Asymptotic Theory of Eigenvectors for Latent Embeddings with Generalized Laplacian Matrices

Jianqing Fan[1], Yingying Fan[2], Jinchi Lv[2], Fan Yang[3], and Diwen Yu[3]

Princeton University[1], University of Southern California[2], and Tsinghua University[3]

Abstract

Laplacian matrices are widely used in practice to capture latent structural information in data, ranging from graphs to manifolds. Their normalization forms naturally induce dependencies among matrix entries, and such dependencies are known to pose significant challenges for advances in random matrix theory (RMT). Motivated by this, we introduce a general class of generalized Laplacian matrices, which includes both the standard Laplacian and random adjacency matrices as special cases, and we develop a new framework—Asymptotic Theory of Eigenvectors for latent embeddings with Generalized Laplacian matrices (ATE-GL)—for studying their spectral properties. Our theory is driven by two key ingredients: the use of generalized quadratic vector equations to handle dependency in RMT, and refined high-order asymptotic expansions for empirical spiked eigenvectors and eigenvalues based on local laws. These results lead to asymptotic normality for both spiked eigenvectors and eigenvalues, enabling precise statistical inference and uncertainty quantification for a broad class of applications involving generalized Laplacian matrices. We also discuss two motivating applications of the ATE-GL framework and demonstrate its effectiveness through numerical examples.

Key words: Graph and manifold embeddings; Asymptotic distributions; Eigenvectors and eigenvalues; Local laws; RMT under dependency; High dimensionality

I. Introduction

Degree-normalized Laplacian-type matrices are a cornerstone of modern network analysis and graph-based learning, as they underpin regularized spectral clustering under degree heterogeneity, diffusion and graph-smoothing estimators, and the normalized adjacency filters that appear in spectral and message-passing graph neural networks. They serve as fundamental objects of study and connect to a multitude of valuable graph properties; see, e.g., Chung (1997); Mohar et al. (1991); Merris (1994); Godsil and Royle (2001) for an overview. Given an $n \times n$ *adjacency matrix* $\widetilde{\mathbf{X}} = (\widetilde{X}_{ij})$ representing a network with n nodes, where the detailed definition of $\widetilde{\mathbf{X}}$ is presented at the beginning of Section II, its (symmetric normalized) Laplacian is defined as

$$\mathbf{I} - \mathbf{D}^{-1/2}\widetilde{\mathbf{X}}\mathbf{D}^{-1/2}, \tag{1}$$

where $\mathbf{I}$ is the identity matrix, and $\mathbf{D} := \operatorname{diag}(d_1, \cdots, d_n)$ is a diagonal matrix with $d_i := \sum_{j\in[n]} \widetilde{X}_{ij}$ denoting the degree of the ith node with $1 \le i \le n$. The Laplacian matrix finds applications in various domains such as information theory, communication, and Ramanujan graphs (Sipser and Spielman, 1996; Lubotzky et al., 1988; Donetti et al., 2006; Hoory et al., 2006); quantum graphs and quantum chaos (Smilansky, 2007; Braunstein et al., 2006; Kurasov, 2008; Kook, 2011); and mathematical biology and chemistry (Trinajstic et al., 1994; Klein, 2002; Xiao and Gutman, 2003; Estrada and Hatano, 2010; Freschi, 2011). Furthermore, the importance of the Laplacian matrix extends to other domains, such as manifold learning, where a similar and related concept is the "transition matrix" derived from the affinity matrix constructed based on a noisy point cloud of the manifold (Hardoon et al., 2004; Lederman and Talmon, 2018; Ding and Wu, 2020).

Statistical inference of network data involves matrices beyond the adjacency and Laplacian matrices. To give an important example, we consider the degree-corrected mixed membership (DCMM) model introduced in Jin

Jianqing Fan is Frederick L. Moore '18 Professor of Finance and Professor of Operations Research and Financial Engineering, Department of Operations Research and Financial Engineering, Princeton University, Princeton, NJ 08544, USA (E-mail: jqfan@princeton.edu). Yingying Fan is Centennial Chair in Business Administration and Professor, Data Sciences and Operations Department, Marshall School of Business, University of Southern California, Los Angeles, CA 90089 (E-mail: *fanyingy@marshall.usc.edu*). Jinchi Lv is Kenneth King Stonier Chair in Business Administration and Professor, Data Sciences and Operations Department, Marshall School of Business, University of Southern California, Los Angeles, CA 90089 (E-mail: *jinchilv@marshall.usc.edu*). Fan Yang is Associate Professor, Yau Mathematical Sciences Center, Tsinghua University, China (E-mail: *fyangmath@mail.tsinghua.edu.cn*). Diwen Yu is Ph.D. candidate, Yau Mathematical Sciences Center, Tsinghua University, China (E-mail: *ydw23@mails.tsinghua.edu.cn*). This work was supported in part by National Key R&D Program of China (No. 2023YFA1010400), and NSF grants EF-2125142 and DMS-2324490.

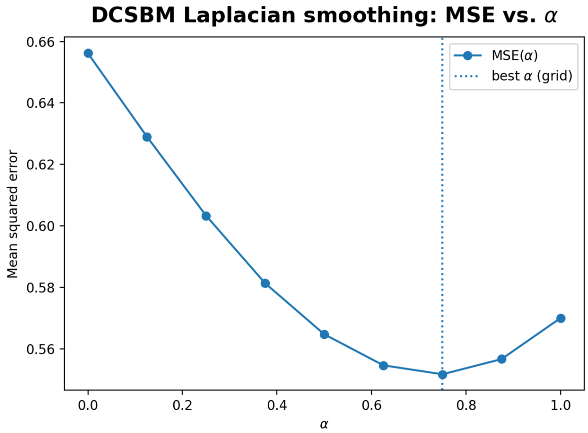

Fig. 1: The MSE as a function of α for regularized Laplacian smoothing. The minimum MSE is attained when $\alpha = 0.75$. The MSE is calculated as the average over 200 independent repetitions.

et al. (2024). The detailed probabilistic model is provided in Example 1. The adjacency matrix $\widetilde{\mathbf{X}}$ of this model is random, with entries given by independent Bernoulli random variables (subject to the symmetry constraint $\widetilde{\mathbf{X}} = \widetilde{\mathbf{X}}^T$). Assume that the graph contains K latent communities. Then the mean structure of $\widetilde{\mathbf{X}}$ admits the representation

$$\mathbf{H} := \mathbb{E}[\widetilde{\mathbf{X}}] = \mathbf{\Theta}\mathbf{\Pi}\mathbf{P}\mathbf{\Pi}^T\mathbf{\Theta} \in \mathbb{R}^{n\times n}, \tag{2}$$

where $\mathbf{\Theta}$ is an $n \times n$ diagonal matrix reflecting the degree heterogeneity of nodes in the random graph, $\mathbf{\Pi}$ is an $n \times K$ whose ith row specifies the community membership probabilities of node i, and $\mathbf{P}$ is a $K \times K$ matrix encoding the connection probabilities between communities. Compared to network models without significant degree heterogeneity, statistical inference for the DCMM model encounters additional complexities due to the presence of matrix $\mathbf{\Theta}$, whose entries can vary wildly in magnitude; see e.g., the related discussions in Fan et al. (2022b); Fan et al. (2022); Bhattacharya et al. (2023). To deal with such an issue, notice that under certain normalization, $\mathbf{\Lambda} := \mathbb{E}[\mathbf{D}]$ is proportional to $\mathbf{\Theta}$. On the other hand, by the law of large numbers (LLN), $\mathbf{\Lambda}$ can be well-approximated by $\mathbf{D}$ for large networks when the node degrees diverge. Thus, dividing $\widetilde{\mathbf{X}}$ by $\mathbf{D}$ on both sides largely removes the intrinsic degree heterogeneity of the DCMM model. This motivates the exploration of the following random matrix

$$\mathbf{D}^{-1}\widetilde{\mathbf{X}}\mathbf{D}^{-1}, \tag{3}$$

which is more suitable for certain applications.

Motivated by the above applications, in this paper we consider the *generalized Laplacian matrices* of large random networks defined as

$$\mathbf{X} \equiv \mathbf{X}(\alpha) := \mathbf{D}^{-\alpha}\widetilde{\mathbf{X}}\mathbf{D}^{-\alpha} \tag{4}$$

for an arbitrary constant $\alpha \in [0, \infty)$. When $\alpha = 0$, $\mathbf{X}$ reduces to the adjacency matrix; when $\alpha = 1/2$, $\mathbf{X}$ becomes the Laplacian matrix (1) (up to a trivial transformation $\mathbf{I} - \mathbf{X}$); when $\alpha = 1$, we obtain random matrix (3). While we acknowledge that $\alpha \in \{0, 1/2, 1\}$ are the most common practical choices, allowing a general exponent α is not only for generality: it provides a *continuous degree-normalization strength* parameter that changes the leading fluctuation terms and hence impacts subsequent inference in a principled way. This will be made clear from our entrywise eigenvector expansion results to be presented in Theorem 3 later. Consequently, the asymptotic variance of eigenvector coordinates and downstream functionals in various statistical tasks depend explicitly on α. Indeed, as demonstrated in Figure 1, where a Laplacian-regularized graph smoothing is investigated via a simulation study, the mean squared error (MSE) attains its minimum when $\alpha = 0.75$, a value that differs from the common choices of 0, $1/2$, and 1. See Section A of the Supplementary Material for the detailed model setting for Figure 1 and additional simulation examples investigating the role of α. Both our theoretical results and simulation studies suggest an *intriguing* phenomenon: α should be regarded as a tuning parameter in downstream statistical analyses, and its optimal value depends on the specific statistical task and often differs from the common choices used in the existing literature.

To facilitate the theoretical analysis, we further assume that $\widetilde{\mathbf{X}}$ can be decomposed into a low-rank signal plus a random noise matrix

$$\widetilde{\mathbf{X}} = \mathbf{H} + \mathbf{W} \quad \text{with} \quad \mathbf{H} := \mathbb{E}[\widetilde{\mathbf{X}}], \tag{5}$$

where the signal matrix $\mathbf{H}$ has rank K and large signal eigenvalues in magnitude, and $\mathbf{W}$ is a random noise matrix with independent (up to symmetry) centered entries. In the Erdős–Rényi–Gilbert random graph with self-loops, one has $\mathbf{H} = p\mathbf{1}\mathbf{1}^T$ with $p \in (0, 1)$, which has rank $K = 1$, while $\mathbf{W}$ is a centered Bernoulli Wigner-type matrix*. For the DCMM model (2), this amounts to assuming that $\mathbf{P}$ in (2) is a $K \times K$ matrix. The decomposition (5) underlies the extensive literature on the spectral analysis of random graphs, where the leading eigenvalues of $\widetilde{\mathbf{X}}$ arise from the low-rank structure of $\mathbf{H}$, while the remaining eigenvalues form a bulk determined by $\mathbf{W}$; see, e.g., Bollobás (2001), van der Hofstad (2016), Oliveira (2009), and Benaych-Georges and Nadakuditi (2011). For many applications, it is also desired to consider the generalized Laplacian random matrices beyond networks, whose entries may take non-zero-one values. For example, one may take a signal-plus-noise model (5), where $\mathbf{W}$ is a Gaussian random matrix. Then we get a model that behaves similarly to dense networks. Our theory can be straightforwardly extended to such non-network settings. However, for clear exposition we focus on the network setting where $\widetilde{\mathbf{X}}$ is a adjacency matrix.

In the existing literature, it is the common practice to use only the leading eigenvalues and eigenvectors of a Laplacian-type operator; this is not just for computational convenience, but also for the reason that they define the embedding coordinates, smoothing directions, and spectral filters used to produce scientific conclusions from a single observed graph. Despite their ubiquity, rigorous distributional characterizations of Laplacian eigen-objects remain limited. Most existing theory is formulated in terms of consistency or global perturbation bounds, and does not provide the entrywise fluctuation results that are needed to quantify uncertainty in node-level outputs (e.g., predicted responses at given nodes, confidence intervals for spectral embeddings, or the calibration of hypothesis tests built from spectral features). This paper aims to fill this gap by leveraging some recent developments in random matrix theory (RMT). We establish asymptotic expansion results for leading eigenvalues and eigenvectors that are designed for statistical use: many downstream procedures in network inference and graph learning can be expressed as smooth functionals of a finite number of leading eigenpairs of a Laplacian-type matrix. Accordingly, the expansions and central limit theorems (CLTs) established in this paper immediately yield asymptotic distributions, and thus uncertainty quantification for node-level outputs of spectral procedures.

We illustrate this principle in Section V in two concrete settings: (i) a one-layer spectral graph neural network (GNN) filter evaluated on a pre-chosen fixed set of nodes, and (ii) Laplacian-regularized graph smoothing, where the estimator admits a decomposition into eigenvalue-filter and eigenvector-coordinate contributions, allowing confidence intervals and variance estimates to be constructed from our established limiting theory. We discuss therein that our theory can help quantify the uncertainties in both the spectral GNN and the Laplacian regularized graph smoothing induced by randomness in the graph; see Section V for more details. Our theory may also enhance the characterization of the community membership probability matrix $\mathbf{\Pi}$ through spectral clustering methods for community detection, a widely used and scalable tool in the literature, as demonstrated in Chaudhuri et al. (2012); Amini et al. (2013); Abbe (2017); Jin (2015); Le et al. (2016); Lei and Rinaldo (2015); Rohe et al. (2011), or may enable hypothesis testing with network data, a prevalent technique utilized in various contexts such as Arias-Castro and Verzelen (2014); Verzelen and Arias-Castro (2015); Bickel and Sarkar (2016); Lei (2016); Wang and Bickel (2017); Fan et al. (2022b); Fan et al. (2022). Due to the length constraint, we leave the investigation of various important applications of our theoretical results to future work.

Our theoretical results extend significantly the previous works Fan et al. (2022a); Fan et al. (2022) to the context of the generalized Laplacian matrix framework. These prior studies established the LLN and CLTs for the spiked sample eigen-pair of the adjacency matrices, which can be viewed as a special case of our results when $\alpha = 0$. Our results also compensate for the results of a recent work Ke and Wang (2025), where entrywise large-deviation bounds for the eigenvectors associated with the largest eigenvalues of the Laplacian matrix for the DCMM model were established through the leave-one-out strategy. Additionally, in Tang and Priebe (2018), the CLTs for the components of eigenvectors pertaining to the adjacency matrix and the Laplacian matrix of a random dot product graph were established, under the assumption of a prior distribution on the mean adjacency matrix.

The existing results in RMT have focused mainly on the setting of independent entries modulo symmetry. However, due to the use of the normalization terms, the generalized Laplacian matrix is an example of a random matrix with dependency, and thus existing RMT theories cannot be directly applied. Compared to previous works

*In the classical definition of Erdős–Rényi–Gilbert random graph, there are no self-loops and $\mathbf{H} = p(\mathbf{1}\mathbf{1}^T - \mathbf{I})$. Our theory can still accommodate this model by setting the diagonal entries of $\mathbf{W}$ to zero. The same treatment applies to other network models without self-loops.

Fan et al. (2022a); Fan et al. (2022); Bhattacharya et al. (2023), our paper introduces important theoretical and technical *innovations*. First, to the best of our knowledge, this paper is the *first* in the literature to establish the limiting distributions of the empirical eigen-pair of the generalized Laplacian matrix (4), featuring general distributions for the entries of $\mathbf{W}$ and severe sparsity. In particular, it is worth noting that our theoretical framework accommodates sparse networks with an average node degree exceeding $(\log n)^8/n$, encompassing both dense and sparse scenarios outlined in Arias-Castro and Verzelen (2014) and Verzelen and Arias-Castro (2015), respectively. Consequently, our setting may also apply to matrix completion problems involving sparse random matrices with missing values that are beyond network models.

Second, similar to Fan et al. (2022a); Fan et al. (2022), our analysis relies on delicate and precise estimates of the resolvent (or Green's function) of matrix $\mathbf{X}$, defined as $(\mathbf{X} - z\mathbf{I})^{-1}$ for $z \in \mathbb{C}$. These estimates are commonly referred to as the *entrywise and anisotropic local laws*. Building on such local laws, we derive asymptotic expansions for the spiked eigenvalues, as well as for individual components and general projections of the spiked eigenvectors of the generalized Laplacian matrix. The desired LLN and CLTs then follow as consequences of these new expansions. Related resolvent-based approaches for establishing local laws and self-consistent equations for resolvents have been extensively developed for random matrix ensembles in the mean-field or uncorrelated settings. In particular, our analysis draws on technical tools for sparse random matrices Erdős et al. (2013); Fan et al. (2022), the isotropic or anisotropic local law frameworks of Knowles and Yin (2017, 2013); Bloemendal et al. (2014), isotropic self-consistent equations He et al. (2018), and the matrix Dyson equation approach for correlated ensembles Ajanki et al. (2019). However, compared to these works, our setting presents *new* technical challenges in establishing sufficiently sharp entrywise and anisotropic local laws and in deriving the limiting distributions. One of the main difficulties lies in handling the strong dependence between the random degree matrix $\mathbf{D}$ and the adjacency matrix $\widetilde{\mathbf{X}}$. To address this issue, we adapt a decorrelation idea inspired by Ke and Wang (2025) within the resolvent and asymptotic expansion framework. Nevertheless, obtaining finer results—such as CLTs for individual eigenvector components—requires new decorrelation techniques *beyond* the standard leave-one-out strategy. More precisely, for each $i \in [n]$, we define a modified degree matrix $\mathbf{D}_{[i]}$ by recomputing the jth diagonal entry for all $j \neq i$ using the adjacency matrix with the ith row and column removed, while keeping the ith diagonal entry unchanged, i.e., $(\mathbf{D}_{[i]})_{ii} = d_i$; see equation (87) for the formal definition. This construction weakens effectively the correlation between $\mathbf{D}_{[i]}$ and $\widetilde{\mathbf{X}}$, allowing us to first establish *almost sharp* entrywise and anisotropic local laws for the resolvent of the intermediate matrix $\mathbf{D}_{[i]}^{-\alpha}\widetilde{\mathbf{X}}\mathbf{D}_{[i]}^{-\alpha}$, and subsequently transfer these results to obtain almost sharp local laws for the resolvent of $\mathbf{X}$ itself. To the best of our knowledge, this decorrelation strategy has *not* appeared in the existing random matrix theory literature and constitutes a novel technical contribution of the current work.

Third, we offer estimates for the asymptotic variances of the related statistics (i.e., empirical eigenvalues, and components and general projections of the empirical eigenvectors) by leveraging rank inference and bias correction techniques inspired by the methodologies presented in Fan et al. (2022a,b). These approaches enable a more applicable analysis of the asymptotic behaviors of the spiked eigenvalues and eigenvectors of the generalized Laplacian matrix. Our findings further contribute to a deeper understanding of the principal components of the generalized Laplacian matrix, unveiling insightful properties and characteristics. For instance, we confirm the intriguing phase transition phenomenon identified in Fan et al. (2022a) within the context of the generalized Laplacian matrix: the variances of the projections of spiked eigenvectors can exhibit distinct orders based on whether the direction of the projection operation aligns with the population spiked eigenvector or not.

The theoretical analysis in this paper relies on a collection of advanced probabilistic tools that have been developed recently in the RMT literature. For comprehensive overviews of modern developments in RMT, see, for example, Anderson et al. (2010); Erdős and Yau (2017); Tao (2012). The asymptotic behavior of spiked empirical eigenvalues and eigenvectors for Wigner and sample covariance matrices has been studied extensively. The spiked sample covariance model was first introduced in Johnstone (2001), and the celebrated Baik–Ben Arous–Péché (BBP) phase transition for spiked eigenvalues in the complex Wishart setting was identified in Baik et al. (2005); Baik and Silverstein (2006). First-order asymptotics for spiked eigenvalues, together with asymptotics for certain projections of the associated spiked eigenvectors, were established in Paul (2007); Nadler (2008); Bloemendal et al. (2016); Koltchinskii and Lounici (2016). Extensions to more general spiked separable covariance structures were subsequently developed in Ding and Yang (2021). The asymptotic distributions of spiked eigenvalues and projections of the corresponding eigenvectors were derived for the spiked Wishart model in Paul (2007), and later extended to broader classes of spiked sample covariance matrices under weak moment assumptions on the matrix entries in Wang and Fan (2017); Bao et al. (2021, 2022). For spiked Wigner-type random matrices, convergence results for spiked eigenvalues and appropriate eigenvector projections were studied in Benaych-Georges and Nadakuditi (2011). The asymptotic distributions of spiked eigenvalues were established in Pizzo et al. (2013); Renfrew and Soshnikov (2013); Knowles and Yin (2013, 2014), while the asymptotic distribution of certain projections of spiked eigenvectors was analyzed in Capitaine and Donati-Martin (2018). Entrywise asymptotic expansions for spiked eigenvectors were obtained in Abbe et al. (2020). More recently, Bhattacharya et al. (2023) derived asymptotic expansions for membership mixing probabilities in the DCMM model. We emphasize that the above list includes only a subset of the most closely related works and is by no means exhaustive. Importantly, *none* of the existing results address the spectral theory of (generalized) graph Laplacian matrices, which is the focus of the present paper. Our generalized Laplacian also exhibits a correlation-type normalization structure, which we further discuss and relate to classical correlation matrix models in Section IV.

The rest of the paper is organized as follows. Section II introduces the model setting. We suggest the new framework of the asymptotic theory of eigenvectors for latent embeddings with generalized Laplacian matrices (ATE-GL) and present the main results in Section III. Section IV details the technical innovations of our new theoretical work at a high level. We showcase two applications of our new asymptotic theory in Section V and provide several simulation examples verifying the theoretical results in Section VI. Section VII discusses some implications and extensions of our work. All the proofs and technical details are provided in the Supplementary Material.

II. Model setting

Consider an undirected graph $\mathcal{N} = (V, E)$, where $V = [n] := \{1, \cdots, n\}$ denotes a set of n nodes and E represents the set of all the network edges. The network edge set E is given by a symmetric random adjacency matrix $\widetilde{\mathbf{X}} = (\widetilde{X}_{ij}) \in \mathbb{R}^{n\times n}$ satisfying that $\widetilde{X}_{ij} = \widetilde{X}_{ji}$ with $1 \le i \ne j \le n$. In particular, the values of $\widetilde{X}_{ij} = 1$ or $\widetilde{X}_{ij} = 0$ correspond to the cases when network nodes i and j are connected or not connected, respectively. In many important network models, $\widetilde{\mathbf{X}}$ is assumed to be a Bernoulli random matrix with independent entries modulo the symmetry and heterogeneous variances. The mean matrix of $\widetilde{\mathbf{X}}$ encodes the interesting community structure of the underlying graph through the low-rank representation. The generalized Laplacian matrix $\mathbf{X}$, which is our primary investigation target, is defined in (4).

We now provide the rigorous definition of the random matrix model we will study.

Definition 1. *Consider an $n \times n$ symmetric random matrix $\widetilde{\mathbf{X}}$ that admits the signal-plus-noise decomposition in (5), where $\mathbf{H} = (H_{ij})_{1\le i,j\le n}$ is a deterministic symmetric signal matrix and $\mathbf{W} = (W_{ij})_{1\le i,j\le n}$ is a symmetric random noise matrix with centered and independent upper triangular entries. We assume that the signal matrix $\mathbf{H}$ has low rank $K \ge 1$. In particular, we allow K to diverge slowly, as specified later in (23). We define*

$$\theta := n^{-2} \sum_{1\le i,j\le n} \mathbb{E}|W_{ij}|^2, \tag{6}$$

which represents the overall "sparsity" level of the matrix $\widetilde{\mathbf{X}}$, and set

$$q := \sqrt{n\theta}. \tag{7}$$

Next, define the node degree matrix

$$\mathbf{D} := \operatorname{diag}(d_i : i \in [n]) \quad \textit{with} \quad d_i := \sum_{1\le j\le n} \widetilde{X}_{ij}, \tag{8}$$

and let $\bar{d} := n^{-1}\sum_{j=1}^n d_j$ be the average degree. We assume that $\mathbf{D}$ is positive definite almost surely.

For a fixed $\alpha \in (0, \infty)$, we define the generalized Laplacian matrix $\mathbf{X} \equiv \mathbf{X}(\alpha)$ as in (4). To streamline the technical presentation, we impose the following regularity conditions with a constant $C_0 > 0$.

(i) The sparsity parameter q defined in (7) satisfies

$$(\log n)^4 \ll q \le C_0 n^{1/2}. \tag{9}$$

(ii) The noise matrix $\mathbf{W}$ satisfies

$$\max_{i,j\in[n]} |W_{ij}| \le C_0 \tag{10}$$

almost surely, and

$$\mathbb{E}W_{ij} = 0, \quad s_{ij} := \mathbb{E}|W_{ij}|^2 \le C_0\theta, \quad \forall i,j \in [n]. \tag{11}$$

(iii) The signal matrix $\mathbf{H}$ *has nonnegative entries and satisfies*

$$\max_{i\in[n]} \theta_i \le C_0 \quad \textit{with} \quad \theta_i := \frac{1}{n\theta} \sum_{1\le j\le n} H_{ij}. \tag{12}$$

Moreover, we introduce a parameter that represents the minimum (expected) node degree, rescaled by the average node degree $n\theta$

$$\beta_n := \min_{i\in[n]} \theta_i. \tag{13}$$

Definition 1 above imposes only general conditions on the means and variances of the entries of $\widetilde{\mathbf{X}}$, together with the uniform upper bound (10). In particular, it accommodates settings in which the entries of $\widetilde{\mathbf{X}}$ are *not* necessarily binary. We also refer the reader to the *arXiv version* of this paper (Fan et al., 2025), which establishes theoretical results in an even more general framework. Therein, the uniform bound (10) is replaced with some suitable moment conditions, and certain regularizations of the node degree matrix $\mathbf{D}$ are allowed.

Despite this level of generality, our primary motivation remains the case in which $\widetilde{\mathbf{X}}$ represents the adjacency matrix of an undirected random graph. In this setting, the entries of $\widetilde{\mathbf{X}}$ are independent (up to symmetry) Bernoulli random variables, and $\widetilde{\mathbf{X}}$ can be written in form (5). If, in addition, the network is sparse—so that an edge between each pair of nodes appears with probability of asymptotic order $\theta \le 1$—then the means and variances of the entries of $\widetilde{\mathbf{X}}$ are typically of order θ. This observation justifies the assumptions in (11) and (12). In addition, the lower bound on q in (9) imposes a sparsity constraint $\theta \gg (\log n)^8/n$ on the network. The parameters θ_i introduced in Condition (iii) quantify the degree of heterogeneity among nodes in network models. Assumption (12) essentially requires that $q^2 = n\theta$ be of the same order as the maximum node degree in the network. Consequently, our results are best suited to networks in which a non-negligible fraction of nodes have large degrees. For networks with only a few high-degree nodes, one may instead rescale the adjacency matrix using a different choice of q, and all results in this paper can be developed in parallel for that setting as well. For definiteness and notational simplicity, we choose to work under assumption (12).

To make the discussion more concrete, we next examine the degree-corrected mixed membership (DCMM) model (Jin et al., 2024) as a representative example.

Example 1 (DCMM model). Assume that random graph $\mathcal{N}$ has some underlying network structure in that there exist K disjoint subsets $C_1, \cdots, C_K$ called the *latent communities* of the network, and each network node $i \in [n]$ has an associated K-dimensional community membership probability vector $\boldsymbol{\pi}_i := (\pi_i(1), \cdots, \pi_i(K))^T$ with

$$\mathbb{P}[i \in C_k] = \pi_i(k) \tag{14}$$

for each $1 \le k \le K$, meaning that each node has mixed membership among the K latent communities. Denote by $\mathbf{\Pi} := (\boldsymbol{\pi}_1, \cdots, \boldsymbol{\pi}_n)^T$ the $n \times K$ matrix of community membership probability vectors. Further, assume that the connection probability of any two nodes $i \ne j \in [n]$ is given by

$$H_{ij} = \mathbb{P}[\widetilde{X}_{ij} = 1] = \vartheta_i \vartheta_j \sum_{k,\, l\in[K]} \pi_i(k)\pi_j(l)p_{kl}, \tag{15}$$

where parameter $\vartheta_i > 0$ represents the degree heterogeneity of each node i, and parameter p_{kl} can be understood as the probability that two nodes in communities C_k and C_l connect to each other with $1 \le k, l \le K$. Rewriting (15) in the matrix form, we have the representation (2), where $\mathbf{H} = (H_{ij})_{1\le i,j\le n}$, $\mathbf{\Theta} := \operatorname{diag}(\vartheta_1, \cdots, \vartheta_n)$, and $\mathbf{P} = (p_{kl}) \in \mathbb{R}^{K\times K}$. The adjacency matrix $\widetilde{\mathbf{X}}$ for the DCMM model then takes form (5) with $\mathbf{H}$ in (2). In this model, the network sparsity parameter θ in (6) is given by

$$\theta := n^{-2} \sum_{i,j\in[n]} \vartheta_i \vartheta_j \sum_{k,\, l\in[K]} \pi_i(k)\pi_j(l)p_{kl}.$$

In particular, when $\sum_{k,l\in[K]} \pi_i(k)\pi_j(l)p_{kl}$ are all of order 1, we have that $\theta \sim (n^{-1}\sum_{i\in[n]} \vartheta_i)^2$.

By the classical law of large numbers (LLN) and central limit theorem (CLT), the entries of the node degree matrix $\mathbf{D}$ given in (8) would concentrate around the deterministic diagonal matrix

$$\mathbf{\Lambda} = \operatorname{diag}(\Lambda_1, \cdots, \Lambda_n) := \mathbb{E}[\mathbf{D}]. \tag{16}$$

In view of (5), (4), and (16), we will consider the spectral decompositions of both $\mathbf{X}$ and its population counterpart

$\mathbf{\Lambda}^{-\alpha}\mathbf{H}\mathbf{\Lambda}^{-\alpha}$ given by

$$\mathbf{X} = \sum_{i\in[n]} \widehat{\delta}_i \widehat{\mathbf{v}}_i \widehat{\mathbf{v}}_i^T \quad \text{and} \quad \mathbf{\Lambda}^{-\alpha}\mathbf{H}\mathbf{\Lambda}^{-\alpha} = \sum_{i\in[K]} \delta_i \mathbf{v}_i \mathbf{v}_i^T, \tag{17}$$

where we arrange the eigenvalues according to the descending order in magnitude with $|\widehat{\delta}_1| \geq \cdots \geq |\widehat{\delta}_n|$ and $|\delta_1| \geq \cdots \geq |\delta_K| > 0$, and $\widehat{\mathbf{v}}_i$'s and $\mathbf{v}_i$'s are the corresponding eigenvectors. Given the empirical and population eigen-decompositions in (17) above, let us define the diagonal matrices of top K eigenvalues

$$\widehat{\mathbf{\Delta}} := \operatorname{diag}(\widehat{\delta}_1, \cdots, \widehat{\delta}_K) \quad \text{and} \quad \mathbf{\Delta} := \operatorname{diag}(\delta_1, \cdots, \delta_K), \tag{18}$$

as well as the corresponding eigenvector matrices

$$\widehat{\mathbf{V}} = (\widehat{\mathbf{v}}_1, \cdots, \widehat{\mathbf{v}}_K) \quad \text{and} \quad \mathbf{V} = (\mathbf{v}_1, \cdots, \mathbf{v}_K). \tag{19}$$

The major goal of this work is to study the asymptotic behavior of $\widehat{\delta}_k$ and $\widehat{\mathbf{v}}_k$ with $1 \leq k \leq K$ and in particular, identify their dependence on the population eigenvalues and eigenvectors δ_k's and $\mathbf{v}_k$'s. We now formally define spiked eigenvalues and eigenvectors.

Definition 2 (Spiked eigenvalues and eigenvectors)**.** *We call δ_k a* spiked eigenvalue *of the population generalized Laplacian $\mathbf{\Lambda}^{-\alpha}\mathbf{H}\mathbf{\Lambda}^{-\alpha}$ if it satisfies*

$$|\delta_k| \gg q^{1-4\alpha}\beta_n^{-2\alpha}, \tag{20}$$

where q and β_n are defined in (7) and (13), respectively. The corresponding empirical eigenvalue $\widehat{\delta}_k$ is referred to as an empirical spiked eigenvalue of $\mathbf{X}$, and the associated population and empirical eigenvectors are called the spiked eigenvectors and empirical spiked eigenvectors, respectively.

As will be shown (cf. Lemma C.1 and Proposition C.1), we have $\|\mathbf{W}\| \lesssim q$ with high probability, and the smallest eigenvalue of $\mathbf{D}$ is of order $\gtrsim \beta_n q^2$. These facts imply that the operator norm of the noise matrix satisfies $\|\mathbf{\Lambda}^{-\alpha}\mathbf{W}\mathbf{\Lambda}^{-\alpha}\| = O(q^{1-4\alpha}\beta_n^{-2\alpha})$ with high probability. Hence, condition (20) ensures that δ_k is a true spike relative to the noise eigenvalues.

The terminology *spiked* used in Definition 2 above is standard in RMT and refers to eigenvalues separated from the bulk spectrum generated by random fluctuations. Such phenomena were first identified and systematically studied in the context of spiked covariance and Wigner-type models, where low-rank perturbations give rise to outlier eigenvalues; see, e.g., Johnstone (2001) and Baik et al. (2005), among others. Closely related results have also been developed for adjacency matrices of random graphs and inhomogeneous random graphs, where the expectation of the adjacency matrix is typically low-rank and the corresponding eigenvalues are outliers; see, for example, Oliveira (2009), Benaych-Georges and Nadakuditi (2011), Rohe et al. (2011), and Lei and Rinaldo (2015). See also the review article Paul and Aue (2014) and the references therein. Definition 2 generalizes this notion to the setting of generalized Laplacian matrices considered here and provides a unified framework encompassing both classical spiked random matrix models and graph-based models.

To facilitate the technical presentation, let us introduce some necessary notation. We focus on the asymptotic regime of network size $n \to \infty$ and refer to a constant whenever it does not depend on parameter n. We will use C to denote a generic large positive constant whose value may change from line to line. Similarly, we will use notation such as ϵ, c, and δ to represent generic small positive constants. For any two sequences a_n and b_n, $a_n = O(b_n)$ (or $b_n = \Omega(a_n)$) means that $|a_n| \leq C|b_n|$ for some constant $C > 0$, whereas $a_n = o(b_n)$ or $|a_n| \ll |b_n|$ means that $|a_n|/|b_n| \to 0$ as $n \to \infty$. We say that $a_n \lesssim b_n$ if $a_n = O(b_n)$ and that $a_n \sim b_n$ if $a_n = O(b_n)$ and $b_n = O(a_n)$. Given a vector $\mathbf{v} = (v_i)_{i=1}^n$, $|\mathbf{v}| \equiv \|\mathbf{v}\| \equiv \|\mathbf{v}\|_2$ denotes the Euclidean norm and $\|\mathbf{v}\|_p$ denotes the L_p-norm. Given a matrix $\mathbf{A} = (A_{ij})$, we denote by $\|\mathbf{A}\|$, $\|\mathbf{A}\|_F$, and $\|\mathbf{A}\|_{\max} := \max_{i,j} |A_{ij}|$ the matrix operator norm, Frobenius norm, and entrywise maximum norm, respectively. We use A^T to denote the transpose of A, and A^* to denote its conjugate transpose. For notational simplicity, we write $\mathbf{A} = O(a_n)$ and $\mathbf{A} = o(a_n)$ to mean that $\|\mathbf{A}\| = O(a_n)$ and $\|\mathbf{A}\| = o(a_n)$, respectively. Moreover, we will use A_{ij} and $\mathbf{A}(k)$ to denote the (i,j)th entry and kth *row vector* of a given matrix $\mathbf{A}$, respectively, and use $v(k)$ to denote the kth component of a given vector $\mathbf{v}$. We will often write an identity matrix of appropriate size as $\mathbf{I}$ without specifying the size in the subscript. Denote by $\mathbf{e}_i$ the unit vector with the ith component being 1 and others being 0. Given any $n \times n$ matrix $\mathbf{A}$ and vectors $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$, we define

$$A_{i\mathbf{v}} := \mathbf{e}_i^T \mathbf{A}\mathbf{v}, \;\; A_{\mathbf{u}i} := \mathbf{u}^T \mathbf{A}\mathbf{e}_i, \;\; A_{\mathbf{u}\mathbf{v}} := \mathbf{u}^T \mathbf{A}\mathbf{v}. \tag{21}$$

Throughout the paper, we will use the notion of "with high probability" as defined below.

Definition 3. *We say that an event Ω holds with high probability (w.h.p.) if for any large constant $D > 0$, $\mathbb{P}(\Omega^c) \leq n^{-D}$ for all large enough n.*

III. ATE-GL FOR LATENT EMBEDDINGS

We now formally introduce the asymptotic framework of eigenvectors for latent embeddings with generalized Laplacian matrices (ATE-GL).

A. Technical conditions and preparation

We impose the following regularity conditions in addition to the basic assumptions in Definitions 1 and 2.

Assumption 1. *Consider the model in Definition 1. Fix $\alpha \in (0, \infty)$ and assume that the following conditions hold for some $1 \leq K_0 \leq K$.*

(i) (Spiked eigenvalues) $\delta_1, \cdots, \delta_{K_0}$ are spiked population eigenvalues in the sense of Definition 2.

(ii) (Eigengap) There exists a constant $\epsilon_0 > 0$ such that

$$\min_{1 \leq k \leq K_0} \frac{|\delta_k|}{|\delta_{k+1}|} > 1 + \epsilon_0, \tag{22}$$

where no eigengap condition is required for the remaining eigenvalues $|\delta_k|$ with $K_0 + 1 \leq k \leq K$.

(iii) (Low-rank signals) The rank K of $\mathbf{H}$ satisfies

$$K \log n \left(\frac{q^{1-4\alpha}}{|\delta_{K_0}| \beta_n^{1+2\alpha}} + \frac{\log n}{q \beta_n^2} + \|\mathbf{V}\|_{\max} \right) \ll q, \tag{23}$$

where $\mathbf{V}$ is given in (19).

Condition (iii) imposes only a mild constraint on the rank K. For example, if $\beta_n \gtrsim 1$, $|\delta_{K_0}| \gtrsim q^{2-4\alpha}$ (which occurs when the K_0th eigenvalue of $\mathbf{H}$ is at least of order $\Omega(n\theta)$), and $|\mathbf{V}|_{\max} \leq q^{-1}$, then (23) requires $K \ll q^2/(\log n)^2 = n\theta/(\log n)^2$. For network models, this means that the number of latent communities K is assumed to be "slightly" below the typical node degree $n\theta$. Next, we comment on the eigengap requirement in Condition (ii) of Assumption 1.

Remark 1. Assumption 1 (iii) requires the spiked eigenvalues of $\mathbf{\Lambda}^{-\alpha}\mathbf{H}\mathbf{\Lambda}^{-\alpha}$ to be well separated. This excludes highly symmetric configurations (for example, planted partition models with equal block sizes) or specially tuned parameter choices for which the spiked eigenvalues have multiplicity greater than one and individual eigenvectors are therefore not identifiable. This assumption serves as an identifiability condition under which the eigenvector-level inference is meaningful. Our main results are formulated for individual empirical eigenvalue-eigenvector pairs $(\widehat{\delta}_k, \widehat{\mathbf{v}}_k)$, and we will also establish entrywise asymptotic expansions and CLTs for the coordinates of $\widehat{\mathbf{v}}_k$. If a population spiked eigenvalue δ_k has multiplicity greater than one, then only the associated eigenspace of $\mathbf{\Lambda}^{-\alpha}\mathbf{H}\mathbf{\Lambda}^{-\alpha}$ is identifiable, while the individual eigenvectors are defined up to arbitrary orthogonal rotations within that subspace. In such a degenerate case, coordinate-wise limits for $\mathbf{v}_k(i)$ and $\widehat{\mathbf{v}}_k(i)$ depend on the choice of basis and are therefore not intrinsically meaningful. In many network models of interest (such as the DCMM model) and in most practical applications, the spiked eigenvalues of $\mathbf{\Lambda}^{-\alpha}\mathbf{H}\mathbf{\Lambda}^{-\alpha}$ are expected to be *simple* under generic, nonsymmetric parameter configurations. Exact multiplicities typically arise only under highly symmetric parameter choices. Moreover, even when exact multiplicities occur, they may be removed by a small perturbation of the parameters or by introducing a mild regularization term. In this sense, the non-degeneracy assumption is standard in statistical network analysis and is generally not regarded as overly restrictive.

Technically, the separation assumption guarantees the stability of the resolvent-based expansions used in our proofs and prevents denominators of form $(z - \delta_k)^{-1}$ and $(\delta_k - \delta_\ell)^{-1}$ from diverging. If some spiked eigenvalues coincide or become arbitrarily close, one would need to replace the eigenvector-level analysis by a matrix-valued perturbation theory for the corresponding invariant subspaces, resulting in a *substantially more* delicate and involved proof. Although such an extension is, in principle, feasible, it lies beyond the scope of the present paper and is therefore deferred to future work.

We now provide the necessary technical preliminaries related to the *generalized quadratic vector equation* (GQVE), a key analytical tool used to characterize the asymptotic limit t_k of the empirical spiked eigenvalue $\widehat{\delta}_k$. Let $\widetilde{\mathbf{M}} \equiv \widetilde{\mathbf{M}}_n(z) = (\widetilde{M}_1(z), \cdots, \widetilde{M}_n(z))^T \in \mathbb{C}^n$ be the z-dependent solution to the generalized QVE

$$\frac{1}{\widetilde{M}_i} = -z - \sum_{j \in [n]} \Lambda_i^{-2\alpha} s_{ij} \Lambda_j^{-2\alpha} \widetilde{M}_j \tag{24}$$

subject to the condition $\operatorname{Im} \widetilde{M}_i(z) \geq 0$ for all $i \in [n]$ and $z \in \mathbb{C}_+$, where $\mathbb{C}_+$ denotes the upper half complex plane, and s_{ij} and Λ_i are defined in (11) and (16), respectively. Next, define the associated deterministic diagonal matrix

$$\widetilde{\boldsymbol{\Upsilon}}(z) := \operatorname{diag}(\widetilde{M}_1(z), \cdots, \widetilde{M}_n(z)). \tag{25}$$

For each $1 \leq k \leq K$, let $\mathbf{V}_{-k}$ be an $n \times (K-1)$ matrix obtained by removing the kth column of $\mathbf{V}$, and $\boldsymbol{\Delta}_{-k}$ a $(K-1) \times (K-1)$ matrix obtained by removing the kth row and column of $\boldsymbol{\Delta}$, where $\mathbf{V}$ and $\boldsymbol{\Delta}$ are given in (19) and (18), respectively. For $1 \leq k \leq K_0$, we introduce a subset

$$\widetilde{\mathcal{I}}_k := \left\{ x \in \mathbb{R} : \frac{|\delta_k|}{1+\epsilon_0/2} \leq |x| \leq (1+\epsilon_0/2)|\delta_k| \right\} \subset \mathbb{R}, \tag{26}$$

where ϵ_0 is the eigengap constant from (22). We then define t_k as the real solution to the nonlinear equation

$$1 + \delta_k \mathbf{v}_k^T \widetilde{\boldsymbol{\Upsilon}}(x) \mathbf{v}_k - \delta_k \mathbf{v}_k^T \widetilde{\boldsymbol{\Upsilon}}(x) \mathbf{V}_{-k} \frac{1}{\boldsymbol{\Delta}_{-k}^{-1} + \mathbf{V}_{-k}^T \widetilde{\boldsymbol{\Upsilon}}(x) \mathbf{V}_{-k}} \mathbf{V}_{-k}^T \widetilde{\boldsymbol{\Upsilon}}(x) \mathbf{v}_k = 0 \tag{27}$$

over $x \in \widetilde{\mathcal{I}}_k$. Using analogous arguments to those in Lemma 3 of Fan et al. (2022a) and Section A.2 of Fan et al. (2022), we obtain the following lemma, which establishes the existence, uniqueness, and leading-order behavior of t_k.

Lemma 1. *Under parts (ii) and (iii) of Assumption 1, for each $1 \leq k \leq K_0$, equation (27) admits a unique solution $x = t_k$ in the subset $\widetilde{\mathcal{I}}_k$, and moreover, it holds that*

$$t_k = \delta_k + O\left(q^{2-8\alpha} \beta_n^{-4\alpha} / |\delta_k|\right). \tag{28}$$

B. Main results

Our first main result shows that the population quantity t_k introduced above serves as the asymptotic limit of the empirical spiked eigenvalue $\widehat{\delta}_k$. For clarity, we introduce the rescaled population eigenvalue

$$\mathsf{d}_k := \delta_k / (q^{1-4\alpha} \beta_n^{-2\alpha}), \quad 1 \leq k \leq K. \tag{29}$$

Note that under Definition 2, the rescaled spiked population eigenvalues satisfy $|\mathsf{d}_k| \gg 1$ for $1 \leq k \leq K_0$. We also define the following error-control parameter

$$\widetilde{\psi}_n(\delta_k) := \frac{1}{|\mathsf{d}_k| \beta_n} + \frac{\sqrt{\log n}}{q \beta_n^2} + \|\mathbf{V}\|_{\max}. \tag{30}$$

Theorem 1. *Under Definition 1 and Assumption 1, it holds w.h.p. that*

$$\frac{|\widehat{\delta}_k - t_k|}{|\delta_k|} = O\left\{ \frac{\sqrt{\log n}\, \widetilde{\psi}_n(\delta_k)}{q} \left(1 + \frac{K}{|\mathsf{d}_k|^4}\right) \right\}, \quad \forall 1 \leq k \leq K_0. \tag{31}$$

Theorem 1 above establishes that the population quantity t_k provides the first-order asymptotic limit of the empirical spiked eigenvalue $\widehat{\delta}_k$. Moreover, recall from (28) that t_k is itself asymptotically close to the population eigenvalue δ_k. Note that under condition (23), the error bound in (31) is of strictly smaller order than that in (28) when $q \gg \sqrt{\log n}\, \widetilde{\psi}_n(\delta_k) |\mathsf{d}_k|^2$. Combining Theorem 1 with (28) yields that

$$|\widehat{\delta}_k - \delta_k| = o(|\delta_k|), \quad \forall 1 \leq k \leq K_0. \tag{32}$$

In view of (26) and (32), $\widehat{\delta}_k$ also lies in the subset $\widetilde{\mathcal{I}}_k$ asymptotically.

Based on (32) and eigengap condition (22), we may define a closed contour $\mathcal{C}_k$ in the complex plane $\mathbb{C}$ such that w.h.p., $\mathcal{C}_k$ encloses only the empirical spiked eigenvalue $\widehat{\delta}_k$ and no other eigenvalues of the generalized Laplacian matrix $\mathbf{X}$. This property allows us to extract information about the empirical spiked eigenvector $\widehat{\mathbf{v}}_k$ by evaluating contour integrals of Green's function (resolvent) $(\mathbf{X} - z\mathbf{I})^{-1}$ of the random matrix $\mathbf{X}$ for $z \in \mathbb{C}$, and applying Cauchy's integral formula. Consequently, our analysis begins with deriving delicate asymptotic expansions for the empirical spiked eigenvectors $\widehat{\mathbf{v}}_k$. These expansions enable us to characterize the asymptotic behavior of the projection of $\widehat{\mathbf{v}}_k$ onto any deterministic unit vector $\mathbf{u} \in \mathbb{R}^n$.

To state our second main result, let us define

$$\widetilde{\boldsymbol{\Upsilon}}_k(z) := \widetilde{\boldsymbol{\Upsilon}}(z) - \widetilde{\boldsymbol{\Upsilon}}(z) \mathbf{V}_{-k} \frac{1}{\boldsymbol{\Delta}_{-k}^{-1} + \mathbf{V}_{-k}^T \widetilde{\boldsymbol{\Upsilon}}(z) \mathbf{V}_{-k}} \mathbf{V}_{-k}^T \widetilde{\boldsymbol{\Upsilon}}(z), \tag{33}$$

where the $n \times n$ deterministic matrix $\widetilde{\mathbf{\Upsilon}}(z)$ is given in (25).

Theorem 2. *Under Definition 1 and Assumption 1, for each $1 \leq k \leq K_0$, it holds w.h.p. that*

$$\left| \mathbf{v}_k^T (\mathbf{D}/\mathbf{\Lambda})^{-\alpha} \widehat{\mathbf{v}}_k - \frac{1}{\sqrt{\delta_k^2 \mathbf{v}_k^T \widetilde{\mathbf{\Upsilon}}_k'(t_k) \mathbf{v}_k}} \right| \lesssim \frac{\sqrt{\log n}\widetilde{\psi}_n(\delta_k)}{q} \left(1 + \frac{K}{|\mathsf{d}_k|^4}\right), \tag{34}$$

where the sign of $\widehat{\mathbf{v}}_k$ is chosen such that $\widehat{\mathbf{v}}_k^T \mathbf{v}_k > 0$. Moreover, for any deterministic unit vector $\mathbf{u} \in \mathbb{R}^n$, it holds w.h.p. that

$$\left| \mathbf{u}^T (\mathbf{D}/\mathbf{\Lambda})^{-\alpha} \widehat{\mathbf{v}}_k + \frac{\delta_k \mathbf{u}^T \widetilde{\mathbf{\Upsilon}}_k(t_k) \mathbf{v}_k}{\sqrt{\delta_k^2 \mathbf{v}_k^T \widetilde{\mathbf{\Upsilon}}_k'(t_k) \mathbf{v}_k}} \right| \lesssim \frac{\sqrt{\log n}\widetilde{\psi}_n(\delta_k)}{q} \left[1 + \frac{K}{|\mathsf{d}_k|^4} + \|\mathbf{u}^T \mathbf{V}_{-k}\| \left(\sqrt{K} + \frac{K}{|\mathsf{d}_k|^2}\right)\right]. \tag{35}$$

The second terms on the left-hand side (LHS) of (34) and (35) admit the following asymptotic expansions

$$\delta_k^2 \mathbf{v}_k^T \widetilde{\mathbf{\Upsilon}}_k'(t_k) \mathbf{v}_k = 1 + O\left(|\mathsf{d}_k|^{-2}\right), \quad \delta_k \mathbf{u}^T \widetilde{\mathbf{\Upsilon}}_k(t_k) \mathbf{v}_k = -\mathbf{u}^T \mathbf{v}_k + O\left(|\mathsf{d}_k|^{-2}\right). \tag{36}$$

Theorem 2 above provides the first-order asymptotic limits for linear projections of the empirical spiked eigenvector $\widehat{\mathbf{v}}_k$ under different weight vectors. Since the node degrees d_i concentrate around their expectations, the random diagonal matrix $\mathbf{D}/\mathbf{\Lambda}$ is approximately the identity matrix plus a small perturbation (recall (8) and (16)). Combining Lemma C.1 in Section C-A of the Supplementary Material with the low-rank structure in (23), we may directly apply (35) and (36) to obtain the corresponding estimate for $\mathbf{u}^T \widehat{\mathbf{v}}_k$:

$$\mathbf{u}^T \widehat{\mathbf{v}}_k = \mathbf{u}^T \mathbf{v}_k + O\left(\frac{\sqrt{\log n}}{q\beta_n} + \frac{1}{|\mathsf{d}_k|^2} + \frac{\sqrt{K \log n}\widetilde{\psi}_n(\delta_k)}{q} \|\mathbf{u}^T \mathbf{V}_{-k}\|\right). \tag{37}$$

As a special but significant case, setting $\mathbf{u} = \mathbf{e}_i$ in (35) already yields the first-order asymptotics of each component $\widehat{v}_k(i)$ of the empirical spiked eigenvector. The following theorem strengthens this observation by providing a *much sharper* entrywise expansion of $\widehat{v}_k(i)$, which will allow us to establish the CLT as $n \to \infty$.

Theorem 3. *Assume that Definition 1 and Assumption 1 hold, and*

$$K\widetilde{\psi}_n(\delta_k)\beta_n \lesssim 1, \quad \|\mathbf{V}\|_{\max} \ll \frac{1}{|\mathsf{d}_k|\beta_n} + \frac{\sqrt{\log n}}{q\beta_n^2} \tag{38}$$

for all $1 \leq k \leq K_0$. Then for each $i \in [n]$, it holds w.h.p. that

$$\begin{aligned} \widehat{v}_k(i) &= (\Lambda_i/d_i)^\alpha v_k(i) + \frac{1}{t_k d_i^\alpha} \sum_{j \in [n]} W_{ij} \Lambda_j^{-\alpha} v_k(j) \\ &+ O\left(\|\mathbf{V}\|_{\max} \left(\frac{\sqrt{K}}{|\mathsf{d}_k|^2} + \frac{K\sqrt{\log n}}{q} \left(\frac{1}{|\mathsf{d}_k|\beta_n} + \frac{\sqrt{\log n}}{q\beta_n^2}\right)\right) + \frac{\sqrt{\log n}}{\sqrt{n}|\mathsf{d}_k|} \left(\frac{1}{|\mathsf{d}_k|} + \frac{\sqrt{\log n}}{q\beta_n}\right)\right), \end{aligned} \tag{39}$$

where we choose the sign of $\widehat{\mathbf{v}}_k$ such that $\widehat{\mathbf{v}}_k^T \mathbf{v}_k > 0$. Consequently, we have that w.h.p.,

$$\begin{aligned} \widehat{v}_k(i) &= v_k(i) - \frac{\alpha v_k(i)}{\Lambda_i} \sum_{j \in [n]} W_{ij} + \frac{1}{t_k} \sum_{j \in [n]} \Lambda_i^{-\alpha} W_{ij} \Lambda_j^{-\alpha} v_k(j) \\ &+ O\left(\|\mathbf{V}\|_{\max} \left(\frac{\sqrt{K}}{|\mathsf{d}_k|^2} + \frac{K\sqrt{\log n}}{q} \left(\frac{1}{|\mathsf{d}_k|\beta_n} + \frac{\sqrt{\log n}}{q\beta_n^2}\right)\right) + \frac{\sqrt{\log n}}{\sqrt{n}|\mathsf{d}_k|} \left(\frac{1}{|\mathsf{d}_k|} + \frac{\sqrt{\log n}}{q\beta_n}\right)\right). \end{aligned} \tag{40}$$

Remark 2. By Definition 3, the componentwise asymptotic expansions for the empirical spiked eigenvectors $\widehat{\mathbf{v}}_k$'s established in (39) and (40) hold with very high probability $1 - O(n^{-D})$ for any large constant $D > 0$. Applying the union bound then yields that these expansions hold simultaneously for all $1 \leq k \leq K_0$ and $i \in [n]$. If one is content with the weaker probability $1 - o(1)$, the error terms can in fact be sharpened by removing certain $\sqrt{\log n}$ factors. We do not pursue this refinement here, since in many applications a uniform bound over all k and i is essential. For this reason, we retain the $\sqrt{\log n}$ factors to ensure that our results remain broadly applicable.

Remark 3. The additional assumption (38) in Theorem 3 is imposed solely to simplify the error terms and make their orders more transparent. Under this condition, the leading-order behavior becomes more readable, allowing us to present a cleaner expression for the remainder. In practice, particularly in network applications, assumption (38) is not restrictive. It is common in these settings to assume that the number of communities K is fixed or grows slowly, and that the spiked eigenvectors are sufficiently delocalized, e.g.,

$$\|\mathbf{V}\|_{\max}^2 \lesssim K/n. \tag{41}$$

For interested readers, we derive in Proposition C.2 (Section C-D of the Supplementary Material) the asymptotic expansion of $d_i^\alpha \widehat{v}_k(i)$ without invoking (38). The primary difference is that, in the absence of this simplifying assumption, the resulting error term is slightly more involved.

Remark 4. With the asymptotic expansions developed in Theorem 3, one can construct both one-sample and two-sample test statistics for the DCMM models introduced in Example 1. A representative application is statistical inference for determining whether two or more nodes in a DCMM network share the same community membership probability vector, or whether two networks possess the same community membership matrix. For the one-sample problem, one may adopt the SIMPLE method proposed in Fan et al. (2022b) or the more general SIMPLE-RC procedure introduced in Fan et al. (2022). More precisely, these works construct test statistics from ratios of empirical eigenvector entries, such as $\widehat{v}_k(i)/\widehat{v}_1(i)$, associated with the adjacency matrix (equivalently, the generalized Laplacian matrix with $\alpha = 0$). The same type of statistics can be naturally extended to our setting by replacing the adjacency matrix eigenvectors with those of the generalized Laplacian matrix corresponding to a potentially nonzero parameter α. The statistical behavior of these statistics can then be analyzed using the asymptotic expansions (39) and (40), following arguments similar to those in Fan et al. (2022b) and Fan et al. (2022). In fact, our framework may offer additional flexibility and potential advantages over the adjacency-matrix-based approaches in Fan et al. (2022b) and Fan et al. (2022), since the presence of the tuning parameter α may allow one to improve the statistical power of the resulting procedures. We also expect that the eigenvector-ratio statistics proposed in Fan et al. (2022b) and Fan et al. (2022) can be further extended to the two-sample setting. Owing to space limitations, however, we leave a systematic investigation of these interesting directions for future work.

It is natural to expect CLT from the asymptotic expansions in Theorem 3. Indeed, (40) shows that the remainder term is (up to factors involving K, $\sqrt{\log n}$, and β_n) of order $(|\mathsf{d}_k|^{-2} + q^{-2})\|\mathbf{V}\|_{\max}$. On the other hand, the leading fluctuation term converges in law to a normal distribution with variance given by

$$\begin{aligned}\sigma_{k,i}^2 &:= \operatorname{var}\Big\{ -\frac{\alpha v_k(i)}{\Lambda_i} \sum_{j\in[n]} W_{ij} + \frac{1}{t_k} \sum_{j\in[n]} \Lambda_i^{-\alpha} W_{ij} \Lambda_j^{-\alpha} v_k(j) \Big\} \\ &= \sum_{j\in[n]} s_{ij} \left(-\frac{\alpha v_k(i)}{\Lambda_i} + \frac{1}{t_k} \Lambda_i^{-\alpha} \Lambda_j^{-\alpha} v_k(j) \right)^2, \end{aligned} \tag{42}$$

which is typically of order $q^{-2}|v_k(i)|^2 + \beta_n^{4\alpha}/(n|\mathsf{d}_k|^2)$. Consequently, whenever the remainder in (40) is asymptotically negligible compared to $\sigma_{k,i}$, a CLT for $\widehat{v}_k(i)$ follows, as stated in the corollary below. A corresponding CLT for $d_i^\alpha \widehat{v}_k(i)$ can also be obtained directly from (39), although we omit the details for the sake of brevity.

Corollary 1. *Under all the conditions of Theorem 3, fix any $1 \le k \le K_0$ and $i \in [n]$. If $\|\mathbf{v}_k\|_\infty \to 0$ and*

$$\|\mathbf{V}\|_{\max} \left(\frac{\sqrt{K}}{|\mathsf{d}_k|^2} + \frac{K\sqrt{\log n}}{q} \left(\frac{1}{|\mathsf{d}_k|\beta_n} + \frac{\sqrt{\log n}}{q\beta_n^2} \right) \right) + \frac{\sqrt{\log n}}{\sqrt{n}|\mathsf{d}_k|} \left(\frac{1}{|\mathsf{d}_k|} + \frac{\sqrt{\log n}}{q\beta_n} \right) \ll \sigma_{k,i}, \tag{43}$$

then we have that $(\widehat{v}_k(i) - v_k(i))/\sigma_{k,i} \xrightarrow{d} \mathcal{N}(0,1)$ as $n \to \infty$. In particular, (43) is satisfied provided that

$$\sigma_{k,i} \gtrsim \frac{\beta_n^{2\alpha}}{\sqrt{n}|\mathsf{d}_k|}, \qquad \sqrt{\log n} \ll \beta_n^{2\alpha}|\mathsf{d}_k|, \quad \log n \ll q\beta_n^{1+2\alpha}, \quad \|\mathbf{V}\|_{\max} \le \frac{a}{\sqrt{n}}, \tag{44}$$

$$K \ll \frac{q\beta_n^{1+2\alpha}}{a\sqrt{\log n}} \wedge \frac{q^2\beta_n^{2+2\alpha}}{a|\mathsf{d}_k|\log n} \wedge \frac{\beta_n^{4\alpha}|\mathsf{d}_k|^2}{a^2} \tag{45}$$

for some parameter $a \ge 1$, possibly depending on n.

Proof. By the classical Lindeberg–Feller CLT (see, e.g., Chung (2001)), we have that

$$\frac{1}{\sigma_{k,i}} \Big\{ -\frac{\alpha v_k(i)}{\Lambda_i} \sum_{j\in[n]} W_{ij} + \frac{1}{t_k} \sum_{j\in[n]} \Lambda_i^{-\alpha} W_{ij} \Lambda_j^{-\alpha} v_k(j) \Big\} \to \mathcal{N}(0,1) \ \text{ in law,}$$

provided that $\|\mathbf{v}_k\|_\infty \to 0$. Together with (43), this leads to the desired CLT for $(\widehat{v}_k(i) - v_k(i))/\sigma_{k,i}$. □

To illustrate the relevance of the conditions in Corollary 1, consider the network setting with $\sum_{j\in[n]} s_{il}|v_k(j)|^2 \gtrsim \theta$, $H_{ij} = O(\theta)$, and assume that (41) holds. In addition, assume that no intrinsic cancellation occurs on the right-hand side of (42), so that $\sigma_{k,i}^2 \gtrsim t_k^{-2}\sum_{j\in[n]} \Lambda_i^{-2\alpha} s_{ij} \Lambda_j^{-2\alpha}|v_k(j)|^2$. Under these assumptions, we can obtain that

$$\sigma_{k,i}^2 \gtrsim \frac{\theta}{|\delta_k|^2 (n\theta)^{4\alpha}} = \frac{\beta_n^{4\alpha}}{n|\mathsf{d}_k|^2} \quad \text{and} \quad |\mathsf{d}_k| \le \frac{\|\mathbf{\Lambda}^{-\alpha}\mathbf{H}\mathbf{\Lambda}^{-\alpha}\|}{q^{1-4\alpha}\beta_n^{-2\alpha}} \lesssim q.$$

Consequently, condition (44) holds with $a = \sqrt{K}$ provided that $q\beta_n^{1+2\alpha} \gg \log n$, $\beta_n^{2\alpha}|\mathsf{d}_k| \gg \sqrt{\log n}$, and K is not too large, specifically

$$K \ll \left(q\beta_n^{2+2\alpha}/\log n\right)^{2/3} \wedge (\beta_n^{2\alpha}|\mathsf{d}_k|).$$

Next, we derive the higher-order asymptotic expansions for both the empirical spiked eigenvalues $\widehat{\delta}_k$ and projections of the empirical spiked eigenvectors $\mathbf{u}^T(\mathbf{D}/\mathbf{\Lambda})^{-\alpha}\widehat{\mathbf{v}}_k$, thereby sharpening the results of Theorems 1 and 2. Equivalently, we identify the leading-order random fluctuations hidden within the error terms of (31) and (35).

Theorem 4. *Under Definition 1 and Assumption 1, for each $1 \le k \le K_0$, it holds w.h.p. that*

$$\begin{aligned}\frac{\widehat{\delta}_k - t_k - A_k}{t_k} &= -2\alpha \mathbf{v}_k^T \frac{\mathbf{D}-\mathbf{\Lambda}}{\mathbf{\Lambda}}\mathbf{v}_k + \frac{1}{t_k}\mathbf{v}_k^T \overline{\mathbf{W}}\mathbf{v}_k + B_k \\ &+ O\left(\frac{1}{|\mathsf{d}_k|^3} + \frac{(\log n)^{3/2}}{q^3\beta_n^3} + \frac{\sqrt{\log n}\widetilde{\psi}_n(\delta_k)}{q}\left(\frac{\sqrt{K}}{|\mathsf{d}_k|^2} + \frac{K\sqrt{\log n}\widetilde{\psi}_n(\delta_k)}{q} + \frac{\log n}{q|\mathsf{d}_k|}\|\mathbf{v}_k\|_\infty^2\right)\right),\end{aligned} \tag{46}$$

where $\overline{\mathbf{W}} := \mathbf{\Lambda}^{-\alpha}\mathbf{W}\mathbf{\Lambda}^{-\alpha}$, the deterministic term A_k is given by

$$A_k = \alpha(2\alpha+1)t_k \mathbb{E}\mathbf{v}_k^T \frac{(\mathbf{D}-\mathbf{\Lambda})^2}{\mathbf{\Lambda}^2}\mathbf{v}_k - 4\alpha \mathbb{E}\mathbf{v}_k^T \frac{\mathbf{D}-\mathbf{\Lambda}}{\mathbf{\Lambda}}\overline{\mathbf{W}}\mathbf{v}_k, \tag{47}$$

and B_k is a centered random error satisfying

$$\operatorname{var}(B_k) \lesssim \frac{\|\mathbf{v}_k\|_\infty^2}{q^4\beta_n^4} + \frac{1}{q^4n^2\beta_n^4} + \frac{\|\mathbf{v}_k\|_\infty^2}{q^2|\mathsf{d}_k|^2\beta_n^2} + \frac{1}{q^2n|\mathsf{d}_k|^2\beta_n^2} + \frac{1}{\sqrt{n}q|\mathsf{d}_k|^4}. \tag{48}$$

Roughly speaking, the asymptotic expansion in Theorem 4 above indicates that the fluctuation of the (rescaled) empirical spiked eigenvalue $(\widehat{\delta}_k - t_k)/t_k$ is dominated, to the leading order, by the random term

$$-2\alpha \mathbf{v}_k^T \frac{\mathbf{D}-\mathbf{\Lambda}}{\mathbf{\Lambda}}\mathbf{v}_k + \frac{1}{t_k}\mathbf{v}_k^T\overline{\mathbf{W}}\mathbf{v}_k.$$

Through a direct calculation, we can obtain its variance as

$$\varsigma_k^2 := \sum_{1\le i\le j\le n}\left(\frac{\mathfrak{S}_{ij}^{\mathbf{v}_k\mathbf{v}_k}}{1+\delta_i^j}\right)^2 s_{ij}, \tag{49}$$

where δ_i^j represents the Kronecker delta, and for any vectors $\mathbf{x} = (x(i))_{i\in[n]}$, $\mathbf{y} = (y(i))_{i\in[n]} \in \mathbb{R}^n$,

$$\mathfrak{S}_{ij}^{\mathbf{x}\mathbf{y}} := -2\alpha\left(\frac{x(i)y(i)}{\Lambda_i} + \frac{x(j)y(j)}{\Lambda_j}\right) + t_k^{-1}\frac{x(i)y(j)+x(j)y(i)}{(\Lambda_i\Lambda_j)^\alpha}, \quad i,j\in[n]. \tag{50}$$

Provided that the variance of the remaining error terms is asymptotically negligible relative to ς_k, we can establish a CLT for the empirical spiked eigenvalue $\widehat{\delta}_k$, as stated in the corollary below.

Corollary 2. *Under Definition 1 and Assumption 1, assume that $\|\mathbf{v}_k\|_\infty \to 0$ and*

$$\begin{aligned}\varsigma_k \gg \frac{1}{|\mathsf{d}_k|^3} &+ \frac{(\log n)^{3/2}}{q^3\beta_n^3} + \frac{\sqrt{\log n}\widetilde{\psi}_n(\delta_k)}{q}\left(\frac{\sqrt{K}}{|\mathsf{d}_k|^2} + \frac{K\sqrt{\log n}\widetilde{\psi}_n(\delta_k)}{q} + \frac{\log n}{q|\mathsf{d}_k|}\|\mathbf{v}_k\|_\infty^2\right) \\ &+ \frac{\|\mathbf{v}_k\|_\infty}{q^2\beta_n^2} + \frac{\|\mathbf{v}_k\|_\infty}{q|\mathsf{d}_k|\beta_n} + \frac{1}{q\sqrt{n}|\mathsf{d}_k|\beta_n}.\end{aligned} \tag{51}$$

Then for each $1 \le k \le K_0$, we have that $(\widehat{\delta}_k - t_k - A_k)/(t_k\varsigma_k) \xrightarrow{d} \mathcal{N}(0,1)$ as $n\to\infty$, where ς_k is given in (49).

Proof. Applying the classical Lindeberg–Feller CLT gives that

$$\frac{1}{\varsigma_k}\left(-2\alpha\mathbf{v}_k^T\frac{\mathbf{D}-\mathbf{\Lambda}}{\mathbf{\Lambda}}\mathbf{v}_k+\frac{1}{t_k}\mathbf{v}_k^T\overline{\mathbf{W}}\mathbf{v}_k\right)\to\mathcal{N}(0,1)\ \text{ in law,}$$

whenever $|\mathbf{v}_k|_\infty \to 0$. Combining this with (51) yields the claimed CLT for $(\widehat{\delta}_k - t_k - A_k)/(t_k\varsigma_k)$. □

Note that Corollary 2 also yields a CLT for $\widehat{\delta}_k - \delta_k - A_k'$ by choosing the bias term as $A_k' = t_k + A_k - \delta_k$. We now provide some insights into the assumptions in Corollary 2 underlying the CLT for $\widehat{\delta}_k$. In the generic case—i.e., when there are no essential cancellations in the expression of ς_k^2—the asymptotic standard deviation ς_k is typically of order $\beta_n^{2\alpha}/(\sqrt{n}|\mathsf{d}_k|)+\|\mathbf{v}_k\|_4^2/q$, where $\|\mathbf{v}_k\|_\infty \ge \|\mathbf{v}_k\|_4^2 = (\sum_i |v_k(i)|^4)^{1/2} \ge n^{-1/2}$. Consequently, if the following conditions hold

$$|\mathsf{d}_k| \gg n^{1/4}/\beta_n^\alpha \quad \text{and} \quad q \gg \frac{K\log n}{\beta_n^{2+2\alpha}}(n^{1/4}+\sqrt{n}\|\mathbf{V}\|_{\max}), \tag{52}$$

then (51) is satisfied, and hence the CLT for $\widehat{\delta}_k$ holds. As shown in Corollary 2, the asymptotic bias is given by the population quantity A_k, which takes the form

$$A_k = \alpha(2\alpha+1)t_k\sum_{i,j\in[n]} s_{ij}\frac{v_k(i)^2}{\Lambda_i^2} - 4\alpha\sum_{i,j\in[n]} s_{ij}\frac{v_k(i)v_k(j)}{\Lambda_i^{1+\alpha}\Lambda_j^\alpha}. \tag{53}$$

In practice, A_k can be estimated by the plug-in estimator $\widehat{A}_k$ defined in (78) below, obtained by substituting empirical counterparts for all population parameters; see the discussion in Section III-C for details.

We now turn to the higher-order asymptotic expansions for the empirical spiked eigenvectors $\widehat{\mathbf{v}}_k$, which will be used to establish the corresponding CLTs. For notational convenience and to streamline the presentation of Theorem 5, we decompose any vector $\mathbf{u}$ into its components that are parallel and orthogonal to $\mathbf{v}_k$, respectively.

Theorem 5. *Assume that Definition 1 and Assumption 1 hold.*

1) For each $1 \le k \le K_0$ and any deterministic unit vector $\mathbf{u}\in\mathbb{R}^n$ such that $\mathbf{u}^T\mathbf{v}_k = 0$, it holds w.h.p. that

$$\begin{aligned}\mathbf{u}^T(\mathbf{D}/\mathbf{\Lambda})^{-\alpha}\widehat{\mathbf{v}}_k - \mathcal{A}_k &= \mathbf{w}^T\left(-2\alpha\frac{\mathbf{D}-\mathbf{\Lambda}}{\mathbf{\Lambda}}+t_k^{-1}\overline{\mathbf{W}}\right)\mathbf{v}_k+\mathcal{B}_k\\ &+O\left(\frac{1}{|\mathsf{d}_k|^3}+\frac{(\log n)^{3/2}}{q^3\beta_n^3}+\frac{K\sqrt{\log n}}{q|\mathsf{d}_k|^2}\widetilde{\psi}_n(\delta_k)+\frac{K^{3/2}\log n}{q^2}\widetilde{\psi}_n(\delta_k)^2\right)\\ &+O\left(\left(\frac{\log n}{q^2\beta_n^2}+\frac{\log n}{q|\mathsf{d}_k|^2}\right)\|\mathbf{v}_k\|_\infty^2+\frac{(\log n)^{3/2}}{q^2|\mathsf{d}_k|\beta_n}\|\mathbf{v}_k\|_\infty^3\right),\end{aligned} \tag{54}$$

where we choose the sign of $\widehat{\mathbf{v}}_k$ such that $\widehat{\mathbf{v}}_k^T\mathbf{v}_k > 0$, and $\mathbf{w}$ is a deterministic vector defined as

$$\mathbf{w} := \left(\mathbf{I}+\mathbf{V}_{-k}\frac{\mathbf{\Delta}_{-k}}{t_k\mathbf{I}-\mathbf{\Delta}_{-k}}\mathbf{V}_{-k}^T\right)\mathbf{u}. \tag{55}$$

The deterministic bias term is given by

$$\mathcal{A}_k = \mathbb{E}\mathbf{w}^T\left(\alpha(2\alpha+1)\frac{(\mathbf{D}-\mathbf{\Lambda})^2}{\mathbf{\Lambda}^2}-\frac{2\alpha}{t_k}\left(\frac{\mathbf{D}-\mathbf{\Lambda}}{\mathbf{\Lambda}}\overline{\mathbf{W}}+\overline{\mathbf{W}}\frac{\mathbf{D}-\mathbf{\Lambda}}{\mathbf{\Lambda}}\right)+\frac{\overline{\mathbf{W}}^2}{t_k^2}\right)\mathbf{v}_k,$$

and the centered random fluctuation $\mathcal{B}_k$ satisfies

$$\operatorname{var}(\mathcal{B}_k) \lesssim \frac{\|\mathbf{v}_k\|_\infty\|\mathbf{w}\|_\infty}{q^4\beta_n^4}+\frac{1}{q^4n^2\beta_n^4}+\frac{1}{q^2|\mathsf{d}_k|^2\beta_n^2}\left(\|\mathbf{v}_k\|_\infty\|\mathbf{w}\|_\infty+\frac{1}{n}\right)+\frac{1}{q\sqrt{n}|\mathsf{d}_k|^4}.$$

2) For $\mathbf{u} = \mathbf{v}_k$ and each $1 \le k \le K_0$, it holds w.h.p. that

$$\begin{aligned}\mathbf{v}_k^T(\mathbf{D}/\mathbf{\Lambda})^{-\alpha}\widehat{\mathbf{v}}_k - \mathbf{v}_k^T(\mathbf{D}/\mathbf{\Lambda})^{-\alpha}\mathbf{v}_k - \mathfrak{A}_k &= \frac{\alpha^2}{2}\mathbf{v}_k^T\left(\frac{\mathbf{D}-\mathbf{\Lambda}}{\mathbf{\Lambda}}\right)^2\mathbf{v}_k - \frac{1}{2t_k^2}\mathbf{v}_k^T\overline{\mathbf{W}}^2\mathbf{v}_k+\mathfrak{B}_k\\ &+O\left(\frac{1}{|\mathsf{d}_k|^4}+\frac{K\log n}{q^2}\widetilde{\psi}_n(\delta_k)^2+\left(\frac{\log n}{q^2\beta_n^2}+\frac{\log n}{q|\mathsf{d}_k|^3}\right)\|\mathbf{v}_k\|_\infty^2\right),\end{aligned} \tag{56}$$

where $\mathfrak{A}_k$ is a deterministic bias term given by

$$\mathfrak{A}_k := (\delta_k^2\mathbf{v}_k^T\widetilde{\mathbf{\Upsilon}}_k'(t_k)\mathbf{v}_k)^{-1/2} - 1 + \frac{1}{2}\mathbf{v}_k^T(t_k^2\widetilde{\mathbf{\Upsilon}}'(t_k)+2t_k\widetilde{\mathbf{\Upsilon}}(t_k)+\mathbf{I})\mathbf{v}_k$$

and $\mathfrak{B}_k$ is a random variable satisfying

$$\mathbb{E}\mathfrak{B}_k^2 \lesssim \left(\frac{1}{q^8\beta_n^6} + \frac{1}{q^2|\mathsf{d}_k|^6} + \frac{1}{q^4|\mathsf{d}_k|^4\beta_n^4}\right) n^2\|\mathbf{v}_k\|_\infty^4.$$

We can also establish the corresponding results for $\mathbf{u}^T\widehat{\mathbf{v}}_k$. More precisely,

3) For each $1 \le k \le K_0$ and any deterministic unit vector $\mathbf{u}$ such that $\mathbf{u}^T\mathbf{v}_k = 0$, it holds w.h.p. that

$$\begin{aligned}\mathbf{u}^T\widehat{\mathbf{v}}_k - \widetilde{\mathcal{A}}_k &= \mathbf{w}^T\left(-2\alpha\frac{\mathbf{D}-\mathbf{\Lambda}}{\mathbf{\Lambda}} + t_k^{-1}\overline{\mathbf{W}}\right)\mathbf{v}_k + \alpha\mathbf{u}^T\frac{\mathbf{D}-\mathbf{\Lambda}}{\mathbf{\Lambda}}\mathbf{v}_k + \widetilde{\mathcal{B}}_k \\ &+ O\left(\frac{1}{|\mathsf{d}_k|^3} + \frac{(\log n)^{3/2}}{q^3\beta_n^3} + \frac{K\sqrt{\log n}}{q|\mathsf{d}_k|^2}\widetilde{\psi}_n(\delta_k) + \frac{K^{3/2}\log n}{q^2}\widetilde{\psi}_n(\delta_k)^2\right) \\ &+ O\left(\left(\frac{\log n}{q^2\beta_n^2} + \frac{\log n}{q|\mathsf{d}_k|^2}\right)\|\mathbf{v}_k\|_\infty^2 + \frac{(\log n)^{3/2}}{q^2|\mathsf{d}_k|\beta_n}\|\mathbf{v}_k\|_\infty^3\right),\end{aligned} \tag{57}$$

where we choose the sign of $\widehat{\mathbf{v}}_k$ such that $\widehat{\mathbf{v}}_k^T\mathbf{v}_k > 0$. The deterministic bias term is given by

$$\widetilde{\mathcal{A}}_k = \mathcal{A}_k + \mathbb{E}\mathbf{u}^T\left(-\frac{\alpha(3\alpha+1)}{2}\frac{(\mathbf{D}-\mathbf{\Lambda})^2}{\mathbf{\Lambda}^2} + \frac{\alpha}{t_k}\frac{\mathbf{D}-\mathbf{\Lambda}}{\mathbf{\Lambda}}\overline{\mathbf{W}}\right)\mathbf{v}_k,$$

and the centered random fluctuation $\widetilde{\mathcal{B}}_k$ satisfies

$$\operatorname{var}(\widetilde{\mathcal{B}}_k) \lesssim \frac{\|\mathbf{v}_k\|_\infty\|\mathbf{w}\|_\infty}{q^4\beta_n^4} + \frac{1}{q^4n^2\beta_n^4} + \frac{1}{q^2|\mathsf{d}_k|^2\beta_n^2}\left(\|\mathbf{v}_k\|_\infty\|\mathbf{w}\|_\infty + \frac{1}{n}\right) + \frac{1}{q\sqrt{n}|\mathsf{d}_k|^4}.$$

4) For $\mathbf{u} = \mathbf{v}_k$ and each $1 \le k \le K_0$, it holds w.h.p. that

$$\begin{aligned}\mathbf{v}_k^T\widehat{\mathbf{v}}_k - 1 - \mathfrak{A}_k &= -\frac{1}{2}\mathbf{v}_k^T\left(\alpha\frac{\mathbf{D}-\mathbf{\Lambda}}{\mathbf{\Lambda}} - \frac{\overline{\mathbf{W}}}{t_k}\right)^2\mathbf{v}_k + \widetilde{\mathfrak{B}}_k \\ &+ O\left(\frac{1}{|\mathsf{d}_k|^4} + \frac{K\log n}{q^2}\widetilde{\psi}_n(\delta_k)^2 + \left(\frac{\log n}{q^2\beta_n^2} + \frac{\log n}{q|\mathsf{d}_k|^3}\right)\|\mathbf{v}_k\|_\infty^2\right),\end{aligned} \tag{58}$$

where $\widetilde{\mathfrak{B}}_k$ is a random variable satisfying

$$\mathbb{E}\widetilde{\mathfrak{B}}_k^2 \lesssim \left(\frac{1}{q^8\beta_n^6} + \frac{1}{q^2|\mathsf{d}_k|^6} + \frac{1}{q^4|\mathsf{d}_k|^4\beta_n^4}\right) n^2\|\mathbf{v}_k\|_\infty^4.$$

Using the higher-order asymptotic expansions established in Theorem 5 above, we are now ready to present more general CLT results for the empirical spiked eigenvectors $\widehat{\mathbf{v}}_k$ (*beyond* the result obtained previously in Corollary 1) under some suitable conditions on q and $|\mathsf{d}_k|$, as stated in the corollary below.

Corollary 3. *Assume that Definition 1 and Assumption 1 hold, and fix any $1 \le k \le K_0$. Assume that*

$$\|\mathbf{v}_k\|_\infty \to 0, \ \sqrt{n} \ll q^2, \ \sqrt{n} \ll |\mathsf{d}_k|^2. \tag{59}$$

For an arbitrary deterministic unit vector $\mathbf{u} \in \mathbb{R}^n$, define

$$\widetilde{\mathbf{w}} := \left(\mathbf{I} + \mathbf{V}_{-k}\frac{t_k}{t_k\mathbf{I} - \mathbf{\Delta}_{-k}}\mathbf{V}_{-k}^T\right)(\mathbf{I} - \mathbf{v}_k\mathbf{v}_k^T)\mathbf{u}. \tag{60}$$

1) Assume further that $|\mathbf{u}^T\mathbf{v}_k| \neq 1$. Define the asymptotic variance

$$\mathfrak{s}_{\mathbf{u},k}^2 := \sum_{i\le j\in[n]}\left(\frac{\mathfrak{S}_{ij}^{\widetilde{\mathbf{w}}\mathbf{v}_k}}{1+\delta_i^j}\right)^2 s_{ij}, \tag{61}$$

where function $\mathfrak{S}$ is given in (50). If the following lower bound holds

$$\begin{aligned}\frac{\mathfrak{s}_{\mathbf{u},k}}{\sqrt{1-|\mathbf{u}^T\mathbf{v}_k|^2}} \gg \frac{1}{|\mathsf{d}_k|^3} + \frac{(\log n)^{3/2}}{q^3\beta_n^3} + \frac{K\sqrt{\log n}}{q|\mathsf{d}_k|^2}\widetilde{\psi}_n(\delta_k) + \frac{K^{3/2}\log n}{q^2}\widetilde{\psi}_n(\delta_k)^2 + \left(\frac{\log n}{q^2\beta_n^2} + \frac{\log n}{q|\mathsf{d}_k|^2}\right)\|\mathbf{v}_k\|_\infty^2 \\ + \frac{(\log n)^{3/2}}{q^2|\mathsf{d}_k|\beta_n}\|\mathbf{v}_k\|_\infty^3 + \left(\frac{1}{q^2\beta_n^2} + \frac{1}{q|\mathsf{d}_k|\beta_n}\right)\sqrt{\|\mathbf{v}_k\|_\infty\|\widetilde{\mathbf{w}}\|_\infty + n^{-1}}\end{aligned}$$

$$+\frac{|\mathbf{u}^T\mathbf{v}_k|}{\sqrt{1-|\mathbf{u}^T\mathbf{v}_k|^2}}\left(\frac{1}{|\mathsf{d}_k|^4}+\frac{K\log n}{q^2}\widetilde{\psi}_n(\delta_k)^2+\left(\frac{\log n}{q^2\beta_n^2}+\frac{\log n}{q|\mathsf{d}_k|^3}\right)\|\mathbf{v}_k\|_\infty^2\right)$$
$$+\frac{|\mathbf{u}^T\mathbf{v}_k|}{\sqrt{1-|\mathbf{u}^T\mathbf{v}_k|^2}}n\|\mathbf{v}_k\|_\infty^2\left(\frac{1}{q^4\beta_n^3}+\frac{1}{q|\mathsf{d}_k|^3}+\frac{1}{q^2|\mathsf{d}_k|^2\beta_n^2}\right),$$

then we have the following CLT

$$\frac{\mathbf{u}^T(\mathbf{D}/\mathbf{\Lambda})^{-\alpha}\widehat{\mathbf{v}}_k-\mathbf{u}^T\mathbf{v}_k\mathbf{v}_k^T(\mathbf{D}/\mathbf{\Lambda})^{-\alpha}\mathbf{v}_k-\mathfrak{D}_{\mathbf{u},k}}{\mathfrak{s}_{\mathbf{u},k}}\xrightarrow{d}\mathcal{N}(0,1)\tag{62}$$

as $n\to\infty$. Here, the sign of $\widehat{\mathbf{v}}_k$ is chosen such that $\widehat{\mathbf{v}}_k^T\mathbf{v}_k>0$, and the asymptotic bias is given by

$$\mathfrak{D}_{\mathbf{u},k}:=\mathbb{E}\widetilde{\mathbf{w}}^T\left(\alpha(2\alpha+1)\frac{(\mathbf{D}-\mathbf{\Lambda})^2}{\mathbf{\Lambda}^2}-\frac{2\alpha}{t_k}\left(\frac{\mathbf{D}-\mathbf{\Lambda}}{\mathbf{\Lambda}}\overline{\mathbf{W}}+\overline{\mathbf{W}}\frac{\mathbf{D}-\mathbf{\Lambda}}{\mathbf{\Lambda}}\right)+\frac{\overline{\mathbf{W}}^2}{t_k^2}\right)\mathbf{v}_k$$
$$+\frac{\alpha^2}{2}\mathbf{u}^T\mathbf{v}_k\mathbb{E}\mathbf{v}_k^T\left(\frac{\mathbf{D}-\mathbf{\Lambda}}{\mathbf{\Lambda}}\right)^2\mathbf{v}_k-\frac{1}{2t_k^2}\mathbf{u}^T\mathbf{v}_k\mathbb{E}\mathbf{v}_k^T\overline{\mathbf{W}}^2\mathbf{v}_k$$
$$+\mathbf{u}^T\mathbf{v}_k\left((\delta_k^2\mathbf{v}_k^T\widetilde{\mathbf{\Upsilon}}_k'(t_k)\mathbf{v}_k)^{-1/2}-1+\frac{1}{2}\mathbf{v}_k^T(t_k^2\widetilde{\mathbf{\Upsilon}}'(t_k)+2t_k\widetilde{\mathbf{\Upsilon}}(t_k)+\mathbf{I})\mathbf{v}_k\right).$$

2) In the special case of $\mathbf{u}=\mathbf{v}_k$, fix $1\le k\le K_0$. Assume that the following lower bound holds

$$s_{\mathbf{v}_k,k}\gg\kappa_{\mathbf{v}_k}^{1/4}+\frac{1}{|\mathsf{d}_k|^4}+\frac{K\log n}{q^2}\widetilde{\psi}_n(\delta_k)^2+\left(\frac{\log n}{q^2\beta_n^2}+\frac{\log n}{q|\mathsf{d}_k|^3}\right)\|\mathbf{v}_k\|_\infty^2\tag{63}$$
$$+n\|\mathbf{v}_k\|_\infty^2\left(\frac{1}{q^4\beta_n^3}+\frac{1}{q|\mathsf{d}_k|^3}+\frac{1}{q^2|\mathsf{d}_k|^2\beta_n^2}\right).$$

Here, the forms of $s_{\mathbf{v}_k,k}$ and $\kappa_{\mathbf{v}_k}$ are involved; explicit expressions are given in (C.112) and (C.113), respectively, in Section C-H of the Supplementary Material. Then the following CLT holds

$$\frac{\mathbf{v}_k^T(\mathbf{D}/\mathbf{\Lambda})^{-\alpha}\widehat{\mathbf{v}}_k-\mathbf{v}_k^T(\mathbf{D}/\mathbf{\Lambda})^{-\alpha}\mathbf{v}_k-\mathfrak{C}_{\mathbf{v}_k,k}}{\mathfrak{s}_{\mathbf{v}_k,k}}\xrightarrow{d}\mathcal{N}(0,1)\tag{64}$$

as $n\to\infty$, where the sign of $\widehat{\mathbf{v}}_k$ is chosen so that $\widehat{\mathbf{v}}_k^T\mathbf{v}_k>0$, and the asymptotic bias is given by

$$\mathfrak{C}_{\mathbf{v}_k,k}:=\frac{\alpha^2}{2}\mathbb{E}\mathbf{v}_k^T\left(\frac{\mathbf{D}-\mathbf{\Lambda}}{\mathbf{\Lambda}}\right)^2\mathbf{v}_k-\frac{\mathbb{E}\mathbf{v}_k^T\overline{\mathbf{W}}^2\mathbf{v}_k}{2t_k^2}+(\delta_k^2\mathbf{v}_k^T\widetilde{\mathbf{\Upsilon}}_k'(t_k)\mathbf{v}_k)^{-1/2}-1$$
$$+\frac{1}{2}\mathbf{v}_k^T(t_k^2\widetilde{\mathbf{\Upsilon}}'(t_k)+2t_k\widetilde{\mathbf{\Upsilon}}(t_k)+\mathbf{I})\mathbf{v}_k.$$

By parts 3) and 4) of Theorem 5, the same CLTs also hold for $\mathbf{u}^T\widehat{\mathbf{v}}_k$ and $\mathbf{v}_k^T\widehat{\mathbf{v}}_k$, with the only difference being the asymptotic variances, whose explicit forms are omitted for brevity.

Remark 5. It is worth mentioning that one can, in fact, derive the asymptotic expansions of even higher order than those in (46) and (57), in which higher-order random fluctuations are extracted explicitly from the error terms. Such refinements would allow one to establish the limiting distributions for $\widehat{\delta}_k-t_k$ and $\mathbf{u}^T\widehat{\mathbf{v}}_k$ under substantially weaker assumptions on q and $|\mathsf{d}_k|$, in particular, for smaller values of these parameters. In principle, our technical framework permits the derivation of arbitrarily high-order asymptotic series for $\widehat{\delta}_k-t_k$ and $\mathbf{u}^T\widehat{\mathbf{v}}_k$ provided that $q\ge n^{\varepsilon}$ and $|\mathsf{d}_k|\ge n^{\varepsilon}$ for some constant $\varepsilon>0$; see, for example, Fan et al. (2022a). However, in contrast to the setting of Fan et al. (2022a), identifying the limiting distributions of these higher-order terms here is considerably more challenging, due to the intrinsic dependence between random matrices $\mathbf{D}$ and $\mathbf{W}$. Owing to space constraints, we leave a systematic investigation of this problem to future work.

For completeness, we finally consider the practical problem of estimating the latent embedding dimensionality K_0, namely, the number of strong spikes.

Theorem 6. *Assume that Definition 1 and Assumption 1 hold, and in addition*

$$|\mathsf{d}_{K_0+1}|\gg1,\quad\left|\frac{\delta_{K_0+1}}{\delta_{K_0+2}}\right|\ge1+\epsilon_0,\quad K\sqrt{\log n}\widetilde{\psi}_n(\delta_{K_0+1})\ll q.\tag{65}$$

Assume futher that K_0 satisfies

$$K_0=\max\{k\in[K]:|\mathsf{d}_k|\ge a_n\}\tag{66}$$

for some deterministic sequence $a_n \to \infty$, and there exists another deterministic sequence $a'_n \to \infty$ such that

$$\limsup_{n\to\infty} \left|\frac{a'_n}{a_n}\right| < 1, \quad \limsup_{n\to\infty} \frac{|\mathsf{d}_{K_0+1}|}{a'_n} < 1. \tag{67}$$

Then the estimator of the latent embedding dimensionality defined as

$$\widehat{K}_0 := \max\{k \in [K] : |\widehat{\delta}_k|/(q^{1-4\alpha}\beta_n^{-2\alpha}) \geq a'_n\} \tag{68}$$

is consistent, i.e., $\mathbb{P}\{\widehat{K}_0 = K_0\} \to 1$ *as* $n \to \infty$.

Theorem 6 above establishes the theoretical validity of estimator $\widehat{K}_0$ defined in (68). We now provide guidance on the choice of the tuning parameter a'_n. Assume that in (67) one may take $a_n = q^{1-4\alpha}\beta_n^{-2\alpha}(\log n)^c$ for some constant $c > 0$, and that condition (iv) of Definition 1 is strengthened to

$$c_0 \leq \max_{i\in[n]} \theta_i \leq C_0. \tag{69}$$

Under these assumptions, we propose to use

$$a'_n = \frac{\check{q}}{(\min_{j\in[n]} d_j)^{2\alpha}} \frac{(\log n)^c}{\log\log n}, \tag{70}$$

where $\check{q}^2 := \max_{j\in[n]} \sum_{l\in[n]} X_{jl}$ denotes the maximum node degree of the network. By a standard concentration inequality (see, e.g., Lemma C.2 below) and condition (69), we have that with probability $1 - o(1)$,

$$\check{q}^2 = (1 + o(1))q^2 \max_{j\in[n]} \theta_j \sim q^2, \quad \min_{j\in[n]} d_j = (1 + o(1))q^2 \min_j \theta_j \sim q^2\beta_n. \tag{71}$$

As a concrete example, following Fan et al. (2022), in the DCMM model (Example 1) one may choose $c = 1/2$ for testing a given pair of nodes and $c = 3/2$ for the group test. Further discussions on the estimation and inference of K_0 in network settings can be found in Fan et al. (2022); Han et al. (2023).

Remark 6. Equation (70) represents a conservative sufficient threshold derived from the uniform worst-case bounds that account for potentially strong degree heterogeneity. When additional structural information about the model is available—for example, when the degree parameters are bounded or appropriately truncated—so that the effective noise scale is of order $\|\mathbf{W}\| \sim q^{1-4\alpha}$, we can use a less conservative threshold such as $q^{1-4\alpha}\log\log n$. This choice continues to satisfy the separation condition required by Theorem 6, while typically leading to improved finite-sample performance in the estimation of K_0.

C. *Estimating population quantities*

For practical implementation including the numerical examples in Section VI-A, we provide a computational algorithm for evaluating the population quantities. Recall from (24) and (25) that the (deterministic) matrix-valued function $\widetilde{\mathbf{\Upsilon}}(z)$ is analytic for sufficiently large $|z|$ and vanishes at infinity. Consequently, $\widetilde{\mathbf{\Upsilon}}(z)$ admits a Laurent series expansion of form

$$\widetilde{\mathbf{\Upsilon}}(z) = \sum_{l=0}^{\infty} \frac{1}{z^l}\mathbf{Y}_l, \tag{72}$$

where $\mathbf{Y}_l$'s are $n \times n$ deterministic diagonal matrices. These matrices can be computed recursively via

$$\mathbf{Y}_0 = 0, \quad \mathbf{Y}_1 = -\mathbf{I}, \quad \text{and} \quad \mathbf{Y}_{l+1}\mathbf{e}_{[n]} = -\sum_{m=0}^{l} \mathbf{Y}_m\mathbf{\Sigma}\mathbf{Y}_{l-m}\mathbf{e}_{[n]}, \tag{73}$$

where $\mathbf{e}_{[n]} := \sum_{i\in[n]} \mathbf{e}_i$ with $\mathbf{e}_i$ denoting the ith canonical basis vector of $\mathbb{R}^n$, and $\mathbf{\Sigma} = (\Sigma_{ij})_{1\leq i,j\leq n}$ is a covariance matrix given by

$$\Sigma_{ij} := \operatorname{var}\left(\mathbf{e}_i^T\mathbf{\Lambda}^{-\alpha}\widetilde{\mathbf{X}}\mathbf{\Lambda}^{-\alpha}\mathbf{e}_j\right) = \Lambda_i^{-2\alpha}\Lambda_j^{-2\alpha}s_{ij}. \tag{74}$$

With this definition of $\mathbf{\Sigma}$, the generalized QVE (24) can be written compactly as

$$z\widetilde{\mathbf{\Upsilon}}(z)\mathbf{e}_{[n]} = -(\mathbf{I} + \widetilde{\mathbf{\Upsilon}}(z)\mathbf{\Sigma}\widetilde{\mathbf{\Upsilon}}(z))\mathbf{e}_{[n]}. \tag{75}$$

Motivated by the theoretical representations (72)–(73) and the approximation analyses in Fan et al. (2022a,b), we adopt a *quadratic approximation* to $\widetilde{M}_i(z)$'s and $\widetilde{\mathbf{\Upsilon}}(z)$, namely,

$$\mathcal{M}_i(z) := -z^{-1} - z^{-3}\Lambda_i^{-2\alpha} \sum_{j\in[n]} \Lambda_j^{-2\alpha} s_{ij} \quad \text{and} \quad \boldsymbol{\mathcal{Y}}(z) := \operatorname{diag}\{\mathcal{M}_1(z), \cdots, \mathcal{M}_n(z)\}. \tag{76}$$

To compute t_k, note that in the defining equation (27), the fractional term is asymptotically negligible compared to the leading term, introducing an error of order $O(|t_k|/|\mathrm{d}_k|^4)$ to the numerical value of t_k. We therefore ignore this term and replace $\widetilde{\mathbf{\Upsilon}}(z)$ by its quadratic approximation $\boldsymbol{\mathcal{Y}}(z)$ introduced in (76), giving rise to the simplified equation

$$1 + \delta_k \mathbf{v}_k^T \boldsymbol{\mathcal{Y}}(x)\mathbf{v}_k = 0. \tag{77}$$

The value of t_k can then be obtained numerically by applying the Newton–Raphson method to solve (77) over $x \in \mathbb{R}$ for each $1 \le k \le K_0$.

The results from Corollary 1 to Corollary 3 suggest that the population spiked eigenvector $\mathbf{v}_k$ can be estimated consistently by its empirical counterpart $\widehat{\mathbf{v}}_k$, and that t_k can be estimated by $\widehat{\delta}_k$. We now discuss the estimation of the noise variances $s_{ij} = \mathbb{E}|W_{ij}|^2$. A naive estimator of s_{ij} is given by $\widehat{W}_{0,ij}^2$, where

$$\widehat{\mathbf{W}}_0 = (\widehat{W}_{0,ij}) := \widetilde{\mathbf{X}} - \mathbf{D}^\alpha \Big(\sum_{k\in[\widehat{K}]} \widehat{\delta}_k \widehat{\mathbf{v}}_k \widehat{\mathbf{v}}_k^T \Big) \mathbf{D}^\alpha$$

and $\widehat{K} = \widehat{K}_0$ is given by (68). In some applications, the naive estimator is *not* accurate enough and a *bias-corrected* one is needed for improved estimation accuracy. To address this issue, we adopt a one-step refinement procedure similar to the bias-correction idea of Fan et al. (2022b), motivated by the higher-order asymptotic expansion of t_k (recall (27), (76), and(77)). The goal is to shrink $\widehat{\delta}_k$ and thereby reduce its bias. The one-step refinement procedure is as follows:

(i) Compute the initial estimator $\widehat{\mathbf{W}}_0$ and estimate s_{ij} by $\widehat{s}_{ij,0} = \widehat{W}_{0,ij}^2$. Let $\widehat{\mathbf{\Sigma}}_0 := (\widehat{s}_{ij,0})_{i,j\in[n]}$, and set the initial estimator of t_k as $\widehat{t}_{k,0} = \widehat{\delta}_k$.

(ii) Estimate the theoretical bias term A_k by

$$\widehat{A}_{k,0} = \alpha(2\alpha+1)\widehat{t}_{k,0} \sum_{i,j\in[n]} \widehat{s}_{ij,0} \frac{\widehat{v}_k(i)^2}{d_i^2} - 4\alpha \sum_{i,j\in[n]} \widehat{s}_{ij,0} \frac{\widehat{v}_k(i)\widehat{v}_k(j)}{d_i^{1+\alpha} d_j^\alpha}. \tag{78}$$

(iii) Update the estimator of t_k via

$$\widehat{t}_{k,1} = \widehat{\delta}_k - \widehat{A}_{k,0}.$$

(iv) Using the initial estimator $\widehat{\mathbf{W}}_0$, update the estimator of δ_k as

$$\widetilde{\delta}_k := \left[\frac{1}{\widehat{t}_{k,1}} + \frac{\widehat{\mathbf{v}}_k^T \operatorname{diag}[(\mathbf{D}^{-\alpha}\widehat{\mathbf{W}}_0\mathbf{D}^{-\alpha})^2]\widehat{\mathbf{v}}_k}{\widehat{t}_{k,1}^3} \right]^{-1}. \tag{79}$$

(v) Update the estimator of $\mathbf{W}$ as

$$\widehat{\mathbf{W}} = (\widehat{W}_{ij}) := \widetilde{\mathbf{X}} - \mathbf{D}^\alpha \Big(\sum_{k\in[\widehat{K}]} \widetilde{\delta}_k \widehat{\mathbf{v}}_k \widehat{\mathbf{v}}_k^T \Big) \mathbf{D}^\alpha,$$

and estimate s_{ij} by $\widehat{s}_{ij} := \widehat{W}_{ij}^2$.

Finally, we propose estimating the asymptotic variances of the eigenvector components $\sigma_{k,i}^2$ and of the eigenvalue ς_k^2 by substituting t_k, $\mathbf{v}_k$, s_{ij}, and $\mathbf{\Lambda}$ with $\widehat{t}_{k,0}$, $\widehat{\mathbf{v}}_k$, $\widehat{s}_{ij}$, and $\mathbf{D}$, respectively, in (42), (49), and (50). In particular, to estimate the population quantity δ_k (as opposed to t_k), we apply the above *bias-correction* procedure (79) based on the empirical spiked eigenvalues $\widehat{\delta}_k$ and the estimator $\widehat{A}_k \equiv \widehat{A}_{k,0}$.

To demonstrate that the CLT results in Corollaries 1 and 2 continue to hold with sample plug-in estimators, we establish end-to-end feasible CLTs in Section E of the Supplementary Material under the degree-corrected stochastic block model (DCSBM). The DCSBM is a special case of the DCMM in Example 1, where the membership probability vector $\boldsymbol{\pi}_i$ for each node i is restricted to the set $\{\mathbf{e}_1, \cdots, \mathbf{e}_K\}$ with $\mathbf{e}_k$ denoting the kth standard basis vector of $\mathbb{R}^K$. In these feasible CLTs, the asymptotic biases and variances in Corollaries 1 and 2 are replaced by their sample counterparts via the plug-in rules, obtained by substituting $(t_k, \mathbf{v}_k, s_{ij}, \mathbf{\Lambda})$ with $(\widehat{t}_{k,0}, \widehat{\mathbf{v}}_k, \widehat{s}_{ij,0}, \mathbf{D})$, respectively. For simplicity, we adopt the naive residual-based estimator $\widehat{s}_{ij,0}$ for the noise variances. The theorems in Section E of the Supplementary Material show that the resulting CLTs remain valid under mild degree heterogeneity. A complete list of assumptions, along with proof sketches, are provided therein.

IV. Technical innovations of our theory

The foundational theoretical contributions of this paper are the higher-order asymptotic expansions established in Theorems 3, 4, and 5. These expansions serve as the key technical tools for deriving the CLTs for both empirical spiked eigenvalues and empirical spiked eigenvectors, as presented in Corollaries 1, 2, and 3, respectively. To clearly highlight our technical innovations, we provide a detailed roadmap of the proof strategies for the main results and discuss the additional mathematical challenges encountered along the way. The complete proofs are deferred to Sections B–F of the the Supplementary Material. Our analysis follows a general line of approach that is standard in the literature on spiked random matrix models and eigenvalue-eigenvector fluctuations, while addressing several new technical difficulties specific to our setting; see, for example, Fan et al. (2022a), Fan et al. (2022), and Ke and Wang (2025).

We begin by analyzing the master equation for the spiked eigenvalues. Our proofs rely primarily on the resolvents (i.e., Green's functions) of the relevant random matrices, defined as

$$\mathbf{G}(z) := \left(\overline{\mathbf{W}} - z(\mathbf{D}/\mathbf{\Lambda})^{2\alpha}\right)^{-1} \quad \text{and} \quad \mathbf{R}(z) := \left(\overline{\mathbf{W}} - z\mathbf{I}\right)^{-1}, \quad z \in \mathbb{C}, \tag{80}$$

where we recall that $\overline{\mathbf{W}} = \mathbf{\Lambda}^{-\alpha}\mathbf{W}\mathbf{\Lambda}^{-\alpha}$. We next focus on the equation governing the behavior of the empirical spiked eigenvalue $\widehat{\delta}_k$, observing that

$$\begin{aligned} \det(\mathbf{X} - \widehat{\delta}_k\mathbf{I}) = 0 &\iff \det(\mathbf{\Lambda}^{-\alpha}\widetilde{\mathbf{X}}\mathbf{\Lambda}^{-\alpha} - \widehat{\delta}_k(\mathbf{D}/\mathbf{\Lambda})^{2\alpha}) = 0 \\ &\iff \det(\mathbf{G}^{-1}(\widehat{\delta}_k) + \mathbf{V}\mathbf{\Delta}\mathbf{V}^T) = 0 \\ &\iff \det(\mathbf{\Delta}^{-1} + \mathbf{V}^T\mathbf{G}(\widehat{\delta}_k)\mathbf{V}) = 0. \end{aligned} \tag{81}$$

To analyze the asymptotic behavior of the empirical spiked eigenvalue $\widehat{\delta}_k$, we introduce the asymptotic limit of the resolvent $\mathbf{G}(z)$, denoted as $\widetilde{\mathbf{\Upsilon}}(z)$ (see (25)). We then replace $\mathbf{G}(z)$ in (81) with $\widetilde{\mathbf{\Upsilon}}(z)$ and obtain a deterministic equation

$$\det(\mathbf{\Delta}^{-1} + \mathbf{V}^T\widetilde{\mathbf{\Upsilon}}(t_k)\mathbf{V}) = 0, \tag{82}$$

which characterizes the asymptotic limit of $\widehat{\delta}_k$, denoted as t_k. To establish the relationship between $\widehat{\delta}_k$ and t_k and derive the asymptotic expansion of $\widehat{\delta}_k$, we subtract the expressions in (81) and (82), and control the error term $\mathbf{V}(\mathbf{G}(z) - \widetilde{\mathbf{\Upsilon}}(z))\mathbf{V}$. This enables us to analyze the asymptotic behavior of the empirical spiked eigenvalues.

Moving on to the empirical spiked eigenvectors, we employ the Cauchy integral formula to extract a specific spiked eigenvector $\widehat{\mathbf{v}}_k$ from the random generalized Laplacian matrix $\mathbf{X}$ using the formula

$$(\mathbf{D}/\mathbf{\Lambda})^{-\alpha}\widehat{\mathbf{v}}_k\widehat{\mathbf{v}}_k^T(\mathbf{D}/\mathbf{\Lambda})^{-\alpha} = -\frac{1}{2\pi i}\oint_{\mathcal{C}_k}(\mathbf{D}/\mathbf{\Lambda})^{-\alpha}(\mathbf{X} - z)^{-1}(\mathbf{D}/\mathbf{\Lambda})^{-\alpha}\mathrm{d}z, \tag{83}$$

where $\mathcal{C}_k$ represents a contour in the complex plane $\mathbb{C}$ that encloses only the eigenvalue $\widehat{\delta}_k$ and no other eigenvalues of random matrix $\mathbf{X}$, and $i = (-1)^{1/2}$ denotes the imaginary unit. By leveraging the Woodbury matrix identity, we can obtain the representation

$$\mathbf{u}^T(\mathbf{D}/\mathbf{\Lambda})^{-\alpha}\widehat{\mathbf{v}}_k\widehat{\mathbf{v}}_k^T(\mathbf{D}/\mathbf{\Lambda})^{-\alpha}\mathbf{v} = -\frac{1}{2\pi i}\oint_{\mathcal{C}_k}\mathbf{u}^T\left(\mathbf{G}(z) - \mathbf{G}(z)\mathbf{V}\frac{1}{\mathbf{\Delta}^{-1} + \mathbf{V}^T\mathbf{G}(z)\mathbf{V}}\mathbf{V}^T\mathbf{G}(z)\right)\mathbf{v}\mathrm{d}z, \tag{84}$$

which expresses the bilinear form $\mathbf{u}^T(\mathbf{D}/\mathbf{\Lambda})^{-\alpha}\widehat{\mathbf{v}}_k\widehat{\mathbf{v}}_k^T(\mathbf{D}/\mathbf{\Lambda})^{-\alpha}\mathbf{v}$ for arbitrary deterministic unit vectors $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$ in terms of resolvent $\mathbf{G}(z)$, enabling us to estimate the projection $\mathbf{u}^T(\mathbf{D}/\mathbf{\Lambda})^{-\alpha}\widehat{\mathbf{v}}_k$. To deduce the asymptotic expansion of $\mathbf{u}^T(\mathbf{D}/\mathbf{\Lambda})^{-\alpha}\widehat{\mathbf{v}}_k$, we replace all occurrences of $\mathbf{G}(z)$ in (84) with $\widetilde{\mathbf{\Upsilon}}(z)$, which provides the relationship between the projection and its asymptotic limit $\mathbf{u}^T(\mathbf{D}/\mathbf{\Lambda})^{-\alpha}\mathbf{v}_k$. Note that rather than working directly with the eigenvectors $\widehat{\mathbf{v}}_k$ of random matrix $\mathbf{X} = \mathbf{D}^{-\alpha}\widetilde{\mathbf{X}}\mathbf{D}^{-\alpha}$, the quantity $\mathbf{u}^T(\mathbf{D}/\mathbf{\Lambda})^{-\alpha}\mathbf{v}_k$ characterizes the asymptotic behavior of the transformed vectors $\mathbf{D}^{-\alpha}\widehat{\mathbf{v}}_k$. This is equivalent to studying the eigenvectors of matrix $\mathbf{D}^{-2\alpha}\widetilde{\mathbf{X}}$, which can be obtained from $\mathbf{X}$ via the similarity transformation induced by $\mathbf{D}^{\alpha}$. To derive the asymptotic expansion of the projections $\mathbf{u}^T\widehat{\mathbf{v}}_k$, we replace $\mathbf{u}$ with $(\mathbf{D}/\mathbf{\Lambda})^{\alpha}\mathbf{u}$ in (84), which leads to

$$\mathbf{u}^T\widehat{\mathbf{v}}_k\widehat{\mathbf{v}}_k^T(\mathbf{D}/\mathbf{\Lambda})^{-\alpha}\mathbf{v} = -\frac{1}{2\pi i}\oint_{\mathcal{C}_k}\mathbf{u}^T(\mathbf{D}/\mathbf{\Lambda})^{\alpha}\left(\mathbf{G}(z) - \mathbf{G}(z)\mathbf{V}\frac{1}{\mathbf{\Delta}^{-1} + \mathbf{V}^T\mathbf{G}(z)\mathbf{V}}\mathbf{V}^T\mathbf{G}(z)\right)\mathbf{v}\mathrm{d}z. \tag{85}$$

Provided that an asymptotic expansion of this modified bilinear form (with $\mathbf{v} = \mathbf{v}_k$) can be established, dividing by $\widehat{\mathbf{v}}_k^T(\mathbf{D}/\mathbf{\Lambda})^{-\alpha}\mathbf{v}_k$, as obtained in part 2) of Theorem 5, yields the desired asymptotic expansions for $\mathbf{u}^T\widehat{\mathbf{v}}_k$.

To present the detailed formulas for the asymptotic expansions and error bounds, we examine the differences between (81) and (82), as well as the error introduced when replacing $\mathbf{G}(z)$ with $\widetilde{\mathbf{\Upsilon}}(z)$ in (84). To determine the leading terms and the order of error terms, we need to characterize the asymptotic behavior of $\mathbf{u}^T(\mathbf{G}(z)-\widetilde{\mathbf{\Upsilon}}(z))\mathbf{v}$ for some deterministic unit vectors $\mathbf{u}$ and $\mathbf{v}$. We expect to get precise estimates of the form

$$|\mathbf{u}^T(\mathbf{G}(z) - \widetilde{\mathbf{\Upsilon}}(z))\mathbf{v}| \leq \epsilon_n(z, \mathbf{u}, \mathbf{v}),$$

where $\epsilon_n > 0$ is a sequence of deterministic error control parameters and small enough compared to the leading terms as random matrix size n increases. Such estimates are referred to as the *anisotropic local laws* in the RMT literature; see, e.g., Bloemendal et al. (2014); Knowles and Yin (2013, 2017). In our context, the required local laws are stated in Section D of the Supplementary Material as Theorems D.2–D.4. These theorems provide the necessary tools to establish the high-order asymptotic expansions and error bounds for the empirical spiked eigenvalues, and the components and projections of the empirical spiked eigenvectors.

To prove the local laws in our paper, we first utilize the local laws of the intermediate matrix $\mathbf{R}(z)$, which are established in Fan et al. (2022) for the specific case when $\alpha = 0$ and $\widetilde{\mathbf{X}}$ is the adjacency matrix of a random graph. Combining the methods in Fan et al. (2022); Erdős et al. (2013), we obtain the corresponding local laws of $\mathbf{R}$ under our setting of $\widetilde{\mathbf{X}}$, which are summarized in Theorem D.1 (see Section D of the Supplementary Material). Then we can derive the local laws of $\mathbf{G}$ from those of $\mathbf{R}$ by controlling the difference $\mathbf{G}-\mathbf{R}$. However, the presence of correlations between random matrices $\widetilde{\mathbf{X}}$ and $\mathbf{D}$ poses a *significant challenge* in extending the local laws of $\mathbf{R}(z)$ to those of $\mathbf{G}(z)$, particularly for Theorem 3, which requires a sufficiently precise estimate for $\mathbf{e}_i^T\mathbf{G}(z)\mathbf{v}$. To overcome such a challenge, we define the resolvent

$$\mathbf{G}_{[i]}(z) = \left(\overline{\mathbf{W}} - z(\mathbf{D}_{[i]}/\mathbf{\Lambda})^{2\alpha}\right)^{-1}, \tag{86}$$

where $\mathbf{D}_{[i]}$ with $i \in [n]$ is a random diagonal matrix with diagonal entries $(D_{[i]})_i = d_i$ and

$$(D_{[i]})_j = \Lambda_j + \sum_{s\in[n]\setminus\{i\}} W_{js}, \quad \forall j \neq i. \tag{87}$$

We derive the local law for $\mathbf{e}_i^T\mathbf{G}\mathbf{v}$ by first obtaining the corresponding local law for $\mathbf{e}_i^T\mathbf{G}_{[i]}\mathbf{v}$, and then controlling the difference between $\mathbf{G}$ and $\mathbf{G}_{[i]}$. The main motivation for this approach is the observation that when we deal with the ith row and column of $\mathbf{G}$, most correlations between $\mathbf{D}$ and $\widetilde{\mathbf{X}}$ come from the entries in the ith row and column of $\widetilde{\mathbf{X}}$. Hence, by introducing $\mathbf{D}_{[i]}$ we can reduce its correlation with $\widetilde{\mathbf{X}}$ greatly, which allows us to prove a sufficiently accurate local law for $\mathbf{e}_i^T\mathbf{G}_{[i]}\mathbf{v}$. On the other hand, the difference between $\mathbf{D}$ and $\mathbf{D}_{[i]}$ is very small, because we have removed only a single entry from $\widetilde{\mathbf{X}}$ in each entry of $\mathbf{D}_{[i]}$, leading to an asymptotically negligible difference. As a result, the difference between $\mathbf{G}_{[i]}$ and $\mathbf{G}$ is also asymptotically negligible. The details of this technical argument can be found in Lemmas D.1 and D.2 (see Section D of the Supplementary Material).

One of the major challenges in our paper is the *insufficiency* of low-order expansions for the asymptotic expansions of $\widehat{\delta}_k$ and $\mathbf{u}^T(\mathbf{D}/\mathbf{\Lambda})^{-\alpha}\widehat{\mathbf{v}}_k$ for general deterministic vector $\mathbf{u} \in \mathbb{R}^n$. During the manipulation of expressions in (81) and (84), we investigate $\mathbf{G}(z)$ through series expansion

$$\mathbf{G}(z) = \left(\overline{\mathbf{W}} - z(\mathbf{D}/\mathbf{\Lambda})^{2\alpha}\right)^{-1} = -(\mathbf{D}/\mathbf{\Lambda})^{-2\alpha}\sum_{l=0}^{\infty} z^{-(l+1)}\left(\overline{\mathbf{W}}(\mathbf{D}/\mathbf{\Lambda})^{-2\alpha}\right)^l. \tag{88}$$

Upon detailed calculations, it is found that to derive the required CLTs under the extra assumption (59), we need to truncate the series expansion (88) at $l = 3$. The inclusion of *higher-order terms* in the expansion allows us to obtain the formulas presented in Theorems 4 and 5. These formulas provide accurate enough approximations to ensure the validity of the CLTs stated in Corollaries 2 and 3.

Through the higher-order asymptotic expansions, we have also confirmed the interesting *phase transition phenomenon* discussed in Fan et al. (2022a), where the limiting distribution of $\mathbf{u}^T\widehat{\mathbf{v}}_k$ depends on the proximity of the deterministic unit vector $\mathbf{u} \in \mathbb{R}^n$ to $\mathbf{v}_k$ (modulo the sign). Qualitatively speaking, if we denote the angle between $\mathbf{u}$ and $\mathbf{v}_k$ as γ, and the angle between $\mathbf{v}_k$ and $\widehat{\mathbf{v}}_k$ as $\Delta\gamma$, then from the Taylor expansion we have that

$$\mathbf{u}^T\widehat{\mathbf{v}}_k = \cos(\gamma + \Delta\gamma) = \mathbf{u}^T\mathbf{v}_k - \sin(\gamma)\Delta\gamma - \frac{1}{2}\cos(\gamma)(\Delta\gamma)^2 + O\left((\Delta\gamma)^3\right).$$

When $\mathbf{u}$ is far away from $\mathbf{v}_k$, the leading term in the representation above is $\sin(\gamma)\Delta\gamma$, which yields a first-order variation, as stated in part 1) of Theorem 5. On the other hand, when $\mathbf{u}$ is close to $\mathbf{v}_k$, the first-order term vanishes, and the leading term becomes second-order, as stated in part 2) of Theorem 5. Such a phase transition phenomenon provides valuable insights into the asymptotic behavior of the projection of the empirical spiked eigenvector in different regimes, and our higher-order asymptotic expansion confirms and quantifies this phenomenon.

Before concluding this section, we relate our generalized Laplacian framework to the classical theory of sample correlation matrices. Let $\mathbf{Y\Sigma}^{1/2}$ be a large $n \times p$ data matrix consisting of n independent and identically distributed (i.i.d.) rows, where each row is a p-dimensional random vector with population covariance matrix $\mathbf{\Sigma}$. The sample covariance matrix is given by $\mathbf{S} = \mathbf{\Sigma}^{1/2}\mathbf{Y}^T\mathbf{Y}\mathbf{\Sigma}^{1/2}$ and the corresponding sample correlation matrix is given by

$$\mathbf{R} = \mathbf{S}_D^{-1/2}\mathbf{S}\mathbf{S}_D^{-1/2}, \tag{89}$$

where $\mathbf{S}_D$ denotes the diagonal matrix of sample variances, i.e., $(S_D)_{ii} = S_{ii}$. The population counterpart of $\mathbf{R}$ is $\mathbf{C} = \mathbf{\Sigma}_D^{-1/2}\mathbf{\Sigma}\,\mathbf{\Sigma}_D^{-1/2}$, where $\mathbf{\Sigma}_D$ is the diagonal matrix of population variances. The sample correlation matrix $\mathbf{R}$ admits a close structural analogy to our generalized Laplacian matrix $\mathbf{X}$. In the random matrix literature, it is common to assume that $\mathbf{Y}$ has independent entries, in which case the sample covariance matrix $\mathbf{S}$ exhibits asymptotic behavior similar to that of Wigner-type random matrices, analogous to the role played by the adjacency-type matrix $\widetilde{\mathbf{X}}$ in our setting. The key *distinction* lies in the normalization: the node degree matrix $\mathbf{D}$ induces an L^1-type normalization of the rows of $\widetilde{\mathbf{X}}$, whereas the diagonal matrix $\mathbf{S}_D$ induces an L^2-type normalization of the columns of $\mathbf{Y\Sigma}^{1/2}$. In both cases, this data-driven diagonal normalization introduces *nontrivial dependencies* between the normalized matrix and the underlying raw matrix, leading to significant analytical challenges. An important *open problem* is to derive precise asymptotic expansions for the spiked eigenvalues and eigenvectors of the sample correlation matrix $\mathbf{R}$, and to relate them to those of its population counterpart $\mathbf{C}$. Motivated by our generalized Laplacian framework, it is also natural to consider "*generalized sample correlation matrices*" in which $\mathbf{\Sigma}_D^{-1/2}$ is replaced with $\mathbf{\Sigma}_D^{-\alpha}$ for a general exponent $\alpha > 0$. We expect that the techniques developed in this paper will be useful in addressing these problems. A systematic investigation of these questions is left for future work.

V. Applications of ATE-GL

In this section, we discuss two applications of our ATE-GL theoretical framework for uncertainty quantification in modern statistical learning. The notation is local to this section. The examples here are intentionally stylized with the purpose of illustrating how our main theory can be propagated to downstream graph-learning quantities. Additional applications specific to network settings, such as inferring the number of communities and testing node community memberships, are discussed briefly in the Introduction and Remark 4. These problems can be addressed in a relatively straightforward manner by following the approaches outlined in the cited references therein; therefore, we omit detailed discussion here.

For simplicity, throughout this section, we assume that K_0 is fixed as $n \to \infty$ and known, $K = O(1)$, $\beta_n \sim 1$, and that the delocalization condition (41) holds. When K_0 is unknown, it can be estimated consistently using the result in Theorem 6. In the interest of space, we include only a sketch of the derivations below to showcase the applications of our core theorems; full proofs will be provided in an extended version of this work.

A. Graph neural networks

One natural application of ATE-GL is to quantify the uncertainty of the spectral graph neural network (GNN) layers (Zhang et al., 2019) that are built from the generalized Laplacian eigenvectors. The spectral GNNs implement graph convolutions by applying filters of the form $\widehat{\mathbf{V}}\mathbf{C}\widehat{\mathbf{V}}^T$ to the node features, where $\widehat{\mathbf{V}}$ collects K_0 spiked eigenvectors of a graph Laplacian or generalized Laplacian ($\mathbf{X}$ using our notation), and $\mathbf{C}$ is a diagonal matrix of learnable filter coefficients (Bruna et al., 2014). In modern applications, the underlying graphs are themselves random and often severely sparse. Most existing analyses of GNNs, however, treat the graph as deterministic, and thus cannot answer the following basic question: *How much variability in the spectral GNN features and predictions is induced purely by randomness in the graph?* Our theory can be directly applied to address this question.

For simplicity, we focus on the standard one-layer spectral graph convolution used in spectral GNNs. Specifically, let $\mathbf{F}^{(0)} = (F_{ir}^{(0)}) \in \mathbb{R}^{n\times d}$ be a fixed (exogenous) node-feature matrix, where the ith row $\mathbf{F}^{(0)}(i,:)$ is the d-dimensional feature vector of node i, and the rth column $\mathbf{F}^{(0)}(:,r)$ is the n-vector of the rth feature. Upon normalization, we assume throughout that $\max_{1\le r\le d} \|\mathbf{F}^{(0)}(:,r)\|_1 \le n$ for all n. Let us consider a single spectral GNN layer applied to $\mathbf{F}^{(0)}$ of the form

$$\widehat{\mathbf{Z}}_j \;=\; \widehat{\mathbf{V}}\,\mathrm{diag}(\mathbf{C}_j)\,\widehat{\mathbf{V}}^T\mathbf{F}^{(0)} \;\in\; \mathbb{R}^{n\times d}, \quad j = 1, \cdots, d_{\text{out}}, \tag{90}$$

where $\mathbf{C}_j = (C_{j,1}, \cdots, C_{j,K_0})^T \in \mathbb{R}^{K_0}$ collects the spectral filter coefficients for the jth output channel, and d_{out} is the number of output channels. The corresponding *population* spectral GNN layer is given by

$$\mathbf{Z}_{j,\star} \;=\; \mathbf{V}\, \text{diag}(\mathbf{C}_j)\, \mathbf{V}^T \mathbf{F}^{(0)} \;\in\; \mathbb{R}^{n\times d}, \quad j = 1, \cdots, d_{\text{out}}. \tag{91}$$

Our general ATE-GL theory can be applied to study the entrywise asymptotic behavior of the difference $\widehat{\mathbf{Z}}_j - \mathbf{Z}_{j,\star}$. In the derivations below, we condition on $\mathbf{C}_j$'s and $\mathbf{F}^{(0)}$ so that we can focus on the randomness from the underlying graph.

We now control the difference

$$\begin{aligned}\widehat{\mathbf{Z}}_j - \mathbf{Z}_{j,\star} &= \sum_{k=1}^{K_0} C_{j,k}\big(\widehat{\mathbf{v}}_k \widehat{\mathbf{v}}_k^T - \mathbf{v}_k \mathbf{v}_k^T\big)\mathbf{F}^{(0)} \\ &= \sum_{k=1}^{K_0} C_{j,k}\Big[\mathbf{v}_k(\widehat{\mathbf{v}}_k - \mathbf{v}_k)^T + (\widehat{\mathbf{v}}_k - \mathbf{v}_k)\mathbf{v}_k^T + (\widehat{\mathbf{v}}_k - \mathbf{v}_k)(\widehat{\mathbf{v}}_k - \mathbf{v}_k)^T\Big]\mathbf{F}^{(0)}.\end{aligned} \tag{92}$$

From the asymptotic expansion established in Theorem 3, for each spiked eigenpair $k \in [K_0]$ and each coordinate $i \in [n]$, we have that

$$\widehat{v}_k(i) = v_k(i) - \frac{\alpha}{\Lambda_i} v_k(i) \sum_{j=1}^n W_{ij} + \frac{1}{t_k}\sum_{j=1}^n \Lambda_i^{-\alpha} W_{ij} \Lambda_j^{-\alpha} v_k(j) + \mathcal{E}_{k,i}, \tag{93}$$

where under conditions (38), (43)–(45), and the delocalization condition (41), it holds that

$$\sigma_{k,i} = o(n^{-1/2}) \quad \text{and} \quad \max_{1\le k\le K_0} \max_{1\le i\le n} |\mathcal{E}_{k,i}| = o_{\mathbb{P}}(r_n \sigma_{k,i}). \tag{94}$$

Here, $r_n \to 0$ denotes a convergence rate such that the remainder terms $\mathcal{E}_{k,i}$ are asymptotically negligible relative to the leading terms. Now, plugging (93) into (92) yields the following expansion for the (i,r)th entry of $\widehat{\mathbf{Z}}_j - \mathbf{Z}_{j,\star}$

$$\big(\widehat{\mathbf{Z}}_j - \mathbf{Z}_{j,\star}\big)_{i,r} = \sum_{k=1}^{K_0} C_{j,k}\Big[v_k(i)\,\Gamma_{k,r} + \Delta_{k,i}\, \mathbf{v}_k^T \mathbf{F}^{(0)}(:,r)\Big] + o_{\mathbb{P}}\big(r_n(\sqrt{n}\sigma_{k,i})\big) \tag{95}$$

for all $i = 1, \ldots, n$ and $r = 1, \ldots, d$, where

$$\Gamma_{k,r} := \sum_{\ell=1}^n \Delta_{k,\ell} F_{\ell,r}^{(0)}, \quad \Delta_{k,i} := -\frac{\alpha}{\Lambda_i} v_k(i) \sum_{\ell=1}^n W_{i\ell} + \frac{1}{t_k}\sum_{\ell=1}^n \Lambda_i^{-\alpha} W_{i\ell} \Lambda_\ell^{-\alpha} v_k(\ell). \tag{96}$$

Each $\Delta_{k,i}$ is a linear functional of the noise entries $\{W_{i\ell}\}$. Hence, as in Corollary 1, under some suitable regularity conditions, the Lindeberg–Feller CLT applies and yields a one-dimensional CLT for $\Delta_{k,i}$ with variance $\sigma_{k,i}^2 = \text{Var}(\Delta_{k,i})$.

For convenience, we denote the leading linear term in (95) as

$$L_{j,i,r}(\mathbf{W}) := \sum_{k=1}^{K_0} C_{j,k}\Big[v_k(i)\,\Gamma_{k,r} + \Delta_{k,i}\, \mathbf{v}_k^T \mathbf{F}^{(0)}(:,r)\Big], \quad i \in [n],\ r \in [d].$$

For a finite node set $S \subset [n]$, let us collect these quantities into a vector

$$L_{j,S}(\mathbf{W}) := \big(L_{j,i,r}(\mathbf{W})\big)_{i\in S,\ 1\le r\le d}$$

and define its covariance matrix as

$$\Sigma_{j,S} := \text{Cov}\big(L_{j,S}(\mathbf{W})\big) \quad \text{with} \quad \big(\Sigma_{j,S}\big)_{(i,r),(i',r')} := \text{Cov}\big(L_{j,i,r}(\mathbf{W}), L_{j,i',r'}(\mathbf{W})\big). \tag{97}$$

It is seen that the covariance entries can be computed explicitly in terms of the variance profile $s_{i\ell} = \mathbb{E}|W_{i\ell}|^2$, the population spiked eigenpairs $\{(\delta_k, \mathbf{v}_k)\}_{k=1}^{K_0}$, and the feature matrix $\mathbf{F}^{(0)}$. In addition, assume that there exists a constant $C > 0$ such that

$$\|\Sigma_{j,S}^{-1/2}\| \cdot r_n(\sqrt{n}\sigma_{k,i}) \le C, \tag{98}$$

which ensures that the higher-order remainder in (95) is asymptotically negligible after normalization. We are now ready to state the multivariate CLT for the spectral GNN outputs.

Theorem 7. *Fix a finite node set $S \subset [n]$ and an output channel $j \in \{1, \ldots, d_{\text{out}}\}$. Assume that the conditions required for the eigenvalue and eigenvector CLTs in Section III hold. In addition, assume that (98) holds, $q \gg (\log n)^4$, $K = O(1)$, $\beta_n \sim 1$, and the delocalization condition (41) is satisfied. Then we have that as $n \to \infty$,*

$$\Sigma_{j,S}^{-1/2}\Big((\widehat{\mathbf{Z}}_j - \mathbf{Z}_{j,\star})_{i,r}\Big)_{i\in S,\ 1\le r\le d} \xrightarrow{d} \mathcal{N}\big(0, \mathbf{I}_{|S|d}\big).$$

Spectral GNNs apply a *pointwise* activation function (e.g., sigmoid) $\sigma : \mathbb{R} \to \mathbb{R}$ to the output $(\widehat{\mathbf{Z}}_1, \cdots, \widehat{\mathbf{Z}}_{d_{\text{out}}})^T$ and may subsequently apply a linear or nonlinear prediction head. We now discuss how the CLT result in Theorem 7 above *propagates* through such post-processing operations. For each channel $j \in \{1, \ldots, d_{\text{out}}\}$, define the post-activation outputs as

$$\widehat{\mathbf{F}}_{\text{out},j} := \sigma(\widehat{\mathbf{Z}}_j), \quad \mathbf{F}_{\text{out},j,\star} := \sigma(\mathbf{Z}_{j,\star}),$$

where activation $\sigma(\cdot)$ acts entrywise. Then for any fixed (i, r), a first-order Taylor expansion gives that

$$\widehat{\mathbf{F}}_{\text{out},j}(i,r) - \mathbf{F}_{\text{out},j,\star}(i,r) = \sigma'\big(\mathbf{Z}_{j,\star}(i,r)\big)\big(\widehat{\mathbf{Z}}_j(i,r) - \mathbf{Z}_{j,\star}(i,r)\big) + o_{\mathbb{P}}\big(|\widehat{\mathbf{Z}}_j(i,r) - \mathbf{Z}_{j,\star}(i,r)|\big).$$

Consequently, the asymptotic normality of $\widehat{\mathbf{F}}_{\text{out},j}(i,r) - \mathbf{F}_{\text{out},j,\star}(i,r)$ follows from Theorem 7 via the delta method. If we further apply a linear prediction head $\mathbf{w} \in \mathbb{R}^{d_{\text{out}}}$ followed by a Lipschitz link function g (e.g., logistic or a softmax component) to get the final output $\widehat{y}(i)$ for node i, i.e., $\widehat{y}(i) = g\big(\mathbf{w}^T\widehat{\mathbf{F}}_{\text{out}}(i,:)\big)$, then the prediction error admits an *analogous* asymptotic expansion. In particular, $\widehat{y}(i) - y_\star(i)$ is asymptotically Gaussian as a consequence of Theorem 7 and another application of the delta method.

B. Laplacian-regularized graph smoothing

The Laplacian-based smoothing/regularization is a standard tool for learning and inference on graphs, and has been used widely in graph regularization and kernels (Smola and Kondor, 2003), Gaussian-field/harmonic-function methods for semi-supervised learning (Zhu et al., 2003), and manifold regularization (Belkin et al., 2006). Assume that we observe a response vector associated with the n nodes, $\mathbf{y} = (y_i)_{i\in[n]} \in \mathbb{R}^n$, which may be fully observed or partially observed with missing entries imputed as zero. Upon normalization, we assume throughout that $\|\mathbf{y}\|_1 \le n$ for all n, and continue to use $\mathbf{X}$ to denote the generalized Laplacian matrix. Many Laplacian-regularized smoothing procedures construct a smoothed prediction vector $\widehat{\mathbf{f}} = (\widehat{f}_1, \ldots, \widehat{f}_n)^T$ via a *spectral filter* of the form

$$\widehat{\mathbf{f}} := g(\mathbf{X})\mathbf{y} = \sum_{i=1}^{n} g(\widehat{\delta}_i)\,(\widehat{\mathbf{v}}_i^T\mathbf{y})\,\widehat{\mathbf{v}}_i,$$

where g is a scalar function applied to the eigenvalues of $\mathbf{X}$. Typical choices of $g(\cdot)$ include the Tikhonov/Laplacian regression $g_{\text{Tik}}(x) = (1 + \lambda\phi(x))^{-1}$ with regularization parameter $\lambda > 0$, and the heat-kernel/diffusion filters $g_{\text{heat}}(x) = e^{-\tau\phi(x)}$ with tuning parameter $\tau > 0$, where ϕ denotes an affine transformation of the spectrum of $\mathbf{X}$ (e.g., for the normalized Laplacian, one typically takes $\phi(t) = 1 - t$). The population counterpart corresponding to $\widehat{\mathbf{f}}$ is obtained by applying the same spectral filter to $\mathbf{\Lambda}^{-\alpha}\mathbf{H}\mathbf{\Lambda}^{-\alpha}$

$$\mathbf{f}_\star := \sum_{k=1}^{K_0} g(\delta_k)\,(\mathbf{v}_k^T\mathbf{y})\,\mathbf{v}_k, \quad f_{\star,i} := \mathbf{e}_i^T\mathbf{f}_\star \quad \text{for } i \in [n]. \tag{99}$$

Let us assume that $g(\cdot)$ is twice continuously differentiable (i.e., C^2) and supported on the subset $I_\epsilon := \{x \in \mathbb{R} : |x| \ge (1+\varepsilon)|\delta_{K_0+1}|\}$, which contains the population spiked eigenvalues $\{\delta_k : 1 \le k \le K_0\}$ if $\varepsilon > 0$ is chosen smaller than constant ε_0 in the eigengap condition (22). Then by Theorem 1 and (28), with high probability set I_ϵ contains all empirical spiked eigenvalues $\{\widehat{\delta}_k : 1 \le k \le K_0\}$, while excluding all non-spiked eigenvalues $\widehat{\delta}_k$ with $k \ge K_0 + 1$.

Let $\mathbf{f}_0 = (f_{0,1}, \cdots, f_{0,n})^T$ denote the underlying true signal on the graph. The target $\mathbf{f}_\star$ separates the error of $\widehat{\mathbf{f}}$ relative to $\mathbf{f}_0$ into two conceptually different components: $\mathbf{f}_\star - \mathbf{f}_0$ and $\widehat{\mathbf{f}} - \mathbf{f}_\star$. The term $\mathbf{f}_\star - \mathbf{f}_0$ is the representation error of the chosen population filter class, which is fixed approximation cost once the filter class is chosen. The term $\widehat{\mathbf{f}} - \mathbf{f}_\star$ is the statistical estimation error caused by replacing $\mathbf{\Lambda}^{-\alpha}\mathbf{H}\mathbf{\Lambda}^{-\alpha}$ with its empirical version. The asymptotic results here quantify this latter component, namely the fluctuation of $\widehat{f}_i$ around $f_{\star,i}$.

We next discuss how to use the general ATE-GL theory to derive a CLT for the prediction error $\widehat{f}_i - f_{\star,i}$ at a fixed node i, which enables uncertainty quantification. Our derivations are based on the eigenvalue and eigenvector expansions and CLTs developed in Section III. In particular,

1) For each $k \in [K_0]$, let ς_k^2 be the variance in (49). By Corollary 2, the spiked eigenvalues satisfy the CLT

$$\frac{\widehat{\delta}_k - t_k - A_k}{t_k \cdot \varsigma_k} \xrightarrow{d} \mathcal{N}(0,1). \tag{100}$$

2) For each $k \in [K_0]$ and $i \in [n]$, let $\sigma_{k,i}^2$ be as in (42). By Corollary 1, we have the entrywise eigenvector CLT

$$\frac{\widehat{v}_k(i) - v_k(i)}{\sigma_{k,i}} \xrightarrow{d} \mathcal{N}(0,1). \tag{101}$$

More precisely, the eigenvalue and eigenvector fluctuations admit the following decompositions

$$(\widehat{\delta}_k - t_k - A_k)/t_k = L_k^{(\delta)}(\mathbf{W}) + R_k^{(\delta)}, \qquad \mathrm{Var}\big(L_k^{(\delta)}(\mathbf{W})\big) = \varsigma_k^2, \qquad R_k^{(\delta)} = o_{\mathbb{P}}(\varsigma_k), \tag{102}$$

$$\widehat{v}_k(i) - v_k(i) = L_{k,i}^{(v)}(\mathbf{W}) + R_{k,i}^{(v)}, \qquad \mathrm{Var}\big(L_{k,i}^{(v)}(\mathbf{W})\big) = \sigma_{k,i}^2, \qquad R_{k,i}^{(v)} = o_{\mathbb{P}}(\sigma_{k,i}). \tag{103}$$

Here, by Theorems 3 and 4, the leading fluctuations are given by

$$L_k^{(\delta)}(\mathbf{W}) := -2\alpha \mathbf{v}_k^T \frac{\mathbf{D}-\mathbf{\Lambda}}{\mathbf{\Lambda}} \mathbf{v}_k + \frac{1}{t_k} \mathbf{v}_k^T \overline{\mathbf{W}} \mathbf{v}_k,$$

$$L_{k,i}^{(v)}(\mathbf{W}) := -\frac{\alpha}{\Lambda_i} v_k(i) \sum_{j=1}^n W_{ij} + \frac{1}{t_k} \sum_{j=1}^n \Lambda_i^{-\alpha} W_{ij} \Lambda_j^{-\alpha} v_k(j),$$

and the remainder terms $R_k^{(\delta)}$ and $R_{k,i}^{(v)}$ collect all higher-order contributions. Furthermore, under Assumption 1, along with the conditions $K = \mathcal{O}(1)$, $\beta_n \sim 1$, and the delocalization condition (41), it holds that

$$\max_{1\le k\le K_0} \varsigma_k = o(n^{-1/2}), \qquad \max_{1\le k\le K_0} \max_{1\le i\le n} \sigma_{k,i} = o(n^{-1/2}).$$

A direct calculation shows that with high probability,

$$\widehat{f}_i - f_{\star,i} = \sum_{k=1}^{K_0} \Big[g(\widehat{\delta}_k) - g(\delta_k)\Big] (\widehat{\mathbf{v}}_k^T \mathbf{y}) \widehat{v}_k(i) + \sum_{k=1}^{K_0} g(\delta_k) \Big[(\widehat{\mathbf{v}}_k^T \mathbf{y} - \mathbf{v}_k^T \mathbf{y}) \widehat{v}_k(i) + (\widehat{v}_k(i) - v_k(i)) \mathbf{v}_k^T \mathbf{y}\Big]. \tag{104}$$

The three bracketed terms above correspond to: i) an eigenvalue-filter fluctuation $g(\widehat{\delta}_k) - g(\delta_k)$, ii) an eigenvector-response inner product fluctuation $\widehat{\mathbf{v}}_k^T \mathbf{y} - \mathbf{v}_k^T \mathbf{y}$, and iii) an eigenvector-coordinate fluctuation $\widehat{v}_k(i) - v_k(i)$, respectively. We will analyze these contributions one by one.

For each k, denote by $B_k := t_k + A_k - \delta_k$ the deterministic bias. Combining (102) with a Taylor expansion of g gives that

$$g(\widehat{\delta}_k) - g(\delta_k) = g'(\delta_k)\big(\widehat{\delta}_k - t_k - A_k\big) + g'(\delta_k) \left[B_k + t_k L_k^{(\delta)}(\mathbf{W})\right] + o_{\mathbb{P}}(|\delta_k| \varsigma_k). \tag{105}$$

Using the expansion from Theorem 3 coupled with the conditions $\|\mathbf{y}\|_1 < n$, (41), and (94), we can deduce that

$$\widehat{\mathbf{v}}_k^T \mathbf{y} - \mathbf{v}_k^T \mathbf{y} = \sum_{j=1}^n y_j (\widehat{v}_k(j) - v_k(j)) = L_k^{(v,1)}(\mathbf{W}) + R_k^{(v,1)}, \tag{106}$$

$$\widehat{v}_k(i) - v_k(i) = L_{k,i}^{(v)}(\mathbf{W}) + R_{k,i}^{(v)}, \tag{107}$$

where $L_k^{(v,1)}$ and $L_{k,i}^{(v)}$ are linear functionals of $\mathbf{W}$ with variances of order $\sum_j y_j^2 \sigma_{k,j}^2$ and $\sigma_{k,i}^2$, respectively. Moreover, it holds that

$$\sup_k |R_k^{(v,1)}| = o_{\mathbb{P}}\big(n \cdot \max_j \sigma_{k,j}\big), \quad \sup_{k,i} |R_{k,i}^{(v)}| = o_{\mathbb{P}}\big(n \cdot \max_i \sigma_{k,i}\big).$$

It can be shown that their contribution to (104) is of order $o_{\mathbb{P}}(\sqrt{n} \cdot \max_{1\le k\le K_0}(\varsigma_k + \max_i \sigma_{k,i}))$. Substituting (105) together with the eigenvector expansions (106) and (107) into (104), it follows that

$$\widehat{f}_i - f_{\star,i} = \underbrace{\sum_{k=1}^{K_0} g'(\delta_k)(\mathbf{v}_k^T \mathbf{y}) B_k v_k(i)}_{=: B_{g,i}} + \sum_{k=1}^{K_0} g'(\delta_k)(\mathbf{v}_k^T \mathbf{y})\big(\widehat{\delta}_k - t_k - A_k\big) v_k(i)$$

$$+\sum_{k=1}^{K_0} g(\delta_k) \sum_{\ell=1}^{n} y_\ell \big(\widehat{v}_k(\ell) - v_k(\ell)\big) v_k(i) + \sum_{k=1}^{K_0} g(\delta_k)(\mathbf{v}_k^T \mathbf{y})\big(\widehat{v}_k(i) - v_k(i)\big)$$
$$+ o_{\mathbb{P}}\Big(\sqrt{n} \cdot \max_{1\le k\le K_0} (\varsigma_k + \max_i \sigma_{k,i})\Big). \tag{108}$$

Here, $B_{g,i}$ is a deterministic bias term, which does not vanish in general. Let us define for each $k \in [K_0]$ and $i \in [n]$,

$$T_{k,i}^{(\mathrm{val})} := g'(\delta_k)(\mathbf{v}_k^T \mathbf{y})\big(\widehat{\delta}_k - t_k - A_k\big) v_k(i),$$
$$T_{k,i}^{(\mathrm{vec},1)} := g(\delta_k) \sum_{\ell=1}^{n} y_\ell \big(\widehat{v}_k(\ell) - v_k(\ell)\big) v_k(i),$$
$$T_{k,i}^{(\mathrm{vec},2)} := g(\delta_k)(\mathbf{v}_k^T \mathbf{y})\big(\widehat{v}_k(i) - v_k(i)\big).$$

Then equation (108) can be rewritten as

$$\widehat{f}_i - f_{\star,i} = B_{g,i} + \sum_{k=1}^{K_0} \Big(T_{k,i}^{(\mathrm{val})} + T_{k,i}^{(\mathrm{vec},1)} + T_{k,i}^{(\mathrm{vec},2)}\Big) + o_{\mathbb{P}}\Big(\sqrt{n} \cdot \max_{1\le k\le K_0} (\varsigma_k + \max_i \sigma_{k,i})\Big).$$

Now, substituting (102) and (103) into the definitions of $T_{k,i}^{(\mathrm{val})}$, $T_{k,i}^{(\mathrm{vec},1)}$, and $T_{k,i}^{(\mathrm{vec},2)}$, we can obtain that

$$\widehat{f}_i - f_{\star,i} = B_{g,i} + L_i(\mathbf{W}) + o_{\mathbb{P}}\Big(\sqrt{n} \cdot \max_{1\le k\le K_0} (\varsigma_k + \max_i \sigma_{k,i})\Big). \tag{109}$$

Here, $L_i(\mathbf{W}) = \sum_{1\le a\le b\le n} c_{ab}^{(i)} W_{ab}$ is a centered linear functional of $\mathbf{W}$, where the coefficients $c_{ab}^{(i)}$ depend deterministically on $(\{v_k, \delta_k\}_{k=1}^{K_0}, g, \mathbf{y}, \Lambda)$ and on the variance profile s_{ab}

$$c_{ab}^{(i)} := \sum_{k=1}^{K_0} g'(\delta_k)(\mathbf{v}_k^T \mathbf{y}) \chi_{ab}^{(k)} + \sum_{k=1}^{K_0} g(\delta_k) \Big[v_k(i) \sum_{\ell=1}^{n} y_\ell\, \xi_{ab}^{(k,\ell)} + (\mathbf{v}_k^T \mathbf{y})\, \xi_{ab}^{(k,i)}\Big]$$

with

$$\chi_{ab}^{(k)} := \frac{t_k}{1 + \mathbf{1}\{a=b\}} S_{ab}^{(\mathbf{v})} \quad \text{with } S_{ab}^{(\mathbf{v})} := -2\alpha\Big(\frac{v_k(a)^2}{\Lambda_a} + \frac{v_k(b)^2}{\Lambda_b}\Big) + \frac{1}{t_k} \frac{2 v_k(a) v_k(b)}{(\Lambda_a \Lambda_b)^\alpha},$$
$$\xi_{ab}^{(k,i)} = \begin{cases} -\alpha\Big(\frac{v_k(a)}{\Lambda_a}\mathbf{1}\{i=a\} + \frac{v_k(b)}{\Lambda_b}\mathbf{1}\{i=b\}\Big) + \frac{\Lambda_a^{-\alpha}\Lambda_b^{-\alpha}}{t_k}\big[v_k(b)\mathbf{1}\{i=a\} + v_k(a)\mathbf{1}\{i=b\}\big], & a<b, \\ \Big(-\frac{\alpha}{\Lambda_a} v_k(a) + \frac{\Lambda_a^{-2\alpha}}{t_k} v_k(a)\Big)\mathbf{1}\{i=a\}, & a=b. \end{cases}$$

Fix a finite subset $S \subset [n]$. Let us collect the linear terms $L_i(\mathbf{W})$'s into a vector $\mathbf{L}_S(\mathbf{W}) := \big(L_i(\mathbf{W})\big)_{i\in S}$, and define its covariance matrix as

$$\big(\Sigma_S\big)_{i,i'} := \sum_{1\le a\le b\le n} s_{ab}\, c_{ab}^{(i)} c_{ab}^{(i')}, \qquad i, i' \in S.$$

Assume further that there exists a constant $C > 0$ such that

$$\|\Sigma_S^{-1/2}\| \cdot \sqrt{n} \max_{1\le k\le K_0} (\varsigma_k + \max_i \sigma_{k,i}) \le C, \tag{110}$$

which ensures that the remainder term in (109) is asymptotically negligible after normalization. Under these conditions, we are now ready to establish a multivariate CLT for the centered prediction errors $\big(\widehat{f}_i - f_{\star,i} - B_{g,i}\big)_{i\in S}$ as $n \to \infty$.

Theorem 8 (Asymptotic normality of Laplacian smoothing)**.** *Fix a finite node set $S \subset [n]$. Assume that the conditions required for the eigenvalue and eigenvector CLTs in Section III hold. In addition, assume that (110) holds, $q \gg (\log n)^4$, $K = O(1)$, $\beta_n \sim 1$, and the delocalization condition (41) is satisfied. Let $g \in C^2(\mathbb{R})$ be supported on $I_\epsilon = \{x \in \mathbb{R} : |x| \ge (1+\varepsilon)|\delta_{K_0+1}|\}$. Then the vector of prediction errors $\big(\widehat{f}_i - f_{\star,i}\big)_{i\in S}$ satisfies that as $n \to \infty$,*

$$\Sigma_S^{-1/2} \big(\widehat{f}_i - f_{\star,i} - B_{g,i}\big)_{i\in S} \xrightarrow{d} \mathcal{N}(0, \mathbf{I}_{|S|}),$$

where the (deterministic) bias term is given by $B_{g,i} = \sum_{k=1}^{K_0} g'(\delta_k)(\mathbf{v}_k^T \mathbf{y})(t_k + A_k - \delta_k) v_k(i)$ with A_k defined *in* (53).

VI. Simulation study

In this section, we conduct a simulation study to verify the asymptotic distributions of the empirical spiked eigenvalues and eigenvectors of the generalized Laplacian matrices established in Section III.

A. Simulation settings

We now describe the simulation design for the generalized Laplacian matrix $\mathbf{X}$. Following Simulation Example 1 in Fan et al. (2022), we first generate an $n \times n$ symmetric random matrix $\widetilde{\mathbf{X}}$ with independent entries up to symmetry according to the model in (5). This setting is based on the widely used mixed membership (MM) model for random adjacency matrices, under which

$$\mathbf{H} := \mathbb{E}[\widetilde{\mathbf{X}}] = \theta \mathbf{\Pi} \mathbf{P} \mathbf{\Pi}^T \in \mathbb{R}^{n \times n}, \tag{111}$$

which is a special case of (2) with $\mathbf{\Theta} = \sqrt{\theta}\mathbf{I}$. This model was adopted in Fan et al. (2022) to study group network testing under non-sharp nulls and weak signals; see Section 5 therein for further details. Specifically, we consider a network of size $n = 3000$ with $K = 5$ communities, where each community contains $n_0 = 300$ pure nodes. Each pure node in the kth community ($1 \leq k \leq K$) has a community membership probability vector $\boldsymbol{\pi} = \mathbf{e}_k$, which is the kth basis vector. The remaining $n - Kn_0$ nodes are divided into four groups of equal size and are treated as mixed (non-pure) nodes. As in Fan et al. (2022), the community membership probability vector for each mixed node in group l ($1 \leq l \leq 4$) is given by $\boldsymbol{\pi} = \mathbf{a}_l$, where $\mathbf{a}_1 = (0.1, 0.6, 0.1, 0.1, 0.1)^T$, $\mathbf{a}_2 = (0.6, 0.1, 0.1, 0.1, 0.1)^T$, $\mathbf{a}_3 = (0.1, 0.1, 0.6, 0.1, 0.1)^T$, and $\mathbf{a}_4 = (1/K, \cdots, 1/K)^T$. This fully specifies the $n \times K$ matrix of community membership probability vectors $\mathbf{\Pi}$ associated with the MM model. It remains to specify the kernel matrix $\mathbf{P}$ associated with the MM model, which is a $K \times K$ nonsingular matrix. We set the diagonal entries of $\mathbf{P}$ to be one, and define its (j,k)th entry as $\rho/|j-k|$ for each $1 \leq j \neq k \leq K$, with $\rho = 0.2$. Finally, we let the sparsity parameter θ vary in $\{0.1, 0.5, 0.9\}$, where smaller values of θ correspond to lower average node degrees and hence weaker signal strength. This completes the specification of the $n \times n$ symmetric random adjacency matrix $\widetilde{\mathbf{X}}$ in (5), with mean matrix given by (111).

We next define the generalized Laplacian matrix $\mathbf{X} = \mathbf{D}^{-\alpha}\widetilde{\mathbf{X}}\mathbf{D}^{-\alpha}$ as in (4), where $\mathbf{D} = \mathrm{diag}\,(d_i : i \in [n])$ is the diagonal degree matrix defined in (8). For *numerical stability*, we introduce a small regularization and instead take $\mathbf{D} = \mathrm{diag}\,(d_i + \lambda : i \in [n])$ with $\lambda = 10^{-4}$, ensuring that $\mathbf{D}$ is nonsingular. To investigate the finite-sample performance of the empirical spiked eigenvalues $\widehat{\delta}_k$'s and spiked eigenvectors $\widehat{\mathbf{v}}_k$'s of $\mathbf{X}$, we generate 500 independent data sets for each combination of (α, θ), where $\alpha \in \{0.25, 1/2, 1, 2\}$ and $\theta \in \{0.1, 0.5, 0.9\}$.

TABLE I: Means and standard deviations (SDs) of the empirical spiked eigenvalue $\widehat{\delta}_k$ (with $k = 1$) corrected by the theoretical quantity A_k. Results are based on 500 replications for the simulation setting in Section VI-A.

α	θ	Empirical Eigenvalue	Asymptotic Eigenvalue	Empirical SD	Asymptotic SD
0.25	0.1	9.7617	9.7592	0.0121	0.0127
	0.5	21.6409	21.6406	0.0118	0.0115
	0.9	29.0056	29.0063	0.0095	0.0100
0.5	0.1	1.0101	1.0100	2.69E-09	2.65E-05
	0.5	1.0016	1.0016	2.37E-10	1.76E-06
	0.9	1.0006	1.0006	7.75E-11	4.89E-07
1	0.1	0.0110	0.0110	2.74E-05	2.94E-05
	0.5	0.0022	0.0022	2.43E-06	2.33E-06
	0.9	0.0012	0.0012	7.86E-07	8.46E-07
2	0.1	1.45E-06	1.45E-06	1.79E-08	1.61E-08
	0.5	1.14E-08	1.14E-08	5.50E-11	5.08E-11
	0.9	1.96E-09	1.96E-09	5.37E-12	5.65E-12

B. Simulation results

The simulation results for the empirical spiked eigenvalues $\widehat{\delta}_k$'s and spiked eigenvectors $\widehat{\mathbf{v}}_k$'s of the generalized Laplacian matrix $\mathbf{X}$ under the setting of Section VI-A are summarized in Figures 2–7 and Tables I–VI.

TABLE II: Means and SDs of the empirical spiked eigenvalue $\widehat{\delta}_k$ (with $k = 2$) corrected by the theoretical quantity A_k. Results are based on 500 replications for the simulation setting in Section VI-A.

α	θ	Empirical Eigenvalue	Asymptotic Eigenvalue	Empirical SD	Asymptotic SD
0.25	0.1	4.1352	4.1090	0.0230	0.0233
	0.5	8.7895	8.7882	0.0202	0.0201
	0.9	11.7246	11.7248	0.0160	0.0162
0.5	0.1	0.4450	0.4427	0.0024	0.0025
	0.5	0.4238	0.4237	0.0010	0.0010
	0.9	0.4213	0.4214	0.0006	0.0006
1	0.1	0.0055	0.0055	4.59E-05	4.66E-05
	0.5	0.0010	0.0010	3.92E-06	3.87E-06
	0.9	0.0006	0.0006	1.43E-06	1.51E-06
2	0.1	8.43E-07	8.33E-07	1.40E-08	1.24E-08
	0.5	6.39E-09	6.39E-09	4.26E-11	4.07E-11
	0.9	1.09E-09	1.09E-09	4.74E-12	4.90E-12

TABLE III: Means and SDs of the empirical spiked eigenvalue $\widehat{\delta}_k$ (with $k = 3$) corrected by the theoretical quantity A_k. Results are based on 500 replications for the simulation setting in Section VI-A.

α	θ	Empirical Eigenvalue	Asymptotic Eigenvalue	Empirical SD	Asymptotic SD
0.25	0.1	3.6481	3.6124	0.0239	0.0246
	0.5	7.6343	7.6327	0.0208	0.0210
	0.9	10.1666	10.1650	0.0163	0.0166
0.5	0.1	0.3876	0.3840	0.0024	0.0025
	0.5	0.3631	0.3631	0.0009	0.0010
	0.9	0.3605	0.3604	0.0006	0.0006
1	0.1	0.0044	0.0043	3.16E-05	3.18E-05
	0.5	0.0008	0.0008	2.65E-06	2.65E-06
	0.9	0.0004	0.0004	1.03E-06	1.03E-06
2	0.1	5.57E-07	5.76E-07	1.99E-07	9.53E-09
	0.5	4.35E-09	4.35E-09	3.31E-11	3.05E-11
	0.9	7.40E-10	7.40E-10	3.65E-12	3.65E-12

TABLE IV: Means and SDs of the empirical spiked eigenvector component $v_k(i)$ (with $k = 1$ and $i = 1$). Results are based on 500 replications for the simulation setting in Section VI-A.

α	θ	Empirical Eigenvector	Asymptotic Eigenvector	Empirical SD	Asymptotic SD
0.25	0.1	-0.01094	-0.01907	0.00115	0.00188
	0.5	-0.01647	-0.01907	0.00066	0.00078
	0.9	-0.01910	-0.01907	0.00054	0.00052
0.5	0.1	-0.00621	-0.01867	0.00078	0.00184
	0.5	-0.01393	-0.01867	0.00059	0.00076
	0.9	-0.01870	-0.01867	0.00053	0.00051
1	0.1	-0.00196	-0.01755	0.00068	0.00176
	0.5	-0.00977	-0.01755	0.00053	0.00073
	0.9	-0.01756	-0.01755	0.00051	0.00050
2	0.1	-0.00018	-0.01316	0.00065	0.00163
	0.5	-0.00405	-0.01316	0.00048	0.00070
	0.9	-0.01310	-0.01316	0.00051	0.00049

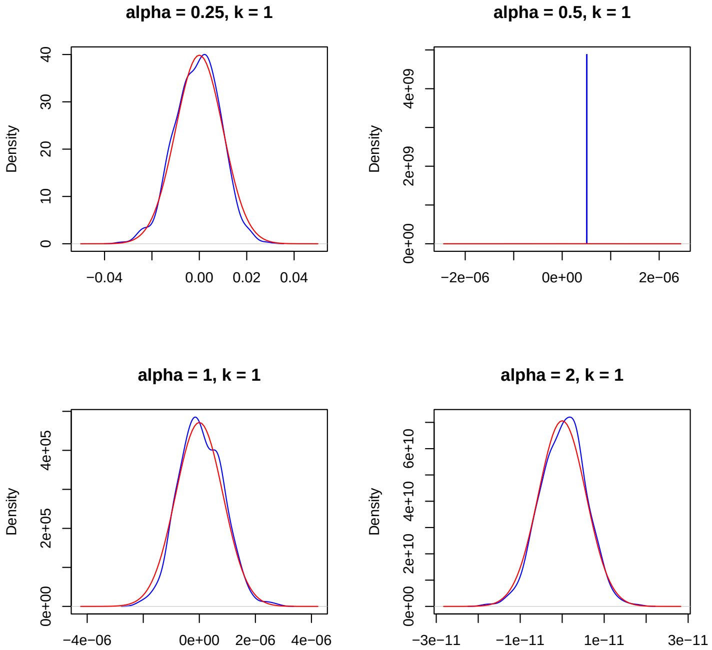

Fig. 2: Kernel density estimates (KDEs) of the distribution of the empirical spiked eigenvalue $\widehat{\delta}_k$ (with $k = 1$) corrected by the theoretical quantity A_k. Results are based on 500 replications for the simulation setting in Section VI-A with $\theta = 0.9$. Both the empirical KDEs (blue) and the target normal densities (red) are centered at the asymptotic limit t_k. The relatively concentrated behavior observed in the top-right panel is due to the extremely small standard deviation (Table I reports an empirical SD = 7.75E-11 versus an asymptotic SD = 4.89E-07), which stems from the fact that the normalized Laplacian has a trivial largest eigenvalue equal to 1.

TABLE V: Means and SDs of the empirical spiked eigenvector component $v_k(i)$ (with $k = 2$ and $i = 1$). Results are based on 500 replications for the simulation setting in Section VI-A.

α	θ	Empirical Eigenvector	Asymptotic Eigenvector	Empirical SD	Asymptotic SD
0.25	0.1	-0.00285	-0.00548	0.00244	0.00396
	0.5	-0.00472	-0.00548	0.00154	0.00174
	0.9	-0.00552	-0.00548	0.00121	0.00121
0.5	0.1	-0.00194	-0.00633	0.00134	0.00380
	0.5	-0.00467	-0.00633	0.00128	0.00167
	0.9	-0.00637	-0.00633	0.00116	0.00116
1	0.1	-0.00082	-0.00777	0.00044	0.00345
	0.5	-0.00425	-0.00777	0.00088	0.00153
	0.9	-0.00780	-0.00777	0.00107	0.00107
2	0.1	-0.00013	-0.01166	0.00018	0.00285
	0.5	-0.00358	-0.01166	0.00052	0.00126
	0.9	-0.01169	-0.01166	0.00088	0.00089

Figures 2–4 depict the distributions of the empirical spiked eigenvalues $\widehat{\delta}_k$ ($1 \leq k \leq 3$), after bias correction by the theoretical quantities A_k, for different values of α in the representative case of $\theta = 0.9$. Each distribution is centered at the corresponding asymptotic limit t_k. As shown in Figures 2–4, the bias-corrected empirical spiked eigenvalues $\widehat{\delta}_k - A_k$ follow closely the target asymptotic distributions established in Corollary 2[†]. We also implemented bias correction using the estimated quantities $\widehat{A}_k$ defined in (78), while still centering at the

[†]In contrast, we observe a noticeable bias in the raw empirical eigenvalues $\widehat{\delta}_k$ without correction, even in relatively dense networks (i.e., for larger values of θ).

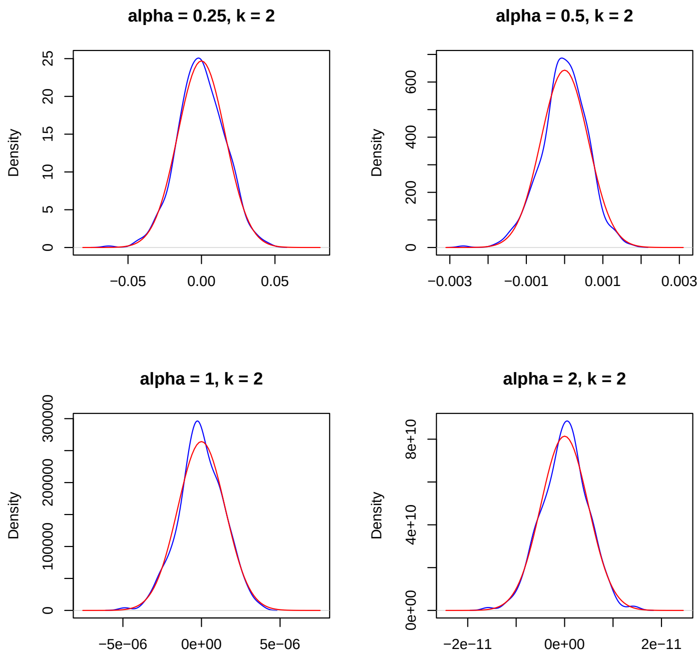

Fig. 3: KDEs of the distribution of the empirical spiked eigenvalue $\widehat{\delta}_k$ (with $k = 2$) corrected by the theoretical quantity A_k. Results are based on 500 replications for the simulation setting in Section VI-A with $\theta = 0.9$. Both the empirical KDEs (blue) and the target normal densities (red) are centered at the asymptotic limit t_k.

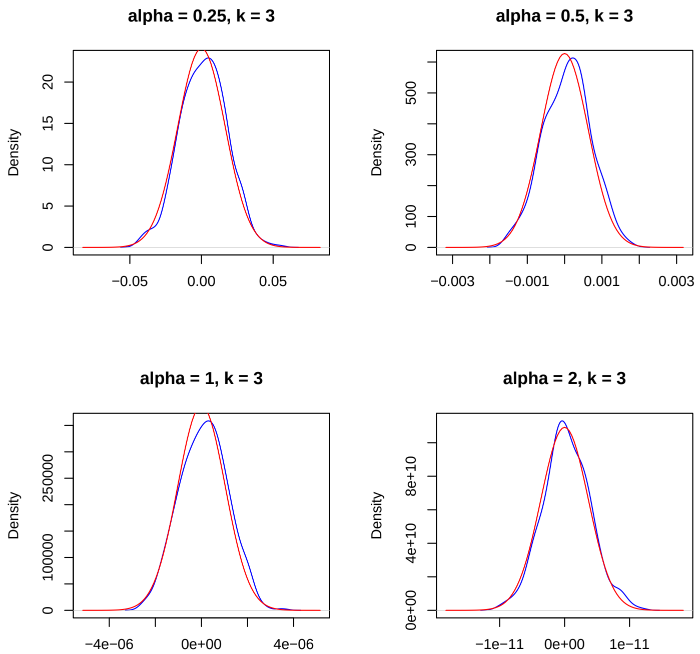

Fig. 4: KDEs of the distribution of the empirical spiked eigenvalue $\widehat{\delta}_k$ (with $k = 3$) corrected by the theoretical quantity A_k. Results are based on 500 replications for the simulation setting in Section VI-A with $\theta = 0.9$. Both the empirical KDEs (blue) and the target normal densities (red) are centered at the asymptotic limit t_k.

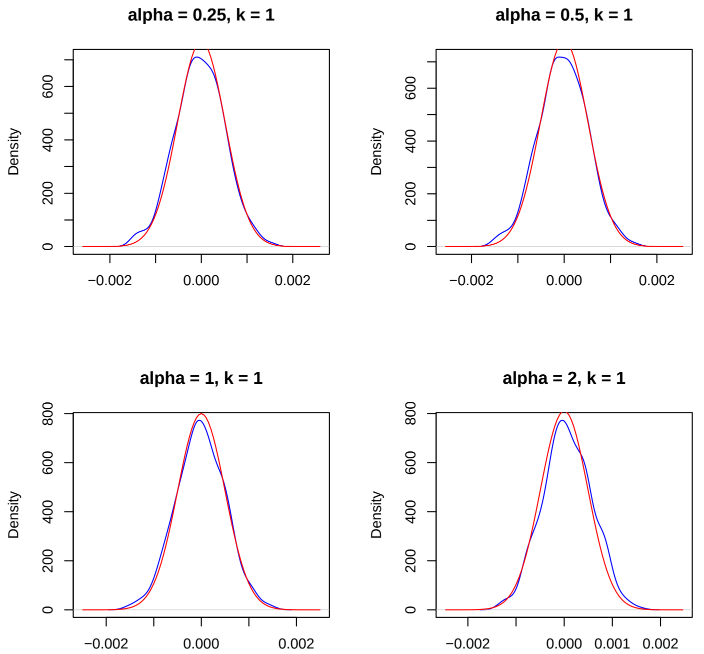

Fig. 5: KDEs of the distribution of the empirical spiked eigenvector component $\widehat{v}_k(i)$ (with $k = 1$ and $i = 1$). Results are based on 500 replications for the simulation setting in Section VI-A with $\theta = 0.9$. Both the empirical KDEs (blue) and the target normal densities (red) are centered at the asymptotic limit $v_k(i)$.

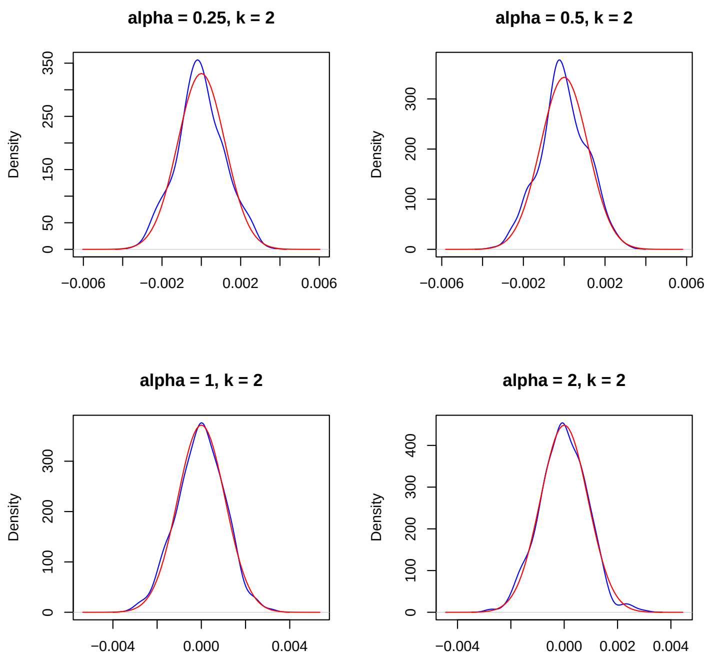

Fig. 6: KDEs of the distribution of the empirical spiked eigenvector component $\widehat{v}_k(i)$ (with $k = 2$ and $i = 1$). Results are based on 500 replications for the simulation setting in Section VI-A with $\theta = 0.9$. Both the empirical KDEs (blue) and the target normal densities (red) are centered at the asymptotic limit $v_k(i)$.

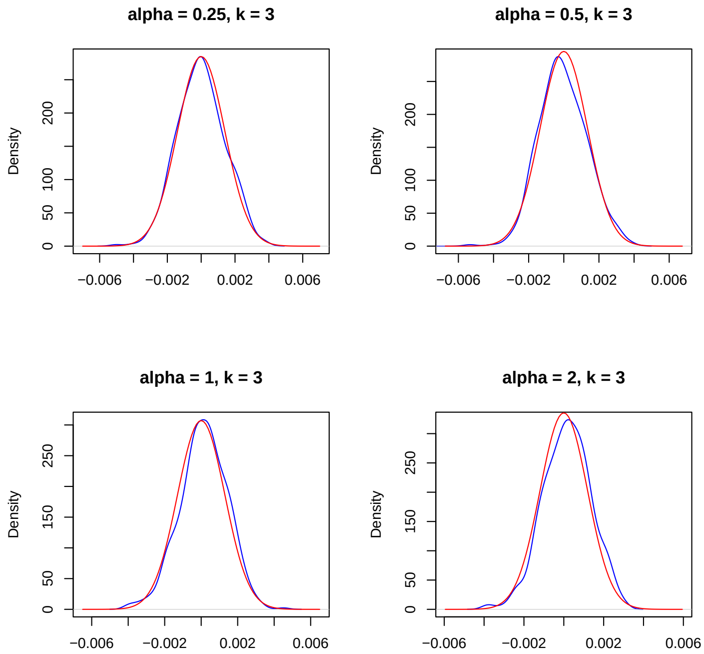

Fig. 7: KDEs of the distribution of the empirical spiked eigenvector component $\widehat{v}_k(i)$ (with $k = 3$ and $i = 1$). Results are based on 500 replications for the simulation setting in Section VI-A with $\theta = 0.9$. Both the empirical KDEs (blue) and the target normal densities (red) are centered at the asymptotic limit $v_k(i)$.

TABLE VI: Means and SDs of the empirical spiked eigenvector component $v_k(i)$ (with $k = 3$ and $i = 1$). Results are based on 500 replications for the simulation setting in Section VI-A.

α	θ	Empirical Eigenvector	Asymptotic Eigenvector	Empirical SD	Asymptotic SD
0.25	0.1	0.00476	0.00848	0.00300	0.00468
	0.5	0.00718	0.00848	0.00187	0.00205
	0.9	0.00848	0.00848	0.00140	0.00140
0.5	0.1	0.00057	0.00519	0.00216	0.00432
	0.5	0.00295	0.00519	0.00267	0.00192
	0.9	0.00496	0.00519	0.00190	0.00135
1	0.1	-0.00002	-0.00083	0.00049	0.00405
	0.5	-0.00029	-0.00083	0.00102	0.00183
	0.9	-0.00070	-0.00083	0.00130	0.00130
2	0.1	0.00006	0.00459	0.00008	0.00371
	0.5	0.00150	0.00459	0.00054	0.00167
	0.9	0.00475	0.00459	0.00120	0.00119

asymptotic limits t_k. The resulting distributions are very similar to those in Figures 2–4; see Figures 11–13 in Section F of the Supplementary Material for details. In addition, we examine an alternative correction strategy that combines the estimated bias $\widehat{A}_k$ with the empirical bias-corrected estimator in (79), targeting at the asymptotic limits δ_k directly. This approach also performs well across different parameter settings; see Figures 14–16 in Section F of the Supplementary Material. Overall, these simulation results showcase the advantages of *both bias-correction ideas* suggested in Section III-C.

Figures 5–7 display the distributions of the empirical spiked eigenvector components $\widehat{v}_k(i)$ $(1 \leq k \leq 3)$ across different values of α, for the representative case of $\theta = 0.9$. Each distribution curve has been centered by the corresponding asymptotic limit $v_k(i)$ as given in Corollary 1. For simplicity, we examine only the representative scenario of $i = 1$. It is interesting to observe that the distributions of the empirical spiked eigenvector components match closely the target asymptotic distributions predicted by Corollary 1.

Tables I–III report the means and standard deviations (SDs) of the empirical spiked eigenvalues $\widehat{\delta}_k$ in comparison to their theoretical (asymptotic) counterparts from Corollary 2, while Tables IV–VI present the means and SDs of the empirical spiked eigenvector components $\widehat{v}_k(i)$ in comparison to their theoretical limits from Corollary 1, across different choices of (α, θ). From these tables, we observe that the asymptotic theory developed in Section III for the empirical spiked eigenvalues and eigenvectors of generalized Laplacian matrices remains largely accurate at the finite-sample level. In particular, the accuracy—both in terms of the mean and variance—improves as the network sparsity parameter θ increases, which is intuitive since larger θ enhances the signal strength in the network model. Overall, these results demonstrate that our asymptotic theory for the generalized Laplacian matrix is *robust* across different values of index $\alpha \in (0, \infty)$, highlighting its practical flexibility and applicability.

VII. Discussions

In this paper, we studied the problem of extending latent embeddings for graphs and manifolds via generalized Laplacian matrices—a class of random matrices that contains both the standard Laplacian and the adjacency matrix as special cases. This class provides flexibility for extracting underlying latent structures in practical applications, while introducing nontrivial theoretical challenges due to the intrinsic dependencies in the matrices. By leveraging generalized quadratic vector equations and local laws, we derived the asymptotic distributions for both empirical spiked eigenvectors and eigenvalues. The proposed ATE-GL framework enables practical, flexible inference and rigorous uncertainty quantification for latent embeddings based on generalized Laplacian matrices.

To streamline the technical analysis, we have focused on unnormalized random matrices with independent entries (up to symmetry). It would be interesting to extend this framework to random matrices $\widetilde{\mathbf{X}}$ with dependencies, which would induce stronger dependencies in the corresponding generalized Laplacian matrices. From a practical perspective, investigating rank inference under the ATE-GL framework is another promising direction. Additionally, eigenvector selection for downstream tasks—such as clustering or local manifold learning—presents an important problem. For specific applications, identifying the optimal parameter $\alpha \in (0, \infty)$ within the ATE-GL framework also warrants further study. These topics are beyond the scope of the current paper but represent interesting directions for future research.

References

Abbe, E. (2017). Community detection and stochastic block models: recent developments. *Journal of Machine Learning Research 18*(1), 6446–6531.

Abbe, E., J. Fan, K. Wang, and Y. Zhong (2020). Entrywise eigenvector analysis of random matrices with low expected rank. *The Annals of statistics 48*(3), 1452–1474.

Ajanki, O. H., L. Erdős, and T. Krüger (2017). Universality for general Wigner-type matrices. *Probability Theory and Related Fields 169*, 667–727.

Ajanki, O. H., L. Erdős, and T. H. Krüger (2019). Stability of the matrix Dyson equation and random matrices with correlations. *Probability Theory and Related Fields 173*(1-2), 293–373.

Amini, A. A., A. Chen, P. J. Bickel, and E. Levina (2013). Pseudo-likelihood methods for community detection in large sparse networks. *The Annals of Statistics 41*(4), 2097–2122.

Anderson, G. W., A. Guionnet, and O. Zeitouni (2010). *An Introduction to Random Matrices*. Number 118. Cambridge University Press.

Arias-Castro, E. and N. Verzelen (2014). Community detection in dense random networks. *The Annals of Statistics 42*(3), 940–969.

Bai, Z. and J. W. Silverstein (2006). *Spectral Analysis of Large Dimensional Random Matrices*, Volume 20. Springer.

Baik, J., G. Ben Arous, and S. Péché (2005). Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices. *The Annals of Probability 33*(5), 1643–1697.

Baik, J. and J. W. Silverstein (2006). Eigenvalues of large sample covariance matrices of spiked population models. *Journal of Multivariate Analysis 97*(6), 1382–1408.

Bao, Z., X. Ding, J. Wang, and K. Wang (2022). Statistical inference for principal components of spiked covariance matrices. *The Annals of Statistics 50*(2), 1144–1169.

Bao, Z., X. Ding, and K. Wang (2021). Singular vector and singular subspace distribution for the matrix denoising model. *The Annals of Statistics 49*(1), 370–392.

Belkin, M., P. Niyogi, and V. Sindhwani (2006). Manifold regularization: A geometric framework for learning from labeled and unlabeled examples. *Journal of Machine Learning Research 7*, 2399–2434.

Benaych-Georges, F. and R. R. Nadakuditi (2011). The eigenvalues and eigenvectors of finite, low rank perturbations of large random matrices. *Advances in Mathematics 227*(1), 494–521.

Bhattacharya, S., J. Fan, and J. Hou (2023). Inferences on mixing probabilities and ranking in mixed-membership models. *arXiv preprint arXiv:2308.14988*.

Bickel, P. J. and P. Sarkar (2016). Hypothesis testing for automated community detection in networks. *Journal of the Royal Statistical Society Series B 78*(1), 253–273.

Bloemendal, A., L. Erdős, A. Knowles, H.-T. Yau, and J. Yin (2014). Isotropic local laws for sample covariance and generalized Wigner matrices. *Electronic Journal of Probability 19*, 1–53.

Bloemendal, A., A. Knowles, H.-T. Yau, and J. Yin (2016). On the principal components of sample covariance matrices. *Probability Theory and Related Fields 164*(1), 459–552.

Bollobás, B. (2001). *Random Graphs* (2 ed.). Cambridge Studies in Advanced Mathematics. Cambridge University Press.

Braunstein, S. L., S. Ghosh, and S. Severini (2006). The Laplacian of a graph as a density matrix: a basic combinatorial approach to separability of mixed states. *Annals of Combinatorics 10*, 291–317.

Bruna, J., W. Zaremba, A. Szlam, and Y. LeCun (2014). Spectral networks and locally connected networks on graphs.

Capitaine, M. and C. Donati-Martin (2018). Non universality of fluctuations of outlier eigenvectors for block diagonal deformations of wigner matrices. *arXiv preprint arXiv:1807.07773*.

Chaudhuri, K., F. Chung, and A. Tsiatas (2012). Spectral clustering of graphs with general degrees in the extended planted partition model. *Journal of Machine Learning Research 23*, 35.1–35.23.

Chung, F. R. (1997). *Spectral Graph Theory*, Volume 92. American Mathematical Society.

Chung, K. L. (2001). *A Course in Probability Theory*. Academic Press.

Ding, X. and H.-T. Wu (2020). On the spectral property of kernel-based sensor fusion algorithms of high dimensional data. *IEEE Transactions on Information Theory 67*(1), 640–670.

Ding, X. and F. Yang (2021). Spiked separable covariance matrices and principal components. *The Annals of Statistics 49*(2), 1113–1138.

Donetti, L., F. Neri, and M. A. Munoz (2006). Optimal network topologies: expanders, cages, Ramanujan graphs, entangled networks and all that. *Journal of Statistical Mechanics: Theory and Experiment 2006*(08), P08007.

Erdős, L., A. Knowles, H.-T. Yau, and J. Yin (2013). Spectral statistics of Erdős–Rényi graphs I: local semicircle law. *The Annals of Probability 41*, 2279–2375.

Erdős, L. and H.-T. Yau (2017). *A Dynamical Approach to Random Matrix Theory*, Volume 28. American Mathematical Society.

Estrada, E. and N. Hatano (2010). Topological atomic displacements, Kirchhoff and Wiener indices of molecules. *Chemical Physics Letters 486*, 166–170.

Fan, J., Y. Fan, X. Han, and J. Lv (2022a). Asymptotic theory of eigenvectors for random matrices with diverging spikes. *Journal of the American Statistical Association 117*, 996–1009.

Fan, J., Y. Fan, X. Han, and J. Lv (2022b). SIMPLE: statistical inference on membership profiles in large networks. *Journal of the Royal Statistical Society Series B 84*, 630–653.

Fan, J., Y. Fan, J. Lv, and F. Yang (2022). SIMPLE-RC: group network inference with non-sharp nulls and weak signals. *arXiv preprint arXiv:2211.00128*.

Fan, J., Y. Fan, J. Lv, F. Yang, and D. Yu (2025). Asymptotic theory of eigenvectors for latent embeddings with generalized Laplacian matrices. *arXiv preprint arXiv:2503.00640*.

Freschi, V. (2011). Improved biological network reconstruction using graph Laplacian regularization. *Journal of Computational Biology 18*(8), 987–996.

Godsil, C. and G. F. Royle (2001). *Algebraic Graph Theory*, Volume 207. Springer Science & Business Media.

Han, X., Q. Yang, and Y. Fan (2023). Universal rank inference via residual subsampling with application to large networks. *The Annals of Statistics 51*(3), 1109–1133.

Hardoon, D. R., S. Szedmak, and J. Shawe-Taylor (2004). Canonical correlation analysis: An overview with application to learning methods. *Neural Computation 16*(12), 2639–2664.

He, Y., A. Knowles, and R. Rosenthal (2018). Isotropic self-consistent equations for mean-field random matrices. *Probability Theory and Related Fields 171*(1), 203–249.

Hoory, S., N. Linial, and A. Wigderson (2006). Expander graphs and their applications. *Bulletin of the American Mathematical Society 43*(4), 439–561.

Jin, J. (2015). Fast community detection by SCORE. *The Annals of Statistics 43*(1), 57–89.

Jin, J., Z. T. Ke, and S. Luo (2024). Mixed membership estimation for social networks. *Journal of Econometrics 239*(2), 105369.

Johnstone, I. M. (2001). On the distribution of the largest eigenvalue in principal components analysis. *The Annals of Statistics 29*(2), 295–327.

Ke, Z. T. and J. Wang (2025). Optimal network membership estimation under severe degree heterogeneity. *Journal of the American Statistical Association 120*, 948–962.

Klein, D. J. (2002). Resistance-distance sum rules. *Croatica Chemica Acta 75*(2), 633–649.

Knowles, A. and J. Yin (2013). The isotropic semicircle law and deformation of Wigner matrices. *Communications on Pure and Applied Mathematics 66*(11), 1663–1749.

Knowles, A. and J. Yin (2014). The outliers of a deformed Wigner matrix. *The Annals of Probability 42*(5), 1980–2031.

Knowles, A. and J. Yin (2017). Anisotropic local laws for random matrices. *Probability Theory and Related Fields 169*, 257–352.

Koltchinskii, V. and K. Lounici (2016). Asymptotics and concentration bounds for bilinear forms of spectral projectors of sample covariance. *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques 52*(4), 1976–2013.

Kook, W. (2011). Combinatorial Green's function of a graph and applications to networks. *Advances in Applied Mathematics 46*, 417–423.

Kurasov, P. (2008). Graph laplacians and topology. *Arkiv för Matematik 46*(1), 95–111.

Le, C. M., E. Levina, and R. Vershynin (2016). Optimization via low-rank approximation for community detection in networks. *The Annals of Statistics 44*(1), 373–400.

Lederman, R. R. and R. Talmon (2018). Learning the geometry of common latent variables using alternating-diffusion. *Applied and Computational Harmonic Analysis 44*(3), 509–536.

Lei, J. (2016). A goodness-of-fit test for stochastic block models. *The Annals of Statistics 44*(1), 401–424.

Lei, J. and A. Rinaldo (2015). Consistency of spectral clustering in stochastic block models. *The Annals of Statistics 43*(1), 215–237.

Lubotzky, A., R. Phillips, and P. Sarnak (1988). Ramanujan graphs. *Combinatorica 8*(3), 261–277.

Merris, R. (1994). Laplacian matrices of graphs: a survey. *Linear Algebra and Its Applications 197*, 143–176.

Mohar, B., Y. Alavi, G. Chartrand, and O. Oellermann (1991). The Laplacian spectrum of graphs. *Graph Theory, Combinatorics, and Applications 2*(871-898), 12.

Nadler, B. (2008). Finite sample approximation results for principal component analysis: A matrix perturbation approach. *The Annals of Statistics 36*(6), 2791–2817.

Oliveira, R. I. (2009). Concentration of the adjacency matrix and of the Laplacian in random graphs with independent edges. *arXiv preprint arXiv:0911.0600*.

Paul, D. (2007). Asymptotics of sample eigenstructure for a large dimensional spiked covariance model. *Statistica Sinica 17*, 1617–1642.

Paul, D. and A. Aue (2014). Random matrix theory in statistics: A review. *Journal of Statistical Planning and Inference 150*, 1–29.

Pizzo, A., D. Renfrew, and A. Soshnikov (2013). On finite rank deformations of wigner matrices. In *Annales de l'IHP Probabilités et statistiques*, Volume 49, pp. 64–94.

Renfrew, D. and A. Soshnikov (2013). On finite rank deformations of Wigner matrices II: Delocalized perturbations. *Random Matrices: Theory and Applications 2*(01), 1250015.

Rohe, K., S. Chatterjee, and B. Yu (2011). Spectral clustering and the high-dimensional stochastic blockmodel. *The Annals of Statistics 39*(4), 1878–1915.

Sipser, M. and D. A. Spielman (1996). Expander codes. *IEEE Transactions on Information Theory 42*(6), 1710–1722.

Smilansky, U. (2007). Quantum chaos on discrete graphs. *Journal of Physics A: Mathematical and Theoretical 40*(27), F621.

Smola, A. J. and R. Kondor (2003). Kernels and regularization on graphs. In B. Schölkopf and M. K. Warmuth (Eds.), *Computational Learning Theory and Kernel Machines, 16th Annual Conference on Computational Learning Theory and 7th Kernel Workshop, COLT/Kernel 2003, Washington, DC, USA, August 24-27, 2003, Proceedings*, Volume 2777 of *Lecture Notes in Computer Science*, pp. 144–158. Springer.

Tang, M. and C. E. Priebe (2018). Limit theorems for eigenvectors of the normalized Laplacian for random graphs. *The Annals of Statistics 46*(5), 2360–2415.

Tao, T. (2012). *Topics in Random Matrix Theory*, Volume 132. American Mathematical Society.

Trinajstic, N., D. Babic, S. Nikolic, D. Plavsic, D. Amic, and Z. Mihalic (1994). The Laplacian matrix in chemistry. *Journal of Chemical Information and Computer Sciences 34*(2), 368–376.

van der Hofstad, R. (2016). *Random Graphs and Complex Networks*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press.

Vershynin, R. (2018). *High-Dimensional Probability: An Introduction with Applications in Data Science*, Volume 47. Cambridge University Press.

Verzelen, N. and E. Arias-Castro (2015). Community detection in sparse random networks. *The Annals of Applied Probability 25*(6), 3465–3510.

Wang, W. and J. Fan (2017). Asymptotics of empirical eigenstructure for high dimensional spiked covariance. *The Annals of Statistics 45*(3), 1342–1374.

Wang, Y. R. and P. J. Bickel (2017). Likelihood-based model selection for stochastic block models. *The Annals of Statistics 45*(2), 500–528.

Weyl, H. (1912). Das asymptotische verteilungsgesetz der eigenwerte linearer partieller differentialgleichungen (mit einer anwendung auf die theorie der hohlraumstrahlung). *Mathematische Annalen 71*, 441–479.

Xiao, W. and I. Gutman (2003). Resistance distance and Laplacian spectrum. *Theoretical Chemistry Accounts 110*, 284–289.

Zhang, S., H. Tong, J. Xu, and R. Maciejewski (2019). Graph convolutional networks: a comprehensive review. *Computational Social Networks 6*(1), 11.

Zhu, X., Z. Ghahramani, and J. D. Lafferty (2003). Semi-supervised learning using Gaussian fields and harmonic functions. In T. Fawcett and N. Mishra (Eds.), *Machine Learning, Proceedings of the Twentieth International Conference (ICML 2003), August 21-24, 2003, Washington, DC, USA*, pp. 912–919. AAAI Press.

APPENDIX A
ADDITIONAL SIMULATION RESULTS INVESTIGATING THE ROLE OF α

In this section of the Supplementary Material, we conduct two additional simulation studies to investigate the role of parameter α in two different statistical learning tasks. The results show that the optimal choice of α for achieving the best statistical performance is task dependent and varies with the degree of network heterogeneity.

A. Rank estimation and community detection under DCSBM

We generate an undirected network on $n = 1,800$ nodes from the degree-corrected stochastic block model (DCSBM) with $K = 3$ communities of equal sizes. Let $g_i \in \{1, \dots, K\}$ be the community label of node i. The degree heterogeneity is modeled via independent and identically distributed (i.i.d.) node weights

$$\nu_i \sim \mathrm{LogNormal}(0, \sigma^2),$$

truncated at $\nu_{\max} = 8$ and renormalized so that $\frac{1}{n}\sum_{i=1}^n \nu_i = 1$. Conditional on (g, ν), the edges are generated independently for $i < j$ according to

$$\mathbb{P}(A_{ij} = 1 \mid g, \theta) = \min\left\{\frac{1}{n}\nu_i\nu_j\, B_{g_i g_j}, 0.98\right\}, \tag{A.1}$$

and we set $A_{ji} = A_{ij}$ and $A_{ii} = 0$. The block connectivity matrix $\mathbf{B}$ is constructed by setting its diagonal entries to a and off-diagonal entries to b. To control both the average degree and the signal strength, we parameterize (a, b) using a target average degree $\bar{d}$ and a gap $\Delta = a - b$

$$b = \bar{d} - \Delta/K, \quad a = b + \Delta.$$

In the simulations, we set $\bar{d} = 180$ and $\Delta = 60$. Define $\mathbf{D}_\varepsilon = \mathrm{diag}(d_i + \varepsilon : i \in [n])$, where $\varepsilon = 10^{-4}$ is added to node degrees d_i's for *numerical stability*. For each $\alpha \in \{0, 0.1, 0.2, \cdots, 1\}$, we form the generalized Laplacian matrix

$$\mathbf{X}_\alpha = \mathbf{D}_\varepsilon^{-\alpha}\mathbf{A}\mathbf{D}_\varepsilon^{-\alpha}.$$

Given $\mathbf{X}_\alpha$, we first perform community detection using the true number of communities K for benchmarking purposes. Specifically, we compute the top K eigenvectors of $\mathbf{X}_\alpha$, optionally apply row normalization to the eigenvector embedding, and then run the K-means clustering to obtain estimated community labels $\widehat{g}^{(\alpha)}$. The row normalization is used widely in degree-heterogeneous networks for community detection; accordingly, we report results for both row-normalized and unnormalized eigenvector embeddings in order to assess whether the effect of α depends on this normalization step. The performance is evaluated using the misclustering rate after optimal label permutation.

We then estimate the number of communities K using the hard-thresholding method introduced in Theorem 6. Specifically, let $\widehat{\delta}_1, \cdots, \widehat{\delta}_{K_{\max}}$ be the leading eigenvalues of $\mathbf{X}_\alpha$, ordered by decreasing magnitude, where we set $K_{\max} = 10$. Define $\bar{q}^2 := n^{-1}\sum_{i=1}^n d_i$ as the average degree. Following Remark 6, we adopt the (finite-sample calibrated) hard threshold $a'_{n,\alpha} := \bar{q}^{1-4\alpha}\log\log n$, where $\log\log n$ may be replaced with some other slowly diverging sequence. We then estimate K by

$$\widehat{K}(\alpha) = \max\left\{k \le K_{\max} : \ |\widehat{\delta}_k| \ge a'_{n,\alpha}\right\}.$$

When $\widehat{K}(\alpha) \ge 2$, we repeat the spectral clustering using $\widehat{K}(\alpha)$ eigenvectors together with row-normalized embeddings, and obtain community labels $\widehat{g}^{(\alpha)}_{\widehat{K}}$. All reported results are averaged over 100 independent replications.

The simulation results with the true number of communities $K = 3$ are summarized in Figure 8. It is seen that even with true K, the optimal parameter α that minimizes the clustering error varies with the degree heterogeneity parameter σ, and is generally different from the commonly used choices $\alpha \in \{0, 0.5, 1\}$. When degree heterogeneity is high (i.e., σ is large), the optimal α depends on whether row-normalized or unnormalized eigenvectors are used for clustering, whereas for small σ the row normalization has little effect on the choice of the optimal α. It is also seen that the minimum achievable clustering error increases as the degree heterogeneity σ increases, which reflects the increased difficulty of community detection under stronger degree heterogeneity. Figure 9 further reports the optimal α that achieves the minimum clustering error when the hard-thresholding estimator $\widehat{K}$ is used. It is seen that moderate values of α, such as 0.3 or 0.5, are needed to achieve the minimum clustering error.

Fixing $\sigma = 0.5$, we conduct a more in-depth analysis on the estimation accuracy of K and the resulting clustering error. Across 100 repetitions, the average minimum degree count is 26.88 and the average maximum

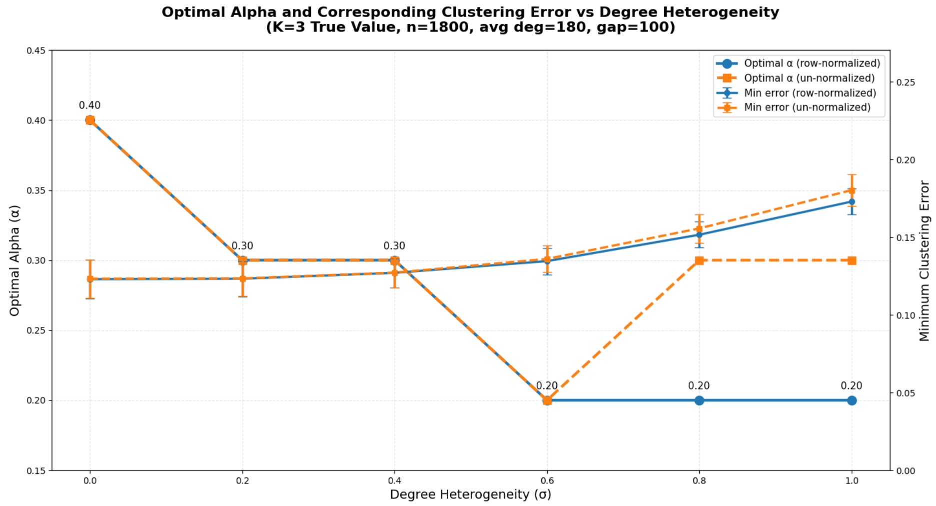

Fig. 8: Blue curves: using the true number of communities $K = 3$ with row-normalized eigenvectors, showing the optimal α and the corresponding minimum clustering error as functions of σ. Orange curves: using the true number of communities $K = 3$ with unnormalized eigenvectors, showing the optimal α and the corresponding minimum clustering error as functions of σ.

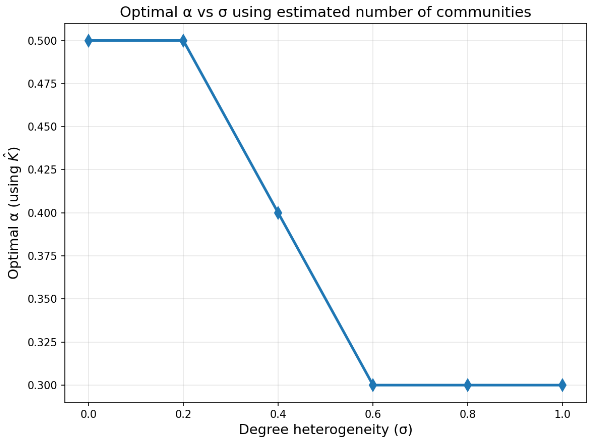

Fig. 9: The optimal α as a function of degree heterogeneity parameter σ with estimated K.

degree count is 817.78, indicating severe degree heterogeneity. The simulation results are summarized in Figure 10. We observe that, under substantial degree heterogeneity, α around 0.3 achieves good estimation results for both K and community labels.

In summary, this simulation study demonstrates that the optimal tuning parameter α depends on the degree heterogeneity parameter σ, and is *generally different* from the canonical values 0, 1/2, and 1.

B. *Laplacian-regularized graph smoothing*

We now present results from a simulation study under the Laplacian graph smoothing framework introduced in Section V-B. Specifically, we generate an undirected graph on $n = 300$ nodes from the DCSBM with $K = 2$ balanced communities. Let $g_i \in \{1, 2\}$ be the community label of node i, and $\nu_i > 0$ the corresponding degree parameter. Conditional on $(g_i, \nu_i)_{i\in[n]}$, the edges are generated independently (modulo symmetry) according to (A.1) with $B_{11} = B_{22} = 20$ and $B_{12} = B_{21} = 5$. The degree heterogeneity is generated by drawing $\nu_i \sim \text{LogNormal}(0, 2^2)$, truncated at 20 and then renormalizing so that $\frac{1}{n}\sum_{i=1}^{n} \nu_i = 1$. We simulate the node responses according to

$$y_i = 2(g_i - 3/2) + \varepsilon_i, \quad \varepsilon_i \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, 1).$$

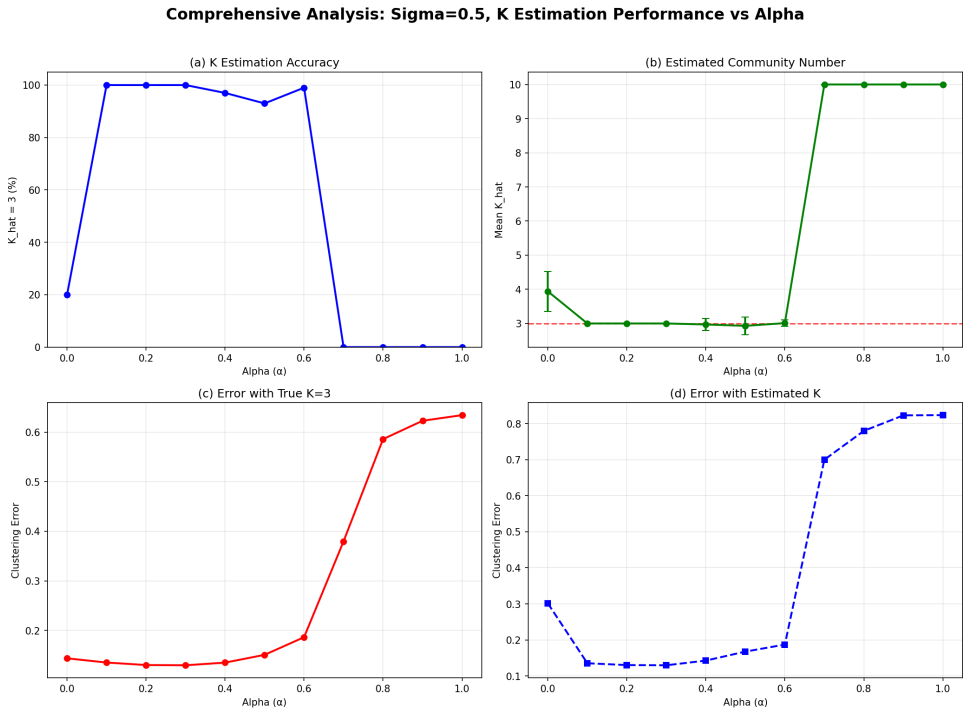

Fig. 10: Panel (a): percentage of correctly estimated $\widehat{K}$. Panel (b): average value of $\widehat{K}$. Panel (c): clustering error using true $K = 3$ with row-normalized eigenvectors. Panel (d): clustering error using $\widehat{K}$ with row-normalized eigenvectors. All results are averaged over 100 independent replications.

For each fixed $\alpha \in [0,1]$, we form the generalized Laplacian matrix $\mathbf{X}_\alpha := \mathbf{D}_\varepsilon^{-\alpha}\mathbf{A}\mathbf{D}_\varepsilon^{-\alpha}$, and define $\mathbf{L}_\alpha := \mathrm{diag}(\mathbf{X}_\alpha \mathbf{1}) - \mathbf{X}_\alpha$, where $\mathbf{1}$ denotes the vector of all ones. We then apply the standard Laplacian smoothing estimator

$$\widehat{f}^{(\alpha)} := \arg\min_{f\in\mathbb{R}^n} \left\{ \|\mathbf{y} - f\|_2^2 + \lambda\, f^T \widetilde{\mathbf{L}}_\alpha f \right\} = (\mathbf{I}_n + \lambda \widetilde{\mathbf{L}}_\alpha)^{-1}\mathbf{y},$$

which is a special case of the spectral filtering framework in Section V-B. Here, we use the *rescaled* Laplacian $\widetilde{\mathbf{L}}_\alpha := \mathbf{L}_\alpha \Big/ \left(\frac{1}{n}\mathrm{tr}(\mathbf{L}_\alpha)\right)$, so that the smoothing strength λ is comparable across different values of α. Since our focus is to evaluate the effect of α, we set $\lambda = 1$.

We evaluate the mean squared error (MSE) of $\widehat{f}^{(\alpha)}$ as an estimator of $f_0 := (2(g_i - 3/2))_{i\in[n]}$, averaged over 200 independent simulations, as a function of α (over a uniform grid on $[0,1]$). Figure 1 reports the resulting MSE curve. We observe that the value of α achieving the smallest MSE is approximately 0.75 on this grid, which does *not* belong to the commonly used choices $\alpha \in \{0, 1/2, 1\}$.

Appendix B
Main results for the rescaled model

To streamline the proofs of the main theoretical results presented in the main text, in this section of the Supplementary Material we clarify the relationship between several key parameters and quantities in the main text and their rescaled counterparts introduced below. Throughout the appendix, the sparsity parameters θ and θ_i, as well as the scaling parameters q and β_n, are *never rescaled.* These parameters are fixed as given in Definition 1. By keeping them at their original scale, we ensure consistency and coherence in the subsequent technical arguments.

To present the rescaled model in detail, let us introduce the following rescaled matrices

$$\widetilde{\mathbf{X}} \to \widetilde{\mathbf{X}}/q, \quad \mathbf{H} \to \mathbf{H}/q, \quad \mathbf{W} \to \mathbf{W}/q. \tag{B.1}$$

From now on, unless stated otherwise, the notation $\widetilde{\mathbf{X}} = (\widetilde{X}_{ij})_{i,j\in[n]}$, $\mathbf{H} = (H_{ij})_{i,j\in[n]}$, $\mathbf{W} = (W_{ij})_{i,j\in[n]}$, and

$s_{ij} := \mathbb{E}|W_{ij}|^2$ will refer to the rescaled quantities. Correspondingly, we define the rescaled diagonal matrix

$$\mathbf{D} := \operatorname{diag}(d_1, \cdots, d_n) = \frac{1}{q\beta_n}\operatorname{diag}\Big(\sum_{j\in[n]} \widetilde{X}_{ij} : i \in [n]\Big). \tag{B.2}$$

We also set

$$\mathbf{\Lambda} := \operatorname{diag}(\Lambda_1, \cdots, \Lambda_n) = \mathbb{E}\mathbf{D}, \tag{B.3}$$

and for each $i \in [n]$, define the random diagonal matrix $\mathbf{D}_{[i]}$ as the matrix with $(D_{[i]})_i = d_i$ and

$$(D_{[i]})_j = \Lambda_j + \frac{1}{q\beta_n}\sum_{s\in[n]\setminus\{i\}} W_{js}, \quad \forall j \neq i. \tag{B.4}$$

Here, we emphasize that $\widetilde{X}_{ij}$ in (B.2) and W_{ij} in (B.4) denote the entries of the *rescaled* matrices $\widetilde{\mathbf{X}}$ and $\mathbf{W}$, respectively. Hence, the relationship between the original $\mathbf{D}$ in (8) and the rescaled $\mathbf{D}$ in (B.2), the relationship between their expectations, and the relationship between the original $\mathbf{D}_{[i]}$ in (87) and the rescaled $\mathbf{D}_{[i]}$ in (B.4) are given by

$$\mathbf{D} \to \frac{\mathbf{D}}{q^2\beta_n}, \quad \mathbf{\Lambda} \to \frac{\mathbf{\Lambda}}{q^2\beta_n}, \quad \mathbf{D}_{[i]} \to \frac{\mathbf{D}_{[i]}}{q^2\beta_n}, \tag{B.5}$$

respectively. Finally, we note that introducing the normalization factor $q^{-2}\beta_n^{-1}$ ensures that the largest diagonal entries of $\mathbf{\Lambda}$ are of order β_n^{-1}, while the smallest ones are of order 1 by part (iii) of Definition 1. Such normalization is convenient for the later technical analyses.

From this point on, $\mathbf{D}$, $\mathbf{\Lambda}$, and $\mathbf{D}_{[i]}$ always refer to their rescaled versions. Consequently, we define the generalized Laplacian matrix as

$$\mathbf{X} := \mathbf{D}^{-\alpha}\widetilde{\mathbf{X}}\mathbf{D}^{-\alpha}, \tag{B.6}$$

and denote by $\overline{\mathbf{W}} := \mathbf{\Lambda}^{-\alpha}\mathbf{W}\mathbf{\Lambda}^{-\alpha}$ for each $\alpha \in (0,\infty)$. We consider both empirical and population eigendecompositions

$$\mathbf{X} = \sum_{i\in[n]} \widehat{\delta}_i \widehat{\mathbf{v}}_i \widehat{\mathbf{v}}_i^T \quad \text{and} \quad \mathbf{\Lambda}^{-\alpha}\mathbf{H}\mathbf{\Lambda}^{-\alpha} = \sum_{i\in[K]} \delta_i \mathbf{v}_i \mathbf{v}_i^T, \tag{B.7}$$

where the eigenvalues are arranged in descending order of magnitude with $|\widehat{\delta}_1| \geq \cdots \geq |\widehat{\delta}_n|$ and $|\delta_1| \geq \cdots \geq |\delta_K| > 0$, and $\widehat{\mathbf{v}}_i$ and $\mathbf{v}_i$ denote the corresponding eigenvectors. We further introduce the diagonal matrices of spiked eigenvalues

$$\widehat{\mathbf{\Delta}} := \operatorname{diag}(\widehat{\delta}_1, \cdots, \widehat{\delta}_K) \quad \text{and} \quad \mathbf{\Delta} := \operatorname{diag}(\delta_1, \cdots, \delta_K), \tag{B.8}$$

and the associated matrices of spiked eigenvectors

$$\widehat{\mathbf{V}} = (\widehat{\mathbf{v}}_1, \cdots, \widehat{\mathbf{v}}_K) \;\text{ and }\; \mathbf{V} = (\mathbf{v}_1, \cdots, \mathbf{v}_K). \tag{B.9}$$

Under the rescaling scheme described above, the relationships between matrices $\mathbf{X}$, $\overline{\mathbf{W}}$, eigenvalues $\widehat{\delta}_k$, δ_k, and their unrescaled counterparts are given by

$$\mathbf{X} \to \frac{\beta_n^{2\alpha}\mathbf{X}}{q^{1-4\alpha}}, \quad \overline{\mathbf{W}} \to \frac{\beta_n^{2\alpha}\overline{\mathbf{W}}}{q^{1-4\alpha}}, \quad \widehat{\delta}_k \to \frac{\beta_n^{2\alpha}\widehat{\delta}_k}{q^{1-4\alpha}}, \quad \delta_k \to \frac{\beta_n^{2\alpha}\delta_k}{q^{1-4\alpha}}, \tag{B.10}$$

while the eigenvectors remain invariant under this transformation. Throughout the remainder of the appendix, the notation $\mathbf{X}$, $\overline{\mathbf{W}}$, $\widehat{\delta}_k$, and δ_k will always refer to these rescaled quantities.

With the rescaled model now fully specified, we are prepared to restate the technical assumptions in a manner consistent with this notation. In particular, under the rescaled setting, the assumptions from Definition 1 are reformulated as follows.

Condition B.1. *Assume that the following regularity conditions hold for some constant $C_0 > 0$.*

(i) The sparsity parameter q satisfies

$$(\log n)^4 \ll q \leq C_0 n^{1/2}. \tag{B.11}$$

(ii) The entries of $\mathbf{W}$ satisfy

$$\max_{i,j\in[n]} |W_{ij}| \leq C_0/q \tag{B.12}$$

almost surely, and

$$\mathbb{E}W_{ij}=0,\quad s_{ij}=\mathbb{E}|W_{ij}|^2\le\frac{C_0}{n}=\frac{C_0\theta}{q^2},\quad \forall i,j\in[n]. \tag{B.13}$$

(iii) The entries of $\mathbf{H}$ *are nonnegative and satisfy*

$$\max_{i\in[n]}\theta_i\le C_0. \tag{B.14}$$

We next restate Assumption 1 under the rescaled model.

Assumption B.1. *Fix* $\alpha\in(0,\infty)$ *and assume that the following conditions hold for some* $1\le K_0\le K$.

(i) (Spiked eigenvalues) $|\delta_k|\gg 1$ *for all* $1\le k\le K_0$.

(ii) (Eigengap) There exists some constant $\epsilon_0>0$ *such that*

$$\min_{1\le k\le K_0}\frac{|\delta_k|}{|\delta_{k+1}|}>1+\epsilon_0. \tag{B.15}$$

(iii) (Low-rank signals) The rank K *of* $\mathbf{H}$ *satisfies that*

$$K\log n\left(\frac{1}{|\delta_{K_0}|\beta_n}+\frac{\log n}{q\beta_n^2}+\|\mathbf{V}\|_{\max}\right)\ll q. \tag{B.16}$$

To present the main results under the rescaled model, we first define the asymptotic limits t_k of the eigenvalues $\widehat{\delta}_k$ in analogy with (27). We begin by introducing the complex-valued vector $\mathbf{M}\equiv\mathbf{M}_n(z)=(M_1(z),\cdots,M_n(z))^T$ $\equiv(M_1,\cdots,M_n)^T$ as the z-dependent solution to the *generalized quadratic vector equation (GQVE)*

$$\frac{1}{M_i}=-z-\sum_{j\in[n]}\Lambda_i^{-2\alpha}s_{ij}\Lambda_j^{-2\alpha}M_j \tag{B.17}$$

with $\operatorname{Im}M_i(z)\ge 0$ for all $i\in[n]$ and $z\in\mathbb{C}_+$, where $\mathbb{C}_+$ denotes the upper half complex plane. The following properties are known; see, for instance, Corollary 1.3 of Ajanki et al. (2017).

1) There exists a probability measure μ_c supported on $\mathbb{R}$ such that $\langle\mathbf{M}\rangle:=n^{-1}\sum_{i\in[n]}M_i(z)$ is the Stieltjes transform of μ_c. Indeed, μ_c is known as the limiting empirical spectral distribution (ESD) of the noise random matrix $\overline{\mathbf{W}}$ (Ajanki et al., 2017).
2) The measure μ_c is absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}$, and its density ρ_c is given by
$$\rho_c(x)=\frac{1}{\pi}\lim_{\eta\to 0+}\operatorname{Im}\langle\mathbf{M}(x+i\eta)\rangle,\quad\forall x\in\mathbb{R}.$$
3) The measure μ_c is compactly supported on $\mathbb{R}$ with support $\operatorname{supp}(\mu_c)\subset[-2\sqrt{\mathfrak{m}},2\sqrt{\mathfrak{m}}]$, where
$$\mathfrak{m}:=\max_{i\in[n]}\sum_{j\in[n]}\Lambda_i^{-\alpha}s_{ij}\Lambda_j^{-\alpha}.$$
4) M_i is itself the Stieltjes transform of a finite measure with the same support as μ_c, and $\sup_{z\in\mathbb{C}_+}|M_i(z)|\lesssim 1$.

We next define the complex-valued deterministic diagonal matrix

$$\mathbf{\Upsilon}(z):=\operatorname{diag}(M_1(z),\cdots,M_n(z)) \tag{B.18}$$

and the associated complex-valued deterministic matrix

$$\mathbf{\Upsilon}_k(z):=\mathbf{\Upsilon}(z)-\mathbf{\Upsilon}(z)\mathbf{V}_{-k}\frac{1}{\mathbf{\Delta}_{-k}^{-1}+\mathbf{V}_{-k}^T\mathbf{\Upsilon}(z)\mathbf{V}_{-k}}\mathbf{V}_{-k}^T\mathbf{\Upsilon}(z) \tag{B.19}$$

for $z\in\mathbb{C}$ and $1\le k\le K$. For notational simplicity, we will omit the dependence on z whenever it is clear from the context. Comparing (24) and (B.17), we observe that

$$\widetilde{\mathbf{\Upsilon}}(z)=\frac{\beta_n^{2\alpha}}{q^{1-4\alpha}}\mathbf{\Upsilon}\left(\frac{\beta_n^{2\alpha}}{q^{1-4\alpha}}z\right)\quad\text{and}\quad\widetilde{\mathbf{\Upsilon}}_k(z)=\frac{\beta_n^{2\alpha}}{q^{1-4\alpha}}\mathbf{\Upsilon}_k\left(\frac{\beta_n^{2\alpha}}{q^{1-4\alpha}}z\right). \tag{B.20}$$

For each $1\le k\le K_0$, define an interval

$$\mathcal{I}_k:=\left\{x\in\mathbb{R}:\frac{|\delta_k|}{1+\epsilon_0/2}\le|x|\le(1+\epsilon_0/2)|\delta_k|\right\}, \tag{B.21}$$

and let $t_k \in \mathbb{R}$ be the unique solution to the nonlinear equation

$$1 + \delta_k \mathbf{v}_k^T \mathbf{\Upsilon}(x) \mathbf{v}_k - \delta_k \mathbf{v}_k^T \mathbf{\Upsilon}(x) \mathbf{V}_{-k} \frac{1}{\mathbf{\Delta}_{-k}^{-1} + \mathbf{V}_{-k}^T \mathbf{\Upsilon}(x) \mathbf{V}_{-k}} \mathbf{V}_{-k}^T \mathbf{\Upsilon}(x) \mathbf{v}_k = 0 \tag{B.22}$$

over $x \in \mathcal{I}_k$. The next lemma corresponds to Lemma 1 in the main text and establishes the existence, uniqueness, and leading-order behavior of t_k for the rescaled model.

Lemma B.1. *Under parts (ii) and (iii) of Assumption B.1, for each $1 \leq k \leq K_0$, equation (B.22) admits a unique solution $x = t_k$ in $\mathcal{I}_k$, and moreover, it holds that*

$$t_k = \delta_k + O(|\delta_k|^{-1}). \tag{B.23}$$

Comparing (27) and (B.22), we obtain the corresponding rescaling relation

$$t_k \to \frac{\beta_n^{2\alpha} t_k}{q^{1-4\alpha}}, \tag{B.24}$$

which is consistent with the rescaling in (B.10). We now introduce three types of resolvents (Green's functions)

$$\mathbf{G}(z) := \left(\overline{\mathbf{W}} - z(\mathbf{D}/\mathbf{\Lambda})^{2\alpha}\right)^{-1}, \quad \mathbf{R}(z) := \left(\overline{\mathbf{W}} - z\mathbf{I}\right)^{-1}, \quad \mathbf{G}_{[i]}(z) = \left(\overline{\mathbf{W}} - z(\mathbf{D}_{[i]}/\mathbf{\Lambda})^{2\alpha}\right)^{-1} \tag{B.25}$$

for $z \in \mathbb{C}$ and $1 \leq i \leq n$. In comparison with (80) and (86), one can observe that the rescaling of the resolvents is given by

$$\mathbf{G}(z) \to \frac{\beta_n^{2\alpha}}{q^{1-4\alpha}} \mathbf{G}\left(\frac{\beta_n^{2\alpha}}{q^{1-4\alpha}} z\right), \quad \mathbf{R}(z) \to \frac{\beta_n^{2\alpha}}{q^{1-4\alpha}} \mathbf{R}\left(\frac{\beta_n^{2\alpha}}{q^{1-4\alpha}} z\right), \quad \mathbf{G}_{[i]}(z) \to \frac{\beta_n^{2\alpha}}{q^{1-4\alpha}} \mathbf{G}_{[i]}\left(\frac{\beta_n^{2\alpha}}{q^{1-4\alpha}} z\right). \tag{B.26}$$

To summarize, the rescaling scheme for the scaled model in this section is specified in (B.1), (B.5), (B.10), (B.24), and (B.26). Throughout this appendix, all symbols $\widetilde{\mathbf{X}}$, $\mathbf{H}$, $\mathbf{W}$, s_{ij}, $\mathbf{D}$, $\mathbf{\Lambda}$, $\mathbf{D}_{[i]}$, $\mathbf{X}$, $\overline{\mathbf{W}}$, $\widehat{\delta}_k$, δ_k, t_k, $\mathbf{G}$, $\mathbf{R}$, and $\mathbf{G}_{[i]}$ refer to their rescaled counterparts.

Finally, Theorems B.1–B.6 below are equivalent to Theorems 1–6 in the main text. Hence, to prove Theorems 1–6, it suffices to establish Theorems B.1–B.6 in the rescaled setting.

Theorem B.1. *Under Condition B.1 and Assumption B.1, it holds w.h.p. that*

$$\frac{|\widehat{\delta}_k - t_k|}{|\delta_k|} = O\left\{\frac{\sqrt{\log n}\,\psi_n(\delta_k)}{q}\left(1 + \frac{K}{|\delta_k|^4}\right)\right\} \tag{B.27}$$

for each $1 \leq k \leq K_0$, where we introduce the shorthand notation

$$\psi_n(\delta_k) := \frac{1}{|\delta_k|\beta_n} + \frac{\sqrt{\log n}}{q\beta_n^2} + \|\mathbf{V}\|_{\max}. \tag{B.28}$$

Theorem B.2. *Under Condition B.1 and Assumption B.1, for each $1 \leq k \leq K_0$, it holds w.h.p. that*

$$\left|\mathbf{v}_k^T (\mathbf{D}/\mathbf{\Lambda})^{-\alpha} \widehat{\mathbf{v}}_k - \frac{1}{\sqrt{\delta_k^2 \mathbf{v}_k^T \mathbf{\Upsilon}_k'(t_k) \mathbf{v}_k}}\right| \lesssim \frac{\sqrt{\log n}\,\psi_n(\delta_k)}{q}\left(1 + \frac{K}{|\delta_k|^4}\right), \tag{B.29}$$

where the sign of $\widehat{\mathbf{v}}_k$ is chosen such that $\widehat{\mathbf{v}}_k^T \mathbf{v}_k > 0$. In addition, for any deterministic unit vector $\mathbf{u} \in \mathbb{R}^n$, it holds w.h.p. that

$$\left|\mathbf{u}^T (\mathbf{D}/\mathbf{\Lambda})^{-\alpha} \widehat{\mathbf{v}}_k + \frac{\delta_k \mathbf{u}^T \mathbf{\Upsilon}_k(t_k) \mathbf{v}_k}{\sqrt{\delta_k^2 \mathbf{v}_k^T \mathbf{\Upsilon}_k'(t_k) \mathbf{v}_k}}\right| \lesssim \frac{\sqrt{\log n}\,\psi_n(\delta_k)}{q}\left[1 + \frac{K}{|\delta_k|^4} + \|\mathbf{u}^T \mathbf{V}_{-k}\|\left(\sqrt{K} + \frac{K}{|\delta_k|^2}\right)\right]. \tag{B.30}$$

Moreover, for the second terms on the left-hand side (LHS) of (B.29) and (B.30), we have

$$\delta_k^2 \mathbf{v}_k^T \mathbf{\Upsilon}_k'(t_k) \mathbf{v}_k = 1 + O(\delta_k^{-2}) \quad \textit{and} \quad \delta_k \mathbf{u}^T \mathbf{\Upsilon}_k(t_k) \mathbf{v}_k = -\mathbf{u}^T \mathbf{v}_k + O(\delta_k^{-2}). \tag{B.31}$$

Theorem B.3. *Assume that Condition B.1 and Assumption B.1 hold, and that*

$$K\psi_n(\delta_k)\beta_n \lesssim 1, \quad \|\mathbf{V}\|_{\max} \ll \frac{1}{|\delta_k|\beta_n} + \frac{\sqrt{\log n}}{q\beta_n^2} \tag{B.32}$$

for all $1 \le k \le K_0$. Then for each $i \in [n]$, it holds w.h.p. that

$$\begin{aligned} \widehat{v}_k(i) &= (\Lambda_i/d_i)^{\alpha} v_k(i) + \frac{1}{t_k d_i^{\alpha}} \sum_{j\in[n]} W_{ij}\Lambda_j^{-\alpha} v_k(j) \\ &+ O\left(\|\mathbf{V}\|_{\max} \left(\frac{\sqrt{K}}{|\delta_k|^2} + \frac{K\sqrt{\log n}}{q}\left(\frac{1}{|\delta_k|\beta_n} + \frac{\sqrt{\log n}}{q\beta_n^2}\right)\right) + \frac{\sqrt{\log n}}{\sqrt{n}|\delta_k|}\left(\frac{1}{|\delta_k|} + \frac{\sqrt{\log n}}{q\beta_n}\right)\right), \end{aligned} \tag{B.33}$$

where the direction of $\widehat{\mathbf{v}}_k$ is chosen so that $\widehat{\mathbf{v}}_k^T \mathbf{v}_k > 0$. Consequently, we have that

$$\begin{aligned} \widehat{v}_k(i) &= v_k(i) - \frac{\alpha v_k(i)}{q\beta_n\Lambda_i} \sum_{j\in[n]} W_{ij} + \frac{1}{t_k}\sum_{j\in[n]} \Lambda_i^{-\alpha} W_{ij}\Lambda_j^{-\alpha} v_k(j) \\ &+ O\left(\|\mathbf{V}\|_{\max} \left(\frac{\sqrt{K}}{|\delta_k|^2} + \frac{K\sqrt{\log n}}{q}\left(\frac{1}{|\delta_k|\beta_n} + \frac{\sqrt{\log n}}{q\beta_n^2}\right)\right) + \frac{\sqrt{\log n}}{\sqrt{n}|\delta_k|}\left(\frac{1}{|\delta_k|} + \frac{\sqrt{\log n}}{q\beta_n}\right)\right). \end{aligned} \tag{B.34}$$

Theorem B.4. *Under Condition B.1 and Assumption B.1, for each $1 \le k \le K_0$, it holds w.h.p. that*

$$\begin{aligned} \widehat{\delta}_k - t_k - A_k &= -2\alpha t_k \mathbf{v}_k^T \frac{\mathbf{D}-\mathbf{\Lambda}}{\mathbf{\Lambda}}\mathbf{v}_k + \mathbf{v}_k^T \overline{\mathbf{W}}\mathbf{v}_k + B_k \\ &+ O\left(\frac{1}{|\delta_k|^2} + \frac{(\log n)^{3/2}|\delta_k|}{q^3\beta_n^3} + \frac{\sqrt{\log n}\psi_n(\delta_k)}{q}\left(\frac{\sqrt{K}}{|\delta_k|} + \frac{K\sqrt{\log n}|\delta_k|}{q}\psi_n(\delta_k) + \frac{\log n}{q}\|\mathbf{v}_k\|_\infty^2\right)\right), \end{aligned} \tag{B.35}$$

where the deterministic term A_k is given by

$$A_k = \alpha(2\alpha+1)t_k \mathbb{E}\mathbf{v}_k^T \frac{(\mathbf{D}-\mathbf{\Lambda})^2}{\mathbf{\Lambda}^2}\mathbf{v}_k - 4\alpha\mathbb{E}\mathbf{v}_k^T\frac{\mathbf{D}-\mathbf{\Lambda}}{\mathbf{\Lambda}}\overline{\mathbf{W}}\mathbf{v}_k,$$

and B_k is a centered random error satisfying

$$\mathrm{var}(B_k) \lesssim \frac{|\delta_k|^2\|\mathbf{v}_k\|_\infty^2}{q^4\beta_n^4} + \frac{|\delta_k|^2}{q^4n^2\beta_n^4} + \frac{\|\mathbf{v}_k\|_\infty^2}{q^2\beta_n^2} + \frac{1}{q^2n\beta_n^2} + \frac{1}{\sqrt{n}q|\delta_k|^2}. \tag{B.36}$$

Theorem B.5. *Assume that Condition B.1 and Assumption B.1 hold.*

1) For each $1 \le k \le K_0$ and any deterministic unit vector $\mathbf{u}$ such that $\mathbf{u}^T\mathbf{v}_k = 0$, it holds w.h.p. that

$$\begin{aligned} \mathbf{u}^T(\mathbf{D}/\mathbf{\Lambda})^{-\alpha}\widehat{\mathbf{v}}_k - \mathcal{A}_k &= \mathbf{w}^T\left(-2\alpha\frac{\mathbf{D}-\mathbf{\Lambda}}{\mathbf{\Lambda}} + t_k^{-1}\overline{\mathbf{W}}\right)\mathbf{v}_k + \mathcal{B}_k \\ &+ O\left(\frac{1}{|\delta_k|^3} + \frac{\sqrt{\log n}}{q|\delta_k|^2\beta_n} + \frac{(\log n)^{3/2}}{q^3\beta_n^3} + \frac{K\sqrt{\log n}}{q|\delta_k|^2}\psi_n(\delta_k) + \frac{K^{3/2}\log n}{q^2}\psi_n(\delta_k)^2\right) \\ &+ O\left(\left(\frac{\log n}{q^2\beta_n^2} + \frac{\log n}{q|\delta_k|^2}\right)\|\mathbf{v}_k\|_\infty^2 + \frac{(\log n)^{3/2}}{q^2|\delta_k|\beta_n}\|\mathbf{v}_k\|_\infty^3\right), \end{aligned} \tag{B.37}$$

where we choose the sign of $\widehat{\mathbf{v}}_k$ such that $\widehat{\mathbf{v}}_k^T\mathbf{v}_k > 0$, and $\mathbf{w}$ is a deterministic vector defined in (55). The deterministic bias term is given by

$$\mathcal{A}_k = \mathbb{E}\mathbf{w}^T\left(\alpha(2\alpha+1)\frac{(\mathbf{D}-\mathbf{\Lambda})^2}{\mathbf{\Lambda}^2} - \frac{2\alpha}{t_k}\left(\frac{\mathbf{D}-\mathbf{\Lambda}}{\mathbf{\Lambda}}\overline{\mathbf{W}} + \overline{\mathbf{W}}\frac{\mathbf{D}-\mathbf{\Lambda}}{\mathbf{\Lambda}}\right) + \frac{\overline{\mathbf{W}}^2}{t_k^2}\right)\mathbf{v}_k,$$

and the centered random fluctuation $\mathcal{B}_k$ satisfies

$$\mathrm{var}(\mathcal{B}_k) \lesssim \frac{\|\mathbf{v}_k\|_\infty\|\mathbf{w}\|_\infty}{q^4\beta_n^4} + \frac{1}{q^4n^2\beta_n^4} + \frac{1}{q^2|\delta_k|^2\beta_n^2}\left(\|\mathbf{v}_k\|_\infty\|\mathbf{w}\|_\infty + \frac{1}{n}\right) + \frac{1}{q\sqrt{n}|\delta_k|^4}.$$

2) For $\mathbf{u} = \mathbf{v}_k$ and each $1 \le k \le K_0$, it holds w.h.p. that

$$\begin{aligned} \mathbf{v}_k^T(\mathbf{D}/\mathbf{\Lambda})^{-\alpha}\widehat{\mathbf{v}}_k - \mathbf{v}_k^T(\mathbf{D}/\mathbf{\Lambda})^{-\alpha}\mathbf{v}_k - \mathfrak{A}_k &= \frac{\alpha^2}{2}\mathbf{v}_k^T\left(\frac{\mathbf{D}-\mathbf{\Lambda}}{\mathbf{\Lambda}}\right)^2\mathbf{v}_k - \frac{1}{2t_k^2}\mathbf{v}_k^T\overline{\mathbf{W}}^2\mathbf{v}_k + \mathfrak{B}_k \\ &+ O\left(\frac{1}{|\delta_k|^4} + \frac{K\log n}{q^2}\psi_n(\delta_k)^2 + \left(\frac{\log n}{q^2\beta_n^2} + \frac{\log n}{q|\delta_k|^3}\right)\|\mathbf{v}_k\|_\infty^2\right), \end{aligned} \tag{B.38}$$

where $\mathfrak{A}_k$ *is a deterministic bias term given by*

$$\mathfrak{A}_k := (\delta_k^2 \mathbf{v}_k^T \boldsymbol{\Upsilon}_k'(t_k)\mathbf{v}_k)^{-1/2} - 1 + \frac{1}{2}\mathbf{v}_k^T(t_k^2\boldsymbol{\Upsilon}'(t_k) + 2t_k\boldsymbol{\Upsilon}(t_k) + \mathbf{I})\mathbf{v}_k$$

and $\mathfrak{B}_k$ *is a random variable satisfying*

$$\mathbb{E}\mathfrak{B}_k^2 \lesssim \left(\frac{1}{q^8\beta_n^6} + \frac{1}{q^2|\delta_k|^6} + \frac{1}{q^4|\delta_k|^4\beta_n^4}\right) n^2\|\mathbf{v}_k\|_\infty^4.$$

We can also establish the corresponding results for $\mathbf{u}^T\widehat{\mathbf{v}}_k$*. More precisely,*

3) For each $1 \leq k \leq K_0$ *and any deterministic unit vector* $\mathbf{u}$ *such that* $\mathbf{u}^T\mathbf{v}_k = 0$*, it holds w.h.p. that*

$$\begin{aligned}\mathbf{u}^T\widehat{\mathbf{v}}_k - \widetilde{\mathcal{A}}_k &= \mathbf{w}^T\left(-2\alpha\frac{\mathbf{D}-\boldsymbol{\Lambda}}{\boldsymbol{\Lambda}} + t_k^{-1}\overline{\mathbf{W}}\right)\mathbf{v}_k + \alpha\mathbf{u}^T\frac{\mathbf{D}-\boldsymbol{\Lambda}}{\boldsymbol{\Lambda}}\mathbf{v}_k + \widetilde{\mathcal{B}}_k\\ &+ O\left(\frac{1}{|\delta_k|^3} + \frac{\sqrt{\log n}}{q|\delta_k|^2\beta_n} + \frac{(\log n)^{3/2}}{q^3\beta_n^3} + \frac{K\sqrt{\log n}}{q|\delta_k|^2}\psi_n(\delta_k) + \frac{K^{3/2}\log n}{q^2}\psi_n(\delta_k)^2\right)\\ &+ O\left(\left(\frac{\log n}{q^2\beta_n^2} + \frac{\log n}{q|\delta_k|^2}\right)\|\mathbf{v}_k\|_\infty^2 + \frac{(\log n)^{3/2}}{q^2|\delta_k|\beta_n}\|\mathbf{v}_k\|_\infty^3\right),\end{aligned} \tag{B.39}$$

where we choose the sign of $\widehat{\mathbf{v}}_k$ *such that* $\widehat{\mathbf{v}}_k^T\mathbf{v}_k > 0$*. The deterministic bias term is given by*

$$\widetilde{\mathcal{A}}_k = \mathcal{A}_k + \mathbb{E}\mathbf{u}^T\left(-\frac{\alpha(3\alpha+1)}{2}\frac{(\mathbf{D}-\boldsymbol{\Lambda})^2}{\boldsymbol{\Lambda}^2} + \frac{\alpha}{t_k}\frac{\mathbf{D}-\boldsymbol{\Lambda}}{\boldsymbol{\Lambda}}\overline{\mathbf{W}}\right)\mathbf{v}_k,$$

and the centered random fluctuation $\widetilde{\mathcal{B}}_k$ *satisfies*

$$\operatorname{var}(\widetilde{\mathcal{B}}_k) \lesssim \frac{\|\mathbf{v}_k\|_\infty\|\mathbf{w}\|_\infty}{q^4\beta_n^4} + \frac{1}{q^4n^2\beta_n^4} + \frac{1}{q^2|\delta_k|^2\beta_n^2}\left(\|\mathbf{v}_k\|_\infty\|\mathbf{w}\|_\infty + \frac{1}{n}\right) + \frac{1}{q\sqrt{n}|\delta_k|^4}.$$

4) For $\mathbf{u} = \mathbf{v}_k$ *and each* $1 \leq k \leq K_0$*, it holds w.h.p. that*

$$\begin{aligned}\mathbf{v}_k^T\widehat{\mathbf{v}}_k - 1 - \mathfrak{A}_k &= -\frac{1}{2}\mathbf{v}_k^T\left(\alpha\frac{\mathbf{D}-\boldsymbol{\Lambda}}{\boldsymbol{\Lambda}} - \frac{\overline{\mathbf{W}}}{t_k}\right)^2\mathbf{v}_k + \widetilde{\mathfrak{B}}_k\\ &+ O\left(\frac{1}{|\delta_k|^4} + \frac{K\log n}{q^2}\psi_n(\delta_k)^2 + \left(\frac{\log n}{q^2\beta_n^2} + \frac{\log n}{q|\delta_k|^3}\right)\|\mathbf{v}_k\|_\infty^2\right),\end{aligned} \tag{B.40}$$

where $\widetilde{\mathfrak{B}}_k$ *is a random variable satisfying*

$$\mathbb{E}\widetilde{\mathfrak{B}}_k^2 \lesssim \left(\frac{1}{q^8\beta_n^6} + \frac{1}{q^2|\delta_k|^6} + \frac{1}{q^4|\delta_k|^4\beta_n^4}\right) n^2\|\mathbf{v}_k\|_\infty^4.$$

Theorem B.6. *Assume that Condition B.1 and Assumption B.1 hold, and in addition,*

$$|\delta_{K_0+1}| \gg 1, \quad \left|\frac{\delta_{K_0+1}}{\delta_{K_0+2}}\right| \geq 1 + \epsilon_0, \quad K\sqrt{\log n}\psi_n(\delta_{K_0+1}) \ll q. \tag{B.41}$$

Assume further that K_0 *satisfies*

$$K_0 = \max\{k \in [K] : |\delta_k| \geq a_n\} \tag{B.42}$$

for some deterministic sequence $a_n \to \infty$*, and that there exists another deterministic sequence* $a_n' \to \infty$ *such that*

$$\limsup_{n\to\infty}\left|\frac{a_n'}{a_n}\right| < 1, \quad \limsup_{n\to\infty}\frac{|\delta_{K_0+1}|}{a_n'} < 1. \tag{B.43}$$

Then the estimator of the latent embedding dimensionality, defined as

$$\widehat{K}_0 := \max\{k \in [K] : |\widehat{\delta}_k| \geq a_n'\}, \tag{B.44}$$

is consistent, i.e., $\mathbb{P}\{\widehat{K}_0 = K_0\} \to 1$ *as* $n \to \infty$*.*

Remark 7. We compare our results to those of Knowles and Yin (2013, 2014), where the authors derived the joint distribution of spiked eigenvalues in the spiked Wigner model. The spiked Wigner model can be viewed as a prototype of the rescaled random matrix model studied in this section: it consists of a low-rank perturbation of an $n \times n$ Wigner matrix H with centered, independent entries (up to the symmetry constraint $H^T = H$)

and identical variance $\mathbb{E}|H_{ij}|^2 = n^{-1}$. Our model generalizes substantially the spiked Wigner setting in several important directions. First, we allow for sparse random matrices satisfying (B.11), whereas the model in Knowles and Yin (2013, 2014) is essentially dense, corresponding to q of order $\sqrt{n}$. Second, instead of assuming identical variances for the entries of $\widetilde{\mathbf{X}}$, we permit heterogeneous variances by imposing only upper bounds on s_{ij} as in (B.13). Third, we introduce an additional degree normalization $\mathbf{D}^{-\alpha}$ for arbitrary $\alpha \geq 0$, which fundamentally changes the spectral structure and has *no* analogue in the classical spiked Wigner model.

Consequently, the results in Knowles and Yin (2013, 2014) are much stronger than those in our Theorems B.1 and B.4. Specifically, they established the asymptotic distribution of spiked eigenvalues arbitrarily close to the BBP transition, namely for signals satisfying $|\delta_k - 1| \gg n^{-1/3}$, where δ_k denotes the population spike. Moreover, they derived the joint distribution of all spiked eigenvalues for arbitrary signal configurations without requiring an eigengap condition such as (B.15); in particular, degenerate or near-degenerate spikes are allowed.

However, such strong results are *not* attainable in our setting, even without the degree normalization (i.e., in the case of $\alpha = 0$). First, in the sparse regime, the BBP transition scale $|\delta_k - 1| \sim n^{-1/3}$ cannot be reached even for the spiked Wigner model. This is because the extreme eigenvalues of a sparse Wigner-type matrix fluctuate on a much larger scale than $n^{-2/3}$, which is the Tracy–Widom scale for dense Wigner matrices. As a result, a signal is expected to generate an outlier eigenvalue only when $|\delta_k - 1|$ is much larger than a power of q^{-1}. To the best of our knowledge, the precise transition threshold in this sparse setting has *not yet* been characterized in the random matrix theory literature. A more fundamental difficulty arises from the potentially highly heterogeneous variance profile $\{s_{ij}\}$. In fact, determining the BBP transition threshold for spiked random matrices with general heterogeneous variances remains an *open problem* in random matrix theory. The key challenge is that the asymptotic location of the empirical spiked eigenvalue—and hence the corresponding phase transition threshold—depends critically on the interaction between the population eigenvector $\mathbf{v}_k$ and the variance profile $\{s_{ij}\}$. Hence, unlike in the spiked Wigner model, the transition behavior is not governed solely by the signal strength δ_k. For this reason, we impose the condition in part (i) of Assumption B.1, which guarantees that each δ_k produces an outlier empirical eigenvalue *without* requiring any specific structural assumptions on the population eigenvector $\mathbf{v}_k$. While it may be possible to extend our theory to partially relax the eigengap condition (B.15) and accommodate (near-)degenerate spiked eigenvalues, a systematic analysis of this regime would require substantially new ideas. We therefore leave this direction for future work, as noted in Remark 1.

It is also worth noting that Knowles and Yin (2013, 2014) do not contain results on the asymptotic distributions of the spiked eigenvectors. In contrast, the corresponding asymptotic theory for both empirical spiked eigenvalues and eigenvectors in the spiked sample covariance models was developed in Bao et al. (2021, 2022), under the assumptions that the spiked signals lie slightly above the BBP transition threshold and that an eigengap condition holds. However, similar to Knowles and Yin (2013, 2014), these works are restricted to the dense regime and assume that the population covariance matrix is a low-rank perturbation of the identity matrix. As discussed above, such assumptions are *fundamentally different* from those in our setting, which involves sparsity, heterogeneous variance profiles, and degree normalization. Consequently, the tools developed in Bao et al. (2021, 2022) *cannot* be applied directly to our model, and we therefore cannot expect to obtain results of comparable strength in terms of the transition sharpness or joint limiting distributions.

Appendix C
Proofs of Theorems B.1–B.6 and Corollary 3

In this section of the Supplementary Material, we provide the complete proofs of our main results, namely Theorems 1–6, as well as Corollary 3. As explained in Section B of the Supplementary Material, Theorems 1–6 are equivalent to Theorems B.1–B.6, which are formulated under the rescaled model. Therefore, it suffices to prove Theorems B.1–B.6. Throughout the following proofs, we work entirely in the rescaled setting introduced in Section B, incorporating the rescalings in (B.1), (B.5), (B.10), (B.24), and (B.26).

A. *Some key tools*

In this subsection, we collect several key tools that will be used in our technical analysis. We begin by establishing a concentration estimate for the rescaled diagonal matrix $\mathbf{D}$ defined in (B.2). To this end, let us introduce the diagonal random error matrix $\boldsymbol{\mathcal{E}}$ given by

$$\boldsymbol{\mathcal{E}} := \operatorname{diag}(\mathcal{E}_1, \cdots, \mathcal{E}_n) = \mathbf{D} - \boldsymbol{\Lambda}. \tag{C.1}$$

In accordance with Definition 3, given a sequence of random variables ξ_n and deterministic control parameters ϕ_n, we write $|\xi_n| \lesssim \phi_n$ if for any large constant $D > 0$, there exists a constant $C > 0$ such that $|\xi_n| \leq C\phi_n$ with probability at least $1 - n^{-D}$.

Lemma C.1. *Under Condition B.1, we have that with high probability,*

$$\max_{i\in[n]} |\mathcal{E}_i| \lesssim \sqrt{\log n}/(q\beta_n). \tag{C.2}$$

(Note that under condition (B.16), we have $q\beta_n \gg \log n$.) Moreover, there exists a constant $C_1 > 0$ (depending on C_0) such that for all $i \in [n]$,

$$C_1^{-1} \le \Lambda_i = \beta_n^{-1}\theta_i \le C_1\beta_n^{-1}. \tag{C.3}$$

As a consequence, there exists a constant $C_2 > 0$ such that with high probability,

$$\|\mathbf{\Lambda}\|, \|\mathbf{D}\| \le C_2\beta_n^{-1}, \quad \|\mathbf{\Lambda}^{-1}\|, \|\mathbf{D}^{-1}\| \le C_2, \quad \textit{and} \quad C_2^{-1} \le \|\mathbf{D}/\mathbf{\Lambda}\| \le C_2, \tag{C.4}$$

and for each fixed $\alpha \in (0, \infty)$ and $i \in [n]$,

$$|(d_i^\alpha - \Lambda_i^\alpha)/\Lambda_i^\alpha| \lesssim \sqrt{\log n}/(q\beta_n) \tag{C.5}$$

with high probability.

The above result follows immediately from the classical Bernstein's inequality stated below.

Lemma C.2 (Bernstein's inequality (Vershynin, 2018))**.** *Let $(x_i)_{i\in[n]}$ be a family of centered independent random variables satisfying that $\max_{i\in[n]} |x_i| \le \phi_n$ for some (n-dependent) parameter $\phi_n > 0$. Then for any $t > 0$, we have*

$$\mathbb{P}\bigg(\sum_{i\in[n]} x_i > t\bigg) \le 2\exp\bigg(-\frac{ct^2}{\sum_{i\in[n]}\mathbb{E}x_i^2 + \phi_n t}\bigg),$$

where $c > 0$ is an absolute constant.

By Lemma C.2 above, we see that w.h.p.,

$$\bigg|\sum_{i\in[n]} x_i\bigg| \lesssim \bigg(\sum_{i\in[n]}\mathbb{E}x_i^2\bigg)^{1/2}(\log n)^{1/2} + \phi_n\log n. \tag{C.6}$$

We now use the above estimate to complete the proof of Lemma C.1.

Proof of Lemma C.1. The inequality (C.3) follows directly from the definition. From (C.6), we can deduce that

$$|\mathcal{E}_i| \lesssim |d_i - \mathbb{E}d_i| = \frac{1}{q\beta_n}\bigg|\sum_{j\in[n]} W_{ij}\bigg| \lesssim \frac{(\log n)^{1/2}}{q\beta_n} + \frac{\log n}{q^2\beta_n} \lesssim \frac{(\log n)^{1/2}}{q\beta_n}$$

with high probability, which establishes (C.2). Using (C.2) and (C.3), along with assumption (B.16), we immediately obtain (C.4). Next, by the Taylor expansion, it holds that

$$\left|\frac{d_i^\alpha - \Lambda_i^\alpha}{\Lambda_i^\alpha}\right| = \alpha t^{\alpha-1}\frac{\mathcal{E}_i}{\Lambda_i},$$

where t is some value between d_j/Λ_j and 1. Equation (C.5) then follows immediately from (C.2) and (C.4). ☐

We next introduce the technical notion of *minors* of matrices as given in the definition below.

Definition 4 (Minors)**.** *Given an $n \times n$ matrix $\mathbf{A} = \mathbf{W}, \overline{\mathbf{W}}, \mathbf{\Upsilon}, \mathbf{\Lambda}$, or $\mathbf{D}$, and a subset $\mathbb{T} \subset [n]$, we define the minor $\mathbf{A}^{(\mathbb{T})} := (A_{ij} : i, j \notin \mathbb{T})$ as a matrix of size $(n - |\mathbb{T}|) \times (n - |\mathbb{T}|)$ defined by removing all rows and columns of $\mathbf{A}$ with indices belonging to $\mathbb{T}$. We keep the names of indices for $\mathbf{A}^{(\mathbb{T})}$, i.e., $A_{ij}^{(\mathbb{T})} = A_{ij}$ for $i, j \notin \mathbb{T}$. Then we define the resolvent minors as*

$$\mathbf{G}^{(\mathbb{T})}(z) := \left[\overline{\mathbf{W}}^{(\mathbb{T})} - z(\mathbf{D}^{(\mathbb{T})}/\mathbf{\Lambda}^{(\mathbb{T})})^{2\alpha}\right]^{-1} \quad \textit{and} \quad \mathbf{R}^{(\mathbb{T})}(z) := \left(\overline{\mathbf{W}}^{(\mathbb{T})} - z\right)^{-1}.$$

For simplicity of notation, we will abbreviate $(\{i\}) \equiv (i)$, $(\{i,j\}) \equiv (ij)$, and $\sum_{i\in[n]}^{(\mathbb{T})} \equiv \sum_{i\in[n]\setminus\mathbb{T}}$. As a convention, we let $A_{ij}^{(\mathbb{T})} = R_{ij}^{(\mathbb{T})} = G_{ij}^{(\mathbb{T})} = 0$ whenever i or j belongs to $\mathbb{T}$.

We define a parameter of order one as

$$\mathfrak{M} := \bigg(\max_{i\in[n]} \Lambda_i^{-\alpha}\sum_{j\in[n]} s_{ij}\Lambda_j^{-\alpha}\bigg) \vee 1, \tag{C.7}$$

where $\vee$ denotes the maximum of two given numbers. By adapting arguments from Erdős et al. (2013) together with estimate (C.5), we can obtain bounds on the operator norms of $\overline{\mathbf{W}}$, $\mathbf{G}$, $\mathbf{R}$, and their minors, as stated in the proposition below.

Proposition C.1. *Under Condition B.1, we have that with high probability,*

$$\max\left\{\|\overline{\mathbf{W}}\|,\ \max_{i\in[n]}\|\overline{\mathbf{W}}^{(i)}\|,\ \max_{i,j\in[n]}\|\overline{\mathbf{W}}^{(ij)}\|\right\}\le 2\sqrt{\mathfrak{M}}+(\log n)^{1+a_0}/\sqrt{q} \tag{C.8}$$

for any small constant $a_0>0$. Consequently, for any $\mathfrak{C}>2\sqrt{\mathfrak{M}}+\kappa$ with some constant $\kappa>0$, there exists a constant $C_3>0$ such that with high probability,

$$\sup_{z\in S(\mathfrak{C})}\left(|z|-2\sqrt{\mathfrak{M}}\right)\max\left\{\|\mathbf{R}(z)\|,\ \max_{i\in[n]}\|\mathbf{R}^{(i)}(z)\|,\ \max_{i,j\in[n]}\|\mathbf{R}^{(ij)}(z)\|\right\}\le C_3, \tag{C.9}$$

$$\sup_{z\in S(\mathfrak{C})}\left(|z|-2\sqrt{\mathfrak{M}}\right)\max\left\{\|\mathbf{G}(z)\|,\ \max_{i\in[n]}\|\mathbf{G}^{(i)}(z)\|,\ \max_{i,j\in[n]}\|\mathbf{G}^{(ij)}(z)\|\right\}\le C_3, \tag{C.10}$$

where the spectral domain is defined as

$$S(\mathfrak{C}):=\{z=E+i\eta:\mathfrak{C}\le|E|\le n^{\mathfrak{C}},\ \eta\ge0\}. \tag{C.11}$$

Proof. The estimate (C.8) can be proved using the same arguments as in the proof of (Erdős et al., 2013, Lemma 4.3); see also Lemma 2 in Fan et al. (2022). The bound (C.9) then follows immediately from (C.8) by definition. Next, in view of (C.4), (C.5), and (B.16), we have that for $\mathbb{T}=\emptyset,\ \{i\}$, or $\{i,j\}$,

$$\|(\mathbf{D}^{(\mathbb{T})}/\mathbf{\Lambda}^{(\mathbb{T})})^{2\alpha}-\mathbf{I}\|\lesssim\sqrt{\log n/(q\beta_n)}\ll1$$

with high probability. This together with (C.8) leads to (C.10), which completes the proof of Proposition C.1. □

Remark 8. Note that we have the following trivial upper bound on the signal strength. Under the rescaling (B.1), each node degree is of order $O(q)$, and it follows that

$$|\delta_1|^2\lesssim\|\mathbf{H}\|^2\le\sum_{i,j\in[n]}H_{ij}^2\le\sum_{i\in[n]}\Bigg(\sum_{j\in[n]}H_{ij}\Bigg)^2\lesssim nq^2\lesssim n^2. \tag{C.12}$$

Hence, since $\mathfrak{C}>2$ in view of (C.7), the spectral domain $S(\mathfrak{C})$ contains all the subsets $\mathcal{I}_k$ defined in (26).

Using the *generalized QVE (GQVE)* in (B.17) together with the definition in (B.18), we can obtain the following estimates on $\mathbf{\Upsilon}(z)$ and its first and second derivatives.

Lemma C.3. *For $z\in S(\mathfrak{C})$ with $\mathfrak{C}>2\sqrt{\mathfrak{M}}+\kappa$ and some constant $\kappa>0$, we have that*

$$\mathbf{\Upsilon}(z)=-z^{-1}+\boldsymbol{\mathcal{E}}_1(z),\quad \mathbf{\Upsilon}'(z)=z^{-2}+\boldsymbol{\mathcal{E}}_2(z),\quad \mathbf{\Upsilon}''(z)=-2z^{-3}+\boldsymbol{\mathcal{E}}_3(z), \tag{C.13}$$

where $\boldsymbol{\mathcal{E}}_1(z)$, $\boldsymbol{\mathcal{E}}_2(z)$, and $\boldsymbol{\mathcal{E}}_3(z)$ are deterministic diagonal matrices satisfying $\|\boldsymbol{\mathcal{E}}_1(z)\|=O(|z|^{-3})$, $\|\boldsymbol{\mathcal{E}}_2(z)\|=O(|z|^{-4})$, and $\|\boldsymbol{\mathcal{E}}_3(z)\|=O(|z|^{-5})$. Moreover, $\boldsymbol{\mathcal{E}}_1(z)$ admits the refined expansion

$$\boldsymbol{\mathcal{E}}_1(z)=\operatorname{diag}\Big(-z^{-3}\sum_{j\in[n]}\Lambda_i^{-2\alpha}s_{ij}\Lambda_j^{-2\alpha}:i\in[n]\Big)+O(|z|^{-5}). \tag{C.14}$$

Proof. This lemma follows from a Laurent series expansion of $\mathbf{\Upsilon}$, analogous to the expansion of $\widetilde{\mathbf{\Upsilon}}$ in (72). □

In light of Lemma C.3 above and the fact of $\mathbf{v}_k^T\mathbf{V}_{-k}=0$, we have that for all $z\in S(\mathfrak{C})$ with $\mathfrak{C}>2\sqrt{\mathfrak{M}}+\kappa$,

$$\mathbf{v}_k^T\mathbf{\Upsilon}(z)\mathbf{v}_k=-z^{-1}+O(|z|^{-3}),\quad \mathbf{V}_{-k}^T\mathbf{\Upsilon}(z)\mathbf{V}_{-k}=-z^{-1}\mathbf{I}+O(|z|^{-3}),\quad \mathbf{v}_k^T\mathbf{\Upsilon}(z)\mathbf{V}_{-k}=O(|z|^{-3}), \tag{C.15}$$

where in the second and third expressions, $O(|z|^{-3})$ denotes a matrix $\boldsymbol{\mathcal{E}}$ and a vector $\boldsymbol{\varepsilon}$ satisfying $\|\boldsymbol{\mathcal{E}}\|=O(|z|^{-3})$ and $|\boldsymbol{\varepsilon}|=O(|z|^{-3})$, respectively. We recall that $\mathbf{v}_k$'s are the eigenvectors of $\mathbf{\Lambda}^{-\alpha}\mathbf{H}\mathbf{\Lambda}^{-\alpha}$ as in (17), and $\mathbf{V}_{-k}$ denotes the $n\times(K-1)$ matrix obtained by removing the kth column of matrix $\mathbf{V}$, as defined below (25).

Our proofs of Theorems B.1–B.6 are based on some *fine* estimates on $\mathbf{G}(z)$, summarized in the following theorem and referred to as the *local laws*.

Theorem C.1. *Assume that Condition [B.1](#) holds and $\mathfrak{C} > 2\sqrt{\mathfrak{M}} + \kappa$ for some constant $\kappa > 0$. Then for any $k \in [n]$ and any large constant $D > 0$, there exists a constant $C > 0$ such that the following events hold with probability at least $1 - n^{-D}$:*

$$\bigcap_{z \in S(\mathfrak{C})} \left\{ \max_{i \in [n]} |\mathbf{e}_i^T (\mathbf{G}(z) - \mathbf{\Upsilon}(z)) \mathbf{v}_k| \le C \frac{\sqrt{\log n}}{|z|} \left(\frac{1}{\sqrt{n}|z|} + \left(\frac{\sqrt{\log n}}{q|z|} + \frac{1}{q\beta_n} \right) \|\mathbf{v}_k\|_\infty \right) \right\}, \tag{C.16}$$

$$\bigcap_{z \in S(\mathfrak{C})} \left\{ \max_{i \in [n]} |\mathbf{e}_i^T (\mathbf{G}(z) - \mathbf{\Upsilon}(z)) \mathbf{V}_{-k}| \le C \frac{\sqrt{K \log n}}{|z|} \left(\frac{1}{\sqrt{n}|z|} + \left(\frac{\sqrt{\log n}}{q|z|} + \frac{1}{q\beta_n} \right) \|\mathbf{V}_{-k}\|_{\max} \right) \right\}, \tag{C.17}$$

$$\bigcap_{z \in S(\mathfrak{C})} \left\{ |\mathbf{v}_k^T (\mathbf{G}(z) - \mathbf{\Upsilon}(z)) \mathbf{v}_k| \le C \frac{\sqrt{\log n}}{q|z|} \left(\frac{1}{|z|\beta_n} + \frac{\sqrt{\log n}}{q\beta_n^2} + \|\mathbf{v}_k\|_\infty \right) \right\}, \tag{C.18}$$

$$\bigcap_{z \in S(\mathfrak{C})} \left\{ \|\mathbf{V}_{-k}^T (\mathbf{G}(z) - \mathbf{\Upsilon}(z)) \mathbf{V}_{-k}\| \le C \frac{K\sqrt{\log n}}{q|z|} \left(\frac{1}{|z|\beta_n} + \frac{\sqrt{\log n}}{q\beta_n^2} + \|\mathbf{V}_{-k}\|_{\max} \right) \right\}, \tag{C.19}$$

$$\bigcap_{z \in S(\mathfrak{C})} \left\{ \|\mathbf{v}_k^T (\mathbf{G}(z) - \mathbf{\Upsilon}(z)) \mathbf{V}_{-k}\| \le C \frac{\sqrt{K \log n}}{q|z|} \left(\frac{1}{|z|\beta_n} + \frac{\sqrt{\log n}}{q\beta_n^2} + \|\mathbf{V}\|_{\max} \right) \right\}. \tag{C.20}$$

Proof. These estimates are direct consequences of Theorems D.2 and D.3 in Section D. ☐

Combining (C.15) with (C.18)–(C.20) and using condition (B.16), we can obtain that for all $|z| \gtrsim |\delta_{K_0}|$,

$$|\mathbf{v}_k^T \mathbf{G}(z) \mathbf{v}_k| \lesssim |z|^{-1}, \quad \|\mathbf{V}_{-k}^T \mathbf{G}(z) \mathbf{V}_{-k}\| \lesssim |z|^{-1}, \tag{C.21}$$

$$\|\mathbf{v}_k^T \mathbf{G}(z) \mathbf{V}_{-k}\| \lesssim \frac{1}{|z|^3} + \frac{\sqrt{K \log n}}{q|z|} \left(\frac{1}{|z|\beta_n} + \frac{\sqrt{\log n}}{q\beta_n^2} + \|\mathbf{V}\|_{\max} \right) \tag{C.22}$$

with high probability.

B. *Proof of Theorem B.1*

The key ingredient of the proof is to show that $\widehat{\delta}_k$ satisfies the same equation as in (B.22) but with $\mathbf{\Upsilon}(x)$ replaced by $\mathbf{G}(x)$; see (C.25) below. Subtracting (B.22) from (C.25) and applying the local laws in (C.18)–(C.20), we can derive the desired conclusion.

More precisely, combining the eigengap condition (B.15) in Assumption B.1 with (C.15), we see that there exists a constant $C > 0$ such that

$$\left\| \delta_k^{-1} (\mathbf{\Delta}_{-k}^{-1} + \mathbf{V}_{-k}^T \mathbf{\Upsilon}(z) \mathbf{V}_{-k})^{-1} \right\| \le C \tag{C.23}$$

for all $z \in \mathcal{I}_k$. Moreover, with the aid of (C.19) and (C.23), we can deduce that w.h.p.,

$$\left\| \frac{1}{\mathbf{\Delta}_{-k}^{-1} + \mathbf{V}_{-k}^T \mathbf{G}(z) \mathbf{V}_{-k}} - \frac{1}{\mathbf{\Delta}_{-k}^{-1} + \mathbf{V}_{-k}^T \mathbf{\Upsilon}(z) \mathbf{V}_{-k}} \right\| \lesssim |\delta_k| \frac{K\sqrt{\log n}}{q} \psi_n(\delta_k).$$

In view of (B.16), this implies that for all $z \in \mathcal{I}_k$,

$$\left\| \frac{1}{\mathbf{\Delta}_{-k}^{-1} + \mathbf{V}_{-k}^T \mathbf{G}(z) \mathbf{V}_{-k}} \right\| \lesssim |\delta_k| \quad \text{w.h.p.} \tag{C.24}$$

With an application of Weyl's inequality (Weyl, 1912) and Proposition C.1, it holds that w.h.p., $\widehat{\delta}_k \in \mathcal{I}_k$ for each $1 \le k \le K_0$. We first make a useful claim that w.h.p., $\widehat{\delta}_k$ satisfies the nonlinear equation

$$1 + \delta_k \mathbf{v}_k^T \mathbf{G}(\widehat{\delta}_k) \mathbf{v}_k - \delta_k \mathbf{v}_k^T \mathbf{G}(\widehat{\delta}_k) \mathbf{V}_{-k} \frac{1}{\mathbf{\Delta}_{-k}^{-1} + \mathbf{V}_{-k}^T \mathbf{G}(\widehat{\delta}_k) \mathbf{V}_{-k}} \mathbf{V}_{-k}^T \mathbf{G}(\widehat{\delta}_k) \mathbf{v}_k = 0. \tag{C.25}$$

In fact, $\widehat{\delta}_k$ is a solution to equation $\det(\mathbf{X} - z\mathbf{I}) = 0$ over $z \in \mathcal{I}_k$. Moreover, for all $|z| \gg 1$, $\mathbf{G}(z)$ exists and is nonsingular with high probability by Proposition C.1. Hence, with the spectral decomposition $\mathbf{\Lambda}^{-\alpha} \mathbf{H} \mathbf{\Lambda}^{-\alpha} = \mathbf{V} \mathbf{\Delta} \mathbf{V}^T$ (recall (17)) and the identity $\det(\mathbf{I} + \mathbf{AB}) = \det(\mathbf{I} + \mathbf{BA})$ for any conformable matrices $\mathbf{A}$ and $\mathbf{B}$, we observe that equation $\det(\mathbf{X} - z\mathbf{I}) = 0$ is equivalent to

$$\det(\mathbf{G}(z)^{-1} + \mathbf{\Lambda}^{-\alpha} \mathbf{H} \mathbf{\Lambda}^{-\alpha}) = 0 \iff \det(\mathbf{\Delta}^{-1} + \mathbf{V}^T \mathbf{G}(z) \mathbf{V}) = 0.$$

Let us write the second equation above as

$$\det\begin{pmatrix}\delta_k^{-1}+\mathbf{v}_k^T\mathbf{G}(z)\mathbf{v}_k & \mathbf{v}_k^T\mathbf{G}(z)\mathbf{V}_{-k}\\ \mathbf{V}_{-k}^T\mathbf{G}(z)\mathbf{v}_k & \mathbf{\Delta}_{-k}^{-1}+\mathbf{V}_{-k}^T\mathbf{G}(z)\mathbf{V}_{-k}\end{pmatrix}=0.$$

Using Schur's formula for the determinant, this equation is equivalent to $\det(\mathbf{\Delta}_{-k}^{-1}+\mathbf{V}_{-k}^T\mathbf{G}(z)\mathbf{V}_{-k})=0$ or

$$1+\delta_k\mathbf{v}_k^T\mathbf{G}(z)\mathbf{v}_k-\delta_k\mathbf{v}_k^T\mathbf{G}(z)\mathbf{V}_{-k}\frac{1}{\mathbf{\Delta}_{-k}^{-1}+\mathbf{V}_{-k}^T\mathbf{G}(z)\mathbf{V}_{-k}}\mathbf{V}_{-k}^T\mathbf{G}(z)\mathbf{v}_k=0.$$

In light of (C.24), we see that matrix $\mathbf{\Delta}_{-k}^{-1}+\mathbf{V}_{-k}^T\mathbf{G}(\widehat{\delta}_k)\mathbf{V}_{-k}$ is nonsingular w.h.p., which entails that equation (C.25) indeed holds w.h.p.

We are now ready to establish (B.27). Subtracting (B.22) from (C.25), we obtain that w.h.p.,

$$\begin{aligned}\mathbf{v}_k^T\big[\mathbf{G}(\widehat{\delta}_k)-\mathbf{\Upsilon}(t_k)\big]\mathbf{v}_k&=\mathbf{v}_k^T\mathbf{G}(\widehat{\delta}_k)\mathbf{V}_{-k}\frac{1}{\mathbf{\Delta}_{-k}^{-1}+\mathbf{V}_{-k}^T\mathbf{G}(\widehat{\delta}_k)\mathbf{V}_{-k}}\mathbf{V}_{-k}^T\mathbf{G}(\widehat{\delta}_k)\mathbf{v}_k\\&\quad-\mathbf{v}_k^T\mathbf{\Upsilon}(t_k)\mathbf{V}_{-k}\frac{1}{\mathbf{\Delta}_{-k}^{-1}+\mathbf{V}_{-k}^T\mathbf{\Upsilon}(t_k)\mathbf{V}_{-k}}\mathbf{V}_{-k}^T\mathbf{\Upsilon}(t_k)\mathbf{v}_k.\end{aligned}\tag{C.26}$$

From (C.18), it holds that w.h.p.,

$$\mathbf{v}_k^T\big[\mathbf{G}(\widehat{\delta}_k)-\mathbf{\Upsilon}(t_k)\big]\mathbf{v}_k=\mathbf{v}_k^T\big[\mathbf{\Upsilon}(\widehat{\delta}_k)-\mathbf{\Upsilon}(t_k)\big]\mathbf{v}_k+O\bigg(\frac{\sqrt{\log n}\psi_n(\delta_k)}{q|\delta_k|}\bigg).\tag{C.27}$$

Using (C.19)–(C.23) and (C.24), we can deduce that w.h.p.,

$$\begin{aligned}&\mathbf{v}_k^T\mathbf{G}(z)\mathbf{V}_{-k}\frac{1}{\mathbf{\Delta}_{-k}^{-1}+\mathbf{V}_{-k}^T\mathbf{G}(z)\mathbf{V}_{-k}}\mathbf{V}_{-k}^T\mathbf{G}(z)\mathbf{v}_k-\mathbf{v}_k^T\mathbf{\Upsilon}(z)\mathbf{V}_{-k}\frac{1}{\mathbf{\Delta}_{-k}^{-1}+\mathbf{V}_{-k}^T\mathbf{\Upsilon}(z)\mathbf{V}_{-k}}\mathbf{V}_{-k}^T\mathbf{\Upsilon}(z)\mathbf{v}_k\\&\lesssim\frac{\sqrt{K\log n}}{q}\psi_n(z)\left(\frac{1}{|z|^3}+\frac{\sqrt{K}}{|z|^5}+\frac{\sqrt{K\log n}}{q|z|}\psi_n(z)\right),\quad\forall z\in\mathcal{I}_k,\end{aligned}\tag{C.28}$$

where we have used the notation in (B.28) with δ_k replaced by a general z, and the asymptotic bound above is understood *implicitly* for the absolute value of the quantity involved (for notational simplicity). Furthermore, with the aid of (C.13) and (C.15), we can rewrite it as

$$\begin{aligned}&\mathbf{v}_k^T\mathbf{\Upsilon}(\widehat{\delta}_k)\mathbf{V}_{-k}\frac{1}{\mathbf{\Delta}_{-k}^{-1}+\mathbf{V}_{-k}^T\mathbf{\Upsilon}(\widehat{\delta}_k)\mathbf{V}_{-k}}\mathbf{V}_{-k}^T\mathbf{\Upsilon}(\widehat{\delta}_k)\mathbf{v}_k\\&=\ \mathbf{v}_k^T\mathbf{\Upsilon}(t_k)\mathbf{V}_{-k}\frac{1}{\mathbf{\Delta}_{-k}^{-1}+\mathbf{V}_{-k}^T\mathbf{\Upsilon}(t_k)\mathbf{V}_{-k}}\mathbf{V}_{-k}^T\mathbf{\Upsilon}(t_k)\mathbf{v}_k+O\bigg(\frac{|\widehat{\delta}_k-t_k|}{|\delta_k|^6}\bigg).\end{aligned}\tag{C.29}$$

Plugging (C.27), (C.28) (with $z=\widehat{\delta}_k$), and (C.29) into (C.26) and using (B.16), we can deduce that

$$\Big|\mathbf{v}_k^T\big[\mathbf{\Upsilon}(\widehat{\delta}_k)-\mathbf{\Upsilon}(t_k)\big]\mathbf{v}_k\Big|\lesssim\frac{\sqrt{\log n}}{q|\delta_k|}\psi_n(\delta_k)\left(1+\frac{K}{|\delta_k|^4}\right)+\frac{|\widehat{\delta}_k-t_k|}{|\delta_k|^6}\quad\text{w.h.p.}\tag{C.30}$$

By Corollary 3.4 in Ajanki et al. (2017), $M_i(z)$ is the Stieltjes transform of a finite measure μ_i on $\mathbb{R}$ given by

$$M_i(z)=\int_{\mathbb{R}}\frac{\mu_i(\mathrm{d}x)}{x-z},\tag{C.31}$$

where the support of μ_i satisfies $\operatorname{supp}\{\mu_i\}\subset[-2\sqrt{\mathfrak{M}},2\sqrt{\mathfrak{M}}]$. Thus, $M_i(x)$ is strictly increasing in x on $(-\infty,-2\sqrt{\mathfrak{M}}]$ and $[2\sqrt{\mathfrak{M}},+\infty)$, respectively. Such property implies that

$$\Big|\mathbf{v}_k^T(\mathbf{\Upsilon}(\widehat{\delta}_k)-\mathbf{\Upsilon}(t_k))\mathbf{v}_k\Big|\gtrsim|\widehat{\delta}_k-t_k|/|\delta_k|^2.$$

Therefore, plugging this result into (C.30) and solving for $|\widehat{\delta}_k-t_k|$ yield the desired conclusion, which completes the proof of Theorem B.1.

C. *Proof of Theorem B.2*

We start by describing the main ideas of the proof. To study the asymptotic behavior of the spiked eigenvectors, we define the contour

$$\mathcal{C}_k:=\{z\in\mathbb{C}:|z-t_k|=ct_k\},\tag{C.32}$$

where $c = c(\epsilon_0) > 0$ is small enough such that $(1 \pm c)t_k \in \mathcal{I}_k$. Under part (iii) of Assumption B.1 and Theorem B.1, contour $\mathcal{C}_k$ encloses $\widehat{\delta}_k$ and no other eigenvalues of $\mathbf{X}$ w.h.p. Then, using Cauchy's integral formula, we can estimate the projections of $\widehat{\mathbf{v}}_k$ by evaluating the loop integral $\oint_{\mathcal{C}_k} \mathbf{u}^T(\mathbf{X} - z\mathbf{I})^{-1}\mathbf{v}dz$ for any deterministic vectors $\mathbf{u}$ and $\mathbf{v}$ in $\mathbb{R}^n$. In particular, by taking $\mathbf{v} = \mathbf{v}_k$, we will obtain an estimate of the quadratic form

$$\mathbf{u}^T(\mathbf{D}/\mathbf{\Lambda})^{-\alpha}\widehat{\mathbf{v}}_k\widehat{\mathbf{v}}_k^T(\mathbf{D}/\mathbf{\Lambda})^{-\alpha}\mathbf{v}_k. \tag{C.33}$$

If we further take $\mathbf{u} = \mathbf{v}_k$, we can get an estimate of $\mathbf{v}_k^T(\mathbf{D}/\mathbf{\Lambda})^{-\alpha}\widehat{\mathbf{v}}_k$. Then dividing (C.33) by $\mathbf{v}_k^T(\mathbf{D}/\mathbf{\Lambda})^{-\alpha}\widehat{\mathbf{v}}_k$ will conclude the proof.

Specifically, we first establish a contour integral representation for the quadratic form (C.33) above. To this end, let us define a new resolvent

$$\begin{aligned}\mathbf{G}_k(z) &:= \left(\mathbf{\Lambda}^{-\alpha}\widetilde{\mathbf{X}}\mathbf{\Lambda}^{-\alpha} - \delta_k\mathbf{v}_k\mathbf{v}_k^T - z(\mathbf{D}/\mathbf{\Lambda})^{2\alpha}\right)^{-1} = \left(\overline{\mathbf{W}} + \mathbf{V}_{-k}\mathbf{\Delta}_{-k}\mathbf{V}_{-k}^T - z(\mathbf{D}/\mathbf{\Lambda})^{2\alpha}\right)^{-1} \\ &= \mathbf{G}(z) - \mathbf{G}(z)\mathbf{V}_{-k}\frac{1}{\mathbf{\Delta}_{-k}^{-1} + \mathbf{V}_{-k}^T\mathbf{G}(z)\mathbf{V}_{-k}}\mathbf{V}_{-k}^T\mathbf{G}(z),\end{aligned} \tag{C.34}$$

where in the last step above, we have used the Woodbury matrix identity

$$(\mathbf{A} + \mathbf{U}\mathbf{B}\mathbf{V})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{U}(\mathbf{B}^{-1} + \mathbf{V}\mathbf{A}^{-1}\mathbf{U})^{-1}\mathbf{V}\mathbf{A}^{-1} \tag{C.35}$$

for any nonsingular matrices $\mathbf{A}, \mathbf{B}$ and any matrices $\mathbf{U}, \mathbf{V}$. Applying (C.35) again, we can write that

$$\begin{aligned}(\mathbf{X} - z\mathbf{I})^{-1} &= (\mathbf{D}/\mathbf{\Lambda})^{\alpha}(\mathbf{G}_k^{-1}(z) + \delta_k\mathbf{v}_k\mathbf{v}_k^T)^{-1}(\mathbf{D}/\mathbf{\Lambda})^{\alpha} \\ &= (\mathbf{D}/\mathbf{\Lambda})^{\alpha}\left(\mathbf{G}_k(z) - \mathbf{G}_k(z)\mathbf{v}_k\frac{1}{\delta_k^{-1} + \mathbf{v}_k^T\mathbf{G}_k(z)\mathbf{v}_k}\mathbf{v}_k^T\mathbf{G}_k(z)\right)(\mathbf{D}/\mathbf{\Lambda})^{\alpha}.\end{aligned}$$

Now, applying Cauchy's integral formula to $\oint_{\mathcal{C}_k} \mathbf{u}^T(\mathbf{X} - z\mathbf{I})^{-1}\mathbf{v}dz$, we can deduce that

$$\begin{aligned}\mathbf{u}^T(\mathbf{D}/\mathbf{\Lambda})^{-\alpha}\widehat{\mathbf{v}}_k\widehat{\mathbf{v}}_k^T(\mathbf{D}/\mathbf{\Lambda})^{-\alpha}\mathbf{v} &= -\frac{1}{2\pi i}\oint_{\mathcal{C}_k}\mathbf{u}^T(\mathbf{D}/\mathbf{\Lambda})^{-\alpha}(\mathbf{X} - z)^{-1}(\mathbf{D}/\mathbf{\Lambda})^{-\alpha}\mathbf{v}dz \\ = -\frac{1}{2\pi i}\oint_{\mathcal{C}_k}\mathbf{u}^T&\left(\mathbf{G}_k(z) - \mathbf{G}_k(z)\mathbf{v}_k\frac{1}{\delta_k^{-1} + \mathbf{v}_k^T\mathbf{G}_k(z)\mathbf{v}_k}\mathbf{v}_k^T\mathbf{G}_k(z)\right)\mathbf{v}dz\end{aligned}$$

for any deterministic vectors $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$. Using Weyl's inequality and Proposition C.1, we obtain that w.h.p.,

$$\left|\lambda_l(\overline{\mathbf{W}} + \mathbf{V}_{-k}\mathbf{\Delta}_{-k}\mathbf{V}_{-k}^T) - \lambda_l(\mathbf{V}_{-k}\mathbf{\Delta}_{-k}\mathbf{V}_{-k}^T)\right| \leq \|\overline{\mathbf{W}}\| = O(1),$$

where $\lambda_l(\cdot)$ denotes the lth eigenvalue of a given symmetric matrix. Then due to the eigengap condition in (B.15), the contour $\mathcal{C}_k$ does not enclose any eigenvalue of $\overline{\mathbf{W}} + \mathbf{V}_{-k}\mathbf{\Delta}_{-k}\mathbf{V}_{-k}^T$ w.h.p., i.e., $\mathbf{G}_k(z)$ is nonsingular in the regime enclosed by $\mathcal{C}_k$. Thus, it holds that w.h.p.,

$$\oint_{\mathcal{C}_k}\mathbf{x}^T\mathbf{G}_k(z)\mathbf{y}dz = 0,$$

which in turn leads to

$$\mathbf{u}^T(\mathbf{D}/\mathbf{\Lambda})^{-\alpha}\widehat{\mathbf{v}}_k\widehat{\mathbf{v}}_k^T(\mathbf{D}/\mathbf{\Lambda})^{-\alpha}\mathbf{v} = \frac{1}{2\pi i}\oint_{\mathcal{C}_k}\frac{\mathbf{u}^T\mathbf{G}_k(z)\mathbf{v}_k\mathbf{v}_k^T\mathbf{G}_k(z)\mathbf{v}}{\delta_k^{-1} + \mathbf{v}_k^T\mathbf{G}_k(z)\mathbf{v}_k}dz. \tag{C.36}$$

It remains to estimate the right-hand side (RHS) of (C.36) above. Since $\mathbf{\Upsilon}$ is a deterministic approximation of $\mathbf{G}(z)$ due to the local laws, $\mathbf{\Upsilon}_k$ in (B.19) is the corresponding deterministic approximation of $\mathbf{G}_k(z)$. We now control the differences between the bilinear forms of $\mathbf{G}_k(z)$ and $\mathbf{\Upsilon}_k(z)$ using the local law established in Theorem D.3. More precisely, applying (C.18) and (C.28), it holds that w.h.p.,

$$\begin{aligned}\mathbf{v}_k^T\left(\mathbf{G}_k(z) - \mathbf{\Upsilon}_k(z)\right)\mathbf{v}_k &\lesssim \frac{\sqrt{\log n}}{q|\delta_k|}\psi_n(\delta_k) + \frac{\sqrt{K\log n}}{q}\psi_n(\delta_k)\left(\frac{1}{|\delta_k|^3} + \frac{\sqrt{K}}{|\delta_k|^5} + \frac{\sqrt{K\log n}}{q|\delta_k|}\psi_n(\delta_k)\right) \\ &\lesssim \frac{\sqrt{\log n}}{q|\delta_k|}\psi_n(\delta_k)\left(1 + \frac{K}{|\delta_k|^4}\right)\end{aligned} \tag{C.37}$$

uniformly in $z \in \mathcal{C}_k$, where we have also used (B.16) in the second step. By (C.13) and (C.23), we have that

$$\mathbf{u}^T\mathbf{\Upsilon}(z)\mathbf{v}_k = -\mathbf{u}^T\mathbf{v}_k/z + O(|z|^{-3}), \quad \|\mathbf{u}^T\mathbf{\Upsilon}(z)\mathbf{V}_{-k}\| \lesssim \|\mathbf{u}^T\mathbf{V}_{-k}\|/|z| + |z|^{-3}, \tag{C.38}$$

$$\mathbf{u}^T\mathbf{\Upsilon}_k(z)\mathbf{v}_k = -\mathbf{u}^T\mathbf{v}_k/z + O(|z|^{-3}), \quad \mathbf{v}_k^T\mathbf{\Upsilon}_k'(z)\mathbf{v}_k = z^{-2} + O(|z|^{-4}). \tag{C.39}$$

Using the local law (D.21) below, we see that the following estimates

$$|\mathbf{u}^T(\mathbf{G}(z) - \mathbf{\Upsilon}(z))\mathbf{v}_k| \lesssim \frac{\sqrt{\log n}}{q|z|}\left(\frac{1}{|z|\beta_n} + \frac{\sqrt{\log n}}{q\beta_n^2} + \|\mathbf{v}_k\|_\infty\right), \tag{C.40}$$

$$\|\mathbf{u}^T(\mathbf{G}(z) - \mathbf{\Upsilon}(z))\mathbf{V}_{-k}\| \lesssim \frac{\sqrt{K\log n}}{q|z|}\left(\frac{1}{|z|\beta_n} + \frac{\sqrt{\log n}}{q\beta_n^2} + \|\mathbf{V}_{-k}\|_{\max}\right) \tag{C.41}$$

hold uniformly in $z \in S(\mathfrak{C})$ w.h.p.

Combining the estimates (C.38)–(C.41) with (C.15), (C.18)–(C.23), and (C.24), we can obtain that for any deterministic unit vector $\mathbf{u}$, with high probability,

$$\begin{aligned}\mathbf{u}^T(\mathbf{G}_k(z) - \mathbf{\Upsilon}_k(z))\mathbf{v}_k &\lesssim \frac{\sqrt{\log n}}{q|\delta_k|}\psi_n(\delta_k) + \frac{K\log n}{q^2|\delta_k|}\psi_n(\delta_k)^2 + \frac{\sqrt{K\log n}}{q|\delta_k|}\psi_n(\delta_k) \qquad \text{(C.42)}\\ &\quad\times\left(\|\mathbf{u}^T\mathbf{V}_{-k}\| + \frac{1}{|\delta_k|^2}\right)\left(1 + \frac{\sqrt{K}}{|\delta_k|^2}\right)\\ &\lesssim \frac{\sqrt{\log n}}{q|\delta_k|}\psi_n(\delta_k)\left[1 + \left(\|\mathbf{u}^T\mathbf{V}_{-k}\| + \frac{1}{|\delta_k|^2}\right)\left(\sqrt{K} + \frac{K}{|\delta_k|^2}\right)\right] \qquad \text{(C.43)}\end{aligned}$$

uniformly in $z \in \mathcal{C}_k$. Combining (C.39) with (C.43) yields that w.h.p.,

$$\mathbf{u}^T\mathbf{G}_k(z)\mathbf{v}_k \lesssim \frac{|\mathbf{u}^T\mathbf{v}_k|}{|\delta_k|} + \frac{1}{|\delta_k|^3} + \frac{\sqrt{\log n}}{q|\delta_k|}\psi_n(\delta_k)\left(1 + \sqrt{K}\|\mathbf{u}^T\mathbf{V}_{-k}\|\right) \tag{C.44}$$

uniformly in $z \in \mathcal{C}_k$.

We now estimate (C.36) for the case of $\mathbf{u} = \mathbf{v} = \mathbf{v}_k$. By (C.39), we see that for all $z \in \mathcal{C}_k$,

$$\mathbf{v}_k^T\mathbf{\Upsilon}_k(z)\mathbf{v}_k = -z^{-1} + O(|z|^{-3}), \tag{C.45}$$

which entails that

$$\min_{z\in\mathcal{C}_k}\left|1 + \delta_k\mathbf{v}_k^T\mathbf{\Upsilon}_k(z)\mathbf{v}_k\right| = \frac{c}{1+c} + o(1). \tag{C.46}$$

With (C.37) and (C.46), we can deduce from (C.36) that w.h.p.,

$$\begin{aligned}\mathbf{v}_k^T(\mathbf{D}/\mathbf{\Lambda})^{-\alpha}\widehat{\mathbf{v}}_k\widehat{\mathbf{v}}_k^T(\mathbf{D}/\mathbf{\Lambda})^{-\alpha}\mathbf{v}_k &= \frac{1}{2\pi i}\oint_{\mathcal{C}_k}\frac{(\mathbf{v}_k^T\mathbf{G}_k(z)\mathbf{v}_k)^2}{\delta_k^{-1} + \mathbf{v}_k^T\mathbf{G}_k(z)\mathbf{v}_k}\mathrm{d}z = \frac{1}{2\pi i\delta_k}\oint_{\mathcal{C}_k}\frac{1}{1 + \delta_k\mathbf{v}_k^T\mathbf{G}_k(z)\mathbf{v}_k}\mathrm{d}z\\ &= \frac{1}{2\pi i\delta_k}\oint_{\mathcal{C}_k}\frac{1}{1 + \delta_k\mathbf{v}_k^T\mathbf{\Upsilon}_k(z)\mathbf{v}_k}\mathrm{d}z + O\left(\frac{\sqrt{\log n}}{q}\psi_n(\delta_k)\left(1 + \frac{K}{|\delta_k|^4}\right)\right)\\ &= \frac{1}{\delta_k^2\mathbf{v}_k^T\mathbf{\Upsilon}_k'(t_k)\mathbf{v}_k} + O\left(\frac{\sqrt{\log n}}{q}\psi_n(\delta_k)\left(1 + \frac{K}{|\delta_k|^4}\right)\right), \qquad \text{(C.47)}\end{aligned}$$

where we have used Cauchy's residue theorem from complex analysis at the pole $z = t_k$ in the last step. Moreover, it follows from (B.23) and (C.39) that

$$\delta_k^2\mathbf{v}_k^T\mathbf{\Upsilon}_k'(t_k)\mathbf{v}_k = \delta_k^2/t_k^2 + O\left(\delta_k^2/t_k^4\right) = 1 + O(\delta_k^{-2}),$$

which yields the first estimate in (B.31). Now, taking the square root of (C.47) and using the first estimate in (B.31), we obtain (B.29). Also, note that plugging (B.23) into the first expression in (C.39) results in the second estimate in (B.31).

We next take $\mathbf{v} = \mathbf{v}_k$ in (C.36). With the aid of (C.44), (C.37), and (C.46), we obtain that w.h.p.,

$$\begin{aligned}\mathbf{u}^T(\mathbf{D}/\mathbf{\Lambda})^{-\alpha}\widehat{\mathbf{v}}_k\widehat{\mathbf{v}}_k^T(\mathbf{D}/\mathbf{\Lambda})^{-\alpha}\mathbf{v}_k &= -\frac{1}{2\pi i}\oint_{\mathcal{C}_k}\frac{\mathbf{u}^T\mathbf{G}_k(z)\mathbf{v}_k}{1 + \delta_k\mathbf{v}_k^T\mathbf{G}_k(z)\mathbf{v}_k}\mathrm{d}z\\ = -\frac{1}{2\pi i}\oint_{\mathcal{C}_k}\frac{\mathbf{u}^T\mathbf{G}_k(z)\mathbf{v}_k}{1 + \delta_k\mathbf{v}_k^T\mathbf{\Upsilon}_k(z)\mathbf{v}_k}\mathrm{d}z + \mathcal{E}_\mathbf{u} &= -\frac{\mathbf{u}^T\mathbf{G}_k(t_k)\mathbf{v}_k}{\delta_k\mathbf{v}_k^T\mathbf{\Upsilon}_k'(t_k)\mathbf{v}_k} + \mathcal{E}_\mathbf{u}, \qquad \text{(C.48)}\end{aligned}$$

where $\mathcal{E}_\mathbf{u}$ is a random error that can be bounded w.h.p. as

$$\mathcal{E}_\mathbf{u} \lesssim \frac{\sqrt{\log n}}{q}\psi_n(\delta_k)\left(1 + \frac{K}{|\delta_k|^4}\right)\left[|\mathbf{u}^T\mathbf{v}_k| + \frac{1}{|\delta_k|^2} + \frac{\sqrt{\log n}}{q}\psi_n(\delta_k)\left(1 + \sqrt{K}\|\mathbf{u}^T\mathbf{V}_{-k}\|\right)\right],$$

and we have used Cauchy's residue theorem at the pole $z = t_k$ in the last step. Then an application of (C.43) and (C.39) yields that w.h.p.,

$$\begin{aligned}
&\mathbf{u}^T(\mathbf{D}/\mathbf{\Lambda})^{-\alpha}\widehat{\mathbf{v}}_k\widehat{\mathbf{v}}_k^T(\mathbf{D}/\mathbf{\Lambda})^{-\alpha}\mathbf{v}_k \\
&= -\frac{\mathbf{u}^T\mathbf{\Upsilon}_k(t_k)\mathbf{v}_k}{\delta_k\mathbf{v}_k^T\mathbf{\Upsilon}_k'(t_k)\mathbf{v}_k} + \mathcal{E}_{\mathbf{u}} + O\left\{\frac{\sqrt{\log n}}{q}\psi_n(\delta_k)\left[1 + \left(\|\mathbf{u}^T\mathbf{V}_{-k}\| + \frac{1}{|\delta_k|^2}\right)\left(\sqrt{K} + \frac{K}{|\delta_k|^2}\right)\right]\right\} \\
&= -\frac{\mathbf{u}^T\mathbf{\Upsilon}_k(t_k)\mathbf{v}_k}{\delta_k\mathbf{v}_k^T\mathbf{\Upsilon}_k'(t_k)\mathbf{v}_k} + O\left\{\frac{\sqrt{\log n}}{q}\psi_n(\delta_k)\left[1 + \frac{K}{|\delta_k|^4} + \|\mathbf{u}^T\mathbf{V}_{-k}\|\left(\sqrt{K} + \frac{K}{|\delta_k|^2}\right)\right]\right\},
\end{aligned} \tag{C.49}$$

where we have used (B.16) to simplify the error term. Dividing (C.49) by (B.29) leads to (B.30). This concludes the proof of Theorem B.2.

D. *Proof of Theorem B.3*

Observe that the estimate in (B.33) is an immediate consequence of the slightly stronger result stated below, once the additional assumption (B.32) is incorporated.

Proposition C.2. *Under Condition B.1 and Assumption B.1, for each $1 \le k \le K_0$ and $i \in [n]$, the following estimate*

$$\begin{aligned}
\widehat{v}_k(i) &= (\Lambda_i/d_i)^\alpha v_k(i) + \frac{1}{t_k d_i^\alpha}\sum_{j\in[n]} W_{ij}\Lambda_j^{-\alpha}v_k(j) \\
&+ O\left(\|\mathbf{v}_k\|_\infty\left(\frac{1}{|\delta_k|^2} + \frac{\sqrt{\log n}\psi_n(\delta_k)}{q}\right) + \|\mathbf{V}_{-k}\|_{\max}\left(\frac{\sqrt{K}}{|\delta_k|^2} + \frac{K\sqrt{\log n}}{q}\psi_n(\delta_k)\right)\right) \\
&+ O\left(\frac{\sqrt{\log n}}{\sqrt{n}|\delta_k|}\left(\frac{1+\sqrt{K}|\delta_k|^{-1}}{|\delta_k|} + \frac{\sqrt{\log n}\beta_n^{-1} + K\sqrt{\log n}\psi_n(\delta_k)}{q}\right)\right)
\end{aligned} \tag{C.50}$$

holds w.h.p.

We now show how (B.34) follows from (B.33). By an application of Lemma C.1 together with the Taylor expansion, we have that with high probability,

$$(d_i/\Lambda_i)^{-\alpha} = 1 - \frac{\alpha}{q\beta_n\Lambda_i}\sum_j W_{ij} + O\left(\frac{\log n}{q^2\beta_n^2}\right).$$

Moreover, by (C.6), it holds with high probability that

$$\sum_{j\in[n]} W_{ij}\Lambda_j^{-\alpha}v_k(j) \lesssim \frac{\log n}{q}\|\mathbf{v}_k\|_\infty + \frac{\sqrt{\log n}}{\sqrt{n}}.$$

Combining these two estimates with (B.33), we obtain (B.34), which completes the proof of Theorem B.3. It remains to establish Proposition C.2.

Proof of Proposition C.2. We take $\mathbf{u} = \mathbf{e}_i^T$ and $\mathbf{v} = \mathbf{v}_k$ in (C.36). First, with the aid of Lemma C.3 and (C.23), we see that that for all $z \in \mathcal{C}_k$,

$$\mathbf{e}_i^T\mathbf{\Upsilon}\mathbf{v}_k \lesssim |\delta_k|^{-1}|\mathbf{v}_k(i)|, \quad |\mathbf{e}_i^T\mathbf{\Upsilon}\mathbf{V}_{-k}| \lesssim \sqrt{K}|\delta_k|^{-1}\|\mathbf{V}_{-k}(i)\|_{\max}, \tag{C.51}$$

$$\mathbf{e}_i^T\mathbf{\Upsilon}_k\mathbf{v}_k \lesssim |\delta_k|^{-1}|\mathbf{v}_k(i)| + \sqrt{K}|\delta_k|^{-3}\|\mathbf{V}_{-k}(i)\|_{\max}. \tag{C.52}$$

Then an application of (C.16)–(C.23), (C.24), and (C.51) yields that w.h.p.,

$$\begin{aligned}
|\mathbf{e}_i^T(\mathbf{G}_k - \mathbf{\Upsilon}_k)\mathbf{v}_k| &\le |\mathbf{e}_i^T(\mathbf{G} - \mathbf{\Upsilon})\mathbf{v}_k| + \left|\mathbf{e}_i^T\mathbf{\Upsilon}\mathbf{V}_{-k}\left(\frac{1}{\mathbf{\Delta}_{-k}^{-1} + \mathbf{V}_{-k}^T\mathbf{G}\mathbf{V}_{-k}} - \frac{1}{\mathbf{\Delta}_{-k}^{-1} + \mathbf{V}_{-k}^T\mathbf{\Upsilon}\mathbf{V}_{-k}}\right)\mathbf{V}_{-k}^T\mathbf{\Upsilon}\mathbf{v}_k\right| \\
&+ \left|\mathbf{e}_i^T\mathbf{\Upsilon}\mathbf{V}_{-k}\frac{1}{\mathbf{\Delta}_{-k}^{-1} + \mathbf{V}_{-k}^T\mathbf{G}\mathbf{V}_{-k}}\mathbf{V}_{-k}^T(\mathbf{\Upsilon} - \mathbf{G})\mathbf{v}_k\right| + \left|\mathbf{e}_i^T(\mathbf{G} - \mathbf{\Upsilon})\mathbf{V}_{-k}\frac{1}{\mathbf{\Delta}_{-k}^{-1} + \mathbf{V}_{-k}^T\mathbf{G}\mathbf{V}_{-k}}\mathbf{V}_{-k}^T\mathbf{G}\mathbf{v}_k\right| \\
&\lesssim \frac{\sqrt{\log n}}{|\delta_k|}\left(\frac{1}{\sqrt{n}|\delta_k|} + \left(\frac{\sqrt{\log n}}{q|\delta_k|} + \frac{1}{q\beta_n}\right)\|\mathbf{v}_k\|_\infty\right) + \|\mathbf{V}_{-k}\|_{\max}\frac{K^{3/2}\sqrt{\log n}}{q|\delta_k|^3}\psi_n(\delta_k)
\end{aligned} \tag{C.53}$$

$$
\begin{aligned}
&+\|\mathbf{V}_{-k}\|_{\max}\frac{K\sqrt{\log n}}{q|\delta_k|}\psi_n(\delta_k)\\
&+\sqrt{K\log n}\left(\frac{1}{\sqrt{n}|\delta_k|}+\left(\frac{\sqrt{\log n}}{q|\delta_k|}+\frac{1}{q\beta_n}\right)\|\mathbf{V}_{-k}\|_{\max}\right)\left(\frac{1}{|\delta_k|^3}+\frac{\sqrt{K\log n}}{q|\delta_k|}\psi_n(\delta_k)\right)\\
\lesssim&\left(1+\frac{\sqrt{K}}{|\delta_k|^2}\right)\left(\frac{\sqrt{\log n}}{\sqrt{n}|\delta_k|^2}+\|\mathbf{V}_{-k}\|_{\max}\frac{K\sqrt{\log n}}{q|\delta_k|}\psi_n(\delta_k)\right)+\frac{\sqrt{\log n}}{|\delta_k|}\left(\frac{\sqrt{\log n}}{q|\delta_k|}+\frac{1}{q\beta_n}\right)\|\mathbf{v}_k\|_\infty\\
&+\|\mathbf{V}_{-k}\|_{\max}\frac{\sqrt{K\log n}}{|\delta_k|^3}\left(\frac{\sqrt{\log n}}{q|\delta_k|}+\frac{1}{q\beta_n}\right)
\end{aligned}\tag{C.54}
$$

uniformly in $z\in\mathcal{C}_k\cup\mathcal{I}_k$, where we have also used condition (B.16) in the last step to simplify the estimate. Using (C.52) and (C.54), along with condition (B.16), we can deduce that w.h.p.,

$$
\begin{aligned}
\mathbf{e}_i^T\mathbf{G}_k(z)\mathbf{v}_k&=\mathbf{e}_i^T\boldsymbol{\Upsilon}_k(z)\mathbf{v}_k+\mathbf{e}_i^T\left[\mathbf{G}_k(z)-\boldsymbol{\Upsilon}_k(z)\right]\mathbf{v}_k\\
&\lesssim\frac{\|\mathbf{v}_k\|_\infty}{|\delta_k|}+\left(1+\frac{\sqrt{K}}{|\delta_k|^2}\right)\frac{\sqrt{\log n}}{\sqrt{n}|\delta_k|^2}+\|\mathbf{V}_{-k}\|_{\max}\left(\frac{\sqrt{K}}{|\delta_k|^3}+\frac{K\sqrt{\log n}}{q|\delta_k|}\psi_n(\delta_k)\right)
\end{aligned}\tag{C.55}
$$

uniformly in $z\in\mathcal{C}_k\cup\mathcal{I}_k$.

We are now ready to establish the asymptotic expansion of $\widehat{\mathbf{v}}_k(i)$. Taking $\mathbf{u}=\mathbf{e}_i$ and $\mathbf{v}=\mathbf{v}_k$ in (C.36) and applying (C.55), (C.37), and (C.46), we can show that w.h.p.,

$$
\begin{aligned}
\mathbf{e}_i^T(\mathbf{D}/\boldsymbol{\Lambda})^{-\alpha}\widehat{\mathbf{v}}_k\widehat{\mathbf{v}}_k^T(\mathbf{D}/\boldsymbol{\Lambda})^{-\alpha}\mathbf{v}_k&=-\frac{1}{2\pi i}\oint_{\mathcal{C}_k}\frac{\mathbf{e}_i^T\mathbf{G}_k(z)\mathbf{v}_k}{1+\delta_k\mathbf{v}_k^T\mathbf{G}_k(z)\mathbf{v}_k}\mathrm{d}z\\
=-\frac{1}{2\pi i}\oint_{\mathcal{C}_k}\frac{\mathbf{e}_i^T\mathbf{G}_k(z)\mathbf{v}_k}{1+\delta_k\mathbf{v}_k^T\boldsymbol{\Upsilon}_k(z)\mathbf{v}_k}\mathrm{d}z+\mathcal{E}_i&=-\frac{\mathbf{e}_i^T\mathbf{G}_k(t_k)\mathbf{v}_k}{\delta_k\mathbf{v}_k^T\boldsymbol{\Upsilon}_k'(t_k)\mathbf{v}_k}+\mathcal{E}_i,
\end{aligned}\tag{C.56}
$$

where $\mathcal{E}_i$ is a random error that can be bounded w.h.p. as

$$
\begin{aligned}
\mathcal{E}_i\lesssim|\delta_k|\frac{\sqrt{\log n}}{q}\psi_n(\delta_k)\left(1+\frac{K}{|\delta_k|^4}\right)&\left[\frac{\|\mathbf{v}_k\|_\infty}{|\delta_k|}+\left(1+\frac{\sqrt{K}}{|\delta_k|^2}\right)\frac{\sqrt{\log n}}{\sqrt{n}|\delta_k|^2}\right.\\
&\left.+\|\mathbf{V}_{-k}\|_{\max}\left(\frac{\sqrt{K}}{|\delta_k|^3}+\frac{K\sqrt{\log n}}{q|\delta_k|}\psi_n(\delta_k)\right)\right].
\end{aligned}
$$

Dividing (C.56) by the estimate of $\mathbf{v}_k^T(\mathbf{D}/\boldsymbol{\Lambda})^{-\alpha}\widehat{\mathbf{v}}_k$ in (B.29) and using (B.31) and (B.23), it holds that w.h.p.,

$$
(d_i/\Lambda_i)^{-\alpha}\widehat{v}_k(i)=-t_k\mathbf{e}_i^T\mathbf{G}_k(t_k)\mathbf{v}_k+\mathcal{E}_i',\tag{C.57}
$$

where $\mathcal{E}_i'$ is a random error satisfying that w.h.p.,

$$
\begin{aligned}
\mathcal{E}_i'\lesssim\left(\frac{1}{|\delta_k|}+\frac{\sqrt{\log n}|\delta_k|\psi_n(\delta_k)}{q}\right)&\left[\frac{\|\mathbf{v}_k\|_\infty}{|\delta_k|}+\left(1+\frac{\sqrt{K}}{|\delta_k|^2}\right)\frac{\sqrt{\log n}}{\sqrt{n}|\delta_k|^2}\right.\\
&\left.+\|\mathbf{V}_{-k}\|_{\max}\left(\frac{\sqrt{K}}{|\delta_k|^3}+\frac{K\sqrt{\log n}\psi_n(\delta_k)}{q|\delta_k|}\right)\right].
\end{aligned}
$$

Further, from the definition (C.34) of $\mathbf{G}_k$, we obtain that w.h.p.,

$$
\begin{aligned}
\mathbf{e}_i^T(\mathbf{G}(t_k)-\mathbf{G}_k(t_k))\mathbf{v}_k&=\mathbf{e}_i^T\mathbf{G}(t_k)\mathbf{V}_{-k}\frac{1}{\boldsymbol{\Delta}_{-k}^{-1}+\mathbf{V}_{-k}^T\mathbf{G}(t_k)\mathbf{V}_{-k}}\mathbf{V}_{-k}^T\mathbf{G}(t_k)\mathbf{v}_k\\
&\lesssim\left(\sqrt{K}\|\mathbf{V}_{-k}\|_{\max}+\frac{\sqrt{K\log n}}{\sqrt{n}|\delta_k|}\right)\left(\frac{1}{|\delta_k|^3}+\frac{\sqrt{K\log n}}{q|\delta_k|}\psi_n(\delta_k)\right),
\end{aligned}
$$

where we have used (C.17) and (C.51) to bound $\mathbf{e}_i^T\mathbf{G}(t_k)\mathbf{V}_{-k}$, (C.22) to bound $\mathbf{V}_{-k}^T\mathbf{G}(t_k)\mathbf{v}_k$, and (C.24) to bound the denominator. Plugging the above estimate into (C.57) and using (B.16), we can deduce that w.h.p.,

$$
\begin{aligned}
(d_i/\Lambda_i)^{-\alpha}\widehat{v}_k(i)=-t_k\mathbf{e}_i^T\mathbf{G}(t_k)\mathbf{v}_k&+O\left(\left(\frac{1}{|\delta_k|^2}+\frac{\sqrt{\log n}\psi_n(\delta_k)}{q}\right)\|\mathbf{v}_k\|_\infty\right)\\
&+O\left(\left(\sqrt{K}\|\mathbf{V}_{-k}\|_{\max}+\frac{\sqrt{K\log n}}{\sqrt{n}|\delta_k|}\right)\left(\frac{1}{|\delta_k|^2}+\frac{\sqrt{K\log n}}{q}\psi_n(\delta_k)\right)\right).
\end{aligned}\tag{C.58}
$$

We next handle the first term on the RHS of (C.58)

$$-t_k\mathbf{e}_i^T\mathbf{G}(t_k)\mathbf{v}_k = (d_i/\Lambda_i)^{-2\alpha}\mathbf{e}_i^T t_k(\mathbf{D}/\mathbf{\Lambda})^{2\alpha}\frac{1}{t_k(\mathbf{D}/\mathbf{\Lambda})^{2\alpha}-\overline{\mathbf{W}}}\mathbf{v}_k$$
$$= (d_i/\Lambda_i)^{-2\alpha}v_k(i) - (d_i/\Lambda_i)^{-2\alpha}\mathbf{e}_i^T\overline{\mathbf{W}}\mathbf{G}(t_k)\mathbf{v}_k. \tag{C.59}$$

Together with (C.58) and (C.5), it yields that w.h.p.,

$$\begin{aligned}\widehat{v}_k(i) &= (\Lambda_i/d_i)^{\alpha}v_k(i) - (\Lambda_i/d_i)^{\alpha}\mathbf{e}_i^T\overline{\mathbf{W}}\mathbf{G}(t_k)\mathbf{v}_k + O\left(\left(\frac{1}{|\delta_k|^2}+\frac{\sqrt{\log n}\psi_n(\delta_k)}{q}\right)\|\mathbf{v}_k\|_\infty\right)\\
&\quad + O\left(\left(\sqrt{K}\|\mathbf{V}_{-k}\|_{\max}+\frac{\sqrt{K\log n}}{\sqrt{n}|\delta_k|}\right)\left(\frac{1}{|\delta_k|^2}+\frac{\sqrt{K\log n}}{q}\psi_n(\delta_k)\right)\right)\\
&= (\Lambda_i/d_i)^{\alpha}v_k(i) - (\Lambda_i/d_i)^{\alpha}\mathbf{e}_i^T\overline{\mathbf{W}}\mathbf{\Upsilon}(t_k)\mathbf{v}_k\\
&\quad + O\left(\|\mathbf{v}_k\|_\infty\left(\frac{1}{|\delta_k|^2}+\frac{\sqrt{\log n}\psi_n(\delta_k)}{q}\right)+\|\mathbf{V}_{-k}\|_{\max}\left(\frac{\sqrt{K}}{|\delta_k|^2}+\frac{K\sqrt{\log n}}{q}\psi_n(\delta_k)\right)\right)\\
&\quad + O\left(\frac{\sqrt{\log n}}{\sqrt{n}|\delta_k|}\left(\frac{1+\sqrt{K}|\delta_k|^{-1}}{|\delta_k|}+\frac{\sqrt{\log n}\beta_n^{-1}+K\sqrt{\log n}\psi_n(\delta_k)}{q}\right)\right),\end{aligned} \tag{C.60}$$

where we have used the local law in Theorem D.4 below in the second step above.

Finally, recalling that $\mathbf{\Upsilon}(z)+z^{-1}=\boldsymbol{\mathcal{E}}_1=O(|z|^{-3})$ by (C.13) and using (C.6), we can obtain that

$$\begin{aligned}\Lambda_i^{\alpha}\mathbf{e}_i^T\overline{\mathbf{W}}\left(\mathbf{\Upsilon}(t_k)+t_k^{-1}\right)\mathbf{v}_k &= \mathbf{e}_i^T\mathbf{W}\mathbf{\Lambda}^{-\alpha}\boldsymbol{\mathcal{E}}_1(t_k)\mathbf{v}_k\\
&= \sum_{j\in[n]}W_{ij}\Lambda_j^{-\alpha}(\mathcal{E}_1(t_k))_{jj}v_k(j) \lesssim \frac{1}{|\delta_k|^3}\left(\frac{\sqrt{\log n}}{\sqrt{n}}+\frac{\log n}{q}\|\mathbf{v}_k\|_\infty\right)\end{aligned} \tag{C.61}$$

with high probability. Therefore, combining (C.4), (C.60), and (C.61) leads to (C.50), which completes the proof of Proposition C.2. ☐

E. *Proof of Theorem B.4*

To derive the asymptotic expansion of the spiked eigenvalue $\widehat{\delta}_k$, we again rely on (C.26), but now use a more accurate approximation of resolvent $\mathbf{G}$ via the Taylor expansion. We begin by applying (C.2), (C.4), (C.8), and (C.10), together with the Taylor expansion, to deduce that w.h.p.,

$$\begin{aligned}\mathbf{G}(z) &= -\left((\mathbf{D}/\mathbf{\Lambda})^{2\alpha}z\right)^{-1}\frac{1}{\mathbf{I}-\overline{\mathbf{W}}(\mathbf{D}/\mathbf{\Lambda})^{-2\alpha}z^{-1}}\\
&= -(\mathbf{D}/\mathbf{\Lambda})^{-2\alpha}z^{-1}-(\mathbf{D}/\mathbf{\Lambda})^{-2\alpha}\overline{\mathbf{W}}(\mathbf{D}/\mathbf{\Lambda})^{-2\alpha}z^{-2}-\overline{\mathbf{W}}^2z^{-3}+O\left(\frac{1}{|\delta_k|^4}+\frac{\sqrt{\log n}}{q|\delta_k|^3\beta_n}\right)\end{aligned} \tag{C.62}$$

uniformly in $z\in\mathcal{C}_k\cup\mathcal{I}_k$. Combining (C.26), (C.28), (C.29), and (C.62), we obtain that w.h.p.,

$$\begin{aligned}t_k^{-1}-\widehat{\delta}_k^{-1} &= \widehat{\delta}_k^{-1}\mathbf{v}_k^T\left((\mathbf{D}/\mathbf{\Lambda})^{-2\alpha}-\mathbf{I}+\widehat{\delta}_k^{-1}(\mathbf{D}/\mathbf{\Lambda})^{-2\alpha}\overline{\mathbf{W}}(\mathbf{D}/\mathbf{\Lambda})^{-2\alpha}+\widehat{\delta}_k^{-2}\overline{\mathbf{W}}^2\right)\mathbf{v}_k\\
&\quad + \mathbf{v}_k^T(t_k^{-1}+\mathbf{\Upsilon}(t_k))\mathbf{v}_k + O\left(\frac{1}{|\delta_k|^4}+\frac{\sqrt{\log n}}{q|\delta_k|^3\beta_n}\right)\\
&\quad + O\left(\frac{\sqrt{K\log n}}{q}\psi_n(\delta_k)\left(\frac{1}{|\delta_k|^3}+\frac{\sqrt{K\log n}}{q|\delta_k|}\psi_n(\delta_k)\right)+\frac{|\widehat{\delta}_k-t_k|}{|\delta_k|^6}\right).\end{aligned}$$

Together with (C.4), (C.8), and (B.27), this yields the following expansion of $\widehat{\delta}_k$ w.h.p.

$$\begin{aligned}\widehat{\delta}_k - t_k &= t_k\mathbf{v}_k^T\left((\mathbf{D}/\mathbf{\Lambda})^{-2\alpha}-\mathbf{I}+\widehat{\delta}_k^{-1}(\mathbf{D}/\mathbf{\Lambda})^{-2\alpha}\overline{\mathbf{W}}(\mathbf{D}/\mathbf{\Lambda})^{-2\alpha}+t_k^{-2}\overline{\mathbf{W}}^2\right)\mathbf{v}_k\\
&\quad + t_k^2\mathbf{v}_k^T(t_k^{-1}+\mathbf{\Upsilon}(t_k))\mathbf{v}_k\\
&\quad + O\left(\frac{1}{|\delta_k|^2}+\frac{\sqrt{\log n}}{q|\delta_k|\beta_n}+\frac{\sqrt{K\log n}|\delta_k|}{q}\psi_n(\delta_k)\left(\frac{1}{|\delta_k|^2}+\frac{\sqrt{K\log n}}{q}\psi_n(\delta_k)\right)\right).\end{aligned} \tag{C.63}$$

Next, using (B.27), (C.8), and Lemma C.1, we obtain that w.h.p.,

$$\left|\left(\frac{t_k}{\widehat{\delta}_k}-1\right)\mathbf{v}_k^T(\mathbf{D}/\mathbf{\Lambda})^{-2\alpha}\overline{\mathbf{W}}(\mathbf{D}/\mathbf{\Lambda})^{-2\alpha}\mathbf{v}_k\right| \lesssim \left(|\mathbf{v}_k^T\overline{\mathbf{W}}\mathbf{v}_k|+\frac{\sqrt{\log n}}{q\beta_n}\right)\frac{\sqrt{\log n}}{q}\psi_n(\delta_k)\left(1+\frac{K}{|\delta_k|^4}\right)$$

$$\lesssim \left(\frac{\log n}{q} \|\mathbf{v}_k\|_\infty^2 + \frac{\sqrt{\log n}}{q\beta_n} \right) \frac{\sqrt{\log n}}{q} \psi_n(\delta_k) \left(1 + \frac{K}{|\delta_k|^4} \right), \quad \text{(C.64)}$$

where in the second step, we have used the high-probability bound

$$\mathbf{v}_k^T \overline{\mathbf{W}} \mathbf{v}_k = \sum_{i,j \in [n]} v_k(i) \Lambda_i^{-\alpha} W_{ij} \Lambda_j^{-\alpha} v_k(j) \lesssim \frac{\log n}{q} \|\mathbf{v}_k\|_\infty^2 + \frac{\sqrt{\log n}}{\sqrt{n}} \quad \text{(C.65)}$$

following from (C.6). Plugging (C.64) into (C.63), and using (B.16) together with definition (B.28), we conclude that w.h.p.,

$$\widehat{\delta}_k - t_k = t_k \mathbf{v}_k^T \left((\mathbf{D}/\mathbf{\Lambda})^{-2\alpha} - \mathbf{I} + t_k^{-1} (\mathbf{D}/\mathbf{\Lambda})^{-2\alpha} \overline{\mathbf{W}} (\mathbf{D}/\mathbf{\Lambda})^{-2\alpha} + t_k^{-2} \overline{\mathbf{W}}^2 \right) \mathbf{v}_k + t_k^2 \mathbf{v}_k^T (t_k^{-1} + \mathbf{\Upsilon}(t_k)) \mathbf{v}_k$$
$$+ O\left(\frac{1}{|\delta_k|^2} + \frac{\sqrt{\log n}}{q|\delta_k|\beta_n} + \frac{\sqrt{\log n}}{q} \psi_n(\delta_k) \left(\frac{\sqrt{K}}{|\delta_k|} + \frac{K\sqrt{\log n}|\delta_k|}{q} \psi_n(\delta_k) + \frac{\log n}{q} \|\mathbf{v}_k\|_\infty^2 \right) \right). \quad \text{(C.66)}$$

We further apply the Taylor expansion to the factors $(\mathbf{D}/\mathbf{\Lambda})^{-2\alpha}$ appearing on the RHS of (C.66). Using Lemma C.1, (C.8), and (C.65), we can deduce that w.h.p,

$$t_k \mathbf{v}_k^T \left((\mathbf{D}/\mathbf{\Lambda})^{-2\alpha} - \mathbf{I} + t_k^{-1} (\mathbf{D}/\mathbf{\Lambda})^{-2\alpha} \overline{\mathbf{W}} (\mathbf{D}/\mathbf{\Lambda})^{-2\alpha} \right) \mathbf{v}_k = -2\alpha t_k \mathbf{v}_k^T \frac{\mathbf{D} - \mathbf{\Lambda}}{\mathbf{\Lambda}} \mathbf{v}_k + \mathbf{v}_k^T \overline{\mathbf{W}} \mathbf{v}_k$$
$$+ \alpha(2\alpha+1) t_k \mathbf{v}_k^T \frac{(\mathbf{D} - \mathbf{\Lambda})^2}{\mathbf{\Lambda}^2} \mathbf{v}_k - 4\alpha \mathbf{v}_k^T \frac{\mathbf{D} - \mathbf{\Lambda}}{\mathbf{\Lambda}} \overline{\mathbf{W}} \mathbf{v}_k + O\left(\frac{(\log n)^{3/2} |\delta_k|}{q^3 \beta_n^3} + \frac{\log n}{q^2 \beta_n^2} \right). \quad \text{(C.67)}$$

On the other hand, using (C.14), it is straightforward to verify that

$$\frac{1}{t_k} \mathbb{E} \mathbf{v}_k^T \overline{\mathbf{W}}^2 \mathbf{v}_k + t_k^2 \mathbf{v}_k^T (t_k^{-1} + \mathbf{\Upsilon}(t_k)) \mathbf{v}_k = O(|\delta_k|^{-3}). \quad \text{(C.68)}$$

Plugging (C.67) and (C.68) into (C.66), and using the following implications of Young's inequality

$$\frac{1}{|\delta_k|^2} + \frac{(\log n)^{3/2} |\delta_k|}{q^3 \beta_n^3} \gtrsim \frac{\sqrt{\log n}}{q |\delta_k| \beta_n}, \quad \frac{1}{|\delta_k|^2} + \frac{(\log n)^{3/2} |\delta_k|}{q^3 \beta_n^3} \gtrsim \frac{\log n}{q^2 \beta_n^2},$$

we can rewrite (C.66) in the form of (B.35), where the centered random error B_k is defined as

$$B_k := \alpha(2\alpha+1) t_k \mathbf{v}_k^T \frac{(\mathbf{D} - \mathbf{\Lambda})^2}{\mathbf{\Lambda}^2} \mathbf{v}_k - 4\alpha \mathbf{v}_k^T \frac{\mathbf{D} - \mathbf{\Lambda}}{\mathbf{\Lambda}} \overline{\mathbf{W}} \mathbf{v}_k + \frac{1}{t_k} \left(\mathbf{v}_k^T \overline{\mathbf{W}}^2 \mathbf{v}_k - \mathbb{E} \mathbf{v}_k^T \overline{\mathbf{W}}^2 \mathbf{v}_k \right) - A_k. \quad \text{(C.69)}$$

It remains to verify that the variances of the random quadratic terms in (C.69) satisfy bound (B.36). We therefore establish variance estimates for

$$\mathbf{u}^T \frac{(\mathbf{D} - \mathbf{\Lambda})^2}{\mathbf{\Lambda}^2} \mathbf{v}_k, \quad \mathbf{u}^T \frac{\mathbf{D} - \mathbf{\Lambda}}{\mathbf{\Lambda}} \overline{\mathbf{W}} \mathbf{v}_k, \quad \mathbf{u}^T \overline{\mathbf{W}} \frac{\mathbf{D} - \mathbf{\Lambda}}{\mathbf{\Lambda}} \mathbf{v}_k, \quad \mathbf{u}^T \overline{\mathbf{W}}^2 \mathbf{v}_k,$$

where $\mathbf{u}$ is an arbitrary deterministic unit vector. We state these bounds in this general form since they will be invoked again in subsequent arguments. To control the variance of $\mathbf{u}^T ((\mathbf{D} - \mathbf{\Lambda})/\mathbf{\Lambda})^2 \mathbf{v}_k$, we first estimate each contribution arising in the variance expansion. Specifically, for each $i, j, l, s \in [n]$, let us consider

$$\mathbb{E} W_{ij} W_{ls} \mathbf{u}^T \frac{(\mathbf{D} - \mathbf{\Lambda})^2}{\mathbf{\Lambda}^2} \mathbf{v}_k.$$

Using (B.12) and (C.4), together with direct computations, it holds that

$$\mathbb{E} W_{ij} W_{ls} \mathbf{u}^T \frac{(\mathbf{D} - \mathbf{\Lambda})^2}{\mathbf{\Lambda}^2} \mathbf{v}_k \lesssim \frac{\|\mathbf{u}\|_\infty \|\mathbf{v}_k\|_\infty}{q^2 n^3 \beta_n^2} \quad \text{if } \{i,j\} \cap \{l,s\} = \emptyset,$$

$$\mathbb{E} W_{ij} W_{is} \mathbf{u}^T \frac{(\mathbf{D} - \mathbf{\Lambda})^2}{\mathbf{\Lambda}^2} \mathbf{v}_k \lesssim \frac{|u(i) v_k(i)|}{q^2 n^2 \beta_n^2} + \frac{\|\mathbf{u}\|_\infty \|\mathbf{v}_k\|_\infty}{q^2 n^3 \beta_n^2} \quad \text{if } j \neq s,$$

$$\mathbb{E} W_{ij}^2 \mathbf{u}^T \frac{(\mathbf{D} - \mathbf{\Lambda})^2}{\mathbf{\Lambda}^2} \mathbf{v}_k = s_{ij} \mathbb{E} \mathbf{v}_k^T \frac{(\mathbf{D} - \mathbf{\Lambda})^2}{\mathbf{\Lambda}^2} \mathbf{v}_k + O\left(\frac{|u(i) v_k(i)| + |u(j) v_k(j)| + n^{-1} \|\mathbf{u}\|_\infty \|\mathbf{v}_k\|_\infty}{q^4 n \beta_n^2} \right).$$

By counting the occurrences of the above cases in $\mathbb{E}(\mathbf{u}^T ((\mathbf{D} - \mathbf{\Lambda})/\mathbf{\Lambda})^2 \mathbf{v}_k)^2$ and summing the corresponding contributions, we deduce that

$$\operatorname{var}\left(\mathbf{u}^T\frac{(\mathbf{D}-\mathbf{\Lambda})^2}{\mathbf{\Lambda}^2}\mathbf{v}_k\right)\lesssim\frac{\|\mathbf{u}\|_\infty\|\mathbf{v}_k\|_\infty}{q^4\beta_n^4}+\frac{1}{q^4n^2\beta_n^4}. \tag{C.70}$$

Using similar arguments, we can derive variance bounds for the remaining quadratic forms. For brevity, we state the results below and omit the technical details

$$\operatorname{var}\left(\mathbf{u}^T\frac{\mathbf{D}-\mathbf{\Lambda}}{\mathbf{\Lambda}}\overline{\mathbf{W}}\mathbf{v}_k\right)+\operatorname{var}\left(\mathbf{u}^T\overline{\mathbf{W}}\frac{\mathbf{D}-\mathbf{\Lambda}}{\mathbf{\Lambda}}\mathbf{v}_k\right)\lesssim\frac{\|\mathbf{u}\|_\infty\|\mathbf{v}_k\|_\infty}{q^2\beta_n^2}+\frac{1}{q^2n\beta_n^2}, \tag{C.71}$$

$$\operatorname{var}\left(\mathbf{u}^T\overline{\mathbf{W}}^2\mathbf{v}_k\right)\lesssim\frac{1}{q\sqrt{n}}+\frac{\|\mathbf{u}\|_\infty\|\mathbf{v}_k\|_\infty}{q^2}. \tag{C.72}$$

Finally, setting $\mathbf{u}=\mathbf{v}_k$ in (C.70)–(C.72) and recalling the definition of B_k in (C.69), we can obtain the desired bound (B.36), thereby completing the proof of Theorem B.4.

F. Proof of Theorem B.5

We now derive the asymptotic expansions for the spiked eigenvector projections $\mathbf{u}^T(\mathbf{D}/\mathbf{\Lambda})^{-\alpha}\widehat{\mathbf{v}}_k$. The argument is still based on (C.36), but we estimate $\mathbf{G}_k$ via the more accurate expansion (C.62). As already observed in Fan et al. (2022a) for the special case of $\alpha=0$, the asymptotic variance of $\mathbf{v}_k^T(\mathbf{D}/\mathbf{\Lambda})^{-\alpha}\widehat{\mathbf{v}}_k$ is much smaller than that of $\mathbf{u}^T(\mathbf{D}/\mathbf{\Lambda})^{-\alpha}\widehat{\mathbf{v}}_k$ when $\mathbf{u}$ is *not* parallel to $\mathbf{v}_k$. Consequently, we treat these two cases separately in the proof. We begin with part 2) of Theorem B.5.

Taking $\mathbf{u}=\mathbf{v}_k$ in (C.36) and applying Cauchy's residue theorem, local law (C.43), bound (C.46), and condition (B.16), we can deduce that w.h.p.,

$$\begin{aligned}\mathbf{v}_k^T(\mathbf{D}/\mathbf{\Lambda})^{-\alpha}\widehat{\mathbf{v}}_k\widehat{\mathbf{v}}_k^T(\mathbf{D}/\mathbf{\Lambda})^{-\alpha}\mathbf{v}_k&=\delta_k^{-2}(\mathbf{v}_k^T\mathbf{\Upsilon}_k'(t_k)\mathbf{v}_k)^{-1}-\frac{1}{2\pi i}\oint_{\mathcal{C}_k}\frac{\mathbf{v}_k^T(\mathbf{G}_k(z)-\mathbf{\Upsilon}_k(z))\mathbf{v}_k}{(1+\delta_k\mathbf{v}_k^T\mathbf{\Upsilon}_k(z)\mathbf{v}_k)^2}\mathrm{d}z\\&\quad+O\left(\frac{1}{|\delta_k|^4}+\frac{\log n}{q^2}\psi_n(\delta_k)^2\right).\end{aligned} \tag{C.73}$$

To estimate $\mathbf{v}_k^T(\mathbf{G}_k(z)-\mathbf{\Upsilon}_k(z))\mathbf{v}_k$, we first use a similar argument as in (C.54) to get that w.h.p.,

$$\begin{aligned}|\mathbf{v}_k^T\left[(\mathbf{G}_k-\mathbf{G})-(\mathbf{\Upsilon}_k-\mathbf{\Upsilon})\right]\mathbf{v}_k|&\le\left|\mathbf{v}_k^T\mathbf{\Upsilon}\mathbf{V}_{-k}\left(\frac{1}{\mathbf{\Delta}_{-k}^{-1}+\mathbf{V}_{-k}^T\mathbf{G}\mathbf{V}_{-k}}-\frac{1}{\mathbf{\Delta}_{-k}^{-1}+\mathbf{V}_{-k}^T\mathbf{\Upsilon}\mathbf{V}_{-k}}\right)\mathbf{V}_{-k}^T\mathbf{\Upsilon}\mathbf{v}_k\right|\\&\quad+\left|\mathbf{v}_k^T\mathbf{\Upsilon}\mathbf{V}_{-k}\frac{1}{\mathbf{\Delta}_{-k}^{-1}+\mathbf{V}_{-k}^T\mathbf{G}\mathbf{V}_{-k}}\mathbf{V}_{-k}^T(\mathbf{\Upsilon}-\mathbf{G})\mathbf{v}_k\right|+\left|\mathbf{v}_k^T(\mathbf{G}-\mathbf{\Upsilon})\mathbf{V}_{-k}\frac{1}{\mathbf{\Delta}_{-k}^{-1}+\mathbf{V}_{-k}^T\mathbf{G}\mathbf{V}_{-k}}\mathbf{V}_{-k}^T\mathbf{G}\mathbf{v}_k\right|\\&\lesssim\frac{K\sqrt{\log n}}{q|\delta_k|^5}\psi_n(\delta_k)+\frac{\sqrt{K\log n}}{q|\delta_k|^3}\psi_n(\delta_k)+\frac{K\log n}{q^2|\delta_k|}\psi_n(\delta_k)^2.\end{aligned} \tag{C.74}$$

where we have used (C.18)–(C.23), (C.24), and (C.15) in the derivation. Next, to estimate $\mathbf{v}_k^T(\mathbf{G}(z)-\mathbf{\Upsilon}(z))\mathbf{v}_k$, we will resort to a refinement of (C.62). Applying (C.2), (C.4), (C.8), and (C.10) along with the Taylor expansion yields that

$$\begin{aligned}\mathbf{G}(z)=&-(\mathbf{D}/\mathbf{\Lambda})^{-2\alpha}z^{-1}-(\mathbf{D}/\mathbf{\Lambda})^{-2\alpha}\overline{\mathbf{W}}(\mathbf{D}/\mathbf{\Lambda})^{-2\alpha}z^{-2}-(\mathbf{D}/\mathbf{\Lambda})^{-2\alpha}\overline{\mathbf{W}}(\mathbf{D}/\mathbf{\Lambda})^{-2\alpha}\overline{\mathbf{W}}(\mathbf{D}/\mathbf{\Lambda})^{-2\alpha}z^{-3}\\&-\overline{\mathbf{W}}^3z^{-4}+O\left(\frac{1}{|\delta_k|^5}+\frac{\sqrt{\log n}}{q|\delta_k|^4\beta_n}\right).\end{aligned} \tag{C.75}$$

Plugging (C.74) and (C.75) into (C.73), and denoting

$$\mathcal{A}(z):=(\mathbf{D}/\mathbf{\Lambda})^{-2\alpha}+(\mathbf{D}/\mathbf{\Lambda})^{-2\alpha}\mathbf{W}(\mathbf{D}/\mathbf{\Lambda})^{-2\alpha}z^{-1}+(\mathbf{D}/\mathbf{\Lambda})^{-2\alpha}(\overline{\mathbf{W}}(\mathbf{D}/\mathbf{\Lambda})^{-2\alpha})^2z^{-2}+\overline{\mathbf{W}}^3z^{-3},$$

it follows that w.h.p.,

$$\begin{aligned}\mathbf{v}_k^T(\mathbf{D}/\mathbf{\Lambda})^{-\alpha}\widehat{\mathbf{v}}_k\widehat{\mathbf{v}}_k^T(\mathbf{D}/\mathbf{\Lambda})^{-\alpha}\mathbf{v}_k&=\delta_k^{-2}(\mathbf{v}_k^T\mathbf{\Upsilon}_k'(t_k)\mathbf{v}_k)^{-1}+\frac{1}{2\pi i}\oint_{\mathcal{C}_k}\frac{\mathbf{v}_k^T(\mathbf{\Upsilon}(z)+z^{-1}\mathcal{A}(z))\mathbf{v}_k}{(1+\delta_k\mathbf{v}_k^T\mathbf{\Upsilon}_k(z)\mathbf{v}_k)^2}\mathrm{d}z\\&\quad+O\left(\frac{1}{|\delta_k|^4}+\frac{K\log n}{q^2}\psi_n(\delta_k)^2\right).\end{aligned} \tag{C.76}$$

Applying Cauchy's residue theorem to the above contour integral leads to

$$\begin{aligned}
&\frac{1}{2\pi i}\oint_{\mathcal{C}_k}\frac{\mathbf{v}_k^T(\boldsymbol{\Upsilon}(z)+z^{-1}\mathcal{A}(z))\mathbf{v}_k}{(1+\delta_k\mathbf{v}_k^T\boldsymbol{\Upsilon}_k(z)\mathbf{v}_k)^2}\mathrm{d}z\\
&=\frac{[\mathbf{v}_k^T(\boldsymbol{\Upsilon}'(t_k)-t_k^{-2}\mathcal{A}(t_k)+t_k^{-1}\mathcal{A}'(t_k))\mathbf{v}_k][\mathbf{v}_k^T\boldsymbol{\Upsilon}_k'(t_k)\mathbf{v}_k]-[\mathbf{v}_k^T(\boldsymbol{\Upsilon}(t_k)+t_k^{-1}\mathcal{A}(t_k))\mathbf{v}_k][\mathbf{v}_k^T\boldsymbol{\Upsilon}_k''(t_k)\mathbf{v}_k]}{\delta_k^2(\mathbf{v}_k^T\boldsymbol{\Upsilon}_k'(t_k)\mathbf{v}_k)^3}\\
&=t_k^2\mathbf{v}_k^T(\boldsymbol{\Upsilon}'(t_k)-t_k^{-2}\mathcal{A}(t_k)+t_k^{-1}\mathcal{A}'(t_k))\mathbf{v}_k+2t_k\mathbf{v}_k^T(\boldsymbol{\Upsilon}(t_k)+t_k^{-1}\mathcal{A}(t_k))\mathbf{v}_k\\
&\quad+O\left(\frac{1}{|\delta_k|^4}+\frac{\sqrt{\log n}}{q|\delta_k|^2\beta_n}\|\mathbf{v}_k\|_\infty+\frac{\log n}{q^2|\delta_k|^2\beta_n^2}+\frac{\log n}{q|\delta_k|^3}\|\mathbf{v}_k\|_\infty^2\right),
\end{aligned}\tag{C.77}$$

where in the second step, we have used (B.23) and the following estimates (all holding with high probability)

$$\mathbf{v}_k^T\boldsymbol{\Upsilon}_k'(t_k)\mathbf{v}_k=t_k^{-2}+O(|\delta_k|^{-4}),\quad \mathbf{v}_k^T\boldsymbol{\Upsilon}_k''(t_k)\mathbf{v}_k=-2t_k^{-3}+O(|\delta_k|^{-5}),\tag{C.78}$$

$$\left|\mathbf{v}_k^T(\boldsymbol{\Upsilon}(t_k)+t_k^{-1}\mathcal{A}(t_k))\mathbf{v}_k\right|\lesssim\frac{1}{|\delta_k|^3}+\frac{\sqrt{\log n}}{q|\delta_k|\beta_n}\|\mathbf{v}_k\|_\infty+\frac{\log n}{q^2|\delta_k|\beta_n^2}+\frac{\log n}{q|\delta_k|^2}\|\mathbf{v}_k\|_\infty^2,\tag{C.79}$$

$$\left|\mathbf{v}_k^T(\boldsymbol{\Upsilon}'(t_k)-t_k^{-2}\mathcal{A}(t_k)+t_k^{-1}\mathcal{A}'(t_k))\mathbf{v}_k\right|\lesssim\frac{1}{|\delta_k|^4}+\frac{\sqrt{\log n}}{q|\delta_k|^2\beta_n}\|\mathbf{v}_k\|_\infty+\frac{\log n}{q^2|\delta_k|^2\beta_n^2}+\frac{\log n}{q|\delta_k|^3}\|\mathbf{v}_k\|_\infty^2,\tag{C.80}$$

obtained from Lemma C.1, Proposition C.1, Lemma C.3, and estimate (C.65). In deriving (C.79)–(C.80), we have also used the following high probability bound

$$\mathbf{v}_k^T\left[(\mathbf{D}/\boldsymbol{\Lambda})^{-2\alpha}-\mathbf{I}\right]\mathbf{v}_k=-\frac{2\alpha}{q\beta_n}\sum_{i,j\in[n]}|v_k(i)|^2W_{ij}+O\left(\frac{\log n}{q^2\beta_n^2}\right)\lesssim\frac{\sqrt{\log n}}{q\beta_n}\|\mathbf{v}_k\|_\infty+\frac{\log n}{q^2\beta_n^2}\tag{C.81}$$

derived from the Taylor expansion, Lemma C.1, and (C.6). Then substituting (C.77) into (C.82) gives that

$$\begin{aligned}
&\mathbf{v}_k^T(\mathbf{D}/\boldsymbol{\Lambda})^{-\alpha}\widehat{\mathbf{v}}_k\widehat{\mathbf{v}}_k^T(\mathbf{D}/\boldsymbol{\Lambda})^{-\alpha}\mathbf{v}_k\\
&=\delta_k^{-2}(\mathbf{v}_k^T\boldsymbol{\Upsilon}_k'(t_k)\mathbf{v}_k)^{-1}+\mathcal{E}_k+O\left(\frac{1}{|\delta_k|^4}+\frac{K\log n}{q^2}\psi_n(\delta_k)^2+\frac{\sqrt{\log n}}{q|\delta_k|^2\beta_n}\|\mathbf{v}_k\|_\infty+\frac{\log n}{q|\delta_k|^3}\|\mathbf{v}_k\|_\infty^2\right),
\end{aligned}\tag{C.82}$$

where the random error $\mathcal{E}_k$ is defined as

$$\begin{aligned}
\mathcal{E}_k&:=\mathbf{v}_k^T\left[t_k^2\boldsymbol{\Upsilon}'(t_k)+2t_k\boldsymbol{\Upsilon}(t_k)+(\mathbf{D}/\boldsymbol{\Lambda})^{-2\alpha}-t_k^{-2}(\mathbf{D}/\boldsymbol{\Lambda})^{-2\alpha}\left(\overline{\mathbf{W}}(\mathbf{D}/\boldsymbol{\Lambda})^{-2\alpha}\right)^2-2t_k^{-3}\overline{\mathbf{W}}^3\right]\mathbf{v}_k\\
&=2\mathbf{v}_k^T\left[(\mathbf{D}/\boldsymbol{\Lambda})^{-\alpha}-\mathbf{I}\right]\mathbf{v}_k+\mathbf{v}_k^T\left[(\mathbf{D}/\boldsymbol{\Lambda})^{-\alpha}-\mathbf{I}\right]^2\mathbf{v}_k+\mathbf{v}_k^T\left[t_k^2\boldsymbol{\Upsilon}'(t_k)+2t_k\boldsymbol{\Upsilon}(t_k)+\mathbf{I}\right]\mathbf{v}_k\\
&\quad-\mathbf{v}_k\left[t_k^{-2}(\mathbf{D}/\boldsymbol{\Lambda})^{-2\alpha}\left(\overline{\mathbf{W}}(\mathbf{D}/\boldsymbol{\Lambda})^{-2\alpha}\right)^2+2t_k^{-3}\overline{\mathbf{W}}^3\right]\mathbf{v}_k.
\end{aligned}$$

We now take the square root of (C.82). In light of Lemma C.3 and Proposition C.1, we have that w.h.p.,

$$\mathbf{v}_k^T\left[t_k^2\boldsymbol{\Upsilon}'(t_k)+2t_k\boldsymbol{\Upsilon}(t_k)+\mathbf{I}\right]\mathbf{v}_k\lesssim|\delta_k|^{-2},\quad\left|\mathbf{v}_k^T(\mathbf{D}/\boldsymbol{\Lambda})^{-2\alpha}\left(\overline{\mathbf{W}}(\mathbf{D}/\boldsymbol{\Lambda})^{-2\alpha}\right)^2\mathbf{v}_k\right|+\left|\mathbf{v}_k^T\overline{\mathbf{W}}^3\mathbf{v}_k\right|\lesssim 1.\tag{C.83}$$

On the other hand, using the Taylor expansion of $(\mathbf{D}/\boldsymbol{\Lambda})^{-\alpha}-\mathbf{I}$, together with Lemma C.1 and estimates analogous to (C.81), we can deduce that w.h.p.,

$$\mathbf{v}_k^T\left[(\mathbf{D}/\boldsymbol{\Lambda})^{-\alpha}-\mathbf{I}\right]\mathbf{v}_k\lesssim\frac{\sqrt{\log n}}{q\beta_n}\|\mathbf{v}_k\|_\infty+\frac{\log n}{q^2\beta_n^2},\quad\mathbf{v}_k^T\left[(\mathbf{D}/\boldsymbol{\Lambda})^{-\alpha}-\mathbf{I}\right]^2\mathbf{v}_k\lesssim\frac{\log n}{q^2\beta_n^2},\tag{C.84}$$

$$\mathbf{v}_k^T\left[(\mathbf{D}/\boldsymbol{\Lambda})^{-\alpha}-\mathbf{I}\right]^2\mathbf{v}_k=\alpha^2\mathbf{v}_k^T\left(\frac{\mathbf{D}-\boldsymbol{\Lambda}}{\boldsymbol{\Lambda}}\right)^2\mathbf{v}_k-\alpha^2(\alpha+1)\mathbf{v}_k^T\left(\frac{\mathbf{D}-\boldsymbol{\Lambda}}{\boldsymbol{\Lambda}}\right)^3\mathbf{v}_k+O\left(\frac{(\log n)^2}{q^4\beta_n^4}\right).\tag{C.85}$$

Taking the square root of (C.82) and using the above bounds yields that

$$\begin{aligned}
&\mathbf{v}_k^T(\mathbf{D}/\boldsymbol{\Lambda})^{-\alpha}\widehat{\mathbf{v}}_k-\mathbf{v}_k^T(\mathbf{D}/\boldsymbol{\Lambda})^{-\alpha}\mathbf{v}_k=\frac{\alpha^2}{2}\mathbf{v}_k^T\left(\frac{\mathbf{D}-\boldsymbol{\Lambda}}{\boldsymbol{\Lambda}}\right)^2\mathbf{v}_k-\frac{\alpha^2(\alpha+1)}{2}\mathbf{v}_k^T\left(\frac{\mathbf{D}-\boldsymbol{\Lambda}}{\boldsymbol{\Lambda}}\right)^3\mathbf{v}_k\\
&-\frac{t_k^{-2}}{2}\mathbf{v}_k^T(\mathbf{D}/\boldsymbol{\Lambda})^{-2\alpha}\left(\overline{\mathbf{W}}(\mathbf{D}/\boldsymbol{\Lambda})^{-2\alpha}\right)^2\mathbf{v}_k-t_k^{-3}\mathbf{v}_k^T\overline{\mathbf{W}}^3\mathbf{v}_k+(\delta_k^2\mathbf{v}_k^T\boldsymbol{\Upsilon}_k'(t_k)\mathbf{v}_k)^{-1/2}-1\\
&+\frac{1}{2}\mathbf{v}_k^T(t_k^2\boldsymbol{\Upsilon}'(t_k)+2t_k\boldsymbol{\Upsilon}(t_k)+\mathbf{I})\mathbf{v}_k+O\left(\frac{1}{|\delta_k|^4}+\frac{K\log n}{q^2}\psi_n(\delta_k)^2+\left(\frac{\log n}{q^2\beta_n^2}+\frac{\log n}{q|\delta_k|^3}\right)\|\mathbf{v}_k\|_\infty^2\right).
\end{aligned}\tag{C.86}$$

This establishes (B.38) in view of the following asymptotic expansion

$$\begin{aligned}&\mathbf{v}_k^T(\mathbf{D}/\mathbf{\Lambda})^{-2\alpha}\overline{\mathbf{W}}(\mathbf{D}/\mathbf{\Lambda})^{-2\alpha}\overline{\mathbf{W}}(\mathbf{D}/\mathbf{\Lambda})^{-2\alpha}\mathbf{v}_k\\ &=\mathbf{v}_k^T\overline{\mathbf{W}}^2\mathbf{v}_k-4\alpha\mathbf{v}_k^T\left(\frac{\mathbf{D}-\mathbf{\Lambda}}{\mathbf{\Lambda}}\right)\overline{\mathbf{W}}^2\mathbf{v}_k-2\alpha\mathbf{v}_k^T\overline{\mathbf{W}}\left(\frac{\mathbf{D}-\mathbf{\Lambda}}{\mathbf{\Lambda}}\right)\overline{\mathbf{W}}\mathbf{v}_k+O\left(\frac{\log n}{q^2\beta_n^2}\right)\end{aligned} \tag{C.87}$$

w.h.p. from Lemma C.1.

It remains to bound the second moments of the cubic terms in (C.86) and (C.87). By arguments analogous to those used in the proof of (C.70) (with technical details omitted), it holds that

$$\mathbb{E}\left(\mathbf{v}_k^T\left(\frac{\mathbf{D}-\mathbf{\Lambda}}{\mathbf{\Lambda}}\right)^3\mathbf{v}_k\right)^2\lesssim\frac{n^2\|\mathbf{v}_k\|_\infty^4}{q^8\beta_n^6}, \tag{C.88}$$

$$\mathbb{E}\left(\mathbf{v}_k^T\left(\frac{\mathbf{D}-\mathbf{\Lambda}}{\mathbf{\Lambda}}\right)\overline{\mathbf{W}}^2\mathbf{v}_k\right)^2+\mathbb{E}\left(\mathbf{v}_k^T\overline{\mathbf{W}}\left(\frac{\mathbf{D}-\mathbf{\Lambda}}{\mathbf{\Lambda}}\right)\overline{\mathbf{W}}\mathbf{v}_k\right)^2\lesssim\frac{n^2\|\mathbf{v}_k\|_\infty^4}{q^4\beta_n^4}, \tag{C.89}$$

$$\mathbb{E}\left(\mathbf{v}_k^T\overline{\mathbf{W}}^3\mathbf{v}_k\right)^2\lesssim\frac{n^2\|\mathbf{v}_k\|_\infty^4}{q^2}. \tag{C.90}$$

Hence, combining (C.86)–(C.90) completes the proof of part 2) of Theorem B.5.

We now proceed to the proof of part 1) of Theorem B.5. We begin by estimating the integral term in (C.36) for a general vector $\mathbf{u}$ and $\mathbf{v}=\mathbf{v}_k$. By Cauchy's residue theorem, it holds that

$$\mathbf{u}^T(\mathbf{D}/\mathbf{\Lambda})^{-\alpha}\widehat{\mathbf{v}}_k\widehat{\mathbf{v}}_k^T(\mathbf{D}/\mathbf{\Lambda})^{-\alpha}\mathbf{v}_k=-\frac{1}{2\pi i}\oint_{\mathcal{C}_k}\frac{\mathbf{u}^T\mathbf{G}_k(z)\mathbf{v}_k}{1+\delta_k\mathbf{v}_k^T\mathbf{G}_k(z)\mathbf{v}_k}\mathrm{d}z. \tag{C.91}$$

For convenience, let us define

$$\mathbf{A}(z):=(\mathbf{D}/\mathbf{\Lambda})^{-2\alpha}+z^{-1}(\mathbf{D}/\mathbf{\Lambda})^{-2\alpha}\overline{\mathbf{W}}(\mathbf{D}/\mathbf{\Lambda})^{-2\alpha}+z^{-2}\overline{\mathbf{W}}^2.$$

In view of (C.62), we have that w.h.p.,

$$\mathbf{G}(z)=-z^{-1}\mathbf{A}(z)+O\left(\frac{1}{|\delta_k|^4}+\frac{\sqrt{\log n}}{q|\delta_k|^3\beta_n}\right) \tag{C.92}$$

uniformly for $z\in\mathcal{C}_k\cup\mathcal{I}_k$. Combining this with (C.43)–(C.46), we can deduce that w.h.p.,

$$\begin{aligned}\frac{\mathbf{u}^T\mathbf{G}_k(z)\mathbf{v}_k}{1+\delta_k\mathbf{v}_k^T\mathbf{G}_k(z)\mathbf{v}_k}=&\frac{\mathbf{u}^T\mathbf{G}_k(z)\mathbf{v}_k}{1+\delta_k\mathbf{v}_k^T\mathbf{\Upsilon}_k(z)\mathbf{v}_k}-\delta_k\frac{z^{-1}\mathbf{u}^T\mathbf{A}(z)\mathbf{v}_k\mathbf{v}_k^T(z^{-1}\mathbf{A}(z)+\mathbf{\Upsilon}(z))\mathbf{v}_k}{(1+\delta_k\mathbf{v}_k^T\mathbf{\Upsilon}_k(z)\mathbf{v}_k)^2}\\ &+O\left(\frac{1}{|\delta_k|^4}+\frac{\sqrt{\log n}}{q|\delta_k|^3\beta_n}\right),\end{aligned} \tag{C.93}$$

again uniformly on $z\in\mathcal{C}_k\cup\mathcal{I}_k$.

To estimate the contour integral of the second term on the RHS of (C.93), we can apply Cauchy's residue theorem to get

$$\begin{aligned}&\frac{\delta_k}{2\pi i}\oint_{\mathcal{C}_k}\frac{z^{-1}\mathbf{u}^T\mathbf{A}(z)\mathbf{v}_k\mathbf{v}_k^T(z^{-1}\mathbf{A}(z)+\mathbf{\Upsilon}(z))\mathbf{v}_k}{(1+\delta_k\mathbf{v}_k^T\mathbf{\Upsilon}_k(z)\mathbf{v}_k)^2}\mathrm{d}z\\ &=\frac{1}{\delta_k(\mathbf{v}_k^T\mathbf{\Upsilon}_k'(t_k)\mathbf{v}_k)^2}\left.\frac{\partial(z^{-1}\mathbf{u}^T\mathbf{A}(z)\mathbf{v}_k\mathbf{v}_k^T(z^{-1}\mathbf{A}(z)+\mathbf{\Upsilon}(z))\mathbf{v}_k)}{\partial z}\right|_{z=t_k}\\ &\quad-\frac{\delta_k}{t_k}\mathbf{u}^T\mathbf{A}(t_k)\mathbf{v}_k\mathbf{v}_k^T(t_k^{-1}\mathbf{A}(t_k)+\mathbf{\Upsilon}(t_k))\mathbf{v}_k\frac{\mathbf{v}_k^T\mathbf{\Upsilon}_k''(t_k)\mathbf{v}_k}{\delta_k^2(\mathbf{v}_k^T\mathbf{\Upsilon}_k'(t_k)\mathbf{v}_k)^3}.\end{aligned} \tag{C.94}$$

With the same arguments as those in (C.79), (C.80), (C.81), and (C.65), we can derive the following bounds w.h.p.

$$\left|\mathbf{v}_k^T(\mathbf{\Upsilon}(t_k)+t_k^{-1}\mathbf{A}(t_k))\mathbf{v}_k\right|\lesssim\frac{1}{|\delta_k|^3}+\frac{\sqrt{\log n}}{q|\delta_k|\beta_n}\|\mathbf{v}_k\|_\infty+\frac{\log n}{q^2|\delta_k|\beta_n^2}+\frac{\log n}{q|\delta_k|^2}\|\mathbf{v}_k\|_\infty^2, \tag{C.95}$$

$$\left|\mathbf{v}_k^T(\mathbf{\Upsilon}'(t_k)-t_k^{-2}\mathbf{A}(t_k)+t_k^{-1}\mathbf{A}'(t_k))\mathbf{v}_k\right|\lesssim\frac{1}{|\delta_k|^4}+\frac{\sqrt{\log n}}{q|\delta_k|^2\beta_n}\|\mathbf{v}_k\|_\infty+\frac{\log n}{q^2|\delta_k|^2\beta_n^2}+\frac{\log n}{q|\delta_k|^3}\|\mathbf{v}_k\|_\infty^2, \tag{C.96}$$

$$\mathbf{u}^T\left[(\mathbf{D}/\mathbf{\Lambda})^{-2\alpha}-\mathbf{I}\right]\mathbf{v}_k\lesssim\frac{\sqrt{\log n}}{q\beta_n}\|\mathbf{v}_k\|_\infty+\frac{\log n}{q^2\beta_n^2},\quad \mathbf{u}^T\overline{\mathbf{W}}\mathbf{v}_k\lesssim\frac{\log n}{q}\|\mathbf{v}_k\|_\infty+\frac{\sqrt{\log n}}{\sqrt{n}}. \tag{C.97}$$

Using these estimates together with (B.23) and (C.78), we can now bound the first term on the RHS of (C.94). Specifically, it holds that w.h.p.,

$$
\begin{aligned}
&\frac{1}{\delta_k(\mathbf{v}_k^T\boldsymbol{\Upsilon}_k'(t_k)\mathbf{v}_k)^2}\frac{\partial(z^{-1}\mathbf{u}^T\mathbf{A}(z)\mathbf{v}_k\mathbf{v}_k^T(z^{-1}\mathbf{A}(z)+\boldsymbol{\Upsilon}(z))\mathbf{v}_k)}{\partial z}\bigg|_{z=t_k}\\
&=-\mathbf{u}^T\mathbf{v}_k\mathbf{v}_k^T\left(t_k\boldsymbol{\Upsilon}(t_k)+\mathbf{A}(t_k)\right)\mathbf{v}_k+\mathbf{u}^T\mathbf{v}_k\mathbf{v}_k^T(t_k^2\boldsymbol{\Upsilon}'(t_k)-\mathbf{A}(t_k)+t_k\mathbf{A}'(t_k))\mathbf{v}_k\\
&\quad+O\left(\left(\frac{1}{|\delta_k|}+\frac{\sqrt{\log n}}{q\beta_n}\|\mathbf{v}_k\|_\infty+\frac{\log n}{q^2\beta_n^2}\right)\left(\frac{1}{|\delta_k|^2}+\frac{\sqrt{\log n}}{q\beta_n}\|\mathbf{v}_k\|_\infty+\frac{\log n}{q^2\beta_n^2}+\frac{\log n}{q|\delta_k|}\|\mathbf{v}_k\|_\infty^2\right)\right)\\
&=\mathbf{u}^T\mathbf{v}_k\mathbf{v}_k^T\left(4\alpha\frac{\mathbf{D}-\boldsymbol{\Lambda}}{\boldsymbol{\Lambda}}-2\alpha(2\alpha+1)\left(\frac{\mathbf{D}-\boldsymbol{\Lambda}}{\boldsymbol{\Lambda}}\right)^2-3t_k^{-1}(\mathbf{D}/\boldsymbol{\Lambda})^{-2\alpha}\overline{\mathbf{W}}(\mathbf{D}/\boldsymbol{\Lambda})^{-2\alpha}-4t_k^{-2}\overline{\mathbf{W}}^2\right)\mathbf{v}_k\\
&\quad+\mathbf{u}^T\mathbf{v}_k\mathbf{v}_k^T\left(t_k^2\boldsymbol{\Upsilon}'(t_k)-t_k\boldsymbol{\Upsilon}(t_k)-2\right)\mathbf{v}_k+O\left(\frac{1}{|\delta_k|^3}+\frac{(\log n)^{3/2}}{q^3\beta_n^3}\right)\\
&\quad+O\left(\frac{\sqrt{\log n}}{q|\delta_k|^2\beta_n}\|\mathbf{v}_k\|_\infty+\left(\frac{\log n}{q|\delta_k|^2}+\frac{\log n}{q^2\beta_n^2}\right)\|\mathbf{v}_k\|_\infty^2+\frac{(\log n)^{3/2}}{q^2|\delta_k|\beta_n}\|\mathbf{v}_k\|_\infty^3\right),
\end{aligned}
\tag{C.98}
$$

where in the second step we have used the Taylor expansion of $(\mathbf{D}/\boldsymbol{\Lambda})^{-2\alpha}$ in $\mathbf{A}(z)$, together with Lemma C.1 and Proposition C.1. Similarly, by resorting to (C.95), (C.97), (C.78), (B.23), Lemma C.1, and Proposition C.1, we can estimate the second term on the RHS of (C.94) w.h.p.

$$
\begin{aligned}
&-\frac{\delta_k}{t_k}\mathbf{u}^T\mathbf{A}(t_k)\mathbf{v}_k\mathbf{v}_k^T\left(t_k^{-1}\mathbf{A}(t_k)+\boldsymbol{\Upsilon}(t_k)\right)\mathbf{v}_k\frac{\mathbf{v}_k^T\boldsymbol{\Upsilon}_k''(t_k)\mathbf{v}_k}{\delta_k^2(\mathbf{v}_k^T\boldsymbol{\Upsilon}_k'(t_k)\mathbf{v}_k)^3}=2\mathbf{u}^T\mathbf{v}_k\mathbf{v}_k^T\left(\mathbf{A}(t_k)+t_k\boldsymbol{\Upsilon}(t_k)\right)\mathbf{v}_k\\
&\quad+O\left(\left(\frac{1}{|\delta_k|}+\frac{\sqrt{\log n}}{q\beta_n}\|\mathbf{v}_k\|_\infty+\frac{\log n}{q^2\beta_n^2}\right)\left(\frac{1}{|\delta_k|^2}+\frac{\sqrt{\log n}}{q\beta_n}\|\mathbf{v}_k\|_\infty+\frac{\log n}{q^2\beta_n^2}+\frac{\log n}{q|\delta_k|}\|\mathbf{v}_k\|_\infty^2\right)\right)\\
&=2\mathbf{u}^T\mathbf{v}_k\mathbf{v}_k\left(-2\alpha\frac{\mathbf{D}-\boldsymbol{\Lambda}}{\boldsymbol{\Lambda}}+\alpha(2\alpha+1)\left(\frac{\mathbf{D}-\boldsymbol{\Lambda}}{\boldsymbol{\Lambda}}\right)^2+t_k^{-1}(\mathbf{D}/\boldsymbol{\Lambda})^{-2\alpha}\overline{\mathbf{W}}(\mathbf{D}/\boldsymbol{\Lambda})^{-2\alpha}+t_k^{-2}\overline{\mathbf{W}}^2\right)\mathbf{v}_k\\
&\quad+2\mathbf{u}^T\mathbf{v}_k\mathbf{v}_k^T(1+t_k\boldsymbol{\Upsilon}(t_k))\mathbf{v}_k+O\left(\frac{1}{|\delta_k|^3}+\frac{(\log n)^{3/2}}{q^3\beta_n^3}\right)\\
&\quad+O\left(\frac{\sqrt{\log n}}{q|\delta_k|^2\beta_n}\|\mathbf{v}_k\|_\infty+\left(\frac{\log n}{q|\delta_k|^2}+\frac{\log n}{q^2\beta_n^2}\right)\|\mathbf{v}_k\|_\infty^2+\frac{(\log n)^{3/2}}{q^2|\delta_k|\beta_n}\|\mathbf{v}_k\|_\infty^3\right).
\end{aligned}
\tag{C.99}
$$

It remains to estimate the contour integral of the first term on the RHS of (C.93). An application of Cauchy's residue theorem gives that w.h.p.,

$$
\begin{aligned}
&\frac{1}{2\pi i}\oint_{\mathcal{C}_k}\frac{\mathbf{u}^T\mathbf{G}_k(z)\mathbf{v}_k}{1+\delta_k\mathbf{v}_k^T\boldsymbol{\Upsilon}_k(z)\mathbf{v}_k}dz=\frac{\mathbf{u}^T\mathbf{G}_k(t_k)\mathbf{v}_k}{\delta_k\mathbf{v}_k^T\boldsymbol{\Upsilon}_k'(t_k)\mathbf{v}_k}\\
&=\frac{1}{\delta_k\mathbf{v}_k^T\boldsymbol{\Upsilon}_k'(t_k)\mathbf{v}_k}\left[\mathbf{u}^T\mathbf{G}(t_k)\mathbf{v}_k-\mathbf{u}^T\mathbf{G}(t_k)\mathbf{V}_{-k}\frac{1}{\boldsymbol{\Delta}_{-k}^{-1}+\mathbf{V}_{-k}^T\mathbf{G}(t_k)\mathbf{V}_{-k}}\mathbf{V}_{-k}^T\mathbf{G}(t_k)\mathbf{v}_k\right]\\
&=\frac{-t_k^{-1}}{\delta_k\mathbf{v}_k^T\boldsymbol{\Upsilon}_k'(t_k)\mathbf{v}_k}\mathbf{u}^T\left(\mathbf{I}-2\alpha\frac{\mathbf{D}-\boldsymbol{\Lambda}}{\boldsymbol{\Lambda}}+\alpha(2\alpha+1)\left(\frac{\mathbf{D}-\boldsymbol{\Lambda}}{\boldsymbol{\Lambda}}\right)^2+t_k^{-1}(\mathbf{D}/\boldsymbol{\Lambda})^{-2\alpha}\overline{\mathbf{W}}(\mathbf{D}/\boldsymbol{\Lambda})^{-2\alpha}+t_k^{-2}\overline{\mathbf{W}}^2\right)\mathbf{v}_k\\
&\quad-t_k^{-1}\mathbf{u}^T\mathbf{V}_{-k}\frac{1}{\boldsymbol{\Delta}_{-k}^{-1}-t_k^{-1}\mathbf{I}}\mathbf{V}_{-k}^T\left(-2\alpha\frac{\mathbf{D}-\boldsymbol{\Lambda}}{\boldsymbol{\Lambda}}+\alpha(2\alpha+1)\left(\frac{\mathbf{D}-\boldsymbol{\Lambda}}{\boldsymbol{\Lambda}}\right)^2+t_k^{-1}(\mathbf{D}/\boldsymbol{\Lambda})^{-2\alpha}\overline{\mathbf{W}}(\mathbf{D}/\boldsymbol{\Lambda})^{-2\alpha}+t_k^{-2}\overline{\mathbf{W}}^2\right)\mathbf{v}_k\\
&\quad+O\left(\frac{1}{|\delta_k|^3}+\frac{(\log n)^{3/2}}{q^3\beta_n^3}+\frac{K\sqrt{\log n}}{q|\delta_k|^2}\psi_n(\delta_k)+\frac{K^{3/2}\log n}{q^2}\psi_n(\delta_k)^2\right),
\end{aligned}
\tag{C.100}
$$

where in the last step we have used (C.78), (C.24), expansion (C.62), the Taylor expansion of $(\mathbf{D}/\boldsymbol{\Lambda})^{-2\alpha}$, Lemma C.1, Proposition C.1, and Lemma C.3, together with local laws (C.19)–(C.22) and (C.41).

Now, combining (C.91), (C.93), (C.94), (C.98), (C.99), and (C.100), we can deduce that w.h.p.,

$$
\begin{aligned}
&\mathbf{u}^T(\mathbf{D}/\boldsymbol{\Lambda})^{-\alpha}\widehat{\mathbf{v}}_k\widehat{\mathbf{v}}_k^T(\mathbf{D}/\boldsymbol{\Lambda})^{-\alpha}\mathbf{v}_k\\
&=\frac{t_k^{-1}}{\delta_k\mathbf{v}_k^T\boldsymbol{\Upsilon}_k'(t_k)\mathbf{v}_k}\mathbf{u}^T\left(\mathbf{I}-2\alpha\frac{\mathbf{D}-\boldsymbol{\Lambda}}{\boldsymbol{\Lambda}}+\alpha(2\alpha+1)\left(\frac{\mathbf{D}-\boldsymbol{\Lambda}}{\boldsymbol{\Lambda}}\right)^2+t_k^{-1}(\mathbf{D}/\boldsymbol{\Lambda})^{-2\alpha}\overline{\mathbf{W}}(\mathbf{D}/\boldsymbol{\Lambda})^{-2\alpha}+t_k^{-2}\overline{\mathbf{W}}^2\right)\mathbf{v}_k
\end{aligned}
$$

$$+ t_k^{-1}\mathbf{u}^T\mathbf{V}_{-k}\frac{1}{\mathbf{\Delta}_{-k}^{-1}-t_k^{-1}\mathbf{I}}\mathbf{V}_{-k}^T\left(-2\alpha\frac{\mathbf{D}-\mathbf{\Lambda}}{\mathbf{\Lambda}}+\alpha(2\alpha+1)\left(\frac{\mathbf{D}-\mathbf{\Lambda}}{\mathbf{\Lambda}}\right)^2+t_k^{-1}(\mathbf{D}/\mathbf{\Lambda})^{-2\alpha}\overline{\mathbf{W}}(\mathbf{D}/\mathbf{\Lambda})^{-2\alpha}+t_k^{-2}\overline{\mathbf{W}}^2\right)\mathbf{v}_k$$
$$-\mathbf{u}^T\mathbf{v}_k\mathbf{v}_k^T\left(t_k^{-1}(\mathbf{D}/\mathbf{\Lambda})^{-2\alpha}\overline{\mathbf{W}}(\mathbf{D}/\mathbf{\Lambda})^{-2\alpha}+2t_k^{-2}\overline{\mathbf{W}}^2\right)\mathbf{v}_k+\mathbf{u}^T\mathbf{v}_k\mathbf{v}_k^T\left(t_k^2\mathbf{\Upsilon}'(t_k)+t_k\mathbf{\Upsilon}(t_k)\right)\mathbf{v}_k$$
$$+O\left(\frac{1}{|\delta_k|^3}+\frac{(\log n)^{3/2}}{q^3\beta_n^3}+\frac{K\sqrt{\log n}}{q|\delta_k|^2}\psi_n(\delta_k)+\frac{K^{3/2}\log n}{q^2}\psi_n(\delta_k)^2\right)$$
$$+O\left(\left(\frac{\log n}{q^2\beta_n^2}+\frac{\log n}{q|\delta_k|^2}\right)\|\mathbf{v}_k\|_\infty^2+\frac{(\log n)^{3/2}}{q^2|\delta_k|\beta_n}\|\mathbf{v}_k\|_\infty^3\right). \tag{C.101}$$

Next, using (C.83)–(C.85), (C.65), Lemma C.1, and Proposition C.1, and applying the Taylor expansion to $\mathbf{v}_k^T[(\mathbf{D}/\mathbf{\Lambda})^{-\alpha}-\mathbf{I}]\mathbf{v}_k$, we can rewrite equation (C.86) into a slightly coarser form

$$\mathbf{v}_k^T(\mathbf{D}/\mathbf{\Lambda})^{-\alpha}\widehat{\mathbf{v}}_k=(\delta_k^2\mathbf{v}_k^T\mathbf{\Upsilon}_k'(t_k)\mathbf{v}_k)^{-1/2}+\frac{1}{2}\mathbf{v}_k^T(t_k^2\mathbf{\Upsilon}'(t_k)+2t_k\mathbf{\Upsilon}(t_k)+\mathbf{I})\mathbf{v}_k \tag{C.102}$$
$$-\alpha\mathbf{v}_k^T\frac{\mathbf{D}-\mathbf{\Lambda}}{\mathbf{\Lambda}}\mathbf{v}_k+\frac{\alpha(2\alpha+1)}{2}\mathbf{v}_k^T\left(\frac{\mathbf{D}-\mathbf{\Lambda}}{\mathbf{\Lambda}}\right)^2\mathbf{v}_k-\frac{t_k^{-2}}{2}\mathbf{v}_k^T\overline{\mathbf{W}}^2\mathbf{v}_k$$
$$+O\left(\frac{1}{|\delta_k|^3}+\frac{(\log n)^{3/2}}{q^3\beta_n^3}+\frac{K\log n}{q^2}\psi_n(\delta_k)^2+\left(\frac{\log n}{q^2\beta_n^2}+\frac{\log n}{q|\delta_k|^3}\right)\|\mathbf{v}_k\|_\infty^2\right),$$

where the resulting error terms are one order larger.

Dividing (C.101) by (C.102) yields the following asymptotic expansion w.h.p.

$$\mathbf{u}^T(\mathbf{D}/\mathbf{\Lambda})^{-\alpha}\widehat{\mathbf{v}}_k-\frac{\mathbf{u}^T\mathbf{v}_k}{t_k(\mathbf{v}_k^T\mathbf{\Upsilon}_k(t_k)\mathbf{v}_k)^{1/2}} \tag{C.103}$$
$$=-\frac{1}{2}\mathbf{u}^T\mathbf{v}_k\mathbf{v}_k^T\left(t_k^2\mathbf{\Upsilon}'(t_k)+2t_k\mathbf{\Upsilon}(t_k)+\mathbf{I}-2\alpha\frac{\mathbf{D}-\mathbf{\Lambda}}{\mathbf{\Lambda}}+\alpha(2\alpha+1)\left(\frac{\mathbf{D}-\mathbf{\Lambda}}{\mathbf{\Lambda}}\right)^2-t_k^{-2}\overline{\mathbf{W}}^2\right)\mathbf{v}_k$$
$$+\mathbf{u}^T\left(-2\alpha\frac{\mathbf{D}-\mathbf{\Lambda}}{\mathbf{\Lambda}}+\alpha(2\alpha+1)\left(\frac{\mathbf{D}-\mathbf{\Lambda}}{\mathbf{\Lambda}}\right)^2+t_k^{-1}(\mathbf{D}/\mathbf{\Lambda})^{-2\alpha}\overline{\mathbf{W}}(\mathbf{D}/\mathbf{\Lambda})^{-2\alpha}+t_k^{-2}\overline{\mathbf{W}}^2\right)\mathbf{v}_k$$
$$+t_k^{-1}\mathbf{u}^T\mathbf{V}_{-k}\frac{1}{\mathbf{\Delta}_{-k}^{-1}-t_k^{-1}\mathbf{I}}\mathbf{V}_{-k}^T\left(-2\alpha\frac{\mathbf{D}-\mathbf{\Lambda}}{\mathbf{\Lambda}}+\alpha(2\alpha+1)\left(\frac{\mathbf{D}-\mathbf{\Lambda}}{\mathbf{\Lambda}}\right)^2+t_k^{-1}(\mathbf{D}/\mathbf{\Lambda})^{-2\alpha}\overline{\mathbf{W}}(\mathbf{D}/\mathbf{\Lambda})^{-2\alpha}+t_k^{-2}\overline{\mathbf{W}}^2\right)\mathbf{v}_k$$
$$-\mathbf{u}^T\mathbf{v}_k\mathbf{v}_k^T\left(t_k^{-1}(\mathbf{D}/\mathbf{\Lambda})^{-2\alpha}\overline{\mathbf{W}}(\mathbf{D}/\mathbf{\Lambda})^{-2\alpha}+2t_k^{-2}\overline{\mathbf{W}}^2\right)\mathbf{v}_k+\mathbf{u}^T\mathbf{v}_k\mathbf{v}_k^T\left(t_k^2\mathbf{\Upsilon}'(t_k)+t_k\mathbf{\Upsilon}(t_k)\right)\mathbf{v}_k$$
$$+O\left(\frac{1}{|\delta_k|^3}+\frac{(\log n)^{3/2}}{q^3\beta_n^3}+\frac{K\sqrt{\log n}}{q|\delta_k|^2}\psi_n(\delta_k)+\frac{K^{3/2}\log n}{q^2}\psi_n(\delta_k)^2\right)$$
$$+O\left(\left(\frac{\log n}{q^2\beta_n^2}+\frac{\log n}{q|\delta_k|^2}\right)\|\mathbf{v}_k\|_\infty^2+\frac{(\log n)^{3/2}}{q^2|\delta_k|\beta_n}\|\mathbf{v}_k\|_\infty^3\right).$$

As a special case, when $\mathbf{u}^T\mathbf{v}_k=0$, reorganizing the terms in (C.103) gives that

$$\mathbf{u}^T(\mathbf{D}/\mathbf{\Lambda})^{-\alpha}\widehat{\mathbf{v}}_k=\mathbf{u}^T\left(\mathbf{I}+\mathbf{V}_{-k}\frac{\mathbf{\Delta}_{-k}}{t_k\mathbf{I}-\mathbf{\Delta}_{-k}}\mathbf{V}_{-k}^T\right)\left(-2\alpha\frac{\mathbf{D}-\mathbf{\Lambda}}{\mathbf{\Lambda}}+\alpha(2\alpha+1)\left(\frac{\mathbf{D}-\mathbf{\Lambda}}{\mathbf{\Lambda}}\right)^2+\frac{\overline{\mathbf{W}}}{t_k}+\frac{\overline{\mathbf{W}}^2}{t_k^2}\right)\mathbf{v}_k$$
$$-\frac{2\alpha}{t_k}\mathbf{u}^T\left(\mathbf{I}+\mathbf{V}_{-k}\frac{\mathbf{\Delta}_{-k}}{t_k\mathbf{I}-\mathbf{\Delta}_{-k}}\mathbf{V}_{-k}^T\right)\left(\frac{\mathbf{D}-\mathbf{\Lambda}}{\mathbf{\Lambda}}\overline{\mathbf{W}}+\overline{\mathbf{W}}\frac{\mathbf{D}-\mathbf{\Lambda}}{\mathbf{\Lambda}}\right)\mathbf{v}_k$$
$$+O\left(\frac{1}{|\delta_k|^3}+\frac{(\log n)^{3/2}}{q^3\beta_n^3}+\frac{K\sqrt{\log n}}{q|\delta_k|^2}\psi_n(\delta_k)+\frac{K^{3/2}\log n}{q^2}\psi_n(\delta_k)^2\right)$$
$$+O\left(\left(\frac{\log n}{q^2\beta_n^2}+\frac{\log n}{q|\delta_k|^2}\right)\|\mathbf{v}_k\|_\infty^2+\frac{(\log n)^{3/2}}{q^2|\delta_k|\beta_n}\|\mathbf{v}_k\|_\infty^3\right).$$

Here, we have also applied the Taylor expansion of $(\mathbf{D}/\mathbf{\Lambda})^{-2\alpha}\overline{\mathbf{W}}(\mathbf{D}/\mathbf{\Lambda})^{-2\alpha}$ and bounded the resulting error terms using Lemma C.1 and (C.65). Combining the above estimate with the variance bounds in (C.70)–(C.72) leads to the desired conclusion in part 1) of Theorem B.5.

Finally, we prove parts 3) and 4) of Theorem B.5. We again rely on the contour integral representation in

(C.36), with $\mathbf{u}$ replaced by $(\mathbf{D}/\mathbf{\Lambda})^\alpha \mathbf{u}$. It follows from Cauchy's residue theorem that

$$\begin{aligned}\mathbf{u}^T\widehat{\mathbf{v}}_k\widehat{\mathbf{v}}_k^T(\mathbf{D}/\mathbf{\Lambda})^{-\alpha}\mathbf{v}_k &= -\frac{1}{2\pi i}\oint_{\mathcal{C}_k}\frac{\mathbf{u}^T(\mathbf{D}/\mathbf{\Lambda})^\alpha\mathbf{G}_k(z)\mathbf{v}_k}{1+\delta_k\mathbf{v}_k^T\mathbf{G}_k(z)\mathbf{v}_k}\mathrm{d}z \\ &= -\frac{1}{2\pi i}\oint_{\mathcal{C}_k}\frac{\mathbf{u}^T\mathbf{G}_k(z)\mathbf{v}_k}{1+\delta_k\mathbf{v}_k^T\mathbf{G}_k(z)\mathbf{v}_k}\mathrm{d}z+\frac{1}{2\pi i}\oint_{\mathcal{C}_k}\frac{\mathbf{u}^T[\mathbf{I}-(\mathbf{D}/\mathbf{\Lambda})^\alpha]\mathbf{G}_k(z)\mathbf{v}_k}{1+\delta_k\mathbf{v}_k^T\mathbf{G}_k(z)\mathbf{v}_k}\mathrm{d}z. \end{aligned} \tag{C.104}$$

The first term on the RHS above has already been treated in the previous proof. The second term on the RHS above can be analyzed by combining the Taylor expansion of $\mathbf{I}-(\mathbf{D}/\mathbf{\Lambda})^\alpha$ with the expansion (C.75) of $\mathbf{G}$ in the case of $\mathbf{u}=\mathbf{v}_k$ (resp. (C.92) in the case of $\mathbf{u}\perp\mathbf{v}_k$), together with arguments analogous to those used between (C.73)–(C.90) for $\mathbf{u}=\mathbf{v}_k$ (resp. (C.91)–(C.103) for $\mathbf{u}\perp\mathbf{v}_k$). We omit the technical details here since they parallel closely with the aforementioned arguments. This analysis results in (B.39) and (B.40), and thus completes the proof of Theorem B.5.

G. Proof of Theorem B.6

From the proof of Theorem B.1 in Section C-B, we see that (B.23) also holds for $k=K_0+1$. Hence, we can write

$$\begin{aligned}&\mathbb{P}[\widehat{K}_0\neq K_0]=\mathbb{P}[|\widehat{\delta}_{K_0}|<a_n']+\mathbb{P}[|\widehat{\delta}_{K_0+1}|\geq a_n'] \\ &=\mathbb{P}\left[1<\frac{a_n'}{|\delta_{K_0}|}+\frac{|\widehat{\delta}_{K_0}-t_{K_0}|}{|\delta_{K_0}|}+O(|\delta_{K_0}|^{-2})\right]+\mathbb{P}\left[\frac{a_n'}{|\delta_{K_0+1}|}\leq 1+\frac{|\widehat{\delta}_{K_0+1}-t_{K_0+1}|}{|\delta_{K_0+1}|}+O(|\delta_{K_0+1}|^{-2})\right] \\ &\to 0, \end{aligned} \tag{C.105}$$

where in the second step we have applied (B.23) for $k=K_0+1$, and in the last step we have used the assumptions $|\delta_{K_0}|\gtrsim a_n$, (B.41), and (B.43), together with estimate (B.27) from Theorem B.1. This establishes the consistency of $\widehat{K}_0$ and concludes the proof of Theorem B.6.

H. Proof of Corollary 3

We remark that the proof of Corollary 3 presented below is conducted in the *original, unrescaled setting*, as opposed to the rescaled framework introduced in Section B. The proof of part 1) of Corollary 3 again follows directly from the classical Lindeberg–Feller central limit theorem (CLT), so we omit the technical details for simplicity. It therefore remains to prove part 2) of Corollary 3. We need only to establish the CLT for

$$\frac{\alpha^2}{2}\mathbf{v}_k^T\left(\frac{\mathbf{D}-\mathbf{\Lambda}}{\mathbf{\Lambda}}\right)^2\mathbf{v}_k-\frac{1}{2t_k^2}\mathbf{v}_k^T\overline{\mathbf{W}}^2\mathbf{v}_k, \tag{C.106}$$

which can be rewritten as

$$\begin{aligned}\frac{\alpha^2}{2}\mathbf{v}_k^T\left(\frac{\mathbf{D}-\mathbf{\Lambda}}{\mathbf{\Lambda}}\right)^2\mathbf{v}_k-\frac{1}{2t_k^2}\mathbf{v}_k^T\overline{\mathbf{W}}^2\mathbf{v}_k &= \frac{\alpha^2}{2}\sum_{i\in[n]}\frac{v_k(i)^2}{\Lambda_i^2}\sum_{j_1,j_2\in[n]}W_{ij_1}W_{ij_2}-\frac{1}{2t_k^2}\sum_{i,j,l\in[n]}v_k(i)v_k(l)\overline{W}_{ij}\overline{W}_{jl} \\ &= \frac{1}{2}\sum_{1\leq i\leq j\leq n}(W_{ij}b_{ij}+W_{ij}^2c_{ij}) \end{aligned} \tag{C.107}$$

with

$$\begin{aligned}b_{ij} &:= \sum_{1\leq l<j}W_{il}f_k(i,j,l)+\sum_{1\leq l<i}W_{jl}f_k(j,i,l), \\ c_{ij} &:=\alpha^2(1+\delta_i^j)^{-1}\left(\frac{v_k(i)^2}{\Lambda_i^2}+\frac{v_k(j)^2}{\Lambda_j^2}\right)-\frac{1}{t_k^2}(1+\delta_i^j)^{-1}\Lambda_i^{-2\alpha}\Lambda_j^{-2\alpha}(v_k(i)^2+v_k(j)^2). \end{aligned}$$

Here, function $f_k(i,j,l)$ is defined as

$$f_k(i,j,l) :=(1+\delta_i^j)^{-1}\left(\alpha^2\Lambda_i^{-2}v_k(i)^2-\frac{2}{t_k^2}\Lambda_i^{-2\alpha}\Lambda_j^{-\alpha}\Lambda_l^{-\alpha}v_k(j)v_k(l)\right).$$

In view of (C.107), the mean of (C.106) can be written as

$$\mathbb{E}\left[\frac{\alpha^2}{2}\mathbf{v}_k^T\left(\frac{\mathbf{D}-\mathbf{\Lambda}}{\mathbf{\Lambda}}\right)^2\mathbf{v}_k-\frac{1}{2t_k^2}\mathbf{v}_k^T\overline{\mathbf{W}}^2\mathbf{v}_k\right]=\frac{1}{2}\sum_{1\leq i\leq j\leq n}s_{ij}c_{ij}. \tag{C.108}$$

Observe that for each integer $t \in [n(n+1)/2]$, there exist unique $i, j \in [n]$ such that $t = i + 2^{-1}j(j-1)$. With such property, let us define a filtration

$$\mathcal{F}_t := \sigma\{W_{ls} : 1 \le l \le s < j \text{ or } 1 \le s \le i \le l = j\}. \tag{C.109}$$

With respect to this filtration, the centered sum

$$\frac{\alpha^2}{2}\mathbf{v}_k^T\left(\frac{\mathbf{D}-\mathbf{\Lambda}}{\mathbf{\Lambda}}\right)^2\mathbf{v}_k - \frac{1}{2t_k^2}\mathbf{v}_k^T\overline{\mathbf{W}}^2\mathbf{v}_k - \mathbb{E}\left[\frac{\alpha^2}{2}\mathbf{v}_k^T\left(\frac{\mathbf{D}-\mathbf{\Lambda}}{\mathbf{\Lambda}}\right)^2\mathbf{v}_k - \frac{1}{2t_k^2}\mathbf{v}_k^T\overline{\mathbf{W}}^2\mathbf{v}_k\right]$$

is in fact a sum of *martingale differences*: for each $1 \le i \le j \le n$,

$$\frac{1}{2}\mathbb{E}[W_{ij}b_{ij} + (W_{ij}^2 - s_{ij})c_{ij}|\mathcal{F}_{i+2^{-1}j(j-1)-1}] = 0. \tag{C.110}$$

Next, we consider the sum of conditional variances

$$\begin{aligned} P_k = P_k(n) &:= \frac{1}{4}\sum_{1\le i\le j\le n}\mathbb{E}[(W_{ij}b_{ij} + (W_{ij}^2 - s_{ij})c_{ij})^2|\mathcal{F}_{i+2^{-1}j(j-1)-1}] \\ &= \frac{1}{4}\sum_{1\le i\le j\le n}\left(s_{ij}b_{ij}^2 + 2\gamma_{ij}b_{ij}c_{ij} + \kappa_{ij}c_{ij}^2\right), \end{aligned} \tag{C.111}$$

where $\gamma_{ij} := \mathbb{E}W_{ij}^3$ and $\kappa_{ij} := \mathbb{E}(W_{ij}^2 - s_{ij})^2$. It is easy to check that the mean of P_k is given by

$$\begin{aligned} \mathfrak{s}_{\mathbf{v}_k,k}^2 := \mathbb{E}P_k =& \frac{1}{4}\sum_{1\le i\le j\le n}s_{ij}\left(\sum_{1\le l<j}s_{il}(2-\delta_i^l)f_k(i,j,l)^2 + \sum_{1\le l<i}s_{jl}(2-\delta_j^l)f_k(j,i,l)^2\right) \\ &+ \frac{1}{4}\sum_{1\le i\le j\le n}\kappa_{ij}c_{ij}^2, \end{aligned} \tag{C.112}$$

and the variance of P_k is given by

$$\begin{aligned} \kappa_{\mathbf{v}_k} := \operatorname{var}(P_k) =& \frac{1}{16}\sum_{i_1,i_2,j_1,j_2\in[n], i_1\le j_1, i_2\le j_2}\mathbb{E}\left((s_{i_1j_1}(b_{i_1j_1}^2 - \mathbb{E}b_{i_1j_1}^2) + 2\gamma_{i_1j_1}b_{i_1j_1}c_{i_1j_1})\right. \\ &\left.\times\,(s_{i_2j_2}(b_{i_2j_2}^2 - \mathbb{E}b_{i_2j_2}^2) + 2\gamma_{i_2j_2}b_{i_2j_2}c_{i_2j_2})\right). \end{aligned} \tag{C.113}$$

We now recall the classical martingale CLT; see, for instance, Lemma 9.12 of Bai and Silverstein (2006). Let $\{Y_t\}$ be a martingale difference sequence with respect to a filtration $\{\mathcal{F}_t\}$. If the following conditions hold:

a) $\frac{\sum_{t\in[T]}\mathbb{E}[Y_t^2|\mathcal{F}_{t-1}]}{\sum_{t\in[T]}\mathbb{E}Y_t^2} \to 1$ in probability,

b) $\frac{\sum_{t\in[T]}\mathbb{E}[Y_t^2 I(|Y_t|/\sqrt{\sum_{t\in[T]}\mathbb{E}Y_t^2} \ge \epsilon)]}{\sum_{t\in[T]}\mathbb{E}Y_t^2} \le \frac{\sum_{t\in[T]}\mathbb{E}Y_t^4}{(\sum_{t\in[T]}\mathbb{E}Y_t^2)^2} \to 0$ for any fixed $\epsilon > 0$,

then we have $\frac{\sum_{t\in[T]}Y_t}{\sqrt{\sum_{t\in[T]}\mathbb{E}Y_t^2}} \to \mathcal{N}(0,1)$ in law as $T \to \infty$, where $I(\cdot)$ denotes the indicator function.

By the assumption $\kappa_{\mathbf{v}_k}^{1/2} \ll \mathfrak{s}_{\mathbf{v}_k,k}^2$, we have that $P_k/\mathbb{E}P_k \to 1$ in probability, which verifies condition (a) above. Thus, it remains to check condition (b) above in order to apply the martingale CLT. From direct calculations using (10) and (11), we can show that

$$\max_{i,j,l\in[n]}|f_k(i,j,l)| \lesssim \left(\frac{1}{q^2\beta_n^2} + \frac{1}{|\mathsf{d}_k|^2}\right)\|\mathbf{v}_k\|_\infty^2, \tag{C.114}$$

$$\max_{i,j\in[n]}|c_{ij}| \lesssim \left(\frac{1}{q^2\beta_n^2} + \frac{1}{|\mathsf{d}_k|^2}\right)\|\mathbf{v}_k\|_\infty^2, \tag{C.115}$$

$$\mathbb{E}b_{ij}^2 \lesssim \left(\frac{1}{q^2\beta_n^2} + \frac{1}{|\mathsf{d}_k|^2}\right)^2\|\mathbf{v}_k\|_\infty^4\frac{i+j}{n}, \tag{C.116}$$

$$\mathbb{E}b_{ij}^3 \lesssim \left(\frac{1}{q^2\beta_n^2} + \frac{1}{|\mathsf{d}_k|^2}\right)^3\|\mathbf{v}_k\|_\infty^6\frac{i+j}{nq}, \tag{C.117}$$

$$\mathbb{E}b_{ij}^4 \lesssim \left(\frac{1}{q^2\beta_n^2} + \frac{1}{|\mathsf{d}_k|^2}\right)^4\|\mathbf{v}_k\|_\infty^8\left(\frac{i+j}{nq^2} + \frac{i^2+j^2}{n^2}\right). \tag{C.118}$$

With the aid of (C.114)–(C.118), we can bound the sum of the fourth moments as

$$\begin{aligned}
&\sum_{1\le i\le j\le n} \mathbb{E}(W_{ij}b_{ij} + (W_{ij}^2 - s_{ij})c_{ij})^4 \\
&= \sum_{1\le i\le j\le n} \left(\mathbb{E}[W_{ij}^4]\mathbb{E}[b_{ij}^4] + 4c_{ij}\mathbb{E}[W_{ij}^3(W_{ij}^2 - s_{ij})]\mathbb{E}[b_{ij}^3] + 6c_{ij}^2\mathbb{E}[W_{ij}^2(W_{ij}^2 - s_{ij})^2]\mathbb{E}[b_{ij}^2] + c_{ij}^4\mathbb{E}[(W_{ij}^2 - s_{ij})^4]\right) \\
&\lesssim \left(\frac{1}{q^2\beta_n^2} + \frac{1}{|\mathsf{d}_k|^2}\right)^4 \|\mathbf{v}_k\|_\infty^8 \sum_{1\le i\le j\le n} \left(\frac{i^2+j^2}{n^3q^2} + \frac{i+j}{n^2q^4} + \frac{1}{nq^6}\right) \\
&\lesssim \frac{n}{q^2}\left(\frac{1}{q^2\beta_n^2} + \frac{1}{|\mathsf{d}_k|^2}\right)^4 \|\mathbf{v}_k\|_\infty^8 \ll \mathfrak{s}_{\mathbf{v}_k,k}^4,
\end{aligned}$$

where in the last step, we have used condition (63). This verifies condition (b) in the martingale CLT. Consequently, an application of the classical martingale CLT yields the desired conclusion in part 2) of Corollary 3. The proof of Corollary 3 is now complete.

Appendix D
Proofs of Local Laws

This section of the Supplementary Material is devoted to establishing the local laws stated in Theorem C.1. We begin by recalling the local laws for $\mathbf{R}(z)$ from the previous work Fan et al. (2022), which show that $\mathbf{\Upsilon}$ defined in (B.18) is the asymptotic limit of $\mathbf{G}(z)$ in various senses, along with precise rates of convergence.

Theorem D.1 (Local laws of $\mathbf{R}$). *Assume that Condition B.1 holds and $\mathfrak{C} > 2\sqrt{\mathfrak{M}} + \kappa$ for some constant $\kappa > 0$. Then for any deterministic unit vectors $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$ and any large constant $D > 0$, there exists a constant $C > 0$ such that the following events*

$$\bigcap_{z\in S(\mathfrak{C})} \left\{ |\mathbf{u}^T(\mathbf{R}(z) - \mathbf{\Upsilon}(z))\mathbf{v}| \le \frac{C\log n}{q|z|^2} \right\}, \tag{D.1}$$

$$\bigcap_{z\in S(\mathfrak{C})} \left\{ \max_{i\in[n]} |\mathbf{e}_i^T(\mathbf{R}(z) - \mathbf{\Upsilon}(z))\mathbf{v}| \le \frac{C}{|z|^2}\left(\sqrt{\frac{\log n}{n}} + \frac{\log n}{q}\|\mathbf{v}\|_\infty\right) \right\}, \tag{D.2}$$

$$\bigcap_{z\in S(\mathfrak{C})} \left\{ \max_{i\in[n]} |\mathbf{e}_i^T\mathbf{\Lambda}^{-\alpha}\mathbf{W}\mathbf{\Lambda}^{-\alpha}(\mathbf{R}(z) - \mathbf{\Upsilon}(z))\mathbf{v}| \le \frac{C}{|z|^2}\left(\sqrt{\frac{\log n}{n}} + \|\mathbf{v}\|_\infty\right) \right\} \tag{D.3}$$

hold with probability at least $1 - n^{-D}$, where $S(\mathfrak{C})$ is defined in (C.11).

Proof. These local laws (D.1), (D.2), and (D.3) correspond to Theorem 11, Proposition 1, and Proposition 2 in Fan et al. (2022). We therefore omit the technical details of proofs here. □

Combining Theorem D.1 with the preliminary estimates in Subsection C-A, we will establish local laws for the resolvent $\mathbf{G}(z)$ (see Theorems D.2–D.4), which form the core technical random matrix theory (RMT) tools used in the proofs of our main results in Section C. First, with the aid of Lemma C.1, we immediately obtain the proposition below.

Proposition D.1. *Under Condition B.1, for any $\mathfrak{C} > 2\sqrt{\mathfrak{M}} + \kappa$ with some constant $\kappa > 0$, the following estimates*

$$\|\mathbf{G}(z) - \mathbf{R}(z)\| \lesssim \frac{\sqrt{\log n}}{q|z|\beta_n}, \tag{D.4}$$

$$\|\mathbf{G}(z) - \mathbf{\Upsilon}(z)\| \lesssim \frac{1}{|z|^2} + \frac{\sqrt{\log n}}{q|z|\beta_n} \tag{D.5}$$

hold uniformly in $z \in S(\mathfrak{C})$ with high probability.

Proof. Recall the simple matrix identity $\mathbf{A}^{-1} = \mathbf{B}^{-1} + \mathbf{A}^{-1}(\mathbf{B} - \mathbf{A})\mathbf{B}^{-1}$ for any nonsingular matrices $\mathbf{A}$ and $\mathbf{B}$. Then an application of (C.5) along with Proposition C.1 gives that with high probability,

$$\|\mathbf{G}(z) - \mathbf{R}(z)\| = \|z\mathbf{G}((\mathbf{D}/\mathbf{\Lambda})^{2\alpha} - \mathbf{I})\mathbf{R}\| \lesssim \frac{\sqrt{\log n}}{q|z|\beta_n}.$$

Similarly, using (C.5), Proposition C.1, and (C.13), we can obtain that

$$\|\mathbf{G}(z)-\mathbf{\Upsilon}(z)\| \le \left\|z^{-1}\mathbf{G}(z)\left[\overline{\mathbf{W}}+(\mathbf{I}-(\mathbf{D}/\mathbf{\Lambda})^{2\alpha})z\right]\right\| + \left\|\mathbf{\Upsilon}(z)+z^{-1}\mathbf{I}\right\| \lesssim \frac{1}{|z|^2}+\frac{\sqrt{\log n}}{q|z|\beta_n},$$

which concludes the proof of Proposition D.1. □

Combining the local laws of $\mathbf{R}$ in Theorem D.1 with Proposition D.1, we can derive immediately some local laws for $\mathbf{G}$. However, these estimates are *not* sharp enough for our purposes. In the remainder of this section, we derive *refined local laws* for $\mathbf{G}$ that give *almost optimal* error bounds. A main difficulty in this analysis is that random matrices $\mathbf{D}$ and $\mathbf{W}$ are *not independent*. A useful observation is that the ith diagonal entry d_i of $\mathbf{D}$ depends primarily on the entries in the ith row and column of $\mathbf{W}$. To decouple this dependence, we introduce the intermediate resolvent $\mathbf{G}_{[i]}$ as in (86). Here, for defining the jth diagonal entry of $\mathbf{D}_{[i]}$, we remove contributions from the ith row and column of $\mathbf{W}$; recall (B.4). Consequently, $\mathbf{D}_{[i]}$ and $\mathbf{G}_{[i]}$ are independent of the entries in the ith row and column of $\mathbf{W}$. By invoking Lemma C.2, we can control the difference between $\mathbf{D}$ and $\mathbf{D}_{[i]}$ as follows.

Lemma D.1. *Under Condition B.1, for each fixed $\alpha \in (0,\infty)$, the following estimates*

$$\|\mathbf{D}-\mathbf{D}_{[i]}\|_F \lesssim \frac{1}{q\beta_n}, \quad \left\|\frac{\mathbf{D}^{\alpha}-\mathbf{D}^{\alpha}_{[i]}}{\mathbf{D}^{\alpha}}\right\|_F \lesssim \frac{1}{q\beta_n}, \quad \left\|\frac{\mathbf{D}^{\alpha}_{[i]}-\mathbf{\Lambda}^{\alpha}}{\mathbf{\Lambda}^{\alpha}}\right\| \lesssim \frac{\sqrt{\log n}}{q\beta_n} \tag{D.6}$$

hold with high probability. Consequently, for any $\mathfrak{C} > 2\sqrt{\mathfrak{M}}+\kappa$ with some constant $\kappa > 0$, the following estimates

$$\max_{i\in[n]}\left\|\mathbf{G}_{[i]}-\mathbf{R}(z)\right\| \lesssim \frac{\sqrt{\log n}}{q|z|\beta_n}, \tag{D.7}$$

$$\max_{i,j\in[n]}\left\|\mathbf{G}^{(j)}_{[i]}-\mathbf{\Upsilon}^{(j)}(z)\right\| \lesssim \frac{1}{|z|^2}+\frac{\sqrt{\log n}}{q|z|\beta_n}, \tag{D.8}$$

$$\max\left\{\max_{i\in[n]}\left\|\mathbf{G}_{[i]}(z)\right\|,\ \max_{i,j\in[n]}\left\|\mathbf{G}^{(j)}_{[i]}(z)\right\|,\ \max_{i,j,k\in[n]}\left\|\mathbf{G}^{(jk)}_{[i]}(z)\right\|\right\} \lesssim \frac{1}{|z|} \tag{D.9}$$

hold uniformly in $z \in S(\mathfrak{C})$ with high probability.

Proof. The third estimate in (D.6) can be proved in the same way as in Lemma C.1, which also entails that $1 \lesssim \|\mathbf{D}_{[i]}\| \lesssim \beta_n^{-1}$ with high probability. Together with (C.4) and a simple application of the mean value theorem, this leads to

$$\|(\mathbf{D}^{\alpha}-\mathbf{D}^{\alpha}_{[i]})\mathbf{D}^{-\alpha}\|_F \lesssim \|\mathbf{D}-\mathbf{D}_{[i]}\|_F$$

with high probability. Hence, the second estimate in (D.6) follows directly from the first estimate in (D.6), which we now establish. By definition, it holds that for any $j \ne i$,

$$d_j-(D_{[i]})_j = \frac{1}{q\beta_n}W_{ij}.$$

From (C.6), we can show that with high probability,

$$\sum_{j\in[n]\setminus\{i\}}|W_{ij}|^2-\sum_{j\in[n]\setminus\{i\}}s_{ij} \lesssim \frac{\log n}{q^2}.$$

It then follows that with high probability,

$$\left\|\mathbf{D}-\mathbf{D}_{[i]}\right\|_F^2 \lesssim \frac{1}{q^2\beta_n^2}\sum_{j\in[n]\setminus\{i\}}|W_{ij}|^2 \lesssim \frac{1}{q^2\beta_n^2},$$

which establishes the first estimate in (D.6). Combining (D.6) with Proposition C.1 immediately yields (D.9) and that with high probability,

$$\|\mathbf{G}(z)-\mathbf{G}_{[i]}(z)\| = \|z\mathbf{G}((\mathbf{D}/\mathbf{\Lambda})^{2\alpha}-(\mathbf{D}_{[i]}/\mathbf{\Lambda})^{2\alpha})\mathbf{R}\| \lesssim \frac{1}{q|z|\beta_n}.$$

Together with (D.4), this implies (D.7). Finally, estimate (D.8) can be established in the same way as for (D.5), which completes the proof of Lemma D.1. □

With Lemma D.1 above in hand, we can bound the difference between the bilinear forms $\mathbf{u}^T\mathbf{G}(z)\mathbf{v}$ and $\mathbf{u}^T\mathbf{G}_{[i]}(z)\mathbf{v}$. We emphasize that in Lemma D.2 below, vectors $\mathbf{u}$ and $\mathbf{v}$ are *not* necessarily deterministic (in contrast to some other results in this paper).

Lemma D.2. *Assume that Condition B.1 holds and $\mathfrak{C} > 2\sqrt{\mathfrak{M}} + \kappa$ for some constant $\kappa > 0$. Then for any (possibly random) vectors $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$ and any $z \in S(\mathfrak{C})$, we have that with high probability,*

$$\left|\mathbf{u}^T(\mathbf{G}(z) - \mathbf{G}_{[i]}(z))\mathbf{v}\right| \lesssim \frac{1}{q\beta_n}|\mathbf{u}| \left(\|\mathbf{G}\mathbf{v}\|_\infty \wedge \|\mathbf{G}_{[i]}\mathbf{v}\|_\infty\right). \tag{D.10}$$

An analogous estimate holds for $\mathbf{G}^{(i)}$, i.e.,

$$\left|\mathbf{u}^T \left(\mathbf{G}^{(i)}(z) - \mathbf{G}^{(i)}_{[i]}(z)\right) \mathbf{v}\right| \lesssim \frac{1}{q\beta_n}|\mathbf{u}| \left(\|\mathbf{G}^{(i)}\mathbf{v}\|_\infty \wedge \|\mathbf{G}^{(i)}_{[i]}\mathbf{v}\|_\infty\right). \tag{D.11}$$

Proof. It follows from the definition that

$$\begin{aligned}
\mathbf{u}^T \left(\mathbf{G}(z) - \mathbf{G}_{[i]}(z)\right) \mathbf{v} &= z\mathbf{u}^T\mathbf{G}(z)\mathbf{\Lambda}^{-2\alpha} \left(\mathbf{D}^{2\alpha} - \mathbf{D}^{2\alpha}_{[i]}\right) \mathbf{G}_{[i]}(z)\mathbf{v} \\
&\lesssim |\mathbf{u}| \left|\mathbf{\Lambda}^{-2\alpha} \left(\mathbf{D}^{2\alpha} - \mathbf{D}^{2\alpha}_{[i]}\right) \mathbf{G}_{[i]}(z)\mathbf{v}\right| \\
&\lesssim |\mathbf{u}| \left\|\mathbf{\Lambda}^{-2\alpha}(\mathbf{D}^{2\alpha} - \mathbf{D}^{2\alpha}_{[i]})\right\|_F \left\|\mathbf{G}_{[i]}(z)\mathbf{v}\right\|_\infty \lesssim \frac{1}{q\beta_n}|\mathbf{u}| \left\|\mathbf{G}_{[i]}(z)\mathbf{v}\right\|_\infty,
\end{aligned}$$

where we have used (C.10) in the second step and (D.6) in the last step. The term $\|\mathbf{G}_{[i]}(z)\mathbf{v}\|_\infty$ above can also be replaced with $\|\mathbf{G}(z)\mathbf{v}\|_\infty$ by writing the first step as

$$z\mathbf{u}^T\mathbf{G}_{[i]}(z)\mathbf{\Lambda}^{-2\alpha} \left(\mathbf{D}^{2\alpha} - \mathbf{D}^{2\alpha}_{[i]}\right) \mathbf{G}(z)\mathbf{v}$$

and using the bound (D.9). Thus, we obtain (D.10). The estimate in (D.11) can be proved in a similar way, which concludes the proof of Lemma D.2. ☐

By Schur's complement formula, we have the resolvent identities collected in the lemma below. The reader can also consult Lemma 3.4 of Erdős et al. (2013) for these identities.

Lemma D.3 (Resolvent identities)**.** *The following resolvent identities hold for $\mathbf{G}(z)$.*

(i) For each $i \in [n]$, we have

$$\frac{1}{G_{ii}} = -z(d_i/\Lambda_i)^{2\alpha} - \overline{W}_{ii} - \sum_{k,l\in[n]}^{(i)} \overline{W}_{ik}\overline{W}_{il}G^{(i)}_{kl}. \tag{D.12}$$

(ii) For each $i \neq j \in [n]$, we have

$$G_{ij} = -G_{ii}\sum_{k\in[n]}^{(i)} \overline{W}_{ik}G^{(i)}_{kj} = G_{ii}G^{(i)}_{jj}\Big(- \overline{W}_{ij} + \sum_{k,l\in[n]}^{(i,j)} \overline{W}_{ik}W_{jl}G^{(ij)}_{kl}\Big). \tag{D.13}$$

(iii) For each $k \in [n] \setminus \{i, j\}$, we have

$$G^{(k)}_{ij} = G_{ij} - \frac{G_{ik}G_{kj}}{G_{kk}}. \tag{D.14}$$

Same identities also hold for $\mathbf{G}_{[i]}$ and $\mathbf{R}$ by replacing $\mathbf{D}$ with $\mathbf{D}_{[i]}$ and $\mathbf{\Lambda}$, respectively.

The proof of the local laws for $\mathbf{G}(z)$ relies on the following *(almost) sharp* estimates on $\mathbf{e}_i^T\mathbf{G}(z)\mathbf{v}$ and $\mathbf{e}_i^T\mathbf{G}_{[j]}(z)\mathbf{v}$, for any $i, j \in [n]$ and deterministic unit vector $\mathbf{v}$.

Lemma D.4. *In the setting of Theorem D.1, for any deterministic unit vector $\mathbf{v} \in \mathbb{R}^n$ and any $z \in S(\mathfrak{C})$, the following estimates*

$$\max_{i\in[n]} \left|\mathbf{e}_i^T\mathbf{G}(z)\mathbf{v}\right| \lesssim \frac{\sqrt{\log n}}{\sqrt{n}|z|^2} + \frac{\|\mathbf{v}\|_\infty}{|z|}, \tag{D.15}$$

$$\max_{i,j\in[n]} \left|\mathbf{e}_i^T\mathbf{G}_{[j]}(z)\mathbf{v}\right| \lesssim \frac{\sqrt{\log n}}{\sqrt{n}|z|^2} + \frac{\|\mathbf{v}\|_\infty}{|z|}, \tag{D.16}$$

$$\max_{i\in[n]}\left|\frac{\mathbf{e}_i^T\mathbf{G}_{[i]}(z)\mathbf{v}}{(G_{[i]})_{ii}}\right| \lesssim \frac{\sqrt{\log n}}{\sqrt{n}|z|}+\|\mathbf{v}\|_\infty \tag{D.17}$$

hold with high probability.

We are now ready to state and prove the *three local laws* for $\mathbf{G}$ in Theorems D.2–D.4 below. These *refined* local law results yield the local laws stated in Theorem C.1.

Theorem D.2. *In the setting of Theorem D.1, for each constant $D>0$, there exists some constant $C>0$ such that for any deterministic unit vector $\mathbf{v}\in\mathbb{C}^n$, the following events*

$$\bigcap_{z\in S(\mathfrak{C})}\left\{\max_{i\in[n]}|\mathbf{e}_i^T(\mathbf{G}(z)-\mathbf{\Upsilon}(z))\mathbf{v}|\le C\frac{\sqrt{\log n}}{|z|}\left(\frac{1}{\sqrt{n}|z|}+\left(\frac{\sqrt{\log n}}{q|z|}+\frac{1}{q\beta_n}\right)\|\mathbf{v}\|_\infty\right)\right\}, \tag{D.18}$$

$$\bigcap_{z\in S(\mathfrak{C})}\left\{\max_{i,j\in[n]}|\mathbf{e}_i^T(\mathbf{G}_{[j]}(z)-\mathbf{\Upsilon}(z))\mathbf{v}|\le C\frac{\sqrt{\log n}}{|z|}\left(\frac{1}{\sqrt{n}|z|}+\left(\frac{\sqrt{\log n}}{q|z|}+\frac{1}{q\beta_n}\right)\|\mathbf{v}\|_\infty\right)\right\}, \tag{D.19}$$

$$\bigcap_{z\in S(\mathfrak{C})}\left\{\max_{i\neq j\in[n]}\left|\mathbf{e}_j^T(\mathbf{G}_{[i]}^{(i)}(z)-\mathbf{\Upsilon}^{(i)}(z))\mathbf{v}\right|\le C\frac{\sqrt{\log n}}{|z|}\left(\frac{1}{\sqrt{n}|z|}+\left(\frac{\sqrt{\log n}}{q|z|}+\frac{1}{q\beta_n}\right)\|\mathbf{v}\|_\infty\right)\right\} \tag{D.20}$$

hold with probability at least $1-n^{-D}$.

Theorem D.3. *In the setting of Theorem D.1, for each constant $D>0$, there exists some constant $C>0$ such that for any deterministic unit vectors $\mathbf{u},\mathbf{v}\in\mathbb{R}^n$, the event*

$$\bigcap_{z\in S(\mathfrak{C})}\left\{|\mathbf{u}^T(\mathbf{G}(z)-\mathbf{\Upsilon}(z))\mathbf{v}|\le C\frac{\sqrt{\log n}}{q|z|}\left(\frac{1}{|z|\beta_n}+\frac{\sqrt{\log n}}{q\beta_n^2}+\|\mathbf{u}\|_\infty\wedge\|\mathbf{v}\|_\infty\right)\right\} \tag{D.21}$$

holds with probability at least $1-n^{-D}$.

Theorem D.4. *In the setting of Theorem D.1, for each constant $D>0$, there exists some constant $C>0$ such that for any deterministic unit vector $\mathbf{v}\in\mathbb{R}^n$, the event*

$$\bigcap_{z\in S(\mathfrak{C})}\left\{\max_{i\in[n]}\left|\mathbf{e}_i^T\overline{\mathbf{W}}(\mathbf{G}-\mathbf{\Upsilon})\mathbf{v}\right|\le C_{10}\left(\left(\frac{1}{|z|}+\frac{\sqrt{\log n}}{q\beta_n}\right)\frac{\sqrt{\log n}}{\sqrt{n}|z|}+\left(\frac{1}{|z|}+\frac{1}{q\beta_n}\right)\frac{\|\mathbf{v}\|_\infty}{|z|}\right)\right\} \tag{D.22}$$

holds with probability at least $1-n^{-D}$.

The remainder of this section is devoted to the proofs of Lemma D.4 and Theorems D.2–D.4 above.

A. *Proof of Lemma D.4*

Denote by $\mathbf{v}^{(i)}$ the vector with components $\mathbf{v}^{(i)}(j)=\mathbf{1}_{j\neq i}v(j)$, i.e., $\mathbf{v}^{(i)}$ is obtained by setting the ith component of $\mathbf{v}$ as zero. Applying (D.13) for $\mathbf{G}_{[i]}$ and recalling the notation in (21), it holds that w.h.p.,

$$\mathbf{e}_i^T\mathbf{G}_{[i]}\mathbf{v}=-(G_{[i]})_{ii}\sum_l^{(i)}\overline{W}_{il}(G_{[i]}^{(i)})_{l\mathbf{v}^{(i)}}+(G_{[i]})_{ii}v(i). \tag{D.23}$$

Note that by the definition in (B.25), $G_{[i]}^{(i)}$ is independent of the entries $\overline{W}_{il}$. Hence, we can apply (C.6) to $\sum_l^{(i)}\overline{W}_{il}(G_{[i]}^{(i)})_{l\mathbf{v}^{(i)}}$ and obtain that w.h.p.,

$$\begin{aligned}\left|\frac{\mathbf{e}_i^T\mathbf{G}_{[i]}(z)\mathbf{v}}{(G_{[i]})_{ii}}\right| &\lesssim \log n\frac{\max_{1\le l\neq i\le n}|(G_{[i]}^{(i)})_{l\mathbf{v}^{(i)}}|}{q}+\sqrt{\log n}\left(\frac{1}{n}\sum_l^{(i)}|(G_{[i]}^{(i)})_{l\mathbf{v}^{(i)}}|^2\right)^{1/2}+\|\mathbf{v}\|_\infty\\ &\lesssim \frac{\log n}{q}\max_{1\le l\neq i\le n}|(G_{[i]}^{(i)})_{l\mathbf{v}^{(i)}}|+\frac{\sqrt{\log n}}{\sqrt{n}|z|}+\|\mathbf{v}\|_\infty,\end{aligned} \tag{D.24}$$

where in the second step, we have used (D.9) to bound $\left(\sum_l^{(i)}|(G_{[i]}^{(i)})_{l\mathbf{v}^{(i)}}|^2\right)^{1/2}$ by $O(|z|^{-1})$. Plugging (D.24) into (D.23) and using again (D.9) to bound $(G_{[i]})_{ii}$, we can show that

$$\left|\mathbf{e}_i^T\mathbf{G}_{[i]}\mathbf{v}\right|\lesssim\frac{\log n}{q|z|}\max_{1\le l\neq i\le n}|(G_{[i]}^{(i)})_{l\mathbf{v}^{(i)}}|+\frac{\sqrt{\log n}}{\sqrt{n}|z|^2}+\frac{\|\mathbf{v}\|_\infty}{|z|}. \tag{D.25}$$

Moreover, applying (D.14) and (D.13) to $\mathbf{G}_{[i]}$ gives that w.h.p.,

$$\begin{aligned}(G_{[i]}^{(i)})_{l\mathbf{v}^{(i)}} &= (G_{[i]})_{l\mathbf{v}^{(i)}} - \frac{(G_{[i]})_{li}(G_{[i]})_{i\mathbf{v}^{(i)}}}{(G_{[i]})_{ii}} \\ &= (G_{[i]})_{l\mathbf{v}^{(i)}} + (G_{[i]})_{i\mathbf{v}^{(i)}} \cdot \sum_k^{(i)} (G_{[i]}^{(i)})_{lk} \overline{W}_{ki} \\ &\lesssim \max_{l\in[n]} |(G_{[i]})_{l\mathbf{v}^{(i)}}| \lesssim \max_{l\in[n]} |(G_{[i]})_{l\mathbf{v}}| + |v(i)|/|z|, \end{aligned} \tag{D.26}$$

where in the third step, we have again applied (C.6) and (D.9) to get that w.h.p.,

$$\sum_k^{(i)} (G_{[i]}^{(i)})_{lk} \overline{W}_{ki} \lesssim \frac{\log n}{q|z|} + \sqrt{\log n}\left(\frac{1}{n}\sum_k^{(i)} |(G_{[i]}^{(i)})_{lk}|^2\right)^{1/2} \lesssim \frac{\log n}{q|z|} + \frac{\sqrt{\log n}}{\sqrt{n}|z|}.$$

Hence, combining (D.25) and (D.26), we can obtain that w.h.p.,

$$\left|\mathbf{e}_i^T \mathbf{G}_{[i]} \mathbf{v}\right| \lesssim \frac{\log n}{q|z|} \max_{l\in[n]} |(G_{[i]})_{l\mathbf{v}}| + \frac{\sqrt{\log n}}{\sqrt{n}|z|^2} + \frac{\|\mathbf{v}\|_\infty}{|z|}. \tag{D.27}$$

On the other hand, an application of Lemma D.2 shows that w.h.p.,

$$|(G_{[i]})_{l\mathbf{v}} - G_{l\mathbf{v}}| \lesssim \frac{1}{q\beta_n} \max_{l\in[n]} |G_{l\mathbf{v}}|.$$

Plugging this bound into (D.27), it follows that w.h.p.,

$$|\mathbf{e}_i^T \mathbf{G}\mathbf{v}| \lesssim \frac{\log n}{q|z|} \max_{l\in[n]} |G_{l\mathbf{v}}| + \frac{\log n}{q^2|z|\beta_n} \max_{l\in[n]} |G_{l\mathbf{v}}| + \frac{\sqrt{\log n}}{\sqrt{n}|z|^2} + \frac{\|\mathbf{v}\|_\infty}{|z|}.$$

By the union bound, this estimate holds uniformly over all $i \in [n]$ w.h.p. Then taking the maximum over $i \in [n]$ on the left-hand side yields that w.h.p.,

$$\max_{i\in[n]} |G_{i\mathbf{v}}| \lesssim \left(\frac{\log n}{q|z|} + \frac{\log n}{q^2|z|\beta_n}\right) \max_{i\in[n]} |G_{i\mathbf{v}}| + \frac{\sqrt{\log n}}{\sqrt{n}|z|^2} + \frac{\|\mathbf{v}\|_\infty}{|z|}.$$

Solving for $\max_{i\in[n]} |G_{i\mathbf{v}}|$ and using the assumption $\log n \ll q\beta_n$ from (B.16), we can arrive at (D.15). Combining (D.15) with Lemma D.2 then leads to (D.16). Finally, applying (D.26) and (D.16) to (D.24), we can derive (D.17), which concludes the proof of Lemma D.4.

B. *Proof of Theorem D.2*

For any $z \in S(\mathfrak{C})$, applying (D.13) to $\mathbf{G}_{[i]}(z)$ gives that

$$\begin{aligned}\mathbf{e}_i^T(\mathbf{G}_{[i]} - \boldsymbol{\Upsilon})\mathbf{v} &= -(G_{[i]})_{ii} \sum_l^{(i)} \overline{W}_{il} (G_{[i]}^{(i)})_{l\mathbf{v}^{(i)}} + v(i)\left((G_{[i]})_{ii} - M_i\right) \\ &\lesssim \frac{1}{|z|}\left(\frac{\log n}{q} \max_{1\le l\neq i\le n} |(G_{[i]}^{(i)})_{l\mathbf{v}^{(i)}}| + \frac{\sqrt{\log n}}{\sqrt{n}|z|}\right) + \left(\frac{\log n}{q|z|^2} + \frac{\sqrt{\log n}}{q|z|\beta_n}\right)\|\mathbf{v}\|_\infty \end{aligned} \tag{D.28}$$

with high probability. Here, in the second step above, we have used a similar argument as in (D.24) based on (C.6) and (D.9), and applied (D.1) and (D.7) to control $(G_{[i]})_{ii} - M_i$. In light of (D.26) and (D.16), it holds further that w.h.p.,

$$|(G_{[i]}^{(i)})_{l\mathbf{v}^{(i)}}| \lesssim \frac{\sqrt{\log n}}{\sqrt{n}|z|^2} + \frac{\|\mathbf{v}\|_\infty}{|z|}. \tag{D.29}$$

Plugging (D.29) into (D.28), we can deduce that for any $z \in S(\mathfrak{C})$,

$$\left|\mathbf{e}_i^T(\mathbf{G}_{[i]} - \boldsymbol{\Upsilon})\mathbf{v}\right| \lesssim \frac{\sqrt{\log n}}{\sqrt{n}|z|^2} + \left(\frac{\log n}{q|z|^2} + \frac{\sqrt{\log n}}{q|z|\beta_n}\right)\|\mathbf{v}\|_\infty \tag{D.30}$$

with high probability. We next upgrade this estimate to hold uniformly in $z \in S(\mathfrak{C})$ by a standard ϵ-net argument. First, using the union bound, we can derive a uniform estimate for all z in an $(n|z|)^{-3}$-net $S(\mathfrak{C}) \cap \{(n|z|)^{-3}\mathbb{Z}^2\}$. Second, by the Lipschitz continuity of $\max_{i\in[n]} |\mathbf{e}_i^T(\mathbf{G}_{[i]}(z) - \boldsymbol{\Upsilon}(z))\mathbf{v}|$ in z (with Lipschitz constant $O(1)$ due to (C.8)), the bound extends uniformly to all $S(\mathfrak{C})$. Consequently, there exists a constant $C > 0$ such that the event

$$\bigcap_{z \in S(\mathfrak{C})} \left\{ \max_{i \in [n]} |\mathbf{e}_i^T (\mathbf{G}_{[i]}(z) - \boldsymbol{\Upsilon}(z))\mathbf{v}| \le C \frac{\sqrt{\log n}}{|z|} \left(\frac{1}{\sqrt{n}|z|} + \left(\frac{\sqrt{\log n}}{q|z|} + \frac{1}{q\beta_n} \right) \|\mathbf{v}\|_\infty \right) \right\} \tag{D.31}$$

holds with probability $\ge 1 - n^{-D}$.

In view of (D.31), the estimate (D.18) for each fixed $z \in S(\mathfrak{C})$ follows immediately from Lemmas D.2 and D.4. With (D.18) at hand, the estimate (D.19) for fixed $z \in S(\mathfrak{C})$ also follows from Lemmas D.2 and D.4. Repeating the ϵ-net argument above then yields the uniform versions (D.18) and (D.19). Finally, for (D.20), we observe that

$$\mathbf{e}_j^T \big(\mathbf{G}_{[i]}^{(i)}(z) - \boldsymbol{\Upsilon}^{(i)}(z)\big)\mathbf{v} = \mathbf{1}_{i \neq j} \mathbf{e}_j^T \big(\mathbf{G}_{[i]}^{(i)}(z) - \boldsymbol{\Upsilon}(z)\big)\mathbf{v}^{(i)}.$$

Applying (D.14) to $\mathbf{G}_{[i]}$, we can obtain that for $j \neq i$,

$$\begin{aligned} \left| \mathbf{e}_j^T (\mathbf{G}_{[i]}(z) - \mathbf{G}_{[i]}^{(i)}(z))\mathbf{v}^{(i)} \right| &= \left| (G_{[i]})_{ji} \right| \left| \frac{(G_{[i]})_{i\mathbf{v}^{(i)}}}{(G_{[i]})_{ii}} \right| \\ &\lesssim \frac{\sqrt{\log n}}{|z|} \left(\frac{1}{\sqrt{n}|z|} + \frac{\sqrt{\log n}}{q|z|} + \frac{1}{q\beta_n} \right) \left(\frac{\sqrt{\log n}}{\sqrt{n}|z|} + \|\mathbf{v}\|_\infty \right) \end{aligned} \tag{D.32}$$

w.h.p., where in the second step, we have used (D.17) to control $\left|(G_{[i]})_{i\mathbf{v}^{(i)}}/(G_{[i]})_{ii}\right|$ and (D.30) with $\mathbf{v} = \mathbf{e}_j$ to control $\left|(G_{[i]})_{ji}\right|$. Combining (D.19) with (D.32) leads to (D.20), which completes the proof of Theorem D.2.

C. *Proof of Theorem D.3*

Denote by $\boldsymbol{\mathcal{E}}' := (\mathbf{D}/\boldsymbol{\Lambda})^{2\alpha} - \mathbf{I}$. By the estimate (C.5), it holds that

$$\|\boldsymbol{\mathcal{E}}'\| \lesssim \sqrt{\log n}/(q\beta_n) \tag{D.33}$$

with high probability. Combining this bound with Theorem D.1 and Proposition D.1, we can show that w.h.p.,

$$\begin{aligned} \mathbf{u}^T (\mathbf{G}(z) - \boldsymbol{\Upsilon}(z))\mathbf{v} = \mathbf{u}^T (\mathbf{G} - \mathbf{R})\mathbf{v} + O\left(\frac{\log n}{q|z|^2} \right) &= z \mathbf{u}^T \mathbf{G} \boldsymbol{\mathcal{E}}' \mathbf{R} \mathbf{v} + O\left(\frac{\log n}{q|z|^2} \right) \\ &= z \mathbf{u}^T \boldsymbol{\Upsilon} \boldsymbol{\mathcal{E}}' \boldsymbol{\Upsilon} \mathbf{v} + O\left(\frac{\sqrt{\log n}}{q|z|^2 \beta_n} + \frac{\log n}{q^2 |z| \beta_n^2} \right). \end{aligned} \tag{D.34}$$

It remains to estimate the first term on the RHS of (D.34) above. With the Taylor expansion of $\boldsymbol{\mathcal{E}}'$, we have that

$$z \mathbf{u}^T \boldsymbol{\Upsilon} \boldsymbol{\mathcal{E}}' \boldsymbol{\Upsilon} \mathbf{v} = z \sum_{i \in [n]} u(i) v(i) \Upsilon_i^2 \mathcal{E}_i' = \frac{z}{q} \sum_{i \in [n]} u(i) v(i) \Upsilon_i^2 \cdot \frac{2\alpha}{\Lambda_i} \sum_{j \in [n]} W_{ij} + O\left(\frac{\log n}{q^2 |z| \beta_n^2} \right) \tag{D.35}$$

w.h.p., where we have used (D.33) and (C.13) in the second step. Next, applying (C.6) to the leading term on the right-hand side above yields that w.h.p.,

$$\begin{aligned} \sum_{i,j \in [n]} u(i) v(i) \Upsilon_i^2 \frac{2\alpha}{\Lambda_i} W_{ij} &\lesssim \frac{1}{|z|^2} \left[\frac{\log n}{q} \|\mathbf{u}\|_\infty \|\mathbf{v}\|_\infty + \sqrt{\log n} \left(\frac{1}{n} \sum_{i,j \in [n]} |u(i)|^2 |v(i)|^2 \right)^{1/2} \right] \\ &\lesssim \frac{\sqrt{\log n}}{|z|^2} \|\mathbf{u}\|_\infty \wedge \|\mathbf{v}\|_\infty, \end{aligned}$$

where $\wedge$ denotes the minimum of two given numbers. Plugging this estimate into (D.35), we can obtain that w.h.p.,

$$\left| z \mathbf{u}^T \boldsymbol{\Upsilon} \boldsymbol{\mathcal{E}}' \boldsymbol{\Upsilon} \mathbf{v} \right| \lesssim \frac{\sqrt{\log n}}{q|z|} \|\mathbf{u}\|_\infty \wedge \|\mathbf{v}\|_\infty + \frac{\log n}{q^2 |z| \beta_n^2}. \tag{D.36}$$

Together with (D.34), this yields (D.21) for each fixed $z \in S(\mathfrak{C})$. Finally, an application of the standard ϵ-net argument extends the estimate uniformly to all $z \in S(\mathfrak{C})$, and completes the proof of Theorem D.3.

D. *Proof of Theorem D.4*

The proof of Theorem D.4 relies on the following concentration estimates for the bilinear forms of centered independent random variables.

Lemma D.5 (Lemma 3.8 of Erdős et al. (2013)). *Let $(x_i)_{i\in[n]}$ and $(y_i)_{i\in[n]}$ be independent families of centered independent random variables, and $(B_{ij})_{i,j\in[n]}$ a family of deterministic numbers. Assume that all components x_i and y_i have variances at most n^{-1}, and satisfy that $\max_{i\in[n]}|x_i|\le\phi_n$ and $\max_{i\in[n]}|y_i|\le\phi_n$ for some (n-dependent) parameter $\phi_n\ge n^{-1/2}$. Then for any deterministic parameter $\xi\ge\log n$, there exists an absolute constant $a>0$ such that the following estimates*

$$\left|\sum_{i\in[n]}\bar{x}_iB_{ii}x_i-\sum_{i\in[n]}(\mathbb{E}|x_i|^2)B_{ii}\right|\le\left(\xi^{1/2}\phi_n+\xi\phi_n^2\right)B_d, \tag{D.37}$$

$$\left|\sum_{i\ne j\in[n]}\bar{x}_iB_{ij}x_j\right|\le\xi^2\left[\phi_nB_o+\frac{1}{n}\Big(\sum_{i\ne j\in[n]}|B_{ij}|^2\Big)^{1/2}\right] \tag{D.38}$$

hold with probability at least $1-\exp(-a\xi)$, where $B_d:=\max_{i\in[n]}|B_{ii}|$ and $B_o:=\max_{i\ne j\in[n]}|B_{ij}|$.

In view of (D.10), it follows that w.h.p.,

$$\mathbf{e}_i^T\overline{\mathbf{W}}(\mathbf{G}-\boldsymbol{\Upsilon})\mathbf{v}=\mathbf{e}_i^T\overline{\mathbf{W}}(\mathbf{G}_{[i]}-\boldsymbol{\Upsilon})\mathbf{v}+O\left[\frac{1}{q\beta_n}\left(\frac{\sqrt{\log n}}{\sqrt{n}|z|^2}+\frac{\|\mathbf{v}\|_\infty}{|z|}\right)\right], \tag{D.39}$$

where we have used (C.8) to bound $\|\overline{\mathbf{W}}\mathbf{e}_i\|$ and (D.15) to bound $\|\mathbf{G}\mathbf{v}\|_{\max}$. By the identity (D.14), it holds that

$$\begin{aligned}\mathbf{e}_i^T\overline{\mathbf{W}}(\mathbf{G}_{[i]}-\boldsymbol{\Upsilon})\mathbf{v}=&\sum_{j\in[n]\setminus\{i\}}\overline{W}_{ij}\left(\mathbf{G}_{[i]}^{(i)}-\boldsymbol{\Upsilon}^{(i)}\right)_{j\mathbf{v}}+\sum_{j\in[n]\setminus\{i\}}\overline{W}_{ij}\frac{(G_{[i]})_{ji}(G_{[i]})_{i\mathbf{v}}}{(G_{[i]})_{ii}}\\&+\overline{W}_{ii}\left((G_{[i]})_{i\mathbf{v}}-\Upsilon_{i\mathbf{v}}\right).\end{aligned} \tag{D.40}$$

We first control the last term above. Combining (B.12) with (D.19), we have that w.h.p.,

$$\overline{W}_{ii}\left((G_{[i]})_{i\mathbf{v}}-\Upsilon_{i\mathbf{v}}\right)\lesssim\frac{\sqrt{\log n}}{q|z|}\left(\frac{1}{\sqrt{n}|z|}+\left(\frac{\sqrt{\log n}}{q|z|}+\frac{1}{q\beta_n}\right)\|\mathbf{v}\|_\infty\right). \tag{D.41}$$

Next, we bound the first term on the right-hand side of (D.40). By an application of (C.6), together with (D.20) to control the L_∞-norm and (D.8) to control the L_2-norm of $(\mathbf{G}_{[i]}^{(i)}-\boldsymbol{\Upsilon}^{(i)})\mathbf{v}$, we can deduce that w.h.p.,

$$\begin{aligned}\sum_{j\in[n]\setminus\{i\}}\overline{W}_{ij}\left(\mathbf{G}_{[i]}^{(i)}-\boldsymbol{\Upsilon}^{(i)}\right)_{j\mathbf{v}}\lesssim&\frac{(\log n)^{3/2}}{q|z|}\left(\frac{1}{\sqrt{n}|z|}+\left(\frac{\sqrt{\log n}}{q|z|}+\frac{1}{q\beta_n}\right)\|\mathbf{v}\|_\infty\right)\\&+\frac{\sqrt{\log n}}{\sqrt{n}}\left(\frac{1}{|z|^2}+\frac{\sqrt{\log n}}{q|z|\beta_n}\right).\end{aligned} \tag{D.42}$$

Finally, we bound the second term on the right-hand side of (D.40). In light of (D.13), we can write it as

$$\sum_{j\in[n]\setminus\{i\}}\overline{W}_{ij}\frac{(G_{[i]})_{ji}(G_{[i]})_{i\mathbf{v}}}{(G_{[i]})_{ii}}=-(G_{[i]})_{i\mathbf{v}}\sum_{j,k\in[n]\setminus\{i\}}\overline{W}_{ij}\overline{W}_{ki}(G_{[i]}^{(i)})_{jk}. \tag{D.43}$$

Then applying (D.37) and (D.38) with $\xi=C\log n$ for a large constant $C>0$, it holds that w.h.p.,

$$\begin{aligned}&\sum_{j,k\in[n]\setminus\{i\}}\overline{W}_{ij}\overline{W}_{ki}(G_{[i]}^{(i)})_{jk}-\sum_{j\in[n]\setminus\{i\}}\Lambda_i^{-4\alpha}\Lambda_j^{-4\alpha}s_{ij}(G_{[i]}^{(i)})_{jj}\\&\lesssim\sqrt{\log n}\frac{\max_{j\in[n]}|(G_{[i]}^{(i)})_{jj}|}{q}+(\log n)^2\frac{\max_{1\le j\ne k\le n}|(G_{[i]}^{(i)})_{jk}|}{q}+\frac{(\log n)^2}{n}\Big(\sum_{j,k\in[n]\setminus\{i\}}|(G_{[i]}^{(i)})_{jk}|^2\Big)^{1/2}\\&\lesssim\frac{\sqrt{\log n}}{q|z|}+\frac{(\log n)^{5/2}}{q|z|}\left(\frac{1}{\sqrt{n}|z|}+\frac{\sqrt{\log n}}{q|z|}+\frac{1}{q\beta_n}\right)+\frac{(\log n)^2}{\sqrt{n}|z|}\lesssim\frac{\sqrt{\log n}}{q|z|}+\frac{(\log n)^{5/2}}{q^2|z|\beta_n}+\frac{(\log n)^2}{\sqrt{n}|z|}.\end{aligned} \tag{D.44}$$

Here, in the second step above, we have used (D.9) to bound $\max_{j\in[n]}|(G_{[i]}^{(i)})_{jj}|$, (D.20) to bound $\max_{1\le j\ne k\le n}|(G_{[i]}^{(i)})_{jk}|$, and again (D.9) to estimate

$$\sum_{j,k\in[n]\setminus\{i\}}|(G_{[i]}^{(i)})_{jk}|^2=\mathrm{tr}[\mathbf{G}_{[i]}^{(i)}(\mathbf{G}_{[i]}^{(i)})^*]=O(n|z|^{-2})\quad\text{w.h.p.}$$

In the last step above, we have also utilized the assumption $q \gg (\log n)^4$. Moreover, it follows from (D.9) that

$$\sum_{j\in[n]\setminus\{i\}} \Lambda_i^{-4\alpha}\Lambda_j^{-4\alpha} s_{ij}(G_{[i]}^{(i)})_{jj} \lesssim |z|^{-1} \quad \text{w.h.p.}$$

Combining this estimate with (D.44) yields that w.h.p.,

$$\sum_{j,k\in[n]\setminus\{i\}} \overline{W}_{ij}\overline{W}_{ki}(G_{[i]}^{(i)})_{jk} \lesssim \left(1+\frac{(\log n)^{5/2}}{q^2\beta_n}\right)\frac{1}{|z|}.$$

Further, plugging it into (D.43) and using (D.16), we can obtain that

$$\sum_{j\in[n]\setminus\{i\}} \overline{W}_{ij}\frac{(G_{[i]})_{ji}(G_{[i]})_{i\mathbf{v}}}{(G_{[i]})_{ii}} \lesssim \left(1+\frac{(\log n)^{5/2}}{q^2\beta_n}\right)\left(\frac{\sqrt{\log n}}{\sqrt{n}|z|^3}+\frac{\|\mathbf{v}\|_\infty}{|z|^2}\right). \tag{D.45}$$

Now, combining (D.41), (D.42), and (D.45), and using the assumptions $q \gg (\log n)^4$ and $q\beta_n \gg \log n$ from (B.16), we can conclude that w.h.p.,

$$\mathbf{e}_i^T\overline{\mathbf{W}}(\mathbf{G}_{[i]}-\boldsymbol{\Upsilon})\mathbf{v} \lesssim \left(\frac{1}{|z|}+\frac{\sqrt{\log n}}{q\beta_n}\right)\frac{\sqrt{\log n}}{\sqrt{n}|z|}+\left(\frac{(\log n)^{3/2}}{q^2\beta_n}+\frac{1}{|z|}\right)\frac{\|\mathbf{v}\|_\infty}{|z|}.$$

Plugging this into (D.39) leads to the estimate (D.22) for each fixed $z \in S(\mathfrak{C})$. Again, a standard ϵ-net argument then extends the estimate uniformly over $z \in S(\mathfrak{C})$, thereby completing the proof of Theorem D.4.

Appendix E
End-to-end CLTs under DCSBM

In this section of the Supplementary Material, we present some end-to-end feasible CLTs for the degree-corrected stochastic block model (DCSBM). For clarity of exposition, we focus on the case of mild degree heterogeneity, under which parameter β_n in (13) is bounded away from zero. Our theory can be readily extended to the regime of $\beta_n = o(1)$, albeit at the cost of additional assumptions and technical arguments. Despite the involved technical details, the underlying logic is straightforward: once a "population" CLT is established, the consistency of the plug-in bias and variance estimators, together with Slutsky's lemma, will yield the desired feasible CLTs.

We first define the model as a special case of Example 1. Let us consider the DCSBM with $K \geq 2$ communities, where K is a fixed integer. Denote by $g_i \in [K]$ the community label of node i, $\vartheta_i > 0$ the corresponding degree parameter (which may depend on n), and $\rho_n \in (0,1)$ a sparsity factor. Let $B \in [0,\infty)^{K\times K}$ be a fixed symmetric matrix. Conditional on $(g_i,\vartheta_i)_{i\in[n]}$, we generate an undirected adjacency matrix $\widetilde{\mathbf{X}} = (\widetilde{X}_{ij})_{i,j\in[n]}$ according to

$$\mathbb{P}\left[\widetilde{X}_{ij}=1 \,\middle|\, (g_i,\vartheta_i)_{i\in[n]}\right] = p_{ij} := \rho_n\cdot\vartheta_i\vartheta_j\cdot B_{g_ig_j}, \quad 1\leq i\leq j\leq n. \tag{E.1}$$

For simplicity of presentation, throughout the following we treat $(g_i,\vartheta_i)_{i\in[n]}$ as fixed parameters and omit the conditioning on them. We assume that $p_{ij} \leq 1-\varepsilon$ uniformly for some fixed constant $\varepsilon \in (0,1)$. Then the mean matrix is $\mathbf{H} := \mathbb{E}[\widetilde{\mathbf{X}}] = (p_{ij})$, the noise matrix is $\mathbf{W} := \widetilde{\mathbf{X}} - \mathbf{H}$, and the variance profile is given by

$$s_{ij} := \mathbb{E}[W_{ij}^2] = \mathrm{Var}(\widetilde{X}_{ij}) = p_{ij}(1-p_{ij}). \tag{E.2}$$

We impose a mild degree-heterogeneity condition to ensure that β_n is bounded away from zero; specifically, we assume that all degree parameters ϑ_i are bounded from below by a positive constant, as summarized below.

Assumption E.1. *Assume that the following conditions hold.*

(i) (Sparsity) The sparsity parameter satisfies $n\rho_n \gg (\log n)^8$.

(ii) (Community balance) There exist some constants $0 < c_\pi \leq C_\pi < \infty$ such that for each $a \in [K]$,

$$c_\pi n \;\leq\; n_a := \#\{i : g_i = a\} \;\leq\; C_\pi n. \tag{E.3}$$

(iii) (Mild degree heterogeneity) There exist some constants $0 < c_\vartheta \leq C_\vartheta < \infty$ such that for all $i \in [n]$,

$$c_\vartheta \leq \vartheta_i \leq C_\vartheta.$$

Under this condition, we have $\Lambda_i = \sum_j p_{ij} \sim n\rho_n$ and $\theta \sim \rho_n$ (recall the definition in (6)), so that $q^2 = n\theta \sim n\rho_n$. Consequently, by definitions (12) and (13), we have $\theta_i = \Lambda_i/(n\theta) \sim 1$ and $\beta_n = \min_i \theta_i \sim 1$.

(iv) (Spiked eigenvalues) The nonzero eigenvalues $(\delta_k)_{k\in[K]}$ *of* $\mathbf{\Lambda}^{-\alpha}\mathbf{H}\mathbf{\Lambda}^{-\alpha}$ *are all spiked in the sense that* $|\delta_k| \gg q^{1-4\alpha}\sqrt{\log n}$ *(equivalently,* $|\mathsf{d}_k| \gg \sqrt{\log n}$ *under notation (29)). Moreover, they satisfy the eigengap condition (22) with* $K_0 = K$.

Under Assumption E.1, the population eigenvectors $\{\mathbf{v}_k\}_{k\in[K]}$ are uniformly delocalized as given by the following lemma.

Lemma E.1 (Delocalization of $\mathbf{v}_k$)**.** *For the DCSBM (E.1), assume that Assumption E.1 holds. Then for each fixed* $k \in [K]$*, there exists a constant* $C > 0$ *such that*

$$\|\mathbf{v}_k\|_\infty \le C/\sqrt{n}. \tag{E.4}$$

Proof. Let $Z \in \{0,1\}^{n\times K}$ be the community membership matrix defined as $Z_{ia} = \mathbf{1}\{g_i = a\}$, and $\mathbf{\Theta} = \operatorname{diag}(\vartheta_1, \ldots, \vartheta_n)$. Define the diagonal matrix $\mathbf{S} := \mathbf{\Theta}\mathbf{\Lambda}^{-\alpha}$. Then the population matrix can be written as

$$\mathbf{\Lambda}^{-\alpha}\mathbf{H}\mathbf{\Lambda}^{-\alpha} = \rho_n\, \mathbf{S}\mathbf{Z}\mathbf{B}\mathbf{Z}^T\mathbf{S}.$$

It follows that each population eigenvector $\mathbf{v}_k$ lies in the column span of $\mathbf{SZ}$. Hence, there exists some vector $\mathbf{u}_k \in \mathbb{R}^K$ such that $\mathbf{v}_k = \mathbf{SZu}_k$. Denote by $\mathbf{A} := \mathbf{Z}^T\mathbf{S}^2\mathbf{Z}$. Since $\mathbf{S}$ is diagonal and $\mathbf{Z}$ has disjoint indicator columns, $\mathbf{A}$ is diagonal with entries

$$A_{aa} = \sum_{i:g_i=a} \vartheta_i^2\, \Lambda_i^{-2\alpha} \sim nq^{-4\alpha},$$

where the scaling follows from (E.3) and the fact that $\Lambda_i \sim q^2$. Using the normalization $\|\mathbf{v}_k\|_2 = 1$, we can obtain that $1 = \mathbf{v}_k^T\mathbf{v}_k = \mathbf{u}_k^T\mathbf{A}\mathbf{u}_k$, which further entails that

$$|u_k(a)| \le 1/\sqrt{A_{aa}} \lesssim q^{2\alpha}/\sqrt{n}, \quad \forall a \in [K].$$

Finally, we can bound the entries of $\mathbf{v}_k$ as

$$|v_k(i)| \le \sum_{a\in[K]} |(\mathbf{SZ})_{ia}||u_k(a)| \lesssim q^{-2\alpha} \max_a |u_k(a)| \lesssim n^{-1/2}.$$

This establishes (E.4) and completes the proof of Lemma E.1. ☐

Remark 9. The following results extend directly to the more general DCMM model described in Example 1, with only minor modifications, provided that conditions (ii) and (iii) in Assumption E.1 are replaced by the following assumptions (recall that $q = \sqrt{n\theta} \gg (\log n)^4$ by Definition 1):

(ii') (Community membership) There exists a constant $0 < C_\pi < \infty$ such that

$$\|(\mathbf{\Pi}^T\mathbf{\Pi})^{-1}\| \;\le\; C_\pi n^{-1}.$$

(iii') (Mild degree heterogeneity) There exist constants $0 < c_\vartheta \le C_\vartheta < \infty$ such that, for all $i \in [n]$,

$$c_\vartheta\sqrt{\theta} \le \vartheta_i \le C_\vartheta\sqrt{\theta}, \quad \text{and} \quad c_\vartheta q^2 \le \Lambda_i = \sum_j p_{ij} \le C_\vartheta q^2.$$

Under these assumptions, definitions (12) and (13) imply that $\theta_i = \Lambda_i/(n\theta) \sim 1$ and $\beta_n = \min_i \theta_i \sim 1$. Moreover, using the argument in Appendix B.5 of Fan et al. (2022b), one obtains

$$\|\mathbf{v}_k\|_\infty \lesssim \|(\mathbf{\Pi}^T\mathbf{\Pi})^{-1}\|^{1/2} \lesssim n^{-1/2}.$$

With this estimate at hand, all of the subsequent arguments extend directly to the present setting, and we therefore omit the routine modifications. As mentioned earlier, we expect that our results can also be extended to DCMMs with severe degree heterogeneity (i.e., $\beta_n = o(1)$), at the cost of additional assumptions and more technical arguments. We do not pursue this direction in the current paper.

Note that under Assumption E.1, the DCSBM satisfies Definition 1 and Assumption 1. Hence, all conclusions in the main text apply to this model. For the remainder of this section, we focus on the central limit theorems (CLTs) stated in Corollaries 1 and 2:

- (Eigenvalue CLT) By Corollary 2, for each $k \in [K]$ we have

$$\frac{\widehat{\delta}_k - t_k - A_k}{t_k \varsigma_k} \xrightarrow{d} \mathcal{N}(0,1). \tag{E.5}$$

- (Eigenvector CLT) By Corollary 1, for each $k \in [K]$ and $i \in [n]$ we have

$$\frac{\widehat{v}_k(i) - v_k(i)}{\sigma_{k,i}} \xrightarrow{d} \mathcal{N}(0,1). \tag{E.6}$$

Analogous arguments can be developed for the CLTs of $\mathbf{u}^T\widehat{\mathbf{v}}_k$ in Corollary 3. However, these extensions involve *substantially more* technical work, and we leave a detailed treatment to future research. We adopt the estimation procedure proposed in Section III-C to estimate the asymptotic bias and variances in (E.5) and (E.6). For simplicity, we assume that the underlying rank K is known and finite.

a) *Residual-based estimation of* s_{ij}. Define the rank-K reconstruction

$$\widehat{\mathbf{H}}_0 := \mathbf{D}^{\alpha}\left(\sum_{k=1}^{K}\widehat{\delta}_k\,\widehat{\mathbf{v}}_k\widehat{\mathbf{v}}_k^T\right)\mathbf{D}^{\alpha}, \quad \widehat{\mathbf{W}}_0 := \widetilde{\mathbf{X}} - \widehat{\mathbf{H}}_0, \quad \widehat{s}_{ij,0} := \widehat{W}_{0,ij}^2. \tag{E.7}$$

b) *Feasible bias estimator.* Take $\widehat{t}_{k,0} = \widehat{\delta}_k$ and compute the feasible bias estimator $\widehat{A}_k \equiv \widehat{A}_{k,0}$ as in (78).

c) *Feasible variance estimators.* Define $\widehat{\varsigma}_k^2$ and $\widehat{\sigma}_{k,i}^2$ via the plug-in rules proposed in the main text:

$$\text{substituting } (t_k, \mathbf{v}_k, s_{ij}, \mathbf{\Lambda}) \text{ with their sample estimates } (\widehat{t}_{k,0}, \widehat{\mathbf{v}}_k, \widehat{s}_{ij,0}, \mathbf{D}). \tag{E.8}$$

The bias and variance expressions in (42), (49), and (53) depend on the noise variances s_{ij} only through weighted sums of form $\sum_{i\le j} a_{ij}\, s_{ij}$, where weights a_{ij} are built from spiked eigenvectors and diagonal powers of $\mathbf{\Lambda}$. In their corresponding sample versions, the weights are constructed from their empirical analogues. To justify the use of the plug-in estimators, we first establish a weighted law of large numbers (LLN) for W_{ij}^2.

Lemma E.2 (Weighted LLN for W_{ij}^2)**.** *Let $\{a_{ij}\}$ be deterministic weights, and $\sigma_n > 0$ a deterministic control parameter such that*

$$\sigma_n^{-2}\sum_{i\le j}|a_{ij}|^2 s_{ij} \xrightarrow{P} 0. \tag{E.9}$$

Then the relative error converges to zero in probability

$$\sigma_n^{-1}\sum_{i\le j}a_{ij}\left(W_{ij}^2 - s_{ij}\right) \xrightarrow{P} 0.$$

Proof. This lemma follows directly from Chebyshev's inequality together with the bound $\operatorname{var}(W_{ij}^2) \lesssim s_{ij}$. □

Let us fix any $k \in [K]$. Corresponding to (42), (49), and (53), we introduce the following deterministic population weight arrays for $i, j \in [n]$,

$$a_{ij}^{(k)} := \frac{\mathfrak{a}_{ij}^{(k)} + \mathfrak{a}_{ji}^{(k)}}{1+\delta_i^j} \quad \text{with} \quad \mathfrak{a}_{ij}^{(k)} := (2\alpha+1)\,t_k\,\frac{v_k(i)^2}{\Lambda_i^2} - 4\frac{v_k(i)v_k(j)}{\Lambda_i^{1+\alpha}\Lambda_j^{\alpha}}, \tag{E.10}$$

$$b_{ij}^{(k)} := \left(\frac{\mathfrak{S}_{ij}^{\mathbf{v}_k\mathbf{v}_k}}{1+\delta_i^j}\right)^2, \quad c_{ij}^{(k)} := \left(-\frac{\alpha v_k(i)}{\Lambda_i} + \frac{1}{t_k}\Lambda_i^{-\alpha}\Lambda_j^{-\alpha}v_k(j)\right)^2. \tag{E.11}$$

With the aid of these definitions, we can write

$$A_k = \alpha\sum_{i\le j}a_{ij}^{(k)}s_{ij}, \quad \varsigma_k^2 = \sum_{i\le j}b_{ij}^{(k)}s_{ij}, \quad \sigma_{k,i}^2 = \sum_j c_{ij}^{(k)}s_{ij}. \tag{E.12}$$

We will further impose the following assumption on these population quantities.

Assumption E.2. *For a given $k \in [K]$ and $i \in [n]$, there exists a constant $c > 0$ such that $\varsigma_k^2 \ge c(n|\mathsf{d}_k|^2)^{-1}$ and $\sigma_{k,i}^2 \ge c(n|\mathsf{d}_k|^2)^{-1}$, where d_k is defined in (29).*

Example 2. We provide a concrete DCSBM example in which Assumption E.2 above holds. Assume that the communities are balanced with $n_a = n/K$, the node degrees are homogeneous with $\vartheta_i \equiv 1$ for all $i \in [n]$, and the block probability matrix takes the diagonal form $\mathbf{B} = \operatorname{diag}(b_1, \cdots, b_K)$ with fixed constants $b_1 > b_2 > \cdots > b_K > 0$. Then the edge probabilities satisfy that $p_{ij} = \rho_n b_{g_i}\mathbf{1}\{g_i = g_j\}$, where g_i represents the community label of node i. Assume further that $\rho_n b_a \le 1-\epsilon$ for all $a \in [K]$ and some constant $\epsilon > 0$. A direct calculation shows that $\Lambda_i = n_a\rho_n b_a$ if $g_i = a$, the K nonzero population eigenvalues satisfy that $\delta_k \sim t_k \sim (n_k\rho_n b_k)^{1-2\alpha}$ for $k \in [K]$, and the corresponding eigenvectors satisfy that $v_k(i) = n_k^{-1/2}\mathbf{1}\{g_i = k\}$ for $k \in [K]$ and $i \in [n]$. It is then straightforward to verify that Assumption E.2 holds in this setting.

Lemma E.3. *Assume that Assumptions E.1 and E.2 hold. Then we have that*

$$\max_{i\le j}|a_{ij}^{(k)}|\lesssim n^{-1}q^{-2-4\alpha},\quad \max_{i\le j}|a_{ij}^{(k)}|^2\cdot\sum_{i\le j}s_{ij}\lesssim n^{-1}q^{-2-8\alpha}. \tag{E.13}$$

Furthermore, the arrays $\{b_{ij}^{(k)}:1\le i\le j\le n\}$ and $\{c_{ij}^{(k)}:j\in[n]\}$ (for a given $i\in[n]$) satisfy that

$$\max_{i\le j}|b_{ij}^{(k)}|\lesssim \frac{1}{nq^2}\varsigma_k^2,\quad \max_{i\le j}|b_{ij}^{(k)}|^2\cdot\sum_{i\le j}s_{ij}\lesssim \frac{1}{nq^2}\varsigma_k^4, \tag{E.14}$$

$$\max_{i\le j}|c_{ij}^{(k)}|\lesssim \frac{1}{q^2}\varsigma_k^2,\quad \max_{j}|c_{ij}^{(k)}|^2\cdot\sum_{j}s_{ij}\lesssim \frac{1}{q^2}\sigma_{k,i}^4. \tag{E.15}$$

Proof. Under Assumption E.1, it holds that $\Lambda_i\sim n\rho_n\sim q^2$ uniformly in i, and

$$t_k=(1+o(1))\delta_k\lesssim\max_i|\Lambda_i|^{-2\alpha}\cdot\|\mathbf{H}\|\lesssim q^{2-4\alpha}. \tag{E.16}$$

Moreover, Lemma E.1 yields that $\|\mathbf{v}_k\|_\infty\lesssim n^{-1/2}$. With these estimates, we can deduce that

$$\left|(2\alpha+1)\,t_k\,\frac{v_k(i)^2}{\Lambda_i^2}\right|\lesssim\frac{|t_k|}{nq^4}\lesssim\frac{1}{nq^{2+4\alpha}},\qquad \left|4\frac{v_k(i)v_k(j)}{\Lambda_i^{1+\alpha}\Lambda_j^{\alpha}}\right|\lesssim\frac{1}{nq^{2+4\alpha}},$$

which together imply the first bound in (E.13). Combining this estimate with $s_{ij}\sim\rho_n\sim\theta$, we can further show that

$$\max_{i\le j}|a_{ij}^{(k)}|^2\cdot\sum_{i\le j}s_{ij}\lesssim(nq^{2+4\alpha})^{-2}\cdot n^2\theta\lesssim n^{-1}q^{-2-8\alpha}.$$

This concludes (E.13). For the proof of (E.14), using $\Lambda_i\sim q^2$ and Lemma E.1, we can obtain that

$$|\mathfrak{S}_{ij}^{\mathbf{v}_k\mathbf{v}_k}|\lesssim\frac{1}{nq^2}+\frac{1}{n|t_k|q^{4\alpha}}\lesssim\frac{1}{n|\mathsf{d}_k|q},$$

where in the second step, we have used (E.16) and definition (29). Combining this bound with Assumption E.2 results in (E.14). The estimate (E.15) follows by a similar argument. □

Lemma E.4. *Assume that Assumption E.1 holds, and define $\widehat{\mathbf{H}}_0$ as in (E.7). Then we have that with high probability,*

$$\max_{k\in[K]}|\widehat{\delta}_k-\delta_k|\lesssim\varepsilon_n|\delta_k|, \tag{E.17}$$

$$\max_{k\in[K]}\|\widehat{\mathbf{v}}_k-\mathbf{v}_k\|_\infty\lesssim\varepsilon_n/\sqrt{n}, \tag{E.18}$$

$$\|\mathbf{H}-\widehat{\mathbf{H}}_0\|_F\lesssim\varepsilon_nq^2,\qquad\|\mathbf{H}-\widehat{\mathbf{H}}_0\|_{\max}\lesssim\varepsilon_n\theta, \tag{E.19}$$

where the error control parameter ε_n is defined as

$$\varepsilon_n:=\sqrt{\log n}\left(q^{-1}+|\mathsf{d}_k|^{-1}\right). \tag{E.20}$$

Proof. Using Lemma 1 and Theorem 1 (under $\beta_n\sim 1$, $q\gg(\log n)^4$, and $|\mathsf{d}_k|\gg 1$), we obtain

$$\frac{|\widehat{\delta}_k-\delta_k|}{|\delta_k|}\lesssim\frac{q^{2-8\alpha}}{|\delta_k|^2}+\frac{\log n}{q^2}=\frac{1}{|\mathsf{d}_k|^2}+\frac{\log n}{q^2}\ll\varepsilon_n, \tag{E.21}$$

which proves (E.17). Next, by (40) and Lemma E.1, with high probability,

$$\widehat{v}_k(i)-v_k(i)=-\frac{\alpha v_k(i)}{\Lambda_i}\sum_{j\in[n]}W_{ij}+\frac{1}{t_k}\sum_{j\in[n]}\Lambda_i^{-\alpha}W_{ij}\Lambda_j^{-\alpha}v_k(j)+O\left(\frac{\log n}{q^2\sqrt{n}}\right). \tag{E.22}$$

Applying Bernstein's inequality (C.6), the first two terms on the right-hand side above are bounded w.h.p. by

$$-\frac{\alpha v_k(i)}{\Lambda_i}\sum_{j\in[n]}W_{ij}+\frac{1}{t_k}\sum_{j\in[n]}\Lambda_i^{-\alpha}W_{ij}\Lambda_j^{-\alpha}v_k(j)\lesssim\frac{\sqrt{n\theta\log n}}{q^2\sqrt{n}}+\frac{\sqrt{n\theta\log n}}{\sqrt{n}|\delta_k|q^{4\alpha}}=\frac{\sqrt{\log n}}{\sqrt{n}}\left(\frac{1}{q}+\frac{1}{|\mathsf{d}_k|}\right)=\frac{\varepsilon_n}{\sqrt{n}}.$$

Substituting this bound into (E.22) yields (E.18). We now prove (E.19). Subtracting $\widehat{\mathbf{H}}_0$ from the population low-rank factorization $\mathbf{H} = \mathbf{\Lambda}^\alpha \big(\sum_{k=1}^K \delta_k\, \mathbf{v}_k \mathbf{v}_k^T\big)\mathbf{\Lambda}^\alpha$, we obtain

$$\mathbf{H} - \widehat{\mathbf{H}}_0 = (\mathbf{\Lambda}^\alpha - \mathbf{D}^\alpha)\Big(\sum_{k=1}^K \delta_k \mathbf{v}_k \mathbf{v}_k^T\Big)\mathbf{\Lambda}^\alpha + \mathbf{D}^\alpha\Big(\sum_{k=1}^K (\delta_k \mathbf{v}_k \mathbf{v}_k^T - \widehat{\delta}_k \widehat{\mathbf{v}}_k \widehat{\mathbf{v}}_k^T)\Big)\mathbf{\Lambda}^\alpha + \mathbf{D}^\alpha\Big(\sum_{k=1}^K \widehat{\delta}_k \widehat{\mathbf{v}}_k \widehat{\mathbf{v}}_k^T\Big)(\mathbf{\Lambda}^\alpha - \mathbf{D}^\alpha). \tag{E.23}$$

We bound the three terms on the right-hand side above. Under Assumption E.1 and by Lemma C.1, it holds that w.h.p.,

$$\|\mathbf{D}^\alpha - \mathbf{\Lambda}^\alpha\| \lesssim \sqrt{\log n}\, q^{2\alpha-1} \quad \text{and} \quad \|\mathbf{D}\| \lesssim q^2. \tag{E.24}$$

Moreover, combining (E.17) and (E.18), we have that w.h.p.,

$$\Big\|\delta_k \mathbf{v}_k \mathbf{v}_k^T - \widehat{\delta}_k \widehat{\mathbf{v}}_k \widehat{\mathbf{v}}_k^T\Big\|_F \le |\delta_k - \widehat{\delta}_k|\, \|\mathbf{v}_k \mathbf{v}_k^T\|_F + |\widehat{\delta}_k|\, \|\mathbf{v}_k \mathbf{v}_k^T - \widehat{\mathbf{v}}_k \widehat{\mathbf{v}}_k^T\|_F \lesssim \varepsilon_n |\delta_k|,$$
$$\Big\|\delta_k \mathbf{v}_k \mathbf{v}_k^T - \widehat{\delta}_k \widehat{\mathbf{v}}_k \widehat{\mathbf{v}}_k^T\Big\|_{\max} \le |\delta_k - \widehat{\delta}_k|\, \|\mathbf{v}_k \mathbf{v}_k^T\|_{\max} + |\widehat{\delta}_k|\, \|\mathbf{v}_k \mathbf{v}_k^T - \widehat{\mathbf{v}}_k \widehat{\mathbf{v}}_k^T\|_{\max} \lesssim \varepsilon_n |\delta_k|/n.$$

Plugging these bounds, together with (E.24), into (E.23) and using (E.16), we conclude that w.h.p.,

$$\|\mathbf{H} - \widehat{\mathbf{H}}_0\|_F \lesssim \varepsilon_n |\delta_k| q^{4\alpha} \lesssim \varepsilon_n q^2, \quad \|\mathbf{H} - \widehat{\mathbf{H}}_0\|_{\max} \lesssim \varepsilon_n |\delta_k| q^{4\alpha}/n \lesssim \varepsilon_n \theta.$$

This concludes the proof of (E.19). ☐

We next show that replacing W_{ij} with the estimator $\widehat{W}_{0,ij}$ does *not* affect the weighted LLN in Lemma E.2.

Lemma E.5 (Weighted LLN for $\widehat{s}_{ij,0} = \widehat{W}^2_{0,ij}$)**.** *Assume that Assumption E.1 holds. Let $\{a_{ij}\}$ be a sequence of deterministic weights satisfying (E.9) and $\max_{i\le j}|a_{ij}| \lesssim \widetilde{\sigma}_n$ for some deterministic control parameter $\widetilde{\sigma}_n > 0$. Then we have that*

$$\frac{\sum_{i\le j} a_{ij}\big(\widehat{s}_{ij,0} - s_{ij}\big)}{\sigma_n q^2 + \widetilde{\sigma}_n q^4} \xrightarrow{P} 0. \tag{E.25}$$

Similarly, if $a_{ij} = a_j \delta_{ij}$ is a sequence of diagonal deterministic weights satisfying (E.9) and $\max_j |a_j| \lesssim \widetilde{\sigma}_n$, then we have that

$$\frac{\sum_j a_j\big(\widehat{s}_{ij,0} - s_{ij}\big)}{\sigma_n q + \widetilde{\sigma}_n q^2} \xrightarrow{P} 0. \tag{E.26}$$

Proof. By definition, we have the expansion

$$\sum_{i\le j} a_{ij}(\widehat{s}_{ij,0} - s_{ij}) = \sum_{i\le j} a_{ij}\big(W_{ij}^2 - s_{ij}\big) + \sum_{i\le j} a_{ij}(\mathbf{H} - \widehat{\mathbf{H}}_0)^2_{ij} + 2\sum_{i\le j} a_{ij} W_{ij}(\mathbf{H} - \widehat{\mathbf{H}}_0)_{ij}.$$

In view of Lemma E.2, the first term on the right-hand side above is $o_P(\sigma_n)$ under condition (E.9). For the second term, using the bound $\max_{i\le j}|a_{ij}| \lesssim \widetilde{\sigma}_n$ together with (E.19), we obtain that w.h.p.,

$$\Big|\sum_{i\le j} a_{ij}(\mathbf{H} - \widehat{\mathbf{H}}_0)^2_{ij}\Big| \lesssim \max_{i\le j}|a_{ij}| \cdot \|\mathbf{H} - \widehat{\mathbf{H}}_0\|_F^2 \lesssim \widetilde{\sigma}_n \cdot (\varepsilon_n q^2)^2 \ll \widetilde{\sigma}_n q^4.$$

For the cross term, by the Cauchy–Schwarz inequality, it holds that

$$\mathbb{E}\Big|\sum_{i\le j} a_{ij} W_{ij}(\mathbf{H} - \widehat{\mathbf{H}}_0)_{ij}\Big|^2 \le \Big(\mathbb{E}\sum_{i\le j}|a_{ij}|^2 W_{ij}^2\Big) \cdot \mathbb{E}\|\mathbf{H} - \widehat{\mathbf{H}}_0\|_F^2 \lesssim (\varepsilon_n q^2)^2 \sum_{i\le j}|a_{ij}|^2 s_{ij} \ll \sigma_n^2 q^4.$$

Combining the above bounds and applying Chebyshev's inequality yield (E.25). The proof of (E.26) is similar and therefore omitted here for simplicity. ☐

We are now ready to state the end-to-end DCSBM results: the plug-in estimators preserve the limiting laws.

Theorem E.1 (Feasible eigenvalue and eigenvector CLTs)**.** *Assume that Assumption E.1 is satisfied and Assumption E.2 holds for a fixed $k \in [K]$ and $i \in [n]$. Construct $(\widehat{A}_{k,0}, \widehat{t}_{k,0}, \widehat{\varsigma}_k^2, \widehat{\sigma}_{k,i}^2)$ according to the procedure described between (E.7) and (E.8).*

(i) Assume that the population eigenvalue CLT (E.5) holds for a fixed $k \in [K]$. In addition, assume that

$$|\mathsf{d}_k| \gg n^{1/4}, \quad q \gg n^{1/4}\log n, \tag{E.27}$$

recalling (52). *Then we have that*

$$\frac{\widehat{A}_{k,0} - A_k}{t_k \varsigma_k} \xrightarrow{P} 0, \quad \frac{\widehat{\varsigma}_k}{\varsigma_k} \xrightarrow{P} 1, \quad \frac{\widehat{t}_{k,0}}{t_k} \xrightarrow{P} 1, \tag{E.28}$$

and consequently,

$$\frac{\widehat{\delta}_k - t_k - \widehat{A}_{k,0}}{\widehat{t}_{k,0}\widehat{\varsigma}_k} \xrightarrow{d} \mathcal{N}(0,1).$$

(ii) Assume that the population eigenvector CLT (E.6) *holds. Then we have that*

$$\widehat{\sigma}_{k,i}/\sigma_{k,i} \xrightarrow{P} 1, \tag{E.29}$$

and consequently,

$$\frac{\widehat{v}_k(i) - v_k(i)}{\widehat{\sigma}_{k,i}} \xrightarrow{d} \mathcal{N}(0,1).$$

Proof. We focus on the proof of part (i); the proof of part (ii) is similar. In particular, the argument for (E.29) parallels that for the convergence $\widehat{\varsigma}_k/\varsigma_k \xrightarrow{P} 1$, and relies on the same technical ingredients developed in Lemmas E.2–E.5. The estimate $\widehat{t}_{k,0}/t_k \xrightarrow{P} 1$ follows directly from Theorem 1 and Lemma 1. It therefore remains to establish the first two convergences in (E.28).

Proof of $(\widehat{A}_{k,0} - A_k)/(t_k\varsigma_k) \xrightarrow{P} 0$. Recall the expression of the bias term A_k in (E.12). We define the plug-in estimates of the weights by

$$\widehat{a}_{ij}^{(k)} := \frac{\widehat{\mathfrak{a}}_{ij}^{(k)} + \widehat{\mathfrak{a}}_{ji}^{(k)}}{1+\delta_i^j} \quad \text{with} \quad \widehat{\mathfrak{a}}_{ij,0}^{(k)} := (2\alpha+1)\widehat{t}_{k,0}\frac{\widehat{v}_k(i)^2}{d_i^2} - 4\frac{\widehat{v}_k(i)\widehat{v}_k(j)}{d_i^{1+\alpha}d_j^{\alpha}}.$$

With these definitions, we can decompose $\widehat{A}_{k,0} - A_k$ as

$$\widehat{A}_{k,0} - A_k = \underbrace{\alpha\sum_{i\le j} a_{ij}^{(k)}\big(\widehat{s}_{ij,0} - s_{ij}\big)}_{=:T_{1,k}} + \underbrace{\alpha\sum_{i\le j}\big(\widehat{a}_{ij,0}^{(k)} - a_{ij}^{(k)}\big)\widehat{s}_{ij,0}}_{=:T_{2,k}}. \tag{E.30}$$

Thus, it suffices to show that $T_{r,k} = o_P(t_k\varsigma_k)$ for $r \in \{1,2\}$.

We first bound $T_{1,k}$. By Lemma E.3, and using the fact that $s_{ij} \sim \theta$ so that $\sum_{i\le j} s_{ij} \sim nq^2$, the deterministic array $\{a_{ij}^{(k)}\}_{i\le j}$ satisfies the weighted LLN condition (E.9) for any $\sigma_n \gg n^{-1/2}q^{-1-4\alpha}$, together with the bound $\max_{i\le j}|a_{ij}^{(k)}| \lesssim \widetilde{\sigma}_n$, where $\widetilde{\sigma}_n = n^{-1}q^{-2-4\alpha}$. Hence, applying Lemma E.5 with weights $a_{ij} = a_{ij}^{(k)}$ gives that

$$T_{1,k} = \alpha\sum_{i\le j} a_{ij}^{(k)}\big(\widehat{s}_{ij,0} - s_{ij}\big) = o_P(n^{-1/2}q^{1-4\alpha}) = o_P\Big(|\delta_k|/(n|\mathsf{d}_k|^2)^{1/2}\Big) = o_P(t_k\varsigma_k), \tag{E.31}$$

where Assumption E.2 is used in the final step. We next bound $T_{2,k}$. Using the estimates (E.17) and (E.18), along with an argument analogous to that in Lemma E.3, we can obtain that w.h.p.,

$$\max_{i,j\in[n]}|\widehat{a}_{ij,0}^{(k)} - a_{ij}^{(k)}| \lesssim \varepsilon_n \max_{i,j\in[n]}|a_{ij}^{(k)}| \lesssim \varepsilon_n n^{-1}q^{-2-4\alpha}. \tag{E.32}$$

On the other hand, recall that $\widehat{s}_{ij,0} = \widehat{W}_{0,ij}^2$, where $\widehat{\mathbf{W}}_0 = \mathbf{W} + (\mathbf{H} - \widehat{\mathbf{H}}_0)$. Consequently, it holds that

$$\begin{aligned}\sum_{i\le j}\widehat{s}_{ij,0} &= \sum_{i\le j}\widehat{W}_{0,ij}^2 \le 2\sum_{i\le j}W_{ij}^2 + 2\sum_{i\le j}(H - \widehat{H}_0)_{ij}^2 \\ &\lesssim_P \sum_{i\le j}s_{ij} + \sum_{i\le j}(W_{ij}^2 - s_{ij}) + (\varepsilon_n q^2)^2 \lesssim_P nq^2,\end{aligned} \tag{E.33}$$

where we have used (E.19) in the third step and applied the LLN from (E.2) in the final step. Combining (E.32) with (E.33), we can show that

$$|T_{2,k}| \lesssim_P \varepsilon_n n^{-1}q^{-2-4\alpha}\sum_{i\le j}\widehat{s}_{ij,0} \lesssim_P \varepsilon_n q^{-4\alpha}.$$

Using the definition of ε_n in (E.20) along with condition (E.27), we can further deduce that

$$|T_{2,k}| \lesssim_P \varepsilon_n q^{-4\alpha} \ll n^{-1/2}q^{1-4\alpha} \lesssim |t_k|\varsigma_k. \tag{E.34}$$

Hence, combining (E.31) and (E.34), we conclude that $(\widehat{A}_{k,0} - A_k)/(t_k \varsigma_k) \xrightarrow{P} 0$.

Proof of $\widehat{\varsigma}_k/\varsigma_k \xrightarrow{P} 1$. We define the plug-in estimate $\widehat{\varsigma}_k^2$ as

$$\widehat{\varsigma}_k^2 = \sum_{1\le i\le j\le n} \widehat{b}_{ij,0}^{(k)}\, \widehat{s}_{ij,0} \quad \text{with} \quad \widehat{b}_{ij,0}^{(k)} := \left(\frac{\widehat{t}_k \widehat{\mathfrak{S}}_{ij}^{\widehat{\mathbf{v}}_k \widehat{\mathbf{v}}_k}}{1+\delta_i^j} \right)^2 . \tag{E.35}$$

By Assumption E.2, it suffices to show that

$$\frac{\widehat{\varsigma}_k^2 - \varsigma_k^2}{\varsigma_k^2} \xrightarrow{P} 0. \tag{E.36}$$

Similar to (E.30), we have the following decomposition

$$\widehat{\varsigma}_k^2 - \varsigma_k^2 = \underbrace{\sum_{i\le j} b_{ij}^{(k)} \big(\widehat{s}_{ij,0} - s_{ij}\big)}_{=:U_{1,k}} + \underbrace{\sum_{i\le j} \big(\widehat{b}_{ij,0}^{(k)} - b_{ij}^{(k)}\big)\widehat{s}_{ij,0}}_{=:U_{2,k}} . \tag{E.37}$$

Hence, it suffices to show that $U_{r,k} = o_P(\varsigma_k^2)$ for $r \in \{1,2\}$.

We first bound $U_{1,k}$. By Lemma E.3, and using the fact that $\sum_{i\le j} s_{ij} \sim nq^2$, the deterministic array $\{b_{ij}^{(k)}\}_{i\le j}$ satisfies the weighted LLN condition (E.9) for any $\sigma_n \gg n^{-1/2} q^{-1} \varsigma_k^2$, together with the uniform bound $\max_{i\le j} |b_{ij}^{(k)}| \lesssim \widetilde{\sigma}_n$, where $\widetilde{\sigma}_n = n^{-1} q^{-2} \varsigma_k^2$. Then an application of Lemma E.5 with weights $a_{ij} = b_{ij}^{(k)}$ leads to

$$U_{1,k} = \sum_{i\le j} b_{ij}^{(k)} \big(\widehat{s}_{ij,0} - s_{ij}\big) = o_P(qn^{-1/2}\varsigma_k^2) = o_P(\varsigma_k^2). \tag{E.38}$$

We next bound $U_{2,k}$. To control the uniform weight error $\max_{i\le j} |\widehat{b}_{ij,0}^{(k)} - b_{ij}^{(k)}|$, note that the map

$$(\mathbf{x}, \mathbf{y}, \Lambda_i, \Lambda_j) \mapsto \mathfrak{S}_{ij}^{\mathbf{x}\mathbf{y}},$$

defined in (50), is smooth in a neighborhood of $(\mathbf{v}_k, \mathbf{v}_k, \Lambda_i, \Lambda_j)$, with denominators bounded away from 0 (since $\Lambda_i, \Lambda_j \sim q^2$). Then using the estimates (C.2) (for $|d_i - \Lambda_i|$), (E.17), and (E.18), together with a mean-value expansion of $\widehat{\mathfrak{S}}_{ij}^{\widehat{\mathbf{v}}_k \widehat{\mathbf{v}}_k} - \mathfrak{S}_{ij}^{\mathbf{v}_k \mathbf{v}_k}$, we can deduce that w.h.p.,

$$\max_{i,j\in[n]} \left| \widehat{\mathfrak{S}}_{ij}^{\widehat{\mathbf{v}}_k \widehat{\mathbf{v}}_k} - \mathfrak{S}_{ij}^{\mathbf{v}_k \mathbf{v}_k} \right| \lesssim \varepsilon_n \max_{i,j\in[n]} |\mathfrak{S}_{ij}^{\mathbf{v}_k \mathbf{v}_k}| \lesssim \varepsilon_n n^{-1} q^{-2} \varsigma_k^2. \tag{E.39}$$

Combining this bound with (E.33), we can obtain that w.h.p.,

$$|U_{2,k}| \le \max_{i\le j} \big|\widehat{b}_{ij,0}^{(k)} - b_{ij}^{(k)}\big| \sum_{i\le j} \widehat{s}_{ij,0} \lesssim_P \varepsilon_n \varsigma_k^2. \tag{E.40}$$

Therefore, combining (E.38) and (E.40), we conclude (E.36), which completes the proof of Theorem E.1. ☐

Appendix F
Additional simulation results

In this section of the Supplementary Material, we present additional simulation results in the *original (unrescaled) setting*. Specifically, Figures 11–13 serve as the counterparts of Figures 2–4, respectively, where the empirical spiked eigenvalue $\widehat{\delta}_k$ is corrected using the data-driven estimator $\widehat{A}_k$ in (78), rather than the theoretical quantity A_k with asymptotic limit t_k. In these figures, the blue curves display the KDEs of the corrected empirical spiked eigenvalues using $\widehat{A}_k$, while the red curves represent the corresponding target normal densities. Both curves are centered at the asymptotic limit t_k. On the other hand, Figures 14–16 correspond to Figures 2–4, respectively, where $\widehat{\delta}_k$ is corrected by $\widehat{A}_k$ in (78) together with the empirical bias correction procedure proposed around (79), targeting at the population quantity δ_k. In these figures, the blue curves represent the KDEs of the empirically bias-corrected spiked eigenvalues, while the red curves display the target normal densities. Both curves are centered at the asymptotic limit δ_k.

Overall, from Figures 11–16, we observe that both approaches—bias correction via $\widehat{A}_k$ toward the population quantity t_k, and the combined correction using $\widehat{A}_k$ together with the empirical bias adjustment toward δ_k—perform well across all tested settings for the empirical spiked eigenvalues $\widehat{\delta}_k$ of the generalized Laplacian matrix $\mathbf{X}$.

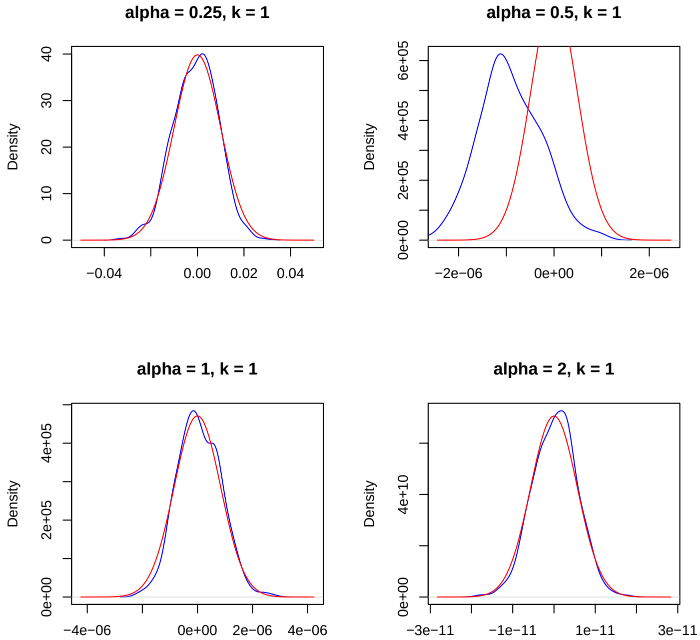

Fig. 11: Kernel density estimates (KDEs) of the distribution of the empirical spiked eigenvalue $\widehat{\delta}_k$ (with $k = 1$) corrected using the estimator $\widehat{A}_k$. Results are based on 500 replications for the simulation setting in Section VI with $\theta = 0.9$. Both the empirical KDEs (blue) and the target normal densities (red) are centered at the asymptotic limit t_k. The relatively concentrated behavior observed in the top-right panel is due to the small empirical standard deviation, which stems from the fact that the normalized Laplacian has a trivial largest eigenvalue equal to 1.

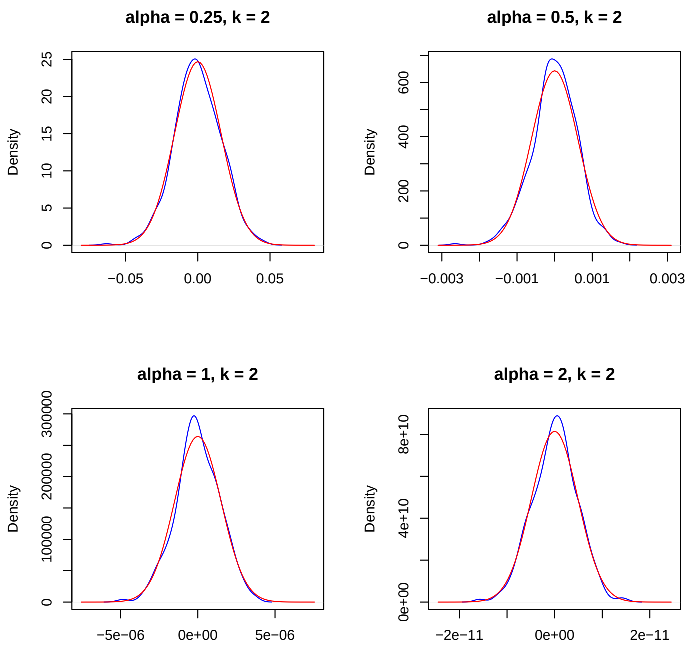

Fig. 12: KDEs of the distribution of the empirical spiked eigenvalue $\widehat{\delta}_k$ (with $k = 2$) corrected using the estimator $\widehat{A}_k$. Results are based on 500 replications for the simulation setting in Section VI with $\theta = 0.9$. Both the empirical KDEs (blue) and the target normal densities (red) are centered at the asymptotic limit t_k.

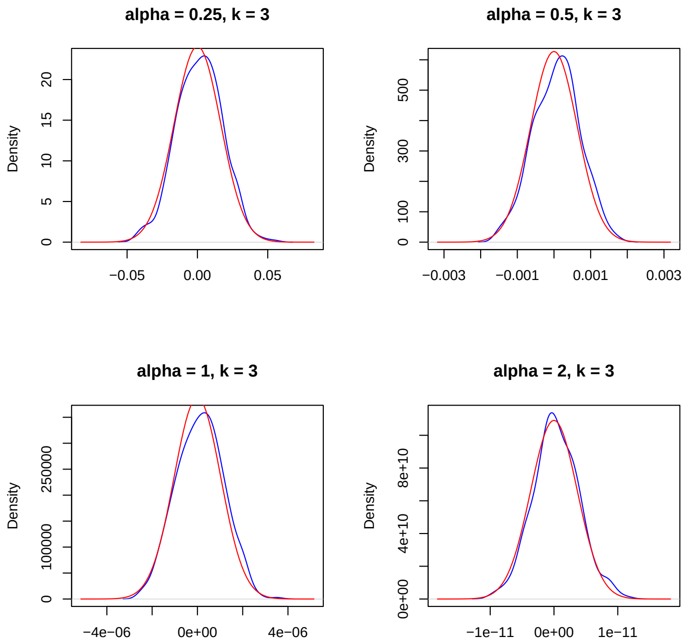

Fig. 13: KDEs of the distribution of the empirical spiked eigenvalue $\widehat{\delta}_k$ (with $k = 3$) corrected using the estimator $\widehat{A}_k$. Results are based on 500 replications for the simulation setting in Section VI with $\theta = 0.9$. Both the empirical KDEs (blue) and the target normal densities (red) are centered at the asymptotic limit t_k.

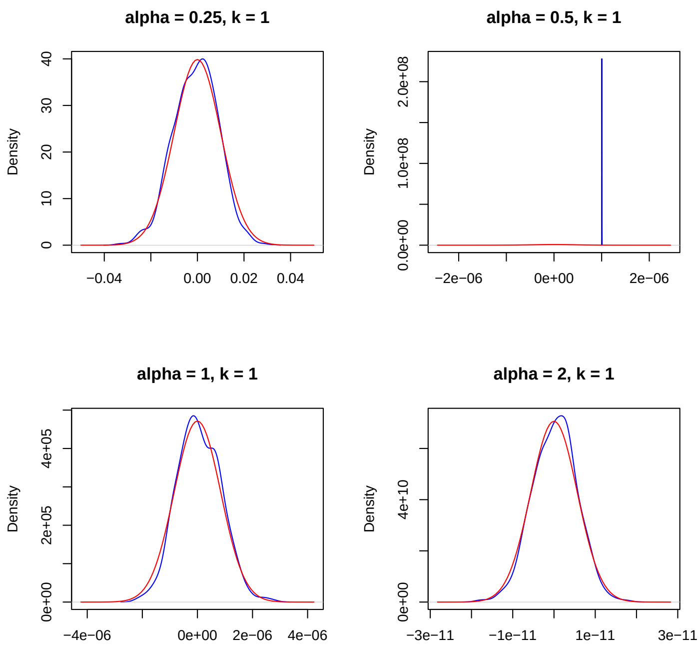

Fig. 14: KDEs of the distribution of the empirical spiked eigenvalue $\widehat{\delta}_k$ (with $k = 1$), corrected by the estimator $\widehat{A}_k$ together with the empirical bias correction procedure proposed around (79). Results are based on 500 replications for the simulation setting in Section VI with $\theta = 0.9$. Both the empirical KDEs (blue) and the target normal densities (red) are centered at δ_k. The concentrated behavior observed in the top-right panel again stems from the fact that the normalized Laplacian has a trivial largest eigenvalue equal to 1.

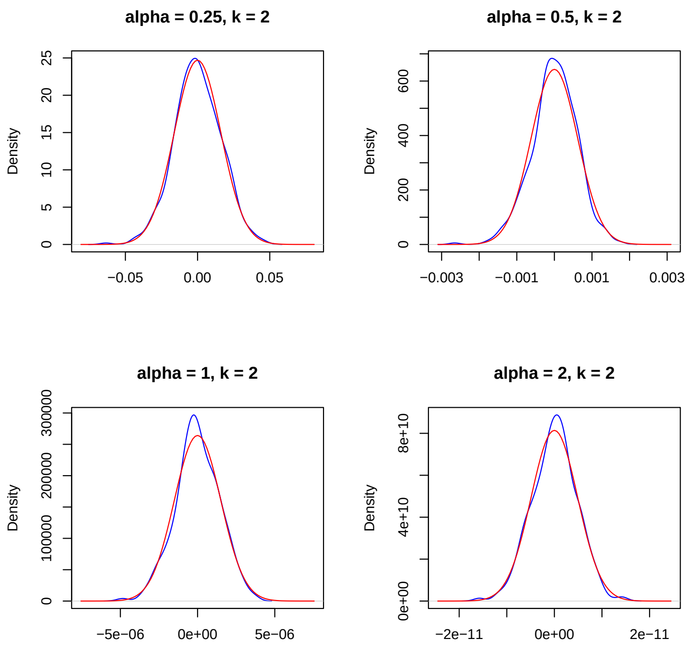

Fig. 15: KDEs of the distribution of the empirical spiked eigenvalue $\widehat{\delta}_k$ (with $k = 2$), corrected by the estimator $\widehat{A}_k$ together with the empirical bias correction procedure proposed around (79). Results are based on 500 replications for the simulation setting in Section VI with $\theta = 0.9$. Both the empirical KDEs (blue) and the target normal densities (red) are centered at δ_k.

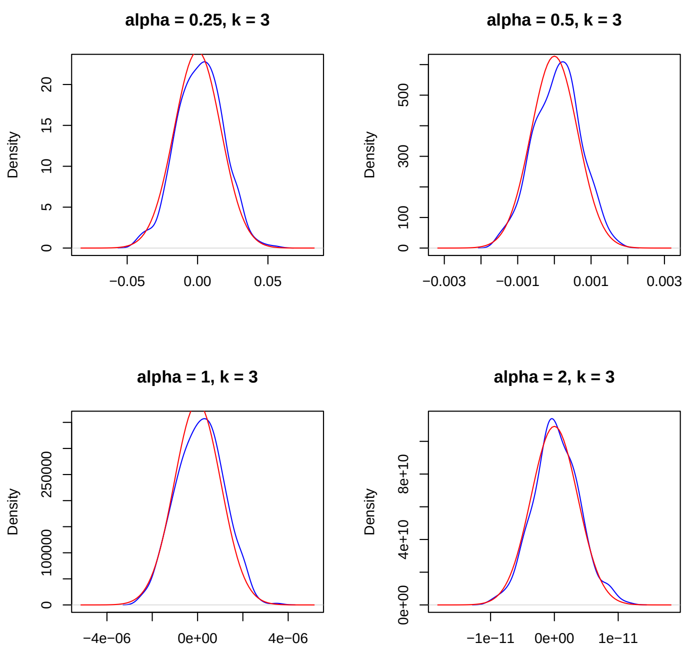

Fig. 16: KDEs of the distribution of the empirical spiked eigenvalue $\widehat{\delta}_k$ (with $k = 3$), corrected by the estimator $\widehat{A}_k$ together with the empirical bias correction procedure proposed around (79). Results are based on 500 replications for the simulation setting in Section VI with $\theta = 0.9$. Both the empirical KDEs (blue) and the target normal densities (red) are centered at δ_k.