

Largest eigenvalues of sparse inhomogeneous Erdős-Rényi graphs

Florent Benaych-Georges Charles Bordenave Antti Knowles

April 25, 2018

Abstract

We consider inhomogeneous Erdős-Rényi graphs. We suppose that the maximal mean degree d satisfies $d \ll \log n$. We characterize the asymptotic behavior of the $n^{1-o(1)}$ largest eigenvalues of the adjacency matrix and its centred version. We prove that these extreme eigenvalues are governed at first order by the largest degrees and, for the adjacency matrix, by the nonzero eigenvalues of the expectation matrix. Our results show that the extreme eigenvalues exhibit a novel behaviour which in particular rules out their convergence to a non-degenerate point process. Together with the companion paper [4], where we analyse the extreme eigenvalues in the complementary regime $d \gg \log n$, this establishes a crossover in the behaviour of the extreme eigenvalues around $d \sim \log n$. Our proof relies on a tail estimate for the Poisson approximation of an inhomogeneous sum of independent Bernoulli random variables, as well as on an estimate on the operator norm of a pruned graph due to Le, Levina, and Vershynin from [15].

1 Introduction

The purpose of the present text is to understand the extreme eigenvalues of the adjacency matrix of an inhomogeneous Erdős-Rényi random graph on n vertices in the regime where the maximal mean degree d satisfies $d \ll \log n$. Heuristically, such eigenvalues arise from three different origins: (i) the edge of the limiting bulk eigenvalue density, (ii) vertices of large degrees, and (iii) outliers associated with nonzero eigenvalues of the expectation matrix. One goal of this paper is a precise understanding of this interplay between random matrices on the one hand and the geometry of random graphs on the other. Such questions have several motivations from applications, such as the estimation of the spectral gap and spectral clustering.

The simplest random graph is the Erdős-Rényi random graph $G(n, d/n)$, where each edge is present independently with probability d/n . In this case it is rather well understood that the behaviour of the extreme eigenvalues in the regime $d \gg \log n$ is governed by random matrix behavior; see [4, 9–12, 14, 17, 20]. In the complementary regime $d \ll \log n$, the main result available up to now was due to Sudakov and Krivelevich [13], who showed that the largest eigenvalue of the adjacency matrix is asymptotically equivalent to the maximum of the maximal mean degree d and the square root of the largest degree (their result holds in fact for all regimes of d).

Our main result is a description of the behaviour of the $n^{1-o(1)}$ largest and smallest eigenvalues of the adjacency matrix A and its centred version $\underline{A} := A - \mathbb{E}A$, for an inhomogeneous Erdős-Rényi random graph whose mean degree d is much smaller than $\log n$. Informally, we prove that the k -th largest eigenvalue eigenvalue of \underline{A} satisfies

$$\lambda_k(\underline{A}) \approx \sqrt{\frac{\log(n/k)}{\log((\log n)/d)}}, \quad k \geq n^{1-\varepsilon}, \quad \varepsilon \in (0, 1). \quad (1.1)$$

Under mild additional assumptions (satisfied for instance by stochastic block models), we show that the same result holds for the eigenvalues of A , with the exception of some outlier eigenvalues whose locations we also characterize.

A consequence of our results, combined with those from the companion paper [4] (see also the recent preprint [14]), where we analyse the extreme eigenvalues in the complementary regime $d \gg \log n$, is a crossover in the behaviour of the extreme eigenvalues around $d \sim \log n$ (the same threshold as for the graph connectivity). Indeed, in [4] we prove that if $d \gg \log n$ then all eigenvalues are asymptotically contained within the support of the semicircle law describing the macroscopic eigenvalue density, while in the current paper we establish for $d \ll \log n$ a novel behaviour of the extreme eigenvalues, which implies that $n^{1-o(1)}$ eigenvalues escape the support of the semicircle law. Their locations are governed by (1.1) and define a distribution that is illustrated in Figure 1 below.

It is helpful to analyse the behaviour of the extreme eigenvalues for $d \ll \log n$ in the context of random matrix theory. Until now, in random matrix theory two different types of universal behaviour at leading order of the extreme eigenvalues have been established, exhibited for instance by light- and heavy-tailed Wigner matrices respectively. After a suitable deterministic rescaling of the matrix, these two classes may be characterized as follows.

- (a) The extreme eigenvalues converge to the edge of the support of the asymptotic bulk spectrum.

(b) The extreme eigenvalues form asymptotically a Poisson point process.

For example, it is known [1, 16, 19] that a Wigner matrix whose entries have tail decay $x^{-\alpha}$ belongs to class (a) if $\alpha > 4$ and to class (b) if $\alpha < 4$. Moreover, as stated above, in the companion paper [4] we prove that the Erdős-Rényi graph belongs to class (a) if $d \gg \log n$. Also, sparse heavy-tailed random matrices exhibit a transition between these classes depending on the sparsity and the tail decay of the entries [5].

A consequence of our results is that, perhaps surprisingly, for $d \ll \log n$, the (possibly inhomogeneous) Erdős-Rényi graph belongs to neither class (a) nor class (b). Instead, the behaviour from (1.1) results in a sharp increase in the density of eigenvalues as one moves towards the centre of the spectrum, which implies that, no matter the rescaling of the spectrum, any nondegenerate limiting point process will be infinite on compact sets.

The proof consists of two main steps. In a first step, we analyse the distribution of the $n^{1-o(1)}$ largest degrees, and prove that the corresponding vertices are with high probability separated by distance at least 3 from each other. The key tool behind this step is an estimate (Theorem 3.1 below) on the tail of a sum of inhomogeneous independent Bernoulli random variables. This estimate may be regarded as an improvement for the tails of the well-known Poisson approximation provided by Le Cam's inequality [3]. In a second step, we compare the $n^{1-o(1)}$ largest eigenvalues of the graph with those of the graph obtained by only keeping the edges incident to the $n^{1-o(1)}$ vertices of largest degree. The latter corresponds to a block-diagonal matrix whose blocks are associated with star graphs of high central degree. This comparison is based on a sharp estimate on the operator norm of the complementary graph due to Le, Levina, and Vershynin [15, Theorem 2.1].

This text is organized as follows. In the remainder of the introduction, we state our main results, which are proved in Section 2. In Section 3 we state and prove the tail estimate for Poisson approximation mentioned above.

Notation The eigenvalues of a Hermitian $n \times n$ matrix H are denoted by $\lambda_1(H) \geq \lambda_2(H) \geq \dots \geq \lambda_n(H)$. Its operator norm is given by $\|H\| = \max(\lambda_1(H), \lambda_1(-H))$. For $p \in [0, 1]$, we denote by $B(p)$ the Bernoulli law with parameter p , i.e. $B(p) = (1-p)\delta_0 + p\delta_1$. We denote by $\text{Bin}(p_1, \dots, p_n)$ the law $B(p_1) * \dots * B(p_n)$. In particular, $\text{Bin}(p, \dots, p)$ is the Binomial distribution with parameters (n, p) . For $x > 0$ use the abbreviation $[x] := \{1, 2, \dots, \lfloor x \rfloor\}$.

1.1 Hypotheses and definitions

Throughout this paper, A is the adjacency matrix of an inhomogeneous (undirected) Erdős-Rényi random graph G with vertex set $[n]$, where the undirected edge $\{i, j\}$ is included with probability $p_{ij} = p_{ji} \in [0, 1]$ independently of the others. Note that we allow loops: there is a loop at vertex i with probability p_{ii} .

The maximal edge probability is

$$p_{\max} := \max_{i \neq j} p_{ij}.$$

The mean degree of the vertex $i \in [n]$ and the maximal mean degree are defined as

$$d_i := \sum_j p_{ij}, \quad d := \max_i d_i$$

respectively.

We always suppose that there are $\kappa > 0$ and $\eta \in (0, 1)$ such that

$$\kappa \leq d \leq \eta \log n \quad \text{and} \quad p_{\max} \leq n^{-1+\eta}. \quad (1.2)$$

As all of our error term controls will be uniform, with quantitative rates of convergence, in the parameters $(p_{ij})_{i,j \in [n]}$ such that (1.2) holds, we introduce the following definitions.

Definition 1.1. (i) An *admissible error function* is a function $\psi(n, \kappa, \eta)$ satisfying

$$\forall \kappa > 0, \quad \lim_{\substack{n \rightarrow \infty \\ \eta \rightarrow 0}} \psi(n, \kappa, \eta) = 0. \quad (1.3)$$

(ii) Given an event E and a condition \mathcal{A} on the parameters $\kappa, \eta, (p_{ij})_{i,j \in [n]}$, we say that, *under \mathcal{A} , E holds with high probability (w.h.p.)* if there is an admissible error function $\psi(n, \kappa, \eta)$ such that

$$\mathbb{P}(E) \geq 1 - \psi(n, \kappa, \eta)$$

for all $\kappa, \eta, (p_{ij})_{i,j \in [n]}$ satisfying \mathcal{A} .

(iii) Given a condition \mathcal{A} on the parameters $\kappa, \eta, (p_{ij})_{i,j \in [n]}$, for two families of random variables $(u_t), (v_t)$ we say that *under \mathcal{A} , for all t , $u_t \sim v_t$* if there is an admissible error function $\psi(n, \kappa, \eta)$ such that

$$\mathbb{P} \left(\left| \frac{v_t}{u_t} - 1 \right| \leq \psi(n, \kappa, \eta) \text{ for all } t \right) \geq 1 - \psi(n, \kappa, \eta)$$

for all $\kappa, \eta, (p_{ij})_{i,j \in [n]}$ satisfying \mathcal{A} .

Let us emphasize that the point in this definition is the uniformity of the error terms in the asymptotic regime where $n \rightarrow \infty$, $d = o(\log n)$, and $p_{\max} = n^{-1+o(1)}$. To simplify presentation, in the following we shall not identify the error functions $\psi(n, \kappa, \eta)$ explicitly, although a careful look at our proofs will easily yield explicit expressions for them.

Finally, for $k \in [n]$ we set

$$L_k := \frac{\log(n/k)}{\log((\log n)/d)}. \quad (1.4)$$

1.2 Relation between the centred adjacency matrix and the largest degrees

For $i \in [n]$, let D_i denote the degree of the vertex i in the graph G . Denote by

$$D_1^\downarrow \geq \dots \geq D_n^\downarrow$$

the decreasingly ordered degrees D_1, \dots, D_n . We also introduce the centred adjacency matrix

$$\underline{A} := A - \mathbb{E}A.$$

By definition, $(\mathbb{E}A)_{ij} = p_{ij}$. The following theorem relates the largest eigenvalues of \underline{A} to the largest degrees, whose behaviour is described in Propositions 1.11 and 1.12 below.

Theorem 1.2. *For any $\varepsilon \in (0, 1)$, under (1.2), w.h.p.,*

$$\max_{k \in [n^{1-\varepsilon}]} \left| \lambda_k(\underline{A}) - \sqrt{D_k^\downarrow} \right| \leq C\sqrt{np_{\max}} + \varepsilon\sqrt{L_1}, \quad (1.5)$$

$$\max_{k \in [n^{1-\varepsilon}]} \left| \lambda_{n+1-k}(\underline{A}) + \sqrt{D_k^\downarrow} \right| \leq C\sqrt{np_{\max}} + \varepsilon\sqrt{L_1}, \quad (1.6)$$

where C is a universal constant and L_1 is defined in (1.4).

The proof of Theorem 1.2 is based on an analysis of the graph spanned by the largest degree vertices, and on [15, Theorem 2.1] due to Le, Levina, and Vershynin on the operator norm of the centred adjacency matrix where all large degree vertices have been removed. The term $\pm\sqrt{D_k^\downarrow}$ arises as an eigenvalue of a star graph with central degree $D_k^\downarrow - O(1)$ (see Definition 2.6 below).

By Proposition 1.11 below, for any $\varepsilon \in (0, 1)$, we have $D_1^\downarrow \leq (1 + \varepsilon)L_1$ w.h.p., which yields the following corollary.

Corollary 1.3. For any $\varepsilon \in (0, 1)$, under (1.2), w.h.p.,

$$\|\underline{A}\| \leq (1 + \varepsilon)\sqrt{L_1} + C\sqrt{np_{\max}}.$$

As explained, for example, in [15], Corollary 1.3 finds applications in the analysis of spectral clustering techniques on random graphs.

Under the additional hypothesis that all vertices have the same mean degree, the behaviour of the largest degrees summarized in Corollary 1.13 below implies that $D_k^\downarrow \sim L_k$, where L_k was defined in (1.4). We deduce the following result.

Corollary 1.4. Let $\varepsilon \in (0, 1)$. Then under the conditions $d_i = d$ for all i , $np_{\max} \leq \eta L_1$, and (1.2), we have for all $k \in [n^{1-\varepsilon}]$

$$\lambda_k(\underline{A}) \sim \sqrt{L_k} \quad \text{and} \quad \lambda_{n+1-k}(\underline{A}) \sim -\sqrt{L_k}. \quad (1.7)$$

Remark 1.5. There is an equivalent way to state Corollary 1.4. Introduce the counting function of the renormalized eigenvalues of $\underline{A} = A - \mathbb{E}A$, defined as

$$N_{\underline{A}}(x) := \#\left\{k \in [n] : \frac{\lambda_k(\underline{A})}{\sqrt{L_1}} \geq x\right\}. \quad (1.8)$$

The first estimate of (1.7) implies that for any $x \in (0, 1)$,

$$\frac{\log N_{\underline{A}}(x)}{\log n} \sim 1 - x^2. \quad (1.9)$$

Indeed, for any $\delta > 0$ small enough, for $k := \lfloor n^{1-x^2-\delta} \rfloor$ and $k' := \lceil n^{1-x^2+\delta} \rceil$. We have

$$k \leq N_{\underline{A}}(x) < k' \iff \lambda_{k'}(\underline{A}) < x\sqrt{L_1} \leq \lambda_k(\underline{A}),$$

which happens w.h.p. by the first estimate of (1.7).

Informally, (1.9) states that $N_{\underline{A}}(x) \approx n^{1-x^2}$, from which we deduce that the density of renormalized eigenvalues $\frac{\lambda_k(\underline{A})}{\sqrt{L_1}}$ at $x \in (0, 1)$ is asymptotically

$$2 \log(n) n^{1-x^2} x. \quad (1.10)$$

See Figure 1 below for an illustration.

Remark 1.6. The estimate (1.9) states there exists no deterministic sequence $\alpha = \alpha_n$ such that the point process

$$\Xi := \{\alpha \lambda_k(\underline{A}) : k \in [n]\}$$

is asymptotically finite and nonzero on compact sets. In particular, Ξ cannot converge to a point process as $n \rightarrow \infty$. Note, however, that our results do not rule out the existence of an affine transformation parametrized by $\alpha = \alpha_n$ and $\beta = \beta_n$ such that the point process $\{\alpha(\lambda_k(\underline{A}) - \beta) : k \in [n]\}$ converges.

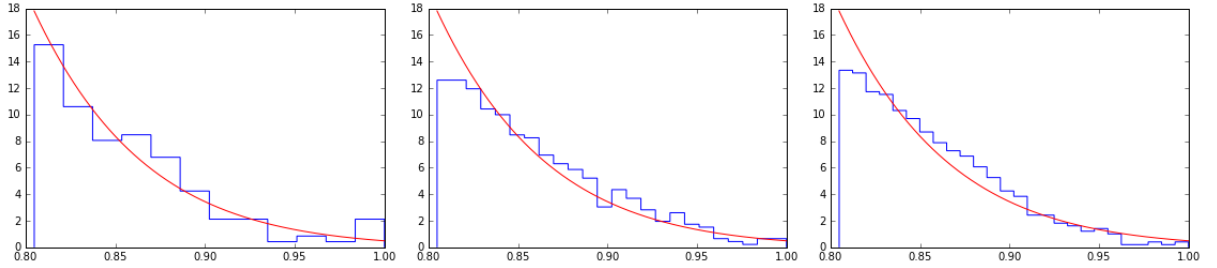


Figure 1: Histogram of the right edge of the spectrum of A in the case of a homogeneous Erdős-Rényi graph and density of (1.10). The eigenvalues are renormalized in such a way that $\lambda_2 = 1$ (λ_1 has been removed) and the histogram as well as the density are normalized in such a way that the total area is 1. Here, $n = 5 \cdot 10^4$ and $d = 0.5$ (left), $d = 1.5$ (centre), $d = 2.5$ (right). We see that as d grows, the empirical density gets more convex at its edge, which agrees with the idea that the semicircle law approximation gets more accurate.

1.3 Consequences for the adjacency matrix

Gershgorin's Circle Theorem [6, Problem VIII.6.3] implies that

$$\|\mathbb{E}A\| \leq d.$$

Then, writing $A = \underline{A} + \mathbb{E}A$, the following corollary is an immediate consequence of Theorem 1.2 and Weyl's inequality (see e.g. [6, Corollary III.2.6]).

Corollary 1.7. (a) *Theorem 1.2 holds with \underline{A} replaced by A and the right-hand sides of (1.5)–(1.6) replaced by $C\sqrt{np_{\max}} + \varepsilon\sqrt{L_1} + d$.*

(b) *Under condition (1.2), w.h.p., for any $k \in [n]$,*

$$|\lambda_k(A) - \lambda_k(\mathbb{E}A)| \leq C\left(\sqrt{L_1} + \sqrt{np_{\max}}\right),$$

for some universal constant C .

Remark 1.8. As $L_1 \sim D_1^\dagger$, for an homogeneous Erdős-Rényi random graph, Corollary 1.7 is consistent with [13, Theorem 1.1] which asserts that $\lambda_1(A) = \|A\| \sim \max\{\sqrt{D_1^\dagger}, d\}$ in all regimes of d .

1.4 Applications to stochastic block models

In the stochastic block model, $\mathbb{E}A$ has bounded rank and all its nonzero eigenvalues are of order d . We denote by $\lambda_1^+ \geq \dots \geq \lambda_{k^+}^+ > 0$ the positive

eigenvalues of \mathbb{E} and $\lambda_1^- \leq \dots \leq \lambda_{k^-}^- < 0$ the negative eigenvalues of $\mathbb{E}A$. For κ the constant of (1.2), we suppose that

$$d \geq \lambda_1^+ \geq \dots \geq \lambda_{k^+}^+ \geq \kappa d, \quad -d \leq \lambda_1^- \leq \dots \leq \lambda_{k^-}^- \leq -\kappa d, \quad k^+ + k^- \leq \kappa^{-1}. \quad (1.11)$$

Then, there is a dichotomy in the behaviour of the $k^+ + k^-$ largest (in absolute value) eigenvalues of A , depending on whether $d \ll \sqrt{L_1}$ or $d \gg \sqrt{L_1}$. Under mild conditions on d , these conditions read $d \ll \sqrt{\log n / \log \log n}$ or $d \gg \sqrt{\log n / \log \log n}$.

Proposition 1.9. *Let $\varepsilon \in (0, 1)$.*

(a) *Under conditions (1.2) and $d \leq \eta \sqrt{\log n / \log \log n}$, w.h.p.*

$$\|A\| \leq (1 + \varepsilon) \sqrt{L_1} + C \sqrt{np_{\max}}. \quad (1.12)$$

Under the additional conditions $d_i = d$ for all i and $np_{\max} \leq \eta L_1$, we have for all $k \in [n^{1-\varepsilon}]$

$$\lambda_k(A) \sim \sqrt{L_k} \quad \text{and} \quad \lambda_{n+1-k}(A) \sim -\sqrt{L_k}. \quad (1.13)$$

(b) *Under conditions (1.2), (1.11), and $d \geq \eta^{-1} \max\{\sqrt{\log n / \log \log n}, \sqrt{np_{\max}}\}$,*

$$\forall i = 1, \dots, k^+, \quad \lambda_i(A) \sim \lambda_i^+, \quad \forall i = 1, \dots, k^-, \quad \lambda_{n+1-i}(A) \sim \lambda_i^- \quad (1.14)$$

and w.h.p.,

$$\max\{|\lambda_{k^++1}(A)|, |\lambda_{n-k^-}(A)|\} \leq (1 + \varepsilon) \sqrt{L_1} + C \sqrt{np_{\max}}. \quad (1.15)$$

Under the additional conditions $d_i = d$ for all i and $np_{\max} \leq \eta L_1$, we have for all $k \in [n^{1-\varepsilon}]$

$$\lambda_{k^++k}(A) \sim \sqrt{L_k} \quad \text{and} \quad \lambda_{n+1-(k^-+k)}(A) \sim -\sqrt{L_k}. \quad (1.16)$$

We remark that in the case (a) of small degree, the nontrivial eigenvalues λ_i^\pm of $\mathbb{E}A$ do not give rise to corresponding eigenvalues of A , and A may be regarded as a perturbation of \underline{A} . In contrast, in the case (b) of large degree, the nontrivial eigenvalues λ_i^\pm of $\mathbb{E}A$ gives rise to associated outlier eigenvalues of A , and A may be regarded as a perturbation of $\mathbb{E}A$. Hence, the spectrum of A retains some information about the spectrum of $\mathbb{E}A$ if and only if $d \gg \sqrt{\log n / \log \log n}$.

Remark 1.10. Remarks 1.5 and 1.6 also hold for the eigenvalue counting measure of A . See Figure 1 for an illustration.

Proof of Proposition 1.9. (a) Introduce $\alpha \in \mathbb{R}$ such that

$$d = \left(\frac{\log n}{\log \log n} \right)^\alpha,$$

so that

$$L_1 = \frac{\log n}{(1 - \alpha) \log \log n + \alpha \log \log \log n}.$$

As soon as $\eta < 1$, under the condition $d \leq \eta \sqrt{\log n / \log \log n}$, we have $\alpha < 1/2$, which implies that

$$(1 - \alpha)L_1 \sim d^{1/\alpha}$$

and that

$$\sqrt{L_1} + d \sim \sqrt{L_1}.$$

For $k \leq n^{1-\varepsilon}$, we have $L_k \geq \varepsilon L_1$, so that we also have

$$\sqrt{L_k} + d \sim \sqrt{L_k}.$$

As a consequence, (1.12) follows from (a) of Corollary 1.7 and, considering A as a perturbation of \underline{A} , (1.13) follows from (1.7) and [6, Corollary III.2.6]. (b) In the same way, (1.15) follows from [6, Corollary III.2.3 and Exercise III.2.4], Corollary 1.3 and (1.11), (1.16) follows from (1.7) and [6, Corollary III.2.3 and Exercise III.2.4] and considering, that time, A as a perturbation of $\mathbb{E}A$, (1.14) follows from [6, Corollary III.2.6], Corollary 1.3 and (1.11). \square

1.5 Behaviour of the largest degrees

In our regime of interest, the largest degrees of the graph play a key role in the analysis of the largest eigenvalues of the adjacency matrix. We now describe their asymptotic behavior.

Proposition 1.11. *Let $\varepsilon \in (0, 1)$. Under (1.2), w.h.p., for any $k \in [n^{1-\varepsilon}]$*

$$D_k^\downarrow \leq (1 + \varepsilon)L_k.$$

For any $x > 0$, we introduce the sets

$$\mathcal{V}_{\geq x} := \{i \in [n] : D_i \geq x\}, \quad \mathcal{V}_{=x} := \{i \in [n] : D_i = x\}. \quad (1.17)$$

These sets are related to the ordered degrees through

$$D_k^\downarrow \geq x \iff \#\mathcal{V}_{\geq x} \geq k. \quad (1.18)$$

Let us introduce the function f on (d, ∞) defined by

$$f(x) = f_d(x) := x \log \left(\frac{x}{d} \right) - (x - d) + \log \sqrt{2\pi x}. \quad (1.19)$$

If Y is a Poisson random variable with mean d , for a large integer x Stirling's approximation gives

$$\mathbb{P}(Y = x) = e^{-f(x) + O(x^{-1})}. \quad (1.20)$$

We shall in fact prove that, roughly speaking, under condition (1.2), we have

$$\#\mathcal{V}_{\geq x} \lesssim ne^{f(x)}, \quad (1.21)$$

which, under the additional assumption $d_i = d$ for all i , can be strengthened to

$$\#\mathcal{V}_{\geq x} \approx ne^{f(x)}. \quad (1.22)$$

This leads us to introduce, for $k \in [n]$, the solution Δ_k of the equation

$$f(\Delta_k) = \log(n/k) \quad \text{and} \quad \Delta_k \geq d.$$

This solution is unique and satisfies

$$\Delta_k \sim L_k. \quad (1.23)$$

(See Lemma 2.1 below for the full details.) The combination of the characterization (1.18) of the largest degrees with estimates (1.21), (1.22) and (1.23) naturally leads to the estimates given in Propositions 1.11 and 1.12.

In the special case of a homogenous Erdős-Rényi graph, the next proposition is essentially contained in [7, Chapter 3]. Hence, our next result may be viewed as a generalization of this result to the inhomogeneous case. It is a more precise version of Proposition 1.11 under the additional assumption that all vertices have the same mean degree.

Proposition 1.12. *Let $\varepsilon \in (0, 1)$.*

(a) *For any $k \in [n^{1-\varepsilon}]$ there exists a deterministic $\Delta'_k \in \{[\Delta_k], \lceil \Delta_k \rceil\}$ such that under the conditions $d_i = d$ for all i and (1.2), w.h.p., for any $k \in [n^{1-\varepsilon}]$, we have*

$$(i) \quad D_k^\downarrow \in \{\Delta'_k, \Delta'_k - 1\},$$

$$(ii) \quad D_k^\downarrow = \lfloor \Delta_k \rfloor \text{ when } \text{dist}(\Delta_k, \mathbb{N}) \geq \varepsilon.$$

(b) Under the conditions $d_i = d$ for all i and (1.2), for all integers $t \in [\varepsilon\Delta_1, \Delta_1 - \varepsilon]$,

$$\#\mathcal{V}_{\geq t} \sim \#\mathcal{V}_{=t} \sim ne^{-f(t)}. \quad (1.24)$$

Under the same conditions, if $t \geq \Delta_1 + \varepsilon$ then, w.h.p., $\#\mathcal{V}_{\geq t} = 0$.

An immediate corollary of Proposition 1.12 and (1.23) is the following.

Corollary 1.13. *Let $\varepsilon \in (0, 1)$. Under the conditions $d_i = d$ for all i and (1.2), for all $k \in [n^{1-\varepsilon}]$,*

$$D_k^\downarrow \sim L_k.$$

Remark 1.14 (Lack of limit point process of largest degrees). Proposition 1.12 (b) shows that, perhaps surprisingly, there is no Poisson point process at the right edge of the multiset of degrees of G . There is instead a sharp transition at Δ_1 : for any integer $t \leq \Delta_1 - \varepsilon$, w.h.p. the number of vertices with degree t is $\gg 1$ and for any integer $t \geq \Delta_1 + \varepsilon$, w.h.p. there is no vertex with degree t .

2 Estimation of the largest degrees and comparison with the eigenvalues

The rest of this paper is devoted to the proofs of our main results.

Throughout this section we use the following conventions about convergence of deterministic quantities. Let u and v be deterministic quantities depending on n and $(p_{ij})_{i,j \in [n]}$. We write $u = o(v)$, or, equivalently, $u \ll v$, whenever $u/v \rightarrow 0$ as $n \rightarrow \infty$ and $\eta \rightarrow 0$, uniformly in $(p_{ij})_{i,j \in [n]}$ satisfying (1.2) and all parameters except ε . We remark that such a convergence can always be upgraded to a quantitative convergence using some admissible error function from Definition 1.1, but for the sake of simplicity we shall not do this.

2.1 Largest degrees: proof of Proposition 1.11 and Proposition 1.12

Recall that the function f was defined in (1.19).

Lemma 2.1. *Let $\varepsilon \in (0, 1)$. For n large enough and η small enough, under condition (1.2), for any $k \in [n^{1-\varepsilon}]$, there exists a unique solution Δ_k of the equation*

$$f(\Delta_k) = \log(n/k) \quad \text{and} \quad \Delta_k \geq d. \quad (2.1)$$

Moreover, under condition (1.2), for any $k \in [n^{1-\varepsilon}]$,

$$\Delta_k \sim L_k. \quad (2.2)$$

Proof. The function f is increasing on (d, ∞) (indeed, $f'(u) = \log(u/d) + 1/(2u)$) and satisfies

$$f(u) = u \log(u/d) + O(u) \quad \text{as } u \rightarrow \infty, \quad (2.3)$$

(more precisely, $u^{-1}(f(u) - u \log(u/d))$ is uniformly bounded on $[d, +\infty)$ by a constant depending only on the κ of condition (1.2)) so that for n large enough, Δ_k is well defined for any $1 \leq k \leq n^{1-\varepsilon}$. Moreover, we have

$$f(u+x) = f(u) + x \log(u/d) + O(|x|/u + x^2/u), \quad (2.4)$$

with an implicit constant in $O(\cdot)$ depending only on a positive lower bound on $1 + \frac{x}{u}$ and unconditional to d . Indeed,

$$f(u+x) - f(u) = (u+x) \log\left(1 + \frac{x}{u}\right) + x \log\left(\frac{u}{d}\right) - x + \frac{1}{2} \log\left(1 + \frac{x}{u}\right). \quad (2.5)$$

Let us now prove (2.2). As both $\Delta_k = \Delta_k(n, d)$ and $L_k = L_k(n, d)$ are deterministic and depend only on n and d , by Definition 1.1, (2.2) reads

$$\lim_{\substack{n \rightarrow \infty \\ \eta \rightarrow 0}} \sup_{\kappa \leq d \leq \eta \log n} \sup_{k \in [n^{1-\varepsilon}]} \left| \frac{\Delta_k(n, d)}{L_k(n, d)} - 1 \right| = 0. \quad (2.6)$$

If it were not the case, there would be an infinite set I of positive integers and some sequences $(\eta_n)_{n \in I}$, $(d_n)_{n \in I}$, $(k_n)_{n \in I}$ satisfying $\kappa \leq d_n \leq \eta_n \log n$, $1 \leq k_n \leq n^{1-\varepsilon}$ and $\eta_n \rightarrow 0$ as $n \in I$ tends to infinity, such that

$$\left| \frac{\Delta_{k_n}(n, d_n)}{L_{k_n}(n, d_n)} - 1 \right| > c \quad (2.7)$$

for some positive constant c . Let us drop the index n from the notation. One first verifies that $\Delta_k/d \rightarrow \infty$ as $n \in I$ grows (by a simple argument by contradiction using (2.1)). Then, introduce $\varphi > 0$ such that $\Delta_k = \varphi L_k$. By (2.3), we have

$$\begin{aligned} \log(n/k) &= f(\Delta_k) \\ &\sim \Delta_k \log(\Delta_k/d) \\ &= \varphi \frac{\log(n/k)}{\log((\log n)/d)} (\log(\log(n/k)/d) + \log \varphi - \log \log((\log n)/d)). \end{aligned}$$

By assumption, $\varepsilon \log n \leq \log(n/k) \leq \log n$ so that

$$\frac{1}{\varphi} \sim 1 + \frac{\log \varphi}{\log((\log n)/d)} + o(1). \quad (2.8)$$

One easily deduces from (2.8) that φ is bounded away from 0 and ∞ , and then that φ tends to one, which contradicts (2.7). Thus (2.6), hence also (2.2), are true. \square

Lemma 2.2. *Suppose that n is large enough and η small enough, and that d satisfies condition (1.2), so that Δ_1 is well defined (see Lemma 2.1). Let $q_1, \dots, q_n > 0$ satisfy $d = \sum_i q_i$, and let X be a random variable with law $\text{Bin}(q_1, \dots, q_n)$. Suppose that $q_{\max} := \max_i q_i \leq d/(\log n)^{5/2}$. Then for any $u \leq \Delta_1$ and x such that $x^2 \leq u$ and $u + x \geq 2d$ is integer,*

$$\mathbb{P}(X \geq u + x) = (1 + \delta) \left(\frac{u}{d}\right)^{-x} e^{-f(u)}. \quad (2.9)$$

where $\delta = O((d + x^2)/u + q_{\max} u^{5/2}/d)$.

Proof. By Theorem 3.1 below, $\mathbb{P}(X \geq u + x) = (1 + \delta)\mathbb{P}(Y = u + x)$ where Y has Poisson distribution with mean d and $\delta = O(d/u + q_{\max} u^{5/2}/d)$. Now, by (1.20), we have

$$\mathbb{P}(Y = u + x) = e^{-f(u+x)+O(1/u)}. \quad (2.10)$$

Then, the estimate (2.4) allows to conclude. \square

We are now ready to prove Proposition 1.11.

Proof of Proposition 1.11. Let us first explain how we can swap the quantifiers *w.h.p.* and *for all $k \in [n^{1-\varepsilon}]$* . Suppose that we have proved that for any fixed $\varepsilon \in (0, 1)$, w.h.p., for each $\ell \in [1/\varepsilon]$,

$$D_k^\downarrow \leq (1 + \varepsilon)L_k \quad \text{for } k = \lfloor n^{1-\ell\varepsilon} \rfloor.$$

Let us then fix $\varepsilon \in (0, 1)$. We know that w.h.p., for each $\ell \in [1/(\varepsilon/2)]$,

$$D_k^\downarrow \leq (1 + \varepsilon/2)L_k \quad \text{for } k = \lfloor n^{1-\ell\varepsilon/2} \rfloor.$$

For any $k \in [n^{1-\varepsilon}]$, one can choose $\ell \in [1/(\varepsilon/2)]$ such that

$$k_- := n^{1-\ell\varepsilon/2} \leq k.$$

It follows that w.h.p.

$$D_k^\downarrow \leq D_{k_-}^\downarrow \leq (1 + \varepsilon/2)L_{k_-}.$$

But for n large enough, for any $k \in [n^{1-\varepsilon}]$,

$$(1 + \varepsilon/2)L_{k_-} \leq (1 + \varepsilon)L_k,$$

so the result is proved.

Thus by the union bound, it suffices to fix $\varepsilon \in (0, 1)$ and prove that for a fixed choice of $\ell \in [1/\varepsilon]$ and $k := \lfloor n^{1-\ell\varepsilon} \rfloor$, w.h.p. we have $D_k^\downarrow \leq (1 + \varepsilon)L_k$.

We note that $D_k^\downarrow \geq t$ is equivalent to $\#\mathcal{V}_{\geq t} \geq k$ (the set $\mathcal{V}_{\geq t}$ has been defined in (1.17)). By (2.4), we have

$$ne^{-f(\Delta_k+x)} = k(1 + o(1))(\Delta_k/d)^{-x}. \quad (2.11)$$

Thus (2.11) applied to $x = \lfloor \Delta_k \rfloor + 2 - \Delta_k$ implies that $ne^{-f(\lfloor \Delta_k \rfloor + 2)} = o(k)$. By (2.9) in Lemma 2.2, it follows that $\mathbb{E}\#\mathcal{V}_{\geq t} = o(k)$ if $t = \lfloor \Delta_k \rfloor + 2$. It remains to use (2.2) and Markov's inequality. \square

Our proof of Proposition 1.12 will require a sharp bound on the variance of $\#\mathcal{V}_{\geq t}$.

Lemma 2.3. *Let D_i denote the degree of the vertex i in the graph G . Then any integer $t \geq 0$,*

$$\text{Var}(\#\mathcal{V}_{\geq t}) \leq \mathbb{E}\#\mathcal{V}_{\geq t} + 3dn \max_i \mathbb{P}(D_i \geq t - 1)^2.$$

Proof. For ease of notation, we set $q_i := \mathbb{P}(D_i \geq t)$. Since $\#\mathcal{V}_{\geq t} = \sum_{i=1}^n \mathbb{1}_{D_i \geq t}$, we have $\mathbb{E}\#\mathcal{V}_{\geq t} = \sum_{i=1}^n q_i$ and

$$\begin{aligned} \text{Var}(\#\mathcal{V}_{\geq t}) &= \sum_{i,j} \text{Cov}(\mathbb{1}_{D_i \geq t}, \mathbb{1}_{D_j \geq t}) \\ &= \sum_i q_i(1 - q_i) + \sum_{i \neq j} \text{Cov}(\mathbb{1}_{D_i \geq t}, \mathbb{1}_{D_j \geq t}). \end{aligned}$$

Hence it suffices to prove that for $i \neq j$,

$$\text{Cov}(\mathbb{1}_{D_i \geq t}, \mathbb{1}_{D_j \geq t}) \leq 3p_{ij} \max_i \mathbb{P}(D_i \geq t - 1)^2.$$

Let us fix $i \neq j$. We have $D_i = \sum_k A_{ik}$ and $D_j = \sum_k A_{jk}$. We introduce the events $C := \{A_{ij} = 0\}$,

$$E_0 := \left\{ \sum_{k \neq j} A_{ik} \geq t \right\}, \quad E := E_0 \cap C, \quad E' := \left\{ \sum_k A_{ik} \geq t \right\},$$

$$F_0 := \left\{ \sum_{k \neq i} A_{jk} \geq t \right\}, \quad F := F_0 \cap C, \quad F' := \left\{ \sum_k A_{jk} \geq t \right\}.$$

Then $E \subset E'$ and $F \subset F'$. Thus by Lemma A.1,

$$\begin{aligned} \text{Cov}(\mathbb{1}_{D_i \geq t}, \mathbb{1}_{D_j \geq t}) &= \text{Cov}(\mathbb{1}_{E'}, \mathbb{1}_{F'}) \\ &= \text{Cov}(\mathbb{1}_E, \mathbb{1}_F) + \mathbb{P}(F' \setminus F)\mathbb{P}(E) + \mathbb{P}(E' \setminus E)\mathbb{P}(F) \\ &= \mathbb{P}(E_0 \cap F_0 \cap C) - \mathbb{P}(E_0 \cap C)\mathbb{P}(F_0 \cap C) + \mathbb{P}(F' \setminus F)\mathbb{P}(E) + \mathbb{P}(E' \setminus E)\mathbb{P}(F) \end{aligned}$$

Then, note that

$$E' \setminus E = (E' \setminus E_0) \cup (E' \setminus C) \subset C^c \cap \left\{ \sum_{k \neq j} A_{ik} \geq t - 1 \right\}$$

and in the same way,

$$F' \setminus F \subset C^c \cap \left\{ \sum_{k \neq i} A_{jk} \geq t - 1 \right\},$$

hence by independence of the events E_0 , F_0 and C ,

$$\begin{aligned} \text{Cov}(\mathbb{1}_{D_i \geq t}, \mathbb{1}_{D_j \geq t}) &= \mathbb{P}(E_0)\mathbb{P}(F_0)\mathbb{P}(C)\mathbb{P}(C^c) + \mathbb{P}(F' \setminus F)\mathbb{P}(E) + \mathbb{P}(E' \setminus E)\mathbb{P}(F) \\ &\leq 3p_{ij}\mathbb{P}\left(\sum_{k \neq j} A_{ik} \geq t - 1\right)\mathbb{P}\left(\sum_{k \neq i} A_{jk} \geq t - 1\right), \end{aligned}$$

which allows to conclude. \square

We are ready to prove Proposition 1.12.

Proof of Proposition 1.12. First we remark that the inversion of w.h.p. and for all for all integers $t \in [\varepsilon\Delta_1, \Delta_1 - \varepsilon]$ for (b) or for all $k \in [n^{1-\varepsilon}]$ for (a) can be treated as in the proof of Proposition 1.11.

- (b) By (2.9) in Lemma 2.2, if $t \geq \Delta_1 + \varepsilon$, $\mathbb{P}(\#\mathcal{V}_{\geq t} \geq 1) \leq \mathbb{E}\#\mathcal{V}_{\geq t} \leq (1+o(1))(d/\Delta_1)^\varepsilon$ tends to 0. Moreover, in the regime $\varepsilon\Delta_1 \leq t \leq \Delta_1 - \varepsilon$, as $ne^{-f(t)}$ goes to infinity, to prove

$$\#\mathcal{V}_{\geq t} \sim ne^{-f(t)}$$

by Markov's inequality it suffices to prove that $\mathbb{E}\#\mathcal{V}_{\geq t} = ne^{-f(t)}(1 + o(1))$ and $\text{Var}(\#\mathcal{V}_{\geq t}) \leq ne^{-f(t)}(1 + o(1))$, which follows directly from (2.9) in Lemma 2.1 and Lemma 2.3.

It remains to prove that $\#\mathcal{V}_{=t} \sim ne^{-f(t)}$ when $\varepsilon\Delta_1 \leq t \leq \Delta_1 - \varepsilon$. We note that

$$\#\mathcal{V}_{=t} = \#\mathcal{V}_{\geq t} - \#\mathcal{V}_{\geq t+1}.$$

From what precedes, it suffices to check that $\#\mathcal{V}_{t+1} = o(ne^{-f(t)})$. The latter is a consequence of (2.9) in Lemma 2.1 which implies that $\mathbb{E}\#\mathcal{V}_{t+1} \leq (1 + o(1))(d/t)ne^{-f(t)}$.

- (a) We note that for any k, t , the claim $D_k^\downarrow \geq t$ is equivalent to $\#\mathcal{V}_{\geq t} \geq k$. Thus (2.11) applied to $x = \lfloor \Delta_k \rfloor - 1 - \Delta_k$ and (b) imply that with high probability $D_k^\downarrow \geq \lfloor \Delta_k \rfloor - 1$. Similarly, (2.11) applied to $x = \lfloor \Delta_k \rfloor + 2 - \Delta_k$ and (b) imply with high probability $D_k^\downarrow \leq \lfloor \Delta_k \rfloor + 1$. Moreover, if $\lfloor \Delta_k \rfloor + 1 - \Delta_k \geq \varepsilon$, then with high probability $D_k^\downarrow \leq \lfloor \Delta_k \rfloor$, while if $\lfloor \Delta_k \rfloor - \Delta_k \leq -\varepsilon$ then with high probability $D_k^\downarrow \geq \lfloor \Delta_k \rfloor$. Note that either $\lfloor \Delta_k \rfloor - \Delta_k \leq -\varepsilon$ or $\lfloor \Delta_k \rfloor + 1 - \Delta_k \geq \varepsilon$ holds when $\varepsilon = 1/2$. \square

2.2 Proof of Theorem 1.2

First it is easy to see, by Weyl's inequality [6, Cor. III.2.6], that we may assume without loss of generality that $p_{ii} = 0$ for all $i \in [n]$. Indeed, changing the values of the p_{ii} amounts, in the right coupling, to change degrees by at most 1 and to add to the matrices A and \underline{A} a perturbation of operator norm at most 1, which can for example, at fixed κ , be absorbed by the first error term if C is large enough. As pointed in introduction, our strategy is to describe the graph spanned by the vertices of high degree. We start with a deviation inequality on the degrees. Define

$$h(x) := (1+x)\log(1+x) - x. \quad (2.12)$$

Lemma 2.4. *For distinct $i_1, \dots, i_k \in [n]$ and $t \geq 0$ we have*

$$\mathbb{P}\left(\sum_{\ell=1}^k D_{i_\ell} \geq k(d+t)\right) \leq \exp\left(-kdh\left(\frac{t}{d}\right) + k^2 p_{\max}\left(\frac{t}{d} + 1\right)^2\right). \quad (2.13)$$

In particular, if $d \ll t$ and $kt/\log(t/d) \ll d^2/p_{\max}$, we have

$$\mathbb{P}\left(\sum_{\ell=1}^k D_{i_\ell} \geq kt\right) \leq \exp\left(-(1+o(1))kt \log\left(\frac{t}{d}\right)\right). \quad (2.14)$$

Proof. We have

$$\mathbb{P}\left(\sum_{\ell=1}^k D_{i_\ell} \geq k(d+t)\right) \leq \mathbb{P}\left(\sum_{\ell=1}^k (D_{i_\ell} - \mathbb{E}D_{i_\ell}) \geq kt\right).$$

Now,

$$\sum_{\ell=1}^k D_{i_\ell} = 2 \sum_{\ell=1}^k \sum_{j=\ell+1}^k A_{i_\ell i_j} + \sum_{\ell=1}^k \sum_{j \notin \{i_1, \dots, i_k\}} A_{i_\ell j} =: 2S_1 + S_2,$$

where S_1 and S_2 are independent. From Chernoff's inequality [8, p. 21], for any $\lambda \geq 0$,

$$\mathbb{P}\left(\sum_{\ell=1}^k (D_{i_\ell} - \mathbb{E}D_{i_\ell}) \geq kt\right) \leq \exp(-\lambda kt + \alpha\phi(\lambda) + \beta\phi(2\lambda)),$$

where $\phi(\lambda) = e^\lambda - \lambda - 1$,

$$\alpha = \sum_{\ell=1}^k \sum_{j \notin \{i_1, \dots, i_k\}} p_{i_\ell j} \quad \text{and} \quad \beta = 2 \sum_{\ell=1}^k \sum_{j=\ell+1}^k p_{i_\ell i_j} = \sum_{\ell=1}^k \sum_{j=1}^k p_{i_\ell i_j}.$$

By hypothesis, $\alpha + \beta \leq kd$ and $\beta \leq k^2 p_{\max}$. We take $\lambda = \log(t/d + 1)$ and use that $\phi(2\lambda) \leq e^{2\lambda}$, we arrive at (2.13). The second claim (2.14) is an immediate consequence of the fact that the function h from (2.12) satisfies $h(x) \sim x \log x$ when x goes to infinity. \square

Lemma 2.5. *There exists a constant $c > 0$ such that the following holds. Let $\varepsilon \in (0, 1)$ and $t = \varepsilon L_1$. For $S \subset \{1, \dots, n\}$, let $\mathcal{N}(S) = \{j : A_{ij} = 1 \text{ for some } i \in S\}$ denote the set of neighbours of elements in S . Then, under (1.2), w.h.p. for any $i \in \mathcal{V}_{\geq t}$ we have*

$$\#[\mathcal{N}(\{i\}) \cap (\mathcal{V}_{\geq t} \cup \mathcal{N}(\mathcal{V}_{\geq t} \setminus \{i\}))] \leq c/\varepsilon.$$

Proof. Fix an integer k . Let $P(1)$ be the probability that there exists a vertex of $\mathcal{V}_{\geq t}$ which is neighbour to at least k other elements of $\mathcal{V}_{\geq t}$. We have

$$\begin{aligned} P(1) &\leq \sum_{\substack{i_0, \dots, i_k \\ \text{distinct}}} \mathbb{P}(D_{i_0} \geq t \text{ and } \forall \ell \in \{1, \dots, k\} : D_{i_\ell} \geq t, A_{i_0 i_\ell} = 1) \\ &\leq \sum_{\substack{i_0, \dots, i_k \\ \text{distinct}}} \mathbb{P}(D_{i_0} \geq t - k) \mathbb{P}(\forall \ell \in \{1, \dots, k\} : D_{i_\ell} \geq t - 1) \prod_{\ell=1}^k p_{i_0 i_\ell} \end{aligned}$$

Since for any fixed i_0 we have

$$\sum_{\substack{i_1, \dots, i_k \\ \text{distinct}}} \prod_{\ell=1}^k p_{i_0 i_\ell} \leq d^k \quad (2.15)$$

we deduce that

$$\begin{aligned} P(1) &\leq nd^k \max_i \mathbb{P}(D_i \geq t - k) \max_{\substack{i_1, \dots, i_k \\ \text{distinct}}} \mathbb{P}(\forall \ell \in \{1, \dots, k\} : D_{i_\ell} \geq t - 1) \\ &\leq nd^k \exp\left(- (1 + o(1))(k + 1)t \log \frac{t}{d}\right) \end{aligned}$$

if $t \gg d$ and $tp_{\max} \ll d^2 \log(t/d)$ (from (2.14)). For $t = \varepsilon L_1 \gg d$ and k fixed such that $k + 1 > a/\varepsilon$ with $a > 1$. We find

$$P(1) \leq n^{1-a(1+o(1))} = o(1).$$

Similarly, let $P(2)$ be the probability that there exists a vertex $i \in \mathcal{V}_{\geq t}$ which is neighbour to at least k elements of $\mathcal{N}(\mathcal{V}_{\geq t} \setminus \{i\}) \setminus \mathcal{V}_{\geq t}$. Then

$$P(2) \leq \sum_{s=1}^k \sum_{\tau} \sum_{\substack{i_0, \dots, i_s, j_1, \dots, j_k \\ \text{distinct}}} \mathbb{P}\left(D_{i_0}, \dots, D_{i_s} \geq t \text{ and } \forall \ell = 1, \dots, k : A_{i_0 j_\ell} = A_{j_\ell i_{\tau(\ell)}} = 1\right),$$

where the second sum is over all surjective maps $\tau : \{1, \dots, k\} \rightarrow \{1, \dots, s\}$.

We deduce that

$$P(2) \leq \sum_{s=1}^k \sum_{\tau} \sum_{\substack{i_0, \dots, i_s, j_1, \dots, j_k \\ \text{distinct}}} \mathbb{P}(D_{i_0} \geq t - s) \mathbb{P}(\forall r \in \{1, \dots, s\} : D_{i_r} \geq t - \#\tau^{-1}(\{r\})) \prod_{\ell=1}^k p_{i_0 j_\ell} p_{j_\ell i_{\tau(\ell)}}.$$

Now, note that for any fixed surjective map $\tau : [k] \rightarrow [s]$,

- for any i_1, \dots, i_s , we have

$$\{\forall r \in \{1, \dots, s\} : D_{i_r} \geq t - \#\tau^{-1}(\{r\})\} \subset \left\{ \sum_{r=1}^s D_{i_r} \geq st - k \right\};$$

- for any fixed j_1, \dots, j_k , we have, as in (2.15),

$$\sum_{\substack{i_1, \dots, i_s \\ \text{distinct}}} \prod_{\ell=1}^k p_{j_\ell i_{\tau(\ell)}} \leq d^s p_{\max}^{k-s}.$$

We deduce, using (2.15) and (2.14) again, that

$$\begin{aligned} P(2) &\leq \sum_{s=1}^k \sum_{\tau} n d^{k+s} p_{\max}^{k-s} \max_i \mathbb{P}(D_i \geq t - k) \max_{\substack{i_1, \dots, i_s \\ \text{distinct}}} \mathbb{P}\left(\sum_{\ell=1}^s D_{i_\ell} \geq st - k\right) \\ &\leq n d^k \sum_{s=1}^k \binom{k}{s} k^{k-s} d^s p_{\max}^{k-s} e^{-(1+o(1))(s+1)t \log(\frac{t}{d})}. \end{aligned}$$

The $o(1)$ is uniform over $1 \leq s \leq k$. Hence,

$$P(2) \leq n d^k e^{-(1+o(1))t \log(\frac{t}{d})} \left(d e^{-(1+o(1))t \log(\frac{t}{d})} + k p_{\max} \right)^k.$$

As above for $t = \varepsilon L_1 \gg d$ and k a fixed integer such that $k + 1 > a/\varepsilon$ with $a > 1$. We find $P(2) = o(1)$. This concludes the proof of the first claim of the lemma. \square

Definition 2.6. A *star graph with central degree D* is a graph with $D + 1$ vertices $\{i_0, i_1, \dots, i_D\}$ and edges $\{\{i_0, i_1\}, \{i_0, i_2\}, \dots, \{i_0, i_D\}\}$.

We may now prove Theorem 1.2.

Proof of Theorem 1.2. Let $0 < \delta < \varepsilon$ and set $t := \delta L_1$. By Proposition 1.11, w.h.p. for any $k \in [n^{1-\varepsilon}]$,

$$D_k^\downarrow \geq t.$$

Let G_\star be the graph obtained from G as follows. The vertex set of G_\star is $[n]$. The edge set of G_\star is the set of edges $\{i, j\}$ of G (i.e. $A_{ij} = 1$) such that $i \in \mathcal{V}_{\geq t}$ and $j \notin \mathcal{V}_{\geq t} \cup \mathcal{N}(\mathcal{V}_{\geq t} \setminus \{i\})$ (where the notation $\mathcal{N}(\cdot)$ was introduced in Lemma 2.5). By construction, G_\star is a disjoint union of isolated vertices and of star graphs with central degrees D_i^\star , $i \in \mathcal{V}_{\geq t}$. By Lemma 2.5, w.h.p., the central degrees of the stars satisfy $D_i - c/\delta \leq D_i^\star \leq D_i$. Let A_\star be the adjacency matrix of G_\star and let $A' = A - A_\star$ be the adjacency matrix of $G \setminus G_\star$.

By Lemma A.2, the nonzero eigenvalues of A_\star are $\pm \sqrt{D_i^\star}$, $i \in \mathcal{V}_{\geq t}$. From what precedes, w.h.p. for all $i \in \mathcal{V}_{\geq t}$,

$$\left| \sqrt{D_i^\star} - \sqrt{D_i} \right| = \frac{D_i - D_i^\star}{\sqrt{D_i} + \sqrt{D_i^\star}} \leq \frac{c}{\delta \sqrt{t}}. \quad (2.16)$$

for c the universal constant of Lemma 2.5.

Besides, w.h.p. the maximal degree in $G \setminus G_\star$ is bounded by $\max(t, c/\delta)$ (which is asymptotically equal to t). Indeed, let $i \in [n]$ be a vertex. If the

degree of i in G is $< t$, then there is nothing to prove (as the degree of i in $G \setminus G_\star$ is bounded by its degree in G), whereas if $D_i \geq t$, then the degree of i in $G \setminus G_\star$ is $D_i - D_i^\star \leq c/\delta$.

By Proposition 1.11, we know that w.h.p., the cardinal number of $\mathcal{V}_{\geq t}$ is at most $n^{1-\delta+o(1)} \leq 10/p_{\max}$, hence by Theorem A.3 of Le, Levina, and Vershynin, w.h.p.

$$\|A' - \mathbb{E}A\| \leq C((np_{\max})^{1/2} + (\delta L_1)^{1/2}). \quad (2.17)$$

Then, one concludes thanks to Weyl's perturbation inequality (see e.g. [6, Corollary III.2.6]), noticing that the constants from (2.16) and (2.17) do not depend on the choice of $\delta > 0$. \square

3 Poisson tail approximation

The following tail asymptotic of $\text{Bin}(p_1, \dots, p_n)$ is stronger than what can be deduced from Le Cam's inequality (see [3]).

Theorem 3.1. *Let X with distribution $\text{Bin}(p_1, \dots, p_n)$, $d = p_1 + \dots + p_n > 0$ and $p_{\max} = \max_i p_i \geq d/n$. Let Y be a Poisson variable with mean d . There exists a universal constant $C > 0$ such that for any integer k satisfying $2d \leq k \leq (d/p_{\max})^{2/5}/C$, we have*

$$\left| \frac{\mathbb{P}(X = k)}{\mathbb{P}(Y = k)} - 1 \right| \leq C \frac{p_{\max} k^{5/2}}{d}, \quad (3.1)$$

and

$$\mathbb{P}(X > k) \leq C \left(\frac{d}{k} + \frac{p_{\max} k^{5/2}}{d} \right) \mathbb{P}(X = k). \quad (3.2)$$

Remark 3.2. The estimate (3.1) is contained in [2, (1.8)], which is however given without proof. It can also be proved using a steepest descent analysis [18]. Here, we give a probabilistic proof of (3.1) based on a comparison principle. Our argument also yields the bound (3.2), which is a key ingredient of the proof of Theorem 1.2.

We first check that Theorem 3.1 holds for standard binomial variable.

Lemma 3.3 (Tails of binomial laws). *Let Z be distributed according to the binomial distribution with parameters $(n, d/n)$. There exists a universal constant $C > 0$ such that for any integer k with $2d \leq k \leq \sqrt{n}/C$,*

$$\left| \frac{\mathbb{P}(Z = k)}{\mathbb{P}(Y = k)} - 1 \right| \leq \frac{Ck^2}{n}, \quad (3.3)$$

and

$$\mathbb{P}(Z > k) \leq \frac{Cd}{k} \mathbb{P}(Z = k). \quad (3.4)$$

Proof. We have

$$\frac{\mathbb{P}(Z = k)}{\mathbb{P}(Y = k)} = \frac{n!}{(n-k)!(n-d)^k} \left(1 - \frac{d}{n}\right)^n e^d$$

using $(n-k)^k \leq \frac{n!}{(n-k)!} \leq n^k$, we get that

$$\left(1 - \frac{k}{n}\right)^k \left(1 - \frac{d}{n}\right)^n e^d \leq \frac{\mathbb{P}(Z = k)}{\mathbb{P}(Y = k)} \leq \left(\frac{n}{n-d}\right)^k \leq e^{kd/(n-d)}. \quad (3.5)$$

Using $\log(1+x) \leq x$, we get $\left(\frac{n}{n-d}\right)^k \leq e^{kd/(n-d)}$. Then, it is easy to see that there is C such that as soon as $d, k \leq \sqrt{n}$ and $n \geq 2$, we have $e^{kd/(n-d)} \leq 1 + C \frac{kd}{n}$, so that

$$\frac{\mathbb{P}(Z = k)}{\mathbb{P}(Y = k)} \leq 1 + C \frac{kd}{n}. \quad (3.6)$$

For the lower bound, first note that there is C such that for any $1 \leq k \leq n$,

$$1 - \frac{Ck^2}{n} \leq \left(1 - \frac{k}{n}\right)^k \quad (3.7)$$

(indeed, it comes down to prove that there is a constant C such that for any $n \geq 1$, for $x \in (0, (Cn)^{-1/2})$, $\log(1 - Cnx^2) \leq nx \log(1 - x)$, which is easily obtained thanks to the series expansion). Then note that there is C such that as soon as $d \leq \sqrt{n}$ and $n \geq 2$,

$$\log\left(1 - \frac{d}{n}\right) + \frac{d}{n} \geq -C \left(\frac{d}{n}\right)^2 \quad \text{and} \quad e^{-nC(d/n)^2} \geq 1 - C \frac{d^2}{n}. \quad (3.8)$$

From (3.5), (3.7) and (3.8), we deduce that

$$\frac{\mathbb{P}(Z = k)}{\mathbb{P}(Y = k)} \geq \left(1 - C \frac{k^2}{n}\right) \left(1 - C \frac{d^2}{n}\right). \quad (3.9)$$

The claim (3.3) then follows from (3.6) and (3.9).

Note that for any integer $j \geq d$,

$$\mathbb{P}(Z = j+1) = \mathbb{P}(Z = j) \frac{(n-j)_+}{n-d} \frac{d}{j+1} \leq \frac{d}{j+1} \mathbb{P}(Z = j).$$

We deduce that for $\varepsilon := \frac{d}{k}$,

$$\mathbb{P}(Z > k) \leq \mathbb{P}(Z = k) \sum_{\ell \geq 1} \varepsilon^\ell = \frac{\varepsilon}{1 - \varepsilon} \mathbb{P}(Z = k).$$

This concludes the proof. \square

The classical Bennett's inequality gives a first tail bound for the distribution $\text{Bin}(p_1, \dots, p_n)$.

Lemma 3.4 (Half of Bennett's inequality). *Let X with distribution $\text{Bin}(p_1, \dots, p_n)$ and set $d = p_1 + \dots + p_n$. Then for any $t \geq 0$, we have*

$$\mathbb{P}(X \geq d(1 + t)) \leq \exp\{-dh(t)\} \quad (3.10)$$

where h was defined in (2.12).

The next crucial lemma will use the Lindeberg's replacement principle to compare the distributions $\text{Bin}(p_1, \dots, p_n)$ and $\text{Bin}(p, \dots, p)$ with $p = \frac{1}{n} \sum_i p_i$.

Lemma 3.5 (Comparison principle). *Let $P = \text{Bin}(p_1, \dots, p_n)$ and $Q = \text{Bin}(d/n, \dots, d/n)$ with $d := \sum_{i=1}^n p_i$. For any integer k , we have*

$$|P(k) - Q(k)| \leq 2dp_{\max} e^{-dh((\frac{k-2-d}{d})_+)} \quad \text{and} \quad |P(k, \infty) - Q(k, \infty)| \leq dp_{\max} e^{-dh((\frac{k-2-d}{d})_+)}.$$

Proof. Fix the integers $1 \leq i, j \leq n$. We write $f(t) = P(t)$ and $f^{ij}(t)$ is the probability of t for the distribution $\text{Bin}((p_l)_{l \neq \{i, j\}})$. We have

$$f(t) = (1-p_i)(1-p_j)f^{ij}(t) + p_i(1-p_j)f^{ij}(t-1) + p_j(1-p_i)f^{ij}(t-1) + p_i p_j f^{ij}(t-2). \quad (3.11)$$

As $(1-p_i)(1-p_j) + p_i(1-p_j) + p_j(1-p_i) + p_i p_j = 1$, we first deduce from (3.11) that for any $k \geq 0$,

$$P(k, \infty) = \sum_{t \geq k} f(t) \geq \sum_{t \geq k} f^{ij}(t). \quad (3.12)$$

We also deduce from (3.11) that $f(t)$ is a polynomial of order 2 in (p_i, p_j) . Interestingly, the term of order 1 is symmetric and equal to

$$(p_i + p_j)(f^{ij}(t-1) - f^{ij}(t)).$$

The second order term is equal to

$$p_i p_j (f^{ij}(t) - 2f^{ij}(t-1) + f^{ij}(t-2)).$$

Now, let (q_1, \dots, q_n) be such that $p_l = q_l$ for $l \notin \{i, j\}$, $q_i + q_j = p_i + p_j$. Hence, if $g(t)$ is the probability of t for $\text{Bin}(q_1, \dots, q_n)$, we find

$$f(t) - g(t) = (p_i p_j - q_i q_j)(f^{ij}(t) - 2f^{ij}(t-1) + f^{ij}(t-2)).$$

and

$$\sum_{t \geq k} (f(t) - g(t)) = (p_i p_j - q_i q_j)(f^{ij}(k-2) - f^{ij}(k-1)).$$

Assume that $p_i p_j$ and $q_i q_j$ are both bounded by dp_{\max}/n . Then from the previous equations and from (3.12), we deduce that

$$|f(k) - g(k)| \leq 2 \frac{dp_{\max}}{n} P(k-2, \infty) \quad \text{and} \quad \left| \sum_{t \geq k} (f(t) - g(t)) \right| \leq \frac{dp_{\max}}{n} P(k-2, \infty).$$

By Lemma 3.4, we find that

$$|f(k) - g(k)| \leq 2 \frac{dp_{\max}}{n} e^{-dh((\frac{k-2-d}{d})_+)} \quad \text{and} \quad \left| \sum_{t \geq k} (f(t) - g(t)) \right| \leq \frac{dp_{\max}}{n} e^{-dh((\frac{k-2-d}{d})_+)}. \quad (3.13)$$

Now, if $(p_1, \dots, p_n) \neq (d/n, \dots, d/n)$ then there exists (i, j) such that $p_i < d/n < p_j$ (since the average d/n is in the convex hull of (p_1, \dots, p_n)). We consider (q_1, \dots, q_n) as above such that $q_i = d/n$ and $p_i < q_j = p_i + p_j - d/n < p_j$. Then the bound (3.13) applies here and $q_{\max} = \max_k q_k \leq p_{\max}$. We may thus repeat the same operation to $p^1 = (q_1, \dots, q_n)$ and get p^2 and so on. After $m \leq n$ iterations, we arrive at $p^m = (d/n, \dots, d/n)$. Summing the m times (3.13) gives

$$|P(k) - Q(k)| \leq 2m \frac{dp_{\max}}{n} e^{-dh((\frac{k-2-d}{d})_+)} \leq 2dp_{\max} e^{-dh((\frac{k-2-d}{d})_+)}.$$

This gives the first claim. The same argument applied to the right-hand side of (3.13) gives the second claim. \square

We are finally ready for the proof of Theorem 3.1

Proof of Theorem 3.1. Let Z be a random variable with distribution $\text{Bin}(d/n, \dots, d/n)$. In view of Lemma 3.3, it sufficient to prove (up to adjusting the universal constant $C > 0$) that

$$|\mathbb{P}(X > k) - \mathbb{P}(Z > k)| + |\mathbb{P}(X = k) - \mathbb{P}(Z = k)| \leq C \frac{p_{\max} k^{\frac{5}{2}}}{d} \mathbb{P}(Y = k).$$

Then, from Lemma 3.5, we deduce that it is sufficient to check that for $k \geq d$,

$$\frac{d e^{-dh\left(\left(\frac{k-2-d}{d}\right)_+\right)}}{\mathbb{P}(Y = k)} \leq C \frac{k^{\frac{5}{2}}}{d}. \quad (3.14)$$

First, from Stirling's formula, we find

$$\mathbb{P}(Y = k) = \frac{d^k}{k!} e^{-d} \geq \frac{e^{-k \log(k/d) + k - d}}{C\sqrt{k}} = \frac{e^{-dh\left(\frac{k-d}{d}\right)}}{C\sqrt{k}}.$$

Secondly, from the convexity of $x \mapsto h(x_+)$, we find

$$dh\left(\left(\frac{k-2-d}{d}\right)_+\right) = dh\left(\left(\frac{k-d}{d} - \frac{2}{d}\right)_+\right) \geq dh\left(\frac{k-d}{d}\right) - 2 \log\left(\frac{k}{d}\right).$$

It follows that

$$\frac{e^{-dh\left(\left(\frac{k-2-d}{d}\right)_+\right)}}{\mathbb{P}(Y = k)} \leq C\sqrt{k} \left(\frac{k}{d}\right)^2.$$

This concludes the proof of (3.14). \square

A Auxiliary results

Lemma A.1. *Let, in a probability space, $E \subset E'$ and $F \subset F'$ be some events. Assume that $(E' \setminus E) \cap F = (F' \setminus F) \cap E = \emptyset$. Then*

$$\text{Cov}(\mathbb{1}_{E'}, \mathbb{1}_{F'}) = \text{Cov}(\mathbb{1}_E, \mathbb{1}_F) + \mathbb{P}(E' \setminus E)\mathbb{P}(F) + \mathbb{P}(F' \setminus F)\mathbb{P}(E).$$

Proof. Set $E'' := E' \setminus E$ and $F'' := F' \setminus F$. We have

$$\begin{aligned} & \text{Cov}(\mathbb{1}_{E'}, \mathbb{1}_{F'}) - \text{Cov}(\mathbb{1}_E, \mathbb{1}_F) \\ &= \mathbb{P}(E' \cap F') - \mathbb{P}(E \cap F) - (\mathbb{P}(E')\mathbb{P}(F') - \mathbb{P}(E)\mathbb{P}(F)) \\ &= \mathbb{P}((E' \cap F') \setminus (E \cap F)) - [(\mathbb{P}(E) + \mathbb{P}(E''))(\mathbb{P}(F) + \mathbb{P}(F'')) - \mathbb{P}(E)\mathbb{P}(F)] \end{aligned}$$

As

$$(E' \cap F') \setminus (E \cap F) = (E \cap F'') \cup (E'' \cap F) \cup (E'' \cap F''),$$

we conclude easily. \square

Lemma A.2. *The adjacency matrix A of a star graph with central degree $D \geq 1$ (see Definition 2.6) is a $(D+1) \times (D+1)$ real symmetric matrix with nonzero eigenvalues $\pm\sqrt{D}$ and associated eigenvectors $(\pm\sqrt{D}, 1, \dots, 1)$ (the first coordinate corresponds to the centre of the star).*

Proof. One notices that the matrix has rank 2 and that the vectors given here are actually eigenvectors for $\pm\sqrt{D}$. \square

The following result is [15, Theorem 2.1]. It concerns general inhomogeneous Erdős-Rényi graphs with mean adjacency matrix $(p_{ij})_{i,j \in [n]}$. We recall that a weighted graph has adjacency matrix A' whose entries are non-negative real numbers, with the entry $A'_{ij} \geq 0$ denoting the weight of the edge $\{i, j\}$.

Theorem A.3. *Set $p_{\max} := \max_{i,j} p_{ij}$ and choose $r \geq 1$. Then the following holds with probability at least $1 - n^{-r}$. Consider a subset of at most $10p_{\max}^{-1}$ vertices and reduce the weights of the edges incident to those vertices in an arbitrary way. Then the adjacency matrix A' of the new (weighted) graph satisfies*

$$\|A' - \mathbb{E}A\| \leq Cr^{3/2}(\sqrt{np_{\max}} + \sqrt{d'}),$$

where C is a constant independent of r , $d' = \max_{1 \leq i \leq n} \|R'_i\|_{\ell^1}$ with R'_1, \dots, R'_n the rows of A' .

References

- [1] A. Auffinger, G. Ben Arous, and S. Péché. Poisson convergence for the largest eigenvalues of heavy tailed random matrices. *Ann. Inst. Henri Poincaré Probab. Stat.*, 45(3):589–610, 2009.
- [2] A. D. Barbour *Topics in Poisson Approximation*, D. N. Shanbhag and C. R. Rao, eds., Handbook of Statistics, Vol. 19 A, 2001 Elsevier Science.
- [3] A. D. Barbour, L. Holst, and S. Janson. *Poisson approximation*, volume 2 of *Oxford Studies in Probability*. The Clarendon Press, Oxford University Press, New York, 1992. Oxford Science Publications.
- [4] F. Benaych-Georges, C. Bordenave, and A. Knowles. Spectral radii of sparse random matrices. Preprint arXiv:1704.02945, 2017.
- [5] F. Benaych-Georges and S. Péché. Localization and delocalization for heavy tailed band matrices. *Ann. Inst. Henri Poincaré Probab. Stat.*, 50(4):1385–1403, 2014.
- [6] R. Bhatia. *Matrix analysis*, volume 169 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 1997.

- [7] B. Bollobás. *Random graphs*, volume 73 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, second edition, 2001.
- [8] S. Boucheron, G. Lugosi, and P. Massart. *Concentration inequalities*. Oxford University Press, Oxford, 2013. A nonasymptotic theory of independence, With a foreword by Michel Ledoux.
- [9] L. Erdős, A. Knowles, H.-T. Yau, and J. Yin, Spectral Statistics of Erdős-Rényi Graphs II: Eigenvalue Spacing and the Extreme Eigenvalues, *Comm. Math. Phys.*, 314:587–640, 2012.
- [10] L. Erdős, A. Knowles, H.-T. Yau, and J. Yin, Spectral Statistics of Erdős-Rényi Graphs I: Local Semicircle Law, *Ann. Prob.*, 41:2279–2375, 2013.
- [11] U. Feige and E. Ofek. Spectral techniques applied to sparse random graphs. *Random Structures Algorithms*, 27(2):251–275, 2005.
- [12] Z. Füredi and J. Komlós. The eigenvalues of random symmetric matrices. *Combinatorica*, 1(3):233–241, 1981.
- [13] M. Krivelevich and B. Sudakov. The largest eigenvalue of sparse random graphs. *Combin. Probab. Comput.*, 12(1):61–72, 2003.
- [14] R. Latała, R. van Handel, and P. Youssef. The dimension-free structure of nonhomogeneous random matrices. *Preprint arXiv:1711.00807*, 2017.
- [15] C. M. L. Le, E. Levina, and R. Vershynin. Concentration and regularization of random graphs. *arXiv:1506.00669*, 2015.
- [16] J. O. Lee and J. Yin. A necessary and sufficient condition for edge universality of Wigner matrices. *Duke Math. J.*, 163(1):117–173, 2014.
- [17] J. O. Lee and K. Schnelli, Local law and Tracy-Widom limit for sparse random matrices, *Preprint arXiv:1605.08767*.
- [18] S. Sodin, *Private communication*.
- [19] A. Soshnikov. Poisson statistics for the largest eigenvalues of Wigner random matrices with heavy tails. *Electr. Comm. Prob.*, 9:82–91, 2004.
- [20] V. H. Vu. Spectral norm of random matrices. *Combinatorica*, 27(6):721–736, 2007.

Florent Benaych-Georges, MAP 5 (CNRS UMR 8145) - Université Paris Descartes, 45 rue des Saints-Pères 75270 Paris cedex 6, France. Email: florent.benaych-georges@parisdescartes.fr.

Charles Bordenave, Institut de Mathématiques (CNRS UMR 5219) - Université Paul Sabatier. 31062 Toulouse cedex 09. France. Email: bordenave@math.univ-toulouse.fr.

Antti Knowles, University of Geneva, Section of Mathematics, 2-4 Rue du Lièvre, 1211 Genève 4, Switzerland. Email: antti.knowles@unige.ch.

Acknowledgements C. B. is supported by grants ANR-14-CE25-0014 and ANR-16-CE40-0024-01. A. K. is supported by the Swiss National Science Foundation and the European Research Council.