Degree Heterogeneity in Higher-Order Networks: Inference in the Hypergraph $\boldsymbol{\beta}$-Model

Sagnik Nandy and Bhaswar B. Bhattacharya

***Abstract*— The $\boldsymbol{\beta}$-model for random graphs is commonly used for representing pairwise interactions in a network with degree heterogeneity. Going beyond pairwise interactions, Stasi et al. (2014) introduced the hypergraph $\boldsymbol{\beta}$-model for capturing degree heterogeneity in networks with higher-order (multi-way) interactions. In this paper we initiate the rigorous study of the hypergraph $\boldsymbol{\beta}$-model with multiple layers, which allows for hyperedges of different sizes across the layers. To begin with, we derive the rates of convergence of the maximum likelihood (ML) estimates and establish their minimax rate optimality. We also derive the limiting distribution of the ML estimates and construct asymptotically valid confidence intervals for the model parameters. Next, we consider the goodness-of-fit problem in the hypergraph $\boldsymbol{\beta}$-model. Specifically, we establish the asymptotic normality of the likelihood ratio (LR) test under the null hypothesis, derive its detection threshold, and also its limiting power at the threshold. Interestingly, the detection threshold of the LR test turns out to be minimax optimal, that is, all tests are asymptotically powerless below this threshold. The theoretical results are further validated in numerical experiments. In addition to developing the theoretical framework for estimation and inference for hypergraph $\boldsymbol{\beta}$-models, the above results fill a number of gaps in the graph $\boldsymbol{\beta}$-model literature, such as the minimax optimality of the ML estimates and the non-null properties of the LR test, which, to the best of our knowledge, have not been studied before.**

***Index Terms*— Graphical models, maximum likelihood (ML) estimation, minimax techniques.**

I. Introduction

THE $\boldsymbol{\beta}$-model is an exponential family distribution on graphs with the degree sequence as the sufficient statistic. Specifically, in the $\boldsymbol{\beta}$-model with vertex set $[n] := \{1, 2, \ldots, n\}$, the edge (i, j) is present independently with probability

$$p_{ij} := \frac{e^{\beta_i + \beta_j}}{1 + e^{\beta_i + \beta_j}}, \tag{I.1}$$

for $1 \leqslant i < j \leqslant n$ and $\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots, \beta_n) \in \mathbb{R}^n$. This model was first considered by Park and Newman [41] and can also be viewed as the undirected version of the p_1-model that appears in the earlier work of Holland and Leinhardt [26]. Thereafter, the $\boldsymbol{\beta}$-model has been widely used for capturing degree heterogeneity in networks (see Blitzstein and Diaconis [7] Chen et al. [11] Graham [23] Jackson et al. [28], among several others). The term $\boldsymbol{\beta}$-model can be attributed to the seminal paper of Chatterjee et al. [10], which provides the theoretical foundations for parameter estimation in this model.

While random graph models, such as the $\boldsymbol{\beta}$-model, are important tools for understanding binary (pairwise) relational data, in many modern applications interactions occur not just between pairs, but among groups of agents. Examples include folksonomy [17], collaboration networks [29], [30], [42], complex ecosystems [24], biological networks [37], [43], circuit design [32], computer vision [1], among others. Hypergraphs provide the natural mathematical framework for modeling such higher-order interactions [4], [5], [6]. Formally, a hypergraph H is denoted by $H = (V(H), E(H))$, where $V(H)$ is the vertex set (the agents in the network) and $E(H)$ is a collection of non-empty subsets of $V(H)$ of cardinality greater than 1. The elements in $E(H)$, referred to as *hyperedges*, represent the interactions among groups of agents. Motivated by the emergence of complex relational data with higher-order structures, there has been a slew of recent results on modeling random hypergraphs, community detection, recovery, clustering, and motif analysis, among others (see [2], [3], [15], [18], [19], [20], [21], [22], [27], [33], [34], [35], [36], [40], [51], [57], [58], [59] and the references therein).

In this paper we study the hypergraph $\boldsymbol{\beta}$-model, introduced by Stasi et al. [45], that allows one to incorporate degree heterogeneity in networks with higher-order structures. Like the graph $\boldsymbol{\beta}$-model (I.1), this is an exponential family on hypergraphs where the (hypergraph) degrees (as defined in (II.3)) are the sufficient statistics. In its general form it allows for layered hypergraphs with hyperedges of different sizes across the layers. To describe the model formally we need a few notations: For $r \geqslant 2$, denote by $\binom{[n]}{r}$ the collection of all r-element subsets of $[n] := \{1, 2, \ldots, n\}$. A hypergraph $H = (V(H), E(H))$ is said to be r-uniform if every element in $E(H)$ has cardinality r. (Clearly, 2-uniform hypergraphs are simple graphs.) We will denote by $\mathcal{H}_{n,r}$ the collection of all r-uniform hypergraphs with vertex set $[n]$ and by $\mathcal{H}_{n,[r]} := \bigcup_{s=2}^{r} \mathcal{H}_{n,s}$, the collection of all hypergraphs with vertex set $[n]$ where every hyperedge has size at most r. Then the *r-layered hypergraph $\boldsymbol{\beta}$-model* is a probability distribution on $\mathcal{H}_{n,[r]}$ defined as follows:

Manuscript received 12 July 2023; revised 2 February 2024; accepted 23 May 2024. Date of publication 11 June 2024; date of current version 16 July 2024. The work of Bhaswar B. Bhattacharya was supported in part by NSF CAREER under Grant DMS 2046393, in part by NSF under Grant DMS 2113771, and in part by the Sloan Research Fellowship. *(Corresponding author: Sagnik Nandy.)*

Sagnik Nandy is with the Department of Statistics, University of Chicago, Chicago, IL 60637 USA (e-mail: nandysagniksxi@gmail.com).

Bhaswar B. Bhattacharya is with the Department of Statistics and Data Science, University of Pennsylvania, Philadelphia, PA 19104 USA.

Communicated by V. Y. F. Tan, Associate Editor for Machine Learning and Statistics.

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TIT.2024.3411523.

Digital Object Identifier 10.1109/TIT.2024.3411523

Definition 1: [45] Fix $r \geqslant 2$ and parameters $\boldsymbol{B} := (\boldsymbol{\beta}_2, \ldots, \boldsymbol{\beta}_r)$, where $\boldsymbol{\beta}_s := (\beta_{s,v})_{v \in [n]} \in \mathbb{R}^n$. The *$r$-layered hypergraph $\boldsymbol{\beta}$-model* is a random hypergraph in $\mathcal{H}_{n,[r]}$, denoted by $\mathsf{H}_{[r]}(n, \boldsymbol{B})$, where, for every $2 \leqslant s \leqslant r$, the hyperedge $\{v_1, v_2, \ldots, v_s\} \in \binom{[n]}{s}$ is present independently with probability:

$$p_{v_1, v_2, \ldots, v_s} := \frac{e^{\beta_{s,v_1} + \ldots + \beta_{s,v_s}}}{1 + e^{\beta_{s,v_1} + \ldots + \beta_{s,v_s}}}. \tag{I.2}$$

This model can be expressed as an exponential family on $\mathcal{H}_{n,[r]}$ with the hypergraph degrees up to order r as the sufficient statistics (see (II.2)). Specifically, the parameter $\beta_{s,u}$ encodes the popularity of the node $u \in [n]$ to form groups of size s, for $2 \leqslant s \leqslant r$. Consequently, $\beta_{s,u}$ controls the local density of hyperedges of size s around the around node u. The model (I.2) includes as a special case the classical graph $\boldsymbol{\beta}$-model (when $r = 2$) and also the r-uniform hypergraph $\boldsymbol{\beta}$-model, where only the hyperedges of size r are present. More formally, given parameters $\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots, \beta_n) \in \mathbb{R}^n$, the *$r$-uniform hypergraph $\boldsymbol{\beta}$-model* is a random hypergraph in $\mathcal{H}_{n,r}$, denoted by $\mathsf{H}_r(n, \boldsymbol{\beta})$, where each r-element hyperedge $\{v_1, v_2, \ldots, v_r\} \in \binom{[n]}{r}$ is present independently with probability:

$$p_{v_1, v_2, \ldots, v_r} := \frac{e^{\beta_{v_1} + \ldots + \beta_{v_r}}}{1 + e^{\beta_{v_1} + \ldots + \beta_{v_r}}}. \tag{I.3}$$

It is worth noting that, since the degrees (defined later in (II.3)) are the sufficient statistics in the aforementioned models, it is enough to observe only the degree sequences (not the entire network) for inference regarding the model parameters. This feature makes the $\boldsymbol{\beta}$-model particularly attractive because collecting information about the entire network can often be difficult for cost or privacy reasons. For example, Elmer et al. [16] (see also Zhang et al. [60]) studied social networks between a group of Swiss students before and during COVID-19 lockdown, where, for privacy reasons, only the total number of connections of each student in the network (that is, the degrees of the vertices) were released. The $\boldsymbol{\beta}$-model is also relevant in the analysis of aggregated relational data, where instead of asking about connections between groups of individuals directly, one collects data on the number of connections of an individual with a given feature (see, for example, Breza et al. [8] and the references therein).

Stasi et al. [45] proposed two algorithms for computing the maximum likelihood (ML) estimates for the hypergraph $\boldsymbol{\beta}$ models described above and reported their empirical performance. However, the statistical properties of the ML estimates in these models have remained unexplored.

A. Summary of Results

In this paper we develop a framework for estimation and inference in the hypergraph $\boldsymbol{\beta}$-model. Along the way, we obtain a number of new results on the graph $\boldsymbol{\beta}$-model as well. The following is a summary of the results:

- *Estimation*: In Section II we derive the rates of convergence of the ML estimates in r-layered hypergraph $\boldsymbol{\beta}$-model (I.2), both in the L_∞ and the L_2 norms.[1] Specifically, we show that given a sample $H_n \sim \mathsf{H}_{[r]}(n, \boldsymbol{B})$ from the r-layered hypergraph $\boldsymbol{\beta}$-model, the ML estimate $\hat{\boldsymbol{B}} = (\hat{\boldsymbol{\beta}}_2, \ldots, \hat{\boldsymbol{\beta}}_r)$ of $\boldsymbol{B}$ satisfies the following:
$$\begin{aligned} \|\hat{\boldsymbol{\beta}}_s - \boldsymbol{\beta}_s\|_2 &\leqslant C_s \sqrt{\frac{1}{n^{s-2}}}, \\ \|\hat{\boldsymbol{\beta}}_s - \boldsymbol{\beta}_s\|_\infty &\leqslant C_s \sqrt{\frac{\log n}{n^{s-1}}}, \end{aligned} \tag{I.4}$$
for some constant $C_s > 0$ (depending on s and $\|\boldsymbol{\beta}_s\|_\infty$), where $2 \leqslant s \leqslant r$, with probability going to 1 (see Theorem 1). These extend the results of Chatterjee et al. [10] on the graph $\boldsymbol{\beta}$-model, where the rate of convergence of the ML estimate was derived only in the L_∞ norm, to the hypergraph case. Next, in Theorem 2 we show that both the rates in (I.4) are, in fact, minimax optimal (up to a $\sqrt{\log n}$ factor for the L_∞ norm). To the best of our knowledge, these are the first results showing the statistical optimality of the ML estimates in the $\boldsymbol{\beta}$-model even for the graph case.
- *Inference*: In Section II-C we derive the asymptotic distribution of the ML estimate $\hat{\boldsymbol{B}}$. In particular, we prove that the finite dimensional distributions of the ML estimate converges to a multivariate Gaussian distribution (see Theorem 3). Moreover, the covariance matrix of the Gaussian distribution can be estimated consistently, using which we can construct asymptotically valid confidence sets for the model parameters (see Theorem 4).
- *Testing*: In Section III we study the goodness-of-fit problem for the hypergraph $\boldsymbol{\beta}$-model, that is, given $\boldsymbol{\gamma} \in \mathbb{R}^n$ we wish to distinguish:
$$H_0 : \boldsymbol{\beta}_s = \boldsymbol{\gamma} \quad \text{versus} \quad H_1 : \boldsymbol{\beta}_s \neq \boldsymbol{\gamma}. \tag{I.5}$$
We show that the likelihood ratio (LR) statistic for this problem (centered and scaled appropriately) is asymptotically normal under H_0 (see Theorem 5 for details). Using this result we construct an asymptotically level α test for (I.5). Next, we study the power properties of this test. In particular, we show that the detection threshold for the LR test in the L_2 norm is $n^{-\frac{2s-3}{4}}$, that is, the LR test is asymptotically powerful/powerless in detecting $\boldsymbol{\gamma}' \in \mathbb{R}^n$ depending on whether $\|\boldsymbol{\gamma}' - \boldsymbol{\gamma}\|_2$ is asymptotically greater/smaller than $n^{-\frac{2s-3}{4}}$, respectively. We also derive the limiting power function of the LR test at the threshold $\|\boldsymbol{\gamma}' - \boldsymbol{\gamma}\|_2 = \Theta(n^{-\frac{2s-3}{4}})$ (see Theorem 6). Further, in Theorem 7 we show that this detection threshold is, in fact, minimax optimal, that is all tests are asymptotically powerless when $\|\boldsymbol{\gamma}' - \boldsymbol{\gamma}\|_2$ is asymptotically smaller than $n^{-\frac{2s-3}{4}}$. In Section III-C we also obtain the detection threshold of the LR test in the L_∞ norm and establish its optimality. Again, these appear to be the first results on the non-null properties of the LR test and its optimality in the $\boldsymbol{\beta}$-model for the graph case itself.

[1] We denote by $\|\boldsymbol{x}\|_\infty$ and $\|\boldsymbol{x}\|_2$, the L_∞ and the L_2 norms of a vector $\boldsymbol{x}$, respectively.

In Section IV we illustrate the finite-sample performances of the proposed methods in simulations.

B. Related Work on the Graph β-Model

As mentioned before, Chatterjee et al. [10] initiated the rigorous study of estimation in the graph β-model. They derived, among others things, the convergence rate of the ML estimate in the L_∞ norm. Thereafter, Rinaldo et al. [44] derived necessary and sufficient conditions for the existence of the ML estimate in terms of the polytope of the degree sequences. Yan and Xu [52] proved the asymptotic normality of ML estimate and later, Yan et al. [54] derived the properties of a moment based estimator. Karwa and Slavkovic [31] studied the problem of estimation in the β-model under privacy constraints.

In the context of hypothesis testing, Mukherjee et al. [38] considered the problem of sparse signal detection in the β-model, that is, testing whether all the node parameters are zero versus whether a (possibly) sparse subset of them are non-zero. Recently, Yan et al. [56] derived the asymptotic properties of the LR test for the goodness-of-fit problem in the graph β-model, under the null hypothesis.

The graph β-model has also been generalized to incorporate additional information, such as covariates, directionality, sparsity, and weights (see Chen et al. [11] Chen and Olvera-Cravioto [12] Graham [23] Hillar and Wibisono [46] Yan and Xu [53] Hillar and Wibisono [25] Wahlström et al. [49], [55] Zhang et al. [60] and the references therein). For other exponential random graph models with functions of the degrees as sufficient statistics, see Mukherjee [39] and Xu and Mukherjee [50].

C. Asymptotic Notation

For positive sequences $\{a_n\}_{n\geqslant 1}$ and $\{b_n\}_{n\geqslant 1}$, $a_n = O(b_n)$ means $a_n \leqslant C_1 b_n$ and $a_n = \Theta(b_n)$ (and equivalently, $a_n \asymp b_n$) means $C_2 b_n \leqslant a_n \leqslant C_1 b_n$, for all n large enough and positive constants C_1, C_2. Similarly, for positive sequences $\{a_n\}_{n\geqslant 1}$ and $\{b_n\}_{n\geqslant 1}$, $a_n \lesssim b_n$ means $a_n \leqslant C_1 b_n$ and $a_n \gtrsim b_n$ means $a_n \geqslant C_2 b_n$ for all n large enough and positive constants C_1, C_2. Moreover, subscripts in the above notation, for example $O_\square$, $\lesssim_\square$, $\gtrsim_\square$, and $\Theta_\square$, denote that the hidden constants may depend on the subscripted parameters. Also, for positive sequences $\{a_n\}_{n\geqslant 1}$ and $\{b_n\}_{n\geqslant 1}$, $a_n \ll b_n$ means $a_n/b_n \to 0$ and $a_n \gg b_n$ means $a_n/b_n \to \infty$, as $n \to \infty$.

II. Maximum Likelihood Estimation in Hypergraph β-Models

In this section we consider the problem of parameter estimation in the hypergraph β-model using the ML method. In Section II-A we derived the rates of the consistency of the ML estimate. The central limit theorem of the ML estimate and the construction of confidence intervals for the model parameters are presented in Section II-C.

A. Rates of Convergence

Given a sample $H_n \sim \mathsf{H}_{n,[r]}(n, \boldsymbol{B})$ from the *r-layered hypergraph β-model*, the likelihood function can be written as follows:

$$L_n(\boldsymbol{B}) = \prod_{2\leqslant s\leqslant r} \prod_{\{v_1,v_2,\ldots,v_s\}\in E(H_n)} \frac{e^{(\beta_{s,v_1}+\ldots+\beta_{s,v_s})}}{1+e^{\beta_{s,v_1}+\ldots+\beta_{s,v_s}}}. \tag{II.1}$$

Therefore, the negative log-likelihood is given by

$$\ell_n(\boldsymbol{B}) := -\log L_n(\boldsymbol{B}) = -\sum_{s=2}^{r}\left\{\sum_{v=1}^{n}\beta_{s,v}d_s(v) - \sum_{\{v_1,v_2,\ldots,v_s\}\in\binom{[n]}{s}} \log\left(1+\exp\left(\beta_{s,v_1}+\ldots+\beta_{s,v_s}\right)\right)\right\}, \tag{II.2}$$

where

$$d_s(v) := \sum_{\boldsymbol{e}\in E(H_n):|\boldsymbol{e}|=s} \mathbb{1}\{v\in\boldsymbol{e}\}, \tag{II.3}$$

is the s-degree of the vertex $v \in [n]$, that is, the number of hyperedges of size s in H_n passing through v. The negative log-likelihood in (II.2) can be re-written as:

$$\ell_n(\boldsymbol{B}) = \sum_{s=2}^{r} \ell_{n,s}(\boldsymbol{\beta}_s), \tag{II.4}$$

where

$$\ell_{n,s}(\boldsymbol{\beta}) := \sum_{\{v_1,v_2,\ldots,v_s\}\in\binom{[n]}{s}} \log\left(1+\exp\left(\beta_{s,v_1}+\ldots+\beta_{s,v_s}\right)\right) - \sum_{v=1}^{n}\beta_{s,v}d_s(v). \tag{II.5}$$

Note that (II.4) is separable in $\boldsymbol{\beta}_2,\ldots,\boldsymbol{\beta}_r$, hence, the ML estimate of $\boldsymbol{B} = (\boldsymbol{\beta}_2,\ldots,\boldsymbol{\beta}_r)$ is given by $\hat{\boldsymbol{B}} = (\hat{\boldsymbol{\beta}}_2,\ldots,\hat{\boldsymbol{\beta}}_r)$, where

$$\hat{\boldsymbol{\beta}}_s := \operatorname{argmin}_{\boldsymbol{\beta}} \ell_{n,s}(\boldsymbol{\beta}). \tag{II.6}$$

Therefore, differentiating (II.2) with respect to $\boldsymbol{\beta}_s$ and setting the gradient to zero we can conclude that the ML estimate $\hat{\boldsymbol{\beta}}_s$ satisfies the following set of gradient equations: For all $v \in [n]$ and $2 \leqslant s \leqslant r$,

$$d_s(v) = \sum_{\{v_2,\ldots,v_s\}\in\binom{[n]\backslash\{v\}}{s-1}} \frac{e^{\hat{\beta}_{s,v}+\hat{\beta}_{s,v_2}+\ldots+\hat{\beta}_{s,v_s}}}{1+e^{\hat{\beta}_{s,v}+\hat{\beta}_{s,v_2}+\ldots+\hat{\beta}_{s,v_s}}}, \tag{II.7}$$

where $\binom{[n]\backslash\{v\}}{s-1}$ denotes the collection of all $(s-1)$-element subsets of $[n]\backslash\{v\}$. Stasi et al. [45] presented two algorithms for computing the ML estimate, namely, an iterative proportional scaling algorithm and a fixed point algorithm, and showed that both algorithms converge if the ML estimate exists.

In this paper we study the asymptotic properties of the ML estimates. In the following theorem we show that the likelihood equations (II.7) have a unique solution with high-probability and derive its rate of convergence. Recall we denote by $\|\boldsymbol{x}\|_\infty$ and $\|\boldsymbol{x}\|_2$, the L_∞ and the L_2 norms of a vector $\boldsymbol{x}$, respectively. Also, denote $\mathcal{B}_M = \{\boldsymbol{x} : \|\boldsymbol{x}\|_\infty \leqslant M\}$, the L_∞ the ball of radius M. Throughout we will assume $\boldsymbol{\beta}_s \in \mathcal{B}_M$, for all $2 \leqslant s \leqslant r$, for some constant $M > 0$.

Theorem 1: Suppose $H_n \sim \mathsf{H}_{n,[r]}(n, \boldsymbol{B})$ is a sample from the r-layered hypergraph $\boldsymbol{\beta}$-model as defined in (I.2). Then with probability $1 - o(1)$ the likelihood equations (II.7) have a unique solution $\hat{\boldsymbol{B}} = (\hat{\boldsymbol{\beta}}_2, \ldots, \hat{\boldsymbol{\beta}}_r)$, that satisfies:

$$\|\hat{\boldsymbol{\beta}}_s - \boldsymbol{\beta}_s\|_2 \lesssim_{s,M} \sqrt{\frac{1}{n^{s-2}}},$$
$$\|\hat{\boldsymbol{\beta}}_s - \boldsymbol{\beta}_s\|_\infty \lesssim_{s,M} \sqrt{\frac{\log n}{n^{s-1}}}, \tag{II.8}$$

for $2 \leqslant s \leqslant r$.

Theorem 1 provides the rates for the ML estimates for the parameters in a r-layered hypergraph $\boldsymbol{\beta}$-model both in the L_∞ and L_2 norms. To interpret the rates in (II.8) note that s-degree of a vertex (recall (II.3)) in the r-layered model $\mathsf{H}_{n,[r]}(n, \boldsymbol{B})$ is $O(n^{s-1})$ with high probability. This means there are essentially $O(n^{s-1})$ independent hyperedges containing information about each parameter in the s-th layer. Hence, each parameter in the s-th layer can be estimated at the rate $1/\sqrt{n^{s-1}}$. Aggregating this over the n coordinates gives the rates in (II.8) for the vector of parameters $\boldsymbol{\beta}_s$ in the s-th layer.

The proof of Theorem 1 is given in Appendix V. The following discussion provides a high-level outline of the proof.

- For the rate in the L_2 norm we first upper bound the gradient of the log-likelihood at the true parameter value. Specifically, we show that $\|\nabla \ell_{n,s}(\boldsymbol{\beta}_s)\|_2^2 = O(n^s)$ with high probability (see Lemma 1 for details). Next, we show that the smallest eigenvalue of the Hessian matrix $\nabla^2 \ell_{n,s}(\boldsymbol{\beta}_s)$ is bounded below by n^{s-1} (up to constants) in a neighborhood of the true parameter (see Lemma 2). Then a Taylor expansion of the log-likelihood around the true parameter, combined with the above estimates, imply the rate in the L_2 norm in (II.8) (see Appendix V-A for details).
- The proof of the rate in the L_∞ norm is more involved. For the graph case, [10] analyzed the fixed point algorithm for solving the ML equations and developed a stability version of the Erdős-Gallai condition (which provides a necessary and sufficient condition for a sequence of numbers to be the degree sequence of a graph) to derive the rate of ML estimate in the L_∞ norm. One of the technical challenges in dealing with the hypergraph case is the absence of Erdős-Gallai-type characterizations of the degree sequence. To circumvent this issue, we take a more analytic approach based on the 'leave-one-out' technique, that appear in the analysis of ranking models [13], [14]. Here the idea is to decompose, for each $u \in [n]$, the log-likelihood function of the s-th layer $\ell_{n,s}$ (recall (II.5)) into two parts: one depending on $\beta_{s,u}$ and the other not depending on it. Using the part of the log-likelihood not depending on $\beta_{s,u}$ we first analyze the properties of the constrained leave-one-out ML estimate, which is the maximizer of the part of the log-likelihood not depending on $\beta_{s,u}$ in a neighborhood of the leave-one-out true parameter. Then from the part of the log-likelihood depending on $\beta_{s,u}$ we obtain, by a Taylor expansion around the true parameter value $\beta_{s,u}$, the L_∞ rate in (II.8) with an extra additive error term which depends on the constrained leave-one-out ML estimate. Using the bound on the latter obtained earlier we show this error term is negligible compared to the L_∞ rate in (II.8).

The following corollary about the r-uniform model is an immediate consequence of Theorem 1. We record it separately for ease of referencing.

Corollary 1: Suppose $H_n \sim \mathsf{H}_{n,r}(n, \boldsymbol{\beta})$ is a sample from the r-uniform hypergraph $\boldsymbol{\beta}$-model as defined in (I.3). Then with probability $1 - o(1)$ the ML estimate $\hat{\boldsymbol{\beta}}$ is unique and the following hold:

$$\|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\|_2 \lesssim_{r,M} \sqrt{\frac{1}{n^{r-2}}},$$
$$\|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\|_\infty \lesssim_{r,M} \sqrt{\frac{\log n}{n^{r-1}}}. \tag{II.9}$$

B. Minimax Rates

In the following theorem we establish the tightness of the rates of ML estimate obtained in the previous section by proving matching lower bounds.

Theorem 2: Suppose $H_n \sim \mathsf{H}_{n,[r]}(n, \boldsymbol{B})$, with $\boldsymbol{B} = (\boldsymbol{\beta}_2, \ldots, \boldsymbol{\beta}_r)$, such that $\boldsymbol{\beta}_s \in \mathcal{B}_M$, for $2 \leqslant s \leqslant r$. Given $\delta \in (0, 1)$ there exists a constant C (depending on M, r, and δ) such that the following holds for estimation in the L_2 norm:

$$\min_{\hat{\boldsymbol{\beta}}} \max_{\boldsymbol{\beta}_s \in \mathcal{B}_M} \mathbb{P}\left(\|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_s\|_2 \geqslant C \sqrt{\frac{1}{n^{s-2}}} \right) \geqslant 1 - \delta. \tag{II.10}$$

Moreover, for estimation in the L_∞ norm the following holds: there exists $n_0 \geqslant 0$ such that for all $n \geqslant n_0$

$$\min_{\hat{\boldsymbol{\beta}}} \max_{\boldsymbol{\beta}_s \in \mathcal{B}_M} \mathbb{P}\left(\|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_s\|_\infty \geqslant C \sqrt{\frac{\log n}{n^{s-1}}} \right) \geqslant 1 - \delta. \tag{II.11}$$

This result shows that the ML estimate is minimax rate optimal in both the L_2 and the L_∞ metrics. The proof of Theorem 2 is given in Appendix VI. The bounds in (II.10) and (II.11) are proved using Fano's lemma. For the bound in (II.10) we construct $2^{\Theta(n)}$ well-separated points in the parameter space which have 'small' average Kulbeck-Leibler (KL) divergence with the origin (see Appendix VI-A) and for the bound in (II.11) we construct n well separated points which have 'small' KL-divergence with the origin (see Appendix VI-B)

C. Central Limit Theorems and Confidence Intervals

The results obtained in the previous section show that the vector ML estimates are consistent in the L_∞-norm. However, for conducting asymptotically precise inference on the individual model parameters, we need to understand the limiting distribution of the ML estimates. In Theorem 3 below we show that the finite dimensional distributions of the ML estimates (appropriately scaled) converge to a multivariate Gaussian distribution. Using this result in Theorem 4 we construct joint confidence sets for any finite collection of parameters. Towards this, for $H_n \sim \mathsf{H}_{n,[r]}(n, \boldsymbol{B})$ denote the variance of the s-degree of the node $v \in [n]$ as:

$$\sigma_s(v)^2 := \mathrm{Var}[d_s(v)] \tag{II.12}$$

$$= \sum_{\{v_2,\ldots,v_s\} \in \binom{[n]\setminus\{v\}}{s-1}} \frac{e^{\beta_{s,v}+\beta_{s,v_2}+\ldots+\beta_{s,v_s}}}{(1+e^{\beta_{s,v}+\beta_{s,v_2}+\ldots+\beta_{s,v_s}})^2}. \tag{II.13}$$

Then we have the following result:

Theorem 3: Suppose $H_n \sim \mathsf{H}_{n,[r]}(n, \boldsymbol{B})$ is a sample from the r-layered hypergraph $\boldsymbol{\beta}$-model as defined in (I.2). Fix any collection of integers $a_1, a_2, \ldots, a_r \geqslant 1$ and any subsets of nodes $J_1, J_2, \ldots, J_r$ with cardinalities $a_1, a_2, \ldots, a_r$, respectively. Then as $n \to \infty$,

$$\begin{pmatrix} [\boldsymbol{D}_2(\hat{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}_2)]_{J_2} \\ [\boldsymbol{D}_3(\hat{\boldsymbol{\beta}}_3 - \boldsymbol{\beta}_3)]_{J_3} \\ \vdots \\ [\boldsymbol{D}_r(\hat{\boldsymbol{\beta}}_r - \boldsymbol{\beta}_r)]_{J_r} \end{pmatrix} \xrightarrow{D} \mathcal{N}_{\sum_{s=2}^r a_s}(\boldsymbol{0}, \boldsymbol{I}), \tag{II.14}$$

where $\boldsymbol{D}_s = \mathrm{diag}\,(\sigma_s(v))_{v\in[n]}$, for $2 \leqslant s \leqslant r$ and for any vector $\boldsymbol{x} \in \mathbb{R}^n$, $[\boldsymbol{x}]_{J_s} = (x_v)_{v\in[J_s]}^\top$.

Remark 1: Observe that since $\boldsymbol{\beta}_s \in \mathcal{B}_M$ for all $2 \leqslant s \leqslant r$, $O(\sigma_s(v)^2) \asymp n^{s-1}$ for all $v \in [n]$. Therefore, the convergence rate of $[(\hat{\boldsymbol{\beta}}_s - \boldsymbol{\beta}_s)]_{J_s}$ is of order $n^{(s-1)/2}$ for all $2 \leqslant s \leqslant r$.

The proof of Theorem 3 is given in Appendix VII-A. The idea of the proof is to linearize $\hat{\beta}_{s,v} - \beta_{s,v}$ in terms of the s-degree of the node $v \in [n]$. Since the s-degree of a node is the sum of independent random variables, applying Lindeberg's CLT gives the result in (II.14). In the special case of the r-uniform model, Theorem 3 can be written in the following simpler form:

Corollary 2: Suppose $H_n \sim \mathsf{H}_{n,r}(n, \boldsymbol{\beta})$ is a sample from the r-uniform hypergraph $\boldsymbol{\beta}$-model as defined in (I.3). For all $v \in [n]$, let

$$\sigma(v)^2 := \sum_{\{v_2,\ldots,v_r\} \in \binom{[n]\setminus\{v\}}{r-1}} \frac{e^{\beta_v+\beta_{v_2}+\ldots+\beta_{v_r}}}{1+e^{\beta_v+\beta_{v_2}+\ldots+\beta_{v_r}}}.$$

Then for any collection of $a \geqslant 1$ indices $J := \{v_1, \cdots, v_a\} \in \binom{[n]}{a}$, as $n \to \infty$,

$$[\boldsymbol{D}]_J([\hat{\boldsymbol{\beta}}]_J - [\boldsymbol{\beta}]_J) \xrightarrow{D} \mathcal{N}_a(\boldsymbol{0}, \boldsymbol{I}),$$

where $\boldsymbol{D} = \mathrm{diag}\,(\sigma(v))_{v\in[n]}$, $[\boldsymbol{D}]_J = \mathrm{diag}\,(\sigma(v))_{v\in J}$, $[\hat{\boldsymbol{\beta}}]_J = (\hat{\beta}_v)_{v\in[J]}^\top$, and $[\boldsymbol{\beta}]_J = (\beta_v)_{v\in[J]}^\top$.

To use the above results to construct confidence sets for the parameters, we need to consistently estimate the elements of the matrix $\boldsymbol{D}_s$. Note that the natural plug-in estimate of $\sigma_s(v)$ is

$$\hat{\sigma}_s(v)^2 := \sum_{\{v_2,\ldots,v_s\} \in \binom{[n]\setminus\{v\}}{s-1}} \frac{e^{\hat{\beta}_{s,v}+\hat{\beta}_{s,v_2}+\ldots+\hat{\beta}_{s,v_s}}}{(1+e^{\hat{\beta}_{s,v}+\hat{\beta}_{s,v_2}+\ldots+\hat{\beta}_{s,v_s}})^2}. \tag{II.15}$$

This estimate turns out to be consistent for $\sigma_s(v)$, leading to the following result (see Appendix VII-B for the proof):

Theorem 4: Suppose $H_n \sim \mathsf{H}_{n,[r]}(n, \boldsymbol{B})$ is a sample from the r-layered hypergraph $\boldsymbol{\beta}$-model as defined in (I.2). Fix any collection of integers $a_1, a_2, \ldots, a_r \geqslant 1$ and any subsets of nodes $J_1, J_2, \ldots, J_r$ with cardinalities $a_1, a_2, \ldots, a_r$, respectively. Then, for all $\alpha \in (0,1)$,

$$\lim_{n\to\infty} \mathbb{P}\left(\left\{\sum_{s=2}^r ([(\hat{\boldsymbol{\beta}}_s - \boldsymbol{\beta}_s)]_{J_s})^\top [\hat{\boldsymbol{D}}_s^2]_{J_s} ([(\hat{\boldsymbol{\beta}}_s - \boldsymbol{\beta}_s)]_{J_s}) \leqslant \chi^2_{\sum_{s=2}^r a_s, 1-\alpha}\right\}\right) = 1-\alpha, \tag{II.16}$$

where $\hat{\boldsymbol{D}}_s^2 = \mathrm{diag}(\hat{\sigma}_s(v)^2)_{v\in[n]}$, $[\hat{\boldsymbol{D}}_s^2]_{J_s} = \mathrm{diag}(\hat{\sigma}_s(v)^2)_{v\in J_s}$, for $2 \leqslant s \leqslant r$, and for $a \geqslant 1$, $\chi^2_{a,1-\alpha}$ is the $(1-\alpha)$-th quantile of the chi-squared distribution with a degrees of freedom.

III. Goodness-of-Fit: Asymptotics of the Likelihood Ratio Test and Minimax Detection Rates

In this section we consider the problem of testing for goodness-of-fit in the hypergraph $\boldsymbol{\beta}$-model. In particular, given $\boldsymbol{\gamma} \in \mathbb{R}^n$ and a sample $H_n \sim \mathsf{H}_{n,[r]}(n, \boldsymbol{B})$, with $\boldsymbol{B} = (\boldsymbol{\beta}_2, \ldots, \boldsymbol{\beta}_r)$, we consider the following hypothesis testing problem: For $2 \leqslant s \leqslant r$,

$$H_0 : \boldsymbol{\beta}_s = \boldsymbol{\gamma} \quad \text{versus} \quad H_1 : \boldsymbol{\beta}_s \neq \boldsymbol{\gamma}. \tag{III.1}$$

This section is organized as follows: In Section III-A we derive the asymptotic distribution and detection rates of the likelihood ratio (LR) test for the problem (III.1). In Section III-B we show that the detection rate of the LR test is minimax optimal for testing in L_2 norm. Rates for testing in L_∞ norm are derived in Section III-C.

A. Asymptotics of the Likelihood Ratio Test

Consider the LR statistic for the testing problem (III.1):

$$\log \Lambda_{n,s} = \ell_{n,s}(\boldsymbol{\gamma}) - \ell_{n,s}(\hat{\boldsymbol{\beta}}_s), \tag{III.2}$$

where $\ell_{n,s}$ is the negative log-likelihood function (II.5) and $\hat{\boldsymbol{\beta}}_s$ is the ML estimate (II.6). The following theorem proves the limiting distribution of the LR statistic (III.2) under H_0.

Theorem 5: Suppose $\boldsymbol{\gamma} \in \mathcal{B}_M$. Then under H_0,

$$\lambda_{n,s} := \frac{2\log \Lambda_{n,s} - n}{\sqrt{2n}} \xrightarrow{D} \mathcal{N}(0,1), \tag{III.3}$$

for $\log \Lambda_{n,s}$ as defined in (III.2).

The proof of Theorem 5 is given in Appendix VIII-A. To prove the result we first expand $\log \Lambda_{n,s}$ around the null parameter $\boldsymbol{\gamma}$ and derive an asymptotic expansion of $\lambda_{n,s}$ in terms of the sum of squares of the s-degree sequence $(d_s(1), d_s(2), \ldots, d_s(n))^\top$ (see (VIII.13)). Since the degrees are asymptotically independent (recall Theorem 3), we can show that the sum of squares of the degrees (appropriately centered and scaled) is asymptotically normal (see Proposition 3), establishing the result in (III.3).

Theorem 5 shows that the LR test

$$\phi_{n,s} := \mathbb{1}\left\{|\lambda_{n,s}| > z_{\alpha/2}\right\}, \tag{III.4}$$

where $z_{\alpha/2}$ is the $(1-\alpha/2)$-th quantile of the standard normal distribution, has asymptotic level α. To study the power of this test consider the following testing problem:

$$H_0 : \boldsymbol{\beta}_s = \boldsymbol{\gamma} \quad \text{versus} \quad H_1 : \boldsymbol{\beta}_s = \boldsymbol{\gamma}', \tag{III.5}$$

where $\boldsymbol{\gamma}' \neq \boldsymbol{\gamma}$ is such that $\|\boldsymbol{\gamma} - \boldsymbol{\gamma}'\|_2 = O(1)$. Recall that $\boldsymbol{d}_s = (d_s(1), d_s(2), \ldots, d_s(n))^\top$ is the vector of s-degrees. Also, $\mathrm{Cov}_{\boldsymbol{\gamma}}[\boldsymbol{d}_s]$ will denote the covariance matrix of the vector of s-degrees (see (VII.2)).

Theorem 6: Suppose (III.3) holds and $\boldsymbol{\gamma}'$ be as in (III.5), then the asymptotic power of the test $\phi_{n,s}$ defined in (III.4) satisfies:

$$\lim_{n\to\infty} \mathbb{E}_{\boldsymbol{\gamma}'}[\phi_{n,s}] = \begin{cases} \alpha & \text{if } \|\boldsymbol{\gamma}' - \boldsymbol{\gamma}\|_2 \ll n^{-\frac{2s-3}{4}}, \\ 1 & \text{if } \|\boldsymbol{\gamma}' - \boldsymbol{\gamma}\|_2 \gg n^{-\frac{2s-3}{4}}. \end{cases} \tag{III.6}$$

Moreover, if $n^{\frac{2s-3}{4}}\|\boldsymbol{\gamma}' - \boldsymbol{\gamma}\|_2 \to \tau \in (0,\infty)$, then there exists $\eta \in (0,\infty)$ depending on τ such that

$$\eta = \lim_{n\to\infty} \frac{(\boldsymbol{\gamma}' - \boldsymbol{\gamma})^\top \mathrm{Cov}_{\boldsymbol{\gamma}}[\boldsymbol{d}_s](\boldsymbol{\gamma}' - \boldsymbol{\gamma})}{\sqrt{n}},$$

where the limit always exists along a subsequence, and

$$\lim_{n\to\infty} \mathbb{E}_{\boldsymbol{\gamma}'}[\phi_{n,s}] = \mathbb{P}\left(\left|\mathcal{N}(-\tfrac{\eta}{\sqrt{2}}, 1)\right| > z_{\alpha/2}\right). \tag{III.7}$$

The proof of Theorem 6 is given in Appendix VIII-B. It entails analyzing the asymptotic distribution of the scaled LR statistic $\lambda_{n,s}$ under H_1 as in (III.5). Specifically, we show that when $\|\boldsymbol{\gamma}' - \boldsymbol{\gamma}\|_2 \ll n^{-\frac{2s-3}{4}}$, then $\lambda_{n,s} \xrightarrow{D} \mathcal{N}(0,1)$, hence the LR test (III.3) is asymptotically powerless in detecting H_1. On the other hand, if $\|\boldsymbol{\gamma}' - \boldsymbol{\gamma}\|_2 \gg n^{-\frac{2s-3}{4}}$, then the $\lambda_{n,s}$ diverges to infinity, hence the LR test is asymptotically powerful. In other words, $n^{-\frac{2s-3}{4}}$ is the detection threshold in the L_2 norm of the LR test. We also derive the limiting power function of the LR test at the threshold $n^{\frac{2s-3}{4}}\|\boldsymbol{\gamma}' - \boldsymbol{\gamma}\|_2 \to \tau \in (0,\infty)$. In this case, $\lambda_{n,s} \xrightarrow{D} \mathcal{N}(-\eta/\sqrt{2}, 1)$, where 'effective signal strength' η is the limit of the scaled Mahalanobis distance between $\boldsymbol{\gamma}$ and $\boldsymbol{\gamma}'$, where the dispersion matrix is the covariance matrix of the degrees. Here, the Mahalanobis distance between two vectors $\boldsymbol{\gamma}$ and $\boldsymbol{\gamma}'$ with the dispersion matrix $\boldsymbol{\Sigma}$ refers to the quantity $(\boldsymbol{\gamma} - \boldsymbol{\gamma}')^\top\boldsymbol{\Sigma}(\boldsymbol{\gamma} - \boldsymbol{\gamma}')$. In the next section we will show that this detection rate is, in fact, minimax optimal.

B. Minimax Detection Rate in the L_2 Norm

In this section we will show that the detection threshold of the LR test obtained in Theorem 6 is information-theoretically tight. To formalize this consider the testing problem: For $\varepsilon > 0$ and $\boldsymbol{\gamma} \in \mathcal{B}_M$,

$$H_0 : \boldsymbol{\beta}_s = \boldsymbol{\gamma} \quad \text{versus} \quad H_1 : \|\boldsymbol{\beta}_s - \boldsymbol{\gamma}\|_2 \geqslant \varepsilon. \tag{III.8}$$

The worst-case risk of a test function ψ_n for the testing problem (III.8) is defined as:

$$\begin{aligned} &\mathcal{R}(\psi_n, \boldsymbol{\gamma}) \\ &= \mathbb{P}_{H_0}(\psi_n = 1) + \sup_{\boldsymbol{\gamma}' \in \mathcal{B}_M : \|\boldsymbol{\gamma}' - \boldsymbol{\gamma}\|_2 \geqslant \varepsilon} \mathbb{P}_{\boldsymbol{\gamma}'}(\psi_n = 0), \end{aligned} \tag{III.9}$$

which is the sum of the Type I error and the maximum possible Type II error of the test function ψ_n. Given $H_n \sim \mathsf{H}_{n,s}(n, \boldsymbol{\beta}_s)$, for some $\boldsymbol{\beta}_s \in \mathcal{B}_M$, and $\varepsilon = \varepsilon_n$ (depending on n), a sequence of test functions ψ_n is said to be *asymptotically powerful* for (III.9), if for all $\boldsymbol{\gamma} \in \mathcal{B}_M$ $\lim_{n\to\infty} \mathcal{R}(\psi_n, \boldsymbol{\gamma}) = 0$. On the other hand, a sequence of test functions ψ_n is said to be *asymptotically powerless* for (III.9), if there exists $\boldsymbol{\gamma} \in \mathcal{B}_M$ such that $\lim_{n\to\infty} \mathcal{R}(\psi_n, \boldsymbol{\gamma}) = 1$.

Theorem 7: Given $H_n \sim \mathsf{H}_{n,s}(n, \boldsymbol{\beta}_s)$ and $\boldsymbol{\gamma} \in \mathcal{B}_M$, consider the testing problem (III.8). Then the following hold:

(a) The LR test (III.4) is asymptotically powerful for (III.8), when $\varepsilon \gg n^{-\frac{2s-3}{4}}$.

(b) On the other hand, all tests are asymptotically powerless for (III.8), when $\varepsilon \ll n^{-\frac{2s-3}{4}}$.

The result in Theorem 7 (a) is a direct consequence of Theorem 6. The proof of Theorem 7 (b) is given in Appendix IX-A. For this we chose $\boldsymbol{\gamma} = \mathbf{0} \in \mathbb{R}^n$ and randomly perturb (that is, randomly add or subtract $\varepsilon/\sqrt{n}$) the coordinates of $\boldsymbol{\gamma}$ to construct $\boldsymbol{\beta}_s \in \mathcal{B}_M$ satisfying $\|\boldsymbol{\beta}_s - \boldsymbol{\gamma}\|_2 \geqslant \varepsilon$. Then a second-moment calculation of the likelihood ratio shows that detecting these two models is impossible for $\varepsilon \ll n^{-\frac{2s-3}{4}}$. These results combined show that $n^{-\frac{2s-3}{4}}$ is the minimax detection rate for the testing problem (III.8) and the LR test attain the minimax rate.

Remark 2 (Comparison Between Testing and Estimation Rates): Recall from (II.8) and (II.10) that the minimax rate of estimating $\hat{\boldsymbol{\beta}}_s$ in the L_2 norm is $n^{-\frac{s-2}{2}}$. On the other hand, Theorem 7 shows that the minimax rate of testing in the L_2 norm is $n^{-\frac{2s-3}{4}} \ll n^{-\frac{s-2}{2}}$. For example, in the graph case (where $s = 2$), the estimation rate is $\Theta(1)$ whereas the rate of testing is $n^{-\frac{1}{4}}$. This is an instance of the well-known phenomenon that high-dimensional estimation is, in general, harder that testing in the squared-error loss.

C. Testing in the L_∞ Norm

In this section we consider the goodness-of-fit problem when the separation is measured in the L_∞ norm. This complements our results on estimation in the L_∞ norm in Theorem 1. Towards this, as in (III.8), consider the testing problem: For $\varepsilon > 0$ and $\boldsymbol{\gamma} \in \mathcal{B}_M$,

$$H_0 : \boldsymbol{\beta}_s = \boldsymbol{\gamma} \quad \text{versus} \quad H_1 : \|\boldsymbol{\beta}_s - \boldsymbol{\gamma}\|_\infty \geqslant \varepsilon. \tag{III.10}$$

In this case the minimax risk of a test function is defined as in (III.9) with the L_2 norm $\|\boldsymbol{\gamma}' - \boldsymbol{\gamma}\|_2$ replaced by the L_∞ norm $\|\boldsymbol{\gamma}' - \boldsymbol{\gamma}\|_\infty$. Then consider the test:

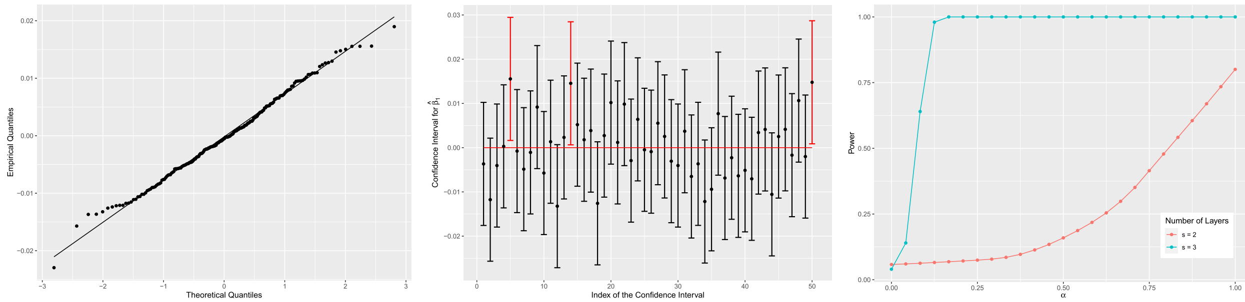
$$\phi_{n,s}^{\max} := \mathbb{1}\left\{\|\hat{\boldsymbol{\beta}}_s - \boldsymbol{\gamma}\|_\infty \geqslant 2C\sqrt{\frac{\log n}{n^{s-1}}}\right\}, \tag{III.11}$$

where $C := C(s, M) > 0$ is chosen according to (II.8) such that

$$\mathbb{P}_{\boldsymbol{\kappa}}\left(\|\hat{\boldsymbol{\beta}}_s - \boldsymbol{\kappa}\|_\infty \leqslant C\sqrt{\frac{\log n}{n^{s-1}}}\right) \to 1,$$

for all $\boldsymbol{\kappa} \in \mathcal{B}_M$. This implies, $\mathbb{E}_{\boldsymbol{\gamma}}[\phi_{n,s}^{\max}] \to 0$. Also, for $\boldsymbol{\gamma}' \in \mathcal{B}_M$ such that $\|\boldsymbol{\gamma} - \boldsymbol{\gamma}'\|_\infty \geqslant \varepsilon$,

$$\begin{aligned} \mathbb{E}_{\boldsymbol{\gamma}'}[\phi_{n,s}^{\max}] &= \mathbb{P}_{\boldsymbol{\gamma}'}\left(\|\hat{\boldsymbol{\beta}}_s - \boldsymbol{\gamma}\|_\infty \geqslant 2C\sqrt{\frac{\log n}{n^{s-1}}}\right) \\ &\geqslant \mathbb{P}_{\boldsymbol{\gamma}'}\left(\|\hat{\boldsymbol{\beta}}_s - \boldsymbol{\gamma}'\|_\infty \leqslant C\sqrt{\frac{\log n}{n^{s-1}}}\right) \\ &\to 1, \end{aligned} \tag{III.12}$$



(a) (b) (c)

Fig. 1. (a) QQ plot of the ML estimate $\hat{\beta}_1$, (b) confidence intervals for β_1, and (c) power of the LR test for the goodness of fit problem (IV.1). The first two plots correspond to the 3-uniform hypergraph $\boldsymbol{\beta}$-model whereas plot (c) corresponds to s-uniform hypergraph $\boldsymbol{\beta}$-model for $s = 2, 3$.

whenever $\varepsilon \gg \sqrt{\log n / n^{s-1}}$. This is because $\|\hat{\boldsymbol{\beta}}_s - \boldsymbol{\gamma}'\|_\infty \leqslant C\sqrt{\log n / n^{s-1}}$ implies,

$$\|\hat{\boldsymbol{\beta}}_s - \boldsymbol{\gamma}\|_\infty \geqslant \|\boldsymbol{\gamma} - \boldsymbol{\gamma}'\|_\infty - \|\hat{\boldsymbol{\beta}}_s - \boldsymbol{\gamma}'\|_\infty \geqslant \varepsilon - C\sqrt{\frac{\log n}{n^{s-1}}} \geqslant 2C\sqrt{\frac{\log n}{n^{s-1}}}, \tag{III.13}$$

whenever $\varepsilon \gg \sqrt{\log n / n^{s-1}}$. This implies that the test $\phi_{n,s}^{\max}$ in (III.11) is asymptotically powerful for (III.10) whenever $\varepsilon \gg \sqrt{\log n / n^{s-1}}$. The following result shows that this rate is optimal (up to a factor of $\sqrt{\log n}$) for testing in the L_∞ norm.

Theorem 8: Given $H_n \sim \mathsf{H}_{n,s}(n, \boldsymbol{\beta}_s)$ and $\boldsymbol{\gamma}, \boldsymbol{\beta}_s \in \mathcal{B}_M$, consider the testing problem (III.10). Then the following hold:

(a) The test $\phi_{n,s}^{\max}$ in (III.12) is asymptotically powerful for (III.10), when $\varepsilon \gg \sqrt{\frac{\log n}{n^{s-1}}}$.

(b) On the other hand, all tests are asymptotically powerless for (III.10), when $\varepsilon \ll \sqrt{\frac{\log n}{n^{s-1}}}$.

The proof of Theorem 8 (b) is given in Appendix IX-B. Note that in this case minimax rates of estimation and testing are the same, since the effect of high-dimensional aggregation does not arise when separation is measured in the L_∞ norm.

IV. Numerical Experiments

In this section we study the performance of the ML estimates and the LR tests discussed above in simulations. To begin with we simulate a 3-uniform hypergraph $\boldsymbol{\beta}$-model $\mathsf{H}_3(n, \boldsymbol{\beta})$, with $n = 400$ vertices and $\boldsymbol{\beta} = \mathbf{0} \in \mathbb{R}^n$. Figure 1(a) shows the quantile-quantile (QQ) plot (over 200 iterations) of the first coordinate of the ML estimate $[\boldsymbol{D}]_1([\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}]_1)$ (where $\hat{\boldsymbol{\beta}}$ is computed using the fixed point algorithm described in [45]. Here $\boldsymbol{D}$ is as defined in Corollary 2). We observe that the empirical quantiles closely follow the quantiles of the standard normal distribution, validating the result in Corollary 2.

In the same setup as above, Figure 1(b) shows the 95% confidence interval for $[\boldsymbol{\beta}]_1$ over 50 iterations. Specifically, we plot the intervals

$$\left[[\hat{\boldsymbol{\beta}}]_1 - \frac{1.96}{[\hat{\boldsymbol{D}}]_1}, [\hat{\boldsymbol{\beta}}]_1 + \frac{1.96}{[\hat{\boldsymbol{D}}]_1} \right],$$

where $\hat{\boldsymbol{D}}$ is the estimate of $\boldsymbol{D}$ as defined in Theorem 4. This figure shows that 47 out of 50 of intervals cover the true parameter, which gives an empirical coverage of $47/50 = 0.94$.

Next, we consider the goodness of fit problem in s-uniform hypergraph $\boldsymbol{\beta}$-model:

$$H_0 : \boldsymbol{\beta}_s = 0 \quad \text{versus} \quad H_1 : \boldsymbol{\beta}_s \neq 0, \tag{IV.1}$$

for $s = 2, 3$. For this we simulate $H_n \sim \mathsf{H}_s(n, \boldsymbol{\gamma})$, with $n = 250$ and $\boldsymbol{\gamma} = \alpha \cdot \boldsymbol{u}$, where $\boldsymbol{u}$ is chosen uniformly at random from the n-dimensional unit sphere and $\alpha \in [0, 1]$. Figure 1(c) shows the empirical power of the LR test (III.4) (over 50 iterations) as α varies over a grid of 25 uniformly spaced values in $[0, 1]$, for $s = 2, 3$. In both cases, as expected, the power increases with α, which, in this case, determines the signal strength. Also, the LR test is more powerful in the 3-uniform case compared to the 2-uniform case. This aligns with conclusions of Theorem 6, which shows that the detection threshold of the LR test in the 3-uniform case is $n^{-\frac{3}{4}}$, while for 2-uniform case it is $n^{-\frac{1}{4}}$. Hence, one expects to see more power at lower signal strengths (smaller α) for $s = 3$ compared to $s = 2$.

V. Proof of Theorem 1

A. Convergence Rate in the L_2 Norm

As mentioned in the Introduction, the proof of Theorem 1 involves showing the following: (1) a concentration bound on the gradient of negative log-likelihood $\ell_{n,s}$ (recall (II.5)) at the true parameter value $\boldsymbol{B} = (\boldsymbol{\beta}_2, \ldots, \boldsymbol{\beta}_r)$, and (2) the strong convexity of $\ell_{n,s}$ in a neighborhood of the true parameter. We begin with the concentration of the gradient $\nabla \ell_{n,s}$ in both the L_2 and the L_∞ norms:

Lemma 1: Suppose the assumptions of Theorem 1 hold. Then for each $2 \leqslant s \leqslant r$, there exists a constant $C > 0$ (depending on r and M) such that the following hold:

$$\|\nabla \ell_{n,s}(\boldsymbol{\beta}_s)\|_2^2 \leqslant C\, n^s \quad \textit{and} \quad \|\nabla \ell_{n,s}(\boldsymbol{\beta}_s)\|_\infty^2 \leqslant C\, n^{s-1} \log n, \tag{V.1}$$

with probability $1 - O\left(\frac{1}{n^2}\right)$.

The next step is to establish the strong convexity of $\ell_{n,s}$. Towards this we need to show that the smallest eigenvalue $\lambda_{\min}(\nabla^2 \ell_{n,s})$ of the Hessian matrix $\nabla^2 \ell_{n,s}$ (appropriately scaled) is bounded away from zero in a neighborhood of the true value $\boldsymbol{\beta}_s$. This is the content of the following lemma, which also establishes a matching upper bound on the largest eigenvalue $\lambda_{\max}(\nabla^2 \ell_{n,s})$ of the Hessian matrix $\nabla^2 \ell_{n,s}$.

Lemma 2: Suppose the assumptions of Theorem 1 hold. Fix $2 \leqslant s \leqslant r$ and a constant $K > 0$. Then there exists constants $C_1', C_2' > 0$ (depending on r and M) such that the following hold:

$$C_1' e^{-s\|\boldsymbol{\beta}-\boldsymbol{\beta}_s\|_\infty} n^{s-1} \leqslant \lambda_{\min}(\nabla^2 \ell_{n,s}(\boldsymbol{\beta})) \leqslant \lambda_{\max}(\nabla^2 \ell_{n,s}(\boldsymbol{\beta})) \leqslant C_2' n^{s-1}. \quad \text{(V.2)}$$

As a consequence, there exists a constants $C_1, C_2 > 0$ (depending on r, K, and M) such that the following hold:

$$C_1 n^{s-1} \leqslant \inf_{\boldsymbol{\beta}:\|\boldsymbol{\beta}-\boldsymbol{\beta}_s\|_2 \leqslant K} \lambda_{\min}(\nabla^2 \ell_{n,s}(\boldsymbol{\beta})) \quad \text{(V.3)}$$
$$\leqslant \sup_{\boldsymbol{\beta}:\|\boldsymbol{\beta}-\boldsymbol{\beta}_s\|_2 \leqslant K} \lambda_{\max}(\nabla^2 \ell_{n,s}(\boldsymbol{\beta})) \leqslant C_2 n^{s-1}. \quad \text{(V.4)}$$

Note that the constant $K > 0$ in the above lemma specifies the width of the neighborhood around $\boldsymbol{\beta}_s$ where the inequality (A.3) holds. The inequality continues to hold for any $K > 0$, but the constants C_1 and C_2 start depending on K. If K is large, the constants C_1 and C_2 become small and large, respectively.

The proofs of Lemma 1 and Lemma 2 are given in Appendix V-A2 and Appendix V-A3, respectively. We first apply these results to prove the rate of convergence in the L_2 norm in Theorem 1.

1) Deriving the L_2 Norm Bound in (II.8): To begin with suppose the ML equations (II.7) have a solution $\hat{\boldsymbol{B}} = (\hat{\boldsymbol{\beta}}_2, \ldots, \hat{\boldsymbol{\beta}}_r)$. This implies, $\nabla \ell_{n,s}(\hat{\boldsymbol{\beta}}_s) = 0$, for $2 \leqslant s \leqslant r$, where $\ell_{n,s}$ is as defined in (II.5). For $2 \leqslant s \leqslant r$ and $0 \leqslant t \leqslant 1$, define

$$\bar{\boldsymbol{\beta}}_s(t) := t\hat{\boldsymbol{\beta}}_s + (1-t)\boldsymbol{\beta}_s,$$

and

$$g_s(t) := (\hat{\boldsymbol{\beta}}_s - \boldsymbol{\beta}_s)^\top \nabla \ell_{n,s}(\bar{\boldsymbol{\beta}}_s(t)).$$

Note that $\nabla \ell_{n,s}(\boldsymbol{\beta}_s(1)) = \nabla \ell_{n,s}(\hat{\boldsymbol{\beta}}_s) = 0$. Hence, by the Cauchy-Schwarz inequality,

$$|g_s(1) - g_s(0)| = |(\hat{\boldsymbol{\beta}}_s - \boldsymbol{\beta}_s)^\top \nabla \ell_{n,s}(\boldsymbol{\beta}_s)| \leqslant \|\hat{\boldsymbol{\beta}}_s - \boldsymbol{\beta}_s\|_2 \cdot \|\nabla \ell_{n,s}(\boldsymbol{\beta}_s)\|_2. \quad \text{(V.5)}$$

Also,

$$g_s'(t) = (\hat{\boldsymbol{\beta}}_s - \boldsymbol{\beta}_s)^\top \nabla^2 \ell_{n,s}(\bar{\boldsymbol{\beta}}_s(t))(\hat{\boldsymbol{\beta}}_s - \boldsymbol{\beta}_s) \quad \text{(V.6)}$$
$$\geqslant \lambda_{\min}(\nabla^2 \ell_{n,s}(\bar{\boldsymbol{\beta}}_s(t)))\|\hat{\boldsymbol{\beta}}_s - \boldsymbol{\beta}_s\|_2^2. \quad \text{(V.7)}$$

We now consider two cases: To begin with assume $s \geqslant 3$. By Lemma 2, given a constant $K > 0$ there exists a constant $C_1 > 0$ (depending on r, K, M) such that

$$\inf_{\boldsymbol{\beta}:\|\boldsymbol{\beta}-\boldsymbol{\beta}_s\|_2 \leqslant K} \lambda_{\min}(\nabla^2 \ell_{n,s}(\boldsymbol{\beta})) \geqslant C_1 n^{s-1}. \quad \text{(V.8)}$$

Note that $\|\bar{\boldsymbol{\beta}}_s(t) - \boldsymbol{\beta}_s\|_2 = t\|\hat{\boldsymbol{\beta}}_s - \boldsymbol{\beta}_s\|_2$. Then

$$\begin{aligned} |g_s(1) - g_s(0)| &\geqslant g_s(1) - g_s(0) \\ &= \int_0^1 g_s'(t)\mathrm{d}t \\ &\geqslant \int_0^{\min\{1, \frac{K}{\|\hat{\boldsymbol{\beta}}_s - \boldsymbol{\beta}_s\|_2}\}} g_s'(t)\mathrm{d}t \\ &\geqslant C_1 n^{s-1}\|\hat{\boldsymbol{\beta}}_s - \boldsymbol{\beta}_s\|_2^2 \min\left\{1, \frac{K}{\|\hat{\boldsymbol{\beta}}_s - \boldsymbol{\beta}_s\|_2}\right\}, \end{aligned}$$

where the last step follows from (V.6) and (V.8). Therefore, by (V.5) and Lemma 1, with probability $1 - O(\frac{1}{n^2})$,

$$\min\{\|\hat{\boldsymbol{\beta}}_s - \boldsymbol{\beta}_s\|_2, K\} \lesssim_{r,K,M} \frac{1}{n^{s-1}} \cdot \|\nabla \ell_{n,s}(\boldsymbol{\beta}_s)\|_2 \lesssim_{r,K,M} \sqrt{\frac{1}{n^{s-2}}}. \quad \text{(V.9)}$$

Since $K > 0$ is fixed and the RHS of (V.9) converges to zero for $s \geqslant 3$, the L_2 norm bound in (II.8) follows, under the assumption that ML equations (II.7) have a solution.

Next, suppose $s = 2$. Since $\|\bar{\boldsymbol{\beta}}_2(t) - \boldsymbol{\beta}_2\|_\infty = t\|\hat{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}_2\|_\infty$. Since $t \in [0,1]$, by Lemma 2,

$$\lambda_{\min}(\nabla^2 \ell_{n,2}(\bar{\boldsymbol{\beta}}_2(t))) \geqslant C_1' e^{-2t\|\hat{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}_2\|_\infty} n \geqslant C_1' e^{-2\|\hat{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}_2\|_\infty} n, \quad \text{(V.10)}$$

for some constant $C_1' > 0$ depending on M. Then

$$\begin{aligned} |g_2(1) - g_2(0)| &\geqslant g_2(1) - g_2(0) \\ &= \int_0^1 g_2'(t)\mathrm{d}t \\ &\geqslant C_1' n\|\hat{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}_2\|_2^2 e^{-2\|\hat{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}_2\|_\infty} \\ &\qquad . \qquad \text{(by (V.6) and (V.10))} \end{aligned}$$

Therefore, by (V.5) and Lemma 1, with probability $1 - o(1)$,

$$\|\hat{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}_2\|_2 e^{-2\|\hat{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}_2\|_\infty} \leqslant \frac{1}{C_1' n} \cdot \|\nabla \ell_{n,2}(\boldsymbol{\beta}_2)\|_2 \leqslant C', \quad \text{(V.11)}$$

for some constant $C' > 0$ depending on M. Hence, if there exists a bounded solution to (II.7), the L_2 norm rate will follow for $s = 2$.

To complete the proof we need to show that bounded solution to equation (II.7) exists. To this end, for $2 \leqslant s \leqslant r$, denote by $\mathcal{D}_s$, the set of all possible degree sequences in an s-uniform hypergraph on n vertices. Moreover, let $\mathcal{R}_s$ be the set of all expected degree sequences in a hypergraph on n vertices sampled from the s-uniform model (I.3). The following result shows that any convex combination of s-degree sequences in $\mathcal{D}_s$ can be reached by the limit of expected degree sequences of the s-uniform hypergraph $\boldsymbol{\beta}$-model. This was proved in the graph case ($s = 2$) by Chatterjee et al. [10, Theorem 1.4]. Here, we show that the same holds for all $2 \leqslant s \leqslant r$.

Proposition 1: Fix $2 \leqslant s \leqslant r$ and let $\mathcal{D}_s$ and $\mathcal{R}_s$ be as defined above. Then $\mathrm{conv}(\mathcal{D}_s) = \bar{\mathcal{R}}_s$, where $\mathrm{conv}(\mathcal{D}_s)$ denotes the convex hull of $\mathcal{D}_s$ and $\bar{\mathcal{R}}_s$ is the closure of $\mathcal{R}_s$.

The proof of the above result is given in Appendix X. Using this proposition we now show the existence of bounded solutions of the ML equations (II.7). Note that by Proposition 1, given $H_n \sim \mathsf{H}_{n,[r]}(n, \boldsymbol{B})$ the s-degree sequence $(d_s(1), \ldots, d_s(n)) \in \mathcal{D}_s \subseteq \bar{\mathcal{R}}_s$. This implies, there exists a sequence $\{\boldsymbol{x}_t\}_{t \geqslant 0} \in \mathcal{R}_s$ satisfying

$$\lim_{t\to\infty} \boldsymbol{x}_t = (d_s(1), \ldots, d_s(n)).$$

Since $\boldsymbol{x}_t \in \mathcal{R}_s$, there exists $\{\hat{\boldsymbol{\beta}}_2^{(t)}, \ldots, \hat{\boldsymbol{\beta}}_r^{(t)}\}$ such that

$$(\boldsymbol{x}_t)_v = \sum_{\{v_2,\ldots,v_s\} \in \binom{[n]\setminus\{v\}}{s-1}} \frac{e^{\hat{\beta}_{s,v}^{(t)} + \hat{\beta}_{s,v_2}^{(t)} + \ldots + \hat{\beta}_{s,v_s}^{(t)}}}{1 + e^{\hat{\beta}_{s,v}^{(t)} + \hat{\beta}_{s,v_2}^{(t)} + \ldots + \hat{\beta}_{s,v_s}^{(t)}}}, \quad \text{(V.12)}$$

for $1 \leqslant v \leqslant n$ and $2 \leqslant s \leqslant r$. In other words, for each $t \geqslant 0$, $\{\hat{\boldsymbol{\beta}}_2^{(t)}, \ldots, \hat{\boldsymbol{\beta}}_r^{(t)}\}$ is a solution of the ML equations (II.7) with $(d_s(1), d_s(2), \ldots, d_s(n))$ replaced by $\boldsymbol{x}_t$. By the previous argument, there exists $C > 0$ (not depending on t) such that with probability $1 - o(1)$,

$$\max_{2 \leqslant s \leqslant r} \|\hat{\boldsymbol{\beta}}_s^{(t)}\|_\infty \leqslant C,$$

for all $t \geqslant 0$. Therefore, the sequence $\{(\hat{\boldsymbol{\beta}}_2^{(t)}, \ldots, \hat{\boldsymbol{\beta}}_r^{(t)})\}_{t \geqslant 0}$ has a limit point. This limit point is a solution to (II.7) (by taking limit as $t \to \infty$ in (V.12)) and is bounded. Finally, since $\ell_{n,s}$ is strongly convex for $\boldsymbol{\beta} \in \mathcal{B}_M$ (see (V.2)), if the gradient equations have a bounded solution, it is the unique minimizer. Therefore, there exists a unique bounded solution to (II.7) which is the minimizer of $\ell_{n,s}$.

2) Proof of Lemma 1: Recalling (II.7) note that, for $v \in [n]$, v-th coordinate of the gradient of $\nabla \ell_{n,s}$ is given by:

$$\nabla \ell_{n,s}(\boldsymbol{\beta}_s)_v = \mathbb{E}[d_s(v)] - d_s(v) \quad \text{(V.13)}$$

where

$$\mathbb{E}[d_s(v)] = \sum_{\{v_2,\ldots,v_s\} \in \binom{[n]\setminus\{v\}}{s-1}} \frac{e^{\beta_{s,v} + \beta_{s,v_2} + \ldots + \beta_{s,v_s}}}{1 + e^{\beta_{s,v} + \beta_{s,v_2} + \ldots + \beta_{s,v_s}}}. \quad \text{(V.14)}$$

Since $d_s(v)$ is the sum of $O(n^{s-1})$ independent random variables bounded by 1, by Hoeffding's inequality and the union bound,

$$\mathbb{P}\left(\|\nabla \ell_{n,s}(\boldsymbol{\beta}_s)\|_\infty^2 \geqslant 4\, C_{s,M} n^{s-1} \log n\right) \leqslant \frac{1}{n^2},$$

for some constant $C_{s,M} > 0$ (depending on s and M). This establishes the second bound in (V.1).

Next, we prove the first bound in (V.1). Denote by $\mathbb{B}^n := \{\boldsymbol{x} \in \mathbb{R}^n : \|\boldsymbol{x}\|_2 \leqslant 1\}$ the unit ball in $\mathbb{R}^n$. By [48, Lemma 5.2], we can construct an $\frac{1}{2}$-net $\mathcal{V}$ of $\mathbb{B}^n$ satisfying $\log|\mathcal{V}| \leqslant C_1 n$ for some constant $C_1 > 0$. By a $\frac{1}{2}$-net of $\mathbb{B}^n$, we refer to a set $\mathcal{V}$, such that for all $u \in \mathbb{B}^n$ there exists $v \in \mathcal{V}$ satisfying $\|u - v\|_2 \leqslant \frac{1}{2}$. Now, for any unit vector $\boldsymbol{a} = (a_1, a_2, \ldots, a_n)^\top \in \mathbb{B}^n$ and the corresponding point $\boldsymbol{b} = (b_1, b_2, \ldots, b_n)^\top \in \mathcal{V}$, recalling (V.13) gives,

$$\begin{aligned}\sum_{v=1}^n a_v \nabla \ell_{n,s}(\boldsymbol{\beta}_s)_v &= \sum_{v=1}^n a_v \left(\mathbb{E}[d_s(v)] - d_s(v)\right) \\ &= \sum_{v=1}^n b_v \left(\mathbb{E}[d_s(v)] - d_s(v)\right) + \Delta, \quad \text{(V.15)}\end{aligned}$$

where

$$\begin{aligned}\Delta &:= \sum_{v=1}^n (a_v - b_v)\left(\mathbb{E}[d_s(v)] - d_s(v)\right) \\ &\leqslant \sqrt{\sum_{v=1}^n (a_v - b_v)^2 \sum_{v=1}^n \left(\mathbb{E}[d_s(v)] - d_s(v)\right)^2} \\ &\leqslant \frac{1}{2}\sqrt{\sum_{v=1}^n \left(\mathbb{E}[d_s(v)] - d_s(v)\right)^2} = \frac{1}{2}\|\nabla \ell_{n,s}(\boldsymbol{\beta}_s)\|_2, \end{aligned} \quad \text{(V.16)}$$

by the Cauchy-Schwarz inequality and the fact that $\|\boldsymbol{a} - \boldsymbol{b}\|_2 \leqslant \frac{1}{2}$. Using the above in (V.15) gives,

$$\begin{aligned}\sum_{v=1}^n a_v \nabla \ell_{n,s}(\boldsymbol{\beta}_s)_v &\leqslant \sum_{v=1}^n b_v \left(\mathbb{E}[d_s(v)] - d_s(v)\right) \\ &\quad + \frac{1}{2}\|\nabla \ell_{n,s}(\boldsymbol{\beta}_s)\|_2. \quad \text{(V.17)}\end{aligned}$$

The inequality (V.17) holds for all $\boldsymbol{a}$ such that $\|\boldsymbol{a}\|_2 = 1$. In particular,

$$\begin{aligned}&\max_{\boldsymbol{a}:\|\boldsymbol{a}\|_2 = 1} \sum_{v=1}^n a_v \nabla \ell_{n,s}(\boldsymbol{\beta}_s)_v \\ &\leqslant \sum_{v=1}^n b_v \left(\mathbb{E}[d_s(v)] - d_s(v)\right) + \frac{1}{2}\|\nabla \ell_{n,s}(\boldsymbol{\beta}_s)\|_2. \quad \text{(V.18)}\end{aligned}$$

By Cauchy-Schwartz inequality, the maximum value in the left hand side is equal to $\|\nabla \ell_{n,s}(\boldsymbol{\beta}_s)\|_2$. Therefore, by maximizing over $\boldsymbol{a} \in \mathbb{B}^n$ and $\boldsymbol{b} \in \mathcal{V}$ on both sides of (V.18) and rearranging the terms,

$$\|\nabla \ell_{n,s}(\boldsymbol{\beta}_s)\|_2 \leqslant 2 \max_{\boldsymbol{b} \in \mathcal{V}} \sum_{v=1}^n b_v \left(\mathbb{E}[d_s(v)] - d_s(v)\right). \quad \text{(V.19)}$$

For $\boldsymbol{e} = (u_1, u_2, \ldots, u_s) \in \binom{[n]}{s}$ denote $\boldsymbol{\beta}_{s,\boldsymbol{e}} = (\beta_{s,u_1}, \beta_{s,u_2}, \ldots, \beta_{s,u_s})^\top$. Hence, by (V.19), Hoeffding's inequality, and union bound,

$$\begin{aligned}&\mathbb{P}\left(\|\nabla \ell_{n,s}(\boldsymbol{\beta}_s)\|_2^2 > 4C^2 n^s\right) \\ &\leqslant \sum_{\boldsymbol{b} \in \mathcal{V}} \mathbb{P}\left(\sum_{v=1}^n b_v \left(\mathbb{E}[d_s(v)] - d_s(v)\right) > 2Cn^{\frac{s}{2}}\right) \\ &= \sum_{\boldsymbol{b} \in \mathcal{V}} \mathbb{P}\left(\sum_{v=1}^n \sum_{\boldsymbol{e} \in \binom{[n]}{s}: v \in \boldsymbol{e}} b_v \left\{\frac{e^{\boldsymbol{\beta}_{s,\boldsymbol{e}}^\top \mathbf{1}}}{1 + e^{\boldsymbol{\beta}_{s,\boldsymbol{e}}^\top \mathbf{1}}} - \mathbb{1}\{\boldsymbol{e} \in E(H_n)\}\right\} \right. \\ &\qquad \left. > 2Cn^{\frac{s}{2}}\right) \\ &\leqslant \sum_{\boldsymbol{b} \in \mathcal{V}} e^{-\frac{2C^2 n}{\sum_{v=1}^n b_v^2}} \leqslant 2^{C_1 n} e^{-2C^2 n} \to 0,\end{aligned}$$

by choosing $C > C_1$ to be large enough. This proves the first inequality in (V.1). $\square$

3) Proof of Lemma 2: For $\boldsymbol{e} = (u_1, u_2, \ldots, u_s) \in \binom{[n]}{s}$ and $\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots, \beta_n) \in \mathbb{R}^n$, denote $\boldsymbol{\beta}_{\boldsymbol{e}} = (\beta_{u_1}, \beta_{u_2}, \ldots, \beta_{u_s})^\top$. Recalling (II.7) note that, the Hessian matrix $\nabla^2 \ell_{n,s}$ can be expressed as:

$$\nabla^2 \ell_{n,s}(\boldsymbol{\beta}) = \sum_{u,v \in [n]} \sum_{\boldsymbol{e} \in \binom{[n]}{s}} \frac{e^{\boldsymbol{\beta}_{\boldsymbol{e}}^\top \mathbf{1}}}{(1 + e^{\boldsymbol{\beta}_{\boldsymbol{e}}^\top \mathbf{1}})^2} \boldsymbol{\eta}_u \boldsymbol{\eta}_v^\top \mathbb{1}\{u, v \in \boldsymbol{e}\},$$

where $\boldsymbol{\eta}_u$ is the u-th basis vector in $\mathbb{R}^n$, for $1 \leqslant u \leqslant n$.

Note that for $\boldsymbol{\beta} \in \mathbb{R}^n$, since $\boldsymbol{\beta}_s \in \mathcal{B}_M$, we have,

$$|\mathbf{1}^\top \boldsymbol{\beta}_{\boldsymbol{e}}| \leqslant s\|\boldsymbol{\beta}\|_\infty \leqslant s\|\boldsymbol{\beta}_s\|_\infty + s\|\boldsymbol{\beta}_s - \boldsymbol{\beta}\|_\infty.$$

Hence,

$$\frac{1}{4}e^{-s(M+\|\boldsymbol{\beta}_s-\boldsymbol{\beta}\|_\infty)} \leqslant \frac{e^{\mathbf{1}^\top\boldsymbol{\beta}_{\boldsymbol{e}}}}{(1+e^{\mathbf{1}^\top\boldsymbol{\beta}_{\boldsymbol{e}}})^2} \leqslant 1. \tag{V.20}$$

For $\boldsymbol{x} \in \mathbb{R}^n$, consider

$$\begin{aligned}
&\boldsymbol{x}^\top\nabla^2\ell_{n,s}(\boldsymbol{\beta})\boldsymbol{x} \\
&= \sum_{u,v\in[n]}\sum_{\boldsymbol{e}\in\binom{[n]}{s}} \frac{e^{\boldsymbol{\beta}_{\boldsymbol{e}}^\top\mathbf{1}}}{(1+e^{\boldsymbol{\beta}_{\boldsymbol{e}}^\top\mathbf{1}})^2} x_u x_v \mathbb{1}\{u,v\in\boldsymbol{e}\} \\
&= \sum_{\boldsymbol{e}\in\binom{[n]}{s}} \frac{e^{\boldsymbol{\beta}_{\boldsymbol{e}}^\top\mathbf{1}}}{(1+e^{\boldsymbol{\beta}_{\boldsymbol{e}}^\top\mathbf{1}})^2}\left(\sum_{u,v\in[n]} x_u x_v \mathbb{1}\{u,v\in\boldsymbol{e}\}\right) \\
&= \sum_{\boldsymbol{e}\in\binom{[n]}{s}} \frac{e^{\boldsymbol{\beta}_{\boldsymbol{e}}^\top\mathbf{1}}}{(1+e^{\boldsymbol{\beta}_{\boldsymbol{e}}^\top\mathbf{1}})^2}\left(\sum_{u\in[n]} x_u \mathbb{1}\{u\in\boldsymbol{e}\}\right)^2 \\
&\geqslant \frac{1}{4}e^{-s(M+\|\boldsymbol{\beta}_s-\boldsymbol{\beta}\|_\infty)} \sum_{\boldsymbol{e}\in\binom{[n]}{s}} \left(\sum_{u\in[n]} x_u \mathbb{1}\{u\in\boldsymbol{e}\}\right)^2,
\end{aligned}$$

where the last step uses (V.20). Observe that for any $\boldsymbol{x} \in \mathbb{R}^n$

$$\sum_{\boldsymbol{e}\in\binom{[n]}{s}} \left(\sum_{u\in[n]} x_u \mathbb{1}\{u\in\boldsymbol{e}\}\right)^2 = \boldsymbol{x}^\top\boldsymbol{L}\boldsymbol{x},$$

where

$$\begin{aligned}
\boldsymbol{L} &:= \sum_{u,v\in[n]}\sum_{\boldsymbol{e}\in\binom{[n]}{s}} \boldsymbol{\eta}_u\boldsymbol{\eta}_v^\top \mathbb{1}\{u,v\in\boldsymbol{e}\} \\
&= \left(\binom{n-1}{s-1} - \binom{n-2}{s-2}\right)\boldsymbol{I}_n + \binom{n-2}{s-2}\mathbf{1}\mathbf{1}^\top, \qquad \text{(V.21)}
\end{aligned}$$

where $\boldsymbol{I}_n$ is the $n\times n$ identity matrix and $\mathbf{1} = (1,1,\ldots,1)^\top$. Similarly, we can show from (V.20) that for any $\boldsymbol{x} \in \mathbb{R}^n$

$$\boldsymbol{x}^\top\nabla^2\ell_{n,s}(\boldsymbol{\beta})\boldsymbol{x} \leqslant \boldsymbol{x}^\top\boldsymbol{L}\boldsymbol{x}.$$

Thus, for $\boldsymbol{\beta} \in \mathbb{R}^n$

$$\begin{aligned}
\frac{1}{4}e^{-s(M+\|\boldsymbol{\beta}_s-\boldsymbol{\beta}\|_\infty)}\lambda_{\min}(\boldsymbol{L}) &\leqslant \lambda_{\min}(\nabla^2\ell_{n,s}(\boldsymbol{\beta})) \\
&\leqslant \lambda_{\max}(\nabla^2\ell_{n,s}(\boldsymbol{\beta})) \\
&\leqslant \lambda_{\max}(\boldsymbol{L}). \qquad \text{(V.22)}
\end{aligned}$$

Note that $\boldsymbol{L}$ is a circulant matrix with 2 non-zero eigenvalues:

$$\binom{n-1}{s-1} \quad \text{and} \quad \binom{n-1}{s-1} - \binom{n-2}{s-2}.$$

Further, there exists constants $C_1'', C_2'' > 0$ (depending on r), such that

$$\binom{n-1}{s-1} \leqslant C_1'' n^{s-1} \text{ and } \binom{n-1}{s-1} - \binom{n-2}{s-2} \geqslant C_2'' n^{s-1}.$$

This implies, from (V.22), that there exists constants C_1', $C_2' > 0$ (depending on r and M) such that (V.2) hold. The result in (V.3) from hold from (V.2) by noting that $\|\boldsymbol{\beta}_s - \boldsymbol{\beta}\|_\infty \leqslant \|\boldsymbol{\beta}_s - \boldsymbol{\beta}\|_2$.

B. Convergence Rate in the L_∞ Norm

Suppose $H_n \sim \mathsf{H}_{n,[r]}(n,\boldsymbol{B})$ be as in the statement of Theorem 1. From the arguments in Appendix V-A we know that, with probability $1-o(1)$, the ML equations (II.7) have a bounded solution $\hat{\boldsymbol{B}} = (\hat{\boldsymbol{\beta}}_2,\ldots,\hat{\boldsymbol{\beta}}_r)$, that is, $\nabla\ell_{n,s}(\hat{\boldsymbol{\beta}}_s) = 0$, for $2 \leqslant s \leqslant r$, and $\max_{2\leqslant s\leqslant r}\|\hat{\boldsymbol{\beta}}_s\|_\infty = O(1)$. To establish the rate in L_∞ norm we decompose the likelihood for the s-th layer as follows.

$$\begin{aligned}
\ell_{n,s}(\boldsymbol{\beta}) &= \sum_{\{v_1,v_2,\ldots,v_s\}\in\binom{[n]}{s}} \log\left(1+e^{\beta_{v_1}+\ldots+\beta_{v_s}}\right) - \sum_{v=1}^n \beta_v d_s(v) \\
&= \sum_{\boldsymbol{e}\in\binom{[n]}{s}} \left\{\log\left(1+e^{\boldsymbol{\beta}_e^\top\mathbf{1}}\right) - \mathbb{1}\{\boldsymbol{e}\in E(H_n)\}\boldsymbol{\beta}_e^\top\mathbf{1}\right\} \\
&= \ell_{n,s}^+(\beta_u|\boldsymbol{\beta}_{\bar{u}}) + \ell_{n,s}^-(\boldsymbol{\beta}_{\bar{u}}), \qquad \text{(V.23)}
\end{aligned}$$

where $\boldsymbol{\beta}_{\bar{u}} = (\beta_1,\ldots,\beta_{u-1},\beta_{u+1},\ldots,\beta_n)$,

$$\begin{aligned}
\ell_{n,s}^+(\beta_u|\boldsymbol{\beta}_{\bar{u}}) &:= \sum_{\boldsymbol{e}\in\binom{[n]}{s}:u\in\boldsymbol{e}} \Big\{\log\left(1+e^{\boldsymbol{\beta}_e^\top\mathbf{1}}\right) \\
&\qquad\qquad - \mathbb{1}\{\boldsymbol{e}\in E(H_n)\}\boldsymbol{\beta}_e^\top\mathbf{1}\Big\}, \\
\ell_{n,s}^-(\boldsymbol{\beta}_{\bar{u}}) &:= \sum_{\boldsymbol{e}\in\binom{[n]}{s}:u\notin\boldsymbol{e}} \Big\{\log\left(1+e^{\boldsymbol{\beta}_e^\top\mathbf{1}}\right) \\
&\qquad\qquad - \mathbb{1}\{\boldsymbol{e}\in E(H_n)\}\boldsymbol{\beta}_e^\top\mathbf{1}\Big\}. \qquad \text{(V.24)}
\end{aligned}$$

Fix a constant $K > 0$ and define

$$\hat{\boldsymbol{\beta}}_{s,\bar{u}}^\circ = \operatorname{argmin}_{\boldsymbol{\beta}_{\bar{u}}:\|\boldsymbol{\beta}_{\bar{u}}-\boldsymbol{\beta}_{s,\bar{u}}\|_2\leqslant K}\ \ell_{n,s}^-(\boldsymbol{\beta}_{\bar{u}}), \tag{V.25}$$

where $\boldsymbol{\beta}_{s,\bar{u}} = (\beta_{s,1},\ldots,\beta_{s,u-1},\beta_{s,u+1},\ldots,\beta_{s,n})$. This is the leave-one-out ML estimate on the constrained set $\|\boldsymbol{\beta}_{\bar{u}} - \boldsymbol{\beta}_{s,\bar{u}}\|_2 \leqslant K$. First we bound the difference (in L_2 norm) of constrained leave-one-out ML estimate defined above and the leave-one-out true parameter $\boldsymbol{\beta}_{s,\bar{u}}$.

Lemma 3: Let $\hat{\boldsymbol{\beta}}_{s,\bar{u}}^\circ$ and $\boldsymbol{\beta}_{s,\bar{u}}$ be as defined above. Then, for $u \in [n]$, with probability $1-o(1)$,

$$\max_{u\in[n]}\|\hat{\boldsymbol{\beta}}_{s,\bar{u}}^\circ - \boldsymbol{\beta}_{s,\bar{u}}\|_2^2 \lesssim_{s,M,K} \frac{1}{n^{s-2}}. \tag{V.26}$$

Proof: To begin with, observe that

$$\begin{aligned}
\ell_{n,s}^-(\boldsymbol{\beta}_{s,\bar{u}}) &\geqslant \ell_{n,s}^-(\hat{\boldsymbol{\beta}}_{s,\bar{u}}^\circ) \\
&= \ell_{n,s}^-(\boldsymbol{\beta}_{s,\bar{u}}) + (\hat{\boldsymbol{\beta}}_{s,\bar{u}}^\circ - \boldsymbol{\beta}_{s,\bar{u}})^\top\nabla\ell_{n,s}^-(\boldsymbol{\beta}_{s,\bar{u}}) \\
&\quad + \frac{1}{2}(\hat{\boldsymbol{\beta}}_{s,\bar{u}}^\circ - \boldsymbol{\beta}_{s,\bar{u}})^\top\nabla^2\ell_{n,s}^-(\tilde{\boldsymbol{\beta}})(\hat{\boldsymbol{\beta}}_{s,\bar{u}}^\circ - \boldsymbol{\beta}_{s,\bar{u}}),
\end{aligned}$$

where $\|\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}_{s,\bar{u}}\|_2 \leqslant \|\hat{\boldsymbol{\beta}}_{s,\bar{u}}^\circ - \boldsymbol{\beta}_{s,\bar{u}}\|_2 \leqslant K$. This implies,

$$\begin{aligned}
&\|\hat{\boldsymbol{\beta}}_{s,\bar{u}}^\circ - \boldsymbol{\beta}_{s,\bar{u}}\|_2 \cdot \|\nabla\ell_{n,s}^-(\boldsymbol{\beta}_{s,\bar{u}})\|_2 \\
&\quad\geqslant -(\hat{\boldsymbol{\beta}}_{s,\bar{u}}^\circ - \boldsymbol{\beta}_{s,\bar{u}})^\top\nabla\ell_{n,s}^-(\boldsymbol{\beta}_{s,\bar{u}}) \\
&\quad\geqslant \frac{1}{2}(\hat{\boldsymbol{\beta}}_{s,\bar{u}}^\circ - \boldsymbol{\beta}_{s,\bar{u}})^\top\nabla^2\ell_{n,s}^-(\tilde{\boldsymbol{\beta}})(\hat{\boldsymbol{\beta}}_{s,\bar{u}}^\circ - \boldsymbol{\beta}_{s,\bar{u}}).
\end{aligned} \tag{V.27}$$

By Lemma 2,

$$\begin{aligned}
&(\hat{\boldsymbol{\beta}}_{s,\bar{u}}^\circ - \boldsymbol{\beta}_{s,\bar{u}})^\top\nabla^2\ell_{n,s}^-(\tilde{\boldsymbol{\beta}})(\hat{\boldsymbol{\beta}}_{s,\bar{u}}^\circ - \boldsymbol{\beta}_{s,\bar{u}}) \\
&\qquad\gtrsim_{s,M,K} \|\hat{\boldsymbol{\beta}}_{s,\bar{u}}^\circ - \boldsymbol{\beta}_{s,\bar{u}}\|^2 n^{s-1}.
\end{aligned} \tag{V.28}$$

Also, by Lemma 1, $\|\nabla \ell^-_{n,s}(\boldsymbol{\beta}_{s,\bar{u}})\|_2^2 \lesssim_{s,M,K} n^s$ with probability $1 - O(\frac{1}{n^2})$. Plugging in the above inequalities in (V.27), and using the union bound we get (V.26). □

Next, we bound the difference between the constrained leave-one-out ML estimate $\hat{\boldsymbol{\beta}}^\circ_{s,\bar{u}}$ and the (unconstrained) leave-one-out ML estimate $\hat{\boldsymbol{\beta}}_{s,\bar{u}} = (\hat{\beta}_{s,1}, \ldots, \hat{\beta}_{s,u-1}, \hat{\beta}_{s,u+1}, \ldots, \hat{\beta}_{s,n})$.

Lemma 4: Let $\hat{\boldsymbol{\beta}}^\circ_{s,\bar{u}}$ and $\hat{\boldsymbol{\beta}}_{s,\bar{u}}$ be as defined above. Then, with probability $1 - o(1)$,

$$\max_{u \in [n]} \|\hat{\boldsymbol{\beta}}^\circ_{s,\bar{u}} - \hat{\boldsymbol{\beta}}_{s,\bar{u}}\|_2^2 \lesssim_{s,M,K} \frac{1}{n} + \frac{\|\hat{\boldsymbol{\beta}}_s - \boldsymbol{\beta}_s\|_\infty^2}{n^{s-1}}. \tag{V.29}$$

Proof: [Please read this proof a bit….there may be some mistakes….I made some major changes.] By the definition of $\hat{\boldsymbol{\beta}}^\circ_{s,\bar{u}}$ (recall (V.25))

$$\begin{aligned}\ell^-_{n,s}(\hat{\boldsymbol{\beta}}_{s,\bar{u}}) &\geqslant \ell^-_{n,s}(\hat{\boldsymbol{\beta}}^\circ_{s,\bar{u}}) \\ &= \ell^-_{n,s}(\hat{\boldsymbol{\beta}}_{s,\bar{u}}) + (\hat{\boldsymbol{\beta}}^\circ_{s,\bar{u}} - \hat{\boldsymbol{\beta}}_{s,\bar{u}})^\top \nabla \ell^-_{n,s}(\hat{\boldsymbol{\beta}}_{s,\bar{u}}) \\ &\quad + (\hat{\boldsymbol{\beta}}^\circ_{s,\bar{u}} - \hat{\boldsymbol{\beta}}_{s,\bar{u}})^\top \nabla \ell^-_{n,s}(\bar{\boldsymbol{\beta}})(\hat{\boldsymbol{\beta}}^\circ_{s,\bar{u}} - \hat{\boldsymbol{\beta}}_{s,\bar{u}}),\end{aligned}$$

where $\|\bar{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_{s,\bar{u}}\|_2 \leqslant \|\hat{\boldsymbol{\beta}}^\circ_{s,\bar{u}} - \hat{\boldsymbol{\beta}}_{s,\bar{u}}\|_2$. Note that $\|\hat{\boldsymbol{\beta}}^\circ_{s,\bar{u}} - \hat{\boldsymbol{\beta}}_{s,\bar{u}}\|_2 = O(1)$, since $\|\hat{\boldsymbol{\beta}}_s\|_\infty = O(1)$ and $\|\hat{\boldsymbol{\beta}}^\circ_{s,\bar{u}}\| = O(1)$. Therefore, we have $K > 0$ such that by Lemma 2,

$$\|\hat{\boldsymbol{\beta}}^\circ_{s,\bar{u}} - \hat{\boldsymbol{\beta}}_{s,\bar{u}}\|_2^2 \lesssim_{s,M,K} \frac{\|\nabla \ell^-_{n,s}(\hat{\boldsymbol{\beta}}_{s,\bar{u}})\|_2^2}{n^{2(s-1)}}. \tag{V.30}$$

Since $\nabla \ell_{n,s}(\hat{\boldsymbol{\beta}}_s) = 0$, that is, $\frac{\partial}{\partial \beta_v} \ell_{n,s}(\hat{\boldsymbol{\beta}}_s) = 0$, for $v \in [n]$. Hence, we have from (V.23),

$$\begin{aligned}\frac{\partial}{\partial \beta_v} \ell^-_{n,s}(\hat{\boldsymbol{\beta}}_{s,\bar{u}}) &= -\frac{\partial}{\partial \beta_v} \ell^+_{n,s}(\hat{\beta}_{s,u}|\hat{\boldsymbol{\beta}}_{s,\bar{u}}) \\ &= -\sum_{\boldsymbol{e} \in \binom{[n]}{s}: \{u,v\} \in \boldsymbol{e}} \{X_{\boldsymbol{e}} - \psi(\mathbf{1}^\top \hat{\boldsymbol{\beta}}_{s,\boldsymbol{e}})\},\end{aligned}$$

where $\psi(x) := \frac{e^x}{1+e^x}$ and $X_{\boldsymbol{e}} = \mathbb{1}\{\boldsymbol{e} \in E(H_n)\}$ for $\boldsymbol{e} \in \binom{[n]}{s}$. This implies,

$$\begin{aligned}&\|\nabla \ell^-_{n,s}(\hat{\boldsymbol{\beta}}_{s,\bar{u}})\|_2^2 \\ &= \sum_{v \in [n] \setminus \{u\}} \left(\sum_{\boldsymbol{e} \in \binom{[n]}{s}: \{u,v\} \in \boldsymbol{e}} \{X_{\boldsymbol{e}} - \psi(\hat{\boldsymbol{\beta}}_{s,\boldsymbol{e}}^\top 1)\} \right)^2 \\ &\lesssim \sum_{v \in [n] \setminus \{u\}} \left[\left(\sum_{\boldsymbol{e} \in \binom{[n]}{s}: \{u,v\} \in \boldsymbol{e}} \{X_{\boldsymbol{e}} - \psi(\mathbf{1}^\top \boldsymbol{\beta}_{s,\boldsymbol{e}})\} \right)^2 \right. \\ &\qquad \left. + \left(\sum_{\boldsymbol{e} \in \binom{[n]}{s}: \{u,v\} \in \boldsymbol{e}} \{\psi(\mathbf{1}^\top \hat{\boldsymbol{\beta}}_{s,\boldsymbol{e}}) - \psi(\mathbf{1}^\top \boldsymbol{\beta}_{s,\boldsymbol{e}})\} \right)^2 \right] \\ &\lesssim_r \sum_{v \in [n] \setminus \{u\}} \left(\sum_{\boldsymbol{e} \in \binom{[n]}{s}: \{u,v\} \in \boldsymbol{e}} \{X_{\boldsymbol{e}} - \psi(\mathbf{1}^\top \boldsymbol{\beta}_{s,\boldsymbol{e}})\} \right)^2 \\ &\qquad + n^{2s-3} \|\hat{\boldsymbol{\beta}}_s - \boldsymbol{\beta}_s\|_\infty^2,\end{aligned} \tag{V.31}$$

using

$$\begin{aligned}|\psi(\mathbf{1}^\top \hat{\boldsymbol{\beta}}_{s,\boldsymbol{e}}) - \psi(\mathbf{1}^\top \boldsymbol{\beta}_{s,\boldsymbol{e}})| &\lesssim |\mathbf{1}^\top \hat{\boldsymbol{\beta}}_{s,\boldsymbol{e}} - \mathbf{1}^\top \boldsymbol{\beta}_{s,\boldsymbol{e}}| \\ &\lesssim_r \|\hat{\boldsymbol{\beta}}_s - \boldsymbol{\beta}_s\|_\infty^2.\end{aligned} \tag{V.32}$$

By (V.30) and (V.31), to prove the result in (V.29) it suffices show the following holds with probability $1 - o(1)$,

$$\begin{aligned}\max_{1 \leqslant u \leqslant n} \sum_{v \in [n] \setminus \{u\}} \left(\sum_{\boldsymbol{e} \in \binom{[n]}{s}: u,v \in \boldsymbol{e}} \left\{ \mathbb{1}\{\boldsymbol{e} \in E(H_n)\} - \psi(\boldsymbol{\beta}_{s,\boldsymbol{e}}^\top \mathbf{1}) \right\} \right)^2 \\ \lesssim n^{s-1}.\end{aligned} \tag{V.33}$$

This is proved in Appendix V-B.1. □

We now apply the above lemmas to derive the bound in the L_∞ norm. To begin with note that since $\ell_{n,s}(\hat{\boldsymbol{\beta}}_s) = \min_{\boldsymbol{\beta}_s} \ell_{n,s}(\boldsymbol{\beta}_s)$,

$$\begin{aligned}\ell^+_{n,s}(\beta_{s,u}|\hat{\boldsymbol{\beta}}_{s,\bar{u}}) + \ell^-_{n,s}(\hat{\boldsymbol{\beta}}_{s,\bar{u}}) &\geqslant \ell_{n,s}(\hat{\boldsymbol{\beta}}_s) \\ &= \ell^+_{n,s}(\hat{\beta}_{s,u}|\hat{\boldsymbol{\beta}}_{s,\bar{u}}) + \ell^-_{n,s}(\hat{\boldsymbol{\beta}}_{s,\bar{u}}).\end{aligned} \tag{V.34}$$

The above inequality implies

$$\begin{aligned}\ell^+_{n,s}(\beta_{s,u}|\hat{\boldsymbol{\beta}}_{s,\bar{u}}) &\geqslant \ell^+_{n,s}(\hat{\beta}_{s,u}|\hat{\boldsymbol{\beta}}_{s,\bar{u}}) \\ &= \ell^+_{n,s}(\beta_{s,u}|\hat{\boldsymbol{\beta}}_{s,\bar{u}}) \\ &\quad + (\hat{\beta}_{s,u} - \beta_{s,u}) \frac{\partial}{\partial \beta_u} \ell^+_{n,s}(\beta_{s,u}|\hat{\boldsymbol{\beta}}_{s,\bar{u}}) \\ &\quad + \frac{1}{2}(\hat{\beta}_{s,u} - \beta_{s,u})^2 \frac{\partial^2}{\partial \beta_u^2} \ell^+_{n,s}(\tilde{\beta}|\hat{\boldsymbol{\beta}}_{s,\bar{u}}),\end{aligned}$$

where $\tilde{\beta}$ is a convex combination of $\hat{\beta}_{s,u}$ and $\beta_{s,u}$. Therefore,

$$(\hat{\beta}_{s,u} - \beta_{s,u})^2 \leqslant \frac{4|\frac{\partial}{\partial \beta_u} \ell^+_{n,s}(\beta_{s,u}|\hat{\boldsymbol{\beta}}_{s,\bar{u}})|^2}{|\frac{\partial^2}{\partial \beta_u^2} \ell^+_{n,s}(\tilde{\beta}|\hat{\boldsymbol{\beta}}_{s,\bar{u}})|^2}. \tag{V.35}$$

From arguments in Appendix V-A we know that with probability $1 - o(1)$, $\|\hat{\boldsymbol{\beta}}_s - \boldsymbol{\beta}_s\|_\infty \leqslant \|\hat{\boldsymbol{\beta}}_s - \boldsymbol{\beta}_s\|_2 \lesssim 1$. Note that for $\boldsymbol{\beta} \in \mathbb{R}^n$ such that $\|\boldsymbol{\beta} - \boldsymbol{\beta}_s\|_\infty \lesssim 1$, we have $\|\boldsymbol{\beta}\|_\infty \lesssim 1$ and hence, $|\mathbf{1}^\top \boldsymbol{\beta}_{\boldsymbol{e}}| \lesssim 1$. This implies, $\psi(\mathbf{1}^\top \boldsymbol{\beta}_{\boldsymbol{e},s})(1 - \psi(\mathbf{1}^\top \boldsymbol{\beta}_{\boldsymbol{e},s})) \gtrsim 1$ and hence,

$$\begin{aligned}\frac{\partial^2}{\partial \beta_u^2} \ell^+_{n,s}(\tilde{\beta}|\hat{\boldsymbol{\beta}}_{s,\bar{u}}) = \sum_{\boldsymbol{e} \in \binom{[n]}{s}: u \in \boldsymbol{e}} \psi(\mathbf{1}^\top \bar{\boldsymbol{\beta}}_{\boldsymbol{e},s})(1 - \psi(\mathbf{1}^\top \bar{\boldsymbol{\beta}}_{\boldsymbol{e},s})) \\ \gtrsim n^{s-1},\end{aligned} \tag{V.36}$$

where $\bar{\boldsymbol{\beta}}_s = (\hat{\beta}_{s,1}, \ldots, \hat{\beta}_{s,u-1}, \beta_{s,u}, \hat{\beta}_{s,u+1}, \ldots, \hat{\beta}_{s,n})^\top$. Hence, (V.35) implies,

$$(\hat{\beta}_{s,u} - \beta_{s,u})^2 \lesssim \frac{|\frac{\partial}{\partial \beta_u} \ell^+_{n,s}(\beta_{s,u}|\hat{\boldsymbol{\beta}}_{s,\bar{u}})|^2}{n^{2s-2}}. \tag{V.37}$$

Now, we bound $|\frac{\partial}{\partial \beta_u} \ell^+_{n,s}(\beta_{s,u}|\hat{\boldsymbol{\beta}}_{s,\bar{u}})|^2$. For this define

$$\bar{\boldsymbol{\beta}}^\circ_s = ([\hat{\boldsymbol{\beta}}^\circ_{s,\bar{u}}]_1, \ldots, [\hat{\boldsymbol{\beta}}^\circ_{s,\bar{u}}]_{u-1}, \beta_{s,u}, [\hat{\boldsymbol{\beta}}^\circ_{s,\bar{u}}]_{u+1}, \ldots, [\hat{\boldsymbol{\beta}}^\circ_{s,\bar{u}}]_n)^\top.$$

Then we have

$$\left|\frac{\partial}{\partial \beta_u}\ell^+_{n,s}(\beta_{s,u}|\hat{\boldsymbol{\beta}}_{s,\bar{u}})\right| = \left|\sum_{\boldsymbol{e}\in\binom{[n]}{s}:u\in\boldsymbol{e}}\{X_{\boldsymbol{e}} - \psi(\mathbf{1}^\top\bar{\boldsymbol{\beta}}_{\boldsymbol{e},s})\}\right| \leqslant T_1(u) + T_2(u) + T_3(u), \tag{V.38}$$

where

$$T_1(u) := \left|\sum_{\boldsymbol{e}\in\binom{[n]}{s}:u\in\boldsymbol{e}}\{X_{\boldsymbol{e}} - \psi(\mathbf{1}^\top\boldsymbol{\beta}_{\boldsymbol{e},s})\}\right|,$$

$$T_2(u) := \left|\sum_{\boldsymbol{e}\in\binom{[n]}{s}:u\in\boldsymbol{e}}\{\psi(\mathbf{1}^\top\bar{\boldsymbol{\beta}}^\circ_{\boldsymbol{e},s}) - \psi(\mathbf{1}^\top\boldsymbol{\beta}_{\boldsymbol{e},s})\}\right|,$$

and

$$T_3(u) := \left|\sum_{\boldsymbol{e}\in\binom{[n]}{s}:u\in\boldsymbol{e}}\{\psi(\mathbf{1}^\top\bar{\boldsymbol{\beta}}^\circ_{\boldsymbol{e},s}) - \psi(\mathbf{1}^\top\bar{\boldsymbol{\beta}}_{\boldsymbol{e},s})\}\right|.$$

Note that since $\{X_{\boldsymbol{e}}\}_{\boldsymbol{e}\in\binom{[n]}{s}}$ are independent and bounded random variables, using Hoeffding's inequality and union bound gives

$$\max_{u\in[n]} T_1(u) \lesssim \sqrt{n^{s-1}\log n},$$

with probability $1 - o(1)$. Next, we consider $T_2(u)$. By Lemma 3, with probability $1 - o(1)$,

$$\begin{aligned}\max_{u\in[n]} T_2(u) &\lesssim \max_{u\in[n]} \sum_{\boldsymbol{e}\in\binom{[n]}{s}:u\in\boldsymbol{e}} \left\{\sum_{v\in\boldsymbol{e}} |\beta_{s,v} - [\hat{\boldsymbol{\beta}}^\circ_{s,\bar{u}}]_v|\right\} \\ &= \max_{u\in[n]} \sum_{v\in[n]\setminus\{u\}} n^{s-2}|\beta_{s,v} - [\hat{\boldsymbol{\beta}}^\circ_{s,\bar{u}}]_v| \\ &\lesssim n^{s-\frac{3}{2}} \max_{u\in[n]} \|\boldsymbol{\beta}_{s,\bar{u}} - \hat{\boldsymbol{\beta}}^\circ_{s,\bar{u}}\|_2 \lesssim \sqrt{n^{s-1}}.\end{aligned}$$

A similar argument shows that, with probability $1 - o(1)$, $\max_{u\in[n]} T_3(u) \lesssim n^{s-\frac{3}{2}}\|\hat{\boldsymbol{\beta}}_{s,\bar{u}} - \hat{\boldsymbol{\beta}}^\circ_{s,\bar{u}}\|_2$. Combining the bounds on T_1, T_2 and T_3 with (V.37) and (V.38) gives, with probability $1 - o(1)$,

$$\|\hat{\boldsymbol{\beta}}_s - \boldsymbol{\beta}_s\|_\infty \lesssim \sqrt{\frac{\log n}{n^{s-1}}} + \frac{\max_{u\in[n]}\|\hat{\boldsymbol{\beta}}_{s,\bar{u}} - \hat{\boldsymbol{\beta}}^\circ_{s,\bar{u}}\|_2}{\sqrt{n}}. \tag{V.39}$$

Applying (V.39) in (V.29) now gives, with probability $1 - o(1)$,

$$\begin{aligned}&\max_{u\in[n]}\|\hat{\boldsymbol{\beta}}^\circ_{s,\bar{u}} - \hat{\boldsymbol{\beta}}_{s,\bar{u}}\|_2 \\ &\lesssim_{s,M,K} \sqrt{\frac{1}{n^{s-1}}} + \frac{\|\hat{\boldsymbol{\beta}}_s - \boldsymbol{\beta}_s\|_\infty}{\sqrt{n}} \\ &\lesssim_{s,M,K} \sqrt{\frac{1}{n^{s-1}}} + \frac{\max_{u\in[n]}\|\hat{\boldsymbol{\beta}}^\circ_{s,\bar{u}} - \hat{\boldsymbol{\beta}}_{s,\bar{u}}\|_2}{n} \lesssim \sqrt{\frac{1}{n^{s-1}}}.\end{aligned}$$

Using this inequality with (V.39) gives, with probability $1 - o(1)$,

$$\|\hat{\boldsymbol{\beta}}_s - \boldsymbol{\beta}_s\|_\infty \lesssim_{s,M,K} \sqrt{\frac{\log n}{n^{s-1}}},$$

establishing the desired bound in (II.8).

1) Proof of (V.33): *Proof:* Denote by $\mathbb{B}^{n-1} = \{\boldsymbol{x} \in \mathbb{R}^{n-1} : \|\boldsymbol{x}\|_2 \leqslant 1\}$. Using [48, Lemma 5.2], we can construct an $\frac{1}{2}$-net $\mathcal{V}_1$ of $\mathbb{B}^{n-1}$ (defined in Section V-A.2) satisfying $\log|\mathcal{V}_1| \leqslant C_2 n$ for some constant $C_2 > 0$. Now, for any $u \in [n]$, any unit vector $\tilde{\boldsymbol{a}} = (\tilde{a}_1, \ldots, \tilde{a}_{u-1}, \tilde{a}_{u+1}, \ldots, \tilde{a}_n)^\top \in \mathbb{B}^{n-1}$ and the corresponding point $\tilde{\boldsymbol{b}} = (\tilde{b}_1, \ldots, \tilde{b}_{u-1}, \tilde{b}_{u+1}, \ldots, \tilde{b}_n)^\top \in \mathcal{V}_1$,

$$\begin{aligned}&\sum_{v\in[n]\setminus\{u\}} \tilde{a}_v \left\{\sum_{\boldsymbol{e}\in\binom{[n]}{s}:u,v\in\boldsymbol{e}} \left\{X_{\boldsymbol{e}} - \frac{e^{\mathbf{1}^\top\boldsymbol{\beta}_{\boldsymbol{e}}}}{1+e^{\mathbf{1}^\top\boldsymbol{\beta}_{\boldsymbol{e}}}}\right\}\right\} \\ &= \sum_{v\in[n]\setminus\{u\}} \tilde{b}_v \left\{\sum_{\boldsymbol{e}\in\binom{[n]}{s}, u,v\in\boldsymbol{e}} \left\{X_{\boldsymbol{e}} - \frac{e^{\mathbf{1}^\top\boldsymbol{\beta}_{\boldsymbol{e}}}}{1+e^{\mathbf{1}^\top\boldsymbol{\beta}_{\boldsymbol{e}}}}\right\}\right\} + \Delta_u,\end{aligned} \tag{V.40}$$

where

$$\Delta_u := \sum_{v\in[n]\setminus\{u\}} (\tilde{a}_v - \tilde{b}_v)\left\{\sum_{\boldsymbol{e}\in\binom{[n]}{s}:u,v\in\boldsymbol{e}} \left\{X_{\boldsymbol{e}} - \frac{e^{\mathbf{1}^\top\boldsymbol{\beta}_{\boldsymbol{e}}}}{1+e^{\mathbf{1}^\top\boldsymbol{\beta}_{\boldsymbol{e}}}}\right\}\right\}.$$

Proceeding as in (V.16), for all $u \in [n]$, we can show

$$|\Delta_u| \leqslant \frac{1}{2}\sqrt{\sum_{v\in[n]\setminus\{u\}} \left\{\sum_{\boldsymbol{e}\in\binom{[n]}{s}:u,v\in\boldsymbol{e}} \left\{X_{\boldsymbol{e}} - \frac{e^{\mathbf{1}^\top\boldsymbol{\beta}_{\boldsymbol{e}}}}{1+e^{\mathbf{1}^\top\boldsymbol{\beta}_{\boldsymbol{e}}}}\right\}\right\}^2}.$$

Maximizing over $\tilde{\boldsymbol{a}} \in \mathbb{B}^{n-1}$ and $\tilde{\boldsymbol{b}} \in \mathcal{V}_1$ on both sides of (V.40) we get

$$\begin{aligned}&\sqrt{\sum_{v\in[n]\setminus\{u\}} \left\{\sum_{\boldsymbol{e}\in\binom{[n]}{s}:u,v\in\boldsymbol{e}} \left\{X_{\boldsymbol{e}} - \frac{e^{\mathbf{1}^\top\boldsymbol{\beta}_{\boldsymbol{e}}}}{1+e^{\mathbf{1}^\top\boldsymbol{\beta}_{\boldsymbol{e}}}}\right\}\right\}^2} \\ &\leqslant 2\max_{\tilde{\boldsymbol{b}}\in\mathcal{V}_1} \sum_{v\in[n]\setminus\{u\}} \tilde{b}_v \left\{\sum_{\boldsymbol{e}\in\binom{[n]}{s}:u,v\in\boldsymbol{e}} \left\{X_{\boldsymbol{e}} - \frac{e^{\mathbf{1}^\top\boldsymbol{\beta}_{\boldsymbol{e}}}}{1+e^{\mathbf{1}^\top\boldsymbol{\beta}_{\boldsymbol{e}}}}\right\}\right\}.\end{aligned} \tag{V.41}$$

As the above relation holds for all $u \in [n]$ we get

$$\begin{aligned}&\sqrt{\max_{u\in[n]}\sum_{v\in[n]\setminus\{u\}} \left\{\sum_{\boldsymbol{e}\in\binom{[n]}{s}:u,v\in\boldsymbol{e}} \left\{X_{\boldsymbol{e}} - \frac{e^{\mathbf{1}^\top\boldsymbol{\beta}_{\boldsymbol{e}}}}{1+e^{\mathbf{1}^\top\boldsymbol{\beta}_{\boldsymbol{e}}}}\right\}\right\}^2} \\ &\leqslant 2\max_{u\in[n]}\max_{\tilde{\boldsymbol{b}}\in\mathcal{V}_1} \sum_{v\in[n]\setminus\{u\}} \tilde{b}_v \left\{\sum_{\boldsymbol{e}\in\binom{[n]}{s}:u,v\in\boldsymbol{e}} \left\{X_{\boldsymbol{e}} - \frac{e^{\mathbf{1}^\top\boldsymbol{\beta}_{\boldsymbol{e}}}}{1+e^{\mathbf{1}^\top\boldsymbol{\beta}_{\boldsymbol{e}}}}\right\}\right\}.\end{aligned} \tag{V.42}$$

Hence, using (V.42), Hoeffding Inequality and union bound we get

$$\mathbb{P}\left(\max_{u\in[n]}\sum_{v\in[n]\setminus\{u\}} \left\{\sum_{\boldsymbol{e}\in\binom{[n]}{s}:u,v\in\boldsymbol{e}} \left\{X_{\boldsymbol{e}} - \frac{e^{\mathbf{1}^\top\boldsymbol{\beta}_{\boldsymbol{e}}}}{1+e^{\mathbf{1}^\top\boldsymbol{\beta}_{\boldsymbol{e}}}}\right\}\right\}^2\right.$$

$$> 4K^2 n^{s-1}\Bigg) \tag{V.43}$$

$$\leqslant \sum_{u=1}^{n} \sum_{\tilde{\boldsymbol{b}} \in \mathcal{V}_1} \mathbb{P}\left(\sum_{v \in [n]\setminus\{u\}} \tilde{b}_v \left\{ \sum_{\boldsymbol{e} \in \binom{[n]}{s}: u, v \in \boldsymbol{e}} \{X_{\boldsymbol{e}} - \frac{e^{\mathbf{1}^\top \boldsymbol{\beta}_{\boldsymbol{e}}}}{1+e^{\mathbf{1}^\top \boldsymbol{\beta}_{\boldsymbol{e}}}} \right\}\right\} > 2Kn^{\frac{s-1}{2}} \right)$$

$$\leqslant \sum_{u=1}^{n} \sum_{\tilde{\boldsymbol{b}} \in \mathcal{V}_1} e^{-\frac{2K^2 n}{\sum_{v=1}^{n-1} \tilde{b}_v^2}}$$

$$\leqslant n\, 2^{C_2 n} e^{-2K^2 n} \to 0,$$

for K large enough. □

VI. Estimation Lower Bounds: Proof of Theorem 2

The lower bound in the L_2 norm is proved in Appendix VI-A and the lower bound in the L_∞ norm is proved Appendix VI-B.

A. Estimation Lower Bound in the L_2 Norm: Proof of (II.10)

For $\boldsymbol{\gamma} \in \mathbb{R}^n$, denote the probability distribution of s-uniform model $\mathsf{H}_s(n, \boldsymbol{\gamma})$ by $\mathbb{P}_{\boldsymbol{\gamma}}$. To prove the result (II.10) recall the following version of Fano's lemma:

Theorem 9 ([47, Theorem 2.5]): Suppose there exists $\boldsymbol{\gamma}^{(0)}, \cdots, \boldsymbol{\gamma}^{(J)} \in \mathbb{R}^n$, with $\|\boldsymbol{\gamma}^{(j)}\| \in \mathcal{B}_M$ for all $0 \leqslant j \leqslant J$, such that

(1) $\|\boldsymbol{\gamma}^{(j)} - \boldsymbol{\gamma}^{(\ell)}\|_2 \geqslant 2\ s > 0$ for all $0 \leqslant j \neq \ell \leqslant J$,
(2) $\frac{1}{J}\sum_{j=1}^{J} \mathsf{KL}(\mathbb{P}_{\boldsymbol{\gamma}^{(j)}} \| \mathbb{P}_{\boldsymbol{\gamma}^{(0)}}) \leqslant \alpha \log J$,

where $\alpha \in (0, 1/8)$. Then

$$\min_{\hat{\boldsymbol{\gamma}}} \max_{\boldsymbol{\gamma}} \mathbb{P}\left(\|\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}\|_2 \geqslant s\right) \geqslant \frac{\sqrt{J}}{\sqrt{J}+1}\left(1 - 2\alpha - \sqrt{\frac{2\alpha}{\log J}}\right). \tag{VI.1}$$

To obtain $\boldsymbol{\gamma}^{(0)}, \ldots, \boldsymbol{\gamma}^{(J)} \in \mathbb{R}^n$ as in the above lemma we will invoke the Gilbert-Varshamov Theorem (see [47, Lemma 2.9]) which states that there exists $\boldsymbol{\omega}^{(0)}, \ldots, \boldsymbol{\omega}^{(J)} \in \{0,1\}^n$, with $J \geqslant 2^{n/8}$, such that $\boldsymbol{\omega}^{(0)} = (0, \cdots, 0)^\top$ and

$$\|\boldsymbol{\omega}^{(j)} - \boldsymbol{\omega}^{(\ell)}\|_1 \geqslant \frac{n}{8}, \tag{VI.2}$$

for all $0 \leqslant j \neq \ell \leqslant J$. For $\boldsymbol{\omega}^{(0)}, \ldots, \boldsymbol{\omega}^{(J)} \in \{0,1\}^n$ as above and $\delta \in (0, 1/8)$ define,

$$\boldsymbol{\gamma}^{(j)} = \varepsilon_n \boldsymbol{\omega}^{(j)}, \quad \text{for } 0 \leqslant j \leqslant J,$$

where $\varepsilon_n = 16Cn^{-\frac{s-1}{2}}$, with $C = C(\delta, s) > 0$ a constant depending on δ and s to be chosen later. By (VI.2) we have

$$\|\boldsymbol{\gamma}^{(j)} - \boldsymbol{\gamma}^{(\ell)}\|_2 \geqslant 2Cn^{-\frac{s-2}{2}}.$$

Now,

$$\mathsf{KL}(\mathbb{P}_{\boldsymbol{\gamma}^{(j)}} \| \mathbb{P}_{\boldsymbol{\gamma}^{(0)}}) = \sum_{t=0}^{s} \binom{\|\boldsymbol{\omega}^{(j)}\|_1}{t}\binom{n - \|\boldsymbol{\omega}^{(j)}\|_1}{s-t}\Big\{\psi(t\varepsilon_n)\log(2\psi(t\varepsilon_n)) + (1-\psi(t\varepsilon_n))\log(2(1-\psi(t\varepsilon_n)))\Big\},$$

where $\psi(x) = \frac{e^x}{1+e^x}$ is the logistic function defined in Lemma 4. By a Taylor expansion, for small enough $x > 0$,

$$\psi(x)\log(2\psi(x)) + (1-\psi(x))\log(2(1-\psi(x))) = \frac{x^2}{8} + O(x^3).$$

Hence, using $\binom{\|\boldsymbol{\omega}^{(j)}\|_1}{t}\binom{n-\|\boldsymbol{\omega}^{(j)}\|_1}{s-t} \lesssim_s n^s$ gives

$$\frac{1}{J}\sum_{j=1}^{J} \mathsf{KL}(\mathbb{P}_{\boldsymbol{\gamma}^{(j)}} \| \mathbb{P}_{\boldsymbol{\gamma}^{(0)}}) \lesssim_s n^s \varepsilon_n^2 \lesssim_s C^2 n \leqslant \delta \log J,$$

for $C = C(\delta, s)$ chosen appropriately. Hence, applying Theorem 9 and taking $J \to \infty$ in (VI.1) gives

$$\min_{\hat{\boldsymbol{\gamma}}} \max_{\boldsymbol{\gamma}} \mathbb{P}\left(\|\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}\|_\infty \geqslant Cn^{-\frac{s-2}{2}}\right) \geqslant 1 - 2\delta.$$

This completes the proof of (II.10).

B. Estimation Lower Bound in L_∞ Norm: Proof of (II.11)

Let us consider $n+1$ points $\boldsymbol{\gamma}^{(0)}, \boldsymbol{\gamma}^{(1)}, \ldots, \boldsymbol{\gamma}^{(n)} \in \mathbb{R}^n$ defined as follows:

$$\boldsymbol{\gamma}^{(j)} = \begin{cases} 2\varepsilon \boldsymbol{e}_j & \text{if } j \in \{1, \ldots, n\}, \\ \mathbf{0} & \text{if } j = 0, \end{cases}$$

where $\boldsymbol{e}_j$ is the j-th unit vector for $j \in [n]$ and $\varepsilon = \tilde{C}\sqrt{\frac{\log n}{n^{s-1}}}$ for a constant $\tilde{C} = \tilde{C}(\delta, s) > 0$ (depending on δ and s) to be chosen later. By definition, we have

$$\|\boldsymbol{\gamma}^{(j)} - \boldsymbol{\gamma}^{(\ell)}\|_\infty \geqslant 2\tilde{C}\sqrt{\frac{\log n}{n^{s-1}}}.$$

Denote the probability distribution of the s-uniform models $\mathsf{H}_s(n, \boldsymbol{\gamma}^{(j)})$ by $\mathbb{P}_{\boldsymbol{\gamma}^{(j)}}$. Observe that for all $j \in \{1, \ldots, n\}$

$$\mathsf{KL}(\mathbb{P}_{\boldsymbol{\gamma}^{(j)}} \| \mathbb{P}_{\boldsymbol{\gamma}^{(0)}}) = \frac{1}{2} \sum_{\boldsymbol{e} \in \binom{[n]}{s}: j \in \boldsymbol{e}} \left[\log\left\{\frac{(1+e^{2\varepsilon})}{2\, e^{2\varepsilon}}\right\} + \log\left\{\frac{1}{2}(1+e^{2\varepsilon})\right\}\right]. \tag{VI.3}$$

By Taylor's theorem, we get

$$\log\left\{\frac{(1+e^{2\varepsilon})}{2\, e^{2\varepsilon}}\right\} + \log\left\{\frac{1}{2}(1+e^{2\varepsilon})\right\} = 4\varepsilon^2 + O(\varepsilon^3) = \frac{4\tilde{C}^2 \log n}{n^{s-1}} + O\left(\frac{(\log n)^{\frac{3}{2}}}{n^{\frac{3}{2}(s-1)}}\right). \tag{VI.4}$$

Hence, from (VI.3),

$$\mathsf{KL}(\mathbb{P}_{\boldsymbol{\gamma}^{(j)}} \| \mathbb{P}_{\boldsymbol{\gamma}^{(0)}}) = L_s \tilde{C}^2 \log n + O\left(\frac{(\log n)^{\frac{3}{2}}}{n^{\frac{1}{2}(s-1)}}\right), \tag{VI.5}$$

for some constant L_s depending on s. Therefore, we have

$$\frac{1}{n}\sum_{j=1}^{n} \mathsf{KL}(\mathbb{P}_{\boldsymbol{\gamma}^{(j)}} \| \mathbb{P}_{\boldsymbol{\gamma}^{(0)}}) \lesssim_s L_s \tilde{C}^2 \log n \leqslant \delta \log n,$$

for $\tilde{C} = \tilde{C}(\delta, s)$ chosen appropriately to ensure $\delta \in (0, 1/8)$. Hence, applying Theorem 9 (with the L_2 norm replaced by the L_∞ norm) and taking large enough n in (VI.1) gives

$$\min_{\hat{\boldsymbol{\gamma}}} \max_{\boldsymbol{\gamma}} \mathbb{P}\left(\|\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}\|_\infty \geqslant \tilde{C}\sqrt{\frac{\log n}{n^{s-1}}}\right) \geqslant 1 - 2\delta. \tag{VI.6}$$

This completes the proof of (II.11).

VII. Proof of Theorem 3 and Theorem 4

We begin with the proof of Theorem 3 in Section VII-A. The proof of Theorem 4 is given in Section VII-B.

A. Proof of Theorem 3

Recall that, for $2 \leqslant s \leqslant r$, $\boldsymbol{d}_s = (d_s(1), d_s(2), \ldots, d_s(n))^\top$ is the vector of s-degrees. The first step in the proof of Theorem 3 is to derive a linearization of $\hat{\boldsymbol{\beta}}_s$ in terms of the s-degrees as in Proposition 2 below. The proof is given in Appendix VII-A.1.

Proposition 2: Fix $2 \leqslant s \leqslant r$. Then under the assumptions of Theorem 3, with probability $1 - o(1)$ as $n \to \infty$,

$$\|\hat{\boldsymbol{\beta}}_s - \boldsymbol{\beta}_s - \boldsymbol{\Sigma}_{n,s}^{-1}(\boldsymbol{d}_s - \mathbb{E}[\boldsymbol{d}_s])\|_\infty = O\left(\frac{\log n}{n^{s-1}}\right), \quad \text{(VII.1)}$$

where $\boldsymbol{\Sigma}_{n,s} = ((\sigma_s(u,v)))_{u,v\in[n]}$ is a $n \times n$ matrix with

$$\sigma_s(u,v) := \sum_{\boldsymbol{e}\in\binom{[n]}{s}:u,v\in\boldsymbol{e}} \frac{e^{\mathbf{1}^\top\boldsymbol{\beta}_{s,\boldsymbol{e}}}}{(1+e^{\mathbf{1}^\top\boldsymbol{\beta}_{s,\boldsymbol{e}}})^2}$$

and

$$\sigma_s(u,u) := \sigma_s(u)^2 = \sum_{\boldsymbol{e}\in\binom{[n]}{s}:u\in\boldsymbol{e}} \frac{e^{\mathbf{1}^\top\boldsymbol{\beta}_{s,\boldsymbol{e}}}}{(1+e^{\mathbf{1}^\top\boldsymbol{\beta}_{s,\boldsymbol{e}}})^2}, \quad \text{(VII.2)}$$

where $\sigma_s(u)^2$ is also defined in (II.12).

Next, define the matrix $\boldsymbol{\Gamma}_{n,s} = (\gamma_s(u,v))_{u,v\in[n]}$ as follows:

$$\gamma_s(u,v) := \frac{\mathbb{1}\{u=v\}}{\sigma_s(u)^2}. \quad \text{(VII.3)}$$

The following lemma shows that it is possible to replace the matrix $\boldsymbol{\Sigma}_{n,s}^{-1}$ in (VII.1) with the matrix $\boldsymbol{\Gamma}_{n,s}$ asymptotically. The proof of the lemma is given in Appendix VII-A.1.

Lemma 5: Suppose $\boldsymbol{\Sigma}_{n,s}$ and $\boldsymbol{\Gamma}_{n,s}$ be as defined in (VII.2) and (VII.3), respectively. Then under the assumptions of Theorem 3,

$$\|\boldsymbol{\Gamma}_{n,s} - \boldsymbol{\Sigma}_{n,s}^{-1}\|_\infty \leqslant O\left(\frac{1}{n^s}\right), \quad \text{(VII.4)}$$

where $\|\boldsymbol{A}\|_\infty = \max_{u,v\in[n]}|a_{u,v}|$ for a matrix $\boldsymbol{A} = ((a_{u,v}))_{u,v\in[n]}$. Furthermore,

$$\begin{aligned}&\|\operatorname{Cov}[(\boldsymbol{\Gamma}_{n,s} - \boldsymbol{\Sigma}_{n,s}^{-1})(\boldsymbol{d}_s - \mathbb{E}[\boldsymbol{d}_s])]\|_\infty \\ &\leqslant \|\boldsymbol{\Gamma}_{n,s} - \boldsymbol{\Sigma}_{n,s}^{-1}\|_\infty + O\left(\frac{1}{n^s}\right). \quad \text{(VII.5)}\end{aligned}$$

To complete the proof of Theorem 3, consider $J_s \in \binom{[n]}{a_s}$, for $a_s \geqslant 1$ fixed. Proposition 2 and Lemma 5 combined implies,

$$\|[(\hat{\boldsymbol{\beta}}_s - \boldsymbol{\beta}_s)]_{J_s} - [\boldsymbol{\Gamma}_{n,s}(\boldsymbol{d}_s - \mathbb{E}[\boldsymbol{d}_s])]_{J_s}\|_\infty = O\left(\frac{1}{\sqrt{n^s}}\right),$$

with probability $1 - o(1)$. Now, recall from the statement of Theorem 3 that $\boldsymbol{D}_s = \operatorname{diag}(\sigma_s(v))_{v\in[n]}$. From (VII.2) observe that $\max_{v\in[n]}\sigma_s(v)^2 \asymp n^{s-1}$, since $\|\boldsymbol{\beta}_s\|_\infty \leqslant M = O(1)$. Hence,

$$\|[\boldsymbol{D}_s(\hat{\boldsymbol{\beta}}_s - \boldsymbol{\beta}_s)]_{J_s} - [\boldsymbol{D}_s(\boldsymbol{\Gamma}_{n,s}(\boldsymbol{d}_s - \mathbb{E}[\boldsymbol{d}_s])]_{J_s}\|_\infty = O\left(\frac{1}{\sqrt{n}}\right).$$

Note that for $v \in J_s$,

$$\sigma_s(v)[\boldsymbol{\Gamma}_{n,s}(\boldsymbol{d}_s - \mathbb{E}[\boldsymbol{d}_s])_v] = \frac{d_s(v) - \mathbb{E}[d_s(v)]}{\sigma_s(v)}. \quad \text{(VII.6)}$$

Therefore, from (VII.6),

$$\begin{aligned}[\boldsymbol{D}_s]_{J_s}([(\hat{\boldsymbol{\beta}}_s - \boldsymbol{\beta}_s)]_{J_s}) &= \left(\left(\frac{d_s(v) - \mathbb{E}[d_s(v)]}{\sqrt{\operatorname{Var}[d_s(v)]}}\right)\right)_{v\in J_s} + O\left(\frac{1}{\sqrt{n}}\right) \\ &\xrightarrow{D} \mathcal{N}_{a_s}(\mathbf{0}, \boldsymbol{I}),\end{aligned}$$

using the central limit theorem for sums of independent bounded random variables. Since $\hat{\boldsymbol{\beta}}_s$ are independent across $2 \leqslant s \leqslant r$, the result in (II.14) follows.

1) Proof of Proposition 2: For $2 \leqslant s \leqslant r$ and $\boldsymbol{e} = (u_1, u_2, \ldots, u_s) \in \binom{[n]}{s}$, let $\boldsymbol{\beta}_{s,\boldsymbol{e}} = (\beta_{s,u_1}, \beta_{s,u_2}, \ldots, \beta_{s,u_s})^\top$ and $\hat{\boldsymbol{\beta}}_{s,\boldsymbol{e}} = (\hat{\beta}_{s,u_1}, \hat{\beta}_{s,u_2}, \ldots, \hat{\beta}_{s,u_s})^\top$. Moreover, $\mathbf{1}$ will denote the vector of ones in the appropriate dimension. To begin with, (II.7) and (V.14) gives, for $v \in [n]$,

$$\begin{aligned}&d_s(v) - \mathbb{E}[d_s(v)] \\ &= \sum_{\boldsymbol{e}\in\binom{[n]}{s}:v\in\boldsymbol{e}} \left\{\frac{e^{\mathbf{1}^\top\hat{\boldsymbol{\beta}}_{s,\boldsymbol{e}}}}{1+e^{\mathbf{1}^\top\hat{\boldsymbol{\beta}}_{s,\boldsymbol{e}}}} - \frac{e^{\mathbf{1}^\top\boldsymbol{\beta}_{s,\boldsymbol{e}}}}{1+e^{\mathbf{1}^\top\boldsymbol{\beta}_{s,\boldsymbol{e}}}}\right\}. \quad \text{(VII.7)}\end{aligned}$$

Note that for $\boldsymbol{e} \in \binom{[n]}{s}$, by a Taylor expansion,

$$\begin{aligned}&\frac{e^{\mathbf{1}^\top\hat{\boldsymbol{\beta}}_{s,\boldsymbol{e}}}}{1+e^{\mathbf{1}^\top\hat{\boldsymbol{\beta}}_{s,\boldsymbol{e}}}} - \frac{e^{\mathbf{1}^\top\boldsymbol{\beta}_{s,\boldsymbol{e}}}}{1+e^{\mathbf{1}^\top\boldsymbol{\beta}_{s,\boldsymbol{e}}}} \\ &= \frac{e^{\mathbf{1}^\top\boldsymbol{\beta}_{s,\boldsymbol{e}}}}{(1+e^{\mathbf{1}^\top\boldsymbol{\beta}_{s,\boldsymbol{e}}})^2}\left(\mathbf{1}^\top\hat{\boldsymbol{\beta}}_{s,\boldsymbol{e}} - \mathbf{1}^\top\boldsymbol{\beta}_{s,\boldsymbol{e}}\right) + T_{s,\boldsymbol{e}}, \quad \text{(VII.8)}\end{aligned}$$

where

$$|T_{s,\boldsymbol{e}}| \leqslant \frac{1}{2}\left|\mathbf{1}^\top\hat{\boldsymbol{\beta}}_{s,\boldsymbol{e}} - \mathbf{1}^\top\boldsymbol{\beta}_{s,\boldsymbol{e}}\right|^2 \lesssim_r \|\hat{\boldsymbol{\beta}}_s - \boldsymbol{\beta}_s\|_\infty^2. \quad \text{(VII.9)}$$

Then, from (VII.7),

$$d_s(v) - \mathbb{E}[d_s(v)] = \left[\boldsymbol{\Sigma}_{n,s}(\hat{\boldsymbol{\beta}}_s - \boldsymbol{\beta}_s)\right]_v + R_{v,s}, \quad \text{(VII.10)}$$

where $R_{v,s} = \sum_{\boldsymbol{e}\in\binom{[n]}{s}:v\in\boldsymbol{e}} T_{s,\boldsymbol{e}}$. From (VII.10), we have

$$\hat{\boldsymbol{\beta}}_s - \boldsymbol{\beta}_s = \boldsymbol{\Sigma}_{n,s}^{-1}(\boldsymbol{d}_s - \mathbb{E}[\boldsymbol{d}_s]) - \boldsymbol{\Sigma}_{n,s}^{-1}\boldsymbol{R}_{n,s}, \quad \text{(VII.11)}$$

where $\boldsymbol{R}_{n,s} = (R_{1,s}, R_{2,s}, \ldots, R_{n,s})^\top$. Note that from (VII.9),

$$|R_{v,s}| \leqslant \sum_{\boldsymbol{e}\in\binom{[n]}{s}:v\in\boldsymbol{e}} |T_{s,\boldsymbol{e}}| \lesssim_r n^{s-1}\|\hat{\boldsymbol{\beta}}_s - \boldsymbol{\beta}_s\|_\infty^2. \quad \text{(VII.12)}$$

To bound $\|\boldsymbol{\Sigma}_{n,s}^{-1}\boldsymbol{R}_{n,s}\|_\infty$, note that for $v \in [n]$,

$$\begin{aligned}&|[\boldsymbol{\Sigma}_{n,s}^{-1}\boldsymbol{R}_{n,s}]_v| \\ &\leqslant |[\boldsymbol{\Gamma}_{n,s}\boldsymbol{R}_{n,s}]_v| + |[(\boldsymbol{\Sigma}_{n,s}^{-1} - \boldsymbol{\Gamma}_{n,s})\boldsymbol{R}_{n,s}]_v|. \quad \text{(VII.13)}\end{aligned}$$

Observe that

$$[\boldsymbol{\Gamma}_{n,s}\boldsymbol{R}_{n,s}]_v = \frac{R_{v,s}}{\sigma_s(v)^2}.$$

Using $\sigma_s(v)^2 \asymp n^{s-1}$, (VII.12), and (II.8) gives,

$$|[\boldsymbol{\Gamma}_{n,s}\boldsymbol{R}_{n,s}]_v| \lesssim \|\hat{\boldsymbol{\beta}}_s - \boldsymbol{\beta}_s\|_\infty^2 = O\left(\frac{\log n}{n^{s-1}}\right),$$

with probability $1 - o(1)$. Further, by Lemma 5, (VII.12), and (II.8),

$$\begin{aligned}|[(\boldsymbol{\Sigma}_{n,s}^{-1} - \boldsymbol{\Gamma}_{n,s})\boldsymbol{R}_{n,s}]_v| &\leqslant \|(\boldsymbol{\Sigma}_{n,s}^{-1} - \boldsymbol{\Gamma}_{n,s})\|_\infty \times n\|\boldsymbol{R}_{n,s}\|_\infty \\ &\lesssim \|\hat{\boldsymbol{\beta}}_s - \boldsymbol{\beta}_s\|_\infty^2 \\ &\leqslant O\left(\frac{\log n}{n^{s-1}}\right),\end{aligned}$$

with probability $1 - o(1)$. Hence, by (VII.11) and (VII.13) the result in (VII.1) follows. □

2) Proof of Lemma 5

Proof of (VII.4): Denote

$$\begin{aligned}\boldsymbol{\Delta}_{n,s} &= \boldsymbol{\Gamma}_{n,s} - \boldsymbol{\Sigma}_{n,s}^{-1} = ((\delta_s(u,v)))_{u,v\in[n]}, \\ \boldsymbol{Z}_{n,s} &= \boldsymbol{I}_n - \boldsymbol{\Sigma}_{n,s}\boldsymbol{\Gamma}_{n,s} = ((z_s(u,v)))_{u,v\in[n]},\end{aligned}$$

and

$$\boldsymbol{\Theta}_{n,s} = \boldsymbol{\Gamma}_{n,s}\boldsymbol{Z}_{n,s} = (\theta_s(u,v))_{u,v\in[n]}.$$

Then

$$\begin{aligned}\boldsymbol{\Delta}_{n,s} &= (\boldsymbol{\Gamma}_{n,s} - \boldsymbol{\Sigma}_{n,s}^{-1})(\boldsymbol{I}_n - \boldsymbol{\Sigma}_{n,s}\boldsymbol{\Gamma}_{n,s}) \\ &\qquad - \boldsymbol{\Gamma}_{n,s}(\boldsymbol{I}_n - \boldsymbol{\Sigma}_{n,s}\boldsymbol{\Gamma}_{n,s}) \\ &= \boldsymbol{\Delta}_{n,s}\boldsymbol{Z}_{n,s} - \boldsymbol{\Theta}_{n,s}. \end{aligned} \tag{VII.14}$$

Hence, for $u, v \in [n]$,

$$\begin{aligned}&\delta_s(u,v) \\ &= \sum_{w=1}^n \delta_s(u,w) z_s(w,v) - \theta_s(u,v) \\ &= \sum_{w=1}^n \delta_s(u,w)\left\{\mathbb{1}\{w=v\} - \sum_{b=1}^n \sigma_s(w,b)\gamma_s(b,v)\right\} \\ &\qquad\qquad - \theta_s(u,v) \\ &= \sum_{w=1}^n \delta_s(u,w)\Bigg\{\mathbb{1}\{w=v\} \\ &\qquad - \sum_{b=1}^n \sigma_s(w,b)\frac{\mathbb{1}\{v=b\}}{\sigma_s(v)^2}\Bigg\} - \theta_s(u,v) \\ &\qquad\qquad\qquad\qquad\qquad\qquad \text{(by (VII.3))} \\ &= \sum_{w=1}^n \delta_s(u,w)\left\{\mathbb{1}\{w=v\} - \frac{\sigma_s(w,v)}{\sigma_s(v)^2}\right\} - \theta_s(u,v) \\ &= -\sum_{w=1}^n \delta_s(u,w)\left\{\mathbb{1}\{w\neq v\}\frac{\sigma_s(w,v)}{\sigma_s(v)^2}\right\} - \theta_s(u,v),\end{aligned} \tag{VII.15}$$

since $\sum_{b\in[n]\setminus\{w\}} \sigma_s(w,b) = \sigma_s(w,w) = \sigma_s(w)^2$. The following lemma bounds the maximum norm of $\boldsymbol{\Theta}_{n,s} = \boldsymbol{\Gamma}_{n,s}\boldsymbol{Z}_{n,s} = ((\theta_s(u,v)))_{u,v\in[n]}$.

Lemma 6: For $u, v, w \in [n]$,

$$\max\{|\theta_s(u,v)|, |\theta_s(u,v) - \theta_s(v,w)|\} \lesssim \frac{\sigma_{s,\max}}{\sigma_{s,\min}^2 n^2}, \tag{VII.16}$$

where $\sigma_{s,\min} := \min_{1\leqslant u<v\leqslant n} \sigma_s(u,v)$ and $\sigma_{s,\max} := \max_{1\leqslant u<v\leqslant n} \sigma_s(u,v)$.

Proof: Note that $\boldsymbol{\Theta}_{n,s} = \boldsymbol{\Gamma}_{n,s}\boldsymbol{Z}_{n,s} = \boldsymbol{\Gamma}_{n,s} - \boldsymbol{\Gamma}_{n,s}\boldsymbol{\Sigma}_{n,s}\boldsymbol{\Gamma}_{n,s}$. This means for $u, v \in [n]$,

$$\theta_s(u,v) = \gamma_s(u,v) - \sum_{x,y\in[n]} \gamma_s(u,x)\sigma_s(x,y)\gamma_s(y,v). \tag{VII.17}$$

Then recalling the definition of $\gamma_s(u,v)$ from (VII.3) gives,

$$\begin{aligned}&\sum_{x,y\in[n]} \gamma_s(u,x)\sigma_s(x,y)\gamma_s(y,v) \\ &= \sum_{x,y\in[n]} \frac{\mathbb{1}\{u=x\}\mathbb{1}\{y=v\}\sigma_s(x,y)}{\sigma_s(u)^2\sigma_s(v)^2} \\ &= \frac{\sigma_s(u,v)}{\sigma_s(u)^2\sigma_s(v)^2}.\end{aligned}$$

Hence, from (VII.3) and (VII.17),

$$|\theta_s(u,v)| = \left|\frac{\sigma_s(u,v)\mathbb{1}\{u\neq v\}}{\sigma_s(u)^2\sigma_s(v)^2}\right| \lesssim \frac{\sigma_{s,\max}}{\sigma_{s,\min}^2 n^2}.$$

This completes the proof of (VII.16). □

Now, for $u \in [n]$, let $\overline{m}, \underline{m} \in [n]$ be such that

$$\begin{aligned}\delta_s(u,\overline{m}) &= \max_{w\in[n]} \delta_s(u,w) \quad \text{and} \\ \delta_s(u,\underline{m}) &= \min_{w\in[n]} \delta_s(u,w).\end{aligned} \tag{VII.18}$$

The following lemma gives bounds on $\delta_s(u,\underline{m})$ and $\delta_s(u,\overline{m})$.

Lemma 7: For $u \in [n]$,

$$\sum_{w=1}^n \delta_s(u,w)\sigma_s(w,u) = 0.$$

This implies, $\delta_s(u,\overline{m}) \geqslant 0$ and $\delta_s(u,\underline{m}) \leqslant 0$.

Proof: Note that $\sum_{w=1}^n \delta_s(u,w)\sigma_s(w,u)$ is the u-th diagonal element of the matrix $\boldsymbol{\Delta}_{n,s}\boldsymbol{\Sigma}_{n,s} = \boldsymbol{\Gamma}_{n,s}\boldsymbol{\Sigma}_{n,s} - \boldsymbol{I}_n$ (recall that $\boldsymbol{\Delta}_{n,s} = \boldsymbol{\Gamma}_{n,s} - \boldsymbol{\Sigma}_{n,s}^{-1}$). Note that the u-th diagonal element of $\boldsymbol{\Gamma}_{n,s}\boldsymbol{\Sigma}_{n,s}$ is given by

$$\sum_{w\in[n]} \gamma_s(u,w)\sigma_s(w,u) = \sum_{w\in[n]} \frac{\mathbb{1}\{u=w\}}{\sigma_s(u)^2}\sigma_s(w,u) = 1,$$

since $\sigma_s(u,u) = \sigma_s(u)^2$. Hence, u-th diagonal element of $\boldsymbol{\Delta}_{n,s}\boldsymbol{\Sigma}_{n,s}$ is zero. □

Now, recalling (VII.15) note that

$$\begin{aligned}&\delta_s(u,\overline{m}) - \delta_s(u,\underline{m}) + (\theta_s(u,\overline{m}) - \theta_s(u,\underline{m})) \\ &= \sum_{w=1}^n \delta_s(u,w)\Bigg\{\frac{\mathbb{1}\{w\neq\underline{m}\}\sigma_s(w,\underline{m})}{\sigma_s(\underline{m})^2} \\ &\qquad - \frac{\mathbb{1}\{w\neq\overline{m}\}\sigma_s(w,\overline{m})}{\sigma_s(\overline{m})^2}\Bigg\} \\ &= \sum_{w=1}^n (\delta_s(u,w) - \delta_s(u,\underline{m}))\Bigg\{\frac{\mathbb{1}\{w\neq\underline{m}\}\sigma_s(w,\underline{m})}{\sigma_s(\underline{m})^2} \\ &\qquad - \frac{\mathbb{1}\{w\neq\overline{m}\}\sigma_s(w,\overline{m})}{\sigma_s(\overline{m})^2}\Bigg\},\end{aligned} \tag{VII.19}$$

since

$$\sum_{w\in[n]\setminus\{\underline{m}\}} \sigma_s(w,\underline{m}) = \sigma_s(\underline{m})^2, \quad \text{and}$$

$$\sum_{w\in[n]\setminus\{\overline{m}\}}\sigma_s(w,\overline{m})=\sigma_s(\overline{m})^2.$$

Define

$$\Omega:=\left\{w\in[n]:\frac{\mathbb{1}\{w\neq\underline{m}\}\sigma_s(w,\underline{m})}{\sigma_s(\underline{m})^2}\geqslant\frac{\mathbb{1}\{w\neq\overline{m}\}\sigma_s(w,\overline{m})}{\sigma_s(\overline{m})^2}\right\},\quad\text{(VII.20)}$$

and $\lambda:=|\Omega|$. Then, we have

$$\begin{aligned}\sum_{w\in\Omega}&(\delta_s(u,w)-\delta_s(u,\underline{m}))\left\{\frac{\mathbb{1}\{w\neq\underline{m}\}\sigma_s(w,\underline{m})}{\sigma_s(\underline{m})^2}-\frac{\mathbb{1}\{w\neq\overline{m}\}\sigma_s(w,\overline{m})}{\sigma_s(\overline{m})^2}\right\}\\&\leqslant(\delta_s(u,\overline{m})-\delta_s(u,\underline{m}))\left\{\frac{\sum_{w\in\Omega}\sigma_s(w,\underline{m})}{\sigma_s(\underline{m})^2}-\frac{\sum_{w\in\Omega}\mathbb{1}\{w\neq\overline{m}\}\sigma_s(w,\overline{m})}{\sigma_s(\overline{m})^2}\right\}.\quad\text{(VII.21)}\end{aligned}$$

Note that

$$\begin{aligned}\frac{\sum_{w\in\Omega}\sigma_s(w,\underline{m})}{\sigma_s(\underline{m})^2}&=\frac{\sum_{w\in\Omega}\sigma_s(w,\underline{m})}{\sum_{w\in\Omega}\sigma_s(w,\underline{m})+\sum_{w\in[n]\setminus(\Omega\bigcup\underline{m})}\sigma_s(w,\underline{m})}\\&=\frac{1}{1+\frac{\sum_{w\in[n]\setminus(\Omega\bigcup\underline{m})}\sigma_s(w,\underline{m})}{\sum_{w\in\Omega}\sigma_s(w,\underline{m})}},\end{aligned}\quad\text{(VII.22)}$$

since $\underline{m}\notin\Omega$. Now, observe that

$$\frac{\sum_{w\in[n]\setminus(\Omega\bigcup\underline{m})}\sigma_s(w,\underline{m})}{\sum_{w\in\Omega}\sigma_s(w,\underline{m})}\geqslant\frac{(n-\lambda-1)\sigma_{s,\min}}{\lambda\sigma_{s,\max}}$$

This implies,

$$\frac{\sum_{w\in\Omega}\sigma_s(w,\underline{m})}{\sigma_s(\underline{m})^2}\leqslant\frac{\lambda\sigma_{s,\max}}{\lambda\sigma_{s,\max}+(n-\lambda-1)\sigma_{s,\min}}.\quad\text{(VII.23)}$$

Similarly,

$$\begin{aligned}\frac{\sum_{w\in\Omega}\mathbb{1}\{w\neq\overline{m}\}\sigma_s(w,\overline{m})}{\sigma_s(\overline{m})^2}&=\frac{\sum_{w\in\Omega}\mathbb{1}\{w\neq\overline{m}\}\sigma_s(w,\overline{m})}{\sum_{w\in[n]}\mathbb{1}\{w\neq\overline{m}\}\sigma_s(w,\overline{m})}\\&=\frac{1}{1+\frac{\sum_{w\in[n]\setminus\Omega}\mathbb{1}\{w\neq\overline{m}\}\sigma_s(w,\overline{m})}{\sum_{w\in\Omega}\mathbb{1}\{w\neq\overline{m}\}\sigma_s(w,\overline{m})}}.\end{aligned}\quad\text{(VII.24)}$$

Therefore, since $\overline{m}\in\Omega$,

$$\frac{\sum_{w\in[n]\setminus\Omega}\mathbb{1}\{w\neq\overline{m}\}\sigma_s(w,\overline{m})}{\sum_{w\in\Omega}\mathbb{1}\{w\neq\overline{m}\}\sigma_s(w,\overline{m})}\leqslant\frac{(n-\lambda)\sigma_{s,\max}}{(\lambda-1)\sigma_{s,\min}}.$$

Hence,

$$\frac{\sum_{w\in\Omega}\mathbb{1}\{w\neq\overline{m}\}\sigma_s(w,\overline{m})}{\sigma_s(\overline{m})^2}\geqslant\frac{(\lambda-1)\sigma_{s,\min}}{(\lambda-1)\sigma_{s,\min}+(n-\lambda)\sigma_{s,\max}}.\quad\text{(VII.25)}$$

Applying (VII.23) and (VII.25) in (VII.21) gives,

$$\begin{aligned}\sum_{w\in\Omega}&(\delta_s(u,w)-\delta_s(u,\underline{m}))\left\{\frac{\mathbb{1}\{w\neq\underline{m}\}\sigma_s(w,\underline{m})}{\sigma_s(\underline{m})^2}-\frac{\mathbb{1}\{w\neq\overline{m}\}\sigma_s(w,\overline{m})}{\sigma_s(\overline{m})^2}\right\}\quad\text{(VII.26)}\\&\leqslant(\delta_s(u,\overline{m})-\delta_s(u,\underline{m}))f(\lambda),\quad\text{(VII.27)}\end{aligned}$$

where

$$f(\lambda)=\frac{\lambda\sigma_{s,\max}}{\lambda\sigma_{s,\max}+(n-1-\lambda)\sigma_{s,\min}}-\frac{(\lambda-1)\sigma_{s,\min}}{(\lambda-1)\sigma_{s,\min}+(n-\lambda)\sigma_{s,\max}}.\quad\text{(VII.28)}$$

Note that $f(\lambda)$ attains maximum at $\lambda=n/2$ over $\lambda\in(1,n-1)$ and

$$f(n/2)=\frac{n\sigma_{s,\max}-(n-2)\sigma_{s,\min}}{n\sigma_{s,\max}+(n-2)\sigma_{s,\min}}.$$

Therefore, from Lemma 6, (VII.19), there exists a constant $C>0$ such that (VII.26),

$$\begin{aligned}\delta_s(u,\overline{m})&-\delta_s(u,\underline{m})\\&\leqslant\frac{n\sigma_{s,\max}-(n-2)\sigma_{s,\min}}{n\sigma_{s,\max}+(n-2)\sigma_{s,\min}}(\delta_s(u,\overline{m})-\delta_s(u,\underline{m}))+\frac{C\sigma_{s,\max}}{\sigma_{s,\min}^2n^2}.\end{aligned}$$

This implies,

$$\begin{aligned}\delta_s(u,\overline{m})-\delta_s(u,\underline{m})&\leqslant\frac{C\sigma_{s,\max}(n\sigma_{s,\max}+(n-2)\sigma_{s,\min})}{2(n-2)\sigma_{s,\min}^3n^2}\\&\lesssim\frac{\sigma_{s,\max}^2}{\sigma_{s,\min}^3n^2}.\end{aligned}$$

Hence, from Lemma 7,

$$\begin{aligned}\max_{1\leqslant w\leqslant n}|\delta_s(u,w)|&\leqslant\delta_s(u,\overline{m})-\delta_s(u,\underline{m})\\&\leqslant\frac{\sigma_{s,\max}^2}{\sigma_{s,\min}^3n^2}\lesssim\frac{1}{n^s},\end{aligned}$$

since $\sigma_{s,\min}\asymp n^{s-2}$ and $\sigma_{s,\max}\asymp n^{s-2}$, using

$$\|\boldsymbol{\beta}_s\|_\infty\leqslant M=O(1).$$

This completes the proof of (VII.4). □

Proof of (VII.5):

Define

$$\begin{aligned}\boldsymbol{U}_{n,s}&=\mathrm{Cov}[(\boldsymbol{\Gamma}_{n,s}-\boldsymbol{\Sigma}_{n,s}^{-1})(\boldsymbol{d}_s-\mathbb{E}[\boldsymbol{d}_s])]\\&=\mathrm{Cov}[\boldsymbol{\Delta}_{n,s}(\boldsymbol{d}_s-\mathbb{E}[\boldsymbol{d}_s])],\end{aligned}\quad\text{(VII.29)}$$

since $\boldsymbol{\Delta}_{n,s}=\boldsymbol{\Gamma}_{n,s}-\boldsymbol{\Sigma}_{n,s}^{-1}$. Observe that

$$\begin{aligned}\boldsymbol{U}_{n,s}&=\boldsymbol{\Delta}_{n,s}\mathbb{E}[(\boldsymbol{d}_s-\mathbb{E}[\boldsymbol{d}_s])(\boldsymbol{d}_s-\mathbb{E}[\boldsymbol{d}_s])^\top]\boldsymbol{\Delta}_{n,s}^\top\\&=\boldsymbol{\Delta}_{n,s}\boldsymbol{\Sigma}_{n,s}\boldsymbol{\Delta}_{n,s}^\top\\&=(\boldsymbol{\Gamma}_{n,s}-\boldsymbol{\Sigma}_{n,s}^{-1})-\boldsymbol{\Gamma}_{n,s}(\boldsymbol{I}_n-\boldsymbol{\Sigma}_{n,s}\boldsymbol{\Gamma}_{n,s})\\&=(\boldsymbol{\Gamma}_{n,s}-\boldsymbol{\Sigma}_{n,s}^{-1})-\boldsymbol{\Theta}_{n,s},\end{aligned}\quad\text{(VII.30)}$$

since $\boldsymbol{\Theta}_{n,s}=\boldsymbol{\Gamma}_{n,s}\boldsymbol{Z}_{n,s}$ and $\boldsymbol{Z}_{n,s}=\boldsymbol{I}_n-\boldsymbol{\Sigma}_{n,s}\boldsymbol{\Gamma}_{n,s}$. By Lemma 6,

$$\|\boldsymbol{\Theta}_{n,s}\|_\infty\lesssim\frac{\sigma_{s,\max}}{\sigma_{s,\min}^2n^2}\lesssim\frac{1}{n^s},\quad\text{(VII.31)}$$

since $\sigma_{s,\min}\asymp n^{s-2}$ and $\sigma_{s,\max}\asymp n^{s-2}$, using $\|\boldsymbol{\beta}_s\|_\infty\leqslant M=O(1)$. By (VII.4), (VII.30), and (VII.31) the result in (VII.5) follows. □

B. Proof of Theorem 4

For $\boldsymbol{x} = (x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$ and $u \in [n]$ define the function

$$g_u(\boldsymbol{x}) = \sum_{\boldsymbol{e} \in \binom{[n]}{s}: u \in \boldsymbol{e}} \frac{e^{\boldsymbol{1}^\top \boldsymbol{x}_{\boldsymbol{e}}}}{(1 + e^{\boldsymbol{1}^\top \boldsymbol{x}_{\boldsymbol{e}}})^2},$$

where $\boldsymbol{x}_{\boldsymbol{e}} = (x_{u_1}, x_{u_2}, \ldots, x_{u_s})$ for $\boldsymbol{e} = (u_1, u_2, \ldots, u_s)$. Then recalling (II.12) and (II.15), $\sigma_s(v)^2 = g_v(\boldsymbol{\beta}_s)$ and $\hat{\sigma}_s(v)^2 = g_v(\hat{\boldsymbol{\beta}}_s)$ for all $v \in [n]$. Hence, by a Taylor expansion,

$$\begin{aligned}
&|\hat{\sigma}_s(v)^2 - \sigma_s(v)^2| \\
&= |g_v(\hat{\boldsymbol{\beta}}_s) - g_v(\boldsymbol{\beta}_s)| \\
&= \left| \sum_{\boldsymbol{e} \in \binom{[n]}{s}: v \in \boldsymbol{e}} \left\{ \frac{e^{\boldsymbol{1}^\top \hat{\boldsymbol{\beta}}_{s,\boldsymbol{e}}}}{(1 + e^{\boldsymbol{1}^\top \hat{\boldsymbol{\beta}}_{s,\boldsymbol{e}}})^2} - \frac{e^{\boldsymbol{1}^\top \boldsymbol{\beta}_{s,\boldsymbol{e}}}}{(1 + e^{\boldsymbol{1}^\top \boldsymbol{\beta}_{s,\boldsymbol{e}}})^2} \right\} \right| \\
&\lesssim_r \|\hat{\boldsymbol{\beta}}_s - \boldsymbol{\beta}_s\|_\infty n^{s-1}. \qquad \text{(VII.32)}
\end{aligned}$$

Recalling the definition of $J_s = \{v_{s,1}, \ldots, v_{s,a_s}\}$ from Theorem 4, this implies

$$\begin{aligned}
&\sum_{s=2}^{r} ([(\hat{\boldsymbol{\beta}}_s - \boldsymbol{\beta}_s)]_{J_s})^\top [\hat{\boldsymbol{D}}_s^2]_{J_s} ([(\hat{\boldsymbol{\beta}}_s - \boldsymbol{\beta}_s)]_{J_s}) \\
&= \sum_{s=2}^{r} \sum_{j=1}^{a_s} \hat{\sigma}_s(v_{a_j})^2 (\hat{\beta}_{s,v_{a_j}} - \beta_{s,v_{a_j}})^2 \\
&= \sum_{s=2}^{r} \sum_{j=1}^{a_s} \sigma_s(v_{a_j})^2 (\hat{\beta}_{s,v_{a_j}} - \beta_{s,v_{a_j}})^2 \\
&+ \sum_{s=2}^{r} \sum_{j=1}^{a_s} (\hat{\sigma}_s(v_{a_j})^2 - \sigma_s(v_{a_j})^2) (\hat{\beta}_{s,v_{a_j}} - \beta_{s,v_{a_j}})^2 \\
&\xrightarrow{D} \chi^2_{\sum_{s=2}^{r} a_s} + o_P(1),
\end{aligned}$$

by Theorem 3, (VII.32) and (II.8). This completes the proof of (II.16).

VIII. Proofs of Theorems 5 and 6

A. Proof of Theorem 5

Suppose $H_n \sim \mathsf{H}_{n,s}(n, \boldsymbol{\gamma})$ for $\boldsymbol{\gamma}$ as in (III.1). Let $\boldsymbol{\Sigma}_{n,s}$ be as defined in (VII.2) with $\boldsymbol{\beta}_s$ replaced by $\boldsymbol{\gamma} = (\gamma_1, \gamma_2, \ldots, \gamma_n)^\top$. Then $\nabla^2 \ell_{n,s}(\boldsymbol{\gamma}) = \boldsymbol{\Sigma}_{n,s}$. By a Taylor expansion,

$$\begin{aligned}
&\ell_{n,s}(\boldsymbol{\gamma}) - \ell_{n,s}(\hat{\boldsymbol{\beta}}_s) \\
&= (\hat{\boldsymbol{\beta}}_s - \boldsymbol{\gamma})^\top \nabla \ell_{n,s}(\boldsymbol{\gamma}) + \frac{1}{2} (\hat{\boldsymbol{\beta}}_s - \boldsymbol{\gamma})^\top \boldsymbol{\Sigma}_{n,s} (\hat{\boldsymbol{\beta}}_s - \boldsymbol{\gamma}) + \mathcal{T}_{n,s}, \qquad \text{(VIII.1)}
\end{aligned}$$

where

$$\mathcal{T}_{n,s} = \mathcal{T}_{n,s}^{(1)} + \mathcal{T}_{n,s}^{(2)} + \mathcal{T}_{n,s}^{(3)}, \qquad \text{(VIII.2)}$$

with

$$\begin{aligned}
\mathcal{T}_{n,s}^{(1)} &:= \frac{1}{6} \sum_{u=1}^{n} \frac{\partial^3 \ell_{n,s}(\boldsymbol{\gamma} + \theta(\hat{\boldsymbol{\beta}}_s - \boldsymbol{\gamma}))}{\partial (\beta_{s,u})^3} (\hat{\beta}_{s,u} - \gamma_u)^3, \\
\mathcal{T}_{n,s}^{(2)} &:= \frac{1}{3} \sum_{1 \leqslant u \neq v \leqslant n} \left\{ \frac{\partial^3 \ell_{n,s}(\boldsymbol{\gamma} + \theta(\hat{\boldsymbol{\beta}}_s - \boldsymbol{\gamma}))}{\partial (\beta_{s,u})^2 \partial \beta_{s,v}} (\hat{\beta}_{s,u} - \gamma_u)^2 (\hat{\beta}_{s,v} - \gamma_v) \right\}, \\
\mathcal{T}_{n,s}^{(3)} &:= \frac{1}{6} \sum_{1 \leqslant u \neq v \neq w \leqslant n} \left\{ \frac{\partial^3 \ell_{n,s}(\boldsymbol{\gamma} + \theta(\hat{\boldsymbol{\beta}}_s - \boldsymbol{\gamma}))}{\partial \beta_{s,u} \partial \beta_{s,v} \partial \beta_{s,w}} (\hat{\beta}_{s,u} - \gamma_u)(\hat{\beta}_{s,v} - \gamma_v)(\hat{\beta}_{s,w} - \gamma_w) \right\},
\end{aligned}$$

for some $\theta \in (0, 1)$.

Now, by arguments as in (VII.11) it follows that

$$\hat{\boldsymbol{\beta}}_s - \boldsymbol{\gamma} = \boldsymbol{\Sigma}_{n,s}^{-1} (\boldsymbol{d}_s - \mathbb{E}_{\boldsymbol{\gamma}}[\boldsymbol{d}_s]) + \boldsymbol{\Sigma}_{n,s}^{-1} \boldsymbol{R}_{n,s}, \qquad \text{(VIII.3)}$$

where $\boldsymbol{R}_{n,s}$ is as defined in (VII.10) and (VII.11) with $\boldsymbol{\beta}_s$ replaced by $\boldsymbol{\gamma}$. Using this and noting that $-\nabla \ell_{n,s}(\boldsymbol{\gamma}) = \boldsymbol{d}_s - \mathbb{E}_{\boldsymbol{\gamma}}[\boldsymbol{d}_s]$,

$$\begin{aligned}
&(\hat{\boldsymbol{\beta}}_s - \boldsymbol{\gamma})^\top \nabla \ell_{n,s}(\boldsymbol{\gamma}) \\
&= (\boldsymbol{d}_s - \mathbb{E}_{\boldsymbol{\gamma}}[\boldsymbol{d}_s])^\top \boldsymbol{\Sigma}_{n,s}^{-1} \nabla \ell_{n,s}(\boldsymbol{\gamma}) + \boldsymbol{R}_{n,s}^\top \boldsymbol{\Sigma}_{n,s}^{-1} \nabla \ell_{n,s}(\boldsymbol{\gamma}) \\
&= -(\boldsymbol{d}_s - \mathbb{E}_{\boldsymbol{\gamma}}[\boldsymbol{d}_s])^\top \boldsymbol{\Sigma}_{n,s}^{-1} (\boldsymbol{d}_s - \mathbb{E}_{\boldsymbol{\gamma}}[\boldsymbol{d}_s]) \\
&\quad - \boldsymbol{R}_{n,s}^\top \boldsymbol{\Sigma}_{n,s}^{-1} (\boldsymbol{d}_s - \mathbb{E}_{\boldsymbol{\gamma}}[\boldsymbol{d}_s]). \qquad \text{(VIII.4)}
\end{aligned}$$

Similarly, using (VIII.3),

$$\begin{aligned}
&(\hat{\boldsymbol{\beta}}_s - \boldsymbol{\gamma})^\top \boldsymbol{\Sigma}_{n,s} (\hat{\boldsymbol{\beta}}_s - \boldsymbol{\gamma}) \\
&= (\boldsymbol{d}_s - \mathbb{E}_{\boldsymbol{\gamma}}[\boldsymbol{d}_s])^\top \boldsymbol{\Sigma}_{n,s}^{-1} (\boldsymbol{d}_s - \mathbb{E}_{\boldsymbol{\gamma}}[\boldsymbol{d}_s]) \\
&\quad + 2 \boldsymbol{R}_{n,s}^\top \boldsymbol{\Sigma}_{n,s}^{-1} (\boldsymbol{d}_s - \mathbb{E}_{\boldsymbol{\gamma}}[\boldsymbol{d}_s]) + \boldsymbol{R}_{n,s}^\top \boldsymbol{\Sigma}_{n,s}^{-1} \boldsymbol{R}_{n,s}. \qquad \text{(VIII.5)}
\end{aligned}$$

Combining (VIII.1), (VIII.4), and (VIII.5) gives,

$$\begin{aligned}
&\ell_{n,s}(\hat{\boldsymbol{\beta}}_s) - \ell_{n,s}(\boldsymbol{\gamma}) \\
&= -\frac{1}{2} (\boldsymbol{d}_s - \mathbb{E}_{\boldsymbol{\gamma}}[\boldsymbol{d}_s])^\top \boldsymbol{\Sigma}_{n,s}^{-1} (\boldsymbol{d}_s - \mathbb{E}_{\boldsymbol{\gamma}}[\boldsymbol{d}_s]) \\
&\quad + \frac{1}{2} \boldsymbol{R}_{n,s}^\top \boldsymbol{\Sigma}_{n,s}^{-1} \boldsymbol{R}_{n,s} + \mathcal{T}_{n,s}. \qquad \text{(VIII.6)}
\end{aligned}$$

We begin by showing that $\boldsymbol{R}_{n,s}^\top \boldsymbol{\Sigma}_{n,s}^{-1} \boldsymbol{R}_{n,s} = o_P(\sqrt{n})$. To this end, (VII.12) and $\sigma_s(u)^2 \asymp n^{s-1}$ gives,

$$\left| \boldsymbol{R}_{n,s}^\top \boldsymbol{\Gamma}_{n,s} \boldsymbol{R}_{n,s} \right| = \left| \sum_{u=1}^{n} \frac{R_{s,u}^2}{\sigma_s(u)^2} \right| \lesssim n^s \|\hat{\boldsymbol{\beta}}_s - \boldsymbol{\beta}_s\|_\infty^4 \lesssim \frac{\log^2 n}{n^{s-2}}, \qquad \text{(VIII.7)}$$

with probability $1 - o(1)$ by (II.8). Next, observe that

$$\begin{aligned}
\left| \boldsymbol{R}_{n,s}^\top \boldsymbol{\Delta}_{n,s} \boldsymbol{R}_{n,s} \right| &\leqslant n \|\boldsymbol{\Delta}_{n,s} \boldsymbol{R}_{n,s}\|_\infty \cdot \|\boldsymbol{R}_{n,s}\|_\infty \\
&\leqslant n^2 \|\boldsymbol{R}_{n,s}\|_\infty^2 \cdot \|\boldsymbol{\Delta}_{n,s}\|_\infty \\
&\lesssim n^s \|\hat{\boldsymbol{\beta}}_s - \boldsymbol{\beta}_s\|_\infty^4 \\
&\qquad \text{(by (VII.4) and (VII.12))} \\
&\lesssim \frac{\log^2 n}{n^{s-2}}, \qquad \text{(VIII.8)}
\end{aligned}$$

with probability $1 - o(1)$ by (II.8). Combining (VIII.7) and (VIII.8) it follows that with probability $1 - o(1)$,

$$\begin{aligned}
\left| \boldsymbol{R}_{n,s}^\top \boldsymbol{\Sigma}_{n,s}^{-1} \boldsymbol{R}_{n,s} \right| &\leqslant \left| \boldsymbol{R}_{n,s}^\top \boldsymbol{\Gamma}_{n,s} \boldsymbol{R}_{n,s} \right| \\
&\quad + \left| \boldsymbol{R}_{n,s}^\top \boldsymbol{\Delta}_{n,s} \boldsymbol{R}_{n,s} \right| \lesssim \frac{\log^2 n}{n^{s-2}} \\
&= o_P(\sqrt{n}). \qquad \text{(VIII.9)}
\end{aligned}$$

This implies, the second term in the RHS of (VIII.6) does not contribute to the CLT of the log-likelihood ratio $\log \Lambda_{n,s}$.

Next, we show that the third term in the RHS of (VIII.6) is $o_P(\sqrt{n})$, hence, it also does not contribute to the CLT of $\log \Lambda_{n,s}$.

Lemma 8: Suppose $s \geqslant 3$ and $\boldsymbol{\gamma} \in \mathcal{B}(M)$. Then $\mathcal{T}_{n,s} = o_P(\sqrt{n})$.

Proof: Define $\tilde{\boldsymbol{\beta}}_s = \boldsymbol{\gamma} + \theta(\hat{\boldsymbol{\beta}}_s - \boldsymbol{\gamma})$, for $\theta \in (0,1)$. Then recalling (VIII.2) observe that

$$\mathcal{T}_{n,s}^{(1)} := \frac{1}{6}\sum_{u=1}^{n}\left\{\left[\sum_{\boldsymbol{e}\in\binom{[n]}{s}:u\in\boldsymbol{e}} \frac{e^{\mathbf{1}^\top \tilde{\boldsymbol{\beta}}_{s,\boldsymbol{e}}}(1-e^{\mathbf{1}^\top \tilde{\boldsymbol{\beta}}_{s,\boldsymbol{e}}})}{(1+e^{\mathbf{1}^\top \tilde{\boldsymbol{\beta}}_{s,\boldsymbol{e}}})^3}\right](\hat{\beta}_{s,u}-\gamma_u)^3\right\}$$

$$\mathcal{T}_{n,s}^{(2)} := \frac{1}{3}\sum_{1\leqslant u\neq v\leqslant n}\left\{\left[\sum_{\boldsymbol{e}\in\binom{[n]}{s}:u,v\in\boldsymbol{e}} \frac{e^{\mathbf{1}^\top \tilde{\boldsymbol{\beta}}_{s,\boldsymbol{e}}}(1-e^{\mathbf{1}^\top \tilde{\boldsymbol{\beta}}_{s,\boldsymbol{e}}})}{(1+e^{\mathbf{1}^\top \tilde{\boldsymbol{\beta}}_{s,\boldsymbol{e}}})^3}\right](\hat{\beta}_{s,u}-\gamma_u)^2(\hat{\beta}_{s,v}-\gamma_v)\right\},$$

$$\mathcal{T}_{n,s}^{(3)} := \frac{1}{6}\sum_{1\leqslant u\neq v\neq w\leqslant n}\left\{\left[\sum_{\boldsymbol{e}\in\binom{[n]}{s}:u,v,w\in\boldsymbol{e}} \frac{e^{\mathbf{1}^\top \tilde{\boldsymbol{\beta}}_{s,\boldsymbol{e}}}(1-e^{\mathbf{1}^\top \tilde{\boldsymbol{\beta}}_{s,\boldsymbol{e}}})}{(1+e^{\mathbf{1}^\top \tilde{\boldsymbol{\beta}}_{s,\boldsymbol{e}}})^3}\right](\hat{\beta}_{s,u}-\gamma_u)(\hat{\beta}_{s,v}-\gamma_v)(\hat{\beta}_{s,w}-\gamma_w)\right\},$$

where

$$\tilde{\boldsymbol{\beta}}_{s,\boldsymbol{e}} = (\tilde{\beta}_{s,u_1}, \tilde{\beta}_{s,u_2}, \ldots, \tilde{\beta}_{s,u_s})^\top, \text{ for } \quad \boldsymbol{e} = (u_1, u_2, \ldots, u_s).$$

Since $\boldsymbol{\gamma} \in \mathcal{B}_M$ and

$$\|\hat{\boldsymbol{\beta}}_s - \boldsymbol{\gamma}\|_\infty \lesssim_{s,M} \sqrt{\log n / n^{s-1}}$$

with probability $1 - o(1)$, $\hat{\boldsymbol{\beta}}_s \in \mathcal{B}_{2M}$ for large n with probability $1 - o(1)$. This implies,

$$\mathcal{T}_{n,s}^{(1)} \lesssim_M n^s \|\hat{\boldsymbol{\beta}}_s - \boldsymbol{\gamma}\|_\infty^3 \lesssim_{M,s} \sqrt{\frac{(\log n)^3}{n^{s-3}}} = o_P(\sqrt{n}), \tag{VIII.10}$$

for $s \geqslant 3$. Similarly, we can show that for $s \geqslant 3$, $\mathcal{T}_{n,s}^{(2)} = o_P(\sqrt{n})$ and $\mathcal{T}_{n,s}^{(3)} = o_P(\sqrt{n})$. This completes the proof of the Lemma 8. □

Remark 3: Note that Lemma 8 assumes that $s \geqslant 3$. This is because when $s = 2$ (that is, the graph case), the proof of Lemma 8 gives the bound $\mathcal{T}_{n,2} = O(\mathrm{polygon}(n)/\sqrt{n})$ which is not $o_P(\sqrt{n})$ (see (VIII.10)). Nevertheless, it follows from the proof of Theorem 1 (a) in Yan et al. [56], where the asymptotic null distribution of the LR test for the graph $\boldsymbol{\beta}$-model was derived, that the result in Lemma 8 also holds when $s = 2$, that is, $\mathcal{T}_{n,2} = o_P(\sqrt{n})$. For this one has to expand $\ell_{n,s}(\hat{\boldsymbol{\beta}}_s) - \ell_{n,s}(\boldsymbol{\gamma})$ up to the fourth order term, and show that the third order term is $o_P(\sqrt{n})$ at the true parameter value and the fourth order term is $o_P(\sqrt{n})$ at an intermediate point. For $s \geqslant 3$, the third order term at an intermediate point is $o_P(\sqrt{n})$, hence, we do not have to consider the fourth order term.

Now, recall the definition of $\log \Lambda_{n,s}$ from (III.2). Then by Lemma 8 and (VIII.6)

$$\frac{2\log\Lambda_{n,s} - n}{\sqrt{2n}} \tag{VIII.11}$$

$$= \frac{(\boldsymbol{d}_s - \mathbb{E}_{\boldsymbol{\gamma}}[\boldsymbol{d}_s])^\top \boldsymbol{\Sigma}_{n,s}^{-1}(\boldsymbol{d}_s - \mathbb{E}_{\boldsymbol{\gamma}}[\boldsymbol{d}_s]) - n}{\sqrt{2n}} + o_P(1). \tag{VIII.12}$$

By the following lemma we can replace $\boldsymbol{\Sigma}_{n,s}^{-1}$ with $\boldsymbol{\Gamma}_{n,s}$ in the RHS above. The proof of the lemma is given in Appendix VIII-A.1.

Lemma 9: For $L > 0$,

$$\mathbb{P}\left(\boldsymbol{d}_s - \mathbb{E}_{\boldsymbol{\gamma}}[\boldsymbol{d}_s])^\top(\boldsymbol{\Sigma}_{n,s}^{-1} - \boldsymbol{\Gamma}_{n,s})(\boldsymbol{d}_s - \mathbb{E}_{\boldsymbol{\gamma}}[\boldsymbol{d}_s]) > L\right) \lesssim \frac{1}{L^2}.$$

This implies, $(\boldsymbol{d}_s - \mathbb{E}_{\boldsymbol{\gamma}}[\boldsymbol{d}_s])^\top(\boldsymbol{\Sigma}_{n,s}^{-1} - \boldsymbol{\Gamma}_{n,s})(\boldsymbol{d}_s - \mathbb{E}_{\boldsymbol{\gamma}}[\boldsymbol{d}_s])$ is bounded in probability.

By Lemma 9 and recalling (VII.3),

$$\begin{aligned}
&\frac{(\boldsymbol{d}_s - \mathbb{E}_{\boldsymbol{\gamma}}[\boldsymbol{d}_s])^\top \boldsymbol{\Sigma}_{n,s}^{-1}(\boldsymbol{d}_s - \mathbb{E}_{\boldsymbol{\gamma}}[\boldsymbol{d}_s])}{\sqrt{2n}} \\
&= \frac{(\boldsymbol{d}_s - \mathbb{E}_{\boldsymbol{\gamma}}[\boldsymbol{d}_s])^\top \boldsymbol{\Gamma}_{n,s}(\boldsymbol{d}_s - \mathbb{E}_{\boldsymbol{\gamma}}[\boldsymbol{d}_s])}{\sqrt{2n}} + o_P(1) \\
&= \frac{1}{\sqrt{2n}}\sum_{u=1}^{n}\frac{(d_s(u) - \mathbb{E}_{\boldsymbol{\gamma}}[d_s(u)])^2}{\sigma_s(u)^2} + o_P(1).
\end{aligned} \tag{VIII.13}$$

Proposition 3 establishes the asymptotic normality of the leading term in the RHS above. The proof is given in Appendix VIII-A.2. □

Proposition 3: Under the assumption of Theorem 5,

$$\frac{1}{\sqrt{2n}}\left\{\sum_{u=1}^{n}\frac{(d_s(u) - \mathbb{E}_{\boldsymbol{\gamma}}[d_s(u)])^2}{\sigma_s(u)^2} - n\right\} \xrightarrow{D} \mathcal{N}(0,1). \tag{VIII.14}$$

The result in (III.3) now follows from (VIII.11), (VIII.13), and Proposition 3.

1) Proof of Lemma 9: To begin with note that

$$\begin{aligned}
&\mathbb{E}_{\boldsymbol{\gamma}}[(\boldsymbol{d}_s - \mathbb{E}_{\boldsymbol{\gamma}}[\boldsymbol{d}_s])^\top(\boldsymbol{\Sigma}_{n,s}^{-1} - \boldsymbol{\Gamma}_{n,s})(\boldsymbol{d}_s - \mathbb{E}_{\boldsymbol{\gamma}}[\boldsymbol{d}_s])] \\
&= \mathrm{tr}(\mathbb{E}_{\boldsymbol{\gamma}}[(\boldsymbol{d}_s - \mathbb{E}_{\boldsymbol{\gamma}}[\boldsymbol{d}_s])(\boldsymbol{d}_s - \mathbb{E}_{\boldsymbol{\gamma}}[\boldsymbol{d}_s])^\top](\boldsymbol{\Sigma}_{n,s}^{-1} - \boldsymbol{\Gamma}_{n,s})) \\
&= \mathrm{tr}(\boldsymbol{I}_n - \boldsymbol{\Sigma}_{n,s}\boldsymbol{\Gamma}_{n,s}) \\
&= n - \sum_{u,v\in[n]}\sigma_s(u,v)\gamma_s(u,v) \\
&= n - \sum_{u,v\in[n]}\sigma_s(u,v)\frac{\mathbb{1}\{u=v\}}{\sigma_s(u)^2} = 0.
\end{aligned}$$

Next, we will show that $\mathrm{Var}_{\boldsymbol{\gamma}}[(\boldsymbol{d}_s - \mathbb{E}_{\boldsymbol{\gamma}}[\boldsymbol{d}_s])^\top(\boldsymbol{\Sigma}_{n,s}^{-1} - \boldsymbol{\Gamma}_{n,s})(\boldsymbol{d}_s - \mathbb{E}_{\boldsymbol{\gamma}}[\boldsymbol{d}_s])] = O(1)$. The result in Lemma 9 then follows by Chebyshev's inequality. Recall that $\boldsymbol{\Delta}_{n,s} := \boldsymbol{\Sigma}_{n,s}^{-1} - \boldsymbol{\Gamma}_{n,s}$. We shall denote the entries of $\boldsymbol{\Delta}_{n,s}$ by $((\delta_{u,v}))$ for $u, v \in [n]$. Then

$$\begin{aligned}
&(\boldsymbol{d}_s - \mathbb{E}_{\boldsymbol{\gamma}}[\boldsymbol{d}_s])^\top(\boldsymbol{\Sigma}_{n,s}^{-1} - \boldsymbol{\Gamma}_{n,s})(\boldsymbol{d}_s - \mathbb{E}_{\boldsymbol{\gamma}}[\boldsymbol{d}_s]) \\
&= \sum_{u,v\in[n]}\delta_{u,v}(d_s(u) - \mathbb{E}_{\boldsymbol{\gamma}}[d_s(u)])(d_s(v) - \mathbb{E}_{\boldsymbol{\gamma}}[d_s(v)]).
\end{aligned}$$

Define $\bar{d}_s(u) := d_s(u) - \mathbb{E}_{\boldsymbol{\gamma}}[d_s(u)]$, for $u \in [n]$. Then

$$\begin{aligned}&\mathrm{Var}_{\boldsymbol{\gamma}}[(\boldsymbol{d}_s - \mathbb{E}_{\boldsymbol{\gamma}}[\boldsymbol{d}_s])^\top(\boldsymbol{\Sigma}_{n,s}^{-1} - \boldsymbol{\Gamma}_{n,s})(\boldsymbol{d}_s - \mathbb{E}_{\boldsymbol{\gamma}}[\boldsymbol{d}_s])]\\ &= \sum_{u,v,u',v' \in [n]} \delta_{u,v}\delta_{u',v'} \mathrm{Cov}_{\boldsymbol{\gamma}}[\bar{d}_s(u)\bar{d}_s(v), \bar{d}_s(u')\bar{d}_s(v')].\end{aligned} \tag{VIII.15}$$

To analyze the RHS of (VIII.15) we consider the following 4 cases.

a) Case 1: $u = v = u' = v'$. In this case we have

$$\mathrm{Cov}_{\boldsymbol{\gamma}}[\bar{d}_s(u)\bar{d}_s(v), \bar{d}_s(u')\bar{d}_s(v')] = \mathrm{Var}_{\boldsymbol{\gamma}}[\bar{d}_s(u)^2].$$

For $\boldsymbol{e} \in \binom{[n]}{s}$, denote $X_{\boldsymbol{e}} := \mathbb{1}\{\boldsymbol{e} \in E(H_n)\}$ and $\bar{X}_{\boldsymbol{e}} := \mathbb{1}\{\boldsymbol{e} \in E(H_n)\} - \mathbb{E}[\mathbb{1}\{\boldsymbol{e} \in E(H_n)\}]$. Since $\{\bar{X}_{\boldsymbol{e}} : \boldsymbol{e} \in \binom{[n]}{s}\}$ are independent and have zero mean, $\{\bar{X}_{\boldsymbol{e}}\bar{X}_{\boldsymbol{e}'} : \boldsymbol{e}, \boldsymbol{e}' \in \binom{[n]}{s}\}$ are pairwise uncorrelated. Hence,

$$\begin{aligned}\mathrm{Var}_{\boldsymbol{\gamma}}[\bar{d}_s(u)^2] &= \mathrm{Var}_{\boldsymbol{\gamma}}\left[\sum_{\boldsymbol{e},\boldsymbol{e}' \in \binom{[n]}{s}: u \in \boldsymbol{e} \bigcap \boldsymbol{e}'} \bar{X}_{\boldsymbol{e}}\bar{X}_{\boldsymbol{e}'}\right]\\ &= \sum_{\boldsymbol{e},\boldsymbol{e}' \in \binom{[n]}{s}: u \in \boldsymbol{e} \bigcap \boldsymbol{e}'} \mathrm{Var}_{\boldsymbol{\gamma}}\left[\bar{X}_{\boldsymbol{e}}\bar{X}_{\boldsymbol{e}'}\right]\\ &= \sum_{\boldsymbol{e} \in \binom{[n]}{s}: u \in \boldsymbol{e}} \mathrm{Var}_{\boldsymbol{\gamma}}[\bar{X}_{\boldsymbol{e}}^2]\\ &\quad + \sum_{\boldsymbol{e} \neq \boldsymbol{e}' \in \binom{[n]}{s}: u \in \boldsymbol{e} \cap \boldsymbol{e}'} \mathrm{Var}_{\boldsymbol{\gamma}}[\bar{X}_{\boldsymbol{e}}]\,\mathrm{Var}_{\boldsymbol{\gamma}}[\bar{X}_{\boldsymbol{e}'}].\end{aligned} \tag{VIII.16}$$

Since $\|\boldsymbol{\gamma}\|_\infty \leqslant M$,

$$\mathrm{Var}_{\boldsymbol{\gamma}}[\bar{X}_{\boldsymbol{e}}] = \mathrm{Var}_{\boldsymbol{\gamma}}[X_{\boldsymbol{e}}] = \frac{e^{\boldsymbol{1}^\top \gamma_{\boldsymbol{e}}}}{(1 + e^{\boldsymbol{1}^\top \boldsymbol{\gamma}_{\boldsymbol{e}}})} \asymp_M 1,$$

where $\gamma_{\boldsymbol{e}} = (\gamma_{u_1}, \gamma_{u_2}, \ldots, \gamma_{u_s})^\top$, for $\boldsymbol{e} = (u_1, u_2, \ldots, u_s)$. Similarly, $\mathrm{Var}_{\boldsymbol{\gamma}}[\bar{X}_{\boldsymbol{e}}^2] \asymp_M 1$. Hence, (VIII.16) implies that

$$\mathrm{Var}_{\boldsymbol{\gamma}}[\bar{d}_s(u)^2] \lesssim_M n^{2s-2}.$$

b) Case 2: $u \neq v = u' = v'$. Observe that

$$\begin{aligned}&\mathrm{Cov}_{\boldsymbol{\gamma}}[\bar{d}_s(u)\bar{d}_s(v), \bar{d}_s(u')\bar{d}_s(v')]\\ &= \mathrm{Cov}_{\boldsymbol{\gamma}}[\bar{d}_s(u)\bar{d}_s(v), \bar{d}_s(v)^2]\\ &= \sum_{\substack{\boldsymbol{e}_1,\boldsymbol{e}_2,\boldsymbol{e}_3,\boldsymbol{e}_4 \in \binom{[n]}{s}\\ u \in \boldsymbol{e}_1, v \in \boldsymbol{e}_1 \cap \boldsymbol{e}_2 \cap \boldsymbol{e}_3}} \left\{\mathbb{E}_{\boldsymbol{\gamma}}[\bar{X}_{\boldsymbol{e}_1}\bar{X}_{\boldsymbol{e}_2}\bar{X}_{\boldsymbol{e}_3}\bar{X}_{\boldsymbol{e}_4}]\right.\\ &\qquad\qquad \left.-\mathbb{E}_{\boldsymbol{\gamma}}[\bar{X}_{\boldsymbol{e}_1}\bar{X}_{\boldsymbol{e}_2}]\mathbb{E}_{\boldsymbol{\gamma}}[\bar{X}_{\boldsymbol{e}_3}\bar{X}_{\boldsymbol{e}_4}]\right\}.\end{aligned}$$

Note that the non-zero contributions in the RHS above come from the terms when $\boldsymbol{e}_i = \boldsymbol{e}_j$ and $\boldsymbol{e}_k = \boldsymbol{e}_\ell$ for $i, j, k, \ell \in \{1, \ldots, 4\}$. Hence,

$$\begin{aligned}&\mathrm{Cov}_{\boldsymbol{\gamma}}[\bar{d}_s(u)\bar{d}_s(v), \bar{d}_s(v)^2]\\ &= \sum_{\boldsymbol{e} \in \binom{[n]}{s}, u,v \in \boldsymbol{e}} \left(\mathbb{E}_{\boldsymbol{\gamma}}[\bar{X}_{\boldsymbol{e}}^4] - (\mathbb{E}_{\boldsymbol{\gamma}}[\bar{X}_{\boldsymbol{e}}^2])^2\right)\\ &\qquad + 2\sum_{\substack{\boldsymbol{e}_1 \neq \boldsymbol{e}_2 \in \binom{[n]}{s}\\ u,v \in \boldsymbol{e}_1, v \in \boldsymbol{e}_2}} \mathbb{E}_{\boldsymbol{\gamma}}[\bar{X}_{\boldsymbol{e}_1}^2]\mathbb{E}[\bar{X}_{\boldsymbol{e}_2}^2]\\ &\lesssim_M n^{2s-3},\end{aligned}$$

since $\mathbb{E}_{\boldsymbol{\gamma}}[\bar{X}_{\boldsymbol{e}}^4] \asymp_M 1$ and $\mathbb{E}_{\boldsymbol{\gamma}}[\bar{X}_{\boldsymbol{e}}^2]) \asymp_M 1$.

c) Case 3: $u \neq v \neq u' = v'$: By similar reasoning as the previous two cases it can be shown that

$$\begin{aligned}&\mathrm{Cov}_{\boldsymbol{\gamma}}[\bar{d}_s(u)\bar{d}_s(v), \bar{d}_s(u')\bar{d}_s(v')]\\ &\qquad = \mathrm{Cov}_{\boldsymbol{\gamma}}[\bar{d}_s^2(u), \bar{d}_s(u')\bar{d}_s(u')] \lesssim_M n^{2s-3}.\end{aligned}$$

d) Case 4: $u \neq v \neq u' \neq v'$. In this case, it can be shown that

$$\mathrm{Cov}_{\boldsymbol{\gamma}}[\bar{d}_s(u)\bar{d}_s(v), \bar{d}_s(u')\bar{d}_s(v')] \lesssim_M n^{2s-4}.$$

Combining the 4 cases and using (VIII.15),

$$\begin{aligned}&\mathrm{Var}_{\boldsymbol{\gamma}}[(\boldsymbol{d}_s - \mathbb{E}_{\boldsymbol{\gamma}}[\boldsymbol{d}_s])^\top(\boldsymbol{\Sigma}_{n,s}^{-1} - \boldsymbol{\Gamma}_{n,s})(\boldsymbol{d}_s - \mathbb{E}_{\boldsymbol{\gamma}}[\boldsymbol{d}_s])]\\ &\lesssim_M \max_{u,v \in [n]} |\delta_{u,v}|^2 n^{2s} = O(1),\end{aligned}$$

where the last step uses (VII.4).

2) Proof of Proposition 3: Suppose $H_n = (V(H_n), E(H_n)) \sim \mathsf{H}_{n,s}(n, \boldsymbol{\gamma})$ for $\boldsymbol{\gamma}$ as in (III.1). For $\boldsymbol{e} = \{v_1, v_2, \ldots, v_s\} \in \binom{[n]}{s}$, denote

$$X_{\boldsymbol{e}} := X_{\{v_1,v_2,\ldots,v_s\}} := \mathbb{1}\{\boldsymbol{e} \in E(H_n)\}.$$

and $\bar{X}_{\boldsymbol{e}} := \mathbb{1}\{\boldsymbol{e} \in E(H_n)\} - \mathbb{E}_{\boldsymbol{\gamma}}[\mathbb{1}\{\boldsymbol{e} \in E(H_n)\}]$. Also, for $u \in [n]$ denote

$$\bar{d}_s(u) = d_s(u) - \mathbb{E}_{\boldsymbol{\gamma}}[d_s(u)] = \sum_{\boldsymbol{e} \in \binom{[n]}{s}: u \in \boldsymbol{e}} \bar{X}_{\boldsymbol{e}}.$$

Observe that

$$\bar{d}_s(u)^2 = \sum_{\boldsymbol{e} \in \binom{[n]}{s}: u \in \boldsymbol{e}} \bar{X}_{\boldsymbol{e}}^2 + \sum_{\boldsymbol{e} \neq \boldsymbol{e}' \in \binom{[n]}{s}: u \in \boldsymbol{e} \cap \boldsymbol{e}'} \bar{X}_{\boldsymbol{e}}\bar{X}_{\boldsymbol{e}'}. \tag{VIII.17}$$

This implies,

$$\begin{aligned}\mathbb{E}_{\boldsymbol{\gamma}}[\bar{d}_s(u)^2] &= \sum_{\boldsymbol{e} \in \binom{[n]}{s}: u \in \boldsymbol{e}} \mathbb{E}_{\boldsymbol{\gamma}}[\bar{X}_{\boldsymbol{e}}^2]\\ &= \sum_{\boldsymbol{e} \in \binom{[n]}{s}: u \in \boldsymbol{e}} \mathrm{Var}_{\boldsymbol{\gamma}}[\bar{X}_{\boldsymbol{e}}^2] = \mathrm{Var}_{\boldsymbol{\gamma}}[d_s(u)] = \sigma_s(u)^2.\end{aligned}$$

Hence,

$$\begin{aligned}&\frac{1}{\sqrt{2n}}\left\{\sum_{u=1}^n \frac{(d_s(u) - \mathbb{E}_{\boldsymbol{\gamma}}[d_s(u)])^2}{\sigma_s(u)^2} - n\right\}\\ &= \frac{1}{\sqrt{2n}}\sum_{u=1}^n \frac{\bar{d}_s(u)^2 - \mathbb{E}_{\boldsymbol{\gamma}}[\bar{d}_s(u)^2]}{\sigma_s(u)^2}\\ &= \frac{1}{\sqrt{2n}}\sum_{u=1}^n \sum_{\boldsymbol{e} \in \binom{[n]}{s}: u \in \boldsymbol{e}} \frac{\bar{X}_{\boldsymbol{e}}^2 - \mathbb{E}_{\boldsymbol{\gamma}}[\bar{X}_{\boldsymbol{e}}^2]}{\sigma_s(u)^2}\\ &\qquad + \frac{1}{\sqrt{2n}}\sum_{u=1}^n \sum_{\boldsymbol{e} \neq \boldsymbol{e}' \in \binom{[n]}{s}: u \in \boldsymbol{e} \cap \boldsymbol{e}'} \frac{\bar{X}_{\boldsymbol{e}}\bar{X}_{\boldsymbol{e}'}}{\sigma_s(u)^2}\\ &\qquad\qquad \text{(by (VIII.17))}\\ &:= T_1 + T_2.\end{aligned} \tag{VIII.18}$$

We will first show that $T_1 = o_P(1)$. Towards this note that

$$T_1 = \frac{s}{\sqrt{2n}} \sum_{\boldsymbol{e} \in \binom{[n]}{s}} \frac{\bar{X}_{\boldsymbol{e}}^2 - \mathbb{E}_{\boldsymbol{\gamma}}[\bar{X}_{\boldsymbol{e}}^2]}{\sigma_s(u)^2}.$$

Since $\{\bar{X}_{\boldsymbol{e}} : \boldsymbol{e} \in \binom{[n]}{s}\}$ are independent,

$$\mathrm{Var}_{\boldsymbol{\gamma}}[T_1] = \frac{s^2}{2\,n} \sum_{\boldsymbol{e}\in\binom{[n]}{s}} \frac{\mathrm{Var}_{\boldsymbol{\gamma}}[\bar{X}_{\boldsymbol{e}}^2]}{\sigma_s(u)^4} \lesssim_M \frac{1}{n^{s-1}},$$

using $\mathrm{Var}_{\boldsymbol{\gamma}}[\bar{X}_{\boldsymbol{e}}^2] \asymp_M 1$ and $\sigma_s(u)^2 \asymp_M n^{s-1}$. This implies, $T_1 = o_P(1)$.

Therefore, from (VIII.18), to prove (VIII.14) it remains to show $T_2 \xrightarrow{D} N(0,1)$. For this we will express T_2 as a sum of a martingale difference sequence. To this end, define the following sequence of sigma-fields: For $u \in [n]$,

$$\mathcal{F}_u := \sigma\left(\bigcup_{v=1}^{u}\{\bar{X}_{\boldsymbol{e}} : v \in \boldsymbol{e}\}\right),$$

is the sigma algebra generated by the collection of random variables $\bigcup_{v=1}^{u}\{\bar{X}_{\boldsymbol{e}} : v \in \boldsymbol{e}\}$. Clearly, $\mathcal{F}_1 \subseteq \mathcal{F}_2 \cdots \subseteq \mathcal{F}_n$, hence $\{\mathcal{F}_u\}_{u\in[n]}$ is a filtration. Now, for $u \in [n]$, define

$$T_{2,u} := \sum_{\substack{\boldsymbol{e},\boldsymbol{e}'\in\binom{[n]}{s}:\boldsymbol{e}\neq\boldsymbol{e}',u\in\boldsymbol{e}\cap\boldsymbol{e}',\\ \boldsymbol{e}\cap\{1,\ldots,u\}\neq\varnothing\\ \text{and } \boldsymbol{e}'\cap\{1,\ldots,u-1\}=\varnothing}} w_{\boldsymbol{e},\boldsymbol{e}'}\bar{X}_{\boldsymbol{e}}\bar{X}_{\boldsymbol{e}'}$$

where $w_{\boldsymbol{e},\boldsymbol{e}'} := \sum_{z\in\boldsymbol{e}\cap\boldsymbol{e}'} \frac{1}{\sigma_s(z)^2}$. Note that $T_{2,u}$ is $\mathcal{F}_u$ measurable and $\mathbb{E}_{\boldsymbol{\gamma}}[T_{2,u}|\mathcal{F}_{u-1}] = 0$, that is, $T_{2,u}$, for $u \in [n]$, is a martingale difference sequence. Also, recalling the definition of T_2 from (VIII.18) observe that

$$\begin{aligned} T_2 &= \frac{1}{\sqrt{2n}} \sum_{u=1}^{n} \sum_{\boldsymbol{e}\neq\boldsymbol{e}'\in\binom{[n]}{s}:u\in\boldsymbol{e}\cap\boldsymbol{e}'} \frac{\bar{X}_{\boldsymbol{e}}\bar{X}_{\boldsymbol{e}'}}{\sigma_s(u)^2} \\ &= \frac{1}{\sqrt{2n}} \sum_{\boldsymbol{e}\neq\boldsymbol{e}'\in\binom{[n]}{s}, \boldsymbol{e}\cap\boldsymbol{e}'\neq\varnothing} w_{\boldsymbol{e},\boldsymbol{e}'}\bar{X}_{\boldsymbol{e}}\bar{X}_{\boldsymbol{e}'} \\ &= \frac{1}{\sqrt{2n}} \sum_{u=1}^{n} T_{2,u}, \end{aligned}$$

that is, T_2 is the sum of a martingale difference sequence. Now, invoking the martingale central theorem [9] it can be shown that $T_2 \xrightarrow{D} N(0,1)$. The details are omitted.

B. Proof of Theorem 6

Suppose $H_n \sim \mathsf{H}_{n,s}(n,\boldsymbol{\gamma}')$ for $\boldsymbol{\gamma}'$ as in (III.5). Then by arguments as in (VIII.6),

$$\begin{aligned} &\ell_{n,s}(\hat{\boldsymbol{\beta}}_s) - \ell_{n,s}(\boldsymbol{\gamma}') \\ &= -\frac{1}{2}(\boldsymbol{d}_s - \mathbb{E}_{\boldsymbol{\gamma}'}[\boldsymbol{d}_s])^\top \overline{\boldsymbol{\Sigma}}_{n,s}^{-1}(\boldsymbol{d}_s - \mathbb{E}_{\boldsymbol{\gamma}'}[\boldsymbol{d}_s]) \\ &+ \frac{1}{2}\boldsymbol{R}_{n,s}^\top \overline{\boldsymbol{\Sigma}}_{n,s}^{-1}\boldsymbol{R}_{n,s} + \mathcal{T}_{n,s}, \end{aligned}$$

where $\overline{\boldsymbol{\Sigma}}_{n,s}$ and $\boldsymbol{R}_{n,s}$ are as defined in (VII.2) and (VII.10), respectively, with $\boldsymbol{\beta}_s$ replaced by $\boldsymbol{\gamma}'$ and $\mathcal{T}_{n,s}$ as defined in (VIII.2) with $\boldsymbol{\gamma}$ replaced by $\boldsymbol{\gamma}'$. Therefore,

$$\begin{aligned} &\ell_{n,s}(\hat{\boldsymbol{\beta}}_s) - \ell_{n,s}(\boldsymbol{\gamma}) \\ &= -\frac{1}{2}(\boldsymbol{d}_s - \mathbb{E}_{\boldsymbol{\gamma}'}[\boldsymbol{d}_s])^\top \overline{\boldsymbol{\Sigma}}_{n,s}^{-1}(\boldsymbol{d}_s - \mathbb{E}_{\boldsymbol{\gamma}'}[\boldsymbol{d}_s]) \\ &+ \frac{1}{2}\boldsymbol{R}_{n,s}^\top \overline{\boldsymbol{\Sigma}}_{n,s}^{-1}\boldsymbol{R}_{n,s} + \mathcal{T}_{n,s} \\ &+ \ell_{n,s}(\boldsymbol{\gamma}') - \ell_{n,s}(\boldsymbol{\gamma}), \end{aligned} \tag{VIII.19}$$

By Taylor expansion,

$$\begin{aligned} &\ell_{n,s}(\boldsymbol{\gamma}') - \ell_{n,s}(\boldsymbol{\gamma}) \\ &= (\boldsymbol{d}_s - \mathbb{E}_{\boldsymbol{\gamma}'}[\boldsymbol{d}_s])^\top(\boldsymbol{\gamma}' - \boldsymbol{\gamma}) + \frac{1}{2}(\boldsymbol{\gamma}' - \boldsymbol{\gamma})^\top \tilde{\boldsymbol{\Sigma}}_{n,s}(\boldsymbol{\gamma}' - \boldsymbol{\gamma}), \end{aligned}$$

where $\tilde{\boldsymbol{\Sigma}}_{n,s}$ is the covariance matrix defined in (VII.2) with $\boldsymbol{\beta}_s$ replaced by $\tilde{\boldsymbol{\gamma}} = \boldsymbol{\gamma}' + \theta(\boldsymbol{\gamma}' - \boldsymbol{\gamma})$ for some $0 < \theta < 1$. Then by arguments as in (VIII.9) and Lemma 8, Lemma 9, (VIII.19) can be written as:

$$\begin{aligned} &\ell_{n,s}(\hat{\boldsymbol{\beta}}_s) - \ell_{n,s}(\boldsymbol{\gamma}) \\ &= -\frac{1}{2}(\boldsymbol{d}_s - \mathbb{E}_{\boldsymbol{\gamma}'}[\boldsymbol{d}_s])^\top \overline{\boldsymbol{\Gamma}}_{n,s}(\boldsymbol{d}_s - \mathbb{E}_{\boldsymbol{\gamma}'}[\boldsymbol{d}_s]) \\ &+ (\boldsymbol{d}_s - \mathbb{E}_{\boldsymbol{\gamma}'}[\boldsymbol{d}_s])^\top(\boldsymbol{\gamma}' - \boldsymbol{\gamma}) \\ &+ \frac{1}{2}(\boldsymbol{\gamma}' - \boldsymbol{\gamma})^\top \tilde{\boldsymbol{\Sigma}}_{n,s}(\boldsymbol{\gamma}' - \boldsymbol{\gamma}) + o_P(\sqrt{n}), \end{aligned} \tag{VIII.20}$$

where $\overline{\boldsymbol{\Gamma}}_{n,s}$ is as defined in (VII.3) with the parameter $\boldsymbol{\beta}_s$ replaced by $\boldsymbol{\gamma}'$.

We begin with the case $\|\boldsymbol{\gamma}' - \boldsymbol{\gamma}\|_2 \ll n^{-\frac{2s-3}{4}}$. In this case, since $\nabla^2\ell_{n,s}(\boldsymbol{\gamma}') = \overline{\boldsymbol{\Sigma}}_{n,s}$, by Lemma 2

$$(\boldsymbol{\gamma}' - \boldsymbol{\gamma})^\top \overline{\boldsymbol{\Sigma}}_{n,s}(\boldsymbol{\gamma}' - \boldsymbol{\gamma}) \asymp n^{s-1}\|\boldsymbol{\gamma}' - \boldsymbol{\gamma}\|_2^2 \ll \sqrt{n}. \tag{VIII.21}$$

Similarly,

$$(\boldsymbol{\gamma}' - \boldsymbol{\gamma})^\top \tilde{\boldsymbol{\Sigma}}_{n,s}(\boldsymbol{\gamma}' - \boldsymbol{\gamma}) \asymp n^{s-1}\|\boldsymbol{\gamma}' - \boldsymbol{\gamma}\|_2^2 \ll \sqrt{n}. \tag{VIII.22}$$

Hence,

$$\mathrm{Var}[(\boldsymbol{d}_s - \mathbb{E}_{\boldsymbol{\gamma}'}[\boldsymbol{d}_s])^\top(\boldsymbol{\gamma}' - \boldsymbol{\gamma})] = (\boldsymbol{\gamma}' - \boldsymbol{\gamma})^\top \overline{\boldsymbol{\Sigma}}_{n,s}(\boldsymbol{\gamma}' - \boldsymbol{\gamma}) \ll n,$$

which implies, $(\boldsymbol{d}_s - \mathbb{E}_{\boldsymbol{\gamma}'}[\boldsymbol{d}_s])^\top(\boldsymbol{\gamma}' - \boldsymbol{\gamma}) = o_P(\sqrt{n})$, since $\mathbb{E}[(\boldsymbol{d}_s - \mathbb{E}_{\boldsymbol{\gamma}'}[\boldsymbol{d}_s])^\top(\boldsymbol{\gamma}' - \boldsymbol{\gamma})] = 0$. Therefore, under H_1 as in (III.5),

$$\begin{aligned} &\frac{2\log\Lambda_{n,s} - n}{\sqrt{2n}} \\ &= \frac{2(\ell_{n,s}(\boldsymbol{\gamma}) - \ell_{n,s}(\hat{\boldsymbol{\beta}}_s)) - n}{\sqrt{2n}} \\ &= \frac{(\boldsymbol{d}_s - \mathbb{E}_{\boldsymbol{\gamma}'}[\boldsymbol{d}_s])^\top \overline{\boldsymbol{\Gamma}}_{n,s}(\boldsymbol{d}_s - \mathbb{E}_{\boldsymbol{\gamma}'}[\boldsymbol{d}_s]) - n}{\sqrt{2n}} + o_P(1) \\ &\qquad\qquad \text{(by (VIII.20), (VIII.21), and (VIII.22))} \\ &\xrightarrow{D} \mathcal{N}(0,1), \end{aligned}$$

by Proposition 3. This proves the first assertion in (III.6).

Next, suppose $\|\boldsymbol{\gamma}' - \boldsymbol{\gamma}\|_2 \gg n^{-\frac{2s-3}{4}}$. In this case, by Lemma 2, $(\boldsymbol{\gamma}' - \boldsymbol{\gamma})^\top \overline{\boldsymbol{\Sigma}}_{n,s}(\boldsymbol{\gamma}' - \boldsymbol{\gamma}) \asymp n^{s-1}\|\boldsymbol{\gamma}' - \boldsymbol{\gamma}\|_2^2 \gg \sqrt{n}$. We will first assume:

$$\sqrt{n} \ll (\boldsymbol{\gamma}' - \boldsymbol{\gamma})^\top \overline{\boldsymbol{\Sigma}}_{n,s}(\boldsymbol{\gamma}' - \boldsymbol{\gamma}) \lesssim n. \tag{VIII.23}$$

Then we have

$$\begin{aligned} &\mathrm{Var}[(\boldsymbol{d}_s - \mathbb{E}_{\boldsymbol{\gamma}'}[\boldsymbol{d}_s])^\top(\boldsymbol{\gamma}' - \boldsymbol{\gamma})] \\ &= (\boldsymbol{\gamma}' - \boldsymbol{\gamma})^\top \overline{\boldsymbol{\Sigma}}_{n,s}(\boldsymbol{\gamma}' - \boldsymbol{\gamma}) = O(n). \end{aligned} \tag{VIII.24}$$

Using this and Proposition 3 it follows that

$$\frac{1}{\sqrt{n}}\left[\frac{1}{2}(\boldsymbol{d}_s - \mathbb{E}_{\boldsymbol{\gamma}'}[\boldsymbol{d}_s])^\top \overline{\boldsymbol{\Gamma}}_{n,s}(\boldsymbol{d}_s - \mathbb{E}_{\boldsymbol{\gamma}'}[\boldsymbol{d}_s])\right.$$

$$+(\boldsymbol{d}_s - \mathbb{E}_{\boldsymbol{\gamma}'}[\boldsymbol{d}_s])^\top(\boldsymbol{\gamma}'-\boldsymbol{\gamma})\Big] \quad \text{(VIII.25)}$$

is bounded in probability. Hence, from (VIII.20),

$$\frac{2\log\Lambda_{n,s}-n}{\sqrt{2n}} = \frac{(\ell_{n,s}(\boldsymbol{\gamma})-\ell_{n,s}(\hat{\boldsymbol{\beta}}_s))-n}{\sqrt{2n}} \to \infty,$$

in probability, since by Lemma 2, $(\boldsymbol{\gamma}'-\boldsymbol{\gamma})^\top\tilde{\boldsymbol{\Sigma}}_{n,s}(\boldsymbol{\gamma}'-\boldsymbol{\gamma}) \asymp n^{s-1}\|\boldsymbol{\gamma}'-\boldsymbol{\gamma}\|_2^2 \gg \sqrt{n}$. This implies, $\mathbb{E}_{\boldsymbol{\gamma}'}[\phi_{n,s}] \to 1$, whenever (VIII.23) holds. Next, we assume

$$(\boldsymbol{\gamma}'-\boldsymbol{\gamma})^\top\overline{\boldsymbol{\Sigma}}_{n,s}(\boldsymbol{\gamma}'-\boldsymbol{\gamma}) \gg n. \quad \text{(VIII.26)}$$

For notational convenience denote $\vartheta_{n,s} := (\boldsymbol{\gamma}'-\boldsymbol{\gamma})^\top\overline{\boldsymbol{\Sigma}}_{n,s}(\boldsymbol{\gamma}'-\boldsymbol{\gamma})$. Then Proposition 3 and (VIII.26) imply that

$$\frac{1}{\sqrt{\vartheta_{n,s}}}\Big[\frac{1}{2}(\boldsymbol{d}_s - \mathbb{E}_{\boldsymbol{\gamma}'}[\boldsymbol{d}_s])^\top\overline{\boldsymbol{\Gamma}}_{n,s}(\boldsymbol{d}_s - \mathbb{E}_{\boldsymbol{\gamma}'}[\boldsymbol{d}_s]) + (\boldsymbol{d}_s - \mathbb{E}_{\boldsymbol{\gamma}'}[\boldsymbol{d}_s])^\top(\boldsymbol{\gamma}'-\boldsymbol{\gamma})\Big]$$

is bounded in probability. Using (VIII.21) and (VIII.22) we also get

$$\frac{(\boldsymbol{\gamma}'-\boldsymbol{\gamma})^\top\tilde{\boldsymbol{\Sigma}}_{n,s}(\boldsymbol{\gamma}'-\boldsymbol{\gamma})}{\sqrt{\vartheta_{n,s}}} \asymp n^{\frac{s-1}{2}}\|\boldsymbol{\gamma}'-\boldsymbol{\gamma}\|_2 \to \infty.$$

This implies, from from (VIII.20),

$$\mathbb{E}_{\boldsymbol{\gamma}'}[\phi_{n,s}] = \mathbb{P}_{\boldsymbol{\gamma}'}\left(\left|\frac{2\log\Lambda_{n,s}-n}{\sqrt{\vartheta_{n,s}}}\right| \geqslant z_{\alpha/2}\sqrt{\frac{2n}{\vartheta_{n,s}}}\right) \to 1.$$

This concludes the proof. This completes the proof of the third assertion in (III.6).

Now, we consider the case $n^{\frac{2s-3}{4}}\|\boldsymbol{\gamma}'-\boldsymbol{\gamma}\|_2 \to \tau \in (0,\infty)$. By Taylor expansion,

$$\begin{aligned}&\ell_{n,s}(\boldsymbol{\gamma}') - \ell_{n,s}(\boldsymbol{\gamma})\\ &= (\boldsymbol{d}_s - \mathbb{E}_{\boldsymbol{\gamma}'}[\boldsymbol{d}_s])^\top(\boldsymbol{\gamma}'-\boldsymbol{\gamma})\\ &+ \frac{1}{2}(\boldsymbol{\gamma}'-\boldsymbol{\gamma})^\top\boldsymbol{\Sigma}_{n,s}(\boldsymbol{\gamma}'-\boldsymbol{\gamma}) + \tilde{\mathcal{T}}_{n,s},\end{aligned} \quad \text{(VIII.27)}$$

where $\boldsymbol{\Sigma}_{n,s}$ is as defined in (VII.2) with $\boldsymbol{\beta}_s$ replaced by $\boldsymbol{\gamma}$ and $\tilde{\mathcal{T}}_{n,s}$ is as defined in (VIII.2) with the parameter $\tilde{\boldsymbol{\gamma}} = \boldsymbol{\gamma}' + \theta(\boldsymbol{\gamma}'-\boldsymbol{\gamma})$ for some $0<\theta<1$. By arguments as in Lemma 8, $\tilde{\mathcal{T}}_{n,s} = o_P(\sqrt{n})$. Then (VIII.9) and Lemma 8, Lemma 9, (VIII.19) can be written as:

$$\begin{aligned}&\ell_{n,s}(\hat{\boldsymbol{\beta}}_s) - \ell_{n,s}(\boldsymbol{\gamma})\\ &= -\frac{1}{2}(\boldsymbol{d}_s - \mathbb{E}_{\boldsymbol{\gamma}'}[\boldsymbol{d}_s])^\top\boldsymbol{\Gamma}_{n,s}(\boldsymbol{d}_s - \mathbb{E}_{\boldsymbol{\gamma}'}[\boldsymbol{d}_s])\\ &+ (\boldsymbol{d}_s - \mathbb{E}_{\boldsymbol{\gamma}'}[\boldsymbol{d}_s])^\top(\boldsymbol{\gamma}'-\boldsymbol{\gamma})\\ &+ \frac{1}{2}(\boldsymbol{\gamma}'-\boldsymbol{\gamma})^\top\boldsymbol{\Sigma}_{n,s}(\boldsymbol{\gamma}'-\boldsymbol{\gamma}) + o_P(\sqrt{n}).\end{aligned} \quad \text{(VIII.28)}$$

Note that $\mathbb{E}[(\boldsymbol{d}_s - \mathbb{E}_{\boldsymbol{\gamma}'}[\boldsymbol{d}_s])^\top(\boldsymbol{\gamma}'-\boldsymbol{\gamma})] = 0$ and by Lemma 2,

$$\begin{aligned}&\mathrm{Var}[(\boldsymbol{d}_s - \mathbb{E}_{\boldsymbol{\gamma}'}[\boldsymbol{d}_s])^\top(\boldsymbol{\gamma}'-\boldsymbol{\gamma})]\\ &= (\boldsymbol{\gamma}'-\boldsymbol{\gamma})^\top\boldsymbol{\Sigma}_{n,s}(\boldsymbol{\gamma}'-\boldsymbol{\gamma}) \asymp_{n,r} \sqrt{n},\end{aligned}$$

when $\|\boldsymbol{\gamma}'-\boldsymbol{\gamma}\|_2 \asymp n^{-\frac{2s-3}{4}}$. Hence, in this case,

$$(\boldsymbol{d}_s - \mathbb{E}_{\boldsymbol{\gamma}'}[\boldsymbol{d}_s])^\top(\boldsymbol{\gamma}'-\boldsymbol{\gamma}) = o_P(\sqrt{n}).$$

This also implies that

$$\eta := \lim_{n\to\infty}\frac{(\boldsymbol{\gamma}'-\boldsymbol{\gamma})^\top\boldsymbol{\Sigma}_{n,s}(\boldsymbol{\gamma}'-\boldsymbol{\gamma})}{\sqrt{n}}$$

exists along a subsequence. (Note that $\mathrm{Cov}_{\boldsymbol{\gamma}}[\boldsymbol{d}_s] = \boldsymbol{\Sigma}_{n,s}$.) Hence, from (VIII.28),

$$\begin{aligned}\frac{2\log\Lambda_{n,s}-n}{\sqrt{2n}} &= \frac{2(\ell_{n,s}(\boldsymbol{\gamma})-\ell_{n,s}(\hat{\boldsymbol{\beta}}_s))-n}{\sqrt{2n}}\\ &= \frac{(\boldsymbol{d}_s - \mathbb{E}_{\boldsymbol{\gamma}'}[\boldsymbol{d}_s])^\top\boldsymbol{\Gamma}_{n,s}(\boldsymbol{d}_s - \mathbb{E}_{\boldsymbol{\gamma}'}[\boldsymbol{d}_s])-n}{\sqrt{2n}}\\ &\quad - \frac{(\boldsymbol{\gamma}'-\boldsymbol{\gamma})^\top\boldsymbol{\Sigma}_{n,s}(\boldsymbol{\gamma}'-\boldsymbol{\gamma})}{\sqrt{2n}} + o_P(1)\\ &\xrightarrow{D} \mathcal{N}(-\tfrac{\eta}{\sqrt{2}}, 1).\end{aligned}$$

This completes the proof of (III.7).

IX. Testing Lower Bounds

In this section we prove the lower bounds for the goodness-of-fit problem in the L_2 and L_∞ norms, that is, Theorem 7 (b) and Theorem 8 (b), respectively. For this, suppose π_n be a prior probability distribution on the alternative H_1 (as in (III.8) or (III.10)). Then the Bayes risk of a test function ψ_n is defined as

$$\mathcal{R}(\psi_n, \boldsymbol{\gamma}, \pi_n) = \mathbb{P}_{H_0}(\psi_n = 1) + \mathbb{E}_{\boldsymbol{\gamma}'\sim\pi_n}\left[\mathbb{P}_{\boldsymbol{\gamma}'}(\psi_n = 0)\right]. \quad \text{(IX.1)}$$

For any prior π_n the worst-case risk of test function ψ_n, as defined in (III.9), can be bounded below as:

Lemma 10: Let $\mathcal{H}_{n,s}$ denote the collection of s-uniform hypergraphs on n vertices. Then

$$\mathcal{R}(\psi_n, \boldsymbol{\gamma}) \geqslant \mathcal{R}(\psi_n, \boldsymbol{\gamma}, \pi_n) \geqslant 1 - \tfrac{1}{2}\sqrt{\mathbb{E}_{H_0}[L_{\pi_n}^2]-1}, \quad \text{(IX.2)}$$

where $L_{\pi_n} = \frac{\mathbb{E}_{\boldsymbol{\gamma}'\sim\pi_n}[\mathbb{P}_{\boldsymbol{\gamma}'}(\omega)]}{\mathbb{P}_{H_0}(\omega)}$, $\omega \in \mathcal{H}_{n,s}$, is the π_n-integrated likelihood ratio.

Proof: Clearly, $\mathcal{R}(\psi_n, \boldsymbol{\gamma}) \geqslant \mathcal{R}(\psi_n, \boldsymbol{\gamma}, \pi_n)$. To show the second inequality in (IX.2) observe that,

$$\begin{aligned}\mathcal{R}(\psi_n, \boldsymbol{\gamma}, \pi_n) &\geqslant \inf_{\psi_n}\{\mathbb{P}_{H_0}(\psi_n=1) + \mathbb{E}_{\boldsymbol{\gamma}'\sim\pi_n}(\mathbb{P}_{\boldsymbol{\gamma}'}(\psi_n=0))\}\\ &\geqslant 1 - \sup_{\psi_n}|\mathbb{P}_{H_0}(\psi_n=1)\\ &\qquad - \mathbb{E}_{\boldsymbol{\gamma}'\sim\pi_n}(\mathbb{P}_{\boldsymbol{\gamma}'}(\psi_n=1))|\\ &\geqslant 1 - \sup_{\omega\in\mathcal{H}_{n,s}}|\mathbb{P}_{H_0}(\omega) - \mathbb{E}_{\boldsymbol{\gamma}'\sim\pi_n}[\mathbb{P}_{\boldsymbol{\gamma}'}(\omega)]|\\ &\geqslant 1 - \frac{1}{2}\sum_{\omega\in\mathcal{H}_{n,s}}\left|\frac{\mathbb{E}_{\boldsymbol{\gamma}'\sim\pi_n}[\mathbb{P}_{\boldsymbol{\gamma}'}(\omega)]}{\mathbb{P}_{H_0}(\omega)} - 1\right|\mathbb{P}_{H_0}(\omega)\\ &= 1 - \tfrac{1}{2}\mathbb{E}_{H_0}|L_{\pi_n}-1|\\ &\geqslant 1 - \tfrac{1}{2}\sqrt{\mathbb{E}_{H_0}[L_{\pi_n}^2]-1},\end{aligned}$$

where the last step uses the Cauchy-Schwarz inequality. □

Therefore, to show all tests are powerless it suffices to construct a prior π_n on H_1 such that $\mathbb{E}_{H_0}[L_{\pi_n}^2] \to 1$. We show this for the L_2 norm in Appendix IX-A and for the L_∞ norm in Appendix IX-B.

A. Testing Lower Bound in L_2 Norm: Proof of Theorem 7 (b)

We choose $\boldsymbol{\gamma} = \mathbf{0}$, $\varepsilon \ll n^{-\frac{2s-3}{4}}$, and construct a prior π_n on H_1 as in (III.8) as follows: Suppose $\boldsymbol{\gamma}' = (\gamma_1', \gamma_2', \ldots, \gamma_n')^\top \in \mathbb{R}^n$ with

$$\gamma_u' = \eta_u \cdot \frac{\varepsilon}{\sqrt{n}},$$

for $u \in [n]$, where $\eta_1, \ldots, \eta_n$ are i.i.d Rademacher random variables, taking values $\{\pm 1\}$ with probability $\frac{1}{2}$. Clearly, $\|\boldsymbol{\gamma} - \boldsymbol{\gamma}'\|_2 = \varepsilon$. Then, for $H \in \mathcal{H}_{n,s}$, the π_n integrated likelihood ratio is given by

$$L_{\pi_n} = \mathbb{E}_{\boldsymbol{\eta}}\left[\frac{\mathbb{P}_{\boldsymbol{\gamma}'}(H)}{\mathbb{P}_{\mathbf{0}}(H)}\right] = \mathbb{E}_{\boldsymbol{\eta}}\left[\prod_{\boldsymbol{e} \in \binom{[n]}{s}} \frac{2\, e^{w_{\boldsymbol{\eta}}(\boldsymbol{e}) X_{\boldsymbol{e}}}}{1 + e^{w_{\boldsymbol{\eta}}(\boldsymbol{e})}}\right],$$

where $X_{\boldsymbol{e}} := \mathbb{1}\{\boldsymbol{e} \in E(H)\}$, $\boldsymbol{\eta} := (\eta_1, \ldots, \eta_n)$, and $w_{\boldsymbol{\eta}}(\boldsymbol{e}) := \frac{\varepsilon}{\sqrt{n}} \sum_{u \in \boldsymbol{e}} \eta_u$, for $\boldsymbol{e} \in \binom{[n]}{s}$. Then

$$L_{\pi_n}^2 = \mathbb{E}_{\boldsymbol{\eta}, \boldsymbol{\eta}'}\left[\prod_{\boldsymbol{e} \in \binom{[n]}{s}} \frac{4\, e^{(w_{\boldsymbol{\eta}}(\boldsymbol{e}) + w_{\boldsymbol{\eta}'}(\boldsymbol{e})) X_{\boldsymbol{e}}}}{(1 + e^{w_{\boldsymbol{\eta}}(\boldsymbol{e})})(1 + e^{w_{\boldsymbol{\eta}'}(\boldsymbol{e})})}\right],$$

where $\eta_1', \ldots, \eta_n'$ are i.i.d Rademacher random variables which are independent of $\eta_1, \ldots, \eta_n$, $\boldsymbol{\eta}' := (\eta_1', \ldots, \eta_n')$, and $w_{\boldsymbol{\eta}'}(\boldsymbol{e}) := \frac{\varepsilon}{\sqrt{n}} \sum_{u \in \boldsymbol{e}} \eta_u'$, for $\boldsymbol{e} \in \binom{[n]}{s}$. Taking expectation with respect to H_0 gives,

$$\begin{aligned}
&\mathbb{E}_{H_0}[L_{\pi_n}^2] \\
&= \mathbb{E}_{\boldsymbol{\eta}, \boldsymbol{\eta}'}\left[\prod_{\boldsymbol{e} \in \binom{[n]}{s}} \frac{2\,(e^{(w_{\boldsymbol{\eta}}(\boldsymbol{e}) + w_{\boldsymbol{\eta}'}(\boldsymbol{e}))} + 1)}{(1 + e^{w_{\boldsymbol{\eta}}(\boldsymbol{e})})(1 + e^{w_{\boldsymbol{\eta}'}(\boldsymbol{e})})}\right] \\
&= \mathbb{E}_{\boldsymbol{\eta}, \boldsymbol{\eta}'}\Bigg[\prod_{\boldsymbol{e} \in \binom{[n]}{s}} 2\,\{\psi(w_{\boldsymbol{\eta}}(\boldsymbol{e}))\psi(w_{\boldsymbol{\eta}'}(\boldsymbol{e})) \\
&\qquad + (1 - \psi(w_{\boldsymbol{\eta}}(\boldsymbol{e})))(1 - \psi(w_{\boldsymbol{\eta}'}(\boldsymbol{e})))\}\Bigg], \qquad \text{(IX.3)}
\end{aligned}$$

where $\psi(x)$ is the logistic function as defined in Lemma 4. Using the Taylor expansions of $\psi(x)$ and $1 - \psi(x)$ around 0, we can show that for all $x \in \mathbb{R}$,

$$\psi(x) \leqslant \frac{1}{2} + \frac{x}{4} + \frac{x^3}{48} \text{ and } 1 - \psi(x) \leqslant \frac{1}{2} - \frac{x}{4} + \frac{x^3}{48}.$$

As a consequence, for $\boldsymbol{e} \in \binom{[n]}{s}$,

$$\begin{aligned}
&2\,\{\psi(w_{\boldsymbol{\eta}}(\boldsymbol{e}))\psi(w_{\boldsymbol{\eta}'}(\boldsymbol{e})) + (1 - \psi(w_{\boldsymbol{\eta}}(\boldsymbol{e})))(1 - \psi(w_{\boldsymbol{\eta}'}(\boldsymbol{e})))\} \\
&\leqslant 1 + \frac{1}{4} w_{\boldsymbol{\eta}}(\boldsymbol{e}) w_{\boldsymbol{\eta}'}(\boldsymbol{e}) + \frac{1}{24}(w_{\boldsymbol{\eta}}(\boldsymbol{e})^3 + w_{\boldsymbol{\eta}'}(\boldsymbol{e})^3) \\
&+ \frac{1}{24^2} w_{\boldsymbol{\eta}}(\boldsymbol{e})^3 w_{\boldsymbol{\eta}'}(\boldsymbol{e})^3.
\end{aligned}$$

Using this bound in (IX.3) gives,

$$\begin{aligned}
&\mathbb{E}_{H_0}[L_{\pi_n}^2] \\
&\leqslant \mathbb{E}_{\boldsymbol{\eta}, \boldsymbol{\eta}'}\Bigg[\prod_{\boldsymbol{e} \in \binom{[n]}{s}} \Bigg(1 + \frac{1}{4} w_{\boldsymbol{\eta}}(\boldsymbol{e}) w_{\boldsymbol{\eta}'}(\boldsymbol{e}) + \frac{1}{24}(w_{\boldsymbol{\eta}}(\boldsymbol{e})^3 \\
&\qquad + w_{\boldsymbol{\eta}'}(\boldsymbol{e})^3) + \frac{1}{24^2} w_{\boldsymbol{\eta}}(\boldsymbol{e})^3 w_{\boldsymbol{\eta}'}(\boldsymbol{e})^3\Bigg)\Bigg] \\
&\leqslant \mathbb{E}_{\boldsymbol{\eta}, \boldsymbol{\eta}'}\Bigg[\exp\Bigg(\sum_{\boldsymbol{e} \in \binom{[n]}{s}} \Bigg\{\frac{1}{4} w_{\boldsymbol{\eta}}(\boldsymbol{e}) w_{\boldsymbol{\eta}'}(\boldsymbol{e}) + \frac{1}{24}(w_{\boldsymbol{\eta}}(\boldsymbol{e})^3 \\
&\qquad + w_{\boldsymbol{\eta}'}(\boldsymbol{e})^3) + \frac{1}{24^2} w_{\boldsymbol{\eta}}(\boldsymbol{e})^3 w_{\boldsymbol{\eta}'}(\boldsymbol{e})^3\Bigg\}\Bigg)\Bigg], \qquad \text{(IX.4)}
\end{aligned}$$

since $1 + x \leqslant e^x$.

Recalling the definition of $w_{\boldsymbol{\eta}}(\boldsymbol{e})$ observe that

$$\left|\sum_{\boldsymbol{e} \in \binom{[n]}{s}} w_{\boldsymbol{\eta}}(\boldsymbol{e})^3\right| \leqslant \frac{\varepsilon^3}{n^{\frac{3}{2}}} \sum_{\boldsymbol{e} \in \binom{[n]}{s}} \left(\sum_{u \in \boldsymbol{e}} |\eta_u|\right)^3 \leqslant \varepsilon^3 n^{s - \frac{3}{2}}.$$

Hence,

$$\mathbb{E}\left[e^{2 \sum_{\boldsymbol{e} \in \binom{[n]}{s}} w_{\boldsymbol{\eta}}(\boldsymbol{e})^3}\right] \leqslant e^{2\varepsilon^3 n^{s - \frac{3}{2}}} \to 1, \qquad \text{(IX.5)}$$

since $\varepsilon \ll n^{-\frac{2s-3}{4}}$ and, for $s \geqslant 2$, $-\frac{s}{2} + \frac{3}{4} > 0$. Similarly, it can be shown that

$$\lim_{n \to \infty} \mathbb{E}\left[e^{2 \sum_{\boldsymbol{e} \in \binom{[n]}{s}} w_{\boldsymbol{\eta}}(\boldsymbol{e})^3 w_{\boldsymbol{\eta}'}(\boldsymbol{e})^3}\right] = 1. \qquad \text{(IX.6)}$$

Then Hölder's inequality applied to (IX.4) followed by (IX.5) and (IX.6) gives

$$\mathbb{E}_{H_0}[L_{\pi}^2] \leqslant \left\{\mathbb{E}_{\boldsymbol{\eta}, \boldsymbol{\eta}'}\left[e^{\frac{3}{4} \sum_{\boldsymbol{e} \in \binom{[n]}{s}} w_{\boldsymbol{\eta}}(\boldsymbol{e}) w_{\boldsymbol{\eta}'}(\boldsymbol{e})}\right]\right\}^{1/3} (1 + o(1)). \qquad \text{(IX.7)}$$

Next, observe that

$$\begin{aligned}
&\sum_{\boldsymbol{e} \in \binom{[n]}{s}} w_{\boldsymbol{\eta}}(\boldsymbol{e}) w_{\boldsymbol{\eta}'}(\boldsymbol{e}) \\
&= \frac{\varepsilon^2}{n}\left\{\sum_{\boldsymbol{e} \in \binom{[n]}{s}} \left(\sum_{u \in \boldsymbol{e}} \eta_u\right)\left(\sum_{v \in \boldsymbol{e}} \eta_v'\right)\right\} \\
&= \frac{\varepsilon^2}{n}\left\{\binom{n-1}{s-1} \sum_{u=1}^n \eta_u \eta_u' + \binom{n-2}{s-2} \sum_{1 \leqslant u \neq v \leqslant n} \eta_u \eta_v'\right\} \\
&\leqslant \varepsilon^2 n^{s-2} \sum_{u=1}^n \eta_u \eta_u' + \varepsilon^2 n^{s-3} \sum_{1 \leqslant u \neq v \leqslant n} \eta_u \eta_v' \\
&= \varepsilon^2 n^{s-2} \sum_{u=1}^n \eta_u \eta_u' \\
&+ \varepsilon^2 n^{s-3}\left\{\left(\sum_{u=1}^n \eta_u\right)\left(\sum_{v=1}^n \eta_v'\right) - \sum_{u=1}^n \eta_u \eta_u'\right\}. \qquad \text{(IX.8)}
\end{aligned}$$

Note that $\varepsilon^2 n^{s-3} |\sum_{u=1}^n \eta_u \eta_u'| \leqslant \varepsilon^2 n^{s-2}$. Hence,

$$\mathbb{E}\left[e^{\frac{9}{4}\varepsilon^2 n^{s-3} \sum_{u=1}^n \eta_u \eta_u'}\right] \lesssim e^{\varepsilon^2 n^{s-2}} \to 1, \qquad \text{(IX.9)}$$

since $\varepsilon \ll n^{-\frac{2s-3}{4}}$. From (IX.7), by Hölder's inequality followed by (IX.8) and (IX.9) gives

$$\mathbb{E}_{H_0}[L_\pi^2] \lesssim_s \left\{\mathbb{E}_{\boldsymbol{\eta},\boldsymbol{\eta}'}\left[e^{\frac{9}{4}\varepsilon^2 n^{s-2}\sum_{u=1}^n \eta_u\eta'_u}\right]\right\}^{1/9} \times \left\{\mathbb{E}_{\boldsymbol{\eta},\boldsymbol{\eta}'}\left[e^{\frac{9}{4}\varepsilon^2 n^{s-3}(\sum_{u=1}^n \eta_u)(\sum_{v=1}^n \eta'_v)}\right]\right\}^{1/9} (1+o(1)). \quad \text{(IX.10)}$$

Denote $X_n := \sum_{u=1}^n \eta_u$ and $Y_n := \sum_{v=1}^n \eta'_u$. Since X_n and Y_n are independent and each of them is a sum of i.i.d. Rademacher random variables,

$$\begin{aligned}\mathbb{E}_{\boldsymbol{\eta},\boldsymbol{\eta}'}\left[e^{\frac{9}{4}\varepsilon^2 n^{s-3}X_nY_n}\right] &= \mathbb{E}\left[\mathbb{E}\left[e^{\frac{9}{4}\varepsilon^2 n^{s-3}X_nY_n}|Y_n\right]\right]\\ &= \mathbb{E}\left[\left(\cosh\left(\frac{9}{4}\varepsilon^2 n^{s-3}Y_n\right)\right)^n\right]\\ &\leqslant \mathbb{E}\left[e^{\frac{81}{16}\varepsilon^4 n^{2s-5}Y_n^2}\right],\end{aligned}$$

where last step uses $\cosh(x) \leqslant e^{x^2}$, for all $x \in \mathbb{R}$. Since $|Y_n| \leqslant n$, this implies,

$$\mathbb{E}_{\boldsymbol{\eta},\boldsymbol{\eta}'}\left[e^{\frac{9}{4}\varepsilon^2 n^{s-3}X_nY_n}\right] \leqslant e^{\frac{81}{16}\varepsilon^4 n^{2s-5}Y_n^2} \leqslant e^{\frac{81}{16}\varepsilon^4 n^{2s-3}} \to 1, \quad \text{(IX.11)}$$

since $\varepsilon \ll n^{-\frac{2s-3}{4}}$. Next, observe that $\eta_u\eta'_u$, for $u = 1, \cdots, n$, are i.i.d. Rademacher random variables. Again using $\cosh(x) \leqslant e^{x^2}$ for all $x \in \mathbb{R}$, we can show that

$$\begin{aligned}&\mathbb{E}_{\boldsymbol{\eta},\boldsymbol{\eta}'}\left[e^{\frac{9}{4}\varepsilon^2 n^{s-2}\sum_{u=1}^n \eta_u\eta'_u}\right]\\ &= \left(\cosh\left(\frac{9}{4}\varepsilon^2 n^{s-2}\right)\right)^n \leqslant e^{\frac{81}{16}\varepsilon^4 n^{2s-3}} \to 1, \quad \text{(IX.12)}\end{aligned}$$

since $\varepsilon \ll n^{-\frac{2s-3}{4}}$. Hence, using (IX.11) and (IX.12) in (IX.10) gives,

$$\lim_{n\to\infty} E_{H_0}[L_\pi^2] = 1.$$

By Lemma 10, this completes the proof of Theorem 7 (b).

B. Testing Lower Bound in L_∞ Norm: Proof of Theorem 8 (b)

We choose $\boldsymbol{\gamma} = \mathbf{0}$, $\varepsilon \ll (\log n/n^{s-1})^{1/2}$, and construct a prior π_n on H_1 as in (III.8) as follows: Suppose $\boldsymbol{\gamma}_u \in \mathbb{R}^n$ with

$$\boldsymbol{\gamma}_u = \varepsilon \boldsymbol{e}_u,$$

for $u \in [n]$, where $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n$ being the canonical basis vectors in $\mathbb{R}^n$. Then π_n assigns probability $1/n$ to each $\boldsymbol{\gamma}_u$. Clearly, $\|\boldsymbol{\gamma} - \boldsymbol{\gamma}_u\|_\infty = \varepsilon$ for all $u \in [n]$. Then, for $H \in \mathcal{H}_{n,s}$, the π_n integrated likelihood ratio is given by

$$L_{\pi_n} = \frac{1}{n}\sum_{u\in[n]}\prod_{\boldsymbol{e}\in\binom{[n]}{s}:u\in\boldsymbol{e}} \frac{2\ e^{\varepsilon X_{\boldsymbol{e}}}}{1+e^{\varepsilon}},$$

where $X_{\boldsymbol{e}} := \mathbf{1}\{\boldsymbol{e} \in E(H)\}$. Then

$$L_{\pi_n}^2 = \frac{1}{n^2}\sum_{u\in[n]}\prod_{\boldsymbol{e}\in\binom{[n]}{s}:u\in\boldsymbol{e}} \frac{4\ e^{2\varepsilon X_{\boldsymbol{e}}}}{(1+e^{\varepsilon})^2} + \frac{1}{n^2}\sum_{u\neq v\in[n]}\prod_{\boldsymbol{e}\in\binom{[n]}{s}:u,v\in\boldsymbol{e}} \frac{4\ e^{2\varepsilon X_{\boldsymbol{e}}}}{(1+e^{\varepsilon})^2}. \quad \text{(IX.13)}$$

Observe that

$$\mathbb{E}_{H_0}\left[\prod_{\boldsymbol{e}\in\binom{[n]}{s}:u\in\boldsymbol{e}} \frac{4\ e^{2\varepsilon X_{\boldsymbol{e}}}}{(1+e^{\varepsilon})^2}\right] = \left(2\,\psi(\varepsilon)^2 + 2\,(1-\psi(\varepsilon))^2\right)^{\binom{n}{s-1}}, \quad \text{(IX.14)}$$

where $\psi(x) = \frac{e^x}{1+e^x}$. Similarly,

$$\mathbb{E}_{H_0}\left[\prod_{\boldsymbol{e}\in\binom{[n]}{s}:u,v\in\boldsymbol{e}} \frac{4\ e^{2\varepsilon X_{\boldsymbol{e}}}}{(1+e^{\varepsilon})^2}\right] = \left(2\,\psi(\varepsilon)^2 + 2\,(1-\psi(\varepsilon))^2\right)^{\binom{n}{s-2}}. \quad \text{(IX.15)}$$

Since $\varepsilon \ll (\log n/n^{s-1})^{1/2}$, a Taylor expansion around zero gives $\psi(\varepsilon) = \frac{1}{2} + \frac{1}{4}\varepsilon + O(\varepsilon^2)$. Hence,

$$2\psi(\varepsilon)^2 + 2(1-\psi(\varepsilon))^2 = 1 + O(\varepsilon^2).$$

Therefore, by (IX.14)-(IX.15) and using $1 + x \leqslant e^x$ gives,

$$\mathbb{E}_{H_0}[L_{\pi_n}^2] \leqslant \frac{1}{n}e^{O(\varepsilon^2 n^{s-1})} + e^{O(\varepsilon^2 n^{s-2})} \to 1,$$

since $\varepsilon \ll (\log n/n^{s-1})^{1/2}$. By Lemma 10, this completes the proof of Theorem 8 (b).

X. Proof of Proposition 1

Define $g = (g_1, g_2, \ldots, g_n) : \mathbb{R}^n \to \mathbb{R}^n$ where $g_u : \mathbb{R}^n \to \mathbb{R}$, for $u \in [n]$, as follows:

$$g_u(\boldsymbol{x}) = \sum_{\boldsymbol{e}\in\binom{[n]}{s}:u\in\boldsymbol{e}} \frac{e^{\boldsymbol{x}_{\boldsymbol{e}}^\top \mathbf{1}}}{1+e^{\boldsymbol{x}_{\boldsymbol{e}}^\top \mathbf{1}}},$$

where $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)^\top$ and $\boldsymbol{x}_{\boldsymbol{e}} = (x_{u_1}, x_{u_2}, \ldots, x_{u_s})^\top$ for $\boldsymbol{e} = (u_1, u_2, \ldots, u_s)$. Observe that $\mathcal{R}_s$ is the range of g. Since the expected degree of a vertex is a weighted combination of all the possible degrees in s-uniform hypergraphs on n vertices, this implies $\bar{\mathcal{R}}_s \subseteq \mathrm{conv}\,(\mathcal{D}_s)$.

To show the other side, for every $\boldsymbol{y} \in \mathbb{R}^n$ we define,

$$f_{\boldsymbol{y}}(\boldsymbol{x}) = \sum_{i=1}^n x_i y_i - \sum_{\{v_1,v_2,\ldots,v_s\}\in\binom{[n]}{s}} \log(1 + e^{x_{v_1}+\ldots+x_{v_s}}).$$

Since the probability of observing an s-uniform hypergraph with parameter $\boldsymbol{x}$ and s-degree sequence $\boldsymbol{d}_s = (d_s(1), \ldots, d_s(n))$ is

$$\frac{e^{\sum_{v=1}^n d_s(v)x_v}}{\prod_{\{v_1,v_2,\ldots,v_s\}\in\binom{[n]}{s}}(1 + e^{x_{v_1}+\ldots+x_{v_s}})}.$$

and is less than 1, taking logarithm on both sides we get $f_{\boldsymbol{d}_s}(\boldsymbol{x}) \leqslant 0$. Further as $f_{\boldsymbol{y}}(\boldsymbol{x})$ depends linearly on $\boldsymbol{y}$, we have $f_{\boldsymbol{y}}(\boldsymbol{x}) \leqslant 0$ for all $\boldsymbol{y} \in \mathrm{conv}\,(\mathcal{D}_s)$ and $\boldsymbol{x} \in \mathbb{R}^n$. Now, let us fix $\boldsymbol{y} \in \mathrm{conv}\,(\mathcal{D}_s)$. It can be shown that the Hessian $\nabla^2 f_{\boldsymbol{y}}(\boldsymbol{x})$ is uniformly bounded, hence, by [10, Lemma 3.1] there exists a sequence $\{\boldsymbol{x}_k\}_{k\geqslant 1}$ such that $\nabla f_{\boldsymbol{y}}(\boldsymbol{x}_k) \to 0$. Observing that $\nabla f_{\boldsymbol{y}}(\boldsymbol{x}_k) = \boldsymbol{y} - g(\boldsymbol{x})$, we get $g(\boldsymbol{x}_k) \to \boldsymbol{y}$. As $\boldsymbol{y} \in \mathrm{conv}\,(\mathcal{D}_s)$ is arbitrary, this implies $\mathrm{conv}\,(\mathcal{D}_s) \subseteq \bar{\mathcal{R}}_s$. $\square$

ACKNOWLEDGMENT

The authors thank the anonymous referees for their insightful comments which improved the quality and the presentation of the article. They also thank Rui Feng for pointing out an error in a previous draft.

REFERENCES

[1] S. Agarwal, J. Lim, L. Zelnik-Manor, P. Perona, D. J. Kriegman, and S. J. Belongie, "Beyond pairwise clustering," in *Proc. Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2005, pp. 838–845.

[2] K. Ahn, K. Lee, and C. Suh, "Community recovery in hypergraphs," *IEEE Trans. Inf. Theory*, vol. 65, no. 10, pp. 6561–6579, Oct. 2019.

[3] K. Balasubramanian, "Nonparametric modeling of higher-order interactions via hypergraphons," *J. Mach. Learn. Res.*, vol. 22, no. 1, pp. 6464–6498, 2021.

[4] F. Battiston et al., "Networks beyond pairwise interactions: Structure and dynamics," *Phys. Rep.*, vol. 874, pp. 1–92, Aug. 2020.

[5] F. Battiston et al., "The physics of higher-order interactions in complex systems," *Nat. Phys.*, vol. 17, no. 10, pp. 1093–1098, 2021.

[6] A. R. Benson, D. F. Gleich, and J. Leskovec, "Higher-order organization of complex networks," *Science*, vol. 353, no. 6295, pp. 163–166, Jul. 2016.

[7] J. Blitzstein and P. Diaconis, "A sequential importance sampling algorithm for generating random graphs with prescribed degrees," *Internet Math.*, vol. 6, no. 4, pp. 489–522, Mar. 2011.

[8] E. Breza, A. G. Chandrasekhar, S. Lubold, T. H. McCormick, and M. Pan, "Consistently estimating network statistics using aggregated relational data," *Proc. Nat. Acad. Sci. USA*, vol. 120, no. 21, May 2023, Art. no. e2207185120.

[9] B. M. Brown, "Martingale central limit theorems," *Ann. Math. Statist.*, vol. 42, no. 1, pp. 59–66, Feb. 1971.

[10] S. Chatterjee, P. Diaconis, and A. Sly, "Random graphs with a given degree sequence," *Ann. Appl. Probab.*, vol. 21, no. 4, pp. 1400–1435, Aug. 2011.

[11] M. Chen, K. Kato, and C. Leng, "Analysis of networks via the sparse β-model," *J. Roy. Stat. Soc. B, Stat. Methodol.*, vol. 83, no. 5, pp. 887–910, 2021.

[12] N. Chen and M. Olvera-Cravioto, "Directed random graphs with given degree distributions," *Stochastic Syst.*, vol. 3, no. 1, pp. 147–186, 2013.

[13] P. Chen, C. Gao, and A. Y. Zhang, "Partial recovery for top-k ranking: Optimality of MLE and SubOptimality of the spectral method," *Ann. Statist.*, vol. 50, no. 3, pp. 1618–1652, Jun. 2022.

[14] Y. Chen, J. Fan, C. Ma, and K. Wang, "Spectral method and regularized MLE are both optimal for top-K ranking," *Ann. Statist.*, vol. 47, no. 4, pp. 2204–2235, Aug. 2019.

[15] I. E. Chien, C.-Y. Lin, and I.-H. Wang, "On the minimax misclassification ratio of hypergraph community detection," *IEEE Trans. Inf. Theory*, vol. 65, no. 12, pp. 8095–8118, Dec. 2019.

[16] T. Elmer, K. Mepham, and C. Stadtfeld, "Students under lockdown: Comparisons of students? Social networks and mental health before and during the COVID-19 crisis in Switzerland," *PLoS ONE*, vol. 15, no. 7, 2020, Art. no. e023633.

[17] G. Ghoshal, V. Zlatić, G. Caldarelli, and M. E. J. Newman, "Random hypergraphs and their applications," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 79, no. 6, Jun. 2009, Art. no. 066118.

[18] D. Ghoshdastidar and A. Dukkipati, "A provable generalized tensor spectral method for uniform hypergraph partitioning," in *Proc. Int. Conf. Mach. Learn.*, 2015, pp. 400–409.

[19] D. Ghoshdastidar and A. Dukkipati, "Spectral clustering using multilinear SVD: Analysis, approximations and applications," in *Proc. AAAI Conf. Artif. Intell.*, vol. 29, 2015.

[20] D. Ghoshdastidar and A. Dukkipati, "Consistency of spectral hypergraph partitioning under planted partition model," *Ann. Statist.*, vol. 45, no. 1, pp. 289–315, Feb. 2017.

[21] D. Ghoshdastidar and A. Dukkipati, "Uniform hypergraph partitioning: Provable tensor methods and sampling techniques," *J. Mach. Learn. Res.*, vol. 18, no. 1, pp. 1638–1678, 2017.

[22] T. Gracious, A. Gupta, and A. Dukkipati, "Neural temporal point process for forecasting higher order and directional interactions," 2023, *arXiv:2301.12210*.

[23] B. S. Graham, "An econometric model of network formation with degree heterogeneity," *Econometrica*, vol. 85, no. 4, pp. 1033–1063, 2017.

[24] J. Grilli, G. Barabás, M. J. Michalska-Smith, and S. Allesina, "Higher-order interactions stabilize dynamics in competitive network models," *Nature*, vol. 548, no. 7666, pp. 210–213, 2017.

[25] C. Hillar and A. Wibisono, "Maximum entropy distributions on graphs," 2013, *arXiv:1301.3321*.

[26] P. W. Holland and S. Leinhardt, "An exponential family of probability distributions for directed graphs," *J. Amer. Stat. Assoc.*, vol. 76, no. 373, pp. 33–50, Mar. 1981.

[27] J. Hu and M. Wang, "Multiway spherical clustering via degree-corrected tensor block models," in *Proc. Int. Conf. Artif. Intell. Statist.*, 2022, pp. 1078–1119.

[28] M. O. Jackson et al., *Social and Economic Networks*, vol. 3. Princeton, NJ, USA: Princeton Univ. Press, 2008.

[29] P. Ji and J. Jin, "Coauthorship and citation networks for statisticians," *Ann. Appl. Statist.*, vol. 10, no. 4, pp. 1779–1812, Dec. 2016.

[30] F. W. Crawford, "Discussion of 'Coauthorship and citation networks for statisticians,'" *Ann. Appl. Statist.*, vol. 10, no. 4, pp. 1827–1834, Dec. 2016.

[31] V. Karwa and A. Slavković, "Inference using noisy degrees: Differentially private β-model and synthetic graphs," *Ann. Statist.*, vol. 44, no. 1, pp. 87–112, Feb. 2016.

[32] G. Karypis and V. Kumar, "Multilevel k-way hypergraph partitioning," in *Proc. Design Autom. Conf.*, 1999, pp. 343–348.

[33] Z. Tracy Ke, F. Shi, and D. Xia, "Community detection for hypergraph networks via regularized tensor power iteration," 2019, *arXiv:1909.06503*.

[34] Q. F. Lotito, F. Musciotto, A. Montresor, and F. Battiston, "Higher-order motif analysis in hypergraphs," *Commun. Phys.*, vol. 5, no. 1, p. 79, Apr. 2022.

[35] Q. Francesco Lotito, F. Musciotto, A. Montresor, and F. Battiston, "Hyperlink communities in higher-order networks," 2023, *arXiv:2303.01385*.

[36] S. Lunagómez, S. Mukherjee, R. L. Wolpert, and E. M. Airoldi, "Geometric representations of random hypergraphs," *J. Amer. Stat. Assoc.*, vol. 112, no. 517, pp. 363–383, Jan. 2017.

[37] T. Michoel and B. Nachtergaele, "Alignment and integration of complex networks by hypergraph-based spectral clustering," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 86, no. 5, Nov. 2012, Art. no. 056111.

[38] R. Mukherjee, S. Mukherjee, and S. Sen, "Detection thresholds for the β-model on sparse graphs," *Ann. Statist.*, vol. 46, no. 3, pp. 1288–1317, Jun. 2018.

[39] S. Mukherjee, "Degeneracy in sparse ERGMs with functions of degrees as sufficient statistics," *Bernoulli*, vol. 26, no. 2, pp. 1016–1043, May 2020.

[40] S. Pal and Y. Zhu, "Community detection in the sparse hypergraph stochastic block model," *Random Struct. Algorithms*, vol. 59, no. 3, pp. 407–463, Oct. 2021.

[41] J. Park and M. E. Newman, "Statistical mechanics of networks," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 70, no. 6, 2004, Art. no. 066117.

[42] A. Patania, G. Petri, and F. Vaccarino, "The shape of collaborations," *EPJ Data Sci.*, vol. 6, no. 1, pp. 1–16, Dec. 2017.

[43] G. Petri et al., "Homological scaffolds of brain functional networks," *J. Roy. Soc. Interface*, vol. 11, no. 101, Dec. 2014, Art. no. 20140873.

[44] A. Rinaldo, S. Petrović, and S. E. Fienberg, "Maximum lilkelihood estimation in the β-model," *Ann. Statist.*, vol. 41, no. 3, pp. 1085–1110, Jun. 2013.

[45] D. Stasi, K. Sadeghi, A. Rinaldo, S. Petrovic, and S. Fienberg, "β models for random hypergraphs with a given degree sequence," in *Proc. 21st Int. Conf. Comput. Statist.*, 2014, pp. 593–600.

[46] S. Stein and C. Leng, "A sparse beta regression model for network analysis," 2020, *arXiv:2010.13604*.

[47] A. B. Tsybakov, "Introduction to nonparametric estimation," in *Springer Series in Statistics*. New York, NY, USA: Springer, 2009.

[48] R. Vershynin, "Introduction to the non-asymptotic analysis of random matrices," 2010, *arXiv:1011.3027*.

[49] J. Wahlström, I. Skog, P. S. La Rosa, P. Händel, and A. Nehorai, "The β-model–maximum likelihood, Cramér–Rao bounds, and hypothesis testing," *IEEE Trans. Signal Process.*, vol. 65, no. 12, pp. 3234–3246, May 2017.

[50] Y. Xu and S. Mukherjee, "Signal detection in degree corrected ERGMs," 2021, *arXiv:2108.09255*.

[51] N. Yadati, M. Nimishakavi, P. Yadav, V. Nitin, A. Louis, and P. Talukdar, "HyperGCN: A new method of training graph convolutional networks on hypergraphs," in *Proc. 33rd Int. Conf. Neural Inf. Process. Syst.*, 2019, pp. 1511–1522.
[52] T. Yan and J. Xu, "A central limit theorem in the β-model for undirected random graphs with a diverging number of vertices," *Biometrika*, vol. 100, no. 2, pp. 519–524, Jun. 2013.
[53] T. Yan, C. Leng, and J. Zhu, "Asymptotics in directed exponential random graph models with an increasing bi-degree sequence," *Ann. Statist.*, vol. 44, no. 1, pp. 31–57, Feb. 2016.
[54] T. Yan, H. Qin, and H. Wang, "Asymptotics in undirected random graph models parameterized by the strengths of vertices," *Statistica Sinica*, vol. 26, pp. 273–293, Jan. 2016.
[55] T. Yan, B. Jiang, S. E. Fienberg, and C. Leng, "Statistical inference in a directed network model with covariates," *J. Amer. Stat. Assoc.*, vol. 114, no. 526, pp. 857–868, Apr. 2019.
[56] T. Yan, Y. Li, J. Xu, Y. Yang, and J. Zhu, "Wilks' theorems in the β-model," 2022, *arXiv:2211.10055*.
[57] J.-G. Young, G. Petri, and T. P. Peixoto, "Hypergraph reconstruction from network data," *Commun. Phys.*, vol. 4, no. 1, p. 135, Jun. 2021.
[58] M. Yuan, R. Liu, Y. Feng, and Z. Shang, "Testing community structure for hypergraphs," *Ann. Statist.*, vol. 50, no. 1, pp. 147–169, Feb. 2022.
[59] Q. Zhang and V. Y. F. Tan, "Exact recovery in the general hypergraph stochastic block model," *IEEE Trans. Inf. Theory*, vol. 69, no. 1, pp. 453–471, Jan. 2023.
[60] M. Shao, Y. Zhang, Q. Wang, Y. Zhang, J. Luo, and T. Yan, "L-2 regularized maximum likelihood for β-model in large and sparse networks," 2021, *arXiv:2110.11856*.

Sagnik Nandy received the B.Stat. and M.Stat. degrees from the Indian Statistical Institute, Kolkata, India, in 2017 and 2019, respectively, and the Ph.D. degree in statistics and data science from the University of Pennsylvania in 2024. He is currently a Kruskal Instructor with the Department of Statistics, University of Chicago, Chicago, IL, USA. His research interests include high dimensional statistics, random matrix theory, and inference on networks.

Bhaswar B. Bhattacharya received the B.Stat. and M.Stat. degrees from the Indian Statistical Institute, Kolkata, India, in 2009 and 2011, respectively, and the Ph.D. degree in statistics from Stanford University in 2016. He is currently an Associate Professor with the Department of Statistics and Data Science, Wharton School, University of Pennsylvania, Philadelphia, PA, USA. His research interests include non-parametric statistics, combinatorial probability, inference on networks, and discrete and computational geometry. He was a recipient of the NSF CAREER Award and the Sloan Research Fellowship.