Available online at www.sciencedirect.com
SCIENCE DIRECT®
LINEAR ALGEBRA AND ITS APPLICATIONS
ELSEVIER Linear Algebra and its Applications 377 (2004) 219–240
www.elsevier.com/locate/laa

Distribution of the eigenvalues of random block-matrices

Marianna Bolla [1]

Institute of Mathematics, Budapest University of Technology and Economics, P.O. Box 91, Bldg. H, V/7, 1521 Budapest, Hungary

Received 21 November 2001; accepted 20 August 2003

Submitted by R.A. Brualdi

Abstract

The asymptotic behaviour of the eigenvalues of random block-matrices is investigated with block sizes tending to infinity in the same order. In the proofs some extended version of Wigner's semi-circle law as well as perturbation results for symmetric matrices are used. The paper also deals with the asymptotic distribution of the protruding eigenvalues and some applications to random graphs.

AMS classification: 15A52; 15A42; 60F05

Keywords: Random matrices; Inequalities involving eigenvalues and eigenvectors; Central limit and other weak theorems

1. Introduction

Since Wigner [9] stated his famous semi-circle law, a variety of similar results has appeared. Some authors weakened Wigner's conditions (e.g., Arnold [1]), while some others investigated the largest eigenvalue that does not obey the semi-circle law. In fact, the weak convergence in probability of Wigner's theorem allows for at most one protruding eigenvalue (see [5]). In the sequel we shall use the following results:

E-mail address: marib@math.bme.hu.
[1] Research supported by the Foundation for Research, Development in Higher Education (FKFP) Grant no. 0058/2001.

doi:10.1016/j.laa.2003.08.015

Theorem FK 1 [5]. *Let w_{ij}, $i \geqslant j$, be independent (not necessarily identically distributed) random variables bounded with a common bound K (i.e., there is a $K > 0$ such that $|w_{ij}| \leqslant K$, $\forall i, j$). Assume that for $i > j$, the w_{ij}'s have a common expectation μ and variance σ^2, further that $\mathrm{E}w_{ii} = \nu$. Define w_{ij} for $i < j$ by $w_{ij} = w_{ji}$.*

The numbers K, μ, σ^2, ν will be kept fixed as the size n of the random symmetric matrix $\mathbf{W} = (w_{ij})$ will tend to infinity. Let $\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_n$ be the eigenvalues of $\mathbf{W}$.

If $\mu > 0$ then the distribution of the largest eigenvalue λ_1 can be approximated in order $1/\sqrt{n}$ by a normal distribution of expectation

$$(n-1)\mu + \nu + \sigma^2/\mu$$

and variance $2\sigma^2$. Further, with probability tending to 1,

$$\max_{i \geqslant 2} |\lambda_i| < 2\sigma\sqrt{n} + \mathrm{O}(n^{1/3}\log n).$$

Theorem FK 2 [5]. *Under the conditions of Theorem 1, if $\mu = 0$ then*

$$\max_{1 \leqslant i \leqslant n} |\lambda_i| = 2\sigma\sqrt{n} + \mathrm{O}(n^{1/3}\log n),$$

with probability tending to 1.

In fact, the same statements could be proved under the conditions of the semicircle law (the entries in and above the main diagonal are independent, those above are identically distributed and those in are also, possibly with another distribution, moreover both have finite moments; otherwise the matrix is symmetric). For independent, Bernoully distributed entries see [6]. Theorems FK1 and FK2 above do not postulate the identical distribution.

Also, if all the entries in and above the main diagonal are independent and normally distributed, the result of Theorem FK2 can be proved with the method of Mehta [7]. In this case, however, the entries in and those above the main diagonal have the same mean and variance, so they are also identically distributed, but not uniformly bounded. In the sequel, whenever we refer to the conditions of Theorem FK1 to hold for a matrix, the conditions of normality or those of Wigner can be substituted for them.

Throughout the paper the following generalization of the above model is investigated. The symmetric $n \times n$ matrix $\mathbf{W}$ is decomposed as the sum of the deterministic matrix $\mathbf{B}$ and the stochastic matrix $\mathbf{P}$. Here $\mathbf{B}$ is a block-diagonal matrix with diagonal blocks $\mathbf{B}_i$, $i = 1, \ldots, k$. The $n_i \times n_i$ symmetric matrix $\mathbf{B}_i$ has the positive real numbers μ_i's as entries, except in its main diagonal where it has entries ν_i's. Here k is a fixed positive integer and $n = \sum_{i=1}^{k} n_i$. The entries in and above the main diagonal of the $n \times n$ random matrix $\mathbf{P}$ are independent, uniformly bounded random variables (their common bound is K) with zero expectation and with the same variance σ^2, further, $\mathbf{P}$ is symmetric.

In this setup K, k, σ^2, μ_i, and ν_i $(i = 1, \ldots, k)$ will be kept fixed while $n_1, \ldots, n_k$ will tend to infinity in the same order. Under these conditions, in Section 2 we give a rough estimate for the eigenvalues of the $n \times n$ random symmetric block-matrix $\mathbf{W}$ as $n \to \infty$. Even from this rough characterization it turns out that there is a spectral gap between the k largest and the remaining eigenvalues of $\mathbf{W}$ (see Theorem 1). In Section 3, we investigate the distance between the corresponding eigenspaces (see Theorem 2). Using these results, in Section 4 we give a finer estimation of the protruding eigenvalues. Namely, it is proved that they asymptotically follow a k-variate normal distribution (see Theorem 4).

In Section 5, some applications and further remarks are included. Such random block matrices frequently occur in the presence of large data structures when the n observations come from, say, k loosely connected strata ($k \ll n$). E.g., the sociologist M.S. Granovetter investigated the strength of weak ties in the American Journal of Sociology 78 (1973) 1360–1380. He proved that loose connections between strongly connected strata can positively help in finding job.

Our data can also be viewed as a random graph $G = (V, \mathbf{W})$, $V = \{v_1, \ldots, v_n\}$ being the vertex set and the entries of the $n \times n$ symmetric matrix $\mathbf{W}$ being real values assigned to the edges as random weights, the diagonal entries belong to the loops. The model $\mathbf{W} = \mathbf{B} + \mathbf{P}$ gives rise to a k-partition, or equivalently, a coloring of the vertices with k different colors in such a way that v_i and v_j have the same color l if and only if the entry in the (i, j) position of $\mathbf{B}$ is also an entry of the block $\mathbf{B}_l$ for some l, $1 \leqslant l \leqslant k$. The edges are colored too, namely, we color an edge by the colors of its endpoints. An edge is called monocolored if both endpoints have the same color, and it is bicolored otherwise. In our model the monocolored edges have weights of positive expectation, while the bicolored ones of zero expectation. Sometimes we have a digraph: for $i \leqslant j$, $w_{ij} > 0$ means an $i \to j$ edge, while $w_{ij} < 0$ stands for a $j \to i$ edge ($w_{ij} = 0$ if and only if there is no edge between i and j, in the case of loops the sign is immaterial), otherwise $\mathbf{W}$ is symmetric.

For example, some special communication, biological, and sociological networks satisfy the above conditions. In practice, to detect the latent block structure, we have to investigate the spectral decomposition of the matrix $\mathbf{W}$. The number of the protruding eigenvalues gives the number of clusters, while on the basis of the corresponding eigenvectors the vertices can be represented in k dimension (see Theorem 3 in Section 3), and a metric classification technique is applicable to them.

Some remarks concerning recent empirical results on the spectra of some “real-world” and sparse graphs are also made (see [4]). In these models, the investigated matrices do not satisfy the conditions of Wigner’s theorem as the entries above the main diagonal are not independent, and even their distribution changes with n. In these cases there is a high peak of the spectral density around zero. To the contrary, in our model the eigenvalues follow the semi-circle law around zero, the only difference is that we have k dominant eigenvalues.

2. Rough characterization of the spectrum

Theorem 1. *Let the random matrix* $\mathbf{W}$ *be the sum of the* $n \times n$ *symmetric matrices* $\mathbf{B}$ *and* $\mathbf{P}$ *defined in the following way. Let* k *be a positive integer,* $\mu_i > 0$, *and* ν_i

$(i = 1, \dots, k)$ be real numbers. The deterministic matrix $\mathbf{B}$ *is the Kronecker-sum of the matrices* $\mathbf{B}_i$, $i = 1, \dots, k$, *where* $\mathbf{B}_i$ *is the* $n_i \times n_i$ *symmetric matrix with non-diagonal entries* μ_i*'s and diagonal ones* ν_i*'s. The entries in and above the main diagonal of the matrix* $\mathbf{P} = (p_{ij})$ *are independent (not necessarily identically distributed) random variables, uniformly bounded with the constant* K, *they have zero expectation and variance* σ^2, *and* $p_{ij} = p_{ji}$ *for* $i > j$. *The numbers* k, K, σ, μ_i, *and* ν_i $(i = 1, \dots, k)$ *will be kept fixed as* $n_1, \dots, n_k$ *will tend to infinity in such a way that* $n/n_i \leqslant C_0$ $(i = 1, \dots, k)$, *with* $n = \sum_{i=1}^{k} n_i$ *and positive constant* C_0 $(C_0 \geqslant k)$.

Then, with probability tending to 1, *for the eigenvalues* $\lambda_1, \dots, \lambda_n$ *of* $\mathbf{W}$ *the following inequalities will hold. There is an ordering of the* k *largest eigenvalues* $\lambda_1, \dots, \lambda_k$ *such that*

$$|\lambda_i - [(n_i - 1)\mu_i + \nu_i]| \leqslant 2\sigma\sqrt{n} + \mathrm{O}(n^{1/3}\log n), \quad i = 1, \dots, k. \tag{1}$$

Among the other eigenvalues, for $i = 1, \dots, k$ *there are* $n_i - 1$, λ_j*'s with*

$$|\lambda_j - [\nu_i - \mu_i]| \leqslant 2\sigma\sqrt{n} + \mathrm{O}(n^{1/3}\log n). \tag{2}$$

Proof. The set of the eigenvalues of the matrix $\mathbf{B}$ is the union of those of $\mathbf{B}_i$'s. It is known that the eigenvalues of $\mathbf{B}_i$ are $\beta_i := (n_i - 1)\mu_i + \nu_i$ and $n_i - 1$ numbers equal to $\nu_i - \mu_i$, $i = 1, \dots, k$.

We can use Weyl's perturbation theorem for the eigenvalues of Hermitian (in our case symmetric) matrices with perturbation $\mathbf{P}$ (see [10]). According to this, if the eigenvalues of $\mathbf{W}$ and $\mathbf{B}$ are matched in descending order, then the difference between the corresponding ones is at most the largest absolute value eigenvalue of $\mathbf{P}$. Since the random matrix $\mathbf{P}$ satisfies the conditions of Theorem FK2, for the eigenvalues of $\mathbf{P}$ we have

$$\varepsilon := \max_{i=1}^{n} |\lambda_i(\mathbf{P})| = 2\sigma\sqrt{n} + \mathrm{O}(n^{1/3}\log n), \tag{3}$$

with probability tending to 1. This finishes the proof. □

Note that the last $n - k$ eigenvalues of $\mathbf{W}$ are much smaller than the k large ones, the gap between the two groups being $\Delta - 2\varepsilon$, where

$$\Delta := \min_{i=1}^{k} [(n_i - 1)\mu_i + \nu_i] - \max_{i=1}^{k} (\nu_i - \mu_i). \tag{4}$$

Especially, if $n_1 = \cdots = n_k = n/k$, $\mu_1 = \cdots = \mu_k = \mu$, and $\nu_1 = \cdots = \nu_k = \nu$, then the k largest eigenvalues are around $(n/k - 1)\mu + \nu$ within a strip of width ε, while the other ones are around $\nu - \mu$ within also an ε-strip, and the gap is $(n/k)\mu - 2\varepsilon \to \infty$ as $n \to \infty$, with probability tending to 1. In the generic case, the spectral gap, $\Delta - 2\varepsilon$ is also of order n. That is, due to the conditions imposed on n_i the relation $n/C_0 \leqslant n_i \leqslant n$ holds, therefore n_i is of the same order as n (we shall denote it by $n_i \asymp n$), and ε is asymptotically equal to $2\sigma\sqrt{n}$ (we shall denote it by $\varepsilon \sim 2\sigma\sqrt{n}$).

No matter how small ε is compared to Δ, it can cause a huge fluctuation (of order $\sqrt{n}$) in the k largest eigenvalues of $\mathbf{W}$. In Theorem 4 it will be shown that in fact, this fluctuation is finite with high probability. To this end, the behaviour of the corresponding eigenvectors will be investigated.

3. Eigenvectors, eigenspaces, and representation

Theorem 2. *Let the random matrix* $\mathbf{W} = \mathbf{B} + \mathbf{P}$ *be the same as in Theorem* 1. *Let* $\mathbf{y}_1, \ldots, \mathbf{y}_n$ *and* $\mathbf{u}_1, \ldots, \mathbf{u}_n$ *be sets of orthonormal eigenvectors corresponding to the eigenvalues* (*in descending order*) *of* $\mathbf{W}$ *and* $\mathbf{B}$, *respectively*, *further* $F := \mathrm{Span}\{\mathbf{u}_1, \ldots, \mathbf{u}_k\} \subset \mathbb{R}^n$ *k-dimensional subspace. Now*, $n_1, \ldots, n_k$ *and*, *of course*, $n = \sum_{i=1}^{k}$ *will tend to infinity under the conditions of Theorem* 1. (*Hence*, *it can be supposed that* Δ *is a great deal larger than* ε, *where* ε *and* Δ *are as introduced in* (3) *and* (4), *respectively*. *Consequently*, F *is well defined*.) *Then*

$$\sum_{i=1}^{k} d^2(\mathbf{y}_i, F) \leqslant k \frac{\varepsilon^2}{(\Delta - \varepsilon)^2} \tag{5}$$

holds with probability tending to 1, *where* d *denotes the Euclidean distance between a vector and a subspace*.

Proof. Let us choose one from the eigenvectors $\mathbf{y}_1, \ldots, \mathbf{y}_k$ of $\mathbf{W}$. Let us denote it simply by $\mathbf{y}$, and the corresponding eigenvalue by λ. We shall estimate the distance between $\mathbf{y}$ and F. For this purpose we expand $\mathbf{y}$ in the basis $\mathbf{u}_1, \ldots, \mathbf{u}_n$ with appropriate real numbers $t_1, \ldots, t_n$:

$$\mathbf{y} = \sum_{i=1}^{n} t_i \mathbf{u}_i .$$

The eigenvalues of the matrix $\mathbf{B}$ corresponding to $\mathbf{u}_1, \ldots, \mathbf{u}_n$ are denoted by $\beta_1, \ldots, \beta_n$, where the k largest eigenvalues $\beta_1, \ldots, \beta_k$ are those defined in the proof of Theorem 1 (we can assume that they are in non-increasing order with the proper reordering of the blocks), and there is a sudden drop following these eigenvalues in the spectrum of $\mathbf{B}$. Then, on the one hand

$$\mathbf{W}\mathbf{y} = (\mathbf{B} + \mathbf{P})\mathbf{y} = \sum_{i=1}^{n} t_i \beta_i \mathbf{u}_i + \mathbf{P}\mathbf{y}, \tag{6}$$

and on the other hand

$$\mathbf{W}\mathbf{y} = \lambda \mathbf{y} = \sum_{i=1}^{n} t_i \lambda \mathbf{u}_i . \tag{7}$$

Equating the right-hand sides of (6) and (7) we get that

$$\sum_{i=1}^{n} t_i (\lambda - \beta_i) \mathbf{u}_i = \mathbf{P}\mathbf{y}.$$

Applying the Pythagorean theorem

$$\sum_{i=1}^{n} t_i^2(\lambda - \beta_i)^2 = \|\mathbf{Py}\|^2 = \mathbf{y}^{\mathrm{T}}\mathbf{P}^{\mathrm{T}}\mathbf{Py} \leqslant \varepsilon^2, \tag{8}$$

as $\|\mathbf{y}\| = 1$ and the largest eigenvalue of $\mathbf{P}^{\mathrm{T}}\mathbf{P}$ is ε^2.

The distance between $\mathbf{y}$ and F is $d^2(\mathbf{y}, F) = \sum_{i=k+1}^{n} t_i^2$. According to Theorem 1 the distance between λ and any β_i $(k+1 \leqslant i \leqslant n)$ can be estimated from below by $\Delta - \varepsilon$, provided $n_1, \ldots, n_k$ are large enough so that $\Delta > \varepsilon$ holds. Therefore,

$$\begin{aligned}(\Delta - \varepsilon)^2 d^2(\mathbf{y}, F) &= (\Delta - \varepsilon)^2 \sum_{i=k+1}^{n} t_i^2 \\ &\leqslant \sum_{i=k+1}^{n} t_i^2(\lambda - \beta_i)^2 \\ &\leqslant \sum_{i=1}^{n} t_i^2(\lambda - \beta_i)^2 \leqslant \varepsilon^2,\end{aligned}$$

where in the last inequality we used (8). From here

$$d^2(\mathbf{y}, F) \leqslant \frac{\varepsilon^2}{(\Delta - \varepsilon)^2} \tag{9}$$

follows.

Applying (9) for the eigenvectors $\mathbf{y}_1, \ldots, \mathbf{y}_k$ of $\mathbf{W}$ and adding the k inequalities together we obtain the upper bound $k\varepsilon^2/(\Delta - \varepsilon)^2$ for the sum of the distances. □

The practical meaning of Theorem 2 is that the eigenvectors of the perturbed matrix are "close" to the corresponding eigenvectors of the original one. In fact, the sum of the squared distances between the eigenvectors corresponding to the k largest eigenvalues of $\mathbf{W}$ and the subspace of the eigenvectors spanned by the k largest eigenvalues of $\mathbf{B}$ can be estimated from above by

$$k\frac{\varepsilon^2}{(\Delta - \varepsilon)^2} = k\frac{1}{(\Delta/\varepsilon - 1)^2} \leqslant k\frac{a}{n}$$

(with some constant a), that tends to zero in order $\mathrm{O}(1/n)$ as $n_1, \ldots, n_k$ and n tends to infinity in the prescribed way, with probability tending to 1. That is, by the definition of ε and Δ the ratio Δ/ε is of order $\sqrt{n}$.

The next theorem is formulated for random graphs and finds a Euclidean representation of the vertices in such a way that the representatives of vertices of the same color are close to each other. First, we introduce some notation.

Let $G = (V, \mathbf{W})$ be a weighted graph with set of vertices V and weight matrix of edges $\mathbf{W}$. Now $\mathbf{W}$ is a random matrix, the same as in Theorem 1, and $\mathbf{y}_1, \ldots, \mathbf{y}_k$ form a set of orthonormal eigenvectors corresponding to the k largest eigenvalues of it (cf. Theorem 2). Further, let $\mathbf{Y}$ denote the $n \times k$ matrix with column vectors

$\mathbf{y}_i$'s and let $\mathbf{x}_1, \ldots, \mathbf{x}_n \in \mathbb{R}^k$ denote the rows of $\mathbf{Y}$. The vectors $\mathbf{x}_i$'s will be called *k-dimensional representatives* of the vertices, briefly, $\mathbf{X} := (\mathbf{x}_1, \ldots, \mathbf{x}_n)$ is *k-dimensional representation* of the random graph. Let c denote the coloring of the vertices induced by the block-matrix $\mathbf{B}$ (see Section 1), where $c(j)$ is the color of vertex j, i.e., $c(j) = i$, if and only if $\sum_{l=1}^{i-1} n_l < j \leqslant \sum_{l=1}^{i} n_l$, $i = 1, \ldots, k$.

The k-variance of the representation $\mathbf{X}$ in the coloring c is defined in [3] as

$$S_k^2(c, \mathbf{X}) = \sum_{i=1}^{k} \sum_{j:c(j)=i} \left\| \mathbf{x}_j - \bar{\mathbf{x}}_i \right\|^2, \tag{10}$$

where $\bar{\mathbf{x}}_i = \sum_{j:c(j)=i} \mathbf{x}_j / n_i$ is the gravity center of the representatives of color (cluster) i.

Theorem 3. *With the above notation and under the conditions of Theorem* 1, *the upper estimate*

$$S_k^2(c, \mathbf{X}) \leqslant k \frac{\varepsilon^2}{(\Delta - \varepsilon)^2}$$

holds with probability tending to 1, *if* $n_1, \ldots, n_k$ *tends to infinity* (*in the same way as in Theorem* 1).

Proof. First of all, it can be easily verified that the eigenvectors of $\mathbf{B}$ corresponding to its k largest eigenvalues are the following

$$u_i(j) = \begin{cases} 1/\sqrt{n_i}, & \text{if } c(j) = i, \\ 0, & \text{otherwise}, \end{cases}$$

where c denotes the coloring induced by $\mathbf{B}$ and $u_i(j)$ denotes the jth coordinate of the eigenvector $\mathbf{u}_i$, $i = 1, \ldots, k$.

It is easy to see the vectors of F—as linear combinations of $\mathbf{u}_1, \ldots, \mathbf{u}_k$—have equal coordinates for vertices with the same color (such vectors are called c-consistent, see [3]). The projection of any $\mathbf{y}$ onto F is a c-consistent vector $\mathbf{u}$, and their distance is

$$\begin{aligned} d^2(\mathbf{y}, F) &= \min_{\mathbf{u}\ c\text{-consistent}} d^2(\mathbf{y}, \mathbf{u}) \\ &= \min_{\mathbf{u}\ c\text{-consistent}} \sum_{i=1}^{k} \sum_{j:c(j)=i} (y(j) - u(i))^2, \end{aligned} \tag{11}$$

where $y(j)$ denotes the jth coordinate of $\mathbf{y}$ and $u(i)$ is the coordinate of $\mathbf{u}$ belonging to the entries of the ith color. Applying Steiner's Theorem we derive that the different coordinates of the optimum c-consistent $\mathbf{u}^*$ are

$$u^*(i) = \frac{\sum_{j:c(j)=i} y(j)}{n_i}, \quad i = 1, \ldots, k. \tag{12}$$

Let us assign such a c-consistent vector $\mathbf{u}_i^*$ to each $\mathbf{y}_i$ that realizes its distance from F. Summing up the distances for $i = 1, \dots, k$, moreover taking into account the relations (5), (11) and (12), and the coordinatewise expansion in the definition (10) of $S_k^2(c, \mathbf{X})$, we get that

$$S_k^2(c, \mathbf{X}) = \sum_{i=1}^{k} \|\mathbf{y}_i - \mathbf{u}_i^*\|^2 = \sum_{i=1}^{k} d^2(\mathbf{y}_i, F) \leqslant k \frac{\varepsilon^2}{(\varDelta - \varepsilon)^2}$$

holds with probability tending to 1, where in the last inequality we used the result of Theorem 2. □

From Theorem 3, we can arrive at the following conclusion: if the edge-weights of a random graph are randomized in such a way that edges connecting vertices of the same color (cluster) have large weights with positive expectation and those connecting different color-clusters have weights of zero expectation, then in this coloring (clustering) the k-variance of the representatives tends to zero, $n \to \infty$. In this situation any metric classification method applied for the representatives will result in the clusters defined by c. In other words, the well classifiable property of the representatives is a necessary condition of the existence of a latent block-structure.

For the time being we focused on the gap between the k largest and the remaining eigenvalues of $\mathbf{W}$. In fact, the gap was due to the gap between the corresponding eigenvalues of $\mathbf{B}$. We learned that the k largest eigenvalues of $\mathbf{B}$ are the numbers $\beta_i = (n_i - 1)\mu_i + \nu_i$ $(i = 1, \dots, k)$, they tend to infinity and because of $n_i \asymp n$, they usually differ from each other in order n.

We shall discuss in details the two extreme *cases*:

(i) all the pairwise distances between the k largest β_i's are of order n,
(ii) $\beta_1 = \cdots = \beta_k$ is a multiple eigenvalue. $(*)$

The intermediate cases when there are some multiple eigenvalues and all the others differ from them and each other in order n can be traced back to these ones. Only the cases when $\beta_i - \beta_j$ is a non-zero constant for some $i \neq j$ are the problematic ones, our statements will not apply to them. To begin with, we need two propositions that are proved similarly to Theorem 2.

Proposition 1. *Let the random matrix* $\mathbf{W}$ *be the same as in Theorem* 1. *Let* $\mathbf{y}_1, \dots, \mathbf{y}_k$ *and* $\mathbf{u}_1, \dots, \mathbf{u}_k$ *be sets of orthonormal eigenvectors corresponding to the k largest eigenvalues* (*in decreasing order*) *of* $\mathbf{W}$ *and* $\mathbf{B}$, *respectively*, *latter ones being* $\beta_1, \dots, \beta_k$. *Suppose that* β_i *is an isolated eigenvalue*, *that is*

$$|\beta_i - \beta_j| \geqslant \delta_i > \varepsilon \quad \text{for all } j \neq i.$$

Then

$$\|\mathbf{y}_i - t_{ii}\mathbf{u}_i\|^2 \leqslant \frac{\varepsilon^2}{(\delta_i - \varepsilon)^2}$$

holds with probability tending to 1, *where* $t_{ii} = \mathbf{y}_i^{\mathrm{T}}\mathbf{u}_i$.

Proof. As in the proof of Theorem 2, we expand $\mathbf{y}_i$ in the basis $\mathbf{u}_1, \dots, \mathbf{u}_n$ with appropriate real numbers $t_{i1}, \dots, t_{in}$:

$$\mathbf{y}_i = \sum_{i=1}^{n} t_{ij}\mathbf{u}_j.$$

Comparing the right-hand sides of equations

$$\mathbf{W}\mathbf{y}_i = (\mathbf{B}+\mathbf{P})\mathbf{y}_i = \sum_{j=1}^{n} t_{ij}\beta_j\mathbf{u}_j + \mathbf{P}\mathbf{y}_i$$

and

$$\mathbf{W}\mathbf{y}_i = \lambda_i\mathbf{y}_i = \sum_{j=1}^{n} t_{ij}\lambda_i\mathbf{u}_j,$$

we get that

$$\sum_{j=1}^{n} t_{ij}(\lambda_i - \beta_j)\mathbf{u}_j = \mathbf{P}\mathbf{y}_i.$$

By the Pythagorean theorem

$$t_{ii}^2(\lambda_i-\beta_i)^2 + \sum_{j\neq i} t_{ij}^2(\lambda_i-\beta_j)^2 = \|\mathbf{P}\mathbf{y}_i\|^2 = \mathbf{y}_i^{\mathrm{T}}\mathbf{P}^{\mathrm{T}}\mathbf{P}\mathbf{y}_i \leqslant \varepsilon^2.$$

Since $\|\mathbf{y}_i - t_{ii}\mathbf{u}_i\|^2 = \sum_{j\neq i}^{n} t_{ij}^2$ and $|\lambda_i - \beta_j| \geqslant \delta_i - \varepsilon$, $(j \neq i)$,

$$(\delta_i-\varepsilon)^2\|\mathbf{y}_i - t_{ii}\mathbf{u}_i\|^2 \leqslant \sum_{j\neq i} t_{ij}^2(\lambda_i-\beta_j)^2 \leqslant \sum_{j=1}^{n} t_{ij}^2(\lambda_i-\beta_j)^2 \leqslant \varepsilon^2,$$

which implies our statement. □

In the Case (i) discussed above the following statement applies to β_i with δ_i of order n, $i = 1, \dots, k$.

Corollary 1. *If δ_i is of order n then $1 - t_{ii} = \mathrm{O}(1/n)$, $|t_{ij}| = \mathrm{O}(1/\sqrt{n})$, $j \neq i$, and $\|\mathbf{y}_i - \mathbf{u}_i\|^2 = \mathrm{O}(1/n)$, with probability tending to 1.*

Proof. By Proposition 1, in case of $\delta_i \asymp n$

$$\|\mathbf{y}_i - t_{ii}\mathbf{u}_i\|^2 \leqslant \frac{\varepsilon^2}{(\delta_i-\varepsilon)^2} = \frac{1}{(\delta_i/\varepsilon - 1)^2} \leqslant a_i/n$$

holds with an appropriate constant a_i. Hence,

$$\|\mathbf{y}_i - t_{ii}\mathbf{u}_i\|^2 = \|\mathbf{y}_i\|^2 - 2t_{ii}\mathbf{y}_i^{\mathrm{T}}\mathbf{u}_i + t_{ii}^2\|\mathbf{u}_i\|^2 = 1 - t_{ii}^2 \leqslant a_i/n,$$

with probability tending to 1. This implies that $t_{ii}^2 \geqslant 1 - a_i/n$, or equivalently, as $t_{ii} > 0$, $t_{ii} \geqslant \sqrt{1 - a_i/n}$. Via Taylor's expansion it follows that $1 - t_{ii} \leqslant a_i/2n$.

To prove the second part of the statement, remember that by Theorem 2 the relation $d^2(\mathbf{y}_i, F) = \mathrm{O}(1/n)$ holds. Then

$$\mathbf{y}_i = \sum_{j=1}^{k} t_{ij}\mathbf{u}_j + \mathbf{r}_i, \tag{13}$$

where the first term is in F, while the last one is orthogonal to F, therefore $\|\mathbf{r}_i\| = \mathrm{O}(1/\sqrt{n})$.

Denote the lth component of $\mathbf{y}_i$ by $y_i(l)$. Then, on the one hand, since $\|\mathbf{y}_i - t_{ii}\mathbf{u}_i\|^2 \leqslant a_i/n$, we have that

$$\sum_{l:c(l)\neq i} y_i^2(l) + \sum_{l:c(l)=i} \left(y_i(l) - \frac{t_{ii}}{\sqrt{n_i}}\right)^2 \leqslant \frac{a_i}{n},$$

consequently,

$$\left[\sum_{l:c(l)=j} y_i(l)\right]^2 \leqslant n_j \sum_{l:c(l)=j} y_i^2(l) \leqslant n_j \frac{a_i}{n}, \quad j \neq i.$$

On the other hand, with the help of (13) we have that

$$t_{ij} = \mathbf{y}_i^{\mathrm{T}}\mathbf{u}_j = \frac{\sum_{l:c(l)=j} y_i(l)}{\sqrt{n_j}}, \quad j = 1, \ldots, k.$$

Hence

$$|t_{ij}| = \frac{|\sum_{l:c(l)=j} y_i(l)|}{\sqrt{n_j}} \leqslant \frac{\sqrt{n_j a_i}/\sqrt{n}}{\sqrt{n_j}} = \frac{\sqrt{a_i}}{\sqrt{n}}, \quad j \neq i,$$

that is $|t_{ij}| = \mathrm{O}(1/\sqrt{n})$.

Eventually,

$$\|\mathbf{y}_i - \mathbf{u}_i\|^2 = \|\mathbf{y}_i\|^2 + \|\mathbf{u}_i\|^2 - 2\mathbf{y}_i^{\mathrm{T}}\mathbf{u}_i = 2(1 - t_{ii}) = \mathrm{O}(1/n)$$

that finishes the proof. □

Proposition 2. *Let the random matrix* $\mathbf{W}$ *be the same as in Theorem* 1. *Let* $\mathbf{y}_1, \ldots, \mathbf{y}_k$ *and* $\mathbf{u}_1, \ldots, \mathbf{u}_k$ *be sets of orthonormal eigenvectors corresponding to the k largest eigenvalues* (*in descending order*) *of* $\mathbf{W}$ *and* $\mathbf{B}$, *respectively. Then a set* $\mathbf{v}_1, \ldots, \mathbf{v}_k$ *of orthonormal vectors within F can be found such that*

$$\sum_{i=1}^{k} \|\mathbf{y}_i - \mathbf{v}_i\|^2 \leqslant 2k\frac{\varepsilon^2}{(\Delta - \varepsilon)^2}$$

holds with probability tending to 1.

Proof. Let us denote the $n \times k$ matrices with column vectors $\mathbf{y}_1, \ldots, \mathbf{y}_k$ and $\mathbf{u}_1, \ldots, \mathbf{u}_k$ by $\mathbf{U}$ and $\mathbf{Y}$, respectively. The desired vectors $\mathbf{v}_1, \ldots, \mathbf{v}_k$ are also put in an $n \times k$ matrix $\mathbf{V}$, and hence, we are looking for $\mathbf{V}$ in the form $\mathbf{V} = \mathbf{UR}$ with an appropriate $k \times k$ orthogonal matrix $\mathbf{R}$. More precisely, we shall find $\mathbf{R}$ such that, with it

$$\sum_{i=1}^{k} \|\mathbf{y}_i - \mathbf{v}_i\|^2 = \operatorname{tr}(\mathbf{Y} - \mathbf{UR})^{\mathrm{T}}(\mathbf{Y} - \mathbf{UR}) \leqslant 2\sum_{i=1}^{k} d^2(\mathbf{y}_i, F) \tag{14}$$

holds. By the additive and cyclic property of the trace operator,

$$\begin{aligned} \operatorname{tr}(\mathbf{Y} - \mathbf{UR})^{\mathrm{T}}(\mathbf{Y} - \mathbf{UR}) &= \operatorname{tr}\mathbf{Y}^{\mathrm{T}}\mathbf{Y} + \operatorname{tr}\mathbf{R}^{\mathrm{T}}\mathbf{U}^{\mathrm{T}}\mathbf{UR} - 2\operatorname{tr}\mathbf{Y}^{\mathrm{T}}\mathbf{UR} \\ &= \operatorname{tr}\mathbf{Y}^{\mathrm{T}}\mathbf{Y} + \operatorname{tr}(\mathbf{U}^{\mathrm{T}}\mathbf{U})(\mathbf{RR}^{\mathrm{T}}) - 2\operatorname{tr}\mathbf{Y}^{\mathrm{T}}\mathbf{UR} \\ &= 2(k - \operatorname{tr}\mathbf{Y}^{\mathrm{T}}\mathbf{UR}) \end{aligned} \tag{15}$$

is obtained, where we used that $\mathbf{Y}^{\mathrm{T}}\mathbf{Y} = \mathbf{U}^{\mathrm{T}}\mathbf{U} = \mathbf{RR}^{\mathrm{T}} = \mathbf{I}_k$. The last quantity in (15) is minimum if $\operatorname{tr}\mathbf{Y}^{\mathrm{T}}\mathbf{UR}$ is maximum as a function of $\mathbf{R}$. But we can apply a lemma (see [2]), according to which $\operatorname{tr}(\mathbf{Y}^{\mathrm{T}}\mathbf{U})\mathbf{R}$ is maximum if $(\mathbf{Y}^{\mathrm{T}}\mathbf{U})\mathbf{R}$ is symmetric, and the maximum is $\sum_{i=1}^{k} s_i$, with s_i's being the singular values of $\mathbf{Y}^{\mathrm{T}}\mathbf{U}$. So, the minimum that can be attained in (15) is equal to

$$2\sum_{i=1}^{k}(1 - s_i). \tag{16}$$

Eventually, the sum of the above distances in (14) can also be written in terms of the singular values $s_1, \ldots, s_k$. As $\mathbf{UU}^{\mathrm{T}}$ is the matrix of orthogonal projection onto F,

$$\begin{aligned} \sum_{i=1}^{k} d^2(\mathbf{y}_i, F) &= \operatorname{tr}(\mathbf{Y} - \mathbf{UU}^{\mathrm{T}}\mathbf{Y})^{\mathrm{T}}(\mathbf{Y} - \mathbf{UU}^{\mathrm{T}}\mathbf{Y}) \\ &= \operatorname{tr}\mathbf{Y}^{\mathrm{T}}\mathbf{Y} - \operatorname{tr}\mathbf{Y}^{\mathrm{T}}\mathbf{UU}^{\mathrm{T}}\mathbf{Y} \\ &= k - \sum_{i=1}^{k} s_i^2 = \sum_{i=1}^{k}(1 - s_i^2). \end{aligned} \tag{17}$$

Comparing (16) and (17), it remains to show that $\sum_{i=1}^{k}(1 - s_i) \leqslant \sum_{i=1}^{k}(1 - s_i^2)$. But s_i's are the singular values of the matrix $\mathbf{U}^{\mathrm{T}}\mathbf{Y}$, therefore denoting by $s_{\max}(\cdot)$ the maximum singular value of the matrix in the argument, we have

$$s_i \leqslant s_{\max}(\mathbf{U}^{\mathrm{T}}\mathbf{Y}) \leqslant s_{\max}(\mathbf{U}) \cdot s_{\max}(\mathbf{Y}) = 1,$$

as all singular values of the matrices $\mathbf{U}$ and $\mathbf{Y}$ are equal to 1. Hence, $s_i \geqslant s_i^2$ implies the desired relation (14). □

This proposition will be applied in the Case (ii). If the largest eigenvalue of $\mathbf{B}$ has multiplicity k, then F is the eigenspace belonging to it. So, $\mathbf{v}_1, \ldots, \mathbf{v}_k$ form a set of orthonormal eigenvectors belonging to this multiple eigenvalue, and the statement of Theorem 2 applies immediately to the paired distances between the eigenvectors of the perturbed matrix and the original one. We shall summarize our conclusions.

Corollary 2. *If $\beta_1 = \cdots = \beta_k$ is a multiple eigenvalue then with the vectors $\mathbf{v}_1, \ldots, \mathbf{v}_k$ of Proposition 2 the relations $\|\mathbf{y}_i - \mathbf{v}_i\|^2 = \mathrm{O}(1/n)$ hold, $i = 1, \ldots, k$. Further, for the coefficients of the decomposition*

$$\mathbf{y}_i = z_{ii}\mathbf{v}_i + \sum_{j \neq i} z_{ij}\mathbf{v}_j + \mathbf{r}_i$$

it holds true that $1 - z_{ii} = \mathrm{O}(1/n)$ and $|z_{ij}| = \mathrm{O}(1/\sqrt{n})$, $j \neq i$, $i = 1, \ldots, k$.

Proof. By Proposition 2 the relation $\sum_{i=1}^{k} \|\mathbf{y}_i - \mathbf{v}_i\|^2 = \mathrm{O}(1/n)$ follows, and it is also true for the individual terms. The decomposition $\mathbf{y}_i = z_{ii}\mathbf{v}_i + \sum_{j \neq i} z_{ij}\mathbf{v}_j + \mathbf{r}_i$ holds true with the same $\mathbf{r}_i$ as in Corollary 1, because the $\mathbf{v}_i$'s span the same subspace as the $\mathbf{u}_i$'s, hence $\|\mathbf{r}_i\|^2 = \mathrm{O}(1/n)$. As

$$\|\mathbf{y}_i - \mathbf{v}_i\|^2 = 2(1 - z_{ii}) = \mathrm{O}(1/n),$$

$1 - z_{ii} = \mathrm{O}(1/n)$ follows.

By the Pythagorean theorem we have

$$1 = z_{ii}^2 + \sum_{j \neq i} z_{ij}^2 + \|\mathbf{r}_i\|^2,$$

that implies $z_{ij}^2 \leqslant 1 - z_{ii}^2 - \|\mathbf{r}_i\|^2 = \mathrm{O}(1/n)$, consequently $|z_{ij}| = \mathrm{O}(1/\sqrt{n})$. This finishes the proof. □

4. Fine characterization of the spectrum

Theorem 4. *With the conditions of Theorem 1, in the Cases (i) and (ii) (cf. (∗)) the asymptotic distribution of the eigenvalue λ_i of $\mathbf{W}$ for any $0 < \delta < 1/2$ can be approximated in order $1/n^{1/2-\delta}$ by a distribution that differs at most in $\mathrm{O}(1)$ from a normal distribution with mean $(n_i - 1)\mu_i + \nu_i + \frac{\sigma^2}{\mu_i}$ and variance $2\sigma^2$ $(i = 1, \ldots, k)$, with probability tending to 1.*

Further, given the t_{ij}'s $(i = 1, \ldots, k;\ j = 1, \ldots, k)$ defined in (13), the distribution of the k largest eigenvalues of $\mathbf{W}$ for any $0 < \delta < 1/2$ can be approximated in order $1/n^{1/2-\delta}$ by a k-variate normal distribution with covariance matrix $2\sigma^2\mathbf{I}_k$ and mean vector $\mathbf{m}$ with ith component

$$m_i = \sum_{j=1}^{k} t_{ij}^2[(n_j - 1)\mu_j + \nu_j + \sigma^2/\mu_j] \quad (i = 1, \ldots, k)$$

with probability tending to 1.

We need some lemmas.

Lemma 1. *Let $X_1, \dots, X_k \sim \mathcal{N}(0, \sigma^2)$ be i.i.d. random variables and the random variables $Z_1, \dots, Z_k$ be uniformly bounded by $n^{-\tau}$. The real number $\tau > 0$ and the positive integer k are fixed. Then*

$$P\left(\left|\sum_{i=1}^{k} Z_i X_i\right| > \frac{1}{n^{\tau-\delta}}\right) \leqslant \sqrt{\frac{2}{\pi}}\frac{\sigma\sqrt{k}}{n^{\delta}}\,\mathrm{e}^{-\frac{n^{2\delta}}{2k\sigma^2}} \quad \textit{for } 0 < \delta < \tau.$$

Proof. With the notation $\mathbf{Z} = (Z_1, \dots, Z_k)$, the conditional distribution of $\sum_{i=1}^{k} Z_i X_i$ conditioned on $\mathbf{Z} = \mathbf{z}$ is $\mathcal{N}(0, \sigma^2 \sum_{i=1}^{k} z_i^2)$ due to the independence of X_i's. Then for the absolute value of the standardized $\sum_{i=1}^{k} Z_i X_i$ the upper bound

$$\begin{aligned}
P\left(\left|\sum_{i=1}^{k} Z_i X_i\right| > \frac{1}{n^{\tau-\delta}} \,\middle|\, \mathbf{Z} = \mathbf{z}\right) &= P\left(\frac{|\sum_{i=1}^{k} z_i X_i|}{\sigma\sqrt{\sum_{i=1}^{k} z_i^2}} > \frac{1}{n^{\tau-\delta}\sigma\sqrt{\sum_{i=1}^{k} z_i^2}}\right)\\
&\leqslant P\left(\frac{|\sum_{i=1}^{k} z_i X_i|}{\sigma\sqrt{\sum_{i=1}^{k} z_i^2}} > \frac{n^{\delta}}{\sigma\sqrt{k}}\right)\\
&\leqslant \sqrt{\frac{2}{\pi}}\frac{\sigma\sqrt{k}}{n^{\delta}}\,\mathrm{e}^{-\frac{n^{2\delta}}{2k\sigma^2}} =: p_n \quad \text{for } 0 < \delta < \tau
\end{aligned}$$

is obtained, where we used a result of [8] for estimating the following probability: if ξ is a standard normal variable then we have

$$P(|\xi| > u) = 2(1 - \Phi(u)) \leqslant 2\frac{1}{\sqrt{2\pi}}\frac{1}{u}\,\mathrm{e}^{-u^2/2} = \sqrt{\frac{2}{\pi}}\frac{1}{u}\,\mathrm{e}^{-u^2/2}, \tag{18}$$

where Φ is the standard normal (Gauss) distribution function. By this, for the unconditional probability

$$\begin{aligned}
P\left(\left|\sum_{i=1}^{k} Z_i X_i\right| > \frac{1}{n^{\tau-\delta}}\right) &= \int_{\mathbb{R}^k} P\left(\left|\sum_{i=1}^{k} Z_i X_i\right| > \frac{1}{n^{\tau-\delta}} \,\middle|\, \mathbf{Z} = \mathbf{z}\right) h(\mathbf{z})\mathrm{d}\mathbf{z}\\
&\leqslant p_n \int_{\mathbb{R}^k} h(\mathbf{z})\mathrm{d}\mathbf{z} = p_n
\end{aligned}$$

holds, whatever the distribution of $\mathbf{Z}$ was. (We suppose that the distribution of $\mathbf{Z}$ is absolute continuous with respect to the Lebesgue measure, with density function $h(\mathbf{z})$.) □

Corollary 3. *Since $\lim_{n\to\infty} p_n = 0$, it follows that for any $\tau > 0$ the random variable $\sum_{i=1}^{k} Z_i X_i$ tends to zero in probability (i.e., with probability tending to 1) as $n \to \infty$.*

Proof. Let $\alpha > 0$ be an arbitrary "small" positive constant and δ be such that $\alpha = 1/n^{\tau-\delta}$. With this, the relation

$$P\left(\left|\sum_{i=1}^{k} Z_i X_i\right| > \alpha\right) \leqslant \sqrt{\frac{2}{\pi}}\frac{\sigma\sqrt{k}}{\alpha n^{\tau}}\,\mathrm{e}^{-\frac{\alpha^2 n^{2\tau}}{2k\sigma^2}}$$

holds, and the right-hand side tends to zero with $n \to \infty$. □

Remark 1. Though we do not use the fact, we note that the statement of Corollary 3 can be sharpened. As $\lim_{n\to\infty} p_n = 0$ in exponential order, $\sum_{n=1}^{\infty} p_n < \infty$, hence by the Borel–Cantelli lemma $\sum_{i=1}^{k} Z_i X_i$ almost surely converges to zero.

Lemma 2. *Let $X_1, \ldots, X_n \sim \mathcal{N}(0, \sigma^2)$ be i.i.d. random variables and the random vector $\mathbf{Z} = (Z_1, \ldots, Z_n)$ be such that $\|\mathbf{Z}\| = \sqrt{\sum_{j=1}^{n} Z_j^2} \leqslant n^{-\tau}$ with probability tending to* 1, $\tau > 0$. *Then*

$$P\left(\left|\sum_{j=1}^{n} Z_j X_j\right| > \frac{1}{n^{\tau-\delta}}\right) \leqslant \sqrt{\frac{2}{\pi}}\frac{\sigma}{n^{\delta}}\,\mathrm{e}^{-\frac{n^{2\delta}}{2\sigma^2}} \quad \textit{for } 0 < \delta < \tau.$$

Proof. The proof is similar to that of Lemma 1. With the same notations $\sum_{j=1}^{n} Z_j X_j$ conditioned on $\mathbf{Z} = \mathbf{z}$ is an $\mathcal{N}(0, \sigma^2 \sum_{j=1}^{n} z_j^2)$ variable. For the absolute value of the standardized $\sum_{j=1}^{n} Z_j X_j$ inequality (18) gives that

$$\begin{aligned}
P\left(\left|\sum_{j=1}^{n} Z_j X_j\right| > \frac{1}{n^{\tau-\delta}} \,\middle|\, \mathbf{Z} = \mathbf{z}\right) &= P\left(\frac{|\sum_{j=1}^{n} z_j X_j|}{\sigma\sqrt{\sum_{j=1}^{n} z_j^2}} > \frac{1}{n^{\tau-\delta}\sigma\sqrt{\sum_{j=1}^{n} z_j^2}}\right) \\
&\leqslant P\left(\frac{|\sum_{j=1}^{n} z_j X_j|}{\sigma\sqrt{\sum_{j=1}^{n} z_j^2}} > \frac{n^{\delta}}{\sigma}\right) \\
&\leqslant \sqrt{\frac{2}{\pi}}\frac{\sigma}{n^{\delta}}\,\mathrm{e}^{-\frac{n^{2\delta}}{2\sigma^2}} =: q_n \quad \text{for } 0 < \delta < \tau.
\end{aligned}$$

For the unconditional probability

$$\begin{aligned}
P\left(\left|\sum_{j=1}^{n} Z_j X_j\right| > \frac{1}{n^{\tau-\delta}}\right) &= \int_{\mathbb{R}^k} P\left(\left|\sum_{j=1}^{n} Z_j X_j\right| > \frac{1}{n^{\tau-\delta}} \,\middle|\, \mathbf{Z} = \mathbf{z}\right) h(\mathbf{z})\mathrm{d}\mathbf{z} \\
&\leqslant q_n \int_{\mathbb{R}^k} h(\mathbf{z})\,\mathrm{d}\mathbf{z} = q_n
\end{aligned}$$

holds, whatever the distribution of $\mathbf{Z}$ was. □

Corollary 4. *Since* $\lim_{n\to\infty} q_n = 0$, *with the argument of Corollary* 3, *it also follows that for any $\tau > 0$ the random variable $\sum_{j=1}^{n} Z_j X_j$ tends to zero in probability as $n \to \infty$.* □

Remark 2. We note that the statement of Corollary 4 can be sharpened. As $\lim_{n\to\infty} q_n = 0$ in exponential order, $\sum_{n=1}^{\infty} q_n < \infty$, hence by the Borel–Cantelli lemma $\sum_{j=1}^{n} Z_j X_j$ almost surely converges to zero.

Lemma 3. *Let the distribution of* ξ_n *be approximated in order* $1/\sqrt{n}$ *by an* $\mathcal{N}(0, \sigma^2)$ *distribution and* ζ_n *be a nonnegative random variable with* $1 - \zeta_n = \mathrm{O}(1/n)$. *Then* $\zeta_n \xi_n$ *asymptotically also follows an* $\mathcal{N}(0, \sigma^2)$ *distribution, in the same order.*

Proof. Let $\xi_n = X_n + \eta_n/\sqrt{n}$, where $X_n \sim \mathcal{N}(0, \sigma^2)$, $|\eta_n| < a$ with some constant a, and $z_n = 1 - \zeta_n$. Then

$$\zeta_n \xi_n = (1 - z_n)\xi_n = \xi_n - z_n(X_n + \eta_n/\sqrt{n}) = \xi_n - z_n X_n - z_n \eta_n/\sqrt{n}.$$

Here the first term is asymptotically normally distributed (in order $1/\sqrt{n}$), the second one differs from zero in order $1/n^{1-\delta}$, $0 < \delta < 1$ by Lemma 1, while the last one is $\mathrm{O}(n^{-3/2})$, with probability tending to 1. This finishes the proof. □

Now, we are ready to present the

Proof of Theorem 4. Observe, that the diagonal blocks of $\mathbf{W}$ satisfy the conditions of Theorem FK1. More precisely, let us decompose the perturbation $\mathbf{P}$ into the perturbation on the diagonal blocks and that on the non-diagonal part: $\mathbf{P} = \mathbf{P}_1 + \mathbf{P}_2$. The k largest eigenvalues of $\mathbf{W}_1 = \mathbf{B} + \mathbf{P}_1$ are—due to [5]—as follows: the largest eigenvalue of the ith block can be approximated in order $1/\sqrt{n_i}$ (or equivalently, in order $1/\sqrt{n}$) by a normal distribution of expectation $(n_i - 1)\mu_i + \nu_i + \sigma^2/\mu_i$ and variance $2\sigma^2$, $i = 1, \ldots, k$, and they are independent of each other (because they depend on independent entries), consequently, they approximately have a k-variate normal distribution with independent components. After putting the random perturbation $\mathbf{P}_2$ on $\mathbf{W}_1$, with computer simulation we still experienced the same shape of covariance ellipsoids (they were spheres, with a practically unchanged variance and mildly shifted mean).

For a precise proof, first we notice that

$$\mathbf{u}_i^{\mathrm{T}} \mathbf{W} \mathbf{u}_i = \mathbf{u}_i^{\mathrm{T}} \mathbf{W}_1 \mathbf{u}_i \quad (i = 1, \ldots, k),$$

hence by Theorem FK1, these numbers approximate the eigenvalues of $\mathbf{W}_1$ in order $1/\sqrt{n_i}$ and they are nearly normally distributed with mean $(n_i - 1)\mu_i + \nu_i + \sigma^2/\mu_i$ and variance $2\sigma^2$, with probability tending to 1. We shall show that in the Cases (i) and (ii) the same numbers give also a good approximation for the eigenvalues $\lambda_1, \ldots, \lambda_k$ of $\mathbf{W}$.

Now, let us investigate λ_i. By the symmetry of $\mathbf{W}$ and also using (13) we have that

$$\lambda_i = \mathbf{y}_i^{\mathrm{T}} \mathbf{W} \mathbf{y}_i = \left(t_{ii}\mathbf{u}_i + \sum_{\substack{j=1 \\ j \neq i}}^{k} t_{ij}\mathbf{u}_j + \mathbf{r}_i \right)^{\mathrm{T}} \mathbf{W} \left(t_{ii}\mathbf{u}_i + \sum_{\substack{j=1 \\ j \neq i}}^{k} t_{ij}\mathbf{u}_j + \mathbf{r}_i \right),$$

hence

$$\lambda_i = t_{ii}^2 \mathbf{u}_i^{\mathrm{T}} \mathbf{W} \mathbf{u}_i + \mathbf{r}_i^{\mathrm{T}} \mathbf{W} \mathbf{r}_i + \sum_{\substack{j=1 \\ j \neq i}}^{k} \sum_{\substack{l=1 \\ l \neq i}}^{k} t_{ij} t_{il} \mathbf{u}_j^{\mathrm{T}} \mathbf{W} \mathbf{u}_l$$
$$+ 2 t_{ii} \sum_{\substack{j=1 \\ j \neq i}}^{k} t_{ij} \mathbf{u}_i^{\mathrm{T}} \mathbf{W} \mathbf{u}_j + 2 t_{ii} \mathbf{u}_i^{\mathrm{T}} \mathbf{W} \mathbf{r}_i + 2 \sum_{\substack{j=1 \\ j \neq i}}^{k} t_{ij} \mathbf{r}_i^{\mathrm{T}} \mathbf{W} \mathbf{u}_j. \tag{19}$$

We shall estimate the above terms one by one in the Case (i).

For the second term in (19)

$$\mathbf{r}_i^{\mathrm{T}} \mathbf{W} \mathbf{r}_i \leqslant \|\mathbf{r}_i\| \|\mathbf{W} \mathbf{r}_i\| \leqslant \|\mathbf{r}_i\| (\|\mathbf{B} \mathbf{r}_i\| + \|\mathbf{P}\| \|\mathbf{r}_i\|) \leqslant \frac{\sqrt{a_i}}{\sqrt{n}} 2\varepsilon \frac{\sqrt{a_i}}{\sqrt{n}} = \mathrm{O}\left(\frac{1}{\sqrt{n}}\right)$$

holds with constant a_i, as $\|\mathbf{P}\| = \varepsilon \sim 2\sigma\sqrt{n}$, $\|\mathbf{r}_i\| = \mathrm{O}(1/\sqrt{n})$, and we apply the Courant–Fischer–Weyl minimax principle for $\|\mathbf{B}\mathbf{r}_i\| \leqslant \varepsilon \|\mathbf{r}_i\|$ using that $\mathbf{r}_i$ is orthogonal to F spanned by the eigenvectors belonging to the k largest eigenvalues of the matrix $\mathbf{B}$.

For the third term in (19)

$$\sum_{\substack{j=1 \\ j \neq i}}^{k} \sum_{\substack{l=1 \\ l \neq i}}^{k} t_{ij} t_{il} \mathbf{u}_j^{\mathrm{T}} \mathbf{W} \mathbf{u}_l = 2 \sum_{\substack{j,l=1 \\ j<l \\ j \neq i \\ l \neq i}}^{k} t_{ij} t_{il} \mathbf{u}_j^{\mathrm{T}} \mathbf{P} \mathbf{u}_l + \sum_{\substack{j=1 \\ j \neq i}}^{k} t_{ij}^2 \mathbf{u}_j^{\mathrm{T}} \mathbf{W} \mathbf{u}_j$$
$$= 2 \sum_{\substack{j,l=1 \\ j<l \\ j \neq i \\ l \neq i}}^{k} t_{ij} t_{il} \left[\frac{1}{\sqrt{n_j n_l}} \sum_{c(x)=j} \sum_{c(y)=l} p_{xy} \right]$$
$$+ \sum_{\substack{j=1 \\ j \neq i}}^{k} t_{ij}^2 \mathbf{u}_j^{\mathrm{T}} \mathbf{W} \mathbf{u}_j, \tag{20}$$

as $\mathbf{u}_j^{\mathrm{T}} \mathbf{W} \mathbf{u}_l = \mathbf{u}_j^{\mathrm{T}} \mathbf{P} \mathbf{u}_l$ holds for $j \neq l$ because of $\mathbf{u}_j^{\mathrm{T}} \mathbf{B} \mathbf{u}_l = \beta_l \mathbf{u}_j^{\mathrm{T}} \mathbf{u}_l = 0$. Here the random variables in brackets follow an $\mathscr{N}(0, \sigma^2)$ distribution due to the central limit theorem as $n_i, n_j \to \infty$, moreover, they are independent of each other for indices $j < l$. (That is, we utilized that the entries p_{xy}'s above the main diagonal of $\mathbf{P}$ are independent.) Since the random variables $|t_{ij}|$ and $|t_{il}|$ are $\mathrm{O}(1/\sqrt{n})$, Lemma 1 applies with $\tau = 1$ to the first term, consequently, it is of order $1/n^{1-\delta}$ with any "small" $\delta > 0$. For the second term we use the argument of [5] that the distribution of $\mathbf{u}_j^{\mathrm{T}} \mathbf{W} \mathbf{u}_j$ is approximated (in order $1/\sqrt{n_j}$) by an $\mathscr{N}((n_j - 1)\mu_j + \nu_j + \sigma^2/\mu_j, 2\sigma^2)$ variable that can be written as

$$\xi_j + (n_j - 1)\mu_j + \nu_j + \sigma^2/\mu_j,$$

where ξ_j has an $\mathscr{N}(0, 2\sigma^2)$ distribution. Then the last sum in (20) can be written as

$$\sum_{\substack{j=1\\ j\neq i}}^{k} t_{ij}^2 \xi_j + \sum_{\substack{j=1\\ j\neq i}}^{k} t_{ij}^2 \left[(n_j - 1)\mu_j + \nu_j + \sigma^2/\mu_j \right].$$

To the first term above Lemma 1 applies with $\tau = 1$, while the second term above is of O(1), since $t_{ij}^2 = \mathrm{O}(1/n)$ for $j \neq i$.

Similarly, for the half of the fourth term in (19)

$$t_{ii} \sum_{\substack{j=1\\ j\neq i}}^{k} t_{ij} \mathbf{u}_i^{\mathrm{T}} \mathbf{W} \mathbf{u}_j = t_{ii} \sum_{\substack{j=1\\ j\neq i}}^{k} t_{ij} \mathbf{u}_i^{\mathrm{T}} \mathbf{P} \mathbf{u}_j = \sum_{\substack{j=1\\ j\neq i}}^{k} t_{ij} \left[\frac{1}{\sqrt{n_i n_j}} \sum_{c(x)=i} \sum_{c(y)=j} p_{xy} \right]$$
$$- (1 - t_{ii}) \sum_{\substack{j=1\\ j\neq i}}^{k} t_{ij} \left[\frac{1}{\sqrt{n_i n_j}} \sum_{c(x)=i} \sum_{c(y)=j} p_{xy} \right] \tag{21}$$

holds. Here, in brackets we have $\mathscr{N}(0, \sigma^2)$ variables and for the sums Lemma 1 applies with $\tau = 1/2$ because of $|t_{ij}| = \mathrm{O}(1/\sqrt{n})$. The multiplication by $1 - t_{ii} = \mathrm{O}(1/n)$ makes the order of the second big term $1/n^{3/2-\delta}$, so the difference is of order $1/n^{1/2-\delta}$.

For the half of the fifth term in (19) we have that

$$t_{ii} \mathbf{u}_i^{\mathrm{T}} \mathbf{W} \mathbf{r}_i = t_{ii} \mathbf{u}_i^{\mathrm{T}} \mathbf{P} \mathbf{r}_i = \sum_{j=1}^{n} r_{ij} \left[\frac{1}{\sqrt{n_i}} \sum_{c(x)=i} p_{xj} \right]$$
$$- (1 - t_{ii}) \sum_{j=1}^{n} r_{ij} \left[\frac{1}{\sqrt{n_i}} \sum_{c(x)=i} p_{xj} \right], \tag{22}$$

where r_{ij} denotes the jth coordinate of the vector $\mathbf{r}_i$. Here we used the fact that $\mathbf{u}_i^{\mathrm{T}} \mathbf{W} \mathbf{r}_i = \mathbf{r}_i^{\mathrm{T}} \mathbf{B} \mathbf{u}_i + \mathbf{u}_i^{\mathrm{T}} \mathbf{P} \mathbf{r}_i = \beta_i \mathbf{r}_i^{\mathrm{T}} \mathbf{u}_i + \mathbf{u}_i^{\mathrm{T}} \mathbf{P} \mathbf{r}_i = \mathbf{u}_i^{\mathrm{T}} \mathbf{P} \mathbf{r}_i$, as $\mathbf{u}_i$ is an eigenvector of $\mathbf{B}$ and $\mathbf{r}_i$ is orthogonal to it. The random variables in brackets are denoted by X_j's and they are $\mathscr{N}(0, \sigma^2)$ variables (by the central limit theorem). For js with $c(j) \neq i$ (out of block i) they are independent (they have no common terms in the sums $\sum_{c(x)=i} p_{xj}$), but for indices $j \neq l$ in the case of $c(j) = c(l) = i$ (within the ith diagonal block) there is exactly one common term ($p_{lj} = p_{jl}$) in the corresponding sums. Lemma 2 is not immediately applicable here, as the terms in brackets are not fully independent. However, with a slight modification, we can use the lemma. The idea is that we substitute the dependent terms with a fewer number of independent ones, that are not identically distributed, still their variances can be estimated as follows.

With the notation of Lemma 2, let Z_j be r_{ij} and $X_1, \ldots, X_n$ be the random variables in brackets. Then $\|\mathbf{Z}\| \leqslant n^{-1/2}$ holds and X_i's are either independent or each pair in the ith block has a common term in their defining sums. So their common distribution is multivariate normal and their covariances are

$$\mathrm{Cov}(X_j, X_l) = \begin{cases} \mathrm{Cov}\Big(\frac{1}{\sqrt{n_i}} p_{lj}, \frac{1}{\sqrt{n_i}} p_{jl}\Big) = \frac{1}{n_i}\sigma^2, & \text{if } c(j) = c(l) = i, \quad j \neq l, \\ 0, & \text{otherwise.} \end{cases}$$

Hence, for fixed $\mathbf{Z} = \mathbf{z}$:

$$\begin{aligned}
\mathrm{Var}\left(\sum_{j=1}^{n} z_j X_j\right) &= \mathrm{Var}\left(\sum_{c(j)=i} z_j X_j\right) + \mathrm{Var}\left(\sum_{c(j)\neq i} z_j X_j\right) \\
&= \sum_{c(j)=i} z_j^2 \sigma^2 + \sum_{\substack{j\neq l \\ c(j)=c(l)=i}} z_j z_l \,\mathrm{Cov}(X_j, X_l) + \sum_{c(j)\neq i} z_j^2 \sigma^2 \\
&= \sigma^2 \sum_{j=1}^{n} z_j^2 + \frac{\sigma^2}{n_i} \sum_{\substack{j\neq l \\ c(j)=c(l)=i}} z_j z_l \\
&= \sigma^2 \sum_{j=1}^{n} z_j^2 + \frac{\sigma^2}{n_i} \left(\sum_{c(j)=c(l)=i} z_j z_l - \sum_{c(j)=i} z_j^2 \right) \\
&\leqslant \sigma^2 \sum_{j=1}^{n} z_j^2 + \frac{\sigma^2}{n_i} \left(n_i \sum_{c(j)=i} z_j^2 - \sum_{c(j)=i} z_j^2 \right) \\
&= \sigma^2 \sum_{j=1}^{n} z_j^2 + \sigma^2 \left(1 - \frac{1}{n_i}\right) \sum_{c(j)=i} z_j^2 \\
&\leqslant 2\sigma^2 \sum_{j=1}^{n} z_j^2.
\end{aligned}$$

Therefore the standard deviation of $\sum_{j=1}^{n} z_j X_j$ can be estimated from above by $\sqrt{2}\sigma \sqrt{\sum_{j=1}^{n} z_j^2}$. Consequently, in the proof of Lemma 2 everything works with $\sqrt{2}\sigma$ instead of σ, and the calculation results in

$$P\left(\left|\sum_{j=1}^{n} Z_j X_j\right| > \frac{1}{n^{\tau-\delta}}\right) \leqslant \sqrt{\frac{2}{\pi}} \frac{\sqrt{2}\sigma}{n^{\delta}} \,\mathrm{e}^{-\frac{n^{2\delta}}{4\sigma^2}},$$

that is the random variable $\sum_{j=1}^{n} Z_j X_j$ tends to zero in probability in order $1/n^{1/2-\delta}$ as $n \to \infty$.

Similar arguments apply to the half of the sixth term in (19)

$$\sum_{\substack{j=1 \\ j\neq i}}^{k} t_{ij} \mathbf{r}_i^{\mathrm{T}} \mathbf{W} \mathbf{u}_j = \sum_{\substack{j=1 \\ j\neq i}}^{k} t_{ij} \mathbf{r}_i^{\mathrm{T}} \mathbf{P} \mathbf{u}_j = \sum_{\substack{j=1 \\ j\neq i}}^{k} t_{ij} \sum_{l=1}^{n} r_{il} \left[\frac{1}{\sqrt{n_j}} \sum_{c(x)=j} p_{lx} \right]. \tag{23}$$

As, by the central limit theorem, the terms in brackets have $\mathscr{N}(0, \sigma^2)$ distribution and the inner sums, with the above modification of Lemma 2, are also of order $1/n^{1/2-\delta}$. Multiplying by $t_{ij} = \mathrm{O}(1/n^{1/2})$ will further decrease the order to $1/n^{1-\delta}$.

It remains to estimate the leading (first) term

$$t_{ii}^2 \mathbf{u}_i^{\mathrm{T}} \mathbf{W} \mathbf{u}_i = t_{ii}^2 \xi_i + t_{ii}^2 [(n_i - 1)\mu_i + \nu_i + \sigma^2/\mu_i], \tag{24}$$

where ξ_i in order $1/\sqrt{n_i}$ has an $\mathscr{N}(0, 2\sigma^2)$ distribution, hence by Lemma 3, $t_{ii}^2 \xi_i$ also has an asymptotically $\mathscr{N}(0, 2\sigma^2)$ distribution in the same order, since $1 - t_{ii}^2 \leqslant a_i/n$ (a_i is a constant, cf. the proof of Corollary 1). As

$$\begin{aligned} t_{ii}^2 [(n_i - 1)\mu_i + \nu_i + \sigma^2/\mu_i] = {} & [(n_i - 1)\mu_i + \nu_i + \sigma^2/\mu_i] \\ & -(1 - t_{ii}^2)[(n_i - 1)\mu_i + \nu_i + \sigma^2/\mu_i], \end{aligned}$$

the constant in the first term on the right-hand side is added to the expected value, and the last term is less than $[(n_i - 1)\mu_i + \nu_i + \sigma^2/\mu_i]a_i/n$, i.e., it is of order O(1). It follows that the random variable $t_{ii}^2 \mathbf{u}_i^{\mathrm{T}} \mathbf{W} \mathbf{u}_i$ differs in at most a constant term from a random variable that asymptotically follows an $\mathscr{N}((n_i - 1)\mu_i + \nu_i + \sigma^2/\mu_i, 2\sigma^2)$ distribution, $i = 1, \dots, k$.

Summarizing, the sum of the six terms in (19)—except the above part of the first term and $k - 1$ subterms in the third one (for $j = 1, \dots, k; \quad j \neq i$) that produce a term of order O(1)—differs in order $1/n^{1/2-\delta}$ from an $\mathscr{N}((n_i - 1)\mu_i + \nu_i + \sigma^2/\mu_i, 2\sigma^2)$ variable, that was to be proved. Further, the conditional distribution of λ_i conditioned on t_{ij}'s is asymptotically $\mathscr{N}(m_i, \sigma^2)$, where $m_i = (n_i - 1)\mu_i + \nu_i + \sigma^2/\mu_i + C_i$ with the constant shift C_i that results in adding together the above O(1) terms:

$$\begin{aligned} C_i = {} & -(1 - t_{ii}^2)[(n_i - 1)\mu_i + \nu_i + \sigma^2/\mu_i] \\ & + \sum_{\substack{j=1 \\ j \neq i}}^{k} t_{ij}^2 [(n_j - 1)\mu_j + \nu_j + \sigma^2/\mu_j]. \end{aligned} \tag{25}$$

That is, m_i satisfies the statement of the theorem. As the constant shift of the means does not change the covariances, conditioned on t_{ij}'s the eigenvalues λ_i's are in order $1/n^{1/2-\delta}$ uncorrelated with variance $2\sigma^2$, this finishes the discussion of the Case (i).

In the Case (ii), if $\beta_1 = \cdots = \beta_k = m$ is a multiple eigenvalue, then according to [5], the random variables $\mathbf{u}_i^{\mathrm{T}} \mathbf{W} \mathbf{u}_i$ ($i = 1, \dots, k$) are asymptotically $\mathscr{N}(m, 2\sigma^2)$ distributed with the same mean and variance, and they are independent of each other.

First we show that the random variables $\mathbf{v}_i^{\mathrm{T}} \mathbf{W} \mathbf{v}_i$'s have the same distribution as $\mathbf{u}_i^{\mathrm{T}} \mathbf{W} \mathbf{u}_i$'s, where the set $\mathbf{v}_1, \dots, \mathbf{v}_k$ of orthonormal vectors is obtained from that of $\mathbf{u}_1, \dots, \mathbf{u}_k$ by the orthogonal rotation $\mathbf{R}$ (cf. Proposition 2). The entries of the latter $k \times k$ orthogonal matrix $\mathbf{R}$ are denoted by r_{ij}.

$$
\begin{aligned}
\mathbf{v}_i^{\mathrm{T}}\mathbf{W}\mathbf{v}_i &= \left(\sum_{j=1}^{k} r_{ji}\mathbf{u}_j^{\mathrm{T}}\right)\mathbf{W}_1\left(\sum_{j=1}^{k} r_{ji}\mathbf{u}_j\right) + \left(\sum_{j=1}^{k} r_{ji}\mathbf{u}_j^{\mathrm{T}}\right)\mathbf{P}_2\left(\sum_{j=1}^{k} r_{ji}\mathbf{u}_j\right) \\
&= \sum_{j=1}^{k} r_{ji}^2\mathbf{u}_j^{\mathrm{T}}\mathbf{W}_1\mathbf{u}_j + 2\sum_{j=1}^{k-1}\sum_{l=j+1}^{k} r_{ji}r_{li}\left[\frac{1}{\sqrt{n_i n_j}}\sum_{c(x)=j}\sum_{c(y)=l} p_{xy}\right],
\end{aligned}
$$

where $\mathbf{W}_1 = \mathbf{B} + \mathbf{P}_1$ and the entries of $\mathbf{P}_2$ in the non-diagonal blocks are identical to those of $\mathbf{P}$ (think of the decomposition $\mathbf{P} = \mathbf{P}_1 + \mathbf{P}_2$ at the beginning of the proof). As the r_{ij}'s are constants, and they are entries of an orthogonal matrix, furthermore the terms $\mathbf{u}_j^{\mathrm{T}}\mathbf{W}_1\mathbf{u}_j$ and those in the brackets are normally distributed (due to the central limit theorem), $\mathbf{v}_i^{\mathrm{T}}\mathbf{W}\mathbf{v}_i$'s will also be normally distributed with the same mean and variance. That is,

$$
E(\mathbf{v}_i^{\mathrm{T}}\mathbf{W}\mathbf{v}_i) = \sum_{j=1}^{k} r_{ji}^2 m + 0 = m
$$

and

$$
\begin{aligned}
\mathrm{Var}(\mathbf{v}_i^{\mathrm{T}}\mathbf{W}\mathbf{v}_i) &= \sum_{j=1}^{k} r_{ji}^4 2\sigma^2 + 4\sigma^2\left(\sum_{j=1}^{k-1}\sum_{l=j+1}^{k} r_{ji}^2 r_{li}^2\right) \\
&= 2\sigma^2\sum_{j=1}^{k} r_{ji}^4 + 2\sigma^2\left(\sum_{j=1}^{k}\sum_{l=1}^{k} r_{ji}^2 r_{li}^2 - \sum_{j=1}^{k} r_{ji}^4\right) \\
&= 2\sigma^2\sum_{j=1}^{k} r_{ji}^2\sum_{l=1}^{k} r_{li}^2 = 2\sigma^2
\end{aligned}
$$

hold true due to the orthogonality of $\mathbf{R}$. To show the independence of $\mathbf{v}_i^{\mathrm{T}}\mathbf{W}\mathbf{v}_i$'s, in view of the normality, it is enough to show that their pairwise covariances are zeros. That is, for $1 \leqslant i < m \leqslant k$ we have that

$$
\begin{aligned}
\mathrm{Cov}(\mathbf{v}_i^{\mathrm{T}}\mathbf{W}\mathbf{v}_i, \mathbf{v}_m^{\mathrm{T}}\mathbf{W}\mathbf{v}_m) &= \sum_{j=1}^{k} r_{ji}^2 r_{jm}^2\mathrm{Var}(\mathbf{u}_j^{\mathrm{T}}\mathbf{W}_1\mathbf{u}_j) \\
&\quad + 2\sum_{j=1}^{k-1}\sum_{l=j+1}^{k} r_{ji}r_{li}r_{jm}r_{lm}\mathrm{Var}(\mathbf{u}_j^{\mathrm{T}}\mathbf{P}_2\mathbf{u}_l) \\
&= 2\sigma^2\sum_{j=1}^{k} r_{ji}^2 r_{jm}^2 \\
&\quad + 4\sigma^2\frac{1}{2}\left(\sum_{j=1}^{k} r_{ji}r_{jm}\sum_{l=1}^{k} r_{li}r_{lm} - \sum_{j=1}^{k} r_{ji}^2 r_{jm}^2\right) = 0.
\end{aligned}
$$

As $\|\mathbf{y}_i - \mathbf{v}_i\|^2 = \mathrm{O}(1/n)$, the same calculation with $\mathbf{v}_i$'s instead of $\mathbf{u}_i$'s can be performed (cf. Corollary 2) that finishes the proof. □

5. Conclusions and further remarks

The following corollary comes out easily from the proof of Theorem 4.

Corollary 5. *Given the eigenvectors, for the eigenvalue λ_i of* $\mathbf{W}$ *the relation*

$$P\left(|\lambda_i - \eta_i| > C_i + \frac{1}{n^{1/2-\delta}}\right) = \mathrm{O}\left(\frac{1}{n}\right)$$

holds, where η_i follows an $\mathscr{N}((n_i - 1)\mu_i + \nu_i + \sigma^2/\mu_i, 2\sigma^2)$ distribution, and C_i is as defined in (25), $i = 1, \dots, k$.

Proof. Let γ_i be a random variable that asymptotically follows an $\mathscr{N}((n_i - 1)\mu_i + \nu_i + \sigma^2/\mu_i, 2\sigma^2)$ distribution, in order $1/n^{1/2-\delta}$, and differs from λ_i in the constant C_i. With it, the argument

$$\begin{aligned} P\left(|\lambda_i - \eta_i| > C_i + \frac{1}{n^{1/2-\delta}}\right) &\leqslant P\left(|\lambda_i - \gamma_i| + |\gamma_i - \eta_i| > C_i + \frac{1}{n^{1/2-\delta}}\right) \\ &\leqslant P\left(C_i + |\gamma_i - \eta_i| > C_i + \frac{1}{n^{1/2-\delta}}\right) \\ &= P\left(|\gamma_i - \eta_i| > \frac{1}{n^{1/2-\delta}}\right) = \mathrm{O}\left(\frac{1}{n}\right), \end{aligned}$$

is valid, where the last equation follows from the proof of the Theorem FK1. □

Corollary 5 shows that with probability tending to 1, the eigenvalue λ_i is within a constant distance from a normally distributed random variable that takes on values with high probability in a bounded region. For example, to the 95% level a confidence interval around β_i with radius $C_i + 1.96\sqrt{2}\sigma + \mathrm{O}(1/\sqrt{n})$ can be formed. Even in the generic case, the O(1) terms can be bounded with a constant C ($C > C_i$), and with this C, the statement of Corollary 5 remains valid. In fact, $C = 2k\sigma$ comes out with a more refined computation that is, indeed, very "small" compared to the size of the problem.

We could further improve even the rough estimate for the eigenvalues, if we gave a better upper bound for the largest absolute value eigenvalue of the between-blocks perturbation $\mathbf{P}_2$. Our conjecture is that it is of order $2\sigma\sqrt{n(k-1)/k}$. That is, following the proof of Theorem 2 in [5] the number of desired walks can be reduced to roughly $[n(k-1)/k]^k$, as only walks affecting consequent vertices of different colors count toward the non-zero expectation, so instead of $2\sigma\sqrt{n}$ we would have $2\sigma\sqrt{n(k-1)/k}$ as upper bound.

We performed this spectral decomposition of weight matrices assigned to the edges of some special graphs describing telecommunication, biological, and sociological networks. In some cases we experienced the same order of magnitude of the eigenvalues as described in Theorems 1 and 2, and the representatives were well classifiable into k clusters with k-variance of order of Theorem 3. In such situations k latent clusters can be assumed and even obtained with the help of computer programs available for matrices of size up to 1000. Here the integer k is chosen by inspection from the spectral gap. We remark that such a block-structure is not obvious for the first sight, partly because of the large size of the problem, and partly because of possible permutations of the rows or/and columns of $\mathbf{W}$.

Physicists [4] found deviations from the semi-circle law if the model is either sparse uncorrelated (the expectation p of the independent and Bernoulli-distributed w_{ij}'s tends to zero in such a way that np tends to a constant) or correlated (w_{ij}'s are formed randomly but not with equal probabilities as the "real-world" graph evolves). The spectral density shows a high peak around zero in these cases. It is not the case in our model. We have k dominant eigenvalues, with high probability they are in a bounded region far away from the others that obediently follow the semi-circle law.

Acknowledgements

I would like to thank Gábor Tusnády and Miklós Simonovits for drawing my attention to this perturbation and random graph problem, Zoltán Füredi, András Krámli, Péter Major for their valuable comments and Gábor Molnár-Sáska for making computer simulations. I also thank an anonymous referee for useful suggestions.

References

[1] L. Arnold, On the asymptotic distribution of the eigenvalues of random matrices, J. Math. Anal. Appl. 20 (1967) 262–268.
[2] M. Bolla, Methods and statistical applications for the singular values decomposition of real matrices (in Hungarian), Ph.d. Thesis, Eötvös Loránd University, Budapest, 1982.
[3] M. Bolla, G. Tusnády, Spectra and optimal partitions of weighted graphs, Discrete Math. 128 (1994) 1–20.
[4] I.J. Farkas, I. Derényi, A.-L. Barabási, T. Vicsek, Spectra of "Real-World" graphs: beyond the semi-circle law, Phys. Rev. E 64 (2001) 026704.
[5] Z. Füredi, J. Komlós, The eigenvalues of random symmetric matrices, Combinatorica 1 (3) (1981) 233–241.
[6] F. Juhász, On the spectrum of a random graph, in: Lovász, et al. (Eds.), Algebraic Methods in Graph Theory, Coll. Math. Soc. J. Bolyai 25, North-Holland, 1981, pp. 313–316.
[7] M.L. Mehta, Random Matrices, Academic Press, New York, 1991.
[8] S. Ross, A First Course in Probability, fifth ed., Prentice-Hall, New York, 1998.
[9] E.P. Wigner, On the distribution of the roots of certain symmetric matrices, Ann. Math. 62 (1958) 325–327.
[10] J.H. Wilkinson, The Algebraic Eigenvalue Problem, Clarendon Press, Oxford, 1965.