 + \|\Sigma_2\|_F^2 \Big\}}{\|\mu\|^2}. \tag{A.16}$$

If $\beta_n^{(2)} = o(1)$, then under the null hypothesis, $\mathrm{Var}(T) = [1 + o(1)] \cdot \Theta_n$.

A.3 The decomposition of V

Lemma A.10. *Let $\{Z_{ir}\}_{1\leq i\leq n, 1\leq r\leq N_i}$ be as in* (A.4)*. Recall that*

$$V = 2\sum_{k=1}^{K}\sum_{i\in S_k}\sum_{j=1}^{p}\Big(\frac{1}{n_k\bar{N}_k} - \frac{1}{n\bar{N}}\Big)^2\Big[\frac{N_iX_{ij}^2}{N_i-1} - \frac{N_iX_{ij}(N_i-X_{ij})}{(N_i-1)^2}\Big] \tag{A.17}$$

$$+ \frac{2}{n^2\bar{N}^2}\sum_{1\leq k\neq \ell\leq K}\sum_{i\in S_k}\sum_{m\in S_\ell}\sum_{j=1}^{p}X_{ij}X_{mj} + 2\sum_{k=1}^{K}\sum_{\substack{i\in S_k, m\in S_k,\\ i\neq m}}\sum_{j=1}^{p}\Big(\frac{1}{n_k\bar{N}_k} - \frac{1}{n\bar{N}}\Big)^2 X_{ij}X_{mj}.$$

Define

$$\theta_i = (\frac{1}{n_k\bar{N}_k} - \frac{1}{n\bar{N}})^2\frac{N_i^3}{N_i-1} \quad \text{for } i\in S_k\ , \quad \text{and let}$$

$$\alpha_{im} = \begin{cases} \frac{2}{n^2\bar{N}^2} & \text{if } i\in S_k, m\in S_\ell, k\neq \ell \\ 2(\frac{1}{n_k\bar{N}_k} - \frac{1}{n\bar{N}})^2 & \text{if } i, m\in S_k \end{cases}$$

If we let

$$A_1 = \sum_i\sum_{r=1}^{N_i}\sum_j\big[\frac{4\theta_i\Omega_{ij}}{N_i} + \sum_{m\in[n]\setminus\{i\}}2\alpha_{im}N_m\Omega_{mj}\big]Z_{ijr}, \tag{A.18}$$

$$A_2 = \sum_i\sum_{r\neq s\in[N_i]}\frac{2\theta_i}{N_i(N_i-1)}\big(\sum_j Z_{ijr}Z_{ijs}\big) \tag{A.19}$$

$$A_3 = \sum_{i\neq m}\sum_{r=1}^{N_i}\sum_{s=1}^{N_m}\alpha_{im}\big(\sum_j Z_{ijr}Z_{mjs}\big), \tag{A.20}$$

then these terms are mean zero, are mutually uncorrelated, and satisfy

$$V = A_1 + A_2 + A_3 + \Theta_{n2} + \Theta_{n3} + \Theta_{n4}. \tag{A.21}$$

A.4 Properties of V

First we control the variance of V.

Lemma A.11. *Let A_1, A_2, and A_3 be defined as in Lemma A.10. Then*

$$\mathrm{Var}(A_1) \lesssim \frac{1}{n\bar{N}}\|\mu\|_3^3 + \sum_k\frac{\|\mu_k\|_3^3}{n_k\bar{N}_k} \lesssim \sum_k\frac{\|\mu_k\|_3^3}{n_k\bar{N}_k}$$

$$\mathrm{Var}(A_2) \lesssim \sum_k\sum_{i\in S_k}\frac{N_i^2\|\Omega_i\|_2^2}{n_k^4\bar{N}_k^4} \lesssim \sum_k\frac{\|\mu_k\|^2}{n_k^2\bar{N}_k^2}$$

$$\mathrm{Var}(A_3) \lesssim \sum_k\frac{\|\mu_k\|^2}{n_k^2\bar{N}_k^2} + \frac{1}{n^2\bar{N}^2}\|\mu\|^2 \lesssim \sum_k\frac{\|\mu_k\|^2}{n_k^2\bar{N}_k^2}.$$

Next we show consistency of V under the null, which is crucial in properly standardizing our test statistic and establishing asymptotic normality.

Proposition A.1. *Recall the definition of* β_n *in* (A.15)*. Suppose that* $\beta_n = o(1)$ *and that the condition* (3.1) *holds. If under the null hypothesis we have*

$$K^2\|\mu\|^4 \gg \sum_k \frac{\|\mu\|^2}{n_k^2\bar{N}_k^2} \vee \sum_k \frac{\|\mu\|_3^3}{n_k\bar{N}_k}, \tag{A.22}$$

then $V/\mathrm{Var}T \to 1$ *in probability.*

To later control the type II error, we must also show that V does not dominate the true variance under the alternative. We first state an intermediate result that is useful throughout.

Lemma A.12. *Suppose that, under either the null or alternative,* $\max_i \|\Omega_i\|_\infty \leq 1 - c_0$ *holds for an absolute constant* $c_0 > 0$*. Then*

$$\mathrm{Var}(T) \gtrsim \Theta_{n2} + \Theta_{n3} + \Theta_{n4}. \tag{A.23}$$

Proposition A.2. *Suppose that under the alternative* (3.1) *holds and*

$$\big(\sum_k \|\mu_k\|^2\big)^2 \gg \sum_k \frac{\|\mu_k\|^2}{n_k^2\bar{N}_k^2} \vee \sum_k \frac{\|\mu_k\|_3^3}{n_k\bar{N}_k}. \tag{A.24}$$

Then $V = O_{\mathbb{P}}(\mathrm{Var}(T))$ *under the alternative.*

We also require versions of Proposition A.1 and Proposition A.2 that hold under weaker conditions in the special case $K = 2$. We omit the proofs as they are similar. Below we use the notation of Section 3.4.

Proposition A.3. *Suppose that* $K = 2$ *and recall the definition of* $\beta_n^{(2)}$ *in A.16. Suppose that* $\beta_n^{(2)} = o(1)$*,* $\min_i N_i \geq 2, \min_i M_i \geq 2$*, and* $\max_i \|\Omega_i\|_\infty \leq 1 - c_0, \max_i \|\Gamma_i\|_\infty \leq 1 - c_0$*. If under the null hypothesis*

$$\|\mu\|^4 \gg \max\Big\{\big(\frac{\|\mu\|_2^2}{n^2\bar{N}^2} + \frac{\|\mu\|_2^2}{m^2\bar{M}_2^2}\big), \big(\frac{\|\mu\|_3^3}{n\bar{N}} + \frac{\|\mu\|_3^3}{m\bar{M}}\big)\Big\}, \tag{A.25}$$

then $V/\mathrm{Var}(T) \to 1$ *in probability.*

Under the alternative we have the following.

Proposition A.4. *Suppose that* $K = 2$*,* $\min_i N_i \geq 2, \min_i M_i \geq 2$*, and* $\max_i \|\Omega_i\|_\infty \leq 1 - c_0, \max_i \|\Gamma_i\|_\infty \leq 1 - c_0$*. If under the alternative*

$$\Big\|\frac{m\bar{M}}{n\bar{N} + m\bar{M}}\eta + \frac{n\bar{N}}{n\bar{N} + m\bar{M}}\theta\Big\|^4 \gg \max\Big\{\big(\frac{\|\eta\|_2^2}{n^2\bar{N}^2} + \frac{\|\theta\|_2^2}{m^2\bar{M}_2^2}\big), \big(\frac{\|\eta\|_3^3}{n\bar{N}} + \frac{\|\theta\|_3^3}{m\bar{M}}\big)\Big\}, \tag{A.26}$$

then $V = O_{\mathbb{P}}(\mathrm{Var}(T))$*.*

In the setting of $K = n$ and utilize the variance estimator V^*. The next results capture the behavior of V^* under the null and alternative. The proofs are given later in this section.

Proposition A.5. *Define*

$$\beta_n^{(n)} = \frac{\sum_i \|\Omega_i\|^3}{n\|\mu\|^2}. \tag{A.27}$$

Suppose that (3.1) *holds,* $\beta_n^{(n)} = o(1)$, *and*

$$n^2\|\mu\|^4 \gg \sum_i \frac{\|\mu\|^2}{N_i^2} \vee \sum_i \frac{\|\mu\|_3^3}{N_i}. \tag{A.28}$$

Then $V^*/\mathrm{Var}(T) \to 1$ *in probability as* $n \to \infty$.

Proposition A.6. *Suppose that under the alternative* (3.1) *holds and*

$$\big(\sum_i \|\Omega_i\|^2\big)^2 \gg \sum_i \frac{\|\Omega_i\|^2}{N_i^2} \vee \sum_i \frac{\|\Omega_i\|_3^3}{N_i}. \tag{A.29}$$

Then $V^* = O_{\mathbb{P}}(\mathrm{Var}(T))$ *under the alternative.*

A.5 Proof of Lemma A.1

We first show that $\mathbb{E}[U_\kappa] = \mathbf{0}_p$ and $\mathbb{E}[U_\kappa U_\zeta'] = \mathbf{0}_{p\times p}$ for $\kappa \neq \zeta$. Note that $\{Z_{ir}\}_{1\leq i\leq n, 1\leq r\leq N_i}$ are independent mean-zero random vectors. It follows that each U_κ is a mean-zero random vector. We then compute $\mathbb{E}[U_{\kappa j_1}U_{\zeta j_2}]$ for $\kappa \neq \zeta$ and all $1 \leq j_1, j_2 \leq p$. By direct calculations,

$$\mathbb{E}[U_{1j}U_{2j_2}] = 2\sum_{(k,i,r,s)}\sum_{(k',i',r')}\Big(\frac{1}{n_k\bar{N}_k} - \frac{1}{n\bar{N}}\Big)(\mu_{k'j} - \mu_j)\frac{N_i}{N_i - 1}\mathbb{E}[Z_{ij_2r}Z_{ij_2s}Z_{i'j_1r'}].$$

If $i' \neq i$, or if $i' = i$ and $r' \notin \{r, s\}$, then $Z_{i'j_1r'}$ is independent of $Z_{ij_2r}Z_{ij_2s}$, and it follows that $\mathbb{E}[Z_{ij_2r}Z_{ij_2s}Z_{i'j_1r'}] = 0$. If $i' = i$ and $r = r'$, then $\mathbb{E}[Z_{ij_2r}Z_{ij_2s}Z_{i'j_1r'}] = \mathbb{E}[Z_{ij_2r}Z_{ij_1r}]\cdot\mathbb{E}[Z_{ij_2s}]$; since $r \neq s$, we also have $\mathbb{E}[Z_{ij_2r}Z_{ij_2s}Z_{i'j_1r'}] = 0$. This proves $\mathbb{E}[U_{1j}U_{2j_*}] = 0$. Since this holds for all $1 \leq j_1, j_2 \leq p$, we immediately have

$$\mathbb{E}[U_1U_2'] = \mathbf{0}_{p\times p}.$$

We can similarly show that $\mathbb{E}[U_\kappa U_\zeta'] = \mathbf{0}_{p\times p}$, for other $\kappa \neq \zeta$. The proof is omitted.

It remains to prove the desirable decomposition of T. Recall that $T = \sum_{j=1}^p T_j$. Write $\rho^2 = \sum_{j=1}^p \rho_j^2$, where $\rho_j^2 = 2\sum_{k=1}^K n_k\bar{N}_k(\mu_{kj} - \mu_j)^2$. It suffices to show that

$$T_j = \rho_j^2 + U_{1j} + U_{2j} + U_{3j} + U_{4j}, \qquad \text{for all } 1 \leq j \leq p. \tag{A.30}$$

To prove (A.30), we need some preparation. Define

$$Y_{ij} := \frac{X_{ij}}{N_i} - \Omega_{ij} = \frac{1}{N_i}\sum_{r=1}^{N_i} Z_{ijr}, \qquad Q_{ij} := Y_{ij}^2 - \mathbb{E}Y_{ij}^2 = Y_{ij}^2 - \frac{\Omega_{ij}(1-\Omega_{ij})}{N_i}. \tag{A.31}$$

With these notations, $X_{ij} = N_i(\Omega_{ij} + Y_{ij})$ and $N_iY_{ij}^2 = N_iQ_{ij} + \Omega_{ij}(1 - \Omega_{ij})$. Moreover, we can use (A.31) to re-write Q_{ij} as a function of $\{Z_{ijr}\}_{1\leq r\leq N_i}$ as follows:

$$Q_{ij} = \frac{1}{N_i^2}\sum_{r=1}^{N_i}[Z_{ijr}^2 - \Omega_{ij}(1 - \Omega_{ij})] + \frac{1}{N_i^2}\sum_{1\leq r\neq s\leq N_i} Z_{ijr}Z_{ijs}.$$

Note that $Z_{ijr} = B_{ijr} - \Omega_{ij}$, where B_{ijr} can only take values in $\{0,1\}$. Hence, $(Z_{ijr}+\Omega_{ij})^2 = (Z_{ijr}+\Omega_{ij})$ always holds. Re-arranging the terms gives $Z_{ijr}^2 - \Omega_{ij}(1-\Omega_{ij}) = (1-2\Omega_{ij})Z_{ijr}$. It follows that

$$Q_{ij} = (1-2\Omega_{ij})\frac{Y_{ij}}{N_i} + \frac{1}{N_i^2}\sum_{1\le r\neq s\le N_i} Z_{ijr}Z_{ijs}. \tag{A.32}$$

This is a useful equality which we will use in the proof below.

We now show (A.30). Fix j and write $T_j = R_j - D_j$, where

$$R_j = \sum_{k=1}^K n_k\bar{N}_k(\hat{\mu}_{kj}-\hat{\mu}_j)^2, \quad \text{and} \quad D_j = \sum_{k=1}^K \sum_{i\in S_k} \xi_k \frac{X_{ij}(N_i - X_{ij})}{n_k\bar{N}_k(N_i-1)}, \quad \text{with } \xi_k = 1 - \frac{n_k\bar{N}_k}{n\bar{N}}$$

First, we study D_j. Note that $X_{ij}(N_{ij}-X_{ij}) = N_i^2(\Omega_{ij}+Y_{ij})(1-\Omega_{ij}-Y_{ij}) = N_i^2\Omega_{ij}(1-\Omega_{ij}) - N_i^2Y_{ij}^2 + N_i^2(1-2\Omega_{ij})Y_{ij}$, where $Y_{ij}^2 = Q_{ij} + N_i^{-1}\Omega_{ij}(1-\Omega_{ij})$. It follows that

$$\frac{X_{ij}(N_{ij}-X_{ij})}{N_i(N_i-1)} = \Omega_{ij}(1-\Omega_{ij}) - \frac{N_iQ_{ij}}{N_i-1} + \frac{N_i}{N_i-1}(1-2\Omega_{ij})Y_{ij}.$$

We apply (A.32) to get

$$\frac{X_{ij}(N_{ij}-X_{ij})}{N_i(N_i-1)} = \Omega_{ij}(1-\Omega_{ij}) + (1-2\Omega_{ij})Y_{ij} - \frac{1}{N_i(N_i-1)}\sum_{1\le r\neq s\le N_i} Z_{ijr}Z_{ijs}. \tag{A.33}$$

It follows that

$$\begin{aligned} D_j = &\sum_{k=1}^K \sum_{i\in S_k} \frac{\xi_k N_i}{n_k\bar{N}_k}\Omega_{ij}(1-\Omega_{ij}) + \sum_{k=1}^K \sum_{i\in S_k} \frac{\xi_k N_i}{n_k\bar{N}_k}(1-2\Omega_{ij})Y_{ij} \\ &- \sum_{k=1}^K \sum_{i\in S_k} \frac{\xi_k}{n_k\bar{N}_k(N_i-1)} \sum_{1\le r\neq s\le N_i} Z_{ijr}Z_{ijs}. \end{aligned} \tag{A.34}$$

Next, we study R_j. Note that $n_k\bar{N}_k(\hat{\mu}_{kj}-\hat{\mu}_j) = \sum_{i\in S_k}(X_{ij} - \bar{N}_k\hat{\mu}_j)$. It follows that

$$R_j = \sum_{k=1}^K \frac{1}{n_k\bar{N}_k}\Big[\sum_{i\in S_k}(X_{ij}-\bar{N}_k\hat{\mu}_j)\Big]^2.$$

Recall that $X_{ij} = N_i(\Omega_{ij}+Y_{ij})$. By direct calculations, $\sum_{i\in S_k} X_{ij} = n_k\bar{N}_k\mu_{kj} + \sum_{i\in S_k} N_iY_{ij}$, and $\hat{\mu}_j = \mu_j + (n\bar{N})^{-1}\sum_{m=1}^n N_mY_{mj}$. We then have the following decomposition:

$$\sum_{i\in S_k}(X_{ij}-\bar{N}_k\hat{\mu}_j) = n_k\bar{N}_k(\mu_{kj}-\mu_j) + \sum_{i\in S_k} N_iY_{ij} - \frac{n_k\bar{N}_k}{n\bar{N}}\Big(\sum_{m=1}^n N_mY_{mj}\Big).$$

Using this decomposition, we can expand $[\sum_{i\in S_k}(X_{ij}-\bar{N}_k\hat{\mu}_j)]^2$ to a total of 6 terms, where 3 are quadratic terms and 3 are cross terms. It yields a decomposition of R_j into 6 terms:

$$R_j = \sum_{k=1}^K n_k\bar{N}_k(\mu_{kj}-\mu_j)^2 + \sum_{k=1}^K \frac{1}{n_k\bar{N}_k}\Big(\sum_{i\in S_k} N_iY_{ij}\Big)^2 + \sum_{k=1}^K \frac{n_k\bar{N}_k}{n^2\bar{N}^2}\Big(\sum_{m=1}^n N_mY_{mj}\Big)^2$$

$$
\begin{aligned}
&+ 2\sum_{k=1}^{K}(\mu_{kj}-\mu_j)\Big(\sum_{i\in S_k} N_iY_{ij}\Big) - 2\sum_{k=1}^{K}\frac{n_k\bar{N}_k}{n\bar{N}}(\mu_{kj}-\mu_j)\Big(\sum_{m=1}^{n} N_mY_{mj}\Big)\\
&- \frac{2}{n\bar{N}}\sum_{k=1}^{K}\Big(\sum_{i\in S_k} N_iY_{ij}\Big)\Big(\sum_{m=1}^{n} N_mY_{mj}\Big)\\
\equiv{}& I_1+I_2+I_3+I_4+I_5+I_6.
\end{aligned}
\tag{A.35}
$$

By definition, $\sum_{k=1}^K n_k\bar{N}_k = n\bar{N}$ and $\sum_{k=1}^K n_k\bar{N}_k\mu_{kj} = n\bar{N}\mu_j$. It follows that

$$
I_3 = \frac{1}{n\bar{N}}\Big(\sum_{m=1}^{n} N_mY_{mj}\Big)^2, \qquad I_5 = 0, \qquad I_6 = -\frac{2}{n\bar{N}}\Big(\sum_{m=1}^{n} N_mY_{mj}\Big)^2 = -2I_3.
$$

It follows that

$$
R_j = I_1 + I_2 - I_3 + I_4. \tag{A.36}
$$

We further simplify I_3. Recall that $\xi_k = 1-(n\bar{N})^{-1}n_k\bar{N}_k$. By direct calculations,

$$
\begin{aligned}
I_3 &= \frac{1}{n\bar{N}}\Big(\sum_{m=1}^{n} N_mY_{mj}\Big)^2 = \frac{1}{n\bar{N}}\Big[\sum_{k=1}^{K}\Big(\sum_{i\in S_k} N_iY_{ij}\Big)\Big]^2\\
&= \frac{1}{n\bar{N}}\sum_{k=1}^{K}\Big(\sum_{i\in S_k} N_iY_{ij}\Big)^2 + \frac{1}{n\bar{N}}\sum_{1\le k\neq \ell\le K}\Big(\sum_{i\in S_k} N_iY_{ij}\Big)\Big(\sum_{m\in S_\ell} N_mY_{mj}\Big)\\
&= \sum_{k=1}^{K}(1-\xi_k)\frac{1}{n_k\bar{N}_k}\Big(\sum_{i\in S_k} N_iY_{ij}\Big)^2 + \underbrace{\frac{1}{n\bar{N}}\sum_{k\neq\ell}\sum_{i\in S_k}\sum_{m\in S_\ell} N_iN_mY_{ij}Y_{mj}}_{J_1}\\
&= I_2 - \sum_{k=1}^{K}\sum_{i\in S_k}\frac{\xi_k}{n_k\bar{N}_k}\Big(\sum_{i\in S_k} N_iY_{ij}\Big)^2 + J_1\\
&= I_2 + J_1 - \sum_{k=1}^{K}\frac{\xi_k}{n_k\bar{N}_k}\Big(\sum_{i\in S_k} N_i^2Y_{ij}^2\Big) - \underbrace{\sum_{k=1}^{K}\frac{\xi_k}{n_k\bar{N}_k}\sum_{\substack{i\in S_k, m\in S_k\\ i\neq m}} N_iN_mY_{ij}Y_{mj}}_{J_2}.
\end{aligned}
\tag{A.37}
$$

By (A.31), $N_iY_{ij}^2 = N_iQ_i + \Omega_{ij}(1-\Omega_{ij})$. We further apply (A.32) to get

$$
N_i^2Y_{ij}^2 = N_i(1-2\Omega_{ij})Y_{ij} + \sum_{1\le r\neq s\le N_i} Z_{ijr}Z_{ijs} + N_i\Omega_{ij}(1-\Omega_{ij}).
$$

It follows that

$$
\begin{aligned}
\sum_{k=1}^{K}\frac{\xi_k}{n_k\bar{N}_k}\Big(\sum_{i\in S_k} N_i^2Y_{ij}^2\Big) &= \underbrace{\sum_{k=1}^{K}\sum_{i\in S_k}\frac{\xi_kN_i}{n_k\bar{N}_k}(1-2\Omega_{ij})Y_{ij}}_{J_3}\\
&\quad + \underbrace{\sum_{k=1}^{K}\sum_{i\in S_k}\frac{\xi_k}{n_k\bar{N}_k}\sum_{r\neq s} Z_{ijr}Z_{ijs}}_{J_4} + \underbrace{\sum_{k=1}^{K}\sum_{i\in S_k}\frac{\xi_kN_i}{n_k\bar{N}_k}\Omega_{ij}(1-\Omega_{ij})}_{J_5}.
\end{aligned}
\tag{A.38}
$$

We plug (A.38) into (A.37) to get $I_3 = I_2 + J_1 - J_2 - J_3 - J_4 - J_5$. Further plugging I_3 into the expression of R_j in (A.36), we have

$$R_j = I_1 + I_4 - J_1 + J_2 + J_3 + J_4 + J_5, \tag{A.39}$$

where I_1 and I_4 are defined in (A.35), J_1-J_2 are defined in (A.37), and J_3-J_5 are defined in (A.38).

Finally, we combine the expressions of D_j and R_j. By (A.34) and the definitions of J_1-J_5,

$$\begin{aligned} D_j &= J_5 + J_3 - \sum_{k=1}^{K} \sum_{i\in S_k} \frac{\xi_k}{n_k \bar{N}_k (N_i - 1)} \sum_{r\neq s} Z_{ijr} Z_{ijs} \\ &= J_5 + J_3 + J_4 - \underbrace{\sum_{k=1}^{K} \sum_{i\in S_k} \frac{\xi_k N_i}{n_k \bar{N}_k (N_i - 1)} \sum_{r\neq s} Z_{ijr} Z_{ijs}}_{J_6}. \end{aligned}$$

Combining it with (A.39) gives $T_j = R_j - D_j = I_1 + I_4 - J_1 + J_2 + J_6$. We further plug in the definition of each term. It follows that

$$\begin{aligned} T_j = &\sum_{k=1}^{K} n_k \bar{N}_k (\mu_{kj} - \mu_j)^2 + 2\sum_{k=1}^{K} \sum_{i\in S_k} (\mu_{kj} - \mu_j) N_i Y_{ij} - \frac{1}{n\bar{N}} \sum_{k\neq \ell} \sum_{i\in S_k, m\in S_\ell} N_i N_m Y_{ij} Y_{mj} \\ &+ \sum_{k=1}^{K} \sum_{\substack{i\in S_k, m\in S_k \\ i\neq m}} \frac{\xi_k}{n_k \bar{N}_k} N_i N_m Y_{ij} Y_{mj} + \sum_{k=1}^{K} \sum_{i\in S_k} \frac{\xi_k N_i}{n_k \bar{N}_k (N_i - 1)} \sum_{r\neq s} Z_{ijr} Z_{ijs}. \end{aligned} \tag{A.40}$$

We plug in $Y_{ij} = N_i^{-1} \sum_{r=1}^{N_i} Z_{ijr}$ and take a sum of $1 \le j \le p$. It gives (A.30) immediately. The proof is now complete. □

A.6 Proof of Lemma A.2

Recall that $\{Z_{ir}\}_{1\le i\le n, 1\le r\le N_i}$ are independent random vectors. Write

$$\mathbf{1}_p' U_1 = 2\sum_{k=1}^{K} \sum_{i\in S_k} \sum_{r=1}^{N_i} (\mu_k - \mu)' Z_{ir}.$$

The covariance matrix of Z_{ir} is $\mathrm{diag}(\Omega_i) - \Omega_i \Omega_i'$. It follows that

$$\begin{aligned} \mathrm{Var}(\mathbf{1}_p' U_1) &= 4\sum_{k=1}^{K} \sum_{i\in S_k} \sum_{r=1}^{N_i} (\mu_k - \mu)' \big[\mathrm{diag}(\Omega_i) - \Omega_i \Omega_i'\big] (\mu_k - \mu) \\ &= 4\sum_{k} (\mu_k - \mu)' \Big[\mathrm{diag}\Big(\sum_{i\in S_k} N_i \Omega_i\Big) - \Big(\sum_{i\in S_k} N_i \Omega_i \Omega_i'\Big)\Big] (\mu_k - \mu) \\ &= 4\sum_{k} (\mu_k - \mu)' \Big[\mathrm{diag}(n_k \bar{N}_k \mu_k) - n_k \bar{N}_k \Sigma_k\Big] (\mu_k - \mu) \end{aligned}$$

$$= 4\sum_k n_k \bar{N}_k \big\| \mathrm{diag}(\mu_k)^{1/2}(\mu_k - \mu)\big\|^2 - 4\sum_k n_k \bar{N}_k \big\|\Sigma_k^{1/2}(\mu_k - \mu)\big\|^2. \quad \text{(A.41)}$$

This proves the first claim. Furthermore, by (A.41),

$$\mathrm{Var}(\mathbf{1}_p' U_1) \le 4\sum_k n_k \bar{N}_k \big\| \mathrm{diag}(\mu_k)^{1/2}(\mu_k - \mu)\big\|^2 \le 4\sum_k n_k \bar{N}_k \|\mathrm{diag}(\mu_k)\| \|\mu_k - \mu\|^2.$$

Note that $\|\mathrm{diag}(\mu_k)\| = \|\mu_k\|_\infty$. Therefore, if $\max_k \|\mu_k\|_\infty = o(1)$, the right hand side above is $o(1) \cdot 4\sum_k n_k \bar{N}_k \|\mu_k - \mu\|^2 = o(\rho^2)$. This proves the second claim. $\square$

A.7 Proof of Lemma A.3

For each $1 \le k \le K$, define a set of index triplets: $\mathcal{M}_k = \{(i, r, s) : i \in S_k, 1 \le r < s \le N_i\}$. Let $\mathcal{M} = \cup_{k=1}^K \mathcal{M}_k$. Write for short $\theta_i = (\frac{1}{n_k \bar{N}_k} - \frac{1}{nN})^2 \frac{N_i^3}{N_i - 1}$, for $i \in S_k$. It is seen that

$$\mathbf{1}_p' U_2 = 2 \sum_{(i,r,s)\in\mathcal{M}} \frac{\sqrt{\theta_i}}{\sqrt{N_i(N_i-1)}} W_{irs}, \qquad \text{with} \quad W_{irs} = \sum_{j=1}^p Z_{ijr} Z_{ijs}.$$

For W_{irs} and $W_{i'r's'}$, if $i \neq i'$, or if $i = i'$ and $\{r, s\} \cap \{r', s'\} = \emptyset$, then these two variables are independent; if $i = i'$, $r = r'$ and $s \neq s'$, then $\mathbb{E}[W_{irs} W_{irs'}] = \sum_{j,j'} \mathbb{E}[Z_{ijr} Z_{ijs} Z_{ij'r} Z_{ij's'}] = \sum_{j,j'} \mathbb{E}[Z_{ijr} Z_{ij'r}] \cdot \mathbb{E}[Z_{ijs}] \cdot \mathbb{E}[Z_{ij's'}] = 0$. Therefore, $\{W_{irs}\}_{(i,r,s)\in\mathcal{M}}$ is a collection of mutually uncorrelated variables. It follows that

$$\mathrm{Var}(\mathbf{1}_p' U_2) = 4 \sum_{(i,r,s)\in\mathcal{M}} \frac{\theta_i}{N_i(N_i-1)} \mathrm{Var}(W_{irs}).$$

It remains to calculate the variance of each W_{irs}. By direction calculations,

$$\begin{aligned}
\mathrm{Var}(W_{irs}) &= \sum_j \mathbb{E}[Z_{ijr}^2 Z_{ijs}^2] + 2\sum_{j<\ell} \mathbb{E}[Z_{ijr} Z_{ijs} Z_{i\ell r} Z_{i\ell s}] \\
&= \sum_j [\Omega_{ij}(1 - \Omega_{ij})]^2 + 2\sum_{j<\ell} (-\Omega_{ij}\Omega_{i\ell})^2 \\
&= \sum_j \Omega_{ij}^2 - 2\sum_j \Omega_{ij}^3 + \Big(\sum_j \Omega_{ij}^2\Big)^2 \\
&= \|\Omega_i\|^2 - 2\|\Omega_i\|_3^3 + \|\Omega_i\|^4
\end{aligned} \quad \text{(A.42)}$$

Since $\max_{ij} \Omega_{ij} \le 1$, we have

$$\|\Omega_i\|^2 - \|\Omega_i\|_3^3 \le \mathrm{Var}(W_{irs}) \le \|\Omega_i\|^2.$$

Therefore,

$$\mathrm{Var}(\mathbf{1}_p' U_2) = 4 \sum_{k=1}^K \sum_{i\in S_k} \sum_{1\le r<s\le N_i} \frac{\theta_i}{N_i(N_i-1)} \mathrm{Var}(W_{irs})$$

$$= 2\sum_{k=1}^{K}\sum_{i\in S_k}\theta_i \mathrm{Var}(W_{irs}) \geq 2\sum_{k=1}^{K}\sum_{i\in S_k}\theta_i\big[\|\Omega_i\|^2 - \|\Omega_i\|_3^3\big] = \Theta_{n2} - A_n,$$

and similarly $\mathrm{Var}(\mathbf{1}_p'U_2) \leq \Theta_{n2}$, which proves the first claim. To prove the second claim, note that $\mathrm{Var}(\mathbf{1}_p'U_2) = \Theta_{n2} + O(A_n)$. By (A.9) and the assumption $\min N_i \geq 2$, we have

$$\begin{aligned} A_n &\lesssim \sum_k \big(\frac{1}{n_k\bar{N}_k} - \frac{1}{n\bar{N}}\big)^2 \sum_{i\in S_k} N_i^2\|\Omega_i\|_3^3 \\ &= \sum_k \big(\frac{1}{n_k\bar{N}_k} - \frac{1}{n\bar{N}}\big)^2 \cdot o\Big(\sum_{i\in S_k} N_i^2\|\Omega_i\|^2\Big) = o(\Theta_{n2}), \end{aligned}$$

which implies that $\mathrm{Var}(\mathbf{1}_pU_2) = [1+o(1)]\Theta_{n2}$, as desired.

□

A.8 Proof of Lemma A.4

For each $1 \leq k < \ell \leq K$, define a set of index quadruples: $\mathcal{J}_{k\ell} = \{(i,r,m,s) : i \in S_k, j \in S_\ell, 1 \leq r \leq N_i, 1 \leq s \leq N_m\}$. Let $\mathcal{J} = \cup_{(k,\ell):1\leq k<\ell\leq K}\mathcal{J}_{k\ell}$. It is seen that

$$\mathbf{1}_p'U_3 = -\frac{2}{n\bar{N}}\sum_{(i,r,m,s)\in\mathcal{J}} V_{irms}, \qquad \text{where } V_{irms} = \sum_{j=1}^{p} Z_{ijr}Z_{mjs}.$$

For V_{irms} and $V_{i'r'm's'}$, if $\{(i,r),(m,s)\}\cap\{(i',r'),(m',s')\} = \emptyset$, then the two variables are independent of each other. If $(i,r) = (i',r')$ and $(m,s) \neq (m',s')$, then $\mathbb{E}[V_{irms}V_{irm's'}] = \sum_{j,j'}\mathbb{E}[Z_{ijr}Z_{mjs}Z_{ij'r}Z_{m'j's'}] = \sum_{j,j'}\mathbb{E}[Z_{ijr}Z_{ij'r}]\cdot\mathbb{E}[Z_{mjs}]\cdot\mathbb{E}[Z_{m'js'}] = 0$. Therefore, the only correlated case is when $(i,r,m,s) = (i',r',m',s')$. This implies that $\{V_{irms}\}_{(i,r,m,s)\in\mathcal{J}}$ is a collection of mutually uncorrelated variables. Therefore,

$$\mathrm{Var}(\mathbf{1}_p'U_3) = \frac{4}{n^2\bar{N}^2}\sum_{(i,r,m,s)\in\mathcal{J}}\mathrm{Var}(V_{irms}).$$

Note that $\mathrm{Var}(V_{irms}) = \mathbb{E}[(\sum_j Z_{ijr}Z_{mjs})^2] = \sum_{j,j'}\mathbb{E}[Z_{ijr}Z_{mjs}Z_{ij'r}Z_{mj's}]$; also, the covariance matrix of Z_{ir} is $\mathrm{diag}(\Omega_i) - \Omega_i\Omega_i'$. It follows that

$$\begin{aligned} \mathrm{Var}(V_{irms}) &= \sum_j \mathbb{E}[Z_{ijr}^2]\cdot\mathbb{E}[Z_{mjs}^2] + \sum_{j\neq j'}\mathbb{E}[Z_{ijr}Z_{ij'r}]\cdot\mathbb{E}[Z_{mjs}Z_{mj's}] \\ &= \sum_j \Omega_{ij}(1-\Omega_{ij})\Omega_{mj}(1-\Omega_{mj}) + \sum_{j\neq j'}\Omega_{ij}\Omega_{ij'}\Omega_{mj}\Omega_{mj'} \\ &= \sum_j \Omega_{ij}\Omega_{mj} - 2\sum_j \Omega_{ij}^2\Omega_{mj}^2 + \sum_{j,j'}\Omega_{ij}\Omega_{ij'}\Omega_{mj}\Omega_{mj'}. \end{aligned} \tag{A.43}$$

Write for short $\delta_{im} = -2\sum_j \Omega_{ij}^2\Omega_{mj}^2 + \sum_{j,j'}\Omega_{ij}\Omega_{ij'}\Omega_{mj}\Omega_{mj'}$.Combining the above gives

$$\mathrm{Var}(\mathbf{1}_p'U_3) = \frac{4}{n^2\bar{N}^2}\sum_{k<\ell}\sum_{i\in S_k}\sum_{m\in S_\ell}\sum_{r=1}^{N_i}\sum_{s=1}^{N_m}\Big(\sum_j \Omega_{ij}\Omega_{mj} + \delta_{im}\Big)$$

$$= \frac{2}{n^2\bar{N}^2}\sum_{k\neq\ell}\sum_{i\in S_k}\sum_{m\in S_\ell}\sum_j N_iN_m\Omega_{ij}\Omega_{mj} + \frac{2}{n^2\bar{N}^2}\sum_{k\neq\ell}\sum_{i\in S_k}\sum_{m\in S_\ell} N_iN_m\delta_{im}. \quad \text{(A.44)}$$

It is easy to see that $|\delta_{im}| \leq \sum_{j,j'}\Omega_{ij}\Omega_{ij'}\Omega_{mj}\Omega_{mj'}$. Also, by the definition of Σ_k in (A.2), we have $\Sigma_k(j,j') = \frac{1}{n_k\bar{N}_k}\sum_{i\in S_k} N_i\Omega_{ij}\Omega_{ij'}$. Using these results, we immediately have

$$\begin{aligned}
\Big|\frac{2}{n^2\bar{N}^2}\sum_{k\neq\ell}\sum_{i\in S_k}\sum_{m\in S_\ell} N_iN_m\delta_{im}\Big| &\leq \frac{2}{n^2\bar{N}^2}\sum_{k\neq\ell}\sum_{i\in S_k}\sum_{m\in S_\ell}\sum_{j,j'} N_iN_m\Omega_{ij}\Omega_{ij'}\Omega_{mj}\Omega_{mj'} \\
&= \frac{2}{n^2\bar{N}^2}\sum_{j,j'}\sum_{k\neq\ell}\Big(\sum_{i\in S_k} N_i\Omega_{ij}\Omega_{ij'}\Big)\Big(\sum_{m\in S_\ell} N_i\Omega_{mj}\Omega_{mj'}\Big) \\
&= \frac{2}{n^2\bar{N}^2}\sum_{j,j'}\sum_{k\neq\ell} n_k\bar{N}_k\Sigma_k(j,j')\cdot n_\ell\bar{N}_\ell\Sigma_\ell(j,j') \\
&= 2\sum_{k\neq\ell}\frac{n_kn_\ell\bar{N}_k\bar{N}_\ell}{n^2\bar{N}^2}\mathbf{1}_p'(\Sigma_k\circ\Sigma_\ell)\mathbf{1}_p =: B_n \qquad \text{(A.45)}
\end{aligned}$$

as desired.

□

A.9 Proof of Lemma A.5

For $1 \leq k \leq K$, define a set of index quadruples: $\mathcal{Q}_k = \{(i,r,m,s) : i \in S_k, m \in S_k, i < m, 1 \leq r \leq N_i, 1 \leq s \leq N_m\}$. Let $\mathcal{Q} = \cup_{k=1}^K \mathcal{Q}_k$. Write $\kappa_{im} = (\frac{1}{n_k\bar{N}_k} - \frac{1}{n\bar{N}})^2 N_iN_m$, for $i \in S_k$ and $m \in S_k$. It is seen that

$$\mathbf{1}_p'U_4 = 2\sum_{(i,r,m,s)\in\mathcal{Q}}\frac{\sqrt{\kappa_{im}}}{\sqrt{N_iN_m}}V_{irms}, \qquad \text{where} \quad V_{irms} = \sum_{j=1}^p Z_{ijr}Z_{mjs}.$$

It is not hard to see that V_{irms} and $V_{i'r'm's'}$ are correlated only if $(i,r,m,s) = (i',r',m',s')$. It follows that

$$\mathrm{Var}(\mathbf{1}_p'U_4) = 4\sum_{(i,r,m,s)\in\mathcal{Q}}\frac{\kappa_{im}}{N_iN_m}\mathrm{Var}(V_{irms}).$$

In the proof of Lemma A.4, we have studied $\mathrm{Var}(V_{irms})$. In particular, by (A.43), we have

$$\mathrm{Var}(V_{irms}) = \sum_j \Omega_{ij}\Omega_{mj} + \delta_{im}, \qquad \text{with} \quad |\delta_{im}| \leq \sum_{j,j'}\Omega_{ij}\Omega_{ij'}\Omega_{mj}\Omega_{mj'}.$$

Thus

$$\begin{aligned}
\mathrm{Var}(\mathbf{1}_p'U_4) &= 4\sum_{k=1}^K\sum_{\substack{i\in S_k, m\in S_k\\ i<m}}\sum_{i=1}^{N_i}\sum_{r=1}^{N_m}\frac{\kappa_{im}}{N_iN_m}\mathrm{Var}(V_{irms}) \\
&= 4\sum_{k=1}^K\sum_{\substack{i\in S_k, m\in S_k\\ i<m}}\kappa_{im}\Big(\sum_j\Omega_{ij}\Omega_{mj} + \delta_{im}\Big)
\end{aligned}$$

$$= 2\sum_{k=1}^{K}\sum_{\substack{i\in S_k, m\in S_k\\ i\neq m}}\sum_j \kappa_{im}\Omega_{ij}\Omega_{mj} \pm 2\sum_k\sum_{i\neq m\in S_k}\kappa_{im}\sum_{j,j'}\Omega_{ij}\Omega_{ij'}\Omega_{mj}\Omega_{mj'},$$
$$= \Theta_{n3} \pm E_n. \tag{A.46}$$

which proves the lemma.

□

A.10 Proof of Lemma A.6

By assumption (3.1), $N_i^3/(N_i-1) \asymp N_i$ and $\left(\frac{1}{n_k\bar{N}_k} - \frac{1}{n\bar{N}}\right)^2 \asymp \frac{1}{n_k^2\bar{N}_k^2}$. First, observe that

$$\begin{aligned}\Theta_{n2} + \Theta_{n4} &= 2\sum_{k=1}^{K}\Big(\frac{1}{n_k\bar{N}_k} - \frac{1}{n\bar{N}}\Big)^2\sum_{i\in S_k}\frac{N_i^3}{N_i-1}\|\Omega_i\|^2 \\ &\quad + 2\sum_{k=1}^{K}\sum_{\substack{i\in S_k, m\in S_k\\ i\neq m}}\sum_j\Big(\frac{1}{n_k\bar{N}_k} - \frac{1}{n\bar{N}}\Big)^2 N_iN_m\Omega_{ij}\Omega_{mj} \\ &\asymp \sum_{k=1}\Big(\frac{1}{n_k\bar{N}_k}\Big)^2\sum_j\sum_{i,m\in S_k}N_i\Omega_{ij}\cdot N_m\Omega_{ij} = \sum_k\|\mu_k\|^2.\end{aligned} \tag{A.47}$$

Recall the definitions of μ_k and μ in (A.2)-(A.3). By direct calculations, we have

$$\begin{aligned}\Theta_{n3} &= 2\sum_j\sum_{k\neq\ell}\Big(\frac{1}{n\bar{N}}\sum_{i\in S_k}N_i\Omega_{ij}\Big)\Big(\frac{1}{n\bar{N}}\sum_{m\in S_\ell}N_m\Omega_{mj}\Big) \\ &= 2\sum_j\sum_{k\neq\ell}\frac{n_k\bar{N}_k}{n\bar{N}}\mu_{kj}\cdot\frac{n_\ell\bar{N}_\ell}{n\bar{N}}\mu_{\ell j} \\ &= 2\sum_{k\neq\ell}\frac{n_kn_\ell\bar{N}_k\bar{N}_\ell}{n^2\bar{N}^2}\cdot\mu_k'\mu_\ell \\ &\leq 2\sum_j\Big(\sum_k\frac{n_k\bar{N}_k}{n\bar{N}}\mu_{kj}\Big)^2 = 2\sum_j\mu_j^2 = 2\|\mu\|^2.\end{aligned} \tag{A.48}$$

By Cauchy–Schwarz,

$$\begin{aligned}\|\mu\|^2 &= \sum_j\left(\sum_k(\frac{n_k\bar{N}_k}{n\bar{N}})\mu_{kj}\right)^2 \\ &\leq \sum_j\left(\sum_k(\frac{n_k\bar{N}_k}{n\bar{N}})^2\right)\cdot\left(\sum_k\mu_{kj}^2\right) \\ &\leq \sum_j\left(\sum_k(\frac{n_k\bar{N}_k}{n\bar{N}})\right)\cdot\left(\sum_k\mu_{kj}^2\right) = \sum_j\sum_k\mu_{kj}^2 = \sum_k\|\mu_k\|^2.\end{aligned} \tag{A.49}$$

Combining (A.47), (A.48), and (A.49) yields

$$c\big(\sum_k\|\mu_k\|^2\big) \leq \Theta_{n2} + \Theta_{n3} + \Theta_{n4} \leq C\big(\sum_k\|\mu_k\|^2\big),$$

for absolute constants $c, C > 0$. This completes the proof. □

A.11 Proof of Lemma A.7

By (3.1), it holds that

$$(\frac{1}{n_k\bar{N}_k} - \frac{1}{n\bar{N}})^2 \asymp \frac{1}{(n_k\bar{N}_k)^2}, \tag{A.50}$$

and moreover, for all $i \in \{1, 2, \ldots, n\}$,

$$\frac{N_i^3}{N_i - 1} \asymp N_i^2. \tag{A.51}$$

Recall the definitions of A_n, B_n, and E_n in (A.8), (A.11), and (A.13), respectively. Note that these are the remainder terms in Lemmas A.3, A.4, and A.5, respectively. Under the null hypothesis (recall $\Theta_{n1} \equiv 0$ under the null),

$$\mathrm{Var}(T) = \Theta_{n2} + \Theta_{n3} + \Theta_{n4} + O(A_n + B_n + E_n). \tag{A.52}$$

It holds that

$$A_n \leq \sum_{k=1}^{K} \Big(\frac{1}{n_k\bar{N}_k}\Big)^2 \sum_{i\in S_k} N_i^2 \|\Omega_i\|_3^3. \tag{A.53}$$

Next, by linearity and the definition of Σ_k, Σ in (A.2), (A.3), respectively,

$$\begin{aligned}
B_n &\leq 2\sum_{k,\ell} \frac{n_k n_\ell \bar{N}_k \bar{N}_\ell}{n^2\bar{N}^2} \mathbf{1}_p'(\Sigma_k \circ \Sigma_\ell)\mathbf{1}_p \\
&\leq 2\mathbf{1}_p'\bigg(\frac{1}{n\bar{N}}\sum_k n_k\bar{N}_k\Sigma_k\bigg) \circ \bigg(\frac{1}{n\bar{N}}\sum_\ell n_\ell\bar{N}_\ell\Sigma_k\ell\bigg)\mathbf{1}_p \\
&= 2\mathbf{1}_p'(\Sigma\circ\Sigma)\mathbf{1}_p = 2\|\Sigma\|_F^2
\end{aligned}$$

By Cauchy–Schwarz,

$$\begin{aligned}
B_n \leq \|\Sigma\|_F^2 &= \sum_{j,j'} \bigg(\sum_k (\frac{n_k\bar{N}_k}{n\bar{N}}\Sigma_k(j,j')\bigg)^2 \\
&\leq \sum_{j,j'} \bigg(\sum_k (\frac{n_k\bar{N}_k}{n\bar{N}})^2\bigg) \cdot \bigg(\sum_k \Sigma_k(j,j')^2\bigg) \\
&\leq \sum_{j,j'} \bigg(\sum_k \frac{n_k\bar{N}_k}{n\bar{N}}\bigg) \cdot \bigg(\sum_k \Sigma_k(j,j')^2\bigg) = \sum_{j,j'}\sum_k \Sigma_k(j,j')^2 = \sum_k \|\Sigma_k\|_F^2.
\end{aligned} \tag{A.54}$$

Next by the definition of Σ_k in (A.2), we have $\Sigma_k(j,j') = \frac{1}{n_k\bar{N}_k}\sum_{i\in S_k} N_i\Omega_{ij}\Omega_{ij'}$. It follows that

$$\begin{aligned}
E_n &\leq \sum_k \sum_{j,j'} \Big(\frac{1}{n_k\bar{N}_k}\sum_{i\in S_k} N_i\Omega_{ij}\Omega_{ij'}\Big)\Big(\frac{1}{n_k\bar{N}_k}\sum_{m\in S_k} N_m\Omega_{mj}\Omega_{mj'}\Big) \\
&= \sum_k\sum_{j,j'} \Sigma_k^2(j,j') = \sum_k \|\Sigma_k\|_F^2.
\end{aligned} \tag{A.55}$$

Next, Lemma A.6 implies that

$$\Theta_{n2} + \Theta_{n3} + \Theta_{n4} \asymp \sum_k \|\mu_k\|^2 = K\|\mu\|^2, \tag{A.56}$$

where we use that the null hypothesis holds. By assumption of the lemma, we have

$$\beta_n = \frac{\max\left\{ \sum_k \sum_{i\in S_k} \frac{N_i^2}{n_k^2 \bar{N}_k^2}\|\Omega_i\|_3^3\,,\ \sum_k \|\Sigma_k\|_F^2 \right\}}{K\|\mu\|^2} = o(1)$$

Combining this with (A.52), (A.53), (A.54), (A.55),and (A.56) completes the proof of the first claim. The second claim follows plugging in $\mu_k = \mu$ for all $k \in \{1, 2, \ldots, K\}$.

□

A.12 Proof of Lemma A.8

By assumption, $N_i^3/(N_i - 1) \asymp N_i, M_i^3/(M_i - 1) \asymp M_i$. By direct calculation,

$$\begin{aligned}\Theta_{n2} + \Theta_{n4} &\asymp \Big[\frac{m\bar{M}}{(n\bar{N} + m\bar{M})n\bar{N}}\Big]^2 \sum_{i,m,j} N_i N_m \Omega_{ij}\Omega_{mj} + \Big[\frac{n\bar{N}}{(n\bar{N} + m\bar{M})m\bar{M}}\Big]^2 \sum_{i,m} N_i N_m \Gamma_{ij}\Gamma_{mj} \\ &= \frac{1}{(n\bar{N} + m\bar{M})^2}\Big((m\bar{M})^2\|\eta\|^2 + n\bar{N}^2\|\theta\|^2\Big). \end{aligned} \tag{A.57}$$

Next

$$\begin{aligned}\Theta_{n3} &= \frac{4}{(n\bar{N} + m\bar{M})^2} \sum_{i\in S_1}\sum_{m\in S_2}\sum_j N_i\Omega_{ij} \cdot N_m\Gamma_{mj} \\ &= \frac{4}{(n\bar{N} + m\bar{M})^2} \cdot n\bar{N}m\bar{M}\langle\theta, \eta\rangle. \end{aligned} \tag{A.58}$$

Combining (A.57) and (A.58) yields

$$\begin{aligned}\Theta_{n2} + \Theta_{n3} + \Theta_{n4} &\asymp \frac{1}{(n\bar{N} + m\bar{M})^2}\big((m\bar{M})^2\|\eta\|^2 + 2n\bar{N}m\bar{M}\langle\theta, \eta\rangle + n\bar{N}^2\|\theta\|^2\big) \\ &= \left\|\frac{m\bar{M}}{n\bar{N} + m\bar{M}}\eta + \frac{n\bar{N}}{n\bar{N} + m\bar{M}}\theta\right\|^2, \end{aligned}$$

which proves the first claim. The second follows by plugging in $\theta = \eta = \mu$ under the null. □

A.13 Proof of Lemma A.9

As in (A.52), we have under the null that

$$\mathrm{Var}(T) = \Theta_{n2} + \Theta_{n3} + \Theta_{n4} + O(A_n + B_n + E_n). \tag{A.59}$$

For general K, observe that the proofs of the bounds

$$A_n \le \sum_{k=1}^K \Big(\frac{1}{n_k\bar{N}_k}\Big)^2 \sum_{i\in S_k} N_i^2\|\Omega_i\|_3^3$$

$$
\begin{aligned}
B_n &\le \sum_{k=1}^{K} \|\Sigma_k\|_F^2 \\
E_n &\le \sum_{k=1}^{K} \|\Sigma_k\|_F^2
\end{aligned}
$$

derived in (A.53), (A.54), and (A.55), only use the assumption that $N_i, M_i \ge 2$ for all i.

Translating these bounds to the notation of the $K = 2$ case, we have

$$
\begin{aligned}
A_n &\le \sum_i N_i^2 \|\Omega_i\|^3 + \sum_i M_i^2 \|\Gamma_i\|^3 \\
B_n &\le \|\Sigma_1\|_F^2 + \|\Sigma_2\|_F^2 \\
E_n &\le \|\Sigma_1\|_F^2 + \|\Sigma_2\|_F^2.
\end{aligned} \tag{A.60}
$$

Furthermore, we know that $\Theta_n \ge c\|\mu\|^2$ under the null by Lemma A.8, for an absolute constant $c > 0$. Combining this with (A.59) and (A.60) completes the proof. □

A.14 Proof of Lemma A.10

Define

$$
\begin{aligned}
V_1 &= 2\sum_{k=1}^{K} \sum_{i \in S_k} \sum_{j=1}^{p} \Big(\frac{1}{n_k \bar{N}_k} - \frac{1}{n\bar{N}}\Big)^2 \Big[\frac{N_i X_{ij}^2}{N_i - 1} - \frac{N_i X_{ij}(N_i - X_{ij})}{(N_i - 1)^2}\Big] \\
V_2 &= \frac{2}{n^2 \bar{N}^2} \sum_{1 \le k \ne \ell \le K} \sum_{i \in S_k} \sum_{m \in S_\ell} \sum_{j=1}^{p} X_{ij} X_{mj} \\
V_3 &= 2\sum_{k=1}^{K} \sum_{\substack{i \in S_k, m \in S_k, \\ i \ne m}} \sum_{j=1}^{p} \Big(\frac{1}{n_k \bar{N}_k} - \frac{1}{n\bar{N}}\Big)^2 X_{ij} X_{mj}.
\end{aligned}
$$

Observe that $V_1 + V_2 + V_3 = V$. Also define

$$
A_{11} = \sum_i \sum_{r=1}^{N_i} \sum_j \big[\frac{4\theta_i \Omega_{ij}}{N_i}\big] Z_{ijr} \tag{A.61}
$$

$$
A_{12} = 2\sum_i \sum_{r=1}^{N_i} \sum_j \big[\sum_{m \in [n] \setminus \{i\}} \alpha_{im} N_m \Omega_{mj}\big] Z_{ijr} \tag{A.62}
$$

and observe that $A_{11} + A_{12} = A_1$.

First, we derive the decomposition of V_1. Recall that

$$
Y_{ij} := \frac{X_{ij}}{N_i} - \Omega_{ij} = \frac{1}{N_i} \sum_{r=1}^{N_i} Z_{ijr}, \qquad Q_{ij} := Y_{ij}^2 - \mathbb{E}Y_{ij}^2 = Y_{ij}^2 - \frac{\Omega_{ij}(1 - \Omega_{ij})}{N_i}. \tag{A.63}
$$

With these notations, $X_{ij} = N_i(\Omega_{ij} + Y_{ij})$ and $N_i Y_{ij}^2 = N_i Q_{ij} + \Omega_{ij}(1 - \Omega_{ij})$.

Write

$$V_1 = 2\sum_{i=1}^{n}\sum_{i=1}^{n}\frac{\theta_i}{N_i}\Delta_{ij}, \qquad \text{where} \quad \Delta_{ij} := \frac{X_{ij}^2}{N_i} - \frac{X_{ij}(N_i - X_{ij})}{N_i(N_i-1)}. \tag{A.64}$$

Note that $X_{ij} = N_i(\Omega_{ij} + Y_{ij})$ and $Y_{ij}^2 = Q_{ij} + N_i^{-1}\Omega_{ij}(1-\Omega_{ij})$. It follows that

$$\frac{X_{ij}^2}{N_i} = N_i\Omega_{ij}^2 + 2N_i\Omega_{ij}Y_{ij} + N_iQ_{ij} + \Omega_{ij}(1-\Omega_{ij}).$$

In (A.32), we have shown that $Q_{ij} = (1-2\Omega_{ij})\frac{Y_{ij}}{N_i} + \frac{1}{N_i^2}\sum_{1\le r\neq s\le N_i} Z_{ijr}Z_{ijs}$. It follows that

$$\frac{X_{ij}^2}{N_i} = N_i\Omega_{ij}^2 + 2N_i\Omega_{ij}Y_{ij} + (1-2\Omega_{ij})Y_{ij} + \frac{1}{N_i}\sum_{1\le r\neq s\le N_i} Z_{ijr}Z_{ijs} + \Omega_{ij}(1-\Omega_{ij}).$$

Additionally, by (A.33),

$$\frac{X_{ij}(N_{ij} - X_{ij})}{N_i(N_i-1)} = \Omega_{ij}(1-\Omega_{ij}) + (1-2\Omega_{ij})Y_{ij} - \frac{1}{N_i(N_i-1)}\sum_{1\le r\neq s\le N_i} Z_{ijr}Z_{ijs}.$$

Combining the above gives

$$\begin{aligned}
\Delta_{ij} &= N_i\Omega_{ij}^2 + 2N_i\Omega_{ij}Y_{ij} + \frac{1}{N_i-1}\sum_{1\le r\neq s\le N_i} Z_{ijr}Z_{ijs} \\
&= N_i\Omega_{ij}^2 + 2\Omega_{ij}\sum_{r=1}^{N_i} Z_{ijr} + \frac{1}{N_i-1}\sum_{1\le r\neq s\le N_i} Z_{ijr}Z_{ijs}.
\end{aligned} \tag{A.65}$$

Recall the definition of Θ_{n2} in (A.7), A_2 in (A.19), and A_{11} in (A.61). We have

$$\begin{aligned}
V_1 &= 2\sum_{k,i\in S_k}\sum_j \frac{\theta_i}{N_i}\big[N_i\Omega_{ij}^2 + 2\Omega_{ij}\sum_{r=1}^{N_i} Z_{ijr} + \frac{1}{N_i-1}\sum_{1\le r\neq s\le N_i} Z_{ijr}Z_{ijs}\big]. \\
&= \Theta_{n2} + \sum_{k,i\in S_k}\sum_j \frac{4\theta_i\Omega_{ij}}{N_i}\sum_{r=1}^{N_i} Z_{ijr} + \sum_{k,i\in S_k}\sum_j \frac{2\theta_i}{N_i(N_i-1)}\sum_{1\le r\neq s\le N_i} Z_{ijr}Z_{ijs} \\
&= \Theta_{n2} + A_{11} + A_2
\end{aligned} \tag{A.66}$$

Next, we have

$$\begin{aligned}
V_2 + V_3 &= \sum_{i\neq m}\alpha_{im}N_iN_m\sum_j\Big[(Y_{ij}+\Omega_{ij})(Y_{mj}+\Omega_{mj})\Big] \\
&= \sum_{i\neq m}\alpha_{im}N_iN_m\sum_j Y_{ij}Y_{mj} + 2\sum_{i\neq m}\alpha_{im}N_iN_m\sum_j Y_{ij}\Omega_{mj} + \sum_{i\neq m}\alpha_{im}N_iN_m\sum_j \Omega_{ij}\Omega_{mj} \\
&= \sum_{i\neq m}\sum_{r=1}^{N_i}\sum_{s=1}^{N_m}\alpha_{im}\big(\sum_j Z_{ijr}Z_{mjs}\big) + 2\sum_i\sum_{r=1}^{N_i}\sum_j\big[\sum_{m\in[n]\setminus\{i\}}\alpha_{im}N_m\Omega_{mj}\big]Z_{ijr} + \Theta_{n3} + \Theta_{n4} \\
&= A_3 + A_{12} + \Theta_{n3} + \Theta_{n4}.
\end{aligned}$$

Hence

$$A_1 + A_2 + A_3 + \Theta_{n2} + \Theta_{n3} + \Theta_{n4} = V,$$

which verifies (A.21). By inspection, we also see that $\mathbb{E}A_b = 0$ for $b \in \{1, 2, 3\}$. That A_1, A_2, A_3 are mutually uncorrelated follows immediately from the linearity of expectation and the fact that the random variables $\{Z_{ijr}\}_{i,r} \cup \{Z_{ijr}Z_{mjs}\}_{(i,r)\neq(m,s)}$ are mutually uncorrelated.

$\square$

A.15 Proof of Lemma A.11

Define

$$\gamma_{irj} = \frac{4\theta_i\Omega_{ij}}{N_i} + \sum_{m\in[n]\setminus\{i\}} 2\alpha_{im}N_m\Omega_{mj} \tag{A.67}$$

and recall that $A_1 = \sum_i \sum_{r\in[N_i]} \sum_j \gamma_{irj} Z_{ijr}$. First we develop a bound on γ_{irj}. Suppose that $i \in S_k$. Then we have

$$\begin{aligned}
\gamma_{irj} &\lesssim \frac{N_i\Omega_{ij}}{n_k^2\bar{N}_k^2} + \sum_{m\in S_k, m\neq i} \frac{N_m\Omega_{mj}}{n_k^2\bar{N}_k^2} + \sum_{k'\in[K]\setminus\{k\}} \sum_{m\in S_{k'}} \frac{N_m\Omega_{mj}}{n^2\bar{N}^2} \\
&\lesssim \frac{\mu_{kj}}{n_k\bar{N}_k} + \frac{\mu_j}{n\bar{N}}.
\end{aligned}$$

Next using properties of the covariance matrix of a multinomial vector, we have

$$\begin{aligned}
\mathrm{Var}(A_1) &= \sum_{i,r\in[N_i]} \mathrm{Var}(\gamma_{ir:}' Z_{i:r}) = \sum_{i,r\in[N_i]} \gamma_{ir:}' \mathrm{Cov}(Z_{i:r})\gamma_{ir:} \\
&\leq \sum_{i,r\in[N_i]} \gamma_{ir:}' \mathrm{diag}(\Omega_{i:})\gamma_{ir:} = \sum_{i,r\in[N_i]} \sum_j \Omega_{ij}\gamma_{irj}^2 \\
&\lesssim \sum_{k,j} \big(\frac{\mu_{kj}}{n_k\bar{N}_k} + \frac{\mu_j}{n\bar{N}}\big)^2 \sum_{i\in S_k, r\in[N_i]} \Omega_{ij} \\
&\lesssim \sum_{k,j} \big(\frac{\mu_{kj}}{n_k\bar{N}_k}\big)^2 n_k\bar{N}_k\mu_{kj} + \sum_{k,j} \big(\frac{\mu_j}{n\bar{N}}\big)^2 n_k\bar{N}_k\mu_{kj} \\
&= \big(\sum_k \frac{\|\mu_k\|_3^3}{n_k\bar{N}_k}\big) + \frac{\|\mu\|_3^3}{n\bar{N}} \lesssim \sum_k \frac{\|\mu_k\|_3^3}{n_k\bar{N}_k},
\end{aligned} \tag{A.68}$$

which proves the first claim. The last inequality follows because by Jensen's inequality (noting that the function $x \mapsto x^3$ is convex for $x \geq 0$),

$$\|\mu\|_3^3 = \sum_j \left(\sum_k \big(\frac{n_k\bar{N}_k}{n\bar{N}}\big)\mu_{kj} \right)^3 \leq \sum_j \sum_k \big(\frac{n_k\bar{N}_k}{n\bar{N}}\big)\mu_{kj}^3 \leq \sum_k \|\mu_k\|_3^3.$$

Next observe that

$$A_2 = \sum_i \sum_{r\neq s} \frac{2\theta_i}{N_i(N_i-1)} W_{irs} \tag{A.69}$$

where recall $W_{irs} = \sum_j Z_{ijr}Z_{ijs}$. Also recall that W_{irs} and $W_{i'r's'}$ are uncorrelated unless $i = i'$ and $\{r, s\} = \{r', s'\}$. By (A.42),

$$
\begin{aligned}
\mathrm{Var}(A_2) &= \sum_i \sum_{r \neq s} \frac{4\theta_i^2}{N_i^2(N_i-1)^2} \mathrm{Var}(W_{irs}) \\
&\lesssim \sum_i \sum_{r \neq s} \frac{4\theta_i^2}{N_i^2(N_i-1)^2} \|\Omega_i\|^2 \\
&\lesssim \sum_k \sum_{i \in S_k} \cdot \Big(\frac{1}{n_k \bar{N}_k} - \frac{1}{n\bar{N}}\Big)^4 \frac{N_i^6}{(N_i-1)^2} \cdot \frac{1}{N_i(N_i-1)} \|\Omega_i\|^2 \\
&\lesssim \sum_k \sum_{i \in S_k} \frac{N_i^2}{n_k^4 \bar{N}_k^4} \|\Omega_i\|^2 \qquad \text{(A.70)}
\end{aligned}
$$

Also observe that

$$
\begin{aligned}
\sum_k \frac{1}{n_k^4 \bar{N}_k^4} \sum_{i \in S_k} N_i^2 \|\Omega_i\|_2^2 &\leq \sum_k \frac{1}{n_k^2 \bar{N}_k^2} \sum_{i,m \in S_k} \Big\langle \Big(\frac{N_i}{n_k \bar{N}_k}\Big)\Omega_i, \Big(\frac{N_m}{n_m \bar{N}_m}\Big)\Omega_m \Big\rangle \\
&= \sum_k \frac{1}{n_k^2 \bar{N}_k^2} \|\mu_k\|^2.
\end{aligned}
$$

This establishes the second claim.

Last we study A_3. Observe that

$$
A_3 = \sum_{i \neq m} \sum_{r=1}^{N_i} \sum_{s=1}^{N_m} \alpha_{im} V_{irms}
$$

where recall $V_{irms} = \sum_j Z_{ijr} Z_{mjs}$. Recall that V_{irms} and $V_{i'r'm's'}$ are uncorrelated unless $(r, s) = (r', s')$ and $\{i, m\} = \{i', m'\}$ .By (A.43),

$$
\begin{aligned}
\mathrm{Var}(A_3) &\lesssim \sum_{i \neq m} \alpha_{im}^2 N_i N_m \sum_j \Omega_{ij} \Omega_{mj} \\
&\lesssim \sum_k \sum_{i \neq m \in S_k} \frac{1}{n_k^4 \bar{N}_k^4} \langle N_i \Omega_i, N_m \Omega_m \rangle + \sum_{k \neq \ell} \sum_{i \in S_k, m \in S_\ell} \frac{1}{n^4 \bar{N}^4} \langle N_i \Omega_i, N_m \Omega_m \rangle \\
&\lesssim \sum_k \frac{\|\mu_k\|^2}{n_k^2 \bar{N}_k^2} + \sum_{k,\ell} \frac{1}{n^4 \bar{N}^4} \langle n_k \bar{N}_k \mu_k, n_\ell \bar{N}_\ell \mu_\ell \rangle \\
&\lesssim \sum_k \frac{\|\mu_k\|^2}{n_k^2 \bar{N}_k^2} + \frac{\|\mu\|^2}{n^2 \bar{N}^2} \lesssim \sum_k \frac{\|\mu_k\|^2}{n_k^2 \bar{N}_k^2}. \qquad \text{(A.71)}
\end{aligned}
$$

In the last line we use that $\|\mu\|^2 \leq 2 \sum \|\mu_k\|^2$ as shown in (A.49). This proves all required claims. $\square$

A.16 Proof of Proposition A.1

Under the null hypothesis, we have $\Theta_{n1} \equiv 0$. Thus, $\mathbb{E}V = \Theta_n$ under the null by Lemma A.10. Under (3.1), we have $\mathrm{Var}(T) = [1 + o(1)]\Theta_n$. Therefore,

$$
\mathbb{E}V = [1 + o(1)]\mathrm{Var}(T), \qquad \text{(A.72)}
$$

so V is asymptotically unbiased under the null. Furthermore, by Lemma A.6, we have

$$\Theta_n \asymp K\|\mu\|^2. \tag{A.73}$$

In Lemma A.11, we showed that

$$\mathrm{Var}(A_2) \lesssim \sum_k \sum_{i\in S_k} \frac{N_i^2\|\Omega_i\|_2^2}{n_k^4 \bar{N}_k^4}$$

We conclude by Lemma A.11 that under the null

$$\mathrm{Var}(V) \lesssim \sum_k \frac{\|\mu\|^2}{n_k^2\bar{N}_k^2} \vee \sum_k \frac{\|\mu\|_3^3}{n_k\bar{N}_k}. \tag{A.74}$$

By Chebyshev's inequality, (A.73), (A.74), and assumption (A.22) of the theorem statement, we have

$$\frac{|V-\mathbb{E}V|}{\mathrm{Var}(T)} \asymp \frac{|V-\mathbb{E}V|}{K\|\mu\|^2} = o_{\mathbb{P}}(1).$$

Thus by (A.72),

$$\frac{V}{\mathrm{Var}(T)} = \frac{(V-\mathbb{E}V)}{\mathrm{Var}(T)} + \frac{\mathbb{E}V}{\mathrm{Var}(T)} = o_{\mathbb{P}}(1) + [1+o(1)],$$

as desired. □

A.17 Proof of Lemma A.12

By Lemmas A.1–A.5, we have

$$\mathrm{Var}(T) = \sum_{a=1}^{4} \mathrm{Var}(\mathbf{1}_p' U_a) \geq (\sum_{a=2}^{4} \Theta_{na}) - (A_n + B_n + E_n). \tag{A.75}$$

Using that $\max_i \|\Omega_i\|_\infty \leq 1-c_0$, we have $\|\Omega_i\|^3 \leq (1-c_0)\|\Omega_i\|^2$, which implies that

$$A_n \leq (1-c_0)\Theta_{n2}. \tag{A.76}$$

Again using $\max_i \|\Omega_i\|_\infty \leq 1-c_0$, as well as $\sum_{j'} \Omega_{ij'} = 1$, we have

$$\begin{aligned}
B_n &= \frac{2}{n^2\bar{N}^2} \sum_{k\neq\ell} \sum_{i\in S_k} \sum_{m\in S_\ell} \sum_{j,j'} N_i N_m \Omega_{ij}\Omega_{ij'}\Omega_{mj}\Omega_{mj'} \\
&\leq (1-c_0)\cdot \frac{2}{n^2\bar{N}^2} \sum_{k\neq\ell} \sum_{i\in S_k} \sum_{m\in S_\ell} \sum_{j,j'} N_i N_m \Omega_{ij}\Omega_{ij'}\Omega_{mj} \\
&= (1-c_0)\cdot \frac{2}{n^2\bar{N}^2} \sum_{k\neq\ell} \sum_{i\in S_k} \sum_{m\in S_\ell} \sum_{j} N_i N_m \Omega_{ij}\Omega_{mj} \\
&\leq (1-c_0)\cdot \Theta_{n3}.
\end{aligned} \tag{A.77}$$

Similarly to control E_n, we again use $\max_i \|\Omega_i\|_\infty \leq 1 - c_0$ and obtain

$$\begin{aligned}
E_n &= 2\sum_k \sum_{\substack{i\in S_k, m\in S_k,\\ i\neq m}} \sum_{1\leq j,j'\leq p} \Big(\frac{1}{n_k\bar{N}_k} - \frac{1}{n\bar{N}}\Big)^2 N_i N_m \Omega_{ij}\Omega_{ij'}\Omega_{mj}\Omega_{mj'} \\
&\leq (1-c_0)\cdot 2\sum_k \sum_{\substack{i\in S_k, m\in S_k,\\ i\neq m}} \sum_{1\leq j,j'\leq p} \Big(\frac{1}{n_k\bar{N}_k} - \frac{1}{n\bar{N}}\Big)^2 N_i N_m \Omega_{ij}\Omega_{ij'}\Omega_{mj} \\
&\leq (1-c_0)\cdot 2\sum_k \sum_{\substack{i\in S_k, m\in S_k,\\ i\neq m}} \sum_{1\leq j\leq p} \Big(\frac{1}{n_k\bar{N}_k} - \frac{1}{n\bar{N}}\Big)^2 N_i N_m \Omega_{ij}\Omega_{mj} \\
&\leq (1-c_0)\cdot \Theta_{n4}. \qquad (A.78)
\end{aligned}$$

Combining (A.75), (A.76), (A.77), and (A.78) finishes the proof.

□

A.18 Proof of Proposition A.2

By Lemmas A.6 and A.12,

$$\mathrm{Var}(T) \gtrsim \Theta_{n2} + \Theta_{n3} + \Theta_{n4} \gtrsim \sum_k \|\mu_k\|^2. \qquad (A.79)$$

By Lemma A.11,

$$\mathrm{Var}(V) \lesssim \sum_k \frac{\|\mu_k\|^2}{n_k^2\bar{N}_k^2} \vee \sum_k \frac{\|\mu_k\|_3^3}{n_k\bar{N}_k} \qquad (A.80)$$

Using a similar argument based on Chebyshev's inequality as in the proof of Proposition A.1 and applying (A.79) and (A.80), we have

$$\frac{|V - \mathbb{E}V|}{\mathrm{Var}(T)} \gtrsim \frac{|V - \mathbb{E}V|}{\sum_k \|\mu_k\|^2} = o_{\mathbb{P}}(1). \qquad (A.81)$$

Next, by Lemma A.10 and (A.79),

$$\mathbb{E}V = \Theta_{n2} + \Theta_{n3} + \Theta_{n4} \lesssim \mathrm{Var}(T). \qquad (A.82)$$

Combining (A.81) and (A.82) finishes the proof. □

A.19 Proof of Proposition A.5

From the proof of Lemma A.10, we have

$$V^* = V_1 = \Theta_{n2} + A_{11} + A_2,$$

and the terms on the right-hand-side are mutually uncorrelated. From (A.68), we have

$$\mathrm{Var}(A_{11}) \lesssim \sum_i \frac{\|\Omega_i\|_3^3}{N_i}$$

$$\mathrm{Var}(A_2) \lesssim \sum_i \frac{\|\Omega_i\|^2}{N_i^2}.$$

Hence

$$\begin{aligned}\mathbb{E}V^* &= \Theta_{n2} \\ \mathrm{Var}(V^*) &\lesssim \sum_i \frac{\|\Omega_i\|_3^3}{N_i} \vee \sum_i \frac{\|\Omega_i\|^2}{N_i^2}.\end{aligned} \tag{A.83}$$

Since $K = n$ and the null hypothesis holds, we have $\Theta_{n1} \equiv \Theta_{n4} \equiv 0$. Moreover, by (A.48), we have

$$\Theta_{n3} \lesssim \|\mu\|^2 \ll \Theta_{n2} \asymp n\|\mu\|^2.$$

It follows that

$$\mathrm{Var}(T) = [1 + o(1)]\Theta_{n2} \asymp n\|\mu\|^2. \tag{A.84}$$

Thus by (A.83) and Chebyshev's inequality, we have

$$\frac{V^*}{\mathrm{Var}(T)} = \frac{V^* - \mathbb{E}V^*}{\mathrm{Var}(T)} + \frac{\mathbb{E}V^*}{\mathrm{Var}(T)} = o_{\mathbb{P}}(1) + 1 + o(1),$$

as desired.

□

A.20 Proof of Proposition A.6

By Lemmas A.6 and A.12,

$$\mathrm{Var}(T) \gtrsim \Theta_{n2} + \Theta_{n3} \gtrsim \sum_i \|\Omega_i\|^2. \tag{A.85}$$

By (A.83),

$$\mathrm{Var}(V^*) \lesssim \sum_i \frac{\|\Omega_i\|^2}{N_i^2} \vee \sum_i \frac{\|\Omega_i\|_3^3}{N_i} \tag{A.86}$$

Using a similar argument based on Chebyshev's inequality as in the proof of Proposition A.1 and applying (A.85) and (A.86), we have

$$\frac{|V^* - \mathbb{E}V^*|}{\mathrm{Var}(T)} \gtrsim \frac{|V^* - \mathbb{E}V^*|}{\sum_i \|\Omega_i\|^2} = o_{\mathbb{P}}(1). \tag{A.87}$$

Next, by Lemma A.10 and (A.85),

$$\mathbb{E}V^* = \Theta_{n2} \lesssim \mathrm{Var}(T). \tag{A.88}$$

Combining (A.81) and (A.88) finishes the proof. □

B Proofs of asymptotic normality results

The goal of this section is to prove Theorems 3.1 and 3.2. The argument relies on the martingale central limit theorem and the lemmas stated below. As a preliminary, we describe a martingale decomposition of T under the null.

Define

$$U = \mathbf{1}_p'(U_3 + U_4), \qquad \text{and} \quad S = \mathbf{1}_p' U_2.$$

By Lemma A.1, we have $T = U + S$ under the null hypothesis. It holds that

$$U = \sum_{i<i'} \sigma_{i,i'} \sum_{r=1}^{N_i} \sum_{s=1}^{N_{i'}} \big(\sum_j Z_{ijr} Z_{i'js} \big). \tag{B.1}$$

where we define

$$\sigma_{i,i'} = \begin{cases} 2\big(\frac{1}{n_k \bar{N}_k} - \frac{1}{n\bar{N}}\big) & \text{if } i, i' \in S_k \text{ for some } k \\ -\frac{2}{n\bar{N}} & \text{else.} \end{cases}$$

Define a sequence of random variables

$$D_{\ell,s} = \sum_{i \in [\ell-1]} \sigma_{i,\ell} \sum_{r=1}^{N_i} \sum_j Z_{ijr} Z_{\ell js} \tag{B.2}$$

indexed by $(\ell, s) \in \{(i, r)\}_{1 \le i \le n, 1 \le r \le N_i}$, where these tuples are placed in lexicographical order. Precisely, we define

$$(\ell_1, s_1) \prec (\ell_2, s_2)$$

if either

- $\ell_1 < \ell_2$, or
- $\ell_1 = \ell_2$ and $s_1 < s_2$.

Observe that

$$\sum_{\ell,s} D_{\ell,s} = U.$$

Next define $\mathcal{F}_{\prec(\ell,s)}$ to be the σ-field generated by $\{Z_{i:r}\}_{(i,r)\prec(\ell,s)}$. Observe that

$$\mathbb{E}[D_{\ell,s} | \mathcal{F}_{\prec(\ell,s)}] = 0,$$

and hence $\{D_{\ell,s}\}$ is a martingale difference sequence. Turning to S, we have

$$S = \sum_{i=1}^{n} \sigma_i \sum_{r<s} \sum_j Z_{ijr} Z_{ijs}. \tag{B.3}$$

where we define

$$\sigma_i = 2\Big(\frac{1}{n_k \bar{N}_k} - \frac{1}{n\bar{N}}\Big) \frac{N_i}{N_i - 1}$$

if $i \in S_k$. Define

$$E_{\ell,s} = \sigma_\ell \sum_{r\in[s-1]} \sum_j Z_{\ell jr} Z_{\ell js}. \tag{B.4}$$

Note that $E_{\ell,1} = 0$. Order (ℓ, s) lexicographically as above, and recall that $\mathcal{F}_{\prec(\ell,s)}$ is the σ-field generated by $\{Z_{i:r}\}_{(i,r)\prec(\ell,s)}$. Observe that

$$\mathbb{E}[E_{\ell,s}|\mathcal{F}_{\prec(\ell,s)}] = 0,$$

and hence $\{E_{\ell,s}\}$ is a martingale difference sequence. We have

$$\sum_{(\ell,s)} \sigma_\ell \sum_{r\in[s-1]} \sum_j Z_{\ell jr} Z_{\ell js} = \sum_{\ell=1}^{n} \sum_{s=1}^{N_\ell} \sigma_\ell \sum_{r\in[s-1]} \sum_j Z_{\ell jr} Z_{\ell js} = S.$$

Define

$$\mathcal{M}_{\ell,s} = D_{\ell,s} + E_{\ell,s}, \quad \widetilde{\mathcal{M}}_{\ell,s} = \frac{\mathcal{M}_{\ell,s}}{\sqrt{\mathrm{Var}(T)}}. \tag{B.5}$$

Thus we obtain the martingale decomposition:

$$T = U + S = \sum_{(\ell,s)} [D_{\ell,s} + E_{\ell,s}] = \sum_{(\ell,s)} \mathcal{M}_{\ell,s}. \tag{B.6}$$

The technical results below are crucial to the proof of Theorem 3.1 given in Section B.1. Theorem 3.2 then follows easily from Theorem 3.1 and Theorem A.1.

Lemma B.1. *Let $\widetilde{\mathcal{M}}_{\ell,s}$ be defined as in* (B.5)*. It holds that*

$$\mathbb{E}\Big[\sum_{(\ell,s)} \mathrm{Var}\big(\widetilde{\mathcal{M}}_{\ell,s}|\mathcal{F}_{\prec(\ell,s)}\big)\Big] = 1.$$

Lemma B.2. *Suppose that* $\min N_i \geq 2$ *and* $\max \|\Omega_i\|_\infty \leq 1 - c_0$*. Under the null hypothesis, it holds that*

$$\mathrm{Var}\bigg(\sum_{(\ell,s)} \mathrm{Var}(D_{\ell,s}|\mathcal{F}_{\prec(\ell,s)})\bigg) \lesssim \big(\sum_k \frac{1}{n_k \bar{N}_k}\big)\|\mu\|_3^3 + K\|\mu\|_4^4.$$

Lemma B.3. *Suppose that* $\min N_i \geq 2$ *and* $\max \|\Omega_i\|_\infty \leq 1 - c_0$*. Under the null hypothesis, it holds that*

$$\sum_{(\ell,s)} \mathbb{E}D_{\ell,s}^4 \lesssim \big(\sum_k \frac{1}{n_k^2 \bar{N}_k^2}\big)\|\mu\|^2 + \big(\sum_k \frac{1}{n_k \bar{N}_k}\big)\|\mu\|_3^3,$$

Lemma B.4. *Suppose that* $\min N_i \geq 2$ *and and* $\max \|\Omega_i\|_\infty \leq 1 - c_0$*. Then we have*

$$\mathrm{Var}\bigg(\sum_{(\ell,s)} \mathrm{Var}(\tilde{E}_{\ell,s}|\mathcal{F}_{\prec(\ell,s)})\bigg) \lesssim \sum_k \sum_{i\in S_k} \frac{N_i^3\|\Omega_i\|_3^3}{n_k^4 \bar{N}_k^4} \vee \sum_k \sum_{i\in S_k} \frac{N_i^4\|\Omega_i\|_4^4}{n_k^4 \bar{N}_k^4} \tag{B.7}$$

Lemma B.5. *Suppose that* $\min N_i \geq 2$ *and and* $\max \|\Omega_i\|_\infty \leq 1 - c_0$. *Then we have*

$$\sum_{(\ell,s)} \mathbb{E}\, E_{\ell,s}^4 \lesssim \sum_k \sum_{i \in S_k} \frac{N_i^2 \|\Omega_i\|^2}{n_k^4 \bar{N}_k^4} \vee \sum_k \sum_{i \in S_k} \frac{N_i^3 \|\Omega_i\|_3^3}{n_k^4 \bar{N}_k^4}$$

Lemma B.6. *Under either the null or alternative, it holds that*

$$\sum_k \sum_{i \in S_k} \frac{N_i^2 \|\Omega_i\|^2}{n_k^4 \bar{N}_k^4} \leq \sum_k \frac{1}{n_k^2 \bar{N}_k^2} \|\mu_k\|^2$$
$$\sum_k \sum_{i \in S_k} \frac{N_i^3 \|\Omega_i\|_3^3}{n_k^4 \bar{N}_k^4} \leq \sum_k \frac{1}{n_k \bar{N}_k} \|\mu_k\|_3^3$$
$$\sum_k \sum_{i \in S_k} \frac{N_i^4 \|\Omega_i\|_4^4}{n_k^4 \bar{N}_k^4} \leq \sum_k \|\mu_k\|_4^4$$

B.1 Proof of Theorem 3.1

By the martingale central limit theorem (see e.g. Hall and Heyde [2014]), we have that $T/\sqrt{\mathrm{Var}(T)} \Rightarrow N(0,1)$ if the following conditions are satisfied:

$$\sum_{(\ell,s)} \mathrm{Var}\big(\widetilde{\mathcal{M}}_{\ell,s} \big| \mathcal{F}_{\prec(\ell,s)}\big) \xrightarrow{\mathbb{P}} 1 \tag{B.8}$$

$$\sum_{(\ell,s)} \mathbb{E}\big[\widetilde{\mathcal{M}}_{\ell,s}^2 \mathbf{1}_{|\widetilde{\mathcal{M}}_{\ell,s}| > \varepsilon} \big| \mathcal{F}_{\prec(\ell,s)}\big] \xrightarrow{\mathbb{P}} 0, \quad \text{for any } \varepsilon > 0. \tag{B.9}$$

It is known that (B.9), which is a Lindeberg-type condition, is implied by the Lyapunov-type condition

$$\sum_{(\ell,s)} \mathbb{E} \widetilde{\mathcal{M}}_{\ell,s}^4 = o(1). \tag{B.10}$$

See e.g. Jin et al. [2018].

Since (3.1) holds,

$$\mathrm{Var}(T) \gtrsim \Theta = \Theta_{n2} + \Theta_{n3} + \Theta_{n4} \gtrsim K \|\mu\|^2. \tag{B.11}$$

Recall that

$$\widetilde{\mathcal{M}}_{\ell,s} = \frac{\mathcal{M}_{\ell,s}}{\mathrm{Var}(T)} = \frac{D_{\ell,s} + E_{\ell,s}}{\mathrm{Var}(T)},$$

Note that (B.8) holds if

$$\mathbb{E}\Bigg[\mathrm{Var}\big(\widetilde{\mathcal{M}}_{\ell,s} \big| \mathcal{F}_{\prec(\ell,s)}\big)\Bigg] \to 1, \quad \text{and} \tag{B.12}$$

$$\mathrm{Var}\Bigg(\mathrm{Var}\big(\widetilde{\mathcal{M}}_{\ell,s} \big| \mathcal{F}_{\prec(\ell,s)}\big)\Bigg) \to 0. \tag{B.13}$$

Recall that (B.12) holds by Lemma B.1.

Next note that

$$\mathbb{E}(D_{\ell,s}E_{\ell,s}|\mathcal{F}_{\prec(\ell,s)}) = 0,$$

by inspection of the expressions for $D_{\ell,s}$ and $E_{\ell,s}$ in (B.2) and (B.4). Therefore

$$\mathrm{Var}(\mathcal{M}_{\ell,s}|\mathcal{F}_{\prec(\ell,s)}) = \mathrm{Var}(D_{\ell,s}|\mathcal{F}_{\prec(\ell,s)}) + \mathrm{Var}(E_{\ell,s}|\mathcal{F}_{\prec(\ell,s)}).$$

Hence by (B.11); Lemmas B.2, B.4 , and B.6; and the assumption (3.4), under the null hypothesis, we have

$$\begin{aligned}\mathrm{Var}\Big(\mathrm{Var}\big(\widetilde{\mathcal{M}}_{\ell,s}|\mathcal{F}_{\prec(\ell,s)}\big)\Big) &\le \frac{1}{\mathrm{Var}(T)^2}\Big[\mathrm{Var}\Big(\mathrm{Var}\big(D_{\ell,s}|\mathcal{F}_{\prec(\ell,s)}\big)\Big) + \mathrm{Var}\Big(\mathrm{Var}\big(E_{\ell,s}|\mathcal{F}_{\prec(\ell,s)}\big)\Big)\Big] \\ &\lesssim \frac{1}{K^2\|\mu\|^4}\Big[\big(\sum_k \frac{1}{n_k\bar{N}_k}\big)\|\mu\|_3^3 + K\|\mu\|_4^4\big)\|\mu\|^2\Big] = o(1).\end{aligned}$$

This proves (B.13). Thus, (B.12) and (B.13) are established, which proves (B.8).

Similarly, (B.10) (and thus (B.9)) holds by (B.11); Lemmas (B.3), (B.5), and (B.6), and the assumption (3.4). Combining (B.8) and (B.9) verifies the conditions of the martingale central limit theorem, so we conclude that $T/\sqrt{\mathrm{Var}(T)} \Rightarrow N(0,1)$. Since $\mathrm{Var}(T) = [1 + o(1)]\Theta_n$ by (3.4) and Lemma A.7, the proof is complete.

□

We record a useful proposition that records the weaker conditions under which $T/\sqrt{\mathrm{Var}(T)}$ is asymptotically normal.

Proposition B.1. *Recall that α_n is defined as*

$$\alpha_n := \max\left\{\sum_{k=1}^K \frac{\|\mu_k\|_3^3}{n_k\bar{N}_k},\quad \sum_{k=1}^K \frac{\|\mu_k\|^2}{n_k^2\bar{N}_k^2}\right\} \Big/ \left(\sum_{k=1}^K \|\mu_k\|^2\right)^2 \tag{B.14}$$

in (3.2)*. If under the null hypothesis,*

$$\alpha_n = \max\left\{\sum_{k=1}^K \frac{\|\mu_k\|_3^3}{n_k\bar{N}_k},\quad \sum_{k=1}^K \frac{\|\mu_k\|^2}{n_k^2\bar{N}_k^2}\right\} \Big/ \left(K\|\mu\|^2\right)^2 \to 0, \qquad \text{and} \quad \frac{\|\mu\|_4^4}{K\|\mu\|^4} \to 0, \tag{B.15}$$

then $T/\sqrt{\mathrm{Var}(T)} \Rightarrow N(0,1)$.

B.1.1 Proof of Theorem 3.2

By our assumptions, Proposition A.1 holds and $V/\mathrm{Var}(T) \to 1$. Thus the variance estimate V is consistent under the null. Theorem 3.2 follows immediately from Slutsky's theorem and Theorem 3.1. □

B.2 Proof of Lemma B.1

By Lemma A.1, S and U are uncorrelated, and it holds that

$$\mathrm{Var}(T) = \mathrm{Var}(S) + \mathrm{Var}(U). \tag{B.16}$$

Next note that

$$\mathbb{E}(D_{\ell,s}E_{\ell,s}|\mathcal{F}_{\prec(\ell,s)}) = 0,$$

by inspection of the expressions for $D_{\ell,s}$ and $E_{\ell,s}$ in (B.2) and (B.4). Therefore

$$\mathrm{Var}(\mathcal{M}_{\ell,s}|\mathcal{F}_{\prec(\ell,s)}) = \mathrm{Var}(D_{\ell,s}|\mathcal{F}_{\prec(\ell,s)}) + \mathrm{Var}(E_{\ell,s}|\mathcal{F}_{\prec(\ell,s)}).$$

Observe that

$$\begin{aligned}\mathbb{E}\Big[\sum_{(\ell,s)}\mathrm{Var}(E_{\ell,s}|\mathcal{F}_{\prec(\ell,s)})\Big] &= \sum_{(\ell,s)}\mathbb{E}E_{\ell,s}^2 = \sum_{(\ell,s)}\sigma_\ell^2 \sum_{r,r'\in[s-1]}\sum_{j,j'}\mathbb{E}[Z_{\ell jr}Z_{\ell js}Z_{\ell j'r'}Z_{\ell j's}] \\ &= \sum_{(\ell,s)}\sigma_\ell^2 \sum_{r\in[s-1]}\sum_{j,j'}\mathbb{E}[Z_{\ell jr}Z_{\ell j'r}Z_{\ell js}Z_{\ell j's}] \\ &= \sum_{\ell=1}^{n}\sigma_\ell^2\sum_{s\in[N_\ell]}\sum_{r\in[s-1]}\mathbb{E}\big(\sum_j Z_{\ell jr}Z_{\ell js}\big)^2 \\ &= \mathrm{Var}(S). \end{aligned} \tag{B.17}$$

The last line is obtained noting that S as defined in (B.3) is a sum of uncorrelated terms over (i, r, s).

Similarly, we have

$$\begin{aligned}\mathbb{E}\Big[\sum_{(\ell,s)}\mathrm{Var}(D_{\ell,s}|\mathcal{F}_{\prec(\ell,s)})\Big] = \mathbb{E}\Big[\sum_{(\ell,s)}\mathbb{E}\big[D_{\ell,s}^2|\mathcal{F}_{\prec(\ell,s)}\big]\Big] &= \sum_{(\ell,s)}\mathbb{E}\big[D_{\ell,s}^2\big] \\ &= \sum_{(\ell,s)}\sum_{i\in[\ell-1]}\sigma_{i,\ell}^2\mathrm{Var}\big(\sum_{r=1}^{N_i}\sum_j Z_{ijr}Z_{\ell js}\big) \\ &= \sum_{\ell}\sum_{i\in[\ell-1]}\sigma_{i,\ell}^2\mathrm{Var}\big(\sum_{r=1}^{N_i}\sum_{s=1}^{N_\ell} Z_{ijr}Z_{\ell js}\big) \\ &= \mathrm{Var}(U). \end{aligned} \tag{B.18}$$

The lemma follows by combining (B.16)–(B.18). □

B.3 Proof of Lemma B.2

Let $M_k = n_k\bar{N}_k$ and $M = n\bar{N}$. Define

$$\Sigma = \frac{1}{M}\sum_k M_k\Sigma_k = \frac{1}{M}\sum_{\ell\in[n]}N_\ell\Omega_{\ell j_1}\Omega_{\ell j_2}. \tag{B.19}$$

Our main goal is to control the conditional variance process. Define

$$\delta_{jj'\ell} = \mathbb{E}Z_{\ell jr}Z_{\ell j'r} = \begin{cases}\Omega_{\ell j}(1-\Omega_{\ell j}) & \text{if } j = j' \\ -\Omega_{\ell j}\Omega_{\ell j'} & \text{else.}\end{cases} \tag{B.20}$$

Observe that

$$\begin{aligned}
\mathrm{Var}(D_{\ell,s}|\mathcal{F}_{\prec(\ell,s)}) &= \mathbb{E}[\sum_{i,i'\in[\ell-1]}\sum_{r,r'}\sum_{j_1,j_2}\sigma_{i\ell}\sigma_{i'\ell}Z_{ij_1r}Z_{\ell j_1 s}Z_{i'j_2r'}Z_{\ell j_2 s}|\mathcal{F}_{\prec(\ell,s)}] \\
&= \sum_{i,i'\in[\ell-1]}\sum_{r,r'}\sum_{j_1,j_2}\sigma_{i\ell}\sigma_{i'\ell}Z_{ij_1r}Z_{i'j_2r'}\mathbb{E}[Z_{\ell j_1 s}Z_{\ell j_2 s}] \\
&= \sum_{i,i'\in[\ell-1]}\sum_{r,r'}\sigma_{i\ell}\sigma_{i'\ell}\sum_{j_1,j_2}\delta_{j_1j_2\ell}Z_{ij_1r}Z_{i'j_2r'}
\end{aligned}$$

Define

$$\alpha_{ii'j_1j_2} = \sum_{\ell>i'} N_\ell\sigma_{i\ell}\sigma_{i'\ell}\delta_{j_1j_2\ell}. \tag{B.21}$$

Thus

$$\begin{aligned}
\sum_{(\ell,s)}\mathrm{Var}(D_{\ell,s}|\mathcal{F}_{\prec(\ell,s)}) &= \sum_{\ell,s}\sum_{i,i'\in[\ell-1]}\sum_{r=1}^{N_i}\sum_{r'=1}^{N_{i'}}\sigma_{i\ell}\sigma_{i'\ell}\sum_{j_1,j_2}\delta_{j_1j_2\ell}\,Z_{ij_1r}Z_{i'j_2r'} \\
&= \sum_i\sum_{r=1}^{N_i}\sum_{r'=1}^{N_i}\sum_{j_1,j_2}\Big(\sum_{\ell>i}N_\ell\sigma_{i\ell}^2\delta_{j_1j_2\ell}\Big)Z_{ij_1r}Z_{i'j_2r'} \\
&\quad + 2\sum_{i<i'}\sum_{r=1}^{N_i}\sum_{r'=1}^{N_{i'}}\sum_{j_1,j_2}\Big(\sum_{\ell>i'}N_\ell\sigma_{i\ell}\sigma_{i'\ell}\delta_{j_1j_2\ell}\Big)Z_{ij_1r}Z_{i'j_2r'} \\
&= \sum_i\sum_{r=1}^{N_i}\sum_{r'=1}^{N_i}\sum_{j_1,j_2}\alpha_{iij_1j_2}Z_{ij_1r}Z_{i'j_2r'} \\
&\quad + 2\sum_{i<i'}\sum_{r=1}^{N_i}\sum_{r'=1}^{N_{i'}}\sum_{j_1,j_2}\alpha_{ii'j_1j_2}Z_{ij_1r}Z_{i'j_2r'}.
\end{aligned}$$

Define

$$\zeta_{iri'r'} = \sum_{j_1,j_2}\alpha_{ii'j_1j_2}Z_{ij_1r}Z_{i'j_2r'}. \tag{B.22}$$

Then

$$\begin{aligned}
\sum_{(\ell,s)}\mathrm{Var}(D_{\ell,s}|\mathcal{F}_{\prec(\ell,s)}) &= \sum_i\sum_{r\in[N_i]}\zeta_{irir} + \Big(2\sum_i\sum_{r<r'\in[N_i]}\zeta_{irir'} + 2\sum_{i<i'}\sum_{r=1}^{N_i}\sum_{r'=1}^{N_{i'}}\zeta_{iri'r'}\Big) \\
&=: V_1 + V_2
\end{aligned}$$

With this decomposition, Lemma B.2 follows directly from Lemmas B.7 and B.8 stated below and proved in the next remainder of this subsection.

Lemma B.7. *It holds that*

$$\mathrm{Var}(V_1) \lesssim \big(\sum_k \frac{1}{M_k}\big)\|\mu\|_3^3.$$

Lemma B.8. *It holds that*

$$\mathrm{Var}(V_2) \lesssim K\|\mu\|_4^4$$

$\square$

B.3.1 Statement and proof of Lemma B.9

The proofs of Lemmas B.7 and B.8 heavily rely on the following intermediate result that bounds the coefficients $\alpha_{ii'j_1j_2}$ in all cases.

Lemma B.9. *It holds that*

$$\alpha_{ii'j_1j_2} \lesssim \begin{cases} \frac{1}{M_k}\mu_{j_1} & \text{if } i,i' \in S_k, j_1 = j_2 \\ \frac{1}{M_k}\Sigma_{kj_1j_2} + \frac{1}{M}\Sigma_{j_1j_2} & \text{if } i,i' \in S_k, j_1 \neq j_2 \\ \frac{1}{M}\mu_{j_1} & \text{if } i \in S_{k_1}, i' \in S_{k_2}, k_1 \neq k_2, j_1 = j_2 \\ \frac{1}{M}\sum_{a=1}^{2}\Sigma_{k_aj_1j_2} + \frac{1}{M}\Sigma_{j_1j_2} & \text{if } i \in S_{k_1}, i' \in S_{k_2}, k_1 \neq k_2, j_1 \neq j_2 \end{cases}$$

Proof. If $j_1 = j_2$ and $i, i' \in S_k$, we have

$$\begin{aligned} |\alpha_{ii'j_1j_1}| &= |\sum_{\ell > i'} N_\ell \sigma_{i\ell}\sigma_{i'\ell}\delta_{j_1j_1\ell}| \leq \sum_{k'=1}^{K}\sum_{\ell\in S_{k'}} N_\ell \sigma_{i\ell}\sigma_{i'\ell}\delta_{j_1j_1\ell} \\ &\lesssim \frac{1}{M_k}\cdot\frac{1}{M_k}\sum_{\ell\in S_k} N_\ell\Omega_{\ell j_1} + \frac{1}{M}\cdot\frac{1}{M}\sum_{\ell\in[n]} N_\ell\Omega_{\ell j_1} \lesssim \frac{1}{M_k}\mu_{j_1} + \frac{1}{M}\mu_{j_1} \lesssim \frac{1}{M_k}\mu_{j_1}. \end{aligned}$$

If $j_1 \neq j_2$ and $i, i' \in S_k$, we have

$$\begin{aligned} |\alpha_{ii'j_1j_2}| &= |\sum_{\ell > i'} N_\ell \sigma_{i\ell}\sigma_{i'\ell}\delta_{j_1j_2\ell}| \leq \sum_{\ell\in[n]} N_\ell |\sigma_{i\ell}\sigma_{i'\ell}|\Omega_{\ell j_1}\Omega_{\ell j_2} \\ &\lesssim \frac{1}{M_k}\cdot\frac{1}{M_k}\sum_{\ell\in S_k} N_\ell\Omega_{\ell j_1}\Omega_{\ell j_2} + \frac{1}{M}\cdot\frac{1}{M}\sum_{\ell\in[n]} N_\ell\Omega_{\ell j_1}\Omega_{\ell j_2} \lesssim \frac{1}{M_k}\Sigma_{kj_1j_2} + \frac{1}{M}\Sigma_{j_1j_2}. \end{aligned}$$

If $i \neq i'$, $j_1 = j_2$, and $i \in S_{k_1}, i' \in S_{k_2}$ where $k_1 \neq k_2$, we have

$$\begin{aligned} |\alpha_{ii'j_1j_1}| &= |\sum_{\ell > i'} N_\ell \sigma_{i\ell}\sigma_{i'\ell}\delta_{j_1j_1\ell}| \leq \sum_{\ell} N_\ell |\sigma_{i\ell}\sigma_{i'\ell}|\Omega_{\ell j_1} \\ &\lesssim \frac{1}{M}\cdot\sum_{a=1}^{2}\frac{1}{M_{k_a}}\sum_{\ell\in S_{k_a}} N_\ell\Omega_{\ell j_1} + \frac{1}{M}\cdot\frac{1}{M}\sum_{\ell\in[n]} N_\ell\Omega_{\ell j_1} = \frac{3}{M}\mu_{j_1}. \end{aligned}$$

If $i \neq i'$, $j_1 \neq j_2$, and $i \in S_{k_1}, i' \in S_{k_2}$ where $k_1 \neq k_2$, we have

$$\begin{aligned} |\alpha_{ii'j_1j_2}| &= |\sum_{\ell > i'} N_\ell \sigma_{i\ell}\sigma_{i'\ell}\delta_{j_1j_2\ell}| \lesssim \sum_{\ell} N_\ell \sigma_{i\ell}\sigma_{i'\ell}\Omega_{\ell j_1}\Omega_{\ell j_2} \\ &\lesssim \frac{1}{M}\cdot\sum_{a=1}^{2}\frac{1}{M_{k_a}}\sum_{\ell\in S_{k_a}} N_\ell\Omega_{\ell j_1}\Omega_{\ell j_2} + \frac{1}{M}\cdot\frac{1}{M}\sum_{\ell\in[n]} N_\ell\Omega_{\ell j_1}\Omega_{\ell j_2} \\ &\leq \frac{1}{M}\sum_{a=1}^{2}\Sigma_{k_aj_1j_2} + \frac{1}{M}\Sigma_{j_1j_2}. \end{aligned}$$

$\square$

B.3.2 Proof of Lemma B.7

We have

$$\mathrm{Var}(V_1) = \sum_{i,r} \mathbb{E}\zeta_{irir}^2.$$

Next by symmetry,

$$\begin{aligned}
\mathbb{E}\zeta_{irir}^2 = & \sum_{j_1,j_2,j_3,j_4} \alpha_{iij_1j_2}\alpha_{iij_3j_4}\,\mathbb{E}Z_{ij_1r}Z_{ij_3r}Z_{ij_2r}Z_{ij_4r} \\
\lesssim & \sum_{j_1} \alpha_{iij_1j_1}^2 \Omega_{ij_1} + \sum_{j_1 \neq j_4} \alpha_{iij_1j_1}\alpha_{iij_1j_4}\,\Omega_{ij_1}\Omega_{ij_4} \\
& + \sum_{j_1 \neq j_3} \alpha_{iij_1j_1}\alpha_{iij_3j_3}\,\Omega_{ij_1}\Omega_{ij_3} + \sum_{j_1 \neq j_2} \alpha_{iij_1j_2}^2\,\Omega_{ij_1}\Omega_{ij_2} \\
& + \sum_{j_1,j_3,j_4 (dist.)} \alpha_{iij_1j_1}\alpha_{iij_3j_4}\,\Omega_{ij_1}\Omega_{ij_3}\Omega_{ij_4} + \sum_{j_1,j_2,j_4 (dist.)} \alpha_{iij_1j_2}\alpha_{iij_1j_4}\Omega_{ij_1}\Omega_{ij_2}\Omega_{ij_4} \\
& + \sum_{j_1,j_2,j_3,j_4 (dist.)} \alpha_{iij_1j_2}\alpha_{iij_3j_4}\Omega_{ij_1}\Omega_{ij_2}\Omega_{ij_3}\Omega_{ij_4} =: \sum_{a=1}^{7} B_{a,i,r}
\end{aligned}$$

Thus

$$\mathrm{Var}(V_1) \lesssim \sum_a \underbrace{\Big(\sum_{i,r} B_{a,i,r} \Big)}_{=:B_a}.$$

We analyze B_1– B_7 separately, bounding the $\alpha_{ii'j_rj_s}$ coefficients using Lemma B.9.

For B_1,

$$\begin{aligned}
B_1 &\lesssim \sum_{i,r}\sum_{j_1} \alpha_{iij_1j_2}^2 \Omega_{ij_1} \lesssim \sum_{k=1}^{k}\sum_{i\in S_k}\sum_{r\in[N_i]}\sum_{j_1} (\frac{1}{M_k}\mu_{j_1})^2 \Omega_{ij_1} \\
&\lesssim \sum_k \sum_{j_1} (\frac{1}{M_k}\mu_{j_1})^2 M_k \mu_{j_1} \lesssim \big(\sum_k \frac{1}{M_k}\big)\|\mu\|_3^3.
\end{aligned} \tag{B.23}$$

For B_2,

$$\begin{aligned}
B_2 &\lesssim \sum_{i,r}\sum_{j_1\neq j_4} \alpha_{iij_1j_1}\alpha_{iij_1j_4}\,\Omega_{ij_1}\Omega_{ij_4} \\
&\lesssim \sum_k \sum_{i\in S_k}\sum_{r\in[N_i]}\sum_{j_1\neq j_4} \frac{1}{M_k}\mu_{j_1} \cdot \big(\frac{1}{M_k}\Sigma_{kj_1j_4} + \frac{1}{M}\Sigma_{j_1j_4}\big) \cdot \Omega_{ij_1}\Omega_{ij_4} \\
&\lesssim \sum_k \sum_{j_1\neq j_4} \frac{1}{M_k}\mu_{j_1} \cdot \big(\frac{1}{M_k}\Sigma_{kj_1j_4} + \frac{1}{M}\Sigma_{j_1j_4}\big) \cdot M_k \Sigma_{kj_1j_4} \\
&\lesssim \sum_k \frac{1}{M_k} \sum_{j_1\neq j_4} \Sigma_{kj_1j_4}^2 \mu_{j_1} + \sum_k \frac{1}{M}\sum_{j_1\neq j_4} \Sigma_{kj_1j_4}\Sigma_{j_1j_4}\mu_{j_1}
\end{aligned}$$

$$\lesssim \sum_k \frac{\mathbf{1}'\Sigma_k^{\circ 2}\mu}{M_k} + \sum_k \frac{\mathbf{1}'(\Sigma_k \circ \Sigma)\mu}{M} = \sum_k \frac{\mathbf{1}'\Sigma_k^{\circ 2}\mu}{M_k}$$

Next,

$$\begin{aligned}
\sum_{j_1 \neq j_4} \Sigma_{kj_1j_4}^2 \mu_{j_1} &= \sum_{j_1 \neq j_4} \frac{1}{M_k^2} \sum_{i,i' \in S_k} N_i N_{i'} \Omega_{ij_1} \Omega_{i'j_1} \Omega_{ij_4} \Omega_{i'j_4} \cdot \mu_{j_1} \\
&\leq \sum_{j_1} \frac{1}{M_k^2} \sum_{i,i' \in S_k} N_i N_{i'} \Omega_{ij_1} \Omega_{i'j_1} \mu_{j_1} \cdot \big(\sum_{j_4} \Omega_{ij_4} \Omega_{i'j_4} \big) \\
&\leq \sum_{j_1} \frac{1}{M_k^2} \sum_{i,i' \in S_k} N_i N_{i'} \Omega_{ij_1} \Omega_{i'j_1} \cdot \mu_{j_1} \\
&\leq \sum_{j_1} \mu_{j_1}^3 = \|\mu\|_3^3, \qquad \text{(B.24)}
\end{aligned}$$

and similarly

$$\begin{aligned}
\sum_{j_1 \neq j_4} \Sigma_{kj_1j_4} \Sigma_{j_1j_4} \mu_{j_1} &= \sum_{j_1 \neq j_4} \frac{1}{M_k M} \sum_{i \in S_k, i' \in [n]} N_i N_{i'} \Omega_{ij_1} \Omega_{i'j_1} \Omega_{ij_4} \Omega_{i'j_4} \cdot \mu_{j_1} \\
&\leq \sum_{j_1} \frac{1}{M_k M} \sum_{i \in S_k, i' \in [n]} N_i N_{i'} \Omega_{ij_1} \Omega_{i'j_1} \mu_{j_1} \\
&= \sum_{j_1} \mu_{j_1}^3 = \|\mu\|_3^3.
\end{aligned}$$

Thus

$$B_2 \lesssim \big(\sum_k \frac{1}{M_k} \big) \|\mu\|_3^3. \qquad \text{(B.25)}$$

For B_3,

$$\begin{aligned}
B_3 &\lesssim \sum_{i,r} \sum_{j_1 \neq j_3} \alpha_{iij_1j_1} \alpha_{iij_3j_3} \Omega_{ij_1} \Omega_{ij_3} \\
&\lesssim \sum_k \sum_{i \in S_k} \sum_{r \in [N_i]} \sum_{j_1 \neq j_3} \frac{1}{M_k} \mu_{j_1} \cdot \frac{1}{M_k} \mu_{j_3} \cdot \Omega_{ij_1} \Omega_{ij_3} \\
&\lesssim \sum_k \sum_{j_1 \neq j_3} \frac{1}{M_k} \mu_{j_1} \cdot \frac{1}{M_k} \mu_{j_3} \cdot M_k \Sigma_{kj_1j_3} \lesssim \sum_k \frac{\mu' \Sigma_k \mu}{M_k}.
\end{aligned}$$

We have by Cauchy-Schwarz,

$$\begin{aligned}
\mu' \Sigma_k \mu &= \frac{1}{M_k} \sum_{i \in S_k} N_i \mu' \Omega_i \Omega_{i'}' \mu \\
&= \frac{1}{M_k} \sum_{i \in S_k} N_i \big(\sum_j \mu_j \Omega_{ij} \big)^2 \\
&\leq \frac{1}{M_k} \sum_{i \in S_k} N_i \big(\sum_j \Omega_{ij} \big) \big(\sum_j \mu_j^2 \Omega_{ij} \big)
\end{aligned}$$

$$= \sum_j \mu_j^3 = \|\mu\|_3^3. \tag{B.26}$$

Thus

$$B_3 \lesssim \big(\sum_k \frac{1}{M_k}\big)\|\mu\|_3^3 \tag{B.27}$$

For B_4,

$$\begin{aligned}
B_4 &\lesssim \sum_{i,r}\sum_{j_1\neq j_2} \alpha_{iij_1j_2}^2\,\Omega_{ij_1}\Omega_{ij_2} \lesssim \sum_k\sum_{i\in S_k}\sum_{r\in[N_i]}\sum_{j_1\neq j_2}\big(\frac{1}{M_k}\Sigma_{kj_1j_2}+\frac{1}{M}\Sigma_{j_1j_2}\big)^2\,\Omega_{ij_1}\Omega_{ij_2}\\
&\lesssim \sum_k\sum_{j_1\neq j_2}\big(\frac{1}{M_k}\Sigma_{kj_1j_2}+\frac{1}{M}\Sigma_{j_1j_2}\big)^2\cdot M_k\Sigma_{kj_1j_2} \lesssim \sum_k \frac{\mathbf{1}'(\Sigma_k^{\circ 3})\mathbf{1}}{M_k} + \sum_k \frac{M_k}{M^2}\mathbf{1}'(\Sigma_k\circ\Sigma^{\circ 2})\mathbf{1}\\
&\lesssim \big(\sum_k \frac{\mathbf{1}'(\Sigma_k^{\circ 3})\mathbf{1}}{M_k}\big) + \frac{1}{M}\mathbf{1}'(\Sigma^{\circ 3})\mathbf{1}.
\end{aligned}$$

First,

$$\begin{aligned}
\mathbf{1}'(\Sigma_k^{\circ 3})\mathbf{1} &= \frac{1}{M_k^3}\sum_{i_1,i_2,i_3\in S_k} N_{i_1}N_{i_2}N_{i_3}\big(\sum_j \Omega_{i_1j}\Omega_{i_2j}\Omega_{i_3j}\big)^2\\
&\le \frac{1}{M_k^3}\sum_{i_1,i_2,i_3\in S_k} N_{i_1}N_{i_2}N_{i_3}\cdot\sum_j \Omega_{i_1j}\Omega_{i_2j}\Omega_{i_3j} = \sum_j \mu_j^3 = \|\mu\|_3^3,
\end{aligned}$$

and similarly,

$$\mathbf{1}'(\Sigma^{\circ 3})\mathbf{1} = \frac{1}{M^3}\sum_{i_1,i_2,i_3\in[n]} N_{i_1}N_{i_2}N_{i_3}\big(\sum_j \Omega_{i_1j}\Omega_{i_2j}\Omega_{i_3j}\big)^2 \le \|\mu\|_3^3.$$

Thus

$$B_4 \lesssim \big(\sum_k \frac{1}{M_k}\big)\|\mu\|_3^3 \tag{B.28}$$

For B_5,

$$\begin{aligned}
B_5 &\lesssim \sum_{i,r}\sum_{j_1,j_3,j_4(dist.)} \alpha_{iij_1j_1}\alpha_{iij_3j_4}\,\Omega_{ij_1}\Omega_{ij_3}\Omega_{ij_4}\\
&\lesssim \sum_k\sum_{i\in S_k} N_i \sum_{j_1,j_3,j_4}\frac{1}{M_k}\mu_{j_1}\cdot\big(\frac{1}{M_k}\Sigma_{kj_3j_4}+\frac{1}{M}\Sigma_{j_3j_4}\big)\cdot\,\Omega_{ij_1}\Omega_{ij_3}\Omega_{ij_4}\\
&\lesssim \sum_k\sum_{i\in S_k}\sum_{j_1,j_3,j_4}\frac{N_i\mu_{j_1}\Sigma_{kj_3j_4}\Omega_{ij_1}\Omega_{ij_3}\Omega_{ij_4}}{M_k^2} + \sum_k\sum_{i\in S_k}\sum_{j_1,j_3,j_4}\frac{N_i\mu_{j_1}\Sigma_{j_3j_4}\Omega_{ij_1}\Omega_{ij_3}\Omega_{ij_4}}{M_kM}\\
&=: B_{51}+B_{52}.
\end{aligned}$$

We have

$$B_{51} = \sum_k \frac{1}{M_k^3}\sum_{i_1,i_2\in S_k}\sum_{j_1,j_3,j_4} N_{i_1}N_{i_2}\mu_{j_1}\Omega_{i_1j_1}\Omega_{i_1j_3}\Omega_{i_2j_3}\Omega_{i_1j_4}\Omega_{i_2j_4}$$

$$
\begin{aligned}
&= \sum_k \frac{1}{M_k^3} \sum_{i_1,i_2\in S_k} N_{i_1}N_{i_2}(\Omega_{i_1}'\mu)\cdot(\Omega_{i_1}'\Omega_{i_2})^2 \\
&\le \sum_k \frac{1}{M_k^3} \sum_{i_1,i_2\in S_k} N_{i_1}N_{i_2}\cdot\Omega_{i_1}'\mu\cdot\Omega_{i_1}'\Omega_{i_2} \\
&= \sum_k \frac{1}{M_k^2}\sum_{i_1} N_{i_1}\mu'\Omega_{i_1}\Omega_{i_1}'\mu = \frac{1}{M_k}\mu'\Sigma_k\mu \le \sum_k \frac{1}{M_k}\|\mu\|_3^3. \qquad \text{(B.29)}
\end{aligned}
$$

In the last line we apply (B.26). Similarly,

$$
\begin{aligned}
B_{52} &= \sum_k \frac{1}{M_k M^2} \sum_{i_1\in S_k, i_2\in[n]} \sum_{j_1,j_3,j_4} N_{i_1}N_{i_2}\mu_{j_1}\Omega_{i_1j_1}\Omega_{i_1j_3}\Omega_{i_2j_3}\Omega_{i_1j_4}\Omega_{i_2j_4} \\
&\le \sum_k \frac{1}{M_k M^2} \sum_{i_1\in S_k, i_2\in[n]} N_{i_1}N_{i_2}\cdot\Omega_{i_1}'\mu\cdot\Omega_{i_1}'\Omega_{i_2} \\
&\le \sum_k \frac{1}{M_k M} \sum_{i_1\in S_k} N_{i_1}\mu'\Omega_{i_1}\Omega_{i_1}'\mu \le \sum_k \frac{1}{M}\|\mu\|_3^3. \qquad \text{(B.30)}
\end{aligned}
$$

Thus

$$
B_5 \lesssim \Big(\sum_k \frac{1}{M_k}\Big)\|\mu\|_3^3. \qquad \text{(B.31)}
$$

For B_6,

$$
\begin{aligned}
B_6 &\lesssim \sum_k \sum_{i\in S_k} \sum_{r\in[N_i]} \sum_{j_1,j_2,j_4(dist.)} \Big(\frac{1}{M_k}\Sigma_{kj_1j_2} + \frac{1}{M}\Sigma_{j_1j_2}\Big)\Big(\frac{1}{M_k}\Sigma_{kj_1j_4} + \frac{1}{M}\Sigma_{j_1j_4}\Big)\Omega_{ij_1}\Omega_{ij_2}\Omega_{ij_4} \\
&\lesssim \sum_k \sum_{i\in S_k} \sum_{r\in[N_i]} \sum_{j_1,j_2,j_4} \frac{\Sigma_{kj_1j_2}^2\Omega_{ij_1}\Omega_{ij_2}\Omega_{ij_4}}{M_k^2} + 2\sum_k \sum_{i\in S_k} \sum_{r\in[N_i]} \sum_{j_1,j_2,j_4} \frac{\Sigma_{kj_1j_2}\Sigma_{j_1j_2}\Omega_{ij_1}\Omega_{ij_2}\Omega_{ij_4}}{M_k M} \\
&\quad + \sum_k \sum_{i\in S_k} \sum_{r\in[N_i]} \sum_{j_1,j_2,j_4} \frac{\Sigma_{j_1j_2}^2\Omega_{ij_1}\Omega_{ij_2}\Omega_{ij_4}}{M^2} =: B_{61} + B_{62} + B_{63}.
\end{aligned}
$$

First,

$$
B_{61} \le \sum_k \sum_{i\in S_k} \sum_{r\in[N_i]} \sum_{j_1,j_2,j_4} \frac{\Sigma_{kj_1j_2}^2\Omega_{ij_1}}{M_k^2} = \sum_k \frac{1}{M_k}\mathbf{1}'\Sigma_k^{\circ 2}\mu \le \sum_k \frac{1}{M_k}\|\mu\|_3^3,
$$

where we applied (B.24). Similarly,

$$
\begin{aligned}
B_{62} &\lesssim \sum_k \frac{1}{M_k}\|\mu\|_3^3, \text{ and} \\
B_{63} &\lesssim \sum_k \frac{1}{M_k}\|\mu\|_3^3.
\end{aligned}
$$

Thus

$$
B_6 \lesssim \Big(\sum_k \frac{1}{M_k}\Big)\|\mu\|_3^3. \qquad \text{(B.32)}
$$

For B_7, we have

$$B_7 \lesssim \sum_{j_1,j_2,j_3,j_4(dist.)} \big(\frac{1}{M_k}\Sigma_{kj_1j_2} + \frac{1}{M}\Sigma_{j_1j_2}\big)\big(\frac{1}{M_k}\Sigma_{kj_3j_4} + \frac{1}{M}\Sigma_{j_3j_4}\big)\Omega_{ij_1}\Omega_{ij_2}\Omega_{ij_3}\Omega_{ij_4}$$
$$\lesssim \sum_k \sum_{i\in S_k} \sum_{r\in[N_i]} \sum_{j_1,j_2,j_3,j_4} \frac{\Sigma_{kj_1j_2}\Sigma_{kj_3j_4}\Omega_{ij_1}\Omega_{ij_2}\Omega_{ij_3}\Omega_{ij_4}}{M_k^2}$$
$$+ 2\sum_k \sum_{i\in S_k} \sum_{r\in[N_i]} \sum_{j_1,j_2,j_3,j_4} \frac{\Sigma_{kj_1j_2}\Sigma_{j_3j_4}\Omega_{ij_1}\Omega_{ij_2}\Omega_{ij_3}\Omega_{ij_4}}{M_k M}$$
$$+ \sum_k \sum_{i\in S_k} \sum_{r\in[N_i]} \sum_{j_1,j_2,j_3,j_4} \frac{\Sigma_{j_1j_2}\Sigma_{j_3j_4}\Omega_{ij_1}\Omega_{ij_2}\Omega_{ij_3}\Omega_{ij_4}}{M^2} =: B_{71} + B_{72} + B_{73}.$$

Note that

$$\Sigma_{kj_1j_2} = \frac{1}{M_k}\sum_{i\in S_k} N_i\Omega_{ij_1}\Omega_{ij_2} \le \frac{1}{M_k}\sum_{i\in S_k} N_i\Omega_{ij_1} = \mu_{j_1}, \text{ and}$$
$$\Sigma_{j_1j_2} = \frac{1}{M}\sum_{i\in[n]} N_i\Omega_{ij_1}\Omega_{ij_2} \le \frac{1}{M}\sum_{i\in[n]} N_i\Omega_{ij_1} = \mu_{j_1}. \tag{B.33}$$

Thus

$$B_{71} \le \sum_k \sum_{i\in S_k} \sum_{r\in[N_i]} \sum_{j_1,j_2,j_3,j_4} \frac{\mu_{j_1}\Sigma_{kj_3j_4}\Omega_{ij_1}\Omega_{ij_2}\Omega_{ij_3}\Omega_{ij_4}}{M_k^2}$$
$$\le \sum_k \sum_{i\in S_k} \sum_{j_1,j_3,j_4} \frac{N_i\mu_{j_1}\Sigma_{kj_3j_4}\Omega_{ij_1}\Omega_{ij_3}\Omega_{ij_4}}{M_k^2} \le \sum_k \frac{1}{M_k}\|\mu\|_3^3$$

where we applied (B.29). Similarly,

$$B_{72} \lesssim \sum_k \frac{1}{M_k}\|\mu\|_3^3, \text{ and}$$
$$B_{73} \lesssim \sum_k \frac{1}{M_k}\|\mu\|_3^3.$$

Thus

$$B_7 \lesssim \big(\sum_k \frac{1}{M_k}\big)\|\mu\|_3^3. \tag{B.34}$$

Combining the results for B_1–B_7 concludes the proof.

□

B.3.3 Proof of Lemma B.8

We have

$$\mathrm{Var}(V_2) \lesssim 4 \sum_{(i,r)\neq(i',r')} \mathbb{E}\zeta_{irir'}^2,$$

where $r \in [N_i]$ and $r \in [N_{i'}]$ in the summation above.

By symmetry, if $(i, r) \neq (i', r')$,

$$
\begin{aligned}
\mathbb{E}\zeta_{iri'r'}^2 &= \sum_{j_1,j_2,j_3,j_4} \alpha_{ii'j_1j_2}\alpha_{ii'j_3j_4}\, \mathbb{E}Z_{ij_1r}Z_{ij_3r}\, \mathbb{E}Z_{i'j_2r'}Z_{i'j_4r'} \\
&\lesssim \sum_{j_1} \alpha_{ii'j_1j_1}^2\, \Omega_{ij_1}\Omega_{i'j_1} + \sum_{j_1\neq j_4} \alpha_{ii'j_1j_1}\alpha_{ii'j_1j_4}\, \Omega_{ij_1}\Omega_{i'j_1}\Omega_{i'j_4} \\
&\quad + \sum_{j_1\neq j_3} \alpha_{ii'j_1j_1}\alpha_{ii'j_3j_3}\, \Omega_{ij_1}\Omega_{ij_3}\Omega_{i'j_1}\Omega_{i'j_3} + \sum_{j_1\neq j_2} \alpha_{ii'j_1j_2}^2\, \Omega_{ij_1}\Omega_{i'j_2} \\
&\quad + \sum_{j_1,j_3,j_4(dist.)} \alpha_{ii'j_1j_1}\alpha_{ii'j_3j_4}\, \Omega_{ij_1}\Omega_{ij_3}\Omega_{i'j_1}\Omega_{i'j_4} + \sum_{j_1,j_2,j_4(dist.)} \alpha_{ii'j_1j_2}\alpha_{ii'j_1j_4}\, \Omega_{ij_1}\Omega_{i'j_2}\Omega_{i'j_4} \\
&\quad + \sum_{j_1,j_2,j_3,j_4(dist.)} \alpha_{ii'j_1j_2}\alpha_{ii'j_3j_4}\Omega_{ij_1}\Omega_{ij_3}\Omega_{i'j_2}\Omega_{i'j_4} =: \sum_a^7 C_{a,i,r}.
\end{aligned} \tag{B.35}
$$

Thus

$$
\mathrm{Var}(V_2) \lesssim \sum_{a=1}^7 \sum_{(i,r)\neq(i',r')} C_{a,i,r} \lesssim \sum_{a=1}^7 \underbrace{\sum_{i,i'} N_iN_{i'}C_{a,i,r}}_{=:C_a}.
$$

Next we analyze $C_1, \ldots, C_7$, bounding the $\alpha_{ii'j_rj_s}$ coefficients using Lemma B.9.

For C_1,

$$
\begin{aligned}
C_1 &\lesssim \sum_k \sum_{i,i'\in S_k} \sum_{j_1} N_iN_{i'}\alpha_{ii'j_1j_1}^2\Omega_{ij_1}\Omega_{i'j_1} + \sum_{k\neq k'} \sum_{i\in S_k, i'\in S_{k'}} \sum_{j_1} N_iN_{i'}\alpha_{ii'j_1j_1}^2\Omega_{ij_1}\Omega_{i'j_1} \\
&\lesssim \sum_k \sum_{i,i'\in S_k} \sum_{j_1} N_iN_{i'}(\frac{1}{M_k}\mu_{j_1})^2\Omega_{ij_1}\Omega_{ij_1} + \sum_{k\neq k'} \sum_{i\in S_k, i'\in S_{k'}} \sum_{j_1}(\frac{1}{M}\mu_{j_1})^2\Omega_{ij_1}\Omega_{i'j_1} \\
&\lesssim \sum_k \sum_{j_1} \mu_{j_1}^4 + \sum_{k\neq k'} \sum_{j_1} \frac{M_kM_{k'}}{M^2}\mu_{j_1}^4 \lesssim K\|\mu\|_4^4.
\end{aligned} \tag{B.36}
$$

For C_2,

$$
\begin{aligned}
C_2 &\lesssim \sum_k \sum_{i,i'\in S_k} N_iN_{i'} \sum_{j_1\neq j_4} \alpha_{ii'j_1j_1}\alpha_{ii'j_1j_4}\, \Omega_{ij_1}\Omega_{i'j_1}\Omega_{i'j_4} \\
&\quad + \sum_{k\neq k'} \sum_{i\in S_k, i'\in S_{k'}} N_iN_{i'} \sum_{j_1\neq j_4} \alpha_{ii'j_1j_1}\alpha_{ii'j_1j_4}\, \Omega_{ij_1}\Omega_{i'j_1}\Omega_{i'j_4} \\
&\lesssim \sum_k \sum_{i,i'\in S_k} N_iN_{i'} \sum_{j_1\neq j_4} \frac{1}{M_k}\mu_{j_1} \cdot \big(\frac{1}{M_k}\Sigma_{kj_1j_4} + \frac{1}{M}\Sigma_{j_1j_4}\big)\, \Omega_{ij_1}\Omega_{i'j_1}\Omega_{i'j_4} \\
&\quad + \sum_{k\neq k'} \sum_{i\in S_k, i'\in S_{k'}} N_iN_{i'} \sum_{j_1\neq j_4} \frac{1}{M}\mu_{j_1} \cdot \big(\frac{1}{M} \sum_{a\in\{k,k'\}} \Sigma_{aj_1j_4} + \frac{1}{M}\Sigma_{j_1j_4}\big)\, \Omega_{ij_1}\Omega_{i'j_1}\Omega_{i'j_4} \\
&\lesssim \sum_k \sum_{i,i'\in S_k} N_iN_{i'} \sum_{j_1\neq j_4} \frac{1}{M_k}\mu_{j_1} \cdot \big(\frac{1}{M_k}\mu_{j_1} + \frac{1}{M}\mu_{j_1}\big)\, \Omega_{ij_1}\Omega_{i'j_1}\Omega_{i'j_4} \\
&\quad + \sum_{k\neq k'} \sum_{i\in S_k, i'\in S_{k'}} N_iN_{i'} \sum_{j_1\neq j_4} \frac{1}{M}\mu_{j_1} \cdot \big(\frac{2}{M}\mu_{j_1} + \frac{1}{M}\mu_{j_1}\big)\, \Omega_{ij_1}\Omega_{i'j_1}\Omega_{i'j_4}
\end{aligned}
$$

$$\lesssim \sum_k \sum_{j_1} \big(\mu_{j_1}^4 + \frac{M_k}{M}\mu_{j_1}^4\big) + \sum_{k\neq k'} \sum_{j_1} \frac{M_k M_{k'}}{M^2}\mu_{j_1}^4 \lesssim K\|\mu\|_4^4. \tag{B.37}$$

where we applied (B.33).

For C_3,

$$\begin{aligned}
C_3 &\lesssim \Big(\sum_k \sum_{i,i'\in S_k} N_i N_{i'} + \sum_{k\neq k'} \sum_{i\in S_k, i'\in S_{k'}} N_i N_{i'}\Big) \sum_{j_1\neq j_3} \alpha_{ii'j_1j_1}\alpha_{ii'j_3j_3}\,\Omega_{ij_1}\Omega_{ij_3}\Omega_{i'j_1}\Omega_{i'j_3} \\
&\lesssim \sum_k \sum_{i,i'\in S_k} N_i N_{i'} \sum_{j_1\neq j_3} \frac{1}{M_k}\mu_{j_1}\cdot\frac{1}{M_k}\mu_{j_3}\cdot\Omega_{ij_1}\Omega_{ij_3}\Omega_{i'j_1}\Omega_{i'j_3} \\
&\quad + \sum_{k\neq k'} \sum_{i\in S_k, i'\in S_{k'}} N_i N_{i'} \sum_{j_1\neq j_3} \frac{1}{M}\mu_{j_1}\cdot\frac{1}{M}\mu_{j_3}\,\Omega_{ij_1}\Omega_{ij_3}\Omega_{i'j_1}\Omega_{i'j_3} \\
&= \sum_k \sum_{j_1\neq j_3} \mu_{j_1}\mu_{j_3}\Sigma_{kj_1j_3}^2 + \sum_{k\neq k'} \sum_{j_1\neq j_3} \frac{M_k M_{k'}}{M^2}\mu_{j_1}\mu_{j_3}\Sigma_{kj_1j_3}\Sigma_{k'j_1j_3} \\
&\leq \big(\sum_k \mu'\Sigma_k^{\circ 2}\mu\big) + \mu'\Sigma^{\circ 2}\mu.
\end{aligned}$$

First, by Cauchy–Schwarz,

$$\begin{aligned}
\mu'\Sigma_k^{\circ 2}\mu &= \frac{1}{M_k^2}\sum_{i,i'\in S_k} N_i N_{i'}\big(\sum_j \mu_j \Omega_{ij}\Omega_{i'j}\big)^2 \\
&= \frac{1}{M_k^2}\sum_{i,i'\in S_k} N_i N_{i'}\big(\sum_j \Omega_{ij}\Omega_{i'j}\big)\sum_j \mu_j^2\Omega_{ij}\Omega_{i'j} \\
&\leq \frac{1}{M_k^2}\sum_{i,i'\in S_k} N_i N_{i'}\sum_j \mu_j^2\Omega_{ij}\Omega_{i'j} = \sum_j \mu_j^4 = \|\mu\|_4^4.
\end{aligned} \tag{B.38}$$

Similarly

$$\mu'\Sigma^{\circ 2}\mu \lesssim \|\mu\|_4^4. \tag{B.39}$$

Hence

$$C_3 \lesssim K\|\mu\|_4^4. \tag{B.40}$$

For C_4,

$$\begin{aligned}
C_4 &\lesssim \Big(\sum_k \sum_{i,i'\in S_k} N_i N_{i'} + \sum_{k\neq k'} \sum_{i\in S_k, i'\in S_{k'}} N_i N_{i'}\Big) \sum_{j_1\neq j_2} \alpha_{ii'j_1j_2}^2\,\Omega_{ij_1}\Omega_{i'j_2} \\
&\lesssim \sum_k \sum_{i,i'\in S_k} N_i N_{i'} \sum_{j_1\neq j_2} \big(\frac{1}{M_k}\Sigma_{kj_1j_2} + \frac{1}{M}\Sigma_{j_1j_2}\big)^2\,\Omega_{ij_1}\Omega_{i'j_2} \\
&\quad + \sum_{k\neq k'} \sum_{i\in S_k, i'\in S_{k'}} N_i N_{i'} \sum_{j_1\neq j_2} \big(\frac{1}{M}\sum_{a\in\{k,k'\}}^{2}\Sigma_{aj_1j_2} + \frac{1}{M}\Sigma_{j_1j_2}\big)^2\,\Omega_{ij_1}\Omega_{i'j_2} \\
&\lesssim \sum_k \sum_{i,i'\in S_k} N_i N_{i'} \sum_{j_1\neq j_2} \big(\frac{1}{M_k^2}\Sigma_{kj_1j_2}^2 + \frac{1}{M^2}\Sigma_{j_1j_2}^2\big)\,\Omega_{ij_1}\Omega_{i'j_2}
\end{aligned}$$

$$+\sum_{k\neq k'}\sum_{i\in S_k,i'\in S_{k'}} N_iN_{i'}\sum_{j_1\neq j_2}\big(\frac{1}{M^2}\sum_{a\in\{k,k'\}}^{2}\Sigma^2_{aj_1j_2}+\frac{1}{M^2}\Sigma^2_{j_1j_2}\big)\,\Omega_{ij_1}\Omega_{i'j_2} =: C_{41}+C_{42}$$

First,

$$\begin{aligned}
C_{41} &\lesssim \sum_k\sum_{i,i'\in S_k} N_iN_{i'}\sum_{j_1\neq j_2}\frac{1}{M_k^2}\Sigma^2_{kj_1j_2}\Omega_{ij_1}\Omega_{i'j_2}+\sum_k\sum_{i,i'\in S_k} N_iN_{i'}\sum_{j_1\neq j_2}\frac{1}{M^2}\Sigma^2_{j_1j_2}\Omega_{ij_1}\Omega_{i'j_2}\\
&\lesssim \sum_k\sum_{j_1\neq j_2}\Sigma^2_{kj_1j_2}\mu_{j_1}\mu_{j_2}+\sum_k\sum_{j_1\neq j_2}\frac{M_k^2}{M^2}\Sigma^2_{j_1j_2}\mu_{j_1}\mu_{j_2}\leq \sum_k \mu'\Sigma_k^{\circ 2}\mu+\sum_k\frac{M_k^2}{M^2}\mu'\Sigma^{\circ 2}\mu.
\end{aligned}$$

Similarly,

$$\begin{aligned}
C_{42} &\lesssim \sum_{k\neq k'}\sum_{j_1\neq j_2}\frac{M_kM_{k'}}{M^2}\Sigma^2_{kj_1j_2}\mu_{j_1}\mu_{j_2}+\sum_{k\neq k'}\sum_{j_1\neq j_2}\frac{M_kM_{k'}}{M^2}\Sigma^2_{j_1j_2}\mu_{j_1}\mu_{j_2}\\
&\lesssim \sum_{k\neq k'}\frac{M_kM_{k'}}{M^2}\big(\mu'\Sigma_k^{\circ 2}\mu+\mu'\Sigma^{\circ 2}\mu\big)
\end{aligned}$$

Combining the previous two displays and applying (B.38) and (B.39), we have

$$C_4\lesssim K\|\mu\|_4^4. \tag{B.41}$$

For C_5,

$$\begin{aligned}
C_5 &\lesssim \Bigg(\sum_k\sum_{i,i'\in S_k} N_iN_{i'}+\sum_{k\neq k'}\sum_{i\in S_k,i'\in S_{k'}} N_iN_{i'}\Bigg)\sum_{j_1,j_3,j_4(dist.)}\alpha_{ii'j_1j_1}\alpha_{ii'j_3j_4}\,\Omega_{ij_1}\Omega_{ij_3}\Omega_{i'j_1}\Omega_{i'j_4}\\
&\lesssim \sum_k\sum_{i,i'\in S_k} N_iN_{i'}\sum_{j_1,j_3,j_4}\frac{1}{M_k}\mu_{j_1}\cdot\big(\frac{1}{M_k}\Sigma_{kj_3j_4}+\frac{1}{M}\Sigma_{j_3j_4}\big)\,\Omega_{ij_1}\Omega_{ij_3}\Omega_{i'j_1}\Omega_{i'j_4}\\
&\quad+\sum_{k\neq k'}\sum_{i\in S_k,i'\in S_{k'}} N_iN_{i'}\sum_{j_1,j_3,j_4}\frac{1}{M}\mu_{j_1}\big(\frac{1}{M}\sum_{a\in\{k,k'\}}^{2}\Sigma_{aj_3j_4}+\frac{1}{M}\Sigma_{j_3j_4}\big)\,\Omega_{ij_1}\Omega_{ij_3}\Omega_{i'j_1}\Omega_{i'j_4}\\
&=\sum_k\sum_{j_1,j_3,j_4}\mu_{j_1}\Sigma_{kj_3j_4}\Sigma_{kj_1j_3}\Sigma_{kj_1j_4}+\sum_k\sum_{j_1,j_3,j_4}\frac{M_k}{M}\mu_{j_1}\Sigma_{j_3j_4}\Sigma_{kj_1j_3}\Sigma_{kj_1j_4}\\
&\quad+2\sum_{k\neq k'}\sum_{j_1,j_3,j_4}\frac{M_kM_{k'}}{M^2}\mu_{j_1}\Sigma_{kj_3j_4}\Sigma_{kj_1j_3}\Sigma_{k'j_1j_4}+\sum_{k\neq k'}\sum_{j_1,j_3,j_4}\frac{M_kM_{k'}}{M^2}\mu_{j_1}\Sigma_{j_3j_4}\Sigma_{kj_1j_3}\Sigma_{k'j_1j_4}\\
&=C_{51}+C_{52}+2C_{53}+C_{54}
\end{aligned}$$

For C_{51}, we have

$$\begin{aligned}
C_{51} &= \sum_k\frac{1}{M_k^3}\sum_{i_1,i_2,i_3\in S_k} N_{i_1}N_{i_2}N_{i_3}\langle\mu\circ\Omega_{i_1},\Omega_{i_2}\rangle\langle\Omega_{i_1},\Omega_{i_3}\rangle\langle\Omega_{i_2},\Omega_{i_3}\rangle\\
&=\sum_k\frac{1}{M_k^2}\sum_{i_1,i_2\in S_k} N_{i_1}N_{i_2}\langle\mu\circ\Omega_{i_1},\Omega_{i_2}\rangle\cdot\langle\Omega_{i_1},\Sigma_k\Omega_{i_2}\rangle\\
&\leq \sum_k\Bigg(\frac{1}{M_k^2}\sum_{i_1,i_2\in S_k} N_{i_1}N_{i_2}\langle\mu\circ\Omega_{i_1},\Omega_{i_2}\rangle^2\Bigg)^{1/2}\Bigg(\frac{1}{M_k^2}\sum_{i_1,i_2\in S_k} N_{i_1}N_{i_2}\langle\Omega_{i_1},\Sigma_k\Omega_{i_2}\rangle^2\Bigg)^{1/2}
\end{aligned}$$

$$=: \sum_k C_{511k}^{1/2} \cdot C_{512k}^{1/2}. \tag{B.42}$$

We have by Cauchy–Schwarz that

$$\begin{aligned} C_{511k} &= \frac{1}{M_k^2} \sum_{i_1,i_2\in S_k} N_{i_1} N_{i_2} \big(\sum_j \mu_j \Omega_{i_1 j} \Omega_{i_2 j} \big)^2 \\ &\le \frac{1}{M_k^2} \sum_{i_1,i_2\in S_k} N_{i_1} N_{i_2} \big(\sum_j \mu_j^2 \Omega_{i_1 j} \Omega_{i_2 j} \big) \big(\sum_j \Omega_{i_1 j} \Omega_{i_2 j} \big) \le \|\mu\|_4^4, \end{aligned}$$

and similarly

$$\begin{aligned} C_{512k} &= \frac{1}{M_k^2} \sum_{i_1,i_2\in S_k} N_{i_1} N_{i_2} \big(\sum_{j_1,j_2} \Omega_{i_1 j_1} \Sigma_{k j_1 j_2} \Omega_{i_2 j_2} \big)^2 \\ &= \frac{1}{M_k^2} \sum_{i_1,i_2} N_{i_1} N_{i_2} \big(\sum_{j_1,j_2} \Omega_{i_1 j_1} \Sigma_{k j_1 j_2}^2 \Omega_{i_2 j_2} \big) \big(\sum_{j_1,j_2} \Omega_{i_1 j_1} \Omega_{i_2 j_2} \big) \\ &\le \frac{1}{M_k^2} \sum_{i_1,i_2} N_{i_1} N_{i_2} \big(\sum_{j_1,j_2} \Omega_{i_1 j_1} \Sigma_{k j_1 j_2}^2 \Omega_{i_2 j_2} \big) = \mu' \, \Sigma_k^{\circ 2} \, \mu \end{aligned} \tag{B.43}$$

Since by Cauchy–Schwarz,

$$\begin{aligned} \mu' \, \Sigma_k^{\circ 2} \, \mu &= \sum_{j_1,j_2} \mu_{j_1} \mu_{j_2} \big(\frac{1}{M_k} \sum_{i\in S_k} N_i \Omega_{i j_1} \Omega_{i j_2} \big)^2 = \frac{1}{M_k^2} \sum_{j_1,j_2} \mu_{j_1} \mu_{j_2} \sum_{i,i'\in S_k} N_i N_{i'} \Omega_{i j_1} \Omega_{i j_2} \Omega_{i' j_1} \Omega_{i' j_2} \\ &= \frac{1}{M_k^2} \sum_{i,i'\in S_k} \big(\sum_j \mu_j \Omega_{ij} \Omega_{i'j} \big)^2 \le \frac{1}{M_k^2} \sum_{i,i'\in S_k} \sum_j \mu_j^2 \Omega_{ij} \Omega_{i'j} \le \|\mu\|_4^4 \end{aligned} \tag{B.44}$$

we have in total $C_{512k} \lesssim K \|\mu\|_4^4$. Combining the result with the bound for C_{511k} implies that

$$C_{51} \lesssim K \|\mu\|_4^4.$$

Next we study C_{52} using a similar argument.

$$\begin{aligned} C_{52} &= \sum_k \sum_{j_1,j_3,j_4} \frac{M_k}{M} \mu_{j_1} \Sigma_{j_3 j_4} \Sigma_{k j_1 j_3} \Sigma_{k j_1 j_4} \\ &= \sum_k \sum_{j_1,j_3,j_4} \frac{M_k}{M} \mu_{j_1} \big(\frac{1}{M} \sum_{i_1\in[n]} N_{i_1} \Omega_{i_1 j_3} \Omega_{i_1 j_4} \big) \big(\frac{1}{M_k} \sum_{i_2\in S_k} N_{i_2} \Omega_{i_2 j_1} \Omega_{i_2 j_3} \big) \big(\frac{1}{M_k} \sum_{i_3\in S_k} N_{i_3} \Omega_{i_3 j_1} \Omega_{i_3 j_4} \big) \\ &= \sum_k \frac{1}{M^2 M_k} \sum_{j_1,j_2,j_3} \sum_{\substack{i_1\in[n] \\ i_2,i_3\in S_k}} N_{i_1} N_{i_2} N_{i_3} \langle \mu \circ \Omega_{i_2}, \Omega_{i_3} \rangle \langle \Omega_{i_1}, \Omega_{i_3} \rangle \langle \Omega_{i_1}, \Omega_{i_2} \rangle \\ &= \sum_k \frac{1}{M^2} \sum_{i_2,i_3\in[S_k]} N_{i_2} N_{i_3} \langle \mu \circ \Omega_{i_2}, \Omega_{i_3} \rangle \langle \Omega_{i_3}, \Sigma \Omega_{i_2} \rangle \\ &\le \sum_k \left(\frac{1}{M^2} \sum_{i_2,i_3\in[S_k]} N_{i_2} N_{i_3} \langle \mu \circ \Omega_{i_2}, \Omega_{i_3} \rangle^2 \right)^{1/2} \left(\frac{1}{M^2} \sum_{i_2,i_3\in[S_k]} N_{i_2} N_{i_3} \langle \Omega_{i_3}, \Sigma \Omega_{i_2} \rangle \right)^{1/2} \end{aligned}$$

$$=: \sum_k C_{521k}^{1/2} C_{522k}^{1/2}. \tag{B.45}$$

Observe that $C_{521k} = C_{511k}$, and thus $C_{521} \lesssim \|\mu\|^4$ by (B.43). With a similar argument as in (B.44) we obtain $C_{522k} \lesssim \|\mu\|_4^4$. Hence we obtain

$$C_{52} \leq \sum_k C_{521k}^{1/2} C_{522k}^{1/2} \lesssim K\|\mu\|_4^4.$$

For C_{53}, we have

$$\begin{aligned}
C_{53} &= \sum_{k \neq k'} \sum_{j_1, j_3, j_4} \frac{M_k M_{k'}}{M^2} \mu_{j_1} \Sigma_{kj_3j_4} \Sigma_{kj_1j_3} \Sigma_{k'j_1j_4} \\
&\leq \sum_k \sum_{j_1, j_3, j_4} \frac{M_k}{M} \mu_{j_1} \Sigma_{kj_3j_4} \Sigma_{kj_1j_3} \Sigma_{j_1j_4} \\
&= \sum_k \sum_{j_1, j_3, j_4} \frac{M_k}{M} \mu_{j_1} \big(\frac{1}{M_k} \sum_{i_1 \in S_k} N_{i_1} \Omega_{i_1j_3} \Omega_{i_1j_4} \big) \big(\frac{1}{M_k} \sum_{i_2 \in S_k} N_{i_2} \Omega_{i_2j_1} \Omega_{i_2j_3} \big) \big(\frac{1}{M} \sum_{i_3 \in [n]} N_{i_3} \Omega_{i_3j_1} \Omega_{i_3j_4} \big) \\
&= \sum_k \frac{1}{M^2 M_k} \sum_{\substack{i_1, i_2 \in S_k \\ i_3 \in [n]}} N_{i_1} N_{i_2} N_{i_3} \langle \mu \circ \Omega_{i_2}, \Omega_{i_3} \rangle \langle \Omega_{i_1}, \Omega_{i_2} \rangle \langle \Omega_{i_1}, \Omega_{i_3} \rangle \\
&= \sum_k \frac{1}{M^2} \sum_{i_2 \in S_k, i_3 \in [n]} N_{i_2} N_{i_3} \langle \mu \circ \Omega_{i_2}, \Omega_{i_3} \rangle \langle \Omega_{i_2}, \Sigma_k \Omega_{i_3} \rangle. \qquad \text{(B.46)}
\end{aligned}$$

We then upper bound the last line using a similar strategy as in that we used for C_{51} and C_{52}, respectively. We omit the details and state the final bound:

$$C_{53} \lesssim K\|\mu\|_4^4 \tag{B.47}$$

Finally for C_{54}, summing over k, k' we obtain

$$C_{54} \leq \sum_{j_1, j_3, j_4} \mu_{j_1} \Sigma_{j_3j_4} \Sigma_{j_1j_3} \Sigma_{j_1j_4} = \frac{1}{M^3} \sum_{i_1, i_2, i_3 \in [n]} N_{i_1} N_{i_2} N_{i_3} \langle \mu \circ \Omega_{i_2}, \Omega_{i_3} \rangle \langle \Omega_{i_1}, \Omega_{i_2} \rangle \langle \Omega_{i_1}, \Omega_{i_3} \rangle. \tag{B.48}$$

We then proceed as in (B.46) to control the right-hand side. We omit the details and state the final bound:

$$C_{54} \lesssim K\|\mu\|_4^4. \tag{B.49}$$

Combining the results for $C_{51}, \ldots, C_{54}$, we see that

$$C_5 \lesssim K\|\mu\|^4.$$

For C_6, we have

$$C_6 \leq \Bigg(\sum_k \sum_{i, i' \in S_k} N_i N_{i'} + \sum_{k \neq k'} \sum_{i \in S_k, i' \in S_{k'}} N_i N_{i'} \Bigg) \sum_{j_1, j_2, j_4} \alpha_{ii'j_1j_2} \alpha_{ii'j_1j_4} \Omega_{ij_1} \Omega_{i'j_2} \Omega_{i'j_4}$$

$$\lesssim \sum_k \sum_{i,i'\in S_k} N_i N_{i'} \sum_{j_1,j_2,j_4} \big(\frac{1}{M_k}\Sigma_{kj_1j_2} + \frac{1}{M}\Sigma_{j_1j_2}\big)\big(\frac{1}{M_k}\Sigma_{kj_1j_4} + \frac{1}{M}\Sigma_{j_1j_4}\big)\,\Omega_{ij_1}\Omega_{i'j_2}\Omega_{i'j_4}$$

$$+ \sum_{\substack{k\neq k'\\ i\in S_k, i'\in S_{k'}\\ j_1,j_2,j_4}} N_i N_{i'}\big(\frac{1}{M}\sum_{a\in\{k,k'\}}^{2} \Sigma_{aj_1j_2} + \frac{1}{M}\Sigma_{j_1j_2}\big)\big(\frac{1}{M}\sum_{a\in\{k,k'\}}^{2} \Sigma_{aj_1j_4} + \frac{1}{M}\Sigma_{j_1j_4}\big)\,\Omega_{ij_1}\Omega_{i'j_2}\Omega_{i'j_4}$$

$$=: C_{61} + C_{62}.$$

For C_{61}, we have

$$\begin{aligned} C_{61} &= \sum_k \sum_{i'\in S_k} N_{i'} \sum_{j_1,j_2,j_4} \frac{1}{M_k}\Sigma_{kj_1j_2}\Sigma_{kj_1j_4}\mu_{j_1}\Omega_{i'j_2}\Omega_{i'j_4} \\ &\quad + 2\sum_k \sum_{i'\in S_k} N_{i'} \sum_{j_1,j_2,j_4} \frac{1}{M}\Sigma_{kj_1j_2}\Sigma_{j_1j_4}\mu_{j_1}\Omega_{i'j_2}\Omega_{i'j_4} \\ &\quad + \sum_k \sum_{i'\in S_k} N_{i'} \sum_{j_1,j_2,j_4} \frac{M_k}{M^2}\Sigma_{j_1j_2}\Sigma_{j_1j_4}\mu_{j_1}\Omega_{i'j_2}\Omega_{i'j_4} =: C_{611} + 2C_{612} + C_{613}. \end{aligned}$$

Relabeling indices, we see that

$$C_{611} = \sum_k \sum_{j_1,j_2,j_4} \mu_{j_1}\Sigma_{kj_1j_2}\Sigma_{kj_1j_4}\Sigma_{kj_2j_4} = C_{51}$$

Hence, $C_{611} \lesssim K\|\mu\|_4^4$. Next,

$$C_{612} \leq \sum_k \frac{M_k}{M} \sum_{j_1,j_2,j_4} \mu_{j_1}\Sigma_{kj_1j_2}\Sigma_{j_1j_4}\Sigma_{kj_2j_4} \lesssim K\|\mu\|^4,$$

where we applied (B.46). Similarly,

$$C_{613} = \sum_k \frac{M_k^2}{M^2} \sum_{j_1,j_2,j_4} \mu_{j_1}\Sigma_{j_1j_2}\Sigma_{j_1j_4}\Sigma_{kj_2j_4} \leq \sum_{j_1,j_2,j_4} \mu_{j_1}\Sigma_{j_1j_2}\Sigma_{j_1j_4}\Sigma_{j_2j_4} \lesssim K\|\mu\|^4,$$

where in the final bound we apply (B.48) and (B.49). Combining the results above for $C_{611}, C_{612}, C_{613}$, we obtain

$$C_{61} \lesssim K\|\mu\|_4^4 \tag{B.50}$$

The argument for C_{62} is very similar, so we omit proof and state the final bound. We have

$$C_{62} \lesssim K\|\mu\|_4.$$

Thus

$$C_6 \lesssim K\|\mu\|_4^4$$

For C_7, we have

$$C_7 \lesssim \Bigg(\sum_k \sum_{i,i'\in S_k} N_i N_{i'} + \sum_{k\neq k'} \sum_{i\in S_k, i'\in S_{k'}} N_i N_{i'}\Bigg) \sum_{j_1,j_2,j_3,j_4} \alpha_{ii'j_1j_2}\alpha_{ii'j_3j_4}\Omega_{ij_1}\Omega_{ij_3}\Omega_{i'j_2}\Omega_{i'j_4}$$

$$\lesssim \sum_k \sum_{i,i'\in S_k} N_i N_{i'} \sum_{j_1,j_2,j_3,j_4} \big(\frac{1}{M_k}\Sigma_{kj_1j_2} + \frac{1}{M}\Sigma_{j_1j_2}\big)\big(\frac{1}{M_k}\Sigma_{kj_3j_4} + \frac{1}{M}\Sigma_{j_3j_4}\big)\Omega_{ij_1}\Omega_{ij_3}\Omega_{i'j_2}\Omega_{i'j_4}$$

$$+ \sum_{k\neq k'} \sum_{\substack{j_1,j_2,j_3,j_4\\ i\in S_k, i'\in S_{k'}}} N_i N_{i'} \big(\frac{1}{M}\sum_{a\in\{k,k'\}}^{2} \Sigma_{aj_1j_2} + \frac{1}{M}\Sigma_{j_1j_2}\big)\big(\frac{1}{M}\sum_{a\in\{k,k'\}}^{2} \Sigma_{aj_3j_4} + \frac{1}{M}\Sigma_{j_3j_4}\big)\Omega_{ij_1}\Omega_{ij_3}\Omega_{i'j_2}\Omega_{i'j_4}$$

$$=: C_{71} + C_{72}$$

Write

$$\begin{aligned}
C_{71} &= \sum_k \sum_{i,i'\in S_k} N_i N_{i'} \sum_{j_1,j_2,j_3,j_4} \frac{1}{M_k^2}\Sigma_{kj_1j_2}\Sigma_{kj_3j_4}\Omega_{ij_1}\Omega_{ij_3}\Omega_{i'j_2}\Omega_{i'j_4} \\
&\quad + 2\sum_k \sum_{i,i'\in S_k} N_i N_{i'} \sum_{j_1,j_2,j_3,j_4} \frac{1}{M_k M}\Sigma_{j_1j_2}\Sigma_{kj_3j_4}\Omega_{ij_1}\Omega_{ij_3}\Omega_{i'j_2}\Omega_{i'j_4} \\
&\quad + \sum_k \sum_{i,i'\in S_k} N_i N_{i'} \sum_{j_1,j_2,j_3,j_4} \frac{1}{M^2}\Sigma_{j_1j_2}\Sigma_{j_3j_4}\Omega_{ij_1}\Omega_{ij_3}\Omega_{i'j_2}\Omega_{i'j_4} =: C_{711} + 2C_{712} + C_{713}.
\end{aligned}$$

For C_{711}, we have

$$\begin{aligned}
C_{711} &= \sum_k \sum_{j_1,j_2,j_3,j_4} \Sigma_{kj_1j_2}\Sigma_{kj_3j_4}\Sigma_{kj_1j_3}\Sigma_{kj_2j_4} \\
&= \sum_k \frac{1}{M_k^4} \sum_{i_1,i_2,i_3,i_4\in S_k} N_{i_1}N_{i_2}N_{i_3}N_{i_4}\langle\Omega_{i_1},\Omega_{i_3}\rangle\langle\Omega_{i_1},\Omega_{i_4}\rangle\langle\Omega_{i_2},\Omega_{i_3}\rangle\langle\Omega_{i_2},\Omega_{i_4}\rangle \\
&= \frac{1}{M_k^2}\sum_k \sum_{i_3,i_4} N_{i_3}N_{i_4}\big(\Omega_{i_3}'\Sigma_k\Omega_{i_4}\big)^2 = \sum_k \frac{1}{M_k^2}\sum_{i_3,i_4} N_{i_3}N_{i_4}\big(\sum_{j,j'}\Omega_{i_3j}'\Sigma_{kjj'}\Omega_{i_4j'}\big)^2 \\
&\le \sum_k \frac{1}{M_k^2}\sum_{i_3,i_4} N_{i_3}N_{i_4}\sum_{j,j'}\Omega_{i_3j}'\Sigma_{kjj'}^2\Omega_{i_4j'} \le \sum_k\sum_{j,j'}\mu_j\Sigma_{kjj'}^2\mu_{j'} \lesssim K\|\mu\|_4^4.
\end{aligned} \tag{B.51}$$

In the last line we applied Cauchy–Schwarz and (B.44). For C_{712}, we have similarly

$$\begin{aligned}
C_{712} &= \sum_k \frac{M_k}{M} \sum_{j_1,j_2,j_3,j_4} \Sigma_{j_1j_2}\Sigma_{kj_3j_4}\Sigma_{kj_1j_3}\Sigma_{kj_2j_4} \\
&= \sum_k \frac{1}{M^2 M_k} \sum_{\substack{i_1\in[n]\\ i_2,i_3,i_4\in S_k}} N_{i_1}N_{i_2}N_{i_3}N_{i_4}\langle\Omega_{i_1},\Omega_{i_3}\rangle\langle\Omega_{i_1},\Omega_{i_4}\rangle\langle\Omega_{i_2},\Omega_{i_3}\rangle\langle\Omega_{i_2},\Omega_{i_4}\rangle \\
&= \sum_k \frac{M_k}{M^2} \sum_{i_1\in[n], i_2\in S_k} N_{i_1}N_{i_2}\langle\Omega_{i_1},\Sigma_k\Omega_{i_2}\rangle^2 \le \sum_k \frac{M_k}{M^2}\sum_{i_1\in[n],i_2\in S_k} N_{i_1}N_{i_2}\sum_{j,j'}\Omega_{i_1j}\Sigma_{kjj'}^2\Omega_{i_2j'} \\
&\le \sum_k \frac{M_k^2}{M^2}\sum_{j,j'}\mu_j\Sigma_{kjj'}^2\mu_{j'} \lesssim K\|\mu\|_4^4.
\end{aligned} \tag{B.52}$$

Next,

$$C_{713} = \sum_k \frac{M_k^2}{M^2} \sum_{j_1,j_2,j_3,j_4} \Sigma_{j_1j_2}\Sigma_{j_3j_4}\Sigma_{kj_1j_3}\Sigma_{kj_2j_4}$$

$$= \sum_k \frac{1}{M^4} \sum_{\substack{i_1,i_2\in[n]\\ i_3,i_4\in S_k}} N_{i_1}N_{i_2}N_{i_3}N_{i_4}\langle\Omega_{i_1},\Omega_{i_3}\rangle\langle\Omega_{i_1},\Omega_{i_4}\rangle\langle\Omega_{i_2},\Omega_{i_3}\rangle\langle\Omega_{i_2},\Omega_{i_4}\rangle,$$

and applying a similar strategy as in (B.51), (B.52) leads to the bound $C_{713} \lesssim K\|\mu\|_4^4$. Thus

$$C_{71} \lesssim K\|\mu\|_4^4.$$

Next , by symmetry and summing over $i \in S_k, i' \in S_{k'}$, we have

$$\begin{aligned}
C_{72} &= \sum_{k\neq k'} \frac{M_k M_{k'}}{M^2} \sum_{j_1,j_2,j_3,j_4} \Big[2\Sigma_{kj_1j_2}\Sigma_{kj_3j_4} + 2\Sigma_{k'j_1j_2}\Sigma_{kj_3j_4} + 4\Sigma_{kj_1j_2}\Sigma_{j_3j_4} + \Sigma_{j_1j_2}\Sigma_{j_3j_4}\Big]\Sigma_{kj_1j_3}\Sigma_{k'j_2j_4} \\
&=: 2C_{721} + 2C_{722} + 4C_{723} + C_{724}
\end{aligned}$$

First,

$$C_{721} \le \sum_k \frac{M_k}{M} \sum_{j_1,j_2,j_3,j_4} \Sigma_{kj_1j_2}\Sigma_{kj_3j_4}\Sigma_{kj_1j_3}\Sigma_{j_2j_4} = C_{712} \lesssim K\|\mu\|_4^4$$

by (B.52). Next,

$$\begin{aligned}
C_{722} &= \sum_{k\neq k'} \frac{M_k M_{k'}}{M^2} \sum_{j_1,j_2,j_3,j_4} \Sigma_{k'j_1j_2}\Sigma_{kj_3j_4}\Sigma_{kj_1j_3}\Sigma_{k'j_2j_4} \\
&\le \sum_{k,k'} \frac{1}{M^2 M_k M_{k'}} \sum_{\substack{i_1,i_2\in S_k\\ i_3,i_4\in S_{k'}}} N_{i_1}N_{i_2}N_{i_3}N_{i_4}\langle\Omega_{i_1},\Omega_{i_3}\rangle\langle\Omega_{i_1},\Omega_{i_4}\rangle\langle\Omega_{i_2},\Omega_{i_3}\rangle\langle\Omega_{i_2},\Omega_{i_4}\rangle \\
&= \sum_{k,k'} \frac{M_k}{M^2 M_{k'}} \sum_{i_3,i_4\in S_{k'}} N_{i_3}N_{i_4}\langle\Omega_{i_3},\Sigma_k\Omega_{i_4}\rangle^2 \le \sum_{k,k'} \frac{M_k}{M^2 M_{k'}} \sum_{i_3,i_4\in S_{k'}} N_{i_3}N_{i_4}\sum_{j,j'}\Omega_{i_3j}\Sigma_{kjj'}^2\Omega_{i_4j'} \\
&\le \sum_{k,k'} \frac{M_k M_{k'}}{M^2}\mu'\Sigma_k^{\circ 2}\mu \le \|\mu\|_4^4, \qquad \text{(B.53)}
\end{aligned}$$

where we applied Cauchy-Schwarz in the penultimate line and (B.44) in the last line.

For C_{723}, we have

$$\begin{aligned}
C_{723} &= \sum_{k\neq k'} \frac{M_k M_{k'}}{M^2} \sum_{j_1,j_2,j_3,j_4} \Sigma_{kj_1j_2}\Sigma_{j_3j_4}\Sigma_{kj_1j_3}\Sigma_{k'j_2j_4} \le \sum_k \frac{M_k}{M} \sum_{j_1,j_2,j_3,j_4} \Sigma_{kj_1j_2}\Sigma_{j_3j_4}\Sigma_{kj_1j_3}\Sigma_{j_2j_4} \\
&= \sum_k \frac{1}{M^3 M_k} \sum_{\substack{i_1,i_3\in S_k\\ i_2,i_4\in[n]}} N_{i_1}N_{i_2}N_{i_3}N_{i_4}\langle\Omega_{i_1},\Omega_{i_3}\rangle\langle\Omega_{i_1},\Omega_{i_4}\rangle\langle\Omega_{i_2},\Omega_{i_3}\rangle\langle\Omega_{i_2},\Omega_{i_4}\rangle \\
&= \sum_k \frac{1}{M^2} \sum_{i_3\in S_k, i_4\in[n]} N_{i_3}N_{i_4}\langle\Omega_{i_3},\Sigma_k\Omega_{i_4}\rangle\langle\Omega_{i_3},\Sigma\Omega_{i_4}\rangle \\
&\le \frac{1}{2}\sum_k \frac{1}{M^2} \sum_{i_3\in S_k, i_4\in[n]} N_{i_3}N_{i_4}\big(\langle\Omega_{i_3},\Sigma_k\Omega_{i_4}\rangle^2 + \langle\Omega_{i_3},\Sigma\Omega_{i_4}\rangle^2\big)
\end{aligned}$$

Using a similar technique as in (B.51)–(B.53) and applying (B.38), (B.39) we obtain

$$C_{723} \lesssim \|\mu\|_4^4.$$

Finally, for C_{724} we have

$$\begin{aligned} C_{724} &= \sum_{k\neq k'} \frac{M_k M_{k'}}{M^2} \sum_{j_1,j_2,j_3,j_4} \Sigma_{j_1j_2}\Sigma_{j_3j_4}\Sigma_{kj_1j_3}\Sigma_{k'j_2j_4} \leq \sum_{j_1,j_2,j_3,j_4} \Sigma_{j_1j_2}\Sigma_{j_3j_4}\Sigma_{j_1j_3}\Sigma_{j_2j_4} \\ &= \frac{1}{M^4} \sum_{i_1,i_2,i_3,i_4\in[n]} N_{i_1}N_{i_2}N_{i_3}N_{i_4}\langle\Omega_{i_1},\Omega_{i_3}\rangle\langle\Omega_{i_1},\Omega_{i_4}\rangle\langle\Omega_{i_2},\Omega_{i_3}\rangle\langle\Omega_{i_2},\Omega_{i_4}\rangle \end{aligned}$$

The details are very similar to (B.51)–(B.53), so we omit them and simply state the final bound:

$$C_{724} \lesssim \|\mu\|_4^4$$

Combining the bounds for $C_{721}, C_{722}, C_{723}$, and C_{724} yields

$$C_7 \lesssim K\|\mu\|_4^4.$$

Combining the bounds for C_1–C_7 proves the result.

□

B.4 Proof of Lemma B.3

We have

$$\begin{aligned} \mathbb{E}D_{\ell,s}^4 &= \mathbb{E}\Big[\big(\sum_{i\in[\ell-1]} \sigma_{i,\ell} \sum_{r=1}^{N_i}\sum_j Z_{ijr}Z_{\ell js}\big)^4\Big] \\ &= \sum_{i_1,i_2,i_3,i_4\in[\ell-1]} \sigma_{i_1\ell}\sigma_{i_2\ell}\sigma_{i_3\ell}\sigma_{i_4\ell} \sum_{\substack{r_1,r_2,r_3,r_4\\ j_1,j_2,j_3,j_4}} \mathbb{E}\big[Z_{i_1j_1r_1}Z_{\ell j_1s}Z_{i_2j_2r_2}Z_{\ell j_2s}Z_{i_3j_3r_3}Z_{\ell j_3s}Z_{i_4j_4r_4}Z_{\ell j_4s}\big] \\ &= \sum_{i_1,i_2,i_3,i_4\in[\ell-1]} \sigma_{i_1\ell}\sigma_{i_2\ell}\sigma_{i_3\ell}\sigma_{i_4\ell} \sum_{\substack{r_1,r_2,r_3,r_4\\ j_1,j_2,j_3,j_4}} \mathbb{E}\big[Z_{i_1j_1r_1}Z_{i_2j_2r_2}Z_{i_3j_3r_3}Z_{i_4j_4r_4}\big]\mathbb{E}\big[Z_{\ell j_1s}Z_{\ell j_2s}Z_{\ell j_3s}Z_{\ell j_4s}\big] \\ &= \sum_{j_1,j_2,j_3,j_4} \mathbb{E}[Z_{\ell j_1s}Z_{\ell j_2s}Z_{\ell j_3s}Z_{\ell j_4s}] \sum_{\substack{i_1,i_2,i_3,i_4\in[\ell-1]\\ r_1,r_2,r_3,r_4}} \sigma_{i_1\ell}\sigma_{i_2\ell}\sigma_{i_3\ell}\sigma_{i_4\ell}\mathbb{E}[Z_{i_1j_1r_1}Z_{i_2j_2r_2}Z_{i_3j_3r_3}Z_{i_4j_4r_4}] \\ &=: \sum_{j_1,j_2,j_3,j_4} \mathbb{E}[Z_{\ell j_1s}Z_{\ell j_2s}Z_{\ell j_3s}Z_{\ell j_4s}]A_{j_1,j_2,j_3,j_4} \end{aligned} \tag{B.54}$$

In the summations above, r_t ranges over $[N_{i_t}]$.

Observe that

$$|\mathbb{E}[Z_{\ell j_1s}Z_{\ell j_2s}Z_{\ell j_3s}Z_{\ell j_4s}]| \lesssim \begin{cases} \Omega_{\ell j_1} & \text{if } j_1=j_2=j_3=j_4 \\ \Omega_{\ell j_1}\Omega_{\ell j_4} & \text{if } j_1=j_2=j_3, j_4\neq j_1 \\ \Omega_{\ell j_1}\Omega_{\ell j_3} & \text{if } j_1=j_2, j_3=j_4, j_1\neq j_3 \\ \Omega_{\ell j_1}\Omega_{\ell j_3}\Omega_{\ell j_4} & \text{if } j_1=j_2, j_1,j_3,j_4 \; dist. \\ \Omega_{\ell j_1}\Omega_{\ell j_2}\Omega_{\ell j_3}\Omega_{\ell j_4} & \text{if } j_1,j_2,j_3,j_4 \; dist. \end{cases} \tag{B.55}$$

Up to permutation of the indices $j_1, \dots, j_4$, this accounts for all possible cases.

To proceed we also bound A_{j_1,j_2,j_3,j_4} by casework on the number of distinct j indices. For brevity we define $\omega_t = (i_t, r_t)$ and slightly abuse notation, letting $Z_{\omega_t,j} = Z_{i_t j r_t}$. Further let $\mathcal{I}_\ell = \{\omega = (i, r) : i \in [\ell], 1 \le r \le N_i\}$. Our goal is to control

$$A_{j_1,j_2,j_3,j_4} = \sum_{\omega_1,\omega_2,\omega_3,\omega_4 \in \mathcal{I}_{\ell-1}} \sigma_{i_1\ell}\sigma_{i_2\ell}\sigma_{i_3\ell}\sigma_{i_4\ell}\mathbb{E}[Z_{\omega_1 j_1} Z_{\omega_2 j_2} Z_{\omega_3 j_3} Z_{\omega_4 j_4}]. \tag{B.56}$$

To do this, we study (B.56) in five cases that cover all possibilities (up to permutation of the indices $j_1, \dots, j_4$).

Case 1: $j_1 = j_2 = j_3 = j_4$. Define $j = j_1$. It holds that

$$\begin{aligned}&\sigma_{i_1\ell}\sigma_{i_2\ell}\sigma_{i_3\ell}\sigma_{i_4\ell}\mathbb{E}[Z_{\omega_1 j} Z_{\omega_2 j} Z_{\omega_3 j} Z_{\omega_4 j}] \\ &\qquad = \begin{cases} \sigma_{i_1\ell}^4\, \mathbb{E}Z_{\omega_1 j}^4 \lesssim \sigma_{i_1\ell}^4\, \Omega_{i_1 j} & \text{if } \omega_1 = \omega_2 = \omega_3 = \omega_4 \\ \sigma_{i_1\ell}^2\sigma_{i_3\ell}^2 \mathbb{E}Z_{\omega_1 j}^2 \mathbb{E}Z_{\omega_3 j}^2 \lesssim \sigma_{i_1\ell}^2\sigma_{i_3\ell}^2 \Omega_{i_1 j}\Omega_{i_3 j} & \text{if } \omega_1 = \omega_2, \omega_3 = \omega_4, \omega_1 \neq \omega_3 \end{cases}\end{aligned} \tag{B.57}$$

Up to permutation of the indices $\omega_1, \dots, \omega_4$, this accounts for all cases such that (B.57) is nonvanishing. To be precise, by symmetry, it also holds that for all permutations $\pi : [4] \to [4]$ that if $\omega_{\pi(1)} = \omega_{\pi(2)}, \omega_{\pi(3)} = \omega_{\pi(4)}, \omega_{\pi(1)} \neq \omega_{\pi(3)}$, then

$$\sigma_{i_1\ell}\sigma_{i_2\ell}\sigma_{i_3\ell}\sigma_{i_4\ell}\mathbb{E}[Z_{\omega_1 j} Z_{\omega_2 j} Z_{\omega_3 j} Z_{\omega_4 j}] \lesssim \sigma_{i_{\pi(1)}\ell}^2 \sigma_{i_{\pi(3)}\ell}^2 \Omega_{i_{\pi(1)} j} \Omega_{i_{\pi(3)} j}.$$

In all other cases besides those considered above, we have

$$\sigma_{i_1\ell}\sigma_{i_2\ell}\sigma_{i_3\ell}\sigma_{i_4\ell}\mathbb{E}[Z_{\omega_1 j} Z_{\omega_2 j} Z_{\omega_3 j} Z_{\omega_4 j}] = 0$$

by independence.

Therefore,

$$A_{jjjj} \lesssim \sum_{\omega \in \mathcal{I}_{\ell-1}} \sigma_{i\ell}^4 \Omega_{ij} + \sum_{\omega_1 \neq \omega_3 \in \mathcal{I}_{\ell-1}} \sigma_{i_1\ell}^2 \sigma_{i_3\ell}^2 \Omega_{i_1 j}\Omega_{i_3 j} \tag{B.58}$$

In the remaining Cases 2–6, we follow the same strategy of writing out bounds for

$$\sigma_{i_1\ell}\sigma_{i_2\ell}\sigma_{i_3\ell}\sigma_{i_4\ell}\mathbb{E}[Z_{\omega_1 j_1} Z_{\omega_2 j_2} Z_{\omega_3 j_3} Z_{\omega_4 j_4}]$$

that cover all nonzero cases, up to permutation of the indices $\omega_1, \dots, \omega_4$.

Case 2: $j_1 = j_2 = j_3, j_1 \neq j_4$. It holds that

$$\begin{aligned}&\sigma_{i_1\ell}\sigma_{i_2\ell}\sigma_{i_3\ell}\sigma_{i_4\ell}\mathbb{E}[Z_{\omega_1 j_1} Z_{\omega_2 j_1} Z_{\omega_3 j_1} Z_{\omega_4 j_4}] \\ &\qquad = \begin{cases} \sigma_{i_1\ell}^4 \mathbb{E}[Z_{\omega_1 j_1}^3 Z_{\omega_1 j_4}] \lesssim \sigma_{i_1\ell}^4\, \Omega_{i_1 j_1}\Omega_{i_1 j_4} & \text{if } \omega_1 = \omega_2 = \omega_3 = \omega_4 \\ \sigma_{i_1\ell}^2\sigma_{i_3\ell}^2 \mathbb{E}Z_{\omega_1 j_1}^2 \mathbb{E}Z_{\omega_3 j_1} Z_{\omega_3 j_4} \lesssim \sigma_{i_1\ell}^2\sigma_{i_3\ell}^2\, \Omega_{i_1 j_1}\Omega_{i_3 j_1}\Omega_{i_3 j_4} & \text{if } \omega_1 = \omega_2, \omega_3 = \omega_4, \omega_1 \neq \omega_3 \end{cases}\end{aligned} \tag{B.59}$$

Up to permutation of the indices $\omega_1, \ldots, \omega_4$, this accounts for all cases such that (B.59) is nonvanishing. Thus

$$A_{j_1,j_1,j_1,j_4} \lesssim \sum_{\omega \in \mathcal{I}_{\ell-1}} \sigma_{i\ell}^4 \Omega_{i_1 j_1} \Omega_{i_1 j_4} + \sum_{\omega_1 \neq \omega_3 \in \mathcal{I}_{\ell-1}} \sigma_{i_1\ell}^2 \sigma_{i_3\ell}^2 \Omega_{i_1 j_1} \Omega_{i_3 j_1} \Omega_{i_3 j_4} \tag{B.60}$$

Case 3: $j_1 = j_2, j_3 = j_4, j_1 \neq j_3$. It holds that

$$\begin{aligned} &\sigma_{i_1\ell}\sigma_{i_2\ell}\sigma_{i_3\ell}\sigma_{i_4\ell}\mathbb{E}[Z_{\omega_1 j_1} Z_{\omega_2 j_1} Z_{\omega_3 j_3} Z_{\omega_4 j_3}] \\ &= \begin{cases} \sigma_{i_1\ell}^4 \mathbb{E} Z_{\omega_1 j_1}^2 Z_{\omega_1 j_3}^2 \lesssim \sigma_{i_1\ell}^4 \Omega_{i_1 j_1} \Omega_{i_1 j_3} & \text{if } \omega_1 = \omega_2 = \omega_3 = \omega_4 \\ \sigma_{i_1\ell}^2 \sigma_{i_3\ell}^2 \mathbb{E} Z_{\omega_1 j_1}^2 \mathbb{E} Z_{\omega_3 j_3}^2 \lesssim \sigma_{i_1\ell}^2 \sigma_{i_3\ell}^2 \Omega_{i_1 j_1} \Omega_{i_3 j_3} & \text{if } \omega_1 = \omega_2, \omega_3 = \omega_4, \omega_1 \neq \omega_3 \\ \sigma_{i_1\ell}^2 \sigma_{i_3\ell}^2 \mathbb{E} Z_{\omega_1 j_1} Z_{\omega_1 j_3} \mathbb{E} Z_{\omega_2 j_1} Z_{\omega_2 j_3} \lesssim \sigma_{i_1\ell}^2 \sigma_{i_3\ell}^2 \Omega_{i_1 j_1} \Omega_{i_1 j_3} \Omega_{i_2 j_1} \Omega_{i_2 j_3} & \text{if } \omega_1 = \omega_3, \omega_2 = \omega_4, \omega_1 \neq \omega_2. \end{cases} \end{aligned} \tag{B.61}$$

Up to permutation of the indices $\omega_1, \ldots, \omega_4$, this accounts for all cases such that (B.61) is nonvanishing. Thus by symmetry,

$$\begin{aligned} A_{j_1,j_1,j_3,j_3} \lesssim &\sum_{\omega \in \mathcal{I}_{\ell-1}} \sigma_{i_1\ell}^4 \Omega_{i_1 j_1} \Omega_{i_1 j_3} + \sum_{\omega_1 \neq \omega_3 \in \mathcal{I}_{\ell-1}} \sigma_{i_1\ell}^2 \sigma_{i_3\ell}^2 \Omega_{i_1 j_1} \Omega_{i_3 j_3} \\ &+ \sum_{\omega_1 \neq \omega_3 \in \mathcal{I}_{\ell-1}} \sigma_{i_1\ell}^2 \sigma_{i_3\ell}^2 \Omega_{i_1 j_1} \Omega_{i_1 j_3} \Omega_{i_3 j_1} \Omega_{i_3 j_3} \end{aligned} \tag{B.62}$$

Case 4: $j_1 = j_2$ and j_1, j_3, j_4 distinct. We have

$$\begin{aligned} &\sigma_{i_1\ell}\sigma_{i_2\ell}\sigma_{i_3\ell}\sigma_{i_4\ell}\mathbb{E}[Z_{\omega_1 j_1} Z_{\omega_2 j_1} Z_{\omega_3 j_3} Z_{\omega_4 j_4}] \\ &= \begin{cases} \sigma_{i_1\ell}^4 \mathbb{E} Z_{\omega_1 j_1}^2 Z_{\omega_1 j_3} Z_{\omega_1 j_4} \lesssim \sigma_{i_1\ell}^4 \Omega_{i_1 j_1} \Omega_{i_1 j_3} \Omega_{i_1 j_4} & \text{if } \omega_1 = \omega_2 = \omega_3 = \omega_4 \\ \sigma_{i_1\ell}^2 \sigma_{i_3\ell}^2 \mathbb{E} Z_{\omega_1 j_1}^2 \mathbb{E} Z_{\omega_3 j_3} Z_{\omega_3 j_4} \lesssim \sigma_{i_1\ell}^2 \sigma_{i_3\ell}^2 \Omega_{i_1 j_1} \Omega_{i_3 j_3} \Omega_{i_3 j_4} & \text{if } \omega_1 = \omega_2, \omega_3 = \omega_4, \omega_1 \neq \omega_3 \\ \sigma_{i_1\ell}^2 \sigma_{i_2\ell}^2 \mathbb{E} Z_{\omega_1 j_1} Z_{\omega_1 j_3} \mathbb{E} Z_{\omega_2 j_1} Z_{\omega_2 j_4} \lesssim \sigma_{i_1\ell}^2 \sigma_{i_2\ell}^2 \Omega_{i_1 j_1} \Omega_{i_1 j_3} \Omega_{i_2 j_1} \Omega_{i_2 j_4} & \text{if } \omega_1 = \omega_3, \omega_2 = \omega_4, \omega_1 \neq \omega_2 \end{cases} \end{aligned} \tag{B.63}$$

Up to permutation of the indices $\omega_1, \ldots, \omega_4$, this accounts for all cases such that (B.63) is nonvanishing. Thus

$$\begin{aligned} A_{j_1,j_1,j_3,j_4} \lesssim &\sum_{\omega \in \mathcal{I}_{\ell-1}} \sigma_{i_1\ell}^4 \Omega_{i_1 j_1} \Omega_{i_1 j_3} \Omega_{i_1 j_4} + \sum_{\omega_1 \neq \omega_3 \in \mathcal{I}_{\ell-1}} \sigma_{i_1\ell}^2 \sigma_{i_3\ell}^2 \Omega_{i_1 j_1} \Omega_{i_3 j_3} \Omega_{i_3 j_4} \\ &\sum_{\omega_1 \neq \omega_3 \in \mathcal{I}_{\ell-1}} \sigma_{i_1\ell}^2 \sigma_{i_2\ell}^2 \Omega_{i_1 j_1} \Omega_{i_1 j_3} \Omega_{i_3 j_1} \Omega_{i_3 j_4}. \end{aligned} \tag{B.64}$$

Case 5: j_1, j_2, j_3, j_4 distinct. For this final case, it holds that

$$\begin{aligned} &\sigma_{i_1\ell}\sigma_{i_2\ell}\sigma_{i_3\ell}\sigma_{i_4\ell}\mathbb{E}[Z_{\omega_1 j_1} Z_{\omega_2 j_2} Z_{\omega_3 j_3} Z_{\omega_4 j_4}] \\ &= \begin{cases} \sigma_{i_1\ell}^4 \mathbb{E} Z_{\omega_1 j_1} Z_{\omega_1 j_2} Z_{\omega_1 j_3} Z_{\omega_1 j_4} \lesssim \sigma_{i_1\ell}^4 \Omega_{i_1 j_1} \Omega_{i_1 j_2} \Omega_{i_1 j_3} \Omega_{i_1 j_4} & \text{if } \omega_1 = \omega_2 = \omega_3 = \omega_4 \\ \sigma_{i_1\ell}^2 \sigma_{i_3\ell}^2 \mathbb{E} Z_{\omega_1 j_1} Z_{\omega_1 j_2} \mathbb{E} Z_{\omega_3 j_3} Z_{\omega_3 j_4} \lesssim \sigma_{i_1\ell}^2 \sigma_{i_3\ell}^2 \Omega_{i_1 j_1} \Omega_{i_1 j_2} \Omega_{i_3 j_3} \Omega_{i_3 j_4} & \text{if } \omega_1 = \omega_2, \omega_3 = \omega_4, \omega_1 \neq \omega_3 \end{cases} \end{aligned}$$

The above accounts for all nonzero cases, up to permutation of $\omega_1, \omega_2, \omega_3, \omega_4$. Hence

$$A_{j_1,j_2,j_3,j_4} \lesssim \sum_{\omega \in \mathcal{I}_{\ell-1}} \sigma_{i_1\ell}^4 \, \Omega_{i_1j_1}\Omega_{i_1j_2}\Omega_{i_1j_3}\Omega_{i_1j_4} + \sum_{\omega_1 \neq \omega_3 \in \mathcal{I}_{\ell-1}} \sigma_{i_1\ell}^2 \sigma_{i_3\ell}^2 \, \Omega_{i_1j_1}\Omega_{i_1j_2}\Omega_{i_3j_3}\Omega_{i_3j_4}. \tag{B.65}$$

Finally we control the fourth moment using the casework above. By (B.54) and symmetry,

$$\begin{aligned}
\mathbb{E}D_{\ell,s}^4 \lesssim &\sum_j \mathbb{E}[Z_{\ell js}Z_{\ell js}Z_{\ell js}Z_{\ell js}]A_{j,j,j,j} + \sum_{j_1 \neq j_4} \mathbb{E}[Z_{\ell j_1 s}Z_{\ell j_1 s}Z_{\ell j_1 s}Z_{\ell j_4 s}]A_{j_1,j_1,j_1,j_4} \\
&+ \sum_{j_1 \neq j_3} \mathbb{E}[Z_{\ell j_1 s}Z_{\ell j_1 s}Z_{\ell j_3 s}Z_{\ell j_3 s}]A_{j_1,j_1,j_3,j_3} + \sum_{j_1,j_3,j_4 \, dist.} \mathbb{E}[Z_{\ell j_1 s}Z_{\ell j_1 s}Z_{\ell j_3 s}Z_{\ell j_4 s}]A_{j_1,j_1,j_3,j_4} \\
&+ \sum_{j_1,j_2,j_3,j_4 \, dist.} \mathbb{E}[Z_{\ell j_1 s}Z_{\ell j_2 s}Z_{\ell j_3 s}Z_{\ell j_4 s}]A_{j_1,j_2,j_3,j_4} \\
=: &\, F_{1\ell s} + F_{2\ell s} + F_{3\ell s} + F_{4\ell s} + F_{5\ell s}
\end{aligned} \tag{B.66}$$

By (B.55), (B.58), (B.60) ,(B.62), (B.64), and (B.65),

$$\begin{aligned}
F_{1\ell s} &\lesssim \sum_j \Omega_{\ell j}\Big(\sum_{\omega \in \mathcal{I}_{\ell-1}} \sigma_{i\ell}^4 \Omega_{ij} + \sum_{\omega_1 \neq \omega_3 \in \mathcal{I}_{\ell-1}} \sigma_{i_1\ell}^2 \sigma_{i_3\ell}^2 \Omega_{i_1j}\Omega_{i_3j}\Big) \\
F_{2\ell s} &\lesssim \sum_{j_1 \neq j_4} \Omega_{\ell j_1}\Omega_{\ell j_4}\Big(\sum_{\omega \in \mathcal{I}_{\ell-1}} \sigma_{i\ell}^4 \Omega_{i_1j_1}\Omega_{i_1j_4} + \sum_{\omega_1 \neq \omega_3 \in \mathcal{I}_{\ell-1}} \sigma_{i_1\ell}^2 \sigma_{i_3\ell}^2 \, \Omega_{i_1j_1}\Omega_{i_3j_1}\Omega_{i_3j_4}\Big) \\
F_{3\ell s} &\lesssim \sum_{j_1 \neq j_3} \Omega_{\ell j_1}\Omega_{\ell j_3}\Big(\sum_{\omega \in \mathcal{I}_{\ell-1}} \sigma_{i_1\ell}^4 \, \Omega_{i_1j_1}\Omega_{i_1j_3} + \sum_{\omega_1 \neq \omega_3 \in \mathcal{I}_{\ell-1}} \sigma_{i_1\ell}^2 \sigma_{i_3\ell}^2 \, \Omega_{i_1j_1}\Omega_{i_3j_3} \\
&\qquad + \sum_{\omega_1 \neq \omega_3 \in \mathcal{I}_{\ell-1}} \sigma_{i_1\ell}^2 \sigma_{i_3\ell}^2 \Omega_{i_1j_1}\Omega_{i_1j_3}\Omega_{i_3j_1}\Omega_{i_3j_3}\Big) \\
F_{4\ell s} &\lesssim \sum_{j_1,j_3,j_4 \, dist.} \Omega_{\ell j_1}\Omega_{\ell j_3}\Omega_{\ell j_4}\Big(\sum_{\omega \in \mathcal{I}_{\ell-1}} \sigma_{i_1\ell}^4 \, \Omega_{i_1j_1}\Omega_{i_1j_3}\Omega_{i_1j_4} + \sum_{\omega_1 \neq \omega_3 \in \mathcal{I}_{\ell-1}} \sigma_{i_1\ell}^2 \sigma_{i_3\ell}^2 \, \Omega_{i_1j_1}\Omega_{i_3j_3}\Omega_{i_3j_4} \\
&\qquad + \sum_{\omega_1 \neq \omega_3 \in \mathcal{I}_{\ell-1}} \sigma_{i_1\ell}^2 \sigma_{i_3\ell}^2 \, \Omega_{i_1j_1}\Omega_{i_1j_3}\Omega_{i_3j_1}\Omega_{i_3j_4}.\Big) \\
F_{5\ell s} &\lesssim \sum_{j_1,j_2,j_3,j_4 \, dist.} \Omega_{\ell j_1}\Omega_{\ell j_2}\Omega_{\ell j_3}\Omega_{\ell j_4}\Big(\sum_{\omega \in \mathcal{I}_{\ell-1}} \sigma_{i_1\ell}^4 \, \Omega_{i_1j_1}\Omega_{i_1j_2}\Omega_{i_1j_3}\Omega_{i_1j_4} \\
&\qquad + \sum_{\omega_1 \neq \omega_3 \in \mathcal{I}_{\ell-1}} \sigma_{i_1\ell}^2 \sigma_{i_3\ell}^2 \, \Omega_{i_1j_1}\Omega_{i_1j_2}\Omega_{i_3j_3}\Omega_{i_3j_4}\Big).
\end{aligned}$$

Define

$$\begin{aligned}
F_{11\ell s} &= \sum_{\omega \in \mathcal{I}_{\ell-1}} \sigma_{i\ell}^4 \sum_j \Omega_{\ell j}\Omega_{ij} \\
F_{21\ell s} &= \sum_{\omega \in \mathcal{I}_{\ell-1}} \sigma_{i\ell}^4 \sum_{j_1 \neq j_4} \Omega_{\ell j_1}\Omega_{\ell j_4}\Omega_{i_1j_1}\Omega_{i_1j_4}
\end{aligned}$$

$$
\begin{aligned}
F_{31\ell s} &= \sum_{\omega\in\mathcal{I}_{\ell-1}} \sigma_{i_1\ell}^4 \sum_{j_1\neq j_3} \Omega_{\ell j_1}\Omega_{\ell j_3}\Omega_{i_1j_1}\Omega_{i_1j_3} \\
F_{41\ell s} &= \sum_{\omega\in\mathcal{I}_{\ell-1}} \sigma_{i_1\ell}^4 \sum_{j_1,j_3,j_4\, dist.} \Omega_{\ell j_1}\Omega_{\ell j_3}\Omega_{\ell j_4}\Omega_{i_1j_1}\Omega_{i_1j_3}\Omega_{i_1j_4} \\
F_{51\ell s} &= \sum_{\omega\in\mathcal{I}_{\ell-1}} \sigma_{i_1\ell}^4 \sum_{j_1,j_2,j_3,j_4\, dist.} \Omega_{\ell j_1}\Omega_{\ell j_2}\Omega_{\ell j_3}\Omega_{\ell j_4}\Omega_{i_1j_1}\Omega_{i_1j_2}\Omega_{i_1j_3}\Omega_{i_1j_4}
\end{aligned}
$$

and

$$
\begin{aligned}
F_{12\ell s} &= \sum_{\omega_1\neq\omega_3\in\mathcal{I}_{\ell-1}} \sigma_{i_1\ell}^2\sigma_{i_3\ell}^2 \sum_{j} \Omega_{\ell j}\Omega_{i_1j}\Omega_{i_3j} \\
F_{22\ell s} &= \sum_{\omega_1\neq\omega_3\in\mathcal{I}_{\ell-1}} \sigma_{i_1\ell}^2\sigma_{i_3\ell}^2 \sum_{j_1\neq j_4} \Omega_{\ell j_1}\Omega_{\ell j_4}\Omega_{i_1j_1}\Omega_{i_3j_1}\Omega_{i_3j_4} \\
F_{32\ell s} &= \sum_{\omega_1\neq\omega_3\in\mathcal{I}_{\ell-1}} \sigma_{i_1\ell}^2\sigma_{i_3\ell}^2 \sum_{j_1\neq j_3} \big[\Omega_{\ell j_1}\Omega_{\ell j_3}\Omega_{i_1j_1}\Omega_{i_3j_3} + \Omega_{\ell j_1}\Omega_{\ell j_3}\Omega_{i_1j_1}\Omega_{i_1j_3}\Omega_{i_3j_1}\Omega_{i_3j_3}\big] \\
F_{42\ell s} &= \sum_{\omega_1\neq\omega_3\in\mathcal{I}_{\ell-1}} \sigma_{i_1\ell}^2\sigma_{i_3\ell}^2 \sum_{j_1,j_3,j_4\, dist.} \big[\Omega_{\ell j_1}\Omega_{\ell j_3}\Omega_{\ell j_4}\Omega_{i_1j_1}\Omega_{i_3j_3}\Omega_{i_3j_4} \\
&\qquad\qquad + \Omega_{\ell j_1}\Omega_{\ell j_3}\Omega_{\ell j_4}\Omega_{i_1j_1}\Omega_{i_1j_3}\Omega_{i_3j_1}\Omega_{i_3j_4}\big] \\
F_{52\ell s} &= \sum_{\omega_1\neq\omega_3\in\mathcal{I}_{\ell-1}} \sigma_{i_1\ell}^2\sigma_{i_3\ell}^2 \sum_{j_1,j_2,j_3,j_4\, dist.} \Omega_{\ell j_1}\Omega_{\ell j_2}\Omega_{\ell j_3}\Omega_{\ell j_4}\Omega_{i_1j_1}\Omega_{i_1j_2}\Omega_{i_3j_3}\Omega_{i_3j_4}
\end{aligned}
$$

Note that $\sum_{x=1}^2 F_{tx\ell s} = F_{t\ell s}$ for all $t\in[5]$. Using the fact that $\sum_j \Omega_{ij}=1$, we have

$$
\sum_t F_{t1\ell s} \lesssim F_{11\ell s} = \sum_{\omega\in\mathcal{I}_{\ell-1}} \sigma_{i\ell}^4 \sum_j \Omega_{\ell j}\Omega_{ij} = \sum_{\omega\in\mathcal{I}_{\ell-1}} \sigma_{i\ell}^4 \langle\Omega_\ell,\Omega_i\rangle. \tag{B.67}
$$

To control $\sum_t F_{t2\ell s}$, observe that, since $\Omega_{ij}\le 1$ for all i,j,

$$
\begin{aligned}
&\sum_j \Omega_{\ell j}\Omega_{i_1j} = \langle\Omega_\ell,\Omega_{i_1}\circ\Omega_{i_3}\rangle \\
&\sum_{j_1\neq j_4} \Omega_{\ell j_1}\Omega_{\ell j_4}\Omega_{i_1j_1}\Omega_{i_3j_1}\Omega_{i_3j_4} \le \langle\Omega_\ell,\Omega_{i_1}\circ\Omega_{i_3}\rangle\cdot\langle\Omega_\ell,\Omega_{i_3}\rangle \\
&\sum_{j_1\neq j_3} \big[\Omega_{\ell j_1}\Omega_{\ell j_3}\Omega_{i_1j_1}\Omega_{i_3j_3} + \Omega_{\ell j_1}\Omega_{\ell j_3}\Omega_{i_1j_1}\Omega_{i_1j_3}\Omega_{i_3j_1}\Omega_{i_3j_3}\big] \le 2\langle\Omega_\ell,\Omega_{i_1}\rangle\cdot\langle\Omega_\ell,\Omega_{i_3}\rangle \\
&\sum_{j_1,j_3,j_4\, dist.} \big[\Omega_{\ell j_1}\Omega_{\ell j_3}\Omega_{\ell j_4}\Omega_{i_1j_1}\Omega_{i_3j_3}\Omega_{i_3j_4} + \Omega_{\ell j_1}\Omega_{\ell j_3}\Omega_{\ell j_4}\Omega_{i_1j_1}\Omega_{i_1j_3}\Omega_{i_3j_1}\Omega_{i_3j_4}\big] \le 2\langle\Omega_\ell,\Omega_{i_1}\rangle\langle\Omega_\ell,\Omega_{i_3}\rangle^2 \\
&\sum_{j_1,j_2,j_3,j_4\, dist.} \Omega_{\ell j_1}\Omega_{\ell j_2}\Omega_{\ell j_3}\Omega_{\ell j_4}\Omega_{i_1j_1}\Omega_{i_1j_2}\Omega_{i_3j_3}\Omega_{i_3j_4} \le \langle\Omega_\ell,\Omega_{i_1}\rangle^2\langle\Omega_\ell,\Omega_{i_3}\rangle^2.
\end{aligned}
$$

These bounds are relatively sharp, and it is clear that the first and third lines dominate. Furthermore as. Hence,

$$
\sum_t F_{t2\ell s} \lesssim F_{12\ell s} + F_{32\ell s} \lesssim \sum_{\omega_1\neq\omega_3\in\mathcal{I}_{\ell-1}} \sigma_{i_1\ell}^2\sigma_{i_3\ell}^2 \big[\langle\Omega_\ell,\Omega_{i_1}\circ\Omega_{i_3}\rangle + \langle\Omega_\ell,\Omega_{i_1}\rangle\cdot\langle\Omega_\ell,\Omega_{i_3}\rangle\big]. \tag{B.68}
$$

Observe that if $\ell \in S_k$, then

$$\sum_{\omega} \sigma_{i\ell}^4 \Omega_{ij} \le \sum_{i \in S_k} \frac{1}{n_k^4 \bar{N}_k^4} N_i \Omega_{ij} + \sum_{k'=1}^{K} \sum_{i \in S_{k'}} \frac{1}{n^4 \bar{N}^4} N_i \Omega_{ij} \tag{B.69}$$

$$\le \frac{1}{n_k^3 \bar{N}_k^3} \mu_{kj} + \frac{1}{n^3 \bar{N}^3} \mu_j, \tag{B.70}$$

and

$$\sum_{\omega} \sigma_{i\ell}^2 \Omega_{ij} \le \sum_{i \in S_k} \frac{1}{n_k^2 \bar{N}_k^2} N_i \Omega_{ij} + \sum_{k'=1}^{K} \sum_{i \in S_{k'}} \frac{1}{n\bar{N}} N_i \Omega_{ij}$$
$$\le \frac{1}{n_k \bar{N}_k} \mu_{kj} + \frac{1}{n\bar{N}} \mu_j.$$

Next,

$$\begin{aligned}
\sum_{(\ell,s)} \sum_t F_{t1\ell s} &\lesssim \sum_{(\ell,s)} \sum_{\omega \in \mathcal{I}_{\ell-1}} \sigma_{i\ell}^4 \langle \Omega_\ell, \Omega_i \rangle. \\
&\lesssim \sum_{(\ell,s)} \sum_j \Omega_{\ell j} \big(\frac{1}{n_k^3 \bar{N}_k^3} \mu_{kj} + \frac{1}{n^3 \bar{N}^3} \mu_j \big) \\
&\lesssim \sum_j \sum_k \frac{1}{n_k^2 \bar{N}_k^2} \mu_{kj}^2 + \sum_j \sum_k \frac{1}{n^2 \bar{N}^2} \mu_j^2 \lesssim \sum_k \frac{1}{n_k^2 \bar{N}_k^2} \|\mu_k\|^2,
\end{aligned} \tag{B.71}$$

where we applied that $\|\mu\|^2 \lesssim \sum_k \|\mu_k\|^2$ (see (A.49)). Furthermore,

$$\begin{aligned}
&\sum_{(\ell,s)} \sum_t F_{t2\ell s} \le \sum_{k=1}^{K} \sum_{\ell \in S_k} N_\ell \sum_{\omega_1, \omega_3} \sigma_{i_1 \ell}^2 \sigma_{i_3 \ell}^2 \big[\langle \Omega_\ell, \Omega_{i_1} \circ \Omega_{i_3} \rangle + \langle \Omega_\ell, \Omega_{i_1} \rangle \cdot \langle \Omega_\ell, \Omega_{i_3} \rangle \big] \\
&\lesssim \sum_k \sum_{\ell \in S_k} N_\ell \Big[\sum_j \Omega_{\ell_j} \big(\frac{1}{n_k \bar{N}_k} \mu_{kj} + \frac{1}{n\bar{N}} \mu_j \big)^2 + \Big(\sum_j \Omega_{\ell j} \cdot \big(\frac{1}{n_k \bar{N}_k} \mu_{kj} + \frac{1}{n\bar{N}} \mu_j \big) \Big)^2 \Big] \\
&\lesssim \sum_k \sum_{\ell \in S_k} N_\ell \sum_j \Omega_{\ell_j} \big(\frac{1}{n_k \bar{N}_k} \mu_{kj} + \frac{1}{n\bar{N}} \mu_j \big)^2
\end{aligned}$$

In the last line we apply Cauchy–Schwarz. Continuing, we have

$$\begin{aligned}
\sum_{(\ell,s)} \sum_t F_{t2\ell s} &\lesssim \sum_k \sum_{\ell \in S_k} N_\ell \sum_j \Omega_{\ell_j} \big(\frac{1}{n_k \bar{N}_k} \mu_{kj} + \frac{1}{n\bar{N}} \mu_j \big)^2 \\
&\lesssim \sum_k \sum_{\ell \in S_k} N_\ell \sum_j \Omega_{\ell_j} \big(\frac{1}{n_k \bar{N}_k} \mu_{kj} \big)^2 + \sum_k \sum_{\ell \in S_k} N_\ell \sum_j \Omega_{\ell_j} \big(\frac{1}{n\bar{N}} \mu_j \big)^2 \\
&\lesssim \sum_k \frac{\|\mu_k\|_3^3}{n_k \bar{N}_k} + \sum_k \frac{\|\mu\|_3^3}{n\bar{N}} \lesssim \sum_k \frac{\|\mu_k\|_3^3}{n_k \bar{N}_k},
\end{aligned} \tag{B.72}$$

where we applied (A.68). Combining (B.66), (B.71) and (B.72), we have

$$\sum_{(\ell,s)} \mathbb{E} D_{\ell,s}^4 \lesssim \sum_{(\ell,s)} \sum_{x=1}^{2} \sum_{t=1}^{5} F_{tx\ell s} \lesssim \sum_k \frac{\|\mu_k\|^2}{n_k^2 \bar{N}_k^2} + \sum_k \frac{\|\mu_k\|_3^3}{n_k \bar{N}_k},$$

as desired.

B.5 Proof of Lemma B.4

$$\mathrm{Var}\Big[\sum_{(\ell,s)}\mathrm{Var}(\tilde{E}_{\ell,s}|\mathcal{F}_{\prec(\ell,s)})\Big]\to 0 \tag{B.73}$$

Next we study (B.73). We have

$$\begin{aligned}\mathrm{Var}(E_{\ell,s}|\mathcal{F}_{\prec(\ell,s)}) = \mathbb{E}[E_{\ell,s}^2|\mathcal{F}_{\prec(\ell,s)}] &= \sigma_\ell^2 \sum_{r,r'\in[s-1]}\sum_{j,j'}\mathbb{E}\big[Z_{\ell jr}Z_{\ell js}Z_{\ell j'r'}Z_{\ell j's}\big|\mathcal{F}_{\prec(\ell,s)}\big]\\ &= \sigma_\ell^2 \sum_{r,r'\in[s-1]}\sum_{j,j'} Z_{\ell jr}Z_{\ell j'r'}\mathbb{E}[Z_{\ell js}Z_{\ell j's}]\\ &= \sigma_\ell^2 \sum_{r,r'\in[s-1]}\sum_{j,j'}\delta_{jj'\ell}Z_{\ell jr}Z_{\ell j'r'}, \end{aligned}\tag{B.74}$$

where we let

$$\delta_{jj'\ell} = \mathbb{E}Z_{\ell js}Z_{\ell j's} = \begin{cases}\Omega_{\ell j}(1-\Omega_{\ell j}) & \text{if } j=j'\\ -\Omega_{\ell j}\Omega_{\ell j'} & \text{else.}\end{cases}\tag{B.75}$$

Define

$$\varphi_{\ell r\ell r'} = \sum_{j,j'}\delta_{jj'\ell}Z_{\ell jr}Z_{\ell j'r'}.\tag{B.76}$$

By (B.74) we have

$$\begin{aligned}\sum_{(\ell,s)}\mathrm{Var}(E_{\ell,s}|\mathcal{F}_{\prec(\ell,s)}) &= \sum_{\ell=1}^{n}\sum_{s=1}^{N_\ell}\sum_{r,r'\in[s-1]}\sigma_\ell^2\,\varphi_{\ell r\ell r'}\\ &= \sum_{\ell=1}^{n}\sum_{s=1}^{N_\ell}\Big[\sum_{r\in[s-1]}\sigma_\ell^2\,\varphi_{\ell r\ell r} + 2\sum_{r<r'\in[s-1]}\sigma_\ell^2\,\varphi_{\ell r\ell r'}\Big]\\ &= \sum_{\ell=1}^{n}\sum_{r=1}^{N_\ell}\sum_{s\in[N_\ell]:s>r}\sigma_\ell^2\,\varphi_{\ell r\ell r} + 2\sum_{\ell=1}^{s}\sum_{r<r'\in[N_\ell]}\sum_{s\in[N_\ell]:s>r'}\sigma_\ell^2\,\varphi_{\ell r\ell r'}\\ &= \sum_{\ell=1}^{n}\sum_{r=1}^{N_\ell}(N_\ell-r)\sigma_\ell^2\,\varphi_{\ell r\ell r} + 2\sum_{\ell=1}^{s}\sum_{r<r'\in[N_\ell]}(N_\ell-r')\sigma_\ell^2\,\varphi_{\ell r\ell r'}\\ &\equiv S_1+S_2.\end{aligned}$$

Observe that S_1 and S_2 are uncorrelated. In addition, the terms in the summation defining S_1 are uncorrelated; the same holds for S_2 also.

First we study S_2. Next,

$$\begin{aligned}\mathbb{E}\varphi_{\ell r\ell r'}^2 &= \sum_{j_1,j_2,j_3,j_4}\delta_{j_1j_2,\ell}\delta_{j_3j_4,\ell}\,\mathbb{E}Z_{\ell j_1r}Z_{\ell j_2r'}Z_{\ell j_3r}Z_{\ell j_4r'}\\ &= \sum_{j_1,j_2,j_3,j_4}\delta_{j_1j_2\ell}\delta_{j_3j_4\ell}\mathbb{E}Z_{\ell j_1r}Z_{\ell j_3r}\mathbb{E}Z_{\ell j_2r'}Z_{\ell j_4r'}.\end{aligned}\tag{B.77}$$

First we study V_2. By casework,

$$
\begin{aligned}
&|\delta_{j_1j_2\ell}\delta_{j_3j_4\ell}\mathbb{E}Z_{\ell j_1 r}Z_{\ell j_3 r}\mathbb{E}Z_{\ell j_2 r'}Z_{\ell j_4 r'}| \qquad\qquad (B.78)\\
&= \begin{cases}
\delta_{jj\ell}^2\mathbb{E}Z_{\ell jr}^2\mathbb{E}Z_{\ell jr'}^2 \lesssim \Omega_{\ell j}^4 & \text{if } j_1=\cdots=j_4\\
\delta_{j_1j_1\ell}\delta_{j_1j_4\ell}|\mathbb{E}Z_{\ell j_1 r}^2\mathbb{E}Z_{\ell j_1 r'}Z_{\ell j_4 r'}| \lesssim \Omega_{\ell j_1}^4\Omega_{\ell j_4}^2 & \text{if } j_1=j_2=j_3, j_1\neq j_4\\
\delta_{j_1j_1\ell}\delta_{j_3j_3\ell}\mathbb{E}Z_{\ell j_1 r}Z_{\ell j_3 r}\mathbb{E}Z_{\ell j_1 r'}Z_{\ell j_3 r'} \lesssim \Omega_{\ell j_1}^3\Omega_{\ell j_3}^3 & \text{if } j_1=j_2, j_3=j_4, j_1\neq j_3\\
\delta_{j_1j_2\ell}^2\mathbb{E}Z_{\ell j_1 r}^2\mathbb{E}Z_{\ell j_2 r'}^2 \lesssim \Omega_{\ell j_1}^3\Omega_{\ell j_2}^3 & \text{if } j_1=j_3, j_2=j_4, j_1\neq j_2\\
\delta_{j_1j_1\ell}\delta_{j_3j_4\ell}\mathbb{E}Z_{\ell j_1 r}Z_{\ell j_3 r}\mathbb{E}Z_{\ell j_1 r'}Z_{\ell j_4 r'} \lesssim \Omega_{\ell j_1}^3\Omega_{\ell j_3}^2\Omega_{\ell j_4}^2 & \text{if } j_1=j_2, j_1,j_3,j_4 \textit{ dist.}\\
\delta_{j_1j_2\ell}\delta_{j_1j_4\ell}\mathbb{E}Z_{\ell j_1 r}^2\mathbb{E}Z_{\ell j_2 r'}Z_{\ell j_4 r'} \lesssim \Omega_{\ell j_1}^3\Omega_{\ell j_2}^2\Omega_{\ell j_4}^2 & \text{if } j_1=j_3, j_1,j_2,j_4 \textit{ dist.}\\
\delta_{j_1j_2\ell}\delta_{j_3j_4\ell}\mathbb{E}Z_{\ell j_1 r}Z_{\ell j_3 r}\mathbb{E}Z_{\ell j_2 r'}Z_{\ell j_4 r'} \lesssim \Omega_{\ell j_1}^2\Omega_{\ell j_2}^2\Omega_{\ell j_3}^2\Omega_{\ell j_4}^2 & \text{if } j_1,j_2,j_3,j_4 \textit{ dist.}
\end{cases}
\end{aligned}
$$

Up to permutation of the indices $j_1,\dots,j_4$, all nonzero terms of (B.77) take one of the forms above. By (B.78) and Cauchy–Schwarz, we have

$$
\mathbb{E}\varphi_{\ell r\ell r'}^2 \lesssim \|\Omega_\ell\|_4^4 + \|\Omega_\ell\|_4^4\|\Omega_\ell\|^2 + 2\|\Omega_\ell\|_3^6 + 2\|\Omega_\ell\|_3^3\|\Omega_\ell\|^4 + \|\Omega_\ell\|^8 \lesssim \|\Omega_\ell\|_4^4. \tag{B.79}
$$

Recalling that $\{\varphi_{\ell r\ell r'}\}_{\ell, r<r'\in[N_\ell]}$ are mutually uncorrelated, it follows that

$$
\begin{aligned}
\mathrm{Var}(S_2) &\lesssim \sum_\ell \sum_{r<r'\in[N_\ell]} (N_\ell - r')^2\sigma_\ell^2\mathbb{E}\varphi_{\ell r\ell r'}^2\\
&\lesssim \sum_\ell \sum_{r<r'\in[N_\ell]} (N_\ell - r')^2\sigma_\ell^4\|\Omega_\ell\|_4^4\\
&\lesssim \sum_k \sum_{\ell\in S_k} N_\ell^4 \cdot \frac{1}{n_k^4\bar{N}_k^4}\|\Omega_\ell\|_4^4.
\end{aligned} \tag{B.80}
$$

Next we study S_1. We have

$$
\mathbb{E}\varphi_{\ell r\ell r}^2 = \sum_{j_1,j_2,j_3,j_4} \delta_{j_1j_2\ell}\delta_{j_3j_4\ell}\mathbb{E}Z_{\ell j_1 r}Z_{\ell j_2 r}Z_{\ell j_3 r}Z_{\ell j_4 r}.
$$

We have the following bounds by casework.

$$
\begin{aligned}
&|\delta_{j_1j_2\ell}\delta_{j_3j_4\ell}\mathbb{E}Z_{\ell j_1 r}Z_{\ell j_2 r}Z_{\ell j_3 r}Z_{\ell j_4 r}| \qquad\qquad (B.81)\\
&= \begin{cases}
\delta_{jj\ell}^2\mathbb{E}Z_{\ell jr}^4 \lesssim \Omega_{\ell j}^3 & \text{if } j_1=\cdots=j_4\\
\delta_{j_1j_1\ell}\delta_{j_1j_4\ell}|\mathbb{E}Z_{\ell j_1 r}^3 Z_{\ell j_4 r}| \lesssim \Omega_{\ell j_1}^3\Omega_{\ell j_4}^2 & \text{if } j_1=j_2=j_3, j_1\neq j_4\\
\delta_{j_1j_1\ell}\delta_{j_3j_3\ell}\mathbb{E}Z_{\ell j_1 r}^2 Z_{\ell j_3 r}^2 \lesssim \Omega_{\ell j_1}^2\Omega_{\ell j_3}^2 & \text{if } j_1=j_2, j_3=j_4, j_1\neq j_3\\
\delta_{j_1j_2\ell}^2\mathbb{E}Z_{\ell j_1 r}^2 Z_{\ell j_2 r}^2 \lesssim \Omega_{\ell j_1}^3\Omega_{\ell j_2}^3 & \text{if } j_1=j_3, j_2=j_4, j_1\neq j_3\\
\delta_{j_1j_1\ell}\delta_{j_3j_4\ell}|\mathbb{E}Z_{\ell j_1 r}^2 Z_{\ell j_3 r}Z_{\ell j_4 r}| \lesssim \Omega_{\ell j_1}^2\Omega_{\ell j_3}^2\Omega_{\ell j_4}^2 & \text{if } j_1=j_2, j_1,j_3,j_4 \textit{ dist.}\\
\delta_{j_1j_2\ell}\delta_{j_1j_4\ell}|\mathbb{E}Z_{\ell j_1 r}^2 Z_{\ell j_2 r}Z_{\ell j_4}| \lesssim \Omega_{\ell j_1}^3\Omega_{\ell j_2}^2\Omega_{\ell j_4}^2 & \text{if } j_1=j_3, j_1,j_2,j_4 \textit{ dist.}\\
\delta_{j_1j_2\ell}\delta_{j_3j_4\ell}|\mathbb{E}Z_{\ell j_1 r}Z_{\ell j_2 r}Z_{\ell j_3 r}Z_{\ell j_4 r}| \lesssim \Omega_{\ell j_1}^2\Omega_{\ell j_2}^2\Omega_{\ell j_3}^2\Omega_{\ell j_4}^2 & \text{if } j_1,j_2,j_3,j_4 \textit{ dist.}
\end{cases}
\end{aligned}
$$

Up to symmetry, this accounts for all possible (nonzero) cases. Hence by Cauchy–Schwarz,

$$
\mathbb{E}\varphi_{\ell r\ell r}^2 \lesssim \|\Omega_\ell\|_3^3 + \|\Omega_\ell\|_3^3\|\Omega_\ell\|^2 + \|\Omega_\ell\|^4 + \|\Omega_\ell\|_3^6 + \|\Omega_\ell\|^6 + \|\Omega_\ell\|_3^3\|\Omega_\ell\|^4 + \|\Omega_\ell\|^8 \lesssim \|\Omega_\ell\|_3^3. \tag{B.82}
$$

Recalling that $\{\varphi_{\ell r\ell r}\}_{\ell,r\in[N_\ell]}$ is an uncorrelated collection of random variables, we have

$$\begin{aligned}\mathrm{Var}(S_1) &\lesssim \sum_{\ell}\sum_{r\in[N_\ell]}(N_\ell-r)^2\sigma_\ell^4\mathbb{E}\varphi_{\ell r\ell r}^2\\ &\lesssim \sum_{\ell}\sum_{r\in[N_\ell]}(N_\ell-r)^2\sigma_\ell^4\|\Omega_\ell\|_3^3\\ &\lesssim \sum_{k}\sum_{\ell\in S_k}N_\ell^3\cdot\frac{1}{n_k^4\bar{N}_k^4}\|\Omega_\ell\|_3^3. \end{aligned}\tag{B.83}$$

Combining (B.83) and (B.80) proves the result. □

B.6 Proof of Lemma B.5

We have

$$\begin{aligned}\mathbb{E}E_{\ell,s}^4 &= \sum_{r_1,r_2,r_3,r_4\in[s-1]}\sigma_\ell^4\sum_{j_1,j_2,j_3,j_4}\mathbb{E}Z_{\ell j_1 r_1}Z_{\ell j_1 s}Z_{\ell j_2 r_2}Z_{\ell j_2 s}Z_{\ell j_3 r_3}Z_{\ell j_3 s}Z_{\ell j_4 r_4}Z_{\ell j_4 s}\\ &= \sigma_\ell^4\sum_{j_1,j_2,j_3,j_4}\Bigg[\mathbb{E}[Z_{\ell j_1 s}Z_{\ell j_2 s}Z_{\ell j_3 s}Z_{\ell j_4 s}]\cdot\underbrace{\sum_{r_1,r_2,r_3,r_4\in[s-1]}\mathbb{E}[Z_{\ell j_1 r_1}Z_{\ell j_2 r_2}Z_{\ell j_3 r_3}Z_{\ell j_4 r_4}]}_{=:B_{\ell,s;j_1,j_2,j_3,j_4}}\Bigg]\end{aligned}\tag{B.84}$$

We have by exhaustive casework that

$$|\mathbb{E}[Z_{\ell j_1 r_1}Z_{\ell j_2 r_2}Z_{\ell j_3 r_3}Z_{\ell j_4 r_4}]|\tag{B.85}$$

$$= \begin{cases}\mathbb{E}Z_{\ell j_1 r_1}^4 \lesssim \Omega_{\ell j_1} & \text{if } {\scriptstyle j_1=j_2=j_3=j_4;\ r_1=r_2=r_3=r_4}\\ \mathbb{E}Z_{\ell j_1 r_1}^2\mathbb{E}Z_{\ell j_1 r_3}^2 \lesssim \Omega_{\ell j_1}^2 & \text{if } {\scriptstyle j_1=j_2=j_3=j_4;\ r_1=r_2,r_3=r_4,r_1\neq r_3}\\ |\mathbb{E}[Z_{\ell j_1 r_1}^3 Z_{\ell j_4 r_1}]| \lesssim \Omega_{\ell j_1}\Omega_{\ell j_4} & \text{if } {\scriptstyle j_1=j_2=j_3,j_1\neq j_4;\ r_1=r_2=r_3=r_4}\\ |\mathbb{E}[Z_{\ell j_1 r_1}^2\mathbb{E}Z_{\ell j_1 r_3}Z_{\ell j_4 r_3}]| \lesssim \Omega_{\ell j_1}^2\Omega_{\ell j_4} & \text{if } {\scriptstyle j_1=j_2=j_3,j_1\neq j_4;\ r_1=r_2,r_3=r_4,r_1\neq r_3}\\ |\mathbb{E}Z_{\ell j_1 r_1}^2 Z_{\ell j_3 r_1}^2| \lesssim \Omega_{\ell j_1}\Omega_{\ell j_3} & \text{if } {\scriptstyle j_1=j_2,j_3=j_4,j_1\neq j_3;\ r_1=r_2=r_3=r_4}\\ |\mathbb{E}[Z_{\ell j_1 r_1}^2 Z_{\ell j_3 r_3}^2]| \lesssim \Omega_{\ell j_1}\Omega_{\ell j_3} & \text{if } {\scriptstyle j_1=j_2,j_3=j_4,j_1\neq j_3;\ r_1=r_2,r_3=r_4,r_1\neq r_3}\\ |\mathbb{E}[Z_{\ell j_1 r_1}Z_{\ell j_3 r_1}\mathbb{E}Z_{\ell j_1 r_2}Z_{\ell j_3 r_2}]| \lesssim \Omega_{\ell j_1}^2\Omega_{\ell j_3}^2 & \text{if } {\scriptstyle j_1=j_2,j_3=j_4,j_1\neq j_3;\ r_1=r_3,r_2=r_4,r_1\neq r_2}\\ |\mathbb{E}[Z_{\ell j_1 r_1}^2 Z_{\ell j_3 r_1}Z_{\ell j_4 r_1}]| \lesssim \Omega_{\ell j_1}\Omega_{\ell j_3}\Omega_{\ell j_4} & \text{if } {\scriptstyle j_1=j_2,j_1,j_3,j_4\ dist.;\ r_1=r_2=r_3=r_4}\\ |\mathbb{E}[Z_{\ell j_1 r_1}^2\mathbb{E}Z_{\ell j_3 r_3}Z_{\ell j_4 r_3}]| \lesssim \Omega_{\ell j_1}\Omega_{\ell j_3}\Omega_{\ell j_4} & \text{if } {\scriptstyle j_1=j_2,j_1,j_3,j_4\ dist.;\ r_1=r_2,r_3=r_4,r_1\neq r_3}\\ |\mathbb{E}[Z_{\ell j_1 r_1}Z_{\ell j_3 r_1}\mathbb{E}Z_{\ell j_1 r_2}Z_{\ell j_4 r_2}]| \lesssim \Omega_{\ell j_1}^2\Omega_{\ell j_3}\Omega_{\ell j_4} & \text{if } {\scriptstyle j_1=j_2,j_1,j_3,j_4\ dist.;\ r_1=r_3,r_2=r_4,r_1\neq r_2}\\ |\mathbb{E}[Z_{\ell j_1 r_1}Z_{\ell j_2 r_1}Z_{\ell j_3 r_1}Z_{\ell j_4 r_1}]| \lesssim \Omega_{\ell j_1}\Omega_{\ell j_2}\Omega_{\ell j_3}\Omega_{\ell j_4} & \text{if } {\scriptstyle j_1,j_2,j_3,j_4\ dist;\ r_1=r_2=r_3=r_4}\\ |\mathbb{E}[Z_{\ell j_1 r_1}Z_{\ell j_2 r_1}\mathbb{E}Z_{\ell j_3 r_3}Z_{\ell j_4 r_3}]| \lesssim \Omega_{\ell j_1}\Omega_{\ell j_2}\Omega_{\ell j_3}\Omega_{\ell j_4} & \text{if } {\scriptstyle j_1,j_2,j_3,j_4\ dist;\ r_1=r_2,r_3=r_4,r_1\neq r_3}\end{cases}$$

Up to permutation of the indices j_1, j_2, j_3, j_4 and r_1, r_2, r_3, r_4, this accounts for all possible cases such that (B.85) is nonzero. Therefore,

$$B_{\ell,s;j_1,j_2,j_3,j_4} \lesssim \begin{cases} s\Omega_{\ell j_1} + s^2\Omega_{\ell j_1}^2 & \text{if } j_1 = j_2 = j_3 = j_4 \\ s\Omega_{\ell j_1}\Omega_{\ell j_4} + s^2\Omega_{\ell j_1}^2\Omega_{\ell j_4} & \text{if } j_1 = j_2 = j_3, j_1 \neq j_4 \\ s\Omega_{\ell j_1}\Omega_{\ell j_3} + s^2\Omega_{\ell j_1}\Omega_{\ell j_3} & \text{if } j_1 = j_2, j_3 = j_4, j_1 \neq j_3 \\ s\Omega_{\ell j_1}\Omega_{\ell j_3}\Omega_{\ell j_4} + s^2\Omega_{\ell j_1}\Omega_{\ell j_3}\Omega_{\ell j_4} & \text{if } j_1 = j_2, j_1, j_3, j_4 \; dist. \\ s\Omega_{\ell j_1}\Omega_{\ell j_2}\Omega_{\ell j_3}\Omega_{\ell j_4} + s^2\Omega_{\ell j_1}\Omega_{\ell j_2}\Omega_{\ell j_3}\Omega_{\ell j_4} & \text{if } j_1, j_2, j_3, j_4 \; dist. \end{cases}$$

Up to permutation of j_1, j_2, j_3, j_4, this accounts for all possible cases. Returning to (B.84), we have by applying (B.55) and the previous display that

$$\begin{aligned} \mathbb{E}E_{\ell,s}^4 \lesssim \sigma_\ell^4 \Bigg(& \sum_j \Omega_{\ell j}(s\Omega_{\ell j} + s^2\Omega_{\ell j}^2) + \sum_{j_1 \neq j_4} \Omega_{\ell j_1}\Omega_{\ell j_4}(s\Omega_{\ell j_1}\Omega_{\ell j_4} + s^2\Omega_{\ell j_1}^2\Omega_{\ell j_4}) \\ & + \sum_{j_1 \neq j_3} \Omega_{\ell j_1}\Omega_{\ell j_3}(s\Omega_{\ell j_1}\Omega_{\ell j_3} + s^2\Omega_{\ell j_1}\Omega_{\ell j_3}) \\ & + \sum_{j_1, j_3, j_4 (dist.)} \Omega_{\ell j_1}\Omega_{\ell j_3}\Omega_{\ell j_4}(s\Omega_{\ell j_1}\Omega_{\ell j_3}\Omega_{\ell j_4} + s^2\Omega_{\ell j_1}\Omega_{\ell j_3}\Omega_{\ell j_4}) \\ & + \sum_{j_1, j_2, j_3, j_4 \; dist.} \Omega_{\ell j_1}\Omega_{\ell j_2}\Omega_{\ell j_3}\Omega_{\ell j_4}(s\Omega_{\ell j_1}\Omega_{\ell j_2}\Omega_{\ell j_3}\Omega_{\ell j_4} + s^2\Omega_{\ell j_1}\Omega_{\ell j_2}\Omega_{\ell j_3}\Omega_{\ell j_4}) \Bigg) \\ \lesssim & \; s\sigma_\ell^4\|\Omega_\ell\|^2 + s^2\sigma_\ell^4\|\Omega_\ell\|_3^3. \end{aligned}$$

In the third line we group the coefficients of s and s^2 and use the fact that $\|\Omega_\ell\|^4 \leq \|\Omega_\ell\|_3^3$ by Cauchy–Schwarz. Therefore

$$\begin{aligned} \sum_{(\ell,s)} \mathbb{E}E_{\ell,s}^4 &\lesssim \sum_{(\ell,s)} s\sigma_\ell^4\|\Omega_\ell\|^2 + \sum_{(\ell,s)} s^2\sigma_\ell^4\|\Omega_\ell\|_3^3 \\ &= \sum_k \sum_{\ell \in S_k} \sum_{s \in [N_\ell]} s\sigma_\ell^4\|\Omega_\ell\|^2 + \sum_k \sum_{\ell \in S_k} \sum_{s \in [N_\ell]} s^2\sigma_\ell^4\|\Omega_\ell\|_3^3 \\ &\lesssim \sum_k \sum_{\ell \in S_k} N_\ell^2 \cdot \frac{1}{n_k^4\bar{N}_k^4}\|\Omega_\ell\|^2 + \sum_k \sum_{\ell \in S_k} N_\ell^3 \cdot \frac{1}{n_k^4\bar{N}_k^4}\|\Omega_\ell\|_3^3, \end{aligned}$$

as desired. □

B.7 Proof of Lemma B.6

We have

$$\begin{aligned} \sum_k \sum_{i \in S_k} \frac{N_i^2\|\Omega_i\|^2}{n_k^4\bar{N}_k^4} &\leq \sum_k \frac{1}{n_k^4\bar{N}_k^4} \sum_{i,m \in S_k} N_i N_m \langle \Omega_i, \Omega_m \rangle \\ &= \sum_k \frac{1}{n_k^2\bar{N}_k^2}\|\mu_k\|^2, \end{aligned}$$

which establishes the first claim.

Similarly,

$$\sum_k \sum_{i\in S_k} \frac{N_i^3\|\Omega_i\|_3^3}{n_k^4\bar{N}_k^4} \le \sum_k \frac{1}{n_k^4\bar{N}_k^4} \sum_{i,m,m'\in S_k} N_i N_m N_{m'} \sum_j \Omega_{ij}\Omega_{mj}\Omega_{m'j}$$
$$\le \sum_k \frac{1}{n_k\bar{N}_k}\|\mu_k\|_3^3,$$

which proves the second claim.

The third claim follows similarly and we omit the proof.

□

C Proofs of other main lemmas and theorems

C.1 Proof of Lemma 2.1

We start from computing $\mathbb{E}[(\hat{\mu}_{kj}-\hat{\mu}_j)^2]$. Write $X_{ij} = N_i(\Omega_{ij}+Y_{ij})$. It follows by elementary calculation that

$$\hat{\mu}_{kj} - \hat{\mu}_j = \mu_{kj} - \mu_j + \Big(\frac{1}{n_k\bar{N}_k} - \frac{1}{n\bar{N}}\Big)\sum_{i\in S_k} N_i Y_{ij} - \sum_{1\le \ell\le K:\ell\neq k} \frac{1}{n_\ell \bar{N}_\ell}\sum_{i\in S_\ell} N_i Y_{ij}.$$

For different k, the variables $\sum_{i\in S_k} N_i Y_{ij}$ are independent of each other. It follows that

$$\begin{aligned}
\mathbb{E}[(\hat{\mu}_{kj} - \hat{\mu}_j)^2] &= (\mu_{kj}-\mu_j)^2 + \Big(\frac{1}{n_k\bar{N}_k} - \frac{1}{n\bar{N}}\Big)^2 \mathbb{E}\Big[\Big(\sum_{i\in S_k} N_i Y_{ij}\Big)^2\Big] + \sum_{\ell:\ell\neq k}\frac{1}{n^2\bar{N}^2}\mathbb{E}\Big[\Big(\sum_{i\in S_\ell} N_i Y_{ij}\Big)^2\Big] \\
&= (\mu_{kj}-\mu_j)^2 + \Big(\frac{1}{n_k\bar{N}_k} - \frac{1}{n\bar{N}}\Big)^2 \sum_{i\in S_k} N_i\Omega_{ij}(1-\Omega_{ij}) + \sum_{\ell:\ell\neq k}\frac{1}{n^2\bar{N}^2}\sum_{i\in S_\ell} N_i\Omega_{ij}(1-\Omega_{ij}) \\
&= (\mu_{kj}-\mu_j)^2 + \frac{1}{n_k^2\bar{N}_k^2}\Big(1-\frac{n_k\bar{N}_k}{n\bar{N}}\Big)\sum_{i\in S_k} N_i\Omega_{ij}(1-\Omega_{ij}) \\
&\quad + \frac{1}{n^2\bar{N}^2}\Big[\Big(1-\frac{n\bar{N}}{n_k\bar{N}_k}\Big)\sum_{i\in S_k} N_i\Omega_{ij}(1-\Omega_{ij}) + \sum_{\ell:\ell\neq k}\sum_{i\in S_\ell} N_i\Omega_{ij}(1-\Omega_{ij})\Big] \\
&= (\mu_{kj}-\mu_j)^2 + \frac{1}{n_k^2\bar{N}_k^2}\Big(1-\frac{n_k\bar{N}_k}{n\bar{N}}\Big)\sum_{i\in S_k} N_i\Omega_{ij}(1-\Omega_{ij}) \\
&\quad - \frac{1}{n\bar{N}n_k\bar{N}_k}\underbrace{\Big[\sum_{i\in S_k} N_i\Omega_{ij}(1-\Omega_{ij}) - \frac{n_k\bar{N}_k}{n\bar{N}}\sum_{\ell=1}^K\sum_{i\in S_\ell} N_i\Omega_{ij}(1-\Omega_{ij})\Big]}_{\delta_{kj}}.
\end{aligned} \tag{C.1}$$

Since X_{ij} follows a binomial distribution, it is easy to see that $\mathbb{E}[X_{ij}] = N_i\Omega_{ij}$ and $\mathbb{E}[X_{ij}^2] = (\mathbb{E}[X_{ij}])^2 + \mathrm{Var}(X_{ij}) = N_i^2\Omega_{ij}^2 + N_i\Omega_{ij}(1-\Omega_{ij})$. Combining them gives

$$\mathbb{E}[X_{ij}(N_i - X_{ij})] = N_i(N_i-1)\Omega_{ij}(1-\Omega_{ij}). \tag{C.2}$$

Define

$$\hat{\zeta}_{kj} = (\hat{\mu}_{kj} - \hat{\mu}_j)^2 - \frac{1}{n_k^2 \bar{N}_k^2}\Big(1 - \frac{n_k \bar{N}_k}{n\bar{N}}\Big) \sum_{i \in S_k} \frac{X_{ij}(N_i - X_{ij})}{N_i - 1},$$

It follows from (C.1)-(C.2) that

$$\mathbb{E}[\hat{\zeta}_{kj}] = (\mu_{kj} - \mu_j)^2 - \frac{1}{n\bar{N} n_k \bar{N}_k} \delta_{kj}. \tag{C.3}$$

We are ready to compute $\mathbb{E}[T]$. By definition, $T = \sum_{j=1}^p \sum_{k=1}^K n_k \bar{N}_k \hat{\zeta}_{kj}$ and $\rho^2 = \sum_{j,k} (\mu_{kj} - \mu_j)^2$. Consequently,

$$\mathbb{E}[T] = \sum_{j=1}^p \sum_{k=1}^K n_k \bar{N}_k \Big[(\mu_{kj} - \mu_j)^2 - \frac{1}{n\bar{N} n_k \bar{N}_k} \delta_{kj}\Big] = \rho^2 - \frac{1}{n\bar{N}} \sum_{j=1}^p \sum_{k=1}^K \delta_{kj}. \tag{C.4}$$

We use the definition of δ_{kj} in (C.1). It is seen that for each $1 \leq j \leq p$,

$$\sum_{k=1}^K \delta_{kj} = \sum_{k=1}^K \sum_{i \in S_k} N_i \Omega_{ij}(1 - \Omega_{ij}) - \Big(\sum_{k=1}^K \frac{n_k \bar{N}_k}{n\bar{N}}\Big) \sum_{\ell=1}^K \sum_{i \in S_\ell} N_i \Omega_{ij}(1 - \Omega_{ij}) = 0. \tag{C.5}$$

Combining (C.4)-(C.5) gives $\mathbb{E}[T] = \rho^2$. This proves the claim. $\square$

C.2 Proof of Theorem 3.3

First we show that

$$\mathrm{Var}(T) \lesssim \Theta_n \tag{C.6}$$

Recall

$$\Theta_{n1} = 4 \sum_{k=1}^K \sum_{j=1}^p n_k \bar{N}_k (\mu_{kj} - \mu_j)^2 \mu_{kj}$$

$$\Theta_{n2} = 2 \sum_{k=1}^K \sum_{i \in S_k} \sum_{j=1}^p \Big(\frac{1}{n_k \bar{N}_k} - \frac{1}{n\bar{N}}\Big)^2 \frac{N_i^3}{N_i - 1} \Omega_{ij}^2$$

$$\Theta_{n3} = \frac{2}{n^2 \bar{N}^2} \sum_{1 \leq k \neq \ell \leq K} \sum_{i \in S_k} \sum_{m \in S_\ell} \sum_{j=1}^p N_i N_m \Omega_{ij} \Omega_{mj}$$

$$\Theta_{n4} = 2 \sum_{k=1}^K \sum_{\substack{i \in S_k, m \in S_k, \\ i \neq m}} \sum_{j=1}^p \Big(\frac{1}{n_k \bar{N}_k} - \frac{1}{n\bar{N}}\Big)^2 N_i N_m \Omega_{ij} \Omega_{mj}.$$

and that $\sum_{a=1}^4 \Theta_{na} = \Theta_n$.

By Lemma A.2, we immediately have

$$\mathrm{Var}(\mathbf{1}_p' U_1) \leq \Theta_{n1}. \tag{C.7}$$

For U_2, it is shown in the Proof of Lemma A.3 that

$$\mathrm{Var}(\mathbf{1}_p' U_2) = 4\sum_{k=1}^{K}\sum_{i\in S_k}\sum_{1\le r<s\le N_i}\frac{\theta_i}{N_i(N_i-1)}\big[\|\Omega_i\|^2 + O(\|\Omega_i\|_3^3)\big].$$

Thus

$$\begin{aligned}\mathrm{Var}(\mathbf{1}_p' U_2) &\lesssim 4\sum_{k=1}^{K}\sum_{i\in S_k}\sum_{1\le r<s\le N_i}\frac{\theta_i}{N_i(N_i-1)}\|\Omega_i\|^2\\ &= 2\sum_{k=1}^{K}\sum_{i\in S_k}\theta_i\|\Omega_i\|^2 = \Theta_{n2}\end{aligned} \tag{C.8}$$

Next we study U_3. Using that $\Omega_{mj'} \le 1$ and $\|\Omega_i\|_1 = 1$, we have

$$\begin{aligned}\sum_{k\neq\ell}\frac{n_k n_\ell \bar{N}_k \bar{N}_\ell}{n^2\bar{N}^2}\mathbf{1}_p'(\Sigma_k\circ\Sigma_\ell)\mathbf{1}_p &= \frac{2}{n^2\bar{N}^2}\sum_{k\neq\ell}\sum_{i\in S_k}\sum_{m\in S_\ell}\sum_{j,j'} N_i N_m \Omega_{ij}\Omega_{ij'}\Omega_{mj}\Omega_{mj'}\\ &\le \frac{2}{n^2\bar{N}^2}\sum_{k\neq\ell}\sum_{i\in S_k}\sum_{m\in S_\ell}\sum_{j} N_i N_m \Omega_{ij}\Omega_{mj}\sum_{j'}\Omega_{ij'}\\ &= \frac{2}{n^2\bar{N}^2}\sum_{k\neq\ell}\sum_{i\in S_k}\sum_{m\in S_\ell}\sum_{j} N_i N_m \Omega_{ij}\Omega_{mj}.\end{aligned}$$

Therefore by Lemma A.4,

$$\mathrm{Var}(\mathbf{1}_p' U_3) \lesssim \frac{2}{n^2\bar{N}^2}\sum_{1\le k\neq\ell\le K}\sum_{i\in S_k}\sum_{m\in S_\ell}\sum_{j=1}^{p} N_i N_m\Omega_{ij}\Omega_{mj} = \Theta_{n3}. \tag{C.9}$$

Similarly for U_4, we have by the Proof of Lemma A.5 that

$$\begin{aligned}\mathrm{Var}(\mathbf{1}_p' U_4) &= 4\sum_{k=1}^{K}\sum_{\substack{i\in S_k, m\in S_k\\ i<m}}\kappa_{im}\Big(\sum_j \Omega_{ij}\Omega_{mj} + \delta_{im}\Big)\\ &\lesssim \sum_{k=1}^{K}\sum_{\substack{i\in S_k, m\in S_k\\ i<m}}\kappa_{im}\sum_j \Omega_{ij}\Omega_{mj} = \Theta_{n4}.\end{aligned} \tag{C.10}$$

Above we use that $|\delta_{im}| \le \sum_j \Omega_{ij}\Omega_{mj}$ and recall that $\kappa_{im} = (\frac{1}{n_k\bar{N}_k} - \frac{1}{n\bar{N}})^2 N_i N_m$.

Observe that by Lemma 2.1,

$$\Theta_{n1} = 4\sum_{k=1}^{K}\sum_{j=1}^{p} n_k\bar{N}_k(\mu_{kj}-\mu_j)^2\mu_{kj} \lesssim \max_k\|\mu_k\|_\infty\cdot\rho^2 = \max_k\|\mu_k\|_\infty\cdot\mathbb{E}\,T. \tag{C.11}$$

Since (3.1) holds, Lemma A.6 applies and

$$\Theta_{n_2} + \Theta_{n3} + \Theta_{n4} \asymp \sum_k \|\mu_k\|^2. \tag{C.12}$$

Combining (C.6), (C.11), and (C.12) proves the theorem. $\square$

C.3 Proof of Theorem 3.4

To prove Theorem 3.4, we must prove the following claims:

(a) Under the alternative hypothesis, $\psi \to \infty$ in probability.

(b) For any fixed $\alpha \in (0,1)$, the level-α DELVE test has an asymptotic level of α and an asymptotic power of 1.

(c) If we choose $\alpha = \alpha_n$ such that $\alpha_n \to 0$ and $1 - \Phi(\mathrm{SNR}_n) = o(\alpha_n)$, where Φ is the CDF of $N(0,1)$, then the sum of type I and type II errors of the DELVE test converges to 0.

We show the first claim, that $\psi \to \infty$, under the alternative hypothesis and the conditions of Theorem 3.4. In particular, recall we assume that

$$\frac{\rho^2}{\sqrt{\sum_{k=1}^K \|\mu_k\|^2}} = \frac{n\bar{N}\|\mu\|^2\omega_n^2}{\sqrt{\sum_{k=1}^K \|\mu_k\|^2}} \to \infty. \tag{C.13}$$

Our first goal is to show that

$$T/\sqrt{\mathrm{Var}(T)} \xrightarrow{\mathbb{P}} \infty \tag{C.14}$$

under the alternative. By Chebyshev's inequality, it suffices to show that

$$\mathbb{E}\,T \gg \sqrt{\mathrm{Var}(T)}. \tag{C.15}$$

By Theorem 3.3,

$$\mathrm{Var}(T) \lesssim \sum_k \|\mu_k\|^2 + \max_k \|\mu_k\|_\infty \cdot \mathbb{E}T = \sum_k \|\mu_k\|^2 + \max_k \|\mu_k\|_\infty \cdot \rho^2 \tag{C.16}$$

By (C.13),

$$\mathbb{E}T = \rho^2 \gg \sqrt{\sum_{k=1}^K \|\mu_k\|^2} \geq \max_{1\leq k\leq K} \|\mu_k\|_\infty.$$

Therefore,

$$\sqrt{\max_{1\leq k\leq K} \|\mu_k\|_\infty} \cdot \rho \ll \rho^2 = \mathbb{E}T. \tag{C.17}$$

Moreover, by (C.13),

$$\sum_k \|\mu_k\|^2 \ll \rho^4 = (\mathbb{E}T)^2. \tag{C.18}$$

Combining (C.16), (C.17), and (C.18) implies (C.14).

Next we show that $V > 0$ with high probability (i.e., with probability tending to 1 as $n\bar{N} \to \infty$). Recall that by Lemmas A.6, A.10, and A.11,

$$\mathbb{E}V = \Theta_{n2} + \Theta_{n3} + \Theta_{n4} \gtrsim \sum_k \|\mu_k\|^2 > 0, \text{ and} \tag{C.19}$$

$$\mathrm{Var}(V) \lesssim \sum_k \frac{\|\mu_k\|^2}{n_k^2 \bar{N}_k^2} \vee \sum_k \frac{\|\mu_k\|_3^3}{n_k \bar{N}_k}. \tag{C.20}$$

Using this, the Markov inequality, and (3.4), we have

$$\mathbb{P}\big(V < \mathbb{E}[V]/2\big) \le \mathbb{P}\big(|V - \mathbb{E}[V]| \ge \mathbb{E}[V]/2\big) \le \frac{4\mathrm{Var}(V)}{(\mathbb{E}[V])^2} = o(1), \tag{C.21}$$

which implies that $V > 0$ with high probability.

To finish the proof of the first claim, note that the assumptions of Proposition A.2 are satisfied and we have $V/\mathrm{Var}(T) = O_{\mathbb{P}}(1)$. By this, (C.14), and (C.21), we have

$$\psi = \frac{T\mathbf{1}_{V>0}}{\sqrt{V}} = \frac{\sqrt{\mathrm{Var}(T)}}{\sqrt{V}} \cdot \frac{T}{\sqrt{\mathrm{Var}(T)}} \cdot \mathbf{1}_{V>0} \gtrsim \frac{T}{\sqrt{\mathrm{Var}(T)}} \to \infty$$

in probability.

The second claim follows directly from the first claim and Theorem 3.2.

To prove the third claim, by Chebyshev's inequality and $T/\sqrt{\mathrm{Var}(T)} \to \infty$, it follows that $T > (1/2)\mathbb{E}T = (1/2)\rho^2$ with high probability as $n\bar{N} \to \infty$. By a similar Chebyshev argument as above, it also holds that $V < (3/2)\mathbb{E}V$ with high probability as $n\bar{N} \to \infty$. Recall that $\mathbb{E}V = \Theta_{n2} + \Theta_{n3} + \Theta_{n4} \lesssim \sum_k \|\mu_k\|^2$ by Lemmas A.6 and A.10. Thus, with high probability as $n\bar{N} \to \infty$, we have

$$\psi = T\mathbf{1}_{V>0}/\sqrt{V} \gtrsim \rho^2/\sqrt{\mathbb{E}V} \gtrsim \frac{n\bar{N}\|\mu\|^2\omega_n^2}{\sqrt{\sum_k \|\mu_k\|^2}} = \mathrm{SNR}_n.$$

Choosing α_n as specified yield the third claim. The proof is complete since all three claims are established.

□

C.4 Proof of Theorem 3.5

Without loss of generality, we assume p is even and write $m = p/2$. Let $\mu \in \mathbb{R}^m$ be a nonnegative vector with $\|\mu\|_1 = 1/2$. Let $\tilde{\mu} = (\mu', \mu')' \in \mathbb{R}^p$. We consider the null hypothesis:

$$H_0: \qquad \Omega_i = \tilde{\mu}, \qquad 1 \le i \le n. \tag{C.22}$$

We pair it with a random alternative hypothesis. Let $b_1, b_2, \ldots, b_m$ be a collection of i.i.d. Rademacher variables. Let $z_1, z_2, \ldots, z_K$ denote an independent collection of i.i.d. Rademacher random variables conditioned on the event $|\sum_k z_k| \le 100\sqrt{K}$. For a properly small sequence $\omega_n > 0$ of positive numbers, let

$$H_1: \qquad \Omega_{ij} = \begin{cases} \mu_j\big(1 + \omega_n (n_k\bar{N}_k)^{-1}\big(\frac{1}{K}\sum_{k\in K} n_k\bar{N}_k\big) z_k b_j\big), & \text{if } 1 \le j \le m, i \in S_k \\ \tilde{\mu}_j\big(1 - \omega_n (n_k\bar{N}_k)^{-1}\big(\frac{1}{K}\sum_{k\in K} n_k\bar{N}_k\big) z_k b_{j-m}\big), & \text{if } m+1 \le j \le 2m, i \in S_k \end{cases} \tag{C.23}$$

In this section we slightly abuse notation, using ω_n to refer to the (deterministic) sequence above and reserving $\omega(\Omega)$ for the random quantity

$$\omega(\Omega) = \sqrt{\frac{1}{n\bar{N}\|\mu\|^2}\sum_{k=1}^{K} n_k\bar{N}_k\|\mu_k - \mu\|^2}. \tag{C.24}$$

As long as

$$\omega_n \leq \frac{\min_k n_k\bar{N}_k}{\frac{1}{K}\sum_{k\in[K]} n_k\bar{N}_k} = \frac{\min_k n_k\bar{N}_k}{n\bar{N}/K},$$

then $\Omega_{ij} \geq 0$ for all $i \in [n], j \in [p]$. Furthermore, for each $1 \leq i \leq n$, we have $\|\Omega_i\|_1 = 2\|\mu\|_1 = 1$. We suppose there exists a constant $c \in (0,1)$ such that

$$cK^{-1}n\bar{N} \leq n_k\bar{N}_k \leq c^{-1}K^{-1}n\bar{N} \quad \text{for all } k \in [K] \tag{C.25}$$

With (C.25) in hand, we may assume without loss of generality that

$$\omega_n \leq c/2 \tag{C.26}$$

This assumption implies that (C.23) is well-defined and moreover $\Omega_{ij} \asymp \mu_j$.

Next we characterize the random quantity $\omega(\Omega)$ in terms of ω_n.

Lemma C.1. *Let $\omega^2(\Omega)$ be as in* (C.24)*. When Ω follows Model* (C.23)*, there exists a constant $c_1 \in (0,1)$ such that $c_1\omega_n^2 \leq \omega^2(\Omega) \leq c_1^{-1}\omega_n^2$ with probability* 1*.*

The proof of Lemma C.1 is given in Section C.4.1. By Lemma C.1, under the model (C.23) it holds with probability 1 that

$$\frac{n\bar{N}\|\mu\|^2\omega^2(\Omega)}{\sqrt{\sum_{k=1}^K \|\mu_k\|^2}} \asymp K^{-1/2}n\bar{N}\|\mu\|\omega_n^2. \tag{C.27}$$

Above we use that $\Omega_{ij} \asymp \mu_j$, since we assume (C.26)

We also require Proposition C.1 below, whose proof is given in Section C.4.2.

Proposition C.1. *Suppose that* (C.25) *and* (C.26) *hold. Consider the pair of hypotheses in* (C.22)*-*(C.23) *and let $\mathbb{P}_0$, and $\mathbb{P}_1$ be the respective probability measures. If*

$$\frac{n\bar{N}\|\mu\|^2\omega^2(\Omega)}{\sqrt{\sum_{k=1}^K \|\mu_k\|^2}} \asymp K^{-1/2}n\bar{N}\|\mu\|\omega_n^2 \to 0,$$

then the chi-square distance between $\mathbb{P}_0$ and $\mathbb{P}_1$ converges to 0*.*

Now we prove Theorem 3.5. Let δ_n denote an arbitrary sequence tending to 0. Without loss of generality, we may assume that $\delta_n \leq c^*$ for a small absolute constant $c^* \in (0,1)$. Note that $K^{-1/2}n\bar{N} \geq 1$ since $K \leq n$. Thus for appropriate choice of sequences of $\mu = \mu_n$ and $\omega_n \leq c/2$ in models (C.22), (C.23) and applying (C.27), we obtain

$$2\delta_n \geq \frac{n\bar{N}\|\mu\|^2\omega^2(\Omega)}{\sqrt{\sum_{k=1}^K \|\mu_k\|^2}} \geq \delta_n. \tag{C.28}$$

Recall the definitions of $\mathcal{Q}^*_{0n}$ and $\mathcal{Q}^*_{1n}$ in (3.8). Let Π denote the distribution on $\xi = \{(N_i, \Omega_i, \ell_i)\} \in \mathcal{Q}^*_{1n}$ induced by (C.23). Let ξ_0 denote the parameter associated to the simple null hypothesis in (C.22) associated to our choice of μ and ω_n satisfying (C.28). We have by standard manipulations,

$$\begin{aligned}
\mathcal{R}(\mathcal{Q}^*_{0n}, \mathcal{Q}^*_{1n}) &:= \inf_{\Psi\in\{0,1\}} \Big\{ \sup_{\xi\in\mathcal{Q}^*_{0n}(c_0,\epsilon_n)} \mathbb{P}_\xi(\Psi=1) + \sup_{\xi\in\mathcal{Q}^*_{1n}(\delta_n;c_0,\epsilon_n)} \mathbb{P}_\xi(\Psi=0) \Big\} \\
&= \inf_{\Psi\in\{0,1\}} \Big\{ \sup_{\xi\in\mathcal{Q}^*_{0n}(c_0,\epsilon_n),\xi'\in\mathcal{Q}^*_{1n}(\delta_n;c_0,\epsilon_n)} \big[\mathbb{P}_\xi(\Psi=1) + \mathbb{P}_\xi(\Psi=0)\big] \\
&\geq \inf_{\Psi\in\{0,1\}} \Big\{ \sup_{\xi\in\mathcal{Q}^*_{0n}(c_0,\epsilon_n)} \mathbb{E}_{\xi'\sim\Pi}\Big[\mathbb{P}_\xi(\Psi=1) + \mathbb{P}_{\xi'}(\Psi=0)\Big]\Big\} \\
&\geq \inf_{\Psi\in\{0,1\}} \Big\{ \mathbb{E}_{\xi'\sim\Pi}\Big[\mathbb{P}_{\xi_0}(\Psi=1) + \mathbb{P}_{\xi'}(\Psi=0)\Big]\Big\} \\
&= \inf_{\Psi\in\{0,1\}} \Big\{ \mathbb{P}_0(\Psi=1) + \mathbb{P}_1(\Psi=0) \Big\}.
\end{aligned}$$

In the last line we recall the definition of $\mathbb{P}_0$ and $\mathbb{P}_1$ in (C.22) and (C.23), noting that for all events E,

$$\mathbb{P}_1(E) = \mathbb{E}_{\xi'\sim\pi}\,\mathbb{P}_{\xi'}(E).$$

Next, by the Neyman–Pearson lemma and the standard inequality $\mathrm{TV}(P,Q) \leq \sqrt{\chi^2(P,Q)}$ (see e.g. Chapter 2 of Tsybakov [2008]),

$$\begin{aligned}
\mathcal{R}(\mathcal{Q}^*_{0n}, \mathcal{Q}^*_{1n}) &\geq \inf_{\Psi\in\{0,1\}} \Big\{ \mathbb{P}_0(\Psi=1) + \mathbb{P}_1(\Psi=0) \Big\} \\
&= 1 - \mathrm{TV}\big(\mathbb{P}_0, \mathbb{P}_1\big) \geq 1 - \sqrt{\chi^2(\mathbb{P}_0, \mathbb{P}_1)}.
\end{aligned}$$

By Proposition C.1, as $\delta_n \to 0$ we have $\chi^2(\mathbb{P}_0, \mathbb{P}_1) \to 0$ and thus $\mathcal{R}(\mathcal{Q}^*_{0n}, \mathcal{Q}^*_{1n}) \to 1$, as desired.

□

C.4.1 Proof of Proposition C.1

Next, we perform a change of parameters that preserves the signal strength and chi-squared distance. The testing problem (C.22) and (C.23) has parameters $\Omega_{ij}, N_i, \bar{N}_k, n_k, n$, and K. Let $\mathbb{P}_0$ and $\mathbb{P}_1$ denote the distributions corresponding to the null and alternative hypotheses, respectively. For each $k \in [K]$, we combine all documents in sample k to obtain new null and alternative distributions $\tilde{\mathbb{P}}_0$ and $\tilde{\mathbb{P}}_1$ with parameters $\tilde{\Omega}_{ij}, \tilde{N}_i, \bar{\tilde{N}}_i, \tilde{n}_i, \tilde{n}$, and $\tilde{K}$ such that

$$\begin{aligned}
&\tilde{K} = K = \tilde{n} && \\
&\tilde{N}_i = n_i \bar{N}_i && \text{for } i \in [\tilde{K}] \\
&\bar{\tilde{N}}_i \equiv \tilde{N}_i && \text{for } i \in [\tilde{K}] \\
&\tilde{n}_i = 1 && \text{for } i \in [\tilde{K}]\ . \qquad \text{(C.29)}
\end{aligned}$$

For notational ease, we define $\tilde{N} := \bar{\tilde{N}} = \frac{1}{K}\sum_{k\in[K]} n_k \bar{N}_k$. Furthermore, we have $\tilde{\Omega}_i = \mu$ for all $i \in [\tilde{n}]$ under the null $\tilde{\Omega}_i = \mu_i$ for all $i \in [\tilde{n}]$ under the alternative. Explicitly, in the reparameterized model, we have the null hypothesis

$$H_0: \qquad \Omega_i = \tilde{\mu}, \qquad 1 \le i \le n. \tag{C.30}$$

and alternative hypothesis

$$H_1: \qquad \Omega_{ij} = \begin{cases} \mu_j\big(1 + \omega_n \tilde{N}_i^{-1} \tilde{N} z_i b_j\big), & \text{if } 1 \le j \le m, \\ \tilde{\mu}_j\big(1 - \omega_n \tilde{N}_i^{-1} \tilde{N} z_i b_{j-m}\big), & \text{if } m+1 \le j \le 2m. \end{cases} \tag{C.31}$$

for all $i \in [\tilde{K}] = [K] = [\tilde{n}]$. Observe that the likelihood ratio is preserved: $\frac{d\mathbb{P}_0}{d\mathbb{P}_1} = \frac{d\tilde{\mathbb{P}}_0}{d\tilde{\mathbb{P}}_1}$ and also $\omega(\Omega) = \omega(\widetilde{\Omega})$. For simplicity we work with this reparameterized model in this proof.

If $z_1, \dots, z_{\tilde{n}}$ are independent Rademacher random variables then with probability at least $1/2$ it holds that

$$\Big|\sum_i z_i\Big| \le 100\sqrt{\tilde{n}} \tag{C.32}$$

by Hoeffding's inequality. Recall that our random model is defined in (C.23) where (i) $z_1, \dots, z_{\tilde{n}}$ are independent Rademacher random variables conditioned on the event $|\sum_i z_i| \le 100\sqrt{\tilde{n}}$, and (ii) $b_1, \dots, b_m$ are independent Rademacher random variables.

Now we study $\omega^2(\widetilde{\Omega})$. For each $1 \le j \le m$, we have $\widetilde{\Omega}_{ij} = \mu_j(1 + \omega_n \tilde{N}_i^{-1} \tilde{N} z_i b_j)$. Define $\eta_j = (\tilde{n}\tilde{N})^{-1} \sum_{i=1}^{\tilde{n}} \tilde{N}_i \widetilde{\Omega}_{ij} = \mu_j(1 + \omega_n \bar{z} b_j)$ for $1 \le j \le m$ and $\eta_j = (\tilde{n}\tilde{N})^{-1} \sum_{i=1}^{\tilde{n}} \tilde{N}_i \widetilde{\Omega}_{ij} = \tilde{\mu}_j(1 - \omega_n \bar{z} b_j)$ for $m < j \le 2m$. We have

$$\begin{aligned} \sum_{i=1}^{\tilde{n}} \sum_{j=1}^{p} \tilde{N}_i (\widetilde{\Omega}_{ij} - \eta_j)^2 &= 2 \sum_{i=1}^{\tilde{n}} \sum_{j=1}^{m} \tilde{N}_i \cdot \mu_j^2 \omega_n^2 \frac{\tilde{N}^2}{\tilde{N}_i^2} (z_i - \bar{z})^2 b_j^2 \\ &= 2\omega_n^2 \tilde{N}^2 \|\mu\|^2 \sum_{i=1}^{\tilde{n}} \tilde{N}_i^{-1} (z_i - \bar{z})^2. \end{aligned}$$

By (C.32), $|\bar{z}| \le 100\sqrt{\tilde{n}}$. Thus $|z_i - \bar{z}| \asymp 1$. Write $\tilde{N}_* = (\tilde{n}^{-1} \sum_{i=1}^{\tilde{n}} \tilde{N}_i^{-1})$. It follows that

$$\sum_{i=1}^{\tilde{n}} \sum_{j=1}^{p} \tilde{N}_i (\widetilde{\Omega}_{ij} - \eta_j)^2 \asymp \omega_n^2 \tilde{N}^2 \|\mu\|^2 \cdot \tilde{n} \tilde{N}_*^{-1}.$$

Note that $\tilde{N} \ge \tilde{N}_*$. Additionally, by assumption (C.25), $\tilde{N}_i \asymp \tilde{N} \le c^{-1} \tilde{N}_*$. It follows that

$$\sum_{i=1}^{\tilde{n}} \sum_{j=1}^{p} \tilde{N}_i (\widetilde{\Omega}_{ij} - \eta_j)^2 \asymp \tilde{n} \tilde{N} \|\mu\|^2 \omega_n^2. \tag{C.33}$$

Moreover, $\|\eta\|^2 = \sum_{j=1}^{p} \mu_j^2 (1 + \omega_n \bar{z} b_j)^2$. By our conditioning on the event in (C.32),

$$|\omega_n \bar{z} b_j| \lesssim \omega_n \tilde{n}^{-1/2}.$$

Since $\omega_n \leq 1$ and $\sum_j b_j = 0$, we have

$$\|\eta\|^2 = \|\mu\|^2 + \sum_{j=1}^{p} \mu_j^2 \omega_n^2 \bar{z}^2 = \|\mu\|^2[1 + O(\tilde{n}^{-1})] \asymp \|\mu\|^2. \tag{C.34}$$

Hence

$$\omega^2(\widetilde{\Omega}) = \omega^2(\Omega) \asymp \omega_n^2, \qquad \text{where recall} \quad \omega(\widetilde{\Omega}) = \frac{\sum_{i=1}^{\tilde{n}} \sum_{j=1}^{p} \tilde{N}_i(\widetilde{\Omega}_{ij} - \eta_j)^2}{\tilde{n}\tilde{N}\|\eta\|^2}. \tag{C.35}$$

This finishes the proof. $\square$

C.4.2 Proof of Proposition C.1

In this proof, we continue to employ the reparametrization in (C.29). As discussed there, this reparametrization preserves the likelihood ratio and thus the chi-square distance.

By definition, $\chi^2(\mathbb{P}_0, \mathbb{P}_1) = \int (\frac{d\mathbb{P}_1}{d\mathbb{P}_0})^2 d\mathbb{P}_0 - 1$. It suffices to show that

$$\int \Big(\frac{d\mathbb{P}_1}{d\mathbb{P}_0}\Big)^2 d\mathbb{P}_0 = 1 + o(1). \tag{C.36}$$

From the density of of multinomial distribution, $d\mathbb{P}_0 = \prod_{i,j} \tilde{\mu}_j^{X_{ij}}$, and $d\mathbb{P}_1 = \mathbb{E}_{b,z}[\prod_{i,j} \widetilde{\Omega}_{ij}^{X_{ij}}]$. It follows that

$$\frac{d\mathbb{P}_1}{d\mathbb{P}_0} = \mathbb{E}_{b,z}\Big[\prod_{i=1}^{\tilde{n}} \prod_{j=1}^{p} \Big(\frac{\widetilde{\Omega}_{ij}}{\tilde{\mu}_j}\Big)^{X_{ij}}\Big].$$

Let $b^{(0)} = (b_1^{(0)}, \ldots, b_m^{(0)})'$ and $z^{(0)} = (z_1^{(0)}, \cdots, z_{\tilde{n}}^{(0)})'$ be independent copies of b and z. We construct $\widetilde{\Omega}_{ij}^{(0)}$ similarly as in (C.31). It is seen that

$$\begin{aligned}
\int \Big(\frac{d\mathbb{P}_1}{d\mathbb{P}_0}\Big)^2 d\mathbb{P}_0 &= \mathbb{E}_X \mathbb{E}_{b,z,b^{(0)},z^{(0)}}\Big[\prod_{i=1}^{\tilde{n}} \prod_{j=1}^{p} \Big(\frac{\widetilde{\Omega}_{ij}\widetilde{\Omega}_{ij}^{(0)}}{\tilde{\mu}_j^2}\Big)^{X_{ij}}\Big] \\
&= \mathbb{E}_{b,z,b^{(0)},z^{(0)}}\Big\{\prod_{i=1}^{\tilde{n}} \mathbb{E}_{X_i}\Big[\prod_{j=1}^{p} \Big(\frac{\widetilde{\Omega}_{ij}\widetilde{\Omega}_{ij}^{(0)}}{\tilde{\mu}_j^2}\Big)^{X_{ij}}\Big]\Big\} \\
&= \mathbb{E}_{b,z,b^{(0)},z^{(0)}}\Big\{\prod_{i=1}^{\tilde{n}} \Big(\sum_{j=1}^{p} \tilde{\mu}_j \cdot \frac{\widetilde{\Omega}_{ij}\widetilde{\Omega}_{ij}^{(0)}}{\tilde{\mu}_j^2}\Big)^{\tilde{N}_i}\Big]\Big\} \\
&= \mathbb{E}[\exp(M)], \quad \text{with} \quad M := \sum_{i=1}^{\tilde{n}} \tilde{N}_i \log\Big(\sum_{j=1}^{p} \tilde{\mu}_j^{-1}\widetilde{\Omega}_{ij}\widetilde{\Omega}_{ij}^{(0)}\Big).
\end{aligned} \tag{C.37}$$

Here, the third line follows from the moment generating function of a multinomial distribution. We plug in the expression of $\widetilde{\Omega}_{ij}$ in (C.23). By direct calculations,

$$\sum_{j=1}^{p} \tilde{\mu}_j^{-1}\widetilde{\Omega}_{ij}\widetilde{\Omega}_{ij}^{(0)} = \sum_{j=1}^{m} \mu_j\big(1 + \omega_n \tilde{N}_i^{-1}\tilde{N} z_i b_j\big)\big(1 + \omega_n \tilde{N}_i^{-1}\tilde{N} z_i^{(0)} b_j^{(0)}\big)$$

$$
\begin{aligned}
&+\sum_{j=1}^{m}\mu_j\big(1-\omega_n\tilde{N}_i^{-1}\tilde{N}z_ib_j\big)\big(1-\omega_n\tilde{N}_i^{-1}\tilde{N}z_i^{(0)}b_j^{(0)}\big)\\
&=2\|\mu\|_1+2\sum_{j=1}^{m}\mu_j\omega_n^2\tilde{N}_i^{-2}\tilde{N}^2z_iz_i^{(0)}b_jb_j^{(0)}\\
&=1+2\sum_{j=1}^{m}\mu_j\omega_n^2\tilde{N}_i^{-2}\tilde{N}^2z_iz_i^{(0)}b_jb_j^{(0)}.
\end{aligned}
$$

We plug it into M and notice that $\log(1+t)\le t$ is always true. It follows that

$$
M\le\sum_{i=1}^{\tilde{n}}\tilde{N}_i\cdot 2\sum_{j=1}^{m}\mu_j\omega_n^2\frac{\tilde{N}^2}{\tilde{N}_i^2}z_iz_i^{(0)}b_jb_j^{(0)}=2\tilde{N}\omega_n^2\Big(\sum_{i=1}^{\tilde{n}}\frac{\tilde{N}}{\tilde{N}_i}z_iz_i^{(0)}\Big)\Big(\sum_{j=1}^{m}\mu_jb_jb_j^{(0)}\Big)=:M^*. \tag{C.38}
$$

We combine (C.38) with (C.37). It is seen that to show (C.36), it suffices to show that

$$
\mathbb{E}[\exp(M^*)]=1+o(1). \tag{C.39}
$$

We now show (C.39). Write $M_1=\sum_{i=1}^{\tilde{n}}(\tilde{N}_i^{-1}\tilde{N})z_iz_i^{(0)}$ and $M_2=\sum_{j=1}^{p}\mu_jb_jb_j^{(0)}$.

Recall that we condition on the event (C.32). By Hoeffding's inequality, Bayes's rule, and (C.32),

$$
\begin{aligned}
\mathbb{P}(|M_1|>t)&=\mathbb{P}\bigg(|\sum_i\frac{\tilde{N}}{\tilde{N}_i}z_iz_i^{(0)}\ge t\;\bigg|\;|\sum_i z_i|\le 100\sqrt{\tilde{n}},|\sum_i z_i^{(0)}|\le 100\sqrt{\tilde{n}}\bigg)\\
&=\frac{\mathbb{P}\big(|\sum_i\frac{\tilde{N}}{\tilde{N}_i}z_iz_i^{(0)}|\ge t\big)}{\mathbb{P}(|\sum_i z_i|\le 100\sqrt{\tilde{n}})\;\mathbb{P}(|\sum_i z_i^{(0)}|\le 100\sqrt{\tilde{n}})}\\
&\le 4\cdot 2\exp\Big(-\frac{t^2}{8\sum_{i=1}^{\tilde{n}}(\tilde{N}_i^{-1}\tilde{N})^2}\Big)\\
&=8\exp\Big(-\frac{t^2}{8\tilde{n}}\Big).
\end{aligned}
$$

for all $t>0$. In the last line, we have used the assumption of $\tilde{N}_i\asymp\tilde{N}$. By Hoeffding's inequality again, we also have

$$
\mathbb{P}(|M_2|>t)\le 2\exp\Big(-\frac{t^2}{8\sum_{j=1}^{p}\mu_j^2}\Big)=2\exp\Big(-\frac{t^2}{8\|\mu\|^2}\Big)
$$

for all $t>0$. Write $s_{\tilde{n}}^2=\sqrt{\tilde{n}}\tilde{N}\omega_n^2\|\mu\|$. It follows that

$$
\begin{aligned}
\mathbb{P}(M^*>t)&=\mathbb{P}\big(2\tilde{N}\omega_n^2M_1M_2>t\big)=\mathbb{P}\big(M_1M_2>t\cdot\sqrt{\tilde{n}}\|\mu\|s_{\tilde{n}}^{-2}\big)\\
&\le\mathbb{P}\big(M_1>\sqrt{t}\cdot\sqrt{\tilde{n}}s_{\tilde{n}}^{-1}\big)+\mathbb{P}\big(M_2>\sqrt{t}\cdot\|\mu\|s_{\tilde{n}}^{-1}\big)\\
&\le 8\exp\Big(-\frac{t}{8s_{\tilde{n}}^2}\Big)+2\exp\Big(-\frac{t}{8s_{\tilde{n}}^2}\Big)\\
&\le 4\exp(-c_1t/s_{\tilde{n}}^2),
\end{aligned} \tag{C.40}
$$

for some constant $c_1>0$. Here, in the last line, we have used the assumption of $\tilde{N}_i\asymp\tilde{N}$.

Let $f(x)$ and $F(x)$ be the density and distribution function of M^*. Write $\bar{F}(x) = 1 - F(x)$. Using integration by part, we have $\mathbb{E}[\exp(M^*)] = \int_0^\infty \exp(x)f(x)dx = -\exp(x)\bar{F}(x)|_0^\infty + \int_0^\infty \exp(x)\bar{F}(x)dx = 1 + \int_0^\infty \exp(x)\bar{F}(x)dx$, provided that the integral exists. As a result, when $s_{\tilde{n}} = o(1)$,

$$\begin{aligned}
\mathbb{E}[\exp(M^*)] - 1 &= \int_0^\infty \exp(t) \cdot \mathbb{P}(M^* > t) \\
&\leq 4 \int_0^\infty \exp\big(-[c_1 s_{\tilde{n}}^{-2} - 1]t\big) dt \\
&\leq 4(c_1 s_{\tilde{n}}^{-1} - 1)^{-1} = 4 s_{\tilde{n}}/(c_1 - s_{\tilde{n}}).
\end{aligned}$$

It implies $\mathbb{E}[\exp(M^*)] = 1 + o(1)$, which is exactly (C.39). This completes the proof. because

$$s_{\tilde{n}}^2 = \sqrt{\tilde{n}}\tilde{N}\omega_n^2\|\mu\| = \frac{n\bar{N}\|\mu\|\omega_n^2}{\sqrt{K}} \asymp \frac{n\bar{N}\|\mu\|\omega_n^2}{\sqrt{\sum_{k\in K}\|\mu_k\|^2}.}$$

□

C.5 Proof of Theorem 3.6

First we show that

$$T/\sqrt{\mathrm{Var}(T)} \Rightarrow N(0,1), \quad \text{and} \tag{C.41}$$

$$V/\mathrm{Var}(T) \to 1. \tag{C.42}$$

If (C.41) and (C.42) hold, then by mimicking the proof of Theorem 3.2, we see that ψ is asymptotically normal and the level-α DELVE test has asymptotic level α. We omit the details as they are quite similar.

Recall the martingale decomposition of T described in Section B. Observe that, under our assumptions, Lemmas B.1–B.6 are valid. Moreover, by Lemmas A.8 and A.12

$$\mathrm{Var}(T) \gtrsim \Theta_{n2} + \Theta_{n3} + \Theta_{n4} \gtrsim \left\| \frac{m\bar{M}}{n\bar{N} + m\bar{M}}\eta + \frac{n\bar{N}}{n\bar{N} + m\bar{M}}\theta \right\|^2. \tag{C.43}$$

Combining (C.43) with Lemmas B.1–B.6 and mimicking the argument in Section B.1 implies that $T/\sqrt{V} \Rightarrow N(0,1)$. Thus (C.41) is established.

Moreover, (C.42) is a direct consequence of our assumptions and Proposition A.3. The claims of Theorem 3.6 regarding the null hypothesis follow.

To prove the claims about the alternative hypothesis, it suffices to show

$$T/\sqrt{\mathrm{Var}(T)} \to \infty, \tag{C.44}$$

$$V > 0 \;\text{ with high probability, and} \tag{C.45}$$

$$V = O_{\mathbb{P}}(\mathrm{Var}(T)). \tag{C.46}$$

Once these claims are established, we prove that $\psi = T\mathbf{1}_{V>0}/\sqrt{V} \to \infty$ under the alternative by mimicking the last step of the proof of Theorem 3.4 in Section C.3. We omit the details as they are very similar.

Note that (C.46) follows directly from our assumptions and Proposition A.4.

As in the proof of Theorem 3.4 in Section C.3, to establish (C.44), it suffices to prove that

$$\mathbb{E}T = \rho^2 \gg \mathrm{Var}(T). \tag{C.47}$$

Our main assumption under the alternative when $K = 2$ is

$$\frac{\|\eta - \theta\|^2}{(\frac{1}{n\bar{N}} + \frac{1}{m\bar{M}}) \max\{\|\eta\|, \|\theta\|\}} \to \infty. \tag{C.48}$$

As shown in Section C.2, we have that

$$\mathrm{Var}(T) \lesssim \Theta_n = \Theta_{n1} + \sum_{t=2}^{4} \Theta_{nt}. \tag{C.49}$$

Applying (C.17) to the first term and Lemma A.8 to the remaining terms, we have

$$\begin{aligned} \mathrm{Var}(T) &\lesssim \max\{\|\eta\|_\infty, \|\theta\|_\infty\} \cdot \rho^2 + \left\| \frac{m\bar{M}}{n\bar{N} + m\bar{M}} \eta + \frac{n\bar{N}}{n\bar{N} + m\bar{M}} \theta \right\|^2 \\ &\lesssim \max\{\|\eta\|, \|\theta\|\} \cdot \rho^2 + \max\{\|\eta\|^2, \|\theta\|^2\} \end{aligned} \tag{C.50}$$

Next, note that

$$\begin{aligned} \rho^2 &= n\bar{N}\|\eta - \mu\|^2 + m\bar{M}\|\theta - \mu\|^2 \\ &= n\bar{N} \left\| \eta - (\frac{n\bar{N}}{n\bar{N} + m\bar{M}} \eta + \frac{m\bar{M}}{n\bar{N} + m\bar{M}} \theta) \right\|^2 \\ &\quad + m\bar{M} \left\| \theta - (\frac{n\bar{N}}{n\bar{N} + m\bar{M}} \eta + \frac{m\bar{M}}{n\bar{N} + m\bar{M}} \theta) \right\|^2 \\ &= n\bar{N} \cdot (\frac{m\bar{M}}{n\bar{N} + m\bar{M}})^2 \|\eta - \theta\|^2 + m\bar{M} \cdot (\frac{n\bar{N}}{n\bar{N} + m\bar{M}})^2 \|\eta - \theta\|^2 \\ &= \frac{n\bar{N} m\bar{M}}{(n\bar{N} + m\bar{M})} \|\eta - \theta\|^2 = (\frac{1}{n\bar{N}} + \frac{1}{m\bar{M}})^{-1} \|\eta - \theta\|^2. \end{aligned} \tag{C.51}$$

By (C.48), (C.50), and (C.51), we have

$$\begin{aligned} \frac{(\mathbb{E}T)^2}{\mathrm{Var}(T)} &\gtrsim \frac{\rho^4}{\max\{\|\eta\|, \|\theta\|\} \cdot \rho^2 + \max\{\|\eta\|^2, \|\theta\|^2\}} \\ &\gtrsim \frac{\|\eta - \theta\|^2}{(\frac{1}{n\bar{N}} + \frac{1}{m\bar{M}}) \max\{\|\eta\|, \|\theta\|\}} + \Big(\frac{\|\eta - \theta\|^2}{(\frac{1}{n\bar{N}} + \frac{1}{m\bar{M}}) \max\{\|\eta\|, \|\theta\|\}}\Big)^2 \to \infty, \end{aligned}$$

which proves (C.47) and thus (C.44).

To prove (C.45), we mimick the Markov argument in (C.21) and use that under our assumptions, $\mathrm{Var}(V)/(\mathbb{E}V)^2 = o(1)$. We omit the details as they are similar. Since we have established (C.44), (C.45), and (C.46), the proof is complete. □

C.6 Proof of Theorem 3.7

Note that $T/\sqrt{\mathrm{Var}(T)} \Rightarrow N(0,1)$ by our assumptions and Proposition B.1. In particular, using that $n \to \infty$ and the monotonicity of the ℓ_p norms we have

$$\frac{\|\mu\|_4^4}{K\|\mu\|^4} = \frac{\|\mu\|_4^4}{n\|\mu\|^4} \leq \frac{1}{n} \cdot \frac{\|\mu\|^4}{\|\mu\|^4} = \frac{1}{n} \to 0.$$

Moreover, $V^*/\mathrm{Var}(T) \to 1$ in probability by Proposition A.5. It follows by Slutsky's theorem that $\psi^* = T/\sqrt{V^*} \Rightarrow N(0,1)$ and that the level-α DELVE test has an asymptotic level α.

To conclude the proof, it suffices to show that $\psi^* \to \infty$ under the alternative. As in the proof of Theorem 3.4, this follows immediately if we can show

$$T/\sqrt{\mathrm{Var}(T)} \to \infty, \tag{C.52}$$

$$V^* > 0 \text{ with high probability, and} \tag{C.53}$$

$$V^* = O_{\mathbb{P}}(\mathrm{Var}(T)). \tag{C.54}$$

Note that (C.52) follows from (C.14), and (C.54) is the content of Proposition A.6. Since our assumptions imply that $\mathbb{E}V^* \gg \sqrt{\mathrm{Var}(V^*)}$, (C.53) follows by a Markov argument as in (C.21).

□

C.7 Proof of Theorem 3.8

We apply Theorem 3.2 to get the asymptotic null distribution. Since $N_i = N$ and $\mu = p^{-1}\mathbf{1}_p$, it is easy to see that Condition 3.2 is satisfied under our assumption of $p = o(N^2 n)$. Therefore, by Theorem 3.2, $\psi^* \to N(0,1)$ under H_0.

We now show the asymptotic alternative distribution. By direct calculations and using $\sum_{i=1}^n \delta_{ij} = 0$ and $\sum_{j=1}^p \delta_{ij} = 0$, we have

$$\sum_{i,j} N_i(\Omega_{ij} - \mu_j)^2 = \frac{nN\beta_n^2}{p}, \qquad \sum_{i,j} N_i(\Omega_{ij} - \mu_j)^2\Omega_{ij} = \frac{nN\beta_n^2}{p^2}, \qquad \sum_i \|\Omega_i\|^2 = \frac{n(1+\beta_n^2)}{p}.$$

We apply Lemmas A.1-A.5 and plug in the above expressions. Let $S = \mathbf{1}_p' U_2$. It follows that

$$T = \frac{nN\beta_n^2}{p} + S + O_{\mathbb{P}}\left(\frac{\sqrt{nN}\beta_n}{p} + \frac{1}{\sqrt{p}}\right), \qquad \text{where } \mathrm{Var}(S) = 2p^{-1}n[1+o(1)]. \tag{C.55}$$

First, we plug in $\beta_n^2 = a\sqrt{2p}/(N\sqrt{n})$. It gives $p^{-1}nN\beta_n^2 = \sqrt{2n/p}$. Second, $p^{-1}\sqrt{nN}\beta_n \asymp (np)^{-1/4}\sqrt{n/p} = o(\sqrt{n/p})$. It follows that

$$T = a\sqrt{2n/p} + S + o_{\mathbb{P}}\big(\sqrt{n/p}\big), \qquad \text{where } \mathrm{Var}(S) = (2n/p)[1+o(1)]. \tag{C.56}$$

Recall the martingale decomposition $S = \sum_{(\ell,s)} E_{\ell,s}$ where $E_{\ell,s}$ is defined in (B.4). Observe that Lemmas B.4 and B.5 hold (even under the alternative). Define $\widetilde{E}_{\ell,s} =$

$E_{\ell,s}/\sqrt{\mathrm{Var}(S)}$. Using $\mathrm{Var}(S) \gtrsim n\sum_i \|\Omega_i\|^2$ and these lemmas, it is straightforward to verify that the following conditions hold:

$$\sum_{(\ell,s)} \mathrm{Var}\big(\widetilde{E}_{\ell,s}\big|\mathcal{F}_{\prec(\ell,s)}\big) \xrightarrow{\mathbb{P}} 1 \tag{C.57}$$

$$\sum_{(\ell,s)} \mathbb{E}\widetilde{E}_{\ell,s}^4 \xrightarrow{\mathbb{P}} 0. \tag{C.58}$$

As in Section B.1, the martingale CLT applies and we have

$$S/\sqrt{\mathrm{Var}(S)} \Rightarrow N(0,1).$$

By C.55,

$$T/\sqrt{\mathrm{Var}(S)} \quad \to \quad N(a,1). \tag{C.59}$$

By Lemma A.3 and (A.84),

$$\mathrm{Var}(S) = [1+o(1)]\Theta_{n2} = [1+o(1)]\mathrm{Var}(T)$$

By Proposition A.6, we have that $V^*/\mathrm{Var}(T) \to 1$ in probability. As a result,

$$V^*/\mathrm{Var}(S) \quad \to \quad 1, \qquad \text{in probability.} \tag{C.60}$$

We combine (C.59) and (C.60) to conclude that $\psi = T/\sqrt{V^*} \to N(a,1)$.

□

D Proofs of the corollaries for text analysis

D.1 Proof of Corollary 4.1

Note that Corollary 4.1 follows immediately from the slightly more general result stated below.

Corollary D.1. *Consider Model* (1.1) *and suppose that* $\Omega = \mu\mathbf{1}_n'$ *under the null hypothesis and that* Ω *satisfies* (4.1) *under the alternative hypothesis. Define* $\xi \in \mathbb{R}^n$ *by* $\xi_i = \bar{N}^{-1}N_i$ *and let* $\widetilde{\Omega} = \Omega[\mathrm{diag}(\xi)]^{1/2}$. *Let* $\lambda_1,\ldots,\lambda_M > 0$ *and* $\widetilde{\lambda}_1,\ldots,\widetilde{\lambda}_M > 0$ *denote the singular values of* Ω *and* $\widetilde{\Omega}$*, respectively, arranged in decreasing order. We further assume that under the alternative hypothesis,*

$$\frac{\bar{N}\cdot\sum_{k=2}^M \widetilde{\lambda}_k^2}{\sqrt{\sum_{k=1}^M \lambda_k^2}} \to \infty. \tag{D.1}$$

For any fixed $\alpha \in (0,1)$*, the level-*α *DELVE test has an asymptotic level* α *and an asymptotic power* 1. *Moreover if* $N_i \asymp \bar{N}$ *for all* i*, we may replace* $\sum_{k=2}^M \widetilde{\lambda}_k^2$ *with* $\sum_{k=2}^M \lambda_k^2$ *in the numerator of* (D.1).

Proof of Corollary D.1. This is a special case of our testing problem with $K = n$. Moreover, $\mu = n^{-1}\Omega\xi$ matches with the definition of μ in (1.2). Therefore, we can apply Theorem 3.7 directly. It remains to verify that the condition

$$\frac{\bar{N} \cdot \sum_{k=2}^{M} \widetilde{\lambda}_k^2}{\sqrt{\sum_{k=1}^{M} \lambda_k^2}} \to \infty \tag{D.2}$$

is sufficient to lead to the condition

$$\frac{n\bar{N}\|\mu\|^2\omega_n^2}{\sqrt{\sum_i \|\Omega_i\|^2}} \to \infty. \tag{D.3}$$

If we show this then Theorem 3.7 applies directly. We first calculate ω_n^2. Recall $\xi_i = N_i/\bar{N}$ for $1 \leq i \leq n$. Write

$$\widetilde{\Omega} = \Omega[\mathrm{diag}(\xi)]^{1/2}, \qquad \widetilde{\xi} = [\mathrm{diag}(\xi)]^{1/2}\mathbf{1}_n.$$

For $K = n$, by (3.13), $\omega_n^2 = \frac{1}{n\bar{N}\|\mu\|^2}\sum_{i=1}^n N_i\|\Omega_i - \mu\|^2$. It follows that

$$\omega_n^2 = \frac{1}{n\|\mu\|^2}\left\|(\Omega - \mu\mathbf{1}_n')[\mathrm{diag}(\xi)]^{1/2}\right\|_F^2 = \frac{1}{n\|\mu\|^2}\|\widetilde{\Omega} - \mu\widetilde{\xi}'\|_F^2. \tag{D.4}$$

Recall that $\widetilde{\lambda_1}, \ldots, \widetilde{\lambda}_M$ are the singular values of $\widetilde{\Omega}$. We apply a well-known result in linear algebra [Horn and Johnson, 1985], namely Weyl's inequality: For any rank-1 matrix Δ, $\|\widetilde{\Omega} - \Delta\|_F^2 \geq \sum_{k\neq 1}\widetilde{\lambda}_k^2$. In (D.4), $\mu\widetilde{\xi}'$ is a rank-1 matrix. It follows that

$$\|\widetilde{\Omega} - \mu\widetilde{\xi}'\|_F^2 \geq \sum_{k=2}^{M}\widetilde{\lambda}_k^2. \tag{D.5}$$

Hence

$$\frac{n\bar{N}\|\mu\|^2\omega_n^2}{\sqrt{\sum_i\|\Omega_i\|^2}} \geq \frac{\bar{N}\cdot\sum_{k=2}^M\widetilde{\lambda}_k^2}{\|\Omega\|_F} = \frac{\bar{N}\cdot\sum_{k=2}^M\widetilde{\lambda}_k^2}{\sqrt{\sum_{k=1}^M\lambda_k^2}},$$

which implies (D.3) by our assumption. The first claim is proved.

Next we prove the second claim. Observe that if $N_i \asymp \bar{N}$, then by Weyl's inequality:

$$\begin{aligned}\omega_n^2 &= \frac{1}{\|\mu\|^2 n\bar{N}}\sum_i N_i\|\Omega_i - \mu|^2 \gtrsim \frac{1}{\|\mu\|^2}\sum_i\|\Omega_i - \mu\|^2 \\ &= \frac{1}{\|\mu\|^2}\|\Omega - \mu\mathbf{1}_n'\|_F^2 \geq \frac{1}{\|\mu\|^2}\sum_{k=2}^M\lambda_k^2.\end{aligned}$$

Thus

$$\frac{n\bar{N}\|\mu\|^2\omega_n^2}{\sqrt{\sum_i\|\Omega_i\|^2}} \geq \frac{\bar{N}\cdot\sum_{k=2}^M\lambda_k^2}{\|\Omega\|_F} = \frac{\bar{N}\cdot\sum_{k=2}^M\lambda_k^2}{\sqrt{\sum_{k=1}^M\lambda_k^2}}.$$

We see that the assumption

$$\frac{\bar{N}\cdot\sum_{k=2}^M\lambda_k^2}{\sqrt{\sum_{k=1}^M\lambda_k^2}} \to \infty \tag{D.6}$$

implies (D.3). The second claim is established and the proof is complete. $\square$

D.2 Proof of Corollary 4.2

Recall the construction of a simple null and simple (random) alternative model from Section C.4.2, specialized below to the case of $K = n$ and $N_i \equiv N$:

$$H_0: \qquad \Omega_i = \tilde{\mu}, \qquad 1 \leq i \leq n. \tag{D.7}$$

$$H_1: \qquad \Omega_{ij} = \begin{cases} \mu_j\big(1 + \omega_n z_i b_j\big), & \text{if } 1 \leq j \leq m \\ \tilde{\mu}_j\big(1 - \omega_n z_i b_{j-m}\big), & \text{if } m+1 \leq j \leq 2m \end{cases} \tag{D.8}$$

where $b_1, \ldots, b_m$ are i.i.d. Rademacher random variables and $z_1, \ldots, z_n$ are i.i.d Rademacher random variables conditioned to satisfy $|\sum_i z_i| \leq 100\sqrt{n}$. Define

$$\tilde{b} = (b_1, \ldots, b_m, b_1, \ldots, b_m)'.$$

To derive the lower bound of Corollary 4.2, we assume without loss of generality that ω_n is a sufficiently small absolute constant.

We claim that H_1 prescribes a topic model with $M = 2$ topics. To see this, under the alternative,

$$\Omega_i = \begin{cases} \mu \circ (\mathbf{1}_p + \omega_n \tilde{b}) & \text{if } z_i = 1 \\ \mu \circ (\mathbf{1}_p - \omega_n \tilde{b}) & \text{if } z_i = -1. \end{cases} \tag{D.9}$$

Moreover, we showed in Section C.4.2 that $\Omega_{ij} \geq 0$ for all i, j and that $\|\Omega_{ij}\|_1 = 1$. From (D.9), we see that $\Omega = AW$ where $A \in \mathbb{R}^{p \times 2}$ and $W \in \mathbb{R}^{2 \times n}$ are defined as follows:

$$A_{:1} = \mu \circ (\mathbf{1}_p + \omega_n \tilde{b}), \quad A_{:2} = \mu \circ (\mathbf{1}_p - \omega_n \tilde{b})$$
$$W_{:i} = \begin{cases} (1,0)' & \text{if } z_i = 1 \\ (0,1)' & \text{if } z_i = -1. \end{cases}$$

Moreover, under the null hypothesis, Ω clearly prescribes a topic model with $K = 1$. Therefore Ω follows the topic model (4.1). Moreover, since $N_i \equiv N$, we have $\Omega[\mathrm{diag}(\xi)]^{1/2} = \Omega$.

By Proposition C.1 specialized to our setting, we know that the χ^2 distance between the null and alternative goes to zero if

$$\sqrt{n} N \|\mu\| \omega_n^2 \to 0.$$

Thus to prove Corollary 4.2 it suffices to show that

$$\frac{N \sum_{k \geq 2}^M \lambda_k^2}{\sqrt{\sum_{k=1}^M \lambda_k^2}} = \frac{N \lambda_2^2}{\sqrt{\sum_{k=1}^M \lambda_k^2}} \gtrsim \sqrt{n} N \|\mu\| \omega_n^2 \tag{D.10}$$

Accordingly we study the second largest singular value of Ω. First we have some preliminary calculations. Let $U = \{i : z_i = 1\}$, and let $V = \{i : z_i = -1\}$. Define

$$u = \mu \circ (\mathbf{1}_p + \omega_n \tilde{b}), \quad \text{and}$$

$$v = \mu \circ (\mathbf{1}_p - \omega_n \tilde{b}).$$

Observe that

$$\langle u, v\rangle = \|\mu\|^2 - \omega_n^2 \|\mu \circ \tilde{b}\|^2 = \|\mu\|^2(1-\omega_n^2).$$

Also, since ω_n is a sufficiently small absolute constant,

$$\|u\|^2 = \|\mu\|^2 + 2\omega_n \langle \mu, \mu \circ \tilde{b}\rangle + \omega_n^2 \|\mu \circ \tilde{b}\|^2 = (1+\omega_n^2)\|\mu\|^2 + 2\omega_n \sum_j \mu_j^2 \tilde{b}_j \gtrsim \|\mu\|^2, \quad \text{and}$$

$$\|v\|^2 = \|\mu\|^2 - 2\omega_n \langle \mu, \mu \circ \tilde{b}\rangle + \omega_n^2 \|\mu \circ \tilde{b}\|^2 = (1+\omega_n^2)\|\mu\|^2 - 2\omega_n \sum_j \mu_j^2 \tilde{b}_j \gtrsim \|\mu\|^2. \quad \text{(D.11)}$$

Again, since we assume that ω_n is a sufficiently small absolute constant,

$$\begin{aligned}\delta^2 := \frac{\langle u, v\rangle^2}{\|u\|^2\|v\|^2} &= \frac{\|\mu\|^4(1-\omega_n^2)^2}{(1+\omega_n^2)^2\|\mu\|^4 - 4\omega_n^2\langle \mu, \mu\circ b\rangle^2} \le \frac{\|\mu\|^4(1-\omega_n^2)^2}{(1+\omega_n^2)^2\|\mu\|^4 - 4\omega_n^2\|\mu\|^4} \\ &= \frac{\|\mu\|^4(1-\omega_n^2)^2}{\|\mu\|^4(1+2\omega_n^2-3\omega_n^4)} = \frac{(1-\omega_n^2)^2}{1+2\omega_n^2-3\omega_n^4} \qquad \text{(D.12)}\end{aligned}$$

Note that

$$\begin{aligned}\|au+bv\|^2 &= a^2\|u\|^2 + 2ab\langle u, v\rangle + b^2\|v\|^2 \ge a^2\|u\|^2 + b^2\|v\|^2 - 2ab\delta\|u\|\|v\| \\ &\ge (1-\delta)\big(a^2\|u\|^2 + b^2\|v\|^2\big) + \|au-bv\|^2 \ge (1-\delta)\big(a^2\|u\|^2+b^2\|v\|^2\big).\end{aligned}$$

By (D.12), we have for ω_n sufficiently small that

$$\begin{aligned}1-\delta &\ge 1 - \frac{1-\omega_n^2}{\sqrt{1+2\omega_n^2-3\omega_n^4}} = \frac{\sqrt{1+2\omega_n^2-3\omega_n^4} - 1 + \omega_n^2}{\sqrt{1+2\omega_n^2-3\omega_n^4}} \\ &\ge \frac{\omega_n^2}{\sqrt{1+2\omega_n^2-3\omega_n^4}} \gtrsim \omega_n^2.\end{aligned}$$

Thus

$$\|au+bv\|^2 \ge \omega_n^2(a^2\|u\|^2 + b^2\|v\|^2) \gtrsim \omega_n^2\|\mu\|^2(a^2+b^2) \qquad \text{(D.13)}$$

Recall that if M is a rank k matrix, then

$$\lambda_k(M) = \sup_{y:\|y\|=1,\, y\in \mathrm{Ker}(M)^\perp} \|My\| = \sup_{y:\|y\|=1,\, y\in \mathrm{Im}(M')} \|My\|. \qquad \text{(D.14)}$$

We have

$$\Omega\Omega' = \sum_{i\in U} uu' + \sum_{i\in V} vv' = |U|uu' + |V|vv'.$$

Let $y \in \mathbb{R}^n$ satisfy $\|y\| = 1$ and $y = \Omega' x$ for some x. We have

$$\Omega y = \Omega\Omega' x = |U|\langle u, x\rangle u + |V|\langle v, x\rangle v.$$

By the previous equation and (D.13),

$$\|\Omega y\|^2 = \|\Omega\Omega' x\|^2 = \Big\| |U|\langle u, x\rangle u + |V|\langle v, x\rangle v \Big\|^2 \gtrsim \omega_n^2 \|\mu\|^2 \big(|U|^2\langle u, x\rangle^2 + |V|^2\langle v, x\rangle^2\big).$$

By our conditioning on z, we have $\min(|U|, |V|) \gtrsim n$. Moreover

$$1 = \|y\|^2 = \|\Omega' x\|^2 = |U|\langle u, x\rangle^2 + |V|\langle v, x\rangle^2.$$

Applying these facts and (D.14), we obtain

$$\lambda_2^2 \geq \|\Omega y\|^2 = \|\Omega\Omega' x\|^2 \gtrsim \omega_n^2 \|\mu\|^2 n \big(|U|\langle u, x\rangle^2 + |V|\langle v, x\rangle^2\big) = \omega_n^2 \|\mu\|^2 n.$$

Next,

$$\sum_{k=1}^{M} \lambda_k^2 = \|\Omega\|_F^2 = \sum_{i\in U} \|u\|^2 + \sum_{i \in V} \|v\|^2 = |U| \cdot \|u\|^2 + |V| \cdot \|v\|^2 \asymp n\|\mu\|^2 \tag{D.15}$$

We conclude that

$$\frac{N \sum_{k\geq 2}^{M} \lambda_k^2}{\sqrt{\sum_{k=1}^{M} \lambda_k^2}} = \frac{N\lambda_2^2}{\sqrt{\sum_{k=1}^{M} \lambda_k^2}} \gtrsim \frac{N \cdot \omega_n^2 \|\mu\|^2 n}{\sqrt{n}\|\mu\|} = \sqrt{n} N \|\mu\| \omega_n^2$$

which establishes (D.10). The proof is complete. □

D.3 Proof of Corollary 4.3

This is a special case of our testing problem with $K = 2$, we can apply Theorem 3.6 directly. It remains to verify that the condition

$$\frac{\beta_n^2 \cdot (\|\eta_S\|_1 + \|\theta_S\|_1)}{\big(\frac{1}{n\bar{N}} + \frac{1}{m\bar{M}}\big) \max\{\|\eta\|, \|\theta\|\}} \to \infty \tag{D.16}$$

is sufficient to yield the condition (3.11) in Theorem 3.6. This is done by calculating $\|\eta - \theta\|^2$ directly. By our sparse model (4.4), for $j \in S$, $|\sqrt{\eta_j} - \sqrt{\theta_j}| \geq \beta_n$. It follows that for $j \in S$,

$$|\eta_j - \theta_j|^2 = (\sqrt{\eta_j} + \sqrt{\theta_j})^2 (\sqrt{\eta_j} - \sqrt{\theta_j})^2 \geq \beta_n^2 (\sqrt{\eta_j} + \sqrt{\theta_j})^2 \geq \beta_n^2 (\eta_j + \theta_j).$$

It follows that

$$\|\eta - \theta\|^2 \geq \beta_n^2 \sum_{j \in S} (\eta_j + \theta_j) \geq \beta_n^2 \big(\|\eta_S\|_1 + \|\theta_S\|_1\big). \tag{D.17}$$

We plug it into (3.11) and see immediately that (D.16) implies this condition. The claim follows directly from Theorem 3.6. □

E A modification of DELVE for finite p

Below we write out the variance of the terms of the raw DELVE statistic under the null, using the proofs of Lemmas A.3–A.5.

$$\mathrm{Var}(\mathbf{1}_p' U_2) = 2\sum_{k=1}^{K}\sum_{i\in S_k}\sum_{1\le r<s\le N_i}(\frac{1}{n_k\bar{N}_k}-\frac{1}{n\bar{N}})^2\frac{N_i^2}{(N_i-1)^2}\left[\|\Omega_i\|^2-2\|\Omega_i\|_3^3+\|\Omega_i\|^4\right] \quad \text{(E.1)}$$

$$\mathrm{Var}(\mathbf{1}_p' U_3) = \frac{2}{n^2\bar{N}^2}\sum_{k\neq\ell}\sum_{i\in S_k}\sum_{m\in S_\ell} N_iN_m\Big(\sum_j \Omega_{ij}\Omega_{mj}-2\sum_j\Omega_{ij}^2\Omega_{mj}^2+\sum_{j,j'}\Omega_{ij}\Omega_{ij'}\Omega_{mj}\Omega_{mj'}\Big)$$

$$\mathrm{Var}(\mathbf{1}_p' U_4) = 2\sum_{k=1}^{K}\sum_{\substack{i\in S_k,m\in S_k\\ i\neq m}}(\frac{1}{n_k\bar{N}_k}-\frac{1}{n\bar{N}})^2 N_iN_m\Big(\sum_j \Omega_{ij}\Omega_{mj}-2\sum_j\Omega_{ij}^2\Omega_{mj}^2+\sum_{j,j'}\Omega_{ij}\Omega_{ij'}\Omega_{mj}\Omega_{mj'}\Big).$$

In this section we develop an unbiased estimator for each term above, which leads to an unbiased estimator of $\mathrm{Var}(T)$ by taking their sum. We require some preliminary results proved later in this section. Recall that Lemma E.2 was established in the proof of Lemma A.1.

Lemma E.1. *If $j\neq j'$, an unbiased estimator of $\Omega_{ij}\Omega_{ij'}$ is*

$$\widehat{\Omega_{ij}\Omega_{ij'}} := \frac{X_{ij}X_{ij'}}{N_i(N_i-1)}$$

Lemma E.2. *An unbiased estimator of Ω_{ij}^2 is*

$$\widehat{\Omega_{ij}^2} := \frac{X_{ij}^2 - X_{ij}}{N_i(N_i-1)}. \quad \text{(E.2)}$$

Lemma E.3. *If $j\neq j'$, an unbiased estimator for $\Omega_{ij}^2\Omega_{ij'}^2$ is*

$$\widehat{\Omega_{ij}^2\Omega_{ij'}^2} = \frac{(X_{ij}^2-X_{ij})(X_{ij'}^2-X_{ij'})}{N_i(N_i-1)(N_i-2)(N_i-3)}$$

Lemma E.4. *An unbiased estimator of Ω_{ij}^3 is*

$$\widehat{\Omega_{ij}^3} := \frac{X_{ij}^3-3X_{ij}^2+2X_{ij}}{N_i(N_i-1)(N_i-2)}. \quad \text{(E.3)}$$

Lemma E.5. *An unbiased estimator of Ω_{ij}^4 is*

$$\widehat{\Omega_{ij}^4} := \frac{X_{ij}^4-3X_{ij}^3-X_{ij}^2+3X_{ij}}{N_i(N_i-1)(N_i-2)(N_i-3)}. \quad \text{(E.4)}$$

Define

$$\widehat{\|\Omega_i\|^2} := \sum_j \widehat{\Omega_{ij}^2}$$

$$\widehat{\|\Omega_i\|_3^3} := \sum_j \widehat{\Omega_{ij}^3}$$

$$\widehat{\|\Omega_i\|^4} := \sum_j \widehat{\Omega_{ij}^4} + \sum_{j \neq j'} \widehat{\Omega_{ij}^2 \Omega_{ij'}^2}. \tag{E.5}$$

Using Lemmas E.1–E.5 and (E.5), we define an unbiased estimator for each term of (E.1). Let $\widehat{\Omega_{ij}} = X_{ij}/N_i$ and define

$$\widehat{\mathrm{Var}(\mathbf{1}_p' U_2)} = 2\sum_{k=1}^{K} \sum_{i \in S_k} \sum_{1 \le r < s \le N_i} (\frac{1}{n_k \bar{N}_k} - \frac{1}{n\bar{N}})^2 \frac{N_i^2}{(N_i-1)^2} \big[\widehat{\|\Omega_i\|^2} - 2\widehat{\|\Omega_i\|_3^3} + \widehat{\|\Omega_i\|^4}\big] \tag{E.6}$$

$$\widehat{\mathrm{Var}(\mathbf{1}_p' U_3)} = \frac{2}{n^2 \bar{N}^2} \sum_{k \neq \ell} \sum_{i \in S_k} \sum_{m \in S_\ell} N_i N_m \Big(\sum_j \widehat{\Omega_{ij}} \widehat{\Omega_{mj}} - 2\sum_j \widehat{\Omega_{ij}^2} \widehat{\Omega_{mj}^2} + \sum_{j,j'} \widehat{\Omega_{ij}\Omega_{ij'}} \widehat{\Omega_{mj}\Omega_{mj'}} \Big)$$

$$\widehat{\mathrm{Var}(\mathbf{1}_p' U_4)} = 2\sum_{k=1}^{K} \sum_{\substack{i \in S_k, m \in S_k \\ i \neq m}} (\frac{1}{n_k \bar{N}_k} - \frac{1}{n\bar{N}})^2 N_i N_m \Big(\sum_j \widehat{\Omega_{ij}} \widehat{\Omega_{mj}} - 2\sum_j \widehat{\Omega_{ij}^2} \widehat{\Omega_{mj}^2} + \sum_{j,j'} \widehat{\Omega_{ij}\Omega_{ij'}} \widehat{\Omega_{mj}\Omega_{mj'}} \Big).$$

Define

$$\widetilde{V} = \widehat{\mathrm{Var}(\mathbf{1}_p' U_2)} + \widehat{\mathrm{Var}(\mathbf{1}_p' U_3)} + \widehat{\mathrm{Var}(\mathbf{1}_p' U_4)}. \tag{E.7}$$

We define *exact DELVE* as $\tilde{\psi} = T/\widetilde{V}^{1/2}$. Combining our results above, we obtain the following.

Proposition E.1. *Consider the statistic $\widetilde{V}$ defined in* (E.7). *Under the null hypothesis, $\widetilde{V}$ is an unbiased estimator for* $\mathrm{Var}(T)$.

With this result in hand, it is possible to derive consistency of $\widetilde{V}$ as an estimator of $\mathrm{Var}(T)$ under certain regularity conditions. We omit the details.

E.1 Proof of Lemma E.1

Recall that B_{ijr} is the Bernoulli random variable $B_{ijr} = Z_{ijr} + \Omega_{ij}$ and satisfies $X_{ijr} = \sum_{r=1}^{N_i} B_{ijr}$. Observe that

$$X_{ij} X_{ij'} = \sum_{r,s} B_{ijr} B_{ij's} = \sum_r B_{ijr} B_{ij'r} + \sum_{r \neq s} B_{ijr} B_{ij's} = 0 + \sum_{r \neq s} B_{ijr} B_{ij's}$$

Thus

$$\mathbb{E} X_{ij} X_{ij'} = N_i (N_i - 1) \Omega_{ij} \Omega_{ij'},$$

and we obtain

$$\widehat{\Omega_{ij}\Omega_{ij'}} = \frac{X_{ij} X_{ij'}}{N_i (N_i - 1)}$$

is an unbiased estimator for $\Omega_{ij}\Omega_{ij'}$, as desired. $\square$

E.2 Proof of Lemma E.3

Note that

$$\begin{aligned}
X_{ij}^2 X_{ij'}^2 &= \big(\sum_r B_{ijr} + \sum_{r\neq s} B_{ijr}B_{ijs}\big)\big(\sum_r B_{ij'r} + \sum_{r\neq s} B_{ij'r}B_{ij's}\big) \\
&= \sum_r B_{ijr}B_{ij'r} + \sum_{r_1\neq r_2} B_{ijr}B_{ij's} + \sum_{r_1\neq s} B_{ijr_1}B_{ijs}\sum_{r_2} B_{ij'r_2} + \sum_{r_1\neq s} B_{ij'r_1}B_{ij's}\sum_{r_2} B_{ijr_2} \\
&\quad + \big(\sum_{r\neq s} B_{ijr}B_{ijs}\big)\big(\sum_{r\neq s} B_{ij'r}B_{ij's}\big) \\
&= \sum_{r_1\neq r_2} B_{ijr}B_{ij's} + \sum_{r_1\neq s} B_{ijr_1}B_{ijs}\sum_{r_2} B_{ij'r_2} + \sum_{r_1\neq s} B_{ij'r_1}B_{ij's}\sum_{r_2} B_{ijr_2} \\
&\quad + \big(\sum_{r\neq s} B_{ijr}B_{ijs}\big)\big(\sum_{r\neq s} B_{ij'r}B_{ij's}\big)
\end{aligned}$$

Since $B_{ijr}B_{ij'r} = 0$, note that

$$\begin{aligned}
(X_{ij}^2 - X_{ij})(X_{ij'}^2 - X_{ij'}) &= \sum_{r_1\neq s_1}\sum_{r_2\neq s_2} B_{ijr_1}B_{ijs_1}B_{ij'r_2}B_{ij's_2} \\
&= \sum_{r_1,s_1,r_2,s_2\, dist.} B_{ijr_1}B_{ijs_1}B_{ij'r_2}B_{ij's_2}.
\end{aligned}$$

Thus

$$\begin{aligned}
\mathbb{E}(X_{ij}^2 - X_{ij})(X_{ij'}^2 - X_{ij'}) &= \sum_{r_1,s_1,r_2,s_2\, dist.} \mathbb{E}\big[B_{ijr_1}B_{ijs_1}B_{ij'r_2}B_{ij's_2}\big] \\
&= N_i(N_i-1)(N_i-2)(N_i-3)\cdot\Omega_{ij}^2\Omega_{ij'}^2.
\end{aligned}$$

It follows that

$$\widehat{\Omega_{ij}^2\Omega_{ij'}^2} = \frac{(X_{ij}^2 - X_{ij})(X_{ij'}^2 - X_{ij'})}{N_i(N_i-1)(N_i-2)(N_i-3)}$$

is an unbiased estimator for $\Omega_{ij}^2\Omega_{ij'}^2$.

□

E.3 Proof of Lemma E.4

Recall that B_{ijr} is the Bernoulli random variable $B_{ijr} = Z_{ijr} + \Omega_{ij}$ and satisfies $X_{ijr} = \sum_{r=1}^{N_i} B_{ijr}$. Observe that

$$X_{ij}^3 = \sum_r B_{ijr} + 3\sum_{r_1\neq r_2} B_{ijr_1}B_{ijr_2} + \sum_{r_1\neq r_2\neq r_3} B_{ijr_1}B_{ijr_2}B_{ijr_3}.$$

Thus

$$\mathbb{E}X_{ij}^3 = N_i\Omega_{ij} + 3N_i(N_i-1)\Omega_{ij}^2 + N_i(N_i-1)(N_i-2)\Omega_{ij}^3.$$

Unbiased estimators for Ω_{ij} and Ω_{ij}^2 are

$$\begin{aligned}&\frac{X_{ij}}{N_i}\\ &\frac{X_{ij}^2}{N_i^2} - \frac{X_{ij}(N_i - X_{ij})}{N_i^2(N_i-1)} = \frac{1}{N_i(N_i-1)}\big(X_{ij}^2 - X_{ij}\big),\end{aligned}$$

respectively. Hence

$$X_{ij}^3 - X_{ij} - 3(X_{ij}^2 - X_{ij}) = X_{ij}^3 - 3X_{ij}^2 + 2X_{ij}$$

is an unbiased estimator for $N_i(N_i-1)(N_i-2)\Omega_{ij}^3$, as desired.

□

E.4 Proof of Lemma E.5

Observe that

$$\begin{aligned}X_{ij}^4 &= \sum_r B_{ijr}^4 + 4\sum_{r_1\neq r_2} B_{ijr_1}^3 B_{ijr_2} + 6\sum_{r_1\neq r_2} B_{ijr_1}^2 B_{ijr_2}^2\\ &\quad + 3\sum_{r_1\neq r_2\neq r_3} B_{ijr_1}^2 B_{ijr_2} B_{ijr_3} + \sum_{r_1\neq r_2\neq r_3\neq r_4} B_{ijr_1}B_{ijr_2}B_{ijr_3}B_{ijr_4}\\ &= \sum_r B_{ijr} + 10\sum_{r_1\neq r_2} B_{ijr_1}B_{ijr_2} + 3\sum_{r_1\neq r_2\neq r_3} B_{ijr_1}B_{ijr_2}B_{ijr_3}\\ &\quad + \sum_{r_1\neq r_2\neq r_3\neq r_4} B_{ijr_1}B_{ijr_2}B_{ijr_3}B_{ijr_4}.\end{aligned}$$

Thus

$$\begin{aligned}\mathbb{E}X_{ij}^4 &= N_i\Omega_{ij} + 10N_i(N_i-1)\Omega_{ij}^2 + 3N_i(N_i-1)(N_i-2)\Omega_{ij}^3\\ &\quad + N_i(N_i-1)(N_i-2)(N_i-3)\Omega_{ij}^4.\end{aligned}$$

Plugging in unbiased estimators for the first three terms, we have

$$X_{ij}^4 - X_{ij} - 10(X_{ij}^2 - X_{ij}) - 3(X_{ij}^3 - 3X_{ij}^2 + 2X_{ij}) = X_{ij}^4 - 3X_{ij}^3 - X_{ij}^2 + 3X_{ij}$$

is an unbiased estimator for $N_i(N_i-1)(N_i-2)(N_i-3)$, as desired. □

References

Ery Arias-Castro and Meng Wang. The sparse poisson means model. *Electronic Journal of Statistics*, 9(2):2170–2201, 2015.

Sivaraman Balakrishnan and Larry Wasserman. Hypothesis testing for high-dimensional multinomials: A selective review. *The Annals of Applied Statistics*, 12(2):727–749, 2018.

Sivaraman Balakrishnan and Larry Wasserman. Hypothesis testing for densities and high-dimensional multinomials: Sharp local minimax rates. *The Annals of Statistics*, 47(4): 1893–1927, 2019.

Bhaswar Bhattacharya and Gregory Valiant. Testing closeness with unequal sized samples. *Advances in Neural Information Processing Systems*, 28, 2015.

David Blei, Andrew Ng, and Michael Jordan. Latent dirichlet allocation. *Journal of Machine Learning Research*, 3:993–1022, 2003.

Siu-On Chan, Ilias Diakonikolas, Paul Valiant, and Gregory Valiant. Optimal algorithms for testing closeness of discrete distributions. In *Proc. 25th symposium on discrete algorithms*. SIAM, 2014.

Judith A Chevalier and Dina Mayzlin. The effect of word of mouth on sales: Online book reviews. *Journal of Marketing Research*, 43(3):345–354, 2006.

Ilias Diakonikolas and Daniel M Kane. A new approach for testing properties of discrete distributions. In *2016 IEEE 57th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 685–694. IEEE, 2016.

David Donoho and Jiashun Jin. Higher criticism for detecting sparse heterogeneous mixtures. *The Annals of Statistics*, pages 962–994, 2004.

David L Donoho and Alon Kipnis. Higher criticism to compare two large frequency tables, with sensitivity to possible rare and weak differences. *The Annals of Statistics*, 50(3): 1447–1472, 2022.

Wael H Gomaa, Aly A Fahmy, et al. A survey of text similarity approaches. *International Journal of Computer Applications*, 68(13):13–18, 2013.

Peter Hall and Christopher C Heyde. *Martingale limit theory and its application*. Academic press, 2014.

Stephen Hansen, Michael McMahon, and Andrea Prat. Transparency and deliberation within the fomc: a computational linguistics approach. *The Quarterly Journal of Economics*, 133(2):801–870, 2018.

Roger Horn and Charles Johnson. *Matrix Analysis*. Cambridge University Press, 1985.

Pengsheng Ji and Jiashun Jin. Coauthorship and citation networks for statisticians. *The Annals of Applied Statistics*, 10(4):1779–1812, 2016.

Jiashun Jin, Zheng Ke, and Shengming Luo. Network global testing by counting graphlets. In *International conference on machine learning*, pages 2333–2341. PMLR, 2018.

Zheng Tracy Ke and Minzhe Wang. Using SVD for topic modeling. *Journal of the American Statistical Association*, pages 1–16, 2022.

Ilmun Kim, Sivaraman Balakrishnan, and Larry Wasserman. Minimax optimality of permutation tests. *The Annals of Statistics*, 50(1):225–251, 2022.

Alon Kipnis. Higher criticism for discriminating word-frequency tables and authorship attribution. *The Annals of Applied Statistics*, 16(2):1236–1252, 2022.

Alon Kipnis and David L Donoho. Two-sample testing of discrete distributions under rare/weak perturbations. In *2021 IEEE Int'l Symposium on Information Theory*. IEEE, 2021.

Aleksandra A Kolodziejczyk, Jong Kyoung Kim, Valentine Svensson, John C Marioni, and Sarah A Teichmann. The technology and biology of single-cell RNA sequencing. *Molecular cell*, 58(4):610–620, 2015.

Xi Y Leung and Yang Yang. Are all five points equal? Scaling heterogeneity in hotel online ratings. *International Journal of Hospitality Management*, 88:102539, 2020.

Dharmendra Maurya. Web data: Amazon movie reviews. Electronic, 2018. https://www.kaggle.com/datasets/dm4006/amazon-movie-reviews/metadata.

Frederick Mosteller and David L Wallace. Inference in an authorship problem: A comparative study of discrimination methods applied to the authorship of the disputed federalist papers. *Journal of the American Statistical Association*, 58(302):275–309, 1963.

Frederick Mosteller and David L Wallace. *Applied Bayesian and classical inference: the case of the Federalist papers*. Springer Science & Business Media, 2012.

A.B. Tsybakov. *Introduction to Nonparametric Estimation*. Springer Series in Statistics. Springer New York, 2008. ISBN 9780387790527. URL `https://books.google.com/books?id=mwB8rUBsbqoC`.

Gregory Valiant and Paul Valiant. An automatic inequality prover and instance optimal identity testing. *SIAM Journal on Computing*, 46(1):429–455, 2017.

Feng Zhu and Xiaoquan Zhang. Impact of online consumer reviews on sales: The moderating role of product and consumer characteristics. *Journal of Marketing*, 74(2):133–148, 2010.