Edge Differentially Private Estimation in the β-model via Jittering and Method of Moments*

Jinyuan Chang[a], Qiao Hu[a], Eric D. Kolaczyk[b], Qiwei Yao[c], and Fengting Yi[a]

[a] School of Statistics, Southwestern University of Finance and Economics, Chengdu, 611130, China
[b] Department of Mathematics and Statistics, Boston University, Boston, MA, 02215, USA
[c] Department of Statistics, London School of Economics and Political Science, London, WC2A 2AE, UK

Abstract

A standing challenge in data privacy is the trade-off between the level of privacy and the efficiency of statistical inference. Here we conduct an in-depth study of this trade-off for parameter estimation in the β-model (Chatterjee, Diaconis and Sly, 2011) for edge differentially private network data released via jittering (Karwa, Krivitsky and Slavković, 2017). Unlike most previous approaches based on maximum likelihood estimation for this network model, we proceed via method of moments. This choice facilitates our exploration of a substantially broader range of privacy levels – corresponding to stricter privacy – than has been to date. Over this new range we discover our proposed estimator for the parameters exhibits an interesting phase transition, with both its convergence rate and asymptotic variance following one of three different regimes of behavior depending on the level of privacy. Because identification of the operable regime is difficult to impossible in practice, we devise a novel adaptive bootstrap procedure to construct uniform inference across different phases. In fact, leveraging this bootstrap we are able to provide for simultaneous inference of all parameters in the β-model (i.e., equal to the number of vertices), which would appear to be the first result of its kind. Numerical experiments confirm the competitive and reliable finite sample performance of the proposed inference methods, next to a comparable maximum likelihood method, as well as significant advantages in terms of computational speed and memory.

Key words: adaptive inference, β-model, bootstrap inference, data privacy, data release mechanism, edge differential privacy, phase transition, Stein's method.

1 Introduction

In this information age, data is one of the most important assets. With ever-advancing machine learning technology, collecting, sharing and using data yield great societal and economic benefits, while the abundance and granularity of personal data bring new risks of potential exposure of sensitive personal or financial information which may lead to adverse consequences. Therefore, continuous and conscientious effort has been made to formulate concepts of sensitivity of the data and privacy guarantee in data usage, and those concepts evolve along with the technological advancement. At present, one of most commonly

*The authors equally contributed to the paper. Chang, Hu and Yi were supported in part by the National Natural Science Foundation of China (grant nos. 71991472, 72125008 and 11871401), and the Joint Lab of Data Science and Business Intelligence at Southwestern University of Finance and Economics. Kolaczyk was supported in part by U.S. National Science Foundation award SES-2120115. Yao was supported in part by the U.K. Engineering and Physical Sciences Research Council (grant nos. EP/V007556/1 and EP/X002195/1). Yi was also supported in part by the Fundamental Research Funds for the Central Universities (grant no. JBK2101013), and the China Postdoctoral Science Foundation (grant no. 2021M692663). The authors thank Vishesh Karwa for sharing code for differentially maximum likelihood estimation in the β-model.

used formulations of data privacy is the so-called differential privacy (Dwork, 2006; Wasserman and Zhou, 2010). This paper is devoted to studying statistical estimation in the context of edge differential privacy for network data.

In network data, individuals (e.g., persons or firms) are typically represented by nodes and their inter-relationships are represented by edges. Therefore, network data often contain sensitive individual information. On the other hand, for analysis purposes the information of interest in the data should be sufficiently preserved. Hence, the primary concern for data privacy is two-folded: (a) to release only a sanitized version of the original network data to protect privacy, and (b) the sanitized data should preserve the information of interest such that analysis based on the sanitized data is still effective.

To protect privacy, the conventional approach is to release some noised version of summary statistics of interest. Normally the summary statistics used are of (much) lower dimension than the original data. In the context of network data, the chosen summary statistics can be the node degree sequence (Karwa and Slavković, 2016) or subgraph counts (Blocki et al., 2013). To achieve differential privacy, only a noised version of the summary statistics is released. The noised version of the statistics is generated based on some appropriate release mechanism, which depends critically on the so-called sensitivity of the adopted statistics. One of the most frequently used data release schemes is the Laplace mechanism of Dwork et al. (2006). See also Section 2 of Wasserman and Zhou (2010), and Section 3 of Karwa and Slavković (2016). Karwa and Slavković (2016) considered edge differential privacy for the β-model (Chatterjee, Diaconis and Sly, 2011), where only the node degree sequence, which is a sufficient statistic, is released with added noise generated from a discrete Laplace mechanism. However, a noisy degree sequence may no longer be a legitimate degree sequence. Even for a legitimate degree sequence, the maximum likelihood estimator (MLE) may not exist. Karwa and Slavković (2016) propose a two step procedure that entails 'de-noising' the noisy sequence first and then estimating the parameters using the de-noised data by MLE.

A radically different approach is to release a noisy version of an entire network. Karwa, Krivitsky and Slavković (2017) offer what they call a generalized random response mechanism for doing so and present empirical results of its use with maximum-likelihood estimation in exponential random graph models. The structure of this release mechanism is the same as the noisy network setting of Chang, Kolaczyk and Yao (2021), where the edge status of each pair of nodes is known only up to some binary noise and method of moments was used to estimate certain network summary statistics. As noted by Chang, Kolaczyk and Yao (2020), this noisy network setting in turn is essentially analogous to the idea of jittering in the analysis of classical Euclidean data, where each original data point is released with added noise. In this paper we study this jittering release mechanism for network data, and we do so in the specific context of parameter estimation in the β-model. However, importantly, we note that unlike approaches based on release of noised versions of specific, pre-determined summary statistics, jittering allows for the possibility of multiple statistics to be calculated and/or quantities to be estimated from the same released network.

Specifically, we conduct an in-depth study on the statistical inference for the β-model based on the edge π-differentially private data generated via jittering, where $\pi > 0$ reflects the privacy level; the smaller π, the greater the level of privacy. Unlike most previous approaches to inference under this model, based on maximum likelihood estimation, we proceed via method of moments. This choice facilitates our exploration of a substantially broader range of privacy levels π than has been to date. Let p be the number of nodes in the network. Our major contributions are as follows. First, we develop the asymptotic theory when $p \to \infty$ and $\pi \to 0$, and find that (i) in order to achieve consistency of the newly proposed moment-based estimator, π should decay to zero slower than $p^{-1/3} \log^{1/6} p$, while (ii) both the convergence rate and the asymptotic variance of our proposed estimator depend intimately on the interplay between p and π. In particular, the asymptotic behavior of these quantities exhibits an interesting phase transition phenomenon, as π decays to zero as a function of p, following one of three different regimes of behavior: $\pi \gg p^{-1/4}$, $\pi \asymp p^{-1/4}$, and $p^{-1/4} \gg \pi \gg p^{-1/3} \log^{1/6} p$. Second, because identification of the operable regime is difficult to impossible in practice, we devise a novel adaptive bootstrap procedure to construct uniform inference across different phases. Third, leveraging this bootstrap we are able to provide for simultaneous inference of all parameters in the β-model (i.e., equal to the number of vertices), which would appear to be the first result of its kind and requires a substantially different and more nuanced technical approach than for finite-dimensional results. Lastly, numerical experiments confirm the competitive and reliable finite sample performance of the proposed inference methods, next to a comparable maximum likelihood method, as well as significant advantages in terms of computational speed and memory.

The rest of the paper is organized as follows. Section 2 introduces the concept of edge π-differential privacy for networks, and the data release mechanism by jittering (Karwa, Krivitsky and Slavković, 2017). Section 3 addresses inference for the β-model based on edge differentially private data, introducing the method of moments estimator and characterizing its asymptotic behavior. Section 4 develops the bootstrap adaptive inference that makes inference feasible in practice (i.e., despite the phase transition), and presents the accompanying results on simultaneous inference. Some numerical results are reported in Section 5. We relegate all the technical proofs to the Appendix.

For any integer $d \geq 1$, we write $[d] = \{1, \ldots, d\}$, and denote by $\mathbf{I}_d$ the identity matrix of order d. We denote by $I(\cdot)$ the indicator function. For a vector $\mathbf{h} = (h_1, \ldots, h_d)^{\mathrm{T}}$, we write $|\mathbf{h}|_1 = \sum_{j=1}^{d} |h_j|$ and $|\mathbf{h}|_\infty = \max_{j \in [d]} |h_j|$ for its L_1-norm and L_∞-norm, respectively. For a countable set $\mathcal{S}$, we use $\#\mathcal{S}$ or $|\mathcal{S}|$ to denote its cardinality. For two sequences of positive numbers $\{a_p\}_{p\geq 1}$ and $\{c_p\}_{p\geq 1}$, we write $a_p \lesssim c_p$ or $c_p \gtrsim a_p$ if $\limsup_{p\to\infty} a_p/c_p < \infty$, and write $a_p \asymp c_p$ if and only if $a_p \lesssim c_p$ and $c_p \lesssim a_p$. We also write $a_p \ll c_p$ or $c_p \gg a_p$ if $\limsup_{p\to\infty} a_p/c_p = 0$.

2 Edge differential privacy

2.1 Definition

We consider simple networks in the sense that there are no self-loops and there exists at most one edge from one node to another for a directed network, and at most one edge between two nodes for an undirected network. Such a network with p nodes can be represented by an adjacency matrix $\mathbf{X} = (X_{i,j})_{p\times p}$, where $X_{i,i} \equiv 0$, and $X_{i,j} = 1$ indicating an edge from the i-th node to the j-th node, and 0 otherwise. For undirected networks, we have $X_{i,j} = X_{j,i}$. In this paper, we always assume that the p nodes are fixed and are labeled as $1, \ldots, p$. Then a simple network can be represented entirely by its adjacency matrix. To simplify statements, we often refer to an adjacency matrix $\mathbf{X}$ as a network.

Let $\mathcal{X}$ be the set consisting of the adjacency matrices of all the simple and directed (or undirected) networks with p nodes. For any $\mathbf{X} = (X_{i,j})_{p\times p} \in \mathcal{X}$ and $\mathbf{Y} = (Y_{i,j})_{p\times p} \in \mathcal{X}$, the Hamming distance between $\mathbf{X}$ and $\mathbf{Y}$ is defined as

$$\delta(\mathbf{X}, \mathbf{Y}) = \#\{(i,j) \in \mathcal{I} : X_{i,j} \neq Y_{i,j}\}\,, \tag{2.1}$$

where $\mathcal{I} = \{(i,j) : 1 \le i \ne j \le p\}$ for directed networks, and $\mathcal{I} = \{(i,j) : 1 \le i < j \le p\}$ for undirected networks. To protect privacy, the original network $\mathbf{X}$ is not released directly. Instead we release a sanitized version $\mathbf{Z} = (Z_{i,j})_{p\times p} \in \mathcal{X}$ of the network, where $\mathbf{Z}$ is generated according to some conditional distribution $Q(\cdot \,|\, \mathbf{X})$. Here Q is also called a release mechanism (Wasserman and Zhou, 2010).

Definition 1 (Edge differential privacy)**.** For any $\pi > 0$, a release mechanism (i.e. a conditional probability distribution) Q satisfies π-edge differential privacy if

$$\sup_{\substack{\mathbf{X},\mathbf{Y}\in\mathcal{X},\\ \delta(\mathbf{X},\mathbf{Y})=1}} \sup_{\substack{\mathbf{Z}\in\mathcal{X},\\ Q(\mathbf{Z}\,|\,\mathbf{X})>0}} \frac{Q(\mathbf{Z}\,|\,\mathbf{Y})}{Q(\mathbf{Z}\,|\,\mathbf{X})} \le e^{\pi}\,. \tag{2.2}$$

The definition above equates privacy with the inability to distinguish two close networks. The privacy parameter π controls the amount of randomness added to released data; the smaller π is the more protection on privacy. Notice that (2.2) is much more stringent than requiring $|Q(\mathbf{Z}\,|\,\mathbf{Y}) - Q(\mathbf{Z}\,|\,\mathbf{X})|$ to be small for any $\mathbf{X},\mathbf{Y} \in \mathcal{X}$ with $\delta(\mathbf{X},\mathbf{Y}) = 1$. In practice π is often chosen to be small. Then it follows from (2.2) that

$$\sup_{\substack{\mathbf{X},\mathbf{Y}\in\mathcal{X},\\ \delta(\mathbf{X},\mathbf{Y})=1}} \sup_{\substack{\mathbf{Z}\in\mathcal{X},\\ Q(\mathbf{Z}\,|\,\mathbf{X})>0}} \frac{|Q(\mathbf{Z}\,|\,\mathbf{Y}) - Q(\mathbf{Z}\,|\,\mathbf{X})|}{Q(\mathbf{Z}\,|\,\mathbf{X})} \le e^{\pi} - 1 \approx \pi\,.$$

Note that multiple notions of privacy have been introduced for networks; see Jiang et al. (2020) for a recent survey. In this paper we focus on the notion of edge differential privacy (e.g., Nissim, Raskhodnikova and Smith (2007)). At the same time, there is a connection between differential privacy and hypothesis testing.

Proposition 1 (Wasserman and Zhou, 2010)**.** *Let the released network* $\mathbf{Z} \sim Q(\cdot\,|\,\mathbf{X})$ *and* Q *satisfy* π*-edge differential privacy for some* $\pi > 0$*. For any given* $i \ne j$*, consider hypotheses* $H_0 : X_{i,j} = 1$ *versus* $H_1 : X_{i,j} = 0$*. Then the power of any test at the significance level* γ *and based on* $\mathbf{Z}$*,* Q *and the distribution of* $\mathbf{X}$ *is bounded from above by* γe^{π}*, provided that* $X_{i,j}$ *is independent of* $\{X_{k,\ell} : (k,\ell) \in \mathcal{I}$ *and* $(k,\ell) \ne (i,j)\}$*.*

Proposition 1 implies that if $\mathbf{Z}$ is released through Q which satisfies π-edge differential privacy and π is sufficiently small, it is virtually impossible to identify whether an edge exists (i.e. $X_{i,j} = 1$) or not (i.e. $X_{i,j} = 0$) in the original network through statistical tests, as the power of any test is bounded by its significance level multiplied by e^{π}. The independence condition in Proposition 1 is satisfied by the Erdös-Rényi class of models for which all edges are independent, including the β-model and the well-known stochastic block model. Proposition 1 follows almost immediately from the Neyman-Pearson lemma for the optimality of likelihood ratio tests for simple null and simple alternative hypotheses. It was first proved by Wasserman and Zhou (2010) with independent observations. Since their proof can be adapted to our setting in a straightforward manner, we omit the details.

For further discussion on differential privacy under more general settings, we refer to Dwork et al. (2006) and Wasserman and Zhou (2010).

2.2 Edge privacy via jittering

Now we introduce the data release mechanism of Karwa, Krivitsky and Slavković (2017), which is formally the same as the noisy network structure adopted in Chang, Kolaczyk and Yao (2021). This approach releases a jittered version of the entire network. The word of 'jittering' means that a small amount of noise is added to every single data point (Hennig, 2007).

For $\mathcal{I}$ specified just after (2.1) above, we define a data release mechanism as follows:

$$Z_{i,j} = X_{i,j} I(\varepsilon_{i,j} = 0) + I(\varepsilon_{i,j} = 1) \tag{2.3}$$

for each $(i,j) \in \mathcal{I}$. In the above expression, $\{\varepsilon_{i,j}\}_{(i,j)\in\mathcal{I}}$ are independent random variables with

$$\mathbb{P}(\varepsilon_{i,j} = 1) = \alpha_{i,j}\,, \quad \mathbb{P}(\varepsilon_{i,j} = 0) = 1 - \alpha_{i,j} - \beta_{i,j} \quad \text{and} \quad \mathbb{P}(\varepsilon_{i,j} = -1) = \beta_{i,j}\,. \tag{2.4}$$

For an undirected network, $Z_{i,j} = Z_{j,i}$ for $j > i$. Then it follows from (2.3) and (2.4) that

$$\mathbb{P}(Z_{i,j} = 1 \,|\, X_{i,j} = 0) = \alpha_{i,j} \quad \text{and} \quad \mathbb{P}(Z_{i,j} = 0 \,|\, X_{i,j} = 1) = \beta_{i,j}\,. \tag{2.5}$$

Furthermore the proposition below follows from (2.2) and (2.5) immediately. See also Proposition 1 of Karwa, Krivitsky and Slavković (2017).

Proposition 2. *The data release mechanism* (2.3) *satisfies* π*-edge differential privacy with*

$$\pi = \max_{(i,j)\in\mathcal{I}} \log\left\{\max\left(\frac{\alpha_{i,j}}{1-\beta_{i,j}}\,, \frac{\beta_{i,j}}{1-\alpha_{i,j}}\,, \frac{1-\alpha_{i,j}}{\beta_{i,j}}\,, \frac{1-\beta_{i,j}}{\alpha_{i,j}}\right)\right\}.$$

Remark 1. It is easy to see from Proposition 2 that the maximum privacy is achieved by setting $\alpha_{i,j} = \beta_{i,j} = 0.5$ for all $(i,j) \in \mathcal{I}$, as then $\pi = 0$. By (2.3) and (2.4), $Z_{i,j} = I(\varepsilon_{i,j} = 1)$ then, i.e. $\mathbf{Z}$ carries no information about $\mathbf{X}$. In order to achieve high privacy, we need to use large $\alpha_{i,j}$ and $\beta_{i,j}$. In Section 3 below we will develop statistical inference approaches for the original network $\mathbf{X}$ based on the released data $\mathbf{Z}$ as $\pi \to 0$. When $\pi \to 0$, it holds that $\pi \asymp \max_{(i,j)\in\mathcal{I}}(1 - \alpha_{i,j} - \beta_{i,j})$ and there exists a constant $\epsilon > 0$ such that $\alpha_{i,j}, \beta_{i,j} \in (\epsilon, 1-\epsilon)$ for all $(i,j) \in \mathcal{I}$.

3 Differentially private inference for the β-model

In this section we introduce a new method-of-moments estimator for the parameters of the network β-model and characterize the asymptotic behavior of this estimator, through which we discover an interesting phase transition.

3.1 The β-model

The so-called β-model for undirected networks is characterized by p parameters $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_p)^{\mathrm{T}} \in \mathbb{R}^p$ which define the probability function

$$\mathbb{P}(X_{i,j} = 1) = \frac{\exp(\theta_i + \theta_j)}{1 + \exp(\theta_i + \theta_j)}\,, \qquad i \neq j\,. \tag{3.1}$$

See Chatterjee, Diaconis and Sly (2011). Then the likelihood function is given by

$$f(\mathbf{X}; \boldsymbol{\theta}) = \prod_{i,j:\, i<j} \frac{\exp\{(\theta_i + \theta_j) X_{i,j}\}}{1 + \exp(\theta_i + \theta_j)} \propto \exp(U_1\theta_1 + \cdots + U_p\theta_p)\,,$$

where $U_i = \sum_{j:\, j\neq i} X_{i,j}$ is the degree of the i-th node. Hence the degree sequence $\mathbf{U} = (U_1, \ldots, U_p)^{\mathrm{T}}$ is a sufficient statistic.

Denote by $\tilde{\boldsymbol\theta}(\mathbf{U}) = \{\tilde\theta_1(\mathbf{U}), \ldots, \tilde\theta_p(\mathbf{U})\}^{\mathrm{T}}$ the maximum likelihood estimator for $\boldsymbol\theta$ based on $\mathbf{U}$. For given degree sequence $\mathbf{U}$, $\tilde{\boldsymbol\theta}(\mathbf{U})$ must satisfy the following moment equations:

$$U_i = \sum_{j:\, j\neq i} \frac{\exp\{\tilde\theta_i(\mathbf{U}) + \tilde\theta_j(\mathbf{U})\}}{1 + \exp\{\tilde\theta_i(\mathbf{U}) + \tilde\theta_j(\mathbf{U})\}}\,, \qquad i \in [p]\,.$$

Unfortunately $\tilde{\boldsymbol\theta}(\mathbf{U})$ may not exist; see Theorem 1 of Karwa and Slavković (2016) for necessary and sufficient conditions for the existence of $\tilde{\boldsymbol\theta}(\mathbf{U})$. When $\tilde{\boldsymbol\theta}(\mathbf{U})$ exists, Chatterjee, Diaconis and Sly (2011) show that

$$|\tilde{\boldsymbol\theta}(\mathbf{U}) - \boldsymbol\theta|_\infty \leq C_* p^{-1/2} \log^{1/2} p \tag{3.2}$$

with probability at least $1 - C_* p^{-2}$, where $C_* > 0$ is a constant depending only on $|\boldsymbol\theta|_\infty$. For any fixed integer $s \geq 1$ and distinct $\ell_1, \ldots, \ell_s \in [p]$, Yan and Xu (2013) establish the asymptotic normality of $\{\tilde\theta_{\ell_1}(\mathbf{U}), \ldots, \tilde\theta_{\ell_s}(\mathbf{U})\}^{\mathrm{T}}$ as $p \to \infty$, which can be used to construct joint confidence regions for $(\theta_{\ell_1}, \ldots, \theta_{\ell_s})^{\mathrm{T}}$. However, to our best knowledge, simultaneous inference for all p parameters in the β-model remains unresolved in the literature.

Karwa and Slavković (2016) consider differentially private maximum likelihood estimation for $\boldsymbol\theta$ based on a noisy version of the degree sequence. With $\pi \asymp (\log p)^{-1/2}$, the noisy degree sequence in their setting is defined as $\mathbf{U} + \mathbf{V}$, where the components of $\mathbf{V} = (V_1, \ldots, V_p)^{\mathrm{T}}$ are drawn independently from a discrete Laplace distribution with the probability mass function $\mathbb{P}(V = v) = (1-\kappa)\kappa^{|v|}/(1+\kappa)$ for any integer v, where $\kappa = \exp(-\pi/2)$. Given observed $\mathbf{U} + \mathbf{V}$, Karwa and Slavković (2016) propose a two-step procedure: (a) find the maximum likelihood estimator $\mathbf{U}^*$ for $\mathbf{U}$ based on $\mathbf{U} + \mathbf{V}$, and (b) estimate $\boldsymbol\theta$ by $\tilde{\boldsymbol\theta}(\mathbf{U}^*)$. Theorem 4 of Karwa and Slavković (2016) shows that $\{\tilde\theta_{\ell_1}(\mathbf{U}^*), \ldots, \tilde\theta_{\ell_s}(\mathbf{U}^*)\}^{\mathrm{T}}$ shares the same asymptotic normality as $\{\tilde\theta_{\ell_1}(\mathbf{U}), \ldots, \tilde\theta_{\ell_s}(\mathbf{U})\}^{\mathrm{T}}$ for any fixed integer $s \geq 1$ and distinct $\ell_1, \ldots, \ell_s \in [p]$.

To appreciate this 'free privacy' result, let us assume first that $|\boldsymbol\theta|_\infty \leq C$ for some universal constant $C > 0$. Then there exists a universal constant $\tilde C > 1$ such that $\tilde C^{-1} p \leq \min_{i\in[p]} U_i \leq \max_{i\in[p]} U_i \leq \tilde C p$ holds almost surely as $p \to \infty$. On the other hand, Lemma C in the supplementary material of Karwa and Slavković (2016) indicates that $|\mathbf{U}^* - \mathbf{U}|_\infty \leq \sqrt{6} p^{1/2} \log^{1/2} p$ holds almost surely as $p \to \infty$, which implies that $\mathbf{U}^*$ is dominated by $\mathbf{U}$. Hence, $\tilde{\boldsymbol\theta}(\mathbf{U}^*)$ shares the same asymptotic distribution as $\tilde{\boldsymbol\theta}(\mathbf{U})$ when $\pi \asymp (\log p)^{-1/2}$. However, the asymptotic behavior of $\tilde{\boldsymbol\theta}(\mathbf{U}^*)$ is unknown when $\pi \ll (\log p)^{-1/2}$.

Our interest in this paper is on differentially private estimation based on released data $\mathbf{Z} = (Z_{i,j})_{p\times p}$ generated by the more general jittering mechanism (2.3). Remark 1 in Section 2.2 shows that $\mathbf{Z}$ is π-differentially private with $\pi \asymp \max_{(i,j):\, i<j}(1 - \alpha_{i,j} - \beta_{i,j})$. To gain more appreciation of the impact of the privacy level π on the efficiency of inference, we introduce a new moment-based estimation for $\boldsymbol\theta$ based on $\mathbf{Z}$. We then establish the asymptotic theory under the setting that $p \to \infty$ and π may vary with respect to p. Of particular interest is the findings when $\pi \to 0$ together with $p \to \infty$. It turns out the asymptotic distribution of the new proposed estimator depends intimately on the interplay between π and p, exhibiting interesting phase transition in the convergence rate and the asymptotic variance as π decays to zero as a function of p. See Theorem 1 and Remark 2(a) in Section 3.3. To overcome the complexity in inference due to the phase transition, a novel bootstrap method is proposed, which provides a uniform

inference regardless different phases. In addition, it also facilitates the simultaneous inference for all the p components of $\boldsymbol{\theta}$ as $p \to \infty$.

3.2 A new moment-based estimator

Under β-model (3.1), $\mathbb{P}(X_{i,j} = 1)/\mathbb{P}(X_{i,j} = 0) = \exp(\theta_i + \theta_j)$ for any $i \neq j$. Hence

$$\frac{\mathbb{P}(X_{i,\ell} = 1)\mathbb{P}(X_{i,j} = 0)\mathbb{P}(X_{\ell,j} = 1)}{\mathbb{P}(X_{i,\ell} = 0)\mathbb{P}(X_{i,j} = 1)\mathbb{P}(X_{\ell,j} = 0)} = \exp(2\theta_\ell)\,, \qquad i \neq j \neq \ell\,. \tag{3.3}$$

Since only the sanitized network $\mathbf{Z} = (Z_{i,j})_{p\times p}$, defined as in (2.3)–(2.5), is available, we represent (3.3) in terms of the probabilities of $Z_{i,j}$. For any $i \neq j$ and $\tau \in \{0, 1\}$, put

$$\varphi_{(i,j),\tau}(x) = (x - \alpha_{i,j})^\tau (1 - \beta_{i,j} - x)^{1-\tau}\,.$$

Then $\mathbb{P}(X_{i,j} = 0) = (1-\alpha_{i,j}-\beta_{i,j})^{-1}\mathbb{E}\{\varphi_{(i,j),0}(Z_{i,j})\}$ and $\mathbb{P}(X_{i,j} = 1) = (1-\alpha_{i,j}-\beta_{i,j})^{-1}\mathbb{E}\{\varphi_{(i,j),1}(Z_{i,j})\}$.

To simplify the notation, we write $\varphi_{(i,j),\tau}(Z_{i,j})$ as $\varphi_{(i,j),\tau}$ for any $i \neq j$ and $\tau \in \{0, 1\}$. Since $Z_{i,j}$ is independent of $\{Z_{\tilde{i},\tilde{j}} : |\{\tilde{i}, \tilde{j}\} \cap \{i, j\}| \leq 1\}$, it follows from (3.3) that

$$\frac{\mathbb{E}\{\varphi_{(i,\ell),1}\varphi_{(i,j),0}\varphi_{(\ell,j),1}\}}{\mathbb{E}\{\varphi_{(i,\ell),0}\varphi_{(i,j),1}\varphi_{(\ell,j),0}\}} = \exp(2\theta_\ell)\,, \qquad i \neq j \neq \ell\,. \tag{3.4}$$

Hence a moment-based estimator for θ_ℓ can be defined as

$$\hat{\theta}_\ell = \frac{1}{2}\log\left(\frac{\hat{\mu}_{\ell,1}}{\hat{\mu}_{\ell,2}}\right), \tag{3.5}$$

where

$$\hat{\mu}_{\ell,1} = \frac{1}{|\mathcal{H}_\ell|}\sum_{(i,j)\in\mathcal{H}_\ell}\varphi_{(i,\ell),1}\varphi_{(i,j),0}\varphi_{(\ell,j),1}\,, \tag{3.6}$$

$$\hat{\mu}_{\ell,2} = \frac{1}{|\mathcal{H}_\ell|}\sum_{(i,j)\in\mathcal{H}_\ell}\varphi_{(i,\ell),0}\varphi_{(i,j),1}\varphi_{(\ell,j),0}\,, \tag{3.7}$$

and $\mathcal{H}_\ell = \{(i,j) : i, j \neq \ell \text{ such that } i < j\}$.

3.3 Asymptotic properties and phase transition

Put $\mathcal{I} = \{(i,j) : 1 \leq i < j \leq p\}$. We always confine $\{(\alpha_{i,j}, \beta_{i,j})\}_{(i,j)\in\mathcal{I}} \in \mathcal{M}(\gamma; C_1, C_2)$ with

$$\mathcal{M}(\gamma; C_1, C_2) = \bigg\{\{(\alpha_{i,j}, \beta_{i,j})\}_{(i,j)\in\mathcal{I}} : C_1 < \alpha_{i,j}, \beta_{i,j} < 0.5, \\ C_2^{-1}\gamma \leq \min_{(i,j)\in\mathcal{I}}(1 - \alpha_{i,j} - \beta_{i,j}) \leq \max_{(i,j)\in\mathcal{I}}(1 - \alpha_{i,j} - \beta_{i,j}) \leq C_2\gamma\bigg\}$$

for some $\gamma \in (0, 1]$, $C_1 \in (0, 0.5)$ and $C_2 > 1$. Our theoretical analysis allows γ to be a constant, or to vary with respect to p. Of particular interest are the cases when $\gamma \to 0$ (at different rates) together with $p \to \infty$. When $\{(\alpha_{i,j}, \beta_{i,j})\}_{(i,j)\in\mathcal{I}} \in \mathcal{M}(\gamma; C_1, C_2)$ for some fixed constants $C_1 \in (0, 0.5)$ and $C_2 > 1$, it follows from Remark 1 in Section 2.2 that the privacy level $\pi \asymp \gamma \to 0$.

3.3.1 Consistency

Proposition 3 below presents the consistency for the moment-based estimator $\hat{\theta}_\ell$ defined in (3.5), which indicates that θ_ℓ can be estimated consistently under the edge π-differential privacy with $\pi \to 0$, as long as $\pi \gg p^{-1/3}\log^{1/6} p$.

Condition 1. There exists a universal constant $C_3 > 0$ such that $|\boldsymbol{\theta}|_\infty \le C_3$.

Proposition 3. *Let Condition* 1 *hold and* $\{(\alpha_{i,j}, \beta_{i,j})\}_{(i,j)\in\mathcal{I}} \in \mathcal{M}(\gamma; C_1, C_2)$ *for two fixed constants* $C_1 \in (0, 0.5)$ *and* $C_2 > 1$*. If* $\gamma \gg p^{-1/3}\log^{1/6} p$*, it then holds that* $\max_{\ell\in[p]} |\hat{\theta}_\ell - \theta_\ell| = O_{\rm p}(\gamma^{-3}p^{-1}\log^{1/2} p) + O_{\rm p}(\gamma^{-1}p^{-1/2}\log^{1/2} p)$.

Notice that the privacy level $\pi \asymp \gamma$. In order to ensure the consistency of $\hat{\theta}_\ell$, the edge differential privacy level π must satisfy condition $\pi \gg p^{-1/3}\log^{1/6} p$. On the other hand, when $\alpha_{i,j} \equiv 0 \equiv \beta_{i,j}$, our estimator (3.5) is essentially constructed based on the original network $\mathbf{X}$. In this case, we can then set $\gamma = 1$, and our estimator shares the same convergence rate of the maximum likelihood estimator of Chatterjee, Diaconis and Sly (2011); see (3.2) in Section 3.1.

3.3.2 Asymptotic normality

Let $\mu_{\ell,1} = |\mathcal{H}_\ell|^{-1}\sum_{(i,j)\in\mathcal{H}_\ell}\mathbb{E}\{\varphi_{(i,\ell),1}\varphi_{(i,j),0}\varphi_{(\ell,j),1}\}$ and $\mu_{\ell,2} = |\mathcal{H}_\ell|^{-1}\sum_{(i,j)\in\mathcal{H}_\ell}\mathbb{E}\{\varphi_{(i,\ell),0}\varphi_{(i,j),1}\varphi_{(\ell,j),0}\}$ for each $\ell \in [p]$, which are, respectively, the population analogues of $\hat{\mu}_{\ell,1}$ and $\hat{\mu}_{\ell,2}$ defined in (3.6) and (3.7). Put $N = (p-1)(p-2)$. Proposition 4 gives the asymptotic expansion of $\hat{\theta}_\ell - \theta_\ell$, which can be obtained from the proof of Theorem 1 below. For any $i \ne \ell$, let

$$\lambda_{i,\ell} = \frac{1}{p-2}\sum_{j:\, j\ne\ell,i}\bigg[\frac{1}{\mu_{\ell,1}}\mathbb{E}\{\varphi_{(\ell,j),1}\}\mathbb{E}\{\varphi_{(i,j),0}\} + \frac{1}{\mu_{\ell,2}}\mathbb{E}\{\varphi_{(\ell,j),0}\}\mathbb{E}\{\varphi_{(i,j),1}\}\bigg]\,. \tag{3.8}$$

Proposition 4. *For any* $i \ne j$*, write* $\mathring{Z}_{i,j} = Z_{i,j} - \mathbb{E}(Z_{i,j})$*. Let Condition* 1 *hold and* $\{(\alpha_{i,j}, \beta_{i,j})\}_{(i,j)\in\mathcal{I}} \in \mathcal{M}(\gamma; C_1, C_2)$ *for two fixed constants* $C_1 \in (0, 0.5)$ *and* $C_2 > 1$*. If* $\gamma \gg p^{-1/3}\log^{1/6} p$*, it then holds that* $\hat{\theta}_\ell - \theta_\ell = \tilde{T}_{\ell,1} + \tilde{T}_{\ell,2} + \tilde{R}_\ell$*, where* $\tilde{T}_{\ell,1} = -N^{-1}\sum_{i,j:\, i\ne j,\, i,j\ne\ell}(\mu_{\ell,1}+\mu_{\ell,2})(2\mu_{\ell,1}\mu_{\ell,2})^{-1}\mathring{Z}_{i,\ell}\mathring{Z}_{\ell,j}\mathring{Z}_{i,j}$ *and* $\tilde{T}_{\ell,2} = (p-1)^{-1}\sum_{i:\, i\ne\ell}\lambda_{i,\ell}\mathring{Z}_{i,\ell}$ *satisfy* $|\tilde{T}_{\ell,1}| = O_{\rm p}(\gamma^{-3}p^{-1})$ *and* $|\tilde{T}_{\ell,2}| = O_{\rm p}(\gamma^{-1}p^{-1/2})$*, and the remainder term* $\tilde{R}_\ell$ *satisfies* $|\tilde{R}_\ell| = O_{\rm p}(\gamma^{-6}p^{-2}) + O_{\rm p}(\gamma^{-2}p^{-1}\log p)$.

The leading term in the asymptotic expansion of $\hat{\theta}_\ell - \theta_\ell$ will be different for different scenarios of γ: $\tilde{T}_{\ell,2}$, a partial sum of independent random variables, serves as the leading term if $\gamma \gg p^{-1/4}$, $\tilde{T}_{\ell,1} + \tilde{T}_{\ell,2}$ is the leading term if $\gamma \asymp p^{-1/4}$, and $\tilde{T}_{\ell,1}$, a generalized U-statistic, is the leading term if $p^{-1/4} \gg \gamma \gg p^{-1/3}\log^{1/6} p$. Such characteristic will lead to a phase transition phenomenon in the limiting distribution of the proposed moment-based estimator. Put

$$b_\ell = \frac{1}{p-1}\sum_{i:\, i\ne\ell}\lambda_{i,\ell}^2{\rm Var}(Z_{i,\ell})\,, \tag{3.9}$$

$$\tilde{b}_\ell = \frac{1}{2N}\bigg(\frac{\mu_{\ell,1}+\mu_{\ell,2}}{\mu_{\ell,1}\mu_{\ell,2}}\bigg)^2\sum_{i,j:\, i\ne j,\, i,j\ne\ell}{\rm Var}(Z_{i,\ell}){\rm Var}(Z_{\ell,j}){\rm Var}(Z_{i,j})\,. \tag{3.10}$$

Theorem 1. *Let Condition 1 hold and* $\{(\alpha_{i,j},\beta_{i,j})\}_{(i,j)\in\mathcal{I}}\in\mathcal{M}(\gamma;C_1,C_2)$ *for two fixed constants* $C_1\in(0,0.5)$ *and* $C_2>1$. *For any fixed integer* $s\ge 1$, *let* $1\le\ell_1<\cdots<\ell_s\le p$. *Then as* $p\to\infty$, *the following three assertions hold.*

(a) *For* $\gamma\gg p^{-1/4}$, $(p-1)^{1/2}\mathrm{diag}(b_{\ell_1}^{-1/2},\ldots,b_{\ell_s}^{-1/2})(\hat\theta_{\ell_1}-\theta_{\ell_1},\ldots,\hat\theta_{\ell_s}-\theta_{\ell_s})^{\mathrm{T}}\to\mathcal{N}(\mathbf{0},\mathbf{I}_s)$ *in distribution.*

(b) *For* $p^{-1/4}\gg\gamma\gg p^{-1/3}\log^{1/6}p$, $N^{1/2}\,\mathrm{diag}(\tilde b_{\ell_1}^{-1/2},\ldots,\tilde b_{\ell_s}^{-1/2})(\hat\theta_{\ell_1}-\theta_{\ell_1},\ldots,\hat\theta_{\ell_s}-\theta_{\ell_s})^{\mathrm{T}}\to\mathcal{N}(\mathbf{0},\mathbf{I}_s)$ *in distribution.*

(c) *For* $\gamma\asymp p^{-1/4}$, $N^{1/2}\,\mathrm{diag}[\{(p-2)b_{\ell_1}+\tilde b_{\ell_1}\}^{-1/2},\ldots,\{(p-2)b_{\ell_s}+\tilde b_{\ell_s}\}^{-1/2}](\hat\theta_{\ell_1}-\theta_{\ell_1},\ldots,\hat\theta_{\ell_s}-\theta_{\ell_s})^{\mathrm{T}}\to\mathcal{N}(\mathbf{0},\mathbf{I}_s)$ *in distribution.*

Remark 2. (a) Theorem 1 presents the asymptotic normality of the proposed estimators when $p\to\infty$ and also, possibly, $\pi\asymp\gamma\to 0$. It can be shown that $b_\ell\asymp\gamma^{-2}$ and $\tilde b_\ell\asymp\gamma^{-6}$. The limiting distribution depends on the relative rates of p and γ intimately; yielding an interesting phase transition phenomenon in the convergence rate. More precisely, when $\gamma\gg p^{-1/4}$ (including the case γ is a fixed constant), we have $|\hat\theta_\ell-\theta_\ell|=O_{\mathrm{p}}(p^{-1/2}\gamma^{-1})$. On the other hand, $|\hat\theta_\ell-\theta_\ell|=O_{\mathrm{p}}(p^{-1/4})$ when $\gamma\asymp p^{-1/4}$, and $O_{\mathrm{p}}(p^{-1}\gamma^{-3})$ when $p^{-1/4}\gg\gamma\gg p^{-1/3}\log^{1/6}p$.

(b) The result for the estimator with the original network $\mathbf{X}$ can be obtained by setting $\gamma=1$ (i.e. $\alpha_{i,j}\equiv 0\equiv\beta_{i,j}$). Then by Theorem 1(a), $p^{1/2}b_\ell^{-1/2}(\hat\theta_\ell-\theta_\ell)\to\mathcal{N}(0,1)$ in distribution, where

$$b_\ell=\frac{1}{p-1}\sum_{i:\,i\neq\ell}\bigg[\frac{1}{p-2}\sum_{j:\,j\neq\ell,i}\frac{\mu_{\ell,1}^{-1}\exp(\theta_\ell+\theta_j)+\mu_{\ell,2}^{-1}\exp(\theta_i+\theta_j)}{\{1+\exp(\theta_\ell+\theta_j)\}\{1+\exp(\theta_i+\theta_j)\}}\bigg]^2\frac{\exp(\theta_i+\theta_\ell)}{\{1+\exp(\theta_i+\theta_\ell)\}^2}\,.$$

(c) To construct confidence intervals for θ_ℓ based on Theorem 1, we would have to overcome two obstacles: (i) to identify the most appropriate phase in terms of relative sizes between γ and p, and (ii) to estimate b_ℓ and $\tilde b_\ell$ which determine the asymptotic variances. For (ii), we give their estimates in the Appendix. Unfortunately (i) is extremely difficult if not impossible, as in practice we only have one γ and one p. Proposition 5 in the Appendix shows that (ii) is only partially attainable, as, for example, b_ℓ cannot be estimated consistently when $p^{-1/4}\lesssim\gamma\lesssim p^{-1/4}\log^{1/4}p$. A novel bootstrap adaptive procedure will be developed in Section 4, which provides a uniform inference procedure when $\gamma\asymp\pi\to 0$ across the three different phases. The inference with γ being a fixed constant can be obtained based on Theorem 1 with the estimated $\hat b_\ell$ specified in the Appendix.

4 Bootstrap adaptive inference

The goal of this section is primarily two-fold: First we construct a novel bootstrap confidence interval for θ_ℓ which is automatically adaptive to the three phases identified in Theorem 1. Second, we leverage the new bootstrap procedure with Gaussian approximation via Stein's method to provide simultaneous inference for all p components of $\boldsymbol{\theta}$ as $p\to\infty$. Additionally, we provide an algorithm for data-adaptive selection of a working parameter in our approach. In the sequel, we always assume that the privacy level $\pi\to 0$ together with the number of nodes $p\to\infty$.

4.1 Bootstrap algorithm and simultaneous inference

Recall $N=(p-1)(p-2)$. As $\gamma\asymp\pi\to 0$, the three asymptotic assertions in Theorem 1 admit a uniform representation:

$$N^{1/2}\,\mathrm{diag}(\nu_{\ell_1}^{-1/2},\ldots,\nu_{\ell_s}^{-1/2})(\hat\theta_{\ell_1}-\theta_{\ell_1},\ldots,\hat\theta_{\ell_s}-\theta_{\ell_s})^{\mathrm{T}}\to\mathcal{N}(\mathbf{0},\mathbf{I}_s)\tag{4.1}$$

in distribution, where $\nu_\ell = (p-2)b_\ell + \tilde{b}_\ell$. Note that $b_\ell \asymp \gamma^{-2}$ and $\tilde{b}_\ell \asymp \gamma^{-6}$. Hence $(p-2)b_\ell/\nu_\ell \to 1$ when $\gamma \gg p^{-1/4}$, and $\tilde{b}_\ell/\nu_\ell \to 1$ when $\gamma \ll p^{-1/4}$. Now we reproduce this structure in a bootstrap world based on the available network $\mathbf{Z}$. The goal is to estimate ν_ℓ adaptively regardless of the decay rate of γ.

Recall $\mathcal{I} = \{(i,j) : 1 \le i < j \le p\}$. For a given constant $\delta \in (0, 0.5)$, we draw bootstrap samples $\mathbf{Z}^\dagger = (Z^\dagger_{i,j})_{p\times p}$ according to

$$Z^\dagger_{i,j} \equiv Z^\dagger_{j,i} = Z_{i,j} I(\eta_{i,j} = 0) + I(\eta_{i,j} = 1)\,, \quad (i,j) \in \mathcal{I}\,, \tag{4.2}$$

where $\{\eta_{i,j}\}_{(i,j)\in\mathcal{I}}$ are independent and identically distributed random variables with $\mathbb{P}(\eta_{i,j} = 0) = 1-2\delta$, $\mathbb{P}(\eta_{i,j} = 1) = \delta$ and $\mathbb{P}(\eta_{i,j} = -1) = \delta$. For $i \neq j$ and $\tau \in \{0,1\}$, put

$$\varphi^\dagger_{(i,j),\tau}(x) = \{x - \delta - \alpha_{i,j}(1-2\delta)\}^\tau \{1 - \delta - \beta_{i,j}(1-2\delta) - x\}^{1-\tau}\,.$$

To simplify the notation, we write $\varphi^\dagger_{(i,j),\tau}(Z^\dagger_{i,j})$ as $\varphi^\dagger_{(i,j),\tau}$ for any $i \neq j$ and $\tau \in \{0,1\}$. Note that $\mathbb{P}(X_{i,j} = 0) = \mathbb{E}\{\varphi^\dagger_{(i,j),0}\}(1-2\delta)^{-1}(1-\alpha_{i,j}-\beta_{i,j})^{-1}$ and $\mathbb{P}(X_{i,j} = 1) = \mathbb{E}\{\varphi^\dagger_{(i,j),1}\}(1-2\delta)^{-1}(1-\alpha_{i,j}-\beta_{i,j})^{-1}$. As $Z^\dagger_{i,j}$ is independent of $\{Z^\dagger_{\tilde{i},\tilde{j}} : |\{\tilde{i},\tilde{j}\} \cap \{i,j\}| \le 1\}$, it follows from (3.3) that

$$\frac{\mathbb{E}\{\varphi^\dagger_{(i,\ell),1}\varphi^\dagger_{(i,j),0}\varphi^\dagger_{(\ell,j),1}\}}{\mathbb{E}\{\varphi^\dagger_{(i,\ell),0}\varphi^\dagger_{(i,j),1}\varphi^\dagger_{(\ell,j),0}\}} = \exp(2\theta_\ell)\,, \quad i \neq j \neq \ell\,, \tag{4.3}$$

which is a bootstrap analogue of (3.4). Similarly, we define a bootstrap estimator for θ_ℓ as:

$$\hat{\theta}^\dagger_\ell = \frac{1}{2}\log\left(\frac{\hat{\mu}^\dagger_{\ell,1}}{\hat{\mu}^\dagger_{\ell,2}}\right), \tag{4.4}$$

where $\hat{\mu}^\dagger_{\ell,1} = |\mathcal{H}_\ell|^{-1}\sum_{(i,j)\in\mathcal{H}_\ell}\varphi^\dagger_{(i,\ell),1}\varphi^\dagger_{(i,j),0}\varphi^\dagger_{(\ell,j),1}$ and $\hat{\mu}^\dagger_{\ell,2} = |\mathcal{H}_\ell|^{-1}\sum_{(i,j)\in\mathcal{H}_\ell}\varphi^\dagger_{(i,\ell),0}\varphi^\dagger_{(i,j),1}\varphi^\dagger_{(\ell,j),0}$. We also define the bootstrap analogues of $\mu_{\ell,1}$, $\mu_{\ell,2}$ and $\lambda_{i,\ell}$ as

$$\mu^\dagger_{\ell,1} = \frac{1}{|\mathcal{H}_\ell|}\sum_{(i,j)\in\mathcal{H}_\ell}\mathbb{E}\{\varphi^\dagger_{(i,\ell),1}\varphi^\dagger_{(i,j),0}\varphi^\dagger_{(\ell,j),1}\}\,, \qquad \mu^\dagger_{\ell,2} = \frac{1}{|\mathcal{H}_\ell|}\sum_{(i,j)\in\mathcal{H}_\ell}\mathbb{E}\{\varphi^\dagger_{(i,\ell),0}\varphi^\dagger_{(i,j),1}\varphi^\dagger_{(\ell,j),0}\}\,,$$
$$\lambda^\dagger_{i,\ell} = \frac{1}{p-2}\sum_{j:\, j\neq \ell, i}\left[\frac{1}{\mu^\dagger_{\ell,1}}\mathbb{E}\{\varphi^\dagger_{(\ell,j),1}\}\mathbb{E}\{\varphi^\dagger_{(i,j),0}\} + \frac{1}{\mu^\dagger_{\ell,2}}\mathbb{E}\{\varphi^\dagger_{(\ell,j),0}\}\mathbb{E}\{\varphi^\dagger_{(i,j),1}\}\right].$$

Then $\hat{\theta}^\dagger_\ell$ admits a similar asymptotic property as (4.1). To present it explicitly, we let

$$\nu^\dagger_\ell = (p-2)b^\dagger_\ell + \tilde{b}^\dagger_\ell\,, \quad \ell \in [p]\,, \tag{4.5}$$

where

$$b^\dagger_\ell = \frac{1}{p-1}\sum_{i:\, i\neq\ell}\lambda^{\dagger,2}_{i,\ell}\mathrm{Var}(Z^\dagger_{i,\ell})\,,$$
$$\tilde{b}^\dagger_\ell = \frac{1}{2N}\left(\frac{\mu^\dagger_{\ell,1}+\mu^\dagger_{\ell,2}}{\mu^\dagger_{\ell,1}\mu^\dagger_{\ell,2}}\right)^2\sum_{i,j:\, i\neq j,\, i,j\neq\ell}\mathrm{Var}(Z^\dagger_{i,\ell})\mathrm{Var}(Z^\dagger_{\ell,j})\mathrm{Var}(Z^\dagger_{i,j})\,.$$

Theorem 2. *Let the conditions of Theorem 1 hold, and $\delta \in (0, c]$ for some constant $c < 0.5$. As $p \to \infty$, $\gamma \to 0$ and $\gamma \gg p^{-1/3} \log^{1/6} p$, it holds that* (a) $N^{1/2} \operatorname{diag}(\nu_{\ell_1}^{\dagger,-1/2}, \ldots, \nu_{\ell_s}^{\dagger,-1/2})(\hat{\theta}_{\ell_1}^{\dagger} - \theta_{\ell_1}, \ldots, \hat{\theta}_{\ell_s}^{\dagger} - \theta_{\ell_s})^{\mathrm{T}} \to \mathcal{N}(\mathbf{0}, \mathbf{I}_s)$ *in distribution for any fixed integer $s \geq 1$ and $1 \leq \ell_1 < \cdots < \ell_s \leq p$, and* (b) $\max_{\ell \in [p]} |\nu_{\ell}^{\dagger} \nu_{\ell}^{-1} - 1| = O(\delta)$, *where ν_ℓ is specified in* (4.1).

Theorem 2 indicates that $\nu_{\ell}^{\dagger}/\nu_\ell \to 1$ for any $\gamma \gg p^{-1/3} \log^{1/6} p$ provided that we set $\delta = o(1)$. For fixed $s \geq 1$ and given $1 \leq \ell_1 < \cdots < \ell_s \leq p$, we can draw bootstrap samples $\mathbf{Z}^{\dagger}$ as in (4.2) with some $\delta = o(1)$, and compute the bootstrap estimate $(\hat{\theta}_{\ell_1}^{\dagger}, \ldots, \hat{\theta}_{\ell_s}^{\dagger})^{\mathrm{T}}$ defined in (4.4) based on $\mathbf{Z}^{\dagger}$. We repeat this procedure M times for some large integer M and compute $\hat{\nu}_{\ell_k}^{\dagger} = NM^{-1} \sum_{m=1}^{M} \{\hat{\theta}_{\ell_k}^{\dagger,(m)} - \bar{\hat{\theta}}_{\ell_k}^{\dagger}\}^2$ with $\bar{\hat{\theta}}_{\ell_k}^{\dagger} = M^{-1} \sum_{m=1}^{M} \hat{\theta}_{\ell_k}^{\dagger,(m)}$ for each $k \in [s]$, where $\{\hat{\theta}_{\ell_1}^{\dagger,(m)}, \ldots, \hat{\theta}_{\ell_s}^{\dagger,(m)}\}^{\mathrm{T}}$ is the associated bootstrap estimate in the m-th repetition. Then a confidence region for $(\theta_{\ell_1}, \ldots, \theta_{\ell_s})^{\mathrm{T}}$ can be constructed based on the asymptotic approximation $N^{1/2} \operatorname{diag}(\hat{\nu}_{\ell_1}^{\dagger,-1/2}, \ldots, \hat{\nu}_{\ell_s}^{\dagger,-1/2})(\hat{\theta}_{\ell_1} - \theta_{\ell_1}, \ldots, \hat{\theta}_{\ell_s} - \theta_{\ell_s})^{\mathrm{T}} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_s)$.

Importantly, we note that in both Theorems 1 and 2, s is fixed when $p \to \infty$. Hence the inference methods presented so far are not applicable to all p components of $\boldsymbol{\theta}$ simultaneously. However, a breakthrough can be had via the Gaussian approximation in Theorem 3 below. To our best knowledge, this is the first method for simultaneous inference for all the p components of $\boldsymbol{\theta}$ in the β-model. Write $\hat{\boldsymbol{\theta}} = (\hat{\theta}_1, \ldots, \hat{\theta}_p)^{\mathrm{T}}$ where $\hat{\theta}_\ell$ is the proposed moment-based estimator given in (3.5) based on the sanitized data $\mathbf{Z}$. As shown in Proposition 4, the leading term of $\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}$ cannot be formulated as a partial sum of independent (or weakly dependent) random vectors, which is different from the standard framework of Gaussian approximation (Chernozhukov, Chetverikov and Kato, 2013; Chang, Chen and Wu, 2021). Hence the existing results of Gaussian approximation cannot be applied directly, which requires significant technical challenge to be overcome in our theoretical analysis. We construct Theorem 3 by Stein's method.

Theorem 3. *Let Condition 1 hold and $\{(\alpha_{i,j}, \beta_{i,j})\}_{(i,j) \in \mathcal{I}} \in \mathcal{M}(\gamma; C_1, C_2)$ for two fixed constants $C_1 \in (0, 0.5)$ and $C_2 > 1$. As $p \to \infty$, let $0 < \delta \ll (p \log p)^{-1}$, $\gamma \to 0$ and $\gamma \gg p^{-1/3} \log^{1/2} p$. Then it holds that $\sup_{\mathbf{u} \in \mathbb{R}^p} |\mathbb{P}\{N^{1/2} (\mathbf{V}^{\dagger})^{-1/2} (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \leq \mathbf{u}\} - \mathbb{P}(\boldsymbol{\xi} \leq \mathbf{u})| \to 0$, where $\mathbf{V}^{\dagger} = \operatorname{diag}(\nu_1^{\dagger}, \ldots, \nu_p^{\dagger})$, and $\boldsymbol{\xi} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_p)$.*

Write $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_p)^{\mathrm{T}} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_p)$. For any $\mathcal{S} = \{\ell_1, \ldots, \ell_s\} \subset [p]$, let $\mathbf{V}_{\mathcal{S}}^{\dagger} = \operatorname{diag}(\nu_{\ell_1}^{\dagger}, \ldots, \nu_{\ell_s}^{\dagger})$, $\hat{\boldsymbol{\theta}}_{\mathcal{S}} = (\hat{\theta}_{\ell_1}, \ldots, \hat{\theta}_{\ell_s})^{\mathrm{T}}$, $\boldsymbol{\theta}_{\mathcal{S}} = (\theta_{\ell_1}, \ldots, \theta_{\ell_s})^{\mathrm{T}}$ and $\boldsymbol{\xi}_{\mathcal{S}} = (\xi_{\ell_1}, \ldots, \xi_{\ell_s})^{\mathrm{T}}$. Following the arguments in the proof of Proposition 1 in the supplementary material of Chang et al. (2017), Theorem 3 implies that

$$\sup_{\mathcal{S}} \sup_{u \in \mathbb{R}} \big|\mathbb{P}\{N^{1/2} |(\mathbf{V}_{\mathcal{S}}^{\dagger})^{-1/2} (\hat{\boldsymbol{\theta}}_{\mathcal{S}} - \boldsymbol{\theta}_{\mathcal{S}})|_\infty \leq u\} - \mathbb{P}(|\boldsymbol{\xi}_{\mathcal{S}}|_\infty \leq u)\big| \to 0$$

as $p \to \infty$. Given $\alpha \in (0, 1)$ and $\mathcal{S} \subset [p]$,

$$\boldsymbol{\Theta}_{\mathcal{S},\alpha} := \left\{\mathbf{a} \in \mathbb{R}^{|\mathcal{S}|} : N^{1/2} |(\mathbf{V}_{\mathcal{S}}^{\dagger})^{-1/2} (\hat{\boldsymbol{\theta}}_{\mathcal{S}} - \mathbf{a})|_\infty \leq \Phi^{-1}\left(\frac{1 + \alpha^{1/|\mathcal{S}|}}{2}\right)\right\} \tag{4.6}$$

is a $100 \cdot \alpha\%$ confidence region for $\boldsymbol{\theta}_{\mathcal{S}}$. We refer to Section 4 of Chang et al. (2018) for applications of this type of confidence region in simultaneous inference. If γ is a fixed constant, Theorem 3 still holds with replacing $\mathbf{V}^{\dagger}$ by $(p-2) \cdot \operatorname{diag}(\hat{b}_1, \ldots, \hat{b}_p)$ where $\hat{b}_\ell$ is given in (I.3) in the Appendix. If we set $\alpha_{i,j} \equiv 0 \equiv \beta_{i,j}$ in the jittering mechanism (2.3)–(2.5), then $\gamma = 1$ in this case and the released data $\mathbf{Z}$ is identical to the original data $\mathbf{X}$. Our simultaneous inference procedure also works in this case.

4.2 Adaptive selection of δ

The tuning parameter δ plays a key role in our simultaneous inference procedure. We propose a data-driven method in Algorithm 1 to select δ. To illustrate the basic idea, we denote by $\nu_\ell^\dagger(\delta)$ the associated $\nu_\ell^\dagger$ defined in (4.5) with δ used in generating the bootstrap samples $\mathbf{Z}^\dagger$ in (4.2). If $\{\nu_\ell\}_{\ell\in\mathcal{S}}$ are known, the ideal selection for the tuning parameter δ should be $\delta_{\rm opt} = \arg\min_\delta \max_{\ell\in\mathcal{S}} |\nu_\ell^\dagger(\delta) - \nu_\ell|$. Unfortunately, $\{\nu_\ell\}_{\ell\in\mathcal{S}}$ are unknown in practice, as they depend on the unknown parameters $\theta_1,\ldots,\theta_p$. A natural idea is to replace the ν_ℓ's by their estimates. Recall $\hat\theta_\ell = 2^{-1}\log(\hat\mu_{\ell,1}\hat\mu_{\ell,2}^{-1})$ with $\hat\mu_{\ell,1} = |\mathcal{H}_\ell|^{-1}\sum_{(i,j)\in\mathcal{H}_\ell}\varphi_{(i,\ell),1}\varphi_{(i,j),0}\varphi_{(\ell,j),1}$ and $\hat\mu_{\ell,2} = |\mathcal{H}_\ell|^{-1}\sum_{(i,j)\in\mathcal{H}_\ell}\varphi_{(i,\ell),0}\varphi_{(i,j),1}\varphi_{(\ell,j),0}$. Due to the nonlinear function $\log(\cdot)$ and the ratio between $\hat\mu_{\ell,1}$ and $\hat\mu_{\ell,2}$, $\hat\theta_\ell$ usually includes some high-order bias term. More specifically,

$$\hat\theta_\ell - \theta_\ell = \frac{\hat\mu_{\ell,1}-\mu_{\ell,1}}{2\mu_{\ell,1}} - \frac{\hat\mu_{\ell,2}-\mu_{\ell,2}}{2\mu_{\ell,2}} + \underbrace{\frac{(\hat\mu_{\ell,2}-\mu_{\ell,2})^2}{4\mu_{\ell,2}^2} - \frac{(\hat\mu_{\ell,1}-\mu_{\ell,1})^2}{4\mu_{\ell,1}^2}}_{\text{high-order bias}} + \hat R_\ell\,,$$

where $\hat R_\ell$ is a negligible term in comparison to the high-order bias. Although the high-order bias has little impact on the estimation of θ_ℓ, it may lead to a bad estimate of ν_ℓ if we just plug-in $\hat\theta_1,\ldots,\hat\theta_p$ in the nonlinear function ν_ℓ which depends on $\theta_1,\ldots,\theta_p$. Hence, when we replace $\{\nu_\ell\}_{\ell\in\mathcal{S}}$ in Algorithm 1, we use their associated estimates with bias-corrected $\hat\theta_1^{\rm bc},\ldots,\hat\theta_p^{\rm bc}$. Based on the optimal $\delta_{\rm opt}$ selected in Algorithm 1, we can replace the values $\nu_{\ell_1}^\dagger,\ldots,\nu_{\ell_s}^\dagger$ in (4.6) by $\hat\nu_{\ell_1}^\dagger(\delta_{\rm opt}),\ldots,\hat\nu_{\ell_s}^\dagger(\delta_{\rm opt})$ specified in Algorithm 1 to construct a $100\cdot\alpha\%$ simultaneous confidence region for $\boldsymbol{\theta}_\mathcal{S}$ in practice.

5 Numerical study

5.1 Simulation

In this section we illustrate the finite sample properties of our proposed method of estimation and inference for the unknown parameters in the β-model by simulation. For a given original network $\mathbf{X}$, we always set $(\alpha_{i,j},\beta_{i,j}) \equiv (\alpha,\beta)$ for any $i,j\in[p]$ and $i\neq j$ in the data release mechanism (2.3) and (2.4) to generate $\mathbf{Z}$. In our simulation, we set $\alpha=\beta\in\{0,0.1,0.2,0.3\}$ and $p\in\{1000,2000\}$. Note that $\mathbf{Z}=\mathbf{X}$ when $\alpha=\beta=0$.

We draw $\theta_1,\ldots,\theta_p$ independently from $\mathcal{N}(0,0.2)$, and then generate the adjacency matrix $\mathbf{X}$ according to β-model (3.1). Based on the released data $\mathbf{Z}$, we applied the moment-based method (3.5) to estimate $\boldsymbol{\theta}=(\theta_1,\ldots,\theta_p)^{\rm T}$, and then calculated the estimation error $L(\hat{\boldsymbol{\theta}}) = p^{-1}|\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}|_2^2$. We also compared our estimator with an adaptation of the maximum likelihood estimator (MLE) of Karwa and Slavković (2016), by correcting the degree sequence $\mathbf{U}_X$ as $(1-\alpha-\beta)^{-1}\{\mathbf{U}_Z-(p-1)\alpha\}$ and then applying the algorithm developed in that same paper. Table 1 reports the averages, medians and standard deviations of the estimation errors over 500 replications. The proposed moment-based estimation performed competitively in relation to the MLE, though the MLE is slightly more accurate overall. However the MLE method is memory-demanding when p is large. For example with $p=1000$ and $\alpha=\beta=0.1$, the step generating a graph with given degree sequence (i.e. Algorithm 2 of Karwa and Slavković (2016)) occupied 3.91 GB memory. In contrast, the newly proposed moment-based estimation only used 38.19 MB memory. Furthermore, the MLE is excessively time-consuming computationally when p is large. See Table 1 for the recorded average CPU times for each realization on an Intel(R) Xeon(R) Platinum 8160 processor

Algorithm 1 Selecting tuning parameter δ

Obtain $\{\hat{\theta}_\ell\}_{\ell=1}^p$, $\{\hat{\mu}_{\ell,1}\}_{\ell=1}^p$ and $\{\hat{\mu}_{\ell,2}\}_{\ell=1}^p$ based on (3.5), (3.6) and (3.7), respectively.

Calculate

$$\hat{\varphi}_{(i,j,\ell),1} = \frac{(1-\alpha_{i,\ell}-\beta_{i,\ell})\exp(\hat{\theta}_i+\hat{\theta}_\ell)}{1+\exp(\hat{\theta}_i+\hat{\theta}_\ell)} \frac{(1-\alpha_{i,j}-\beta_{i,j})}{1+\exp(\hat{\theta}_i+\hat{\theta}_j)} \frac{(1-\alpha_{\ell,j}-\beta_{\ell,j})\exp(\hat{\theta}_\ell+\hat{\theta}_j)}{1+\exp(\hat{\theta}_\ell+\hat{\theta}_j)},$$

$$\hat{\varphi}_{(i,j,\ell),2} = \frac{(1-\alpha_{i,\ell}-\beta_{i,\ell})}{1+\exp(\hat{\theta}_i+\hat{\theta}_\ell)} \frac{(1-\alpha_{i,j}-\beta_{i,j})\exp(\hat{\theta}_i+\hat{\theta}_j)}{1+\exp(\hat{\theta}_i+\hat{\theta}_j)} \frac{(1-\alpha_{\ell,j}-\beta_{\ell,j})}{1+\exp(\hat{\theta}_\ell+\hat{\theta}_j)},$$

repeat

leave out one $(i,j) \in \mathcal{H}_\ell$ randomly and denote by $\mathcal{H}_\ell^-$ the set including the rest elements in $\mathcal{H}_\ell$

calculate $\tilde{\mu}_{\ell,1} = |\mathcal{H}_\ell^-|^{-1}\sum_{(i,j)\in\mathcal{H}_\ell^-}\hat{\varphi}_{(i,j,\ell),1}$ and $\tilde{\mu}_{\ell,2} = |\mathcal{H}_\ell^-|^{-1}\sum_{(i,j)\in\mathcal{H}_\ell^-}\hat{\varphi}_{(i,j,\ell),2}$, which provide the estimates of $\mu_{\ell,1}$ and $\mu_{\ell,2}$, respectively

calculate $\text{bias}_\ell = 4^{-1}\tilde{\mu}_{\ell,2}^{-2}(\hat{\mu}_{\ell,2}-\tilde{\mu}_{\ell,2})^2 - 4^{-1}\tilde{\mu}_{\ell,1}^{-2}(\hat{\mu}_{\ell,1}-\tilde{\mu}_{\ell,1})^2$

until M replicates obtained, for a large integer M, and get $\text{bias}_\ell^{(1)},\ldots,\text{bias}_\ell^{(M)}$

approximate the high-order bias in $\hat{\theta}_\ell$ by $\widehat{\text{bias}}_\ell = M^{-1}\sum_{m=1}^M \text{bias}_\ell^{(m)}$, and obtain $\hat{\theta}_\ell^{\text{bc}} = \hat{\theta}_\ell - \widehat{\text{bias}}_\ell$, the bias-correction for $\hat{\theta}_\ell$

calculate $\tilde{\mu}_{\ell,1}^{\text{bc}} = |\mathcal{H}_\ell|^{-1}\sum_{(i,j)\in\mathcal{H}_\ell}\tilde{\varphi}_{(i,j,\ell),1}$ and $\tilde{\mu}_{\ell,2}^{\text{bc}} = |\mathcal{H}_\ell|^{-1}\sum_{(i,j)\in\mathcal{H}_\ell}\tilde{\varphi}_{(i,j,\ell),2}$, where $\tilde{\varphi}_{(i,j,\ell),1}$ and $\tilde{\varphi}_{(i,j,\ell),2}$ are defined in the same manner as $\hat{\varphi}_{(i,j,\ell),1}$ and $\hat{\varphi}_{(i,j,\ell),2}$, respectively, with replacing $\{\hat{\theta}_\ell\}_{\ell=1}^p$ by $\{\hat{\theta}_\ell^{\text{bc}}\}_{\ell=1}^p$

calculate $\hat{\nu}_\ell^{\text{bc}} = (p-2)\hat{b}_\ell^{\text{bc}} + \hat{\bar{b}}_\ell^{\text{bc}}$, where $\hat{b}_\ell^{\text{bc}}$ and $\hat{\bar{b}}_\ell^{\text{bc}}$ are defined in the same manner of $\hat{b}_\ell$ and $\hat{\bar{b}}_\ell$ specified as (I.3) in the Appendix with replacing $(\hat{\mu}_{\ell,1},\hat{\mu}_{\ell,2},\{\hat{\theta}_k\}_{k=1}^p)$ by $(\tilde{\mu}_{\ell,1}^{\text{bc}},\tilde{\mu}_{\ell,2}^{\text{bc}},\{\hat{\theta}_k^{\text{bc}}\}_{k=1}^p)$

repeat

given $\delta > 0$ and draw bootstrap samples $\mathbf{Z}^\dagger = (Z_{i,j}^\dagger)_{p\times p}$ as in (4.2), calculate the bootstrap estimate $\hat{\theta}_\ell^\dagger$ defined in (4.4) based on the bootstrap samples $\mathbf{Z}^\dagger$

until M replicates obtained, for a large integer M, and get $\hat{\theta}_\ell^{\dagger,(1)},\ldots,\hat{\theta}_\ell^{\dagger,(M)}$

calculate $\hat{\nu}_\ell^\dagger(\delta) = p^2M^{-1}\sum_{m=1}^M\{\hat{\theta}_\ell^{\dagger,(m)} - \bar{\hat{\theta}}_\ell^\dagger\}^2$ with $\bar{\hat{\theta}}_\ell^\dagger = M^{-1}\sum_{m=1}^M\hat{\theta}_\ell^{\dagger,(m)}$

select $\delta_{\text{opt}} = \arg\min_\delta \max_{\ell\in\mathcal{S}}|\hat{\nu}_\ell^\dagger(\delta) - \hat{\nu}_\ell^{\text{bc}}|$

(2.10GHz). With $p = 1000$, the average required CPU time for computing the MLE once is over 471 minutes with the original data $\mathbf{X}$ (i.e. $\alpha = \beta = 0$) and is almost double with the sanitized data $\mathbf{Z}$ (i.e. $\alpha, \beta > 0$). It is practically infeasible to conduct the simulation (with replications) for all scenarios with $p = 2000$, for which we only report the results with $\alpha = \beta = 0$ with the average CPU time 5095 minutes per estimation.

Based on our moment-based estimator $\hat{\boldsymbol{\theta}}$, we also constructed the simultaneous confidence regions (4.6) for all the p components $\theta_1,\ldots,\theta_p$. To determine the tuning parameter δ, we applied the data-driven Algorithm 1 with $M = 500$. Table 2 lists the relative frequencies, in 500 replications for each settings, of the occurrence of the event that the constructed confidence region contains the true value of $\boldsymbol{\theta}$. At each of the three nominal levels, those relative frequencies are always close to the corresponding nominal level.

5.2 Real data analysis

Facebook, a social networking site launched in February 2004, now overwhelms numerous aspects of everyday life, and has become an immensely popular societal obsession. The Facebook friendships define a network of undirected edges that connect individual users. In this section, we analyze a small

Table 1: Estimation errors of the proposed moment-based estimation and the maximum likelihood estimation for $\boldsymbol{\theta}$ in the β-model (3.1). Also reported are the average CPU times (in minutes) for completing the estimation once for each of the two methods.

		Proposed method				Maximum likelihood estimation			
		$\alpha = \beta$				$\alpha = \beta$			
p	Summary statistics	0	0.1	0.2	0.3	0	0.1	0.2	0.3
1000	Average	0.0041	0.0065	0.0118	0.0274	0.0062	0.0065	0.0116	0.0272
	Median	0.0041	0.0065	0.0117	0.0274	0.0041	0.0064	0.0116	0.0267
	Standard deviation	0.0002	0.0003	0.0006	0.0012	0.0085	0.0003	0.0005	0.0026
	Time (min)	1.0340	1.0439	0.9191	0.8540	471.6290	881.2221	772.3427	774.1594
2000	Average	0.0020	0.0032	0.0058	0.0133	0.0058	NA	NA	NA
	Median	0.0020	0.0032	0.0058	0.0133	0.0043	NA	NA	NA
	Standard deviation	0.0001	0.0001	0.0002	0.0004	0.0019	NA	NA	NA
	Time (min)	4.2333	4.8707	3.7540	3.7256	5095.0520	NA	NA	NA

Table 2: Empirical frequencies of the constructed simultaneous confidence regions for $\boldsymbol{\theta}$ covering the truth in the β-model (3.1).

p	Level	$\alpha = \beta = 0$	$\alpha = \beta = 0.1$	$\alpha = \beta = 0.2$	$\alpha = \beta = 0.3$
1000	90%	0.876	0.868	0.910	0.888
	95%	0.932	0.928	0.958	0.948
	99%	0.984	0.982	0.982	0.992
2000	90%	0.900	0.876	0.898	0.896
	95%	0.950	0.956	0.946	0.952
	99%	0.988	0.990	0.996	0.992

Facebook friendship network dataset available at `http://wwwlovre.appspot.com/support.jsp`. The network consists of 334 nodes and 2218 edges.

We fit the β-model to this network. As an illustration on the impact of the 'jittering', we identify the nodes with the associated parameters equal to 0 based on both the original network and some sanitized versions. More specifically, we first consider the multiple hypothesis tests:

$$H_{0,\ell} : \theta_\ell = 0 \quad \text{versus} \quad H_{1,\ell} : \theta_\ell \neq 0$$

for $1 \leq \ell \leq 334$. The moment-based estimate $\hat{\boldsymbol{\theta}} = (\hat{\theta}_1, \ldots, \hat{\theta}_{334})^{\mathrm{T}}$ based on the original data $\mathbf{X}$ is calculated according to (3.5). Theorem 1 indicates that the p-value for the ℓ-th test is given by $2\{1 - \Phi(\sqrt{333}\hat{b}_\ell^{-1/2}|\hat{\theta}_\ell|)\}$ with $\hat{b}_\ell$ defined as in (I.3). Note that $\hat{\theta}_{\ell_1}$ and $\hat{\theta}_{\ell_2}$ are asymptotically independent for any $\ell_1 \neq \ell_2$. The BH procedure (Benjamini et al., 1995) at the rate 1% for the 334 multiple tests identifies the 10 nodal parameters $(\theta_2, \theta_{21}, \theta_{33}, \theta_{51}, \theta_{78}, \theta_{186}, \theta_{202}, \theta_{211}, \theta_{263}, \theta_{272})$ being not significantly different from 0. Put $\mathcal{S} = \{2, 21, 33, 51, 78, 186, 202, 211, 263, 272\}$. We consider now the testing for the single hypothesis setting

$$H_0 : \boldsymbol{\theta}_{\mathcal{S}} = \mathbf{0} \quad \text{versus} \quad H_1 : \boldsymbol{\theta}_{\mathcal{S}} \neq \mathbf{0} \tag{5.1}$$

based on both the original network $\mathbf{X}$ and its sanitized versions $\mathbf{Z}$ via jittering mechanism (2.3) with $\alpha = \beta = 0.1, 0.2$ and 0.3. Let $\boldsymbol{\zeta}_1, \ldots, \boldsymbol{\zeta}_{1000}$ be independent and $\mathcal{N}(\mathbf{0}, \mathbf{I}_{10})$. By Theorem 3, the p-value of the test for (5.1) based on $\mathbf{Z}$ is approximately $1000^{-1}\sum_{m=1}^{1000} I\{|\boldsymbol{\zeta}_m|_\infty \geq \sqrt{333 \times 332}|\widehat{\mathbf{V}}_{\mathcal{S}}^{-1/2}\hat{\boldsymbol{\theta}}_{\mathcal{S}}^{(\mathbf{Z})}|_\infty\}$, where $\hat{\boldsymbol{\theta}}_{\mathcal{S}}^{(\mathbf{Z})}$ is the estimate of $\boldsymbol{\theta}_{\mathcal{S}}$ based on $\mathbf{Z}$ by the moment-based method (3.5), and $\widehat{\mathbf{V}}_{\mathcal{S}}$ is the estimate of the asymptotic covariance of $\sqrt{333 \times 332}\{\hat{\boldsymbol{\theta}}_{\mathcal{S}}^{(\mathbf{Z})} - \boldsymbol{\theta}_{\mathcal{S}}\}$. When $\alpha = \beta = 0$, $\mathbf{Z} = \mathbf{X}$, the p-value for testing (5.1) based on $\mathbf{X}$ is then 0.1019. As the test based on $\mathbf{Z}$ depends on a particular realization when $\alpha = \beta = 0.1$, 0.2 and 0.3, we repeat the test 500 times for each setting. The average p-values of those 500 tests (based on $\mathbf{Z}$) with $\alpha = \beta = 0.1$, 0.2 and 0.3 are, respectively, 0.1276, 0.1522 and 0.1874, which are reasonably close to the p-value based on $\mathbf{X}$. The standard errors of the 500 p-values are 0.0795, 0.1281 and 0.1408, respectively, for $\alpha = \beta = 0.1, 0.2$ and 0.3.

This small illustration suggests that, with increasing edge noise (and hence increasing privacy), the resulting p-value is increasingly over-estimated with increasing standard error. Both trends are to be expected – since with increasing edge-noise the signal will be weakened – and merit future study.

Appendix

I Estimating the asymptotic variances in Theorem 1

If we know the decay rate of γ falls into which region, we can construct the confidence region of $(\theta_{\ell_1}, \ldots, \theta_{\ell_s})^{\mathrm{T}}$ based on Theorem 1 with the plug-in method. To do this, we need to estimate b_{ℓ_k}'s and $\tilde{b}_{\ell_k}$'s first. By (3.8), we can estimate $\lambda_{i,\ell}$ as

$$\hat{\lambda}_{i,\ell} = \frac{1}{p-2}\sum_{j:\, j\neq \ell, i}\bigg\{\frac{1}{\hat{\mu}_{\ell,1}}\varphi_{(\ell,j),1}\varphi_{(i,j),0} + \frac{1}{\hat{\mu}_{\ell,2}}\varphi_{(\ell,j),0}\varphi_{(i,j),1}\bigg\} \tag{I.1}$$

with $\hat{\mu}_{\ell,1}$ and $\hat{\mu}_{\ell,2}$ specified in (3.6) and (3.7), respectively. By the definition of $Z_{i,j}$, we have $\mathrm{Var}(Z_{i,j}) = \{\alpha_{i,j} + (1-\beta_{i,j})\exp(\theta_i+\theta_j)\}\{1-\alpha_{i,j}+\beta_{i,j}\exp(\theta_i+\theta_j)\}\{1+\exp(\theta_i+\theta_j)\}^{-2}$ for any $i \neq j$. We can estimate $\mathrm{Var}(Z_{i,j})$ as

$$\widehat{\mathrm{Var}}(Z_{i,j}) = \frac{\alpha_{i,j} + (1-\beta_{i,j})\exp(\hat{\theta}_i+\hat{\theta}_j)}{1+\exp(\hat{\theta}_i+\hat{\theta}_j)} \cdot \frac{1-\alpha_{i,j}+\beta_{i,j}\exp(\hat{\theta}_i+\hat{\theta}_j)}{1+\exp(\hat{\theta}_i+\hat{\theta}_j)}\,. \tag{I.2}$$

Based on (I.1) and (I.2), we can estimate b_ℓ and $\tilde{b}_\ell$, respectively, as

$$\hat{b}_\ell = \frac{1}{p-1}\sum_{i:\, i\neq \ell} \hat{\lambda}_{i,\ell}^2 \widehat{\mathrm{Var}}(Z_{i,\ell})\,, \tag{I.3}$$

$$\hat{\tilde{b}}_\ell = \frac{1}{2N}\bigg(\frac{\hat{\mu}_{\ell,1}+\hat{\mu}_{\ell,2}}{\hat{\mu}_{\ell,1}\hat{\mu}_{\ell,2}}\bigg)^2 \sum_{i,j:\, i\neq j,\, i,j\neq \ell} \widehat{\mathrm{Var}}(Z_{i,\ell})\widehat{\mathrm{Var}}(Z_{\ell,j})\widehat{\mathrm{Var}}(Z_{i,j})\,.$$

The convergence rates of such estimates are presented in Proposition 5.

Proposition 5. *Under Condition* 1 *and* $\{(\alpha_{i,j},\beta_{i,j})\}_{(i,j)\in\mathcal{I}} \in \mathcal{M}(\gamma; C_1, C_2)$ *for two fixed constants* $C_1 \in (0, 0.5)$ *and* $C_2 > 1$*, if* $\gamma \gg p^{-1/3}\log^{1/6} p$*, it holds that* $|\hat{b}_\ell {b_\ell}^{-1} - 1| = O_{\mathrm{p}}(\gamma^{-4}p^{-1}\log p) + O_{\mathrm{p}}(\gamma^{-2}p^{-1/2}\log^{1/2} p)$ *and* $|\hat{\tilde{b}}_\ell \tilde{b}_\ell^{-1} - 1| = O_{\mathrm{p}}(\gamma^{-3}p^{-1}\log^{1/2} p) + O_{\mathrm{p}}(\gamma^{-1}p^{-1/2}\log^{1/2} p)$ *for any given* $\ell \in [p]$.

For any fixed integer $s \geq 1$, Theorem 1 and Proposition 5 imply that $p^{1/2}\mathrm{diag}(\hat{b}_{\ell_1}^{-1/2},\ldots,\hat{b}_{\ell_s}^{-1/2})(\hat{\theta}_{\ell_1} - \theta_{\ell_1},\ldots,\hat{\theta}_{\ell_s} - \theta_{\ell_s})^{\mathrm{T}} \to_d \mathcal{N}(\mathbf{0}, \mathbf{I}_s)$ if $\gamma \gg p^{-1/4}\log^{1/4} p$, and $p\,\mathrm{diag}(\hat{\tilde{b}}_{\ell_1}^{-1/2},\ldots,\hat{\tilde{b}}_{\ell_s}^{-1/2})(\hat{\theta}_{\ell_1} - \theta_{\ell_1},\ldots,\hat{\theta}_{\ell_s} - \theta_{\ell_s})^{\mathrm{T}} \to_d \mathcal{N}(\mathbf{0}, \mathbf{I}_s)$ if $p^{-1/3}\log^{1/6} p \ll \gamma \ll p^{-1/4}$. Unfortunately, such plug-in method does not work in the scenario $p^{-1/4} \lesssim \gamma \lesssim p^{-1/4}\log^{1/4} p$ since $\hat{b}_\ell$ is no longer a valid estimate for b_ℓ. On the other hand, it is difficult to judge which regime the decay rate of γ falls into in practice with finite samples. Hence, the plug-in method is powerless practically.

II Technique Proofs

In the sequel, we use C and $\tilde{C}$ to denote generic positive finite universal constants that may be different in different uses.

A Proof of Proposition 3

For any $\ell \in [p]$, let $\psi_1(i,j;\ell) = \varphi_{(i,\ell),1}\varphi_{(i,j),0}\varphi_{(\ell,j),1}$ and $\psi_2(i,j;\ell) = \varphi_{(i,\ell),0}\varphi_{(i,j),1}\varphi_{(\ell,j),0}$. Write $\mathring{\psi}_1(i,j;\ell) = \psi_1(i,j;\ell) - \mathbb{E}\{\psi_1(i,j;\ell)\}$ and $\mathring{\psi}_2(i,j;\ell) = \psi_2(i,j;\ell) - \mathbb{E}\{\psi_2(i,j;\ell)\}$. Define

$$\hat{\zeta}_\ell = \frac{\sum_{(i,j)\in\mathcal{H}_\ell}\psi_1(i,j;\ell)}{\sum_{(i,j)\in\mathcal{H}_\ell}\psi_2(i,j;\ell)} \quad \text{and} \quad \zeta_\ell = \frac{\sum_{(i,j)\in\mathcal{H}_\ell}\mathbb{E}\{\psi_1(i,j;\ell)\}}{\sum_{(i,j)\in\mathcal{H}_\ell}\mathbb{E}\{\psi_2(i,j;\ell)\}}\,. \tag{A.1}$$

To prove Proposition 3, we need the following lemma.

Lemma 1. *Under Condition* 1 *and* $\{(\alpha_{i,j},\beta_{i,j})\}_{(i,j)\in\mathcal{I}} \in \mathcal{M}(\gamma; C_1, C_2)$ *for two fixed constants* $C_1 \in (0, 0.5)$ *and* $C_2 > 1$*, it holds that* $\max_{\ell\in[p]} ||\mathcal{H}_\ell|^{-1}\sum_{(i,j)\in\mathcal{H}_\ell}\mathring{\psi}_1(i,j;\ell)| = O_{\mathrm{p}}(p^{-1}\log^{1/2} p) + O_{\mathrm{p}}(\gamma^2 p^{-1/2}\log^{1/2} p) + O_{\mathrm{p}}(\gamma p^{-1}\log p) = \max_{\ell\in[p]} ||\mathcal{H}_\ell|^{-1}\sum_{(i,j)\in\mathcal{H}_\ell}\mathring{\psi}_2(i,j;\ell)|$.

Proof. Define $\mathcal{F}_\ell = \{Z_{i,\ell}, Z_{\ell,j} : (i,j) \in \mathcal{H}_\ell\}$. Then $\mathbb{E}\{\psi_1(i,j;\ell) \,|\, \mathcal{F}_\ell\} = \varphi_{(i,\ell),1}\varphi_{(\ell,j),1}\mathbb{E}\{\varphi_{(i,j),0}\}$ and

$$\frac{1}{|\mathcal{H}_\ell|}\sum_{(i,j)\in\mathcal{H}_\ell} \mathring{\psi}_1(i,j;\ell) = \underbrace{\frac{1}{|\mathcal{H}_\ell|}\sum_{(i,j)\in\mathcal{H}_\ell}[\psi_1(i,j;\ell) - \mathbb{E}\{\psi_1(i,j;\ell)\,|\,\mathcal{F}_\ell\}]}_{I_{\ell,1,1}} + \underbrace{\frac{1}{|\mathcal{H}_\ell|}\sum_{(i,j)\in\mathcal{H}_\ell}[\mathbb{E}\{\psi_1(i,j;\ell)\,|\,\mathcal{F}_\ell\} - \mathbb{E}\{\psi_1(i,j;\ell)\}]}_{I_{\ell,1,2}}\,.$$

In the sequel, we will specify the convergence rates of $\max_{\ell\in[p]}|I_{\ell,1,1}|$ and $\max_{\ell\in[p]}|I_{\ell,1,2}|$, respectively.

Convergence rate of $\max_{\ell\in[p]}|I_{\ell,1,1}|$. Conditional on $\mathcal{F}_\ell$, we know that $\{\psi_1(i,j;\ell)\}_{(i,j)\in\mathcal{H}_\ell}$ is an independent sequence. For any $(i,j)\in\mathcal{H}_\ell$, write $\sigma^2_{(i,j),\ell,1} = \mathrm{Var}\{\psi_1(i,j;\ell)\,|\,\mathcal{F}_\ell\} = \varphi^2_{(i,\ell),1}\varphi^2_{(\ell,j),1}\mathrm{Var}(Z_{i,j})$. Due to $\max_{(i,j)\in\mathcal{H}_\ell}|\psi_1(i,j;\ell)|\le C$ and $\max_{(i,j)\in\mathcal{H}_\ell}\sigma^2_{(i,j),\ell,1}\le C$, by Bernstein inequality, we have that $\mathbb{P}(|I_{\ell,1,1}|>u\,|\,\mathcal{F}_\ell)\lesssim\exp(-Cp^2u^2)$ for any $u=o(1)$, which implies that $\mathbb{P}(|I_{\ell,1,1}|>u) = \mathbb{E}\{\mathbb{P}(|I_{\ell,1,1}|>u\,|\,\mathcal{F}_\ell)\}\lesssim\exp(-Cp^2u^2)$ for any $u=o(1)$. Therefore, we have $\max_{\ell\in[p]}|I_{\ell,1,1}| = O_{\mathrm{p}}(p^{-1}\log^{1/2}p)$.

Convergence rate of $\max_{\ell\in[p]}|I_{\ell,1,2}|$. Define $\mathring{\varphi}_{(i,j),\tau} = \varphi_{(i,j),\tau} - \mathbb{E}\{\varphi_{(i,j),\tau}\}$. Due to $\psi_1(i,j;\ell) = \psi_1(j,i;\ell)$ for any $i\ne j$, it then holds that

$$\begin{aligned}(p-1)(p-2)I_{\ell,1,2} &= \sum_{i,j:\,i\ne j,\,i,j\ne\ell}\big[\varphi_{(i,\ell),1}\varphi_{(\ell,j),1} - \mathbb{E}\{\varphi_{(i,\ell),1}\varphi_{(\ell,j),1}\}\big]\mathbb{E}\{\varphi_{(i,j),0}\}\\ &= \underbrace{2\sum_{i,j:\,i\ne j,\,i,j\ne\ell}\mathring{\varphi}_{(i,\ell),1}\mathbb{E}\{\varphi_{(\ell,j),1}\}\mathbb{E}\{\varphi_{(i,j),0}\}}_{I_{\ell,1,2}(1)} + \underbrace{\sum_{i,j:\,i\ne j,\,i,j\ne\ell}\mathring{\varphi}_{(i,\ell),1}\mathring{\varphi}_{(\ell,j),1}\mathbb{E}\{\varphi_{(i,j),0}\}}_{I_{\ell,1,2}(2)}\,.\end{aligned}$$

For $I_{\ell,1,2}(1)$, we have $I_{\ell,1,2}(1) = \sum_{i:\,i\ne\ell}\mathring{\varphi}_{(i,\ell),1}[2\sum_{j:\,j\ne i,\ell}\mathbb{E}\{\varphi_{(\ell,j),1}\}\mathbb{E}\{\varphi_{(i,j),0}\}] =: \sum_{i:\,i\ne\ell}\mathring{\varphi}_{(i,\ell),1}A_{i,\ell}$. By Condition 1, $\min_{\ell\in[p]}\min_{i:\,i\ne\ell}A_{i,\ell}\asymp p\gamma^2\asymp\max_{\ell\in[p]}\max_{i:\,i\ne\ell}A_{i,\ell}$. Note that $\max_{\ell\in[p]}\max_{i:\,i\ne\ell}\mathrm{Var}\{\mathring{\varphi}_{(i,\ell),1}\}\le C$. Given ℓ, since $\{\mathring{\varphi}_{(i,\ell),1}\}_{i:\,i\ne\ell}$ is an independent sequence, by Bernstein inequality, $\mathbb{P}\{|I_{\ell,1,2}(1)|>u\}\lesssim\exp(-Cp^{-3}\gamma^{-4}u^2)$ for any $u=o(p^2\gamma^2)$. Thus, $\max_{\ell\in[p]}|I_{\ell,1,2}(1)| = O_{\mathrm{p}}(\gamma^2p^{3/2}\log^{1/2}p)$. For $I_{\ell,1,2}(2)$, letting $B_{(i,j),\ell} = \gamma^{-1}\mathbb{E}\{\varphi_{(i,j),0}\}$, then $\gamma^{-1}I_{\ell,1,2}(2) = \sum_{i,j:\,i\ne j,\,i,j\ne\ell}\mathring{\varphi}_{(i,\ell),1}\mathring{\varphi}_{(\ell,j),1}B_{(i,j),\ell}$. Under Condition 1, $\min_{\ell\in[p]}\min_{i,j:\,i\ne j,\,i,j\ne\ell}B_{(i,j),\ell}\asymp 1\asymp\max_{\ell\in[p]}\max_{i,j:\,i\ne j,\,i,j\ne\ell}B_{(i,j),\ell}$. By the decoupling inequalities of de la Peña and Montgomery-Smith (1995) and Theorem 3.3 of Giné, Latała and Zinn (2000), we have that $\max_{\ell\in[p]}\mathbb{P}\{|\sum_{i,j:\,i\ne j,\,i,j\ne\ell}\mathring{\varphi}_{(i,\ell),1}\mathring{\varphi}_{(\ell,j),1}B_{(i,j),\ell}|>u\}\lesssim\exp(-Cp^{-1}u)$ for any $p\ll u\ll p^2$, which implies $\max_{\ell\in[p]}|I_{\ell,1,2}(2)| = O_{\mathrm{p}}(\gamma p\log p)$. Hence, $\max_{\ell\in[p]}|I_{\ell,1,2}| = O_{\mathrm{p}}(\gamma^2p^{-1/2}\log^{1/2}p) + O_{\mathrm{p}}(\gamma p^{-1}\log p)$. We establish the first result. Similarly, we can also prove another result. □

Now we begin to show Proposition 3. By Condition 1, $\min_{\ell\in[p]}\mu_{\ell,1}\asymp\gamma^3\asymp\max_{\ell\in[p]}\mu_{\ell,2}$. Notice that $\mu_{\ell,1} = |\mathcal{H}_\ell|^{-1}\sum_{(i,j)\in\mathcal{H}_\ell}\mathbb{E}\{\psi_1(i,j;\ell)\}$ and $\mu_{\ell,2} = |\mathcal{H}_\ell|^{-1}\sum_{(i,j)\in\mathcal{H}_\ell}\mathbb{E}\{\psi_2(i,j;\ell)\}$. Since $\gamma\gg p^{-1/3}\log^{1/6}p$, Lemma 1 implies that $\max_{\ell\in[p]}||\mathcal{H}_\ell|^{-1}\sum_{(i,j)\in\mathcal{H}_\ell}\mathring{\psi}_1(i,j;\ell)| = o_{\mathrm{p}}(\gamma^3) = \max_{\ell\in[p]}||\mathcal{H}_\ell|^{-1}\sum_{(i,j)\in\mathcal{H}_\ell}\mathring{\psi}_2(i,j;\ell)|$. By (A.1), it holds that $\hat{\zeta}_\ell - \zeta_\ell = \mu^{-1}_{\ell,2}|\mathcal{H}_\ell|^{-1}\sum_{(i,j)\in\mathcal{H}_\ell}\mathring{\psi}_1(i,j;\ell) - \mu_{\ell,1}\mu^{-2}_{\ell,2}|\mathcal{H}_\ell|^{-1}\sum_{(i,j)\in\mathcal{H}_\ell}\mathring{\psi}_2(i,j;\ell) + R_{\ell,1}$, where $\max_{\ell\in[p]}|R_{\ell,1}| = O_{\mathrm{p}}(\gamma^{-6}p^{-2}\log p) + O_{\mathrm{p}}(\gamma^{-2}p^{-1}\log p)$. Thus, $\max_{\ell\in[p]}|\hat{\zeta}_\ell - \zeta_\ell| = O_{\mathrm{p}}(\gamma^{-3}p^{-1}\log^{1/2}p) + O_{\mathrm{p}}(\gamma^{-1}p^{-1/2}\log^{1/2}p) = o_{\mathrm{p}}(1)$. Recall that $\zeta_\ell = \mu_{\ell,1}/\mu_{\ell,2}$. Since $\theta_\ell = \log(\zeta_\ell)/2$ and $\hat{\theta}_\ell = \log(\hat{\zeta}_\ell)/2$, by

Taylor expansion, we have that

$$\begin{aligned}\hat{\theta}_\ell - \theta_\ell &= \frac{1}{2\zeta_\ell}(\hat{\zeta}_\ell - \zeta_\ell) + R_{\ell,2} \\ &= \frac{1}{2\mu_{\ell,1}}\cdot\frac{1}{|\mathcal{H}_\ell|}\sum_{(i,j)\in\mathcal{H}_\ell}\mathring{\psi}_1(i,j;\ell) - \frac{1}{2\mu_{\ell,2}}\cdot\frac{1}{|\mathcal{H}_\ell|}\sum_{(i,j)\in\mathcal{H}_\ell}\mathring{\psi}_2(i,j;\ell) + R_{\ell,3}\,,\end{aligned} \tag{A.2}$$

where $\max_{\ell\in[p]}|R_{\ell,2}| = O_{\rm p}(\gamma^{-6}p^{-2}\log p) + O_{\rm p}(\gamma^{-2}p^{-1}\log p) = \max_{\ell\in[p]}|R_{\ell,3}|$. Therefore, $\max_{\ell\in[p]}|\hat{\theta}_\ell - \theta_\ell| = O_{\rm p}(\gamma^{-3}p^{-1}\log^{1/2}p) + O_{\rm p}(\gamma^{-1}p^{-1/2}\log^{1/2}p) = o_{\rm p}(1)$. We complete the proof of Proposition 3. □

B Proof of Theorem 1

Recall $\mathring{\psi}_1(i,j;\ell) = \psi_1(i,j;\ell) - \mathbb{E}\{\psi_1(i,j;\ell)\}$ and $\mathring{\psi}_2(i,j;\ell) = \psi_2(i,j;\ell) - \mathbb{E}\{\psi_2(i,j;\ell)\}$. Note that $\gamma \gg p^{-1/3}\log^{1/6}p$. Write $N = (p-1)(p-2)$. Given $\ell\in[p]$, following the proof of Lemma 1, it holds that

$$\begin{aligned}\frac{1}{|\mathcal{H}_\ell|}\sum_{(i,j)\in\mathcal{H}_\ell}\mathring{\psi}_1(i,j;\ell) = &\underbrace{\frac{1}{|\mathcal{H}_\ell|}\sum_{(i,j)\in\mathcal{H}_\ell}[\psi_1(i,j;\ell) - \mathbb{E}\{\psi_1(i,j;\ell)\,|\,\mathcal{F}_\ell\}]}_{I_{\ell,1,1}=O_{\rm p}(p^{-1})} \\ &+ \underbrace{\frac{2}{N}\sum_{i,j:\,i\neq j,\,i,j\neq\ell}\mathring{\varphi}_{(i,\ell),1}\mathbb{E}\{\varphi_{(\ell,j),1}\}\mathbb{E}\{\varphi_{(i,j),0}\}}_{N^{-1}I_{\ell,1,2}(1)=O_{\rm p}(\gamma^2p^{-1/2})} + R_{\ell,4}\,, \\ \frac{1}{|\mathcal{H}_\ell|}\sum_{(i,j)\in\mathcal{H}_\ell}\mathring{\psi}_2(i,j;\ell) = &\underbrace{\frac{1}{|\mathcal{H}_\ell|}\sum_{(i,j)\in\mathcal{H}_\ell}[\psi_2(i,j;\ell) - \mathbb{E}\{\psi_2(i,j;\ell)\,|\,\mathcal{F}_\ell\}]}_{I_{\ell,2,1}=O_{\rm p}(p^{-1})} \\ &+ \underbrace{\frac{2}{N}\sum_{i,j:\,i\neq j,\,i,j\neq\ell}\mathring{\varphi}_{(i,\ell),0}\mathbb{E}\{\varphi_{(\ell,j),0}\}\mathbb{E}\{\varphi_{(i,j),1}\}}_{N^{-1}I_{\ell,2,2}(1)=O_{\rm p}(\gamma^2p^{-1/2})} + R_{\ell,5}\,,\end{aligned}$$

where $|R_{\ell,4}| = O_{\rm p}(\gamma p^{-1}) = |R_{\ell,5}|$. Also, for given $\ell\in[p]$, by (A.2), we have

$$\hat{\theta}_\ell - \theta_\ell = \underbrace{\frac{I_{\ell,1,1}}{2\mu_{\ell,1}} - \frac{I_{\ell,2,1}}{2\mu_{\ell,2}}}_{T_{\ell,1}} + \underbrace{\frac{I_{\ell,1,2}(1)}{2\mu_{\ell,1}N} - \frac{I_{\ell,2,2}(1)}{2\mu_{\ell,2}N}}_{T_{\ell,2}} + R_{\ell,6}\,, \tag{B.1}$$

where $|R_{\ell,6}| = O_{\rm p}(\gamma^{-6}p^{-2}) + O_{\rm p}(\gamma^{-2}p^{-1})$.

B.1 Case 1: $\gamma \gg p^{-1/4}$

Notice that $T_{\ell,1} = O_{\rm p}(\gamma^{-3}p^{-1})$. We then have $\hat{\theta}_\ell - \theta_\ell = \mu_{\ell,1}^{-1}N^{-1}\sum_{i,j:\,i\neq j,\,i,j\neq\ell}\mathring{\varphi}_{(i,\ell),1}\mathbb{E}\{\varphi_{(\ell,j),1}\}\mathbb{E}\{\varphi_{(i,j),0}\} - \mu_{\ell,2}^{-1}N^{-1}\sum_{i,j:\,i\neq j,\,i,j\neq\ell}\mathring{\varphi}_{(i,\ell),0}\mathbb{E}\{\varphi_{(\ell,j),0}\}\mathbb{E}\{\varphi_{(i,j),1}\} + R_{\ell,7}$, where $|R_{\ell,7}| = O_{\rm p}(\gamma^{-3}p^{-1})$. We define $\lambda_{i,\ell}^* = \gamma_{\min}(p-2)^{-1}\sum_{j:\,j\neq\ell,i}[\mu_{\ell,1}^{-1}\mathbb{E}\{\varphi_{(\ell,j),1}\}\mathbb{E}\{\varphi_{(i,j),0}\} + \mu_{\ell,2}^{-1}\mathbb{E}\{\varphi_{(\ell,j),0}\}\mathbb{E}\{\varphi_{(i,j),1}\}]$ with $\gamma_{\min} = \min_{i,j:\,i\neq j}(1-\alpha_{i,j}-\beta_{i,j})$. Under Condition 1, $\min_{\ell\in[p]}\min_{i:\,i\neq\ell}\lambda_{i,\ell}^* \asymp 1 \asymp \max_{\ell\in[p]}\max_{i:\,i\neq\ell}\lambda_{i,\ell}^*$. Recall $\mathring{\varphi}_{(i,\ell),1} = Z_{i,\ell} - \mathbb{E}(Z_{i,\ell})$ and $\mathring{\varphi}_{(i,\ell),0} = -Z_{i,\ell} + \mathbb{E}(Z_{i,\ell})$. Then $\gamma_{\min}(\hat{\theta}_\ell - \theta_\ell) = (p-1)^{-1}\sum_{i:\,i\neq\ell}\lambda_{i,\ell}^*\{Z_{i,\ell} - \mathbb{E}(Z_{i,\ell})\} + R_{\ell,8}$ with $|R_{\ell,8}| = O_{\rm p}(\gamma^{-2}p^{-1})$. Let $b_{\ell,*} = (p-1)^{-1}\sum_{i:\,i\neq\ell}\lambda_{i,\ell}^{*,2}{\rm Var}(Z_{i,\ell})$. Given s different $\ell_1,\ldots,\ell_s$, we define an s-dimensional vector $\mathbf{w}_i = (W_{i,1},\ldots,W_{i,s})^{\rm T}$ with $W_{i,j} = \lambda_{i,\ell_j}^*\{Z_{i,\ell_j} - \mathbb{E}(Z_{i,\ell_j})\}$ for any $j\in[s]$. Then it holds that $\gamma_{\min}(\hat{\theta}_{\ell_1}-\theta_{\ell_1},\ldots,\hat{\theta}_{\ell_s}-\theta_{\ell_s})^{\mathrm{T}} = p^{-1}\sum_{i:\, i\neq \ell_1,\ldots,\ell_s}\mathbf{w}_i + \mathbf{r}$ with $|\mathbf{r}|_\infty = O_{\rm p}(\gamma^{-2}p^{-1})$. Notice that $p^{-1}\sum_{i:\, i\neq\ell_1,\ldots,\ell_s}\mathrm{Var}(\mathbf{w}_i) = \mathrm{diag}(b_{\ell_1,*},\ldots,b_{\ell_s,*}) + O(p^{-1})$. Since $\{\mathbf{w}_i\}_{i\neq\ell_1,\ldots,\ell_s}$ is an independent sequence, by the Central Limit Theorem, $(p-1)^{1/2}\gamma_{\min}\mathrm{diag}(b_{\ell_1,*}^{-1/2},\ldots,b_{\ell_s,*}^{-1/2})(\hat{\theta}_{\ell_s}-\theta_{\ell_1},\ldots,\hat{\theta}_{\ell_s}-\theta_{\ell_s})^{\mathrm{T}} \xrightarrow{d} \mathcal{N}(\mathbf{0},\mathbf{I}_s)$. Notice that $b_{\ell,*} = \gamma_{\min}^2 b_\ell$. We complete the proof of Case 1. □

B.2 Case 2: $p^{-1/4} \gg \gamma \gg p^{-1/3}\log^{1/6} p$

Notice that $T_{\ell,2} = O_{\rm p}(\gamma^{-1}p^{-1/2})$. Write $\mathring{Z}_{i,j} = Z_{i,j} - \mathbb{E}(Z_{i,j})$. Then it holds that

$$\hat{\theta}_\ell - \theta_\ell = \underbrace{-\frac{1}{N}\sum_{i,j:\, i\neq j,\, i,j\neq \ell}\bigg\{\frac{\varphi_{(i,\ell),1}\varphi_{(\ell,j),1}}{2\mu_{\ell,1}} + \frac{\varphi_{(i,\ell),0}\varphi_{(\ell,j),0}}{2\mu_{\ell,2}}\bigg\}\mathring{Z}_{i,j}}_{J_\ell} + R_{\ell,9} \tag{B.2}$$

with $|R_{\ell,9}| = O_{\rm p}(\gamma^{-6}p^{-2}) + O_{\rm p}(\gamma^{-1}p^{-1/2})$. Notice that

$$\begin{aligned} J_\ell = &-\frac{1}{N}\sum_{i,j:\, i\neq j,\, i,j\neq\ell}\bigg\{\frac{\mathring{\varphi}_{(i,\ell),1}\mathring{\varphi}_{(\ell,j),1}}{2\mu_{\ell,1}} + \frac{\mathring{\varphi}_{(i,\ell),0}\mathring{\varphi}_{(\ell,j),0}}{2\mu_{\ell,2}}\bigg\}\mathring{Z}_{i,j} \\ &\underbrace{-\frac{2}{N}\sum_{i,j:\, i\neq j,\, i,j\neq\ell}\bigg[\frac{\mathbb{E}\{\varphi_{(\ell,j),1}\}}{2\mu_{\ell,1}}\mathring{\varphi}_{(i,\ell),1} + \frac{\mathbb{E}\{\varphi_{(\ell,j),0}\}}{2\mu_{\ell,2}}\mathring{\varphi}_{(i,\ell),0}\bigg]\mathring{Z}_{i,j}}_{J_{\ell,1}} \\ &\underbrace{-\frac{1}{N}\sum_{i,j:\, i\neq j,\, i,j\neq\ell}\bigg[\frac{\mathbb{E}\{\varphi_{(i,\ell),1}\}\mathbb{E}\{\varphi_{(\ell,j),1}\}}{2\mu_{\ell,1}} + \frac{\mathbb{E}\{\varphi_{(i,\ell),0}\}\mathbb{E}\{\varphi_{(\ell,j),0}\}}{2\mu_{\ell,2}}\bigg]\mathring{Z}_{i,j}}_{J_{\ell,2}}\,. \end{aligned} \tag{B.3}$$

Under Condition 1, $\min_{\ell\in[p]}\min_{i,j:\, i\neq j,\, i,j\neq\ell}[\mathbb{E}\{\varphi_{(i,\ell),1}\}\mathbb{E}\{\varphi_{(\ell,j),1}\}(2\mu_{\ell,1})^{-1} + \mathbb{E}\{\varphi_{(i,\ell),0}\}\mathbb{E}\{\varphi_{(\ell,j),0}\}(2\mu_{\ell,2})^{-1}] \asymp \gamma^{-1} \asymp \max_{\ell\in[p]}\max_{i,j:\, i\neq j,\, i,j\neq\ell}[\mathbb{E}\{\varphi_{(i,\ell),1}\}\mathbb{E}\{\varphi_{(\ell,j),1}\}(2\mu_{\ell,1})^{-1} + \mathbb{E}\{\varphi_{(i,\ell),0}\}\mathbb{E}\{\varphi_{(\ell,j),0}\}(2\mu_{\ell,2})^{-1}]$. It follows from Bernstein inequality that $|J_{\ell,2}| = O_{\rm p}(\gamma^{-1}p^{-1})$. For $J_{\ell,1}$, we can reformulate it as follows:

$$\begin{aligned} J_{\ell,1} = &\underbrace{\frac{1}{p-1}\sum_{i:\, i\neq\ell}\mathring{\varphi}_{(i,\ell),1}\bigg[\frac{1}{p-2}\sum_{j:\, j\neq i,\ell}\frac{\mathbb{E}\{\varphi_{(\ell,j),1}\}}{\mu_{\ell,1}}\mathring{Z}_{i,j}\bigg]}_{J_{\ell,1}(1)} \\ &+ \underbrace{\frac{1}{p-1}\sum_{i:\, i\neq\ell}\mathring{\varphi}_{(i,\ell),0}\bigg[\frac{1}{p-2}\sum_{j:\, j\neq i,\ell}\frac{\mathbb{E}\{\varphi_{(\ell,j),0}\}}{\mu_{\ell,2}}\mathring{Z}_{i,j}\bigg]}_{J_{\ell,1}(2)}\,. \end{aligned} \tag{B.4}$$

Due to $\min_{\ell\in[p]}\min_{j:\, j\neq\ell}\mathbb{E}\{\varphi_{(\ell,j),1}\}\mu_{\ell,1}^{-1} \asymp \gamma^{-2} \asymp \max_{\ell\in[p]}\max_{j:\, j\neq\ell}\mathbb{E}\{\varphi_{(\ell,j),1}\}\mu_{\ell,1}^{-1}$, by Bernstein inequality, $\mathbb{P}[(p-2)^{-1}|\sum_{j:\, j\neq i,\ell}\mathbb{E}\{\varphi_{(\ell,j),1}\}\mu_{\ell,1}^{-1}\mathring{Z}_{i,j}| > u] \lesssim \exp(-C\gamma^4 pu^2)$ for any $u = o(1)$. Given sufficiently large $C_* > 0$, define $\mathcal{E}_\ell(C_*) = \{\max_{i:\, i\neq\ell}|(p-2)^{-1}\sum_{j:\, j\neq i,\ell}\mathbb{E}\{\varphi_{(\ell,j),1}\}\mu_{\ell,1}^{-1}\mathring{Z}_{i,j}| \leq C_*\gamma^{-2}p^{-1/2}\log^{1/2}p\}$. Then

$$\mathbb{P}\{|J_{\ell,1}(1)| > u\} \leq \mathbb{P}\Bigg\{\Bigg|\frac{1}{p-1}\sum_{i:\, i\neq\ell}\mathring{\varphi}_{(i,\ell),1}\underbrace{\bigg[\frac{1}{p-2}\sum_{j:\, j\neq i,\ell}\frac{\mathbb{E}\{\varphi_{(\ell,j),1}\}}{\mu_{\ell,1}}\mathring{Z}_{i,j}\bigg]}_{\tilde{A}_{i,\ell}}\Bigg| > u\,, \mathcal{E}_\ell(C_*)\Bigg\} + p^{-C}\,. \tag{B.5}$$

Notice that the constant C in (B.5) can be sufficiently large if we select sufficiently large C_*. Let $\mathcal{F}_{-\ell} = \{Z_{i,j} : i, j \neq \ell\}$. Conditional on $\mathcal{F}_{-\ell}$, Bernstein inequality implies that

$$\mathbb{P}\Bigg\{\Bigg|\frac{1}{p-1}\sum_{i:\, i\neq\ell}\mathring{\varphi}_{(i,\ell),1}\Bigg[\frac{1}{p-2}\sum_{j:\, j\neq i,\ell}\frac{\mathbb{E}\{\varphi_{(\ell,j),1}\}}{\mu_{\ell,1}}\mathring{Z}_{i,j}\Bigg]\Bigg| > u\,, \mathcal{E}_\ell(C_*)\,\Bigg|\,\mathcal{F}_{-\ell}\Bigg\}$$
$$\lesssim \exp\bigg\{-\frac{Cp^2u^2}{\tilde{C}\sum_{i:\, i\neq\ell}\tilde{A}_{i,\ell}^2 + pu\max_{i:\, i\neq\ell}|\tilde{A}_{i,\ell}|}\bigg\} I\bigg(\max_{i:\, i\neq\ell}|\tilde{A}_{i,\ell}| \leq \frac{C_*\log^{1/2}p}{\gamma^2p^{1/2}}\bigg)$$

for any $u > 0$. Selecting $u = C_*\gamma^{-2}p^{-1}\log p$, we have

$$\mathbb{P}\Bigg\{\Bigg|\frac{1}{p-1}\sum_{i:\, i\neq\ell}\mathring{\varphi}_{(i,\ell),1}\Bigg[\frac{1}{p-2}\sum_{j:\, j\neq i,\ell}\frac{\mathbb{E}\{\varphi_{(\ell,j),1}\}}{\mu_{\ell,1}}\mathring{Z}_{i,j}\Bigg]\Bigg| > \frac{C_*\log p}{\gamma^2 p}\,, \mathcal{E}_\ell(C_*)\,\Bigg|\,\mathcal{F}_{-\ell}\Bigg\}$$
$$\lesssim p^{-C}\cdot I\bigg(\max_{i:\, i\neq\ell}|\tilde{A}_{i,\ell}| \leq \frac{C_*\log^{1/2}p}{\gamma^2p^{1/2}}\bigg) \leq p^{-C}\,,$$

which implies $\mathbb{P}\{|(p-1)^{-1}\sum_{i:\, i\neq\ell}\mathring{\varphi}_{(i,\ell),1}[(p-2)^{-1}\sum_{j:\, j\neq i,\ell}\mathbb{E}\{\varphi_{(\ell,j),1}\}\mu_{\ell,1}^{-1}\mathring{Z}_{i,j}]| > C_*\gamma^{-2}p^{-1}\log p\,, \mathcal{E}_\ell(C_*)\} \lesssim p^{-C} \to 0$. Here the constant C can be sufficiently large if we select sufficiently large C_*. Thus, (B.5) implies $\max_{\ell\in[p]}|J_{\ell,1}(1)| = O_{\rm p}(\gamma^{-2}p^{-1}\log p)$. Analogously, we also have $\max_{\ell\in[p]}|J_{\ell,1}(2)| = O_{\rm p}(\gamma^{-2}p^{-1}\log p)$. By (B.4), $\max_{\ell\in[p]}|J_{\ell,1}| = O_{\rm p}(\gamma^{-2}p^{-1}\log p)$. Due to $\mathring{\varphi}_{(i,j),1} = \mathring{Z}_{i,j}$ and $\mathring{\varphi}_{(i,j),0} = -\mathring{Z}_{i,j}$, by (B.2) and (B.3), we have $\hat{\theta}_\ell - \theta_\ell = -N^{-1}\sum_{i,j:\, i\neq j,\, i,j\neq\ell}\{(2\mu_{\ell,1})^{-1} + (2\mu_{\ell,2})^{-1}\}\mathring{Z}_{i,\ell}\mathring{Z}_{\ell,j}\mathring{Z}_{i,j} + R_{\ell,10}$ with $|R_{\ell,10}| = O_{\rm p}(\gamma^{-6}p^{-2}) + O_{\rm p}(\gamma^{-1}p^{-1/2})$, which implies $-2\mu_{\ell,1}\mu_{\ell,2}(\hat{\theta}_\ell - \theta_\ell)/(\mu_{\ell,1} + \mu_{\ell,2}) = N^{-1}\sum_{i,j:\, i\neq j,\, i,j\neq\ell}\mathring{Z}_{i,\ell}\mathring{Z}_{\ell,j}\mathring{Z}_{i,j} + R_{\ell,11} =: \Delta_\ell + R_{\ell,11}$ with $|R_{\ell,11}| = O_{\rm p}(\gamma^{-3}p^{-2}) + O_{\rm p}(\gamma^2p^{-1/2})$. In the sequel, we will specify the limiting distribution of $\sqrt{N}(\Delta_{\ell_1}, \ldots, \Delta_{\ell_s})$ for given s different $\ell_1, \ldots, \ell_s$. For given $k \in [s]$, we have

$$\Delta_{\ell_k} = \underbrace{\frac{1}{N}\sum_{\substack{i,j:\, i\neq j,\\ i,j\neq\ell_1,\ldots,\ell_s}}\mathring{Z}_{i,\ell_k}\mathring{Z}_{\ell_k,j}\mathring{Z}_{i,j}}_{M_{\ell_k,1}} + \underbrace{\frac{1}{N}\sum_{\substack{i,j:\, i\neq j,\, i,j\neq\ell_k,\\ \{i,j\}\cap\{\ell_1,\ldots,\ell_{k-1},\ell_{k+1},\ldots,\ell_s\}\neq\emptyset}}\mathring{Z}_{i,\ell_k}\mathring{Z}_{\ell_k,j}\mathring{Z}_{i,j}}_{M_{\ell_k,2}}\,. \tag{B.6}$$

Notice that $N\cdot M_{\ell_k,2} = \sum_{k':\, k'\neq k}\sum_{j:\, j\neq\ell_1,\ldots,\ell_s}\mathring{Z}_{\ell_{k'},\ell_k}\mathring{Z}_{\ell_k,j}\mathring{Z}_{\ell_{k'},j} + \sum_{k':\, k'\neq k}\sum_{i:\, i\neq\ell_1,\ldots,\ell_s}\mathring{Z}_{i,\ell_k}\mathring{Z}_{\ell_k,\ell_{k'}}\mathring{Z}_{i,\ell_{k'}} + \sum_{k',k'':\, k'\neq k'',\, k',k''\neq k}\mathring{Z}_{\ell_{k'},\ell_k}\mathring{Z}_{\ell_k,\ell_{k''}}\mathring{Z}_{\ell_{k'},\ell_{k''}} = 2\sum_{k':\, k'\neq k}\sum_{i:\, i\neq\ell_1,\ldots,\ell_s}\mathring{Z}_{i,\ell_k}\mathring{Z}_{\ell_k,\ell_{k'}}\mathring{Z}_{i,\ell_{k'}} + \sum_{k',k'':\, k'\neq k'',\, k',k''\neq k}\mathring{Z}_{\ell_{k'},\ell_k}\mathring{Z}_{\ell_k,\ell_{k''}}\mathring{Z}_{\ell_{k'},\ell_{k''}}$. Since $\max_{i,j:\, i\neq j}|\mathring{Z}_{i,j}| \leq C$, $\max_{k\in[s]}\max_{k',k'':\, k',k''\neq k,\, k'\neq k''}|\mathring{Z}_{\ell_{k'},\ell_k}\mathring{Z}_{\ell_k,\ell_{k''}}\mathring{Z}_{\ell_{k'},\ell_{k''}}| \lesssim 1$, which implies $M_{\ell_k,2} = 2\sum_{k':\, k'\neq k}\mathring{Z}_{\ell_k,\ell_{k'}}\sum_{i:\, i\neq\ell_1,\ldots,\ell_s}N^{-1}\mathring{Z}_{i,\ell_k}\mathring{Z}_{i,\ell_{k'}} + R_{\ell_k,12}$ with $|R_{\ell_k,12}| = O(N^{-1})$. For given k' such that $k' \neq k$, since $\{\mathring{Z}_{i,\ell_k}\mathring{Z}_{i,\ell_{k'}}\}_{i\neq\ell_1,\ldots,\ell_s}$ is an independent sequence, by Bernstein inequality, we have $\mathbb{P}\{(p-s)^{-1}|\sum_{i:\, i\neq\ell_1,\ldots,\ell_s}\mathring{Z}_{i,\ell_k}\mathring{Z}_{i,\ell_{k'}}| > u\} \lesssim \exp(-Cpu^2)$ for any $u = o(1)$. Therefore, it holds that $\max_{k',k:\, k'\neq k}|N^{-1}\sum_{i:\, i\neq\ell_1,\ldots,\ell_s}\mathring{Z}_{i,\ell_k}\mathring{Z}_{i,\ell_{k'}}| = O_{\rm p}(p^{-3/2})$, which implies $M_{\ell_k,2} = O_{\rm p}(p^{-3/2})$. By (B.6), $\Delta_{\ell_k} = M_{\ell_k,1} + R_{\ell_k,12}$ with $|R_{\ell_k,12}| = O_{\rm p}(p^{-3/2})$.

Given $\ell_1, \ldots, \ell_s$, write $b_{\ell_k,**} = N^{-1}\sum_{i,j:\, i\neq j,\, i,j\neq\ell_1,\ldots,\ell_s}{\rm Var}(Z_{i,\ell_k}){\rm Var}(Z_{\ell_k,j}){\rm Var}(Z_{i,j})$ and $f(t_1, \ldots, t_s) = \mathbb{E}\{\exp(\iota N^{1/2}\sum_{k=1}^s t_k b_{\ell_k,**}^{-1/2}M_{\ell_k,1})\}$ with $\iota^2 = -1$. Let $\mathcal{F}^*_{\ell_1,\ldots,\ell_s} = \cup_{k=1}^s\{Z_{i,\ell_k}, Z_{\ell_k,j} : i, j \neq \ell_k\}$. Then

$$\mathbb{E}\Bigg\{\exp\bigg(\iota N^{1/2}\sum_{k=1}^s t_k b_{\ell_k,**}^{-1/2}M_{\ell_k,1}\bigg)\,\Bigg|\,\mathcal{F}^*_{\ell_1,\ldots,\ell_s}\Bigg\}$$

$$= \mathbb{E}\bigg[\exp\bigg\{\sum_{\substack{i,j:\, i<j,\\ i,j\neq \ell_1,\ldots,\ell_s}} \iota\bigg(\sum_{k=1}^{s}\frac{2t_k b_{\ell_k,**}^{-1/2}}{\sqrt{N}}\mathring{Z}_{i,\ell_k}\mathring{Z}_{\ell_k,j}\bigg)\mathring{Z}_{i,j}\bigg\}\,\bigg|\,\mathcal{F}^{*}_{\ell_1,\ldots,\ell_s}\bigg]$$

$$= \prod_{\substack{i,j:\, i<j,\\ i,j\neq \ell_1,\ldots,\ell_s}} \mathbb{E}\bigg[\exp\bigg\{\iota\bigg(\sum_{k=1}^{s}\frac{2t_k b_{\ell_k,**}^{-1/2}}{\sqrt{N}}\mathring{Z}_{i,\ell_k}\mathring{Z}_{\ell_k,j}\bigg)\mathring{Z}_{i,j}\bigg\}\,\bigg|\,\mathcal{F}^{*}_{\ell_1,\ldots,\ell_s}\bigg].$$

By Taylor expansion, $\exp\{\iota(\sum_{k=1}^{s} 2t_k b_{\ell_k,**}^{-1/2}N^{-1/2}\mathring{Z}_{i,\ell_k}\mathring{Z}_{\ell_k,j})\mathring{Z}_{i,j}\} = 1+\iota(\sum_{k=1}^{s} 2t_k b_{\ell_k,**}^{-1/2}N^{-1/2}\mathring{Z}_{i,\ell_k}\mathring{Z}_{\ell_k,j})\mathring{Z}_{i,j} - 2^{-1}(\sum_{k=1}^{s} 2t_k b_{\ell_k,**}^{-1/2}N^{-1/2}\mathring{Z}_{i,\ell_k}\mathring{Z}_{\ell_k,j})^2\mathring{Z}_{i,j}^2 + \tilde{R}_{i,j}$ with $|\tilde{R}_{i,j}| \le CN^{-3/2}(|t_1|+\cdots+|t_s|)^3$, which implies $\mathbb{E}[\exp\{\iota(\sum_{k=1}^{s} 2t_k b_{\ell_k,**}^{-1/2}N^{-1/2}\mathring{Z}_{i,\ell_k}\mathring{Z}_{\ell_k,j})\mathring{Z}_{i,j}\}\,|\,\mathcal{F}^{*}_{\ell_1,\ldots,\ell_s}] = 1-2^{-1}(\sum_{k=1}^{s} 2t_k b_{\ell_k,**}^{-1/2}N^{-1/2}\mathring{Z}_{i,\ell_k}\mathring{Z}_{\ell_k,j})^2\mathrm{Var}(Z_{i,j})+\tilde{R}^{*}_{i,j}$ for any $i,j\neq \ell_1,\ldots,\ell_k$, where $|\tilde{R}^{*}_{i,j}| \le CN^{-3/2}(|t_1|+\cdots+|t_s|)^3$. Due to $|\prod_{k=1}^{m} z_k - \prod_{k=1}^{m} w_k| \le \sum_{k=1}^{m}|z_k-w_k|$ for any $z_k, w_k\in\mathbb{C}$ with $|z_k|\le 1$ and $|w_k|\le 1$, $|\mathbb{E}\{\exp(\iota N^{1/2}\sum_{k=1}^{s} t_k b_{\ell_k,**}^{-1/2}M_{\ell_k,1})|\mathcal{F}^{*}_{\ell_1,\ldots,\ell_s}\} - \prod_{i,j:\, i<j,\, i,j\neq\ell_1,\ldots,\ell_s}\{1-2^{-1}(\sum_{k=1}^{s} 2t_k b_{\ell_k,**}^{-1/2}N^{-1/2}\mathring{Z}_{i,\ell_k}\mathring{Z}_{\ell_k,j})^2\mathrm{Var}(Z_{i,j})\}| \lesssim N^{-1/2}(|t_1|+\cdots+|t_s|)^3$. It also holds that $|\prod_{i,j:\, i<j,\, i,j\neq\ell_1,\ldots,\ell_s}\exp\{-2^{-1}(\sum_{k=1}^{s} 2t_k b_{\ell_k,**}^{-1/2}N^{-1/2}\mathring{Z}_{i,\ell_k}\mathring{Z}_{\ell_k,j})^2\mathrm{Var}(Z_{i,j})\} - \prod_{i,j:\, i<j,\, i,j\neq\ell_1,\ldots,\ell_s}\{1-2^{-1}(\sum_{k=1}^{s} 2t_k b_{\ell_k,**}^{-1/2}N^{-1/2}\mathring{Z}_{i,\ell_k}\mathring{Z}_{\ell_k,j})^2\mathrm{Var}(Z_{i,j})\}| \lesssim N^{-1}(|t_1|+\cdots+|t_s|)^3$, which implies that

$$\bigg|\exp\bigg\{-\frac{1}{2}\sum_{i,j:\, i<j,\, i,j\neq\ell_1,\ldots,\ell_s}\bigg(\sum_{k=1}^{s} 2t_k b_{\ell_k,**}^{-1/2}N^{-1/2}\mathring{Z}_{i,\ell_k}\mathring{Z}_{\ell_k,j}\bigg)^2\mathrm{Var}(Z_{i,j})\bigg\} - \mathbb{E}\bigg\{\exp\bigg(\iota N^{1/2}\sum_{k=1}^{s} t_k b_{\ell_k,**}^{-1/2}M_{\ell_k,1}\bigg)\,\bigg|\,\mathcal{F}^{*}_{\ell_1,\ldots,\ell_s}\bigg\}\bigg| \lesssim N^{-1/2}(|t_1|+\cdots+|t_s|)^3\,.$$

Define $Q = \sum_{i,j:\, i\neq j,\, i,j\neq\ell_1,\ldots,\ell_s}(\sum_{k=1}^{s} t_k b_{\ell_k,**}^{-1/2}N^{-1/2}\mathring{Z}_{i,\ell_k}\mathring{Z}_{\ell_k,j})^2\mathrm{Var}(Z_{i,j}) - \sum_{k=1}^{s} t_k^2$. We have $f(t_1,\ldots,t_s) = \exp(-\sum_{k=1}^{s} t_k^2)\mathbb{E}\{\exp(-Q)\} + R$ with $|R| \lesssim N^{-1/2}(|t_1|+\cdots+|t_s|)^3$. In the sequel, we will show $f(t_1,\ldots,t_s)\to\exp(-\sum_{k=1}^{s} t_k^2)$ as $p\to\infty$. Note that $Q = \sum_{k=1}^{s} N^{-1}b_{\ell_k,**}^{-1}\sum_{i,j:\, i\neq j,\, i,j\neq\ell_1,\ldots,\ell_s} t_k^2\{\mathring{Z}_{i,\ell_k}^2\mathring{Z}_{\ell_k,j}^2 - \mathbb{E}(\mathring{Z}_{i,\ell_k}^2\mathring{Z}_{\ell_k,j}^2)\}\mathrm{Var}(Z_{i,j}) + \sum_{k,k':\, k\neq k'}(Nb_{\ell_k,**}^{1/2}b_{\ell_{k'},**}^{1/2})^{-1}\sum_{i,j:\, i\neq j,\, i,j\neq\ell_1,\ldots,\ell_s} t_k t_{k'}\mathring{Z}_{i,\ell_k}\mathring{Z}_{i,\ell_{k'}}\mathring{Z}_{\ell_k,j}\mathring{Z}_{\ell_{k'},j}\mathrm{Var}(Z_{i,j})$. For given k and k' such that $k\neq k'$, by the decoupling inequalities of de la Peña and Montgomery-Smith (1995) and Theorem 3.3 of Giné, Latała and Zinn (2000), $\mathbb{P}[|N^{-1}\sum_{i,j:\, i\neq j,\, i,j\neq\ell_1,\ldots,\ell_s}\{\mathring{Z}_{i,\ell_k}^2\mathring{Z}_{\ell_k,j}^2 - \mathbb{E}(\mathring{Z}_{i,\ell_k}^2\mathring{Z}_{\ell_k,j}^2)\}\mathrm{Var}(Z_{i,j})| > u] \lesssim \exp(-Cpu)$ and $\mathbb{P}[|N^{-1}\sum_{i,j:\, i\neq j,\, i,j\neq\ell_1,\ldots,\ell_s}\mathring{Z}_{i,\ell_k}\mathring{Z}_{i,\ell_{k'}}\mathring{Z}_{\ell_k,j}\mathring{Z}_{\ell_{k'},j}\mathrm{Var}(Z_{i,j})| > u] \lesssim \exp(-Cpu)$ for any $u\to 0$ but $pu\to\infty$, which implies $\mathbb{P}\{|Q| > (|t_1|+\cdots+|t_s|)^2u\} \lesssim \exp(-Cpu)$ for any $u\to 0$ but $pu\to\infty$. For sufficiently large $C_*>0$, define $\mathcal{E}(C_*) = \{|Q|\le C_*p^{-1}\log p\}$. Then $f(t_1,\ldots,t_s) = \exp(-\sum_{k=1}^{s} t_k^2)(\mathbb{E}[\exp(-Q)I\{\mathcal{E}(C_*)\}] + \mathbb{E}[\exp(-Q)I\{\mathcal{E}(C_*)^c\}]) + R$ with $|R| \lesssim N^{-1/2}(|t_1| + \cdots + |t_s|)^3$. As $p\to\infty$, due to the facts $0 \le \mathbb{E}[\exp(-Q-\sum_{k=1}^{s} t_k^2)I\{\mathcal{E}(C_*)^c\}] \le \mathbb{P}\{\mathcal{E}(C_*)^c\}\to 0$ and $1 \leftarrow \exp(-C_*p^{-1}\log p)\mathbb{P}\{\mathcal{E}(C_*)\} \le \mathbb{E}[\exp(-Q)I\{\mathcal{E}(C_*)\}] \le \exp(C_*p^{-1}\log p)\mathbb{P}\{\mathcal{E}(C_*)\}\to 1$, we have $f(t_1,\ldots,t_s)\to\exp(-\sum_{k=1}^{s} t_k^2)$. Since $\sqrt{2}\mu_{\ell_1,1}\mu_{\ell_1,2}b_{\ell_k,**}^{-1/2}(\mu_{\ell_1,1}+\mu_{\ell_1,2})^{-1}\tilde{b}_{\ell_k}^{1/2}\to 1$ for $\tilde{b}_{\ell_k}$ specified in (3.10), we complete the proof of Case 2. □

B.3 Case 3: $\gamma \asymp p^{-1/4}$

Note that $\mathring{\varphi}_{(i,j),1} = \mathring{Z}_{i,j}$ and $\mathring{\varphi}_{(i,j),0} = -\mathring{Z}_{i,j}$. By (B.1), $\hat{\theta}_\ell - \theta_\ell = (p-1)^{-1}\sum_{i:\, i\neq\ell}\lambda_{i,\ell}\mathring{Z}_{i,\ell} + J_\ell + R_{\ell,6}$, where $|R_{\ell,6}| = O_{\mathrm{p}}(p^{-1/2})$, and $\lambda_{i,\ell}$ and J_ℓ are specified in (3.8) and (B.2), respectively. Recall $N=(p-1)(p-2)$. As shown in Section B.2 that $J_\ell = -N^{-1}\sum_{i,j:\, i\neq j,\, i,j\neq\ell}\{(2\mu_{\ell,1})^{-1}+(2\mu_{\ell,2})^{-1}\}\mathring{Z}_{i,\ell}\mathring{Z}_{\ell,j}\mathring{Z}_{i,j} + O_{\mathrm{p}}(p^{-1/2}\log p)$, then $\hat{\theta}_\ell - \theta_\ell = (p-1)^{-1}\sum_{i:\, i\neq\ell}\lambda_{i,\ell}\mathring{Z}_{i,\ell} - N^{-1}\sum_{i,j:\, i\neq j,\, i,j\neq\ell}\{(2\mu_{\ell,1})^{-1}+(2\mu_{\ell,2})^{-1}\}\mathring{Z}_{i,\ell}\mathring{Z}_{\ell,j}\mathring{Z}_{i,j} + R_{\ell,13}$ with $|R_{\ell,13}| = O_{\mathrm{p}}(p^{-1/2}\log p)$. Define $\tilde{\lambda}_{i,\ell} = 2\lambda_{i,\ell}\mu_{\ell,1}\mu_{\ell,2}/(\mu_{\ell,1}+\mu_{\ell,2})$. Then $-2\mu_{\ell,1}\mu_{\ell,2}(\mu_{\ell,1}+\mu_{\ell,2})^{-1}(\hat{\theta}_\ell-\theta_\ell) =$

$-(p-1)^{-1}\sum_{i:\,i\neq\ell}\tilde{\lambda}_{i,\ell}\mathring{Z}_{i,\ell}+N^{-1}\sum_{i,j:\,i\neq j,\,i,j\neq\ell}\mathring{Z}_{i,\ell}\mathring{Z}_{\ell,j}\mathring{Z}_{i,j}+R_{\ell,14}$ with $|R_{\ell,14}|=O_{\rm p}(p^{-5/4}\log p)$. Given s different $\ell_1,\ldots,\ell_s$, as shown in Section B.2, $N^{-1}\sum_{i,j:\,i\neq j,\,i,j\neq\ell_k}\mathring{Z}_{i,\ell_k}\mathring{Z}_{\ell_k,j}\mathring{Z}_{i,j}=M_{\ell_k,1}+O_{\rm p}(p^{-3/2})$, where $M_{\ell_k,1}$ is specified in (B.6). We also have $-(p-1)^{-1}\sum_{i:\,i\neq\ell_k}\tilde{\lambda}_{i,\ell_k}\mathring{Z}_{i,\ell_k}=-(p-1)^{-1}\sum_{i:\,i\neq\ell_1,\ldots,\ell_s}\tilde{\lambda}_{i,\ell_k}\mathring{Z}_{i,\ell_k}+O(p^{-3/2})=:M^*_{\ell_k,1}+O(p^{-3/2})$. Then $-2\mu_{\ell_k,1}\mu_{\ell_k,2}(\mu_{\ell_k,1}+\mu_{\ell_k,2})^{-1}(\hat{\theta}_{\ell_k}-\theta_{\ell_k})=M_{\ell_k,1}+M^*_{\ell_k,1}+O_{\rm p}(p^{-5/4}\log p)$. Let $b_{\ell_k,***}=2\{\mu_{\ell_k,1}\mu_{\ell_k,2}/(\mu_{\ell_k,1}+\mu_{\ell_k,2})\}^2(p-2)b_{\ell_k}$ for any $k\in[s]$, where b_{ℓ_k} is defined in (3.9). Write $\check{b}_{\ell_k}=b_{\ell_k,**}+b_{\ell_k,***}$ with $b_{\ell_k,**}$ specified in Section B.2. Let $\check{f}(t_1,\ldots,t_s)=\mathbb{E}[\exp\{\iota N^{1/2}\sum_{k=1}^s t_k\check{b}_{\ell_k}^{-1/2}(M_{\ell_k,1}+M^*_{\ell_k,1})\}]$ with $\iota^2=-1$. Recall that $\mathcal{F}^*_{\ell_1,\ldots,\ell_s}=\cup_{k=1}^s\{Z_{i,\ell_k},Z_{\ell_k,j}:i,j\neq\ell_k\}$. Following the same argument used in Section B.2, we can show $|\exp\{-2^{-1}\sum_{i,j:\,i<j,\,i,j\neq\ell_1,\ldots,\ell_s}(\sum_{k=1}^s 2t_k\check{b}_{\ell_k}^{-1/2}N^{-1/2}\mathring{Z}_{i,\ell_k}\mathring{Z}_{\ell_k,j})^2{\rm Var}(Z_{i,j})\}$ $-\mathbb{E}\{\exp(\iota N^{1/2}\sum_{k=1}^s t_k\check{b}_{\ell_k}^{-1/2}M_{\ell_k,1})|\mathcal{F}^*_{\ell_1,\ldots,\ell_s}\}|\lesssim N^{-1/2}(|t_1|+\ldots+|t_s|)^3$, which implies that $\check{f}(t_1,\ldots,t_s)=\mathbb{E}[\exp(\iota N^{1/2}\sum_{k=1}^s t_k\check{b}_{\ell_k}^{-1/2}M^*_{\ell_k,1})\times\exp\{-\sum_{i,j:\,i\neq j,\,i,j\neq\ell_1,\ldots,\ell_s}(\sum_{k=1}^s t_k\check{b}_{\ell_k}^{-1/2}N^{-1/2}\mathring{Z}_{i,\ell_k}\mathring{Z}_{\ell_k,j})^2{\rm Var}(Z_{i,j})\}]+O(N^{-1/2})$ for any given $(t_1,\ldots,t_s)$. Define $\tilde{Q}=\sum_{i,j:\,i\neq j,\,i,j\neq\ell_1,\ldots,\ell_s}(\sum_{k=1}^s t_k\check{b}_{\ell_k}^{-1/2}N^{-1/2}\mathring{Z}_{i,\ell_k}\mathring{Z}_{\ell_k,j})^2{\rm Var}(Z_{i,j})$ $-\sum_{k=1}^s t_k^2 b_{\ell_k,**}/\check{b}_{\ell_k}$. Due to $\tilde{Q}=\sum_{k=1}^s t_k^2(N\check{b}_{\ell_k})^{-1}\sum_{i,j:\,i\neq j,\,i,j\neq\ell_1,\ldots,\ell_s}\{\mathring{Z}^2_{i,\ell_k}\mathring{Z}^2_{\ell_k,j}-\mathbb{E}(\mathring{Z}^2_{i,\ell_k}\mathring{Z}^2_{\ell_k,j})\}{\rm Var}(Z_{i,j})+\sum_{k,k':\,k\neq k'}(N\check{b}_{\ell_k}^{1/2}\check{b}_{\ell_{k'}}^{1/2})^{-1}\sum_{i,j:\,i\neq j,\,i,j\neq\ell_1,\ldots,\ell_s}t_kt_{k'}\mathring{Z}_{i,\ell_k}\mathring{Z}_{i,\ell_{k'}}\mathring{Z}_{\ell_k,j}\mathring{Z}_{\ell_{k'},j}{\rm Var}(Z_{i,j})$, using the technique for specifying the upper bound of $\mathbb{P}\{|Q|>(|t_1|+\cdots+|t_s|)^2u\}$ in Section B.2, we have $\mathbb{P}\{|\tilde{Q}|>(|t_1|+\cdots+|t_s|)^2u\}\lesssim\exp(-Cpu)$ for any $u\to0$ but $pu\to\infty$. For sufficiently large $C_*>0$, define $\tilde{\mathcal{E}}(C_*)=\{|\tilde{Q}|\leq C_*p^{-1}\log p\}$. Notice that

$$
\begin{aligned}
&\exp\bigg(\iota N^{1/2}\sum_{k=1}^s t_k\check{b}_{\ell_k}^{-1/2}M^*_{\ell_k,1}\bigg)\exp\bigg\{-\sum_{\substack{i,j:\,i\neq j,\\ i,j\neq\ell_1,\ldots,\ell_s}}\bigg(\sum_{k=1}^s\frac{t_k\check{b}_{\ell_k}^{-1/2}}{\sqrt{N}}\mathring{Z}_{i,\ell_k}\mathring{Z}_{\ell_k,j}\bigg)^2{\rm Var}(Z_{i,j})\bigg\}\\
&\quad=\exp\bigg(\iota N^{1/2}\sum_{k=1}^s t_k\check{b}_{\ell_k}^{-1/2}M^*_{\ell_k,1}\bigg)\exp\bigg(-\tilde{Q}-\sum_{k=1}^s\frac{t_k^2 b_{\ell_k,**}}{\check{b}_{\ell_k}}\bigg)I\{\tilde{\mathcal{E}}(C_*)\}\\
&\qquad+\exp\bigg(\iota N^{1/2}\sum_{k=1}^s t_k\check{b}_{\ell_k}^{-1/2}M^*_{\ell_k,1}\bigg)\exp\bigg(-\tilde{Q}-\sum_{k=1}^s\frac{t_k^2 b_{\ell_k,**}}{\check{b}_{\ell_k}}\bigg)I\{\tilde{\mathcal{E}}(C_*)^c\}\,.
\end{aligned}
$$

Then we have $\check{f}(t_1,\ldots,t_s)=\exp(-\sum_{k=1}^s t_k^2 b_{\ell_k,**}\check{b}_{\ell_k}^{-1})\mathbb{E}[\exp(\iota N^{1/2}\sum_{k=1}^s t_k\check{b}_{\ell_k}^{-1/2}M^*_{\ell_k,1})\exp(-\tilde{Q})I\{\tilde{\mathcal{E}}(C_*)\}]$ $+O(p^{-C})+O(N^{-1/2})$. Note that $(M^*_{\ell_1,1},\ldots,M^*_{\ell_s,1})=-\gamma_{\min}^{-1}{\rm diag}\{2\mu_{\ell_1,1}\mu_{\ell_1,2}(\mu_{\ell_1,1}+\mu_{\ell_1,2})^{-1},\ldots,2\mu_{\ell_s,1}\mu_{\ell_s,2}(\mu_{\ell_s,1}+\mu_{\ell_s,2})^{-1}\}(p-1)^{-1}\sum_{i:\,i\neq\ell_1,\ldots,\ell_s}\mathbf{w}_i$, where $\gamma_{\min}=\min_{i,j:\,i\neq j}(1-\alpha_{i,j}-\beta_{i,j})$ and $\mathbf{w}_i$ is specified in Section B.1. As shown in Section B.2 that $(p-1)^{-1/2}\gamma_{\min}^{-1}{\rm diag}(b_{\ell_1}^{-1/2},\ldots,b_{\ell_s}^{-1/2})\sum_{i:\,i\neq\ell_1,\ldots,\ell_s}\mathbf{w}_i\xrightarrow{d}\mathcal{N}(\mathbf{0},\mathbf{I}_s)$ with b_{ℓ_1} defined as (3.9), then $N^{1/2}{\rm diag}\{(2\mu_{\ell_1,1}\mu_{\ell_1,2})^{-1}(\mu_{\ell_1,1}+\mu_{\ell_1,2})(p-2)^{-1/2}b_{\ell_1}^{-1/2},\ldots,(2\mu_{\ell_s,1}\mu_{\ell_s,2})^{-1}(\mu_{\ell_s,1}+\mu_{\ell_s,2})(p-2)^{-1/2}b_{\ell_s}^{-1/2}\}(M^*_{\ell_1,1},\ldots,M^*_{\ell_s,1})\xrightarrow{d}\mathcal{N}(\mathbf{0},\mathbf{I}_s)$. By the Dominated Convergence Theorem, $\check{f}(t_1,\ldots,t_s)\to\exp(-\sum_{k=1}^s t_k^2)$ for any given $(t_1,\ldots,t_s)$. Hence, $N^{1/2}\,{\rm diag}[\{(p-2)b_{\ell_1}+\tilde{b}_{\ell_1}\}^{-1/2},\ldots,\{(p-2)b_{\ell_s}+\tilde{b}_{\ell_s}\}^{-1/2}](\hat{\theta}_{\ell_1}-\theta_{\ell_1},\ldots,\hat{\theta}_{\ell_s}-\theta_{\ell_s})^{\rm T}\xrightarrow{d}\mathcal{N}(\mathbf{0},\mathbf{I}_s)$. We complete the proof of Case 3. □

C Proof of Proposition 5

To construct Proposition 5, we need the following lemmas.

Lemma 2. *Under Condition* 1 *and* $\{(\alpha_{i,j},\beta_{i,j})\}_{(i,j)\in\mathcal{I}}\in\mathcal{M}(\gamma;C_1,C_2)$ *for two fixed constants* $C_1\in(0,0.5)$ *and* $C_2>1$, *if* $\gamma\gg p^{-1/3}\log^{1/6}p$, *it holds that* $\max_{i:\,i\neq\ell}|\hat{\lambda}_{i,\ell}-\lambda_{i,\ell}|=O_{\rm p}(\gamma^{-3}p^{-1/2}\log^{1/2}p)$ *for any given* $\ell\in[p]$, *where* $\hat{\lambda}_{i,\ell}$ *is defined in* (I.1).

Proof. Due to $\hat{\mu}_{\ell,1}-\mu_{\ell,1}=|\mathcal{H}_\ell|^{-1}\sum_{(i,j)\in\mathcal{H}_\ell}\mathring{\psi}_1(i,j;\ell)$ and $\hat{\mu}_{\ell,2}-\mu_{\ell,2}=|\mathcal{H}_\ell|^{-1}\sum_{(i,j)\in\mathcal{H}_\ell}\mathring{\psi}_2(i,j;\ell)$, following the proof of Lemma 1, $|\hat{\mu}_{\ell,1}-\mu_{\ell,1}|=O_{\rm p}(p^{-1})+O_{\rm p}(\gamma^2p^{-1/2})=|\hat{\mu}_{\ell,2}-\mu_{\ell,2}|$ for any given $\ell\in[p]$. Given (ℓ,i) such that $i\neq\ell$, we know $\{\varphi_{(\ell,j),1}\varphi_{(i,j),0}\}_{j:\,j\neq\ell,i}$ is an independent and bounded sequence. By Bernstein inequality, $\max_{i:\,i\neq\ell}|(p-2)^{-1}\sum_{j:\,j\neq\ell,i}[\varphi_{(\ell,j),1}\varphi_{(i,j),0}-\mathbb{E}\{\varphi_{(\ell,j),1}\}\mathbb{E}\{\varphi_{(i,j),0}\}]|=O_{\rm p}(p^{-1/2}\log^{1/2}p)=\max_{i:\,i\neq\ell}|(p-2)^{-1}\sum_{j:\,j\neq\ell,i}[\varphi_{(\ell,j),0}\varphi_{(i,j),1}-\mathbb{E}\{\varphi_{(\ell,j),0}\}\mathbb{E}\{\varphi_{(i,j),1}\}]|$. Notice that $\mu_{\ell,1}\asymp\gamma^3\asymp\mu_{\ell,2}$. Based on the definition of $\lambda_{i,\ell}$ and $\hat{\lambda}_{i,\ell}$ given, respectively, in (3.8) and (I.1), we complete the proof. $\square$

Lemma 3. *Under Condition* 1 *and* $\{(\alpha_{i,j},\beta_{i,j})\}_{(i,j)\in\mathcal{I}}\in\mathcal{M}(\gamma;C_1,C_2)$ *for two fixed constants* $C_1\in(0,0.5)$ *and* $C_2>1$*, if* $\gamma\gg p^{-1/3}\log^{1/6}p$*, it holds that* $\max_{i:\,i\neq\ell}|\widehat{\rm Var}(Z_{i,\ell})-{\rm Var}(Z_{i,\ell})|=O_{\rm p}(\gamma^{-1}p^{-1/2}\log^{1/2}p)+O_{\rm p}(\gamma^{-3}p^{-1}\log^{1/2}p)$ *for any given* $\ell\in[p]$*, where* $\widehat{\rm Var}(Z_{i,\ell})$ *is defined in* (I.2).

Proof. Define $f(x)=(1+e^x)^{-1}$ for $x\in\mathbb{R}$. We know $\sup_{x\in\mathbb{R}}|f'(x)|\leq C$. As shown in Proposition 3 that $\max_{\ell\in[p]}|\hat{\theta}_\ell-\theta_\ell|=O_{\rm p}(\gamma^{-1}p^{-1/2}\log^{1/2}p)+O_{\rm p}(\gamma^{-3}p^{-1}\log^{1/2}p)$, then we have $\max_{i:\,i\neq\ell}|\{1+\exp(\hat{\theta}_i+\hat{\theta}_\ell)\}^{-1}-\{1+\exp(\theta_i+\theta_\ell)\}^{-1}|\lesssim\max_{\ell\in[p]}|\hat{\theta}_\ell-\theta_\ell|=O_{\rm p}(\gamma^{-3}p^{-1}\log^{1/2}p)+O_{\rm p}(\gamma^{-1}p^{-1/2}\log^{1/2}p)$. Based on the definition of $\widehat{\rm Var}(Z_{i,\ell})$ and ${\rm Var}(Z_{i,\ell})$, we complete the proof. $\square$

Now we begin to prove Proposition 5. Recall that $b_\ell=(p-1)^{-1}\sum_{i:\,i\neq\ell}\lambda_{i,\ell}^2{\rm Var}(Z_{i,\ell})$ and $\hat{b}_\ell=(p-1)^{-1}\sum_{i:\,i\neq\ell}\hat{\lambda}_{i,\ell}^2\widehat{\rm Var}(Z_{i,\ell})$. Then $\hat{b}_\ell-b_\ell=(p-1)^{-1}\sum_{i:\,i\neq\ell}(\hat{\lambda}_{i,\ell}^2-\lambda_{i,\ell}^2){\rm Var}(Z_{i,\ell})+(p-1)^{-1}\sum_{i:\,i\neq\ell}\lambda_{i,\ell}^2\{\widehat{\rm Var}(Z_{i,\ell})-{\rm Var}(Z_{i,\ell})\}+(p-1)^{-1}\sum_{i:\,i\neq\ell}(\hat{\lambda}_{i,\ell}^2-\lambda_{i,\ell}^2)\{\widehat{\rm Var}(Z_{i,\ell})-{\rm Var}(Z_{i,\ell})\}$. Note that $\lambda_{i,\ell}\asymp\gamma^{-1}$ and ${\rm Var}(Z_{i,\ell})\asymp1$. By Lemmas 2 and 3, $|\hat{b}_\ell-b_\ell|\lesssim\max_{i:\,i\neq\ell}|\hat{\lambda}_{i,\ell}^2-\lambda_{i,\ell}^2|+\gamma^{-2}\max_{i:\,i\neq\ell}|\widehat{\rm Var}(Z_{i,\ell})-{\rm Var}(Z_{i,\ell})|=O_{\rm p}(\gamma^{-6}p^{-1}\log p)+O_{\rm p}(\gamma^{-4}p^{-1/2}\log^{1/2}p)$. Since $b_\ell\asymp\gamma^{-2}$, we have $|\hat{b}_\ell/b_\ell-1|=O_{\rm p}(\gamma^{-4}p^{-1}\log p)+O_{\rm p}(\gamma^{-2}p^{-1/2}\log^{1/2}p)$. Analogously, we have $|\hat{\tilde{b}}_\ell-\tilde{b}_\ell|=O_{\rm p}(\gamma^{-9}p^{-1}\log^{1/2}p)+O_{\rm p}(\gamma^{-7}p^{-1/2}\log^{1/2}p)$. Recall that $\tilde{b}_\ell\asymp\gamma^{-6}$. It holds that $|\hat{\tilde{b}}_\ell/\tilde{b}_\ell-1|=O_{\rm p}(\gamma^{-3}p^{-1}\log^{1/2}p)+O_{\rm p}(\gamma^{-1}p^{-1/2}\log^{1/2}p)$. We complete the proof of Proposition 5. $\square$

D Proof of Theorem 2

The proof of Part (a) is almost identical to that of Theorem 1 given in Section B. We only prove Part (b). Recall $\nu_\ell^\dagger=(p-2)b_\ell^\dagger+\tilde{b}_\ell^\dagger$ and $\nu_\ell=(p-2)b_\ell+\tilde{b}_\ell$. Then $\nu_\ell^\dagger-\nu_\ell=(p-2)(b_\ell^\dagger-b_\ell)+\tilde{b}_\ell^\dagger-\tilde{b}_\ell$. Note that $\lambda_{i,\ell}^\dagger=(1-2\delta)^{-1}\lambda_{i,\ell}$, $\mu_{\ell,1}^\dagger=(1-2\delta)^3\mu_{\ell,1}$ and $\mu_{\ell,2}^\dagger=(1-2\delta)^3\mu_{\ell,2}$. Then $b_\ell^\dagger=(p-1)^{-1}(1-2\delta)^{-2}\sum_{i:i\neq\ell}\lambda_{i,\ell}^2{\rm Var}(Z_{i,\ell}^\dagger)$ and $\tilde{b}_\ell^\dagger=\{2N(1-2\delta)^6\}^{-1}\{(\mu_{\ell,1}+\mu_{\ell_2})/(\mu_{\ell_1}\mu_{\ell_2})\}^2\sum_{i,j:i\neq j,i,j\neq\ell}{\rm Var}(Z_{i,\ell}^\dagger){\rm Var}(Z_{\ell,j}^\dagger){\rm Var}(Z_{i,j}^\dagger)$. Recall $\delta\in(0,c]$ with $c<0.5$. For any $i\neq\ell$, noticing that $|Z_{i,\ell}|\leq C$, we have that $|{\rm Var}(Z_{i,\ell}^\dagger)(1-2\delta)^{-2}-{\rm Var}(Z_{i,\ell})|=\delta(1-\delta)(1-2\delta)^{-2}\lesssim\delta$. Under Condition 1, we have $\min_{\ell\in[p]}\min_{i:\,i\neq\ell}\lambda_{i,\ell}\asymp\gamma^{-1}\asymp\max_{\ell\in[p]}\max_{i:\,i\neq\ell}\lambda_{i,\ell}$ and $\min_{\ell\in[p]}\mu_{\ell,1}\asymp\gamma^3\asymp\max_{\ell\in[p]}\mu_{\ell,2}$. Then $(p-2)\max_{\ell\in[p]}|b_\ell^\dagger-b_\ell|\lesssim p\delta\gamma^{-2}$ and $\max_{\ell\in[p]}|\tilde{b}_\ell^\dagger-\tilde{b}_\ell|\lesssim\delta\gamma^{-6}$, which implies $\max_{\ell\in[p]}|\nu_\ell^\dagger-\nu_\ell|\lesssim p\delta\gamma^{-2}+\delta\gamma^{-6}$. For any $\ell\in[p]$, notice that $\nu_\ell\asymp p\gamma^{-2}$ if $\gamma\gtrsim p^{-1/4}$ and $\nu_\ell\asymp\gamma^{-6}$ if $p^{-1/4}\gg\gamma\gg p^{-1/3}\log^{1/6}p$. Then $\max_{\ell\in[p]}|\nu_\ell^\dagger\nu_\ell^{-1}-1|=O(\delta)$. $\square$

E Proof of Theorem 3

As shown in (B.1), it holds that

$$\hat{\theta}_\ell-\theta_\ell=\underbrace{\frac{I_{\ell,1,1}}{2\mu_{\ell,1}}-\frac{I_{\ell,2,1}}{2\mu_{\ell,2}}}_{T_{\ell,1}}+\underbrace{\frac{I_{\ell,1,2}(1)}{2\mu_{\ell,1}N}-\frac{I_{\ell,2,2}(1)}{2\mu_{\ell,2}N}}_{T_{\ell,2}}+R_{\ell,6}\,,\tag{E.1}$$

where $\max_{\ell\in[p]}|R_{\ell,6}|=O_{\rm p}(\gamma^{-6}p^{-2}\log p)+O_{\rm p}(\gamma^{-2}p^{-1}\log p)$. Write $N=(p-1)(p-2)$. As shown in Section B.2, $T_{\ell,1}=-N^{-1}\sum_{i,j:\,i\neq j,\,i,j\neq\ell}\{(2\mu_{\ell,1})^{-1}+(2\mu_{\ell,2})^{-1}\}\mathring{Z}_{i,\ell}\mathring{Z}_{\ell,j}\mathring{Z}_{i,j}+J_{\ell,1}+J_{\ell,2}$, where $\max_{\ell\in[p]}|J_{\ell,1}+J_{\ell,2}|=O_{\rm p}(\gamma^{-2}p^{-1}\log p)$. Recall that $T_{\ell,2}=(p-1)^{-1}\sum_{i:\,i\neq\ell}\lambda_{i,\ell}\mathring{Z}_{i,\ell}$ with $\lambda_{i,\ell}$ specified in (3.8). By (E.1), we have

$\hat{\theta}_\ell - \theta_\ell = -N^{-1}\sum_{i,j:\,i\neq j,\,i,j\neq \ell}\{(2\mu_{\ell,1})^{-1}+(2\mu_{\ell,2})^{-1}\}\mathring{Z}_{i,\ell}\mathring{Z}_{\ell,j}\mathring{Z}_{i,j} + (p-1)^{-1}\sum_{i:\,i\neq\ell}\lambda_{i,\ell}\mathring{Z}_{i,\ell} + R_{\ell,15}$, where $\max_{\ell\in[p]}|R_{\ell,15}| = O_{\rm p}(\gamma^{-6}p^{-2}\log p) + O_{\rm p}(\gamma^{-2}p^{-1}\log p)$. Recall $\nu_\ell = (p-2)b_\ell + \tilde{b}_\ell \asymp p\gamma^{-2}+\gamma^{-6}$. Then $N^{1/2}\nu_\ell^{-1/2}(\hat{\theta}_\ell - \theta_\ell) = -N^{-1/2}\sum_{i,j:\,i\neq j,\,i,j\neq\ell}\nu_\ell^{-1/2}[\{(2\mu_{\ell,1})^{-1}+(2\mu_{\ell,2})^{-1}\}\mathring{Z}_{i,\ell}\mathring{Z}_{\ell,j}\mathring{Z}_{i,j} - \lambda_{i,\ell}\mathring{Z}_{i,\ell}] + R_{\ell,16}$, where $\max_{\ell\in[p]}|R_{\ell,16}| = O_{\rm p}(p^{-1/2}\gamma^{-1}\log p)$ if $\gamma \gg p^{-1/4}$, and $\max_{\ell\in[p]}|R_{\ell,16}| = O_{\rm p}(p^{-1}\gamma^{-3}\log p)$ if $p^{-1/4}\gtrsim\gamma\gg p^{-1/3}\log^{1/6}p$. Given (i,j) such that $i\neq j$, we define $Y_{(i,j),\ell} = -\nu_\ell^{-1/2}[\{(2\mu_{\ell,1})^{-1}+(2\mu_{\ell,2})^{-1}\}\mathring{Z}_{i,\ell}\mathring{Z}_{\ell,j}\mathring{Z}_{i,j} - \lambda_{i,\ell}\mathring{Z}_{i,\ell}]$ for any $\ell\neq i,j$, and $Y_{(i,j),\ell}=0$ for $\ell = i$ or j. Then $N^{1/2}\nu_\ell^{-1/2}(\hat{\theta}_\ell-\theta_\ell) = N^{-1/2}\sum_{i,j:\,i\neq j}Y_{(i,j),\ell} + R_{\ell,16} =: h_\ell + R_{\ell,16}$. Write $\mathbf{V} = {\rm diag}(\nu_1,\ldots,\nu_p)$, $\mathbf{h} = (h_1,\ldots,h_p)^{\rm T}$ and $\tilde{\mathbf{r}} = (R_{1,16},\ldots,R_{p,16})^{\rm T}$. Then

$$N^{1/2}\mathbf{V}^{-1/2}(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}) = \mathbf{h} + \tilde{\mathbf{r}}\,. \tag{E.2}$$

Lemma 4 gives the covariance matrix of the leading term $\mathbf{h}$, whose proof is given in the supplementary material.

Lemma 4. *It holds that* ${\rm Cov}(\mathbf{h}) = \mathbf{B}$ *where* $\mathbf{B} = (B_{\ell_1,\ell_2})_{p\times p}$ *satisfies* $\max_{1\leq\ell_1\neq\ell_2\leq p}|B_{\ell_1,\ell_2}|\lesssim p^{-1}$ *and* $B_{\ell,\ell}=1$ *for any* $\ell\in[p]$.

Now we begin to prove Theorem 3. Define $\varrho = \sup_{\mathbf{u}\in\mathbb{R}^p}|\mathbb{P}(\mathbf{h}\leq\mathbf{u}) - \mathbb{P}(\boldsymbol{\xi}\leq\mathbf{u})|$ with $\boldsymbol{\xi}\sim\mathcal{N}(\mathbf{0},\mathbf{I}_p)$. By (E.2), we have $\mathbb{P}\{N^{1/2}\mathbf{V}^{-1/2}(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})\leq\mathbf{u}\} = \mathbb{P}(\mathbf{h}+\tilde{\mathbf{r}}\leq\mathbf{u}, |\tilde{\mathbf{r}}|_\infty\leq\epsilon) + \mathbb{P}(\mathbf{h}+\tilde{\mathbf{r}}\leq\mathbf{u}, |\tilde{\mathbf{r}}|_\infty>\epsilon)$ for any $\epsilon>0$, which implies $\mathbb{P}\{N^{1/2}\mathbf{V}^{-1/2}(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})\leq\mathbf{u}\}\leq\mathbb{P}(\mathbf{h}\leq\mathbf{u}+\epsilon)+\mathbb{P}(|\tilde{\mathbf{r}}|_\infty>\epsilon)$ and $\mathbb{P}\{N^{1/2}\mathbf{V}^{-1/2}(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})\leq\mathbf{u}\}\geq\mathbb{P}(\mathbf{h}\leq\mathbf{u}-\epsilon, |\tilde{\mathbf{r}}|_\infty\leq\epsilon)\geq\mathbb{P}(\mathbf{h}\leq\mathbf{u}-\epsilon)-\mathbb{P}(|\tilde{\mathbf{r}}|_\infty>\epsilon)$. Therefore, $\mathbb{P}\{N^{1/2}\mathbf{V}^{-1/2}(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})\leq\mathbf{u}\}-\mathbb{P}(\boldsymbol{\xi}\leq\mathbf{u})\leq\mathbb{P}(\mathbf{h}\leq\mathbf{u}+\epsilon)-\mathbb{P}(\boldsymbol{\xi}\leq\mathbf{u}+\epsilon)+\mathbb{P}(\boldsymbol{\xi}\leq\mathbf{u}+\epsilon)-\mathbb{P}(\boldsymbol{\xi}\leq\mathbf{u})+\mathbb{P}(|\tilde{\mathbf{r}}|_\infty>\epsilon)\leq\varrho+\mathbb{P}(\boldsymbol{\xi}\leq\mathbf{u}+\epsilon)-\mathbb{P}(\boldsymbol{\xi}\leq\mathbf{u})+\mathbb{P}(|\tilde{\mathbf{r}}|_\infty>\epsilon)$ and $\mathbb{P}\{N^{1/2}\mathbf{V}^{-1/2}(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})\leq\mathbf{u}\}-\mathbb{P}(\boldsymbol{\xi}\leq\mathbf{u})\geq\mathbb{P}(\mathbf{h}\leq\mathbf{u}-\epsilon)-\mathbb{P}(\boldsymbol{\xi}\leq\mathbf{u}-\epsilon)+\mathbb{P}(\boldsymbol{\xi}\leq\mathbf{u}-\epsilon)-\mathbb{P}(\boldsymbol{\xi}\leq\mathbf{u})-\mathbb{P}(|\tilde{\mathbf{r}}|_\infty>\epsilon)\geq-\varrho+\mathbb{P}(\boldsymbol{\xi}\leq\mathbf{u}-\epsilon)-\mathbb{P}(\boldsymbol{\xi}\leq\mathbf{u})-\mathbb{P}(|\tilde{\mathbf{r}}|_\infty>\epsilon)$, which implies that

$$\begin{aligned}\sup_{\mathbf{u}\in\mathbb{R}^p}&\big|\mathbb{P}\{N^{1/2}\mathbf{V}^{-1/2}(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})\leq\mathbf{u}\}-\mathbb{P}(\boldsymbol{\xi}\leq\mathbf{u})\big|\\ &\leq\varrho+\sup_{\mathbf{u}\in\mathbb{R}^p}\big|\mathbb{P}(\boldsymbol{\xi}\leq\mathbf{u}+\epsilon)-\mathbb{P}(\boldsymbol{\xi}\leq\mathbf{u})\big|+\mathbb{P}(|\tilde{\mathbf{r}}|_\infty>\epsilon)\\ &\leq\varrho+C\epsilon\log^{1/2}p+\mathbb{P}(|\tilde{\mathbf{r}}|_\infty>\epsilon)\,.\end{aligned}\tag{E.3}$$

The last step is based on Nazarov's inequality. Recall $|\tilde{\mathbf{r}}|_\infty = O_{\rm p}(p^{-1/2}\gamma^{-1}\log p)+O_{\rm p}(p^{-1}\gamma^{-3}\log p)$. Since $\gamma\gg p^{-1/3}\log^{1/2}p$, then $p^{-1/2}\gamma^{-1}\log^{3/2}p + p^{-1}\gamma^{-3}\log^{3/2}p = o(1)$. There exists $\epsilon\to0$ such that $\epsilon\log^{1/2}p\to0$ and $p^{-1/2}\gamma^{-1}\log p+p^{-1}\gamma^{-3}\log p = o(\epsilon)$. For such selected ϵ, we have $C\epsilon\log^{1/2}p+\mathbb{P}(|\tilde{\mathbf{r}}|_\infty>\epsilon)\to0$ as $p\to\infty$.

In order to construct Theorem 3, we first need to show $\varrho\to0$ as $p\to\infty$. Let $\mathbf{g}\sim\mathcal{N}(\mathbf{0},\mathbf{B})$ with $\mathbf{B}$ specified in Lemma 4, and define $\bar{\varrho} = \sup_{\mathbf{u}\in\mathbb{R}^p}|\mathbb{P}(\mathbf{h}\leq\mathbf{u})-\mathbb{P}(\mathbf{g}\leq\mathbf{u})|$. Then $\varrho\leq\bar{\varrho}+\sup_{\mathbf{u}\in\mathbb{R}^p}|\mathbb{P}(\boldsymbol{\xi}\leq\mathbf{u})-\mathbb{P}(\mathbf{g}\leq\mathbf{u})|$. Recall that $\boldsymbol{\xi}\sim\mathcal{N}(\mathbf{0},\mathbf{I}_p)$ and $\mathbf{g}\sim\mathcal{N}(\mathbf{0},\mathbf{B})$ with $|\mathbf{I}_p-\mathbf{B}|_\infty\lesssim p^{-1}$, we have $\sup_{\mathbf{u}\in\mathbb{R}^p}|\mathbb{P}(\boldsymbol{\xi}\leq\mathbf{u})-\mathbb{P}(\mathbf{g}\leq\mathbf{u})|\lesssim|\mathbf{I}_p-\mathbf{B}|_\infty^{1/3}\log^{2/3}p\lesssim p^{-1/3}\log^{2/3}p\to0$. To show $\varrho\to0$, it suffices to show $\bar{\varrho}\to0$. Define $\varrho_* = \sup_{\mathbf{u}\in\mathbb{R}^p, v\in[0,1]}|\mathbb{P}(\sqrt{v}\mathbf{h}+\sqrt{1-v}\mathbf{g}\leq\mathbf{u})-\mathbb{P}(\mathbf{g}\leq\mathbf{u})|$ with $\mathbf{g}\sim\mathcal{N}(\mathbf{0},\mathbf{B})$. It is obvious that $\bar{\varrho}\leq\varrho_*$. We only need to show $\varrho_*\to0$ as $p\to\infty$. Let $\beta:=\phi\log p$. For a given $\mathbf{u} = (u_1,\ldots,u_p)^{\rm T}\in\mathbb{R}^p$, we define

$$F_\beta(\mathbf{y}) := \beta^{-1}\log\bigg[\sum_{\ell=1}^p\exp\{\beta(y_\ell-u_\ell)\}\bigg]\tag{E.4}$$

for any $\mathbf{y} = (y_1,\ldots,y_p)^{\rm T}\in\mathbb{R}^p$. Such defined function $F_\beta(\mathbf{y})$ satisfies the property $0\leq F_\beta(\mathbf{y})-$

$\max_{\ell\in[p]}(y_\ell - u_\ell) \le \beta^{-1}\log p = \phi^{-1}$ for any $\mathbf{y}\in\mathbb{R}^p$. Select a thrice continuously differentiable function $f_0:\mathbb{R}\to[0,1]$ whose derivatives up to the third order are all bounded such that $f_0(t)=1$ for $t\le 0$ and $f_0(t)=0$ for $t\ge 1$. Define $f(t):=f_0(\phi t)$ for any $t\in\mathbb{R}$, and $q(\mathbf{y}):=f\{F_\beta(\mathbf{y})\}$ for any $\mathbf{y}\in\mathbb{R}^p$. To simplify the notation, we write $q_\ell(\mathbf{y})=\partial q(\mathbf{y})/\partial y_\ell$, $q_{\ell k}(\mathbf{y})=\partial^2 q(\mathbf{y})/\partial y_\ell\partial y_k$ and $q_{\ell kl}(\mathbf{y})=\partial^3 q(\mathbf{y})/\partial y_\ell \partial y_k \partial y_l$. Let $\tilde{\mathbf{g}}$ be an independent copy of $\mathbf{g}$. Define $\mathcal{T}:=q(\sqrt{v}\mathbf{h}+\sqrt{1-v}\mathbf{g})-q(\tilde{\mathbf{g}})$. Write $\boldsymbol{\delta}=\sqrt{v}\mathbf{h}+\sqrt{1-v}\mathbf{g}$. Notice that $\mathbb{P}(\boldsymbol{\delta}\le\mathbf{u}-\phi^{-1})\le\mathbb{P}\{F_\beta(\boldsymbol{\delta})\le 0\}\le\mathbb{E}\{q(\boldsymbol{\delta})\}\le\mathbb{P}\{F_\beta(\tilde{\mathbf{g}})\le\phi^{-1}\}+\mathbb{E}(\mathcal{T})\le\mathbb{P}(\tilde{\mathbf{g}}\le\mathbf{u}+\phi^{-1})+|\mathbb{E}(\mathcal{T})|\le \mathbb{P}(\tilde{\mathbf{g}}\le\mathbf{u}-\phi^{-1})+C\phi^{-1}\log^{1/2}p+|\mathbb{E}(\mathcal{T})|$ and $\mathbb{P}(\boldsymbol{\delta}\le\mathbf{u}-\phi^{-1})\ge\mathbb{P}(\tilde{\mathbf{g}}\le\mathbf{u}-\phi^{-1})-C\phi^{-1}\log^{1/2}p-|\mathbb{E}(\mathcal{T})|$, then we have

$$\varrho_*\le C\phi^{-1}\log^{1/2}p+\sup_{v\in[0,1]}|\mathbb{E}(\mathcal{T})|\,. \tag{E.5}$$

In the sequel, we will give an upper bound for $\sup_{v\in[0,1]}|\mathbb{E}(\mathcal{T})|$.

To do this, we first generate mean zero normal distributed random variables $\{V_{(i,j),\ell}\}_{i,j,\ell:\,i\ne j\ne\ell}$ that independent of the sequence $\{Y_{(i,j),\ell}\}_{i,j,\ell:\,i\ne j\ne\ell}$ such that $\mathrm{Cov}\{V_{(i_1,j_1),\ell_1},V_{(i_2,j_2),\ell_2}\}=\mathrm{Cov}\{Y_{(i_1,j_1),\ell_1},Y_{(i_2,j_2),\ell_2}\}$ for any (i_1,j_1,ℓ_1) and (i_2,j_2,ℓ_2) such that $i_1\ne j_1\ne\ell_1$ and $i_2\ne j_2\ne\ell_2$. We set $V_{(i,j),\ell}=0$ if $\ell=i$ or j. Let $\{W_{(i,j),\ell}\}_{i,j,\ell:\,i\ne j\ne\ell}$ be an independent copy of $\{V_{(i,j),\ell}\}_{i,j,\ell:\,i\ne j\ne\ell}$. We also set $W_{(i,j),\ell}=0$ if $\ell=i$ or j. For each (i,j) such that $i\ne j$, define three p-dimensional vectors $\mathbf{y}_{(i,j)}=\{Y_{(i,j),1},\ldots,Y_{(i,j),p}\}^{\mathrm{T}}$, $\mathbf{v}_{(i,j)}=\{V_{(i,j),1},\ldots,V_{(i,j),p}\}^{\mathrm{T}}$ and $\mathbf{w}_{(i,j)}=\{W_{(i,j),1},\ldots,W_{(i,j),p}\}^{\mathrm{T}}$. We know that $N^{-1/2}\sum_{i,j:\,i\ne j}\mathbf{v}_{(i,j)}\sim\mathcal{N}(\mathbf{0},\mathbf{B})$ and $N^{-1/2}\sum_{i,j:\,i\ne j}\mathbf{w}_{(i,j)}\sim\mathcal{N}(\mathbf{0},\mathbf{B})$. We let $\mathbf{g}=N^{-1/2}\sum_{i,j:\,i\ne j}\mathbf{v}_{(i,j)}$ and $\tilde{\mathbf{g}}=N^{-1/2}\sum_{i,j:\,i\ne j}\mathbf{w}_{(i,j)}$.

Define $\mathbf{c}(t)=\sum_{i,j:\,i\ne j}\mathbf{c}_{(i,j)}(t)$ for any $t\in[0,1]$, where $\mathbf{c}_{(i,j)}(t):=N^{-1/2}[\sqrt{t}\{\sqrt{v}\mathbf{y}_{(i,j)}+\sqrt{1-v}\mathbf{v}_{(i,j)}\}+\sqrt{1-t}\mathbf{w}_{(i,j)}]$. Write $\mathbf{c}_{(i,j)}(t)=\{c_{(i,j),1}(t),\ldots,c_{(i,j),p}(t)\}^{\mathrm{T}}$. Then $\mathbf{c}(1)=\sqrt{v}\mathbf{h}+\sqrt{1-v}\mathbf{g}$ and $\mathbf{c}(0)=\tilde{\mathbf{g}}$. Define $\dot{\mathbf{c}}_{(i,j)}(t):=N^{-1/2}[t^{-1/2}\{\sqrt{v}\mathbf{y}_{(i,j)}+\sqrt{1-v}\mathbf{v}_{(i,j)}\}-(1-t)^{-1/2}\mathbf{w}_{(i,j)}]$ and write $\dot{\mathbf{c}}_{(i,j)}(t)=\{\dot{c}_{(i,j),1}(t),\ldots,\dot{c}_{(i,j),p}(t)\}^{\mathrm{T}}$. Then

$$\mathcal{T}=\int_0^1\frac{\mathrm{d}q\{\mathbf{c}(t)\}}{\mathrm{d}t}\mathrm{d}t=\frac{1}{2}\sum_{i,j:\,i\ne j}\sum_{\ell=1}^p\int_0^1 q_\ell\{\mathbf{c}(t)\}\dot{c}_{(i,j),\ell}(t)\,\mathrm{d}t\,,$$

which implies $2\mathbb{E}(\mathcal{T})=\sum_{i,j:\,i\ne j}\sum_{\ell=1}^p\int_0^1\mathbb{E}[q_\ell\{\mathbf{c}(t)\}\dot{c}_{(i,j),\ell}(t)]\,\mathrm{d}t$. For given (i,j) such that $i\ne j$, and ℓ, we have $\dot{c}_{(i,j),\ell}(t)=N^{-1/2}[t^{-1/2}\{\sqrt{v}Y_{(i,j),\ell}+\sqrt{1-v}V_{(i,j),\ell}\}-(1-t)^{-1/2}W_{(i,j),\ell}]$. Recall $Y_{(i,j),\ell}=V_{(i,j),\ell}=W_{(i,j),\ell}=0$ for $\ell=i$ or j. It then holds that

$$\mathbb{E}(\mathcal{T})=\frac{1}{2}\sum_{i,j:\,i\ne j}\sum_{\ell:\,\ell\ne i,j}\int_0^1\mathbb{E}[q_\ell\{\mathbf{c}(t)\}\dot{c}_{(i,j),\ell}(t)]\,\mathrm{d}t\,. \tag{E.6}$$

Notice that $Y_{(i,j),\ell}$ is a function of $\{\mathring{Z}_{i,\ell},\mathring{Z}_{\ell,j},\mathring{Z}_{i,j}\}$. Given (i,j,ℓ) such that $i\ne j\ne\ell$, we first consider $\{\mathring{Z}_{i,\ell},\mathring{Z}_{\ell,j},\mathring{Z}_{i,j}\}$ will appear in which $Y_{(i',j'),\ell'}$'s. Recall $Y_{(i',j'),\ell'}=-\nu_{\ell'}^{-1/2}[\{(2\mu_{\ell',1})^{-1}+(2\mu_{\ell',2})^{-1}\}\mathring{Z}_{i',\ell'}\mathring{Z}_{\ell',j'}\mathring{Z}_{i',j'}-\lambda_{i',\ell'}\mathring{Z}_{i',\ell'}]$. Then $\mathring{Z}_{i,\ell}$ will appear in $Y_{(i',j'),\ell'}$ such that either $(i',\ell')=(i,\ell)$, $(j',\ell')=(\ell,i)$ or $(i',j')=(i,\ell)$ holds. Since $\mathring{Z}_{\ell,i}=\mathring{Z}_{i,\ell}$, we know $\mathring{Z}_{i,\ell}$ is also not independent of $Y_{(i',j'),\ell'}$ such that either $(i',\ell')=(\ell,i)$, $(j',\ell')=(i,\ell)$ or $(i',j')=(\ell,i)$ holds. Given i and ℓ such that $i\ne\ell$, let $\mathcal{S}_*(i,\ell)=\{(i',j',\ell'):\{i',j'\}=\{i,\ell\}\}\cup\{(i',j',\ell'):\{j',\ell'\}=\{i,\ell\}\}\cup\{(i',j',\ell'):\{\ell',i'\}=\{i,\ell\}\}$. Then we have $\mathring{Z}_{i,\ell}$ is independent of $\{Y_{(i',j'),\ell'}\}_{(i',j',\ell')\notin\mathcal{S}_*(i,\ell)}$. For any (i,j,ℓ) such that $i\ne j\ne\ell$, define $\mathcal{S}(i,j,\ell)=\mathcal{S}_*(i,j)\cup\mathcal{S}_*(j,\ell)\cup\mathcal{S}_*(\ell,i)$. We know $Y_{(i,j),\ell}$ is independent of $\{Y_{(i',j'),\ell'}\}_{(i',j',\ell')\notin\mathcal{S}(i,j,\ell)}$. For any (i,j,ℓ) and (i',j',ℓ') such that $i\ne j\ne\ell$ and $i'\ne j'\ne\ell'$, let $a_{(i',j'),\ell'}^{(i,j,\ell)}=I\{(i',j',\ell')\in$

$\mathcal{S}(i,j,\ell)\}$. For given (i,j,ℓ) such that $i \neq j \neq \ell$, we set $a^{(i,j,\ell)}_{(i',j'),\ell'} = 0$ if $\ell' \in \{i',j'\}$. Write $\mathbf{a}^{(i,j,\ell)}_{(i',j')} = \{a^{(i,j,\ell)}_{(i',j'),1},\ldots,a^{(i,j,\ell)}_{(i',j'),p}\}^{\mathrm{T}}$. and define $\mathbf{c}^{(i,j),\ell}(t) = \sum_{i',j':\,i'\neq j'} \mathbf{c}_{(i',j')}(t)\circ\mathbf{a}^{(i,j,\ell)}_{(i',j')}$, where $\circ$ denotes the Hadamard product. Let $\mathbf{c}^{-(i,j),\ell}(t) = \mathbf{c}(t) - \mathbf{c}^{(i,j),\ell}(t)$. We can see that $\mathbf{c}^{-(i,j),\ell}(t)$ is independent of $\{\mathring{Z}_{i,\ell},\mathring{Z}_{\ell,j},\mathring{Z}_{i,j}\}$. Write $\mathbf{c}^{(i,j),\ell}(t) = \{c^{(i,j),\ell}_1(t),\ldots,c^{(i,j),\ell}_p(t)\}^{\mathrm{T}}$. It follows from Taylor expansion that

$$
\begin{aligned}
&\int_0^1 \mathbb{E}[q_\ell\{\mathbf{c}(t)\}\dot{c}_{(i,j),\ell}(t)]\,\mathrm{d}t \\
&= \underbrace{\int_0^1 \mathbb{E}[q_\ell\{\mathbf{c}^{-(i,j),\ell}(t)\}\dot{c}_{(i,j),\ell}(t)]\,\mathrm{d}t}_{\mathrm{I}_1(i,j,\ell)} + \sum_{k=1}^p \underbrace{\int_0^1 \mathbb{E}[q_{\ell k}\{\mathbf{c}^{-(i,j),\ell}(t)\}\dot{c}_{(i,j),\ell}(t)c^{(i,j),\ell}_k(t)]\,\mathrm{d}t}_{\mathrm{I}_2(i,j,\ell,k)} \qquad \text{(E.7)} \\
&\quad + \sum_{k,l=1}^p \underbrace{\int_0^1\int_0^1 (1-\tau)\mathbb{E}[q_{\ell kl}\{\mathbf{c}^{-(i,j),\ell}(t)+\tau\mathbf{c}^{(i,j),\ell}(t)\}\dot{c}_{(i,j),\ell}(t)c^{(i,j),\ell}_k(t)c^{(i,j),\ell}_l(t)]\,\mathrm{d}\tau\mathrm{d}t}_{\mathrm{I}_3(i,j,\ell,k,l)}\,.
\end{aligned}
$$

Together with (E.6), we have $2\mathbb{E}(\mathcal{T}) = \sum_{i,j:\,i\neq j}\sum_{\ell:\,\ell\neq i,j}\mathrm{I}_1(i,j,\ell) + \sum_{i,j:\,i\neq j}\sum_{\ell:\,\ell\neq i,j}\sum_{k=1}^p \mathrm{I}_2(i,j,\ell,k) + \sum_{i,j:\,i\neq j}\sum_{\ell:\,\ell\neq i,j}\sum_{k,l=1}^p \mathrm{I}_3(i,j,\ell,k,l)$. As shown in the supplementary material, $\sum_{i,j:\,i\neq j}\sum_{\ell:\,\ell\neq i,j}\mathrm{I}_1(i,j,\ell) = 0$, $\sum_{i,j:\,i\neq j}\sum_{\ell:\,\ell\neq i,j}\sum_{k=1}^p |\mathrm{I}_2(i,j,\ell,k)| \lesssim p^{-1/2}\phi^3\log^{7/2}p$, and $\sum_{i,j:\,i\neq j}\sum_{\ell:\,\ell\neq i,j}\sum_{k,l=1}^p |\mathrm{I}_3(i,j,\ell,k,l)| \lesssim p^{-1/2}\phi^3\log^{7/2}p$. Then $\sup_{v\in[0,1]}|\mathbb{E}(\mathcal{T})| \lesssim p^{-1/2}\phi^3\log^{7/2}p$. Together with (E.5), selecting $\phi = p^{1/8}\log^{-3/4}p$, we have $\varrho_* = \sup_{\mathbf{u}\in\mathbb{R}^p,v\in[0,1]}|\mathbb{P}(\sqrt{v}\mathbf{h}+\sqrt{1-v}\mathbf{g}\le\mathbf{u}) - \mathbb{P}(\mathbf{g}\le\mathbf{u})| \lesssim \phi^{-1}\log^{1/2}p + p^{-1/2}\phi^3\log^{7/2}p \asymp p^{-1/8}\log^{5/4}p$. Hence, $\sup_{\mathbf{u}\in\mathbb{R}^p}|\mathbb{P}\{N^{1/2}\mathbf{V}^{-1/2}(\hat{\boldsymbol\theta}-\boldsymbol\theta)\le\mathbf{u}\} - \mathbb{P}(\boldsymbol\xi\le\mathbf{u})| \to 0$ as $p\to\infty$. On the other hand, since $\delta \ll (p\log p)^{-1}$, we have $N^{1/2}|\{(\mathbf{V}^\dagger)^{-1/2}-\mathbf{V}^{-1/2}\}(\hat{\boldsymbol\theta}-\boldsymbol\theta)|_\infty = o_{\mathrm{p}}\{(\log p)^{-1/2}\}$. Following the same arguments for deriving (E.3), we have $\sup_{\mathbf{u}\in\mathbb{R}^p}|\mathbb{P}\{N^{1/2}(\mathbf{V}^\dagger)^{-1/2}(\hat{\boldsymbol\theta}-\boldsymbol\theta)\le\mathbf{u}\} - \mathbb{P}(\boldsymbol\xi\le\mathbf{u})| \to 0$ as $p\to\infty$. We complete the proof of Theorem 3. □

References

Blocki, J., Blum, A., Datta, A. and Sheffet, O. (2013). Differentially private data analysis of social networks via restricted sensitivity. *Proceedings of the 4th Conference on Innovations in Theoretical Computer Science*, 87–96.

Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society Series B*, **57**, 289–300.

Chang, J., Chen, X. and Wu, M. (2021). Central limit theorems for high dimensional dependent data. Available at arXiv:2104.12929.

Chang, J., Kolaczyk, E. D. and Yao, Q. (2020). Discussion of 'Network cross-validation by edge sampling'. *Biometrika*, 107(2), 277-280.

Chang, J., Kolaczyk, E. D. and Yao, Q. (2021). Estimation of subgraph densities in noisy networks. *Journal of the American Statistical Association*, in press.

Chang, J., Qiu, Y., Yao, Q. and Zou, T. (2018). Confidence regions for entries of a large precision matrix. *Journal of Econometrics*, **206**, 57–82.

Chang, J., Zheng, C., Zhou, W.-X. and Zhou, W. (2017). Simulation-based hypothesis testing of high dimensional means under covariance heterogeneity. *Biometrics*, **73**, 1300–1310.

Chatterjee, S., Diaconis, P. and Sly, A. (2011). Random graphs with a given degree sequence. *The Annals* *of Applied Probability*, **21**, 1400–1435.

Chernozhukov, V., Chetverikov, D. and Kato, K. (2013). Gaussian approximations and multiplier bootstrap in high dimensions. *The Annals of Statistics*, **41**, 2786–2819.

de la Peña, V. and Montgomery-Smith, S. (1995). Decoupling inequalities for the tail probabilities of multivariate U-statistics. *The Annals of Probability*, **23**, 806–816.

Dwork, C. (2006). Dierential privacy. *33rd International Colloquium on Automata, languages and programming*, 1-12.

Dwork, C., McSherry, F., Nissim, K. and Smith, A. (2006). Calibrating noise to sensitivity in private data analysis. *Proceedings of the 3rd Theory of Cryptography Conference*, 265–248.

Giné, E., Latała, R. and Zinn, J. (2000). Exponential and moment inequalities for U-statistics. In *High Dimensional Probability* II (E. Giné, D. M. Mason and J. A. Wellner, eds.) Progress in Probability, vol 47. Birkhäuser, Boston, MA.

Hennig, C. (2007). Cluster-wise assessment of cluster stability. *Computational Statistics & Data Analysis*, **52**, 258–271.

Jiang, H., Pei, J., Yu, D., Yu, J., Gong, B. and Cheng, X. (2020). Differential privacy and its applications in social network analysis: A survey. arXiv e-prints, arXiv-2010.

Karwa, V., Krivitsky, P. N. and Slavković, A. (2017). Sharing social network data: differentially private estimation of exponential family random graph models. *Journal of the Royal Statistical Society*, **C**, **66**, 481–500.

Karwa, V. and Slavković, A. (2016). Inference using noisy degrees: differentially private β-model and synthetic graphs. *The Annals of Statistics*, **44**, 87–112.

Liu, P., Xu, Y., Jiang, Q., Tang, Y., Guo, Y., Wang, L. and Li, X. (2020). Local differential privacy for social network publishing. *Neurocomputing*, **391**, 273–279.

Nissim, K., Raskhodnikova, S. and Smith, A. (2007). Smooth sensitivity and sampling in private data analysis. In Proceedings of the Thirty-ninth Annual ACM symposium on Theory of Computing (pp. 75-84).

Rider, P. R. (1960). Variance of the median of small samples from several special populations. *Journal of the American Statistical Association*, **55**, 148–150.

Wasserman, L. and Zhou, S. (2010). A statistical framework for differential privacy. *Journal of the American Statistical Association*, **105**, 375–397.

Yan, T. and Xu, J. (2013). A central limit theorem in the β-model for undirected random graphs with a diverging number of vertices. *Biometrika*, **100**, 519–524.

Supplementary Material for "Edge Differentially Private Estimation in the β-model via Jittering and Method of Moments"

In the sequel, we use C and $\tilde{C}$ to denote generic positive finite universal constants that may be different in different uses.

S1 Proof of Lemma 4

For any $\ell_1, \ell_2 \in [p]$, we have

$$
\begin{aligned}
B_{\ell_1,\ell_2} = & \underbrace{\frac{1}{N} \sum_{i,j:\, i\neq j,\, i,j\neq \ell_1,\ell_2} \mathrm{Cov}\{Y_{(i,j),\ell_1}, Y_{(i,j),\ell_2}\}}_{B^{(1)}_{\ell_1,\ell_2}} \\
& + \underbrace{\frac{1}{N} \sum_{\substack{i_1,j_1,i_2,j_2:\, i_1\neq j_1,\, i_2\neq j_2, \\ (i_1,j_1)\neq(i_2,j_2),\, i_1,j_1\neq \ell_1,\, i_2,j_2\neq \ell_2}} \mathrm{Cov}\{Y_{(i_1,j_1),\ell_1}, Y_{(i_2,j_2),\ell_2}\}}_{B^{(2)}_{\ell_1,\ell_2}} \,.
\end{aligned}
$$

Since $\mathbb{E}\{Y_{(i,j),\ell}\} = 0$ for any (i,j,ℓ) such that $i \neq j$ and $\ell \neq i,j$, it holds that

$$
\begin{aligned}
\nu_{\ell_1}^{1/2}\nu_{\ell_2}^{1/2}&\mathrm{Cov}\{Y_{(i_1,j_1),\ell_1}, Y_{(i_2,j_2),\ell_2}\} = \nu_{\ell_1}^{1/2}\nu_{\ell_2}^{1/2}\mathbb{E}\{Y_{(i_1,j_1),\ell_1} Y_{(i_2,j_2),\ell_2}\} \\
&= \bigg(\frac{\mu_{\ell_1,1}+\mu_{\ell_1,2}}{2\mu_{\ell_1,1}\mu_{\ell_1,2}}\bigg)\bigg(\frac{\mu_{\ell_2,1}+\mu_{\ell_2,2}}{2\mu_{\ell_2,1}\mu_{\ell_2,2}}\bigg)\mathbb{E}(\mathring{Z}_{i_1,\ell_1}\mathring{Z}_{\ell_1,j_1}\mathring{Z}_{i_1,j_1}\mathring{Z}_{i_2,\ell_2}\mathring{Z}_{\ell_2,j_2}\mathring{Z}_{i_2,j_2}) \\
&\quad - \bigg(\frac{\mu_{\ell_2,1}+\mu_{\ell_2,2}}{2\mu_{\ell_2,1}\mu_{\ell_2,2}}\bigg)\lambda_{i_1,\ell_1}\mathbb{E}(\mathring{Z}_{i_1,\ell_1}\mathring{Z}_{i_2,\ell_2}\mathring{Z}_{\ell_2,j_2}\mathring{Z}_{i_2,j_2}) \\
&\quad - \bigg(\frac{\mu_{\ell_1,1}+\mu_{\ell_1,2}}{2\mu_{\ell_1,1}\mu_{\ell_1,2}}\bigg)\lambda_{i_2,\ell_2}\mathbb{E}(\mathring{Z}_{i_2,\ell_2}\mathring{Z}_{i_1,\ell_1}\mathring{Z}_{\ell_1,j_1}\mathring{Z}_{i_1,j_1}) + \lambda_{i_1,\ell_1}\lambda_{i_2,\ell_2}\mathbb{E}(\mathring{Z}_{i_1,\ell_1}\mathring{Z}_{i_2,\ell_2})
\end{aligned}
$$

for any $(i_1, j_1, \ell_1, i_2, j_2, \ell_2)$ such that $i_1 \neq j_1$, $i_2 \neq j_2$, $i_1, j_1 \neq \ell_1$ and $i_2, j_2 \neq \ell_2$.

Case 1. $(i_1, j_1) = (i_2, j_2)$. In this case, we write $i_1 = i_2 = i$ and $j_1 = j_2 = j$. For any (i, j, ℓ_1, ℓ_2) such that $i \neq j$, $\ell_1 \neq \ell_2$ and $i, j \neq \ell_1, \ell_2$, we have $\mathbb{E}(\mathring{Z}_{i,\ell_1}\mathring{Z}_{\ell_1,j}\mathring{Z}_{i,\ell_2}\mathring{Z}_{\ell_2,j}\mathring{Z}^2_{i,j}) = 0$, $\mathbb{E}(\mathring{Z}_{i,\ell_1}\mathring{Z}_{i,\ell_2}\mathring{Z}_{\ell_2,j}\mathring{Z}_{i,j}) = 0$, $\mathbb{E}(\mathring{Z}_{i,\ell_2}\mathring{Z}_{i,\ell_1}\mathring{Z}_{\ell_1,j}\mathring{Z}_{i,j}) = 0$ and $\mathbb{E}(\mathring{Z}_{i,\ell_1}\mathring{Z}_{i,\ell_2}) = 0$, which implies

$$
B^{(1)}_{\ell_1,\ell_2} = 0 \tag{S.1.1}
$$

for any $\ell_1 \neq \ell_2$. For $\ell_1 = \ell_2$, we write $\ell_1 = \ell_2 = \ell$. Then $\nu_\ell \mathrm{Cov}\{Y_{(i,j),\ell}, Y_{(i,j),\ell}\} = \{(\mu_{\ell,1} + \mu_{\ell,2})/(2\mu_{\ell,1}\mu_{\ell,2})\}^2$ $\mathrm{Var}(Z_{i,\ell})\mathrm{Var}(Z_{\ell,j})\mathrm{Var}(Z_{i,j}) + \lambda^2_{i,\ell}\mathrm{Var}(Z_{i,\ell})$ for any (i,j,ℓ) such that $i \neq j$ and $i, j \neq \ell$. Thus, for any $\ell \in [p]$, we have

$$
\begin{aligned}
\nu_\ell B^{(1)}_{\ell,\ell} = & \frac{1}{N}\bigg(\frac{\mu_{\ell,1}+\mu_{\ell,2}}{2\mu_{\ell,1}\mu_{\ell,2}}\bigg)^2 \sum_{i,j:\, i\neq j,\, i,j\neq \ell} \mathrm{Var}(Z_{i,\ell})\mathrm{Var}(Z_{\ell,j})\mathrm{Var}(Z_{i,j}) \\
& + \frac{1}{N}\sum_{i,j:\, i\neq j,\, i,j\neq \ell} \lambda^2_{i,\ell}\mathrm{Var}(Z_{i,\ell}) = \frac{\tilde{b}_\ell}{2} + b_\ell \,.
\end{aligned} \tag{S.1.2}
$$

Case 2. $i_1 = i_2$, $j_1 \neq j_2$. In this case, we write $i_1 = i_2 = i$. For any $(i, j_1, j_2, \ell_1, \ell_2)$ such that

$j_1, j_2 \neq i$, $j_1 \neq j_2$, $\ell_1 \neq \ell_2$, $i, j_1 \neq \ell_1$ and $i, j_2 \neq \ell_2$, we have $\mathbb{E}(\mathring{Z}_{i,\ell_1}\mathring{Z}_{\ell_1,j_1}\mathring{Z}_{i,j_1}\mathring{Z}_{i,\ell_2}\mathring{Z}_{\ell_2,j_2}\mathring{Z}_{i,j_2}) = \mathrm{Var}(Z_{i,\ell_1})\mathrm{Var}(Z_{\ell_1,\ell_2})\mathrm{Var}(Z_{i,\ell_2})I(j_2 = \ell_1, j_1 = \ell_2)$, $\mathbb{E}(\mathring{Z}_{i,\ell_1}\mathring{Z}_{i,\ell_2}\mathring{Z}_{\ell_2,j_2}\mathring{Z}_{i,j_2}) = 0$, $\mathbb{E}(\mathring{Z}_{i,\ell_2}\mathring{Z}_{i,\ell_1}\mathring{Z}_{\ell_1,j_1}\mathring{Z}_{i,j_1}) = 0$ and $\mathbb{E}(\mathring{Z}_{i,\ell_1}\mathring{Z}_{i,\ell_2}) = 0$, which implies that

$$\begin{aligned} &\frac{\nu_{\ell_1}^{1/2}\nu_{\ell_2}^{1/2}}{N}\sum_{\substack{i,j_1,j_2:\, i\neq j_1,j_2,\, j_1\neq j_2,\\ i,j_1\neq \ell_1,\, i,j_2\neq \ell_2}} \mathrm{Cov}\{Y_{(i,j_1),\ell_1}, Y_{(i,j_2),\ell_2}\} \\ &= \bigg(\frac{\mu_{\ell_1,1}+\mu_{\ell_1,2}}{2\mu_{\ell_1,1}\mu_{\ell_1,2}}\bigg)\bigg(\frac{\mu_{\ell_2,1}+\mu_{\ell_2,2}}{2\mu_{\ell_2,1}\mu_{\ell_2,2}}\bigg)\frac{1}{N}\sum_{i:\, i\neq \ell_1,\ell_2}\mathrm{Var}(Z_{i,\ell_1})\mathrm{Var}(Z_{\ell_1,\ell_2})\mathrm{Var}(Z_{i,\ell_2}) \asymp \gamma^{-6}p^{-1} \end{aligned} \tag{S.1.3}$$

for any $\ell_1 \neq \ell_2$. For $\ell_1 = \ell_2$, we write $\ell_1 = \ell_2 = \ell$. It then holds that $\mathbb{E}(\mathring{Z}_{i,\ell}^2\mathring{Z}_{\ell,j_1}\mathring{Z}_{i,j_1}\mathring{Z}_{\ell,j_2}\mathring{Z}_{i,j_2}) = 0$, $\mathbb{E}(\mathring{Z}_{i,\ell}^2\mathring{Z}_{\ell,j_2}\mathring{Z}_{i,j_2}) = 0$, $\mathbb{E}(\mathring{Z}_{i,\ell}^2\mathring{Z}_{\ell,j_1}\mathring{Z}_{i,j_1}) = 0$ and $\mathbb{E}(\mathring{Z}_{i,\ell}^2) = \mathrm{Var}(Z_{i,\ell})$ for any (i, j_1, j_2, ℓ) such that $j_1, j_2 \neq i$, $j_1 \neq j_2$ and $i, j_1, j_2 \neq \ell$, which implies that for any $\ell \in [p]$, we have

$$\begin{aligned} &\frac{\nu_\ell}{N}\sum_{\substack{i,j_1,j_2:\, i\neq j_1,j_2,\, j_1\neq j_2,\\ i,j_1,j_2\neq \ell}} \mathrm{Cov}\{Y_{(i,j_1),\ell}, Y_{(i,j_2),\ell}\} \\ &= \frac{1}{N}\sum_{\substack{i,j_1,j_2:\, i\neq j_1,j_2,\, j_1\neq j_2,\\ i,j_1,j_2\neq \ell}} \lambda_{i,\ell}^2\mathbb{E}(\mathring{Z}_{i,\ell}^2) = \frac{p-3}{p-1}\sum_{i:\, i\neq \ell}\lambda_{i,\ell}^2\mathrm{Var}(Z_{i,\ell}) = (p-3)b_\ell\,. \end{aligned} \tag{S.1.4}$$

Case 3. $i_1 \neq i_2$, $j_1 = j_2$. In this case, we write $j_1 = j_2 = j$. For any $(i_1, i_2, j, \ell_1, \ell_2)$ such that $i_1, i_2 \neq j$, $i_1 \neq i_2$, $\ell_1 \neq \ell_2$, $j, i_1 \neq \ell_1$ and $j, i_2 \neq \ell_2$, we have $\mathbb{E}(\mathring{Z}_{i_1,\ell_1}\mathring{Z}_{\ell_1,j}\mathring{Z}_{i_1,j}\mathring{Z}_{i_2,\ell_2}\mathring{Z}_{\ell_2,j}\mathring{Z}_{i_2,j}) = \mathrm{Var}(Z_{j,\ell_1})\mathrm{Var}(Z_{\ell_1,\ell_2})\mathrm{Var}(Z_{j,\ell_2})I(i_2 = \ell_1, i_1 = \ell_2)$, $\mathbb{E}(\mathring{Z}_{i_1,\ell_1}\mathring{Z}_{i_2,\ell_2}\mathring{Z}_{\ell_2,j}\mathring{Z}_{i_2,j}) = 0$, $\mathbb{E}(\mathring{Z}_{i_2,\ell_2}\mathring{Z}_{i_1,\ell_1}\mathring{Z}_{\ell_1,j}\mathring{Z}_{i_1,j}) = 0$ and $\mathbb{E}(\mathring{Z}_{i_1,\ell_1}\mathring{Z}_{i_2,\ell_2}) = \mathrm{Var}(Z_{\ell_1,\ell_2})I(i_2 = \ell_1, i_1 = \ell_2)$, which implies that

$$\begin{aligned} &\frac{\nu_{\ell_1}^{1/2}\nu_{\ell_2}^{1/2}}{N}\sum_{\substack{i_1,i_2,j:\, j\neq i_1,i_2,\, i_1\neq i_2,\\ j,i_1\neq \ell_1,\, j,i_2\neq \ell_2}} \mathrm{Cov}\{Y_{(i_1,j),\ell_1}, Y_{(i_2,j),\ell_2}\} \\ &= \bigg(\frac{\mu_{\ell_1,1}+\mu_{\ell_1,2}}{2\mu_{\ell_1,1}\mu_{\ell_1,2}}\bigg)\bigg(\frac{\mu_{\ell_2,1}+\mu_{\ell_2,2}}{2\mu_{\ell_2,1}\mu_{\ell_2,2}}\bigg)\frac{1}{N}\sum_{j:\, j\neq \ell_1,\ell_2}\mathrm{Var}(Z_{j,\ell_1})\mathrm{Var}(Z_{\ell_1,\ell_2})\mathrm{Var}(Z_{j,\ell_2}) \\ &\quad + \frac{\lambda_{\ell_1,\ell_2}^2}{N}\sum_{j:\, j\neq \ell_1,\ell_2}\mathrm{Var}(Z_{\ell_1,\ell_2}) \asymp \gamma^{-6}p^{-1} + \gamma^{-2}p^{-1} \asymp \gamma^{-6}p^{-1} \end{aligned} \tag{S.1.5}$$

for any $\ell_1 \neq \ell_2$. For $\ell_1 = \ell_2$, we write $\ell_1 = \ell_2 = \ell$. It then holds that $\mathbb{E}(\mathring{Z}_{i_1,\ell}\mathring{Z}_{\ell,j}^2\mathring{Z}_{i_1,j}\mathring{Z}_{i_2,\ell}\mathring{Z}_{i_2,j}) = 0$, $\mathbb{E}(\mathring{Z}_{i_1,\ell}\mathring{Z}_{i_2,\ell}\mathring{Z}_{\ell,j}\mathring{Z}_{i_2,j}) = 0$, $\mathbb{E}(\mathring{Z}_{i_2,\ell}\mathring{Z}_{i_1,\ell}\mathring{Z}_{\ell,j}\mathring{Z}_{i_1,j}) = 0$ and $\mathbb{E}(\mathring{Z}_{i_1,\ell}\mathring{Z}_{i_2,\ell}) = 0$ for any (i_1, i_2, j, ℓ) such that $i_1, i_2 \neq j$, $i_1 \neq i_2$ and $j, i_1, i_2 \neq \ell$, which implies that for any $\ell \in [p]$, it holds that

$$\frac{\nu_\ell}{N}\sum_{\substack{i_1,i_2,j:\, j\neq i_1,i_2,\, i_1\neq i_2,\\ j,i_1,i_2\neq \ell}} \mathrm{Cov}\{Y_{(i_1,j),\ell}, Y_{(i_2,j),\ell}\} = 0\,. \tag{S.1.6}$$

Case 4. $i_1 \neq i_2$, $j_1 \neq j_2$. For any $\ell_1 \neq \ell_2$, it holds that $\mathbb{E}(\mathring{Z}_{i_1,\ell_1}\mathring{Z}_{\ell_1,j_1}\mathring{Z}_{i_1,j_1}\mathring{Z}_{i_2,\ell_2}\mathring{Z}_{\ell_2,j_2}\mathring{Z}_{i_2,j_2}) = \mathrm{Var}(Z_{i_1,\ell_1})$ $\mathrm{Var}(Z_{\ell_1,\ell_2})\mathrm{Var}(Z_{i_1,\ell_2})I(i_1 \neq \ell_2, j_2 = i_1, i_2 = \ell_1, j_1 = \ell_2) + \mathrm{Var}(Z_{\ell_1,\ell_2})\mathrm{Var}(Z_{\ell_1,i_2})\mathrm{Var}(Z_{\ell_2,i_2})I(i_1 = \ell_2, i_2 \neq \ell_1, j_2 = \ell_1, i_2 = j_1)$, $\mathbb{E}(\mathring{Z}_{i_1,\ell_1}\mathring{Z}_{i_2,\ell_2}\mathring{Z}_{\ell_2,j_2}\mathring{Z}_{i_2,j_2}) = 0$, $\mathbb{E}(\mathring{Z}_{i_2,\ell_2}\mathring{Z}_{i_1,\ell_1}\mathring{Z}_{\ell_1,j_1}\mathring{Z}_{i_1,j_1}) = 0$ and $\mathbb{E}(\mathring{Z}_{i_1,\ell_1}\mathring{Z}_{i_2,\ell_2}) = \mathrm{Var}(Z_{\ell_1,\ell_2})I(i_1 = \ell_2, i_2 = \ell_1)$ for any $(i_1, j_1, i_2, j_2, \ell_1, \ell_2)$ such that $i_1 \neq j_1$, $i_1, j_1 \neq \ell_1$, $i_2 \neq j_2$, $i_2, j_2 \neq \ell_2$,

$i_1 \neq i_2$ and $j_1 \neq j_2$, which implies that

$$
\begin{aligned}
&\frac{\nu_{\ell_1}^{1/2}\nu_{\ell_2}^{1/2}}{N} \sum_{\substack{i_1,j_1,i_2,j_2:\, i_1\neq j_1,\, i_2\neq j_2,\\ i_1\neq i_2,\, j_1\neq j_2,\, i_1,j_1\neq \ell_1,\, i_2,j_2\neq \ell_2}} \mathrm{Cov}\{Y_{(i_1,j_1),\ell_1}, Y_{(i_2,j_2),\ell_2}\} \\
&= \bigg(\frac{\mu_{\ell_1,1}+\mu_{\ell_1,2}}{2\mu_{\ell_1,1}\mu_{\ell_1,2}}\bigg)\bigg(\frac{\mu_{\ell_2,1}+\mu_{\ell_2,2}}{2\mu_{\ell_2,1}\mu_{\ell_2,2}}\bigg)\frac{1}{N}\sum_{i_1:\, i_1\neq \ell_1,\ell_2} \mathrm{Var}(Z_{i_1,\ell_1})\mathrm{Var}(Z_{\ell_1,\ell_2})\mathrm{Var}(Z_{i_1,\ell_2}) \\
&\quad + \bigg(\frac{\mu_{\ell_1,1}+\mu_{\ell_1,2}}{2\mu_{\ell_1,1}\mu_{\ell_1,2}}\bigg)\bigg(\frac{\mu_{\ell_2,1}+\mu_{\ell_2,2}}{2\mu_{\ell_2,1}\mu_{\ell_2,2}}\bigg)\frac{1}{N}\sum_{i_2:\, i_2\neq \ell_1,\ell_2} \mathrm{Var}(Z_{\ell_1,\ell_2})\mathrm{Var}(Z_{\ell_1,i_2})\mathrm{Var}(Z_{\ell_2,i_2}) \qquad \text{(S.1.7)} \\
&\quad + \frac{\lambda_{\ell_1,\ell_2}^2}{N}\sum_{j_1,j_2:\, j_1\neq j_2,\, j_1,j_2\neq \ell_1,\ell_2} \mathrm{Var}(Z_{\ell_1,\ell_2}) \asymp \gamma^{-6}p^{-1}+\gamma^{-2}
\end{aligned}
$$

for any $\ell_1 \neq \ell_2$. For $\ell_1 = \ell_2$, we write $\ell_1 = \ell_2 = \ell$. It holds that $\mathbb{E}(\mathring{Z}_{i_1,\ell}\mathring{Z}_{\ell,j_1}\mathring{Z}_{i_1,j_1}\mathring{Z}_{i_2,\ell}\mathring{Z}_{\ell,j_2}\mathring{Z}_{i_2,j_2}) = \mathrm{Var}(Z_{i_1,\ell})\mathrm{Var}(Z_{\ell,i_2})\mathrm{Var}(Z_{i_1,i_2})I(i_1 = j_2, i_2 = j_1)$, $\mathbb{E}(\mathring{Z}_{i_1,\ell}\mathring{Z}_{i_2,\ell}\mathring{Z}_{\ell,j_2}\mathring{Z}_{i_2,j_2}) = 0$, $\mathbb{E}(\mathring{Z}_{i_2,\ell}\mathring{Z}_{i_1,\ell}\mathring{Z}_{\ell,j_1}\mathring{Z}_{i_1,j_1}) = 0$ and $\mathbb{E}(\mathring{Z}_{i_1,\ell}\mathring{Z}_{i_2,\ell}) = 0$ for any (i_1,j_1,i_2,j_2,ℓ) such that $i_1 \neq j_1$, $i_2 \neq j_2$, $i_1 \neq i_2$, $j_1 \neq j_2$ and $i_1,i_2,j_1,j_2 \neq \ell$, which implies that for any $\ell \in [p]$,

$$
\begin{aligned}
&\frac{\nu_\ell}{N}\sum_{\substack{i_1,j_1,i_2,j_2:\, i_1\neq j_1,\, i_2\neq j_2,\\ i_1\neq i_2,\, j_1\neq j_2,\, i_1,j_1,i_2,j_2\neq \ell}} \mathrm{Cov}\{Y_{(i_1,j_1),\ell}, Y_{(i_2,j_2),\ell}\} \\
&= \bigg(\frac{\mu_{\ell,1}+\mu_{\ell,2}}{2\mu_{\ell,1}\mu_{\ell,2}}\bigg)^2\frac{1}{N}\sum_{i_1,i_2:\, i_1\neq i_2,\, i_1,i_2\neq \ell} \mathrm{Var}(Z_{i_1,\ell})\mathrm{Var}(Z_{\ell,i_2})\mathrm{Var}(Z_{i_1,i_2}) = \frac{\tilde{b}_\ell}{2}\,.
\end{aligned} \qquad \text{(S.1.8)}
$$

For any $\ell_1 \neq \ell_2$, notice that $\nu_\ell \asymp p\gamma^{-2} + \gamma^{-6}$ and

$$
\begin{aligned}
B_{\ell_1,\ell_2}^{(2)} &= \frac{1}{N}\sum_{\substack{i_1,j_1,i_2,j_2:\, i_1\neq j_1,\, i_2\neq j_2,\\ (i_1,j_1)\neq(i_2,j_2),\, i_1,j_1\neq \ell_1,\, i_2,j_2\neq \ell_2}} \mathrm{Cov}\{Y_{(i_1,j_1),\ell_1}, Y_{(i_2,j_2),\ell_2}\} \\
&= \frac{1}{N}\sum_{\substack{i,j_1,j_2:\, j_1\neq j_2,\, i\neq j_1,j_2,\\ i,j_1\neq \ell_1,\, i,j_2\neq \ell_2}} \mathrm{Cov}\{Y_{(i,j_1),\ell_1}, Y_{(i,j_2),\ell_2}\} + \frac{1}{N}\sum_{\substack{i_1,i_2,j:\, i_1\neq i_2,\, j\neq i_1,i_2,\\ i_1,j\neq \ell_1,\, i_2,j\neq \ell_2}} \mathrm{Cov}\{Y_{(i_1,j),\ell_1}, Y_{(i_2,j),\ell_2}\} \\
&\quad + \frac{1}{N}\sum_{\substack{i_1,j_1,i_2,j_2:\, i_1\neq j_1,\, i_2\neq j_2,\\ i_1\neq i_2,\, j_1\neq j_2,\, i_1,j_1\neq \ell_1,\, i_2,j_2\neq \ell_2}} \mathrm{Cov}\{Y_{(i_1,j_1),\ell_1}, Y_{(i_2,j_2),\ell_2}\}\,.
\end{aligned}
$$

It follows from (S.1.3), (S.1.5) and (S.1.7) that $\max_{1\leq \ell_1\neq \ell_2\leq p}|B_{\ell_1,\ell_2}^{(2)}| \lesssim p^{-1}$. Analogously, by (S.1.4), (S.1.6) and (S.1.8), it holds that $B_{\ell,\ell}^{(2)} = \nu_\ell^{-1}\{(p-3)b_\ell + \tilde{b}_\ell/2\}$ for any $\ell \in [p]$. Notice that $B_{\ell_1,\ell_2} = B_{\ell_1,\ell_2}^{(1)} + B_{\ell_1,\ell_2}^{(2)}$ for any $\ell_1,\ell_2 \in [p]$, together with (S.1.1) and (S.1.2), we have $\max_{1\leq \ell_1\neq \ell_2\leq p}|B_{\ell_1,\ell_2}| \lesssim p^{-1}$ and $B_{\ell,\ell} = \nu_\ell^{-1}\{(p-2)b_\ell + \tilde{b}_\ell\} = 1$ for any $\ell \in [p]$. We complete the proof of Lemma 4. □

S2 To prove $\mathrm{I}_1(i,j,\ell) = 0$ for any $i \neq j \neq \ell$

To simplify the notation, we write $\mathbf{c}(t)$, $\mathbf{c}^{-(i,j),\ell}(t)$, $c_k^{(i,j),\ell}(t)$, $c_{(i,j),\ell}(t)$ and $\dot{c}_{(i,j),\ell}(t)$ as $\mathbf{c}$, $\mathbf{c}^{-(i,j),\ell}$, $c_k^{(i,j),\ell}$, $c_{(i,j),\ell}$ and $\dot{c}_{(i,j),\ell}$, respectively. Notice that $q_\ell\{\mathbf{c}^{-(i,j),\ell}\}\dot{c}_{(i,j),\ell} = N^{-1/2}q_\ell\{\mathbf{c}^{-(i,j),\ell}\}[t^{-1/2}\{\sqrt{v}Y_{(i,j),\ell} + \sqrt{1-v}V_{(i,j),\ell}\} - (1-t)^{-1/2}W_{(i,j),\ell}]$. Since $Y_{(i,j),\ell}$ is independent of $\mathbf{c}^{-(i,j),\ell}$, then $\mathbb{E}[q_\ell\{\mathbf{c}^{-(i,j),\ell}\}Y_{(i,j),\ell}] = 0$. We notice that $V_{(i',j'),\ell'}$ with $i' \neq j' \neq \ell'$ included in $\mathbf{c}^{-(i,j),\ell}$ satisfies $|\{i',j',\ell'\} \cap \{i,j,\ell\}| \leq 1$. Since

$\mathrm{Cov}\{V_{(i',j'),\ell'},V_{(i,j),\ell}\} = \mathrm{Cov}\{Y_{(i',j'),\ell'},Y_{(i,j),\ell}\}$, the proof of Lemma 4 indicates that $\mathrm{Cov}\{V_{(i',j'),\ell'},V_{(i,j),\ell}\} = 0$ for any $i'\neq j'\neq \ell'$ such that $|\{i',j',\ell'\}\cap\{i,j,\ell\}|\leq 1$. Recall that $\{V_{(i,j),\ell}\}_{i,j,\ell:\,i\neq j\neq \ell}$ are normal random variables. Thus, $V_{(i,j),\ell}$ is independent of $\mathbf{c}^{-(i,j),\ell}$. Then $\mathbb{E}[q_\ell\{\mathbf{c}^{-(i,j),\ell}\}V_{(i,j),\ell}]=0$. Analogously, we know that $W_{(i,j),\ell}$ is also independent of $\mathbf{c}^{-(i,j),\ell}$, which implies $\mathbb{E}[q_\ell\{\mathbf{c}^{-(i,j),\ell}\}W_{(i,j),\ell}]=0$. Hence, $\mathrm{I}_1(i,j,\ell)\equiv 0$ for any $i\neq j\neq \ell$.

S3 To prove $\sum_{i,j:\,i\neq j}\sum_{\ell:\,\ell\neq i,j}\sum_{k=1}^p|\mathrm{I}_2(i,j,\ell,k)|\lesssim p^{-1/2}\phi^3\log^{7/2}p$

To simplify the notation, we write $\mathbf{c}(t)$, $\mathbf{c}^{-(i,j),\ell}(t)$, $c_k^{(i,j),\ell}(t)$, $c_{(i,j),\ell}(t)$ and $\dot c_{(i,j),\ell}(t)$ as $\mathbf{c}$, $\mathbf{c}^{-(i,j),\ell}$, $c_k^{(i,j),\ell}$, $c_{(i,j),\ell}$ and $\dot c_{(i,j),\ell}$, respectively. Notice that $c_k^{(i,j),\ell}=\sum_{i',j':\,i'\neq j'}c_{(i',j'),k}a_{(i',j'),k}^{(i,j,\ell)}$. Thus,

$$\mathrm{I}_2(i,j,\ell,k)=\sum_{i',j':\,i'\neq j'}\int_0^1\mathbb{E}\big[q_{\ell k}\{\mathbf{c}^{-(i,j),\ell}\}\dot c_{(i,j),\ell}c_{(i',j'),k}a_{(i',j'),k}^{(i,j,\ell)}\big]\,\mathrm{d}t\,. \tag{S.3.1}$$

We will consider $\mathrm{I}_2(i,j,\ell,k)$ in two cases: (i) $k\neq i,j,\ell$, and (ii) $k\in\{i,j,\ell\}$, in Sections S3.1 and S3.2, respectively. Then $\sum_{i,j:\,i\neq j}\sum_{\ell:\,\ell\neq i,j}\sum_{k=1}^p|\mathrm{I}_2(i,j,\ell,k)| = \sum_{i,j:\,i\neq j}\sum_{\ell:\,\ell\neq i,j}\sum_{k:\,k\neq i,j,\ell}|\mathrm{I}_2(i,j,\ell,k)| + \sum_{i,j:\,i\neq j}\sum_{\ell:\,\ell\neq i,j}|\mathrm{I}_2(i,j,\ell,i)| + \sum_{i,j:\,i\neq j}\sum_{\ell:\,\ell\neq i,j}|\mathrm{I}_2(i,j,\ell,j)| + \sum_{i,j:\,i\neq j}\sum_{\ell:\,\ell\neq i,j}|\mathrm{I}_2(i,j,\ell,\ell)|$. As we will show in Sections S3.1 and S3.2, it holds that $\sum_{i,j:\,i\neq j}\sum_{\ell:\,\ell\neq i,j}\sum_{k=1}^p|\mathrm{I}_2(i,j,\ell,k)|\lesssim p^{-1/2}\phi^3\log^{7/2}p$.

S3.1 Case 1: $k\neq i,j,\ell$

Recall $a_{(i',j'),k}^{(i,j,\ell)}=I\{(i',j',k)\in\mathcal{S}(i,j,\ell)\}$ and $k\neq i,j,\ell$. Then $a_{(i',j'),k}^{(i,j,\ell)}=1$ if and only if $\{i',j'\}\subset\{i,j,\ell\}$. It follows from (S.3.1) that

$$\begin{aligned}\mathrm{I}_2(i,j,\ell,k)=&\int_0^1\mathbb{E}\big[q_{\ell k}\{\mathbf{c}^{-(i,j),\ell}\}\dot c_{(i,j),\ell}c_{(i,\ell),k}\big]\,\mathrm{d}t+\int_0^1\mathbb{E}\big[q_{\ell k}\{\mathbf{c}^{-(i,j),\ell}\}\dot c_{(i,j),\ell}c_{(\ell,i),k}\big]\,\mathrm{d}t\\&+\int_0^1\mathbb{E}\big[q_{\ell k}\{\mathbf{c}^{-(i,j),\ell}\}\dot c_{(i,j),\ell}c_{(i,j),k}\big]\,\mathrm{d}t+\int_0^1\mathbb{E}\big[q_{\ell k}\{\mathbf{c}^{-(i,j),\ell}\}\dot c_{(i,j),\ell}c_{(j,i),k}\big]\,\mathrm{d}t\\&+\int_0^1\mathbb{E}\big[q_{\ell k}\{\mathbf{c}^{-(i,j),\ell}\}\dot c_{(i,j),\ell}c_{(j,\ell),k}\big]\,\mathrm{d}t+\int_0^1\mathbb{E}\big[q_{\ell k}\{\mathbf{c}^{-(i,j),\ell}\}\dot c_{(i,j),\ell}c_{(\ell,j),k}\big]\,\mathrm{d}t\,.\end{aligned} \tag{S.3.2}$$

Notice that $\dot c_{(i,j),\ell}c_{(i,\ell),k}=N^{-1}[t^{-1/2}\{\sqrt{v}Y_{(i,j),\ell}+\sqrt{1-v}V_{(i,j),\ell}\}-(1-t)^{-1/2}W_{(i,j),\ell}][\sqrt{t}\{\sqrt{v}Y_{(i,\ell),k}+\sqrt{1-v}V_{(i,\ell),k}\}+\sqrt{1-t}W_{(i,\ell),k}]$ and $\{Y_{(i,j),\ell},V_{(i,j),\ell},W_{(i,j),\ell}\}$ is independent of $\mathbf{c}^{-(i,j),\ell}$. It then holds that

$$\begin{aligned}&N\cdot\mathbb{E}[q_{\ell k}\{\mathbf{c}^{-(i,j),\ell}\}\dot c_{(i,j),\ell}c_{(i,\ell),k}]\\&=v\mathbb{E}\big[q_{\ell k}\{\mathbf{c}^{-(i,j),\ell}\}Y_{(i,j),\ell}Y_{(i,\ell),k}\big]+(1-v)\mathbb{E}\big[q_{\ell k}\{\mathbf{c}^{-(i,j),\ell}\}V_{(i,j),\ell}V_{(i,\ell),k}\big]\\&\quad-\mathbb{E}\big[q_{\ell k}\{\mathbf{c}^{-(i,j),\ell}\}W_{(i,j),\ell}W_{(i,\ell),k}\big]\,.\end{aligned} \tag{S.3.3}$$

Recall $Y_{(i,\ell),k}$ is a function of $\{\mathring{Z}_{i,k},\mathring{Z}_{k,\ell},\mathring{Z}_{i,\ell}\}$. We have shown that $\mathbf{c}^{-(i,j),\ell}$ is independent of $\mathring{Z}_{i,\ell}$. We will remove the components in $\mathbf{c}^{-(i,j),\ell}$ that depends on $\mathring{Z}_{i,k}$ and $\mathring{Z}_{k,\ell}$. As we have shown that $\mathring{Z}_{i,j}$ is independent of $\{Y_{(i',j'),\ell'}\}_{(i',j',\ell')\notin\mathcal{S}_*(i,j)}$ for any given i and j, we define $\mathbf{c}^{(i,j),\ell,(i,\ell),k}=\sum_{i',j':\,i'\neq j'}\mathbf{c}_{(i',j')}\circ\{\mathbf{a}_{(i',j')}^{(i,j,\ell)}+\mathbf{a}_{(i',j')}^{(i,k)}+\mathbf{a}_{(i',j')}^{(k,\ell)}-\mathbf{a}_{(i',j')}^{(i,j,\ell)}\circ\mathbf{a}_{(i',j')}^{(i,k)}-\mathbf{a}_{(i',j')}^{(i,j,\ell)}\circ\mathbf{a}_{(i',j')}^{(k,\ell)}-\mathbf{a}_{(i',j')}^{(i,k)}\circ\mathbf{a}_{(i',j')}^{(k,\ell)}+\mathbf{a}_{(i',j')}^{(i,j,\ell)}\circ\mathbf{a}_{(i',j')}^{(i,k)}\circ\mathbf{a}_{(i',j')}^{(k,\ell)}\}$, where $\mathbf{a}_{(i',j')}^{(i,j)}=\{a_{(i',j'),1}^{(i,j)},\ldots,a_{(i',j'),p}^{(i,j)}\}^{\mathrm{T}}$ with $a_{(i',j'),\ell'}^{(i,j)}=I\{(i',j',\ell')\in\mathcal{S}_*(i,j)\}$ for $\ell'\neq i',j'$ and $a_{(i',j'),\ell'}^{(i,j)}=0$ for $\ell'\in\{i',j'\}$. We know $Y_{(i,j),\ell}Y_{(i,\ell),k}$ is independent of $\mathbf{c}-\mathbf{c}^{(i,j),\ell,(i,\ell),k}$. Recall $\mathbf{c}^{(i,j),\ell}=\sum_{i',j':\,i'\neq j'}\mathbf{c}_{(i',j')}\circ\mathbf{a}_{(i',j')}^{(i,j,\ell)}$. Then $\mathbf{c}^{(i,j),\ell,(i,\ell),k}-\mathbf{c}^{(i,j),\ell}=\{\mathbf{c}_{(i,k)}+\mathbf{c}_{(k,i)}\}\circ(\mathbf{1}-\mathbf{e}_\ell-\mathbf{e}_j)+\{\mathbf{c}_{(\ell,k)}+\mathbf{c}_{(k,\ell)}\}\circ(\mathbf{1}-\mathbf{e}_i-\mathbf{e}_j)+\sum_{m\neq i,j,\ell,k}\{c_{(k,m),i}+$

$c_{(m,k),i}\}\mathbf{e}_i + \sum_{m\neq i,j,\ell,k}\{c_{(k,m),\ell} + c_{(m,k),\ell}\}\mathbf{e}_\ell + \sum_{m\neq i,j,\ell,k}\{c_{(i,m),k} + c_{(m,i),k} + c_{(\ell,m),k} + c_{(m,\ell),k}\}\mathbf{e}_k$, where $\mathbf{1}$ is a p-dimensional vector with all components being 1. Let $\mathbf{c}^{-(i,j),\ell,(i,\ell),k} = \mathbf{c} - \mathbf{c}^{(i,j),\ell,(i,\ell),k}$. Recall $\mathbf{c}^{-(i,j),\ell} = \mathbf{c} - \mathbf{c}^{(i,j),\ell}$. Then $\mathbf{c}^{(i,j),\ell,(i,\ell),k-(i,j),\ell} := \mathbf{c}^{-(i,j),\ell} - \mathbf{c}^{-(i,j),\ell,(i,\ell),k} = \mathbf{c}^{(i,j),\ell,(i,\ell),k} - \mathbf{c}^{(i,j),\ell}$. Write $\mathbf{c}^{(i,j),\ell,(i,\ell),k-(i,j),\ell} = \{c_1^{(i,j),\ell,(i,\ell),k-(i,j),\ell}, \ldots, c_p^{(i,j),\ell,(i,\ell),k-(i,j),\ell}\}^{\mathrm{T}}$. We know that $Y_{(i,j),\ell}$ and $Y_{(i,\ell),k}$ are independent of $\mathbf{c}^{-(i,j),\ell,(i,\ell),k}$. Based on the proof of Lemma 4, we also have $V_{(i,j),\ell}, V_{(i,\ell),k}, W_{(i,j),\ell}, W_{(i,\ell),k}$ are independent of $\mathbf{c}^{-(i,j),\ell,(i,\ell),k}$. By Taylor expansion, it holds that $\mathbb{E}[q_{\ell k}\{\mathbf{c}^{-(i,j),\ell}\}Y_{(i,j),\ell}Y_{(i,\ell),k}] = \mathbb{E}[q_{\ell k}\{\mathbf{c}^{-(i,j),\ell,(i,\ell),k}\}]\mathbb{E}\{Y_{(i,j),\ell}Y_{(i,\ell),k}\} + \sum_{m=1}^p \int_0^1 \mathbb{E}[q_{\ell k m}\{\mathbf{c}^{-(i,j),\ell,(i,\ell),k} + \tau\mathbf{c}^{(i,j),\ell,(i,\ell),k-(i,j),\ell}\}Y_{(i,j),\ell}Y_{(i,\ell),k} c_m^{(i,j),\ell,(i,\ell),k-(i,j),\ell}]\,\mathrm{d}\tau$. Notice that $\mathbb{E}\{Y_{(i,j),\ell}Y_{(i,\ell),k}\} = \mathbb{E}\{V_{(i,j),\ell}V_{(i,\ell),k}\} = \mathbb{E}\{W_{(i,j),\ell}W_{(i,\ell),k}\}$. Then (S.3.3) implies that

$$
\begin{aligned}
&N\cdot\mathbb{E}[q_{\ell k}\{\mathbf{c}^{-(i,j),\ell}\}\dot{c}_{(i,j),\ell}c_{(i,\ell),k}] \\
&= \sum_{m=1}^p \int_0^1 \mathbb{E}\big[q_{\ell k m}\{\mathbf{c}^{-(i,j),\ell,(i,\ell),k} + \tau\mathbf{c}^{(i,j),\ell,(i,\ell),k-(i,j),\ell}\}c_m^{(i,j),\ell,(i,\ell),k-(i,j),\ell} \qquad \text{(S.3.4)}\\
&\qquad\qquad \times\{vY_{(i,j),\ell}Y_{(i,\ell),k} + (1-v)V_{(i,j),\ell}V_{(i,\ell),k} - W_{(i,j),\ell}W_{(i,\ell),k}\}\big]\,\mathrm{d}\tau\,.
\end{aligned}
$$

Define $\mathcal{E}_1 = \{|Y_{(i,j),\ell}| \vee |V_{(i,j),\ell}| \vee |W_{(i,j),\ell}| \le B \text{ for any } i\neq j\neq \ell\}$ for some $B>0$. Write $\nu = p\gamma^{-2} + \gamma^{-6}$. Then $\nu \asymp \max_{s\in[p]}\nu_s \asymp \min_{s\in[p]}\nu_s$. As shown in Case 1 in the proof of Lemma 4, $\mathrm{Var}\{V_{(i,j),\ell}\} \asymp \nu^{-1}\gamma^{-6} \asymp \mathrm{Var}\{W_{(i,j),\ell}\}$. Recall $V_{(i,j),\ell}$ and $W_{(i,j),\ell}$ are normal random variables with mean zero. Thus, $\max_{i,j,\ell:\, i\neq j\neq \ell}|V_{(i,j),\ell}| = \nu^{-1/2}\gamma^{-3}\cdot O_{\mathrm{p}}(\log^{1/2}p) = \max_{i,j,\ell:\, i\neq j\neq k}|W_{(i,j),\ell}|$. Notice that $\max_{i,j,\ell:\, i\neq j\neq \ell}|Y_{(i,j),\ell}| \lesssim \nu^{-1/2}\gamma^{-3}$. Recall $\nu = p\gamma^{-2}+\gamma^{-6}$. Then $\nu^{-1/2}\gamma^{-3}\lesssim 1$. If we select $B = C_*\log^{1/2}p$ for sufficiently large C_*, it holds that $\mathbb{P}(\mathcal{E}_1^c)\lesssim p^{-C}$. Here C can be sufficiently large if we select sufficiently large C_*. Notice that $\mathbf{c}^{-(i,j),\ell,(i,\ell),k} + \tau\mathbf{c}^{(i,j),\ell,(i,\ell),k-(i,j),\ell} = \mathbf{c} - \mathbf{c}^{(i,j),\ell} + (\tau-1)\mathbf{c}^{(i,j),\ell,(i,\ell),k-(i,j),\ell}$. Restricted on $\mathcal{E}_1$, it holds that

$$
\begin{aligned}
|\mathbf{c}^{(i,j),\ell}|_\infty \le \frac{54B}{\sqrt{N}} &+ \sum_{\substack{s_1,s_2:\, s_1\neq s_2\\ s_1,s_2\in\{i,j,\ell\}}}\bigg|\frac{1}{\sqrt{N}}\sum_{m:\, m\neq i,j,\ell,k} Y_{(s_1,m),s_2}\bigg| + \sum_{\substack{s_1,s_2:\, s_1\neq s_2\\ s_1,s_2\in\{i,j,\ell\}}}\bigg|\frac{1}{\sqrt{N}}\sum_{m:\, m\neq i,j,\ell,k} Y_{(m,s_1),s_2}\bigg| \\
&+ \sum_{\substack{s_1,s_2:\, s_1\neq s_2\\ s_1,s_2\in\{i,j,\ell\}}}\bigg|\frac{1}{\sqrt{N}}\sum_{m:\, m\neq i,j,\ell,k} V_{(s_1,m),s_2}\bigg| + \sum_{\substack{s_1,s_2:\, s_1\neq s_2\\ s_1,s_2\in\{i,j,\ell\}}}\bigg|\frac{1}{\sqrt{N}}\sum_{m:\, m\neq i,j,\ell,k} V_{(m,s_1),s_2}\bigg| \qquad \text{(S.3.5)}\\
&+ \sum_{\substack{s_1,s_2:\, s_1\neq s_2\\ s_1,s_2\in\{i,j,\ell\}}}\bigg|\frac{1}{\sqrt{N}}\sum_{m:\, m\neq i,j,\ell,k} W_{(s_1,m),s_2}\bigg| + \sum_{\substack{s_1,s_2:\, s_1\neq s_2\\ s_1,s_2\in\{i,j,\ell\}}}\bigg|\frac{1}{\sqrt{N}}\sum_{m:\, m\neq i,j,\ell,k} W_{(m,s_1),s_2}\bigg|
\end{aligned}
$$

and

$$
\begin{aligned}
&|\mathbf{c}^{(i,j),\ell,(i,\ell),k-(i,j),\ell}|_\infty \\
&\le \frac{24B}{\sqrt{N}} + \sum_{\substack{s_1,s_2:\, s_1\neq s_2\\ s_1,s_2\in\{i,\ell,k\}}}\bigg|\frac{1}{\sqrt{N}}\sum_{m:\, m\neq i,j,\ell,k} Y_{(s_1,m),s_2}\bigg| + \sum_{\substack{s_1,s_2:\, s_1\neq s_2\\ s_1,s_2\in\{i,\ell,k\}}}\bigg|\frac{1}{\sqrt{N}}\sum_{m:\, m\neq i,j,\ell,k} Y_{(m,s_1),s_2}\bigg| \\
&\quad + \sum_{\substack{s_1,s_2:\, s_1\neq s_2\\ s_1,s_2\in\{i,\ell,k\}}}\bigg|\frac{1}{\sqrt{N}}\sum_{m:\, m\neq i,j,\ell,k} V_{(s_1,m),s_2}\bigg| + \sum_{\substack{s_1,s_2:\, s_1\neq s_2\\ s_1,s_2\in\{i,\ell,k\}}}\bigg|\frac{1}{\sqrt{N}}\sum_{m:\, m\neq i,j,\ell,k} V_{(m,s_1),s_2}\bigg| \qquad \text{(S.3.6)}
\end{aligned}
$$

$$+\sum_{\substack{s_1,s_2:\, s_1\neq s_2\\ s_1,s_2\in\{i,\ell,k\}}}\bigg|\frac{1}{\sqrt{N}}\sum_{m:\, m\neq i,j,\ell,k}W_{(s_1,m),s_2}\bigg|+\sum_{\substack{s_1,s_2:\, s_1\neq s_2\\ s_1,s_2\in\{i,\ell,k\}}}\bigg|\frac{1}{\sqrt{N}}\sum_{m:\, m\neq i,j,\ell,k}W_{(m,s_1),s_2}\bigg|,$$

which implies that

$$\begin{aligned}
&|\mathbf{c}^{(i,j),\ell}-(\tau-1)\mathbf{c}^{(i,j),\ell,(i,\ell),k-(i,j),\ell}|_\infty\\
&\le\frac{78B}{\sqrt{N}}+24\max_{\substack{s_1,s_2:\, s_1\neq s_2\\ s_1,s_2\in\{i,j,\ell,k\}}}\bigg\{\bigg|\frac{1}{\sqrt{N}}\sum_{m:\, m\neq i,j,\ell,k}Y_{(s_1,m),s_2}\bigg|+\bigg|\frac{1}{\sqrt{N}}\sum_{m:\, m\neq i,j,\ell,k}Y_{(m,s_1),s_2}\bigg|\bigg\}\\
&\quad+24\max_{\substack{s_1,s_2:\, s_1\neq s_2\\ s_1,s_2\in\{i,j,\ell,k\}}}\bigg\{\bigg|\frac{1}{\sqrt{N}}\sum_{m:\, m\neq i,j,\ell,k}V_{(s_1,m),s_2}\bigg|+\bigg|\frac{1}{\sqrt{N}}\sum_{m:\, m\neq i,j,\ell,k}V_{(m,s_1),s_2}\bigg|\bigg\}\qquad\text{(S.3.7)}\\
&\quad+24\max_{\substack{s_1,s_2:\, s_1\neq s_2\\ s_1,s_2\in\{i,j,\ell,k\}}}\bigg\{\bigg|\frac{1}{\sqrt{N}}\sum_{m:\, m\neq i,j,\ell,k}W_{(s_1,m),s_2}\bigg|+\bigg|\frac{1}{\sqrt{N}}\sum_{m:\, m\neq i,j,\ell,k}W_{(m,s_1),s_2}\bigg|\bigg\}
\end{aligned}$$

under $\mathcal{E}_1$. Recall $\nu=p\gamma^{-2}+\gamma^{-6}\asymp\max_{s\in[p]}\nu_s\asymp\min_{s\in[p]}\nu_s$. Given $s_1,s_2\in\{i,j,\ell,k\}$ such that $s_1\neq s_2$, we have $N^{-1/2}\sum_{m:\, m\neq i,j,\ell,k}Y_{(s_1,m),s_2}=-\nu_{s_2}^{-1/2}\{(2\mu_{s_2,1})^{-1}+(2\mu_{s_2,2})^{-1}\}N^{-1/2}\sum_{m:\, m\neq i,j,\ell,k}\mathring{Z}_{s_1,s_2}\mathring{Z}_{s_2,m}\mathring{Z}_{s_1,m}+N^{-1/2}(p-4)\nu_{s_2}^{-1/2}\lambda_{s_1,s_2}\mathring{Z}_{s_1,s_2}$. Then $|N^{-1/2}\sum_{m:\, m\neq i,j,\ell,k}Y_{(s_1,m),s_2}|\lesssim\nu^{-1/2}\gamma^{-3}|N^{-1/2}\sum_{m:\, m\neq i,j,\ell,k}\mathring{Z}_{s_2,m}\mathring{Z}_{s_1,m}|$ $+\,\nu^{-1/2}\gamma^{-1}$. Notice that $\{\mathring{Z}_{s_1,m}\mathring{Z}_{s_2,m}\}_{m:\, m\neq i,j,k,\ell}$ is an independent sequence. By Bernstein inequality, we have $\mathbb{P}(|N^{-1/2}\sum_{m:\, m\neq i,j,\ell,k}\mathring{Z}_{s_2,m}\mathring{Z}_{s_1,m}|>u)\lesssim\exp(-Cpu^2)$ for any $u=o(1)$, which implies that

$$\begin{aligned}
&\max_{i,j,\ell,k:\, i\neq j\neq\ell\neq k}\ \max_{\substack{s_1,s_2:\, s_1\neq s_2\\ s_1,s_2\in\{i,j,\ell,k\}}}\bigg|\frac{1}{\sqrt{N}}\sum_{m:\, m\neq i,j,\ell,k}Y_{(s_1,m),s_2}\bigg|\\
&\qquad\le C\nu^{-1/2}\gamma^{-1}+\nu^{-1/2}\gamma^{-3}\cdot O_{\rm p}\bigg(\frac{\log^{1/2}p}{p^{1/2}}\bigg).\qquad\text{(S.3.8)}
\end{aligned}$$

Analogously, we also have $\max_{i,j,\ell,k:\, i\neq j\neq\ell\neq k}\max_{s_1,s_2:\, s_1\neq s_2,\, s_1,s_2\in\{i,j,\ell,k\}}|N^{-1/2}\sum_{m:\, m\neq i,j,\ell,k}Y_{(m,s_1),s_2}|\le C\nu^{-1/2}\gamma^{-1}+\nu^{-1/2}\gamma^{-3}\cdot O_{\rm p}(p^{-1/2}\log^{1/2}p)$. As shown in Cases 1 and 2 in the proof of Lemma 4, we can obtain that $\mathrm{Cov}\{V_{(s_1,m),s_2},V_{(s_1,m),s_2}\}=\nu_{s_2}^{-1}\{(\mu_{s_2,1}+\mu_{s_2,2})/(2\mu_{s_2,1}\mu_{s_2,2})\}^2\mathrm{Var}(Z_{s_1,s_2})\mathrm{Var}(Z_{s_2,m})\mathrm{Var}(Z_{s_1,m})+\nu_{s_2}^{-1}\lambda_{s_1,s_2}^2\mathrm{Var}(Z_{s_1,s_2})$ and $\mathrm{Cov}\{V_{(s_1,m),s_2},V_{(s_1,m'),s_2}\}=\nu_{s_2}^{-1}\lambda_{s_1,s_2}^2\mathrm{Var}(Z_{s_1,s_2})$ for any $m\neq m'$. Then we have $\mathrm{Var}\{\sum_{m:\, m\neq i,j,\ell,k}V_{(s_1,m),s_2}\}\asymp p\nu^{-1}\gamma^{-2}(\gamma^{-4}+p)\asymp p$. Since $\{V_{(s_1,m),s_2}\}_{m:\, m\neq i,j,\ell,k}$ are normal random variables with mean zero, then it holds that $N^{-1/2}\sum_{m:\, m\neq i,j,\ell,k}V_{(s_1,m),s_2}$ is also a normal random variable with mean zero and variance $N^{-1}\mathrm{Var}\{\sum_{m:\, m\neq i,j,\ell,k}V_{(s_1,m),s_2}\}\asymp p^{-1}$. Therefore,

$$\max_{i,j,\ell,k:\, i\neq j\neq\ell\neq k}\ \max_{\substack{s_1,s_2:\, s_1\neq s_2\\ s_1,s_2\in\{i,j,\ell,k\}}}\bigg|\frac{1}{\sqrt{N}}\sum_{m:\, m\neq i,j,\ell,k}V_{(s_1,m),s_2}\bigg|=O_{\rm p}\bigg(\frac{\log^{1/2}p}{p^{1/2}}\bigg).\qquad\text{(S.3.9)}$$

Also, as shown in Case 3 in the proof of Lemma 4, $\mathrm{Cov}\{V_{(m,s_1),s_2},V_{(m',s_1),s_2}\}=0$ for any $m\neq m'$. Then we have $N^{-1/2}\sum_{m:\, m\neq i,j,\ell,k}V_{(m,s_1),s_2}$ is also a normal random variable with mean zero and variance $N^{-1}\sum_{m:\, m\neq i,j,\ell,k}\mathrm{Var}\{V_{(m,s_1),s_2}\}\asymp p^{-1}\nu^{-1}\gamma^{-6}$. Notice that $\nu\asymp p\gamma^{-2}+\gamma^{-6}$. Then it holds that

$N^{-1}\sum_{m:\, m\neq i,j,\ell,k}\mathrm{Var}\{V_{(m,s_1),s_2}\}\lesssim p^{-1}$, which implies that

$$\max_{i,j,\ell,k:\, i\neq j\neq \ell\neq k}\ \max_{\substack{s_1,s_2:\, s_1\neq s_2\\ s_1,s_2\in\{i,j,\ell,k\}}}\bigg|\frac{1}{\sqrt{N}}\sum_{m:\, m\neq i,j,\ell,k}V_{(m,s_1),s_2}\bigg| = O_{\rm p}\bigg(\frac{\log^{1/2}p}{p^{1/2}}\bigg)\,. \tag{S.3.10}$$

Identically, $\max_{i,j,\ell,k:\, i\neq j\neq\ell\neq k}\max_{s_1,s_2:\, s_1\neq s_2\,, s_1,s_2\in\{i,j,\ell,k\}}|N^{-1/2}\sum_{m:\, m\neq i,j,\ell,k}W_{(s_1,m),s_2}| = O_{\rm p}(p^{-1/2}\log^{1/2}p)$. and $\max_{i,j,\ell,k:\, i\neq j\neq\ell\neq k}\max_{s_1,s_2:\, s_1\neq s_2\,, s_1,s_2\in\{i,j,\ell,k\}}|N^{-1/2}\sum_{m:\, m\neq i,j,\ell,k}W_{(m,s_1),s_2}| = O_{\rm p}(p^{-1/2}\log^{1/2}p)$. We define

$$\begin{aligned}\mathcal{E}_2(Y) = \Bigg\{&\max_{\substack{s_1,s_2:\, s_1\neq s_2\\ s_1,s_2\in\{i,j,\ell,k\}}}\bigg|\frac{1}{\sqrt{N}}\sum_{m:\, m\neq i,j,\ell,k}Y_{(s_1,m),s_2}\bigg|\leq\frac{C_{**}\log^{1/2}p}{p^{1/2}}\ \text{ and}\\ &\max_{\substack{s_1,s_2:\, s_1\neq s_2\\ s_1,s_2\in\{i,j,\ell,k\}}}\bigg|\frac{1}{\sqrt{N}}\sum_{m:\, m\neq i,j,\ell,k}Y_{(m,s_1),s_2}\bigg|\leq\frac{C_{**}\log^{1/2}p}{p^{1/2}}\ \text{ for any } i\neq j\neq\ell\neq k\Bigg\}\end{aligned}$$

for sufficiently large $C_{**}>0$. We can also define $\mathcal{E}_2(V)$ and $\mathcal{E}_2(W)$ in the same manner. Let

$$\mathcal{E}_2=\mathcal{E}_2(Y)\cap\mathcal{E}_2(V)\cap\mathcal{E}_2(W)\,. \tag{S.3.11}$$

Notice that $\nu^{-1/2}\gamma^{-1}\lesssim p^{-1/2}$ and $\nu^{-1/2}\gamma^{-3}\lesssim 1$. It holds that $\mathbb{P}(\mathcal{E}_2^c)\lesssim p^{-C}$. Here C can be sufficiently large if we select sufficiently large C_{**}. Recall that $B=C_*\log^{1/2}p$. Restricted on $\mathcal{E}_1\cap\mathcal{E}_2$, by (S.3.7), we have $\max_{i,j,\ell,k:\, i\neq j\neq\ell\neq k}|\mathbf{c}^{(i,j),\ell}-(\tau-1)\mathbf{c}^{(i,j),\ell,(i,\ell),k-(i,j),\ell}|_\infty\leq 78C_*N^{-1/2}\log^{1/2}p+144C_{**}p^{-1/2}\log^{1/2}p$. As $p\to\infty$, it holds that $78C_*N^{-1/2}\log^{1/2}p+144C_{**}p^{-1/2}\log^{1/2}p<3/(4\beta)$ with $\beta=\phi\log p$, which implies that

$$\max_{i,j,\ell,k:\, i\neq j\neq\ell\neq k}|\mathbf{c}^{(i,j),\ell}-(\tau-1)\mathbf{c}^{(i,j),\ell,(i,\ell),k-(i,j),\ell}|_\infty<\frac{3}{4\beta} \tag{S.3.12}$$

under $\mathcal{E}_1\cap\mathcal{E}_2$.

As shown in Lemma A.5 of Chernozhukov, Chetverikov and Kato (2013), there exists $U_{\ell km}(\mathbf{v})$ such that $|q_{\ell km}(\mathbf{v})|\leq U_{\ell km}(\mathbf{v})$ for any $\mathbf{v}\in\mathbb{R}^p$, where $\sum_{\ell,k,m=1}^p U_{\ell km}(\mathbf{v})\lesssim\phi\beta^2$ for any $\mathbf{v}\in\mathbb{R}^p$. Thus, (S.3.4) leads to $\sum_{\ell:\,\ell\neq i,j}\sum_{k:\, k\neq i,j,\ell}|N\cdot\mathbb{E}[q_{\ell k}\{\mathbf{c}^{-(i,j),\ell}\}\dot{c}_{(i,j),\ell}c_{(i,\ell),k}]|\lesssim T_{i,j,1}+T_{i,j,2}$ with $T_{i,j,1}=\log p\cdot\sum_{\ell:\,\ell\neq i,j}\sum_{k:\, k\neq i,j,\ell}\sum_{m=1}^p\int_0^1\mathbb{E}[I(\mathcal{E}_1\cap\mathcal{E}_2)\cdot U_{\ell km}\{\mathbf{c}^{-(i,j),\ell,(i,\ell),k}+\tau\mathbf{c}^{(i,j),\ell,(i,\ell),k-(i,j),\ell}\}|c_m^{(i,j),\ell,(i,\ell),k-(i,j),\ell}|]\,\mathrm{d}\tau$ and $T_{i,j,2}=\sum_{\ell:\,\ell\neq i,j}\sum_{k:\, k\neq i,j,\ell}\sum_{m=1}^p\int_0^1\mathbb{E}[I(\mathcal{E}_1^c\cup\mathcal{E}_2^c)\cdot U_{\ell km}\{\mathbf{c}^{-(i,j),\ell,(i,\ell),k}+\tau\mathbf{c}^{(i,j),\ell,(i,\ell),k-(i,j),\ell}\}\{|Y_{(i,j),\ell}||Y_{(i,\ell),k}|+|V_{(i,j),\ell}||V_{(i,\ell),k}|+|W_{(i,j),\ell}||W_{(i,\ell),k}|\}|c_m^{(i,j),\ell,(i,\ell),k-(i,j),\ell}|]\,\mathrm{d}\tau$. Together with (S.3.12), Lemma A.6 of Chernozhukov, Chetverikov and Kato (2013) implies that, restricted on $\mathcal{E}_1\cap\mathcal{E}_2$, $U_{\ell km}(\mathbf{c})\lesssim U_{\ell km}\{\mathbf{c}^{-(i,j),\ell,(i,\ell),k}+\tau\mathbf{c}^{(i,j),\ell,(i,\ell),k-(i,j),\ell}\}\lesssim U_{\ell km}(\mathbf{c})$ for any $t\in[0,1]$. Thus, $T_{i,j,1}\lesssim\phi\beta^2\log p\,\mathbb{E}[I(\mathcal{E}_1\cap\mathcal{E}_2)\max_{k,\ell:\, k\neq\ell}|\mathbf{c}^{(i,j),\ell,(i,\ell),k-(i,j),\ell}|_\infty]$. Restricted on $\mathcal{E}_1\cap\mathcal{E}_2$, it follows from (S.3.6) that $\max_{k,\ell:\, k\neq\ell}|\mathbf{c}^{(i,j),\ell,(i,\ell),k-(i,j),\ell}|_\infty\leq 24p^{-1/2}C_*\log^{1/2}p+6p^{-1/2}C_{**}\log^{1/2}p$, which implies

$$T_{i,j,1}\lesssim\frac{\phi\beta^2\log^{3/2}p}{p^{1/2}}\,. \tag{S.3.13}$$

For $T_{i,j,2}$, since $U_{\ell km}\{\mathbf{c}^{-(i,j),\ell,(i,\ell),k}(t)+\tau\mathbf{c}^{(i,j),\ell,(i,\ell),k-(i,j),\ell}(t)\}\lesssim\phi\beta^2$, by Cauchy-Schwarz inequality, it holds that

$$T_{i,j,2}\lesssim\phi\beta^2\sum_{\ell:\,\ell\neq i,j}\sum_{k:\, k\neq i,j,\ell}\sum_{m=1}^p\mathbb{E}\big[I(\mathcal{E}_1^c\cup\mathcal{E}_2^c)|c_m^{(i,j),\ell,(i,\ell),k-(i,j),\ell}|$$

$$\times \{|Y_{(i,j),\ell}||Y_{(i,\ell),k}| + |V_{(i,j),\ell}||V_{(i,\ell),k}| + |W_{(i,j),\ell}||W_{(i,\ell),k}|\}\big]$$
$$\lesssim p^3\phi\beta^2\mathbb{P}^{1/2}(\mathcal{E}_1^c \cup \mathcal{E}_2^c)$$
$$\times \max_{\ell,k:\,\ell\neq k} \mathbb{E}^{1/2}\big[|\mathbf{c}^{(i,j),\ell,(i,\ell),k-(i,j),\ell}|_\infty^2\{|Y_{(i,j),\ell}|^2|Y_{(i,\ell),k}|^2 + |V_{(i,j),\ell}|^2|V_{(i,\ell),k}|^2$$
$$+ |W_{(i,j),\ell}|^2|W_{(i,\ell),k}|^2\}\big]$$
$$\lesssim p^3\phi\beta^2\mathbb{P}^{1/2}(\mathcal{E}_1^c \cup \mathcal{E}_2^c) \max_{\ell,k:\,\ell\neq k} \mathbb{E}^{1/4}\big\{|\mathbf{c}^{(i,j),\ell,(i,\ell),k-(i,j),\ell}|_\infty^4\big\}\mathbb{E}^{1/8}\{|Y_{(i,j),\ell}|^8\}\mathbb{E}^{1/8}\{|Y_{(i,\ell),k}|^8\}$$
$$+ p^3\phi\beta^2\mathbb{P}^{1/2}(\mathcal{E}_1^c \cup \mathcal{E}_2^c) \max_{\ell,k:\,\ell\neq k} \mathbb{E}^{1/4}\big\{|\mathbf{c}^{(i,j),\ell,(i,\ell),k-(i,j),\ell}|_\infty^4\big\}\mathbb{E}^{1/8}\{|V_{(i,j),\ell}|^8\}\mathbb{E}^{1/8}\{|V_{(i,\ell),k}|^8\}\,,$$

where the last step is based on the fact $\{V_{(i,j),\ell}, V_{(i,\ell),k}\}$ and $\{W_{(i,j),\ell}, W_{(i,\ell),k}\}$ are identically distributed. Notice that $\max_{i,j,\ell:\,i\neq j\neq \ell}|Y_{(i,j),\ell}| \lesssim \nu^{-1/2}\gamma^{-3} \lesssim 1$ and $V_{(i,j),\ell}$ is a normal distributed random variable with $\mathrm{Var}\{V_{(i,j),\ell}\} \asymp \nu^{-1}\gamma^{-6} \lesssim 1$. Thus $T_{i,j,2} \lesssim p^3\phi\beta^2\mathbb{P}^{1/2}(\mathcal{E}_1^c \cup \mathcal{E}_2^c)\max_{\ell,k:\,\ell\neq k}\mathbb{E}^{1/4}\{|\mathbf{c}^{(i,j),\ell,(i,\ell),k-(i,j),\ell}|_\infty^4\}$. Following the same arguments for (S.3.8), (S.3.9) and (S.3.10), (S.3.6) leads to $\max_{\ell,k:\,\ell\neq k}\mathbb{E}^{1/4}\{|\mathbf{c}^{(i,j),\ell,(i,\ell),k-(i,j),\ell}|_\infty^4\} \lesssim 1$, which implies that $T_{i,j,2} \lesssim p^3\phi\beta^2\mathbb{P}^{1/2}(\mathcal{E}_1^c \cup \mathcal{E}_2^c)$. Recall that $\mathbb{P}(\mathcal{E}_1^c \cup \mathcal{E}_2^c) \lesssim p^{-C}$ and C can be sufficiently large if we select sufficiently large C_* and C_{**} in the definition of $\mathcal{E}_1$ and $\mathcal{E}_2$. Hence, with suitable selection of (C_*, C_{**}), we have $T_{i,j,2} \lesssim p^{-1/2}\phi\beta^2\log^{3/2}p$. Together with (S.3.13), it holds that $\sum_{\ell:\,\ell\neq i,j}\sum_{k:\,k\neq i,j,\ell}|\int_0^1 \mathbb{E}[q_{\ell k}\{\mathbf{c}^{-(i,j),\ell}\}\dot{c}_{(i,j),\ell}c_{(i,\ell),k}]\,\mathrm{d}t| \lesssim p^{-5/2}\phi\beta^2\log^{3/2}p$. Following the same arguments, we can bound the other terms in (S.3.2). Recall $\beta = \phi\log p$. Hence, $\sum_{i,j:\,i\neq j}\sum_{\ell:\,\ell\neq i,j}\sum_{k:\,k\neq i,j,\ell}|\mathrm{I}_2(i,j,\ell,k)| \lesssim p^{-1/2}\phi\beta^2\log^{3/2}p = p^{-1/2}\phi^3\log^{7/2}p$.

S3.2 Case 2: $k = i, j, \ell$

We first consider the case with $k = i$. Notice that $c_{(i',j'),i} = 0$ if $i' = i$ or $j' = i$. By (S.3.1), it holds that

$$\mathrm{I}_2(i,j,\ell,i) = \int_0^1 \mathbb{E}\big[q_{\ell i}\{\mathbf{c}^{-(i,j),\ell}\}\dot{c}_{(i,j),\ell}c_{(\ell,j),i}\big]\,\mathrm{d}t + \int_0^1 \mathbb{E}\big[q_{\ell i}\{\mathbf{c}^{-(i,j),\ell}\}\dot{c}_{(i,j),\ell}c_{(j,\ell),i}\big]\,\mathrm{d}t$$
$$+ \sum_{m:\,m\neq i,j,\ell}\int_0^1 \mathbb{E}\big[q_{\ell i}\{\mathbf{c}^{-(i,j),\ell}\}\dot{c}_{(i,j),\ell}c_{(m,j),i}\big]\,\mathrm{d}t + \sum_{m:\,m\neq i,j,\ell}\int_0^1 \mathbb{E}\big[q_{\ell i}\{\mathbf{c}^{-(i,j),\ell}\}\dot{c}_{(i,j),\ell}c_{(m,\ell),i}\big]\,\mathrm{d}t$$
$$+ \sum_{m:\,m\neq i,j,\ell}\int_0^1 \mathbb{E}\big[q_{\ell i}\{\mathbf{c}^{-(i,j),\ell}\}\dot{c}_{(i,j),\ell}c_{(j,m),i}\big]\,\mathrm{d}t + \sum_{m:\,m\neq i,j,\ell}\int_0^1 \mathbb{E}\big[q_{\ell i}\{\mathbf{c}^{-(i,j),\ell}\}\dot{c}_{(i,j),\ell}c_{(\ell,m),i}\big]\,\mathrm{d}t\,.$$

Note that $\dot{c}_{(i,j),\ell} = N^{-1/2}[t^{-1/2}\{\sqrt{v}Y_{(i,j),\ell}+\sqrt{1-v}V_{(i,j),\ell}\}-(1-t)^{-1/2}W_{(i,j),\ell}]$, $c_{(j,\ell),i} = N^{-1/2}[\sqrt{t}\{\sqrt{v}Y_{(j,\ell),i} + \sqrt{1-v}V_{(j,\ell),i}\} + \sqrt{1-t}W_{(j,\ell),i}]$. Recall that we have shown $\{Y_{(i,j),\ell}, V_{(i,j),\ell}, W_{(i,j),\ell}\}$ is independent of $\mathbf{c}^{-(i,j),\ell}$. Following the same arguments in Section S2 to show $V_{(i,j),\ell}$ is independent of $\mathbf{c}^{-(i,j),\ell}$, we also have $\{V_{(j,\ell),i}, W_{(j,\ell),i}\}$ is independent of $\mathbf{c}^{-(i,j),\ell}$. Notice that $Y_{(j,\ell),i}$ is a function of $(\mathring{Z}_{j,i}, \mathring{Z}_{i,\ell}, \mathring{Z}_{j,\ell})$. Thus $Y_{(j,\ell),i}$ is also independent of $\mathbf{c}^{-(i,j),\ell}$. Since $\{Y_{(i,j),\ell}\}_{i,j,\ell:\,i\neq j\neq\ell}$ is independent of $\{V_{(i,j),\ell}\}_{i,j,\ell:\,i\neq j\neq\ell}$ and $\{W_{(i,j),\ell}\}_{i,j,\ell:\,i\neq j\neq\ell}$ is an independent copy of $\{V_{(i,j),\ell}\}_{i,j,\ell:\,i\neq j\neq\ell}$, it holds that $N\mathbb{E}[q_{\ell i}\{\mathbf{c}^{-(i,j),\ell}\}\dot{c}_{(i,j),\ell}c_{(j,\ell),i}] = \mathbb{E}[q_{\ell i}\{\mathbf{c}^{-(i,j),\ell}\}][v\mathbb{E}\{Y_{(i,j),\ell}Y_{(j,\ell),i}\} + (1-v)\mathbb{E}\{V_{(i,j),\ell}V_{(j,\ell),i}\} - \mathbb{E}\{W_{(i,j),\ell}W_{(j,\ell),i}\}] = 0$. Analogously, we also have $\mathbb{E}[q_{\ell i}\{\mathbf{c}^{-(i,j),\ell}\}\dot{c}_{(i,j),\ell}c_{(\ell,j),i}]$=0. Then

$$\mathrm{I}_2(i,j,\ell,i) = \sum_{m:\,m\neq i,j,\ell}\int_0^1 \mathbb{E}\big[q_{\ell i}\{\mathbf{c}^{-(i,j),\ell}\}\dot{c}_{(i,j),\ell}c_{(m,j),i}\big]\,\mathrm{d}t + \sum_{m:\,m\neq i,j,\ell}\int_0^1 \mathbb{E}\big[q_{\ell i}\{\mathbf{c}^{-(i,j),\ell}\}\dot{c}_{(i,j),\ell}c_{(m,\ell),i}\big]\,\mathrm{d}t \tag{S.3.14}$$

$$+\sum_{m:\,m\neq i,j,\ell}\int_0^1 \mathbb{E}\big[q_{\ell i}\{\mathbf{c}^{-(i,j),\ell}\}\dot{c}_{(i,j),\ell}c_{(j,m),i}\big]\,\mathrm{d}t+\sum_{m:\,m\neq i,j,\ell}\int_0^1 \mathbb{E}\big[q_{\ell i}\{\mathbf{c}^{-(i,j),\ell}\}\dot{c}_{(i,j),\ell}c_{(\ell,m),i}\big]\,\mathrm{d}t\,.$$

In the sequel, we only need to bound each term in (S.3.14). For $\mathbb{E}[q_{\ell i}\{\mathbf{c}^{-(i,j),\ell}\}\dot{c}_{(i,j),\ell}c_{(\ell,m),i}]$ with $m\neq i,j,\ell$, it holds that $N\mathbb{E}[q_{\ell i}\{\mathbf{c}^{-(i,j),\ell}\}\dot{c}_{(i,j),\ell}c_{(\ell,m),i}]=v\mathbb{E}[q_{\ell i}\{\mathbf{c}^{-(i,j),\ell}\}Y_{(i,j),\ell}Y_{(\ell,m),i}]+(1-v)\mathbb{E}[q_{\ell i}\{\mathbf{c}^{-(i,j),\ell}\}V_{(i,j),\ell}V_{(\ell,m),i}]$ $-\mathbb{E}[q_{\ell i}\{\mathbf{c}^{-(i,j),\ell}\}W_{(i,j),\ell}W_{(\ell,m),i}]$. Following the same arguments in Section S2 to show $V_{(i,j),\ell}$ is independent of $\mathbf{c}^{-(i,j),\ell}$, we also have $\{V_{(\ell,m),i},W_{(\ell,m),i}\}$ is independent of $\mathbf{c}^{-(i,j),\ell}$. Thus,

$$N\cdot\mathbb{E}[q_{\ell i}\{\mathbf{c}^{-(i,j),\ell}\}\dot{c}_{(i,j),\ell}c_{(\ell,m),i}]=v\mathbb{E}[q_{\ell i}\{\mathbf{c}^{-(i,j),\ell}\}Y_{(i,j),\ell}Y_{(\ell,m),i}]-v\mathbb{E}[q_{\ell i}\{\mathbf{c}^{-(i,j),\ell}\}]\mathbb{E}\{Y_{(i,j),\ell}Y_{(\ell,m),i}\}\,. \tag{S.3.15}$$

Notice that $Y_{(\ell,m),i}$ is a function of $\{\mathring{Z}_{\ell,i},\mathring{Z}_{i,m},\mathring{Z}_{\ell,m}\}$, and $c_{(i',j'),\ell'}$'s involving $\mathring{Z}_{i,\ell}$ are not included in $\mathbf{c}^{-(i,j),\ell}$. Similar to the strategy used in Section S3.1, we can remove $c_{(i',j'),\ell'}$'s that related to $\{\mathring{Z}_{\ell,m},\mathring{Z}_{i,m}\}$ from $\mathbf{c}^{-(i,j),\ell}$, and denote by $\mathbf{c}^{-(i,j),\ell}$. We define $\mathbf{c}^{(i,j),\ell,(\ell,m),i}=\sum_{i',j':\,i'\neq j'}\mathbf{c}_{(i',j')}\circ\{\mathbf{a}^{(i,j,\ell)}_{(i',j')}+\mathbf{a}^{(\ell,m)}_{(i',j')}+\mathbf{a}^{(i,m)}_{(i',j')}-\mathbf{a}^{(i,j,\ell)}_{(i',j')}\circ\mathbf{a}^{(\ell,m)}_{(i',j')}-\mathbf{a}^{(i,j,\ell)}_{(i',j')}\circ\mathbf{a}^{(i,m)}_{(i',j')}-\mathbf{a}^{(\ell,m)}_{(i',j')}\circ\mathbf{a}^{(i,m)}_{(i',j')}+\mathbf{a}^{(i,j,\ell)}_{(i',j')}\circ\mathbf{a}^{(\ell,m)}_{(i',j')}\circ\mathbf{a}^{(i,m)}_{(i',j')}\}$. We know that $Y_{(i,j),\ell}Y_{(\ell,m),i}$ is independent of $\mathbf{c}-\mathbf{c}^{(i,j),\ell,(\ell,m),i}$. Recall $\mathbf{c}^{(i,j),\ell}=\sum_{i',j':\,i'\neq j'}\mathbf{c}_{(i',j')}\circ\mathbf{a}^{(i,j,\ell)}_{(i',j')}$. Then $\mathbf{c}^{(i,j),\ell,(\ell,m),i}-\mathbf{c}^{(i,j),\ell}=\{\mathbf{c}_{(\ell,m)}+\mathbf{c}_{(m,\ell)}\}\circ(\mathbf{1}-\mathbf{e}_i-\mathbf{e}_j)+\{\mathbf{c}_{(i,m)}+\mathbf{c}_{(m,i)}\}\circ(\mathbf{1}-\mathbf{e}_\ell-\mathbf{e}_j)+\sum_{u\neq i,j,\ell,m}\{c_{(u,m),i}+c_{(m,u),i}\}\mathbf{e}_i+\sum_{u\neq i,j,\ell,m}\{c_{(u,m),\ell}+c_{(m,u),\ell}\}\mathbf{e}_\ell+\sum_{u\neq i,j,\ell,m}\{c_{(u,\ell),m}+c_{(\ell,u),m}+c_{(u,i),m}+c_{(i,u),m}\}\mathbf{e}_m$. Let $\mathbf{c}^{-(i,j),\ell,(\ell,m),i}=\mathbf{c}-\mathbf{c}^{(i,j),\ell,(\ell,m),i}$ and $\mathbf{c}^{(i,j),\ell,(\ell,m),i-(i,j),\ell}=\mathbf{c}^{-(i,j),\ell}-\mathbf{c}^{-(i,j),\ell,(\ell,m),i}$. Then $\mathbf{c}^{(i,j),\ell,(\ell,m),i-(i,j),\ell}=\mathbf{c}^{(i,j),\ell,(\ell,m),i}-\mathbf{c}^{(i,j),\ell}$. By Taylor expansion, $\mathbb{E}[q_{\ell i}\{\mathbf{c}^{-(i,j),\ell}\}Y_{(i,j),\ell}Y_{(\ell,m),i}]=\mathbb{E}[q_{\ell i}\{\mathbf{c}^{-(i,j),\ell,(\ell,m),i}\}]\mathbb{E}\{Y_{(i,j),\ell}Y_{(\ell,m),i}\}+\sum_{s=1}^p\int_0^1\mathbb{E}[q_{\ell is}\{\mathbf{c}^{-(i,j),\ell,(\ell,m),i}+\tau\mathbf{c}^{(i,j),\ell,(\ell,m),i-(i,j),\ell}\}Y_{(i,j),\ell}Y_{(\ell,m),i}c_s^{(i,j),\ell,(\ell,m),i-(i,j),\ell}]\,\mathrm{d}\tau$. Together with (S.3.15), it holds that

$$\begin{aligned}\big|N\cdot\mathbb{E}[q_{\ell i}\{\mathbf{c}^{-(i,j),\ell}\}\dot{c}_{(i,j),\ell}c_{(\ell,m),i}]\big|\le{}&\underbrace{\big|\mathbb{E}[q_{\ell i}\{\mathbf{c}^{-(i,j),\ell,(\ell,m),i}\}]-\mathbb{E}[q_{\ell i}\{\mathbf{c}^{-(i,j),\ell}\}]\big|\big|\mathbb{E}\{Y_{(i,j),\ell}Y_{(\ell,m),i}\}\big|}_{R_1(i,j,\ell,m)}\\&+\underbrace{\sum_{s=1}^p\int_0^1\mathbb{E}[|q_{\ell is}\{\mathbf{c}^{-(i,j),\ell,(\ell,m),i}+\tau\mathbf{c}^{(i,j),\ell,(\ell,m),i-(i,j),\ell}\}|\times|Y_{(i,j),\ell}||Y_{(\ell,m),i}||c_s^{(i,j),\ell,(\ell,m),i-(i,j),\ell}|]\,\mathrm{d}\tau}_{R_2(i,j,\ell,m)}\,.\end{aligned}$$

Note that $\mathbf{c}^{-(i,j),\ell,(\ell,m),i}+\tau\mathbf{c}^{(i,j),\ell,(\ell,m),i-(i,j),\ell}=\mathbf{c}-\mathbf{c}^{(i,j),\ell}+(\tau-1)\mathbf{c}^{(i,j),\ell,(\ell,m),i-(i,j),\ell}$. Following the identical arguments in Section S3.1 for bounding the term on the right-hand side of (S.3.4), we have $\sum_{i:\,i\neq j,m}\sum_{\ell:\,\ell\neq i,j,m}R_2(i,j,\ell,m)\lesssim p^{-1/2}\phi\beta^2\log^{3/2}p$. By Taylor expansion, $\mathbb{E}[q_{\ell i}\{\mathbf{c}^{-(i,j),\ell}\}]-\mathbb{E}[q_{\ell i}\{\mathbf{c}^{-(i,j),\ell,(\ell,m),i}\}]$ $=\sum_{s=1}^p\int_0^1\mathbb{E}[q_{\ell is}\{\mathbf{c}^{-(i,j),\ell,(\ell,m),i}+\tau\mathbf{c}^{(i,j),\ell,(\ell,m),i-(i,j),\ell}\}c_s^{(i,j),\ell,(\ell,m),i-(i,j),\ell}]\,\mathrm{d}\tau$. Thus, $\sum_{i:\,i\neq j,m}\sum_{\ell:\,\ell\neq i,j,m}R_1(i,j,\ell,m)$ $\lesssim p^{-1/2}\phi\beta^2\log^{3/2}p$. Then $\sum_{i:\,i\neq j,m}\sum_{\ell:\,\ell\neq i,j,m}|N\cdot\mathbb{E}[q_{\ell i}\{\mathbf{c}^{-(i,j),\ell}\}\dot{c}_{(i,j),\ell}c_{(\ell,m),i}]|\lesssim p^{-1/2}\phi\beta^2\log^{3/2}p$, which implies $\sum_{j=1}^p\sum_{i:\,i\neq j}\sum_{\ell:\,\ell\neq i,j}|\sum_{m:\,m\neq i,j,\ell}\int_0^1\mathbb{E}[q_{\ell i}\{\mathbf{c}^{-(i,j),\ell}\}\dot{c}_{(i,j),\ell}c_{(\ell,m),i}]\,\mathrm{d}t|\lesssim p^{-1/2}\phi\beta^2\log^{3/2}p$. Analogously, we can obtain the same result for other terms in (S.3.14). Recall $\beta=\phi\log p$. Hence, we have that $\sum_{j=1}^p\sum_{i:\,i\neq j}\sum_{\ell:\,\ell\neq i,j}|\mathrm{I}_2(i,j,\ell,i)|\lesssim p^{-1/2}\phi^3\log^{7/2}p$. We also have $\sum_{j=1}^p\sum_{i:\,i\neq j}\sum_{\ell:\,\ell\neq i,j}|\mathrm{I}_2(i,j,\ell,j)|\lesssim p^{-1/2}\phi^3\log^{7/2}p$. and $\sum_{j=1}^p\sum_{i:\,i\neq j}\sum_{\ell:\,\ell\neq i,j}|\mathrm{I}_2(i,j,\ell,\ell)|\lesssim p^{-1/2}\phi^3\log^{7/2}p$.

S4 To prove $\sum_{i,j:\, i\neq j}\sum_{\ell:\, \ell\neq i,j}\sum_{k,l=1}^p |\mathrm{I}_3(i,j,\ell,k,l)| \lesssim p^{-1/2}\phi^3\log^{7/2} p$

To simplify the notation, we write $\mathbf{c}(t)$, $\mathbf{c}^{-(i,j),\ell}(t)$, $c_k^{(i,j),\ell}(t)$, $c_{(i,j),\ell}(t)$ and $\dot{c}_{(i,j),\ell}(t)$ as $\mathbf{c}$, $\mathbf{c}^{-(i,j),\ell}$, $c_k^{(i,j),\ell}$, $c_{(i,j),\ell}$ and $\dot{c}_{(i,j),\ell}$, respectively. Define $\mathcal{E}=\{|Y_{(i,j),\ell}|\vee|V_{(i,j),\ell}|\vee|W_{(i,j),\ell}|\leq p^{1/2}/(4\beta)$ for any $i\neq j\neq \ell\}$. We then have

$$\mathrm{I}_3(i,j,\ell,k,l) = \underbrace{\int_0^1\int_0^1(1-\tau)\mathbb{E}[I(\mathcal{E}^c)q_{\ell kl}\{\mathbf{c}^{-(i,j),\ell}+\tau\mathbf{c}^{(i,j),\ell}\}\dot{c}_{(i,j),\ell}c_k^{(i,j),\ell}c_l^{(i,j),\ell}]\,\mathrm{d}\tau\mathrm{d}t}_{\mathrm{I}_{3,1}(i,j,\ell,k,l)}$$
$$+\underbrace{\int_0^1\int_0^1(1-\tau)\mathbb{E}[I(\mathcal{E})q_{\ell kl}\{\mathbf{c}^{-(i,j),\ell}+\tau\mathbf{c}^{(i,j),\ell}\}\dot{c}_{(i,j),\ell}c_k^{(i,j),\ell}c_l^{(i,j),\ell}]\,\mathrm{d}\tau\mathrm{d}t}_{\mathrm{I}_{3,2}(i,j,\ell,k,l)}\,.$$

Let $\omega(t)=1/(\sqrt{t}\wedge\sqrt{1-t})$ for any $t\in(0,1)$. Notice that $\dot{c}_{(i,j),\ell}=N^{-1/2}[t^{-1/2}\{\sqrt{v}Y_{(i,j),\ell}+\sqrt{1-v}V_{(i,j),\ell}\}-(1-t)^{-1/2}W_{(i,j),\ell}]$. Then $\max_{i,j,\ell:\, i\neq j\neq\ell}|\dot{c}_{(i,j),\ell}|\lesssim p^{-1}\omega(t)\max_{i,j,\ell:\, i\neq j\neq\ell}\{|Y_{(i,j),\ell}|\vee|V_{(i,j),\ell}|\vee|W_{(i,j),\ell}|\}$. On the other hand, same as (S.3.5), it holds that

$$\begin{aligned}\max_{i,j,\ell:\, i\neq j\neq\ell}|\mathbf{c}^{(i,j),\ell}|_\infty \leq{}& \frac{54}{\sqrt{N}}\max_{i,j,\ell:\, i\neq j\neq\ell}\big\{|Y_{(i,j),\ell}|\vee|V_{(i,j),\ell}|\vee|W_{(i,j),\ell}|\big\}\\
&+\max_{i,j,\ell:\, i\neq j\neq\ell}\sum_{\substack{s_1,s_2:\, s_1\neq s_2\\ s_1,s_2\in\{i,j,\ell\}}}\bigg|\frac{1}{\sqrt{N}}\sum_{m:\, m\neq i,j,\ell}Y_{(s_1,m),s_2}\bigg|\\
&+\max_{i,j,\ell:\, i\neq j\neq\ell}\sum_{\substack{s_1,s_2:\, s_1\neq s_2\\ s_1,s_2\in\{i,j,\ell\}}}\bigg|\frac{1}{\sqrt{N}}\sum_{m:\, m\neq i,j,\ell}Y_{(m,s_1),s_2}\bigg|\\
&+\max_{i,j,\ell:\, i\neq j\neq\ell}\sum_{\substack{s_1,s_2:\, s_1\neq s_2\\ s_1,s_2\in\{i,j,\ell\}}}\bigg|\frac{1}{\sqrt{N}}\sum_{m:\, m\neq i,j,\ell}V_{(s_1,m),s_2}\bigg|\\
&+\max_{i,j,\ell:\, i\neq j\neq\ell}\sum_{\substack{s_1,s_2:\, s_1\neq s_2\\ s_1,s_2\in\{i,j,\ell\}}}\bigg|\frac{1}{\sqrt{N}}\sum_{m:\, m\neq i,j,\ell}V_{(m,s_1),s_2}\bigg|\\
&+\max_{i,j,\ell:\, i\neq j\neq\ell}\sum_{\substack{s_1,s_2:\, s_1\neq s_2\\ s_1,s_2\in\{i,j,\ell\}}}\bigg|\frac{1}{\sqrt{N}}\sum_{m:\, m\neq i,j,\ell}W_{(s_1,m),s_2}\bigg|\\
&+\max_{i,j,\ell:\, i\neq j\neq\ell}\sum_{\substack{s_1,s_2:\, s_1\neq s_2\\ s_1,s_2\in\{i,j,\ell\}}}\bigg|\frac{1}{\sqrt{N}}\sum_{m:\, m\neq i,j,\ell}W_{(m,s_1),s_2}\bigg|\,.\end{aligned}\tag{S.4.1}$$

We define

$$\begin{aligned}\mathcal{E}(Y)=\bigg\{&\max_{\substack{s_1,s_2:\, s_1\neq s_2\\ s_1,s_2\in\{i,j,\ell\}}}\bigg|\frac{1}{\sqrt{N}}\sum_{m:\, m\neq i,j,\ell}Y_{(s_1,m),s_2}\bigg|\leq\frac{C_*\log^{1/2}p}{p^{1/2}}\ \text{ and}\\
&\max_{\substack{s_1,s_2:\, s_1\neq s_2\\ s_1,s_2\in\{i,j,\ell\}}}\bigg|\frac{1}{\sqrt{N}}\sum_{m:\, m\neq i,j,\ell}Y_{(m,s_1),s_2}\bigg|\leq\frac{C_*\log^{1/2}p}{p^{1/2}}\ \text{ for any } i\neq j\neq\ell\bigg\}\end{aligned}$$

for sufficiently large $C_* > 0$. We can also define $\mathcal{E}(V)$ and $\mathcal{E}(W)$ in the same manner. Let $\tilde{\mathcal{E}} = \mathcal{E}(Y) \cap \mathcal{E}(V) \cap \mathcal{E}(W)$. Using the same arguments in Section S3.1 to derive the upper bound of $\mathbb{P}(\mathcal{E}_2^c)$ for $\mathcal{E}_2$ specified in (S.3.11), it holds that $\mathbb{P}(\tilde{\mathcal{E}}^c) \lesssim p^{-C}$. Here C can be sufficiently large if we select sufficiently large C_*. Hence, restricted on $\tilde{\mathcal{E}}$, we have $\max_{i,j,\ell:\, i\neq j\neq \ell} |\mathbf{c}^{(i,j),\ell}|_\infty \lesssim p^{-1} \max_{i,j,\ell:\, i\neq j\neq \ell}\{|Y_{(i,j),\ell}| \vee |V_{(i,j),\ell}| \vee |W_{(i,j),\ell}|\} + p^{-1/2}\log^{1/2} p$, which implies that $\max_{i,j,\ell:\, i\neq j\neq \ell} |\dot{c}_{(i,j),\ell}||\mathbf{c}^{(i,j),\ell}|_\infty^2 \lesssim p^{-3}\omega(t) \max_{i,j,\ell:\, i\neq j\neq \ell}\{|Y_{(i,j),\ell}|^3 \vee |V_{(i,j),\ell}|^3 \vee |W_{(i,j),\ell}|^3\} + p^{-2}\omega(t)\log p \max_{i,j,\ell:\, i\neq j\neq \ell}\{|Y_{(i,j),\ell}| \vee |V_{(i,j),\ell}| \vee |W_{(i,j),\ell}|\}$ under $\tilde{\mathcal{E}}$. As shown in Lemma A.5 of Chernozhukov, Chetverikov and Kato (2013), there exists $U_{\ell kl}(\mathbf{v})$ such that $|q_{\ell kl}(\mathbf{v})| \leq U_{\ell kl}(\mathbf{v})$ for any $\mathbf{v} \in \mathbb{R}^p$, where $\sum_{\ell,k,l=1}^p U_{\ell kl}(\mathbf{v}) \lesssim \phi\beta^2$ for any $\mathbf{v} \in \mathbb{R}^p$. Then

$$\begin{aligned}\sum_{\ell,k,l=1}^p |\mathrm{I}_{3,1}(i,j,\ell,k,l)| &\lesssim \phi\beta^2 p^3 \int_0^1 \mathbb{E}\bigg\{I(\mathcal{E}^c) \max_{i,j,\ell:\, i\neq j\neq \ell} |\dot{c}_{(i,j),\ell}||\mathbf{c}^{(i,j),\ell}|_\infty^2\bigg\}\,\mathrm{d}t \\ &= \phi\beta^2 p^3 \int_0^1 \mathbb{E}\bigg\{I(\mathcal{E}^c \cap \tilde{\mathcal{E}}) \max_{i,j,\ell:\, i\neq j\neq \ell} |\dot{c}_{(i,j),\ell}||\mathbf{c}^{(i,j),\ell}|_\infty^2\bigg\}\,\mathrm{d}t \\ &\quad + \phi\beta^2 p^3 \int_0^1 \mathbb{E}\bigg\{I(\mathcal{E}^c \cap \tilde{\mathcal{E}}^c) \max_{i,j,\ell:\, i\neq j\neq \ell} |\dot{c}_{(i,j),\ell}||\mathbf{c}^{(i,j),\ell}|_\infty^2\bigg\}\,\mathrm{d}t\,.\end{aligned} \tag{S.4.2}$$

Notice that $\max_{i,j,\ell:\, i\neq j\neq \ell} |Y_{(i,j),\ell}| \leq C$, and $V_{(i,j),\ell}$ and $W_{(i,j),\ell}$ are normal random variables. It holds that

$$\mathbb{P}\bigg[\max_{i,j,\ell:\, i\neq j\neq \ell}\{|Y_{(i,j),\ell}| \vee |V_{(i,j),\ell}| \vee |W_{(i,j),\ell}|\} > u\bigg] \leq Cp^3\exp(-Cu^2) \tag{S.4.3}$$

for any $u > 0$. We have $\int_0^1 \mathbb{E}\{I(\mathcal{E}^c \cap \tilde{\mathcal{E}}) \max_{i,j,\ell:\, i\neq j\neq \ell} |\dot{c}_{(i,j),\ell}||\mathbf{c}^{(i,j),\ell}|_\infty^2\}\,\mathrm{d}t \lesssim p^{-3}\mathbb{E}[I(\mathcal{E}^c)\max_{i,j,\ell:\, i\neq j\neq \ell}\{|Y_{(i,j),\ell}|^3 \vee |V_{(i,j),\ell}|^3 \vee |W_{(i,j),\ell}|^3\}] + p^{-2}\log p\,\mathbb{E}[I(\mathcal{E}^c)\max_{i,j,\ell:\, i\neq j\neq \ell}\{|Y_{(i,j),\ell}| \vee |V_{(i,j),\ell}| \vee |W_{(i,j),\ell}|\}]$. By Cauchy-Schwarz inequality, we have $\mathbb{E}[I(\mathcal{E}^c)\max_{i,j,\ell:\, i\neq j\neq \ell}\{|Y_{(i,j),\ell}| \vee |V_{(i,j),\ell}| \vee |W_{(i,j),\ell}|\}] \leq \mathbb{P}^{1/2}(\mathcal{E}^c) \cdot \mathbb{E}^{1/2}[\max_{i,j,\ell:\, i\neq j\neq \ell}\{|Y_{(i,j),\ell}|^2 \vee |V_{(i,j),\ell}|^2 \vee |W_{(i,j),\ell}|^2\}] \lesssim p^3\exp(-Cp\beta^{-2})$. Analogously, we have $\mathbb{E}[I(\mathcal{E}^c)\max_{i,j,\ell:\, i\neq j\neq \ell}\{|Y_{(i,j),\ell}|^3 \vee |V_{(i,j),\ell}|^3 \vee |W_{(i,j),\ell}|^3\}] \lesssim p^3\exp(-Cp\beta^{-2})$. Then it holds that

$$\int_0^1 \mathbb{E}\bigg\{I(\mathcal{E}^c \cap \tilde{\mathcal{E}}) \max_{i,j,\ell:\, i\neq j\neq \ell} |\dot{c}_{(i,j),\ell}||\mathbf{c}^{(i,j),\ell}|_\infty^2\bigg\}\,\mathrm{d}t \lesssim p\log p \cdot \exp(-Cp\beta^{-2})\,. \tag{S.4.4}$$

By Cauchy-Schwarz inequality, it holds that

$$\begin{aligned}&\mathbb{E}\bigg\{I(\mathcal{E}^c \cap \tilde{\mathcal{E}}^c) \max_{i,j,\ell:\, i\neq j\neq \ell} |\dot{c}_{(i,j),\ell}||\mathbf{c}^{(i,j),\ell}|_\infty^2\bigg\} \\ &\qquad \leq \mathbb{P}^{1/2}(\tilde{\mathcal{E}}^c) \cdot \mathbb{E}^{1/2}\bigg\{\max_{i,j,\ell:\, i\neq j\neq \ell} |\dot{c}_{(i,j),\ell}|^2|\mathbf{c}^{(i,j),\ell}|_\infty^4\bigg\} \\ &\qquad \leq \mathbb{P}^{1/2}(\tilde{\mathcal{E}}^c) \cdot \mathbb{E}^{1/4}\bigg\{\max_{i,j,\ell:\, i\neq j\neq \ell} |\dot{c}_{(i,j),\ell}|^4\bigg\} \cdot \mathbb{E}^{1/4}\bigg\{\max_{i,j,\ell:\, i\neq j\neq \ell} |\mathbf{c}^{(i,j),\ell}|_\infty^8\bigg\}\,.\end{aligned}$$

Notice that $\mathbb{P}(\tilde{\mathcal{E}}^c) \lesssim p^{-C}$, where C can be sufficiently large if we select sufficiently large C_* in the definition of $\tilde{\mathcal{E}}$. Thus, with suitable selection of C_*, we have

$$\int_0^1 \mathbb{E}\bigg\{I(\mathcal{E}^c \cap \tilde{\mathcal{E}}^c) \max_{i,j,\ell:\, i\neq j\neq \ell} |\dot{c}_{(i,j),\ell}||\mathbf{c}^{(i,j),\ell}|_\infty^2\bigg\}\,\mathrm{d}t \lesssim \frac{1}{p^{11/2}}\,. \tag{S.4.5}$$

Together with (S.4.4) and $\beta = \phi \log p$, (S.4.2) implies that

$$
\begin{aligned}
\sum_{i,j} \sum_{\ell,k,l=1}^{p} |\mathrm{I}_{3,1}(i,j,\ell,k,l)| &\lesssim \frac{\phi\beta^2}{p^{1/2}} + \phi\beta^2 p^6 \log p \cdot \exp(-Cp\beta^{-2}) \\
&= \frac{\phi^3 \log^2 p}{p^{1/2}} + \phi^3 p^6 \log^3 p \cdot \exp\bigg(-\frac{Cp}{\phi^2 \log^2 p}\bigg). \qquad \text{(S.4.6)}
\end{aligned}
$$

In the sequel, we begin to consider $\mathrm{I}_{3,2}(i,j,\ell,k,l)$. We have that

$$
\begin{aligned}
&|\mathrm{I}_{3,2}(i,j,\ell,k,l)| \\
\lesssim\ & \int_0^1 \int_0^1 (1-\tau)\mathbb{E}[I(\mathcal{E}\cap\tilde{\mathcal{E}})U_{\ell kl}\{\mathbf{c}^{-(i,j),\ell} + \tau\mathbf{c}^{(i,j),\ell}\}|\dot{c}_{(i,j),\ell}||c_k^{(i,j),\ell}||c_l^{(i,j),\ell}|]\,\mathrm{d}\tau\mathrm{d}t \qquad \text{(S.4.7)} \\
&+ \int_0^1 \int_0^1 (1-\tau)\mathbb{E}[I(\mathcal{E}\cap\tilde{\mathcal{E}}^c)U_{\ell kl}\{\mathbf{c}^{-(i,j),\ell} + \tau\mathbf{c}^{(i,j),\ell}\}|\dot{c}_{(i,j),\ell}||c_k^{(i,j),\ell}||c_l^{(i,j),\ell}|]\,\mathrm{d}\tau\mathrm{d}t\,.
\end{aligned}
$$

Notice that

$$
\begin{aligned}
&\int_0^1 \int_0^1 (1-\tau)\mathbb{E}[I(\mathcal{E}\cap\tilde{\mathcal{E}}^c)U_{\ell kl}\{\mathbf{c}^{-(i,j),\ell} + \tau\mathbf{c}^{(i,j),\ell}\}|\dot{c}_{(i,j),\ell}||c_k^{(i,j),\ell}||c_l^{(i,j),\ell}|]\,\mathrm{d}\tau\mathrm{d}t \\
&\qquad \lesssim \phi\beta^2 \int_0^1 \mathbb{E}[I(\mathcal{E}\cap\tilde{\mathcal{E}}^c)|\dot{c}_{(i,j),\ell}||c_k^{(i,j),\ell}||c_l^{(i,j),\ell}|]\,\mathrm{d}t \\
&\qquad \lesssim \phi\beta^2 \int_0^1 \mathbb{E}\bigg\{I(\mathcal{E}\cap\tilde{\mathcal{E}}^c)\max_{i,j,\ell:\, i\neq j\neq \ell}|\dot{c}_{(i,j),\ell}||\mathbf{c}^{(i,j),\ell}|_\infty^2\bigg\}\,\mathrm{d}t\,.
\end{aligned}
$$

Same as (S.4.5), we have

$$
\begin{aligned}
\sum_{i,j:\, i\neq j}\sum_{\ell:\, \ell\neq i,j}\sum_{k,l=1}^{p} \int_0^1 \int_0^1 &(1-\tau)\mathbb{E}[I(\mathcal{E}\cap\tilde{\mathcal{E}}^c)U_{\ell kl}\{\mathbf{c}^{-(i,j),\ell} + \tau\mathbf{c}^{(i,j),\ell}\} \\
&\times |\dot{c}_{(i,j),\ell}||c_k^{(i,j),\ell}||c_l^{(i,j),\ell}|]\,\mathrm{d}\tau\mathrm{d}t \lesssim \frac{\phi^3\log^2 p}{p^{1/2}}\,. \qquad \text{(S.4.8)}
\end{aligned}
$$

Restricted on $\mathcal{E}\cap\tilde{\mathcal{E}}$, (S.4.1) implies that $\max_{i,j,\ell:\, i\neq j\neq\ell}|\mathbf{c}^{(i,j),\ell}|_\infty \lesssim p^{-1/2}\beta^{-1} + p^{-1/2}\log^{1/2} p \leq 3/(4\beta)$ for sufficiently large p if $\phi \ll p^{1/2}\log^{-3/2} p$. Lemma A.6 of Chernozhukov, Chetverikov and Kato (2013) implies that, restricted on $\mathcal{E}\cap\tilde{\mathcal{E}}$, $U_{\ell kl}(\mathbf{c}) \lesssim U_{\ell kl}\{\mathbf{c}^{-(i,j),\ell} + \tau\mathbf{c}^{(i,j),\ell}\} \lesssim U_{\ell kl}(\mathbf{c})$ for any $t\in[0,1]$. We then have

$$
\begin{aligned}
&\int_0^1 \int_0^1 (1-\tau)\mathbb{E}[I(\mathcal{E}\cap\tilde{\mathcal{E}})U_{\ell kl}\{\mathbf{c}^{-(i,j),\ell} + \tau\mathbf{c}^{(i,j),\ell}\}|\dot{c}_{(i,j),\ell}||c_k^{(i,j),\ell}||c_l^{(i,j),\ell}|]\,\mathrm{d}\tau\mathrm{d}t \\
&\qquad \lesssim \int_0^1 \underbrace{\mathbb{E}\{I(\mathcal{E}\cap\tilde{\mathcal{E}})U_{\ell kl}(\mathbf{c})|\dot{c}_{(i,j),\ell}||c_k^{(i,j),\ell}||c_l^{(i,j),\ell}|\}}_{R_3(i,j,\ell,k,l)}\,\mathrm{d}t\,.
\end{aligned}
$$

Notice that

$$c_k^{(i,j),\ell} = \begin{cases} c_{(i,j),k} + c_{(j,i),k} + c_{(j,\ell),k} + c_{(\ell,j),k} + c_{(\ell,i),k} + c_{(i,\ell),k}\,, & \text{if } k \neq i,j,\ell\,, \\ \sum_{m\neq i,j,\ell} \{c_{(m,j),i} + c_{(j,m),i}\} + \sum_{m\neq i,\ell} \{c_{(m,\ell),i} + c_{(\ell,m),i}\}\,, & \text{if } k = i\,, \\ \sum_{m\neq i,j,\ell} \{c_{(m,i),j} + c_{(i,m),j}\} + \sum_{m\neq j,\ell} \{c_{(m,\ell),j} + c_{(\ell,m),j}\}\,, & \text{if } k = j\,, \\ \sum_{m\neq i,j,\ell} \{c_{(m,j),\ell} + c_{(j,m),\ell}\} + \sum_{m\neq i,\ell} \{c_{(m,i),\ell} + c_{(i,m),\ell}\}\,, & \text{if } k = \ell\,. \end{cases}$$

Since $\sum_{\ell,k,l=1}^{p} U_{\ell kl}(\mathbf{v}) \lesssim \phi\beta^2$ for any $\mathbf{v} \in \mathbb{R}^p$, we have

$$\begin{aligned} \sum_{\ell} \sum_{k,l:\, k,l\neq i,j,\ell} R_3(i,j,\ell,k,l) &\lesssim \phi\beta^2 \cdot \mathbb{E}\bigg\{ \max_{\ell:\, \ell\neq i,j} |\dot{c}_{(i,j),\ell}| \cdot \max_{\ell:\, \ell\neq i,j} \max_{k:\, k\neq i,j,\ell} |c_k^{(i,j),\ell}|^2 \bigg\} \\ &\lesssim \frac{\phi\beta^2\omega(t)}{p^3} \cdot \mathbb{E}\bigg[\max_{i,j,\ell:\, i\neq j\neq \ell} \big\{ |Y_{(i,j),\ell}|^3 \vee |V_{(i,j),\ell}|^3 \vee |W_{(i,j),\ell}|^3 \big\} \bigg] \\ &\lesssim \frac{\phi\beta^2\omega(t)\log^{3/2} p}{p^3}\,, \end{aligned}$$

which implies that

$$\begin{aligned} \sum_{i,j:\, i\neq j} \sum_{\ell:\, \ell\neq i,j} \sum_{k,l:\, k,l\neq i,j,\ell} \int_0^1 \int_0^1 (1-\tau)\mathbb{E}[I(\mathcal{E}\cap\tilde{\mathcal{E}}) U_{\ell kl}\{\mathbf{c}^{-(i,j),\ell} + \tau\mathbf{c}^{(i,j),\ell}\} \\ \times |\dot{c}_{(i,j),\ell}||c_k^{(i,j),\ell}||c_l^{(i,j),\ell}|]\,\mathrm{d}\tau\mathrm{d}t \\ \lesssim \frac{\phi\beta^2\log^{3/2} p}{p} = \frac{\phi^3\log^{7/2} p}{p}\,. \end{aligned} \tag{S.4.9}$$

If $k \in \{i,j,\ell\}$, we have

$$\begin{aligned} |c_k^{(i,j),\ell}| \lesssim & \sum_{\substack{s_1,s_2:\, s_1\neq s_2 \\ s_1,s_2\in\{i,j,\ell\}}} \bigg| \frac{1}{\sqrt{N}} \sum_{m:\, m\neq i,j,\ell} Y_{(s_1,m),s_2} \bigg| + \sum_{\substack{s_1,s_2:\, s_1\neq s_2 \\ s_1,s_2\in\{i,j,\ell\}}} \bigg| \frac{1}{\sqrt{N}} \sum_{m:\, m\neq i,j,\ell} Y_{(m,s_1),s_2} \bigg| \\ & + \sum_{\substack{s_1,s_2:\, s_1\neq s_2 \\ s_1,s_2\in\{i,j,\ell\}}} \bigg| \frac{1}{\sqrt{N}} \sum_{m:\, m\neq i,j,\ell} V_{(s_1,m),s_2} \bigg| + \sum_{\substack{s_1,s_2:\, s_1\neq s_2 \\ s_1,s_2\in\{i,j,\ell\}}} \bigg| \frac{1}{\sqrt{N}} \sum_{m:\, m\neq i,j,\ell} V_{(m,s_1),s_2} \bigg| \\ & + \sum_{\substack{s_1,s_2:\, s_1\neq s_2 \\ s_1,s_2\in\{i,j,\ell\}}} \bigg| \frac{1}{\sqrt{N}} \sum_{m:\, m\neq i,j,\ell} W_{(s_1,m),s_2} \bigg| + \sum_{\substack{s_1,s_2:\, s_1\neq s_2 \\ s_1,s_2\in\{i,j,\ell\}}} \bigg| \frac{1}{\sqrt{N}} \sum_{m:\, m\neq i,j,\ell} W_{(m,s_1),s_2} \bigg|\,. \end{aligned}$$

Restricted on $\tilde{\mathcal{E}}$, it holds that $\max_{i,j,\ell:\, i\neq j\neq\ell} \max_{k\in\{i,j,\ell\}} |c_k^{(i,j),\ell}| \lesssim p^{-1/2}\log^{1/2} p$. Also, notice that $\mathbf{c}^{-(i,j),\ell} = \mathbf{c} - \mathbf{c}^{(i,j),\ell}$. Lemma A.6 of Chernozhukov, Chetverikov and Kato (2013) implies that, restricted on $\mathcal{E}\cap\tilde{\mathcal{E}}$, $U_{\ell kl}(\mathbf{c}) \lesssim U_{\ell kl}\{\mathbf{c}^{-(i,j),\ell}\} \lesssim U_{\ell kl}(\mathbf{c})$ for any $t \in [0,1]$. Then

$$\sum_{i,\ell} \sum_{l:\, l\neq i,j,\ell} R_3(i,j,\ell,i,l) \lesssim \frac{\phi\beta^2\log^{1/2} p}{p^{1/2}} \cdot \mathbb{E}\bigg\{ \max_{i,j,\ell:\, i\neq j\neq\ell} |\dot{c}_{(i,j),\ell}| \cdot \max_{i,j,\ell:\, i\neq j\neq\ell} \max_{l:\, l\neq i,j,\ell} |c_l^{(i,j),\ell}| \bigg\}$$

$$\lesssim \frac{\phi\beta^2\omega(t)\log^{1/2}p}{p^{3/2}} \cdot \mathbb{E}\bigg[\max_{i,j,\ell:\, i\neq j\neq \ell}\big\{|Y_{(i,j),\ell}|^2 \vee |V_{(i,j),\ell}|^2 \vee |W_{(i,j),\ell}|^2\big\}\bigg]$$
$$\lesssim \frac{\phi\beta^2\omega(t)\log^{3/2}p}{p^{3/2}}\,.$$

Then $\sum_{i,j:\,i\neq j}\sum_{\ell:\,\ell\neq i,j}\sum_{l:\,l\neq i,j,\ell}\int_0^1\int_0^1(1-\tau)\mathbb{E}[I(\mathcal{E}\cap\tilde{\mathcal{E}})U_{\ell il}\{\mathbf{c}^{-(i,j),\ell}+\tau\mathbf{c}^{(i,j),\ell}\}|\dot{c}_{(i,j),\ell}||c_i^{(i,j),\ell}||c_l^{(i,j),\ell}|]\,\mathrm{d}\tau\mathrm{d}t \lesssim p^{-1/2}\phi\beta^2\log^{3/2}p = p^{-1/2}\phi^3\log^{7/2}p$. We can also obtain the same bound for $k=j$ and ℓ. Thus,

$$\sum_{i,j:\,i\neq j}\sum_{\ell:\,\ell\neq i,j}\sum_{k:\,k=i,j,\ell}\sum_{l:\,l\neq i,j,\ell}\int_0^1\int_0^1(1-\tau)\mathbb{E}[I(\mathcal{E}\cap\tilde{\mathcal{E}})U_{\ell kl}\{\mathbf{c}^{-(i,j),\ell}+\tau\mathbf{c}^{(i,j),\ell}\}$$
$$\times|\dot{c}_{(i,j),\ell}||c_k^{(i,j),\ell}||c_l^{(i,j),\ell}|]\,\mathrm{d}\tau\mathrm{d}t \lesssim \frac{\phi^3\log^{7/2}p}{p^{1/2}}\,. \tag{S.4.10}$$

If $k,l\in\{i,j,\ell\}$, we have

$$\sum_{i,\ell}R_3(i,j,\ell,i,i) \lesssim \frac{\phi\beta^2\log p}{p}\cdot\mathbb{E}\bigg\{\max_{i,j,\ell:\,i\neq j\neq\ell}|\dot{c}_{(i,j),\ell}|\bigg\}$$
$$\lesssim \frac{\phi\beta^2\omega(t)\log p}{p^2}\cdot\mathbb{E}\bigg[\max_{i,j,\ell:\,i\neq j\neq\ell}\big\{|Y_{(i,j),\ell}|\vee|V_{(i,j),\ell}|\vee|W_{(i,j),\ell}|\big\}\bigg]$$
$$\lesssim \frac{\phi\beta^2\omega(t)\log^{3/2}p}{p^2}$$

and

$$\sum_{i,j,\ell}R_3(i,j,\ell,i,j) \lesssim \frac{\phi\beta^2\log p}{p}\cdot\mathbb{E}\bigg\{\max_{i,j,\ell:\,i\neq j\neq\ell}|\dot{c}_{(i,j),\ell}|\bigg\}$$
$$\lesssim \frac{\phi\beta^2\omega(t)\log p}{p^2}\cdot\mathbb{E}\bigg[\max_{i,j,\ell:\,i\neq j\neq\ell}\big\{|Y_{(i,j),\ell}|\vee|V_{(i,j),\ell}|\vee|W_{(i,j),\ell}|\big\}\bigg]$$
$$\lesssim \frac{\phi\beta^2\omega(t)\log^{3/2}p}{p^2}\,.$$

Thus, $\sum_{i,j:\,i\neq j}\sum_{\ell:\,\ell\neq i,j}\sum_{k,l:\,k,l=i,j,\ell}\int_0^1\int_0^1(1-\tau)\mathbb{E}[I(\mathcal{E}\cap\tilde{\mathcal{E}})U_{\ell kl}\{\mathbf{c}^{-(i,j),\ell}+\tau\mathbf{c}^{(i,j),\ell}\}|\dot{c}_{(i,j),\ell}||c_k^{(i,j),\ell}||c_l^{(i,j),\ell}|]\,\mathrm{d}\tau\mathrm{d}t \lesssim p^{-1}\phi^3\log^{7/2}p$. Together with (S.4.9) and (S.4.10), we have

$$\sum_{i,j:\,i\neq j}\sum_{\ell:\,\ell\neq i,j}\sum_{k,l=1}^{p}\int_0^1\int_0^1(1-\tau)\mathbb{E}[I(\mathcal{E}\cap\tilde{\mathcal{E}})U_{\ell kl}\{\mathbf{c}^{-(i,j),\ell}+\tau\mathbf{c}^{(i,j),\ell}\}$$
$$\times|\dot{c}_{(i,j),\ell}||c_k^{(i,j),\ell}||c_l^{(i,j),\ell}|]\,\mathrm{d}\tau\mathrm{d}t \lesssim \frac{\phi^3\log^{7/2}p}{p^{1/2}}\,. \tag{S.4.11}$$

Combining (S.4.8) and (S.4.11), (S.4.7) implies that $\sum_{i,j:\,i\neq j}\sum_{\ell:\,\ell\neq i,j}\sum_{k,l=1}^{p}|\mathrm{I}_{3,2}(i,j,\ell,k,l)| \lesssim p^{-1/2}\phi^3\log^{7/2}p$. Together with (S.4.6), it holds that $\sum_{i,j:\,i\neq j}\sum_{\ell:\,\ell\neq i,j}\sum_{k,l=1}^{p}|\mathrm{I}_3(i,j,\ell,k,l)| \lesssim p^{-1/2}\phi^3\log^{7/2}p+\phi^3p^6\log^3p\cdot\exp\{-Cp\phi^{-2}(\log p)^{-2}\}$. If we select $\phi\ll p^{1/2}\log^{-3/2}p$, we have $\sum_{i,j:\,i\neq j}\sum_{\ell:\,\ell\neq i,j}\sum_{k,l=1}^{p}|\mathrm{I}_3(i,j,\ell,k,l)| \lesssim p^{-1/2}\phi^3\log^{7/2}p$.