Bernoulli **26**(3), 2020, 2400–2435
https://doi.org/10.3150/20-BEJ1201

Local law and Tracy–Widom limit for sparse stochastic block models

JONG YUN HWANG*, JI OON LEE† and WOOSEOK YANG‡

Department of Mathematical Sciences, KAIST, Daejeon, 34141, Korea.
E-mail: *jyh9006@kaist.ac.kr; †jioon.lee@kaist.edu; ‡ws.yang@kaist.ac.kr

We consider the spectral properties of sparse stochastic block models, where N vertices are partitioned into K balanced communities. Under an assumption that the intra-community probability and inter-community probability are of similar order, we prove a local semicircle law up to the spectral edges, with an explicit formula on the deterministic shift of the spectral edge. We also prove that the fluctuation of the extremal eigenvalues is given by the GOE Tracy–Widom law after rescaling and centering the entries of sparse stochastic block models. Applying the result to sparse stochastic block models, we rigorously prove that there is a large gap between the outliers and the spectral edge without centering.

Keywords: local law; random matrices; stochastic block model; Tracy–Widom distribution

1. Introduction

Understanding an underlying network structure is one of the key problems in data science. Many real world data sets can be viewed as networks of interacting nodes, and a common approach to analyze the structure of the network is to find which nodes share similar properties so that they can be grouped into a community. Clustering, or community detection, to recognize such communities from given data sets is thus a natural and fundamental problem.

Community detection problem is vital in understanding the real-world networks. In biology and bioinformatics, community detection appears in finding functional modules in protein-protein interaction networks [7], functional mapping of metabolic network [18,19], analyzing gene expression data [9,25] and more. Community detection problems also naturally arise in social networks. The “friendships” networks of Facebook, the online social network, was studied, including anonymous Facebook users in one hundred American universities [33,34]. Communities of the network were identified, and it was found that the community structure depends strongly on their offline network, such as class year or House affiliation. There have been studies on community structures of other social networks, such as scientific collaboration networks [16, 30].

The stochastic block model (SBM) is one of the simplest models of the network with communities. First appeared in the study of social networks [21], it consists of N vertices partitioned into disjoint subsets $C_1, \dots, C_K$ and a $K \times K$ symmetric matrix P of edge probabilities. The model appears in various fields of study, and numerous results have been obtained for community detection in SBM, including various algorithms [3,16,17,20,26], phase transitions [2], and fundamental limits [31]. We refer to [1] for history and recent developments of the community detection problem and stochastic block models.

The spectral method is one of the most well-known approaches for the community detection of SBM. In this method, the adjacency matrix of a given graph is considered, whose extremal eigenvalues and corresponding eigenvectors contain the information on the ground truth of the model. In the simplest example of an SBM with two communities of equal size, if we denote the $N \times N$ adjacency matrix by A, the probability matrix P is a 2×2 matrix, and the expected adjacency matrix $\mathbb{E}A$ has four blocks, that is,

$$\mathbb{E}A = \left(\begin{array}{c|c} P_{11} & P_{12} \\ \hline P_{21} & P_{22} \end{array}\right).$$

If $P_{11} = P_{22} = p_s$ and $P_{12} = P_{21} = p_d$, the first two eigenvalues of $\mathbb{E}A$ are $N(p_s + p_d)/2$ and $N(p_s - p_d)/2$, and the eigenvalue 0 has multiplicity $N - 2$. If the difference $A - \mathbb{E}A$ is small, then the eigenstructure of A is not much different from that of $\mathbb{E}A$, and one can recover the community structure from the second eigenvector of A. The spectral method is also useful in determining the number of communities K when it is not known a priori [5,28].

In the spectral method, the perturbation $H := A - \mathbb{E}A$, which called centered SBM, is a random matrix, and its property, especially the behavior of its largest eigenvalue, can be precisely predicted by results from random matrix theory when P does not depend on N. However, H is different from Wigner matrices in two aspects: (1) the variances of entries are not identical, and (2) the matrix is sparse. (See Assumption 2.1 for more detail on the sparsity.) The first aspect is due to that the intra-community probability p_s and the inter-community probability p_d are different from each other and hence the random variables have different variances. The second aspect is common in many real data, since the expected degree is much smaller than N and the edge probability decays as N grows. For sparse random matrices with identical off-diagonal entries, which correspond to Erdős–Rényi graphs, the spectral properties were obtained in [10,12, 27]. One of the most notable aspects of sparse random matrices is that the deterministic shift of their largest eigenvalues are much larger than the size of the Tracy–Widom fluctuation. Thus, as discussed in Remark 2.14 of [27], in cases where the intra-community probability p_s and the inter-community probability p_d are both small and close to each other, we can predict that the algorithms for the community detection should reflect the shift of the largest eigenvalues if p_s, $p_d \ll N^{-1/3}$. However, to our best knowledge, it has not been proved for sparse SBM.

In this paper, we consider the spectral properties of sparse SBM with K communities. We assume that the communities are of equal size, or balanced, with $P_{ii} = p_s$ and $P_{ij} = p_d$ for $i \neq j$. We further assume that the model is moderately sparse as in Assumption 2.1. Our main contributions are

(1) proof of local semicircle law for the centered sparse SBM that is believed to be optimal up to the edge of the spectrum (Theorems 2.6 and 2.8),
(2) proof of the Tracy–Widom limit for the shifted, rescaled largest eigenvalue of the centered sparse SBM (Theorem 2.13), and
(3) application to the (non-centered) sparse SBM (Theorem 2.15).

The local semicircle law, the estimates on the resolvent of Wigner type matrices, has been the starting point in the local spectral analysis of Wigner matrices [14,15] and Erdős–Rényi graphs [10,12]. We follow the classical strategy based on Schur complement formula and self-consistent equations as in [14,15], which leads us to a weak local law for the resolvent entries (Theorem 2.8). Since the weak local law is not sufficient for the proof of other properties such as the Tracy–Widom fluctuation of the extremal eigenvalues, we improve it to prove the strong local law for the normalized trace of the resolvent (Theorem 2.6), which is optimal up to the edge of the spectrum, by adapting the strategy of [23,27].

The proof of the Tracy–Widom limit of the extremal eigenvalues is based on the Green function comparison method that utilizes a continuous interpolation as in [27]. With the continuous flow, we can track the change of the normalized trace over time, which is offset by the (deterministic) shift of the spectral edge.

When applying the local spectral properties of Wigner matrices or Erdős–Rényi graphs to the SBM, one of the main technical challenges stems from that the entries in the SBM are not identically distributed, especially the means of the entries are not equal, and thus the results from random matrix theory are not directly applicable. While the difficulty can be overcome by algorithms as in [16], it requires a priori knowledge on the number of clusters K. In this paper, we handle the issue by proving that there is a gap of order 1 between K-th largest eigenvalue and $(K+1)$-st one, which is much larger than the gap between the $(K+1)$-st and the $(K+2)$-nd, when the number of clusters is K. This results justifies the use of the spectral method for community detection even when the SBM is sparse.

In the proof of the local law, as in [27], we choose a polynomial $P(m)$ of the normalized trace m of the Green function, based on a recursive moment estimate. However, the fluctuation averaging mechanism, which was intrinsic in the analysis of Erdős–Rényi graph, is much more complicated for the SBM due to the lack of the symmetry. Technically, it means that we need to separate the off-diagonal elements H_{ij} into two cases in the cumulant expansion – one with when i and j are in the same community and the other when i and j are in different communities. With the separation, the two indices i and j do not decouple completely in the cumulant expansion, and we resolve the problem by replacing the diagonal entries of the Green function by m and further expanding the error terms generated by the replacement. Due to this additional expansion, we need to consider a more complicated polynomial P than the Erdős–Rényi case in [27], and the analysis is more involved for the limiting distribution.

This paper is organized as follows: In Section 2, we introduce our model and state the main results. In Section 3, we describe the outline of our proof. In Section 4, we prove a weak local semicircle law, which is used as an a priori estimate in the proof of our main results. In Section 5, we prove the strong local law by using the recursive moment estimates. In Section 6, we prove the Tracy–Widom limit of the largest eigenvalue using the Green function comparison method. Some technical lemmas in the proof are proved in Supplement [24].

Remark 1.1 (Notational remark). We use the symbols $O(\cdot)$ and $o(\cdot)$ for the standard big-O and little-o notation. The notations O, o, $\ll$, $\gg$ always refer to the limit $N \to \infty$ unless otherwise stated. Here, the notation $a \ll b$ means $a = o(b)$. We use c and C to denote positive constants that do not depend on N. Their values may change from line to line. For summation index, we use $\mathrm{i} \sim \mathrm{j}$ if i and j are within the same group. We write $a \sim b$ if there is $C \geq 1$ such that $C^{-1}|b| \leq |a| \leq C|b|$. Throughout this paper, we denote $z = E + \mathrm{i}\eta \in \mathbb{C}^+$ where $E = \operatorname{Re} z$ and $\eta = \operatorname{Im} z$.

2. Definition and main results

2.1. Models and notations

Let H be an $N \times N$ symmetric matrix with K^2 blocks of same size. The blocks are based on the partition of the vertex set $[N] := \{1, 2, \dots, N\}$,

$$[N] = V_1 \cup V_2 \cup \cdots \cup V_K, \tag{2.1}$$

where $|V_i| = N/K$. For $i, j \in \{1, 2, \dots N\}$, we may consider two types of the edge probability H_{ij}, depending on whether i and j are within the same vertex set V_ℓ or not. More precisely, we consider the sparse block matrix model satisfying the following assumption.

Assumption 2.1 (Balanced generalized sparse random matrix). Fix any small $\phi > 0$. We assume that $H = (H_{ij})$ is a real $N \times N$ block random matrix with K balanced communities with $1 \le K \le N$, whose diagonal entries are almost surely zero and whose off-diagonal entries are independently distributed random variables, up to symmetry constraint $H_{ij} = H_{ji}$. We suppose that each H_{ij} satisfies the moment conditions

$$\mathbb{E}H_{ij} = 0, \qquad \mathbb{E}|H_{ij}|^2 = \sigma_{ij}^2, \qquad \mathbb{E}|H_{ij}|^k \le \frac{(Ck)^{ck}}{Nq^{k-2}}, \quad (k \ge 2), \tag{2.2}$$

with sparsity parameter q satisfying

$$N^{\phi} \le q \le N^{1/2}. \tag{2.3}$$

Here, we further assume the normalization condition

$$\sum_{i=1}^{N} \sigma_{ij}^2 = 1. \tag{2.4}$$

We denote by $\kappa_{ij}^{(k)}$ the k-th cumulant of H_{ij}. Under the moment condition (2.2),

$$\kappa_{ij}^{(1)} = 0, \qquad \left|\kappa_{ij}^{(k)}\right| \le \frac{(2Ck)^{2(c+1)k}}{Nq^{k-2}}, \quad (k \ge 2). \tag{2.5}$$

For our model with the block structure, we abbreviate $\kappa_{ij}^{(k)}$ as

$$\kappa_{ij}^{(k)} = \begin{cases} \kappa_s^{(k)} & \text{if } i \text{ and } j \text{ are within same community,} \\ \kappa_d^{(k)} & \text{otherwise.} \end{cases} \tag{2.6}$$

We will also use the normalized cumulants, $s^{(k)}$, by setting

$$s_{(\cdot)}^{(1)} := 0, \qquad s_{(\cdot)}^{(k)} := Nq^{k-2}\kappa_{(\cdot)}^{(k)}, \quad (k \ge 2). \tag{2.7}$$

We notice that we assume $H_{ii} = 0$ a.s., although this condition can be easily removed. For convenience, we define the parameters ζ and $\xi^{(4)}$ as

$$\zeta := \frac{s_s^{(2)} - s_d^{(2)}}{K} = \frac{N(\kappa_s^{(2)} - \kappa_d^{(2)})}{K}, \qquad \xi^{(4)} := \frac{s_s^{(4)} + (K-1)s_d^{(4)}}{K}. \tag{2.8}$$

The prominent example of a balanced generalized sparse random matrix is the case where H_{ij} is given by the Bernoulli random variable A_{ij} with probability p_s or p_d, depending on whether i and j are within same vertex set or not, respectively. For this reason, in this paper, we oftentimes use the term 'centered generalized stochastic block model' or 'cgSBM' as a representative of the balanced generalized sparse random matrix.

In the rest of this subsection, we introduce some notations of basic definitions.

Definition 2.2 (High probability events). We say that an N-dependent event $\Omega \equiv \Omega^{(N)}$ holds with high probability if for any (large) $D > 0$,

$$\mathbb{P}(\Omega^c) \le N^{-D},$$

for $N \ge N_0(D)$ sufficiently large.

Definition 2.3 (Stochastic domination). Let $X \equiv X^{(N)}$, $Y \equiv Y^{(N)}$ be N-dependent non-negative random variables. We say that X stochastically dominates Y if, for all small $\epsilon > 0$ and large $D > 0$,

$$\mathbb{P}\big(X^{(N)} > N^{\epsilon} Y^{(N)}\big) \le N^{-D}, \tag{2.9}$$

for sufficiently large $N \ge N_0(\epsilon, D)$, and we write $X \prec Y$. When $X^{(N)}$ and $Y^{(N)}$ depend on a parameter $u \in U$, then we say $X(u) \prec Y(u)$ uniformly in $u \in U$ if the threshold $N_0(\epsilon, D)$ can be chosen independently of u.

Throughout this paper, we choose $\epsilon > 0$ sufficiently small. (More precisely, it is smaller than $\phi/10$, where $\phi > 0$ is the fixed parameter in Assumption 2.1 below.)

Definition 2.4 (Stieltjes transform). For given a probability measure ν, we define the Stieltjes transforms of ν as

$$m_\nu(z) := \int \frac{\nu(\mathrm{d}x)}{x - z}, \quad \big(z \in \mathbb{C}^+\big)$$

For example, the Stieltjes transform of the *semicircle measure*,

$$\varrho(\mathrm{d}x) := \frac{1}{2\pi}\sqrt{\big(4 - x^2\big)_+}\,\mathrm{d}x,$$

is given by

$$m_{sc}(z) = \int \frac{\varrho(\mathrm{d}x)}{x - z} = \frac{-z + \sqrt{z^2 - 4}}{2},$$

where the argument of $\sqrt{z^2-4}$ is chosen so that $m_{sc}(z) \in \mathbb{C}^+$ for $z \in \mathbb{C}^+$ and $\sqrt{z^2-4} \sim z$ as $z \to \infty$. Clearly, we have

$$m_{sc}(z) + m_{sc}(z)^{-1} + z = 0.$$

Definition 2.5 (Green function (Resolvent)). Given a real symmetric matrix H we define its *Green function* or *resolvent*, $G(z)$, and the normalized trace of its Green function, m^H, by

$$G^H(z) \equiv G(z) := (H - zI)^{-1}, \qquad m^H(z) \equiv m(z) := \frac{1}{N} \operatorname{Tr} G^H(z), \tag{2.10}$$

where $z = E + \mathrm{i}\eta \in \mathbb{C}^+$ and I is the $N \times N$ identity matrix.

Denoting by $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_N$ the ordered eigenvalues of H, we note that m^H is the Stieltjes transform of the empirical eigenvalue measure of H, μ^H, defined as

$$\mu^H := \frac{1}{N} \sum_{i=1}^{N} \delta_{\lambda_i}.$$

Finally, we introduce the following domains in the upper-half plane

$$\mathcal{E} := \{ z = E + \mathrm{i}\eta \in \mathbb{C}^+ : |E| < 3, 0 < \eta \leq 3 \}, \tag{2.11}$$

$$\mathcal{D}_\ell := \{ z = E + \mathrm{i}\eta \in \mathbb{C}^+ : |E| < 3, N^{-1+\ell} < \eta \leq 3 \}. \tag{2.12}$$

2.2. Main results

Our first main result is the local law for m^H, the normalized trace of $G^H(z)$, up to the spectral edges.

Theorem 2.6 (Strong local law). *Let H satisfy Assumption 2.1 with $\phi > 0$. Then, there exist an algebraic function $\widetilde{m} : \mathbb{C}^+ \to \mathbb{C}^+$ and the deterministic number $2 \leq L < 3$ such that the following hold*:

(1) *The function $\widetilde{m}$ is the Stieltjes transform of a deterministic probability measure $\widetilde{\rho}_t$, that is, $\widetilde{m}(z) = m_{\widetilde{\rho}}(z)$. The measure ρ is supported on $[-L, L]$ and $\widetilde{\rho}$ is absolutely continuous with respect to Lebesgue measure with a strictly positive density on $(-L, L)$.*
(2) *The function $\widetilde{m} \equiv \widetilde{m}(z)$, $z \in \mathbb{C}^+$, is a solution to the polynomial equation*

$$\begin{aligned} P_{1,z}(\widetilde{m}) &:= 1 + z\widetilde{m} + \widetilde{m}^2 + q^{-2}\left(\frac{s_s^{(4)} + (K-1)s_d^{(4)}}{K}\right)\widetilde{m}^4 \\ &= 1 + z\widetilde{m} + \widetilde{m}^2 + q^{-2}\xi^{(4)}\widetilde{m}^4 = 0. \end{aligned} \tag{2.13}$$

(3) *The normalized trace $m(z)$ of the Green function $G(z)$ satisfies the local law*

$$\left|m(z) - \widetilde{m}(z)\right| \prec \frac{1}{q^2} + \frac{1}{N\eta}, \tag{2.14}$$

uniformly on the domain $\mathcal{E}$.

The function $\widetilde{m}$ was first introduced in [27] to consider a correction term to the semicircle measure in the sparse setting. Some properties of probability measure $\widetilde{\rho}$ and its Stieltjes transform $\widetilde{m}$ are collected in Lemma A.2.

From the local law in (2.14), we can easily prove the following estimates on the *local density of states of H*. For $E_1 < E_2$ define

$$\mathfrak{n}(E_1, E_2) := \frac{1}{N}\left|\{i : E_1 < \lambda_i < E_2\}\right|, \qquad n_{\widetilde{\rho}}(E_1, E_2) := \int_{E_1}^{E_2} \widetilde{\rho}(x)\,\mathrm{d}x.$$

Corollary 2.7 (Integrated density of states). *Suppose that H satisfies Assumption 2.1 with $\phi > 0$. Let $E_1, E_2 \in \mathbb{R}$, $E_1 < E_2$. Then,*

$$\left|\mathfrak{n}(E_1, E_2) - n_{\widetilde{\rho}}(E_1, E_2)\right| \prec \frac{E_1 - E_2}{q^2} + \frac{1}{N}. \tag{2.15}$$

Corollary 2.7 easily follows from Theorem 2.6 by applying the Helffer–Sjöstrand calculus. We refer to Section 7.1 of [11] for more detail.

The proof of Theorem 2.6 is based on the following a priori estimates on entries of the resolvents, which we call the weak local semicircle law. While the weak law for m is indeed weaker than the strong local law, Theorem 2.6, we have here an entrywise law, which is believed to be optimal.

Theorem 2.8 (Weak local semicircle law). *Suppose H satisfies Assumption 2.1. Define the spectral parameter $\psi(z)$ by*

$$\psi(z) := \frac{1}{q} + \frac{1}{\sqrt{N\eta}}, \qquad (z = E + \mathrm{i}\eta).$$

Then for any sufficiently small ℓ, the events

$$\max_{i \neq j}\left|G_{ij}(z)\right| \prec \psi(z) \tag{2.16}$$

$$\max_{i \neq j}\left|G_{ii}(z) - m\right| \prec \psi(z), \tag{2.17}$$

and

$$\left|m(z) - m_{sc}(z)\right| \prec \frac{1}{\sqrt{q}} + \frac{1}{(N\eta)^{1/3}} \tag{2.18}$$

hold uniformly on the domain $\mathcal{D}_\ell$.

Remark 2.9. We can extend Theorem 2.8 to domain $\mathcal{E}$ as in the proof of Theorem 2.8 in [12]. However, we do not pursue the direction in this paper.

An immediate consequence of (2.17) of Theorem 2.8 is the complete delocalization of the eigenvectors.

Corollary 2.10. *Suppose that H satisfies Assumption 2.1 with $\phi > 0$. Denote by (u_i^H) the ℓ^2-normalized eigenvectors of H. Then,*

$$\max_{1 \le i \le N} \left\| u_i^H \right\|_\infty \prec \frac{1}{\sqrt{N}}. \tag{2.19}$$

For the proof, we refer to the proof of Corollary 3.2 of [14].

Together with the weak local semicircle law, a standard application of the moment method yields the following weak bound on $\|H\|$; see, *for example*, Lemma 4.3 of [12] and Lemma 7.2 of [14].

Lemma 2.11. *Suppose that H satisfies Assumption 2.1 with $\phi > 0$. Then,*

$$\bigl| \|H\| - 2 \bigr| \prec \frac{1}{q^{1/2}}. \tag{2.20}$$

From the strong law, we can sharpen the estimate (2.20) by containing the deterministic refinement to the semicircle law.

Theorem 2.12. *Suppose that H satisfies Assumption 2.1 with $\phi > 0$. Then,*

$$\bigl| \|H\| - L \bigr| \prec \frac{1}{q^4} + \frac{1}{N^{2/3}}, \tag{2.21}$$

where $\pm L$ are the endpoints of the support of the measure $\widetilde{\rho}$ given by

$$L = 2 + \frac{\xi^{(4)}}{q^2} + O\bigl(q^{-4}\bigr). \tag{2.22}$$

Our last main result states that the fluctuations of the rescaled largest eigenvalue of the centered generalized stochastic block model are given by the Tracy–Widom law when the sparsity parameter q satisfies $q \gg N^{1/6}$.

Theorem 2.13. *Suppose that H satisfies Assumption 2.1 with $\phi > 1/6$. Denote by λ_1^H the largest eigenvalue of H. Then*

$$\lim_{N \to \infty} \mathbb{P}\bigl(N^{2/3}\bigl(\lambda_1^H - L\bigr) \le s\bigr) = F_1(s) \tag{2.23}$$

where L is given in (2.22) and F_1 is the cumulative distribution function of the GOE Tracy–Widom law.

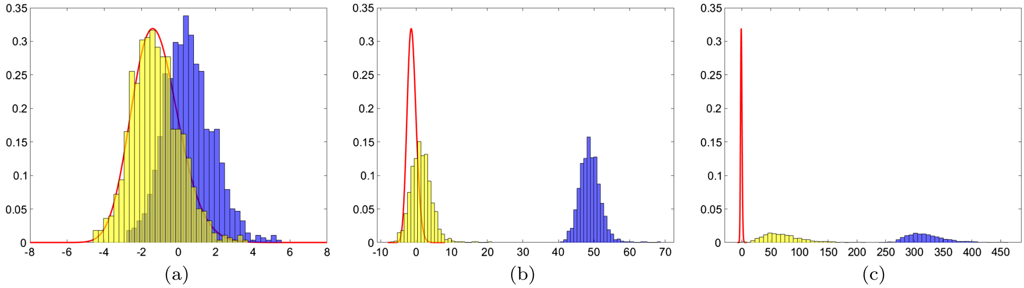


Figure 1. ESD of the largest eigenvalues shifted by 2 (blue) and by L (yellow) plotted against the Tracy–Widom law (red line): cgSBM with $N = 27{,}000$, $K = 3$ and (a) $p_s = 0.03$, $p_d = 0.01$; (b) $p_s = 0.009$, $p_d = 0.006$; (c) $p_s = 0.002$, $p_d = 0.001$.

In Figure 1, we plot histograms of shifted and rescaled largest eigenvalues of the $27{,}000 \times 27{,}000$ cgSBM against the Tracy–Widom distribution when p_s and p_d are (a) between $N^{-1/3}$ and $N^{-2/3}$, (b) between $N^{-2/3}$ and $N^{-7/9}$, and (c) less then $\log N/N$. The red line is the Tracy–Widom distribution and the blue histogram is the eigenvalue histogram shifted by 2 and the yellow one is histogram shifted by L. In the case (a), we see that the deterministic shift given by $L - 2 = O(q^{-2})$ is essential so that the empirical eigenvalue distribution of the sample matrices follows the Tracy–Widom distribution when we shift it by L instead of 2.

In (b), Figure 1(b) shows that the empirical distribution of the sample matrices does not follow the Tracy–Widom law. In [22], it was shown that for Erdős–Rényi graph, there exists a transition from Tracy–Widom to Gaussian fluctuations when $p \sim N^{-2/3}$. We expect that the fluctuation of extreme eigenvalues of cgSBM will also follow the Gaussian distribution in this regime. Finally, in (c) where $p_s, p_d \ll \frac{\log N}{N}$, cgSBM contains an isolated vertex almost surely and thus is disconnected. Due to this disconnectedness, properties of a cgSBM will be entirely changed. In this case, as shown in Figure 1(c), the empirical distribution follows neither the Tracy–Widom distribution nor the Gaussian distribution.

2.3. Applications to the adjacency matrix of the sparse SBM

When $K = 1$, in [12], it was proved that the largest and second largest eigenvalues are separated by a gap of order one when $p \geq (1+\epsilon)/N$. We expect that the similar result holds for SBM with fixed number of clusters. We also introduce the conjecture which stated in [1].

Conjecture 2.14. *Let A be the cgSBM with K clusters, intra-community probability p_s and the inter-community probability p_d. Define $SNR = \frac{N(p_s-p_d)^2}{K(p_s+(K-1)p_d)}$. Then for any $K \geq 2$, if $SNR > 1$ (the Kesten–Stigum (KS) threshold), it is possible to detect communities in polynomial time.*

Now consider an adjacency matrix of the cgSBM with N vertices and K communities. To make bulk eigenvalues lie in an order one interval, we may rescale this matrix ensemble and we are led to the following random matrix ensemble. Let A be a real symmetric $N \times N$ matrix whose entries, A_{ij}, are independent random variables satisfy

$$\mathbb{P}\left(A_{ij} = \frac{1}{\sigma}\right) = \begin{cases} p_s & (i \sim j) \\ p_d & (i \nsim j), \end{cases}$$
$$\mathbb{P}(A_{ij} = 0) = \begin{cases} 1 - p_s & (i \sim j) \\ 1 - p_d & (i \nsim j), \end{cases} \qquad \mathbb{P}(A_{ii} = 0) = 1, \tag{2.24}$$

where $\sigma^2 := \frac{N}{K} p_s(1 - p_s) + \frac{N(K-1)}{K} p_d(1 - p_d)$. Then we can get $\widetilde{A} := A - \mathbb{E}A$ which is a centered matrix obtained from A. Here, $\widetilde{A}_{ij}$ have the distribution

$$\begin{aligned} &\mathbb{P}\left(\widetilde{A}_{ij} = \frac{1 - p_s}{\sigma}\right) = p_s, \qquad \mathbb{P}\left(\widetilde{A}_{ij} = -\frac{p_s}{\sigma}\right) = 1 - p_s, \quad (i \sim j) \\ &\mathbb{P}\left(\widetilde{A}_{ij} = \frac{1 - p_d}{\sigma}\right) = p_d, \qquad \mathbb{P}\left(\widetilde{A}_{ij} = -\frac{p_d}{\sigma}\right) = 1 - p_d, \quad (i \nsim j), \\ &\mathbb{P}(\widetilde{A}_{ii} = 0) = 1. \end{aligned} \tag{2.25}$$

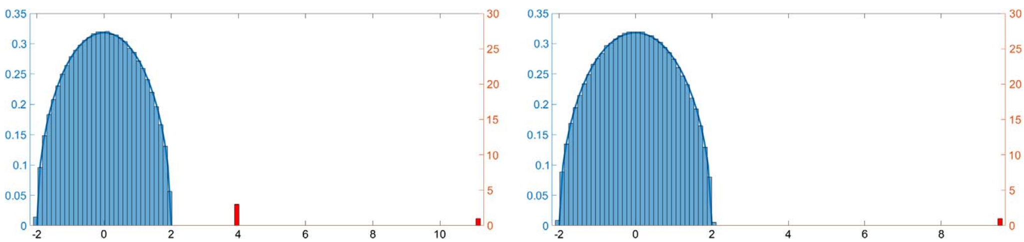
When p_s and p_d have order $N^{-1+2\phi}$, it can be easily shown that $\widetilde{A}$ satisfies Assumption 2.1 with $q \sim N^{\phi}$. Now we may apply Weyl's inequality to $A = \mathbb{E}A + \widetilde{A}$ and get the following conjecture.

Conjecture 2.15. *Fix $\phi > 0$. Let A and $\widetilde{A}$ satisfy (2.24) and (2.25) with $N^{-1+2\phi} \le p_s, p_d \le N^{-2\phi}$ and fixed number of cluster K. Then there is a constant c such that*

$$\lambda_N^A \le \cdots \le \lambda_{K+1}^A \le \lambda_1^{\widetilde{A}} \le 2 + c \le \lambda_K^A \le \cdots \le \lambda_1^A. \tag{2.26}$$

with high probability. In other words, there is an order one gap between the K largest eigenvalues and other eigenvalues.

In Figure 2(a), as expected in Conjecture 2.15, we can see a gap between the outliers and the bulk. However, if K increases in proportion to N, the gap may disappear. For simplicity,



(a) (b)

Figure 2. Empirical distributions of the eigenvalues of 8000×8000 balanced SBM with (a) 4 clusters and (b) 20 clusters. Red bar shows the number of outliers at that position.

assume that $p_s = (c+1)p_d$ and $p_d \sim N^{-1+\theta}$. Then nonzero eigenvalues of $\mathbb{E}A$ are $\frac{(c+K)N^\theta}{K\sigma}$ and $\frac{cN^\theta}{K\sigma}$ of multiplicity 1 and $K-1$, respectively. Since $\sigma^2 \sim O(N^\theta)$ in this case, the largest eigenvalue is approximately order of $N^{\theta/2}$ and second one is order of $N^{\theta/2}K^{-1}$. Therefore, as in Figure 2(b), if $K \gg N^{\theta/2}$, we may not observe the gap between outliers and the bulk of the spectrum, i.e., we cannot detect the proper community structure. Note that this result coincides with Conjecture 2.14 since SNR $\gg 1$ if and only if $K \ll N^{\theta/2}$.

Remark 2.16. We can extend Conjecture 2.15 to more general matrix ensemble, which satisfies the condition that $A - \mathbb{E}A$ follows Assumption 2.1 and entries of $\mathbb{E}A$ is order of $N^{-1+\epsilon}$ for any small ϵ. We further remark that all our results also hold for complex Hermitian balanced generalized sparse random matrices without any change except that the limiting edge fluctuation is given by GUE Tracy–Widom law.

3. Strategy and outline of the proof

In this section, we briefly outline the strategy of our proofs for the results in Section 2.

3.1. Main strategy for the proof

As illustrated in Section 3 of [27], a good estimate on the expectation of a sufficiently high power of the quadratic polynomial $1+zm+m^2$ is enough for the proof of the strong local law. To obtain such an estimate, we expand the term $1+zm$ by using a simple identity

$$1 + zG_{ii} = \sum_{k=1}^{N} H_{ik}G_{ki}. \tag{3.1}$$

In the expansion, which was called the resolvent expansion in [27], it is not easy to fully expand the terms with high powers, and the main idea in [27] was to introduce the recursive moment estimate that estimates $\mathbb{E}|1+zm+m^2|^D$ by the lower moments $\mathbb{E}|1+zm+m^2|^{D-\ell}$ for $\ell \geq 1$. When used together with the resolvent expansion, it makes the tracking of the higher order terms much simpler.

In the actual estimate of the moments, we use a generalized version of Stein's lemma, which was introduced in [32]. It was used in the study of the linear eigenvalue statistics of random matrices [8,29] and also the joint convergence of the largest eigenvalue and the linear statistics [4].

Lemma 3.1 (Cumulant expansion, generalized Stein's lemma). *Fix $\ell \in \mathbb{N}$ and let $F \in C^{\ell+1}(\mathbb{R};\mathbb{C}^+)$. Let Y be a centered random variable with finite moments to order $\ell+2$. Then,*

$$\mathbb{E}\big[YF(Y)\big] = \sum_{r=1}^{\ell} \frac{\kappa^{(r+1)}(Y)}{r!}\mathbb{E}\big[F^{(r)}(Y)\big] + \mathbb{E}\big[\Omega_\ell\big(YF(Y)\big)\big], \tag{3.2}$$

where $\mathbb{E}$ denotes the expectation with respect to Y, $\kappa^{(r+1)}(Y)$ denotes the $(r+1)$-st cumulant of Y and $F^{(r)}$ denotes the r-th derivative of the function F. The error term $\Omega_\ell(YF(Y))$ satisfies

$$\mathbb{E}\big[\Omega_\ell\big(YF(Y)\big)\big] \le C_\ell \mathbb{E}\big[|Y|^{\ell+2}\big] \sup_{|t|\le Q} \big|F^{(\ell+1)}(t)\big| + C_\ell \mathbb{E}[|Y|^{\ell+2}\mathbb{1}\big(|Y|>Q\big) \sup_{t\in\mathbb{R}} \big|F^{(\ell+1)}(t)\big|, \tag{3.3}$$

where $Q>0$ is an arbitrary fixed cutoff and C_ℓ satisfies $C_\ell \le (C\ell)^\ell/\ell!$ for some numerical constant C.

For more detail of the actual application of the methods explained in this subsection, we refer to [23,27].

3.2. Weak local semicircle law

The first obstacle we encounter in the proof of the strong local law is the lack of a priori estimates in the expansion. We follow the conventional strategy, developed in [14,15], based on Schur complement formula and self-consistent equations. However, the bound in the weak local semicircle law is not as strong as the one obtained in [15], due to the sparsity, but comparable with the weak local law in [12]. The weak law, Theorem 2.8, is proved in Section 4.

3.3. Strong local law

The main technical difficulty of the proof the strong local law lies in that the entries of H are not identically distributed. For example, if we use the cumulant expansion on the right-hand side of (3.1), we get

$$\begin{aligned}\mathbb{E}[H_{ik}G_{ki}] &= \sum_{r=1}^{\ell} \frac{\kappa_{ik}^{(r+1)}}{r!} \mathbb{E}\big[\partial_{ik}^{(r+1)} G_{ki}\big] + \mathbb{E}\big[\Omega_\ell(H_{ik}G_{ki})\big] \\ &= -\mathbb{E}\big[\kappa_{ik}^{(2)} G_{kk}G_{ii} + \kappa_{ik}^{(2)} G_{ik}^2\big] + \sum_{r=2}^{\ell} \frac{\kappa_{ik}^{(r+1)}}{r!} \mathbb{E}\big[\partial_{ik}^{(r+1)} G_{ki}\big] + \mathbb{E}\big[\Omega_\ell(H_{ik}G_{ki})\big],\end{aligned}$$

where Ω_ℓ is the error term in the generalized Stein's lemma, Lemma 3.1. In the homogeneous case where $\kappa_{ik}^{(2)}$ are identical, the first term in the right-hand side reduces to m^2 after averaging over indices i and k. In our model, however, $\kappa_{ik}^{(2)}$ depends on the choice of i and k and hence the indices do not decouple even after the expansion. If the weak law were good enough so that the error from the substitution of G_{kk} by m is negligible, the analysis might work even with the absence of the decoupling mechanism, but Theorem 2.8 is not enough in that purpose, and moreover, it is believed to be optimal.

In this paper, exploiting the community structure of the model, we decompose the sum into two parts, one for the case where i and k are in the same community and the other where i and k are not in the same community. For each sum, we replace the diagonal entries of the resolvent G by m if the error from the replacement is negligible. If the error is too large, we decompose it again by applying the community structure. The number of the diagonal entries of G increases in each step, and the terms with enough diagonal entries can be handled by the substitution (by m) even with our local law. The detail can be found in Appendix C of [24], especially in Section C.7.

3.4. Tracy–Widom limit and Green function comparison

The derivation of the Tracy–Widom fluctuation of the extremal eigenvalues from the strong local law is now a standard procedure in random matrix theory. In this paper, we follow the approach in [27], based on the Dyson matrix flow. For the sake of completeness, we briefly outline the main ideas of the proof.

To prove the Tracy–Widom fluctuation, we first notice that we can obtain the distribution of the largest eigenvalue of H from the expectation of a function of $\operatorname{Im} m(z)$. Then, we use the Green function comparison method to compare the edge statistics of the centered generalized stochastic block model and generalized Wigner matrix whose first and second moments match with our model. More precisely, for a given centered generalized stochastic block model H_0, we consider the *Dyson matrix flow* with initial condition H_0 defined by

$$H_t := \mathrm{e}^{-t/2} H_0 + \sqrt{1-\mathrm{e}^{-t}}\, W^G, \quad (t \ge 0), \tag{3.4}$$

where W^G is a generalized Gaussian Wigner matrix independent of H_0, with vanishing diagonal entries. The local edge statistics of W^G with vanishing diagonal follows the GOE Tracy–Widom statistics; see Lemma 3.5 of [27] and Theorem 2.7 of [6].

Along the flow, we track the change of the expectation of a function of $\operatorname{Im} m(z)$. Taking the deterministic shift of the edge into consideration, we find that fluctuation of the extremal eigenvalues of H_t do not change over t, which establishes the Tracy–Widom fluctuation for H_0. The analysis along the proof requires the strong local law for the normalized trace of the Green function of H_t, defined as

$$G_t(z) = (H_t)^{-1}, \qquad m_t(z) = \frac{1}{N}\sum_{i=1}^{N} (G_t)_{ii}(z), \quad \bigl(z \in \mathbb{C}^+\bigr). \tag{3.5}$$

We note that H_t is also a balanced generalized sparse random matrix. To check this, let $\kappa_{t,ij}^{(k)}$ be the k-th cumulant of $(H_t)_{ij}$. Then, by the linearity of the cumulants under the addition of independent random variables, we have $\kappa_{t,(\cdot)}^{(1)} = 0$, $\kappa_{t,(\cdot)}^{(2)} \le \frac{C}{N}$ and $\kappa_{t,(\cdot)}^{(k)} = \mathrm{e}^{-kt/2}\kappa_{(\cdot)}^{(k)}$ for $k \ge 3$. In particular, we have the bound

$$\bigl|\kappa_{t,(\cdot)}^{(k)}\bigr| \le \mathrm{e}^{-t}\frac{(Ck)^{ck}}{N q_t^{k-2}}, \quad (k \ge 3), \tag{3.6}$$

where we introduced the time-dependent parameter

$$q_t := q\mathrm{e}^{t/2}, \qquad \zeta_t := \frac{N}{K}\big(\kappa_{t,s}^{(2)} - \kappa_{t,d}^{(2)}\big). \tag{3.7}$$

We also define a polynomial $P_{z,t}$ of m with parameters z and t by

$$\begin{aligned} P_{z,t}(m) :=& \big(1 + zm + m^2 + \mathrm{e}^{-t} q_t^{-2} \xi^{(4)} m^4\big)\big((z + m + \zeta_t m)^2 - \zeta_t\big(1 + mz + m^2\big)\big) \\ =&: P_{1,z,t}(m) P_{2,z,t}(m). \end{aligned} \tag{3.8}$$

We generalize Theorem 2.6 as follows.

Proposition 3.2. *Let H_0 satisfy Assumption 2.1 with $\phi > 0$. Then, for any $t \geq 0$, there exist deterministic number $2 \leq L_t < 3$ and an algebraic function $\widetilde{m}_t : \mathbb{C}^+ \to \mathbb{C}^+$ such that the following hold*:

(1) *The function $\widetilde{m}_t$ is the Stieltjes transform of a deterministic probability measure $\widetilde{\rho}_t$, i.e., $\widetilde{m}_t(z) = m_{\widetilde{\rho}_t}(z)$. The measure ρ_t is supported on $[-L_t, L_t]$ and $\widetilde{\rho}_t$ is absolutely continuous with respect to Lebesgue measure with a strictly positive density on $(-L_t, L_t)$.*
(2) *The function $\widetilde{m}_t \equiv \widetilde{m}_t(z)$, $z \in \mathbb{C}^+$, is a solution to the polynomial equation*

$$\begin{aligned} P_{1,t,z}(\widetilde{m}_t) :=& 1 + z\widetilde{m}_t + \widetilde{m}_t^2 + e^{-t} q_t^{-2} \left(\frac{s_s^{(4)} + (K-1)s_d^{(4)}}{K}\right) \widetilde{m}_t^4 \\ =& 1 + z\widetilde{m}_t + \widetilde{m}_t^2 + \mathrm{e}^{-2t} q^{-2} \xi^{(4)} \widetilde{m}_t^4 = 0. \end{aligned} \tag{3.9}$$

(3) *The normalized trace of the Green function satisfies the local law*

$$\big|m_t(z) - \widetilde{m}_t(z)\big| \prec \frac{1}{q_t^2} + \frac{1}{N\eta}, \tag{3.10}$$

uniformly on the domain $\mathcal{E}$ and uniformly in $t \in [0, 6\log N]$.

Remark 3.3. Several properties of $\widetilde{m}_t$, defined in Proposition 3.2, are crucial in the proof of the Tracy–Widom fluctuation, especially the square-root decay at the edge of the spectrum and the deterministic shift of the edge, where the upper edge of the support of $\widetilde{\rho}_t$ given by

$$L_t = 2 + e^{-t} q_t^{-2} \xi^{(4)} + O\big(e^{-2t} q_t^{-4}\big). \tag{3.11}$$

In Appendix A of [24], we collect some important properties of $\widetilde{m}_t$ and some basic properties of m_{sc}, the Stieltjes transform of the semicircle measure.

4. Proof of weak local law

4.1. Preliminaries

In this section, we prove Theorem 2.8. Unlike strong local semicircle law for cgSBM, we can prove the weak local semicircle law under weaker condition. More precisely, we do not need any assumption about the community structure. Therefore, in this section, we consider generalized sparse random matrices defined as follows.

Assumption 4.1 (Generalized sparse random matrix). Fix sufficiently small $\phi > 0$. A generalized sparse random matrix, $H = (H)_{ij}$ is a symmetric $N \times N$ matrix whose diagonal entries are almost surely zero and whose off-diagonal entries are independent, up to symmetry constraint $H_{ij} = H_{ji}$, random variables. We further assume that each H_{ij} satisfy the moment conditions

$$\mathbb{E}H_{ij} = 0, \qquad \mathbb{E}|H_{ij}|^2 = \sigma_{ij}^2, \qquad \mathbb{E}|H_{ij}|^k \le \frac{(Ck)^{ck}}{Nq^{k-2}}, \quad (k \ge 2), \tag{4.1}$$

with sparsity parameter q satisfying

$$N^{\phi} \le q \le N^{1/2}. \tag{4.2}$$

We further assume the normalize condition that

$$\sum_{i=1}^{N} \sigma_{ij}^2 = 1. \tag{4.3}$$

Recall that

$$\mathcal{D}_\ell := \{ z = E + \mathrm{i}\eta \in \mathbb{C}^+ : |E| < 3, N^{-1+\ell} < \eta \le 3 \}.$$

Throughout this section, we use the factor N^ϵ and allow ϵ to increase by a tiny amount from line to line to absorb numerical constants in the estimates. Moreover, we choose ℓ satisfying $4\phi \le \ell \le 1$ and $\epsilon > 0$ strictly smaller than the fixed parameter $\phi > 0$ appearing in (4.2). If we take ϕ sufficiently small, then Theorem 2.8 states that (2.16), (2.17) and (2.18) holds on $\mathcal{D}_\ell$ for sufficiently small ℓ. In other words we can claim that Theorem 2.8 valid on $\mathcal{D}_\ell$ for any (small) ℓ.

We define the z-dependent quantities

$$v_k := G_{kk} - m_{sc}, \qquad [v] := \frac{1}{N} \sum_{k=1}^{N} v_k = m - m_{sc}.$$

Our goal is to estimate the following quantities,

$$\Lambda_d := \max_k |v_k| = \max_k |G_{kk} - m_{sc}|, \qquad \Lambda_o := \max_{k \ne l} |G_{kl}|, \qquad \Lambda := |m - m_{sc}|. \tag{4.4}$$

Definition 4.2 (Minors). Consider general matrices whose indices lie in subsets of $\{1, \dots, N\}$. For $T \subset \{1, \dots, N\}$ we define $H^{(T)}$ as the $(N - |T|) \times (N - |T|)$ matrix

$$H^{(T)} = (H_{ij})_{i,j \in \{1,\dots,N\} \setminus T}.$$

It is important to keep the original values of the matrix indices in the minor $H^{(T)}$, not to identify $\{1,\dots,N\}\backslash T$ with $\{1,\dots,N-|T|\}$. We set

$$\sum_i^{(T)} := \sum_{i:i\notin T}.$$

If $T=\{a\}$, we abbreviate $(\{a\})$ by (a) in the above definition; similarly, write (ab) instead of $(\{a,b\})$. We also define the Green function of $H^{(T)}$ as

$$G_{ij}^{(T)}(z) := \big(H^{(T)}-z\big)_{ij}^{-1}.$$

Definition 4.3 (Partial expectation). Let $X \equiv X(H)$ be a random variable and $\mathbf{h}_i = (H_{ij})_{j=1}^N$. For $i \in \{1,\dots,N\}$ we define the operations $\mathbb{E}_i$ and $\mathbb{IE}_i$ through

$$\mathbb{E}_i X := \mathbb{E}(X|\mathbf{h}_i), \qquad \mathbb{IE}_i X := X - \mathbb{E}_i X. \tag{4.5}$$

For $T \subset \{1,\dots,N\}$, we introduce the following notations:

$$Z_{ij}^{(T)} := \sum_{k,l}^{T} H_{ik} G_{kl}^{(T)} H_{lj}, \qquad K_{ij}^{(T)} := H_{ij} - z\delta_{ij} - Z_{ij}^{(T)}. \tag{4.6}$$

We abbreviate

$$Z_i := \mathbb{IE}_i Z_{ii}^{(i)} = \mathbb{IE}_i \sum_{k,l}^{(i)} H_{ik} G_{kl}^{(i)} H_{li}. \tag{4.7}$$

The following formulas with these notations were proved in Lemma 4.2 of [14].

Lemma 4.4 (Self-consistent permutation formulas). *For any Hermitian matrix H and $T \subset \{1,\dots,N\}$ the following identities hold. If $i,j,k \notin T$ and $i,j \neq k$, then*

$$G_{ij}^{(T)} = G_{ij}^{(Tk)} + G_{ik}^{(T)} G_{kj}^{(T)} \big(G_{kk}^{(T)}\big)^{-1}. \tag{4.8}$$

If $i,j \notin T$ satisfy $i \neq j$, then

$$G_{ii}^{(T)} = \big(K_{ii}^{(iT)}\big)^{-1} = \big(H_{ii} - z - Z_{ii}^{(iT)}\big)^{-1}, \tag{4.9}$$

$$G_{ij}^{(T)} = -G_{jj}^{(T)} G_{ii}^{(jT)} K_{ij}^{(ijT)} = -G_{ii}^{(T)} G_{jj}^{(iT)} K_{ij}^{(ijT)}. \tag{4.10}$$

We also have the *Wald identity*

$$\sum_j |G_{ij}|^2 = \frac{\operatorname{Im} G_{ii}}{\eta}. \tag{4.11}$$

The following estimate provide the bound for the matrix element of H which follows from the Markov's inequality and the moment conditions.

Lemma 4.5. *We have*

$$|H_{ij}| \prec \frac{1}{q}.$$

4.2. Self-consistent perturbation equations

Following [13], we define the following quantities:

$$A_i := \sigma_{ii}^2 G_{ii} + \sum_{j \neq i} \sigma_{ij}^2 \frac{G_{ij} G_{ji}}{G_{ii}}, \tag{4.12}$$

$$\Upsilon_i := A_i + h_{ii} - Z_i \tag{4.13}$$

and recall the definition of Z_i in (4.7). Using (4.8) and (4.9), we can easily obtain the self-consistent equations for the deviation m_{sc} of the diagonal matrix elements of the Green function;

$$v_i = G_{ii} - m_{sc} = \frac{1}{-z - m_{sc} - (\sum_j \sigma_{ij}^2 v_j - \Upsilon_i)} - m_{sc}. \tag{4.14}$$

Now we define the exceptional (bad) event

$$\mathbf{B} = \mathbf{B}(z) := \{\Lambda_d(z) + \Lambda_o(z) \geq (\log N)^{-2}\}, \tag{4.15}$$

and the control parameter

$$\Psi(z) := \sqrt{\frac{\Lambda(z) + \operatorname{Im} m_{sc}(z)}{N\eta}}. \tag{4.16}$$

On $\mathbf{B}^c$, we have $\Psi(z) \leq CN^{-2\phi}$ by definition of $\mathcal{D}_\ell$. We collect some basic properties of the Green function in the following elementary lemma which were proved in Lemma 3.5 of [15] and Lemma 3.12 of [13].

Lemma 4.6. *Let $\mathbb{T}$ be a subset of $\{1, \dots, N\}$ and $i \notin \mathbb{T}$. Then there exists a constant $C = C_{\mathbb{T}}$ depending on $|\mathbb{T}|$, such that the following hold in $\mathbf{B}^c$*

$$\left|G_{kk}^{(\mathbb{T})} - m_{sc}\right| \leq \Lambda_d + C\Lambda_o^2 \quad \text{for all } k \notin \mathbb{T} \tag{4.17}$$

$$\frac{1}{C} \leq \left|G_{kk}^{(\mathbb{T})}\right| \leq C \quad \text{for all } k \notin \mathbb{T} \tag{4.18}$$

$$\max_{k \neq l} \left|G_{kl}^{(\mathbb{T})}\right| \leq C\Lambda_o \tag{4.19}$$

$$\max_i |A_i| \leq \frac{C}{N} + C\Lambda_o^2 \tag{4.20}$$

for any fixed $|\mathbb{T}|$ and for sufficiently large N.

We note that all quantities depend on the spectral parameter z and the estimates are uniform in $z = E + \mathrm{i}\eta$.

4.3. Estimate of the exceptional events and analysis of the self-consistent equation

We introduce three lemmas which estimate some exceptional events. Their proofs are given in Appendix B.1 of [24].

Lemma 4.7. *For fixed $z \in D_\ell$ and any small $\epsilon > 0$, we have on $\mathbf{B}^c$ with high probability*

$$\Lambda_o \le C\left(\frac{N^\epsilon}{q} + N^\epsilon \Psi\right). \tag{4.21}$$

Lemma 4.8. *For any $z \in D_\ell$, we have on $\mathbf{B}^c$ with high probability*

$$|Z_i| \le C\left(\frac{N^\epsilon}{q} + N^\epsilon \Psi\right), \tag{4.22}$$

$$\left|Z_{(ij)}^{ij}\right| \le C\left(\frac{N^\epsilon}{q} + N^\epsilon \Psi\right) \quad (i \ne j). \tag{4.23}$$

Lemma 4.9. *For any $z \in D_\ell$, we have on $\mathbf{B}^c$ with high probability*

$$|\Upsilon_i| \le C\left(\frac{N^\epsilon}{q} + N^\epsilon \Psi\right). \tag{4.24}$$

We define the events

$$\begin{aligned}
\Omega_h &:= \left\{\max_{1\le i,j\le N} |H_{ij}| \ge \frac{N^\epsilon}{q}\right\} \cup \left\{\left|\sum_{i=1}^N H_{ii}\right| \ge N^\epsilon\left(\frac{1}{q} + 1\right)\right\}, \\
\Omega_d &:= \left\{\max_i |Z_i| \ge C\left(\frac{N^\epsilon}{q} + N^\epsilon \Psi\right)\right\}, \\
\Omega_o &:= \left\{\max_{i\ne j} \left|Z_{ij}^{(ij)}\right| \ge C\left(\frac{N^\epsilon}{q} + N^\epsilon \Psi\right)\right\},
\end{aligned} \tag{4.25}$$

and let

$$\Omega(z) := \Omega_h \cup \left[(\Omega_d \cup \Omega_o) \cap \mathbf{B}^c\right]. \tag{4.26}$$

Then by (4.22), (4.23) and the large deviation estimate (B.2) we can show that Ω^c holds with high probability.

From (4.14), we obtain the following equation for v_i

$$v_i = m_{sc}^2\Big(\sum_j \sigma_{ij}^2 v_j - \Upsilon_i\Big) + m_{sc}^3\Big(\sum_j \sigma_{ij}^2 v_j - \Upsilon_i\Big)^2 + O\Big(\sum_j \sigma_{ij}^2 v_j - \Upsilon_i\Big)^3. \tag{4.27}$$

By assumption, $\sum_j \sigma_{ij}^2 = 1$, $\mathbf{e} = (1, 1, \dots, 1)$ is the unique eigenvector of $B = (\sigma_{ij})$ with simple eigenvalue 1. Define the parameter

$$g = g(z) := \max\big\{\delta_+, \big|1 - \operatorname{Re} m_{sc}^2(z)\big|\big\}. \tag{4.28}$$

and we recall the following basic lemma that was proven in Lemma 4.8 of [13].

Lemma 4.10. *The matrix $I - m_{sc}^2(z)B$ is invertible on the subspace orthogonal to $\mathbf{e}$. let $\mathbf{u}$ be a vector which is orthogonal to $\mathbf{e}$ and let*

$$\mathbf{w} = \big(I - m_{sc}^2(z)B\big)\mathbf{u},$$

then

$$\|\mathbf{u}\|_\infty \le \frac{C \log N}{g(z)} \|\mathbf{w}\|_\infty$$

for some constant C that only depends on δ_-.

We introduce the following lemma which estimates the deviation of v_i from its average $[v]$.

Lemma 4.11. *Fix $z \in D_\ell$. If in some set Ξ it holds that*

$$\Lambda_d \le \frac{q}{(\log N)^{3/2}}, \tag{4.29}$$

then in the set $\Xi \cap \mathbf{B}^c$, we have

$$\begin{aligned} \max_i \big|v_i - [v]\big| &\le \frac{C \log N}{g}\bigg(\Lambda^2 + \frac{N^\epsilon}{q} + N^\epsilon \Psi + \frac{(\log N)^2}{g^2}\Big(\frac{N^\epsilon}{q} + N^\epsilon \Psi\Big)^2\bigg) \\ &\le \frac{C \log N}{g^3}\Big(\Lambda^2 + \frac{N^\epsilon}{q} + N^\epsilon \Psi\Big), \end{aligned} \tag{4.30}$$

with high probability.

With Lemma 4.11, we can show the following.

Lemma 4.12. *Fix $z \in D_\ell$. Define $[Z] := N^{-1} \sum_{i=1}^N Z_i$. Then in the set $\mathbf{B}^c$ we have*

$$\big(1 - m_{sc}^2\big)[v] = m_{sc}^3 [v]^2 + m_{sc}^2 [Z] + O\Big(\frac{\Lambda^2}{\log N}\Big) + O\bigg((\log N)^3 \Big(\frac{N^\epsilon}{q} + N^\epsilon \Psi\Big)^2\bigg), \tag{4.31}$$

with high probability.

Proofs of Lemmas 4.11 and 4.12 are given in Appendices B.2 and B.3, respectively.

4.4. Dichotomy estimate for Λ and continuity argument

In $\mathbf{B}^c$, $\max_i |Z_i| = O(\frac{N^\epsilon}{q} + N^\epsilon \Psi)$ holds with high probability. Therefore using $\frac{N^\epsilon}{q} + N^\epsilon\Psi \le N^{-\epsilon}$, with high probability, we have

$$(1 - m_{sc}^2)[v] = m_{sc}^3[v]^2 + O\left(\frac{\Lambda^2}{\log N}\right) + O\left(\frac{N^\epsilon}{q} + N^\epsilon \Psi\right). \tag{4.32}$$

Lemma 4.13. *Let $\eta \ge 2$. Then for $z \in D_\ell$ we have*

$$\Lambda_d(z) + \Lambda_o(z) \le C\left(\frac{N^\epsilon}{q} + \frac{N^\epsilon}{\sqrt{N}}\right) \le (\log N)^{-2}. \tag{4.33}$$

Lemma 4.13 is an initial estimates on Λ_d and Λ_o for large $\eta \sim 1$ to get the continuity argument started.

We further introduce the following notations

$$\alpha := \left|\frac{1 - m_{sc}^2}{m_{sc}^3}\right|, \qquad \beta := \frac{N^\epsilon}{\sqrt{q}} + \frac{N^\epsilon}{(N\eta)^{1/3}}, \tag{4.34}$$

where α and β depend on the parameter z. For any $z \in D_\ell$ we have the bound $\beta \le N^{-\frac{1}{2}\phi}$.

From Lemma A.1, it follows that for any $z \in D_\ell$ there is a constant $K \ge 1$ such that

$$\frac{1}{K}\sqrt{\kappa + \eta} \le \alpha(z) \le K\sqrt{\kappa + \eta}. \tag{4.35}$$

Since $\sqrt{\kappa+\eta}$ is increasing and $\beta(E + i\eta)$ is decreasing in η, we know that, for any fixed E and $U > 1$, $\sqrt{\kappa + \eta} = 2U^2 K\beta(E + i\eta)$ has a unique solution $\tilde{\eta} = \tilde{\eta}(U, E)$ which satisfies $\tilde{\eta} \ll 1$.

Lemma 4.14 (Dichotomy). *There exist a constant U_0 such that for any fixed $U \ge U_0$, there exists constant $C_1(U)$ such that the following hold for any $z \in D_\ell$.*

$$\Lambda(z) \le U\beta(z) \quad \text{or} \quad \Lambda(z) \ge \frac{\alpha(z)}{U} \quad \text{if } \eta \ge \tilde{\eta}(U, E) \tag{4.36}$$

$$\Lambda(z) \le C_1(U)\beta(z) \quad \text{if } \eta < \tilde{\eta}(U, E) \tag{4.37}$$

on $\mathbf{B}^c(z)$ with high probability.

Proofs of Lemma 4.13 and 4.14 are given in B.4.

Now choose a decreasing finite sequence $\eta_k \in D_\ell$, $k = 1, 2, \ldots, k_0$, with $k_0 \le CN^8$, $|\eta_k - \eta_{k+1}| \le N^{-8}$, $\eta_1 = 2$, and $\eta_{k_0} = N^{-1+l}$. Fix E with $|E| \le 3$ and set $z_k = E + i\eta_k$. We fix $U \ge U_0$ and recall the definition of $\tilde{\eta}$ from Lemma 4.14.

Consider the first case of z_1. For large N, it is easy to show that $\eta_1 \geq \tilde{\eta}$ for any $|E| \leq 3$. Therefore Lemma 4.13 and Lemma 4.14 imply that $\mathbf{B}^c(z_1)$ and $\Lambda(z_1) \leq U\beta(z_1)$ hold with high probability. For general k we have the following:

Lemma 4.15. *Define* $\Omega_k := \mathbf{B}^c(z_k) \cap \{\Lambda(z_k) \leq C^{(k)}(U)\beta(z_k)\}$ *where*

$$C^{(k)}(U) = \begin{cases} U & \text{if } \eta_k \geq \tilde{\eta}(U, E) \\ C_1(U) & \text{if } \eta_k \leq \tilde{\eta}(U, E). \end{cases} \tag{4.38}$$

Then

$$\mathbb{P}\big(\Omega_k^c\big) \leq 2kN^{-D}. \tag{4.39}$$

Proof of the Lemma 4.15 is given in Appendix B.5 of [24]. Now we are ready to prove our main theorem.

Proof of Theorem 2.8. Take a lattice $L \subset D_\ell$ such that $|L| \leq CN^6$ and for any $z \in D_\ell$ there exist $\tilde{z} \in L$ satisfying $|z - \tilde{z}| \leq N^{-3}$. From the Lipschitz continuity of the map $z \mapsto G_{ij}(z)$ and $z \mapsto m_{sc}(z)$ with a Lipschitz constant bounded by $\eta^{-2} \leq N^2$, we have

$$\big|G_{ij}(z) - G_{ij}(\tilde{z})\big| \leq \frac{|z - \tilde{z}|}{\eta^2} \leq \frac{1}{N}. \tag{4.40}$$

We also have

$$\big|m(z) - m(\tilde{z})\big| \leq \frac{|z - \tilde{z}|}{\eta^2} \leq \frac{1}{N}. \tag{4.41}$$

By Lemma 4.15, we have for some large constant C

$$\mathbb{P}\left[\bigcap_{\tilde{z} \in L} \big\{\Lambda(\tilde{z}) \leq C\beta(\tilde{z})\big\}\right] \geq 1 - N^{-D}. \tag{4.42}$$

Hence with (4.40) and $\beta \gg N^{-1}$ we find

$$\mathbb{P}\left[\bigcup_{z \in D_\ell} \big\{\Lambda(z) > C\beta(z)\big\}\right] \leq N^{-D}, \tag{4.43}$$

for some constant C. Using similar argument, we can also get

$$\mathbb{P}\left[\bigcap_{z \in D_\ell} \mathbf{B}^c(z)\right] \leq N^{-D}. \tag{4.44}$$

In other words, we proved (2.18). Using (4.21), (4.30) and (4.39) with similar lattice arguments, we can conclude the proof of Theorem 2.8. □

5. Proof of Proposition 3.2 and Theorem 2.12

5.1. Proof of Proposition 3.2

In this section, we prove Proposition 3.2. Recall the subdomain $\mathcal{D}_\ell$ of $\mathcal{E}$ and the definition of the polynomial $P \equiv P_{t,z}$, $P_1 \equiv P_{1,t,z}$ and $P_2 \equiv P_{2,t,z}$ in (3.8). We have the following lemma called recursive moment estimate.

Lemma 5.1 (Recursive moment estimate). *Fix $\phi > 0$ and fix any $t \geq 0$. Let H_0 satisfies Assumption 2.1. Then, for any $D > 10$ and small $\epsilon > 0$, the normalized trace of the Green function, $m_t \equiv m_t(z)$, of the matrix H_t satisfies*

$$\begin{aligned}
&\mathbb{E}\big|P(m_t)\big|^{2D} \\
&\quad \leq N^{\epsilon}\mathbb{E}\Bigg[\bigg(\frac{1}{q_t^4} + \frac{\operatorname{Im} m_t}{N\eta}\bigg)\big|P(m_t)\big|^{2D-1}\Bigg] + N^{-\epsilon/8}q_t^{-1}\mathbb{E}\big[|m_t - \widetilde{m}_t|^2\big|P(m_t)\big|^{2D-1}\big] \\
&\qquad + N^{\epsilon}q_t^{-1}\sum_{s=2}^{2D}\sum_{s'=0}^{s-2}\mathbb{E}\Bigg[\bigg(\frac{\operatorname{Im} m_t}{N\eta}\bigg)^{2s-s'-2}\big|P'(m_t)\big|^{s'}\big|P(m_t)\big|^{2D-s}\Bigg] \\
&\qquad + N^{\epsilon}\sum_{s=2}^{2D}\mathbb{E}\Bigg[\bigg(\frac{1}{N\eta} + \frac{1}{q_t}\bigg(\frac{\operatorname{Im} m_t}{N\eta}\bigg)^{1/2} + \frac{1}{q_t^2}\bigg)\bigg(\frac{\operatorname{Im} m_t}{N\eta}\bigg)^{s-1}\big|P'(m_t)\big|^{s-1}\big|P(m_t)\big|^{2D-s}\Bigg] \\
&\qquad + N^{\epsilon}q_t^{-8D},
\end{aligned} \tag{5.1}$$

uniformly on the domain $\mathcal{D}_\ell$, for sufficiently large N.

We give the detailed proof of Lemma 5.1 in Appendix C of [24]. In this section, we only sketch the idea of the proof. We estimate the expectation of $|P(m_t)|^{2D}$ using Lemma 3.1. For example, consider the cumulant expansion of $1 + zm$, a part of $P_{1,t,z}(m)$, computed by

$$\mathbb{E}[1 + zm] = \mathbb{E}\Bigg[\frac{1}{N}\sum_{i=1}^{N}(1 + zG_{ii})\Bigg] = \mathbb{E}\Bigg[\frac{1}{N}\sum_{i=1}^{N}(HG)_{ii}\Bigg] = \mathbb{E}\Bigg[\frac{1}{N}\sum_{i,j}(H_{ji}G_{ji})\Bigg],$$

where we used the definition of Green function to get the second equation. Then by Lemma 3.1, we get

$$\begin{aligned}
\mathbb{E}\Bigg[\frac{1}{N}\sum_{i,j}(H_{ji}G_{ji})\Bigg] &= \frac{1}{N}\sum_{i\neq j}\kappa_{ij}^{(2)}\mathbb{E}\big[(\partial_{ij}G_{ij})\big] \\
&= -\frac{1}{N}\mathbb{E}\Bigg[\sum_{i,j}\kappa_{ij}^{(2)}G_{ii}G_{jj}\Bigg] + \frac{1}{N}\mathbb{E}\Bigg[\sum_{i}\kappa_{ii}^{(2)}G_{ii}^2\Bigg] - \frac{1}{N}\mathbb{E}\Bigg[\sum_{i\neq j}\kappa_{ij}^{(2)}G_{ij}^2\Bigg],
\end{aligned}$$

and it can be easily shown that the terms containing at least one off-diagonal Green function entries are sufficiently small. Thus, the main order term we need to estimate is

$$\frac{1}{N}\mathbb{E}\Big[\sum_{i,j}\kappa_{ij}^{(2)}G_{ii}G_{jj}\Big]=\frac{1}{N}\mathbb{E}\Big[\sum_{i,j}\kappa_{d}^{(2)}G_{ii}G_{jj}\Big]+\frac{1}{N}\mathbb{E}\Big[\sum_{i\sim j}\big(\kappa_{s}^{(2)}-\kappa_{d}^{(2)}\big)G_{ii}G_{jj}\Big]$$
$$=\mathbb{E}\big[(1-\zeta)m^2\big]+\frac{\zeta K}{N^2}\mathbb{E}\Big[\sum_{i\sim j}G_{ii}G_{jj}\Big].$$

To compute with sufficiently small error, using Lemma 3.1 again, we multiply z to the second term to obtain

$$\frac{\zeta K}{N^2}\mathbb{E}\Big[\sum_{i\sim j}zG_{ii}G_{jj}\Big]=\frac{\zeta K}{N^2}\mathbb{E}\Big[\sum_{i\sim j}\Big(\sum_{k}H_{ik}G_{ki}-1\Big)G_{jj}\Big]=\sum_{r=1}^{l}\frac{1}{r!}\mathbb{E}J_r-\zeta\mathbb{E}[m]+O(\Phi_\epsilon),$$

where

$$J_r=\frac{\zeta K}{N^2}\sum_{i\sim j}\sum_{k\neq i}\kappa_{ik}^{(r+1)}\mathbb{E}\big[\big(\partial_{ik}^{r}G_{ik}G_{jj}\big)\big],$$

and $O(\Phi_\epsilon)$ is a sufficiently small error term defined by the right side of (5.1). One of the main order term which only consists of the diagonal entries of the Green function is

$$\mathbb{E}J_1=\frac{\zeta K}{N^2}\sum_{i\sim j}\sum_{k}\kappa_{d}^{(2)}\mathbb{E}\big[(G_{ii}G_{jj}G_{kk})\big]-\frac{\zeta K}{N^2}\sum_{i\sim j\sim k}\big(\kappa_{s}^{(2)}-\kappa_{d}^{(2)}\big)\mathbb{E}\big[(G_{ii}G_{jj}G_{kk})\big]. \tag{5.2}$$

The first term of the right-hand side of (5.2) can be estimated by

$$-\frac{\zeta K}{N^2}\sum_{i\sim j}\sum_{k\neq i}\kappa_{d}^{(2)}\mathbb{E}\big[(G_{ii}G_{jj}G_{kk})\big]=-\frac{\zeta(1-\zeta)K}{N^2}\sum_{i\sim j}\mathbb{E}[mG_{ii}G_{jj}].$$

Thus, if we can estimate the second term of the right-hand side of (5.2) with sufficiently small error, then we get the good estimation for

$$\frac{\zeta K}{N^2}\mathbb{E}\Big[\sum_{i\sim j}zG_{ii}G_{jj}\Big]+\frac{\zeta(1-\zeta)K}{N^2}\sum_{i\sim j}\mathbb{E}[mG_{ii}G_{jj}]=\frac{\zeta K}{N^2}\mathbb{E}\Big[\big(z+(1-\zeta)m\big)\sum_{i\sim j}G_{ii}G_{jj}\Big].$$

Still it is not easy to handle the second term of the right-hand side of (5.2) due to its community structure. We abbreviate

$$\hat{J}:=\frac{\zeta K}{N^2}\sum_{i\sim j\sim k}\big(\kappa_{s}^{(2)}-\kappa_{d}^{(2)}\big)\mathbb{E}\big[(zG_{ii}G_{jj}G_{kk})\big],$$

and use Lemma 3.1 once more. With some expansions and calculations, one can get the estimation for $\hat{J}$ and with small error terms. We apply a similar strategy to the expectation of $|P(m_t)|^{2D}$ and we can obtain the recursive moment estimate stated in (5.1).

With Lemma 5.1, we can prove Proposition 3.2.

Proof of Proposition 3.2 and Theorem 2.6. Fix $t \in [0, 6\log N]$. Let $\widetilde{m}_t$ be the solution w_t of the equation $P_{1,t,z}(w_t) = 0$. One can show the first two parts directly from the properties of $\widetilde{m}_t$ and its Stieltjes inversion ρ_t; see Appendix A of [24]. It remains to prove the third part of the proposition. Since $|m_t| \sim 1$, $|\widetilde{m}_t| \sim 1$, there exist positive constants c_1 and c_2 satisfying

$$\frac{1}{c_1} \le P_2(\widetilde{m}_t) \le c_1, \qquad \frac{1}{c_2} + O\left(\frac{1}{q^2}\right) \le P''(\widetilde{m}_t) \le c_2 + O\left(\frac{1}{q^2}\right). \tag{5.3}$$

We introduce the following z- and t-dependent deterministic parameters

$$\alpha_1(z) := \operatorname{Im} \widetilde{m}_t(z), \qquad \alpha_2(z) := P'\big(\widetilde{m}_t(z)\big), \qquad \beta := \frac{1}{N\eta} + \frac{1}{q_t^2}, \tag{5.4}$$

with $z = E + \mathrm{i}\eta$. We note that

$$|\alpha_2| \ge \big|P_2(\widetilde{m}_t)\big|\big|P_1'(\widetilde{m}_t)\big| \ge \frac{1}{c_1} \operatorname{Im} P_1'(\widetilde{m}_t) \ge \frac{1}{c_1} \operatorname{Im} \widetilde{m}_t = \frac{1}{c_1}\alpha_1$$

Further let

$$\Lambda_t(z) := \big|m_t(z) - \widetilde{m}_t(z)\big|, \quad \big(z \in \mathbb{C}^+\big). \tag{5.5}$$

Note that from weak local law for the cgSBM (2.18), we have that $\Lambda_t(z) \prec 1$ uniformly on $\mathcal{D}_\ell$. Since $P_1(\widetilde{m}_t) = 0$, we have $P'(\widetilde{m}_t) = P_1'(\widetilde{m}_t)P_2(\widetilde{m}_t)$. Similar as in the proof of Lemma 5.1 of [27], we have

$$|\alpha_2| = \big|P'(\widetilde{m}_t)\big| = \big|P_1'(\widetilde{m}_t)\|P_2(\widetilde{m}_t)\big| \sim \sqrt{\kappa_t + \eta}.$$

Recall that Young's inequality states that for any $a, b > 0$ and $x, y > 1$ with $x^{-1} + y^{-1} = 1$,

$$ab \le \frac{a^x}{x} + \frac{b^y}{y}. \tag{5.6}$$

Let $D \ge 10$. Choose any small $\epsilon > 0$. The strategy is now as follows. We apply Young's inequality (5.6) to split up all the terms on the right-hand side of (5.1) and absorb resulting factors of $\mathbb{E}|P(m_t)|^{2D}$ into the left hand side. For the first term on the right of (5.1), we get, upon using applying (5.6) with $x = 2D$ and $y = 2D/(2D-1)$, that

$$\begin{aligned} &N^{\epsilon}\left(\frac{\operatorname{Im} m_t}{N\eta} + q_t^{-4}\right)\left|P(m_t)\right|^{2D-1} \\ &\quad \le N^{\epsilon}\frac{\alpha_1 + \Lambda_t}{N\eta}\left|P(m_t)\right|^{2D-1} + N^{\epsilon}q_t^{-4}\left|P(m_t)\right|^{2D-1} \\ &\quad \le \frac{N^{(2D+1)\epsilon}}{2D}\beta^{2D}(\alpha_1 + \Lambda_t)^{2D} + \frac{N^{(2D+1)\epsilon}}{2D}q_t^{-8D} + \frac{2(2D-1)}{2D}N^{-\frac{\epsilon}{2D-1}}\left|P(m_t)\right|^{2D}, \end{aligned} \tag{5.7}$$

since $(N\eta)^{-1} \le \beta$ and note that the last term can be absorbed into the left-hand side of (5.1). The same idea can be applied to the second term on the right-hand side of (5.1). Hence, we have

$$N^{-\epsilon/8}q_t^{-1}\Lambda_t^2\left|P(m_t)\right|^{2D-1} \le \frac{N^{-(D/4-1)\epsilon}}{2D}q_t^{-2D}\Lambda_t^{4D} + \frac{2D-1}{2D}N^{-\frac{\epsilon}{2D-1}}\left|P(m_t)\right|^{2D}. \tag{5.8}$$

To handle the other terms, we Taylor expand $P'(m_t)$ around $\widetilde{m}_t$ as

$$\left|P'(m_t) - \alpha_2 - P''(\widetilde{m}_t)(m_t - \widetilde{m}_t)\right| \le C\Lambda_t^2. \tag{5.9}$$

Therefore, for some constant C_1, we get

$$\left|P'(m)\right| \le |\alpha_2| + C_1\Lambda_t, \tag{5.10}$$

for all $z \in \mathcal{D}_\ell$, with high probability. Note that for any fixed $s \ge 2$,

$$\begin{aligned} (\alpha_1 + \Lambda_t)^{2s-s'-2}\left(|\alpha_2| + C_1\Lambda_t\right)^{s'} &\le N^{\epsilon/2}(\alpha_1 + \Lambda_t)^{s-1}\left(|\alpha_2| + C_1\Lambda_t\right)^{s-1} \\ &\le N^{\epsilon}(\alpha_1 + \Lambda_t)^{s/2}\left(|\alpha_2| + C_1\Lambda_t\right)^{s/2} \end{aligned}$$

with high probability, uniformly in $\mathcal{D}_\ell$, since $\alpha_1 \le c_1|\alpha_2| \le C$ and $\Lambda_t \prec 1$. Also note that $2s - s' - 2 \ge s$ since $s' \le s - 2$. Therefore for the second line in (5.1), for $2 \le s \le 2D$,

$$\begin{aligned} &N^{\epsilon}q_t^{-1}\left(\frac{\operatorname{Im} m_t}{N\eta}\right)^{2s-s'-2}\left|P'(m_t)\right|^{s'}\left|P(m_t)\right|^{2D-s} \\ &\quad \le N^{\epsilon}q_t^{-1}\beta^s(\alpha_1 + \Lambda_t)^{2s-s'-2}\left(|\alpha_2| + C_1\Lambda_t\right)^{s'}\left|P(m_t)\right|^{2D-s} \\ &\quad \le N^{2\epsilon}q_t^{-1}\beta^s(\alpha_1 + \Lambda_t)^{s/2}\left(|\alpha_2| + C_1\Lambda_t\right)^{s/2}\left|P(m_t)\right|^{2D-s} \\ &\quad \le N^{2\epsilon}q_t^{-1}\frac{s}{2D}\beta^{2D}(\alpha_1 + \Lambda_t)^{D}\left(|\alpha_2| + C_1\Lambda_t\right)^{D} + N^{2\epsilon}q_t^{-1}\frac{2D-s}{2D}\left|P(m_t)\right|^{2D} \end{aligned} \tag{5.11}$$

uniformly on $\mathcal{D}_\ell$ with high probability. Similarly, for the last term in (5.1), for $2 \le s \le 2D$, we obtain

$$
\begin{aligned}
&N^{\epsilon}\left(\frac{1}{N\eta}+\frac{1}{q_t}\left(\frac{\operatorname{Im} m_t}{N\eta}\right)^{1/2}+\frac{1}{q_t^2}\right)\left(\frac{\operatorname{Im} m_t}{N\eta}\right)^{s-1}\left|P'(m_t)\right|^{s-1}\left|P(m_t)\right|^{2D-s}\\
&\quad\le N^{2\epsilon}\beta\cdot\beta^{s-1}(\alpha_1+\Lambda_t)^{s/2}\big(|\alpha_2|+C_1\Lambda_t\big)^{s/2}\left|P(m_t)\right|^{2D-s}\\
&\quad\le \frac{s}{2D}\big(N^{2\epsilon}N^{\frac{(2D-s)\epsilon}{4D^2}}\big)^{\frac{2D}{s}}\beta^{2D}(\alpha_1+\Lambda_t)^{D}\big(|\alpha_2|+C_1\Lambda_t\big)^{D}\\
&\qquad+\frac{2D-s}{2D}\big(N^{-\frac{(2D-s)\epsilon}{4D^2}}\big)^{\frac{2D}{2D-s}}\left|P(m_t)\right|^{2D}\\
&\quad\le N^{(2D+1)\epsilon}\beta^{2D}(\alpha_1+\Lambda_t)^{D}\big(|\alpha_2|+C_1\Lambda_t\big)^{D}+N^{-\frac{\epsilon}{2D}}\left|P(m_t)\right|^{2D},
\end{aligned}
\tag{5.12}
$$

uniformly on $\mathcal{D}_\ell$ with high probability where we used

$$
\frac{1}{N\eta}+\frac{1}{q_t}\left(\frac{\operatorname{Im} m_t}{N\eta}\right)^{1/2}+\frac{1}{q_t^2}\prec\beta.
\tag{5.13}
$$

From (5.1), (5.7), (5.8), (5.11) and (5.12)

$$
\begin{aligned}
\mathbb{E}\big[\left|P(m_t)\right|^{2D}\big]&\le N^{(2D+1)\epsilon}\mathbb{E}\big[\beta^{2D}(\alpha_1+\Lambda_t)^{D}\big(|\alpha_2|+C_1\Lambda_t\big)^{D}\big]+\frac{N^{(2D+1)\epsilon}}{2D}q_t^{-8D}\\
&\quad+\frac{N^{-(D/4-1)\epsilon}}{2D}q_t^{-2D}\mathbb{E}\big[\Lambda_t^{4D}\big]+CN^{-\frac{\epsilon}{2D}}\mathbb{E}\big[\left|P(m_t)\right|^{2D}\big],
\end{aligned}
\tag{5.14}
$$

for all $z\in\mathcal{D}_\ell$. Since the last term can be absorbed into the left-hand side, we eventually find

$$
\begin{aligned}
&\mathbb{E}\big[\left|P(m_t)\right|^{2D}\big]\\
&\quad\le CN^{(2D+1)\epsilon}\mathbb{E}\big[\beta^{2D}(\alpha_1+\Lambda_t)^{D}\big(|\alpha_2|+C_1\Lambda_t\big)^{D}\big]\\
&\qquad+C\frac{N^{(2D+1)\epsilon}}{2D}q_t^{-8D}+C\frac{N^{-(D/4-1)\epsilon}}{2D}q_t^{-2D}\mathbb{E}\big[\Lambda_t^{4D}\big]\\
&\quad\le N^{3D\epsilon}\beta^{2D}|\alpha_2|^{2D}+N^{3D\epsilon}\beta^{2D}\mathbb{E}\big[\Lambda_t^{2D}\big]+N^{3D\epsilon}q_t^{-8D}+N^{-D\epsilon/8}q_t^{-2D}\mathbb{E}\big[\Lambda_t^{4D}\big],
\end{aligned}
\tag{5.15}
$$

uniformly on $\mathcal{D}_\ell$, where we used $\alpha_1 \le c_1|\alpha_2|$ and the inequality

$$
(a+b)^x \le 2^{x-1}\big(a^x+b^x\big)
\tag{5.16}
$$

for any $a, b \ge 0$ and $x \ge 1$ with $D > 10$ to get the last line.

Now, we aim to control Λ_t in terms of $|P(m_t)|$. For that, from the third order Taylor expansion of $P(m_t)$ around $\widetilde{m}_t$ to get

$$
\left|P(m_t)-P'(\widetilde{m}_t)(m_t-\widetilde{m}_t)-\frac{1}{2}P''(\widetilde{m}_t)(m_t-\widetilde{m}_t)^2\right|\le C\Lambda_t^3,
\tag{5.17}
$$

since $P(\widetilde{m}_t)=0$ and $P'''(\widetilde{m}_t)\sim 1$. Then using $\Lambda_t \prec 1$ and $P''(\widetilde{m}_t)\geq C+O(q_t^{-2})$ we obtain

$$\Lambda_t^2 \prec |\alpha_2|\Lambda_t + \big|P(m_t)\big|, \quad (z\in\mathcal{D}_\ell). \tag{5.18}$$

Taking the $2D$-power of (5.18), using (5.16) again, and taking the expectation, we get

$$\mathbb{E}\big[\Lambda_t^{2D}\big] \leq 4^{2D}N^{\epsilon/2}|\alpha_2|^{2D}\mathbb{E}\big[\Lambda_t^{2D}\big] + 4^{2D}N^{\epsilon/2}\mathbb{E}\big[\big|P(m_t)\big|^{2D}\big] \tag{5.19}$$

Replacing form (5.15) for $\mathbb{E}[|P(m_t)|^{2D}]$, for sufficiently large N, we obtain

$$\begin{aligned}\mathbb{E}\big[\Lambda_t^{4D}\big] \leq{}& N^{\epsilon}|\alpha_2|^{2D}\mathbb{E}\big[\Lambda_t^{2D}\big] + N^{(3D+1)\epsilon}\beta^{2D}|\alpha_2|^{2D} + N^{(3D+1)\epsilon}\beta^{2D}\mathbb{E}\big[\Lambda_t^{2D}\big] \\ &+ N^{(3D+1)\epsilon}q_t^{-8D} + N^{-D\epsilon/8+\epsilon}q_t^{-2D}\mathbb{E}\big[\Lambda_t^{4D}\big],\end{aligned} \tag{5.20}$$

uniformly on $\mathcal{D}_\ell$. Using Schwarz inequality for the first term and the third term on the right, absorbing the terms $o(1)\mathbb{E}[\Lambda_t^{4D}]$ into the left side and using (5.13) we get

$$\mathbb{E}\big[\Lambda_t^{4D}\big] \leq N^{2\epsilon}|\alpha_2|^{4D}N^{(3D+2)\epsilon}\beta^{2D}\big|\alpha_2^{2D}\big| + N^{(3D+2)\epsilon}\beta^{4D}, \tag{5.21}$$

uniformly on $\mathcal{D}_\ell$. This estimate can be fed back into (5.15), to get the bound

$$\begin{aligned}\mathbb{E}\big[\big|P(m_t)\big|^{2D}\big] &\leq N^{3D\epsilon}\beta^{2D}|\alpha_2|^{2D} + N^{3D\epsilon}\beta^{4D}\mathbb{E}\big[\Lambda^{2D}\big] + N^{(3D+1)\epsilon}\beta^{4D} + q_t^{-2D}|\alpha_2|^{4D} \\ &\leq N^{5D\epsilon}\beta^{2D}|\alpha_2|^{2D} + N^{5D\epsilon}\beta^{4D} + q_t^{-2D}|\alpha_2|^{4D},\end{aligned} \tag{5.22}$$

uniformly on $\mathcal{D}_\ell$ for sufficiently large N.

For any fixed $z\in\mathcal{D}_\ell$, Markov's inequality then yields $|P(m_t)|\prec|\alpha_2|\beta+\beta^2+q_t^{-1}|\alpha_2|^2$. Then we can obtain from the Taylor expansion of $P(m_t)$ around $\widetilde{m}_t$ in (5.17) that

$$\left|\alpha_2(m_t-\widetilde{m}_t)+\frac{P''(\widetilde{m})}{2}(m_t-\widetilde{m}_t)^2\right| \prec \psi\Lambda_t^2 + |\alpha_2|\beta+\beta^2+q_t^{-1}|\alpha_2|^2, \tag{5.23}$$

for each fixed $z\in\mathcal{D}_\ell$, where ψ, defined in Theorem 2.8, satisfies $\psi\geq q_t^{-2}$. Uniformity in z is easily achieved using a lattice argument and the Lipschitz continuity of $m_t(z)$ and $\widetilde{m}_t(z)$ on $\mathcal{D}_\ell$. Furthermore, for any (small) $\epsilon>0$ and (large) D there is an event $\widetilde{\Xi}$ with $\mathbb{P}(\widetilde{\Xi})\geq 1-N^D$ such that for all $z\in\mathcal{D}_\ell$,

$$\left|\alpha_2(m_t-\widetilde{m}_t)+\frac{P''(\widetilde{m})}{2}(m_t-\widetilde{m}_t)^2\right| \leq N^\epsilon\psi\Lambda_t^2 + N^\epsilon|\alpha_2|\beta+N^\epsilon\beta^2+N^\epsilon q_t^{-1}|\alpha_2|^2, \tag{5.24}$$

on $\widetilde{\Xi}$, for N sufficiently large.

Recall that there exists a constant $C_0>1$ which satisfies $C_0^{-1}\sqrt{\kappa_t(E)+\eta}\leq|\alpha_2|\leq C_0\sqrt{\kappa_t(E)+\eta}$, where we can choose C_0 uniform in $z\in\mathcal{D}_\ell$. Note that, for a fixed E, $\beta=\beta(E+\mathrm{i}\eta)$ is a decreasing function of η whereas $\sqrt{\kappa_t(E)+\eta}$ is increasing. Hence there is $\widetilde{\eta}_0\equiv\widetilde{\eta}_0(E)$ such that $\sqrt{\kappa(E)+\widetilde{\eta}_0}=C_0q_t\beta(E+\mathrm{i}\widetilde{\eta}_0)$. We consider the subdomain $\widetilde{\mathcal{D}}\subset\mathcal{D}_\ell$ defined by

$$\widetilde{\mathcal{D}} := \big\{z=E+\mathrm{i}\eta\in\mathcal{D}_\ell : \eta>\widetilde{\eta}_0(E)\big\}. \tag{5.25}$$

On this subdomain $\widetilde{\mathcal{D}}$, $\beta \le q_t^{-1}|\alpha_2|$, hence we get from (5.24) that there is a high probability event $\widetilde{\Xi}$ such that

$$\left|\alpha_2(m_t - \widetilde{m}_t) + \frac{P''(\widetilde{m})}{2}(m_t - \widetilde{m}_t)^2\right| \le o(1)\Lambda_t^2 + 3N^{\epsilon} q_t^{-1}|\alpha_2|^2$$

and thus

$$|\alpha_2|\Lambda_t \le \left(\frac{c_2}{2} + o(1)\right)\Lambda_t^2 + 3N^{\epsilon} q_t^{-1}|\alpha_2|^2$$

uniformly on $\widetilde{\mathcal{D}}$ on $\widetilde{\Xi}$. Hence, on $\widetilde{\Xi}$, we have either

$$|\alpha_2| \le 2c_2\Lambda_t \quad \text{or} \quad \Lambda_t \le 6N^{\epsilon} q_t^{-1}|\alpha_2|, \quad (z \in \widetilde{\mathcal{D}}). \tag{5.26}$$

When $\eta = 3$, it is easy to see that

$$|\alpha_2| \ge \left|P_2(\widetilde{m})\|P_1'(\widetilde{m})\right| \ge \frac{1}{c_1}\left(|z + 2\widetilde{m}| - C\frac{1}{q^2}\right) \ge \frac{\eta}{c_1} = \frac{3}{c_1} \gg 6N^{\epsilon} q_t^{-1}|\alpha_2|, \tag{5.27}$$

for sufficiently large N. From the a priory estimate, we know that $|\Lambda_t| \prec \psi$, we hence find that

$$\Lambda_t \le 6N^{\epsilon} q_t^{-1}|\alpha_2|, \tag{5.28}$$

holds for $z \in \widetilde{\mathcal{D}}$ on the event $\widetilde{\Xi}$. Putting (5.28) back into (5.15), we obtain that

$$\begin{aligned}\mathbb{E}\left[\left|P(m_t)\right|^{2D}\right] &\le N^{4D\epsilon}\beta^{2D}|\alpha_2|^{2D} + N^{3D\epsilon} q_t^{-8D} + q_t^{-6D}|\alpha_2|^{4D} \\ &\le N^{6D\epsilon}\beta^{2D}|\alpha_2|^{2D} + N^{6D\epsilon}\beta^{4D},\end{aligned} \tag{5.29}$$

for any small $\epsilon > 0$, and large D, uniformly on $\widetilde{\mathcal{D}}$. For $z \in \mathcal{D}_\ell \backslash \widetilde{\mathcal{D}}$, it is direct to check the estimate $\mathbb{E}[|P(m_t)|^{2D}] \le N^{6D\epsilon}\beta^{2D}|\alpha_2|^{2D} + N^{6D\epsilon}\beta^{4D}$. Using a lattice argument and the Lipschitz continuity, we find from a union bound that for any small $\epsilon > 0$ and large D there exists an event Ξ with $\mathbb{P}(\Xi) \ge 1 - N^{-D}$ such that

$$\left|\alpha_2(m_t - \widetilde{m}_t) + \frac{P''(\widetilde{m})}{2}(m_t - \widetilde{m}_t)^2\right| \le N^{\epsilon}\psi\Lambda_t^2 + N^{\epsilon}|\alpha_2|\beta + N^{\epsilon}\beta^2 \tag{5.30}$$

on Ξ, uniformly on $\mathcal{D}_\ell$ for any sufficiently large N.

Recall that for fixed E, $\beta = \beta(E + \mathrm{i}\eta)$ is a decreasing function of η, $\sqrt{\kappa_t(E) + \eta}$ is an increasing function of η, and $\eta_0 \equiv \eta_0(E)$ satisfies that $\sqrt{\kappa(E) + \eta_0} = 10C_0N^{\epsilon}\beta(E + \mathrm{i}\eta_0)$. Further notice that $\eta_0(E)$ is a continuous function. We consider the following subdomains of $\mathcal{E}$:

$$\begin{aligned}\mathcal{E}_1 &:= \left\{z = E + \mathrm{i}\eta \in \mathcal{E} : \eta \le \eta_0(E), 10N^{\epsilon} \le N\eta\right\}, \\ \mathcal{E}_2 &:= \left\{z = E + \mathrm{i}\eta \in \mathcal{E} : \eta > \eta_0(E), 10N^{\epsilon} \le N\eta\right\}, \\ \mathcal{E}_3 &:= \left\{z = E + \mathrm{i}\eta \in \mathcal{E} : 10N^{\epsilon} \ge N\eta\right\}.\end{aligned}$$

We consider the cases $z \in \mathcal{E}_1$, $z \in \mathcal{E}_2$ and $z \in \mathcal{E}_3$, and split the stability analysis accordingly. Let Ξ be a high probability event such that (5.30) holds. Note that we can choose ℓ sufficiently small that satisfies $\mathcal{D}_\ell \supset \mathcal{E}_1 \cup \mathcal{E}_2$.

Case 1: If $z \in \mathcal{E}_1$, we note that $|\alpha_2| \le C_0\sqrt{\kappa(E)+\eta} \le 10C_0^2 N^\epsilon \beta(E+\mathrm{i}\eta)$. Then, we find that

$$\left|\frac{P''(\widetilde{m})}{2}\right|\Lambda_t^2 \le |\alpha_2|\Lambda_t + N^\epsilon\psi\Lambda_t^2 + N^\epsilon|\alpha_2|\beta + N^\epsilon\beta^2$$
$$\le 10C_0^2N^\epsilon\beta\Lambda_t + N^\epsilon\psi\Lambda_t^2 + \big(10C_0^2N^\epsilon+1\big)N^\epsilon\beta^2,$$

on Ξ. Hence, there is some finite constant C such that on Ξ, we have $\Lambda_t \le CN^\epsilon\beta$, $z \in \mathcal{E}_1$.

Case 2: If $z \in \mathcal{E}_2$, we obtain that

$$|\alpha_2|\Lambda_t \le \left(\left|\frac{P''(\widetilde{m})}{2}\right| + N^\epsilon\psi\right)\Lambda_t^2 + |\alpha_2|N^\epsilon\beta + N^\epsilon\beta^2, \tag{5.31}$$

on Ξ. We then notice that $C_0|\alpha_2| \ge \sqrt{\kappa_t(E)+\eta} \ge 10C_0N^\epsilon\beta$, *that is*, $N^\epsilon\beta \le |\alpha_2|/10$, so that

$$|\alpha_2|\Lambda_t \le \left(\left|\frac{P''(\widetilde{m})}{2}\right| + N^\epsilon\psi\right)\Lambda_t^2 + \big(1+N^{-\epsilon}\big)|\alpha_2|\beta \le c_2\Lambda_t^2 + \big(1+N^{-\epsilon}\big)|\alpha_2|\beta \tag{5.32}$$

on Ξ, where we used that $N^\epsilon\psi \le 1$. Hence, on Ξ, we have either

$$|\alpha_2| \le 2c_2\Lambda_t \quad \text{or} \quad \Lambda_t \le 3N^\epsilon\beta. \tag{5.33}$$

We use the dichotomy argument and the continuity argument similarly to the strategy to get (5.28). Since $3N^\epsilon\beta \le |\alpha_2|/8$ on $\mathcal{E}_2$, by continuity, we find that on the event Ξ, $\Lambda_t \le 3N^\epsilon\beta$ for $z \in \mathcal{E}_2$.

Case 3: For $z \in \mathcal{E}_3$ we use that $|m_t'(z)| \le \frac{\operatorname{Im} m_t(z)}{\operatorname{Im} z}$, $z \in \mathbb{C}^+$. Note that m_t is a Stieltjes transform of a probability measure. Set $\widetilde{\eta} := 10N^{-1+\epsilon}$ and observe that

$$\big|m_t(E+\mathrm{i}\eta)\big| \le \int_\eta^{\widetilde{\eta}} \frac{s\operatorname{Im} m_t(E+\mathrm{i}s)}{s^2}\,\mathrm{d}s + \Lambda_t(E+\mathrm{i}\widetilde{\eta}) + \big|\widetilde{m}_t(E+\mathrm{i}\widetilde{\eta})\big|. \tag{5.34}$$

It is easy to check that $s \to s\operatorname{Im} m_t(E+\mathrm{i}s)$ is monotone increasing. Thus, we find that

$$\big|m_t(E+\mathrm{i}\eta)\big| \le \frac{2\widetilde{\eta}}{\eta}\operatorname{Im} m_t(E+\mathrm{i}\widetilde{\eta}) + \Lambda_t(E+\mathrm{i}\widetilde{\eta}) + \big|\widetilde{m}_t(E+\mathrm{i}\widetilde{\eta})\big|$$
$$\le C\frac{N^\epsilon}{N\eta}\big(\operatorname{Im}\widetilde{m}_t(E+\mathrm{i}\widetilde{\eta}) + \Lambda_t(E+\mathrm{i}\widetilde{\eta})\big) + \big|\widetilde{m}_t(E+\mathrm{i}\widetilde{\eta})\big|, \tag{5.35}$$

for some C where we used $\widetilde{\eta} = 10N^{-1+\epsilon}$ to obtain the second inequality. Since $z = E+\mathrm{i}\widetilde{\eta} \in \mathcal{E}_1 \cup \mathcal{E}_2$, we have $\Lambda_t(E+\mathrm{i}\widetilde{\eta}) \le CN^\epsilon\beta(E+\mathrm{i}\widetilde{\eta}) \le C$ on Ξ. Since $\widetilde{m}_t$ is uniformly bounded on $\mathcal{E}$, we obtain that $\Lambda_t \le CN^\epsilon\beta$ on Ξ, for all $z \in \mathcal{E}_3$.

To sum up, we get $\Lambda_t \prec \beta$ uniformly on $\mathcal{E}$ for fixed $t \in [0, 6\log N]$. Choosing $t = 0$, we have proved Theorem 2.6. Now we use the continuity of the Dyson matrix flow to prove that this result holds for all $t \in [0, 6\log N]$. Consider a lattice $\mathcal{L} \subset [0, 6\log N]$ with spacings of order N^{-3}. Then we obtain that $\Lambda_t \prec \beta$, uniformly on $\mathcal{E}$ and on $\mathcal{L}$, by a union bound. Thus, by continuity, we can extend the conclusion to all $t \in [0, 6\log N]$ and conclude the proof of Proposition 3.2. □

5.2. Proof of Theorem 2.12

Theorem 2.12 follows directly from the following result.

Lemma 5.2. *Suppose that H_0 satisfy Assumption 2.1 with $\phi > 0$. Then,*

$$\big|\|H_t\| - L_t\big| \prec \frac{1}{q_t^4} + \frac{1}{N^{2/3}}, \tag{5.36}$$

uniformly in $t \in [0, 6\log N]$.

For the proof of Lemma 5.2, the similar strategy to the ones in [27] and [23] can be applied. We establish the upper bound on the largest eigenvalue of H_t, using a stability analysis starting from (5.15) and the fact that $\alpha_1(z) = \operatorname{Im} \widetilde{m}_t$ behaves as $\eta/\sqrt{\kappa_t(E)+\eta}$, for $E \geq L_+$. The detailed proof is given in Appendix D of [24].

6. Proof of Tracy–Widom limit

In this section, we prove the Theorem 2.13, the Tracy–Widom limiting distribution of the largest eigenvalue. Following the idea from [15], we consider the imaginary part of the normalized trace of the Green function $m \equiv m^H$ of H. For $\eta > 0$, define

$$\theta_\eta(y) = \frac{\eta}{\pi(y^2+\eta^2)}, \quad (y \in \mathbb{R}). \tag{6.1}$$

From the definition of the Green function, one can easily check that

$$\operatorname{Im} m(E + \mathrm{i}\eta) = \frac{\pi}{N} \operatorname{Tr} \theta_\eta(H - E). \tag{6.2}$$

The first proposition in this section shows how we can approximate the distribution of the largest eigenvalue by using the Green function. Recall that L_+ is the right endpoint of the deterministic probability measure in Theorem 2.6.

Proposition 6.1. *Let H satisfy Assumption 2.1, with $\phi > 1/6$. Denote by λ_1^H the largest eigenvalue of H. Fix $\epsilon > 0$ and let $E \in \mathbb{R}$ be such that $|E - L| \leq N^{-2/3+\epsilon}$. Set $E_+ := L + 2N^{-2/3+\epsilon}$ and define $\chi_E := \mathbb{1}_{[E,E_+]}$. Let $\eta_1 := N^{-2/3-3\epsilon}$ and $\eta_2 := N^{-2/3-9\epsilon}$. Let $K : \mathbb{R} \to [0,\infty)$ be a smooth function satisfying*

$$K(x) = \begin{cases} 1 & \text{if } |x| < 1/3, \\ 0 & \text{if } |x| > 2/3, \end{cases} \tag{6.3}$$

which is a monotone decreasing on $[0, \infty)$. *Then*, *for any* $D > 0$,

$$\mathbb{P}\big(\lambda_1^H \le E - \eta_1\big) - N^{-D} < \mathbb{E}\big[K\big(\mathrm{Tr}(\chi_E * \theta_{\eta_2})(H)\big)\big] < \mathbb{P}\big(\lambda_1^H \le E + \eta_1\big) + N^{-D} \quad (6.4)$$

for N *sufficiently large*, *with* θ_{η_2}.

We refer to Proposition 7.1 of [27] for the proof. We remark that the lack of the improved local law near the lower edge does not alter the proof of Proposition 6.1.

Define W^G be a $N \times N$ generalized Wigner matrix independent of H with Gaussian entries W_{ij}^G satisfying

$$\mathbb{E}W_{ij}^G = 0, \qquad \mathbb{E}\big|W_{ij}^G\big|^2 = \mathbb{E}|H_{ij}|^2,$$

and denote by $m^G \equiv m^{W^G}$ the normalized trace of its Green function. The following is the Green function comparison for our model.

Proposition 6.2. *Under the assumptions of Proposition* 6.1 *the following holds*. *Let* $\epsilon > 0$ *and set* $\eta_0 = N^{-2/3-\epsilon}$. *Let* $E_1, E_2 \in \mathbb{R}$ *satisfy* $|E_1|, |E_2| \le N^{-2/3+\epsilon}$. *Consider a smooth function* $F : \mathbb{R} \to \mathbb{R}$ *such that*

$$\max_{x \in \mathbb{R}} \big|F^{(l)}(x)\big|\big(|x| + 1\big)^{-C} \le C, \quad \big(l \in [1, 11]\big). \quad (6.5)$$

Then, *for any sufficiently small* $\epsilon > 0$, *there exists* $\delta > 0$ *such that*

$$\left|\mathbb{E}F\left(N\int_{E_1}^{E_2} \mathrm{Im}\, m(x + L_+ + \mathrm{i}\eta_0)\, dx\right) - \mathbb{E}F\left(N\int_{E_1}^{E_2} \mathrm{Im}\, m^G(x + \lambda_+ + \mathrm{i}\eta_0)\, dx\right)\right| \le N^{-\delta}, \quad (6.6)$$

for large enough N.

From Theorem 2.7 of [6], we know that the largest eigenvalue of the generalized Wigner matrix follows the Tracy–Widom distribution. Thus, Proposition 6.2 directly implies Theorem 2.13, the Tracy–Widom limit for the largest eigenvalue. A detailed proof is found, *for example*, with the same notation in [27], Section 7.

In the remainder of the section, we prove Proposition 6.2. We begin by the following application of the generalized Stein lemma.

Lemma 6.3. *Fix* $\ell \in \mathbb{N}$ *and let* $F \in C^{\ell+1}(\mathbb{R}; \mathbb{C}^+)$. *Let* $Y \equiv Y_0$ *be a random variable with finite moments to order* $\ell + 2$ *and let* W *be a Gaussian random variable independent of* Y. *Assume that* $\mathbb{E}[Y] = \mathbb{E}[W] = 0$ *and* $\mathbb{E}[Y^2] = \mathbb{E}[W^2]$. *Introduce*

$$Y_t := \mathrm{e}^{-t/2} Y_0 + \sqrt{1 - \mathrm{e}^{-t}}\, W, \quad (6.7)$$

and let $\dot{Y}_t \equiv \mathrm{d}Y_t/\mathrm{d}t$. *Then,*

$$\mathbb{E}\big[\dot{Y}_t F(Y_t)\big] = -\frac{1}{2}\sum_{r=2}^{\ell} \frac{\kappa^{(r+1)}(Y_0)}{r!}\mathrm{e}^{-\frac{(r+1)t}{2}}\mathbb{E}\big[F^{(r)}(Y_t)\big] + \mathbb{E}\big[\Omega_\ell\big(\dot{Y}_t F(Y_t)\big)\big], \tag{6.8}$$

where $\mathbb{E}$ *denotes the expectation with respect to* Y *and* W, $\kappa^{(r+1)}(Y)$ *denotes the* $(r+1)$*-th cumulant of* Y *and* $F^{(r)}$ *denotes the* r*-th derivative of the function* F. *The error term* Ω_ℓ *in* (6.8) *satisfies*

$$\begin{aligned}\big|\mathbb{E}\big[\Omega_\ell\big(\dot{Y}_t F(Y_t)\big)\big]\big| &\le C_\ell \mathbb{E}\big[|Y_t|^{|\ell+2}\big] \sup_{|x|\le Q}\big|F^{(\ell+1)}(x)\big| \\ &\quad + C_\ell \mathbb{E}\big[|Y_t|^{\ell+2}\mathbf{1}\big(|Y_t|>Q\big)\big]\sup_{x\in\mathbb{R}}\big|F^{(\ell+1)}(x)\big|,\end{aligned} \tag{6.9}$$

where $Q>0$ *is an arbitrary fixed cutoff and* C_ℓ *satisfies* $C_\ell \le \frac{(C\ell)^\ell}{\ell!}$ *for some numerical constant* C.

Proof of Proposition 6.2. Fix a (small) $\epsilon>0$. Consider $x\in[E_1,E_2]$. For simplicity, let

$$G \equiv G_t(x+L_t+\mathrm{i}\eta_0), \qquad m \equiv m_t(x+L_t+\mathrm{i}\eta_0), \tag{6.10}$$

with $\eta_0 = N^{-2/3-\epsilon}$, and define

$$X \equiv X_t := N\int_{E_1}^{E_2} \operatorname{Im} m(x+L_t+\mathrm{i}\eta_0)\,\mathrm{d}x. \tag{6.11}$$

Note that $X \prec N^\epsilon$ and $|F^{(l)}(X)| \prec N^{C\epsilon}$ for $l\in[1,11]$. From (3.11) we can obtain that

$$L_t = 2+\mathrm{e}^{-t}\xi^{(4)}q_t^{-2} + O\big(\mathrm{e}^{-2t}q_t^{-4}\big), \qquad \dot{L}_t = -2\mathrm{e}^{-t}\xi^{(4)}q_t^{-2} + O\big(\mathrm{e}^{-2t}q_t^{-4}\big),$$

with $q_t = \mathrm{e}^{t/2}q_0$, where $\dot{L}_t$ denotes the derivative with respect to t of L_t. Let $z = x+L_t+\mathrm{i}\eta_0$ and $G\equiv G(z)$. Differentiating $F(X)$ with respect to t, we obtain

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}F(X) &= \mathbb{E}\bigg[F'(X)\frac{\mathrm{d}X}{\mathrm{d}t}\bigg] = \mathbb{E}\bigg[F'(X)\operatorname{Im}\int_{E_1}^{E_2}\sum_{i=1}^{N}\frac{\mathrm{d}G_{ii}}{\mathrm{d}t}\,\mathrm{d}x\bigg] \\ &= \mathbb{E}\bigg[F'(X)\operatorname{Im}\int_{E_1}^{E_2}\bigg(\sum_{i,j,k}\dot{H}_{jk}\frac{\partial G_{ii}}{\partial H_{jk}} + \dot{L}_t\sum_{1\le i,j\le N}G_{ij}G_{ji}\bigg)\mathrm{d}x\bigg],\end{aligned} \tag{6.12}$$

where by definition

$$\dot{H}_{jk} \equiv (\dot{H}_t)_{jk} = -\frac{1}{2}\mathrm{e}^{-t/2}(H_0)_{jk} + \frac{\mathrm{e}^{-t}}{2\sqrt{1-\mathrm{e}^{-t}}}W^G_{jk}. \tag{6.13}$$

Thus, we find that

$$
\begin{aligned}
&\sum_{i,j,k} \mathbb{E}\left[\dot{H}_{jk} F'(X) \frac{\partial G_{ii}}{\partial H_{jk}}\right] \\
&\quad = -2 \sum_{i,j,k} \mathbb{E}\big[\dot{H}_{jk} F'(X) G_{ij} G_{ki}\big] \\
&\quad = \frac{\mathrm{e}^{-t}}{N} \sum_{r=2}^{\ell} \frac{q_t^{-(r-1)}}{r!} \sum_i \sum_{j \neq k} \mathbb{E}\big[s_{(jk)}^{(r+1)} \partial_{jk}^r \big(F'(X) G_{ij} G_{ki}\big)\big] + O\big(N^{1/3+C\epsilon}\big),
\end{aligned} \tag{6.14}
$$

for $\ell = 10$, by Lemma 6.3, where we use the short hand $\partial_{jk} = \partial / \partial H_{jk}$. Here, the error term $O(N^{1/3+C\epsilon})$ in (6.14) corresponds to Ω_ℓ in (6.8), which is $O(N^{C\epsilon} N^2 q_t^{-10})$ for $X = H_{ji}$. To estimate the right-hand side of (2.6), we use the following lemma whose proof is in Appendix E of [24].

Lemma 6.4. *For an integer $r \geq 2$, let*

$$
A_r := \frac{\mathrm{e}^{-t}}{N} \frac{q_t^{-(r-1)}}{r!} \sum_i \sum_{j \neq k} \mathbb{E}\big[s_{(jk)}^{(r+1)} \partial_{jk}^r \big(F'(X) G_{ij} G_{ki}\big)\big]. \tag{6.15}
$$

Then, for any $r \neq 3$,

$$
A_r = O\big(N^{2/3-\epsilon'}\big), \tag{6.16}
$$

and

$$
A_3 = 2\mathrm{e}^{-t} \xi^{(4)} q_t^{-2} \sum_{i,j} \mathbb{E}\big[F'(X) G_{ij} G_{ji}\big] + O\big(N^{2/3-\epsilon'}\big). \tag{6.17}
$$

Assuming Lemma 6.4, we find that there exists $\epsilon' > 2\epsilon$ such that, for all $t \in [0, 6 \log N]$,

$$
\sum_{i,j,k} \mathbb{E}\left[\dot{H}_{jk} F'(X) \frac{\partial G_{ii}}{\partial H_{jk}}\right] = -\dot{L}_t \sum_{i,j} \mathbb{E}\big[G_{ij} G_{ji} F'(X)\big] + O\big(N^{2/3-\epsilon'}\big), \tag{6.18}
$$

which implies that the right-hand side of (6.12) is $O(N^{-\epsilon'/2})$. Integrating Equation (6.12) from $t = 0$ to $t = 6 \log N$, we get

$$
\begin{aligned}
&\left| \mathbb{E} F\left(N \int_{E_1}^{E_2} \operatorname{Im} m(x + L_t + \mathrm{i}\eta_0)\, \mathrm{d}x\right)_{t=0} - \mathbb{E} F\left(N \int_{E_1}^{E_2} \operatorname{Im} m(x + L_t + \mathrm{i}\eta_0)\, \mathrm{d}x\right)_{t=6\log N} \right| \\
&\quad \leq N^{-\epsilon'/4}.
\end{aligned}
$$

By comparing largest eigenvalues of H and λ_i^G, we can get desired result. Let $\lambda_i(6\log N)$ be the i-th largest eigenvalue of $H_{6\log N}$ and λ_i^G the i-th largest eigenvalue of W^G, then $|\lambda_i(6\log N) - \lambda_i^G| \prec N^{-3}$. Then we find that

$$\left|\operatorname{Im} m|_{t=6\log N} - \operatorname{Im} m^G\right| \prec N^{-5/3}. \tag{6.19}$$

This completes the proof of Proposition 6.2. □

Acknowledgements

This work was partially supported by the Samsung Science and Technology Foundation project number SSTF-BA1402-04.

Supplementary Material

Supplementary to "Local law and Tracy–Widom limit for sparse stochastic block models" (DOI: 10.3150/20-BEJ1201SUPP; .pdf). In the supplementary material [24], we will provide some important properties of the deterministic refinement of Wigner's semicircle law, and the proofs of lemmas, including the recursive moment estimates, the bound on the operator norm $\|H\|$, and technical lemmas in Section 4 and Section 6.

References

[1] Abbe, E. (2017). Community detection and stochastic block models: Recent developments. *J. Mach. Learn. Res.* **18** Paper No. 177, 86. MR3827065

[2] Abbe, E., Bandeira, A.S. and Hall, G. (2016). Exact recovery in the stochastic block model. *IEEE Trans. Inf. Theory* **62** 471–487. MR3447993 https://doi.org/10.1109/TIT.2015.2490670

[3] Abbe, E. and Sandon, C. (2018). Proof of the achievability conjectures for the general stochastic block model. *Comm. Pure Appl. Math.* **71** 1334–1406. MR3812075 https://doi.org/10.1002/cpa.21719

[4] Baik, J., Lee, J.O. and Wu, H. (2018). Ferromagnetic to paramagnetic transition in spherical spin glass. *J. Stat. Phys.* **173** 1484–1522. MR3878351 https://doi.org/10.1007/s10955-018-2150-6

[5] Bickel, P.J. and Sarkar, P. (2016). Hypothesis testing for automated community detection in networks. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **78** 253–273. MR3453655 https://doi.org/10.1111/rssb.12117

[6] Bourgade, P., Erdös, L. and Yau, H.-T. (2014). Edge universality of beta ensembles. *Comm. Math. Phys.* **332** 261–353. MR3253704 https://doi.org/10.1007/s00220-014-2120-z

[7] Chen, J. and Yuan, B. (2006). Detecting functional modules in the yeast protein-protein interaction network. *Bioinformatics* **22** 2283–2290.

[8] Chung, H.W. and Lee, J.O. (2018). Weak detection of signal in the spiked wigner model. arXiv:1809.10827.

[9] Cline, M.S., Smoot, M., Cerami, E., Kuchinsky, A., Landys, N., Workman, C., Christmas, R., Avila-Campilo, I., Creech, M., Gross, B., Hanspers, K., Isserlin, R., Kelley, R., Killcoyne, S., Lotia, S., Maere, S., Morris, J., Ono, K., Pavlovic, V., Pico, A.R., Vailaya, A., Wang, P.-L., Adler, A., Conklin, B.R., Hood, L., Kuiper, M., Sander, C., Schmulevich, I., Schwikowski, B., Warner, G.J., Ideker, T. and Bader, G.D. (2007). Integration of biological networks and gene expression data using Cytoscape. *Nat. Protoc.* **2** 2366–2382. https://doi.org/10.1038/nprot.2007.324

[10] Erdős, L., Knowles, A., Yau, H.-T. and Yin, J. (2012). Spectral statistics of Erdős–Rényi Graphs II: Eigenvalue spacing and the extreme eigenvalues. *Comm. Math. Phys.* **314** 587–640. MR2964770 https://doi.org/10.1007/s00220-012-1527-7
[11] Erdős, L., Knowles, A., Yau, H.-T. and Yin, J. (2013). The local semicircle law for a general class of random matrices. *Electron. J. Probab.* **18** no. 59, 58. MR3068390 https://doi.org/10.1214/EJP.v18-2473
[12] Erdős, L., Knowles, A., Yau, H.-T. and Yin, J. (2013). Spectral statistics of Erdős–Rényi graphs I: Local semicircle law. *Ann. Probab.* **41** 2279–2375. MR3098073 https://doi.org/10.1214/11-AOP734
[13] Erdős, L., Yau, H.-T. and Yin, J. (2011). Universality for generalized Wigner matrices with Bernoulli distribution. *J. Comb.* **2** 15–81. MR2847916 https://doi.org/10.4310/JOC.2011.v2.n1.a2
[14] Erdős, L., Yau, H.-T. and Yin, J. (2012). Bulk universality for generalized Wigner matrices. *Probab. Theory Related Fields* **154** 341–407. MR2981427 https://doi.org/10.1007/s00440-011-0390-3
[15] Erdős, L., Yau, H.-T. and Yin, J. (2012). Rigidity of eigenvalues of generalized Wigner matrices. *Adv. Math.* **229** 1435–1515. MR2871147 https://doi.org/10.1016/j.aim.2011.12.010
[16] Girvan, M. and Newman, M.E.J. (2002). Community structure in social and biological networks. *Proc. Natl. Acad. Sci. USA* **99** 7821–7826. MR1908073 https://doi.org/10.1073/pnas.122653799
[17] Guédon, O. and Vershynin, R. (2016). Community detection in sparse networks via Grothendieck's inequality. *Probab. Theory Related Fields* **165** 1025–1049. MR3520025 https://doi.org/10.1007/s00440-015-0659-z
[18] Guimerà, R. and Amaral, L.A.N. (2015). Functional cartography of complex metabolic networks. *Nature* **433** 895–900.
[19] Guimerà, R. and Amaral, L.A.N. (2015). Cartography of complex networks: Modules and universal roles. *J. Stat. Mech.* **02001**.
[20] Hajek, B., Wu, Y. and Xu, J. (2016). Achieving exact cluster recovery threshold via semidefinite programming. *IEEE Trans. Inf. Theory* **62** 2788–2797. MR3493879 https://doi.org/10.1109/TIT.2016.2546280
[21] Holland, P.W., Laskey, K.B. and Leinhardt, S. (1983). Stochastic blockmodels: First steps. *Soc. Netw.* **5** 109–137. MR0718088 https://doi.org/10.1016/0378-8733(83)90021-7
[22] Huang, J., Landon, B. and Yau, H.-T. (2017). Transition from Tracy–Widom to Gaussian fluctuations of extremal eigenvalues of sparse Erdős–Rényi graphs. Preprint, arXiv:1712.03936.
[23] Hwang, J.Y., Lee, J.O. and Schnelli, K. (2019). Local law and Tracy–Widom limit for sparse sample covariance matrices. *Ann. Appl. Probab.* **29** 3006–3036. MR4019881 https://doi.org/10.1214/19-AAP1472
[24] Hwang, J.Y., Lee, J.O. and Yang, W. (2020). Supplement to "Local law and Tracy–Widom limit for sparse stochastic block models." https://doi.org/10.3150/20-BEJ1201SUPP
[25] Jiang, D., Tang, C. and Zhang, A. (2004). Cluster analysis for gene expression data: A survey. *IEEE Trans. Knowl. Data Eng.* **16** 1370–1386.
[26] Krzakala, F., Moore, C., Mossel, E., Neeman, J., Sly, A., Zdeborová, L. and Zhang, P. (2013). Spectral redemption in clustering sparse networks. *Proc. Natl. Acad. Sci. USA* **110** 20935–20940. MR3174850 https://doi.org/10.1073/pnas.1312486110
[27] Lee, J.O. and Schnelli, K. (2018). Local law and Tracy–Widom limit for sparse random matrices. *Probab. Theory Related Fields* **171** 543–616. MR3800840 https://doi.org/10.1007/s00440-017-0787-8
[28] Lei, J. (2016). A goodness-of-fit test for stochastic block models. *Ann. Statist.* **44** 401–424. MR3449773 https://doi.org/10.1214/15-AOS1370
[29] Lytova, A. and Pastur, L. (2009). Central limit theorem for linear eigenvalue statistics of random matrices with independent entries. *Ann. Probab.* **37** 1778–1840. MR2561434 https://doi.org/10.1214/09-AOP452

[30] Newman, M.E.J. (2001). The structure of scientific collaboration networks. *Proc. Natl. Acad. Sci. USA* **98** 404–409. MR1812610 https://doi.org/10.1073/pnas.021544898

[31] Rohe, K., Chatterjee, S. and Yu, B. (2011). Spectral clustering and the high-dimensional stochastic blockmodel. *Ann. Statist.* **39** 1878–1915. MR2893856 https://doi.org/10.1214/11-AOS887

[32] Stein, C.M. (1981). Estimation of the mean of a multivariate normal distribution. *Ann. Statist.* **9** 1135–1151. MR0630098

[33] Traud, A., Mucha, P. and Porter, M. (2012). Social structure of Facebook networks. *Phys. A* **391** 4165–4180.

[34] Traud, A.L., Kelsic, E.D., Mucha, P.J. and Porter, M.A. (2011). Comparing community structure to characteristics in online collegiate social networks. *SIAM Rev.* **53** 526–543. MR2834086 https://doi.org/10.1137/080734315

Received September 2019 and revised January 2020