

# A Unified Framework for Correlation Mining in Ultra-High Dimension

Yun Wei, Bala Rajaratnam, and Alfred O. Hero<sup>id</sup>, *Life Fellow, IEEE*

**Abstract**—Many applications benefit from theory relevant to the identification of variables having large correlations or partial correlations in high dimension. Recently there has been progress in the ultra-high dimensional setting when the sample size  $n$  is fixed and the dimension  $p$  tends to infinity. Despite these advances, the correlation screening framework suffers from practical, methodological and theoretical deficiencies. For instance, previous correlation screening theory requires that the population covariance matrix be sparse and block diagonal. This block sparsity assumption is however restrictive in practical applications. As a second example, correlation and partial correlation screening requires the estimation of dependence measures, which can be computationally prohibitive. In this paper, we propose a unifying approach to correlation and partial correlation mining that is not restricted to block diagonal correlation structure, thus yielding a methodology that is suitable for modern applications. By making connections to random geometric graphs, the number of highly correlated or partial correlated variables are shown to have compound Poisson finite-sample characterizations, which hold for both the finite  $p$  case and when  $p$  tends to infinity. The unifying framework also demonstrates a duality between correlation and partial correlation screening with theoretical and practical consequences.

**Index Terms**—Correlation analysis, compound poisson, fixed  $n$  large  $p$  asymptotics, random geometric graph, Stein's method.

## I. INTRODUCTION

THIS paper considers the problem of identifying high correlations and partial correlations in modern ultra high dimensional setting. In particular we study the problem of screening  $n$  identically distributed  $p$ -variate samples for variables that have high correlation or high partial correlation with at least one other variable when the sample size  $n \leq C_0 \ln p$

for some constant  $C_0$ . In the screening framework one applies a threshold to the sample correlation matrix or the sample partial correlation matrix to detect variables with at least one significant correlation. The threshold serves to separate signal from noise. Correlation and partial correlation screening in ultra-high dimensions with few samples arises frequently in applications where the per-sample cost of collecting high dimensional data is much more costly than the per-variable cost. For example, in genomic correlation screening the cost of high throughput RNAseq assays is decreasing faster than the cost of biological samples [1]. In such situations  $p$  is much larger than  $n$ .

The ultra-high dimensional regime when  $n \leq C_0 \ln p$  is challenging since the number of samples is insufficient to guarantee reliability of commonly applied statistical methods. For example, one way to undertake partial correlation screening is to first estimate the population covariance matrix, then obtain the inverse, from which a partial correlation matrix can be estimated. However, to get a reliable estimate of a general covariance matrix, the number of samples  $n$  must be at least  $\Omega(p)$  as shown in [2, Section 5.4.3]. Even if the covariance matrix has a special structure like sparsity, covariance estimation requires the number of samples be of order  $\Omega(\ln p)$ . The reader is referred to [3], [4], [5], [6], [7], [8], [9], [10], [11], [12], [13], [14], [15], [16] and the references therein for related work in modern high dimensional covariance selection and estimation.

While estimating the correlation matrix or partial correlation matrix is challenging in ultra-high dimensions, recent work [17] and [18] has shown that it is possible to accurately test the number of highly (partial) correlated variables under a false positive probability constraint; in particular the probability that a variable is spuriously (partially) correlated with at least one other variable. While correlation screening finds variables that have a high marginal correlation with at least another variable, partial correlation screening identifies variables that have high conditional correlations with one other variable conditioned on the rest. In [17], the ultra-high dimensional correlation screening problem is studied under a row-sparsity assumption on the population covariance matrix. A phase transition in the number of false positive correlations was characterized as a function of the correlation threshold and the true covariance. In the case of block sparse covariance, the critical phase transition threshold becomes independent of the true covariance. In [18] the partial correlation screening problem was studied, and similar phase transition results as

Manuscript received 4 March 2021; revised 8 March 2022; accepted 2 August 2022. Date of publication 19 August 2022; date of current version 22 December 2022. This work was supported in part by the U.S. Army Research Office under Grant W911NF-15-1-0479 and Grant W911NF-19-1-0269 and in part by the U.S. National Science Foundation under Grant NSF DMS 1976787, Grant NSF DMS 1916787, and Grant NSF CCF 1934568. (Corresponding author: Alfred O. Hero.)

Yun Wei was with the Department of Mathematics, University of Michigan, Ann Arbor, MI 48109 USA. He is now with the Department of Statistical Science, Duke University, Durham, NC 27707 USA (e-mail: yun.wei@duke.edu).

Bala Rajaratnam is with the Department of Statistics, University of California at Davis, Davis, CA 95616 USA (e-mail: brajaratnam@ucdavis.edu).

Alfred O. Hero is with the Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, MI 48109 USA (e-mail: hero@eecs.umich.edu).

Communicated by M. Davenport, 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.2022.3200577>.

Digital Object Identifier 10.1109/TIT.2022.3200577

in correlation screening [17] were obtained under the block-sparse assumption on the population covariance matrix. The survey [19] reviews the correlation and partial correlation screening problems. A follow on work [20] of [17] applies a similar framework to the spurious correlations problem and the low-rank detection problem. The reader is referred to [5], [6], [7], [11], [12], [21], [22] and the references therein for additional work related to high dimensional sample correlation matrices.

Despite these advances in correlation and partial correlation screening, the screening framework proposed in [17] and [18] has some serious methodological, theoretical and practical shortcomings. For instance, results for partial correlation screening impose a restrictive block sparsity assumption on the true underlying correlation matrix. The block sparsity condition in [18] requires that only a small group of the variables have correlation within the blocks and have no correlations with variables outside the block. This assumption is severely restrictive for cases where variables have correlations within a group and also correlations with variables outside their respective groups. Furthermore, expressions for false discovery probabilities in [17] and [18] require estimating dependence functionals. Estimating such functionals lead to computationally prohibitive non-parametric estimation, rendering the screening methodology disconnected from the very setting it was designed for.

In this paper we propose a unifying framework for correlation and partial correlation screening that delivers a practical and scalable variable selection in the ultra-high dimensional regime. By making novel connections to random geometric graphs [23], we demonstrate that the distribution of the number of discoveries beyond a certain threshold is approximated by a compound Poisson distribution, with different parameters in the regimes when  $p$  is finite and when  $p$  approaches  $\infty$ . To the best of our knowledge, such characterization has not previously appeared in the literature. Furthermore, our results are proved in greater generality by relaxing the block-sparse assumption to a new sparsity condition, defined as  $(\tau, \kappa)$  sparsity in Section II-C, on the population covariance matrix. The block-sparse assumption is a special case of the  $(\tau, \kappa)$  sparsity assumption. The characterizations established in this paper depend on the covariance matrix only through the  $(\tau, \kappa)$  sparsity condition. Moreover, the assumptions of  $(\tau, \kappa)$  sparsity allows us to formulate unified theorems, which covers both the cases of correlation and of partial correlation screening.

The theory in this paper is directly relevant to hypothesis testing concerning the empirical degree distribution of a correlation graph. This topic arises in a wide spectrum of areas including graph mining, network science, social science, and natural sciences [24], [25], [26]. Variables having strong sample correlations will appear in the correlation graph as vertices having positive vertex degree. As one sweeps over vertex degree values, the histogram of vertex degrees specifies the empirical degree distribution of the graph. From this perspective, this paper provides compound Poisson characterizations of the empirical degree distribution for large correlation graphs under more realistic sparsity conditions on

the population covariance. The expressions that are derived from our theorems also provide approximations to family-wise error rates associated with false discoveries of vertices of degree exceeding a specified fixed degree.

Finite sample results for controlling the probability of discovering a (false) partial correlation in high-dimensional thresholded covariance settings have been elusive for the better part of the last decade and a half. Indeed, evaluating expressions for such probabilities in the fixed  $n$  setting are known to be a notoriously difficult problem. This difficulty is in part attributed to the dependence of such false positive probabilities on the unknown covariance parameter. Previous work has instead provided expressions for the probability that two distinct connectivity components of the partial correlation graph are falsely joined (see [27] and the references therein for more detail). Controlling such probabilities implicitly assumes that the covariance parameter is block diagonal. Such an assumption is tantamount to requiring that the true partial correlation graph is not fully connected, a restrictive assumption in many application areas. In contrast, the  $(\tau, \kappa)$  sparsity condition introduced in this paper allows the underlying graph to be fully connected. Moreover, we provide finite sample results for controlling the probability that a (partial) correlation is falsely discovered under the above-mentioned  $(\tau, \kappa)$  notion of sparsity.

#### A. Contribution

We summarize the principal contributions of the paper. As above  $p$  denotes the number of variables and  $n$  denotes the number of samples.

- 1) The paper presents a unified and complete asymptotic analysis of the star subgraph counts, the counts of vertices of a given degree and the counts of vertices above a given degree in the random graphs obtained by thresholding the sample correlation and sample partial correlation. This unification of different types of random counts represents an improvement over previous work [17], [18] where only the counts of vertices above a given degree are studied.
- 2) We approximate the full distributions of the random counts for finite  $p$  and as  $p \rightarrow \infty$ . In addition, we characterize the first and second moments of these random counts. This is a generalization of previous results [17], [18] that only established approximations for the mean number of random counts and for the probabilities that these counts were positive.
- 3) We obtain a compound Poisson characterization of the distributions of the random counts. The compound Poisson limit and approximation are well approximated by the standard Poisson limit when  $n$  is moderately large (Section V). This result corrects and refines the claim in [18] that erroneously asserted a Poisson limit.
- 4) The theory in this paper is developed under a novel sparsity condition on the population dispersion matrix. This sparsity condition, called  $(\tau, \kappa)$  sparsity in Sec. II-C, is weaker than previously assumed conditions, which makes our theory more broadly applicable. Specifically, while the block sparsity condition in previous work [9],

[18], [28] imposes that correlation can only occur locally in small blocks of variables, the  $(\tau, \kappa)$  sparsity condition relaxes this condition to more general global correlation patterning.

Some of the broader implications of the technical contributions of this paper are described below.

Previous conditions on population partial correlation networks assume they are of lower dimension. In particular, conditions such as block sparsity do not allow for completely connected partial correlation graphs which involve all  $p$  variables. Such restrictive assumptions are difficult to validate and rule out many realistic population (partial) correlation structures. Overcoming this hurdle has been an open problem for several years. The newly introduced  $(\tau, \kappa)$  sparsity condition on the population covariance matrix settles this longstanding problem by successfully allowing for completely connected partial correlation graphs over the entire set of features.

Historically, the literature on correlation estimation and graphical models has separated the treatments of covariance graph models and undirected graphical models (or inverse covariance graph models) [29]. Unifying the two classes of statistical models has been an open problem for the better part of almost 3 decades. While this separate treatment may be appropriate in low dimensional settings when there are few variables, it is not immediately obvious which of the two frameworks is appropriate for a given data set in modern ultra-high dimensional regimes. To our knowledge, the framework in this paper is the first in the graphical model or correlation graph estimation literature to propose methodology which brings both approaches under one umbrella.

The results in the paper also have relevance to applications. Recall that our Poisson and compound Poisson expressions effectively describe the number of false discoveries and hence allow us to obtain results for the familywise error rate (FWER) or k-FWER, that is the probability of obtaining  $k$  or more false discoveries. Note that we can also obtain the marginal distributions of correlation estimates, which in turn allow us to obtain expressions for p-values for testing correlation estimates. These marginal p-values allow us to establish FDR control too using either the Benjamini-Hochberg [30] or Benjamini-Yakutelli procedures [31]. In summary, the correlations screening framework is sufficiently rich that it allows us to undertake statistical error control in terms of FWER, k-FWER and FDR. This is one of the main strengths of our results: a rigorous inferential framework in the ultra-high dimensional setting.

The remainder of this paper is organized as follows. Section II outlines the framework and presents our main theorem which characterizes the compound Poisson approximations when  $p$  approaches  $\infty$ . In Section III an approximating theorem when  $p$  is finite is presented, based on which the main theorem follows. Section IV covers convergence of moments. Section V provides explicit expressions for the parameters of the compound Poisson characterizations. Notation and symbols used in this paper are collected in the Section A of the Appendix. Most of the technical proofs and auxiliary results are given in the Appendix.

## II. MAIN RESULTS

### A. Framework

Available is a data matrix consisting of multivariate samples

$$\mathbf{X} = [\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \dots, \mathbf{x}^{(n)}]^\top = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_p] \in \mathbb{R}^{n \times p}, \quad (1)$$

where  $\{\mathbf{x}^{(i)}\}_{i=1}^n \subset \mathbb{R}^p$  are samples from a  $p$ -dimensional distribution. We assume that the  $n \times p$  data matrix  $\mathbf{X}$  follows a vector elliptically contoured distribution [32], [33], [34]. A random matrix  $\mathbf{X} \in \mathbb{R}^{n \times p}$  is vector-elliptical with positive definite covariance or dispersion parameter  $\Sigma \in \mathbb{R}^{p \times p}$  and location parameter  $\mu$  if its density satisfies

$$f_{\mathbf{X}}(\mathbf{X}) = \det(\Sigma)^{-n/2} g(\text{tr}((\mathbf{X} - \mathbf{1}\mu^\top)\Sigma^{-1}(\mathbf{X}^\top - \mathbf{1}\mu^\top))), \quad (2)$$

for a shape function  $g: \mathbb{R} \rightarrow [0, \infty)$  such that  $\int f_{\mathbf{X}}(\mathbf{X}) = 1$ . In (2),  $\mathbf{1}$  is a column vector with all elements equal to 1. We use the shorthand  $\mathbf{X} \sim \mathcal{VE}(\mu, \Sigma, g)$  to denote that  $\mathbf{X}$  follows a vector elliptically contoured distribution with density (2). Note that the rows  $\{\mathbf{x}^{(i)}\}_{i=1}^n$  of  $\mathbf{X}$  are uncorrelated but not necessarily independent [34]. An example of a vector-elliptical distributed is the matrix normal distribution, for which the rows  $\{\mathbf{x}^{(i)}\}_{i=1}^n \subset \mathbb{R}^p$  are i.i.d. samples from  $\mathcal{N}(\mu, \Sigma)$ . Specifically, the matrix normal density is obtained when, in (2),  $g(w) = g_0(w) = (2\pi)^{-\frac{np}{2}} \exp(-\frac{1}{2}w)$  and in this case  $\mathbf{X} \sim \mathcal{VE}(\mu, \Sigma, g_0)$ .

Given a data matrix  $\mathbf{X} \sim \mathcal{VE}(\mu, \Sigma, g)^1$ , the sample mean  $\bar{\mathbf{x}}$  is given as a row vector

$$\bar{\mathbf{x}} = \frac{1}{n} \sum_{i=1}^n \mathbf{x}^{(i)} = \frac{1}{n} \mathbf{X}^\top \mathbf{1}.$$

The sample covariance matrix  $\mathbf{S}$  is

$$\mathbf{S} = \frac{1}{n-1} \sum_{i=1}^n (\mathbf{x}^{(i)} - \bar{\mathbf{x}})(\mathbf{x}^{(i)} - \bar{\mathbf{x}})^\top = \frac{1}{n-1} \mathbf{X}^\top \left( \mathbf{I}_n - \frac{1}{n} \mathbf{1}\mathbf{1}^\top \right) \mathbf{X}. \quad (3)$$

The sample correlation matrix  $\mathbf{R}$  is defined as:

$$\mathbf{R} = \text{diag}(\mathbf{S})^{-\frac{1}{2}} \mathbf{S} \text{diag}(\mathbf{S})^{-\frac{1}{2}}, \quad (4)$$

where  $\text{diag}(\mathbf{A})$  for a matrix  $\mathbf{A} \in \mathbb{R}^{n \times n}$  is the diagonal part of  $\mathbf{A}$  and  $\mathbf{B}^{-1/2}$  for a diagonal matrix  $\mathbf{B}$  is a diagonal matrix formed by raising every diagonal element of  $\mathbf{B}$  to the power  $-1/2$ . Since  $\mathbf{R}$  is not invertible, we define the sample partial correlation matrix  $\mathbf{P}$  by

$$\mathbf{P} = \text{diag}(\mathbf{R}^\dagger)^{-\frac{1}{2}} \mathbf{R}^\dagger \text{diag}(\mathbf{R}^\dagger)^{-\frac{1}{2}}, \quad (5)$$

where  $\mathbf{R}^\dagger$  is the Moore-Penrose pseudo-inverse of  $\mathbf{R}$ .

Let  $\Psi = (\Psi_{ij})_{i,j \in [p]}$  be generic notation for a correlation-type matrix like  $\mathbf{R}$  or  $\mathbf{P}$ . Given a threshold  $\rho \in [0, 1)$  define the undirected graph induced by thresholding  $\Psi$ , denoted by  $\mathcal{G}_\rho(\Psi)$ , as follows. The vertex set of graph  $\mathcal{G}_\rho(\Psi)$  is  $V(\Psi) = [p] := \{1, 2, \dots, p\}$  and the edge set is  $E(\Psi) \subset V(\Psi) \times V(\Psi)$ , with  $(i, j) \in E(\Psi)$  if  $|\Psi_{ij}| \geq \rho$ , where  $(i, j)$  denotes an edge between  $i$  and  $j$  ( $i \neq j$ ). We call  $\mathcal{G}_\rho(\Psi)$  the empirical

<sup>1</sup>In previous work [17], [18] it was assumed that the samples  $\mathbf{x}^{(i)}$  are i.i.d. elliptical contoured distributed. This condition is in fact insufficient and the stronger vector elliptical contoured distribution condition (2) is required.

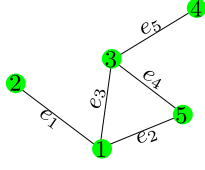


Fig. 1. A graph with 5 vertices and 5 edges.

correlation graph and the empirical partial correlation graph, respectively, when  $\Psi = \mathbf{R}$  and  $\Psi = \mathbf{P}$ . Let  $\Phi^{(\Psi)}(\rho)$  be the adjacency matrices associated with the graph  $\mathcal{G}_\rho(\Psi)$ , with elements  $\Phi_{ij}^{(\Psi)}(\rho) := 1(|\Psi_{ij}| \geq \rho)$  for  $i \neq j$ , where  $1(\cdot)$  is the indicator function. The dependence of  $\Phi^{(\Psi)}(\rho)$  and  $\Phi_{ij}^{(\Psi)}(\rho)$  on  $\rho$  will be suppressed when it is clear from the context.

The focus of this paper is correlation and partial screening, which counts the number of vertices of prescribed degree, the number of star subgraphs, or the number of edges in  $\mathcal{G}_\rho(\Psi)$ . The objective is to characterize the distributions of these counting statistics. More specifically, for the graph  $\mathcal{G}_\rho(\Psi)$  with  $\Psi = \mathbf{R}$  or  $\Psi = \mathbf{P}$ , the degree of vertex  $i$  is defined as  $\sum_{j=1, j \neq i}^p \Phi_{ij}^{(\Psi)}(\rho)$ . For  $1 \leq \delta \leq p-1$ , the total number of vertices with degree exactly  $\delta$  (at least  $\delta$ ), denoted by  $N_{V_\delta}^{(\Psi)}$  ( $N_{V_\delta}^{(\Psi)}$ ), is of particular interest. Note that if a vertex has degree exactly  $\delta$ , then there exists a star subgraph with  $\delta$  edges centered at that vertex. Consequently the number of star subgraphs are important in the analysis of  $N_{V_\delta}^{(\Psi)}$  and  $N_{V_\delta}^{(\Psi)}$ , hence our interest in the number of star subgraphs. For  $2 \leq \delta \leq p-1$ , the number of subgraphs in  $\mathcal{G}_\rho(\Psi)$  that are isomorphic to  $\Gamma_\delta$  is denoted by  $N_{E_\delta}^{(\Psi)}$ , where  $\Gamma_\delta$  denotes a star shaped graph with  $\delta$  edges. In the case when  $\delta = 1$ , we define  $N_{E_1}^{(\Psi)}$  to be twice the number of edges in  $\mathcal{G}_\rho(\Psi)$ .  $N_{E_\delta}^{(\Psi)}$  is referred as the *star subgraph counts*.

*Example II.1:* Figure 1 represents an empirical partial correlation graph. For this graph the number of vertices of degree 2 is  $N_{V_2}^{(\mathbf{P})} = 1$  and the number of vertices of degree at least 2 is  $N_{V_2}^{(\mathbf{P})} = 3$ . The number of subgraphs isomorphic to  $\Gamma_3$  is  $N_{E_3}^{(\mathbf{P})} = 2$ . The number of connected vertices is  $N_{V_1}^{(\mathbf{P})} = 5$ , and  $N_{E_1}^{(\mathbf{P})} = 10$  as there are 5 edges.

Consider now the case where the sample size  $n$  is fixed and there exists a sequence of data matrices  $\mathbf{X} \in \mathbb{R}^{n \times p}$  with increasing dimension  $p$ . Following the procedure described in the paragraph after (5), we obtain a sequence of random graphs  $\mathcal{G}_\rho(\Psi)$  with increasing number of vertices. This paper derives finite sample compound Poisson characterizations of the distributions of the 6 random quantities  $\{N_{E_\delta}^{(\Psi)}, N_{V_\delta}^{(\Psi)}, N_{V_\delta}^{(\Psi)} : \Psi \in \{\mathbf{R}, \mathbf{P}\}\}$ , for  $p$  finite and as  $p \rightarrow \infty$ , for suitably chosen  $\rho$ , under a sparsity assumption on the dispersion parameter  $\Sigma$ . Such characterizations can be used to test the sparsity structure of the dispersion parameter  $\Sigma$  or to guide the choice of the threshold  $\rho$  [17], [18]. Throughout the rest of the paper we use  $\bar{N}_\delta$  to denote a generic random variable equal to one of the 6 quantities  $\{N_{E_\delta}^{(\Psi)}, N_{V_\delta}^{(\Psi)}, N_{V_\delta}^{(\Psi)} : \Psi \in \{\mathbf{R}, \mathbf{P}\}\}$ . By abuse of terminology, we refer to  $\bar{N}_\delta$  generically as vertex counts. We reiterate that in this paper the number  $n$  of samples is

fixed and the number  $p$  of variables could either be finite or tend to infinity.

### B. A Unified Theorem

In this subsection we present a unified theorem that establishes that  $\bar{N}_\delta$  converges in distribution to a compound Poisson distribution when  $p \rightarrow \infty$ . We begin by defining necessary quantities and then we state our main theorem.

For any positive number  $\lambda$  and a probability distribution  $\zeta$  supported on positive integers, let  $\text{CP}(\lambda, \zeta)$  denote the corresponding compound Poisson distribution. Specifically,  $\text{CP}(\lambda, \zeta)$  is the distribution of  $Z = \sum_{i=1}^N Z_i$ , where  $N$  is distributed as a Poisson random variable with mean  $\lambda$ ,  $Z_i \stackrel{\text{i.i.d.}}{\sim} \zeta$  and  $N$  is independent of each  $Z_i$ . Here the random variable  $N$  is the number of occurrences of increments and  $\zeta$  is the distribution of each increment. The parameter  $\lambda$  and  $\zeta$  are often referred to as the arrival rate and the increment distribution, respectively.

As the parameters of the compound Poisson distribution in the next theorem involve a random geometric graph, we define relevant notation. Given a set of points  $\{v_i\}_{i=1}^\delta$  in  $\mathbb{R}^{n-2}$ , denote by  $\mathbf{Ge}(\{v_i\}_{i=1}^\delta, r)$  the geometric graph with radius  $r$ , defined as follows. The vertex set of the graph is  $\{v_i\}_{i=1}^\delta$ , and there is an edge between  $v_i$  and  $v_j$  if  $\|v_i - v_j\|_2 \leq r$ . The graph is called a random geometric graph when the vertices of the geometric graph are random. A *universal vertex* is a vertex of an undirected graph that is adjacent to all other vertices of the graph [35]. Denote by  $\mathbf{NUV}(\{v_i\}_{i=1}^\delta, r)$  the number of universal vertices in  $\mathbf{Ge}(\{v_i\}_{i=1}^\delta, r)$ . Denote  $B^{n-2}$  the unit sphere in  $\mathbb{R}^{n-2}$  and denote  $\mathbf{unif}(B^{n-2})$  the uniform distribution on  $B^{n-2}$ . Let  $\{\tilde{u}_i\}_{i=1}^\delta$  be i.i.d. from  $\mathbf{unif}(B^{n-2})$ . For the random geometric graph  $\mathbf{Ge}(\{\tilde{u}_i\}_{i=1}^\delta, 1)$ , we denote the probability that there are exactly  $\ell - 1$  universal vertices by

$$\alpha_\ell := \mathbb{P}(\mathbf{NUV}(\{\tilde{u}_i\}_{i=1}^\delta, 1) = \ell - 1), \quad \forall \ell \in [\delta + 1], \quad (6)$$

and define a probability distribution  $\zeta_{n,\delta}$  on  $[\delta + 1]$ :

$$\zeta_{n,\delta}(\ell) := (\alpha_\ell / \ell) / \left( \sum_{s=1}^{\delta+1} (\alpha_s / s) \right), \quad \forall \ell \in [\delta + 1]. \quad (7)$$

As will be shown in the next theorem,  $\zeta_{n,\delta}$  is the increment distribution of a compound Poisson approximation to  $\bar{N}_\delta$  when  $p \rightarrow \infty$ .

We also introduce the following sparsity conditions: a matrix is said to be row- $\kappa$  sparse if every row has at most  $\kappa$  nonzero elements. This is a weaker sparsity condition than the block sparsity condition of [18] (see also Subsection II-C). The next definition is a stronger sparsity condition than row- $\kappa$  sparsity but remains weaker than block sparsity.

*Definition II.2 (( $\tau, \kappa$ ) Sparsity):* A  $p$  by  $p$  dimensional symmetric matrix is said to be  $(\tau, \kappa)$  sparse if it is row- $\kappa$  sparse and its lower  $p - \tau$  by  $p - \tau$  block is diagonal.

Another relevant quantity is the normalized determinant defined as follows:

*Definition II.3 (Normalized Determinant):* For any symmetric, positive definite matrix  $\mathbf{A} \in \mathbb{R}^{p \times p}$ , its normalized

determinant  $\mu(\mathbf{A})$  is defined by

$$\mu(\mathbf{A}) := \prod_{i=1}^p \frac{\lambda_i(\mathbf{A})}{\lambda_p(\mathbf{A})} = \frac{\det(\mathbf{A})}{(\lambda_p(\mathbf{A}))^p},$$

where  $\lambda_1(\mathbf{A}) \leq \lambda_2(\mathbf{A}) \leq \dots \leq \lambda_p(\mathbf{A})$  are the eigenvalues of  $\mathbf{A}$ .

For  $\mathcal{I} \subset [p]$  denote by  $\mathbf{A}_{\mathcal{I}}$  the set of all  $|\mathcal{I}| \times |\mathcal{I}|$  submatrices of  $\mathbf{A} \in \mathbb{R}^{p \times p}$  obtained by extracting the corresponding rows and columns indexed by  $\mathcal{I}$ . The set  $\mathbf{A}_{\mathcal{I}}$  contains  $|\mathcal{I}|!$  matrices that are all equivalent up to a permutation applied simultaneously to both rows and columns. Define the *local normalized determinant of degree  $m$*  of a matrix  $\mathbf{A} \in \mathbb{R}^{p \times p}$  to be  $\mu_m(\mathbf{A}) = \min\{\mu(\mathbf{A}_{\mathcal{I}}) : \mathcal{I} \subset [p], |\mathcal{I}| = m\}$ . Note that  $\mu(\mathbf{A}_{\mathcal{I}})$  is well defined since  $\mu(\cdot)$  is invariant to simultaneous application of a permutation to both rows and columns. For symmetric positive definite  $\mathbf{A} \in \mathbb{R}^{p \times p}$  further define the *inverse local normalized determinant*

$$\mu_{n,m}(\mathbf{A}) := \begin{cases} [\mu_m(\mathbf{A})]^{-\frac{n-1}{2}}, & \mathbf{A} \text{ not diagonal,} \\ 1, & \mathbf{A} \text{ diagonal.} \end{cases} \quad (8)$$

By definition  $\mu(\mathbf{A}) \in (0, 1]$  and  $\mu_{n,m}(\mathbf{A}) \in [1, \infty)$ .

Denote  $\Gamma(x)$  the gamma function and let  $a_n := \frac{\Gamma((n-1)/2)}{(n-2)\sqrt{\pi}\Gamma((n-2)/2)}$ . With the above definitions in place, we now state our main theorem: when  $p \rightarrow \infty$ , if the threshold  $\rho$  approaches 1 at a particular rate, then the sequence of vertex counts  $\bar{N}_{\delta}$  converges in distribution to a compound Poisson distribution.

**Theorem II.4 (Compound Poisson limit):** Let  $n \geq 4$  and  $\delta$  be fixed positive integers. Let  $\mathbf{X} \sim \mathcal{VE}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, g)$ . Assume that the threshold  $\rho$  is a function of  $p$  that satisfies  $a_n 2^{\frac{n}{2}} p^{1+\frac{1}{\delta}} (1 - \rho)^{\frac{n-2}{2}} \rightarrow e_{n,\delta}$  as  $p \rightarrow \infty$ , where  $e_{n,\delta}$  is some positive finite constant that possibly depends on  $n$  and  $\delta$ . Denote  $\lambda_{n,\delta}(e_{n,\delta}) = \frac{1}{\delta!} (e_{n,\delta})^{\delta} \sum_{\ell=1}^{\delta+1} \frac{\alpha_{\ell}}{\ell}$ . Suppose  $\boldsymbol{\Sigma}$ , after some row-column permutation, is  $(\tau_p, \kappa_p)$  sparse with  $\lim_{p \rightarrow \infty} \frac{\tau_p}{p} + \mu_{n,2\delta+2}(\boldsymbol{\Sigma}) \frac{\kappa_p}{p} \rightarrow 0$ . Then  $\bar{N}_{\delta}$ , a generic random variable in the set  $\{N_{E_{\delta}}^{(\Psi)}, N_{V_{\delta}}^{(\Psi)}, N_{V_{\delta}}^{(\Psi)} : \Psi \in \{\mathbf{R}, \mathbf{P}\}\}$ , satisfies:

$$\bar{N}_{\delta} \xrightarrow{\mathcal{D}} \text{CP}(\lambda_{n,\delta}(e_{n,\delta}), \boldsymbol{\zeta}_{n,\delta}) \text{ as } p \rightarrow \infty. \quad (9)$$

If only the vertex counts in the empirical correlation graph is of interest, then the  $(\tau, \kappa)$  sparsity assumption can be relaxed to row- $\kappa$  sparsity.

**Lemma II.5 (Compound Poisson Limit in Empirical Correlation Graph):** Let  $n \geq 4$  and  $\delta$  be fixed positive integers. Let  $\mathbf{X} \sim \mathcal{VE}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, g)$ . Assume that the threshold  $\rho$  is a function of  $p$  that satisfies  $a_n 2^{\frac{n}{2}} p^{1+\frac{1}{\delta}} (1 - \rho)^{\frac{n-2}{2}} \rightarrow e_{n,\delta}$  as  $p \rightarrow \infty$ , where  $e_{n,\delta}$  is some positive finite constant that possibly depends on  $n$  and  $\delta$ . Denote  $\lambda_{n,\delta}(e_{n,\delta}) = \frac{1}{\delta!} (e_{n,\delta})^{\delta} \sum_{\ell=1}^{\delta+1} \frac{\alpha_{\ell}}{\ell}$ . Suppose  $\boldsymbol{\Sigma}$  is row- $\kappa_p$  sparse with  $\lim_{p \rightarrow \infty} \mu_{n,2\delta+2}(\boldsymbol{\Sigma}) \frac{\kappa_p}{p} \rightarrow 0$ . Then  $\bar{N}_{\delta}$ , a generic random variable in the set  $\{N_{E_{\delta}}^{(\mathbf{R})}, N_{V_{\delta}}^{(\mathbf{R})}, N_{V_{\delta}}^{(\mathbf{R})}\}$ , satisfies:

$$\bar{N}_{\delta} \xrightarrow{\mathcal{D}} \text{CP}(\lambda_{n,\delta}(e_{n,\delta}), \boldsymbol{\zeta}_{n,\delta}) \text{ as } p \rightarrow \infty. \quad (10)$$

**Remark II.6:** The condition  $a_n 2^{\frac{n}{2}} p^{1+\frac{1}{\delta}} (1 - \rho)^{\frac{n-2}{2}} \rightarrow e_{n,\delta} > 0$  is equivalent to

$$p^{\frac{2}{n-2}(1+\frac{1}{\delta})} (1 - \rho) \rightarrow \left( \frac{e_{n,\delta}}{a_n 2^{\frac{n}{2}}} \right)^{\frac{2}{n-2}} = \frac{1}{2} \left( \frac{e_{n,\delta}}{2a_n} \right)^{\frac{2}{n-2}},$$

which indicates that  $\rho \rightarrow 1$  at rate  $p^{-\frac{2}{n-2}(1+\frac{1}{\delta})}$ . As will be discussed in more detail in Remark IV.2, this rate is in fact both necessary and sufficient for the expected counts  $\mathbb{E}\bar{N}_{\delta}$  to converge to a non-trivial limit. If  $\rho$  does not converge to 1, or converges to 1 at a slower rate, then  $\mathbb{E}\bar{N}_{\delta}$  diverges to  $\infty$ , while if  $\rho$  converges to 1 at a faster rate then  $\mathbb{E}\bar{N}_{\delta}$  converges to 0. This particular rate on  $\rho$  is consistent with the rate of the existing Poisson approximation results in random geometric graphs [23] as will be discussed in Section III-B. A sequence of correlation thresholds  $\rho = \rho_p$  that satisfies this condition is

$$\rho_p = 1 - \frac{1}{2} \left( \frac{e_{n,\delta}}{2a_n p^{1+\frac{1}{\delta}}} \right)^{\frac{2}{n-2}}. \quad (11)$$

Observe that Theorem II.4 and Lemma II.5 hold for any mean  $\boldsymbol{\mu}$  and any shaping function  $g$  when  $\mathbf{X} \sim \mathcal{VE}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, g)$ . We will provide intuition for this invariance property in Remark III.3.

The proofs of Theorem II.4 and Lemma II.5 will be presented in Subsection III-E. We will call  $\text{CP}(\lambda_{n,\delta}(e_{n,\delta}), \boldsymbol{\zeta}_{n,\delta})$  the *limiting compound Poisson distribution, approximation or characterization*. Since  $\bar{N}_{\delta}$  and  $\text{CP}(\lambda_{n,\delta}(e_{n,\delta}), \boldsymbol{\zeta}_{n,\delta})$  are discrete, (9) is equivalent to:

$$d_{\text{TV}}(\mathcal{L}(\bar{N}_{\delta}), \text{CP}(\lambda_{n,\delta}(e_{n,\delta}), \boldsymbol{\zeta}_{n,\delta})) \rightarrow 0 \text{ as } p \rightarrow \infty,$$

where  $\mathcal{L}(\cdot)$  represents the probability distribution of the argument, and  $d_{\text{TV}}(\cdot, \cdot)$  is the total variation distance between two probability distributions. A variant of Theorem II.4 for finite  $p$  (Theorem III.11), which establishes an upper bound on the total variation distance between  $\mathcal{L}(\bar{N}_{\delta})$  and a compound Poisson distribution, will be presented in Subsection III-E. Theorem II.4 and Theorem III.11 specify asymptotic compound Poisson limits and non-asymptotic bounds on the full distribution of vertex counts. These limits correct and extend the Poisson limits that were falsely claimed to hold for all finite  $n, \delta$ , although we note that the compound Poisson limit can be well approximated by the Poisson limit in the case of moderately large  $n$  or large  $\delta$  (see Sec. V).

In Theorem II.4 the  $(\tau, \kappa)$  sparsity condition and the condition  $\mu_{n,m}(\boldsymbol{\Sigma}) < \infty$  are assumed. We elaborate on these two conditions in the next two subsections.

### C. $(\tau, \kappa)$ Sparsity

The matrix (12) below is an example of a  $(\tau, \kappa)$  sparse matrix with  $\tau = 2, \kappa = 3$ . This  $5 \times 5$  symmetric matrix is  $(2, 3)$  sparse since each of the first 2 rows has at most 3 nonzero elements and the lower  $3 \times 3$  block is diagonal.

$$\begin{pmatrix} 5 & 0 & 2 & 0 & 1 \\ 0 & 8 & 3 & 0 & 0 \\ 2 & 3 & 6 & 0 & 0 \\ 0 & 0 & 0 & 7 & 0 \\ 1 & 0 & 0 & 0 & 8 \end{pmatrix} \quad (12)$$

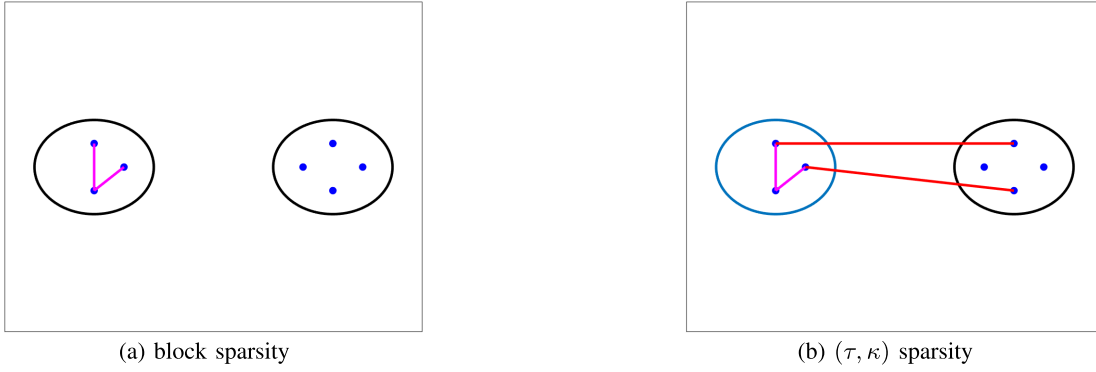


Fig. 2. Diagram of the correlation graph  $\mathcal{G}_0(\tilde{\Sigma})$  for a  $p = 7$  dimensional distribution with two different  $7 \times 7$  correlation matrices. The left panel is associated with a block-3 sparse assumption on  $\tilde{\Sigma}$ . Only the  $\tau = 3$  variables in the group inside the left circle are correlated: there are no correlations (edges) between the remaining 4 variables in the right circle and there are no correlations across the two sets of variables in the different circles. The right panel is associated with  $(\tau, \kappa) = (3, 3)$  sparsity on  $\tilde{\Sigma}$ , where two additional edges, representing correlations between variables, exist across the two groups.

If the adjacency matrix of a graph  $(\mathcal{V}, \mathcal{E})$  is  $(\tau, \kappa)$  sparse, then the vertices  $\mathcal{V}$  can be partitioned into two disjoint subsets  $\mathcal{V}_1$  and  $\mathcal{V}_2$  with the following properties: 1)  $|\mathcal{V}_1| \leq \tau$ ; 2) there is no edge between any two vertices in  $\mathcal{V}_2$ ; 3) the degree of any vertex in  $\mathcal{V}_1$  is no more than  $\kappa - 1$ ; 4) edges connecting vertex in  $\mathcal{V}_1$  and  $\mathcal{V}_2$  may exist.

When the dispersion parameter  $\Sigma$  is row- $\kappa$  sparse, the authors of [17] studied the mean of the quantities  $N_{E_1}^{(R)}$  and  $N_{V_1}^{(R)}$  and they obtained limits of  $\mathbb{P}(N_{E_1}^{(R)} > 0)$  or  $\mathbb{P}(N_{V_1}^{(R)} > 0)$  when  $p \rightarrow \infty$  while fixing  $n$ . In [18] these results were extended to empirical partial correlation graphs when the dispersion parameter  $\Sigma$  is assumed to be block- $\tau$  sparse up to a row-column permutation, i.e., there exists a permutation matrix  $T$  such that

$$T\Sigma T^\top = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & D_{p-\tau} \end{pmatrix} \quad (13)$$

where  $\Sigma_{12} = \Sigma_{21}^\top = \mathbf{0} \in \mathbb{R}^{\tau \times (p-\tau)}$  and  $D_{p-\tau} \in \mathbb{R}^{(p-\tau) \times (p-\tau)}$  is some diagonal matrix. In Theorem II.4 of this paper  $\Sigma$  is assumed to be  $(\tau, \kappa)$  sparse after row-column permutation, i.e. there exists a permutation matrix  $T$  such that (13) holds with  $D_{p-\tau} \in \mathbb{R}^{(p-\tau) \times (p-\tau)}$  some diagonal matrix and with the first  $\tau$  rows  $(\Sigma_{11} \ \Sigma_{12})$  being row- $\kappa$  sparse.

It is clear that the  $(\tau, \kappa)$  sparsity condition is more general than the block sparsity condition as that there is no restriction on  $\Sigma_{12} = 0$ . Indeed, every block- $\tau$  sparse matrix is  $(\tau, \kappa)$  sparse with  $\kappa = \tau$ . Nevertheless, the  $(\tau, \kappa)$  sparsity with  $\kappa = \tau$  allows non-zeros in the top-right submatrix, which permits more possible correlations between the variables relative to the block- $\tau$  sparsity in correlation graphical models. To see this, consider the associated graphical model  $\mathcal{G}_0(\tilde{\Sigma})$  for a correlation matrix  $\tilde{\Sigma}$ . Recall in Section II-A we define  $\mathcal{G}_\rho(\cdot)$  as the graph with adjacency matrix obtained by thresholding a matrix with  $\rho$ . In Figure 2, nodes represent the variables and edges represent the correlation between variables. The left panel is a graphical model associated with the block-3 sparse assumption, while the right panel satisfies  $(\tau, \kappa)$  sparsity with  $(\tau, \kappa) = (3, 3)$ . The later has more correlations (the red edges) across the two sets of variables in the 2 circles. The  $(\tau, \kappa)$  sparsity condition with  $\kappa > \tau$  allows additional

correlations between variables. Notably,  $(\tau, \kappa)$  sparsity allows the underlying graphical model to be connected as shown in the following example.

*Example II.7 (A Connected Correlation Graph That is  $(\tau, \kappa)$  Sparse):* Start with a set of  $p$  vertices  $[p]$  with no edges. Denote the neighborhood of vertex  $i$  to be  $\text{NB}(i)$ . We construct a connected  $(\tau, \kappa)$  sparse graph on  $[p]$  by creating neighborhoods of the first  $\tau$  vertices according to the following rule: for  $i \in [\tau - 1]$ ,

$$\text{NB}(i) = \{i + 1\} \cup \bigcup_{j \in [\kappa - 2]} \{\tau + (i - 1)(\kappa - 2) + j\},$$

$$\text{NB}(\tau) = \bigcup_{j \in [\kappa - 1]} \{\tau + (\tau - 1)(\kappa - 2) + j\}.$$

With this construction, each vertex among the first  $\tau$  vertices is connected to  $\kappa - 1$  other vertices and the remaining  $p - \tau$  vertices  $[p] \setminus [\tau]$  do not connect to each other. As a result the associated adjacency matrix is  $(\tau, \kappa)$  sparse. Moreover, as long as

$$\tau + \tau(\kappa - 2) + 1 = \tau(\kappa - 1) + 1 \geq p, \quad (14)$$

the graph is connected. Now, let  $\Sigma$  be a matrix with diagonal 1 and the locations of the off-diagonal nonzero elements specified by the graph constructed. For simplicity, choose all the off-diagonal nonzero elements to be  $\xi$ . If  $\xi$  is small enough, then  $\Sigma$  is close to the identity matrix, hence positive definite and is therefore a *bone fide* correlation matrix. We have thus constructed a  $(\tau, \kappa)$  sparse correlation matrix whose associated graphical model is connected provided (14) is true. Now observe that by specifying  $\tau = p^\alpha$  and  $\kappa = p^\beta$  with  $\alpha + \beta > 1$ , (14) is indeed satisfied for large  $p$ . Moreover, if  $\alpha, \beta \in (0, 1)$  then  $\tau, \kappa = o(p)$ . By choosing  $\xi$  as a decreasing function of  $p$ , the condition number of  $\Sigma$  will not increase too fast w.r.t.  $p$  so that the condition that  $\lim_{p \rightarrow \infty} \frac{\tau}{p} + \mu_{n, 2\delta+2}(\Sigma) \frac{\kappa}{p} \rightarrow 0$  in Theorem II.4 is satisfied.

On the other hand,  $(\tau, \kappa)$  sparsity is a stronger assumption than row- $\kappa$  sparsity, since every  $(\tau, \kappa)$  sparse matrix is row- $\kappa$  sparse.  $(\tau, \kappa)$  sparsity is thus an intermediate level of sparsity lying between block sparsity and row sparsity.

*Remark II.8:* In Theorem II.4 we supposed that the dispersion parameter  $\Sigma$ , after some row-column permutation, is  $(\tau, \kappa)$  sparse. In this paper we are interested in  $\bar{N}_\delta$ , which is invariant under permutation of the  $p$  variables. Since permutation of the variables is equivalent to row-column permutation of the dispersion parameter  $\Sigma$ , without loss of generality, we can assume that the variables have been permuted such that  $\Sigma$  is  $(\tau, \kappa)$  sparse.

#### D. Local Normalized Determinant

In this subsection, we elaborate on the normalized determinant  $\mu(\cdot)$  and the inverse local normalized determinant defined in Subsection II-B. The normalized determinant measures the closeness between the identity matrix and the dispersion parameter  $\Sigma$ . This quantity plays an important role in Theorem II.4.

Observe that  $\mu(\mathbf{A}) \in (0, 1]$  and that,  $\mu(\mathbf{A}) = 1$  if and only if  $\mathbf{A}$  is a multiple of  $\mathbf{I}_p$ . Moreover  $\mu(\mathbf{A})$  is close to 1, and hence bounded away from 0, as long as all eigenvalues concentrate around a positive number. Below is an example of a sequence of symmetric positive definite matrices with well-concentrated eigenvalues such that their normalized determinants are uniformly bounded away from 0.

*Example II.9:* Let  $\{\alpha_i\}_{i=1}^p$  be a positive real sequence. Let  $\{\beta_i\}_{i=1}^\infty$  be a positive, decreasing sequence such that  $\sum_{i=1}^\infty \beta_i < \infty$ . Consider  $\mathbf{A} \in \mathbb{R}^{p \times p}$  a symmetric positive definite matrix with eigenvalues  $\lambda_i = \alpha_p \exp(-\beta_i)$  for  $1 \leq i \leq p-1$  and  $\lambda_p = \alpha_p$ . Then

$$\mu(\mathbf{A}) = \exp\left(-\sum_{i=1}^{p-1} \beta_i\right).$$

Consider now the case that  $p$  is increasing and consider a sequence of matrices  $\mathbf{A}$  of increasing dimension with the above properties. For this sequence  $\mu(\mathbf{A}) \geq \exp\left(-\sum_{i=1}^\infty \beta_i\right) > 0$ , i.e.  $\mu(\mathbf{A})$  is bounded uniformly away from 0.

It follows by the interlacing property (see Theorem 8.1.7 in [36]) that  $\mu_m(\mathbf{A})$  is decreasing with respect to  $m \in [p]$  for any symmetric positive definite matrix  $\mathbf{A} \in \mathbb{R}^{p \times p}$ . Thus the inverse local normalized determinant  $\mu_{n,m}(\mathbf{A})$  is increasing with respect to  $m \in [p]$ .

It turns out that the inverse local normalized determinant of the dispersion matrix  $\Sigma$  will play an important role in our study of the distribution of  $\bar{N}_\delta$ . Indeed, when  $\delta \geq 2$ ,  $N_{E_\delta}^{(\mathbf{R})}$  can be represented as a sum of indicator functions of the event that a subgraph of  $\delta + 1$  vertices is isomorphic to  $\Gamma_\delta$  (see (48) for the precise formula). Each term in the summation involves only  $\delta + 1$  variables, and thus each pair of two such terms involves at most  $2(\delta + 1)$  variables. Hence  $\mu_{2\delta+2}(\Sigma)$  or  $\mu_{n,2\delta+2}(\Sigma)$  controls the correlation between two indicator terms in the summation of  $N_{E_\delta}^{(\mathbf{R})}$ , which determines the convergence of  $N_{E_\delta}^{(\mathbf{R})}$  to a compound Poisson distribution.

In Theorem II.4 we assume that  $\mu_{n,2\delta+2}(\Sigma) \frac{\kappa_p}{p} \rightarrow 0$ , which holds when  $\mu_{n,2\delta+2}(\Sigma)$  is either bounded or increasing with

rate  $o(\frac{p}{\kappa_p})$ . In the Section B of the Appendix, Lemma A.1 provides a bound on  $\mu_{n,2\delta+2}(\Sigma)$  in terms of the condition number and the eigenvalues of  $\Sigma$ .

### III. COMPOUND POISSON CHARACTERIZATIONS FOR FINITE $p$

In this section we establish a compound Poisson approximation for  $\mathcal{L}(\bar{N}_\delta)$  for finite  $p$ , which is then used to prove the asymptotic Theorem II.4. In Subsection III-A Z-score type representations are introduced, which are used in the subsequent development in Subsection III-B. Subsection III-B provides an equivalent formulation of the empirical correlation and partial correlation graphs in terms of random geometric graphs. Against the backdrop of the first two subsections, Subsection III-C presents the compound Poisson approximation for star subgraph counts  $N_{E_\delta}^{(\mathbf{R})}$ . Subsection III-D demonstrates that all 6 quantities in  $\{N_{E_\delta}^{(\Psi)}, N_{\check{V}_\delta}^{(\Psi)}, N_{V_\delta}^{(\Psi)} : k \in \{\mathbf{R}, \mathbf{P}\}\}$  are close in  $L^1$  distance. Combining results in Subsection III-C and Subsection III-D, a compound Poisson characterization for  $\bar{N}_\delta$  for finite  $p$  is obtained in Subsection III-E, which is then used to deduce Theorem II.4 in Subsection III-F.

#### A. Z-Score Type Representations of Sample Correlation and Partial Correlation

In this subsection, we define the U-scores and the Y-scores for the sample correlation and the sample partial correlation, respectively. These Z-type scores will serve as the vertex set on which random geometric graphs are constructed in Subsection III-B.

The matrix of Z-scores  $\mathbf{Z} = [\mathbf{z}_1, \dots, \mathbf{z}_p] \in \mathbb{R}^{n \times p}$  associated with the data matrix  $\mathbf{X}$  is defined by

$$\mathbf{z}_i := \frac{\mathbf{x}_i - \mathbf{1}\bar{x}_i}{\sqrt{S_{ii}(n-1)}},$$

where  $\bar{x}_i$  is the  $i$ -th coordinate of the sample mean  $\bar{\mathbf{x}}$ . The Z-scores specify the sample correlation matrix by the relation  $\mathbf{R} = \mathbf{Z}^\top \mathbf{Z}$ . Since  $\mathbf{1}^\top \mathbf{z}_i = 0$  for every  $i \in [p]$ , we consider an equivalent but lower-dimensional type of Z-scores called the U-scores [17], [18], obtained by projecting each  $\mathbf{z}_i$  on the subspace  $\{\mathbf{w} \in \mathbb{R}^n : \mathbf{1}^\top \mathbf{w} = 0\}$ . Specifically, consider an orthogonal  $n \times n$  matrix  $\mathbf{H} = [n^{-\frac{1}{2}}\mathbf{1}, \mathbf{H}_{2:n}]^\top$ . The matrix  $\mathbf{H}_{2:n} \in \mathbb{R}^{n \times (n-1)}$  can be obtained by Gramm-Schmidt orthogonalization and satisfies the properties

$$\mathbf{1}^\top \mathbf{H}_{2:n} := \mathbf{0}, \quad \mathbf{H}_{2:n}^\top \mathbf{H}_{2:n} = \mathbf{I}_{n-1}.$$

Define the matrix of U-scores as  $\mathbf{U} = [\mathbf{u}_1, \dots, \mathbf{u}_p]$  with  $\mathbf{u}_i = \mathbf{H}_{2:n}^\top \mathbf{z}_i \in \mathbb{R}^{n-1}$ . It is clear that

$$\mathbf{R} = \mathbf{U}^\top \mathbf{U}. \quad (15)$$

The normalized outer product of  $\mathbf{U}$ , defined by

$$\mathbf{B} := \frac{n-1}{p} \mathbf{U} \mathbf{U}^\top \in \mathbb{R}^{(n-1) \times (n-1)} \quad (16)$$

will play an important role in the analysis of the empirical partial correlation graph.

*Lemma III.1:* Let  $\mathbf{X} \sim \mathcal{VE}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, g)$  and  $p \geq n$ . Then  $\mathbf{B}$  is invertible a.s.

Proofs of all the lemmas and propositions in this paper can be found in the Appendix. By Lemma 1 in [18], provided  $\mathbf{B}$  is invertible, the Moore-Penrose pseudo-inverse of  $\mathbf{R}$  can be expressed as follows:

$$\mathbf{R}^\dagger = \mathbf{U}^\top [\mathbf{U}\mathbf{U}^\top]^{-2} \mathbf{U} = \left( \frac{p}{n-1} \right)^2 \mathbf{U}^\top \mathbf{B}^{-2} \mathbf{U}. \quad (17)$$

It follows from Lemma III.1 that equation (17) holds a.s. Define  $\mathbf{Y} = \mathbf{B}^{-1} \mathbf{U}$  and observe that

$$\mathbf{R}^\dagger = \left( \frac{p}{n-1} \right)^2 \mathbf{Y}^\top \mathbf{Y} \quad \text{a.s.} \quad (18)$$

Further define  $\mathbf{y}_i := \bar{\mathbf{y}}_i / \|\bar{\mathbf{y}}_i\|_2$  for  $i \in [p]$  and  $\mathbf{Y} := [\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_p] \in \mathbb{R}^{(n-1) \times p}$ . Then

$$\mathbf{P} = \mathbf{Y}^\top \mathbf{Y} \quad \text{a.s.} \quad (19)$$

by equation (5) and (18).  $\mathbf{Y}$  in (19) is referred to as the Y-scores representation of the sample partial correlation matrix, as in [18].

By (15) and (19), the set of U-scores and Y-scores summarize all the information about the sample correlations and partial correlations, and hence they capture all the information about the vertex counts  $\bar{N}_\delta$ . We will further elaborate on this point in Subsection III-B. Observing that  $\mathbf{Y}$  is a function of  $\mathbf{U}$ , the distribution of  $\bar{N}_\delta$  is uniquely determined by  $\mathbf{U}$ . The remainder of this subsection focuses on the distribution of the U-scores.

Denote  $S^{n-2}$  the unit sphere in  $\mathbb{R}^{n-1}$  and denote  $\text{unif}(S^{n-2})$  the uniform distribution on  $S^{n-2}$ . Let  $\{\tilde{\mathbf{x}}^{(i)}\}_{i=1}^{n-1} \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$  and

$$\tilde{\mathbf{X}} = [\tilde{\mathbf{x}}^{(1)}, \dots, \tilde{\mathbf{x}}^{(n-1)}]^\top = [\tilde{\mathbf{x}}_1, \dots, \tilde{\mathbf{x}}_p] \in \mathbb{R}^{(n-1) \times p}.$$

In the next lemma we characterize the distribution of  $\mathbf{U}$  in terms of the distribution of  $\tilde{\mathbf{X}}$ .

*Lemma III.2:* Assume  $\mathbf{X} \sim \mathcal{VE}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, g)$ .

(a) The matrix  $\mathbf{U}$  of U-scores has the same distribution as  $\left[ \frac{\tilde{\mathbf{x}}_1}{\|\tilde{\mathbf{x}}_1\|_2}, \dots, \frac{\tilde{\mathbf{x}}_p}{\|\tilde{\mathbf{x}}_p\|_2} \right]$ .

(b) For each  $i \in [p]$ ,  $\mathbf{u}_i$  is distributed as  $\text{unif}(S^{n-2})$ . Moreover, if  $\Sigma_{ij} = 0$ , then  $\mathbf{u}_i$  and  $\mathbf{u}_j$  are independent.

*Remark III.3:* When  $\Sigma_{ij} = 0$ , the  $i$ -th and the  $j$ -th columns of the data matrix are uncorrelated but are not necessarily independent. However Lemma III.2 (b) establishes that in fact the corresponding U-scores  $\mathbf{u}_i$  and  $\mathbf{u}_j$  are independent. This is a consequence of the fact that the entire data matrix follows the vector elliptically contoured distribution  $\mathbf{X} \sim \mathcal{VE}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, g)$ . Another important implication of Lemma III.2 is that the distribution of  $\mathbf{U}$ , hence the distribution of  $\bar{N}_\delta$ , is invariant to the mean  $\boldsymbol{\mu}$  and the shaping function  $g$ .

Let  $\sigma^{n-2}$  be the spherical measure on  $S^{n-2}$ , that is  $\sigma^{n-2}$  is the uniform probability measure corresponding to  $\text{unif}(S^{n-2})$ . Denote by  $f_{\mathbf{u}_{j_1}, \mathbf{u}_{j_2}, \dots, \mathbf{u}_{j_m}}$  the joint density of the  $j_1$ -th,  $j_2$ -th,  $\dots$ ,  $j_m$ -th column of  $\mathbf{U}$  with respect to the product measure  $\otimes^m \sigma^{n-2} := \underbrace{\sigma^{n-2} \otimes \sigma^{n-2} \otimes \dots \otimes \sigma^{n-2}}_m$ .

The next lemma establishes that  $f_{\mathbf{u}_{j_1}, \mathbf{u}_{j_2}, \dots, \mathbf{u}_{j_m}}$  is bounded by the inverse local normalized determinant  $\mu_{n,m}(\boldsymbol{\Sigma})$ .

*Lemma III.4:* Assume  $\mathbf{X} \sim \mathcal{VE}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, g)$ . Consider a set of distinct indexes  $\mathcal{J} = \{j_i : 1 \leq i \leq m\} \subset [p]$ .

(a) The inverse local normalized determinant satisfies

$$\mu_{n,m}(\boldsymbol{\Sigma}_{\mathcal{J}}) \leq \mu_{n,m}(\boldsymbol{\Sigma}).$$

(b) The joint density of any subset of  $m$  columns of U-scores w.r.t.  $\otimes^m \sigma^{n-2}$  is upper bounded by  $\mu_{n,m}(\boldsymbol{\Sigma})$ :  $\forall \mathbf{v}_i \in S^{n-2}, \forall i \in [m]$ ,

$$f_{\mathbf{u}_{j_1}, \mathbf{u}_{j_2}, \dots, \mathbf{u}_{j_m}}(\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_m) \leq \mu_{n,m}(\boldsymbol{\Sigma}_{\mathcal{J}}) \leq \mu_{n,m}(\boldsymbol{\Sigma}).$$

(c) Let  $h : (S^{n-2})^m \rightarrow \mathbb{R}$  be a nonnegative Borel measurable function. Then

$$\begin{aligned} & \mathbb{E}h(\mathbf{u}_{j_1}, \mathbf{u}_{j_2}, \dots, \mathbf{u}_{j_m}) \\ & \leq \mu_{n,m}(\boldsymbol{\Sigma}_{\mathcal{J}}) \mathbb{E}h(\mathbf{u}'_{j_1}, \mathbf{u}'_{j_2}, \dots, \mathbf{u}'_{j_m}) \\ & \leq \mu_{n,m}(\boldsymbol{\Sigma}) \mathbb{E}h(\mathbf{u}'_{j_1}, \mathbf{u}'_{j_2}, \dots, \mathbf{u}'_{j_m}), \end{aligned}$$

where  $\{\mathbf{u}'_{j_\ell}\}_{\ell=1}^m$  are i.i.d. uniformly distributed on  $S^{n-2}$ .

According to Lemma III.4 (c), when calculating the expectation of nonnegative function of any  $m$  columns of  $\mathbf{U}$ , one may always assume that the associated columns  $\{\mathbf{u}_j\}$  are i.i.d.  $\text{unif}(S^{n-2})$  up to an additional multiplicative factor  $\mu_{n,m}(\boldsymbol{\Sigma})$ .

### B. Random Pseudo Geometric Graphs

In this subsection we define random pseudo geometric graphs and provide an equivalent formulation of the empirical correlation and partial correlation graphs in terms of these graphs, for which the vertex sets are, respectively, the U and Y scores. We also define the increment distribution of the compound Poisson that approximates  $\bar{N}_\delta$  when  $p$  is finite.

By the fact that  $\mathbf{R} = \mathbf{U}^\top \mathbf{U}$  and the fact that columns of  $\mathbf{U}$  have Euclidean norm 1,

$$R_{ij} = \mathbf{u}_i^\top \mathbf{u}_j = 1 - \frac{\|\mathbf{u}_i - \mathbf{u}_j\|_2^2}{2} = \frac{\|\mathbf{u}_i + \mathbf{u}_j\|_2^2}{2} - 1. \quad (20)$$

For a threshold  $\rho \in [0, 1)$ , define  $r_\rho := \sqrt{2(1-\rho)} \in (0, \sqrt{2}]$ . As shown in [17] and [18], by (20),

$$\{|R_{ij}| \geq \rho\} = \{\|\mathbf{u}_i + \mathbf{u}_j\|_2 \leq r_\rho\} \cup \{\|\mathbf{u}_i - \mathbf{u}_j\|_2 \leq r_\rho\}. \quad (21)$$

An analogous argument yields the following for the empirical partial correlation graph,

$$\{|P_{ij}| \geq \rho\} = \{\|\mathbf{y}_i + \mathbf{y}_j\|_2 \leq r_\rho\} \cup \{\|\mathbf{y}_i - \mathbf{y}_j\|_2 \leq r_\rho\}. \quad (22)$$

Based on (21), we now introduce novel geometric connections between empirical correlation graphs and random geometric graphs. Recall  $\{|R_{ij}| \geq \rho\}$  is the event that the magnitude of the sample correlation between the  $i$ -th and  $j$ -th variables exceeds the threshold  $\rho$ , or equivalently, the event that there exists a edge connecting the  $i$ -th and  $j$ -th vertices in the empirical correlation graph  $\mathcal{G}_\rho(\mathbf{R})$ . Equation (21) indicates that  $\{|R_{ij}| \geq \rho\}$  is the same as the event that the associated U-scores for the  $i$ -th and  $j$ -th variables,  $\mathbf{u}_i$  and  $\mathbf{u}_j$ , lie in some geometric set on  $S^{n-2} \times S^{n-2}$ . This insight provides an equivalent way to construct  $\mathcal{G}_\rho(\mathbf{R})$  through the U-scores. The empirical partial correlation graph  $\mathcal{G}_\rho(\mathbf{P})$  may similarly be

constructed through the Y-scores based on (22). These remarks are formalized in the next few paragraphs.

**Definition III.5 (Pseudo Geometric Graph):** Given  $m \geq 2$  and a set of points  $\{\mathbf{v}_i\}_{i=1}^m$  in  $\mathbb{R}^{\mathcal{N}}$ , denote by  $\mathbf{PGe}(\{\mathbf{v}_i\}_{i=1}^m, r; \mathcal{N})$  the pseudo geometric graph with radius  $r$ , defined as follows. The vertex set of the graph is  $\{\mathbf{v}_i\}_{i=1}^m$ , and there is an edge between  $\mathbf{v}_i$  and  $\mathbf{v}_j$  if  $\mathbf{dist}(\mathbf{v}_i, \mathbf{v}_j) := \min\{\|\mathbf{v}_i - \mathbf{v}_j\|_2, \|\mathbf{v}_i + \mathbf{v}_j\|_2\} \leq r$ . The graph is called a random pseudo geometric graph when the vertices are random.

It is easy to verify that  $\mathbf{dist}(\cdot, \cdot)$  has the following properties: for  $\forall \mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3 \in \mathbb{R}^m$ ,

- 1)  $\mathbf{dist}(\mathbf{v}_1, \mathbf{v}_2) \geq 0$ ;
- 2)  $\mathbf{dist}(\mathbf{v}_1, \mathbf{v}_2) = 0$  if only if  $\mathbf{v}_1 = \mathbf{v}_2$  or  $\mathbf{v}_1 = -\mathbf{v}_2$ ;
- 3)  $\mathbf{dist}(\mathbf{v}_1, \mathbf{v}_2) = \mathbf{dist}(\mathbf{v}_2, \mathbf{v}_1)$  and  $\mathbf{dist}(\mathbf{v}_1, \mathbf{v}_2) = \mathbf{dist}(\mathbf{v}_1, -\mathbf{v}_2)$
- 4)  $\mathbf{dist}(\mathbf{v}_1, \mathbf{v}_2) \leq \mathbf{dist}(\mathbf{v}_1, \mathbf{v}_3) + \mathbf{dist}(\mathbf{v}_3, \mathbf{v}_2)$ .

That is,  $\mathbf{dist}(\cdot, \cdot)$  is a pseudo metric on  $\mathbb{R}^{\mathcal{N}}$ , hence the name pseudo geometric graph in Definition III.5.

With the above definitions, and by the discussions preceding Definition III.5, the empirical correlation graph  $\mathcal{G}_\rho(\mathbf{R})$  is isomorphic to  $\mathbf{PGe}(\{\mathbf{u}_i\}_{i=1}^p, r_\rho)$ , the random pseudo geometric graph generated by U-scores. Consequently  $\{N_{V_\delta}^{(\mathbf{R})}, N_{V_\delta}^{(\mathbf{R})}, N_{E_\delta}^{(\mathbf{R})}\}$  are the numbers of vertices or subgraphs in  $\mathbf{PGe}(\{\mathbf{u}_i\}_{i=1}^p, r_\rho)$ . An analogous analysis applies to the empirical partial correlation graph and  $\mathbf{PGe}(\{\mathbf{y}_i\}_{i=1}^p, r_\rho)$ . This equivalent construction indicates that the distribution of each of the 3 quantities  $\{N_{E_\delta}^{(\Psi)}, N_{V_\delta}^{(\Psi)}, N_{V_\delta}^{(\Psi)}\}$  with  $\Psi = \mathbf{R}$  ( $\Psi = \mathbf{P}$ ) only depends on the pairwise pseudo distances  $\mathbf{dist}(\cdot, \cdot)$  between columns of  $\mathbf{U}$  ( $\mathbf{Y}$ ).

The random pseudo geometric graph in Definition III.5 is similar to the random geometric graph introduced in [23]. In particular, studied in the monograph [23] is the number of *induced* subgraphs isomorphic to a given graph, typical vertex degrees, and other graphical quantities of random geometric graphs. As a specific example, the rate  $a_n 2^{n/2} p^{1+\frac{1}{\delta}} (1 - \rho)^{\frac{n-2}{2}} \rightarrow e(n, \delta)$  as  $p \rightarrow \infty$  in Theorem II.4 is equivalent to  $2a_n p^{1+\frac{1}{\delta}} r_\rho^{n-2} \rightarrow e(n, \delta)$ , which is consistent with existing Poisson approximation for random geometric graph [23, Theorem 3.4]. The differences between our random pseudo geometric graph in Definition III.5 and the random geometric graphs defined in [23] are: 1) our graph has vertices  $\mathbf{u}_i$  lying on the unit sphere instead of on the entire Euclidean space; 2) our graph is induced by distance  $\mathbf{dist}(\cdot, \cdot)$  instead of the Euclidean distance. Another key difference is that vertices  $\mathbf{u}_i$  are not necessarily independent in Definition III.5. Indeed in our model, the correlations between vertices  $\mathbf{u}_i$  are encoded by a sparse matrix  $\Sigma$  (cf. Lemma III.2 (a)), whereas in [23] the vertices associated with the random geometric graph are assumed to be i.i.d. In [23] it was stated (Example after [23, Corollary 3.6]) without proof that the number of vertices with degree at least 3 was approximately compound Poisson. A similar compound Poisson limit is established in Theorem II.4. There is recent work on testing whether a given graph is a realization of an Erdős-Rényi random graph or a realization of a random geometric graph with vertices i.i.d. uniformly distributed on the sphere [37]. Therein, the authors

study the asymptotics of a count statistic, the (signed) triangles, on a random geometric graph induced by thresholding the distances between  $p$  i.i.d. uniformly distributed points on the  $n$  dimensional sphere as  $n$  goes to  $\infty$ . In contrast, this paper studies the large  $p$  asymptotics of vertex degree counts for dependent non-uniform points on a sphere of fixed dimension  $n$ .

Recall that  $\mathbf{NUV}(\{\mathbf{v}_i\}_{i=1}^m, r; \mathcal{N})$  denotes the number of universal vertices (vertices adjacent to all vertices except itself) in the geometric graph  $\mathbf{Ge}(\{\mathbf{v}_i\}_{i=1}^m, r; \mathcal{N})$ . Analogously, denote by  $\mathbf{PNUV}(\{\mathbf{v}_i\}_{i=1}^m, r; \mathcal{N})$  the number of universal vertices in the pseudo geometric graph  $\mathbf{PGe}(\{\mathbf{v}_i\}_{i=1}^m, r; \mathcal{N})$ .

Denote by  $\deg(\cdot)$  the degree of a given vertex in the graph. Consider  $\{\mathbf{u}'_i\}_{i=1}^{\delta+1} \stackrel{\text{i.i.d.}}{\sim} \mathbf{unif}(S^{n-2})$ . Denote the conditional probability that there are  $\ell$  universal vertices in  $\mathbf{PGe}(\{\mathbf{u}'_i\}_{i=1}^{\delta+1}, r_\rho)$  by

$$\alpha_{n,\delta}(\ell, r_\rho) := \mathbb{P}(\mathbf{PNUV}(\{\mathbf{u}'_i\}_{i=1}^{\delta+1}, r_\rho) = \ell \mid \deg(\mathbf{u}'_1) = \delta), \quad (23)$$

where  $\ell \in [\delta + 1]$ . The conditional probability  $\alpha_{n,\delta}(\ell, r_\rho)$  depends on  $n$ ,  $\delta$  and the threshold  $\rho$  and is abbreviated as  $\alpha(\ell, r_\rho)$  when there is no risk of confusion. Define a probability distribution  $\zeta_{n,\delta,\rho}$  supported on  $[\delta + 1]$  with

$$\zeta_{n,\delta,\rho}(\ell) := \frac{\alpha(\ell, r_\rho)/\ell}{\sum_{\ell=1}^{\delta+1} (\alpha(\ell, r_\rho)/\ell)}. \quad (24)$$

It will be shown in the next three subsections that the probability distribution  $\zeta_{n,\delta,\rho}$  is the increment distribution of the compound Poisson approximation to  $\bar{N}_\delta$  for finite  $p$ .

### C. Closeness of the Star Subgraph Counts to a Compound Poisson

In this subsection we present a proposition that establishes an upper bound on the total variation between  $\mathcal{L}(N_{E_\delta}^{(\mathbf{R})})$  and the compound Poisson distribution.

We first introduce some notation. Let  $\text{SC}(r, \mathbf{w})$  be the spherical cap with radius  $r$  at the center  $\mathbf{w} \in S^{n-2}$ . Formally,

$$\text{SC}(r, \mathbf{w}) = \{\mathbf{x} \in S^{n-2} : \|\mathbf{x} - \mathbf{w}\|_2 \leq r\}. \quad (25)$$

Define  $P_n(r) := \frac{\text{Area}(\text{SC}(r, \mathbf{w}))}{\text{Area}(S^{n-2})}$ , where  $\text{Area}(\cdot)$  is the area of a subset of  $S^{n-2}$ .  $P_n(r)$  is the normalized area of the spherical cap with radius  $r$ . As is shown in (2.6) in [17],

$$P_n(r) = \frac{b_n}{2} \int_{1-\frac{r^2}{2}}^1 (1-u^2)^{\frac{n-4}{2}} du, \quad \text{when } r \in [0, \sqrt{2}], \quad (26)$$

where  $b_n = \frac{2\Gamma((n-1)/2)}{\sqrt{\pi}\Gamma((n-2)/2)}$ . It follows by simple calculation that when  $\sqrt{2} < r \leq 2$

$$P_n(r) = 1 - \frac{\text{Area}(\text{SC}(\sqrt{4-r^2}, \mathbf{w}))}{\text{Area}(S^{n-2})} = 1 - P_n(\sqrt{4-r^2}),$$

and  $P_n(r) = 1$  when  $r > 2$ . Further properties of  $P_n(r)$  are summarized in Lemma A.22 in Section J of the Appendix. Define the rate parameter

$$\lambda_{p,n,\delta,\rho} := \binom{p}{1} \binom{p-1}{\delta} (2P_n(r_\rho))^\delta \sum_{\ell=1}^{\delta+1} \frac{\alpha(\ell, r_\rho)}{\ell}. \quad (27)$$

The next proposition establishes that the distribution of  $N_{E_\delta}^{(\mathbf{R})}$  is approximated by a compound Poisson distribution  $\text{CP}(\lambda_{p,n,\delta,\rho}, \zeta_{n,\delta,\rho})$  with rate parameter  $\lambda_{p,n,\delta,\rho}$  and dispersion parameter  $\zeta_{n,\delta,\rho}$ . Here  $C$  and  $c$ , and their subscripted versions, denote generic positive constants that only depend on their subscripts.

**Proposition III.6 (Compound Poisson Approximation for Subgraph Counts):** Let  $p \geq n \geq 4$ ,  $\delta \in [p-1]$  and  $\gamma > 0$  be given. Suppose  $\mathbf{X} \sim \mathcal{VE}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, g)$ . Suppose  $2p^{1+\frac{1}{\delta}}P_n(r_\rho) \leq \gamma$ , and  $\boldsymbol{\Sigma}$  is row- $\kappa$  sparse. Then

$$\begin{aligned} & d_{\text{TV}} \left( \mathcal{L} \left( N_{E_\delta}^{(\mathbf{R})} \right), \text{CP}(\lambda_{p,n,\delta,\rho}, \zeta_{n,\delta,\rho}) \right) \\ & \leq C_{n,\delta,\gamma} (C'_{\delta,\gamma})^{\mu_{n,\delta+1}(\boldsymbol{\Sigma}) \frac{\kappa-1}{p}} \\ & \quad \left( \mu_{n,2\delta+2}(\boldsymbol{\Sigma}) \frac{\kappa}{p} \left( 1 + \mu_{n,2\delta+2}(\boldsymbol{\Sigma}) \left( \frac{\kappa}{p} \right)^2 \right) + p^{-\frac{1}{\delta}} \right). \end{aligned}$$

where  $C_{n,\delta,\gamma}$  and  $C'_{\delta,\gamma}$  are two constants.

**Remark III.7:** The condition  $2p^{1+\frac{1}{\delta}}P_n(r_\rho) \leq \gamma$  specifies an implicit lower bound on the threshold  $\rho$ . To obtain an explicit lower bound, observe that the condition  $2a_n p^{1+\frac{1}{\delta}} \left( \sqrt{2(1-\rho)} \right)^{n-2} \leq \gamma$  is sufficient for  $2p^{1+\frac{1}{\delta}}P_n(r_\rho) \leq \gamma$ , by Lemma A.22 (a) in Section J of the Appendix. Solving for  $\rho$ , we then obtain

$$\rho \geq 1 - \frac{1}{2} \left( \frac{\gamma}{2a_n p^{1+\frac{1}{\delta}}} \right)^{\frac{2}{n-2}}. \quad (28)$$

The condition (28) is a non-asymptotic version of (11). But the requirement on  $\rho$  from  $2p^{1+\frac{1}{\delta}}P_n(r_\rho) \leq \gamma$  is weaker than (28) since the bound in Lemma A.22 (a) is not tight when  $p$  is finite. For the upper bound in Proposition III.6 to be small,  $p$  should be relatively large and,  $\boldsymbol{\Sigma}$  should have relatively small  $\mu_{n,2\delta+2}(\boldsymbol{\Sigma})$ , and  $\boldsymbol{\Sigma}$  should be row- $\kappa$  sparse with relative small sparsity level  $\kappa/p$ . In Lemma A.1 in Section B of the Appendix, we bound  $\mu_{n,2\delta+2}(\boldsymbol{\Sigma})$  in terms of the condition number and eigenvalues of  $\boldsymbol{\Sigma}$ . In the special case when  $\boldsymbol{\Sigma}$  is diagonal,  $\mu_{n,2\delta+2}(\boldsymbol{\Sigma}) \kappa/p = 1/p$ . In the case that  $\mu_{n,2\delta+2}(\boldsymbol{\Sigma}) \kappa/p$  is small, say  $\mu_{n,2\delta+2}(\boldsymbol{\Sigma}) \kappa/p < 1$ , in the upper bound the term  $\mu_{n,2\delta+2}(\boldsymbol{\Sigma}) (\kappa/p)^2$  can be dropped, resulting in an additional constant factor, since  $\mu_{n,2\delta+2}(\boldsymbol{\Sigma}) (\kappa/p)^2 < 1$ . In other words the effective upper bound is  $\mu_{n,2\delta+2}(\boldsymbol{\Sigma}) \kappa/p + p^{-\frac{1}{\delta}}$ , neglecting the coefficients depending on  $n$ ,  $\delta$  and  $\gamma$ . Respective expressions for  $C'_{\delta,\gamma}$  and  $C_{n,\delta,\rho}$  in Proposition III.6 are presented in equations (91) and (92) in the Appendix.

Proposition III.6 states that for given  $n$ ,  $p$ ,  $\delta$  and  $\gamma$ , if the threshold  $\rho$  is properly chosen, and  $\boldsymbol{\Sigma}$  is row- $\kappa$  sparse and has small  $\mu_{n,2\delta+2}(\boldsymbol{\Sigma})$ , then the distribution of  $N_{E_\delta}^{(\mathbf{R})}$  can be well approximated by the compound Poisson  $\text{CP}(\lambda_{p,n,\delta,\rho}, \zeta_{n,\delta,\rho})$ . We will call  $\text{CP}(\lambda_{p,n,\delta,\rho}, \zeta_{n,\delta,\rho})$  the *non-asymptotic compound Poisson distribution*. In the Appendix we provide an informal argument (Section D) to motivate Proposition III.6 in addition to a complete proof (Section E).

#### D. A Portmanteau Result and Bounds on Pairwise Total Variations

In this subsection upper bounds for pairwise total variation distances and  $L^1$  distances among  $\{N_{E_\delta}^{(\Psi)}, N_{V_\delta}^{(\Psi)}, N_{V_\delta}^{(\mathbf{P})} : k \in \{\mathbf{R}, \mathbf{P}\}\}$  are obtained. Some intuition is presented before stating the portmanteau result Proposition III.9.

**Lemma III.8:** Consider  $\delta \in [p-2]$ .

$$\begin{aligned} N_{E_\delta}^{(\mathbf{R})} - (\delta+1)N_{E_{\delta+1}}^{(\mathbf{R})} & \leq N_{V_\delta}^{(\mathbf{R})} \leq N_{V_\delta}^{(\mathbf{R})} \leq N_{E_\delta}^{(\mathbf{R})}, \\ N_{E_\delta}^{(\mathbf{P})} - (\delta+1)N_{E_{\delta+1}}^{(\mathbf{P})} & \leq N_{V_\delta}^{(\mathbf{P})} \leq N_{V_\delta}^{(\mathbf{P})} \leq N_{E_\delta}^{(\mathbf{P})}. \end{aligned}$$

It follows directly from Lemma III.8 that for  $\tilde{N}_\delta \in \{N_{V_\delta}^{(\mathbf{R})}, N_{V_\delta}^{(\mathbf{P})}\}$ ,

$$\mathbb{E} \left| \tilde{N}_\delta - N_{E_\delta}^{(\mathbf{R})} \right| \leq (\delta+1) \mathbb{E} N_{E_{\delta+1}}^{(\mathbf{R})}. \quad (29)$$

As a result, if  $\mathbb{E} N_{E_{\delta+1}}^{(\mathbf{R})}$  is small, then  $N_{V_\delta}^{(\mathbf{R})}$  and  $N_{V_\delta}^{(\mathbf{P})}$  are close to  $N_{E_\delta}^{(\mathbf{R})}$  in  $L^1$  norm.

To intuitively see why the vertex counts in the empirical partial correlation graph are close to those in the empirical correlation graph, consider large  $p$  and suppose that  $\boldsymbol{\Sigma}$  is diagonal. Then by Lemma III.2 (b),  $\{\mathbf{u}_i\}_{i=1}^p$  are i.i.d.  $\text{unif}(S^{n-2})$ . According to the law of large numbers, for large  $p$ ,

$$\mathbf{B} = \frac{n-1}{p} \sum_{i=1}^p \mathbf{u}_i \mathbf{u}_i^\top \approx (n-1) \mathbb{E} \mathbf{u}_i \mathbf{u}_i^\top = \mathbf{I}_{n-1}, \quad (30)$$

which implies that  $\bar{\mathbf{Y}} = \mathbf{B}^{-1} \mathbf{U} \approx \mathbf{U}$ , so that  $\|\bar{\mathbf{y}}_i\|_2 \approx \|\mathbf{u}_i\|_2 = 1$ , and thus

$$\mathbf{Y} = \left[ \frac{\bar{\mathbf{y}}_1}{\|\bar{\mathbf{y}}_1\|_2}, \dots, \frac{\bar{\mathbf{y}}_p}{\|\bar{\mathbf{y}}_p\|_2} \right] \approx \bar{\mathbf{Y}} \approx \mathbf{U}. \quad (31)$$

Recall in Subsection III-B that  $N_{E_\delta}^{(\mathbf{R})}$  and  $N_{E_\delta}^{(\mathbf{P})}$  are the numbers of subgraphs isomorphic to  $\Gamma_\delta$  associated with the random pseudo geometric graphs induced respectively by  $\mathbf{U}$  and  $\mathbf{Y}$ . Hence by (31),  $N_{E_\delta}^{(\mathbf{R})} \approx N_{E_\delta}^{(\mathbf{P})}$ . Following the same reasoning,  $N_{V_\delta}^{(\mathbf{R})} \approx N_{V_\delta}^{(\mathbf{P})}$  and  $N_{V_\delta}^{(\mathbf{R})} \approx N_{V_\delta}^{(\mathbf{P})}$ . These informal arguments are formalized by Proposition III.9, which establishes that all 6 quantities  $\{N_{E_\delta}^{(\Psi)}, N_{V_\delta}^{(\Psi)}, N_{V_\delta}^{(\mathbf{P})} : k \in \{\mathbf{R}, \mathbf{P}\}\}$  are mutually close in  $L^1$  norm and their distributions are mutually close in total variation. Proposition III.9 is therefore called a ‘‘Portmanteau result’’.

**Proposition III.9 (Portmanteau Result):** Let  $p \geq n \geq 4$  and  $\mathbf{X} \sim \mathcal{VE}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, g)$ . Let  $\delta \in [p-1]$ . Suppose  $2p^{1+\frac{1}{\delta}}P_n(r_\rho) \leq \gamma$ .

(a) Suppose  $\boldsymbol{\Sigma}$  is row- $\kappa$  sparse. Then for  $\tilde{N}_\delta \in \{N_{V_\delta}^{(\mathbf{R})}, N_{V_\delta}^{(\mathbf{P})}\}$ ,

$$\begin{aligned} & d_{\text{TV}} \left( \mathcal{L}(\tilde{N}_\delta), \mathcal{L}(N_{E_\delta}^{(\mathbf{R})}) \right) \\ & \leq \mathbb{E} \left| \tilde{N}_\delta - N_{E_\delta}^{(\mathbf{R})} \right| \\ & \leq \frac{(\delta+1)^2}{\delta!} \gamma^{\delta+1} \left( 1 + \mu_{n,\delta+2}(\boldsymbol{\Sigma}) \frac{\kappa-1}{p} \right) p^{-\frac{1}{\delta}}. \end{aligned}$$

(b) Suppose  $\boldsymbol{\Sigma}$ , after some row-column permutation, is  $(\tau, \kappa)$  sparse with  $\tau \leq \frac{p}{2}$ . Suppose the condition

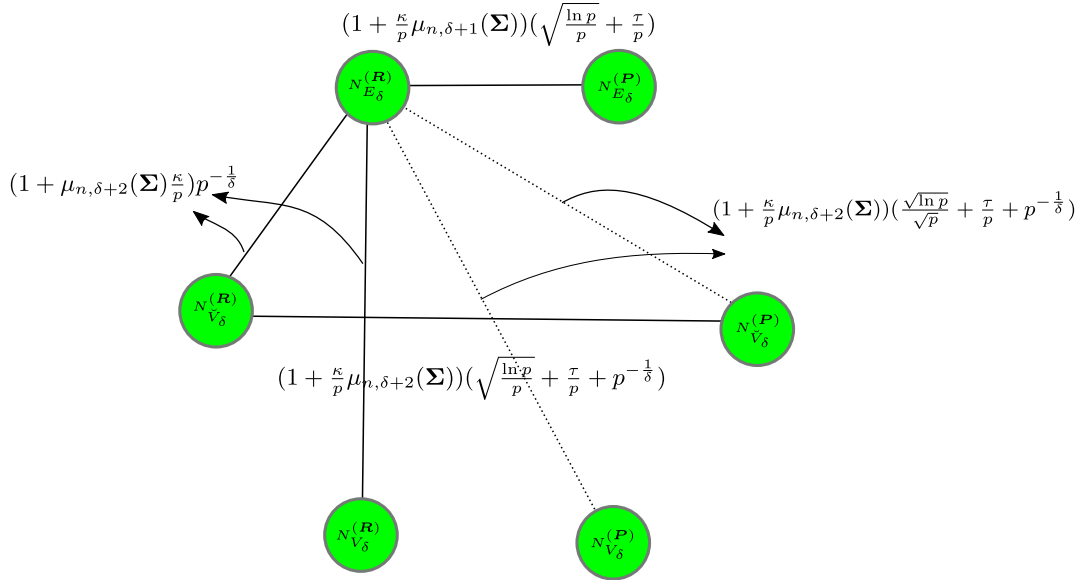


Fig. 3. Illustration of the pairwise total variation distances between the vertex counts associated with the empirical correlation and partial correlation graph established by the portmanteau result Proposition III.9. The 4 solid edges correspond to existence of direct upper bounds on the total variation distances between two types of vertex counts, where the weights correspond respectively to the 4 upper bounds (neglecting constant coefficients) in Proposition III.9. Dashed edges correspond to indirect upper bounds of the associated total variation distances obtained by applying the triangular inequalities, with weights computed from the solid path connecting the two vertices.

$\left(\sqrt{\frac{n-1}{p}} + \sqrt{\frac{\delta \ln p}{p}}\right) \leq c$  is satisfied for some positive and sufficiently small constant  $c$ . Then

$$\begin{aligned} & d_{\text{TV}} \left( \mathcal{L} \left( N_{E_\delta}^{(P)} \right), \mathcal{L} \left( N_{E_\delta}^{(R)} \right) \right) \\ & \leq \mathbb{E} \left| N_{E_\delta}^{(P)} - N_{E_\delta}^{(R)} \right| \\ & \leq C_{E_\delta}^{(P)} \left( 1 + \frac{\kappa-1}{p} \mu_{n,\delta+1}(\Sigma) \right) \left( \sqrt{\frac{\ln p}{p}} + \frac{\tau}{p} \right), \quad (32) \end{aligned}$$

where  $C_{E_\delta}^{(P)}$  is a constant depending on only  $n, \delta$  and  $\gamma$ .

(c) Suppose the same conditions as in part (b) hold. Then

$$\begin{aligned} & d_{\text{TV}} \left( \mathcal{L} \left( N_{V_\delta}^{(P)} \right), \mathcal{L} \left( N_{V_\delta}^{(R)} \right) \right) \\ & \leq \mathbb{E} \left| N_{V_\delta}^{(P)} - N_{V_\delta}^{(R)} \right| \\ & \leq C_{V_\delta}^{(P)} \left( 1 + \frac{\kappa-1}{p} \mu_{n,\delta+2}(\Sigma) \right) \left( \sqrt{\frac{\ln p}{p}} + \frac{\tau}{p} + p^{-\frac{1}{\delta}} \right), \quad (33) \end{aligned}$$

where  $C_{V_\delta}^{(P)}$  is a constant depending on only  $n, \delta$  and  $\gamma$ .

*Remark III.10:* In Proposition III.9 (b) and (c), the condition

$$\left( \sqrt{\frac{n-1}{p}} + \sqrt{\frac{\delta \ln p}{p}} \right) \leq c \quad (34)$$

explicitly specifies a lower bound on  $p$ . Specifically, observe that the left side of (34) is a decreasing function of  $p$ , and its limit is 0 when  $p \rightarrow \infty$ . Thus the smallest positive integer value of  $p$  satisfying (34) exists, denoted by  $p_0$ , which is a function of  $n, \delta$  and  $c$ . Therefore (34) is equivalent to requiring  $p \geq p_0$ . The inequalities (32) and (33) are valid as long as  $c$  is less than finite constant, given respectively

in (110) and (115) in the Appendix. The row- $\kappa$  sparsity condition on  $\Sigma$  guarantees that the vertex counts associated with the empirical correlation graph are close in  $L^1$  norm to  $N_{E_\delta}^{(R)}$  by Proposition III.9 (a). In Proposition III.9 (b) and (c) the stronger condition of  $(\tau, \kappa)$  sparsity suffices to establish that the vertex counts of  $G_\rho(\mathbf{P})$  are close in  $L^1$  norm to those of  $G_\rho(\mathbf{R})$ . All 3 upper bounds in the proposition are bounded by  $\left( 1 + \frac{\kappa-1}{p} \mu_{n,\delta+2}(\Sigma) \right) \left( \sqrt{\frac{\ln p}{p}} + \frac{\tau}{p} + p^{-\frac{1}{\delta}} \right)$ , up to a multiplicative constant depending on only  $n, \delta$  and  $\gamma$ . The interpretation of this upper bound is similar to the second paragraph in Remark III.7. Respective expressions for  $C_{E_\delta}^{(P)}$  and  $C_{V_\delta}^{(P)}$  are presented in equations (112) and (118) in the Appendix.

Figure 3 illustrates the relations established by the portmanteau result Proposition III.9. Dashed edges correspond to indirect upper bounds of the total variation distances between vertices, with weights computed from solid path connecting the two vertices. For instance the weight of dash edge between  $N_{E_\delta}^{(R)}$  and  $N_{V_\delta}^{(P)}$  is computed from

$$\begin{aligned} & d_{\text{TV}} \left( \mathcal{L} \left( N_{V_\delta}^{(P)} \right), \mathcal{L} \left( N_{E_\delta}^{(R)} \right) \right) \\ & \leq \mathbb{E} \left| N_{E_\delta}^{(R)} - N_{V_\delta}^{(P)} \right| \\ & \leq \mathbb{E} \left| N_{E_\delta}^{(R)} - N_{V_\delta}^{(R)} \right| + \mathbb{E} \left| N_{E_\delta}^{(R)} - N_{E_\delta}^{(P)} \right| \\ & \leq \mathbb{E} \left| N_{E_\delta}^{(R)} - N_{V_\delta}^{(R)} \right| + \mathbb{E} \left| N_{V_\delta}^{(R)} - N_{V_\delta}^{(P)} \right| + \mathbb{E} \left| N_{E_\delta}^{(R)} - N_{E_\delta}^{(P)} \right| \\ & \leq C_{n,\delta,\gamma} \left( 1 + \frac{\kappa-1}{p} \mu_{n,\delta+2}(\Sigma) \right) \left( \sqrt{\frac{\ln p}{p}} + \frac{\tau}{p} + p^{-\frac{1}{\delta}} \right), \end{aligned}$$

where the first step follows from Lemma A.24 in Section J in the Appendix, the second step follows from Lemma III.8, and

the last step follows from Proposition III.9, with

$$C_{n,\delta,\gamma} = \left( C_{E_\delta}^{(\mathbf{P})} + C_{\tilde{V}_\delta}^{(\mathbf{P})} + \frac{(\delta+1)^2}{\delta!} \gamma^{\delta+1} \right).$$

Figure 3 graphically illustrates the implication of Proposition III.9 that all 6 quantities  $\{N_{E_\delta}^{(\Psi)}, N_{\tilde{V}_\delta}^{(\Psi)}, N_{V_\delta}^{(\Psi)} : k \in \{\mathbf{R}, \mathbf{P}\}\}$  are mutually close in total variation. As a result, the closeness of one quantity among the 6 to some distribution in total variation implies the closeness of all 6 quantities to that same distribution. By Proposition III.6,  $\mathcal{L}(N_{E_\delta}^{(\mathbf{R})})$  is close to the non-asymptotic compound Poisson distribution  $\text{CP}(\lambda_{p,n,\delta,\rho}, \zeta_{n,\delta,\rho})$  in total variation, which implies all 6 quantities  $\{N_{E_\delta}^{(\Psi)}, N_{\tilde{V}_\delta}^{(\Psi)}, N_{V_\delta}^{(\Psi)} : k \in \{\mathbf{R}, \mathbf{P}\}\}$  are close in total variation to  $\text{CP}(\lambda_{p,n,\delta,\rho}, \zeta_{n,\delta,\rho})$ . We state this formally as Theorem III.11 in the next subsection.

### E. A Unified Theorem for Finite $p$

The following theorem is a variant of Theorem II.4 for finite  $p$ . It states that if the threshold  $\rho$  is properly chosen, and  $\Sigma$  satisfies the  $(\tau, \kappa)$  sparsity condition, then  $\mathcal{L}(\tilde{N}_\delta)$  can be approximated by a compound Poisson distribution.

**Theorem III.11 (Compound Poisson Approximation for Finite  $p$ ):** Let  $n \geq 4$ ,  $\delta \in [p-1]$ , and  $\gamma > 0$  be given. Consider  $\mathbf{X} \sim \mathcal{VE}(\mu, \Sigma, g)$ . Suppose  $2p^{1+\frac{1}{\delta}} P_n(r_\rho) \leq \gamma$ . Suppose  $\Sigma$ , after some row-column permutation, is  $(\tau, \kappa)$  sparse with  $\tau \leq \frac{p}{2}$  and  $\mu_{n,2\delta+2}(\Sigma) \frac{\kappa}{p} < 1$ . Suppose the condition  $\left( \sqrt{\frac{n-1}{p}} + \sqrt{\frac{\delta \ln p}{p}} \right) \leq c$  is satisfied for some positive and sufficiently small constant  $c$ . Then  $\tilde{N}_\delta$ , a generic random variable in the set  $\{N_{E_\delta}^{(\Psi)}, N_{\tilde{V}_\delta}^{(\Psi)}, N_{V_\delta}^{(\Psi)} : k \in \{\mathbf{R}, \mathbf{P}\}\}$ , satisfies:

$$\begin{aligned} & d_{\text{TV}}(\mathcal{L}(\tilde{N}_\delta), \text{CP}(\lambda_{p,n,\delta,\rho}, \zeta_{n,\delta,\rho})) \\ & \leq C_{n,\delta,\gamma} \left( \mu_{n,2\delta+2}(\Sigma) \frac{\kappa}{p} + p^{-\frac{1}{\delta}} + E(p, \tau) \right), \end{aligned} \quad (35)$$

where

$$E(p, \tau) = \begin{cases} 0 & \text{if } \tilde{N}_\delta = N_{E_\delta}^{(\mathbf{R})}, N_{\tilde{V}_\delta}^{(\mathbf{R})} \text{ or } N_{V_\delta}^{(\mathbf{R})}, \\ \sqrt{\frac{\ln p}{p}} + \frac{\tau}{p} & \text{if } \tilde{N}_\delta = N_{E_\delta}^{(\mathbf{P})}, N_{\tilde{V}_\delta}^{(\mathbf{P})} \text{ or } N_{V_\delta}^{(\mathbf{P})}. \end{cases}$$

If only the vertex counts in the empirical correlation graph is of interest, then the  $(\tau, \kappa)$  sparsity assumption can be relaxed to row- $\kappa$  sparsity.

**Lemma III.12 (Compound Poisson Approximation in Empirical Correlation Graph):** Let  $n \geq 4$ ,  $\delta \in [p-1]$ , and  $\gamma > 0$  be given. Consider  $\mathbf{X} \sim \mathcal{VE}(\mu, \Sigma, g)$ . Suppose  $2p^{1+\frac{1}{\delta}} P_n(r_\rho) \leq \gamma$ . Suppose  $\Sigma$  is row- $\kappa$  sparse with  $\mu_{n,2\delta+2}(\Sigma) \frac{\kappa}{p} < 1$ . Then  $\tilde{N}_\delta$ , a generic random variable in the set  $\{N_{E_\delta}^{(\mathbf{R})}, N_{\tilde{V}_\delta}^{(\mathbf{R})}, N_{V_\delta}^{(\mathbf{R})}\}$ , satisfies

$$\begin{aligned} & d_{\text{TV}}(\mathcal{L}(\tilde{N}_\delta), \text{CP}(\lambda_{p,n,\delta,\rho}, \zeta_{n,\delta,\rho})) \\ & \leq C_{n,\delta,\gamma} \left( \mu_{n,2\delta+2}(\Sigma) \frac{\kappa}{p} + p^{-\frac{1}{\delta}} \right). \end{aligned} \quad (36)$$

**Remark III.13:** The assumption  $\mu_{n,2\delta+2}(\Sigma) \frac{\kappa}{p} < 1$  is only used to obtain simpler expressions for the upper bounds in (35) and (36). Without this assumption, similar inequalities

hold with the upper bounds that are the sum the upper bounds in Proposition III.6 and Proposition III.9. See the third paragraph in Remark III.7 for additional discussions. Observe that Theorem III.11 and Lemma III.12 hold for any mean  $\mu$  and any shaping function  $g$  when  $\mathbf{X} \sim \mathcal{VE}(\mu, \Sigma, g)$ . This is a consequence of the invariance property of the U-scores distribution (see Remark III.3). In particular, none of the constants in Theorem III.11 depend on  $\mu$  or  $g$ .

Theorem III.11 and Lemma III.12 directly follow from Proposition III.6 and Proposition III.9 and hence their proofs are omitted. Theorem III.11 and Lemma III.12 provide an approximation for the family-wise error rate (FWER) [18]

$$\mathbb{P}(\tilde{N}_\delta > 0) \approx 1 - e^{-\lambda_{p,n,\delta,\rho}}.$$

We end this subsection by making a comparison between Theorem III.11 and Theorem II.4. Theorem III.11 and Theorem II.4 provide compound Poisson approximations to  $\tilde{N}_\delta$  respectively when  $p$  is finite and when  $p \rightarrow \infty$ . By taking the limit as  $p \rightarrow \infty$ , we obtain simpler formulae for parameters of the approximation. Specifically the increment distribution  $\zeta_{n,\delta,\rho}$  of the non-asymptotic compound Poisson distribution in Theorem III.11 depends on conditional probabilities in the random pseudo geometric graph as in (23). On the other hand, the increment distribution  $\zeta_{n,\delta}$  of the limiting compound Poisson distribution in Theorem II.4 depends on probabilities in the random geometric graph as in (6), which is relatively simpler. For instance, when  $\delta = 2$ , an analytical formula for  $\zeta_{n,2}$  can be obtained (see Example V.10). Obtaining an analytical formula for  $\zeta_{n,2,\rho}$  does not seem straightforward.

Despite the fact that the limiting compound Poisson distribution in Theorem II.4 is relatively simpler than Theorem III.11, it has the disadvantage that the approximation is not accurate unless  $p$  is large. Moreover, Theorem III.11 is stronger than Theorem II.4 in the sense that Theorem III.11 provides explicit upper bounds for the approximation errors, while Theorem II.4 simply provides a limit but no convergence rates. For further discussion see Remark III.17 in the next subsection.

### F. Proof of Theorem II.4

In this subsection we present results on the limit of  $\text{CP}(\lambda_{p,n,\delta,\rho}, \zeta_{n,\delta,\rho})$  as  $p \rightarrow \infty, \rho \rightarrow 1$ , followed by a proof of Theorem II.4.

To study the limiting distribution of  $\text{CP}(\lambda_{p,n,\delta,\rho}, \zeta_{n,\delta,\rho})$ , which requires the limit of the parameter  $\alpha(\ell, r_\rho)$ , the next two lemmas are useful.

**Lemma III.14:** Consider  $r < 2/\sqrt{5}$  and  $\delta \geq 1$ . Suppose  $\{\mathbf{u}'_i\}_{i=1}^{\delta+1} \stackrel{\text{i.i.d.}}{\sim} \text{unif}(S^{n-2})$ . Then for any  $\ell \in [\delta+1]$ ,

$$\begin{aligned} & \mathbb{P}(\text{NUV}(\{\mathbf{u}'_i\}_{i=1}^{\delta+1}, r) = \ell | \deg(\mathbf{u}'_{\delta+1}) = \delta) \\ & = \mathbb{P}(\text{PNUV}(\{\mathbf{u}'_i\}_{i=1}^{\delta+1}, r) = \ell | \deg(\mathbf{u}'_{\delta+1}) = \delta), \end{aligned} \quad (37)$$

where  $\deg(\mathbf{u}'_{\delta+1})$  on the left (right) side is the degree of vertex  $\mathbf{u}'_{\delta+1}$  in the corresponding random (pseudo) geometric graph.

Lemma III.14 establishes that the conditional distributions of the number of universal vertices in the random geometric graph and the random pseudo geometric graph are identical.

By Lemma III.14 and (23),

$$\alpha(\ell, r_\rho) = \mathbb{P}(\mathbf{NUV}(\{\mathbf{u}'_i\}_{i=1}^{\delta+1}, r_\rho) = \ell | \deg(\mathbf{u}'_{\delta+1}) = \delta) \quad (38)$$

when  $r_\rho < 2/\sqrt{5}$  or equivalently  $\rho > 3/5$ . Therefore to study the limit of  $\alpha(\ell, r_\rho)$  when  $r_\rho \rightarrow 0$  or equivalently  $\rho \rightarrow 1$ , the limiting form of the right hand side of (38) is required, given by the following lemma.

**Lemma III.15:** Let  $\delta \geq 1$  and  $n \geq 3$ . Suppose  $\{\mathbf{u}'_i\}_{i=1}^{\delta+1} \stackrel{\text{i.i.d.}}{\sim} \mathbf{unif}(S^{n-2})$  and  $\{\tilde{\mathbf{u}}_i\}_{i=1}^\delta \stackrel{\text{i.i.d.}}{\sim} \mathbf{unif}(B^{n-2})$ . Then for any  $\ell \in [\delta + 1]$ ,

$$\lim_{r \rightarrow 0} \mathbb{P}(\mathbf{NUV}(\{\mathbf{u}'_i\}_{i=1}^{\delta+1}, r) = \ell | \deg(\mathbf{u}'_{\delta+1}) = \delta) = \mathbb{P}(\mathbf{NUV}(\{\tilde{\mathbf{u}}_i\}_{i=1}^\delta, 1) = \ell - 1). \quad (39)$$

One immediate consequence of Lemma III.15 and (38) is that  $\lim_{\rho \rightarrow 1} \alpha(\ell, r_\rho) = \alpha_\ell$  for any  $\ell \in [\delta + 1]$ . As  $p \rightarrow \infty$ , the condition  $2p^{1+\frac{1}{\delta}}P_n(r_\rho) \leq \gamma$  in Theorem III.11 entails  $\rho \rightarrow 1$ . The following lemma states if the rate of  $\rho \rightarrow 1$  is coupled with the rate  $p \rightarrow \infty$ , then the non-asymptotic compound Poisson distribution  $\text{CP}(\lambda_{p,n,\delta,\rho}, \zeta_{n,\delta,\rho})$  converges in distribution to the limiting compound Poisson distribution  $\text{CP}(\lambda_{n,\delta}(e_{n,\delta}), \zeta_{n,\delta})$ .

**Lemma III.16:** Suppose as  $p \rightarrow \infty$ ,  $\rho \rightarrow 1$  such that  $a_n 2^{\frac{n}{2}} p^{1+\frac{1}{\delta}} (1 - \rho)^{\frac{n-2}{2}} \rightarrow e_{n,\delta}$ , where  $e_{n,\delta}$  is some positive constant that possibly depends on  $n$  and  $\delta$ . Then

$$\text{CP}(\lambda_{p,n,\delta,\rho}, \zeta_{n,\delta,\rho}) \xrightarrow{\mathcal{D}} \text{CP}(\lambda_{n,\delta}(e_{n,\delta}), \zeta_{n,\delta}). \quad (40)$$

With the above results established we are in a position to prove Theorem II.4 and Lemma II.5.

*Proof of Theorem II.4 and Lemma II.5:* Theorem II.4 directly follows from Theorem III.11 and Lemma III.16. Lemma II.5 directly follows from Lemma III.12 and Lemma III.16.  $\square$

**Remark III.17:** In (135) of the proof of Lemma III.16, it is shown that part of the upper bound for its error rate is of the order  $1 - \rho$ . This particular rate, however, decreases as  $n$  increases. More concisely, if one chooses  $\rho$  according to (11) then  $1 - \rho$  is of the order  $p^{-(1+\frac{1}{\delta})\frac{2}{n-2}}$ . Hence the convergence to  $\text{CP}(\lambda_{n,\delta}(e_{n,\delta}), \zeta_{n,\delta})$  in Theorem II.4 is only accurate for large  $p$ . On the other hand, the upper bound in Theorem III.11 only depends on  $\rho$  through  $2p^{1+\frac{1}{\delta}}P_n(r_\rho) \leq \gamma$ , which holds for small  $p$ , again if  $\rho$  is chosen according to (11) (see the first paragraph of Remark III.7 for related discussion). Hence Theorem III.11 provides an accurate approximation to  $\tilde{N}_\delta$  even for small  $p$ . The accuracy of these approximations for various values of  $\rho$  and  $n$  is numerically illustrated in Figure 8 in Section K of the Appendix.

#### IV. CONVERGENCE OF MOMENTS

Moment expressions are useful for a number of reasons, including characterizing the behavior of phase transition thresholds and the expected number of false discoveries [17]. Theorems II.4 and III.11 only prescribe the distribution of  $\tilde{N}_\delta$  but not the moments. In this subsection, we present approximations to the first moment and second moment of  $\tilde{N}_\delta$  for finite  $p$ .

Let  $Z \sim \text{CP}(\lambda_{p,n,\delta,\rho}, \zeta_{n,\delta,\rho})$ . Then we can represent  $Z = \sum_{i=1}^N Z_i$ , where  $N$  is distributed as a Poisson with mean  $\lambda_{p,n,\delta,\rho}$ ,  $Z_i \stackrel{\text{i.i.d.}}{\sim} \zeta_{n,\delta,\rho}$  and  $N$  is independent of each  $Z_i$ . The first two moments of  $Z$  are:

$$\mathbb{E}Z = \mathbb{E}N\mathbb{E}Z_1 = \binom{p}{1} \binom{p-1}{\delta} (2P_n(r_\rho))^\delta, \quad (41)$$

$$\begin{aligned} \mathbb{E}Z^2 &= \mathbb{E}N\mathbb{E}Z_1^2 + (\mathbb{E}N\mathbb{E}Z_1)^2 \\ &= \binom{p}{1} \binom{p-1}{\delta} (2P_n(r_\rho))^\delta \sum_{\ell=1}^{\delta+1} \ell \alpha(\ell, r_\rho) \\ &\quad + \left( \binom{p}{1} \binom{p-1}{\delta} (2P_n(r_\rho))^\delta \right)^2. \end{aligned} \quad (42)$$

The next lemma provides an upper bound for the difference between the first moment of  $N_{E_\delta}^{(\mathbf{R})}$  and the first moment of the compound Poisson specified in (41).

**Lemma IV.1:** Let  $p \geq n \geq 4$ ,  $\delta \in [p-1]$  and  $\gamma > 0$  be given. Suppose  $\mathbf{X} \sim \mathcal{VE}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, g)$ . Suppose  $2p^{1+\frac{1}{\delta}}P_n(r_\rho) \leq \gamma$ , and  $\boldsymbol{\Sigma}$  is row- $\kappa$  sparse. Then

$$\left| \mathbb{E}N_{E_\delta}^{(\mathbf{R})} - \mathbb{E}Z \right| \leq \frac{(\delta+1)}{2((\delta-1)!)} \gamma^\delta \mu_{n,\delta+1}(\boldsymbol{\Sigma}) \frac{\kappa-1}{p}.$$

**Remark IV.2:** Lemma IV.1 implies that the condition on  $\rho$  assumed in Theorem II.4 is in fact necessary and sufficient for the mean to converge to a finite and strictly positive limit. For simplicity we specialize to the case that  $\boldsymbol{\Sigma}$  is diagonal, which is row- $\kappa$  sparse with  $\kappa = 1$ . By Lemma IV.1 and (41),

$$\mathbb{E}N_{E_\delta}^{(\mathbf{R})} = \mathbb{E}Z = \binom{p}{1} \binom{p-1}{\delta} (2P_n(r_\rho))^\delta,$$

which via Stirling approximation behaves as  $\frac{1}{\delta!} (2p^{1+\frac{1}{\delta}}P_n(r_\rho))^\delta$  for large  $p$ . By Lemma A.22 (b) we further have, again for large  $p$ ,

$$\mathbb{E}N_{E_\delta}^{(\mathbf{R})} \approx \frac{1}{\delta!} \left( 2p^{1+\frac{1}{\delta}} a_n (1 - \rho)^{\frac{n-2}{2}} \right)^\delta.$$

Thus, in order that the mean count have a non-degenerate limit when  $p \rightarrow \infty$ ,  $2p^{1+\frac{1}{\delta}} a_n (1 - \rho)^{\frac{n-2}{2}}$  must converge to some strictly positive and finite constant value. If  $\rho$  does not converge to 1, or if it converges to 1 at a slower rate, then  $\mathbb{E}\tilde{N}_\delta$  diverges to  $\infty$ , while if  $\rho$  converges to 1 at a faster rate then  $\mathbb{E}\tilde{N}_\delta$  converges to 0. This is reflected in the phase transition phenomenon discussed in [17] and [18]. Thus the rate on  $\rho$  in Theorem II.4 is sharp.

By combining the preceding lemma and the portman-teau result of Proposition III.9 one immediately obtains approximations for the first moment of all 6 quantities  $\{N_{E_\delta}^{(\Psi)}, N_{\tilde{V}_\delta}^{(\Psi)}, N_{V_\delta}^{(\Psi)} : \Psi \in \{\mathbf{R}, \mathbf{P}\}\}$  for finite  $p$ .

We also characterize the second moment approximation of  $N_{E_\delta}^{(\mathbf{R})}$  in the next lemma under the same conditions as Lemma IV.1.

**Proposition IV.3 (Second Moment Bounds for Subgraph Counts):** Let  $p \geq n \geq 4$ ,  $\delta \in [p-1]$  and  $\gamma > 0$  be given. Suppose  $\mathbf{X} \sim \mathcal{VE}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, g)$ . Suppose  $2p^{1+\frac{1}{\delta}}P_n(r_\rho) \leq \gamma$ , and  $\boldsymbol{\Sigma}$  is row- $\kappa$  sparse. Then

$$\left| \mathbb{E} \left( N_{E_\delta}^{(\mathbf{R})} \right)^2 - \mathbb{E}Z^2 \right| \leq C_{n,\delta,\gamma} \left( \mu_{n,2\delta+2}(\boldsymbol{\Sigma}) \frac{\kappa}{p} + p^{-1/\delta} \right).$$

We extend the preceding proposition to other quantities in  $\{N_{E_\delta}^{(\Psi)}, N_{\tilde{V}_\delta}^{(\Psi)}, N_{V_\delta}^{(\Psi)} : \Psi \in \{\mathbf{R}, \mathbf{P}\}\}$  by generalizing Proposition III.9 to  $L^1$  distance between the square of the quantities.

**Proposition IV.4 (Portmanteau Result for  $L^2$  Distance):** Let  $p \geq n \geq 4$  and  $\mathbf{X} \sim \mathcal{VE}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, g)$ . Let  $\delta \in [p-1]$ . Suppose  $2p^{1+\frac{1}{\delta}}P_n(r_\rho) \leq \gamma$ .

- (a) Suppose  $\boldsymbol{\Sigma}$  is row- $\kappa$  sparse. Then for  $\tilde{N}_\delta \in \{N_{\tilde{V}_\delta}^{(\mathbf{R})}, N_{V_\delta}^{(\mathbf{R})}\}$ ,

$$\begin{aligned} & \mathbb{E} \left| \left( \tilde{N}_\delta \right)^2 - \left( N_{E_\delta}^{(\mathbf{R})} \right)^2 \right| \\ & \leq C_{n,\delta,\gamma} \left( 1 + \mu_{n,2\delta+3}(\boldsymbol{\Sigma}) \frac{\kappa-1}{p} \right) p^{-1/\delta}. \end{aligned}$$

- (b) Suppose  $\boldsymbol{\Sigma}$ , after some row-column permutation, is  $(\tau, \kappa)$  sparse with  $\tau \leq \frac{p}{2}$ . Moreover,  $\left( \sqrt{\frac{n-1}{p}} + \sqrt{\frac{\delta \ln p}{p}} \right) \leq c$  hold for some positive and sufficiently small constant  $c$ . Then

$$\begin{aligned} & \mathbb{E} \left| \left( N_{E_\delta}^{(\mathbf{P})} \right)^2 - \left( N_{E_\delta}^{(\mathbf{R})} \right)^2 \right| \\ & \leq C_{n,\delta,\gamma} \left( 1 + \mu_{n,2\delta+2}(\boldsymbol{\Sigma}) \frac{\kappa-1}{p} \right) \left( \frac{\sqrt{\ln p}}{\sqrt{p}} + \frac{\tau}{p} \right). \end{aligned}$$

- (c) Suppose the same conditions as in part (b) hold. Then for  $\tilde{N}_\delta \in \{N_{\tilde{V}_\delta}^{(\mathbf{P})}, N_{V_\delta}^{(\mathbf{P})}\}$

$$\begin{aligned} & \mathbb{E} \left| \left( \tilde{N}_\delta \right)^2 - \left( N_{E_\delta}^{(\mathbf{P})} \right)^2 \right| \\ & \leq C_{n,\delta,\gamma} \left( 1 + \mu_{n,2\delta+3}(\boldsymbol{\Sigma}) \frac{\kappa-1}{p} \right) p^{-1/\delta}. \end{aligned}$$

By applying triangle inequalities to Proposition IV.4 (b) and (c), one obtain for  $\tilde{N}_\delta \in \{N_{\tilde{V}_\delta}^{(\mathbf{P})}, N_{V_\delta}^{(\mathbf{P})}\}$

$$\begin{aligned} & \mathbb{E} \left| \left( \tilde{N}_\delta \right)^2 - \left( N_{E_\delta}^{(\mathbf{R})} \right)^2 \right| \\ & \leq C_{n,\delta,\gamma} \left( 1 + \mu_{n,2\delta+3}(\boldsymbol{\Sigma}) \frac{\kappa-1}{p} \right) \left( \frac{\sqrt{\ln p}}{\sqrt{p}} + \frac{\tau}{p} + p^{-1/\delta} \right). \end{aligned}$$

Thus we have established the  $L^1$  distance between the square of each term in  $\{N_{E_\delta}^{(\Psi)}, N_{\tilde{V}_\delta}^{(\Psi)}, N_{V_\delta}^{(\Psi)} : \Psi \in \{\mathbf{R}, \mathbf{P}\}\}$  and  $\left( N_{E_\delta}^{(\mathbf{R})} \right)^2$ . By combining Proposition IV.3 and Proposition IV.4 one immediately obtains approximations to the second moment of  $\tilde{N}_\delta$  for finite  $p$ .

While this section focuses on the approximation to the first and second moments of  $\tilde{N}_\delta$  for finite  $p$ , their limits can also be obtained when  $p \rightarrow \infty$  and  $\rho \rightarrow 1$  at the rate specified in Theorem II.4.

## V. ANALYTICAL EXPRESSIONS FOR THE COMPOUND POISSON PARAMETERS

Recall that the limiting compound Poisson distribution in Theorem II.4 is in terms of  $\alpha_\ell$ , while the non-asymptotic compound Poisson distributions in Theorem III.11 is in terms of  $\alpha(\ell, r_\rho)$ . Moreover the second moment approximation established in Proposition IV.3 also involves the term  $\alpha(\ell, r_\rho)$

due to (42). Since  $\alpha_\ell$  and  $\alpha(\ell, r_\rho)$  are expressed in terms of random (pseudo) geometric graphs, they are tedious to compute. In subsection V-A, we obtain simple analytical approximations to these quantities for moderately large  $n$  or  $\delta$ . In subsection V-B, exact analytical formulae for these quantities are established for the special case  $\delta = 1$  and  $\delta = 2$ .

### A. Approximations to $\alpha_\ell$ and $\alpha(\ell, r_\rho)$ When $\delta \geq 2$

In this subsection we show that when  $n$  or  $\delta$  is moderately large,  $\alpha_1 \approx 1$  and  $\alpha_\ell \approx 0$  for  $2 \leq \ell \leq \delta + 1$  (parallel results for  $\alpha(\ell, r_\rho)$  are  $\alpha(1, r_\rho) \approx 1$  and  $\alpha(\ell, r_\rho) \approx 0$  for  $2 \leq \ell \leq \delta + 1$ ). These approximations yield simple formulae for parameters of the limiting and non-asymptotic compound Poisson distributions, and for parameters of the second moment (42). Importantly, the compound Poisson distributions are well approximated by Poisson distributions for  $n, \delta$  moderately large.

To show that  $\alpha_\ell$  is small for  $\ell \geq 2$  and  $\alpha_1 \approx 1$ , it suffices to establish a vanishing upper bound on  $\sum_{\ell=2}^{\delta+1} \alpha_\ell$ , since  $\sum_{\ell=1}^{\delta+1} \alpha_\ell = 1$ . By definition of  $\alpha_\ell$ ,  $\sum_{\ell=2}^{\delta+1} \alpha_\ell = \mathbb{P}(\mathbf{NUV}(\{\tilde{\mathbf{u}}_i\}_{i=1}^\delta, 1) \geq 1)$ , where  $\{\tilde{\mathbf{u}}_i\}_{i=1}^\delta \stackrel{\text{i.i.d.}}{\sim} \mathbf{unif}(B^{n-2})$ . The next lemma establishes an upper bound on the geometric quantity  $\mathbb{P}(\mathbf{NUV}(\{\tilde{\mathbf{u}}_i\}_{i=1}^\delta, 1) \geq 1)$ .

**Lemma V.1:** Let  $n \geq 4$  and  $\delta \geq 2$ .

- (a) Consider  $\{\tilde{\mathbf{u}}_i\}_{i=1}^\delta \stackrel{\text{i.i.d.}}{\sim} \mathbf{unif}(B^{n-2})$ . Then

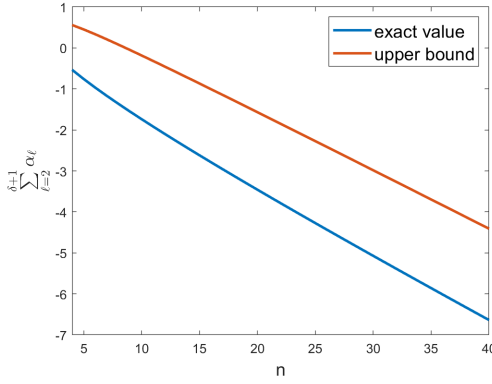
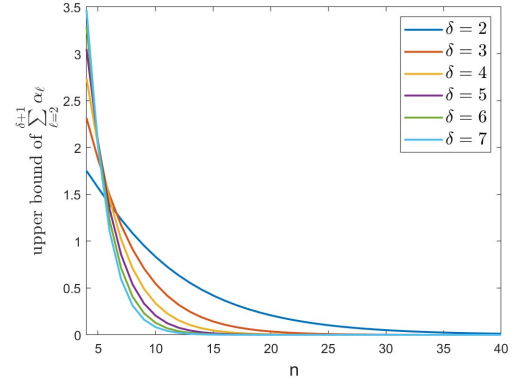
$$\begin{aligned} & \sum_{\ell=2}^{\delta+1} \alpha_\ell \\ & = \mathbb{P}(\mathbf{NUV}(\{\tilde{\mathbf{u}}_i\}_{i=1}^\delta, 1; n-2) \geq 1) \\ & \leq \delta(n-2) \int_0^1 \left( 1 - \frac{r^2}{4} \right)^{\frac{(n-2)(\delta-1)}{2}} r^{n-3} dr \\ & = \delta(n-2) 2^{n-3} B\left(\frac{1}{4}; \frac{n-2}{2}, \frac{(n-2)(\delta-1)}{2} + 1\right), \end{aligned} \tag{43}$$

where  $B(\cdot; \cdot, \cdot)$  is the incomplete beta function.

- (b)

$$\begin{aligned} & \int_0^1 \left( 1 - \frac{r^2}{4} \right)^{\frac{(n-2)(\delta-1)}{2}} r^{n-3} dr \\ & \leq \begin{cases} \left(\frac{4}{5}\right)^{\frac{(n-2)\delta-1}{2}} + \left(1 - \sqrt{\frac{4}{5}}\right) \left(\frac{3}{4}\right)^{\frac{(n-2)(\delta-1)}{2}}, & \delta = 2, 3, \\ \exp\left(\frac{1}{4}\right) \left(\frac{\delta-1}{\delta}\right)^{\frac{(n-2)(\delta-1)}{2}} \left(\frac{4}{\delta}\right)^{\frac{n-3}{2}}, & \delta \geq 4. \end{cases} \end{aligned}$$

Lemma V.1 (a) establishes an upper bound for the probability that there is at least one universal vertex in the random geometric graph generated by the uniform distribution in the unit ball. Lemma V.1 (b) provides an explicit upper bound for part (a) and this upper bound provides the following insight when  $n$  or  $\delta$  is large. In particular, when  $\delta$  is fixed,  $\mathbb{P}(\mathbf{NUV}(\{\tilde{\mathbf{u}}_i\}_{i=1}^\delta, 1; n-2) \geq 1)$  decays exponentially as  $n$  increases. While  $n$  is fixed, it decays at rate  $\delta^{-\frac{n-3}{2}}$  as  $\delta$  increases. Such decay rates also apply to  $\alpha_\ell$  for  $\ell \geq 2$  and  $|\alpha_1 - 1|$ .

(a) Log-scale comparison of the decay when  $\delta = 2$ 

(b) Family of upper bounds

Fig. 4. (a) is a comparison in the log-scale between the upper bound on  $\sum_{\ell=2}^{\delta+1} \alpha_\ell$  by (43) with  $\delta = 2$  and the exact value of  $\sum_{\ell=2}^{\delta+1} \alpha_\ell$  by Example V.9. (b) is the plot of the upper bound on  $\sum_{\ell=2}^{\delta+1} \alpha_\ell$  as a function of  $n$  for  $\delta$  from 2 to 7.

As illustrated in Figure 4 (a), when  $\delta = 2$ ,  $\sum_{\ell=2}^{\delta+1} \alpha_\ell$  and the upper bound in (43) both decay to 0 exponentially fast as  $n$  increases. This diagram demonstrates that the upper bound in (43) captures the decay rate of  $\sum_{\ell=2}^{\delta+1} \alpha_\ell$ . Figure 4 (b) plots the upper bounds in (43) as functions of  $n$  for fixed  $\delta$ . As is clear from the plot, as long as the number of samples  $n$  is above 40,  $\sum_{\ell=2}^{\delta+1} \alpha_\ell \approx 0$ , which yields  $\alpha_\ell \approx 0$  for  $2 \leq \ell \leq \delta + 1$  and  $\alpha_1 \approx 1$ . Moreover, as  $\delta$  increases, the number of samples  $n$  required for  $\sum_{\ell=2}^{\delta+1} \alpha_\ell \approx 0$  decreases.

Denote the Dirac Distribution at  $a$  by  $\text{Dirac}(a)$ . The next lemma establishes approximations to the limiting compound Poisson distribution in Theorem II.4, with the approximation errors upper bounded in terms of  $\sum_{\ell=2}^{\delta+1} \alpha_\ell$ .

**Lemma V.2 (Approximation to the Limiting Compound Poisson Distribution):** Consider  $n \geq 4$  and  $\delta \geq 2$ . Let  $e_{n,\delta}$  be the same as in Theorem II.4.

- (a) The increment distribution satisfies  $d_{\text{TV}}(\zeta_{n,\delta}, \text{Dirac}(1)) \leq \sum_{\ell=2}^{\delta+1} \alpha_\ell$ .
- (b) The arrival rate satisfies

$$\left| \lambda_{n,\delta}(e_{n,\delta}) - \frac{(e_{n,\delta})^\delta}{\delta!} \right| \leq \frac{3}{2} \frac{(e_{n,\delta})^\delta}{\delta!} \sum_{\ell=2}^{\delta+1} \alpha_\ell.$$

- (c) The limiting compound Poisson distribution satisfies

$$d_{\text{TV}}\left(\text{CP}(\lambda_{n,\delta}, \zeta_{n,\delta}), \text{Pois}\left(\frac{(e_{n,\delta})^\delta}{\delta!}\right)\right) \leq \frac{5}{2} \frac{(e_{n,\delta})^\delta}{\delta!} \sum_{\ell=2}^{\delta+1} \alpha_\ell.$$

The proof of Lemma V.2 (c) relies on a general upper bound of the total variation distance between two compound Poisson distributions (Lemma A.25 in Section J of the Appendix). Lemma A.25 may be of independent interest in compound Poisson approximation to non-Poissonian distributions (see [38] and the references therein). From Lemma V.2 (a) and Lemma V.1, the total variation distance between the increment distribution and Dirac distribution at 1 decays exponentially as  $n$  increases and decays at rate  $\delta^{-\frac{n-3}{2}}$  as  $\delta$  increases. Provided that the threshold  $\rho$  is chosen such that  $\frac{(e_{n,\delta})^\delta}{\delta!}$  is not large, the upper bounds in Lemma V.2 (b) and (c) have the same interpretation as part (a). In that case, the limiting compound Poisson are well approximated by the Poisson

distribution  $\text{Pois}\left(\frac{(e_{n,\delta})^\delta}{\delta!}\right)$ . By combining Lemma V.2 (c) and Theorem II.4, we immediately obtain the following result on Poisson approximation to  $\mathcal{L}(\bar{N}_\delta)$  as  $p \rightarrow \infty$ .

**Proposition V.3 (Poisson Approximation as  $p \rightarrow \infty$ ):** Suppose all the conditions in Theorem II.4 hold. Then as  $p \rightarrow \infty$ ,

$$d_{\text{TV}}\left(\mathcal{L}(\bar{N}_\delta), \text{Pois}\left(\frac{(e_{n,\delta})^\delta}{\delta!}\right)\right) \leq \frac{5}{2} \frac{(e_{n,\delta})^\delta}{\delta!} \sum_{\ell=2}^{\delta+1} \alpha_\ell.$$

We now turn our attention to the parameter  $\alpha(\ell, r_\rho)$  of the non-asymptotic compound Poisson distribution. Our results on  $\alpha(\ell, r_\rho)$  are similar to those on  $\alpha_\ell$  but the proofs are more technical. To establish that  $\alpha(1, r_\rho) \approx 1$  and  $\alpha(\ell, r_\rho) \approx 0$  for  $2 \leq \ell \leq \delta + 1$ , it suffices to obtain a vanishing upper bound on  $\sum_{\ell=2}^{\delta+1} \alpha(\ell, r_\rho)$ . By (38), when  $\rho > 3/5$ ,  $\sum_{\ell=2}^{\delta+1} \alpha(\ell, r_\rho) = \mathbb{P}(\text{NUV}(\{\mathbf{u}'_i\}_{i=1}^{\delta+1}, r; n-2) \geq 2 | \deg(\mathbf{u}'_{\delta+1}) = \delta)$ , where  $\{\mathbf{u}'_i\}_{i=1}^{\delta+1} \stackrel{\text{i.i.d.}}{\sim} \text{unif}(S^{n-2})$ . The next lemma is an analogous result to Lemma V.1.

**Lemma V.4:** Let  $n \geq 4$  and  $\delta \geq 2$ .

- (a) Consider  $\{\mathbf{u}'_i\}_{i=1}^{\delta+1} \stackrel{\text{i.i.d.}}{\sim} \text{unif}(S^{n-2})$ . Then for  $0 < r < \sqrt{2}$ ,

$$\begin{aligned} & \mathbb{P}(\text{NUV}(\{\mathbf{u}'_i\}_{i=1}^{\delta+1}, r; n-2) \geq 2 | \deg(\mathbf{u}'_{\delta+1}) = \delta) \\ & \leq \bar{h}\left(\frac{1}{\sqrt{1-r^2/4}}, n, \delta\right) \delta(n-2) \\ & \quad \int_0^1 \left(1 - \left(\frac{r_1}{2}\right)^2\right)^{\frac{(n-2)(\delta-1)}{2}} r_1^{n-3} dr_1 \\ & = \bar{h}\left(\frac{1}{\sqrt{1-r^2/4}}, n, \delta\right) \delta(n-2) 2^{n-3} \\ & \quad B\left(\frac{1}{4}; \frac{n-2}{2}, \frac{(n-2)(m-1)}{2} + 1\right), \end{aligned}$$

where  $\bar{h}(x, n, \delta) = x^{n+\delta-5} x^{(n-2)(\delta-1)}$ .

- (b) When  $r \leq \begin{cases} 2\sqrt{1 - \sqrt{1 - 1/5}}, & \delta = 2, 3 \\ 2\sqrt{1 - \sqrt{1 - 1/\delta}}, & \delta \geq 4 \end{cases}$ , the upper bound in part (a) is upper bounded by

$$\begin{cases} \delta(n-2) \left\{ \left( \frac{\sqrt{5}}{2} \right)^{\frac{\delta-2}{2}} \left( \frac{2}{\sqrt{5}} \right)^{\frac{(n-2)(\delta-1)}{2}} \right. \\ \quad \left. + \left( 1 - \frac{2}{\sqrt{5}} \right) \left( \frac{\sqrt{5}}{2} \right)^{\frac{n+\delta-5}{2}} \left( \frac{3\sqrt{5}}{8} \right)^{\frac{(n-2)(\delta-1)}{2}} \right\}, & \delta = 2, 3, \\ \delta(n-2) \exp\left(\frac{1}{4}\right) \left( \sqrt{\frac{\delta}{\delta-1}} \right)^{\frac{\delta-2}{2}} \\ \quad \left( \sqrt{\frac{\delta-1}{\delta}} \right)^{\frac{(n-2)(\delta-1)}{2}} \left( \frac{4}{\sqrt{\delta(\delta-1)}} \right)^{\frac{n-3}{2}}, & \delta \geq 4. \end{cases}$$

- (c) When  $\rho > 3/5$ ,  $\sum_{\ell=2}^{\delta+1} \alpha(\ell, r_\rho)$  is upper bounded by the upper bound in part (a) with  $r$  replaced by  $r_\rho$ .

$$\text{When } \rho \geq \begin{cases} 4/\sqrt{5} - 1, & \delta = 2, 3 \\ 2\sqrt{1 - 1/\delta} - 1, & \delta \geq 4 \end{cases}, \sum_{\ell=2}^{\delta+1} \alpha(\ell, r_\rho)$$

is upper bounded by the upper bound in part (b) with  $r$  replaced by  $r_\rho$ .

Lemma V.4 (a) establishes an upper bound for the conditional probability that there are at least two universal vertices, conditioned on the existence of one universal vertex, in the random geometric graph over points generated by the uniform distribution on the sphere. The interpretations of Lemma V.4 (b), (c) are similar to those of Lemma V.1 (b) and  $\sum_{\ell=2}^{\delta+1} \alpha_\ell$ , and are therefore omitted.

The next lemma establishes approximations to the non-asymptotic compound Poisson distribution in Theorem III.11 and to its second moment, with the approximation errors upper bounded in terms of  $\sum_{\ell=2}^{\delta+1} \alpha(\ell, r_\rho)$ . Denote  $\bar{\lambda}_{p,n,\delta,\rho} := \binom{p}{1} \binom{p-1}{\delta} (2P_n(r_\rho))^\delta$ .

**Lemma V.5 (Approximation to the Non-Asymptotic Compound Poisson Distribution):** Consider  $n \geq 4$  and  $\delta \geq 2$ .

- (a) The increment distribution satisfies  $d_{TV}(\zeta_{n,\delta,\rho}, \text{Dirac}(1)) \leq \sum_{\ell=2}^{\delta+1} \alpha(\ell, r_\rho)$ .  
 (b) The arrival rate satisfies  $|\lambda_{p,n,\delta,\rho} - \bar{\lambda}_{p,n,\delta,\rho}| \leq \frac{3}{2} \bar{\lambda}_{p,n,\delta,\rho} \sum_{\ell=2}^{\delta+1} \alpha(\ell, r_\rho)$ .  
 (c) The non-asymptotic compound Poisson distribution satisfies

$$\begin{aligned} d_{TV}(\text{CP}(\lambda_{p,n,\delta,\rho}, \zeta_{n,\delta,\rho}), \text{Pois}(\bar{\lambda}_{p,n,\delta,\rho})) \\ \leq \frac{5}{2} \bar{\lambda}_{p,n,\delta,\rho} \sum_{\ell=2}^{\delta+1} \alpha(\ell, r_\rho). \end{aligned}$$

- (d) Let  $Z \sim \text{CP}(\lambda_{p,n,\delta,\rho}, \zeta_{n,\delta,\rho})$ . Then the second moment of the non-asymptotic compound Poisson  $\mathbb{E}Z^2$  in (42) satisfies

$$|\mathbb{E}Z^2 - (\bar{\lambda}_{p,n,\delta,\rho} + (\bar{\lambda}_{p,n,\delta,\rho})^2)| \leq \frac{3}{2} \bar{\lambda}_{p,n,\delta,\rho} \sum_{\ell=2}^{\delta+1} \alpha(\ell, r_\rho).$$

The proof of Lemma V.5 and its interpretation are analogous to those of Lemma V.2, and are therefore omitted. By combining Lemma V.5 (c) and Theorem III.11, we immediately obtain the following result on Poisson approximation to  $\mathcal{L}(\bar{N}_\delta)$  for finite  $p$ . One can derive an equivalent proof of the following proposition by first applying Poisson approximation

from Chen-Stein's method (e.g. [39, Theorem 1]) to obtain an upper bound  $d_{TV}(\mathcal{L}(N_{E_\delta}^{(\mathbf{R})}), \text{Pois}(\bar{\lambda}_{p,n,\delta,\rho}))$ , which is then combined with the portmanteau result Proposition III.9.

**Proposition V.6 (Poisson Approximation for Finite  $p$ ):** Suppose all the conditions in Theorem III.11 hold. Then  $\bar{\lambda}_{p,n,\delta,\rho} \leq \frac{\gamma_\delta}{\delta!}$  and

$$d_{TV}(\mathcal{L}(\bar{N}_\delta), \text{Pois}(\bar{\lambda}_{p,n,\delta,\rho})) \leq$$

$$C_{n,\delta,\gamma} \left( \mu_{n,2\delta+2}(\Sigma) \frac{\kappa}{p} + p^{-\frac{1}{\delta}} + E(p, \tau) \right) + \frac{5}{2} \frac{\gamma_\delta}{\delta!} \sum_{\ell=2}^{\delta+1} \alpha(\ell, r_\rho),$$

where  $E(p, \tau)$  is defined in Theorem III.11.

We have thus established that the limiting compound Poisson distribution  $\text{CP}(\lambda_{n,\delta}(e_{n,\delta}), \zeta_{n,\delta})$  in Theorem II.4 can be approximated by  $\text{Pois}(\frac{(e_{n,\delta})^\delta}{\delta!})$ , and that the non-asymptotic compound Poisson distribution  $\text{CP}(\lambda_{p,n,\delta,\rho}, \zeta_{n,\delta,\rho})$  in Theorem III.11 can be approximated by  $\text{Pois}(\binom{p}{1} \binom{p-1}{\delta} (2P_n(r_\rho))^\delta)$  for sufficiently large  $n$  or  $\delta$ . By combining these results with Theorem III.11 and Theorem II.4 we then obtain Poisson approximations to  $\mathcal{L}(\bar{N}_\delta)$ . In Section K of the Appendix, Figure 9 provides numerical simulations to demonstrate the accuracy of the Poisson approximation to  $\mathcal{L}(\bar{N}_\delta)$ . See Figure 9 and the ensuing discussions for more details.

#### B. Exact Formulae for $\alpha_\ell$ and $\alpha(\ell, \rho)$ When $\delta = 1$ and $\delta = 2$

In this subsection we provide analytical expressions for  $\alpha_\ell$  and  $\alpha(\ell, r_\rho)$  when  $\delta = 1$  and expressions for  $\alpha_\ell$  when  $\delta = 2$ .

**Example V.7 ( $\alpha_\ell$  and  $\alpha(\ell, r_\rho)$  When  $\delta = 1$ ):** When  $\delta = 1$ ,  $\alpha_2 = 1$  since in the random geometric graph  $\mathbf{Ge}(\{\tilde{u}_i\}_{i=1}^1; 1, n-2)$  the number of universal vertices (vertices of degree 0) is 1. In this case  $\alpha_1 = 1 - \alpha_2 = 0$ . Similarly,  $\alpha(2, r_\rho) = 1$  when  $\delta = 1$  since in the random pseudo geometric graph  $\mathbf{PGe}(\{\mathbf{u}'_i\}_{i=1}^2; r_\rho, n-1)$  the number of universal vertices is 2 as long as there exists one universal vertex. In this case  $\alpha(1, r_\rho) = 1 - \alpha(2, r_\rho) = 0$ .

**Remark V.8 (Compound Poisson Approximations When  $\delta = 1$ ):** Using the results for  $\alpha_\ell$  in Example V.7, we obtained  $\zeta_{n,1} = \text{Dirac}(2)$  and  $\lambda_{n,1}(e_{n,1}) = \frac{1}{2}e_{n,1}$ . Then the limiting compound Poisson distribution in Theorem II.4 is  $\text{CP}(\frac{1}{2}e_{n,1}, \text{Dirac}(2))$ . On the other hand, by the results for  $\alpha(\ell, r_\rho)$  in Example V.7,  $\zeta_{n,1,\rho} = \text{Dirac}(2)$  and  $\lambda_{p,n,1,\rho} = p(p-1)P_n(r_\rho)$ . Therefore the non-asymptotic compound Poisson distribution in Theorem III.11 is  $\text{CP}(p(p-1)P_n(r_\rho), \text{Dirac}(2))$ . The intuition for the result that the increment distributions  $\zeta_{n,1}$  and  $\zeta_{n,1,\rho}$  are  $\text{Dirac}(2)$  is as follows. Recall that  $N_{E_1}^{(\Psi)}$  is twice of the number of edges. Thus  $N_{E_1}^{(\Psi)}$  always increases by 2 whenever there is a new edge and  $N_{V_1}^{(\Psi)}$  always increases by 2 since the increment is always a new pair of vertices of degree 1. Note that  $N_{V_1}^{(\Psi)}$  has increment close to 2 since  $N_{V_1}^{(\Psi)} \approx N_{V_1}^{(\Psi)}$  by Lemma III.8. Given that the increment distributions  $\zeta_{n,1}$  and  $\zeta_{n,1,\rho}$  are  $\text{Dirac}(2)$ , we have the following equivalent Poisson approximations to  $\bar{N}_1/2$ :  $\mathcal{L}(\bar{N}_1/2) \approx \text{Pois}(p(p-1)P_n(r_\rho))$  and  $\bar{N}_1/2 \xrightarrow{D} \text{Pois}(\frac{1}{2}e_{n,1})$  as  $p \rightarrow \infty$  and  $a_n 2^{\frac{n}{2}} p^2 (1 - \rho)^{\frac{n-2}{2}} \rightarrow e_{n,1}$ .

As a comparison, Proposition 1 and its proof in [17] under row- $\kappa$  sparsity condition established that  $N_{E_1}^{(\mathbf{R})}/2$  converges to a Poisson distribution and obtained the limits of  $\mathbb{E}N_{V_1}^{(\mathbf{R})}$  and  $\mathbb{P}(N_{V_1}^{(\mathbf{R})} > 0)$ . Proposition 1 and Proposition 3 in [18] under block sparsity condition extended the results in [17] to the corresponding versions in the empirical partial correlation graph, i.e. the same conclusions hold with  $\mathbf{R}$  replaced by  $\mathbf{P}$ . Our results in Theorems III.11, II.4 (and Lemmas III.12, II.5) with  $\delta = 1$  characterize the full distributions of the 6 quantities  $\{N_{E_\delta}^{(\Psi)}, N_{V_\delta}^{(\Psi)}, N_{V_\delta}^{(\Psi)} : \Psi \in \{\mathbf{R}, \mathbf{P}\}\}$ , and our results in Section IV characterize their first and second moments, which together contain and extend the aforementioned previous results. Moreover, our results are established with weaker sparsity assumptions and provide concise formulae for the parameters.

We next explicitly characterize  $\alpha_\ell$  when  $\delta = 2$ .

**Lemma V.9** ( $\alpha_\ell$  When  $\delta = 2$ ): When  $\delta = 2$ ,  $\alpha_2 = 0$ ,  $\alpha_3 = \frac{3}{2}I_{\frac{3}{4}}(\frac{n-1}{2}, \frac{1}{2})$  and  $\alpha_1 = 1 - \alpha_3$ , where  $I_x(a, b)$  is the regularized incomplete Beta function.

**Remark V.10** (Limiting Compound Poisson Approximation When  $\delta = 2$ ): When  $\delta = 2$ , by Lemma V.9,  $\alpha_2 = 0$ ,  $\alpha_3 = \frac{3}{2}I_{\frac{3}{4}}(\frac{n-1}{2}, \frac{1}{2})$  and  $\alpha_1 = 1 - \alpha_3$ . Then  $\sum_{\ell=1}^3 \alpha_\ell/\ell = 1 - I_{\frac{3}{4}}(\frac{n-1}{2}, \frac{1}{2})$ . Thus, the parameters for  $\text{CP}(\lambda_{n,2}(e_{n,2}), \zeta_{n,2})$  in Theorem II.4 are

$$\begin{aligned}\zeta_{n,2}(1) &= \frac{1 - \frac{3}{2}I_{\frac{3}{4}}(\frac{n-1}{2}, \frac{1}{2})}{1 - I_{\frac{3}{4}}(\frac{n-1}{2}, \frac{1}{2})}, \quad \zeta_{n,2}(2) = 0, \\ \zeta_{n,2}(3) &= \frac{\frac{1}{2}I_{\frac{3}{4}}(\frac{n-1}{2}, \frac{1}{2})}{1 - I_{\frac{3}{4}}(\frac{n-1}{2}, \frac{1}{2})}\end{aligned}\quad (44)$$

and

$$\lambda_{n,2}(e_{n,2}) = \frac{1}{2}(e_{n,2})^2 \left(1 - I_{\frac{3}{4}}\left(\frac{n-1}{2}, \frac{1}{2}\right)\right).$$

Note that Proposition 1 in [18] states that for any fixed  $n$  and  $\delta$ ,  $\mathbb{P}(N_{V_\delta}^{(\mathbf{R})} > 0)$  converges to  $\mathbb{P}(N_\delta^* > 0)$  when  $p \rightarrow \infty$ , where  $N_\delta^*$  is the Poisson random variable specified in the proposition. Since the increment distribution  $\zeta_{n,2} \neq \text{Dirac}(1)$  by (44), the limit of  $\mathbb{P}(N_{V_\delta}^{(\mathbf{R})} > 0)$  is  $\mathbb{P}(\text{CP}(\lambda_{n,2}(e_{n,2}), \zeta_{n,2}) > 0)$  and hence [18, Proposition 1] is incorrect. However, as shown in Subsection V-A the result [18, Proposition 1] is still useful when  $n$  or  $\delta$  is large. Specifically, the distribution of each of the count variables  $\{N_{E_\delta}^{(\Psi)}, N_{V_\delta}^{(\Psi)}, N_{V_\delta}^{(\Psi)} : i \in \{\mathbf{R}, \mathbf{P}\}, \Psi \in \{\mathbf{R}, \mathbf{P}\}\}$  can be accurately approximated by a Poisson distribution but with an additional error term depending on  $n$  and  $\delta$ . This error term is small for large  $n$  or large  $\delta$  (see Proposition V.3 and Proposition V.6).

## VI. CONCLUSION AND DISCUSSIONS

In this paper, we studied the number of highly connected vertices in both the empirical correlation graph and the empirical partial correlation graph by adopting unified framework. More specifically, we showed that the distributions of the number of hubs  $N_{V_\delta}^{(\Psi)}$  or  $N_{V_\delta}^{(\Psi)}$  and the star subgraph counts  $N_{E_\delta}^{(\Psi)}$  are close to common compound Poisson, when  $p$  is finite and when  $p$  approaches infinity. We also establish that their

first and second moments are close to that of the compound Poisson distribution. The parameters in the compound Poisson distributions are characterized in terms of random geometric graphs and random pseudo geometric graphs. The parameters are also approximated by simple formulae, which implies that the approximating compound Poisson distributions can be further approximated by Poisson distributions for reasonably large sample size  $n$  or a reasonably hub degree  $\delta$ .

There are multiple avenues for future research. Numerical experiments suggest that the results in this paper hold beyond the  $(\tau, \kappa)$  sparsity condition. A future line of work is to characterize the compound Poisson approximations for weaker sparsity conditions. Generalization of the moment convergence results to beyond the second moments are also of interest.

## APPENDIX

### A. List of Symbols in the Main Text

- $\|\cdot\|_2$  : Euclidean norm.
- $|\cdot|$  : the cardinality of a set.
- $C$  and  $c$  : denote generic positive universal constants that often differ from line to line.  $C$  and  $c$  with subscripts are positive constants depending on the parameter in their subscripts and may differ from line to line.
- $\mathbf{X}$  :  $n \times p$  matrix of observations. It has density  $f_{\mathbf{X}}(\mathbf{X})$  given in (2).
- $\mathbf{R}$  :  $p \times p$  sample correlation matrix. Also denoted  $\Psi^{(\mathbf{R})}$ .
- $\mathbf{P}$  :  $p \times p$  sample partial correlation matrix. Also denoted  $\Psi^{(\mathbf{P})}$ .
- $\rho \in [0, 1]$  : screening threshold applied to elements of matrices  $\mathbf{R}$  or  $\mathbf{P}$ .
- $\Phi^{(\Psi)}$  : the binary indicator matrix of the non-zero entries of the thresholded sample correlation or sample partial correlation matrix  $\Psi$  with entries  $\Phi_{ij}^{(\Psi)} = \Phi_{ij}^{(\Psi)}(\rho) = 1(|\Psi_{ij}| \geq \rho)$ , where  $\Psi \in \{\mathbf{R}, \mathbf{P}\}$ .
- $\mathcal{G}_\rho(\Psi)$  : the graph associated with adjacency matrix  $\Phi^{(\Psi)}$ . Specifically, it is the empirical correlation graph when  $\Psi = \mathbf{R}$  or, equivalently,  $\Psi = \mathbf{R}$  and it is the empirical partial correlation graph when  $\Psi = \mathbf{P}$  or, equivalently,  $\Psi = \mathbf{P}$  (see Subsection II-A).
- $N_{V_\delta}^{(\Psi)}$  : number of vertices of degree at least  $\delta$  in  $\mathcal{G}_\rho(\Psi)$ .
- $N_{\tilde{V}_\delta}^{(\Psi)}$  : number of vertices of degree exactly  $\delta$  in  $\mathcal{G}_\rho(\Psi)$ . The quantity  $\{N_{\tilde{V}_\delta}^{(\Psi)}\}_{\Psi=0}^{n-1}$  is the empirical vertex degree distribution associated with the sample correlation graph ( $\Psi = \mathbf{R}$ ) or partial correlation graph ( $\Psi = \mathbf{P}$ ).
- $\Gamma_\delta$  : a star graph with  $\delta$  edges.
- $N_{E_\delta}^{(\Psi)}$  : the number of subgraphs isomorphic to  $\Gamma_\delta$  in  $\mathcal{G}_\rho(\Psi)$  when  $\delta \geq 2$ .  $N_{E_\delta}^{(\Psi)}$  is defined as twice the number of edges in  $\mathcal{G}_\rho(\Psi)$  when  $\delta = 1$ .
- $\bar{N}_\delta$  : a generic random variable denoting any of the quantities  $\{N_{E_\delta}^{(\Psi)}, N_{V_\delta}^{(\Psi)}, N_{V_\delta}^{(\Psi)} : \Psi \in \{\mathbf{R}, \mathbf{P}\}\}$ .
- $\mathbf{U} = [\mathbf{u}_1, \dots, \mathbf{u}_p] : (n-1) \times p$  matrix of correlation U-scores such that  $\mathbf{U}^\top \mathbf{U} = \mathbf{R}$ .
- $\mathbf{Y} = [\mathbf{y}_1, \dots, \mathbf{y}_p] : (n-1) \times p$  matrix of partial correlation Y-scores such that  $\mathbf{Y}^\top \mathbf{Y} = \mathbf{P}$ .
- $r_\rho = \sqrt{2(1-\rho)}$  : spherical cap radius parameter in terms of  $\rho$ .

- $S^{n-2}$  : unit sphere in  $\mathbb{R}^{n-1}$ , i.e.  $S^{n-2} = \{\mathbf{w} \in \mathbb{R}^{n-1} : \|\mathbf{w}\|_2 = 1\}$ . Its area is denoted by  $\text{Area}(S^{n-2})$  or  $|S^{n-2}|$ .
- $B^{n-2}$  : unit ball in  $\mathbb{R}^{n-2}$ , i.e.  $B^{n-2} = \{\mathbf{w} \in \mathbb{R}^{n-2} : \|\mathbf{w}\|_2 \leq 1\}$ . Its volume is denoted by  $|B^{n-2}|$ .
- $\text{unif}(S^{n-2})$  : uniform distribution on  $S^{n-2}$ .
- $\text{unif}(B^{n-2})$  : uniform distribution in  $B^{n-2}$ .
- $P_n(r)$  : normalised area of a spherical cap on  $S^{n-2}$  with radius  $r$  (see (26)).
- $\text{Ge}(\{\mathbf{v}_i\}_{i=1}^\delta, r)$  : the geometric graph generated by a set of points  $\{\mathbf{v}_i\}_{i=1}^\delta \subset \mathbb{R}^{n-2}$  with ball radius parameter  $r$  (see Subsection II-A).
- Universal vertex : a vertex of an undirected graph that is adjacent to all other vertices
- $\text{NUV}(\{\mathbf{v}_i\}_{i=1}^\delta, r)$  : the number of universal vertices in  $\text{Ge}(\{\mathbf{v}_i\}_{i=1}^\delta, r)$ .
- $\text{PGe}(\{\mathbf{v}_i\}_{i=1}^m, r; \mathcal{N})$  : the pseudo geometric graph generated by  $\{\mathbf{v}_i\}_{i=1}^m \subset \mathbb{R}^{\mathcal{N}}$  with radius parameter  $r$  (see Definition III.5).
- $\text{PNUV}(\{\mathbf{v}_i\}_{i=1}^m, r; \mathcal{N})$  : the number of universal vertices in  $\text{PGe}(\{\mathbf{v}_i\}_{i=1}^m, r; \mathcal{N})$ .
- $\{\mathbf{u}'_i\}_i$  : a sequence of i.i.d. random vectors drawn from  $\text{unif}(S^{n-2})$ , the uniform distribution on  $S^{n-2}$ .
- $\alpha(\ell, r_\rho) = \alpha_{n,\delta}(\ell, r_\rho)$  : the conditional probability that there are  $\ell$  universal vertices in  $\text{PGe}(\{\mathbf{u}'_i\}_{i=1}^{\delta+1}, r_\rho)$  (see (23)).
- $\zeta_{n,\delta,\rho}$  : the discrete distribution supported on  $[\delta+1]$  with  $\zeta_{n,\delta,\rho}(\ell) \propto \alpha(\ell, r_\rho)/\ell$  (see (24)).
- $\text{CP}(\lambda, \zeta)$  : the compound Poisson distribution, where the arrival rate  $\lambda$  is the rate for the underlying Poisson random variable and the increment distribution  $\zeta$  is the distribution of each increment.
- $\text{CP}(\lambda_{p,n,\delta,\rho}, \zeta_{n,\delta,\rho})$  : the non-asymptotic compound Poisson distribution in Theorem III.11.
- $\xrightarrow{\mathcal{D}}$  : denotes convergence in distribution.
- $\lambda_{p,n,\delta,\rho}$  : a quantity defined as  $\binom{p-1}{\delta} (2P_n(r_\rho))^\delta \sum_{\ell=1}^{\delta+1} \frac{\alpha(\ell, r_\rho)}{\ell}$ . It is the arrival rate of  $\text{CP}(\lambda_{p,n,\delta,\rho}, \zeta_{n,\delta,\rho})$ .
- $\{\tilde{\mathbf{u}}_i\}_i$  : a sequence of i.i.d. random vectors drawn from  $\text{unif}(B^{n-2})$ , the uniform distribution in  $B^{n-2}$ .
- $\alpha_\ell$  : the probability that there are exactly  $\ell-1$  universal vertices in  $\text{Ge}(\{\tilde{\mathbf{u}}_i\}_{i=1}^\delta, 1)$  (see (6)).  $\alpha_\ell = \lim_{\rho \rightarrow 1} \alpha(\ell, r_\rho)$  (see Lemma III.15 and the paragraph beneath it).
- $\zeta_{n,\delta}$  : the discrete distribution supported on  $[\delta+1]$  with  $\zeta_{n,\delta}(\ell) \propto \alpha_\ell/\ell$  (see (7)).  $\zeta_{n,\delta}(\ell) = \lim_{\rho \rightarrow 1} \zeta_{n,\delta,\rho}(\ell)$  (see Lemma III.16).
- $\lambda_{n,\delta}(e_{n,\delta})$  : a quantity defined as  $\frac{1}{\delta!} (e_{n,\delta})^\delta \sum_{\ell=1}^{\delta+1} \frac{\alpha_\ell}{\ell}$ .  $\lambda_{n,\delta}(e_{n,\delta}) = \lim_{p \rightarrow \infty} \lambda_{p,n,\delta,\rho}$  when  $\rho$  satisfies  $a_n 2^{\frac{n}{2}} p^{1+\frac{1}{\delta}} (1-\rho)^{\frac{n-2}{2}} \rightarrow e_{n,\delta}$ , where  $a_n = \frac{\Gamma((n-1)/2)}{(n-2)\sqrt{\pi}\Gamma((n-2)/2)}$  and  $e_{n,\delta}$  is some positive constant (see Lemma III.16).
- $\text{CP}(\lambda_{n,\delta}, \zeta_{n,\delta})$  : the limiting compound Poisson distribution in Theorem II.4. Moreover,  $\text{CP}(\lambda_{p,n,\delta,\rho}, \zeta_{n,\delta,\rho}) \xrightarrow{\mathcal{D}} \text{CP}(\lambda_{n,\delta}(e_{n,\delta}), \zeta_{n,\delta})$ , when  $p \rightarrow \infty$  and  $\rho$  satisfies  $a_n 2^{\frac{n}{2}} p^{1+\frac{1}{\delta}} (1-\rho)^{\frac{n-2}{2}} \rightarrow e_{n,\delta}$ . (see Lemma III.16)
- $\mu(\mathbf{A})$  : normalized determinant of square matrix  $\mathbf{A}$  (see Definition II.3).

- $\mu_m(\mathbf{A})$  : local normalized determinant of degree  $m$  (see the paragraph after Definition II.3).
- $\mu_{n,m}(\mathbf{A})$  : inverse local normalized determinant. It is powers of  $\mu_m(\mathbf{A})$  (see (8)).
- $C_\delta^<$  : the index set defined as  $\{\vec{i} = (i_0, i_1, \dots, i_\delta) \in [p]^{\delta+1} : i_1 < i_2 < \dots < i_\delta, \text{ and } i_\ell \neq i_0, \forall \ell \in [\delta]\}$ . Each  $\vec{i} \in C_\delta^<$  corresponds to a group of  $\delta$  vertices indexed by  $\{i_j\}_{j=1}^\delta$  and a center indexed by  $i_0$ .
- $\Phi_{\vec{i}}^{(\mathbf{R})}$  : defined as  $\prod_{j=1}^\delta \Phi_{i_0 i_j}^{(\mathbf{R})}$ . It is the indicator random variable that vertex  $i_0$  is adjacent to each vertex  $i_j$  for  $j \in [\delta]$  in  $\mathcal{G}_\rho(\mathbf{R})$ . Equivalently, it is the indicator function of the event that there exist a star subgraph in  $\mathcal{G}_\rho(\mathbf{R})$  with center at  $i_0$  and leaves  $\{i_j\}_{j=1}^\delta$ .
- $S_{\vec{i}}$  : defined as  $\{\vec{j} \in C_\delta^< \setminus \{\vec{i}\} : \bigcup_{\ell=0}^\delta \{j_\ell\} = \bigcup_{\ell=0}^\delta \{i_\ell\}\}$  for  $\vec{i} \in C_\delta^<$ . It is the set of indices in  $C_\delta^<$  that share the same vertices with  $\vec{i}$  but with center different from  $i_0$ .
- $U_{\vec{i}}$  : defined as  $\sum_{\vec{j} \in S_{\vec{i}}} \Phi_{\vec{j}}^{(\mathbf{R})}$ . It is the sum of highly dependent terms of  $\Phi_{\vec{i}}^{(\mathbf{R})}$ . (see the discussions preceding and following (50)).

## B. Controlling Local Normalized Determinant by Extreme Eigenvalues

**Lemma A.1:** Let  $\mathbf{A}$  be a symmetric positive definite matrix. Consider a sequence of symmetric positive definite matrices  $\Sigma \in \mathbb{R}^{p \times p}$  with increasing dimension  $p$ .

- (a)  $\mu_{n,m}(\mathbf{A})$  is bounded by powers of the largest local condition number:

$$\mu_{n,m}(\mathbf{A}) \leq \begin{cases} \max_{\mathcal{I} \subset [p]} \left( \frac{\lambda_{\max}(\mathbf{A}_{\mathcal{I}})}{\lambda_{\min}(\mathbf{A}_{\mathcal{I}})} \right)^{\frac{m(n-1)}{2}}, & \mathbf{A} \text{ not diagonal,} \\ 1, & \mathbf{A} \text{ diagonal.} \end{cases}$$

- (b)  $\mu_{n,m}(\mathbf{A})$  is bounded by powers of the condition number:

$$\mu_{n,m}(\mathbf{A}) \leq \begin{cases} \left( \frac{\lambda_{\max}(\mathbf{A})}{\lambda_{\min}(\mathbf{A})} \right)^{\frac{m(n-1)}{2}}, & \mathbf{A} \text{ not diagonal,} \\ 1, & \mathbf{A} \text{ diagonal.} \end{cases}$$

- (c) If  $\lambda_{\min}(\Sigma) \geq \underline{\lambda}$  and  $\lambda_{\max}(\Sigma) \leq \bar{\lambda}$  for all  $p$ , then

$$\mu_{n,m}(\Sigma) \leq \begin{cases} \left( \frac{\bar{\lambda}}{\underline{\lambda}} \right)^{\frac{m(n-1)}{2}}, & \Sigma \text{ not diagonal,} \\ 1, & \Sigma \text{ diagonal.} \end{cases}$$

- (d) Let  $M > 0$  be a constant. Suppose for all  $p$ ,  $\sup_{1 \leq i \leq p} \Sigma_{ii} \leq M$ . Moreover suppose  $\lambda_{\min}(\Sigma) \geq \underline{\lambda}$  for all  $p$ . Then

$$\mu_{n,m}(\Sigma) \leq \begin{cases} \left( \frac{Mm}{\underline{\lambda}} \right)^{\frac{m(n-1)}{2}}, & \Sigma \text{ not diagonal,} \\ 1, & \Sigma \text{ diagonal.} \end{cases}$$

*Proof:* (a) It follows directly by definition of  $\mu_{n,m}(\mathbf{A})$  and  $\mu(\mathbf{A}_{\mathcal{I}}) \geq \left( \frac{\lambda_{\min}(\mathbf{A}_{\mathcal{I}})}{\lambda_{\max}(\mathbf{A}_{\mathcal{I}})} \right)^m$ .

(b) Since  $\mu_{n,m}(\mathbf{A})$  is increasing in  $m$  as discussed after Example II.9,  $\mu_{n,m}(\mathbf{A}) \leq \mu_{n,p}(\mathbf{A})$ . The proof is then complete by applying (a) to  $\mu_{n,p}(\mathbf{A})$ .

(c) It follows directly from (b).

(d) Let  $\mathcal{I} \subset [p]$  with  $|\mathcal{I}| = m$ . Since  $\Sigma_{\mathcal{I}}$  is symmetric positive definite,  $|(\Sigma_{\mathcal{I}})_{ij}| \leq \sqrt{(\Sigma_{\mathcal{I}})_{ii}(\Sigma_{\mathcal{I}})_{jj}} \leq M$  and thus  $\lambda_{\max}(\Sigma_{\mathcal{I}}) = \|\Sigma_{\mathcal{I}}\|_2 \leq \|\Sigma_{\mathcal{I}}\|_F \leq Mm$ . By the interlacing property (see Theorem 8.1.7 in [36]),  $\lambda_{\min}(\Sigma_{\mathcal{I}}) \geq \lambda_{\min}(\Sigma)$ . The proof is then completed by applying part (b).  $\square$

### C. Proofs in Subsection III-A

#### 1) Proofs of Lemma III.1 and Lemma III.2:

*Proof of Lemma III.1:* It suffices to prove that  $\mathbf{U}$  has rank  $n-1$  a.s. By Lemma III.2 (a), it suffices to prove that  $\left[\frac{\tilde{\mathbf{x}}_1}{\|\tilde{\mathbf{x}}_1\|_2}, \dots, \frac{\tilde{\mathbf{x}}_p}{\|\tilde{\mathbf{x}}_p\|_2}\right]$  has rank  $n-1$  a.s. Note that  $\tilde{\mathbf{X}}$  is of rank  $n-1$  a.s., since it has a density with respect to Lebesgue measure on  $\mathbb{R}^{(n-1)p}$  and  $p \geq n$ . Then  $\left[\frac{\tilde{\mathbf{x}}_1}{\|\tilde{\mathbf{x}}_1\|_2}, \dots, \frac{\tilde{\mathbf{x}}_p}{\|\tilde{\mathbf{x}}_p\|_2}\right]$  have rank  $n-1$  a.s. as well since it is obtained by normalizing the columns of  $\tilde{\mathbf{X}}$ .  $\square$

*Proof of Lemma III.2:* (a) Consider  $\hat{\mathbf{x}}^{(i)} \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(\mathbf{0}, \Sigma)$ . Then  $\hat{\mathbf{X}} := [\hat{\mathbf{x}}^{(1)}, \dots, \hat{\mathbf{x}}^{(n)}]^\top = [\hat{\mathbf{x}}_1, \dots, \hat{\mathbf{x}}_p] \in \mathbb{R}^{n \times p}$  has the distribution  $\mathcal{VE}(\mathbf{0}, \Sigma, g_0)$  where  $g_0(w) = (2\pi)^{-\frac{np}{2}} \exp(-\frac{1}{2}w)$ . The sample mean  $\hat{\mathbf{x}}$  of  $\hat{\mathbf{X}}$  is given as a row vector

$$\hat{\mathbf{x}} = \frac{1}{n} \sum_{i=1}^n \hat{\mathbf{x}}^{(i)} = \frac{1}{n} \hat{\mathbf{X}}^\top \mathbf{1} = (\hat{x}_1, \dots, \hat{x}_p),$$

where  $\hat{x}_i := n^{-1} \sum_{j=1}^p \hat{X}_{ij}$ . The sample covariance matrix  $\hat{\mathbf{S}}$  of  $\hat{\mathbf{X}}$  is

$$\begin{aligned} \hat{\mathbf{S}} &= \frac{1}{n-1} \sum_{i=1}^n (\hat{\mathbf{x}}^{(i)} - \hat{\mathbf{x}})(\hat{\mathbf{x}}^{(i)} - \hat{\mathbf{x}})^\top \\ &= \frac{1}{n-1} \hat{\mathbf{X}}^\top \left( \mathbf{I} - \frac{1}{n} \mathbf{1}\mathbf{1}^\top \right) \hat{\mathbf{X}}. \end{aligned}$$

The Z-scores of  $\hat{\mathbf{X}}$  is

$$\hat{\mathbf{z}}_i := \frac{\hat{\mathbf{x}}_i - \mathbf{1}\hat{x}_i}{\sqrt{\hat{S}_{ii}(n-1)}} = \frac{(\mathbf{I} - \frac{1}{n} \mathbf{1}\mathbf{1}^\top) \hat{\mathbf{x}}_i}{\sqrt{\hat{S}_{ii}(n-1)}}.$$

Recall  $\mathbf{H}_{2:n} \in \mathbb{R}^{n \times (n-1)}$  is a matrix such that  $n \times (n-1)$  matrix  $\mathbf{H} = [n^{-\frac{1}{2}} \mathbf{1}, \mathbf{H}_{2:n}]^\top$  is orthogonal, which satisfies  $\mathbf{H}_{2:n} \mathbf{H}_{2:n}^\top = \mathbf{I} - \frac{1}{n} \mathbf{1}\mathbf{1}^\top$ . The U-scores of  $\hat{\mathbf{X}}$  is denoted  $\hat{\mathbf{U}} = [\hat{\mathbf{u}}_1, \dots, \hat{\mathbf{u}}_p]$  with  $\hat{\mathbf{u}}_i = \mathbf{H}_{2:n}^\top \hat{\mathbf{z}}_i$  for  $i \in [p]$ .

By the theorem in Section 6 of [34], the distribution of  $\mathbf{U}$  is invariant to  $g$  and  $\mu$ . In particular,  $\mathbf{U}$  has the same distribution as  $\hat{\mathbf{U}}$ . It suffices to study the distribution of  $\hat{\mathbf{U}}$ .

The proof of Theorem 3.3.2 in [40] establishes that

$$\tilde{\mathbf{X}} = \mathbf{H}_{2:n}^\top \hat{\mathbf{X}} \in \mathbb{R}^{(n-1) \times p}$$

has i.i.d. rows  $\{\tilde{\mathbf{x}}^{(i)}\}_{i=1}^{n-1}$  distributed as  $\mathcal{N}(\mathbf{0}, \Sigma)$ , and

$$\begin{aligned} \hat{\mathbf{S}} &= \frac{1}{n-1} \hat{\mathbf{X}}^\top \left( \mathbf{I} - \frac{1}{n} \mathbf{1}\mathbf{1}^\top \right) \hat{\mathbf{X}} \\ &= \frac{1}{n-1} \left( \mathbf{H}_{2:n}^\top \hat{\mathbf{X}} \right)^\top \mathbf{H}_{2:n}^\top \hat{\mathbf{X}} \\ &= \frac{1}{n-1} \tilde{\mathbf{X}}^\top \tilde{\mathbf{X}}. \end{aligned}$$

Then  $\hat{S}_{ii} = \|\tilde{\mathbf{x}}_i\|^2/(n-1)$  where  $\tilde{\mathbf{x}}_i$  is the  $i$ -th column of  $\tilde{\mathbf{X}}$ . Thus

$$\hat{\mathbf{u}}_i = \mathbf{H}_{2:n}^\top \frac{(\mathbf{I} - \frac{1}{n} \mathbf{1}\mathbf{1}^\top) \hat{\mathbf{x}}_i}{\sqrt{\hat{S}_{ii}(n-1)}} = \frac{\mathbf{H}_{2:n}^\top \mathbf{H}_{2:n} \mathbf{H}_{2:n}^\top \hat{\mathbf{x}}_i}{\sqrt{\|\tilde{\mathbf{x}}_i\|^2}} = \frac{\tilde{\mathbf{x}}_i}{\|\tilde{\mathbf{x}}_i\|_2}.$$

The proof is then complete by observing that  $\tilde{\mathbf{X}}$  and  $\tilde{\mathbf{X}}$  has the same distribution.

(b) Part (b) directly follows from part (a).  $\square$

2) *Proof of Lemma III.4:* Part (a) of Lemma III.4 follows from  $\mu_m(\Sigma_{\mathcal{J}}) = \mu(\Sigma_{\mathcal{J}}) \geq \mu_m(\Sigma)$ . Part (c) immediately follows from part (b) and in the remainder of this subsection we give the proof of part (b).

It suffices to prove the following statement:  $\forall \mathbf{v}_i \in S^{n-2}, \forall i \in [m]$ ,

$$f_{\mathbf{u}_{j_1}, \mathbf{u}_{j_2}, \dots, \mathbf{u}_{j_m}}(\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_m) \leq \mu_{n,m}(\Sigma_{\mathcal{J}}).$$

For notation convenience, we only present the proof for  $m = p$  and  $\mathcal{J} = [p]$  since the proof of the general  $m$  and  $\mathcal{J}$  follows the same proof procedure. When  $m = p$  and  $\mathcal{J} = [p]$ , the statement of Lemma III.4 (b) becomes:

The joint density of columns of U-scores w.r.t.  $\otimes^p \sigma^{n-1}$  is upper bounded by  $\mu_{n,p}(\Sigma)$ :

$$f_{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_p}(\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_p) \leq \mu_{n,p}(\Sigma), \quad \forall \mathbf{v}_i \in S^{n-2}, \quad \forall i \in [p]. \quad (45)$$

*Proof of (45):* Recall  $\{\tilde{\mathbf{x}}^{(i)}\}_{i=1}^{n-1} \subset \mathbb{R}^p$ , the rows of  $\tilde{\mathbf{X}}$ , are i.i.d. copy of  $\mathcal{N}(\mathbf{0}, \Sigma)$ ;  $\{\tilde{\mathbf{x}}_i\}_{i=1}^p$  are the columns of  $\tilde{\mathbf{X}}$ ;  $\mathbf{u}_i := \frac{\tilde{\mathbf{x}}_i}{\|\tilde{\mathbf{x}}_i\|_2} \in \mathbb{R}^{n-1}$  has distribution  $\text{unif}(S^{n-2})$  for  $i \in [p]$ .

When  $\Sigma$  is symmetric positive definite and diagonal,  $\{\tilde{\mathbf{x}}_i\}_{i=1}^p$  are independent, which implies that  $\{\mathbf{u}_i\}_{i=1}^p$  are independent. Thus in this case, the joint density of columns of U-scores w.r.t.  $\otimes^p \sigma^{n-1}$  is 1.

Consider general symmetric positive definite  $\Sigma$ . The probability density of  $\tilde{\mathbf{X}}$  w.r.t. the Lebesgue measure on  $\mathbb{R}^{(n-1)p}$  is

$$f_{\tilde{\mathbf{X}}}(\tilde{\mathbf{X}}) = A \exp \left( -\frac{1}{2} \sum_{j=1}^{n-1} \left( \tilde{\mathbf{x}}^{(j)} \right)^\top \Sigma^{-1} \tilde{\mathbf{x}}^{(j)} \right),$$

where the constant  $A = \det(\Sigma)^{-\frac{n-1}{2}} (2\pi)^{-\frac{(n-1)p}{2}}$ . Use the spherical transform for each column  $\tilde{\mathbf{x}}_i = (\tilde{\mathbf{X}}_{ji} : 1 \leq j \leq n-1)^\top$  for  $1 \leq i \leq p$ ,

$$\begin{cases} \tilde{\mathbf{X}}_{1i} = R_i \cos(\theta_{1i}), \\ \tilde{\mathbf{X}}_{2i} = R_i \sin(\theta_{1i}) \cos(\theta_{2i}), \\ \vdots \\ \tilde{\mathbf{X}}_{(n-2)i} = R_i \sin(\theta_{1i}) \sin(\theta_{2i}) \cdots \sin(\theta_{(n-3)i}) \cos(\theta_{(n-2)i}), \\ \tilde{\mathbf{X}}_{(n-1)i} = R_i \sin(\theta_{1i}) \sin(\theta_{2i}) \cdots \sin(\theta_{(n-3)i}) \sin(\theta_{(n-2)i}), \end{cases}$$

where for each  $i \in [p]$ :  $R_i \geq 0, \theta_{ji} \in [0, \pi]$  for  $1 \leq j \leq n-3$  and  $\theta_{(n-2)i} \in [0, 2\pi)$ .

Denote  $\mathbf{R} = (R_i : 1 \leq i \leq p)$  and  $\mathbf{\Theta} = (\theta_{ji} : 1 \leq i \leq p, 1 \leq j \leq (n-2))$ . Then the joint density of  $(\mathbf{R}, \mathbf{\Theta})$  is:

$$f_{\mathbf{R}, \mathbf{\Theta}}(\mathbf{R}, \mathbf{\Theta}) = A e^{-\frac{1}{2} \sum_{j=1}^{n-1} (\mathbf{h}^{(j)})^\top \mathbf{\Sigma}^{-1} \mathbf{h}^{(j)}} \prod_{i=1}^p \left( R_i^{n-2} \prod_{j=1}^{n-2} \sin^{n-2-j}(\theta_{ji}) \right),$$

where for  $1 \leq j \leq n-2$ ,

$$\mathbf{h}^{(j)} = \left( R_i \cos(\theta_{ji}) \prod_{q=1}^{j-1} \sin(\theta_{qi}) : 1 \leq i \leq p \right)^\top \in \mathbb{R}^p$$

and

$$\mathbf{h}^{(n-1)} = \left( R_i \prod_{q=1}^{n-2} \sin(\theta_{qi}) : 1 \leq i \leq p \right)^\top \in \mathbb{R}^p.$$

Then the density of  $\mathbf{\Theta}$  is:

$$\begin{aligned} f_{\mathbf{\Theta}}(\mathbf{\Theta}) &= A \left( \prod_{i=1}^p \prod_{j=1}^{n-2} \sin^{n-2-j}(\theta_{ji}) \right) \int_{[0, \infty)^p} e^{-\frac{1}{2} \sum_{j=1}^{n-1} (\mathbf{h}^{(j)})^\top \mathbf{\Sigma}^{-1} \mathbf{h}^{(j)}} \prod_{i=1}^p R_i^{n-2} dR_i \\ &\leq A \left( \prod_{i=1}^p \prod_{j=1}^{n-2} \sin^{n-2-j}(\theta_{ji}) \right) \int_{[0, \infty)^p} \exp \left( -\frac{1}{2} \lambda_{\min}(\mathbf{\Sigma}^{-1}) \sum_{j=1}^{n-1} \|\mathbf{h}^{(j)}\|_2^2 \right) \prod_{i=1}^p R_i^{n-2} dR_i \\ &= A \left( \prod_{i=1}^p \prod_{j=1}^{n-2} \sin^{n-2-j}(\theta_{ji}) \right) \int_{[0, \infty)^p} \exp \left( -\frac{1}{2} [\lambda_{\max}(\mathbf{\Sigma})]^{-1} \sum_{i=1}^p R_i^2 \right) \prod_{i=1}^p R_i^{n-2} dR_i \\ &= A \left( \prod_{i=1}^p \prod_{j=1}^{n-2} \sin^{n-2-j}(\theta_{ji}) \right) \left( \int_{[0, \infty)} \exp \left( -\frac{1}{2} [\lambda_{\max}(\mathbf{\Sigma})]^{-1} R_1^2 \right) R_1^{n-2} dR_1 \right)^p \\ &\stackrel{(m)}{=} A \left( \prod_{i=1}^p \prod_{j=1}^{n-2} \sin^{n-2-j}(\theta_{ji}) \right) \left( [\lambda_{\max}(\mathbf{\Sigma})]^{\frac{n-1}{2}} \Gamma \left( \frac{n-1}{2} \right) 2^{\frac{n-3}{2}} \right)^p \\ &\stackrel{(mm)}{=} \left[ \frac{(\lambda_{\max}(\mathbf{\Sigma}))^p}{\det(\mathbf{\Sigma})} \right]^{\frac{n-1}{2}} \frac{1}{(\text{Area}(S^{n-2}))^p} \prod_{i=1}^p \prod_{j=1}^{n-2} \sin^{n-2-j}(\theta_{ji}), \end{aligned}$$

where equality (m) follows from the integration of Chi distribution with degree  $n-1$ , and equality (mm) follows from  $|S^{(n-2)}| = 2\pi^{\frac{n-1}{2}}/\Gamma((n-1)/2)$ . The proof is complete by noticing  $f_{\mathbf{\Theta}}(\mathbf{\Theta})$  is joint density of columns of U-scores expressed in spherical coordinate and

$\frac{1}{(\text{Area}(S^{n-2}))^p} \prod_{i=1}^p \prod_{j=1}^{n-2} \sin^{n-2-j}(\theta_{ji})$  is the joint distribution of  $p$  independent  $\text{unif}(S^{n-2})$  expressed as spherical coordinates.  $\square$

#### D. An Informal Derivation of Proposition III.6

In this subsection we provide motivations to the following two questions: 1) why is the distribution of the star subgraph counts  $N_{E_\delta}^{(\mathbf{R})}$  approximately compound Poisson? 2) what are the associated parameters of the compound Poisson approximation to  $N_{E_\delta}^{(\mathbf{R})}$ ?

We first introduce some notation to answer the first question. Denote

$$C_\delta^< := \{\vec{i} = (i_0, i_1, \dots, i_\delta) \in [p]^{\delta+1} : i_1 < i_2 < \dots < i_\delta, \text{ and } i_\ell \neq i_0, \forall \ell \in [\delta]\}. \quad (46)$$

Each  $\vec{i} \in C_\delta^<$  corresponds to a group of  $\delta$  vertices indexed by  $\{i_j\}_{j=1}^\delta$  and a center indexed by  $i_0$ . For  $\vec{i} \in C_\delta^<$ , denote by

$$\Phi_{\vec{i}}^{(\mathbf{R})} = \prod_{j=1}^\delta \Phi_{i_0 i_j}^{(\mathbf{R})} = 1 \left( \bigcap_{j=1}^\delta \{\text{dist}(\mathbf{u}_{i_0}, \mathbf{u}_{i_j}) \leq r_\rho\} \right) \quad (47)$$

the indicator function of the event that vertex  $i_0$  is connected to each vertex  $i_j$  for  $j \in [\delta]$  in the empirical correlation graph. Equivalently, when  $\delta \geq 2$ ,  $\Phi_{\vec{i}}^{(\mathbf{R})}$  is the indicator function of the event that there exists a star subgraph in  $\mathcal{G}_\rho(\mathbf{R})$  with center at  $i_0$  and with leaves  $\{i_j\}_{j=1}^\delta$ . Then by definition

$$N_{E_\delta}^{(\mathbf{R})} = \sum_{\vec{i} \in C_\delta^<} \Phi_{\vec{i}}^{(\mathbf{R})}. \quad (48)$$

If  $\Phi_{\vec{i}}^{(\mathbf{R})}$  for different  $\vec{i} \in C_\delta^<$  are independent or weakly dependent, then the distribution of  $N_{E_\delta}^{(\mathbf{R})}$ , as a sum of independent or weakly dependent indicator random variables, might be expected to be approximately Poisson. This, however, is not the case due to high dependency among many terms in the summation. Specifically, for any  $\vec{i} \in C_\delta^<$ ,  $\Phi_{\vec{i}}^{(\mathbf{R})}$  is highly dependent on  $\Phi_{\vec{j}}^{(\mathbf{R})}$  for any  $\vec{j} \in S_{\vec{i}}$  where

$$S_{\vec{i}} := \left\{ \vec{j} \in C_\delta^< \setminus \{\vec{i}\} : \bigcup_{\ell=0}^\delta \{j_\ell\} = \bigcup_{\ell=0}^\delta \{i_\ell\} \right\}. \quad (49)$$

$S_{\vec{i}}$  is the set of indexes sharing the same vertices with  $\vec{i}$  but with center different from  $i_0$  and thus  $|S_{\vec{i}}| = \delta$ . Indeed, provided  $\Phi_{\vec{i}}^{(\mathbf{R})} = 1$ , which is equivalent to  $\text{dist}(\mathbf{u}_{i_0}, \mathbf{u}_{i_j}) \leq r_\rho$  for  $\forall j \in [\delta]$ , we have: for  $\forall 2 \leq j \leq \delta$ ,

$$\text{dist}(\mathbf{u}_{i_1}, \mathbf{u}_{i_j}) \leq \text{dist}(\mathbf{u}_{i_1}, \mathbf{u}_{i_0}) + \text{dist}(\mathbf{u}_{i_0}, \mathbf{u}_{i_j}) \leq 2r_\rho.$$

That is,  $\{\mathbf{u}_{i_j}\}_{j=2}^{\delta+1}$  are all close to  $\mathbf{u}_{i_1}$  and hence it is likely there are edges connecting them. In other words it is likely for  $\vec{i} = (i_1, i_0, \dots, i_\delta)$ ,  $\Phi_{\vec{i}}^{(\mathbf{R})} = 1$ , where we without loss of generality assume  $i_0 < i_j$  for  $2 \leq j \leq \delta$ . Let

$$U_{\vec{i}} = \sum_{\vec{j} \in S_{\vec{i}}} \Phi_{\vec{j}}^{(\mathbf{R})} \quad (50)$$

be the sum of highly dependent terms of  $\Phi_{\vec{j}}^{(\mathbf{R})}$ . In summary, if there is an increment for  $N_{E_\delta}^{(\mathbf{R})}$ , say  $\Phi_{\vec{i}}^{(\mathbf{R})} = 1$ , there is a

certain probability that  $U_{\vec{i}}$  is greater than 0 due to the high dependence, causing each increment of  $N_{E_\delta}^{(R)}$  to be greater than 1 with a certain probability. This is typical behavior for a compound Poisson random variable.

To answer the second question, we next informally derive the parameters of the compound Poisson approximation to  $N_{E_\delta}^{(R)}$  in the special case when  $\Sigma$  is diagonal. We require some additional definitions. Let  $[\vec{i}] = \{i_0, i_1, \dots, i_\delta\}$  be the unordered set of indexes, which is referred as *index group*, of any  $\vec{i} \in C_\delta^<$  and define  $[C_\delta^<] := \{[\vec{i}] : \vec{i} \in C_\delta^<\}$ . It follows that  $|[C_\delta^<]| = \binom{p}{\delta+1}$ , where  $|\cdot|$  is the cardinality of a set. For a given group of  $\delta + 1$  indexes  $[\vec{i}]$ ,  $\Phi_{\vec{i}}^{(R)} + U_{\vec{i}}$  is the increment associated to this group and its value is between 0 and  $\delta + 1$ . In the following informal argument we assume that the event that two different index groups both have non-zero increments has probability zero. This assumption will be verified in the proof of Proposition III.6. Consequently the probability of increment size  $\ell$  for  $\ell \geq 1$  is proportional to the expectation of the fraction of the number of groups with increment  $\ell$ :

$$\begin{aligned} & \mathbb{E} \frac{1}{|[C_\delta^<]|} \sum_{[\vec{i}] \in [C_\delta^<]} 1 \left( \Phi_{\vec{i}}^{(R)} + U_{\vec{i}} = \ell \right) \\ &= \frac{1}{|[C_\delta^<]|} \frac{1}{\ell} \mathbb{E} \sum_{[\vec{i}] \in [C_\delta^<]} \left( \Phi_{\vec{i}}^{(R)} + U_{\vec{i}} \right) 1 \left( \Phi_{\vec{i}}^{(R)} + U_{\vec{i}} = \ell \right) \\ &= \frac{1}{|[C_\delta^<]|} \frac{1}{\ell} \mathbb{E} \sum_{\vec{i} \in C_\delta^<} \Phi_{\vec{i}}^{(R)} 1 \left( \Phi_{\vec{i}}^{(R)} + U_{\vec{i}} = \ell \right) \\ &= \frac{1}{|[C_\delta^<]|} \frac{1}{\ell} \sum_{\vec{i} \in C_\delta^<} \mathbb{P} \left( \Phi_{\vec{i}}^{(R)} = 1 \right) \mathbb{P} \left( \Phi_{\vec{i}}^{(R)} + U_{\vec{i}} = \ell | \Phi_{\vec{i}}^{(R)} = 1 \right). \end{aligned} \quad (51)$$

Since  $\Sigma$  is assumed to be diagonal,  $\{\mathbf{u}_i\}_{i=1}^p$  are i.i.d.  $\text{unif}(S^{n-2})$  by Lemma III.2 (b) and hence (51) becomes

$$\begin{aligned} & \mathbb{E} \frac{1}{|[C_\delta^<]|} \sum_{[\vec{i}] \in [C_\delta^<]} 1 \left( \Phi_{\vec{i}}^{(R)} + U_{\vec{i}} = \ell \right) \\ &= \frac{1}{|[C_\delta^<]|} \frac{1}{\ell} |C_\delta^<| (2P_n(r_\rho))^\delta \alpha(\ell, r_\rho) \\ &= \frac{\delta + 1}{\ell} (2P_n(r_\rho))^\delta \alpha(\ell, r_\rho), \end{aligned}$$

where  $\alpha(\ell, r_\rho)$  is defined in (23) and  $\mathbb{P} \left( \Phi_{\vec{i}}^{(R)} = 1 \right) = (2P_n(r_\rho))^\delta$  by conditioning on  $\mathbf{u}_{i_0}$ . As a consequence, the probability of increment size  $\ell$  for  $\ell \geq 1$  is:

$$\begin{aligned} & \frac{\delta + 1}{\ell} (2P_n(r_\rho))^\delta \alpha(\ell, r_\rho) / \sum_{\ell=1}^{\delta+1} \left( \frac{\delta + 1}{\ell} (2P_n(r_\rho))^\delta \alpha(\ell, r_\rho) \right) \\ &= \zeta_{n,\delta,\rho}(\ell), \end{aligned}$$

where the equation follows from (24). This argument implies that  $\zeta_{n,\delta,\rho}$  is the increment distribution of the compound Poisson approximation.

Since the mean of a compound Poisson distribution is the product of the arrival rate and the mean of the increment

distribution, the arrival rate  $\lambda$  of the compound Poisson approximation satisfies the following mean constraint:

$$\lambda \mathbb{E} \zeta_{n,\delta,\rho} = \mathbb{E} N_{E_\delta}^{(R)},$$

where  $\mathbb{E} \zeta_{n,\delta,\rho}$  is the mean of  $\zeta_{n,\delta,\rho}$ . One can easily verify  $\mathbb{E} \zeta_{n,\delta,\rho} = 1 / \sum_{\ell=1}^{\delta+1} (\alpha(\ell, r_\rho) / \ell)$  and  $\mathbb{E} N_{E_\delta}^{(R)} = \binom{p}{1} \binom{p-1}{\delta} \mathbb{E} \Phi_{\vec{i}}^{(R)} = \binom{p}{1} \binom{p-1}{\delta} (2P_n(r_\rho))^\delta$  since  $\Sigma$  is diagonal. Hence the arrival rate for the compound Poisson is

$$\lambda = \binom{p}{1} \binom{p-1}{\delta} (2P_n(r_\rho))^\delta \sum_{\ell=1}^{\delta+1} \frac{\alpha(\ell, r_\rho)}{\ell} = \lambda_{p,n,\delta,\rho}. \quad (52)$$

This informal argument motivates the compound Poisson approximation  $\text{CP}(\lambda_{p,n,\delta,\rho}, \zeta_{n,\delta,\rho})$  to  $N_{E_\delta}^{(R)}$  when  $\Sigma$  is diagonal.

### E. Proof of Proposition III.6

1) *Auxiliary Lemmas for Proposition III.6:* Recall for any  $\delta \geq 1$ ,  $C_\delta^<$  is defined in (46). For  $\vec{i} \in C_\delta^<$ , define a symmetric positive definite matrix  $\Sigma_{\vec{i}} \in \mathbb{R}^{(\ell+1) \times (\ell+1)}$  to be the submatrix of  $\Sigma$ , consisting of rows and columns  $\Sigma$  indexed by the ordered components  $(i_0, i_1, \dots, i_\ell)$  of  $\vec{i}$ . Let  $[\vec{i}] = \{i_0, i_1, \dots, i_\ell\}$  be the unordered set of indices of any  $\vec{i} \in C_\delta^<$ . Then  $\Sigma_{\vec{i}} \in \Sigma_{[\vec{i}]}$  and  $\mu_{n,\ell+1}(\Sigma_{\vec{i}}) = \mu_{n,\ell+1}(\Sigma_{[\vec{i}]})$ , where  $\Sigma_{[\vec{i}]}$  and  $\mu_{n,\ell+1}(\Sigma_{[\vec{i}]})$  are defined in the paragraph after Definition II.3.

*Lemma A.2:* Suppose  $\mathbf{X} \sim \mathcal{VE}(\mu, \Sigma, g)$ . Let  $\ell \in [p-1]$ . Consider  $\vec{i} = (i_0, i_1, \dots, i_\ell) \in C_\delta^<$ .

$$\mathbb{E} \prod_{q=1}^{\ell} \Phi_{i_0 i_q}^{(R)} \leq \mu_{n,\ell+1}(\Sigma_{\vec{i}}) (2P_n(r_\rho))^\ell.$$

Moreover, when  $\Sigma_{\vec{i}}$  is diagonal, in the last expression the equality holds and  $\mu_{n,\ell+1}(\Sigma_{\vec{i}}) = 1$ .

*Proof:*  $\prod_{q=1}^{\ell} \Phi_{i_0 i_q}^{(R)}$  is a nonnegative Borel Measurable function of  $\mathbf{u}_j$  for  $j \in [\vec{i}]$ . By Lemma III.4 (c), it suffices to show

$$\mathbb{E} \prod_{q=1}^{\ell} \Phi_{i_0 i_q}^{(R)} \leq (2P_n(r))^\ell$$

for the case  $\mathbf{u}_j$  for  $j \in [\vec{i}]$  are  $\ell+1$  independent  $\text{unif}(S^{n-2})$ . The last inequality holds with equality, which follows from that the terms in the product on the left hand side are independent conditioned on  $\mathbf{u}_{i_0}$ .  $\square$

Lemma A.2 suggests differentiating whether  $\Sigma_{\vec{i}}$  is diagonal or not since  $\mu_{n,\ell+1}(\Sigma_{\vec{i}}) = 1$  when  $\Sigma_{\vec{i}}$  is diagonal. The next lemma establishes an upper bound on the number of  $\vec{i} \in C_\delta^<$  such that  $\Sigma_{\vec{i}}$  is not diagonal.

*Lemma A.3:* Let  $\Sigma$  be row- $\kappa$  sparse. Let  $\delta \in [p-1]$ . Then

$$\sum_{\substack{\vec{i} \in C_\delta^< \\ \Sigma_{\vec{i}} \text{ not diagonal}}} 1 \leq \frac{\delta(\delta+1)}{2} (\kappa-1) \binom{p}{\delta} \leq \frac{(\delta+1)}{2((\delta-1)!)} p^{\delta(\kappa-1)}.$$

*Proof:* Note that

$$\sum_{\substack{\vec{i} \in C_\delta^< \\ \Sigma_{\vec{i}} \text{ diagonal}}} 1 \geq \frac{1}{\delta!} p(p-\kappa) \dots (p-\delta\kappa),$$

where the  $\frac{1}{\delta!}$  is due to in our definition  $\vec{i}$  the index  $i_1 < \dots < i_\delta$  are sorted. Then

$$\begin{aligned} \sum_{\substack{\vec{i} \in C_\delta^< \\ \Sigma_{\vec{i}} \text{ not diagonal}}} 1 &\leq \binom{p}{1} \binom{p-1}{\delta} - \frac{1}{\delta!} p \prod_{\ell=1}^{\delta} (p-\ell\kappa) \\ &\leq \frac{\delta(\delta+1)}{2} (\kappa-1) \binom{p}{\delta}, \end{aligned}$$

where the last inequality follows from Lemma A.26 (b).  $\square$

Note  $\kappa = 1$ , the Lemma A.3 shows  $\sum_{\substack{\vec{i} \in C_\delta^< \\ \Sigma_{\vec{i}} \text{ not diagonal}}} 1 = 0$ ,

which means  $\Sigma$  is diagonal matrix. Next we present a lemma to bound  $\sum_{\vec{i} \in C_\ell^<} \mu_{n,\ell+1}(\Sigma_{\vec{i}})$ .

*Lemma A.4:*

$$\sum_{\vec{i} \in C_\ell^<} \mu_{n,\ell+1}(\Sigma_{\vec{i}}) \leq \frac{p^{\ell+1}}{\ell!} \left( 1 + \ell^2 \mu_{n,\ell+1}(\Sigma) \frac{\kappa-1}{p} \right).$$

*Proof:*

$$\begin{aligned} \sum_{\vec{i} \in C_\ell^<} \mu_{n,\ell+1}(\Sigma_{\vec{i}}) &= \sum_{\substack{\vec{i} \in C_\ell^< \\ \Sigma_{\vec{i}} \text{ diagonal}}} 1 + \sum_{\substack{\vec{i} \in C_\ell^< \\ \Sigma_{\vec{i}} \text{ not diagonal}}} \mu_{n,\ell+1}(\Sigma_{\vec{i}}) \\ &\leq \binom{p}{1} \binom{p-1}{\ell} + \mu_{n,\ell+1}(\Sigma) \sum_{\substack{\vec{i} \in C_\ell^< \\ \Sigma_{\vec{i}} \text{ not diagonal}}} 1 \\ &\leq \frac{p^{\ell+1}}{\ell!} \left( 1 + \ell^2 \mu_{n,\ell+1}(\Sigma) \frac{\kappa-1}{p} \right), \end{aligned}$$

where the first inequality follows from the Lemma III.4 (a), and the second inequality follows from Lemma A.3.  $\square$

*Lemma A.5:* Let  $\mathbf{X} \sim \mathcal{VE}(\mu, \Sigma, g)$ . Let  $\{i_q\}_{q=0}^\alpha, \{j_q\}_{q=0}^\beta \subset [p]$  be respectively a sequence of  $\alpha+1$  and  $\beta+1$  distinct integers. Let  $m \in [\min\{\alpha, \beta\}]$ . Suppose  $i_q = j_q$  for  $q \in [m]$  and  $i_q \neq j_{q'}$  for  $q, q' \notin [m]$ . Denote  $\mathcal{I} = \bigcup_{q=0}^\alpha \{i_q\} \cup \left( \bigcup_{q'=0}^\beta \{j_{q'}\} \right)$  and then  $|\mathcal{I}| = \alpha + \beta - m + 2$ .

(a) Then

$$\begin{aligned} &\mathbb{E} \left( \prod_{q=1}^\alpha \Phi_{i_0 i_q}^{(\mathbf{R})} \right) \left( \prod_{q'=1}^\beta \Phi_{j_0 j_{q'}}^{(\mathbf{R})} \right) \\ &\leq \mu_{n,|\mathcal{I}|}(\Sigma_{\mathcal{I}}) (2P_n(r_\rho))^{\alpha+\beta-m} (2P_n(2r_\rho)). \end{aligned} \quad (53)$$

(b) Then

$$\begin{aligned} &\mathbb{E} \Phi_{i_0 j_0}^{(\mathbf{R})} \left( \prod_{q=1}^\alpha \Phi_{i_0 i_q}^{(\mathbf{R})} \right) \left( \prod_{q'=1}^\beta \Phi_{j_0 j_{q'}}^{(\mathbf{R})} \right) \\ &\leq \mu_{n,|\mathcal{I}|}(\Sigma_{\mathcal{I}}) (2P_n(r_\rho))^{\alpha+\beta-m+1} \sum_{\ell=2}^{m+2} \alpha_{n,m+1}(\ell, r_\rho) \\ &\leq \mu_{n,|\mathcal{I}|}(\Sigma_{\mathcal{I}}) (2P_n(r_\rho))^{\alpha+\beta-m+1}. \end{aligned} \quad (54)$$

The inequality (54) also holds with  $m = 0$ .

*Proof:* (a) By Lemma III.4 (c), it suffices to prove (53) without  $\mu_{n,|\mathcal{I}|}(\Sigma_{\mathcal{I}})$  for the case that  $\{\mathbf{u}_j\}$  for  $j \in \mathcal{I}$  are independent  $\text{unif}(S^{n-2})$ . Conditioned on  $\mathbf{u}_{i_0}$  and  $\mathbf{u}_{j_0}$ ,  $\{\Phi_{i_0 i_q}^{(\mathbf{R})} \Phi_{j_0 j_{q'}}^{(\mathbf{R})}\}_{q=1}^m$  are i.i.d.,  $\{\Phi_{i_0 i_q}^{(\mathbf{R})}\}_{q=m+1}^\alpha \cup \{\Phi_{j_0 j_{q'}}^{(\mathbf{R})}\}_{q'=m+1}^\beta$  are i.i.d. and moreover, every term in  $\{\Phi_{i_0 i_q}^{(\mathbf{R})} \Phi_{j_0 j_{q'}}^{(\mathbf{R})}\}_{q=1}^m$  is independent of every term in  $\{\Phi_{i_0 i_q}^{(\mathbf{R})}\}_{q=m+1}^\alpha \cup \{\Phi_{j_0 j_{q'}}^{(\mathbf{R})}\}_{q'=m+1}^\beta$ . Thus

$$\begin{aligned} &\mathbb{E} \left[ \left( \prod_{q=1}^\alpha \Phi_{i_0 i_q}^{(\mathbf{R})} \right) \left( \prod_{q'=1}^\beta \Phi_{j_0 j_{q'}}^{(\mathbf{R})} \right) \middle| \mathbf{u}_{i_0}, \mathbf{u}_{j_0} \right] \\ &= \mathbb{E} \left[ \prod_{q=1}^m \Phi_{i_0 i_q}^{(\mathbf{R})} \Phi_{j_0 j_{q'}}^{(\mathbf{R})} \middle| \mathbf{u}_{i_0}, \mathbf{u}_{j_0} \right] \left( \mathbb{E} \left[ \Phi_{i_0 i_\alpha}^{(\mathbf{R})} 1(\alpha > m) + \Phi_{j_0 j_\beta}^{(\mathbf{R})} 1(\beta > m) \middle| \mathbf{u}_{i_0}, \mathbf{u}_{j_0} \right] \right)^{\alpha+\beta-2m} \\ &= \mathbb{E} \left[ \prod_{q=1}^m \Phi_{i_0 i_q}^{(\mathbf{R})} \Phi_{j_0 j_{q'}}^{(\mathbf{R})} \middle| \mathbf{u}_{i_0}, \mathbf{u}_{j_0} \right] (2P_n(r_\rho))^{\alpha+\beta-2m} \end{aligned} \quad (55)$$

$$= \left( \mathbb{E} \left[ \Phi_{i_0 i_q}^{(\mathbf{R})} \Phi_{j_0 j_{q'}}^{(\mathbf{R})} \middle| \mathbf{u}_{i_0}, \mathbf{u}_{j_0} \right] \right)^m (2P_n(r_\rho))^{\alpha+\beta-2m} \quad (56)$$

where for the first equality the convention  $0^0 = 1$  is used if  $\alpha = \beta = m$ . Notice (56) also holds for  $m = 0$ .

Denote  $\overline{\text{SC}}(r, \mathbf{w}) = \text{SC}(r, \mathbf{w}) \cup \text{SC}(r, -\mathbf{w})$ . Then conditioned on  $\mathbf{u}_{i_0}$  and  $\mathbf{u}_{j_0}$ ,

$$\begin{aligned} &\Phi_{i_0 i_1}^{(\mathbf{R})} \Phi_{j_0 j_1}^{(\mathbf{R})} \\ &= 1(\mathbf{u}_{i_1} \in \overline{\text{SC}}(r, \mathbf{u}_{i_0}) \cap \overline{\text{SC}}(r, \mathbf{u}_{j_0})) \\ &= \eta 1(\mathbf{u}_{i_1} \in \overline{\text{SC}}(r, \mathbf{u}_{i_0}) \cap \overline{\text{SC}}(r, \mathbf{u}_{j_0})), \end{aligned}$$

where  $\eta = 1(\|\mathbf{u}_{i_0} - \mathbf{u}_{j_0}\|_2 \leq 2r_\rho \text{ or } \|\mathbf{u}_{i_0} + \mathbf{u}_{j_0}\|_2 \leq 2r_\rho)$ , and the last equality follows by noticing that  $\overline{\text{SC}}(r_\rho, \mathbf{u}_{i_0}) \cap \overline{\text{SC}}(r_\rho, \mathbf{u}_{j_0})$  is non-empty only when  $\|\mathbf{u}_{i_0} - \mathbf{u}_{j_0}\|_2 \leq 2r_\rho$  or  $\|\mathbf{u}_{i_0} + \mathbf{u}_{j_0}\|_2 \leq 2r_\rho$ . Plugging the above inequality into (56), we obtain

$$\begin{aligned} &\mathbb{E} \left[ \left( \prod_{q=1}^\alpha \Phi_{i_0 i_q}^{(\mathbf{R})} \right) \left( \prod_{q'=1}^\beta \Phi_{j_0 j_{q'}}^{(\mathbf{R})} \right) \middle| \mathbf{u}_{i_0}, \mathbf{u}_{j_0} \right] \\ &= \eta \left( \mathbb{E} \left[ \Phi_{i_0 i_1}^{(\mathbf{R})} \Phi_{j_0 j_1}^{(\mathbf{R})} \middle| \mathbf{u}_{i_0}, \mathbf{u}_{j_0} \right] \right)^m (2P_n(r_\rho))^{\alpha+\beta-2m} \\ &\leq \eta (2P_n(r_\rho))^m (2P_n(r_\rho))^{\alpha+\beta-2m}. \end{aligned}$$

The result then follows by taking expectation w.r.t.  $\mathbf{u}_{i_0}$  and  $\mathbf{u}_{j_0}$ .

(b) Similar to the proof of (a), it suffices to prove (54) without  $\mu_{n,|\mathcal{I}|}(\Sigma_{\mathcal{I}})$  for the case  $\mathbf{u}_j$  for  $j \in \mathcal{I}$  are independent  $\text{unif}(S^{n-2})$ . Conditioned on  $\mathbf{u}_{i_0}$  and  $\mathbf{u}_{j_0}$ ,

$$\begin{aligned} &\mathbb{E} \left[ \Phi_{i_0 j_0}^{(\mathbf{R})} \left( \prod_{q=1}^\alpha \Phi_{i_0 i_q}^{(\mathbf{R})} \right) \left( \prod_{q'=1}^\beta \Phi_{j_0 j_{q'}}^{(\mathbf{R})} \right) \middle| \mathbf{u}_{i_0}, \mathbf{u}_{j_0} \right] \\ &= \Phi_{i_0 j_0}^{(\mathbf{R})} \mathbb{E} \left[ \prod_{q=1}^m \Phi_{i_0 i_q}^{(\mathbf{R})} \Phi_{j_0 j_{q'}}^{(\mathbf{R})} \middle| \mathbf{u}_{i_0}, \mathbf{u}_{j_0} \right] (2P_n(r_\rho))^{\alpha+\beta-2m}. \end{aligned}$$

where the equality follows from (55). Then

$$\begin{aligned}
& \mathbb{E} \left[ \Phi_{i_0 j_0}^{(\mathbf{R})} \left( \prod_{q=1}^{\alpha} \Phi_{i_0 i_q}^{(\mathbf{R})} \right) \left( \prod_{q'=1}^{\beta} \Phi_{j_0 j_{q'}}^{(\mathbf{R})} \right) \right] \\
&= \mathbb{E} \left[ \prod_{q=0}^m \Phi_{i_0 i_q}^{(\mathbf{R})} \prod_{q'=0}^m \Phi_{j_0 j_{q'}}^{(\mathbf{R})} \right] (2P_n(r_\rho))^{\alpha+\beta-2m} \\
&= \mathbb{P} \left( \prod_{q=0}^m \Phi_{i_0 i_q}^{(\mathbf{R})} = 1 \right) \mathbb{P} \left( \prod_{q'=0}^m \Phi_{j_0 j_{q'}}^{(\mathbf{R})} = 1 \mid \prod_{q=0}^m \Phi_{i_0 i_q}^{(\mathbf{R})} = 1 \right) \\
&\quad (2P_n(r_\rho))^{\alpha+\beta-2m} \\
&\stackrel{(*)}{\leq} (2P_n(r_\rho))^{\alpha+\beta-m+1} \sum_{\ell=2}^{m+2} \alpha_{n,m+1}(\ell, r_\rho) \\
&\leq (2P_n(r_\rho))^{\alpha+\beta-m+1},
\end{aligned}$$

where step (\*) follows from the definition of  $\alpha_{n,\delta}(\ell, r_\rho)$  in (23). Notice (56) also holds for  $m = 0$ .  $\square$

2) *Lemmas on Double Summations:* Denote  $\vec{i} \cup \vec{j} = [\vec{i}] \cup [\vec{j}]$  for any  $\vec{i} \in C_q^<$  and any  $\vec{j} \in C_\delta^<$ . Consider any  $\theta_{\vec{i}, \vec{j}}$  that is a non-negative function of  $\mathbf{u}_\ell$  for  $\ell \in \vec{i} \cup \vec{j}$  defined for  $\vec{i} \in C_q^<$  and  $\vec{j} \in C_\delta^<$  with  $1 \leq \delta \leq q \leq p-1$ . In this section an upper bound on  $\mathbb{E} \sum_{\vec{i} \in C_q^<} \sum_{\vec{j} \in C_\delta^<} \theta_{\vec{i}, \vec{j}}$  is presented. The results in this subsection will be used in the proofs of Proposition III.6 and Proposition IV.4.

For  $i \in [p]$ , let

$$\mathcal{NZ}(i) := \{m \in [p] : \Sigma_{im} \neq 0\} \quad (57)$$

denote the index of the variables that has non zero correlation with the  $i$ -th variable. For  $\vec{i} \in C_q^<$ , define  $\mathcal{NZ}(\vec{i}) := \bigcup_{\ell=0}^q \mathcal{NZ}(i_\ell)$ . Since  $\Sigma$  is row- $\kappa$  sparse, for any  $\vec{i} \in C_q^<$ ,  $|\mathcal{NZ}(\vec{i})| \leq (q+1)\kappa$ , and

$$p_{\vec{i}} := |p \setminus \mathcal{NZ}(\vec{i})| \geq p - (q+1)\kappa. \quad (58)$$

Note that  $p_{\vec{i}}$  denotes the number of variables that are independent of variables with index in  $[\vec{i}]$ .

For  $\vec{i} \in C_q^<$ , define

$$J_{\vec{i}} := \left\{ \vec{j} \in C_\delta^< : \bigcup_{\ell=0}^{\delta} \{j_\ell\} \subset \bigcup_{\ell=0}^q \{i_\ell\} \right\}, \quad (59)$$

$$T_{\vec{i}} := \left\{ \vec{j} \in C_\delta^< : \left( \bigcup_{\ell=0}^{\delta} \{j_\ell\} \right) \cap (\mathcal{NZ}(\vec{i})) = \emptyset \right\}, \quad (60)$$

$$N_{\vec{i}} := C_q^< \setminus J_{\vec{i}} \setminus T_{\vec{i}}. \quad (61)$$

Here  $J_{\vec{i}}$  is the set of indices in  $C_\delta^<$  consisting of coordinates as subsets of  $[\vec{i}]$ ;  $T_{\vec{i}}$  is the set of indices in  $C_\delta^<$  consisting of coordinates outside the neighborhood of  $\vec{i}$ ;  $N_{\vec{i}}$  is the set of “correlated but not highly correlated” indices in  $C_\delta^<$ , i.e. the set of indices of which at least one coordinate is in the neighborhood of  $\vec{i}$ , but excluding those sets of indices of which the set of coordinates are subsets as that of  $\vec{i}$ .

The strategy is to decompose

$$\begin{aligned}
& \mathbb{E} \sum_{\vec{i} \in C_q^<} \sum_{\vec{j} \in C_\delta^<} \theta_{\vec{i}, \vec{j}} \\
&= \mathbb{E} \sum_{\vec{i} \in C_q^<} \sum_{\vec{j} \in J_{\vec{i}}} \theta_{\vec{i}, \vec{j}} + \mathbb{E} \sum_{\vec{i} \in C_q^<} \sum_{\vec{j} \in T_{\vec{i}}} \theta_{\vec{i}, \vec{j}} + \mathbb{E} \sum_{\vec{i} \in C_q^<} \sum_{\vec{j} \in N_{\vec{i}}} \theta_{\vec{i}, \vec{j}}
\end{aligned}$$

and bound each of the three terms.

The next result is an upper bound on the first two terms.

*Lemma A.6:* Let  $p \geq n \geq 4$  and  $\mathbf{X} \sim \mathcal{VE}(\mu, \Sigma, g)$ . Suppose  $\Sigma$  is row- $\kappa$  sparse. Consider any  $\theta_{\vec{i}, \vec{j}}$  that is a non-negative function of  $\mathbf{u}_\ell$  for  $\ell \in \vec{i} \cup \vec{j}$  defined for  $\vec{i} \in C_q^<$  and  $\vec{j} \in C_\delta^<$  with  $1 \leq \delta \leq q \leq p-1$ .

(a) Suppose there exist positive constants  $a, z$  such that  $\mathbb{E} \theta_{\vec{i}, \vec{j}} \leq \mu_{n,q+1}(\Sigma_{\vec{i}}) a z^q$  for any  $\vec{j} \in J_{\vec{i}}$ . Then

$$\begin{aligned}
& \sum_{\vec{i} \in C_q^<} \sum_{\vec{j} \in J_{\vec{i}}} \mathbb{E} \theta_{\vec{i}, \vec{j}} \\
& \leq a p (p z)^q \frac{q+1}{\delta! (q-\delta)!} \left( 1 + q^2 \mu_{n,q+1}(\Sigma) \frac{\kappa-1}{p} \right)
\end{aligned}$$

(b) Suppose there exist positive constants  $a, z$  such that  $\mathbb{E} \theta_{\vec{i}, \vec{j}} \leq \mu_{n,q+\delta+2}(\Sigma_{\vec{i} \cup \vec{j}}) a z^{q+\delta}$  for any  $\vec{j} \in T_{\vec{i}}$ . Then

$$\begin{aligned}
& \sum_{\vec{i} \in C_q^<} \sum_{\vec{j} \in T_{\vec{i}}} \mathbb{E} \theta_{\vec{i}, \vec{j}} \\
& \leq a p^2 (p z)^{q+\delta} \frac{3}{\delta! (q-1)!} \left( 1 + \mu_{n,q+\delta+2}(\Sigma) \frac{\kappa-1}{p} \right).
\end{aligned}$$

*Proof:* (a) Since  $|J_{\vec{i}}| = \binom{q+1}{1} \binom{q}{\delta}$

$$\begin{aligned}
& \sum_{\vec{i} \in C_q^<} \sum_{\vec{j} \in J_{\vec{i}}} \mathbb{E} \theta_{\vec{i}, \vec{j}} \\
& \leq a z^q \binom{q+1}{1} \binom{q}{\delta} \sum_{\vec{i} \in C_q^<} \mu_{n,q+1}(\Sigma_{\vec{i}}) \\
& \leq a p (p z)^q \frac{q+1}{\delta! (q-\delta)!} \left( 1 + q^2 \mu_{n,q+1}(\Sigma) \frac{\kappa-1}{p} \right),
\end{aligned}$$

where the last step follows from Lemma A.4.

(b)

$$\begin{aligned}
& \sum_{\vec{i} \in C_q^<} \sum_{\vec{j} \in T_{\vec{i}}} \mathbb{E} \theta_{\vec{i}, \vec{j}} \\
& \leq a z^{q+\delta} \sum_{\vec{i} \in C_q^<} \sum_{\vec{j} \in T_{\vec{i}}} \mu_{n,q+\delta+2}(\Sigma_{\vec{i} \cup \vec{j}}) \\
& \leq a z^{q+\delta} \left( \sum_{\substack{\vec{i} \in C_q^< \\ \Sigma_{\vec{i}} \text{ diagonal}}} \sum_{\substack{\vec{j} \in T_{\vec{i}} \\ \Sigma_{\vec{j}} \text{ diagonal}}} 1 + \mu_{n,q+\delta+2}(\Sigma) \sum_{\substack{\vec{i} \in C_q^< \\ \Sigma_{\vec{i}} \text{ diagonal}}} \right. \\
& \quad \left. \sum_{\substack{\vec{j} \in T_{\vec{i}} \\ \Sigma_{\vec{j}} \text{ not diagonal}}} 1 + \mu_{n,q+\delta+2}(\Sigma) \sum_{\substack{\vec{i} \in C_q^< \\ \Sigma_{\vec{i}} \text{ not diagonal}}} \sum_{\substack{\vec{j} \in T_{\vec{i}} \\ \Sigma_{\vec{j}} \text{ not diagonal}}} 1 \right)
\end{aligned}$$

$$\leq az^{q+\delta} \left( \binom{p}{1} \binom{p-1}{q} \binom{p}{1} \binom{p-1}{\delta} + \mu_{n,q+\delta+2}(\Sigma) \binom{p}{1} \binom{p-1}{q} \sum_{\substack{\vec{j} \in C_q^\delta \\ \Sigma_{\vec{j}} \text{ not diagonal}}} 1 + \mu_{n,q+\delta+2}(\Sigma) \sum_{\substack{\vec{i} \in C_q^\delta \\ \Sigma_{\vec{i}} \text{ not diagonal}}} \binom{p}{1} \binom{p-1}{\delta} \right) \\ \leq ap^2(pz)^{q+\delta} \frac{3}{\delta!(q-1)!} \left( 1 + \mu_{n,q+\delta+2}(\Sigma) \frac{\kappa-1}{p} \right),$$

where the second inequality follows from that for  $\vec{j} \in T_{\vec{i}}$ ,  $\Sigma_{\vec{i} \cup \vec{j}}$  is diagonal if and only if  $\Sigma_{\vec{i}}$  and  $\Sigma_{\vec{j}}$  are both diagonal; and the last step follows from Lemma A.3.  $\square$

To control  $\mathbb{E} \sum_{\vec{i} \in C_q^\delta} \sum_{\vec{j} \in N_{\vec{i}}} \theta_{\vec{i},\vec{j}}$ , we further partition  $N_{\vec{i}}$  into 6 subsets as follows. For  $\vec{i} \in C_q^\delta$  with  $q \geq \delta$ , define

$$\begin{aligned} \mathcal{K}_1(\vec{i}) &:= \left\{ \vec{j} \in N_{\vec{i}} : j_0 = i_0 \right\}, \\ \mathcal{K}_2(\vec{i}) &:= \left\{ \vec{j} \in N_{\vec{i}} : j_0 \neq i_0, j_0 \in \bigcup_{\ell=1}^q \{i_\ell\}, i_0 \in \bigcup_{\ell=1}^\delta \{j_\ell\} \right\}, \\ \mathcal{K}_3(\vec{i}) &:= \left\{ \vec{j} \in N_{\vec{i}} : j_0 \neq i_0, j_0 \notin \bigcup_{\ell=1}^q \{i_\ell\}, i_0 \in \bigcup_{\ell=1}^\delta \{j_\ell\} \right\}, \\ \mathcal{K}_4(\vec{i}) &:= \left\{ \vec{j} \in N_{\vec{i}} : j_0 \neq i_0, j_0 \in \bigcup_{\ell=1}^q \{i_\ell\}, i_0 \notin \bigcup_{\ell=1}^\delta \{j_\ell\} \right\}, \\ \mathcal{K}_5(\vec{i}) &:= \left\{ \vec{j} \in N_{\vec{i}} : j_0 \neq i_0, j_0 \notin \bigcup_{\ell=1}^q \{i_\ell\}, i_0 \notin \bigcup_{\ell=1}^\delta \{j_\ell\}, \right. \\ &\quad \left. \left| \left( \bigcup_{\ell=1}^q \{i_\ell\} \right) \cap \left( \bigcup_{\ell=1}^\delta \{j_\ell\} \right) \right| \geq 1 \right\}, \\ \mathcal{K}_6(\vec{i}) &:= \left\{ \vec{j} \in N_{\vec{i}} : j_0 \neq i_0, j_0 \notin \bigcup_{\ell=1}^q \{i_\ell\}, i_0 \notin \bigcup_{\ell=1}^\delta \{j_\ell\}, \right. \\ &\quad \left. \left( \bigcup_{\ell=1}^q \{i_\ell\} \right) \cap \left( \bigcup_{\ell=1}^\delta \{j_\ell\} \right) = \emptyset \right\}. \end{aligned}$$

Then  $N_{\vec{i}} = \bigcup_{w=1}^6 \mathcal{K}_w(\vec{i})$ . Let  $D_{\vec{i}}^m = \{ \vec{j} \in N_{\vec{i}} : |(\bigcup_{\ell=1}^q \{i_\ell\}) \cap (\bigcup_{\ell=1}^\delta \{j_\ell\})| = m \}$ . We are now in a good position to present a lemma on  $\mathbb{E} \sum_{\vec{i} \in C_q^\delta} \sum_{\vec{j} \in N_{\vec{i}}} \theta_{\vec{i},\vec{j}}$ .

**Lemma A.7:** Let  $p \geq n \geq 4$  and  $\mathbf{X} \sim \mathcal{VE}(\mu, \Sigma, g)$ . Suppose  $\Sigma$  is row- $\kappa$  sparse. Consider any  $\theta_{\vec{i},\vec{j}}$  that is a non-negative function of  $\mathbf{u}_\ell$  for  $\ell \in \vec{i} \cup \vec{j}$  defined for  $\vec{i} \in C_q^\delta$  and  $\vec{j} \in C_q^\delta$  with  $1 \leq \delta \leq q \leq p-1$ . Suppose there exist positive constants  $a, b, z$  such that  $\theta_{\vec{i},\vec{j}}$  satisfies:  $\forall 0 \leq m \leq \delta-1, \forall w \in \{1, 3, 4\}$ ,

$$\mathbb{E} \theta_{\vec{i},\vec{j}} \leq \mu_{n,|\vec{i} \cup \vec{j}|}(\Sigma_{\vec{i} \cup \vec{j}}) az^{q+\delta-m}, \quad \forall \vec{j} \in \mathcal{K}_w(\vec{i}) \cap D_{\vec{i}}^m;$$

$$\forall 0 \leq m \leq \delta-2,$$

$$\mathbb{E} \theta_{\vec{i},\vec{j}} \leq \mu_{n,|\vec{i} \cup \vec{j}|}(\Sigma_{\vec{i} \cup \vec{j}}) az^{q+\delta-m-1}, \quad \forall \vec{j} \in \mathcal{K}_2(\vec{i}) \cap D_{\vec{i}}^m;$$

$$\forall 1 \leq m \leq \delta,$$

$$\mathbb{E} \theta_{\vec{i},\vec{j}} \leq \mu_{n,|\vec{i} \cup \vec{j}|}(\Sigma_{\vec{i} \cup \vec{j}}) abz^{q+\delta-m}, \quad \forall \vec{j} \in \mathcal{K}_5(\vec{i}) \cap D_{\vec{i}}^m;$$

and

$$\mathbb{E} \theta_{\vec{i},\vec{j}} \leq \mu_{n,|\vec{i} \cup \vec{j}|}(\Sigma_{\vec{i} \cup \vec{j}}) az^{q+\delta}, \quad \forall \vec{j} \in \mathcal{K}_6(\vec{i}).$$

Then

$$\begin{aligned} &\sum_{\vec{i} \in C_q^\delta} \sum_{\vec{j} \in N_{\vec{i}}} \mathbb{E} \theta_{\vec{i},\vec{j}} \\ &\leq ap(pz)^{q+1} \eta_1 \frac{(1+pz)^{\delta-1} \delta(4+b/z)}{(\delta-1)!} + \\ &\quad ap^2(pz)^{q+\delta} \frac{(\delta+1)(q+1)}{\delta!q!} \mu_{n,q+\delta+2}(\Sigma) \frac{\kappa-1}{p}, \end{aligned}$$

$$\text{where } \eta_1 = \left( 1 + \frac{\mu_{n,q+\delta+1}(\Sigma) 3q^2(\kappa-1)}{p} \right).$$

*Proof:* Since

$$N_{\vec{i}} = \bigcup_{w=1}^6 \mathcal{K}_w(\vec{i}),$$

$$\sum_{\vec{i} \in C_q^\delta} \sum_{\vec{j} \in N_{\vec{i}}} \mathbb{E} \theta_{\vec{i},\vec{j}} = \sum_{w=1}^6 \sum_{\vec{i} \in C_q^\delta} I_w(\vec{i}), \quad (62)$$

with

$$I_w(\vec{i}) := \sum_{\vec{j} \in \mathcal{K}_w(\vec{i})} \mathbb{E} \theta_{\vec{i},\vec{j}}.$$

**Case 1:**  $p \geq q + \delta + 2$

Obviously  $\mathcal{K}_1(\vec{i}) = \bigcup_{m=0}^{\delta-1} (\mathcal{K}_1(\vec{i}) \cap D_{\vec{i}}^m)$ . Then for any  $\vec{i} \in C_q^\delta$  satisfying  $\Sigma_{\vec{i}}$  diagonal,

$$\begin{aligned} &\left| \vec{j} \in \mathcal{K}_1(\vec{i}) \cap D_{\vec{i}}^m : \Sigma_{\vec{i} \cup \vec{j}} \text{ diagonal} \right| \\ &= \binom{q}{m} \frac{1}{(\delta-m)!} \sum_{j_1 \in [p] \setminus \mathcal{NZ}(\vec{i})} \sum_{\substack{j_2 \in [p] \setminus \mathcal{NZ}(\vec{i}) \\ j_2 \notin \mathcal{NZ}(j_1)}} \\ &\quad \sum_{\substack{j_3 \in [p] \setminus \mathcal{NZ}(\vec{i}) \\ j_3 \notin \bigcup_{\ell=1}^2 \mathcal{NZ}(j_\ell)}} \dots \sum_{\substack{j_{\delta-m} \in [p] \setminus \mathcal{NZ}(\vec{i}) \\ j_{\delta-m} \notin \bigcup_{\ell=1}^{\delta-m-1} \mathcal{NZ}(j_\ell)}} 1 \\ &\geq \binom{q}{m} \frac{1}{(\delta-m)!} \prod_{\ell=0}^{\delta-m-1} (p_i - \ell\kappa), \end{aligned} \quad (63)$$

where in the first inequality we assume without loss of generality that the components of  $\vec{j}$  distinct from  $\vec{i}$  are

$j_1, j_2, \dots, j_{\delta-m}$ . Then,

$$\begin{aligned}
& \left| \vec{j} \in \mathcal{K}_1(\vec{i}) \cap D_i^m : \Sigma_{\vec{i} \cup \vec{j}} \text{ not diagonal} \right| \\
& \leq \binom{q}{m} \binom{p-1-q}{\delta-m} - \binom{q}{m} \frac{1}{(\delta-m)!} \prod_{\ell=0}^{\delta-m-1} (p_i - \ell \kappa) \\
& = \frac{\binom{q}{m}}{(\delta-m)!} \left( \prod_{\ell=0}^{\delta-m-1} (p-1-q-\ell) - \prod_{\ell=0}^{\delta-m-1} (p_i - \ell) + \right. \\
& \quad \left. \prod_{\ell=0}^{\delta-m-1} (p_i - \ell) - \prod_{\ell=0}^{\delta-m-1} (p_i - \ell \kappa) \right) \\
& \leq \binom{q}{m} \frac{1}{(\delta-m)!} (\delta-m) p^{\delta-m-1} (q+\delta)(\kappa-1), \quad (64)
\end{aligned}$$

where the first inequality follows from (63), and the second inequality follows from Lemma A.26 (a), (b) and (81).

Then

$$\begin{aligned}
& \sum_{\vec{i} \in C_q^<} I_1(\vec{i}) \\
& = \sum_{m=0}^{\delta-1} \left( \sum_{\substack{\vec{i} \in C_q^< \\ \Sigma_{\vec{i}} \text{ diagonal}}} \sum_{\substack{\vec{j} \in \mathcal{K}_1(\vec{i}) \cap D_i^m \\ \Sigma_{\vec{i} \cup \vec{j}} \text{ diagonal}}} + \sum_{\substack{\vec{i} \in C_q^< \\ \Sigma_{\vec{i}} \text{ diagonal}}} \sum_{\substack{\vec{j} \in \mathcal{K}_1(\vec{i}) \cap D_i^m \\ \Sigma_{\vec{i} \cup \vec{j}} \text{ not diagonal}}} \right) \mathbb{E} \theta_{\vec{i}, \vec{j}} \\
& \quad + \sum_{\substack{\vec{i} \in C_q^< \\ \Sigma_{\vec{i}} \text{ not diagonal}}} \sum_{\substack{\vec{j} \in \mathcal{K}_1(\vec{i}) \cap D_i^m}} \mathbb{E} \theta_{\vec{i}, \vec{j}} \\
& \leq \sum_{m=0}^{\delta-1} \left( \sum_{\substack{\vec{i} \in C_q^< \\ \Sigma_{\vec{i}} \text{ diagonal}}} \sum_{\substack{\vec{j} \in \mathcal{K}_1(\vec{i}) \cap D_i^m \\ \Sigma_{\vec{i} \cup \vec{j}} \text{ diagonal}}} + \right. \\
& \quad \left. \mu_{n,q+\delta+1}(\Sigma) \left( \sum_{\substack{\vec{i} \in C_q^< \\ \Sigma_{\vec{i}} \text{ diagonal}}} \sum_{\substack{\vec{j} \in \mathcal{K}_1(\vec{i}) \cap D_i^m \\ \Sigma_{\vec{i} \cup \vec{j}} \text{ not diagonal}}} + \right. \right. \\
& \quad \left. \left. \sum_{\substack{\vec{i} \in C_q^< \\ \Sigma_{\vec{i}} \text{ not diagonal}}} \sum_{\substack{\vec{j} \in \mathcal{K}_1(\vec{i}) \cap D_i^m}} \right) \right) a z^{q+\delta-m} \\
& \leq \sum_{m=0}^{\delta-1} a z^{q+\delta-m} \left( \binom{p}{1} \binom{p-1}{q} \binom{q}{m} \binom{p-1-q}{\delta-m} + \right. \\
& \quad \mu_{n,q+\delta+1}(\Sigma) p \binom{p-1}{q} \binom{q}{m} \frac{\delta-m}{(\delta-m)!} p^{\delta-m-1} (q+\delta)(\kappa-1) \\
& \quad \left. + \mu_{n,q+\delta+1}(\Sigma) \frac{(q+1)}{2((q-1)!)} p^q (\kappa-1) \binom{q}{m} \binom{p-1-q}{\delta-m} \right) \\
& \leq a p (p z)^{q+1} \eta_1 \frac{1}{(\delta-1)!} \sum_{m=0}^{\delta-1} \frac{(\delta-1)!}{m! (\delta-1-m)!} (p z)^{\delta-1-m} \\
& = a p (p z)^{q+1} \eta_1 \frac{1}{(\delta-1)!} (1 + p z)^{\delta-1}, \quad (65)
\end{aligned}$$

where the first inequality follows from  $\mu_{n,q+\delta-m+1}(\Sigma) \leq \mu_{n,q+\delta+1}(\Sigma)$ , and the second inequality follows from Lemma A.3 and (64).

Obviously,

$$\begin{aligned}
\mathcal{K}_2(\vec{i}) &= \bigcup_{m=0}^{\delta-2} \left( \mathcal{K}_2(\vec{i}) \cap D_i^m \right), \\
\mathcal{K}_3(\vec{i}) &= \bigcup_{m=0}^{\delta-1} \left( \mathcal{K}_3(\vec{i}) \cap D_i^m \right), \\
\mathcal{K}_4(\vec{i}) &= \bigcup_{m=0}^{\delta-1} \left( \mathcal{K}_4(\vec{i}) \cap D_i^m \right), \\
\mathcal{K}_5(\vec{i}) &= \bigcup_{m=1}^{\delta} \left( \mathcal{K}_5(\vec{i}) \cap D_i^m \right).
\end{aligned}$$

Then, following a similar analysis to  $\mathcal{K}_1(\vec{i})$ , additionally with Lemma A.5, we obtain, (66)–(69), shown at the bottom of the next page.

The detailed derivation of the above inequalities are omitted for the sake of brevity.

Observe that

$$\begin{aligned}
\mathcal{K}_6(\vec{i}) &= \left\{ \vec{j} \in C_q^< : \left( \bigcup_{\ell=0}^q \{i_\ell\} \right) \cap \left( \bigcup_{\ell=0}^{\delta} \{j_\ell\} \right) = \emptyset, \right. \\
& \quad \left. \exists \ell \in [\delta] \cup \{0\} \text{ such that } j_\ell \in \mathcal{NZ}(\vec{i}) \right\}.
\end{aligned}$$

Then,

$$\begin{aligned}
|\mathcal{K}_6(\vec{i})| &= \binom{p-1-q}{1} \binom{p-2-q}{\delta} - \binom{p_i}{1} \binom{p_i-1}{\delta} \\
& \leq \frac{1}{\delta!} (\delta+1) p^\delta (q+1) (\kappa-1), \quad (70)
\end{aligned}$$

where the last inequality follows from Lemma A.26 (a) and (81). Thus,

$$\begin{aligned}
& \sum_{\vec{i} \in C_q^<} I_6(\vec{i}) \\
& \leq \binom{p}{1} \binom{p-1}{q} \frac{1}{\delta!} (\delta+1) p^\delta (q+1) (\kappa-1) \mu_{n,q+\delta+2}(\Sigma) a z^{q+\delta}, \\
& \leq a p^2 (p z)^{q+\delta} \frac{(\delta+1)(q+1)}{\delta! q!} \mu_{n,q+\delta+2}(\Sigma) \frac{\kappa-1}{p}, \quad (71)
\end{aligned}$$

where the first inequality follows from (70) and Lemma III.4 (a).

**Case 2:**  $p < q + \delta + 2$

We have imposed the condition  $p \geq 2\delta + 2$  to derive (65), (66), (67), (68), (69) and (71). However, one can verify directly that these inequalities also hold when  $p < q + \delta + 2$ . We omit these tedious verifications here and take it for granted (65), (66), (67), (68), (69) and (71) holds for all  $1 \leq \delta \leq q \leq p-1$ .

Thus combining (62), (65), (66), (67), (68), (69) and (71), yield

$$\begin{aligned}
& \sum_{\vec{i} \in C_q^<} \sum_{\vec{j} \in N_i^*} \mathbb{E} \theta_{\vec{i}, \vec{j}} \\
& \leq a p (p z)^{q+1} \left( 1 + \mu_{n,q+\delta+1}(\Sigma) (3q^2) \frac{\kappa-1}{p} \right) (1 + p z)^{\delta-1} \delta \frac{4+b/z}{(\delta-1)!} \\
& \quad + a p^2 (p z)^{q+\delta} \frac{(\delta+1)(q+1)}{\delta! q!} \mu_{n,q+\delta+2}(\Sigma) \frac{\kappa-1}{p}.
\end{aligned}$$

□

Combining Lemma A.6 and Lemma A.7 immediately yields the following lemma.

**Lemma A.8:** Let  $p \geq n \geq 4$  and  $\mathbf{X} \sim \mathcal{VE}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, g)$ . Suppose  $\boldsymbol{\Sigma}$  is row- $\kappa$  sparse. Consider any  $\theta_{\vec{i}, \vec{j}}$  that is a non-negative function of  $\mathbf{u}_\ell$  for  $\ell \in \vec{i} \cup \vec{j}$  defined for  $\vec{i} \in C_q^<$  and  $\vec{j} \in C_\delta^<$  with  $1 \leq \delta \leq q \leq p-1$ . Suppose there exist  $a, z, b$  such that all conditions in Lemma A.6 and in Lemma A.7 hold. Moreover suppose  $b/z \leq c_{n, \delta, q}$  for some positive constant  $c_{n, \delta, q}$  that only depends on  $n, q$  and  $\delta$ . Then

$$\begin{aligned} & \sum_{\vec{i} \in C_q^<} \sum_{\vec{j} \in C_\delta^<} \mathbb{E} \theta_{\vec{i}, \vec{j}} \\ & \leq C_{n, q, \delta} \left( p^{1+\frac{1}{\delta}} z \right)^q \left( 1 + (p^{1+\frac{1}{\delta}} z)^\delta \right) (1 + pz)^{\delta-1} (1 + \\ & \quad \mu_{n, q+\delta+2}(\boldsymbol{\Sigma}) \frac{\kappa-1}{p}) a p^{1-\frac{q}{\delta}}. \end{aligned}$$

3) *Proof of Proposition III.6:* In what follows in this subsection, for the sake of brevity, we write  $C^<$  for  $C_\delta^<$ , and write  $\Phi_{\vec{i}}$  for  $\Phi_{\vec{i}}^{(R)}$ , for any  $\vec{i} \in C^<$ .

*Proof of Proposition III.6:* Recall

$$N_{E_\delta}^{(R)} = \sum_{\vec{i} \in C_\delta^<} \prod_{j=1}^{\delta} \Phi_{i_0 i_j}^{(R)} = \sum_{\vec{i} \in C_\delta^<} \Phi_{\vec{i}}.$$

To apply a Compound Poisson Approximation result, some additional notation need to be introduced. For  $\vec{i} \in C^<$ , let  $S_{\vec{i}}$  be defined in (49), and let  $T_{\vec{i}}, N_{\vec{i}}$  be defined respectively as in (60), (61) with  $q = \delta$ . Here  $T_{\vec{i}}$  is the set of indices consisting of coordinates outside the neighborhood of  $\vec{i}$ ;  $N_{\vec{i}}$  is the set of “correlated but not highly correlated” indices, i.e. the set of indices of which at least one component but not every component is in the neighborhood of  $\vec{i}$ . Denote

$$W_{\vec{i}} = \sum_{\vec{j} \in T_{\vec{i}}} \Phi_{\vec{j}}, \quad Z_{\vec{i}} = \sum_{\vec{j} \in N_{\vec{i}}} \Phi_{\vec{j}}, \quad (72)$$

and recall  $U_{\vec{i}} = \sum_{\vec{j} \in S_{\vec{i}}} \Phi_{\vec{j}}^{(R)}$  is defined in (50). Then  $W_{\vec{i}}$  is independent of  $U_{\vec{i}}$  and  $\Phi_{\vec{i}}$ . Further denote

$$\begin{aligned} \lambda_0 &= \sum_{\vec{i} \in C^<} \mathbb{E} \left( \frac{\Phi_{\vec{i}}}{\Phi_{\vec{i}} + U_{\vec{i}}} 1(\Phi_{\vec{i}} + U_{\vec{i}} \geq 1) \right), \\ \zeta_{0\ell} &= \frac{1}{\lambda_0 \ell} \sum_{\vec{i} \in C^<} \mathbb{E} (\Phi_{\vec{i}} 1(\Phi_{\vec{i}} + U_{\vec{i}} = \ell)), \quad \forall \ell \geq 1 \end{aligned} \quad (73)$$

and a probability distribution  $\zeta_0$  on positive integers with  $\zeta_0(\ell) = \zeta_{0\ell}$ . The mean of  $\zeta_0$  is  $\mathbb{E} \zeta_0 = \sum_{\ell \geq 1} \ell \zeta_{0\ell}$ . Moreover, let  $b_1 = \sum_{\vec{i} \in C^<} \mathbb{E} \Phi_{\vec{i}} \mathbb{E} (\Phi_{\vec{i}} + U_{\vec{i}} + Z_{\vec{i}})$  and

$$b_2 = \sum_{\vec{i} \in C^<} \mathbb{E} (\Phi_{\vec{i}} Z_{\vec{i}}). \quad (74)$$

In this proof we write  $\lambda$  and  $\zeta$  for  $\lambda_{p, n, \delta, \rho}$  and  $\zeta_{n, \delta, \rho}(\ell)$  respectively when there is no confusion. By the compound Poisson Stein’s approximation, i.e. (5.19) and (5.16) in [38],

$$\begin{aligned} & d_{TV} \left( \mathcal{L} \left( N_{E_\delta}^{(R)} \right), \text{CP}(\lambda, \zeta) \right) \\ & \leq e^{\lambda_0} (b_1 + b_2 + \lambda_0 d_W(\zeta'_0, \zeta') \mathbb{E} \zeta_0 + |\lambda_0 \mathbb{E} \zeta_0 - \lambda \mathbb{E} \zeta|), \end{aligned} \quad (75)$$

where  $\zeta'_0(\ell) = \ell \zeta_{0\ell} / \mathbb{E} \zeta_0$  and  $\zeta'(\ell) = \ell \zeta(\ell) / \mathbb{E} \zeta$  for  $\ell \in \mathbb{Z}_+$ , the set of all positive integers. In (75), the distance  $d_W$  is the Wasserstein  $L_1$  metric on probability measures over the set of positive integers  $\mathbb{Z}_+$ :

$$d_W(P, Q) = \sup_{f \in \text{Lip}_1} \left| \int f dP - \int f dQ \right|$$

where  $\text{Lip}_1 = \{f : |f(r) - f(s)| \leq |r - s|, r, s \in \mathbb{Z}_+\}$ .

By Lemma A.23,

$$\begin{aligned} & \lambda_0 d_W(\zeta'_0, \zeta') \mathbb{E} \zeta_0 \\ & \leq \lambda_0 \mathbb{E} \zeta_0 \frac{\delta}{2} \sum_{\ell=1}^{\delta+1} \ell \left| \frac{\lambda_0 \zeta_{0\ell}}{\lambda_0 \mathbb{E} \zeta_0} - \frac{\lambda \zeta(\ell)}{\lambda \mathbb{E} \zeta} \right| \\ & = \frac{\delta}{2} \sum_{\ell=1}^{\delta+1} \ell \left| (\lambda_0 \zeta_{0\ell} - \lambda \zeta(\ell)) + (\lambda \mathbb{E} \zeta - \lambda_0 \mathbb{E} \zeta_0) \frac{\lambda \zeta(\ell)}{\lambda \mathbb{E} \zeta} \right| \\ & \leq \frac{\delta}{2} \sum_{\ell=1}^{\delta+1} \ell |\lambda_0 \zeta_{0\ell} - \lambda \zeta(\ell)| + \frac{\delta}{2} |\lambda \mathbb{E} \zeta - \lambda_0 \mathbb{E} \zeta_0| \sum_{\ell=1}^{\delta+1} \ell \frac{\lambda \zeta(\ell)}{\lambda \mathbb{E} \zeta} \\ & \leq \delta \sum_{\ell=1}^{\delta+1} \ell |\lambda_0 \zeta_{0\ell} - \lambda \zeta(\ell)|. \end{aligned}$$

Plugging the above inequalities into (75),

$$\begin{aligned} & d_{TV} \left( \mathcal{L} \left( N_{E_\delta}^{(R)} \right), \text{CP}(\lambda, \zeta) \right) \\ & \leq e^{\lambda_0} \left( b_1 + b_2 + (\delta + 1) \sum_{\ell=1}^{\delta+1} \ell |\lambda_0 \zeta_{0\ell} - \lambda \zeta(\ell)| \right). \end{aligned} \quad (76)$$

It remains to estimate the quantities in the right hand side of (76).

$$\sum_{\vec{i} \in C_q^<} I_2(\vec{i}) \leq ap(pz)^{q+1} \left( 1 + \mu_{n, q+\delta}(\boldsymbol{\Sigma}) (3q^2) \frac{\kappa-1}{p} \right) \frac{1}{(\delta-2)!} (1+pz)^{\delta-2} 1(\delta \geq 2), \quad (66)$$

$$\sum_{\vec{i} \in C_q^<} I_3(\vec{i}) \leq ap(pz)^{q+1} \left( 1 + \mu_{n, q+\delta+1}(\boldsymbol{\Sigma}) (3q^2) \frac{\kappa-1}{p} \right) \frac{1}{(\delta-1)!} (1+pz)^{\delta-1}, \quad (67)$$

$$\sum_{\vec{i} \in C_q^<} I_4(\vec{i}) \leq ap(pz)^{q+1} \left( 1 + \mu_{n, 2\delta+1}(\boldsymbol{\Sigma}) (3q^2) \frac{\kappa-1}{p} \right) \frac{1}{(\delta-1)!} (1+pz)^{\delta-1}, \quad (68)$$

$$\sum_{\vec{i} \in C_q^<} I_5(\vec{i}) \leq a \left( \frac{b}{z} \right) p(pz)^{q+1} \left( 1 + \mu_{n, q+\delta+1}(\boldsymbol{\Sigma}) (3\delta^2) \frac{\kappa-1}{p} \right) \frac{1}{(\delta-1)!} (1+pz)^{\delta-1}. \quad (69)$$

**Part I. Upper bound for  $\lambda_0$  and  $\sum_{\ell=1}^{\delta+1} \ell |\lambda_0 \zeta_{0\ell} - \lambda \zeta(\ell)|$**   
 For  $\ell \in [\delta + 1]$ ,

$$\begin{aligned}
 & |\lambda_0 \zeta_{0\ell} - \lambda \zeta(\ell)| \\
 &= \left| \lambda_0 \zeta_{0\ell} - \frac{1}{\ell} \binom{p}{1} \binom{p-1}{\delta} (2P_n(r_\rho))^\delta \alpha(\ell, r_\rho) \right| \\
 &\leq \frac{1}{\ell} \sum_{\vec{i} \in C^<} |\mathbb{E}(\Phi_{\vec{i}} 1(\Phi_{\vec{i}} + U_{\vec{i}} = \ell)) - (2P_n(r_\rho))^\delta \alpha(\ell, r_\rho)| \\
 &= \frac{1}{\ell} \sum_{\substack{\vec{i} \in C^< \\ \Sigma_{\vec{i}} \text{ not diagonal}}} |\mathbb{E}(\Phi_{\vec{i}} 1(\Phi_{\vec{i}} + U_{\vec{i}} = \ell)) - (2P_n(r_\rho))^\delta \alpha(\ell, r_\rho)| \\
 &\leq \frac{1}{\ell} (\mu_{n,\delta+1}(\Sigma) + 1) (2P_n(r_\rho))^\delta \alpha(\ell, r_\rho) \sum_{\substack{\vec{i} \in C^< \\ \Sigma_{\vec{i}} \text{ not diagonal}}} 1 \\
 &\leq \frac{\alpha(\ell, r_\rho)}{\ell} \mu_{n,\delta+1}(\Sigma) \gamma^\delta \frac{(\delta+1) \kappa - 1}{(\delta-1)! p}, \quad (77)
 \end{aligned}$$

where the first inequality follows from the definition of  $\zeta_{0\ell}$  in (73), the second inequality follows from Lemma A.2 and Lemma III.4 (a), and the last inequality follows from Lemma A.3 and  $\mu_{n,\delta+1}(\Sigma) \geq 1$ .

Then,

$$\begin{aligned}
 & |\lambda_0 \mathbb{E} \zeta_0 - \lambda \mathbb{E} \zeta| \\
 &\leq \sum_{\ell=1}^{\delta+1} \ell |\lambda_0 \zeta_{0\ell} - \lambda \zeta(\ell)| \leq \mu_{n,\delta+1}(\Sigma) \gamma^\delta \frac{(\delta+1) \kappa - 1}{(\delta-1)! p}, \quad (78)
 \end{aligned}$$

where the last inequality follows from (77). As an immediate consequences,

$$\begin{aligned}
 \lambda_0 &\leq \lambda_0 \mathbb{E} \zeta_0 \\
 &\leq |\lambda_0 \mathbb{E} \zeta_0 - \lambda \mathbb{E} \zeta| + \lambda \mathbb{E} \zeta \\
 &\leq \mu_{n,\delta+1}(\Sigma) \gamma^\delta \frac{(\delta+1) \kappa - 1}{(\delta-1)! p} + \binom{p}{1} \binom{p-1}{\delta} (2P_n(r_\rho))^\delta \\
 &\leq \gamma^\delta \frac{(\delta+1)}{(\delta-1)!} \left( \mu_{n,\delta+1}(\Sigma) \frac{\kappa - 1}{p} + 1 \right), \quad (79)
 \end{aligned}$$

where the third inequality follows from (78).

**Part II. Upper bound for  $b_1$**

Since  $N_{\vec{i}} \cup S_{\vec{i}} \cup \{\vec{i}\} = C^< \setminus T_{\vec{i}}$ ,

$$b_1 = \sum_{\vec{i} \in C^<} \sum_{\vec{j} \in C^< \setminus T_{\vec{i}}} \mathbb{E} \Phi_{\vec{i}} \mathbb{E} \Phi_{\vec{j}}. \quad (80)$$

Given  $\vec{i} \in C^<$ , by (58) with  $q = \delta$ ,

$$p_{\vec{i}} := \left| [p] \setminus \mathcal{N} \mathcal{Z} \left( \begin{smallmatrix} \vec{i} \\ i \end{smallmatrix} \right) \right| \geq p - (\delta + 1) \kappa. \quad (81)$$

Since  $|T_{\vec{i}}| = p_{\vec{i}} \binom{p_{\vec{i}}-1}{\delta}$ ,

$$\begin{aligned}
 & |C^< \setminus T_{\vec{i}}| \\
 &= p \binom{p-1}{\delta} - p_{\vec{i}} \binom{p_{\vec{i}}-1}{\delta} \\
 &\leq \frac{1}{\delta!} (\delta + 1) \left( \prod_{\alpha=0}^{\delta-1} (p - \alpha) \right) (p - p_{\vec{i}}) \leq \frac{(\delta+1)^2}{\delta!} p^\delta \kappa, \quad (82)
 \end{aligned}$$

where the first inequality follows from Lemma A.26 (a).

One straightforward upper bound is

$$\begin{aligned}
 b_1 &= \sum_{\vec{i} \in C^<} \sum_{\vec{j} \in C^< \setminus T_{\vec{i}}} \mathbb{E} \left( \prod_{l=1}^{\delta} \Phi_{i_0 i_l}^{(R)} \right) \mathbb{E} \left( \prod_{l'=1}^{\delta} \Phi_{j_0 j_{l'}}^{(R)} \right) \\
 &\leq p \binom{p-1}{\delta} \frac{(\delta+1)^2}{\delta!} p^\delta \kappa (\mu_{n,\delta+1}(\Sigma))^2 (2P_n(r_\rho))^{2\delta} \\
 &\leq (\mu_{n,\delta+1}(\Sigma))^2 \frac{(\delta+1)^2}{(\delta!)^2} \left( 2p^{1+\frac{1}{\delta}} P_n(r_\rho) \right)^{2\delta} \frac{\kappa}{p},
 \end{aligned}$$

where the first inequality follows from Lemma A.2, Lemma III.4 (a) and (82). The  $(\mu_{n,\delta+1}(\Sigma))^2$  in the above upper bound is not very satisfactory, and can be improved by a more involved analysis.

Observe for given  $\vec{i} \in C^<$ ,

$$\begin{aligned}
 & \left| \{ \vec{j} \in C^< \setminus T_{\vec{i}} : \Sigma_{\vec{j}} \text{ not diagonal} \} \right| \\
 &= |C^< \setminus T_{\vec{i}}| - \left( \left| \{ \vec{j} \in C^< : \Sigma_{\vec{j}} \text{ diagonal} \} \right| - \left| \{ \vec{j} \in T_{\vec{i}} : \Sigma_{\vec{j}} \text{ diagonal} \} \right| \right) \\
 &\leq \frac{1}{\delta!} (\delta + 1) \left( \prod_{\alpha=0}^{\delta-1} (p - \alpha) \right) (p - p_{\vec{i}}) - \left( \left| \{ \vec{j} \in C^< : \Sigma_{\vec{j}} \text{ diagonal} \} \right| - \left| \{ \vec{j} \in T_{\vec{i}} : \Sigma_{\vec{j}} \text{ diagonal} \} \right| \right), \quad (83)
 \end{aligned}$$

where the inequality follows from (82). Then, (84) and (85), shown at the bottom of the next page, where the second equality follows by writing (84) as a telescoping sum with the convention that the summation over  $j_{-1}$  for  $m = 0$  and the summation over  $j_{\delta+1}$  for  $m = \delta$  vanish, and the third equality follows from changing the order of the summation for  $m \geq 1$ . Plugging (85) into (83),

$$\begin{aligned}
 & \left| \{ \vec{j} \in C^< \setminus T_{\vec{i}} : \Sigma_{\vec{j}} \text{ not diagonal} \} \right| \\
 &\leq \frac{1}{\delta!} (\delta + 1) (p - p_{\vec{i}}) \left( \prod_{\alpha=0}^{\delta-1} (p - \alpha) - \prod_{\alpha=0}^{\delta-1} (p_{\vec{i}} - 1 - \alpha) \right) \\
 &\quad \left( \prod_{\alpha=1}^{\delta} (p_{\vec{i}} - \alpha) - \prod_{\alpha=1}^{\delta} (p_{\vec{i}} - \alpha \kappa) \right) \\
 &\leq \frac{1}{\delta!} (\delta + 1) (p - p_{\vec{i}}) (\delta p^{\delta-1} (p - p_{\vec{i}} + 1) + \frac{\delta(\delta+1)}{2} p^{\delta-1} (\kappa - 1)) \\
 &\leq \frac{3\delta(\delta+1)^3}{\delta!} p^{\delta-1} \kappa^2, \quad (86)
 \end{aligned}$$

where the second inequality follows from Lemma A.26 (a) and Lemma A.26 (b), and the last inequality follows from (81).

Then for any  $\vec{i} \in C^<$ ,

$$\begin{aligned}
 & \sum_{\vec{j} \in C^< \setminus T_{\vec{i}}} \mu_{n,\delta+1}(\Sigma_{\vec{j}}) \\
 &\leq \sum_{\substack{\vec{j} \in C^< \setminus T_{\vec{i}} \\ \Sigma_{\vec{j}} \text{ diagonal}}} 1 + \mu_{n,\delta+1}(\Sigma) \sum_{\substack{\vec{j} \in C^< \setminus T_{\vec{i}} \\ \Sigma_{\vec{j}} \text{ not diagonal}}} 1 \\
 &\leq \frac{(\delta+1)^2}{\delta!} p^\delta \kappa + \mu_{n,\delta+1}(\Sigma) \frac{3\delta(\delta+1)^3}{\delta!} p^{\delta-1} \kappa^2,
 \end{aligned}$$

$$\leq \frac{3\delta(\delta+1)^3}{\delta!} p^\delta \kappa \left(1 + \mu_{n,\delta+1}(\Sigma) \frac{\kappa}{p}\right), \quad (87)$$

where the first inequality follows from Lemma III.4 (a), and the second inequality follows from (82), (86).

Then following (80),

$$\begin{aligned} b_1 &\leq \sum_{\vec{i} \in C^<} \sum_{\vec{j} \in C^< \setminus T_{\vec{i}}} \mu_{n,\delta+1}(\Sigma_{\vec{i}}) \mu_{n,\delta+1}(\Sigma_{\vec{j}}) (2P_n(r_\rho))^{2\delta} \\ &\leq \frac{p^{\delta+1}}{\delta!} \left(1 + \delta^2 \mu_{n,\delta+1}(\Sigma) \frac{\kappa-1}{p}\right) \frac{3\delta(\delta+1)^3}{\delta!} p^\delta \kappa (1 + \mu_{n,\delta+1}(\Sigma) \frac{\kappa}{p}) (2P_n(r_\rho))^{2\delta} \\ &\leq \left(3 \frac{\delta^3(\delta+1)^3}{(\delta!)^2} (2p^{1+\frac{1}{\delta}} P_n(r_\rho))^{2\delta}\right) \frac{\kappa}{p} \left(1 + \mu_{n,\delta+1}(\Sigma) \frac{\kappa}{p}\right)^2, \end{aligned} \quad (88)$$

where the first inequality follows from Lemma A.2, the second inequality follows from Lemma A.4 and (87).

### Part III. Upper bound for $b_2$

Let  $\mathcal{K}_w(\vec{i})$  and  $D_i^n$  be the same as in Subsection E.2 with  $q = \delta$ . By Lemma III.4 (c), Lemma A.2 and Lemma A.5, it follows that the conditions in Lemma A.7 with  $q = \delta$  and  $\theta_{\vec{i},\vec{j}} = \Phi_{\vec{i}} \Phi_{\vec{j}}$  are satisfied with  $a = 1$ ,  $b = 2P_n(2r_\rho)1(\delta \geq 2) + 2P_n(r_\rho)1(\delta = 1)$  and  $z = 2P_n(r_\rho)$ . Moreover,  $b/z \leq 2^{n-2}1(\delta \geq 2) + 1$  by Lemma A.22 (d). Thus by Lemma A.7 with  $q = \delta$  and  $\theta_{\vec{i},\vec{j}} = \Phi_{\vec{i}} \Phi_{\vec{j}}$ ,

$$\begin{aligned} b_2 &\leq p(2pP_n(r_\rho))^{\delta+1} \eta_1 (1 + 2pP_n(r_\rho))^{\delta-1} \delta \frac{5 + 2^{n-1}1(\delta \geq 2)}{(\delta-1)!} + \\ &\quad \frac{(\delta+1)^2}{(\delta!)^2} p^2 (2pP_n(r_\rho))^{2\delta} \mu_{n,2\delta+2}(\Sigma) \frac{\kappa-1}{p}. \quad (89) \\ &\leq \gamma^{\delta+1} p^{-\frac{1}{\delta}} \eta_1 \left(1 + \gamma p^{-\frac{1}{\delta}}\right)^{\delta-1} \delta \frac{5 + 2^{n-1}1(\delta \geq 2)}{(\delta-1)!} + \\ &\quad \frac{(\delta+1)^2}{(\delta!)^2} \gamma^{2\delta} \mu_{n,2\delta+2}(\Sigma) \frac{\kappa-1}{p}, \quad (90) \end{aligned}$$

---


$$\begin{aligned} &\left| \{\vec{j} \in C^< : \Sigma_{\vec{j}} \text{ diagonal}\} \right| - \left| \{\vec{j} \in T_{\vec{i}} : \Sigma_{\vec{j}} \text{ diagonal}\} \right| \\ &= \frac{1}{\delta!} \sum_{j_0=1}^p \sum_{j_1 \in [p] \setminus \mathcal{NZ}(j_0)} \cdots \sum_{\substack{j_\delta \in [p] \setminus \bigcup_{l=0}^{\delta-1} \mathcal{NZ}(j_l)}} 1 - \frac{1}{\delta!} \sum_{\substack{j_0 \in [p] \\ j_0 \notin \mathcal{NZ}(\vec{i})}} \sum_{\substack{j_1 \in [p] \setminus \mathcal{NZ}(j_0) \\ j_1 \notin \mathcal{NZ}(\vec{i})}} \cdots \sum_{\substack{j_\delta \in [p] \setminus \bigcup_{l=0}^{\delta-1} \mathcal{NZ}(j_l) \\ j_\delta \notin \mathcal{NZ}(\vec{i})}} 1 \quad (84) \\ &= \frac{1}{\delta!} \sum_{m=0}^{\delta} \left( \sum_{\substack{j_0 \in [p] \\ j_0 \notin \mathcal{NZ}(\vec{i})}} \sum_{\substack{j_1 \in [p] \setminus \mathcal{NZ}(j_0) \\ j_1 \notin \mathcal{NZ}(\vec{i})}} \cdots \right. \\ &\quad \left. \sum_{\substack{j_{m-1} \in [p] \setminus \bigcup_{l=0}^{m-2} \mathcal{NZ}(j_l) \\ j_{m-1} \notin \mathcal{NZ}(\vec{i})}} \sum_{\substack{j_m \in [p] \setminus \bigcup_{l=0}^{m-1} \mathcal{NZ}(j_l) \\ j_m \notin \mathcal{NZ}(\vec{i})}} \sum_{\substack{j_{m+1} \in [p] \setminus \bigcup_{l=0}^m \mathcal{NZ}(j_l) \\ j_{m+1} \notin \mathcal{NZ}(\vec{i})}} \cdots \sum_{\substack{j_\delta \in [p] \setminus \bigcup_{l=0}^{\delta-1} \mathcal{NZ}(j_l) \\ j_\delta \notin \mathcal{NZ}(\vec{i})}} 1 \right) \\ &= \frac{1}{\delta!} \sum_{\substack{j_0 \in [p] \\ j_0 \notin \mathcal{NZ}(\vec{i})}} \sum_{j_1 \in [p] \setminus \mathcal{NZ}(j_0)} \cdots \sum_{\substack{j_\delta \in [p] \setminus \bigcup_{l=0}^{\delta-1} \mathcal{NZ}(j_l)}} 1 + \frac{1}{\delta!} \sum_{m=1}^{\delta} \left( \right. \\ &\quad \left. \sum_{\substack{j_m \in \mathcal{NZ}(\vec{i})}} \sum_{\substack{j_0 \in [p] \\ j_0 \notin \mathcal{NZ}(\vec{i}) \\ j_0 \notin \mathcal{NZ}(j_m)}} \sum_{\substack{j_1 \in [p] \setminus \mathcal{NZ}(j_0) \\ j_1 \notin \mathcal{NZ}(j_m)}} \cdots \sum_{\substack{j_{m-1} \in [p] \setminus \bigcup_{l=0}^{m-2} \mathcal{NZ}(j_l) \\ j_{m-1} \notin \mathcal{NZ}(j_m)}} \sum_{\substack{j_{m+1} \in [p] \setminus \bigcup_{l=0}^m \mathcal{NZ}(j_l)}} \cdots \sum_{\substack{j_\delta \in [p] \setminus \bigcup_{l=0}^{\delta-1} \mathcal{NZ}(j_l)}} 1 \right) \\ &\geq \frac{1}{\delta!} (p - p_{\vec{i}}) \prod_{\beta=1}^{\delta} (p - \beta\kappa) + \frac{1}{\delta!} \sum_{m=1}^{\delta} (p - p_{\vec{i}}) \left( \prod_{\alpha=1}^m (p_{\vec{i}} - \alpha\kappa) \right) \left( \prod_{\beta=m+1}^{\delta} (p - \beta\kappa) \right) \\ &\geq \frac{(\delta+1)}{\delta!} (p - p_{\vec{i}}) \prod_{\alpha=1}^{\delta} (p_{\vec{i}} - \alpha\kappa), \quad (85) \end{aligned}$$

where  $\eta_1 = \left(1 + \mu_{n,2\delta+2}(\Sigma) (3\delta^2)^{\frac{\kappa-1}{p}}\right)$ , and the last step follows from  $2p^{1+\frac{1}{\delta}} P_n(r_\rho) \leq \gamma$ .

By combining (76), (78), (79), (88), (90), together with the assumption  $2p^{1+\frac{1}{\delta}} P_n(r_\rho) \leq \gamma$ ,

$$\begin{aligned} & d_{\text{TV}} \left( \mathcal{L} \left( N_{E_\delta}^{(\mathbf{R})} \right), \text{CP}(\lambda, \zeta) \right) \\ & \leq C_{n,\delta,\gamma} (C'_{\delta,\gamma})^{\mu_{n,\delta+1}(\Sigma)(\kappa-1)/p} \left( p^{-\frac{1}{\delta}} + \right. \\ & \quad \left. \mu_{n,2\delta+2}(\Sigma) \kappa/p \left( 1 + \mu_{n,2\delta+2}(\Sigma) (\kappa/p)^2 \right) \right), \end{aligned}$$

where

$$C'_{\delta,\gamma} = \exp \left( \gamma^\delta \frac{\delta+1}{(\delta-1)!} \right), \quad (91)$$

and

$$C_{n,\delta,\gamma} = C \frac{\delta^6 + \delta^2 2^{n-1} 1(\delta \geq 2)}{\delta!} \gamma^{\delta+1} (1+\gamma)^\delta C'_{\delta,\gamma}. \quad (92)$$

#### F. Proofs in Subsection III-D

##### 1) Proof of Lemma III.8:

*Proof of Lemma III.8:*  $N_{\check{V}_\delta}^{(\mathbf{R})} \leq N_{V_\delta}^{(\mathbf{R})} \leq N_{E_\delta}^{(\mathbf{R})}$  follows trivially from their definitions. It remains to show

$$N_{E_\delta}^{(\mathbf{R})} \leq (\delta+1) N_{E_{\delta+1}}^{(\mathbf{R})} + N_{\check{V}_\delta}^{(\mathbf{R})}. \quad (93)$$

To see this, consider  $\delta \geq 2$  and any vertex  $i$  and denote its degree by  $m$ . If  $m < \delta$ , then it contributes zero to both sides of (93). If  $m = \delta$ , then it contributes 1 to both sides of (93). If  $m > \delta$ , it contributes  $\binom{m}{\delta}$  to left hand side of (93), while contributes  $(\delta+1)\binom{m}{\delta+1} = (m-\delta)\binom{m}{\delta}$ . The above observation proves (93). The case  $\delta = 1$  is similar and is omitted.

The above proof indeed applies to any graph and, in particular, the empirical partial correlation graph. So the second equation in the statement of the lemma holds.  $\square$

2) *Proof of Proposition III.9 (a):* By (29), it suffices to establish an upper bound on  $\mathbb{E} N_{E_{\delta+1}}^{(\mathbf{R})}$ .

*Lemma A.9:* Let  $\mathbf{X} \sim \mathcal{VE}(\mu, \Sigma, g)$ . Suppose  $\Sigma$  is row- $\kappa$  sparse. Let  $\ell \in [p-1]$ . Then

$$\mathbb{E} N_{E_\ell}^{(\mathbf{R})} \leq \frac{1}{\ell!} \left( 1 + \ell^2 \mu_{n,\ell+1}(\Sigma) \frac{\kappa-1}{p} \right) p (2p P_n(r_\rho))^\ell.$$

*Proof:*

$$\begin{aligned} \mathbb{E} N_{E_\ell}^{(\mathbf{R})} &= \sum_{\vec{i} \in C_\ell^<} \mathbb{E} \prod_{j=1}^{\ell} \Phi_{i_0 i_j}^{(\mathbf{R})} \\ &\leq \sum_{\vec{i} \in C_\ell^<} \mu_{n,\ell+1}(\Sigma_{\vec{i}}) (2P_n(r_\rho))^\ell \\ &\leq \frac{1}{\ell!} \left( 1 + \ell^2 \mu_{n,\ell+1}(\Sigma) \frac{\kappa-1}{p} \right) p (2p P_n(r_\rho))^\ell, \end{aligned}$$

where the first inequality follows from Lemma A.2, and the second inequality follows from Lemma A.4.  $\square$

*Proof of Proposition III.9 (a):* It follows from (29), Lemma A.9 and Lemma A.24.  $\square$

3) *Proof of Proposition III.9 (b):* Similar to (47) and (48), denote

$$\Phi_{\vec{i}}^{(\mathbf{R})} = \prod_{j=1}^{\delta} \Phi_{i_0 i_j}^{(\mathbf{R})} = 1 \left( \bigcap_{j=1}^{\delta} \{\text{dist}(\mathbf{u}_{i_0}, \mathbf{u}_{i_j}) \leq r_\rho\} \right).$$

Then by definition

$$N_{E_\delta}^{(\mathbf{P})} = \sum_{\vec{i} \in C_\delta^<} \Phi_{\vec{i}}^{(\mathbf{P})}. \quad (94)$$

By (48) and (94),

$$\left| N_{E_\delta}^{(\mathbf{P})} - N_{E_\delta}^{(\mathbf{R})} \right| \leq \sum_{\vec{i} \in C_\delta^<} |\Phi_{\vec{i}}^{(\mathbf{P})} - \Phi_{\vec{i}}^{(\mathbf{R})}|.$$

The next three lemmas establish upper bound on  $|\Phi_{\vec{i}}^{(\mathbf{P})} - \Phi_{\vec{i}}^{(\mathbf{R})}|$ .

We may suppose  $\Sigma$  is  $(\tau, \kappa)$  sparse throughout this proof and the proof of Proposition III.9 (c) since the conclusion is invariant to permutation of the variables by Remark II.8. As a result, the U-scores may be partitioned into  $\tilde{\mathbf{U}} \in \mathbb{R}^{(n-1) \times \tau}$  consisting of the first  $\tau$  columns and  $\check{\mathbf{U}} \in \mathbb{R}^{(n-1) \times (p-\tau)}$  consisting the remaining  $p-\tau$  columns.

Denote  $[\tau] = \{1, 2, \dots, \tau\}$ . Define a matrix  $\check{\mathbf{B}}$  by

$$\check{\mathbf{B}} = \frac{n-1}{p-\tau} \check{\mathbf{U}} [\check{\mathbf{U}}]^\top = \frac{n-1}{p-\tau} \sum_{i \in [p] \setminus [\tau]} \mathbf{u}_i \mathbf{u}_i^\top. \quad (95)$$

Denote  $\check{\mathbf{Q}} = \sqrt{n-1} \check{\mathbf{U}}$ . Observe that  $\check{\mathbf{Q}}$  has exactly  $p-\tau$  independent columns and each column  $\sqrt{n-1} \mathbf{u}_i \sim \text{unif}(\sqrt{n-1} S^{n-2})$ . These observations immediately give us part (a) of the following.

*Lemma A.10:* Let  $\{\mathbf{u}_\alpha\}_{\alpha=1}^p$  be columns of  $\mathbf{U}$  defined in Section III-A. Let  $\check{\mathbf{B}}$  be defined as in equation (95).

- (a) Suppose  $\Sigma$  is  $(\tau, \kappa)$  sparse.  $\check{\mathbf{B}} = \frac{1}{p-\tau} \mathbf{Q} \mathbf{Q}^\top$ , where  $\mathbf{Q} \in \mathbb{R}^{(n-1) \times (p-\tau)}$  has independent columns with each column distributed as  $\text{unif}(\sqrt{n-1} S^{n-2})$ .
- (b)  $|\lambda_{\max} \left( \frac{p}{p-\tau} \mathbf{B} \right) - \lambda_{\max}(\check{\mathbf{B}})| \leq \frac{n-1}{p-\tau} \tau$ , and  $\lambda_{\min} \left( \frac{p}{p-\tau} \mathbf{B} \right) \geq \lambda_{\min}(\check{\mathbf{B}})$ .

*Proof:* (b) Recall  $\mathbf{B} = \frac{n-1}{p} \sum_{i=1}^p \mathbf{u}_i \mathbf{u}_i^\top$ . Then,

$$\frac{p}{p-\tau} \mathbf{B} - \check{\mathbf{B}} = \frac{n-1}{p-\tau} \sum_{i \in [\tau]} \mathbf{u}_i \mathbf{u}_i^\top.$$

By Lemma A.27 (a), we have:

$$\begin{aligned} |\lambda_{\max} \left( \frac{p}{p-\tau} \mathbf{B} \right) - \lambda_{\max}(\check{\mathbf{B}})| &\leq \left\| \frac{n-1}{p-\tau} \sum_{i \in [\tau]} \mathbf{u}_i \mathbf{u}_i^\top \right\|_2 \\ &\leq \frac{n-1}{p-\tau} \sum_{i \in [\tau]} \|\mathbf{u}_i \mathbf{u}_i^\top\|_2 \\ &\leq \frac{n-1}{p-\tau} \tau, \end{aligned}$$

where for the last inequality, we use the fact that  $\mathbf{u}_i \in S^{n-2}$ . Moreover, by Lemma A.27 (c), we get  $\lambda_{\min} \left( \frac{p}{p-\tau} \mathbf{B} \right) \geq \lambda_{\min}(\check{\mathbf{B}})$ .  $\square$

Denote  $h_0(\check{\mathbf{B}}) = \frac{\lambda_{\max}(\check{\mathbf{B}}) + \frac{n-1}{p-\tau}\tau}{\lambda_{\min}(\check{\mathbf{B}})} = \frac{S_{\max}(\check{\mathbf{B}}) + \frac{n-1}{p-\tau}\tau}{S_{\min}(\check{\mathbf{B}})}$  to be the perturbational condition number of  $\check{\mathbf{B}}$ , where  $\lambda_{\max}(\check{\mathbf{B}})$ ,  $\lambda_{\min}(\check{\mathbf{B}})$ ,  $S_{\max}(\check{\mathbf{B}})$ , and  $S_{\min}(\check{\mathbf{B}})$  are respectively the largest eigenvalue, smallest eigenvalue, largest singular value and smallest singular value of  $\check{\mathbf{B}}$ .

*Lemma A.11:* Suppose  $p \geq n$ . Let  $\{\mathbf{u}_\alpha\}_{\alpha=1}^p$  and  $\{\mathbf{y}_\alpha\}_{\alpha=1}^p$  be defined as in Section III-A. Consider distinct  $i, j$  satisfying  $1 \leq i, j \leq p$ . Then with probability 1,

$$\frac{1}{h_0(\check{\mathbf{B}})} \|\mathbf{u}_i - \mathbf{u}_j\|_2 \leq \|\mathbf{y}_i - \mathbf{y}_j\|_2 \leq h_0(\check{\mathbf{B}}) \|\mathbf{u}_i - \mathbf{u}_j\|_2,$$

and

$$\frac{1}{h_0(\check{\mathbf{B}})} \|\mathbf{u}_i + \mathbf{u}_j\|_2 \leq \|\mathbf{y}_i + \mathbf{y}_j\|_2 \leq h_0(\check{\mathbf{B}}) \|\mathbf{u}_i + \mathbf{u}_j\|_2.$$

*Proof:* Recall  $\mathbf{y}_\alpha = \bar{\mathbf{y}}_\alpha / \|\bar{\mathbf{y}}_\alpha\|_2$  and  $\bar{\mathbf{y}}_\alpha = \mathbf{B}^{-1} \mathbf{u}_\alpha$  a.s., for  $\alpha = i, j$ . Apply the upper bound in Lemma A.28,

$$\begin{aligned} \|\mathbf{y}_i - \mathbf{y}_j\|_2 &\leq \frac{\lambda_{\max}(\mathbf{B}^{-1})}{\lambda_{\min}(\mathbf{B}^{-1})} \|\mathbf{u}_i - \mathbf{u}_j\|_2 \quad \text{a.s.} \\ &= \frac{\lambda_{\max}\left(\frac{p}{p-\tau} \mathbf{B}\right)}{\lambda_{\min}\left(\frac{p}{p-\tau} \mathbf{B}\right)} \|\mathbf{u}_i - \mathbf{u}_j\|_2 \\ &\leq \frac{\lambda_{\max}(\check{\mathbf{B}}) + \frac{n-1}{p-\tau}\tau}{\lambda_{\min}(\check{\mathbf{B}})} \|\mathbf{u}_i - \mathbf{u}_j\|_2, \end{aligned}$$

where the last inequality follows from Lemma A.10 (b). The lower bound of the first desired display follows similarly, by the lower bound in Lemma A.28.

The second desired expression follows analogously.  $\square$

For  $\{i, j\} \in [p]$  with  $i \neq j$ ,  $q \in \{-1, +1\}$ , define

$$\begin{aligned} S_{ij}^{(q)}(r_\rho) &= \{\|\mathbf{y}_i - q\mathbf{y}_j\|_2 \leq r_\rho\}, \\ F_{ij}^{(q)}(r_\rho) &= \{\|\mathbf{u}_i - q\mathbf{u}_j\|_2 \leq r_\rho\}, \\ G_{ij}^{(q)}(r_\rho) &= \left\{ \|\mathbf{u}_i - q\mathbf{u}_j\|_2 \leq \frac{1}{h_0(\check{\mathbf{B}})} r_\rho \right\}, \\ H_{ij}^{(q)}(r_\rho) &= \{\|\mathbf{u}_i - q\mathbf{u}_j\|_2 \leq h_0(\check{\mathbf{B}}) r_\rho\}. \end{aligned} \quad (96)$$

Define  $F_{ij}(r_\rho) = F_{ij}^{(-1)}(r_\rho) \cup F_{ij}^{(+1)}(r_\rho)$ .  $G_{ij}(r_\rho)$ ,  $H_{ij}(r_\rho)$ ,  $S_{ij}(r_\rho)$  are defined similarly. Using these notation, then  $\Phi_{ij}^{(\mathbf{P})}(\rho) = 1(S_{ij}(r_\rho))$ , and  $\Phi_{ij}^{(\mathbf{R})}(\rho) = 1(F_{ij}(r_\rho))$ . For  $\vec{i} \in C_\delta^<$ , denote

$$H_{\vec{i}}(r_\rho) = \bigcap_{\ell=1}^{\delta} H_{i_0 i_\ell}(r_\rho), \quad H_{\vec{i}, -m}(r_\rho) = \bigcap_{\substack{\ell=1 \\ \ell \neq m}}^{\delta} H_{i_0 i_\ell}(r_\rho).$$

When it is clear from the context, the dependence of  $r_\rho$  for the above quantities will be suppressed. By Lemma A.11, with probability 1,

$$G_{ij}^{(q)} \subset S_{ij}^{(q)} \subset H_{ij}^{(q)}, \quad G_{ij}^{(q)} \subset F_{ij}^{(q)} \subset H_{ij}^{(q)}. \quad (97)$$

*Lemma A.12:* Suppose  $p \geq n$ . Consider  $\delta \in [p-1]$ . For any  $\vec{i} \in C_\delta^<$ , with probability 1,

$$\left| \Phi_{\vec{i}}^{(\mathbf{P})} - \Phi_{\vec{i}}^{(\mathbf{R})} \right| \leq \xi_{\vec{i}},$$

where

$$\xi_{\vec{i}} := 1 \left( \bigcup_{m=1}^{\delta} \left( (H_{i_0 i_m} \setminus G_{i_0 i_m}) \cap H_{\vec{i}, -m} \right) \right).$$

*Proof:* Notice  $\Phi_{\vec{i}}^{(\mathbf{R})} = 1 \left( \bigcap_{m=1}^{\delta} F_{i_0 i_m} \right)$  and  $\Phi_{\vec{i}}^{(\mathbf{P})} = 1 \left( \bigcap_{m=1}^{\delta} S_{i_0 i_m} \right)$ . Let  $\Delta$  denote the symmetrization difference of two sets. Then

$$\left| \Phi_{\vec{i}}^{(\mathbf{P})} - \Phi_{\vec{i}}^{(\mathbf{R})} \right| = 1 \left( \left( \bigcap_{m=1}^{\delta} F_{i_0 i_m} \right) \Delta \left( \bigcap_{m=1}^{\delta} S_{i_0 i_m} \right) \right) \leq \xi_{\vec{i}},$$

where the inequality follows from (97) and Lemma A.29 (a).  $\square$

To obtain an upper bound on the expectation of  $\xi_{\vec{i}}$ , we first bound the expectation on a high-probability set. Define the set  $\mathcal{E}(t)$ , with  $t$  being a parameter to be determined, by

$$\begin{aligned} \mathcal{E}(t) &= \left\{ \left[ 1 - C_1 \left( \sqrt{\frac{n-1}{p-\tau}} + \frac{t}{\sqrt{p-\tau}} \right) \right]^2 \leq \lambda_{\min}(\check{\mathbf{B}}) \right\} \cap \\ &\quad \cap \left\{ \lambda_{\max}(\check{\mathbf{B}}) \leq \left[ 1 + C_1 \left( \sqrt{\frac{n-1}{p-\tau}} + \frac{t}{\sqrt{p-\tau}} \right) \right]^2 \right\}, \end{aligned} \quad (98)$$

to be the set such that (184) in Lemma A.30 holds, i.e. the constant  $C_1$  in  $\mathcal{E}(t)$  is the same constant as  $C$  in (184). By Lemma A.10 (a) and Lemma A.30,

$$\mathbb{P}(\mathcal{E}^c(t)) \leq 2 \exp(-c_1 t^2). \quad (99)$$

Since  $\tau \leq \frac{p}{2}$ ,

$$\frac{n-1}{p-\tau} \tau \leq 2(n-1) \frac{\tau}{p}, \quad (100)$$

and

$$C_1 \left( \sqrt{\frac{n-1}{p-\tau}} + \frac{t}{\sqrt{p-\tau}} \right) \leq \sqrt{2} C_1 \left( \sqrt{\frac{n-1}{p}} + \frac{t}{\sqrt{p}} \right). \quad (101)$$

Moreover, on  $\mathcal{E}(t)$ , and assuming

$$\sqrt{2} C_1 \left( \sqrt{\frac{n-1}{p}} + \frac{t}{\sqrt{p}} \right) \leq \frac{1}{2}, \quad (102)$$

one has

$$\begin{aligned} h_0(\mathbf{B}) &\leq \frac{\left( 1 + C_1 \left( \sqrt{\frac{n-1}{p-\tau}} + \frac{t}{\sqrt{p-\tau}} \right) \right)^2 + \frac{n-1}{p-\tau} \tau}{\left( 1 - C_1 \left( \sqrt{\frac{n-1}{p-\tau}} + \frac{t}{\sqrt{p-\tau}} \right) \right)^2} \\ &\leq 1 + 16\sqrt{2} C_1 \left( \sqrt{\frac{n-1}{p}} + \frac{t}{\sqrt{p}} \right) + 8(n-1) \frac{\tau}{p} := \theta_1(t), \end{aligned} \quad (103)$$

where the second inequality follows from (100), (101) and Lemma A.26 (c).

For  $\vec{i} \in C_\delta^<$ , denote

$$F_{\vec{i}}(r_\rho) = \bigcap_{\ell=1}^{\delta} F_{i_0 i_\ell}(r_\rho), \quad F_{\vec{i}, -m}(r_\rho) = \bigcap_{\substack{\ell=1 \\ \ell \neq m}}^{\delta} F_{i_0 i_\ell}(r_\rho).$$

**Lemma A.13:** Let  $p \geq n \geq 4$ ,  $\delta \in [p-1]$  and  $\mathbf{X} \sim \mathcal{VE}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, g)$ . Suppose  $\boldsymbol{\Sigma}$ , after some row-column permutation, is  $(\tau, \kappa)$  sparse with  $\tau \leq \frac{p}{2}$ . Let  $t$  be any positive number, and suppose (102) holds. Then for any  $\vec{i} \in C_\delta^<$ , with probability 1,

$$\xi_{\vec{i}}^{-1}(\mathcal{E}(t)) \leq \eta_{\vec{i}}(t),$$

where

$$\eta_{\vec{i}}(t) := \mathbb{1} \left( \bigcup_{m=1}^{\delta} \left( F_{\vec{i},-m}(\theta_1(t)r_\rho) \cap \left( F_{i_0 i_m}(\theta_1(t)r_\rho) \setminus F_{i_0 i_m} \left( \frac{r_\rho}{\theta_1(t)} \right) \right) \right) \right). \quad (104)$$

Moreover,

$$\mathbb{E} \mathbb{1} \left( \left( F_{i_0 i_m}(\theta_1(t)r_\rho) \setminus F_{i_0 i_m} \left( \frac{r_\rho}{\theta_1(t)} \right) \right) \cap F_{\vec{i},-m}(\theta_1(t)r_\rho) \right) \leq \mu_{n,\delta+1}(\boldsymbol{\Sigma}_{\vec{i}}) 2\theta_2(t) (2P_n(\theta_1(t)r_\rho))^{\delta-1}, \quad (105)$$

where  $\theta_2(t) := P_n(r_\rho \theta_1(t)) - P_n \left( \frac{r_\rho}{\theta_1(t)} \right)$ . Then

$$\begin{aligned} \mathbb{E} \eta_{\vec{i}}(t) &\leq \delta \mu_{n,\delta+1}(\boldsymbol{\Sigma}_{\vec{i}}) 2\theta_2(t) (2P_n(\theta_1(t)r_\rho))^{\delta-1} \\ &\leq \mu_{n,\delta+1}(\boldsymbol{\Sigma}_{\vec{i}}) \delta n(\theta_1(t))^{n\delta} \theta_3(t) (2P_n(r_\rho))^\delta, \end{aligned}$$

where  $\theta_3(t) := \left( \theta_1(t) - \frac{1}{\theta_1(t)} \right)$ .

*Proof:* By (103),  $H_{ij}(r_\rho) \cap \mathcal{E}(t) \subset F_{ij}(\theta_1(t)r_\rho)$  and  $G_{ij}(r_\rho) \cap \mathcal{E}(t) \supset F_{ij} \left( \frac{r_\rho}{\theta_1(t)} \right)$ . Then

$$\xi_{\vec{i}}^{-1}(\mathcal{E}(t)) \leq \eta_{\vec{i}}(t). \quad (106)$$

$$\begin{aligned} \mathbb{E} \mathbb{1} \left( \left( F_{i_0 i_m}(\theta_1(t)r_\rho) \setminus F_{i_0 i_m} \left( \frac{r_\rho}{\theta_1(t)} \right) \right) \cap F_{\vec{i},-m}(\theta_1(t)r_\rho) \right) &\leq \mu_{n,\delta+1}(\boldsymbol{\Sigma}_{\vec{i}}) \mathbb{P} \left( \left( \bigcup_{q \in \{-1, +1\}} \left\{ \frac{r_\rho}{\theta_1(t)} < \|\mathbf{u}'_{i_0} - q\mathbf{u}'_{i_m}\|_2 \leq \theta_1(t)r_\rho \right\} \right) \cap \left( \bigcap_{\substack{\alpha=1 \\ \alpha \neq m}}^{\delta} \left( \bigcup_{q \in \{-1, +1\}} \{\|\mathbf{u}'_{i_0} - q\mathbf{u}'_{i_\alpha}\|_2 \leq \theta_1(t)r_\rho\} \right) \right) \right) \right), \end{aligned} \quad (107)$$

where the last inequality follows from Lemma III.4 (c) with

$$\mathbf{u}'_{i_0}, \mathbf{u}'_{i_1}, \dots, \mathbf{u}'_{i_\delta} \stackrel{\text{i.i.d.}}{\sim} \text{unif}(S^{n-2}).$$

For any  $\mathbf{w} \in S^{n-2}$ , define  $\Omega_{\mathbf{w}}^{(q)} := \{\mathbf{v} \in S^{n-2} : \frac{1}{\theta_1(t)}r_\rho < \|\mathbf{v} - q\mathbf{w}\|_2 \leq r_\rho \theta_1(t)\}$ . Then

$$\begin{aligned} \mathbb{P} \left( \mathbf{u}'_{i_m} \in \bigcup_{q \in \{-1, +1\}} \Omega_{\mathbf{w}}^{(q)} \right) &= 2 \left( P_n(r_\rho \theta_1(t)) - P_n \left( \frac{1}{\theta_1(t)}r_\rho \right) \right). \end{aligned}$$

By conditioning on  $\mathbf{u}'_{i_0}$ , the term in right hand side of (107) equals to

$$\mu_{n,\delta+1}(\boldsymbol{\Sigma}_{\vec{i}}) 2\theta_2(t) (2P_n(\theta_1(t)r_\rho))^{\delta-1},$$

which then proves (105).

By the union bound,

$$\begin{aligned} \mathbb{E} \eta_{\vec{i}}(t) &\leq \sum_{m=1}^{\delta} \mathbb{E} \mathbb{1} \left( \left( F_{i_0 i_m}(\theta_1(t)r_\rho) \setminus F_{i_0 i_m} \left( \frac{r_\rho}{\theta_1(t)} \right) \right) \cap F_{\vec{i},-m}(\theta_1(t)r_\rho) \right) \\ &\stackrel{(*)}{\leq} \delta \mu_{n,\delta+1}(\boldsymbol{\Sigma}_{\vec{i}}) 2\theta_2(t) (2P_n(\theta_1(t)r_\rho))^{\delta-1} \\ &\stackrel{(**)}{\leq} \delta \mu_{n,\delta+1}(\boldsymbol{\Sigma}_{\vec{i}}) 2(n-2)P_n(r_\rho) (\theta_1(t))^{n-3} \theta_3(t) (2P_n(r_\rho \theta_1(t)))^{\delta-1} \\ &\stackrel{(***)}{\leq} \delta \mu_{n,\delta+1}(\boldsymbol{\Sigma}_{\vec{i}}) 2nP_n(r_\rho) (\theta_1(t))^{n-3} \theta_3(t) ((\theta_1(t))^{n-2} 2P_n(r_\rho))^{\delta-1}, \end{aligned}$$

where  $(*)$  follows from (105),  $(**)$  follows from Lemma A.22 (c), and  $(***)$  follows from Lemma A.22 (d).  $\square$

**Lemma A.14:** Let  $p \geq n \geq 4$ ,  $\delta \in [p-1]$  and  $\mathbf{X} \sim \mathcal{VE}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, g)$ . Let  $t$  be any positive number, and suppose (102) holds. Suppose  $\boldsymbol{\Sigma}$ , after some row-column permutation, is  $(\tau, \kappa)$  sparse with  $\tau \leq \frac{p}{2}$ . Then

$$\left| N_{E_\delta}^{(P)} - N_{E_\delta}^{(R)} \right| \mathbb{1}(\mathcal{E}(t)) \leq \sum_{\vec{i} \in C_\delta^<} \eta_{\vec{i}}(t)$$

and

$$\begin{aligned} \mathbb{E} \sum_{\vec{i} \in C_\delta^<} \eta_{\vec{i}}(t) &\leq \frac{Cn^2}{(\delta-1)!} \eta_3(\theta_1(t))^{n\delta} \left( \sqrt{\frac{1}{p}} + \frac{t}{\sqrt{p}} + \frac{\tau}{p} \right) p (2pP_n(r_\rho))^\delta, \end{aligned}$$

where  $C$  is a universal constant, and  $\eta_3 := 1 + \delta^2 \frac{\kappa-1}{p} \mu_{n,\delta+1}(\boldsymbol{\Sigma})$ .

*Proof:*

$$\begin{aligned} \left| N_{E_\delta}^{(P)} - N_{E_\delta}^{(R)} \right| \mathbb{1}(\mathcal{E}(t)) &\leq \sum_{\vec{i} \in C_\delta^<} \mathbb{E} \left| \Phi_{\vec{i}}^{(P)} - \Phi_{\vec{i}}^{(R)} \right| \mathbb{1}(\mathcal{E}(t)) \\ &\leq \sum_{\vec{i} \in C_\delta^<} \eta_{\vec{i}}(t), \end{aligned}$$

where the last inequality follows from Lemma A.12 and Lemma A.13.

By Lemma A.13,

$$\begin{aligned} \sum_{\vec{i} \in C_\delta^<} \mathbb{E} \eta_{\vec{i}} &\leq \sum_{\vec{i} \in C_\delta^<} \mu_{n,\delta+1}(\boldsymbol{\Sigma}_{\vec{i}}) \delta n(\theta_1(t))^{n\delta} \theta_3(t) (2P_n(r_\rho))^\delta \\ &\leq \frac{p^{\delta+1}}{\delta!} \left( 1 + \delta^2 \frac{\kappa-1}{p} \mu_{n,\delta+1}(\boldsymbol{\Sigma}) \right) \delta n(\theta_1(t))^{n\delta} \theta_3(t) (2P_n(r_\rho))^\delta \\ &\leq \frac{Cn}{(\delta-1)!} \eta_3(\theta_1(t))^{n\delta} \left( \sqrt{\frac{n}{p}} + \frac{t}{\sqrt{p}} + n \frac{\tau}{p} \right) \left( 2p^{1+\frac{1}{\delta}} P_n(r_\rho) \right)^\delta, \end{aligned} \quad (108)$$

where the third inequality follows from Lemma A.4 and the last inequality follows from Lemma A.26 (d) and (103).  $\square$

**Lemma A.15:** Let  $p \geq n \geq 4$ ,  $\delta \in [p-1]$  and  $\mathbf{X} \sim \mathcal{VE}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, g)$ . Suppose  $\boldsymbol{\Sigma}$ , after some row-column permutation, is  $(\tau, \kappa)$  sparse with  $\tau \leq \frac{p}{2}$ . Suppose  $2p^{1+\frac{1}{\delta}} P_n(r_\rho) \leq \gamma$

and  $\left(\sqrt{\frac{n-1}{p}} + \sqrt{\frac{\delta \ln p}{p}}\right) \leq c$  hold for some positive and sufficiently small universal constant  $c$ . Then

$$\begin{aligned} & \mathbb{E} \left| N_{E_\delta}^{(\mathbf{P})} - N_{E_\delta}^{(\mathbf{R})} \right| \\ & \leq C_{E_\delta}^{(\mathbf{P})} \left( 1 + \frac{\kappa-1}{p} \mu_{n,\delta+1}(\Sigma) \right) \left( \sqrt{\frac{\ln p}{p}} + \frac{\tau}{p} \right), \end{aligned}$$

where  $C_{E_\delta}^{(\mathbf{P})}$  is defined in (112).

*Proof:*

$$\begin{aligned} & \mathbb{E} \left| N_{E_\delta}^{(\mathbf{P})} - N_{E_\delta}^{(\mathbf{R})} \right| \\ & \leq \mathbb{E} \left| N_{E_\delta}^{(\mathbf{P})} - N_{E_\delta}^{(\mathbf{R})} \right| 1(\mathcal{E}(t)) + \binom{p}{1} \binom{p-1}{\delta} \mathbb{P}(\mathcal{E}^c(t)), \\ & \leq \mathbb{E} \left| N_{E_\delta}^{(\mathbf{P})} - N_{E_\delta}^{(\mathbf{R})} \right| 1(\mathcal{E}(t)) + \frac{p^{\delta+1}}{\delta!} 2 \exp(-c_1 t^2), \quad (109) \end{aligned}$$

where the first inequality follows from  $0 \leq N_{E_\delta}^{(\Psi)} \leq \binom{p}{1} \binom{p-1}{\delta}$  for both  $\Psi = \mathbf{R}$  and  $\Psi = \mathbf{P}$ , and the second inequality follows from (99).

Choose  $t = c_\delta \sqrt{\ln p}$  with  $c_\delta = \sqrt{\frac{5\delta}{2c_1}} \geq \sqrt{\left(\frac{3}{2} + \delta\right) / c_1}$  such that

$$2 \exp(-c_1 t^2) \leq 2 \exp\left(-\left(\frac{3}{2} + \delta\right) \ln p\right) = \frac{2}{p^{\frac{3}{2} + \delta}}.$$

Moreover, for any

$$c < \frac{1}{2 \max\left\{\sqrt{\frac{5}{2c_1}}, 1\right\} \sqrt{2} C_1}, \quad (110)$$

the inequality

$$\left( \sqrt{\frac{n-1}{p}} + \sqrt{\frac{\delta \ln p}{p}} \right) \leq c$$

implies

$$\sqrt{2} C_1 \left( \sqrt{\frac{n-1}{p}} + c_\delta \sqrt{\frac{\ln p}{p}} \right) \leq \frac{1}{2}, \quad (111)$$

which is (102) with  $t = c_\delta \sqrt{\ln p}$ . Then apply Lemma A.14 with  $t = c_\delta \sqrt{\ln p}$  to (109),

$$\begin{aligned} & \mathbb{E} \left| N_{E_\delta}^{(\mathbf{P})} - N_{E_\delta}^{(\mathbf{R})} \right| \\ & \leq \frac{C n^2}{(\delta-1)!} \eta_3 \left( \theta_1 \left( c_\delta \sqrt{\ln p} \right) \right)^{n\delta} \left( \sqrt{\frac{1}{p}} + \frac{\sqrt{\delta \ln p}}{\sqrt{p}} + \frac{\tau}{p} \right) \gamma^\delta \\ & \quad + \frac{2}{\delta! \sqrt{p}} \\ & \leq \frac{C n^2 \sqrt{\delta}}{(\delta-1)!} \eta_3 (\theta_1 (c_\delta \sqrt{\ln p}))^{n\delta} \left( \frac{\sqrt{\ln p}}{\sqrt{p}} + \frac{\tau}{p} \right) \gamma^\delta + \frac{2}{\delta! \sqrt{p}} \\ & \leq C_{E_\delta}^{(\mathbf{P})} \left( 1 + \frac{\kappa-1}{p} \mu_{n,\delta+1}(\Sigma) \right) \left( \sqrt{\frac{\ln p}{p}} + \frac{\tau}{p} \right), \end{aligned}$$

where

$$\begin{aligned} C_{E_\delta}^{(\mathbf{P})} &= \frac{C n^2 \delta^{\frac{5}{2}}}{(\delta-1)!} \left( \theta_1 \left( c_\delta \sqrt{\ln p} \right) \right)^{n\delta} \gamma^\delta + \frac{2}{\delta! \sqrt{\ln p}} \\ &\leq \frac{C n^2 \delta^{\frac{5}{2}}}{(\delta-1)!} (4n+5)^{n\delta} \gamma^\delta + \frac{2}{\delta!}, \quad (112) \end{aligned}$$

where the last step follows from  $\theta_1(c_\delta \sqrt{\ln p}) \leq 9+4(n-1) = 4n+5$  by (111) and  $\tau \leq p/2$ .  $\square$

*Proof of Proposition III.9 (b):* It follows directly from Lemma A.15 and Lemma A.24.  $\square$

4) *Proof of Proposition III.9 (c):* By Lemma III.8,

$$\begin{aligned} & N_{E_\delta}^{(\mathbf{P})} - (\delta+1) N_{E_{\delta+1}}^{(\mathbf{P})} - N_{E_\delta}^{(\mathbf{R})} \\ & \leq N_{\check{V}_\delta}^{(\mathbf{P})} - N_{\check{V}_\delta}^{(\mathbf{R})} \\ & \leq N_{E_\delta}^{(\mathbf{P})} - N_{E_\delta}^{(\mathbf{R})} + (\delta+1) N_{E_{\delta+1}}^{(\mathbf{R})}, \end{aligned}$$

which implies

$$\begin{aligned} & \left| N_{\check{V}_\delta}^{(\mathbf{P})} - N_{\check{V}_\delta}^{(\mathbf{R})} \right| \\ & \leq \left| N_{E_\delta}^{(\mathbf{P})} - N_{E_\delta}^{(\mathbf{R})} \right| + (\delta+1) \left| N_{E_{\delta+1}}^{(\mathbf{P})} - N_{E_{\delta+1}}^{(\mathbf{R})} \right| + (\delta+1) N_{E_{\delta+1}}^{(\mathbf{R})}. \quad (113) \end{aligned}$$

*Lemma A.16:* Let  $p \geq n \geq 4$ ,  $\delta \in [p-1]$  and  $\mathbf{X} \sim \mathcal{VE}(\mu, \Sigma, g)$ . Suppose  $\Sigma$ , after some row-column permutation, is  $(\tau, \kappa)$  sparse with  $\tau \leq \frac{p}{2}$ . Suppose  $2p^{1+\frac{1}{\delta}} P_n(r_p) \leq \gamma$  and  $\left(\sqrt{\frac{n-1}{p}} + \sqrt{\frac{\delta \ln p}{p}}\right) \leq c$  hold for some positive and sufficiently small constant  $c$ . Then

$$\begin{aligned} & \mathbb{E} \left| N_{\check{V}_\delta}^{(\mathbf{P})} - N_{\check{V}_\delta}^{(\mathbf{R})} \right| \\ & \leq C_{\check{V}_\delta}^{(\mathbf{P})} \left( 1 + \frac{\kappa-1}{p} \mu_{n,\delta+2}(\Sigma) \right) \left( \sqrt{\frac{\ln p}{p}} + \frac{\tau}{p} + p^{-\frac{1}{\delta}} \right) \end{aligned}$$

where  $C_{\check{V}_\delta}^{(\mathbf{P})}$  is defined in (118).

*Proof:* Let  $\mathcal{E}(t)$  be the same as in (98) with  $t$  to be determined. Consider  $\delta \in [p-2]$ .

$$\begin{aligned} & \mathbb{E} \left| N_{\check{V}_\delta}^{(\mathbf{P})} - N_{\check{V}_\delta}^{(\mathbf{R})} \right| \\ & \leq \mathbb{E} \left| N_{\check{V}_\delta}^{(\mathbf{P})} - N_{\check{V}_\delta}^{(\mathbf{R})} \right| 1(\mathcal{E}(t)) + p \mathbb{P}(\mathcal{E}^c(t)) \\ & \leq \mathbb{E} \left| N_{E_\delta}^{(\mathbf{P})} - N_{E_\delta}^{(\mathbf{R})} \right| 1(\mathcal{E}(t)) + (\delta+1) \mathbb{E} \left| N_{E_{\delta+1}}^{(\mathbf{P})} - N_{E_{\delta+1}}^{(\mathbf{R})} \right| 1(\mathcal{E}(t)) \\ & \quad + (\delta+1) \mathbb{E} N_{E_{\delta+1}}^{(\mathbf{R})} + 2p e^{-c_1 t^2}, \quad (114) \end{aligned}$$

where the first inequality follows from  $0 \leq N_{\check{V}_\delta}^{(\Psi)} \leq p$  for  $\Psi = \mathbf{R}$  and  $\mathbf{P}$ , the second inequality follows from (113) and (99). If  $\delta = p-1$ , then

$$\begin{aligned} & \mathbb{E} \left| N_{\check{V}_{p-1}}^{(\mathbf{P})} - N_{\check{V}_{p-1}}^{(\mathbf{R})} \right| \\ & \leq \mathbb{E} \left| N_{\check{V}_{p-1}}^{(\mathbf{P})} - N_{\check{V}_{p-1}}^{(\mathbf{R})} \right| 1(\mathcal{E}(t)) + p \mathbb{P}(\mathcal{E}^c(t)) \\ & \leq \mathbb{E} \left| N_{E_{p-1}}^{(\mathbf{P})} - N_{E_{p-1}}^{(\mathbf{R})} \right| 1(\mathcal{E}(t)) + 2p \exp(-c_1 t^2), \end{aligned}$$

which shows that (114) also holds for  $\delta = p-1$  with the convention  $N_{E_p}^{(\mathbf{P})} = N_{E_p}^{(\mathbf{R})} = 0$ .

Choose  $t = \sqrt{\frac{3}{c_1} \ln p} := c_2 \sqrt{\ln p}$ , such that  $p \exp(-c_1 t^2) = \frac{1}{p^2}$ . Moreover, for any

$$c < \frac{1}{2 \max\{c_2, 1\} \sqrt{2} C_1}, \quad (115)$$

the inequality

$$\left( \sqrt{\frac{n-1}{p}} + \sqrt{\frac{\ln p}{p}} \right) \leq c$$

implies

$$\sqrt{2}C_1 \left( \sqrt{\frac{n-1}{p}} + c_2 \sqrt{\frac{\ln p}{p}} \right) \leq \frac{1}{2}, \quad (116)$$

which is (102) with  $t = c_2 \sqrt{\ln p}$ . With  $t = c_2 \sqrt{\ln p}$  Lemma A.14 becomes:

$$\begin{aligned} & \mathbb{E} \left| N_{V_\delta}^{(\mathbf{P})} - N_{V_\delta}^{(\mathbf{R})} \right| 1(\mathcal{E}(t)) \\ & \leq \frac{Cn^2}{(\delta-1)!} \eta_3(\theta_1(c_2 \sqrt{\ln p}))^{n\delta} \left( \frac{\sqrt{\ln p}}{\sqrt{p}} + \frac{\tau}{p} \right) p(2pP_n(r_\rho))^\delta, \end{aligned} \quad (117)$$

Then for  $\delta \in [p-1]$  applying (117) with  $\delta, \delta+1$  and Lemma A.9 to (114), together with  $\mu_{n,\delta+1}(\Sigma) \leq \mu_{n,\delta+2}(\Sigma)$ , shown in the equation at the bottom of the next page.

where in the last inequality

$$\begin{aligned} & C_{\tilde{V}_\delta}^{(\mathbf{P})} \\ & = \frac{Cn^2(\delta+1)^2}{(\delta-1)!} \left( \theta_1(c_2 \sqrt{\ln p}) \right)^{n(\delta+1)} \gamma^\delta (1+\gamma) \left( 1 + \frac{\delta+1}{\delta} \gamma p^{-\frac{1}{\delta}} \right) + \frac{2}{p^{2-\frac{1}{\delta}}} \\ & \leq \frac{Cn^2(\delta+1)^2}{(\delta-1)!} (4n+5)^{n(\delta+1)} \gamma^\delta (1+\gamma) \left( 1 + \frac{\delta+1}{\delta} \gamma p^{-\frac{1}{\delta}} \right) + 2. \end{aligned} \quad (118)$$

where the last step follows from  $\theta_1(c_2 \sqrt{\ln p}) \leq 9+4(n-1) = 4n+5$  by (116) and  $\tau \leq p/2$ .  $\square$

*Proof of Proposition III.9 (c):* Lemmas A.16 and A.24 complete the proof of Proposition III.9 (c).  $\square$

### G. Proofs in Subsection III-E

1) *Proof of Lemma III.14:* To utilize the notation we have defined in this paper, we make the following adjustments to the notation throughout this subsection. In this proof it suffices to prove the conclusion for any  $\delta+1$  i.i.d. random points from  $\text{unif}(S^{n-2})$ . Without loss of generality assume in this subsection that the first  $\delta+1$  U-scores  $\{\mathbf{u}_i\}_{i=1}^{\delta+1}$  are independent. Another adjustment is to replace  $r$  by  $r_\rho$ . With these adjustments Lemma III.14 is equivalent to the following: when  $r_\rho < 2/\sqrt{5}$ ,  $\delta \geq 1$ , for any  $\ell \in [\delta+1]$ ,

$$\begin{aligned} & \mathbb{P}(\text{NUV}(\{\mathbf{u}_i\}_{i=1}^{\delta+1}, r_\rho) = \ell | \deg(\mathbf{u}_{\delta+1}) = \delta) \\ & = \mathbb{P}(\text{PNUV}(\{\mathbf{u}_i\}_{i=1}^{\delta+1}, r_\rho) = \ell | \deg(\mathbf{u}_{\delta+1}) = \delta). \end{aligned} \quad (119)$$

Take  $\vec{i} = (\delta+1, 1, \dots, \delta)$ . Recall the notation  $\vec{i}, \Phi_{\vec{i}} = \Phi_{\vec{i}}^{(\mathbf{R})}, U_{\vec{i}}$  are defined in Subsection III-C, where the dependence of  $\mathbf{R}$  in  $\Phi_{\vec{i}}^{(\mathbf{R})}$  is suppressed throughout this subsection for the sake of clearer exposition. Then  $\text{PNUV}(\{\mathbf{u}_i\}_{i=1}^{\delta+1}, r_\rho) = U_{\vec{i}} + \Phi_{\vec{i}}$ . Moreover, the event  $\{\deg(\mathbf{u}_{\delta+1}) = \delta\}$  in  $\text{PGe}(\{\mathbf{u}_i\}_{i=1}^{\delta+1}, r_\rho)$  is the same as  $\{\Phi_{\vec{i}} = 1\}$ . Define  $F_{ij}^{(q)} = \{\|\mathbf{u}_i - q\mathbf{u}_j\|_2 \leq r_\rho\}$ . Then  $1\left(\bigcap_{j=1}^{\delta} F_{j(\delta+1)}^{(+1)}\right)$  is the indicator function of the event that the degree of vertex  $\mathbf{u}_{\delta+1}$  in  $\text{Ge}(\{\mathbf{u}_i\}_{i=1}^{\delta+1}, r_\rho)$  is  $\delta$ . Hence

Lemma III.14 is also equivalent to the following: when  $r_\rho < 2/\sqrt{5}$ ,  $\delta \geq 1$ , for any  $\ell \in [\delta+1]$ ,

$$\begin{aligned} & \mathbb{P}(U_{\vec{i}} + \Phi_{\vec{i}} = \ell | \Phi_{\vec{i}} = 1) \\ & = \mathbb{P}\left(\text{NUV}(\{\mathbf{u}_i\}_{i=1}^{\delta+1}, r_\rho) = \ell \mid 1\left(\bigcap_{j=1}^{\delta} F_{j(\delta+1)}^{(+1)}\right) = 1\right). \end{aligned} \quad (120)$$

*Proof of (120):* For  $\vec{q} = (q_1, q_2, \dots, q_\delta) \in \{-1, +1\}^\delta$ , denote  $F_{\delta+1}^{(\vec{q})} = \bigcap_{j=1}^{\delta} F_{j(\delta+1)}^{(q_j)}$ . Observe that

$$\{\Phi_{\vec{i}} = 1\} = \bigcap_{j=1}^{\delta} \bigcup_{q_j \in \{+1, -1\}} F_{j(\delta+1)}^{(q_j)} = \bigcup_{\vec{q} \in \{-1, +1\}^\delta} F_{\delta+1}^{(\vec{q})}.$$

Since  $r_\rho < 2/\sqrt{5} < \sqrt{2}$ ,  $F_{j(\delta+1)}^{(-1)}$  and  $F_{j(\delta+1)}^{(+1)}$  are disjoint for every  $j \in [\delta]$ , which implies  $F_{\delta+1}^{(\vec{q})}$  for different  $\vec{q} \in \{-1, +1\}^\delta$  are disjoint. Hence,

$$\mathbb{P}(\Phi_{\vec{i}} = 1, U_{\vec{i}} = \ell - 1) = \sum_{\vec{q} \in \{-1, +1\}^\delta} \mathbb{P}(F_{\delta+1}^{(\vec{q})}, U_{\vec{i}} = \ell - 1). \quad (121)$$

Next observe that  $1(F_{\delta+1}^{(\vec{q})})$  is a function of  $\mathbf{u}_1, \dots, \mathbf{u}_{\delta+1}$ , and hence it has the same distribution when replacing  $\mathbf{u}_i$  by  $-\mathbf{u}_i$  for any  $i \in [\delta]$ . Moreover, replacing  $\mathbf{u}_i$  by  $-\mathbf{u}_i$  for any  $i \in [\delta]$  wouldn't change  $U_{\vec{i}}$ . As a result, (121) implies

$$\mathbb{P}(\Phi_{\vec{i}} = 1, U_{\vec{i}} = \ell - 1) = 2^\delta \mathbb{P}\left(F_{\delta+1}^{(\vec{q}_0)}, U_{\vec{i}} = \ell - 1\right), \quad (122)$$

where  $\vec{q}_0 = (+1, +1, \dots, +1)$  is the vector in  $\mathbb{R}^\delta$  with all its components +1.

Consider  $\omega \in F_{\delta+1}^{(\vec{q}_0)}$ . Then  $\Phi_{\vec{i}}(\omega) = 1$  or equivalently,  $\Phi_{i(\delta+1)}^{(\mathbf{R})}(\omega) = 1$  for any  $i \in [\delta]$ . Then

$$U_{\vec{i}}(\omega) = \sum_{\vec{j} \in S_{\vec{i}}} \Phi_{\vec{j}}(\omega) = \sum_{i=1}^{\delta} \prod_{\substack{j=1 \\ j \neq i}}^{\delta+1} \Phi_{ij}^{(\mathbf{R})}(\omega) = \sum_{i=1}^{\delta} \prod_{\substack{j=1 \\ j \neq i}}^{\delta} \Phi_{ij}^{(\mathbf{R})}(\omega). \quad (123)$$

Since for any distinct  $i, j \in [\delta]$ ,  $\|\mathbf{u}_i(\omega) - \mathbf{u}_j(\omega)\|_2 \leq \|\mathbf{u}_i - \mathbf{u}_{\delta+1}(\omega)\|_2 + \|\mathbf{u}_i(\omega) - \mathbf{u}_{\delta+1}(\omega)\|_2 \leq 2r_\rho < 4/\sqrt{5}$ ,  $\|\mathbf{u}_i(\omega) + \mathbf{u}_j(\omega)\|_2 = \sqrt{4 - \|\mathbf{u}_i(\omega) - \mathbf{u}_j(\omega)\|_2^2} > 2/\sqrt{5} > r_\rho$ . Thus  $\Phi_{ij}^{(\mathbf{R})}(\omega) = 1_{F_{ij}^{(+1)}}(\omega)$ . That is, in the set  $F_{\delta+1}^{(\vec{q}_0)}$ , (123) becomes

$$U_{\vec{i}} = \sum_{i=1}^{\delta} \prod_{\substack{j=1 \\ j \neq i}}^{\delta} 1(F_{ij}^{(+1)}) = \text{NUV}(\{\mathbf{u}_i\}_{i=1}^{\delta}, r_\rho), \quad (124)$$

which implies

$$\begin{aligned} & (\Phi_{\vec{i}} + U_{\vec{i}}) 1\left(F_{\delta+1}^{(\vec{q}_0)}\right) = (1 + \text{NUV}(\{\mathbf{u}_i\}_{i=1}^{\delta}, r_\rho)) 1\left(F_{\delta+1}^{(\vec{q}_0)}\right) \\ & = \text{NUV}(\{\mathbf{u}_i\}_{i=1}^{\delta+1}, r_\rho) 1\left(F_{\delta+1}^{(\vec{q}_0)}\right). \end{aligned} \quad (125)$$

Thus,

$$\begin{aligned}
& \mathbb{P}(U_i + \Phi_i = \ell | \Phi_i = 1) \\
&= \frac{2^\delta \mathbb{P}(F_{\delta+1}^{(\bar{q}_0)}, U_i = \ell - 1)}{\mathbb{P}(\Phi_i = 1)} \\
&= \frac{2^\delta \mathbb{P}(F_{\delta+1}^{(\bar{q}_0)}, U_i = \ell - 1)}{(2P_n(r_\rho))^\delta} \\
&= \frac{\mathbb{P}(F_{\delta+1}^{(\bar{q}_0)}, \mathbf{NUV}(\{\mathbf{u}_i\}_{i=1}^{\delta+1}, r_\rho) = \ell)}{(P_n(r_\rho))^\delta} \\
&= \mathbb{P}\left(\mathbf{NUV}(\{\mathbf{u}_i\}_{i=1}^{\delta+1}, r_\rho) = \ell \mid 1 \left( \bigcap_{j=1}^{\delta} F_{j(\delta+1)}^{(+1)} \right) = 1\right),
\end{aligned}$$

where the first equality follows from (122), the second equality follows from Lemma A.2, and the third equality follows from (125).  $\square$

## 2) Proofs of Lemma III.15 and Lemma III.16:

*Proof of Lemma III.15:* In the set  $\{\deg(\mathbf{u}'_{\delta+1}) = \delta\}$ , it follows that

$$\mathbf{NUV}(\{\mathbf{u}'_i\}_{i=1}^{\delta+1}, r) = \mathbf{NUV}(\{\mathbf{u}'_i\}_{i=1}^{\delta}, r) + 1.$$

Thus, (126), shown at the bottom of the next page, where the last equality follows from that  $T(\mathbf{u}'_{\delta+1})$  as a function of the random variable  $\mathbf{u}'_{\delta+1}$ , due to the rotation invariance property of the distribution  $\mathbf{unif}(S^{n-2})$ , equals to  $T(\mathbf{v}_0)$  a.s. with  $\mathbf{v}_0 = (1, 0, 0, \dots, 0) \in S^{n-2}$ .

Under the condition  $\mathbf{u}'_{\delta+1} = \mathbf{v}_0$ ,  $1(\{\deg(\mathbf{u}'_{\delta+1}) = \delta\}) = \prod_{i=1}^{\delta} 1(\mathbf{u}'_i \in \text{SC}(r, \mathbf{v}_0))$ , where  $\text{SC}(r, \mathbf{v}_0)$  is defined in (25). Consider the coordinate system for  $\mathbf{u}'_i = (u'_{ji} : 1 \leq j \leq n-1)^\top$  in the region  $\text{SC}(r, \mathbf{v}_0)$ :

for  $1 \leq i \leq \delta$ ,

$$\begin{cases} u'_{1i} = 1 - \frac{r^2 r_i^2}{2}, \\ u'_{2i} = r r_i \sqrt{1 - \frac{r^2 r_i^2}{4}} \cos(\theta_{2i}), \\ \vdots, \\ u'_{ji} = r r_i \sqrt{1 - \frac{r^2 r_i^2}{4}} \cos(\theta_{ji}) \prod_{m=2}^{j-1} \sin(\theta_{mi}), \\ \vdots \\ u'_{(n-2)i} = r r_i \sqrt{1 - \frac{r^2 r_i^2}{4}} \cos(\theta_{(n-2)i}) \prod_{m=2}^{n-3} \sin(\theta_{mi}), \\ u'_{(n-1)i} = r r_i \sqrt{1 - \frac{r^2 r_i^2}{4}} \prod_{m=2}^{n-2} \sin(\theta_{mi}), \end{cases}$$

where for each  $i \in [\delta]$ :

$r_i \in [0, 1], \theta_{ji} \in [0, \pi]$  for  $2 \leq j \leq n-3$  and  $\theta_{(n-2)i} \in [0, 2\pi)$ . (127)

Then,

$$\begin{aligned}
& r^{-(n-2)\delta} T(\mathbf{v}_0) \\
&= \frac{\mathbb{E} \prod_{i=1}^{\delta} 1(\mathbf{u}'_i \in \text{SC}(r, \mathbf{v}_0)) 1(\mathbf{NUV}(\{\mathbf{u}'_i\}_{i=1}^{\delta}, r) = \ell - 1)}{r^{(n-2)\delta}} \\
&\stackrel{(*)}{=} \frac{r^{-(n-2)\delta}}{|S^{n-2}|^\delta} \int \cdots \int_{\Omega_0} 1(\mathbf{NUV}(\{\mathbf{u}'_i\}_{i=1}^{\delta}, r) = \ell - 1) \prod_{i=1}^{\delta} \\
&\quad \left( r^{n-2} r_i^{n-3} \left(1 - \frac{r^2 r_i^2}{4}\right)^{\frac{n-4}{2}} dr_i \prod_{j=2}^{n-2} (\sin^{n-2-j}(\theta_{ji}) d\theta_{ji}) \right) \\
&= \frac{1}{|S^{n-2}|^\delta} \int \cdots \int_{\Omega_0} 1(\mathbf{NUV}(\{\mathbf{u}'_i\}_{i=1}^{\delta}, r) = \ell - 1) \\
&\quad \prod_{i=1}^{\delta} \left( r_i^{n-3} \left(1 - \frac{r^2 r_i^2}{4}\right)^{\frac{n-4}{2}} dr_i \prod_{j=2}^{n-2} (\sin^{n-2-j}(\theta_{ji}) d\theta_{ji}) \right), \tag{128}
\end{aligned}$$

where  $\Omega_0$  in equality (\*) is the region described in (127). Denote by  $f(r)$  the integrand in (128).  $f(r)$  is a function of

$$\begin{aligned}
& \mathbb{E} |N_{\check{\mathbf{V}}_\delta}^{(\mathbf{P})} - N_{\check{\mathbf{V}}_\delta}^{(\mathbf{R})}| \\
&\leq \frac{Cn^2}{(\delta-1)!} \left(1 + \delta^2 \frac{\kappa-1}{p} \mu_{n,\delta+2}(\Sigma)\right) \left(\theta_1(c_2 \sqrt{\ln p})\right)^{n\delta} \left(\frac{\sqrt{\ln p}}{\sqrt{p}} + \frac{\tau}{p}\right) p(2pP_n(r_\rho))^\delta + \\
&\quad (\delta+1) \frac{Cn^2}{\delta!} \left(1 + (\delta+1)^2 \frac{\kappa-1}{p} \mu_{n,\delta+2}(\Sigma)\right) \left(\theta_1(c_2 \sqrt{\ln p})\right)^{n(\delta+1)} \left(\frac{\sqrt{\ln p}}{\sqrt{p}} + \frac{\tau}{p}\right) p(2pP_n(r_\rho))^{\delta+1} \\
&\quad + (\delta+1) \frac{1}{(\delta+1)!} \left(1 + (\delta+1)^2 \mu_{n,\delta+2}(\Sigma) \frac{\kappa-1}{p}\right) p(2pP_n(r_\rho))^{\delta+1} + \frac{2}{p^2} \\
&\leq \frac{Cn^2}{(\delta-1)!} \left(1 + (\delta+1)^2 \frac{\kappa-1}{p} \mu_{n,\delta+2}(\Sigma)\right) \left(\theta_1(c_2 \sqrt{\ln p})\right)^{n(\delta+1)} \left(\frac{\sqrt{\ln p}}{\sqrt{p}} + \frac{\tau}{p}\right) \gamma^\delta \left(1 + \frac{\delta+1}{\delta} \gamma p^{-\frac{1}{\delta}}\right) \\
&\quad + \frac{1}{\delta!} \left(1 + (\delta+1)^2 \mu_{n,\delta+2}(\Sigma) \frac{\kappa-1}{p}\right) \gamma^{\delta+1} p^{-\frac{1}{\delta}} + \frac{2}{p^2} \\
&\leq C_{\check{\mathbf{V}}_\delta}^{(\mathbf{P})} \left(1 + \frac{\kappa-1}{p} \mu_{n,\delta+2}(\Sigma)\right) \left(\sqrt{\frac{\ln p}{p}} + \frac{\tau}{p} + p^{-\frac{1}{\delta}}\right),
\end{aligned}$$

$r_i, \theta_{ji}$  for  $2 \leq j \leq n-2$  and  $1 \leq i \leq \delta$ , of which the dependences are suppressed.

Note  $\mathbf{NUV}(\{\mathbf{v}_i\}_{i=1}^\delta, r)$  is a function of  $(1(\|\mathbf{v}_i - \mathbf{v}_j\|_2 \leq r) : 1 \leq i < j \leq \delta)$  and it does not depend on the specific location of each vertex. Thus in (128)  $\mathbf{NUV}(\{\mathbf{u}'_i\}_{i=1}^\delta, r)$  is a function of  $(1(\|\mathbf{u}'_i - \mathbf{u}'_j\|_2 < r) : 1 \leq i < j \leq \delta)$  since  $\|\mathbf{u}'_i - \mathbf{u}'_j\|_2 = r$ , as a set of Lebesgue measure 0, contributes nothing to the integral. We then write

$$\begin{aligned} & \mathbf{NUV}(\{\mathbf{u}'_i\}_{i=1}^\delta, r) \\ &= \chi(1(\|\mathbf{u}'_i - \mathbf{u}'_j\|_2 < r) : 1 \leq i < j \leq \delta) \\ &= \chi\left(1\left(\frac{1}{r}\|\mathbf{u}'_i - \mathbf{u}'_j\|_2 < 1\right) : 1 \leq i < j \leq \delta\right). \end{aligned}$$

Intrinsically,  $1(\|\mathbf{u}'_i - \mathbf{u}'_j\|_2 < r)$  is the indicator random variable about whether there is an edge between vertex  $i$  and  $j$ , and the function  $\chi$  is the function that takes all edge information among  $\delta$  vertices as input and outputs the number of universal vertices.

Then as  $r \rightarrow 0^+$ ,

$$\begin{aligned} & \lim_{r \rightarrow 0^+} f(r) \\ &= \prod_{i=1}^\delta \left( r_i^{n-3} \prod_{j=2}^{n-2} (\sin^{n-2-j}(\theta_{ji})) \right) \\ & \times \lim_{r \rightarrow 0^+} 1\left(\chi\left(1\left(\frac{1}{r}\|\mathbf{u}'_i - \mathbf{u}'_j\|_2 < 1\right) : 1 \leq i < j \leq \delta\right) = \ell - 1\right). \end{aligned} \quad (129)$$

Observe

$$\begin{aligned} & \lim_{r \rightarrow 0^+} \left( \frac{1}{r} \|\mathbf{u}'_i - \mathbf{u}'_j\|_2 \right)^2 \\ &= (r_i \cos(\theta_{2i}) - r_j \cos(\theta_{2j}))^2 + \\ & \sum_{q=3}^{n-2} \left( r_i \prod_{m=2}^{q-1} \sin(\theta_{mi}) \cos(\theta_{qi}) - r_j \prod_{m=2}^{q-1} \sin(\theta_{mj}) \cos(\theta_{qj}) \right)^2 \\ & + \left( r_i \prod_{m=2}^{n-2} \sin(\theta_{mi}) - r_j \prod_{m=2}^{n-2} \sin(\theta_{mj}) \right)^2. \end{aligned} \quad (130)$$

On  $\Omega_0$ , for  $1 \leq i \leq \delta$ , define

$$\begin{cases} \tilde{u}'_{1i} = r_i \cos(\theta_{2i}), \\ \tilde{u}'_{ji} = r_i \cos(\theta_{(j+1)i}) \prod_{m=2}^j \sin(\theta_{mi}), \text{ for } 2 \leq j \leq n-3 \\ \tilde{u}'_{(n-2)i} = r_i \prod_{m=2}^{n-2} \sin(\theta_{mi}), \end{cases}$$

$$(131) \quad \mathbb{P}(\deg(\mathbf{u}'_{\delta+1}) = \delta, \mathbf{NUV}(\{\mathbf{u}'_i\}_{i=1}^{\delta+1}, r) = \ell)$$

$$\begin{aligned} &= \mathbb{E}1\left(\{\deg(\mathbf{u}'_{\delta+1}) = \delta\} \cap \{\mathbf{NUV}(\{\mathbf{u}'_i\}_{i=1}^\delta, r) = \ell - 1\}\right) \\ &= \mathbb{E}\left(\mathbb{E}\left(1\left(\{\deg(\mathbf{u}'_{\delta+1}) = \delta\} \cap \{\mathbf{NUV}(\{\mathbf{u}'_i\}_{i=1}^\delta, r) = \ell - 1\}\right) \middle| \mathbf{u}'_{\delta+1}\right)\right) \\ &:= \mathbb{E}T(\mathbf{u}'_{\delta+1}) \\ &= \mathbb{E}T(\mathbf{v}_0), \end{aligned} \quad (126)$$

and  $\tilde{\mathbf{u}}'_i = (\tilde{u}'_{ji} : 1 \leq j \leq n-2) \in B^{n-2}$ . Then by (130)

$$\lim_{r \rightarrow 0^+} \frac{1}{r} \|\mathbf{u}'_i - \mathbf{u}'_j\|_2 = \|\tilde{\mathbf{u}}'_i - \tilde{\mathbf{u}}'_j\|_2,$$

which, together with (129), imply

$$\begin{aligned} & \lim_{r \rightarrow 0^+} f(r) \\ &= \prod_{i=1}^\delta \left( r_i^{n-3} \prod_{j=2}^{n-2} (\sin^{n-2-j}(\theta_{ji})) \right) \\ & 1(\chi(1(\|\tilde{\mathbf{u}}'_i - \tilde{\mathbf{u}}'_j\|_2 < 1) : 1 \leq i < j \leq \delta) = \ell - 1) \\ &= \prod_{i=1}^\delta \left( r_i^{n-3} \prod_{j=2}^{n-2} (\sin^{n-2-j}(\theta_{ji})) \right) \\ & 1(\chi(1(\|\tilde{\mathbf{u}}'_i - \tilde{\mathbf{u}}'_j\|_2 \leq 1) : 1 \leq i < j \leq \delta) = \ell - 1) \\ &= \prod_{i=1}^\delta \left( r_i^{n-3} \prod_{j=2}^{n-2} (\sin^{n-2-j}(\theta_{ji})) \right) \\ & 1(\mathbf{NUV}(\{\tilde{\mathbf{u}}'_i\}_{i=1}^\delta, 1) = \ell - 1), \end{aligned}$$

where the second equality holds *a.s.* with respect to the Lebesgue measure on  $\Omega_0$ ,

Moreover,  $|f(r)| \leq 1$ , which is integrable over the bounded set  $\Omega_0$ . Applying Dominated Convergence Theorem to (128),

$$\begin{aligned} & \lim_{r \rightarrow 0^+} r^{-(n-2)\delta} T(\mathbf{v}_0) \\ &= \frac{1}{|S^{n-2}|^\delta} \int \cdots \int_{\Omega_0} 1(\mathbf{NUV}(\{\tilde{\mathbf{u}}'_i\}_{i=1}^\delta, 1) = \ell - 1) \\ & \prod_{i=1}^\delta r_i^{n-3} dr_i \prod_{j=2}^{n-2} \sin^{n-2-j}(\theta_{ji}) d\theta_{ji} \\ &= \frac{|B^{n-2}|^\delta}{|S^{n-2}|^\delta} \mathbb{P}(\mathbf{NUV}(\{\tilde{\mathbf{u}}'_i\}_{i=1}^\delta, 1) = \ell - 1), \end{aligned} \quad (132)$$

where the parametrization (131) and the region  $\Omega_0$  coincide with the spherical coordinates for  $B^{n-2}$ .

Thus

$$\begin{aligned} & \lim_{r \rightarrow 0^+} \mathbb{P}(\mathbf{NUV}(\{\mathbf{u}'_i\}_{i=1}^{\delta+1}, r) = \ell | \deg(\mathbf{u}'_{\delta+1}) = \delta) \\ &= \lim_{r \rightarrow 0^+} \frac{\mathbb{P}(\deg(\mathbf{u}'_{\delta+1}) = \delta, \mathbf{NUV}(\{\mathbf{u}'_i\}_{i=1}^{\delta+1}, r) = \ell)}{r^{(n-2)\delta}} \frac{r^{(n-2)\delta}}{(P_n(r))^\delta} \\ &= \frac{|B^{n-2}|^\delta}{|S^{n-2}|^\delta} \mathbb{P}(\mathbf{NUV}(\{\tilde{\mathbf{u}}'_i\}_{i=1}^\delta, 1) = \ell - 1) \frac{1}{(a_n)^\delta} \\ &= \frac{1}{(a_n)^\delta} \frac{|B^{n-2}|^\delta}{|S^{n-2}|^\delta} \mathbb{P}(\mathbf{NUV}(\{\tilde{\mathbf{u}}'_i\}_{i=1}^\delta, 1) = \ell - 1) \\ &= \mathbb{P}(\mathbf{NUV}(\{\tilde{\mathbf{u}}'_i\}_{i=1}^\delta, 1) = \ell - 1), \end{aligned}$$

where the second equality follows from (126), (132) and Lemma A.22 (b).  $\square$

*Proof of Lemma III.16:* By (38), Lemma III.15 and (6),

$$\lim_{\rho \rightarrow 1^-} \alpha(\ell, r_\rho) = \lim_{r_\rho \rightarrow 0^+} \alpha(\ell, r_\rho) = \alpha_\ell, \quad \forall \ell \in [\delta + 1], \quad (133)$$

and thus,

$$\lim_{\rho \rightarrow 1^-} \zeta_{n,\delta,\rho}(\ell) = \zeta_{n,\delta}(\ell), \quad \forall \ell \in [\delta + 1]. \quad (134)$$

Note that

$$\begin{aligned} & \left| 2p^{1+\frac{1}{\delta}} P_n(r_\rho) - e_{n,\delta} \right| \\ & \leq \left| 2p^{1+\frac{1}{\delta}} P_n(r_\rho) - 2a_n p^{1+\frac{1}{\delta}} r_\rho^{n-2} \right| + \left| 2a_n p^{1+\frac{1}{\delta}} r_\rho^{n-2} - e_{n,\delta} \right| \\ & \leq 2a_n p^{1+\frac{1}{\delta}} r_\rho^{n-2} \left( \frac{n-4}{8} r_\rho^2 \right) + \left| 2a_n p^{1+\frac{1}{\delta}} r_\rho^{n-2} - e_{n,\delta} \right| \\ & = 2a_n p^{1+\frac{1}{\delta}} r_\rho^{n-2} \left( \frac{n-4}{4} (1-\rho) \right) + \left| 2a_n p^{1+\frac{1}{\delta}} r_\rho^{n-2} - e_{n,\delta} \right|. \end{aligned} \quad (135)$$

where the second inequality follows from Lemma A.22 (a). Then the preceding expression and (133) yield

$$\begin{aligned} \lim_{p \rightarrow \infty} \lambda_{p,n,\delta,\rho} &= \lim_{p \rightarrow \infty} \frac{1}{\delta!} p^{\delta+1} (2P_n(r_\rho))^\delta \sum_{\ell=1}^{\delta+1} \frac{\alpha(\ell, r_\rho)}{\ell} \\ &= \lim_{p \rightarrow \infty} \frac{1}{\delta!} (e_{n,\delta})^\delta \sum_{\ell=1}^{\delta+1} \frac{\alpha_\ell}{\ell} \\ &= \lambda_{n,\delta}(e_{n,\delta}). \end{aligned} \quad (136)$$

(134) and (136) immediately yield the conclusion.  $\square$

## H. Proofs in Section IV

### 1) Proofs of Lemma IV.1 and Proposition IV.3:

*Proof of Lemma IV.1:*

$$\begin{aligned} & \mathbb{E} N_{E_\delta}^{(\mathbf{R})} - \binom{p}{1} \binom{p-1}{\delta} (2P_n(r_\rho))^\delta \\ &= \sum_{\substack{\vec{i} \in C_\delta^< \\ \Sigma_{\vec{i}} \text{ not diagonal}}} \left( \mathbb{E} \prod_{j=1}^{\delta} \Phi_{i_0 i_j}^{(\mathbf{R})} - (2P_n(r_\rho))^\delta \right) \\ &\leq (\mu_{n,\delta+1}(\Sigma) - 1) (2P_n(r_\rho))^\delta \sum_{\substack{\vec{i} \in C_\delta^< \\ \Sigma_{\vec{i}} \text{ not diagonal}}} 1, \end{aligned} \quad (137)$$

where the first inequality follows from Lemma III.4 (c) and Lemma A.2. By (137),

$$\mathbb{E} N_{E_\delta}^{(\mathbf{R})} - \binom{p}{1} \binom{p-1}{\delta} (2P_n(r_\rho))^\delta \geq -(2P_n(r_\rho))^\delta \sum_{\substack{\vec{i} \in C_\delta^< \\ \Sigma_{\vec{i}} \text{ not diagonal}}} 1.$$

Combining the preceding two expressions,

$$\begin{aligned} & \left| \mathbb{E} N_{E_\delta}^{(\mathbf{R})} - \binom{p}{1} \binom{p-1}{\delta} (2P_n(r_\rho))^\delta \right| \\ & \leq \max\{1, \mu_{n,\delta+1}(\Sigma) - 1\} (2P_n(r_\rho))^\delta \sum_{\substack{\vec{i} \in C_\delta^< \\ \Sigma_{\vec{i}} \text{ not diagonal}}} 1 \\ & \leq \mu_{n,\delta+1}(\Sigma) (2P_n(r_\rho))^\delta \frac{(\delta+1)}{2((\delta-1)!)} p^\delta (\kappa-1) \\ & \leq \frac{(\delta+1)}{2((\delta-1)!)} \gamma^\delta \mu_{n,\delta+1}(\Sigma) \frac{\kappa-1}{p}, \end{aligned}$$

where the second inequality follows from Lemma A.3.  $\square$

*Proof of Proposition IV.3:* Recall for  $\vec{i} \in C_\delta^<$ ,  $\Phi_{\vec{i}}^{(\mathbf{R})}$  is defined in (47),  $U_{\vec{i}}$  is defined in (50), and  $Z_{\vec{i}}$  and  $W_{\vec{i}}$  are defined in (72). Then,

$$N_{E_\delta}^{(\mathbf{R})} = \sum_{\vec{i} \in C_\delta^<} \Phi_{\vec{i}}^{(\mathbf{R})} = \Phi_{\vec{i}}^{(\mathbf{R})} + U_{\vec{i}} + Z_{\vec{i}} + W_{\vec{i}}.$$

Then,

$$\left( N_{E_\delta}^{(\mathbf{R})} \right)^2 = \sum_{\vec{i} \in C_\delta^<} \Phi_{\vec{i}}^{(\mathbf{R})} \left( \Phi_{\vec{i}}^{(\mathbf{R})} + U_{\vec{i}} + Z_{\vec{i}} + W_{\vec{i}} \right). \quad (138)$$

**Step 1:**

Since  $\Phi_{\vec{i}}^{(\mathbf{R})} + U_{\vec{i}}$  takes value in  $[\delta+1] \cup \{0\}$ ,

$$\begin{aligned} & \mathbb{E} \Phi_{\vec{i}}^{(\mathbf{R})} \left( \Phi_{\vec{i}}^{(\mathbf{R})} + U_{\vec{i}} \right) \\ &= \sum_{\ell=1}^{\delta+1} \ell \mathbb{E} \Phi_{\vec{i}}^{(\mathbf{R})} \mathbf{1} \left( \Phi_{\vec{i}}^{(\mathbf{R})} + U_{\vec{i}} = \ell \right) \\ &= \sum_{\ell=1}^{\delta+1} \ell \mathbb{P} \left( \Phi_{\vec{i}}^{(\mathbf{R})} = 1 \right) \mathbb{P} \left( \Phi_{\vec{i}}^{(\mathbf{R})} + U_{\vec{i}} = \ell \mid \Phi_{\vec{i}}^{(\mathbf{R})} = 1 \right). \end{aligned}$$

For  $\vec{i} \in C_\delta^<$  such that  $\Sigma_{\vec{i}}$  diagonal, we have  $\mathbb{P} \left( \Phi_{\vec{i}}^{(\mathbf{R})} = 1 \right) = (2P_n(r_\rho))^\delta$  by Lemma A.2 and moreover  $\mathbb{P} \left( \Phi_{\vec{i}}^{(\mathbf{R})} + U_{\vec{i}} = \ell \mid \Phi_{\vec{i}}^{(\mathbf{R})} = 1 \right) = \alpha(\ell, r_\rho)$  by (38). Thus in this case,

$$\mathbb{E} \Phi_{\vec{i}}^{(\mathbf{R})} \left( \Phi_{\vec{i}}^{(\mathbf{R})} + U_{\vec{i}} \right) = (2P_n(r_\rho))^\delta \sum_{\ell=1}^{\delta+1} \ell \alpha(\ell, r_\rho).$$

Moreover, when  $\vec{i} \in C_\delta^<$  such that  $\Sigma_{\vec{i}}$  is not diagonal, by Lemma III.4 (c)

$$\mathbb{E} \Phi_{\vec{i}}^{(\mathbf{R})} \left( \Phi_{\vec{i}}^{(\mathbf{R})} + U_{\vec{i}} \right) \leq \mu_{n,\delta+1}(\Sigma) (2P_n(r_\rho))^\delta \sum_{\ell=1}^{\delta+1} \ell \alpha(\ell, r_\rho).$$

For the ease of notation, denote  $\eta_5 = \binom{p}{1} \binom{p-1}{\delta} (2P_n(r_\rho))^\delta$ . Then by the preceding two expressions,

$$\begin{aligned} & \sum_{\vec{i} \in C_\delta^<} \mathbb{E} \Phi_{\vec{i}}^{(\mathbf{R})} \left( \Phi_{\vec{i}}^{(\mathbf{R})} + U_{\vec{i}} \right) - \eta_5 \sum_{\ell=1}^{\delta+1} \ell \alpha(\ell, r_\rho) \\ &= \sum_{\substack{\vec{i} \in C_\delta^< \\ \Sigma_{\vec{i}} \text{ not diagonal}}} \left( \mathbb{E} \Phi_{\vec{i}}^{(\mathbf{R})} \left( \Phi_{\vec{i}}^{(\mathbf{R})} + U_{\vec{i}} \right) - \right. \end{aligned}$$

$$\begin{aligned}
& (2P_n(r_\rho))^\delta \sum_{\ell=1}^{\delta+1} \ell \alpha(\ell, r_\rho) \Big) \\
& \leq (\mu_{n,\delta+1}(\Sigma) - 1) (2P_n(r_\rho))^\delta \sum_{\ell=1}^{\delta+1} \ell \alpha(\ell, r_\rho) \sum_{\substack{\vec{i} \in C_\delta^< \\ \Sigma_{\vec{i}} \text{ not diagonal}}} 1.
\end{aligned} \tag{139}$$

By (139),

$$\begin{aligned}
& \sum_{\vec{i} \in C_\delta^<} \mathbb{E} \Phi_{\vec{i}}^{(\mathbf{R})} \left( \Phi_{\vec{i}}^{(\mathbf{R})} + U_{\vec{i}} \right) - \eta_5 \sum_{\ell=1}^{\delta+1} \ell \alpha(\ell, r_\rho) \\
& \geq - (2P_n(r_\rho))^\delta \sum_{\ell=1}^{\delta+1} \ell \alpha(\ell, r_\rho) \sum_{\substack{\vec{i} \in C_\delta^< \\ \Sigma_{\vec{i}} \text{ not diagonal}}} 1.
\end{aligned}$$

By combining the preceding two expressions,

$$\begin{aligned}
& \left| \sum_{\vec{i} \in C_\delta^<} \mathbb{E} \Phi_{\vec{i}}^{(\mathbf{R})} \left( \Phi_{\vec{i}}^{(\mathbf{R})} + U_{\vec{i}} \right) - \eta_5 \sum_{\ell=1}^{\delta+1} \ell \alpha(\ell, r_\rho) \right| \\
& \leq \mu_{n,\delta+1}(\Sigma) (2P_n(r_\rho))^\delta \sum_{\ell=1}^{\delta+1} \ell \alpha(\ell, r_\rho) \sum_{\substack{\vec{i} \in C_\delta^< \\ \Sigma_{\vec{i}} \text{ not diagonal}}} 1 \\
& \leq \mu_{n,\delta+1}(\Sigma) (2P_n(r_\rho))^\delta \sum_{\ell=1}^{\delta+1} \ell \alpha(\ell, r_\rho) \frac{(\delta+1)}{2((\delta-1)!)} p^\delta (\kappa-1) \\
& \leq \mu_{n,\delta+1}(\Sigma) (2p^{1+1/\delta} P_n(r_\rho))^\delta \frac{(\delta+1)^2}{2((\delta-1)!)} \frac{\kappa-1}{p} \tag{140}
\end{aligned}$$

where the second inequality follows from Lemma A.3.

**Step 2:**

$$\begin{aligned}
& \sum_{\vec{i} \in C_\delta^<} \mathbb{E} \Phi_{\vec{i}}^{(\mathbf{R})} W_{\vec{i}} - \left( \binom{p}{1} \binom{p-1}{\delta} (2P_n(r_\rho))^\delta \right)^2 \\
& = \sum_{\vec{i} \in C_\delta^<} \sum_{\vec{j} \in T_{\vec{i}}} \mathbb{E} \Phi_{\vec{i}}^{(\mathbf{R})} \Phi_{\vec{j}}^{(\mathbf{R})} - \sum_{\vec{i} \in C_\delta^<} \sum_{\vec{j} \in C_\delta^<} (2P_n(r_\rho))^{2\delta} \\
& = \sum_{\vec{i} \in C_\delta^<} \sum_{\vec{j} \in T_{\vec{i}}} \mathbb{E} \Phi_{\vec{i}}^{(\mathbf{R})} \mathbb{E} \Phi_{\vec{j}}^{(\mathbf{R})} - \sum_{\vec{i} \in C_\delta^<} \sum_{\vec{j} \in T_{\vec{i}}} (2P_n(r_\rho))^{2\delta} \\
& \quad - \sum_{\vec{i} \in C_\delta^<} \sum_{\vec{j} \in C_\delta^< \setminus T_{\vec{i}}} (2P_n(r_\rho))^{2\delta} \\
& = - \sum_{\vec{i} \in C_\delta^<} \sum_{\vec{j} \in C_\delta^< \setminus T_{\vec{i}}} (2P_n(r_\rho))^{2\delta} \sum_{\substack{\vec{i} \in C_\delta^< \\ \Sigma_{\vec{i}} \text{ not diagonal}}} \sum_{\vec{j} \in T_{\vec{i}}} \left( \mathbb{E} \Phi_{\vec{i}}^{(\mathbf{R})} \Phi_{\vec{j}}^{(\mathbf{R})} \right. \\
& \quad \left. - (2P_n(r_\rho))^{2\delta} \right) \\
& \quad + \sum_{\substack{\vec{i} \in C_\delta^< \\ \Sigma_{\vec{i}} \text{ diagonal}}} \sum_{\substack{\vec{j} \in T_{\vec{i}} \\ \Sigma_{\vec{j}} \text{ not diagonal}}} \left( \mathbb{E} \Phi_{\vec{i}}^{(\mathbf{R})} \Phi_{\vec{j}}^{(\mathbf{R})} - (2P_n(r_\rho))^{2\delta} \right), \tag{141}
\end{aligned}$$

where the last equality follows from  $\mathbb{E} \Phi_{\vec{i}}^{(\mathbf{R})} = (2P_n(r_\rho))^\delta$  for  $\vec{i} \in \left\{ \vec{j} \in C_\delta^< : \Sigma_{\vec{j}} \text{ diagonal} \right\}$  by Lemma A.2. Then by (141),

$$\begin{aligned}
& \sum_{\vec{i} \in C_\delta^<} \mathbb{E} \Phi_{\vec{i}}^{(\mathbf{R})} W_{\vec{i}} - \eta_5^2 \\
& \leq (\mu_{n,2\delta+2}(\Sigma) - 1) (2P_n(r_\rho))^{2\delta} \left( \sum_{\substack{\vec{i} \in C_\delta^< \\ \Sigma_{\vec{i}} \text{ not diagonal}}} \sum_{\vec{j} \in T_{\vec{i}}} 1 + \right. \\
& \quad \left. \sum_{\substack{\vec{i} \in C_\delta^< \\ \Sigma_{\vec{i}} \text{ diagonal}}} \sum_{\substack{\vec{j} \in T_{\vec{i}} \\ \Sigma_{\vec{j}} \text{ not diagonal}}} 1 \right),
\end{aligned}$$

where the inequality follows from Lemma III.4 (c). On the other hand, by (141),

$$\begin{aligned}
& \sum_{\vec{i} \in C_\delta^<} \mathbb{E} \Phi_{\vec{i}}^{(\mathbf{R})} W_{\vec{i}} - \left( \binom{p}{1} \binom{p-1}{\delta} (2P_n(r_\rho))^\delta \right)^2 \\
& \geq - (2P_n(r_\rho))^{2\delta} \left( \sum_{\substack{\vec{i} \in C_\delta^< \\ \Sigma_{\vec{i}} \text{ not diagonal}}} \sum_{\vec{j} \in T_{\vec{i}}} 1 + \sum_{\substack{\vec{i} \in C_\delta^< \\ \Sigma_{\vec{i}} \text{ diagonal}}} \sum_{\substack{\vec{j} \in T_{\vec{i}} \\ \Sigma_{\vec{j}} \text{ not diagonal}}} 1 \right) \\
& \quad - \sum_{\vec{i} \in C_\delta^<} \sum_{\vec{j} \in C_\delta^< \setminus T_{\vec{i}}} (2P_n(r_\rho))^{2\delta}.
\end{aligned}$$

Combining the preceding two displays,

$$\begin{aligned}
& \left| \sum_{\vec{i} \in C_\delta^<} \mathbb{E} \Phi_{\vec{i}}^{(\mathbf{R})} W_{\vec{i}} - \left( \binom{p}{1} \binom{p-1}{\delta} (2P_n(r_\rho))^\delta \right)^2 \right| \\
& \leq \mu_{n,2\delta+2}(\Sigma) (2P_n(r_\rho))^{2\delta} \left( \sum_{\substack{\vec{i} \in C_\delta^< \\ \Sigma_{\vec{i}} \text{ not diagonal}}} \sum_{\vec{j} \in T_{\vec{i}}} 1 + \right. \\
& \quad \left. \sum_{\substack{\vec{i} \in C_\delta^< \\ \Sigma_{\vec{i}} \text{ diagonal}}} \sum_{\substack{\vec{j} \in T_{\vec{i}} \\ \Sigma_{\vec{j}} \text{ not diagonal}}} 1 \right) \\
& \quad + \sum_{\vec{i} \in C_\delta^<} \sum_{\vec{j} \in C_\delta^< \setminus T_{\vec{i}}} (2P_n(r_\rho))^{2\delta} \\
& \leq \mu_{n,2\delta+2}(\Sigma) (2P_n(r_\rho))^{2\delta} \left( 2 \sum_{\substack{\vec{i} \in C_\delta^< \\ \Sigma_{\vec{i}} \text{ not diagonal}}} \sum_{\vec{j} \in C_\delta^<} 1 \right) + \\
& \quad \sum_{\vec{i} \in C_\delta^<} \sum_{\vec{j} \in C_\delta^< \setminus T_{\vec{i}}} (2P_n(r_\rho))^{2\delta} \\
& \stackrel{(*)}{\leq} \mu_{n,2\delta+2}(\Sigma) (2P_n(r_\rho))^{2\delta} 2 \binom{p}{1} \binom{p-1}{\delta} \frac{(\delta+1)}{2((\delta-1)!)} p^\delta (\kappa-1) \\
& \quad + \binom{p}{1} \binom{p-1}{\delta} \frac{(\delta+1)^2}{\delta!} p^\delta \kappa (2P_n(r_\rho))^{2\delta}
\end{aligned}$$

$$\leq \frac{2(\delta+1)^2}{(\delta!)^2} (2p^{1+1/\delta} P_n(r_\rho))^{2\delta} \mu_{n,2\delta+2}(\Sigma) \frac{\kappa}{p}, \quad (142)$$

where step (\*) follows from Lemma A.3 and (82).

**Step 3:**

Notice  $\sum_{\vec{i} \in C_\delta^<} \mathbb{E} \Phi_{\vec{i}}^{(\mathbf{R})} Z_{\vec{i}} = b_2$  as in (74) and thus satisfies the bound (89). Then by (138),

$$\begin{aligned} & \left| \mathbb{E} \left( N_{E_\delta}^{(\mathbf{R})} \right)^2 - \eta_5 \sum_{\ell=1}^{\delta+1} \ell \alpha(\ell, r_\rho) - \eta_5^2 \right| \\ & \leq \left| \sum_{\vec{i} \in C_\delta^<} \mathbb{E} \Phi_{\vec{i}}^{(\mathbf{R})} \left( \Phi_{\vec{i}}^{(\mathbf{R})} + U_{\vec{i}} \right) - \eta_5 \sum_{\ell=1}^{\delta+1} \ell \alpha(\ell, r_\rho) \right| \\ & \quad + \left| \sum_{\vec{i} \in C_\delta^<} \mathbb{E} \Phi_{\vec{i}}^{(\mathbf{R})} W_{\vec{i}} - \eta_5^2 \right| + b_2 \\ & \leq C_{n,\delta} \left( (2p^{1+1/\delta} P_n(r_\rho))^\delta (1 + 2p^{1+1/\delta} P_n(r_\rho))^\delta \mu_{n,2\delta+2}(\Sigma) \frac{\kappa}{p} \right. \\ & \quad \left. + p(2p P_n(r_\rho))^{\delta+1} (1 + 2p P_n(r_\rho))^\delta \right), \end{aligned} \quad (143)$$

where the last inequality follows from (140), (142) and (89). The proof is then completed by using the inequality  $2p^{1+1/\delta} P_n(r_\rho) \leq \gamma$ .  $\square$

2) *Proof of Proposition IV.4:*

*Proof of Proposition IV.4 (a):*

By taking the square of each of the terms in Lemma III.8,

$$\begin{aligned} \left( N_{E_\delta}^{(\mathbf{R})} \right)^2 - 2(\delta+1) N_{E_\delta}^{(\mathbf{R})} N_{E_{\delta+1}}^{(\mathbf{R})} & \leq \left( N_{V_\delta}^{(\mathbf{R})} \right)^2 \leq \left( N_{V_\delta}^{(\mathbf{R})} \right)^2 \\ & \leq \left( N_{E_\delta}^{(\mathbf{R})} \right)^2, \end{aligned}$$

which then implies for  $\bar{N}_\delta \in \{N_{V_\delta}^{(\mathbf{R})}, N_{V_\delta}^{(\mathbf{R})}\}$

$$\begin{aligned} \left| (\bar{N}_\delta)^2 - \left( N_{E_\delta}^{(\mathbf{R})} \right)^2 \right| & \leq 2(\delta+1) N_{E_\delta}^{(\mathbf{R})} N_{E_{\delta+1}}^{(\mathbf{R})} \\ & = 2(\delta+1) \sum_{\vec{i} \in C_{\delta+1}^<} \sum_{\vec{j} \in C_\delta^<} \Phi_{\vec{i}}^{(\mathbf{R})} \Phi_{\vec{j}}^{(\mathbf{R})}. \end{aligned} \quad (144)$$

It suffices to establish an upper bound on  $\mathbb{E} N_{E_{\delta+1}}^{(\mathbf{R})} N_{E_\delta}^{(\mathbf{R})} = \mathbb{E} \sum_{\vec{i} \in C_{\delta+1}^<} \sum_{\vec{j} \in C_\delta^<} \Phi_{\vec{i}}^{(\mathbf{R})} \Phi_{\vec{j}}^{(\mathbf{R})}$ .

Observe for  $\vec{j} \in J_{\vec{i}}$ ,

$$\mathbb{E} \Phi_{\vec{i}}^{(\mathbf{R})} \Phi_{\vec{j}}^{(\mathbf{R})} \leq \mathbb{E} \Phi_{\vec{i}}^{(\mathbf{R})} \leq \mu_{n,\delta+2}(\Sigma_{\vec{i}}) (2P_n(r_\rho))^{\delta+1}.$$

For  $\vec{j} \in T_{\vec{i}}$ ,  $[\vec{j}] \cap [\vec{i}] = \emptyset$ . Thus, if  $\Sigma_{\vec{i} \cup \vec{j}}$  is diagonal,  $\mathbb{E} \Phi_{\vec{i}}^{(\mathbf{R})} \Phi_{\vec{j}}^{(\mathbf{R})} = \mathbb{E} \Phi_{\vec{i}}^{(\mathbf{R})} \mathbb{E} \Phi_{\vec{j}}^{(\mathbf{R})} = (2P_n(r_\rho))^{2\delta+1}$  by Lemma A.2. Then, for the general case when  $\Sigma_{\vec{i} \cup \vec{j}}$  is not necessarily diagonal, by Lemma III.4 (c),

$$\mathbb{E} \Phi_{\vec{i}}^{(\mathbf{R})} \Phi_{\vec{j}}^{(\mathbf{R})} \leq \mu_{n,2\delta+3}(\Sigma_{\vec{i} \cup \vec{j}}) (2P_n(r_\rho))^{2\delta+1}$$

By Lemma III.4 (c), Lemma A.2 and Lemma A.5, it is straightforward that the conditions in Lemma A.7 with  $q = \delta+1$  and  $\theta_{\vec{i},\vec{j}} = \Phi_{\vec{i}}^{(\mathbf{R})} \Phi_{\vec{j}}^{(\mathbf{R})}$  are satisfied with  $a = 1$ ,  $b = 2P_n(2r_\rho)1(\delta \geq 2) + 2P_n(r_\rho)1(\delta = 1)$  and  $z = 2P_n(r_\rho)$ . Moreover,  $b/z \leq 2^{n-2}1(\delta \geq 2) + 1$  by Lemma A.22 (d).

Thus, Lemma A.8 with  $q = \delta+1$  and  $\theta_{\vec{i},\vec{j}} = \Phi_{\vec{i}}^{(\mathbf{R})} \Phi_{\vec{j}}^{(\mathbf{R})}$ ,  $a = 1$ ,  $b = 2P_n(2r_\rho)1(\delta \geq 2) + 2P_n(r_\rho)1(\delta = 1)$  and  $z = 2P_n(r_\rho)$ , together with the fact that  $pz \leq p^{1+\frac{1}{\delta}}z = 2p^{1+\frac{1}{\delta}}P_n(r_\rho) \leq \gamma$ , yield

$$\begin{aligned} \mathbb{E} N_{E_{\delta+1}}^{(\mathbf{R})} N_{E_\delta}^{(\mathbf{R})} & = \mathbb{E} \sum_{\vec{i} \in C_{\delta+1}^<} \sum_{\vec{j} \in C_\delta^<} \Phi_{\vec{i}}^{(\mathbf{R})} \Phi_{\vec{j}}^{(\mathbf{R})} \\ & \leq C_{n,\delta,\gamma} \left( 1 + \mu_{n,2\delta+3}(\Sigma) \frac{\kappa-1}{p} \right) p^{-1/\delta}. \end{aligned} \quad (145)$$

The proof is then complete by the preceding expression and (144).  $\square$

We now present a few lemmas that are used in the proof of Proposition IV.4 (b) and (c). Recall that  $F_{ij}(r_\rho)$ ,  $H_{ij}(r_\rho)$ ,  $G_{ij}(r_\rho)$ ,  $F_i(r_\rho)$  are defined in Section F.3.

**Lemma A.17:** Suppose  $p \geq n$ .  $1 \leq \delta \leq q \leq p-1$ . Then for any  $\vec{i} \in C_q^<$ ,  $\vec{j} \in C_\delta^<$ , with probability 1,

$$\left| \Phi_{\vec{i}}^{(\mathbf{P})} \Phi_{\vec{j}}^{(\mathbf{P})} - \Phi_{\vec{i}}^{(\mathbf{R})} \Phi_{\vec{j}}^{(\mathbf{R})} \right| \leq \xi_{\vec{i},\vec{j}},$$

where

$$\begin{aligned} \xi_{\vec{i},\vec{j}} & := 1 \left( \bigcup_{m=1}^q \left( (H_{i_0 i_m} \setminus G_{i_0 i_m}) \cap H_{\vec{i},-m} \cap H_{\vec{j}} \right) \cup \right. \\ & \quad \left. \bigcup_{\ell=1}^\delta \left( (H_{j_0 j_\ell} \setminus G_{j_0 j_\ell}) \cap H_{\vec{j},-\ell} \cap H_{\vec{i}} \right) \right). \end{aligned}$$

*Proof:*

$$\begin{aligned} & \left| \Phi_{\vec{i}}^{(\mathbf{P})} \Phi_{\vec{j}}^{(\mathbf{P})} - \Phi_{\vec{i}}^{(\mathbf{R})} \Phi_{\vec{j}}^{(\mathbf{R})} \right| \\ & = 1 \left( \left( \bigcap_{m=1}^q F_{i_0 i_m} \cap \bigcap_{\ell=1}^\delta F_{j_0 j_\ell} \right) \triangle \left( \bigcap_{m=1}^q S_{i_0 i_m} \cap \bigcap_{\ell=1}^\delta S_{j_0 j_\ell} \right) \right) \\ & \leq \xi_{\vec{i},\vec{j}} \end{aligned}$$

where the inequality follows from (97) and Lemma A.29 (a).  $\square$

**Lemma A.18:** Let  $p \geq n \geq 4$ ,  $1 \leq \delta \leq q \leq p-1$  and  $\mathbf{X} \sim \mathcal{V}\mathcal{E}(\mu, \Sigma, g)$ . Suppose  $\Sigma$ , after some row-column permutation, is  $(\tau, \kappa)$  sparse with  $\tau \leq \frac{p}{2}$ . Let  $t$  be any positive number, and suppose (102) holds. Then for any  $\vec{i} \in C_q^<$ ,  $\vec{j} \in C_\delta^<$ , with probability 1,

$$\xi_{\vec{i},\vec{j}} 1(\mathcal{E}(t)) \leq \eta_{\vec{i},\vec{j}}(t),$$

where

$$\begin{aligned} \eta_{\vec{i},\vec{j}}(t) & := 1 \left( \bigcup_{m=1}^q \left( \left( F_{i_0 i_m}(\theta_1(t)r_\rho) \setminus F_{i_0 i_m} \left( \frac{r_\rho}{\theta_1(t)} \right) \right) \cap \right. \right. \\ & \quad \left. \left. F_{\vec{i},-m}(\theta_1(t)r_\rho) \cap F_{\vec{j}}(\theta_1(t)r_\rho) \right) \cup \right. \\ & \quad \left. \bigcup_{\ell=1}^\delta \left( \left( F_{j_0 j_\ell}(\theta_1(t)r_\rho) \setminus F_{j_0 j_\ell} \left( \frac{r_\rho}{\theta_1(t)} \right) \right) \cap \right. \right. \\ & \quad \left. \left. F_{\vec{j},-\ell}(\theta_1(t)r_\rho) \cap F_{\vec{i}}(\theta_1(t)r_\rho) \right) \right). \end{aligned}$$

*Proof:* By (103),  $H_{ij}(r_\rho) \cap \mathcal{E}(t) \subset F_{ij}(\theta_1(t)r_\rho)$  and  $G_{ij}(r_\rho) \cap \mathcal{E}(t) \supset F_{ij}\left(\frac{r_\rho}{\theta_1(t)}\right)$ . Then

$$\xi_{i,j}^{-1}(\mathcal{E}(t)) \leq \eta_{i,j}^{-1}(t).$$

□

**Lemma A.19:** Let  $p \geq n \geq 4$ ,  $1 \leq \delta \leq q \leq p-1$  and  $\mathbf{X} \sim \mathcal{V}\mathcal{E}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, g)$ . Suppose  $\boldsymbol{\Sigma}$ , after some row-column permutation, is  $(\tau, \kappa)$  sparse with  $\tau \leq \frac{p}{2}$ . Let  $t$  be any positive number, and suppose (102) holds. Suppose additionally  $2p^{1+\frac{1}{\delta}}P_n(r_\rho) \leq \gamma$ . Then

$$\begin{aligned} & \mathbb{E} \sum_{\vec{i} \in C_q^<} \sum_{\vec{j} \in C_\delta^<} \eta_{\vec{i},\vec{j}}^{-1}(t) \\ & \leq C_{n,q,\delta,\gamma} (\theta_1(t))^{n(2\delta+q)} \eta_6 \left( \sqrt{\frac{1}{p}} + \frac{t}{\sqrt{p}} + \frac{\tau}{p} \right) p^{1-\frac{\delta}{2}}, \end{aligned}$$

where  $\eta_6 := 1 + \mu_{n,q+\delta+2}(\boldsymbol{\Sigma}) \frac{\kappa-1}{p}$ .

*Proof:* Note

$$\begin{aligned} & \sum_{\vec{i} \in C_q^<} \sum_{\vec{j} \in C_\delta^<} \eta_{\vec{i},\vec{j}}^{-1}(t) \\ & = \left( \sum_{\vec{i} \in C_q^<} \sum_{\vec{j} \in J_i^<} + \sum_{\vec{i} \in C_q^<} \sum_{\vec{j} \in T_i^<} + \sum_{\vec{i} \in C_q^<} \sum_{\vec{j} \in N_i^<} \right) \eta_{\vec{i},\vec{j}}^{-1}(t). \end{aligned}$$

**Step 1:**  $\vec{j} \in T_i^<$

By the union bound for indicator functions,

$$\eta_{\vec{i},\vec{j}}^{-1}(t) \leq \eta_i^{-1}(t) 1\left(F_{\vec{j}}(\theta_1(t)r_\rho)\right) + \eta_{\vec{j}}^{-1}(t) 1\left(F_i(\theta_1(t)r_\rho)\right), \quad (146)$$

where  $\eta_i^{-1}(t)$  is defined in (104) with  $\delta$  replaced by  $q$ . Then for  $\vec{j} \in T_i^<$

$$\mathbb{E} \eta_{\vec{i},\vec{j}}^{-1}(t) \leq \mathbb{E} \eta_i^{-1}(t) \mathbb{P}\left(F_{\vec{j}}(\theta_1(t)r_\rho)\right) + \mathbb{E} \eta_{\vec{j}}^{-1}(t) \mathbb{P}\left(F_i(\theta_1(t)r_\rho)\right). \quad (147)$$

Moreover, for  $\vec{j} \in T_i^<$ ,  $\boldsymbol{\Sigma}_{\vec{i} \cup \vec{j}}$  is diagonal if and only if  $\boldsymbol{\Sigma}_{\vec{i}}$  and  $\boldsymbol{\Sigma}_{\vec{j}}$  are both diagonal.

Now suppose  $\boldsymbol{\Sigma}_{\vec{i} \cup \vec{j}}$  is diagonal. By conditioning on  $\mathbf{u}_{j_0}$

$$\begin{aligned} \mathbb{P}\left(F_{\vec{j}}(\theta_1(t)r_\rho)\right) &= (2P_n(\theta_1(t)r_\rho))^\delta, \\ \mathbb{P}\left(F_i(\theta_1(t)r_\rho)\right) &= (2P_n(\theta_1(t)r_\rho))^q. \end{aligned}$$

The preceding expression, (147), Lemma A.13 applied to  $\mathbb{E} \eta_{\vec{j}}^{-1}(t)$ , and Lemma A.13 with  $\delta = q$  applied to  $\mathbb{E} \eta_i^{-1}(t)$  yield

$$\begin{aligned} & \mathbb{E} \eta_{\vec{i},\vec{j}}^{-1}(t) \\ & \leq (\delta + q) 2\theta_2(t) (2P_n(\theta_1(t)r_\rho))^{q+\delta-1}, \end{aligned}$$

where  $\theta_2(t) := P_n(r_\rho \theta_1(t)) - P_n\left(\frac{r_\rho}{\theta_1(t)}\right)$ .

For the general case that  $\boldsymbol{\Sigma}_{\vec{i} \cup \vec{j}}$  is not necessarily diagonal, by Lemma III.4 (c), for any  $\vec{j} \in T_i^<$

$$\begin{aligned} & \mathbb{E} \eta_{\vec{i},\vec{j}}^{-1}(t) \\ & \leq \mu_{n,q+\delta+2}(\boldsymbol{\Sigma}_{\vec{i} \cup \vec{j}}) (\delta + q) 2\theta_2(t) (2P_n(\theta_1(t)r_\rho))^{q+\delta-1}. \end{aligned}$$

Then the condition in Lemma A.6 (b) is satisfied with  $\theta_{i,j}^{-1} = \eta_{i,j}^{-1}$ ,  $z = 2P_n(r_\rho \theta_1(t))$  and  $a = (\delta + q) \frac{P_n(r_\rho \theta_1(t)) - P_n\left(\frac{r_\rho}{\theta_1(t)}\right)}{P_n(r_\rho \theta_1(t))}$ .

**Step 2:**  $\vec{j} \in J_i^<$

(146) implies

$$\begin{aligned} & \eta_{\vec{i},\vec{j}}^{-1}(t) \\ & \leq \eta_i^{-1}(t) + \\ & \sum_{\ell=1}^{\delta} 1\left(F_{j_0 j_\ell}(\theta_1(t)r_\rho) \setminus F_{j_0 j_\ell}\left(\frac{r_\rho}{\theta_1(t)}\right)\right) 1(F_i(\theta_1(t)r_\rho)). \end{aligned} \quad (148)$$

For  $\vec{j} \in J_i^<$ ,  $j_0, j_\ell \in [\vec{i}]$ . If  $i_0 \in \{j_0, j_\ell\}$ , without loss of generality, say  $i_0 = j_\ell$  and  $j_0 = i_\alpha$  for some  $1 \leq \alpha \leq q$ . Then

$$\begin{aligned} & \mathbb{E} 1\left(F_{j_0 j_\ell}(\theta_1(t)r_\rho) \setminus F_{j_0 j_\ell}\left(\frac{r_\rho}{\theta_1(t)}\right)\right) 1(F_i(\theta_1(t)r_\rho)) \\ & = \mathbb{P}\left(\left(F_{i_0 i_\alpha}(\theta_1(t)r_\rho) \setminus F_{i_0 i_\alpha}\left(\frac{r_\rho}{\theta_1(t)}\right)\right) \cap F_{i,-\alpha}(\theta_1(t)r_\rho)\right) \\ & \leq \mu_{n,q+1}(\boldsymbol{\Sigma}_{\vec{i}}) 2\theta_2(t) (2P_n(\theta_1(t)r_\rho))^{q-1}, \end{aligned} \quad (149)$$

where the last step follows from Lemma A.13 with  $\delta$  replace by  $q$ .

If  $i_0 \notin \{j_0, j_\ell\}$ , without loss of generality, let  $j_0 = i_\alpha$ ,  $j_\ell = i_\beta$  for some  $1 \leq \alpha \neq \beta \leq q$ . Suppose for now that  $\boldsymbol{\Sigma}_{\vec{i}}$  is diagonal, and then

$$\begin{aligned} & \mathbb{E} 1\left(F_{j_0 j_\ell}(\theta_1(t)r_\rho) \setminus F_{j_0 j_\ell}\left(\frac{r_\rho}{\theta_1(t)}\right)\right) 1(F_i(\theta_1(t)r_\rho)) \\ & = \mathbb{E} 1\left(F_{i_\alpha i_\beta}(\theta_1(t)r_\rho) \setminus F_{i_\alpha i_\beta}\left(\frac{r_\rho}{\theta_1(t)}\right)\right) 1(F_i(\theta_1(t)r_\rho)) \\ & \stackrel{(*)}{=} (2P_n(\theta_1(t)r_\rho))^{q-2} \mathbb{E} 1\left(F_{i_\alpha i_\beta}(\theta_1(t)r_\rho) \setminus F_{i_\alpha i_\beta}\left(\frac{r_\rho}{\theta_1(t)}\right)\right) \\ & \quad 1(F_{i_0 i_\alpha}(\theta_1(t)r_\rho)) 1(F_{i_0 i_\beta}(\theta_1(t)r_\rho)) \\ & \stackrel{(**)}{\leq} 2 \left(P_n(\theta_1(t)r_\rho) - P_n\left(\frac{r_\rho}{\theta_1(t)}\right)\right) (2P_n(\theta_1(t)r_\rho))^{q-1}, \end{aligned}$$

where the step (\*) follows from conditioning on  $\mathbf{u}_{i_0}, \mathbf{u}_{i_\alpha}, \mathbf{u}_{i_\beta}$ , and the step (\*\*) follows from dropping the term  $1(F_{i_0 i_\beta}(\theta_1(t)r_\rho))$  and then conditioning on  $\mathbf{u}_{i_\alpha}$ . For the general case that  $\boldsymbol{\Sigma}_{\vec{i}}$  is not necessarily diagonal, by Lemma III.4 (c),

$$\begin{aligned} & \mathbb{E} 1\left(F_{j_0 j_\ell}(\theta_1(t)r_\rho) \setminus F_{j_0 j_\ell}\left(\frac{r_\rho}{\theta_1(t)}\right)\right) 1(F_i(\theta_1(t)r_\rho)) \\ & \leq \mu_{n,q+1}(\boldsymbol{\Sigma}_{\vec{i}}) 2\theta_2(t) (2P_n(\theta_1(t)r_\rho))^{q-1}. \end{aligned} \quad (150)$$

By combining (148), (149), (150) and Lemma A.13 with  $\delta$  replace by  $q$ ,

$$\mathbb{E} \eta_{\vec{i},\vec{j}}^{-1}(t) \leq (q + \delta) \mu_{n,q+1}(\boldsymbol{\Sigma}_{\vec{i}}) 2\theta_2(t) (2P_n(\theta_1(t)r_\rho))^{q-1}.$$

Then the conditions in Lemma A.6 (a) with  $\theta_{i,j}^{-1} = \eta_{i,j}^{-1}$ ,  $z = 2P_n(r_\rho \theta_1(t))$  and  $a = (q + \delta) \frac{P_n(\theta_1(t)r_\rho) - P_n\left(\frac{r_\rho}{\theta_1(t)}\right)}{P_n(\theta_1(t)r_\rho)}$  is satisfied.

**Step 3:**  $\vec{j} \in N_i^<$

It is straightforward by Lemma III.4 (c), Lemma A.2 and Lemma A.5 that the conditions in Lemma A.7 with  $\theta_{i,j}^{-1} = \eta_{i,j}^{-1}$

are satisfied with  $a = a_1 = (q + \delta) \frac{P_n(\theta_1(t)r_\rho) - P_n\left(\frac{r_\rho}{\theta_1(t)}\right)}{P_n(\theta_1(t)r_\rho)}$ ,  $b = 2P_n(2r_\rho\theta_1(t))1(\delta \geq 2) + 2P_n(r_\rho\theta_1(t))1(\delta = 1)$  and  $z = 2P_n(r_\rho\theta_1(t))$ . Moreover,  $b/z \leq 2^{n-2}1(\delta \geq 2) + 1$  by Lemma A.22 (d).

By Lemma A.8 with  $\theta_{\vec{i}, \vec{j}} = \eta_{\vec{i}, \vec{j}}$ ,  $a = (q + \delta) \frac{P_n(\theta_1(t)r_\rho) - P_n\left(\frac{r_\rho}{\theta_1(t)}\right)}{P_n(\theta_1(t)r_\rho)}$ ,  $b = 2P_n(2r_\rho\theta_1(t))1(\delta \geq 2) + 2P_n(r_\rho\theta_1(t))1(\delta = 1)$  and  $z = 2P_n(r_\rho\theta_1(t))$

$$\begin{aligned} & \sum_{\vec{i} \in C_q^<} \sum_{\vec{j} \in C_\delta^<} \mathbb{E} \eta_{\vec{i}, \vec{j}} \\ & \leq C_{n,q,\delta} \left(p^{1+\frac{1}{\delta}} z\right)^q \left(1 + (p^{1+\frac{1}{\delta}} z)^\delta\right) (1 + pz)^{\delta-1} \eta_6 a p^{1-\frac{q}{p}}. \end{aligned} \quad (151)$$

#### Step 4

Observe that

$$pz \leq p^{1+\frac{1}{\delta}} z \leq (\theta_1(t))^{n-2} 2p^{1+\frac{1}{\delta}} P_n(r_\rho) \leq (\theta_1(t))^{n-2} \gamma \quad (152)$$

where the second inequality follows from Lemma A.22 (d). Moreover, by Lemma A.22 (c) and the fact that  $\theta_1(t) \geq 1$ ,

$$\begin{aligned} a & \leq (q + \delta) \frac{P_n(\theta_1(t)r_\rho) - P_n\left(\frac{r_\rho}{\theta_1(t)}\right)}{P_n(r_\rho)} \\ & \leq (q + \delta)(n - 2) (\theta_1(t))^{n-3} \left(\theta_1(t) - \frac{1}{\theta_1(t)}\right). \end{aligned} \quad (153)$$

Plugging (152) and (153) into (151) and by the fact that  $\theta_1(t) \geq 1$ ,

$$\begin{aligned} & \sum_{\vec{i} \in C_q^<} \sum_{\vec{j} \in C_\delta^<} \eta_{\vec{i}, \vec{j}} \\ & \leq C_{n,q,\delta,\gamma} (\theta_1(t))^{n(2\delta+q)} \eta_6 \left(\theta_1(t) - \frac{1}{\theta_1(t)}\right) p^{1-\frac{q}{p}} \\ & \leq C_{n,q,\delta,\gamma} (\theta_1(t))^{n(2\delta+q)} \eta_6 \left(\sqrt{\frac{n}{p}} + \frac{t}{\sqrt{p}} + n\frac{\tau}{p}\right) p^{1-\frac{q}{p}}, \end{aligned}$$

where the last inequality follows from Lemma A.26 (d) and (103).  $\square$

**Lemma A.20:** Let  $p \geq n \geq 4$  and  $\mathbf{X} \sim \mathcal{VE}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, g)$ . Suppose  $\boldsymbol{\Sigma}$ , after some row-column permutation, is  $(\tau, \kappa)$  sparse with  $\tau \leq \frac{p}{2}$ . Consider  $1 \leq \delta \leq p - 2$  and let  $q \in \{\delta, \delta + 1\}$ . Suppose  $2p^{1+\frac{1}{\delta}} P_n(r_\rho) \leq \gamma$  and  $\left(\sqrt{\frac{n-1}{p}} + \sqrt{\frac{\delta \ln p}{p}}\right) \leq c$  hold for some positive and sufficiently small universal constant  $c$ . Then

$$\begin{aligned} & \mathbb{E} \left| N_{E_q}^{(\mathbf{P})} N_{E_\delta}^{(\mathbf{P})} - N_{E_q}^{(\mathbf{R})} N_{E_\delta}^{(\mathbf{R})} \right| \\ & \leq C_{n,\delta,\gamma} \left(1 + \mu_{n,q+\delta+2}(\boldsymbol{\Sigma}) \frac{\kappa - 1}{p}\right) \left(\frac{\sqrt{\ln p}}{\sqrt{p}} + \frac{\tau}{p}\right) p^{1-\frac{q}{p}}. \end{aligned}$$

*Proof:* For  $\boldsymbol{\Psi} \in \{\mathbf{R}, \mathbf{P}\}$ ,

$$N_{E_q}^{(\boldsymbol{\Psi})} N_{E_\delta}^{(\boldsymbol{\Psi})} = \sum_{\vec{i} \in C_q^<} \sum_{\vec{j} \in C_\delta^<} \Phi_{\vec{i}}^{(\boldsymbol{\Psi})} \Phi_{\vec{j}}^{(\boldsymbol{\Psi})}.$$

Thus,

$$\begin{aligned} & \mathbb{E} \left| N_{E_q}^{(\mathbf{P})} N_{E_\delta}^{(\mathbf{P})} - N_{E_q}^{(\mathbf{R})} N_{E_\delta}^{(\mathbf{R})} \right| \\ & \leq \mathbb{E} \sum_{\vec{i} \in C_q^<} \sum_{\vec{j} \in C_\delta^<} \left| \Phi_{\vec{i}}^{(\mathbf{P})} \Phi_{\vec{j}}^{(\mathbf{P})} - \Phi_{\vec{i}}^{(\mathbf{R})} \Phi_{\vec{j}}^{(\mathbf{R})} \right| 1(\mathcal{E}(t)) + \\ & \quad \binom{p}{1} \binom{p-1}{q} \binom{p}{1} \binom{p-1}{\delta} \mathbb{P}(\mathcal{E}^c(t)) \\ & \leq \mathbb{E} \sum_{\vec{i} \in C_q^<} \sum_{\vec{j} \in C_\delta^<} \eta_{\vec{i}, \vec{j}} + \frac{p^{q+\delta+2}}{\delta! q!} 2 \exp(-c_1 t^2) \end{aligned} \quad (154)$$

where the first inequality follows from  $0 \leq N_{E_\delta}^{(\boldsymbol{\Psi})} \leq \binom{p}{1} \binom{p-1}{\delta}$  for both  $\boldsymbol{\Psi} = \mathbf{R}$  and  $\boldsymbol{\Psi} = \mathbf{P}$ , and the second inequality follows from Lemma A.17, Lemma A.18 and (99).

Choose  $t = s_0 \sqrt{\ln p}$  with  $s_0 = \sqrt{\left(\frac{9}{2} + 2\delta\right) / c_1}$ . Notice that  $s_0 \geq \sqrt{\left(\frac{3}{2} + q + \delta + \frac{q}{\delta}\right) / c_1}$  since  $q \in \{\delta, \delta + 1\}$ . Then

$$\begin{aligned} 2 \exp(-c_1 t^2) & \leq 2 \exp\left(-\left(\frac{3}{2} + q + \delta + \frac{q}{\delta}\right) \ln p\right) \\ & = \frac{2}{p^{\frac{3}{2} + q + \delta + \frac{q}{\delta}}}. \end{aligned}$$

Moreover, for any  $c < \frac{1}{2 \max\left\{\sqrt{\left(\frac{9}{2} + 2\delta\right) / c_1}, 1\right\}} \sqrt{2} C_1$ ,

$$\left(\sqrt{\frac{n-1}{p}} + \sqrt{\frac{\delta \ln p}{p}}\right) \leq c$$

implies

$$\sqrt{2} C_1 \left(\sqrt{\frac{n-1}{p}} + s_0 \sqrt{\frac{\ln p}{p}}\right) \leq \frac{1}{2}, \quad (155)$$

which is (102) with  $t = s_0 \sqrt{\ln p}$ . Then apply Lemma A.19 with  $t = s_0 \sqrt{\ln p}$  to (154),

$$\begin{aligned} & \mathbb{E} \left| N_{E_q}^{(\mathbf{P})} N_{E_\delta}^{(\mathbf{P})} - N_{E_q}^{(\mathbf{R})} N_{E_\delta}^{(\mathbf{R})} \right| \\ & \leq C_{n,q,\delta,\gamma} (\theta_1(s_0 \ln p))^{n(2\delta+q)} \eta_6 \left(\sqrt{\frac{1}{p}} + \frac{s_0 \sqrt{\ln p}}{\sqrt{p}} + \frac{\tau}{p}\right) p^{1-\frac{q}{p}} \\ & \quad + \frac{2}{\delta! q! \sqrt{p}} p^{1-\frac{q}{p}} \\ & \leq C_{n,q,\delta,\gamma} \eta_6 \left(s_0 \frac{\sqrt{\ln p}}{\sqrt{p}} + \frac{\tau}{p}\right) p^{1-\frac{q}{p}} + \frac{2}{\delta! q! \sqrt{p}} p^{1-\frac{q}{p}} \\ & \leq C_{n,q,\delta,\gamma} \eta_6 \left(\frac{\sqrt{\ln p}}{\sqrt{p}} + \frac{\tau}{p}\right) p^{1-\frac{q}{p}} \\ & \leq C_{n,\delta,\gamma} \eta_6 \left(\frac{\sqrt{\ln p}}{\sqrt{p}} + \frac{\tau}{p}\right) p^{1-\frac{q}{p}}, \end{aligned}$$

where the second inequality follows from  $\theta_1(s_0 \sqrt{\ln p}) \leq 9 + 4(n-1) = 4n+5$  by (155) and  $\tau \leq p/2$ ; and the last step follows from  $q \in \{\delta, \delta + 1\}$ .  $\square$

*Proof of Proposition IV.4 (b) and (c):* (b) It follows directly from Lemma A.20 with  $q = \delta$ .

(c) By taking square of each terms in Lemma III.8,

$$\begin{aligned} \left(N_{E_\delta}^{(\mathbf{P})}\right)^2 - 2(\delta + 1) N_{E_\delta}^{(\mathbf{P})} N_{E_{\delta+1}}^{(\mathbf{P})} & \leq \left(N_{\check{V}_\delta}^{(\mathbf{P})}\right)^2 \leq \left(N_{V_\delta}^{(\mathbf{P})}\right)^2 \\ & \leq \left(N_{E_\delta}^{(\mathbf{P})}\right)^2, \end{aligned}$$

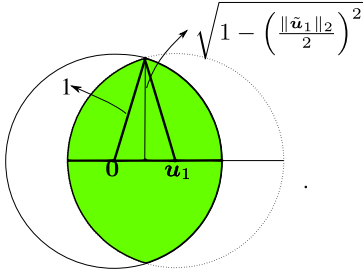


Fig. 5. The solid circle represents the unit Euclidean ball  $B_2^n$  in  $\mathbb{R}^n$  while the dash circle represents the unit ball centered at  $\tilde{\mathbf{u}}_1$ . Their intersection is the green region, which is contained in the ball with center at  $\tilde{\mathbf{u}}_1/2$  and with radius  $\sqrt{1 - \left(\frac{\|\tilde{\mathbf{u}}_1\|_2}{2}\right)^2}$ .

which then implies for  $\bar{N}_\delta \in \{N_{\tilde{V}_\delta}^{(\mathbf{P})}, N_{V_\delta}^{(\mathbf{P})}\}$

$$\begin{aligned} & \left| \bar{N}_\delta - \left( N_{E_\delta}^{(\mathbf{P})} \right)^2 \right| \\ & \leq 2(\delta + 1) \left( N_{E_{\delta+1}}^{(\mathbf{P})} N_{E_\delta}^{(\mathbf{P})} - N_{E_{\delta+1}}^{(\mathbf{R})} N_{E_\delta}^{(\mathbf{R})} \right) + 2(\delta + 1) N_{E_{\delta+1}}^{(\mathbf{R})} N_{E_\delta}^{(\mathbf{R})}. \end{aligned}$$

By Lemma A.20 with  $q = \delta + 1$ ,

$$\begin{aligned} & \mathbb{E} \left| N_{E_{\delta+1}}^{(\mathbf{P})} N_{E_\delta}^{(\mathbf{P})} - N_{E_{\delta+1}}^{(\mathbf{R})} N_{E_\delta}^{(\mathbf{R})} \right| \\ & \leq C_{n,\delta,\gamma} \left( 1 + \mu_{n,2\delta+3}(\Sigma) \frac{\kappa - 1}{p} \right) \left( \frac{\sqrt{\ln p}}{\sqrt{p}} + \frac{\tau}{p} \right) p^{-\frac{1}{\delta}}. \end{aligned}$$

The proof is then completed by combining the preceding two displays, (145) and the fact that

$$\left( \frac{\sqrt{\ln p}}{\sqrt{p}} + \frac{\tau}{p} \right) \leq 1.$$

□

## I. Proofs in Section V

### 1) Proof of Lemma V.1 and Lemma V.2:

*Proof of Lemma V.1:* (a) Denote by  $\deg(\cdot)$  the degree of a vertex in  $\mathbf{Ge}(\{\tilde{\mathbf{u}}_i\}_{i=1}^m, 1; \mathcal{N})$ . Then by union bound,

$$\begin{aligned} & \mathbb{P}(\mathbf{NUV}(\{\tilde{\mathbf{u}}_i\}_{i=1}^\delta, 1; n-2) \geq 1) \\ & \leq \delta \mathbb{P}(\deg(\tilde{\mathbf{u}}_1) = \delta - 1) \\ & = \delta \mathbb{E}(\mathbb{P}(\deg(\tilde{\mathbf{u}}_1) = \delta - 1 | \tilde{\mathbf{u}}_1)) \\ & = \delta \mathbb{E}(\mathbb{P}(\|\tilde{\mathbf{u}}_1 - \tilde{\mathbf{u}}_2\|_2 \leq 1, \dots, \|\tilde{\mathbf{u}}_1 - \tilde{\mathbf{u}}_\delta\|_2 \leq 1 | \tilde{\mathbf{u}}_1)) \\ & = \delta \mathbb{E}(\mathbb{P}(\|\tilde{\mathbf{u}}_1 - \tilde{\mathbf{u}}_2\|_2 \leq 1 | \tilde{\mathbf{u}}_1)^{\delta-1}), \end{aligned} \quad (156)$$

where the last equality follows by conditional independence.

As illustrated in Figure 5,  $\mathbb{P}(\|\tilde{\mathbf{u}}_1 - \tilde{\mathbf{u}}_2\|_2 \leq 1 | \tilde{\mathbf{u}}_1)$  is the ratio between Lebesgue measure of green region and  $|B_2^{n-2}|$ . Moreover, the Lebesgue measure of the green region is less than

$$\left( 1 - \left( \frac{\|\tilde{\mathbf{u}}_1\|_2}{2} \right)^2 \right)^{\frac{n-2}{2}} |B_2^{n-2}|.$$

Then

$$\mathbb{P}(\|\tilde{\mathbf{u}}_1 - \tilde{\mathbf{u}}_2\|_2 \leq 1 | \tilde{\mathbf{u}}_1) \leq \left( 1 - \left( \frac{\|\tilde{\mathbf{u}}_1\|_2}{2} \right)^2 \right)^{\frac{n-2}{2}} \quad a.s. \quad (157)$$

By combining (156) and (157),

$$\begin{aligned} & \mathbb{P}(\mathbf{NUV}(\{\tilde{\mathbf{u}}_i\}_{i=1}^\delta, 1; n-2) \geq 1) \\ & \leq \delta \mathbb{E} \left( 1 - \left( \frac{\|\tilde{\mathbf{u}}_1\|_2}{2} \right)^2 \right)^{\frac{(n-2)(\delta-1)}{2}} \\ & = \delta(n-2) \int_0^1 \left( 1 - \frac{r^2}{4} \right)^{\frac{(n-2)(\delta-1)}{2}} r^{n-3} dr \\ & = \delta(n-2) 2^{n-3} B \left( \frac{1}{4}; \frac{n-2}{2}, \frac{(n-2)(\delta-1)}{2} + 1 \right), \end{aligned} \quad (158)$$

where the first equality follows from expressing the integral in polar coordinates, and the last step follows from changing the variables  $r = 2\sqrt{y}$ .

(b) Denote  $f(r; \alpha, \beta) = \left( 1 - \frac{r^2}{4} \right)^\alpha r^\beta$ . Then it is easy to verify that for any  $\alpha, \beta > 0$ ,

$$\begin{aligned} & \max_{r \in [0,1]} f(r; \alpha, \beta) \\ & = \begin{cases} f(1; \alpha, \beta) = \left( \frac{3}{4} \right)^\alpha & \text{if } 3\beta \geq 2\alpha, \\ f\left(\sqrt{\frac{4\beta}{2\alpha+\beta}}; \alpha, \beta\right) = \left( \frac{2\alpha}{2\alpha+\beta} \right)^\alpha \left( \frac{4\beta}{2\alpha+\beta} \right)^{\frac{\beta}{2}} & \text{if } 3\beta \leq 2\alpha. \end{cases} \end{aligned} \quad (159)$$

Moreover,  $f(r; \alpha, \beta)$  is increasing on  $[0, 1]$  if  $3\beta \geq 2\alpha$ .

Let  $\alpha = \frac{(n-2)(\delta-1)}{2}$  and  $\beta = n-3$ . If  $\delta = 2$ , then for any  $n \geq 4$ ,  $3\beta \geq 2\alpha$  is satisfied. Then since  $f(r; \alpha, \beta)$  is increasing on  $[0, 1]$ ,

$$\begin{aligned} & \int_0^1 f(r; \alpha, \beta) dr \\ & \leq \sqrt{\frac{4}{5}} f\left(\sqrt{\frac{4}{5}}; \alpha, \beta\right) + \left( 1 - \sqrt{\frac{4}{5}} \right) f(1; \alpha, \beta). \end{aligned} \quad (160)$$

If  $\delta = 3$ , then for any  $n \geq 5$ ,  $3\beta \geq 2\alpha$  is satisfied and hence (160) holds. For  $n = 4$ ,  $3\beta \leq 2\alpha$  is satisfied and by (159),

$$\int_0^1 f(r; \alpha, \beta) dr \leq f\left(\sqrt{\frac{4\beta}{2\alpha+\beta}}; \alpha, \beta\right) = f\left(\sqrt{\frac{4}{5}}; \alpha, \beta\right). \quad (161)$$

If  $\delta \geq 4$ , it is easy to see for any  $n \geq 4$ ,  $3\beta \leq 2\alpha$  holds. By (159)

$$\begin{aligned} & \int_0^1 f(r; \alpha, \beta) dr \\ & \leq f\left(\sqrt{\frac{4\beta}{2\alpha+\beta}}; \alpha, \beta\right) \\ & = \left( \frac{\delta-1}{\delta} \right)^{\frac{(n-2)(\delta-1)}{2}} \left( \frac{4}{\delta} \right)^{\frac{n-3}{2}} \left( \frac{n-2}{n-2-\frac{1}{\delta}} \right)^{\frac{\delta-1}{2}} \times \\ & \quad \left( \left( \frac{n-2}{n-2-\frac{1}{\delta}} \right)^{\delta-1} \left( \frac{n-3}{n-2-\frac{1}{m}} \right) \right)^{\frac{n-3}{2}} \\ & \leq \exp\left(\frac{1}{4}\right) \left( \frac{\delta-1}{\delta} \right)^{\frac{(n-2)(\delta-1)}{2}} \left( \frac{4}{\delta} \right)^{\frac{n-3}{2}}, \end{aligned} \quad (162)$$

where the last step follows from  $\left(\frac{n-2}{n-2-\frac{1}{\delta}}\right)^{\delta-1} \left(\frac{n-3}{n-2-\frac{1}{\delta}}\right) \leq 1$  and  $\left(\frac{n-2}{n-2-\frac{1}{\delta}}\right)^{\frac{\delta-1}{2}} \leq \exp\left(\frac{1}{4}\right)$ .

Then (160), (161), (162) and the fact that  $f\left(\sqrt{\frac{4}{5}}; \alpha, \beta\right) = \left(\frac{4}{5}\right)^{\frac{(n-2)\delta-1}{2}}$  yield the conclusion.  $\square$

*Proof of Lemma V.2:* (a) Notice that

$$d_{TV}(\zeta_{n,\delta}, \text{Dirac}(1)) = \frac{1}{2} \sum_{\ell=1}^{\delta+1} |\zeta_{n,\delta}(\ell) - \text{Dirac}(1)(\ell)|$$

$$= \sum_{\ell=2}^{\delta+1} \zeta_{n,\delta}(\ell) \quad (163)$$

$$= \frac{\sum_{\ell=2}^{\delta+1} (\alpha_\ell / \ell)}{\alpha_1 + \sum_{\ell=2}^{\delta+1} (\alpha_\ell / \ell)}$$

$$\leq \frac{\sum_{\ell=2}^{\delta+1} \alpha_\ell}{\alpha_1 + \sum_{\ell=2}^{\delta+1} \alpha_\ell}$$

$$= \sum_{\ell=2}^{\delta+1} \alpha_\ell. \quad (164)$$

(b) It follows from that

$$\left| \lambda_{n,\delta}(e_{n,\delta}) - \frac{1}{\delta!} (e_{n,\delta})^\delta \right| = \frac{1}{\delta!} (e_{n,\delta})^\delta \left| \sum_{\ell=1}^{\delta+1} (\alpha_\ell / \ell) - 1 \right|$$

$$\leq \frac{1}{\delta!} (e_{n,\delta})^\delta \frac{3}{2} \sum_{\ell=2}^{\delta+1} \alpha_\ell.$$

(c) It follows directly by part (a), part (b) and Lemma A.25.  $\square$

2) *Proof of Lemma V.4:*

*Proof of Lemma V.4:* (a) Denote

$$I := \mathbb{P}(\text{NUV}(\{\mathbf{u}'_i\}_{i=1}^{\delta+1}, r; n-2) \geq 2 | \deg(\mathbf{u}'_{\delta+1}) = \delta).$$

Then by the union bound

$$I = \mathbb{P}\left(\bigcup_{i=1}^{\delta} \{\deg(\mathbf{u}'_i) = \delta\} | \deg(\mathbf{u}'_{\delta+1}) = \delta\right)$$

$$\leq \delta \mathbb{P}(\deg(\mathbf{u}'_1) = \delta | \deg(\mathbf{u}'_{\delta+1}) = \delta)$$

$$= \delta \mathbb{P}(\deg(\mathbf{u}'_1) = \delta, \deg(\mathbf{u}'_{\delta+1}) = \delta) / \mathbb{P}(\deg(\mathbf{u}'_{\delta+1}) = \delta). \quad (165)$$

Notice that

$$\mathbb{P}(\deg(\mathbf{u}'_{\delta+1}) = \delta) = \mathbb{E} \prod_{i=1}^{\delta} \mathbb{P}(\mathbf{u}'_i \in \text{SC}(r, \mathbf{u}'_{\delta+1}) | \mathbf{u}'_{\delta+1})$$

$$= (P_n(r))^\delta, \quad (166)$$

where  $\text{SC}(r, \mathbf{u}'_{\delta+1})$  and  $P_n(r)$  are defined in (25) and the paragraph after (25). Moreover

$$\mathbb{P}(\deg(\mathbf{u}'_1) = \delta, \deg(\mathbf{u}'_{\delta+1}) = \delta)$$

$$= \mathbb{E} \mathbb{P}(\deg(\mathbf{u}'_1) = \delta, \deg(\mathbf{u}'_{\delta+1}) = \delta | \mathbf{u}'_1, \mathbf{u}'_{\delta+1})$$

$$= \mathbb{E} 1(\|\mathbf{u}'_1 - \mathbf{u}'_{\delta+1}\|_2 \leq r) \times$$

$$\prod_{i=2}^{\delta} \mathbb{P}(\mathbf{u}'_i \in \text{SC}(r, \mathbf{u}'_1) \cap \text{SC}(r, \mathbf{u}'_{\delta+1}) | \mathbf{u}'_1, \mathbf{u}'_{\delta+1})$$

$$\leq \mathbb{E} 1(\|\mathbf{u}'_1 - \mathbf{u}'_{\delta+1}\|_2 \leq r) (P_n(h(r, \|\mathbf{u}'_1 - \mathbf{u}'_{\delta+1}\|_2)))^{\delta-1} \quad (167)$$

where the last inequality follows from Lemma A.21 with

$$h(r, d) = \sqrt{2 - \frac{2 - r^2}{\sqrt{1 - (d/2)^2}}}.$$

Observing the random quantity in the expectation of (167) only depends the distance between  $\|\mathbf{u}'_1 - \mathbf{u}'_{\delta+1}\|_2$ , replace  $\mathbf{u}'_{\delta+1}$  with  $\mathbf{v}_0 = (1, 0, \dots, 0)$  will not change its value. Then

$$\mathbb{P}(\deg(\mathbf{u}'_1) = \delta, \deg(\mathbf{u}'_{\delta+1}) = \delta)$$

$$\leq \mathbb{E} 1(\|\mathbf{u}'_1 - \mathbf{v}_0\|_2 \leq r) (P_n(h(r, \|\mathbf{u}'_1 - \mathbf{v}_0\|_2)))^{\delta-1}. \quad (168)$$

Use the following coordinate system for each  $\mathbf{u}'_1 = (u_{j1} : 1 \leq j \leq n-1)^\top$  in the region  $\text{SC}(r, \mathbf{v}_0)$ :

$$\begin{cases} u_{11} = 1 - \frac{r^2 r_1^2}{2}, \\ u_{21} = r_1 r \sqrt{1 - \frac{r^2 r_1^2}{4}} \cos(\theta_2), \\ \vdots \\ u_{j1} = r_1 r \sqrt{1 - \frac{r^2 r_1^2}{4}} \cos(\theta_j) \prod_{m=2}^{j-1} \sin(\theta_m), \\ \vdots \\ u_{(n-2)1} = r_1 r \sqrt{1 - \frac{r^2 r_1^2}{4}} \sin(\theta_2) \cdots \sin(\theta_{n-3}) \cos(\theta_{n-2}), \\ u_{(n-1)1} = r_1 r \sqrt{1 - \frac{r^2 r_1^2}{4}} \sin(\theta_2) \cdots \sin(\theta_{n-3}) \sin(\theta_{n-2}), \end{cases}$$

where

$$r_1 \in [0, 1], \theta_j \in [0, \pi] \text{ for } 2 \leq j \leq n-3 \text{ and } \theta_{n-2} \in [0, 2\pi]. \quad (169)$$

Then the right hand side of (168) become

$$\mathbb{E} 1(\|\mathbf{u}'_1 - \mathbf{v}_0\|_2 \leq r) (P_n(h(r, \|\mathbf{u}'_1 - \mathbf{v}_0\|_2)))^{\delta-1}$$

$$= \frac{1}{\text{Area}(S^{n-2})} \int_0^1 (P_n(h(r, r_1 r)))^{\delta-1} r^{n-2} r_1^{n-3} \left(1 - \frac{r^2 r_1^2}{4}\right)^{\frac{n-4}{2}} dr_1 \times$$

$$\prod_{j=2}^{n-3} \int_0^\pi \sin^{n-2-j}(\theta_j) d\theta_j$$

$$= \frac{1}{\int_0^\pi \sin^{n-3}(\theta) d\theta} \int_0^1 (P_n(h(r, r_1 r)))^{\delta-1} r^{n-2} r_1^{n-3} \left(1 - \frac{r^2 r_1^2}{4}\right)^{\frac{n-4}{2}} dr_1$$

$$= \frac{r^{n-2}}{B(\frac{n-2}{2}, \frac{1}{2})} \int_0^1 (P_n(h(r, r_1 r)))^{\delta-1} r_1^{n-3} \left(1 - \frac{r^2 r_1^2}{4}\right)^{\frac{n-4}{2}} dr_1$$

$$\leq \frac{r^{n-2}}{B(\frac{n-2}{2}, \frac{1}{2})} \int_0^1 (P_n(h(r, r_1 r)))^{\delta-1} r_1^{n-3} dr_1. \quad (170)$$

Plug (166), (168) and (170) into (165) and we obtain

$$I \leq \delta \frac{r^{n-2}}{B(\frac{n-2}{2}, \frac{1}{2}) P_n(r)} \int_0^1 \left( \frac{P_n(h(r, r_1 r))}{P_n(r)} \right)^{\delta-1} r_1^{n-3} dr_1$$

$$= \delta(n-2) \frac{a_n r^{n-2}}{P_n(r)} \int_0^1 \left( \frac{P_n(h(r, r_1 r))}{P_n(r)} \right)^{\delta-1} r_1^{n-3} dr_1 \quad (171)$$

where the equality follows from  $a_n = \frac{1}{(n-2)B(\frac{n-2}{2}, \frac{1}{2})}$ . By Lemma A.22 (a),

$$\frac{a_n r^{n-2}}{P_n(r)} \leq \frac{1}{\left(1 - \frac{r^2}{4}\right)^{\frac{n-4}{2}}}. \quad (172)$$

Since when  $0 < r_1 < 1$ ,  $0 < h(r, d)/r < 1$ , by Lemma A.22 (e),

$$\begin{aligned} & \frac{P_n(h(r, r_1 r))}{P_n(r)} \\ & \leq \left(\frac{h(r, r_1 r)}{r}\right)^{n-2} \left(\frac{1 - \frac{h^2(r, r_1 r)}{4}}{1 - \frac{r^2}{4}}\right)^{\frac{n-4}{2}} \\ & \leq \left(1 - \left(\frac{r_1}{2}\right)^2\right)^{\frac{n-2}{2}} \frac{1}{\sqrt{1 - \left(\frac{r_1}{2}\right)^2}} \left(\frac{1 - \frac{h^2(r, r_1 r)}{4}}{(1 - \frac{r^2}{4})\sqrt{1 - \left(\frac{r_1}{2}\right)^2}}\right)^{\frac{n-4}{2}} \end{aligned} \quad (173)$$

where the second inequality follows from

$$\begin{aligned} \left(\frac{h(r, r_1 r)}{r}\right)^2 &= \frac{1}{\sqrt{1 - \left(\frac{r_1}{2}\right)^2}} \left(\frac{-2\left(\frac{r_1}{2}\right)^2}{1 + \sqrt{1 - \left(\frac{r_1}{2}\right)^2}} + 1\right) \\ &\leq \frac{1}{\sqrt{1 - \left(\frac{r_1}{2}\right)^2}} \left(-\left(\frac{r_1}{2}\right)^2 + 1\right). \end{aligned}$$

Since  $h^2(r, r_1 r)$  is decreasing function of  $r_1 \in [0, 1]$ , (173) become

$$\begin{aligned} & \frac{P_n(h(r, r_1 r))}{P_n(r)} \\ & \leq \left(1 - \left(\frac{r_1}{2}\right)^2\right)^{\frac{n-2}{2}} \frac{1}{\sqrt{1 - \left(\frac{r_1}{2}\right)^2}} \left(\frac{1 - \frac{h^2(r, r)}{4}}{(1 - \frac{r^2}{4})\sqrt{1 - \left(\frac{r_1}{2}\right)^2}}\right)^{\frac{n-4}{2}} \\ & \leq \left(1 - \left(\frac{r_1}{2}\right)^2\right)^{\frac{n-2}{2}} \frac{1}{\sqrt{1 - \left(\frac{r_1}{2}\right)^2}} \left(\frac{1}{(1 - \frac{r^2}{4})}\right)^{\frac{n-4}{2}}, \end{aligned} \quad (174)$$

where the second inequality follows from

$$1 - \frac{h^2(r, r)}{4} \leq \sqrt{1 - \frac{r^2}{4}}.$$

Plugging (172) and (174) into (171),

$$\begin{aligned} & I \\ & \leq \delta(n-2) \frac{1}{\left(1 - \frac{r^2}{4}\right)^{\frac{n+\delta-5}{2}}} \left(\frac{1}{(1 - \frac{r^2}{4})}\right)^{\frac{(n-4)(\delta-1)}{2}} \times \\ & \quad \int_0^1 \left(1 - \left(\frac{r_1}{2}\right)^2\right)^{\frac{(n-2)(\delta-1)}{2}} r_1^{n-3} dr_1 \\ & = \bar{h}\left(\frac{1}{\sqrt{1 - r^2/4}}, n, \delta\right) \delta(n-2) \times \\ & \quad \int_0^1 \left(1 - \left(\frac{r_1}{2}\right)^2\right)^{\frac{(n-2)(\delta-1)}{2}} r_1^{n-3} dr_1. \end{aligned}$$

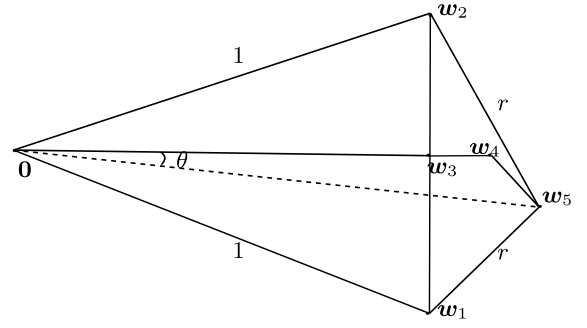


Fig. 6.  $\mathbf{0}$  is the origin in  $\mathbb{R}^{n-2}$  and  $\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_4, \mathbf{w}_5$  are on  $S^{n-2}$ .  $\mathbf{w}_3$  is the midpoint of  $z_1$  and  $z_2$ , while  $z_4$  is the midpoint of the shortest arc on  $S^{n-2}$  connecting  $\mathbf{w}_1$  and  $\mathbf{w}_2$ .  $\mathbf{w}_5$  is one of the two intersection points of the boundary  $\text{SC}(r, \mathbf{w}_1)$  and the boundary of  $\text{SC}(r, \mathbf{w}_2)$ . The angle between line segment  $\mathbf{0}\mathbf{w}_4$  and  $\mathbf{0}\mathbf{w}_5$  is  $\theta$ .

(b) Since  $\frac{1}{\sqrt{1-r^2/4}}$  is decreasing and  $\bar{h}(x, n, \delta)$  as a function of  $x$  is increasing,

$$\begin{aligned} & \bar{h}\left(\frac{1}{\sqrt{1-r^2/4}}, n, \delta\right) \\ & \leq \begin{cases} \bar{h}\left(\left(\frac{5}{4}\right)^{\frac{1}{4}}, n, \delta\right), & \delta = 2, 3 \\ \bar{h}\left(\left(\frac{\delta}{\delta-1}\right)^{\frac{1}{4}}, n, \delta\right), & \delta \geq 4 \end{cases} \\ & = \begin{cases} \left(\sqrt{\frac{5}{4}}\right)^{\frac{n+\delta-5}{2}} \left(\sqrt{\frac{5}{4}}\right)^{\frac{(n-2)(\delta-1)}{2}}, & \delta = 2, 3 \\ \left(\sqrt{\frac{\delta}{\delta-1}}\right)^{\frac{n+\delta-5}{2}} \left(\sqrt{\frac{\delta}{\delta-1}}\right)^{\frac{(n-2)(\delta-1)}{2}}, & \delta \geq 4. \end{cases} \end{aligned} \quad (175)$$

Then the proof is complete by combining part (a), Lemma V.1 (b) and (175).

(c) Similar to (164), we have

$$\begin{aligned} & d_{\text{TV}}(\zeta_{n, \delta, \rho}, \text{Dirac}(1)) \\ & \leq \sum_{\ell=2}^{\delta+1} \alpha(\ell, r_\rho) \\ & = \mathbb{P}(\mathbf{NUV}(\{\mathbf{u}'_i\}_{i=1}^{\delta+1}, r_\rho; n-2) \geq 2 | \deg(\mathbf{u}'_{\delta+1}) = \delta). \end{aligned}$$

where the equality follows from (38). Then the conclusion follows from part (a) and part (b) since  $r_\rho$  satisfies the condition there.  $\square$

**Lemma A.21:** Let  $n \geq 3$  and  $0 < r < \sqrt{2}$ . If  $\mathbf{w}_1$  and  $\mathbf{w}_2$  are two points in  $S^{n-2}$  with  $\|\mathbf{w}_1 - \mathbf{w}_2\|_2 = d$  satisfying  $2 - 2\sqrt{1 - (d/2)^2} < r^2$ , then

$$\mathbb{P}(\mathbf{u}'_1 \in \text{SC}(r, \mathbf{w}_1) \cap \text{SC}(r, \mathbf{w}_2)) \leq P_n(h(r, d))$$

where  $\mathbf{u}'_1$  has distribution  $\text{unif}(S^{n-2})$  and

$$h(r, d) = \sqrt{2 - \frac{2 - r^2}{\sqrt{1 - (d/2)^2}}}.$$

*Proof:* The proof is based on Figure 6 and we use  $|\cdot|$  to represent the length of a line segment in this proof. In the right triangle  $\mathbf{0}\mathbf{w}_3\mathbf{w}_1$ , the line segment  $\mathbf{0}\mathbf{w}_3$  has length  $|\mathbf{0}\mathbf{w}_3| = \sqrt{1 - (d/2)^2}$ . In the right triangle  $\mathbf{w}_1\mathbf{w}_3\mathbf{w}_5$ ,

$|w_3 w_5| = \sqrt{r^2 - (d/2)^2}$ . In the triangle  $0w_3w_5$ , by the law of Cosines,

$$\cos(\theta) = \frac{2 - r^2}{2\sqrt{1 - (\frac{d}{2})^2}}.$$

Then in the isosceles triangle, the line segment  $w_4w_5$  has length

$$|w_4w_5| = 2\sin(\theta/2) = \sqrt{2(1 - \cos(\theta))} = \sqrt{2 - \frac{2 - r^2}{\sqrt{1 - (\frac{d}{2})^2}}} = h(r, d).$$

It is easy to deduce that  $|w_1w_4| = \sqrt{2 - 2\sqrt{1 - (d/2)^2}}$ . The condition  $2 - 2\sqrt{1 - (d/2)^2} < r^2$  entails that  $SC(r, w_1) \cap SC(r, w_2) \neq \emptyset$  and that  $|w_1w_4| < |w_4w_5| = h(r, d)$ . In this case  $SC(r, w_1) \cap SC(r, w_2) \subset SC(h(r, d), w_4)$ . Thus

$$\mathbb{P}(u'_1 \in SC(r, w_1) \cap SC(r, w_2)) \leq \mathbb{P}(u'_1 \in SC(h(r, d), w_4)) = P_n(h(r, d)).$$

□

### 3) Proof of Lemma V.9:

*Proof of Lemma V.9:* When  $\delta = 2$ ,  $\alpha_2 = 0$  since either both vertices have degree 1 or none. Moreover,

$$\begin{aligned} \alpha_3 &= \mathbb{P}(\|\tilde{u}_1 - \tilde{u}_2\|_2 \leq 1) \\ &= \mathbb{E}\mathbb{P}(\|\tilde{u}_1 - \tilde{u}_2\|_2 \leq 1 | \tilde{u}_1) \\ &\stackrel{(*)}{=} \mathbb{E} \frac{2}{\text{Vol}(B^{n-2})} \frac{\pi^{(n-3)/2}}{\Gamma(\frac{n-3}{2} + 1)} \int_0^{\arccos(\frac{\|\tilde{u}_1\|_2}{2})} \sin^{n-2}(\theta) d\theta \\ &\stackrel{(**)}{=} \frac{2}{\text{Vol}(B^{n-2})} \frac{\pi^{(n-3)/2}}{\Gamma(\frac{n-3}{2} + 1)} \frac{\text{Area}(S^{n-3})}{\text{Vol}(B^{n-2})} \times \\ &\quad \int_0^1 r^{n-3} \int_0^{\arccos(\frac{r}{2})} \sin^{n-2}(\theta) d\theta dr \\ &= \frac{2(n-2)}{B(\frac{n-1}{2}, \frac{1}{2})} \int_0^1 r^{n-3} \int_0^{\arccos(\frac{r}{2})} \sin^{n-2}(\theta) d\theta dr \quad (176) \end{aligned}$$

where step (\*) follows from the Subsection “Volume of a hyperspherical cap” from [41] and  $\text{Vol}(B^{n-2})$  is the volume of  $B^{n-2}$ , step (\*\*) follows by observing the random quantity only depends on  $\tilde{u}_1$  through its Euclidean norm, and in the last step  $B(\cdot, \cdot)$  is the Beta function. By Fubini’s Theorem

$$\begin{aligned} &\int_0^1 r^{n-3} \int_0^{\arccos(\frac{r}{2})} \sin^{n-2}(\theta) d\theta dr \\ &= \int_0^{\frac{\pi}{3}} \int_0^1 r^{n-3} \sin^{n-2}(\theta) dr d\theta + \\ &\quad \int_{\frac{\pi}{3}}^{\frac{\pi}{2}} \int_0^{2\cos(\theta)} r^{n-3} \sin^{n-2}(\theta) dr d\theta \\ &= \frac{3}{2(n-2)} \int_0^{\frac{\pi}{3}} \sin^{n-2}(\theta) d\theta \end{aligned}$$

Plugging the preceding formula into (176),  $\alpha_3 = \frac{3}{2} I_{\frac{3}{4}}(\frac{n-1}{2}, \frac{1}{2})$ , where  $I_x(a, b)$  is the regularized incomplete Beta function.  $\alpha_1 = 1 - \alpha_3$  follows from  $\alpha_2 = 0$ . □

### J. Auxiliary Lemmas

*Lemma A.22:* Let  $P_n(r)$  be defined as in Section III-C. Suppose  $n \geq 4$ .

(a) Recall  $a_n = \frac{b_n}{2(n-2)} = \frac{\Gamma((n-1)/2)}{(n-2)\sqrt{\pi}\Gamma((n-2)/2)} \leq 1$ . Then

$$a_n r^{n-2} \left(1 - \frac{\min\{r^2, 4\}}{4}\right)^{\frac{n-4}{2}} \leq P_n(r) \leq a_n r^{n-2}$$

and

$$-\frac{n-4}{8} a_n r^{n-2} \min\{r^2, 4\} \leq P_n(r) - a_n r^{n-2} \leq 0. \quad (177)$$

(b)  $\lim_{r \rightarrow 0} P_n(r) / (a_n r^{n-2}) = 1$ .

(c) Let  $0 \leq \beta < 1 < \alpha$  and  $0 < r \leq 2$ . Then

$$P_n(\alpha r) - P_n(\beta r) \leq (n-2)P_n(r)\alpha^{n-3}(\alpha - \beta).$$

(d) Consider  $\alpha > 1$  and  $r > 0$ . Then

$$P_n(\alpha r) \leq \alpha^{n-2} P_n(r).$$

(e) Consider  $0 < \beta < 1$  and  $0 < r < 2$ . Then

$$P_n(\beta r) \leq \beta^{n-2} \left(\frac{1 - \frac{\beta^2 r^2}{4}}{1 - \frac{r^2}{4}}\right)^{\frac{n-4}{2}} P_n(r).$$

*Proof:* (a) It is easy to verify

$$P'_n(x) = \begin{cases} \frac{b_n}{2} x^{n-3} (1 - \frac{x^2}{4})^{\frac{n-4}{2}} & x < 2, \\ 0 & x \geq 2. \end{cases} \quad (178)$$

Consider  $r > 0$ . Then

$$\frac{P_n(r)}{a_n r^{n-2}} = \frac{P'_n(\xi)}{(n-2)a_n \xi^{n-3}} = \left(1 - \frac{\xi^2}{4}\right)^{\frac{n-4}{2}} \quad (179)$$

where in the first equality  $\xi \in (0, \min\{r, 2\})$  due to the Cauchy Mean Value Theorem and  $P_n(r) \neq 0$ , and the second equality follows from (178). Equation (179) directly implies

$$\left(1 - \frac{(\min\{r, 2\})^2}{4}\right)^{\frac{n-4}{2}} \leq \frac{P_n(r)}{a_n r^{n-2}} \leq 1.$$

(b) It follows directly by taking limit  $r \rightarrow 0^+$  in (179).

(c) Since  $0 \leq \beta < 1 < \alpha$  and  $0 < r \leq 2$ ,  $P_n(\alpha r) - P_n(\beta r) > 0$  and  $P_n(r) > 0$ . Then

$$\begin{aligned} &\frac{P_n(\alpha r) - P_n(\beta r)}{P_n(r)} \\ &= \frac{(P_n(\alpha r) - P_n(\beta r)) - (P_n(\alpha \cdot 0) - P_n(\beta \cdot 0))}{P_n(r) - P_n(0)} \\ &= \frac{\frac{d}{dr}(P_n(\alpha r) - P_n(\beta r))|_{r=\xi}}{\frac{d}{dr}P_n(r)|_{r=\xi}} \\ &= \frac{\alpha^{n-2} \left(1 - \frac{\alpha^2 \xi^2}{4}\right)^{\frac{n-4}{2}} - \beta^{n-2} \left(1 - \frac{\beta^2 \xi^2}{4}\right)^{\frac{n-4}{2}}}{\left(1 - \frac{\xi^2}{4}\right)^{\frac{n-4}{2}}} \\ &\leq \alpha^{n-2} - \beta^{n-2} \\ &\leq (n-2)\alpha^{n-3}(\alpha - \beta), \end{aligned} \quad (180)$$

where the second equality follows from the Cauchy Mean Value Theorem with  $\xi \in (0, r)$ , the third equality follows from (178) together with the fact that the numerator has to be positive, which imply  $\alpha\xi < 2$ , the first inequality follows from  $0 \leq \beta < 1 < \alpha$ , and the last inequality follows from mean value theorem.

(d) When  $r \geq 2$ ,  $P_n(\alpha r) = P_n(r) = 1$  and the conclusion holds trivially. The case  $0 < r < 2$  follows from (180) with  $\beta = 0$ .

(e) Consider  $0 < \beta < 1$  and  $0 < r < 2$ . Then

$$\frac{P_n(\beta r)}{P_n(r)} = \frac{\beta^{n-2} \left(1 - \frac{\beta^2 \xi^2}{4}\right)^{\frac{n-4}{2}}}{\left(1 - \frac{\xi^2}{4}\right)^{\frac{n-4}{2}}} \leq \frac{\beta^{n-2} \left(1 - \frac{\beta^2 r^2}{4}\right)^{\frac{n-4}{2}}}{\left(1 - \frac{r^2}{4}\right)^{\frac{n-4}{2}}},$$

where the equality follows from Cauchy Mean Value Theorem with  $\xi \in (0, r)$ .  $\square$

**Lemma A.23:** Consider  $Z_1$  and  $Z_2$  be two discrete random variable support on  $[\delta]$ . Then

$$d_W(\mathcal{L}(Z_1), \mathcal{L}(Z_2)) \leq \frac{\delta-1}{2} \sum_{\ell=1}^{\delta} |\mathbb{P}(Z_1 = \ell) - \mathbb{P}(Z_2 = \ell)|.$$

*Proof:* By Remark 2.19 (iii) of Section 2.2 in [42],

$$\begin{aligned} d_W(\mathcal{L}(Z_1), \mathcal{L}(Z_2)) &= \sum_{i=1}^{\delta-1} |\mathbb{P}(Z_1 \leq i) - \mathbb{P}(Z_2 \leq i)| \\ &\leq \sum_{i=1}^{\delta-1} \sum_{j=i}^{\delta} |\mathbb{P}(Z_1 = j) - \mathbb{P}(Z_2 = j)|. \end{aligned}$$

On the other hand, from the above equality,

$$\begin{aligned} d_W(\mathcal{L}(Z_1), \mathcal{L}(Z_2)) &= \sum_{i=1}^{\delta-1} |\mathbb{P}(Z_1 \geq i+1) - \mathbb{P}(Z_2 \geq i+1)| \\ &\leq \sum_{i=1}^{\delta-1} \sum_{j=i+1}^{\delta} |\mathbb{P}(Z_1 = j) - \mathbb{P}(Z_2 = j)|. \end{aligned}$$

Averaging the above two inequalities yields the desired conclusion.  $\square$

**Lemma A.24:** For any integer-valued random variable  $Z_1$  and  $Z_2$ ,

$$d_{TV}(\mathcal{L}(Z_1), \mathcal{L}(Z_2)) \leq \mathbb{P}(Z_1 \neq Z_2) \leq \mathbb{E}|Z_1 - Z_2|.$$

*Proof:*

$$\begin{aligned} &d_{TV}(\mathcal{L}(Z_1), \mathcal{L}(Z_2)) \\ &= \max_{A \text{ Borel measurable}} |\mathbb{P}(Z_1 \in A) - \mathbb{P}(Z_2 \in A)| \\ &= \max_{A \text{ Borel measurable}} |\mathbb{P}(Z_1 \in A, Z_1 \neq Z_2) - \mathbb{P}(Z_2 \in A, Z_1 \neq Z_2)| \\ &\leq \mathbb{P}(Z_1 \neq Z_2) \\ &= \mathbb{P}(|Z_1 - Z_2| \geq 1) \\ &\leq \mathbb{E}|Z_1 - Z_2|. \end{aligned}$$

**Lemma A.25:** Consider two compound Poisson distributions  $CP(\lambda_1, \zeta_1)$  and  $CP(\lambda_2, \zeta_2)$ . Then

$$\begin{aligned} &d_{TV}(CP(\lambda_1, \zeta_1), CP(\lambda_2, \zeta_2)) \\ &\leq \min\{\lambda_1, \lambda_2\} d_{TV}(\zeta_1, \zeta_2) + d_{TV}(\text{Pois}(\lambda_1), \text{Pois}(\lambda_2)) \\ &\leq \min\{\lambda_1, \lambda_2\} d_{TV}(\zeta_1, \zeta_2) + \\ &\quad \min\left\{|\lambda_1 - \lambda_2|, \sqrt{\frac{2}{e}} |\sqrt{\lambda_1} - \sqrt{\lambda_2}|\right\}. \end{aligned}$$

*Proof:* By triangular inequality,

$$\begin{aligned} &d_{TV}(CP(\lambda_1, \zeta_1), CP(\lambda_2, \zeta_2)) \\ &\leq d_{TV}(CP(\lambda_1, \zeta_1), CP(\lambda_1, \zeta_2)) + \\ &\quad d_{TV}(CP(\lambda_1, \zeta_2), CP(\lambda_2, \zeta_2)). \end{aligned} \quad (181)$$

We will bound the two terms in the upper bound separately.

**Step 1:**  $d_{TV}(CP(\lambda_1, \zeta_1), CP(\lambda_1, \zeta_2))$

Consider  $Z_1 = \sum_{i=1}^N Y_i$  and  $Z_2 = \sum_{i=1}^N Y'_i$ , where  $N \sim \text{Pois}(\lambda_1)$ ,  $\{Y_i\} \stackrel{\text{i.i.d.}}{\sim} \zeta_1$ ,  $\{Y'_i\} \stackrel{\text{i.i.d.}}{\sim} \zeta_2$ , and  $N$  is independent of  $\{Y_i\}, \{Y'_i\}$ . Then

$$\begin{aligned} d_{TV}(CP(\lambda_1, \zeta_1), CP(\lambda_1, \zeta_2)) &= d_{TV}(\mathcal{L}(Z_1), \mathcal{L}(Z_2)) \\ &\leq 1 - \mathbb{P}(Z_1 = Z_2) \\ &= 1 - \mathbb{E}\mathbb{P}(Z_1 = Z_2 | N) \\ &\leq 1 - \mathbb{E}(\mathbb{P}(Y_1 = Y'_1))^N, \end{aligned}$$

where the first inequality follows from Lemma A.24. Since the above inequality holds for any coupling  $(Y_1, Y'_1)$  of  $\zeta_1$  and  $\zeta_2$ , by Proposition 4.7 in [43] taking the infimum of all the coupling then yields

$$\begin{aligned} d_{TV}(CP(\lambda_1, \zeta_1), CP(\lambda_1, \zeta_2)) &\leq 1 - \mathbb{E}(1 - d_{TV}(\zeta_1, \zeta_2))^N \\ &\leq \lambda_1 d_{TV}(\zeta_1, \zeta_2). \end{aligned}$$

**Step 2:**  $d_{TV}(CP(\lambda_1, \zeta_2), CP(\lambda_2, \zeta_2))$

Consider  $Z_1 = \sum_{i=1}^{N_1} Y'_i$  and  $Z_2 = \sum_{i=1}^N Y'_i$ , where  $N_1 \sim \text{Pois}(\lambda_1)$ ,  $N_2 \sim \text{Pois}(\lambda_2)$ ,  $\{Y'_i\} \stackrel{\text{i.i.d.}}{\sim} \zeta_2$ , and  $\{Y'_i\}$  is independent of  $N_1, N_2$ . Then

$$\begin{aligned} d_{TV}(CP(\lambda_1, \zeta_2), CP(\lambda_2, \zeta_2)) &= d_{TV}(\mathcal{L}(Z_1), \mathcal{L}(Z_2)) \\ &\leq \mathbb{P}(Z_1 \neq Z_2) \\ &= \mathbb{P}(N_1 \neq N_2), \end{aligned}$$

where the first inequality follows from Lemma A.24. Since the above inequality holds for any coupling  $(N_1, N_2)$  of  $\text{Pois}(\lambda_1)$  and  $\text{Pois}(\lambda_2)$ , by Proposition 4.7 in [43] taking the infimum of all the coupling then yields  $d_{TV}(CP(\lambda_1, \zeta_2), CP(\lambda_2, \zeta_2)) \leq d_{TV}(\text{Pois}(\lambda_1), \text{Pois}(\lambda_2))$ .

Plugging step 1 and step 2 into (181),

$$\begin{aligned} &d_{TV}(CP(\lambda_1, \zeta_1), CP(\lambda_2, \zeta_2)) \\ &\leq \lambda_1 d_{TV}(\zeta_1, \zeta_2) + d_{TV}(\text{Pois}(\lambda_1), \text{Pois}(\lambda_2)). \end{aligned}$$

By symmetry property, the first term in the above upper bound can be replace by  $\min\{\lambda_1, \lambda_2\} d_{TV}(\zeta_1, \zeta_2)$ . The proof is then completed by applying equation (2.2) of [44] to bound  $d_{TV}(\text{Pois}(\lambda_1), \text{Pois}(\lambda_2))$ .  $\square$

**Lemma A.26:** (a) Let  $p, p', m$  be positive integers such that  $p \geq p'$ . Then

$$\prod_{i=0}^m (p-i) - \prod_{i=0}^m (p'-i) \leq (m+1) \left( \prod_{i=0}^{m-1} (p-i) \right) (p-p').$$

(b) Let  $p, \delta, \kappa$  be positive integers such that  $\delta \leq p-1$ . Then

$$\prod_{\ell=1}^{\delta} (p-\ell) - \prod_{\ell=1}^{\delta} (p-\ell\kappa) \leq \frac{\delta(\delta+1)}{2} (\kappa-1) \prod_{\ell=1}^{\delta-1} (p-\ell).$$

(c)  $\left(\frac{1+x}{1-x}\right)^2$  is increasing function on  $[0, \frac{1}{2}]$  and  $\left(\frac{1+x}{1-x}\right)^2 \leq 1 + 16x$  for  $0 \leq x \leq \frac{1}{2}$ . Then

(d)  $1 + x - \frac{1}{1+x} \leq 2x$  for any  $x \geq 0$ .

*Proof:* (c) and (d) are simple quadratic inequalities and hence their proofs are omitted. (a) Let  $f(x) = \prod_{i=0}^m (x-i)$ . When  $p' \geq m$ ,  $f'(x) \leq (m+1) \prod_{i=0}^{m-1} (p-i)$  and the conclusion then follows by the mean value theorem. When  $p' \leq m-1$ ,

$$f(p) - f(p') \leq f(p) \leq (p-p') \prod_{i=0}^{m-1} (p-i).$$

(b) Let  $f(x) = \prod_{\ell=1}^{\delta} (p-\ell x)$ . When  $p < \delta\kappa$ ,

$$\begin{aligned} f(1) - f(\kappa) &\leq f(1) \leq (\delta\kappa - \delta) \prod_{\ell=1}^{\delta-1} (p-\ell) \\ &\leq \frac{\delta(\delta+1)}{2} (\kappa-1) \prod_{\ell=1}^{\delta-1} (p-\ell). \end{aligned}$$

When  $p \geq \delta\kappa$ ,  $f'(x) \geq -\frac{\delta(\delta+1)}{2} \prod_{\ell=1}^{\delta-1} (p-\ell)$  for  $x \in [1, \kappa]$ . Then the conclusion follows by the mean value theorem.  $\square$

**Lemma A.27 (Perturbation Theory):** Consider  $\mathbf{D} \in \mathbb{S}^n$  and  $\mathbf{E} \in \mathbb{S}^n$ , where  $\mathbb{S}^n$  is the set of all real symmetric matrices of dimension  $n \times n$ . Let  $\{\lambda_i(\cdot)\}_{i=1}^n$  be the eigenvalues of corresponding matrix such that  $\lambda_1(\cdot) \geq \lambda_2(\cdot) \geq \dots \geq \lambda_n(\cdot)$ .

(a)

$$|\lambda_i(\mathbf{D} + \mathbf{E}) - \lambda_i(\mathbf{D})| \leq \|\mathbf{E}\|_2 \quad (i = 1, 2, \dots, n)$$

(b) Assume  $\mathbf{E} = \omega \mathbf{x} \mathbf{x}^\top$ , where  $\mathbf{x} \in S^{n-1}$ . If  $\omega \geq 0$ , then

$$\lambda_i(\mathbf{D} + \mathbf{E}) \in [\lambda_i(\mathbf{D}), \lambda_{i-1}(\mathbf{D})], \quad (i = 2, 3, \dots, n),$$

while if  $\omega \leq 0$ , then

$$\lambda_i(\mathbf{D} + \mathbf{E}) \in [\lambda_{i+1}(\mathbf{D}), \lambda_i(\mathbf{D})], \quad (i = 1, 2, \dots, n-1).$$

In either case, there exist nonnegative  $m_1, m_2, \dots, m_n$  such that

$$\lambda_i(\mathbf{D} + \mathbf{E}) = \lambda_i(\mathbf{D}) + m_i \omega, \quad (i = 1, 2, \dots, n)$$

with  $m_1 + m_2 + \dots + m_n = 1$ .

(c) Assume  $\mathbf{E} = \sum_{i=1}^m \omega_i \mathbf{x}_i \mathbf{x}_i^\top$ , where  $\{\mathbf{x}_i\}_{i=1}^m \subset S^{n-1}$  and  $\omega_i \geq 0$  for all  $i$ . Then

$$\lambda_n(\mathbf{D} + \mathbf{E}) \geq \lambda_n(\mathbf{D}).$$

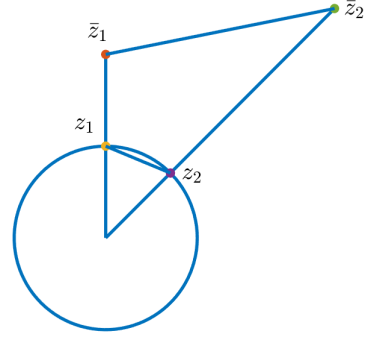


Fig. 7.  $\mathbf{w}_1$  and  $\mathbf{w}_2$  are the normalized vector of  $\bar{\mathbf{w}}_1$  and  $\bar{\mathbf{w}}_2$  respectively.

*Proof:* (a) and (b) is Corollary 8.1.6 and Theorem 8.1.8 in [36]. (c) follows by induction on the smallest eigenvalue using part (b) for  $\omega \geq 0$ .  $\square$

**Lemma A.28:** Let  $\mathbf{x}_1, \mathbf{x}_2$  be two vectors on  $S^{n-1}$ , and  $\mathbf{D} \in \mathbb{R}^{n \times n}$  be an invertible matrix. Let  $S_{\max}(\mathbf{D})$  and  $S_{\min}(\mathbf{D})$  be respectively the largest and smallest singular value of  $\mathbf{D}$ . Define  $\bar{\mathbf{w}}_i = \mathbf{D} \mathbf{x}_i$  and  $\mathbf{w}_i = \bar{\mathbf{w}}_i / \|\bar{\mathbf{w}}_i\|_2$ , ( $i = 1, 2$ ). Then,

$$\frac{S_{\min}(\mathbf{D})}{S_{\max}(\mathbf{D})} \|\mathbf{x}_1 - \mathbf{x}_2\|_2 \leq \|\mathbf{w}_1 - \mathbf{w}_2\|_2 \leq \frac{S_{\max}(\mathbf{D})}{S_{\min}(\mathbf{D})} \|\mathbf{x}_1 - \mathbf{x}_2\|_2$$

*Proof: Part I (Upper Bound)*

Denote  $\angle(\cdot, \cdot)$  the angle between two vectors. By the Law of Cosines,

$$\begin{aligned} \cos(\angle(\mathbf{w}_1, \mathbf{w}_2)) &= \frac{\|\mathbf{w}_1\|_2^2 + \|\mathbf{w}_2\|_2^2 - \|\mathbf{w}_1 - \mathbf{w}_2\|_2^2}{2 \times \|\mathbf{w}_1\|_2 \times \|\mathbf{w}_2\|_2} \\ &= \frac{2 - \|\mathbf{w}_1 - \mathbf{w}_2\|_2^2}{2}, \end{aligned}$$

and

$$\cos(\angle(\bar{\mathbf{w}}_1, \bar{\mathbf{w}}_2)) = \frac{\|\bar{\mathbf{w}}_1\|_2^2 + \|\bar{\mathbf{w}}_2\|_2^2 - \|\bar{\mathbf{w}}_1 - \bar{\mathbf{w}}_2\|_2^2}{2 \times \|\bar{\mathbf{w}}_1\|_2 \times \|\bar{\mathbf{w}}_2\|_2}.$$

Observing  $\angle(\mathbf{w}_1, \mathbf{w}_2) = \angle(\bar{\mathbf{w}}_1, \bar{\mathbf{w}}_2)$ , the right hand sides of the above two equations are equal. Solving for  $\|\mathbf{w}_1 - \mathbf{w}_2\|_2$ , we get

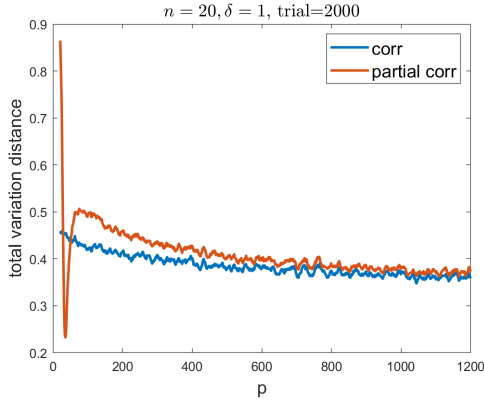
$$\begin{aligned} \|\mathbf{w}_1 - \mathbf{w}_2\|_2^2 &= \frac{\|\bar{\mathbf{w}}_1 - \bar{\mathbf{w}}_2\|_2^2}{\|\bar{\mathbf{w}}_1\|_2 \|\bar{\mathbf{w}}_2\|_2} + \left( 2 - \frac{\|\bar{\mathbf{w}}_2\|_2}{\|\bar{\mathbf{w}}_1\|_2} - \frac{\|\bar{\mathbf{w}}_1\|_2}{\|\bar{\mathbf{w}}_2\|_2} \right) \\ &\leq \frac{\|\bar{\mathbf{w}}_1 - \bar{\mathbf{w}}_2\|_2^2}{\|\bar{\mathbf{w}}_1\|_2 \|\bar{\mathbf{w}}_2\|_2}. \end{aligned}$$

Therefore,

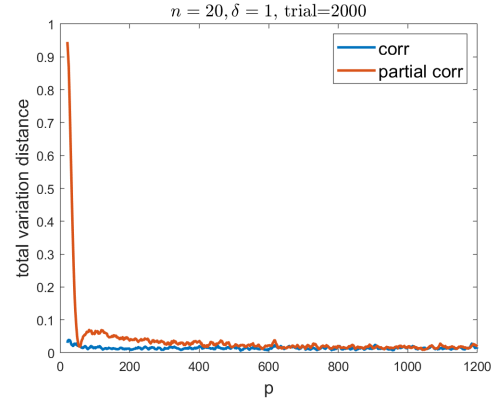
$$\begin{aligned} \|\mathbf{w}_1 - \mathbf{w}_2\|_2 &\leq \frac{\|\bar{\mathbf{w}}_1 - \bar{\mathbf{w}}_2\|_2}{\sqrt{\|\bar{\mathbf{w}}_1\|_2 \|\bar{\mathbf{w}}_2\|_2}} \\ &\leq \frac{S_{\max}(\mathbf{D}) \|\mathbf{x}_1 - \mathbf{x}_2\|_2}{\sqrt{S_{\min}(\mathbf{D}) \|\mathbf{x}_1\|_2 S_{\min}(\mathbf{D}) \|\mathbf{x}_2\|_2}} \\ &= \frac{S_{\max}(\mathbf{D})}{S_{\min}(\mathbf{D})} \|\mathbf{x}_1 - \mathbf{x}_2\|_2. \end{aligned}$$

**Part II (Lower Bound)**

Define  $\bar{\mathbf{x}}_i = \mathbf{D}^{-1} \mathbf{w}_i$ , ( $i = 1, 2$ ). Note for  $\forall i \in \{1, 2\}$ ,  $\mathbf{x}_i$  and  $\bar{\mathbf{x}}_i$  are parallel to each other, since  $\mathbf{x}_i = \mathbf{D}^{-1} \bar{\mathbf{w}}_i$  and  $\bar{\mathbf{w}}_i$  is



(a) limiting compound Poisson



(b) finite p compound Poisson approximation

Fig. 8. The vertical axis of (a) is  $d_{TV}(\mathcal{L}(N_{V_1}^{(\Psi)}), \text{CP}(\lambda_{20,1}(1), \zeta_{20,1}))$  and that of (b) is  $d_{TV}(\mathcal{L}(N_{V_1}^{(\Psi)}), \text{CP}(\lambda_{p,20,1,\rho}, \zeta_{20,1,\rho}))$ . The theoretical convergence results have respectively been established in Theorem II.4 and Theorem III.11. For both plots the samples are independently generated according to  $\mathcal{N}(\mathbf{0}, \Sigma)$  with  $\Sigma$  being a  $(\tau = p^{0.6}, \kappa = p^{0.8})$  sparse matrix for each  $p$ . The parameters are  $n = 20$ ,  $\delta = 1$  and the threshold  $\rho$  is chosen according to (11) with  $e_{n,\delta} = 1$ . The blue curve is for the empirical correlation graph ( $\Psi = \mathbf{R}$ ) and the red curve is for the empirical partial correlation graph ( $\Psi = \mathbf{P}$ ). Note since  $\delta = 1$ ,  $\zeta_{20,1} = \text{Dirac}(2)$  is  $\zeta_{20,1,\rho}$ , by Example V.7. As demonstrated by the plots, for both the empirical correlation and partial correlation graphs, the total variations in (a) decrease very slowly while the total variations in (b) converge to 0 very fast, which has also been analytically discussed in Remark III.17. Our observation that the non-asymptotic compound Poisson distribution provides a better fit to the numerical simulations for small  $p$  is a caveat to practitioners who may be tempted to use the Poisson approximation.

parallel to  $\mathbf{w}_i$ . Thus, we conclude  $\mathbf{x}_i = \bar{\mathbf{x}}_i / \|\bar{\mathbf{x}}_i\|_2$ , ( $i = 1, 2$ ). Reversing the role of  $\mathbf{x}_i$  and  $\mathbf{w}_i$  in Part I, one has

$$\|\mathbf{x}_1 - \mathbf{x}_2\|_2 \leq \frac{S_{\max}(\mathbf{D}^{-1})}{S_{\min}(\mathbf{D}^{-1})} \|\mathbf{w}_1 - \mathbf{w}_2\|_2.$$

The lower bound follows from the relation  $\frac{S_{\max}(\mathbf{D}^{-1})}{S_{\min}(\mathbf{D}^{-1})} = \frac{S_{\max}(\mathbf{D})}{S_{\min}(\mathbf{D})}$ .  $\square$

**Lemma A.29:** Let  $\{\mathcal{D}_i\}_{i=1}^m$ ,  $\{\mathcal{F}_i\}_{i=1}^m$ ,  $\{\mathcal{G}_i\}_{i=1}^m$  and  $\{\mathcal{H}_i\}_{i=1}^m$  be sets satisfying

$$\mathcal{G}_i \subset \mathcal{D}_i \subset \mathcal{H}_i, \quad \mathcal{G}_i \subset \mathcal{F}_i \subset \mathcal{H}_i, \quad (i = 1, 2, \dots, m).$$

Then

(a)

$$\begin{aligned} \left( \bigcap_{i=1}^m \mathcal{D}_i \right) \triangle \left( \bigcap_{i=1}^m \mathcal{F}_i \right) &\subset \bigcup_{i=1}^m (\mathcal{H}_i \setminus \mathcal{G}_i) \cap \left( \bigcap_{j=1}^m \mathcal{H}_j \right) \\ &= \bigcup_{i=1}^m \left( (\mathcal{H}_i \setminus \mathcal{G}_i) \cap \left( \bigcap_{\substack{j=1 \\ j \neq i}}^m \mathcal{H}_j \right) \right). \end{aligned}$$

(b)

$$\left( \bigcup_{i=1}^m \mathcal{D}_i \right) \triangle \left( \bigcup_{i=1}^m \mathcal{F}_i \right) \subset \bigcup_{i=1}^m (\mathcal{H}_i \setminus \mathcal{G}_i).$$

*Proof:* (a) Obviously,

$$\bigcap_{i=1}^m \mathcal{G}_i \subset \bigcap_{i=1}^m \mathcal{D}_i \subset \bigcap_{i=1}^m \mathcal{H}_i, \quad \bigcap_{i=1}^m \mathcal{G}_i \subset \bigcap_{i=1}^m \mathcal{F}_i \subset \bigcap_{i=1}^m \mathcal{H}_i. \quad (182)$$

Thus,

$$\left( \bigcap_{i=1}^m \mathcal{D}_i \right) \triangle \left( \bigcap_{i=1}^m \mathcal{F}_i \right) \subset \left( \bigcap_{i=1}^m \mathcal{H}_i \right) \setminus \left( \bigcap_{i=1}^m \mathcal{G}_i \right).$$

Take  $\forall \omega \in (\bigcap_{i=1}^m \mathcal{H}_i) \setminus (\bigcap_{i=1}^m \mathcal{G}_i)$ , we know  $\omega \in \bigcap_{i=1}^m \mathcal{H}_i$  and  $\omega \notin \bigcap_{i=1}^m \mathcal{G}_i$ . The later fact shows  $\exists j$  (which depends on  $\omega$ ) such that  $\omega \notin \mathcal{G}_j$ . Then,

$$\omega \in \left( \bigcap_{i=1}^m \mathcal{H}_i \right) \setminus \mathcal{G}_j \subset \mathcal{H}_j \setminus \mathcal{G}_j \subset \bigcup_{i=1}^m (\mathcal{H}_i \setminus \mathcal{G}_i). \quad (183)$$

The proof is completed by combining (182) and (183).

(b)

$$\begin{aligned} \left( \bigcup_{i=1}^m \mathcal{D}_i \right) \triangle \left( \bigcup_{i=1}^m \mathcal{F}_i \right) &= \left( \bigcup_{i=1}^m \mathcal{D}_i \right)^c \triangle \left( \bigcup_{i=1}^m \mathcal{F}_i \right)^c \\ &= \left( \bigcap_{i=1}^m \mathcal{D}_i^c \right) \triangle \left( \bigcap_{i=1}^m \mathcal{F}_i^c \right) \\ &\subset \bigcup_{i=1}^m (\mathcal{G}_i^c \setminus \mathcal{H}_i^c) \\ &= \bigcup_{i=1}^m (\mathcal{H}_i \setminus \mathcal{G}_i), \end{aligned}$$

where the inclusion step follows from (a).  $\square$

**Lemma A.30:** Let  $\mathbf{Q} \in \mathbb{R}^{n \times m}$  ( $n \leq m$ ), with each column  $\mathbf{q}_i$  being i.i.d.  $\text{unif}(\sqrt{n}S^{n-1})$ . Let  $\lambda_{\min}$  and  $\lambda_{\max}$  be respectively the largest and smallest eigenvalue of  $\frac{1}{m}\mathbf{Q}\mathbf{Q}^\top$ . Then with probability at least  $1 - 2\exp(-ct^2)$ ,

$$\begin{aligned} \left[ 1 - C \left( \sqrt{\frac{n}{m}} + \frac{t}{\sqrt{m}} \right) \right]^2 &\leq \lambda_{\min} \\ &\leq \lambda_{\max} \leq \left[ 1 + C \left( \sqrt{\frac{n}{m}} + \frac{t}{\sqrt{m}} \right) \right]^2, \end{aligned} \quad (184)$$

where  $c, C$  are absolute constants.

*Proof:* Let  $S_{\max}$ ,  $S_{\min}$  be respectively the largest and smallest singular value of  $\mathbf{Q}$ . Since columns  $\{\mathbf{q}_i\}_{i=1}^m$  are isotropic random vectors with subgaussian norm (or  $\psi_2$  norm) being a constant, by applying Theorem 5.39 in [2] to  $\mathbf{Q}^\top$ ,

$$\sqrt{m} - C(\sqrt{n} + t) \leq S_{\min} \leq S_{\max} \leq \sqrt{m} + C(\sqrt{n} + t), \quad (185)$$

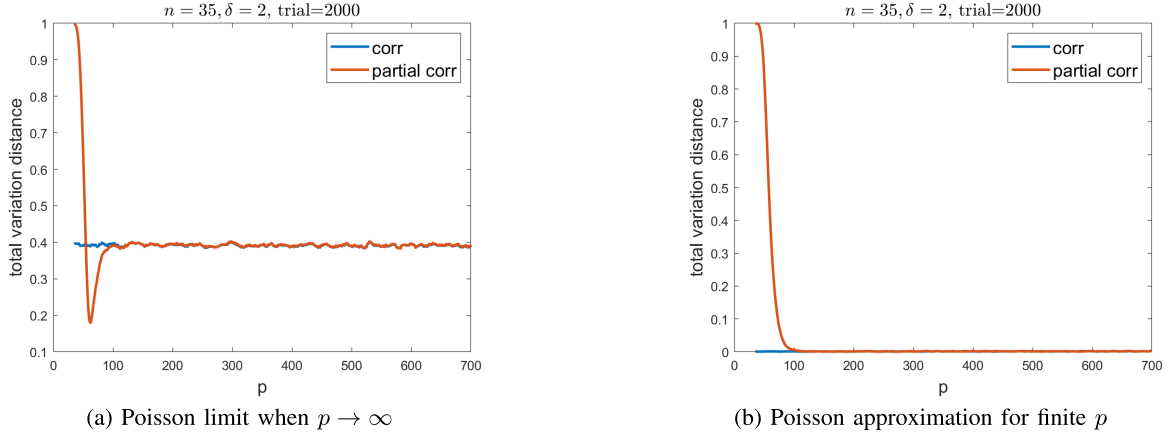


Fig. 9. The vertical axis of (a) is  $d_{TV} \left( \mathcal{L} \left( N_{V_\delta}^{(\Psi)} \right), \text{Pois} \left( \frac{(e_{n,\delta})^\delta}{\delta!} \right) \right)$ , where we replaced  $\text{CP}(\lambda_{n,\delta}(e_{n,\delta}), \zeta_{n,\delta})$  in Theorem II.4 by its approximation  $\text{Pois} \left( \frac{(e_{n,\delta})^\delta}{\delta!} \right)$  as in Proposition V.3. The vertical axis of (b) is  $d_{TV} \left( N_{V_\delta}^{(\Psi)}, \text{Pois} \left( \binom{p}{\delta} (2P_n(r_\rho))^\delta \right) \right)$ , where we replaced  $\text{CP}(\lambda_{p,n,\delta,\rho}, \zeta_{n,\delta,\rho})$  in Theorem III.11 by its approximation  $\text{Pois} \left( \binom{p}{\delta} (2P_n(r_\rho))^\delta \right)$  as in Proposition V.6. For both plots the samples are independently generated according to  $\mathcal{N}(\mathbf{0}, \Sigma)$  with  $\Sigma$  being a  $(\tau = p^{0.6}, \kappa = p^{0.8})$  sparse matrix for each  $p$ . The parameters are  $n = 35$ ,  $\delta = 2$  and the threshold  $\rho$  is chosen according to (11) with  $e_{n,\delta} = 1$ . As demonstrated by the plots, for both the empirical correlation and partial correlation graphs, the total variations in (a) decrease very slowly while the total variations in (b) converge to 0 very fast. The fast convergence in Figure 9 (b) verifies the validity of using Poisson distribution  $\text{Pois} \left( \binom{p}{\delta} (2P_n(r_\rho))^\delta \right)$  to approximate the distribution of random quantities in  $\{N_{E_\delta}^{(\Psi)}, N_{V_\delta}^{(\Psi)}, N_{V_\delta}^{(\Psi)} : \Psi \in \{\mathbf{R}, \mathbf{P}\}\}$  for large  $n$ . We now discuss the slow convergence behavior of Figure 9 (a). Note that  $n = 35$  is large enough to guarantee that  $\sum_{\ell=2}^{\delta+1} \alpha_\ell$  is small as indicated by Figure 4 (b), which implies that  $\text{CP}(\lambda_{n,\delta}(e_{n,\delta}), \zeta_{n,\delta})$  is well approximated by  $\text{Pois} \left( \frac{(e_{n,\delta})^\delta}{\delta!} \right)$  by Lemma V.2 (c). As a result, the extremely slow decrease in Figure 9 (a) is not because of using the Poisson approximation, but is due to the slow convergence of Theorem II.4, which has been extensively discussed in Remark III.17. This specific example additionally indicates that the slow convergence of Theorem II.4 is due to slow convergence of  $\lambda_{p,n,\delta,\rho} \rightarrow \lambda_{n,\delta}$  since the increment distribution in this large  $n$  case are both close to  $\text{Dirac}(1)$ .

holds with probability at least  $1 - 2 \exp(-ct^2)$ , where  $c, C$  are absolute constants. The proof is completed by

$$\lambda_{\max} = \frac{1}{m} S_{\max}^2, \quad \lambda_{\min} = \frac{1}{m} S_{\min}^2.$$

□

### K. Numerical Simulations and Experiments

Figures 8a-9b show comparisons between theory and simulation for the zero mean multivariate Gaussian case with  $n < p$ . In particular, they show empirically the speed of convergence to the compound Poisson limit as  $p$  increases.

### REFERENCES

- [1] A. Hero and B. Rajaratnam, "Large-scale correlation mining for biomolecular network discovery," in *Big Data Over Networks*, S. Cui, A. Hero, Z.-Q. Luo, and J. M. Moura, Eds. Cambridge, U.K.: Cambridge Univ. Press, 2016, pp. 409–435.
- [2] R. Vershynin, "Introduction to the non-asymptotic analysis of random matrices," in *Compressed Sensing: Theory and Practice*, Y. C. Eldar and G. Kutyniok, Eds. Cambridge, U.K.: Cambridge Univ. Press, 2012, pp. 210–268.
- [3] A. P. Dawid and S. L. Lauritzen, "Hyper Markov laws in the statistical analysis of decomposable graphical models," *Ann. Statist.*, vol. 21, no. 3, pp. 1272–1317, Sep. 1993.
- [4] G. Letac and H. Massam, "Wishart distributions for decomposable graphs," *Ann. Statist.*, vol. 35, no. 3, pp. 1278–1323, Jul. 2007.
- [5] A. Anandkumar, L. Tong, and A. Swami, "Detection of Gauss–Markov random fields with nearest-neighbor dependency," *IEEE Trans. Inf. Theory*, vol. 55, no. 2, pp. 816–827, Feb. 2009.
- [6] R. M. Castro, G. Lugosi, and P. A. Savalle, "Detection of correlations with adaptive sensing," *IEEE Trans. Inf. Theory*, vol. 60, no. 12, pp. 7913–7927, Dec. 2014.
- [7] K. M. Tan, P. London, K. Mohan, S.-I. Lee, M. Fazel, and D. Witten, "Learning graphical models with hubs," *J. Mach. Learn. Res.*, vol. 15, no. 95, pp. 3297–3331, 2014.
- [8] K. Khare, S.-Y. Oh, and B. Rajaratnam, "A convex pseudolikelihood framework for high dimensional partial correlation estimation with convergence guarantees," *J. Roy. Stat. Soc., B, Stat. Methodol.*, vol. 77, no. 4, pp. 803–825, Sep. 2015.
- [9] H. Firouzi, A. O. Hero, and B. Rajaratnam, "Two-stage sampling, prediction and adaptive regression via correlation screening," *IEEE Trans. Inf. Theory*, vol. 63, no. 1, pp. 698–714, Jan. 2016.
- [10] O. Dalal and B. Rajaratnam, "Sparse Gaussian graphical model estimation via alternating minimization," *Biometrika*, vol. 104, no. 2, pp. 379–395, Jun. 2017.
- [11] J. Heydari and A. Tajer, "Quickest search for local structures in random graphs," *IEEE Trans. Signal Inf. Process. Netw.*, vol. 3, no. 3, pp. 526–538, Sep. 2017.
- [12] D. A. Tarzanagh and G. Michailidis, "Estimation of graphical models through structured norm minimization," *J. Mach. Learn. Res.*, vol. 18, no. 1, pp. 1–48, 2018.
- [13] R. F. Engle, O. Ledoit, and M. Wolf, "Large dynamic covariance matrices," *J. Bus. Econ. Statist.*, vol. 37, no. 2, pp. 363–375, Apr. 2019.
- [14] X. Cao, K. Khare, and M. Ghosh, "Posterior graph selection and estimation consistency for high-dimensional Bayesian DAG models," *Ann. Statist.*, vol. 47, no. 1, pp. 319–348, Feb. 2019.
- [15] H. Li, A. Aue, D. Paul, J. Peng, and P. Wang, "An adaptable generalization of hotelling's  $T^2$  test in high dimension," *Ann. Statist.*, vol. 48, no. 3, pp. 1815–1847, 2020.
- [16] O. Ledoit and M. Wolf, "Analytical nonlinear shrinkage of large-dimensional covariance matrices," *Ann. Statist.*, vol. 48, no. 5, pp. 3043–3065, Oct. 2020.
- [17] A. Hero and B. Rajaratnam, "Large-scale correlation screening," *J. Amer. Statist. Assoc.*, vol. 106, no. 496, pp. 1540–1552, 2011.
- [18] A. Hero and B. Rajaratnam, "Hub discovery in partial correlation graphs," *IEEE Trans. Inf. Theory*, vol. 58, no. 9, pp. 6064–6078, Sep. 2012.
- [19] A. O. Hero and B. Rajaratnam, "Foundational principles for large-scale inference: Illustrations through correlation mining," *Proc. IEEE*, vol. 104, no. 1, pp. 93–110, Jan. 2016.
- [20] K. Zhang, "Spherical cap packing asymptotics and rank-extreme detection," *IEEE Trans. Inf. Theory*, vol. 63, no. 7, pp. 4572–4584, Jul. 2017.

- [21] E. Arias-Castro, S. Bubeck, and G. Lugosi, "Detection of correlations," *Ann. Statist.*, vol. 40, no. 1, pp. 412–435, Feb. 2012.
- [22] J. Fan, Q.-M. Shao, and W.-X. Zhou, "Are discoveries spurious? Distributions of maximum spurious correlations and their applications," *Ann. Statist.*, vol. 46, no. 3, p. 989, Jun. 2018.
- [23] M. Penrose, *Random Geometric Graphs* (Oxford Studies in Probability), vol. 5. New York, NY, USA: Oxford Univ. Press, 2003.
- [24] D. Chakrabarti and C. Faloutsos, "Graph mining: Laws, generators, and algorithms," *ACM Comput. Surveys*, vol. 38, no. 1, p. 2, Jun. 2006.
- [25] E. D. Kolaczyk and G. Csárdi, *Statistical Analysis of Network Data With R* (Use R!), vol. 65. New York, NY, USA: Springer, 2014.
- [26] E. D. Kolaczyk, *Statistical Analysis of Network Data* (Springer Series in Statistics), vol. 69. New York, NY, USA: Springer, 2009.
- [27] O. Banerjee, L. El Ghaoui, and A. D'Aspremont, "Model selection through sparse maximum likelihood estimation for multivariate Gaussian or binary data," *J. Mach. Learn. Res.*, vol. 9, pp. 485–516, Mar. 2008.
- [28] J. Fan and J. Lv, "Sure independence screening for ultrahigh dimensional feature space," *J. Roy. Statist. Soc. B, Stat. Methodol.*, vol. 70, no. 5, pp. 849–911, 2008.
- [29] D. R. Cox and N. Wermuth, *Multivariate Dependencies: Models, Analysis and Interpretation*. London, U.K.: Chapman Hall, 1996.
- [30] Y. Benjamini and Y. Hochberg, "Controlling the false discovery rate: A practical and powerful approach to multiple testing," *J. Roy. Stat. Soc., B, Methodol.*, vol. 57, no. 1, pp. 289–300, Aug. 1995.
- [31] Y. Benjamini and D. Yekutieli, "The control of the false discovery rate in multiple testing under dependency," *Ann. Statist.*, vol. 29, no. 4, pp. 1165–1188, Aug. 2001.
- [32] P. David, "Spherical matrix distributions and a multivariate model," *J. Roy. Stat. Soc., B, Methodol.*, vol. 39, no. 2, pp. 254–261, Jan. 1977.
- [33] T. W. Anderson and K.-T. Fang, "Theory and applications of elliptically contoured and related distributions," Stanford Univ CA Dept Statist., Stanford, CA, USA, Tech. Rep. 24, 1990.
- [34] T. W. Anderson, "Nonnormal multivariate distributions: Inference based on elliptically contoured distributions," Dept. Statist., Stanford Univ., Stanford, CA, USA, Tech. Rep. 28, 1992.
- [35] F. Larrión, C. P. de Mello, A. Morgana, V. Neumann-Lara, and M. A. Pizaña, "The clique operator on cographs and serial graphs," *Discrete Math.*, vol. 282, nos. 1–3, pp. 183–191, May 2004.
- [36] G. H. Golub and C. F. Van Loan, *Matrix Computations* (Studies in the Mathematical Sciences), vol. 3, 4th ed. Baltimore, MD, USA: Johns Hopkins Univ. Press, 2012.
- [37] S. Bubeck, J. Ding, R. Eldan, and M. Z. Rácz, "Testing for high-dimensional geometry in random graphs," *Random Struct. Algorithms*, vol. 49, no. 3, pp. 503–532, Oct. 2016.
- [38] A. D. Barbour, "Topics in Poisson approximation," *Handbook Statist.*, vol. 19, pp. 79–115, Jan. 2001.
- [39] R. Arratia, L. Goldstein, and L. Gordon, "Poisson approximation and the Chen-Stein method," *Stat. Sci.*, vol. 5, no. 4, pp. 403–424, Nov. 1990.
- [40] T. W. Anderson, *An Introduction to Multivariate Statistical Analysis* (Probability and Statistics), 3rd ed. Hoboken, NJ, USA: Wiley, 2003.
- [41] S. Li, "Concise formulas for the area and volume of a hyperspherical cap," *Asian J. Math. Statist.*, vol. 4, no. 1, pp. 66–70, 2011.
- [42] C. Villani, *Topics in Optimal Transportation* (Graduate Studies in Mathematics), vol. 58. Providence, RI, USA: American Mathematical Society, 2003.
- [43] D. A. Levin and Y. Peres, *Markov Chains and Mixing Times*, vol. 107. Providence, RI, USA: American Mathematical Society, 2017.
- [44] J. A. Adell and P. Jodrá, "Exact Kolmogorov and total variation distances between some familiar discrete distributions," *J. Inequalities Appl.*, vol. 2006, no. 1, p. 64307, 2006.

**Yun Wei** received the B.S. degree in scientific computing from Nankai University in 2013 and the Ph.D. degree in applied mathematics from the University of Michigan in 2020. He is currently a Post-Doctoral Associate with the Department of Statistical Science, Duke University. His research interests include mixture models, high dimensional statistics, information theory, and their applications.

**Bala Rajaratnam** received the B.Sc. (Hons.)/M.Sc. degree from the University of the Witwatersrand, Johannesburg, South Africa, in 1999, and the M.S./Ph.D. degree from Cornell University, Ithaca, NY, USA, in 2006. He is currently a Faculty Member at the Department of Statistics, University of California at Davis, Davis, CA, USA. His research interests include graphical models, machine learning, data science, high-dimensional inference, positivity, signal processing, spatio-temporal and environmental modeling, and financial engineering. He was a recipient of several awards and recognitions including two Federal CAREER Awards, the National Science Foundation (NSF) CAREER Award, and the Defense Advanced Research Projects Agency (DARPA) Young Faculty Award.

**Alfred O. Hero** (Life Fellow, IEEE) received the B.S. degree (*summa cum laude*) from Boston University in 1980 and the Ph.D. degree from Princeton University in 1984, both in electrical engineering. He was the Founding Co-Director of the University's Michigan Institute for Data Science (MIDAS). He is currently the John H. Holland Distinguished University Professor of electrical engineering and computer science and the R. Jamison and Betty Williams Professor of engineering at the University of Michigan, Ann Arbor. His primary appointment is at the Department of Electrical Engineering and Computer Science and also has appointments, by courtesy, at the Department of Biomedical Engineering and the Department of Statistics. His recent research interests are in modeling high dimensional spatio-temporal data, multimodal data integration, statistical signal processing, and machine learning. He is a fellow of the Society for Industrial and Applied Mathematics (SIAM). He has served as the President for the IEEE Signal Processing Society and as a member for the IEEE Board of Directors. He has received numerous awards for his research and service to the profession including several best paper awards, the 2013 IEEE Signal Processing Society Technical Achievement Award, the 2015 Society Award from the IEEE Signal Processing Society, and the 2020 Fourier Award from the IEEE. He received the 2017 Stephen S. Attwood Excellence in Engineering Award and the 2018 H. Scott Fogler Award for Professional Leadership and Service from the University of Michigan. He was the Co-General Chair of the 2019 IEEE Information Theory Symposium. He is a Section Editor of the *SIAM Journal on Mathematics of Data Science* and a Senior Editor of the *IEEE JOURNAL OF SELECTED TOPICS IN SIGNAL PROCESSING*.