THE FIRST EIGENVALUE OF RANDOM GRAPHS

SVANTE JANSON

To Béla Bollobás on his 60th birthday

Abstract. We extend a result by Füredi and Komlós and show that the first eigenvalue of a random graph is asymptotically normal, both for $G_{n,p}$ and $G_{n,m}$, provided $np \geq n^\delta$ or $m/n \geq n^\delta$ for some $\delta > 0$. The asymptotic variance is of order p for $G_{n,p}$, and n^{-1} for $G_{n,m}$. This gives a (partial) solution to a problem raised by Krivelevich and Sudakov.

The formula for the asymptotic mean involves a mysterious power series.

1. Introduction

Füredi and Komlós [2] investigated the eigenvalues of random symmetric matrices. In particular, their result shows that for constant $p \in (0,1)$, the first eigenvalue λ_1 of the adjacency matrix of the random graph $G_{n,p}$ is asymptotically normal, with

$$\lambda_1(G_{n,p}) - (n-1)p - 1 + p \xrightarrow{\mathrm{d}} N\big(0, 2p(1-p)\big) \qquad \text{as } n \to \infty. \tag{1.1}$$

In fact, [2] showed that the random fluctuation of of $\lambda_1(G_{n,p})$ asymptotically can be completely explained by the fluctuation of the number of edges in $G_{n,p}$. More precisely, they showed that if $e(G_{n,p}) \sim N\left(\binom{n}{2}, p\right)$ is the number of edges in $G_{n,p}$, then

$$\lambda_1(G_{n,p}) - \frac{2e(G_{n,p})}{n} - (1-p) = O_p\left(n^{-1/2}\right) \xrightarrow{\mathrm{p}} 0,$$

which immediately implies (1.1) by the central limit theorem.

This suggests studying $\lambda_1(G_{n,p})$ conditioned on a given $e(G_{n,p})$, or, equivalently, $\lambda_1(G_{n,m})$, where m is a given function of n. Assume first, in analogy to the case studied by Füredi and Komlós, that $m/\binom{n}{2} \to p$, with $p \in (0,1)$ fixed. We will show that then $\lambda_1(G_{n,m})$ too is asymptotically normal, but with an asymptotic variance of order only n^{-1}.

We will also extend the results to $p \to 0$ and $m/\binom{n}{2} \to 0$, as long as $np \gg n^\delta$ and $m \gg n^{1+\delta}$ for some $\delta > 0$.

Krivelevich and Sudakov [6] have found the first order asymptotics of $\lambda_1(G_{n,p})$ for all $p = p(n)$; in particular, for p in the range treated here, their result gives $\lambda_1(G_{n,p})/(np) \xrightarrow{\mathrm{p}} 1$. They leave the question of the limit distribution as an open problem, which we thus (partially) answer. Note also the large deviation result by Alon, Krivelevich and Vu [1].

Our main results are the following. Here and elsewhere in this paper, $(a_i)_{i=0}^\infty$ is a certain sequence of integers, defined in Section 4. We have computed a_j for $j \leq 10$ by calculations with `Pascal` and `Maple` and found (unless we made a mistake)

$$A(z) := \sum_0^\infty a_j z^j = 1 + z + z^2 + z^5 + z^7 + 5z^8 + 2z^9 + 17z^{10} + \dots$$

Date: August 4, 2003; revised February 17, 2005.

No simple form is evident.

Theorem 1.1. *Suppose that $n \to \infty$, $p \to p_0 \in [0,1)$ and $n^{1-\delta}p \to \infty$, for some fixed $\delta > 0$. Let, for some integer J with $2J+1 \geq 1/\delta$,*

$$\alpha_{n,p} := (n-2)p + \sum_{j=1}^{J} a_j (np)^{1-j} = \sum_{j=0}^{J} a_j (np)^{1-j} - 2p.$$

Then

$$p^{-1/2}\bigl(\lambda_1(G_{n,p}) - \alpha_{n,p}\bigr) \xrightarrow{\mathrm{d}} N\bigl(0, 2(1-p_0)\bigr).$$

Theorem 1.2. *Suppose that $n \to \infty$, $m/\binom{n}{2} \to p_0 \in [0,1)$ and $n^{-1-\delta}m \to \infty$ for some fixed $\delta > 0$. Let, for some integer J with $2J \geq 1/\delta$,*

$$\alpha_{n,m} := \frac{2m}{n} + \sum_{j=1}^{J} a_j \Bigl(\frac{2m}{n}\Bigr)^{1-j} - \frac{2m}{n^2} = \frac{2m}{n}\biggl(\sum_{j=0}^{J} a_j \Bigl(\frac{2m}{n}\Bigr)^{-j} - \frac{1}{n}\biggr).$$

Then

$$n^{1/2}\bigl(\lambda_1(G_{n,m}) - \alpha_{n,m}\bigr) \xrightarrow{\mathrm{d}} N\bigl(0, 2(1-p_0)^2\bigr).$$

Note that J is chosen such that terms $a_j(np)^{1-j}$ or $a_j(2m/n)^{1-j}$ with $j > J$ can be ignored.

The definition of $(a_i)_{i=0}^{\infty}$ in Section 4 is rather involved, and we find the numbers a_j quite mysterious. Lemma 3.1 exhibits the combinatorial significance of these numbers perhaps better than the theorems above. Nevertheless we are lacking a simple combinatorial interpretation of a_j, and leave it as an open problem to understand these numbers better.

Theorem 1.1 follows easily from Theorem 1.2. We will, however, prove both in parallel by the same method. Not surprisingly, the details are somewhat simpler for Theorem 1.1, but we will see that with our methods, the difference is not great.

Remark 1.3. Also for $p \to 0$, the random variation of $\lambda_1(G_{n,p})$ is explained by the variation of the number of edges $e(G_{n,p})$ in the sense of linear regression. I.e., we have $\lambda_1(G_{n,p}) = a(n,p)e(G_{n,p}) + b(n,p) + R$ for certain constants $a(n,p)$ and $b(n,p)$ and a random error term R such that $p^{-1/2}R \xrightarrow{\mathrm{p}} 0$ while $p^{-1/2}a(n,p)(e(G_{n,p}) - \mathbb{E}\, e(G_{n,p}))$ converges in distribution.

For $G_{n,m}$, where the number of edges is constant and explains nothing, the proof shows that the variation is explained in this way by the number of paths of length 2 (or, equivalently, by the sum of the squares of the vertex degrees).

The proof uses the traditional method of computing the trace of a suitable power of the adjacency matrix as the number of closed walks of a given length in the graph. This number is closely related to subgraph counts, and we use methods from [3] to find the required asymptotics.

We consider the case $np \gg n^{\delta}$ ($m/n \gg n^{\delta}$) for some $\delta > 0$. It turns out that the smaller δ is, the higher matrix powers and the longer walks have to be employed (otherwise we cannot ignore the other eigenvalues); we also need more terms in the sums defining $\alpha_{n,p}$ and $\alpha_{n,m}$. We thus give general arguments treating arbitrarily long walks below. If we restricted ourselves to, say, $p \geq n^{-1/2}$, we would only have to consider a few small values of this length, and the general arguments could be replaced by explicit calculations, which would make the proof simpler but perhaps less interesting.

Remark 1.4. Note that we only study the case when p or m is so large that there is a large gap between the first and second eigenvalue. It seems that different methods are needed in the case of sparser graphs. Perhaps the methods of [6] could be useful.

Remark 1.5. Füredi and Komlós [2] studied more general random symmetric matrices where the entries are not restricted to 0 and 1. We leave it to the reader to extend the results of this paper to such matrices.

Remark 1.6. Note that $\lambda_1(G) \geq 2e(G)/n$ for every graph G with n vertices, since $\mathbf{v}A\mathbf{v}^t = 2e(G)/n$ if $\mathbf{v} = n^{-1/2}(1, \dots, 1)$ and A is the adjacency matrix of G. The results above show that, with high probability, we almost have equality for the random graphs studied here, which witnesses that the eigenvector for λ_1 is close to $\mathbf{v}$.

If X_n are random variables and c_n positive numbers, we write $X_n = o_p(c_n)$ if $X_n/c_n \xrightarrow{\text{p}} 0$, and $X_n = O_p(c_n)$ if the sequence X_n/c_n is stochastically bounded (tight).

If H is a graph, $v(H)$, $e(H)$ and $\operatorname{aut}(H)$ denote the numbers of vertices, edges and automorphisms of H.

Acknowledgement. Part of this research was done on a previous visit to Cambridge. I thank Béla Bollobás for interesting discussions.

2. Matrices

We denote the eigenvalues (with multiplicities) of a real symmetric matrix M by $\lambda_1(M) \geq \lambda_2(M) \geq \dots \geq \lambda_\nu(M)$. For a graph G, we similarly denote the eigenvalues of its adjacency matrix by $\lambda_1(G) \geq \dots$.

The algebraic part of our proofs is the following lemma.

Lemma 2.1. *Let M be a real symmetric matrix and let $T_k := \operatorname{Tr}(M^k) = \sum_i \lambda_i(M)^k$. Suppose that, for some even $k \geq 2$ and $\mu > 0$,*

$$\lambda_1(M) \geq \mu, \tag{2.1}$$

$$T_k \leq \mu^k(1 + 2^{-k}). \tag{2.2}$$

Then

$$T_k\Big(2 - \frac{T_{k-2}T_{k+2}}{T_k^2}\Big) \leq \lambda_1(M)^k \leq T_k.$$

Proof. Let $\delta_i = \lambda_i/\lambda_1$, $1 \leq i \leq \nu$, where ν is the size of M. First, by (2.1) and (2.2),

$$1 + \sum_{i=2}^{\nu} \delta_i^k = T_k/\lambda_1^k \leq 1 + 2^{-k}.$$

Hence $|\delta_i| \leq 1/2$ for $i \geq 2$. In particular,

$$(1 - \delta_i^2)^2 \geq \big(\tfrac{3}{4}\big)^2 > \tfrac{1}{2} \geq 2\delta_i^2.$$

Consequently,

$$\begin{aligned} T_{k-2}T_{k+2}-T_k^2 &= \sum_{i,j=1}^{\nu}\bigl(\lambda_i^{k-2}\lambda_j^{k+2}-\lambda_i^k\lambda_j^k\bigr)=\sum_{i<j}\bigl(\lambda_i^{k-2}\lambda_j^{k+2}+\lambda_i^{k+2}\lambda_j^{k-2}-2\lambda_i^k\lambda_j^k\bigr)\\ &=\sum_{i<j}\lambda_i^{k-2}\lambda_j^{k-2}\bigl(\lambda_i^2-\lambda_j^2\bigr)^2\ge\sum_{j=2}^{\nu}\lambda_1^{k+2}\lambda_j^{k-2}\bigl(1-\delta_j^2\bigr)^2\\ &\ge\sum_{j=2}^{\nu}\lambda_1^{k+2}\lambda_j^{k-2}\cdot 2\delta_j^2=2\lambda_1^k\sum_{j=2}^{\nu}\lambda_j^k=2\lambda_1^k\bigl(T_k-\lambda_1^k\bigr)\\ &\ge T_k\bigl(T_k-\lambda_1^k\bigr). \end{aligned}$$

The left inequality follows. The right one is immediate. □

Lemma 2.2. *Let M_n, $n\ge 1$, be random symmetric matrices (of arbitrary sizes), and let $T_{k,n}:=\operatorname{Tr}(M_n^k)$. Suppose that $\mu_n>0$ are real numbers such that for every $\eta>0$*

$$\mathbb{P}\bigl(\lambda_1(M_n)\ge(1-\eta)\mu_n\bigr)\to 1 \qquad \text{as } n\to\infty, \tag{2.3}$$

and that Y is a random variable and $\varepsilon_n\to 0$ are positive numbers such that

$$\varepsilon_n^{-1}\Bigl(\frac{T_{k,n}}{\mu_n^k}-1\Bigr)\xrightarrow{\mathrm{d}} kY \qquad \text{as } n\to\infty, \tag{2.4}$$

jointly for three fixed consecutive even values of k. Then

$$\varepsilon_n^{-1}\Bigl(\frac{\lambda_1(M_n)}{\mu_n}-1\Bigr)\xrightarrow{\mathrm{d}} Y \qquad \text{as } n\to\infty.$$

Proof. Write

$$T_{k,n}=\mu_n^k\bigl(1+\varepsilon_n kY_{k,n}\bigr). \tag{2.5}$$

Thus $Y_{k,n}\xrightarrow{\mathrm{d}} Y$ jointly for three even values of k, say $k=m-2$, m and $m+2$, and hence $(m-2)Y_{m-2,n}+(m+2)Y_{m+2,n}-2mY_{m,n}\xrightarrow{\mathrm{p}}0$. Then

$$\begin{aligned} Q_n:=\frac{T_{m-2,n}T_{m+2,n}}{T_{m,n}^2}&=\frac{\bigl(1+\varepsilon_n(m-2)Y_{m-2,n}\bigr)\bigl(1+\varepsilon_n(m+2)Y_{m+2,n}\bigr)}{\bigl(1+\varepsilon_n mY_{m,n}\bigr)^2}\\ &=1+\varepsilon_n\bigl((m-2)Y_{m-2,n}+(m+2)Y_{m+2,n}-2mY_{m,n}\bigr)+o_p(\varepsilon_n)\\ &=1+o_p(\varepsilon_n). \end{aligned} \tag{2.6}$$

(The reader that prefers may use the Skorohod representation theorem [5, Theorem 4.30] and assume for simplicity that $Y_{k,n}\to Y$ a.s. for $k=m-2,m,m+2$; then o_p may be replaced by o.)

Moreover, with $\widetilde{\mu}:=\mu_n(1-\eta)$, where $\eta>0$ is so small that $(1-\eta)^{-k}<1+2^{-k}$,

$$\mathbb{P}\bigl(T_{k,n}\le\widetilde{\mu}^k(1+2^{-k})\bigr)=\mathbb{P}\bigl(1+\varepsilon_n kY_n\le(1-\eta)^k(1+2^{-k})\bigr)\to 1 \qquad \text{as } n\to\infty.$$

Since $\mathbb{P}\bigl(\lambda_1(M_n)\ge\widetilde{\mu}\bigr)\to 1$ as $n\to\infty$ by (2.3), we see that with probability tending to 1 as $n\to\infty$, M_n satisfies the assumptions of Lemma 2.1 (with m and $\widetilde{\mu}$) and thus

$$\mathbb{P}\bigl[T_{m,n}(2-Q_n)\le\lambda_1(M_n)^m\le T_{m,n}\bigr]\to 1. \tag{2.7}$$

Combined with (2.6), this yields

$$\lambda_1(M_n)^m=T_{m,n}\bigl(1+o_p(\varepsilon_n)\bigr)$$

and thus

$$\lambda_1(M_n) = T_{m,n}^{1/m}\bigl(1+o_p(\varepsilon_n)\bigr) = \mu_n\bigl(1+\varepsilon_n m Y_{m,n}\bigr)^{1/m}\bigl(1+o_p(\varepsilon_n)\bigr)$$
$$= \mu_n\bigl(1+\varepsilon_n Y_{m,n}+o_p(\varepsilon_n)\bigr).$$

The result follows. □

We apply Lemma 2.2 to $G_{n,p}$ and $G_{n,m}$, with $\mu_n = \alpha_{n,p}$ and $\mu_n = \alpha_{n,m}$, respectively. Note that $\alpha_{n,p} = np\bigl(1+o(1)\bigr)$ and $\alpha_{n,m} = \frac{2m}{n}\bigl(1+o(1)\bigr)$. By Remark 1.6, $\lambda_1(G_{n,m}) \ge 2m/n$, and (2.3) follows. For $G_{n,p}$, similarly, $\lambda_1 \ge 2e(G_{n,p})/n$ and $2e(G_{n,p})/(n^2p) \xrightarrow{\text{p}} 1$ by the law of large numbers; again (2.3) follows.

Note further that if M is the adjacency matrix of a graph G, then $\operatorname{Tr}(M^k)$ equals the number of closed walks of length k in G; i.e. sequences $v_0,\dots,v_k$ of vertices such that $v_0 = v_k$ and v_{i-1} and v_i are adjacent for $1 \le i \le k$; we denote this number by $W_k(G)$. Theorems 1.1 and 1.2 therefore follow by Lemma 2.2 from the following two lemmas. (The assumptions $k \ge 6/\delta$ are made for convenience and could be weakened. However, the results are not true for, say, $k = 2$ or $k = 4$, even for constant p.)

Lemma 2.3. *Under the hypotheses of Theorem 1.1, if $Y \sim N\bigl(0, 2(1-p_0)\bigr)$, then for every $k \ge 6/\delta$,*

$$np^{1/2}\left(\frac{W_k(G_{n,p})}{\alpha_{n,p}^k}-1\right) \xrightarrow{\text{d}} kY$$

and the convergence holds jointly for any set of such k.

Lemma 2.4. *Under the hypotheses of Theorem 1.2, if $Y \sim N\bigl(0, 2(1-p_0)^2\bigr)$, then for every $k \ge 6/\delta$,*

$$2mn^{-1/2}\left(\frac{W_k(G_{n,m})}{\alpha_{n,m}^k}-1\right) \xrightarrow{\text{d}} kY$$

and the convergence holds jointly for any set of such k.

3. Random graphs

We prove Lemmas 2.3 and 2.4 using the orthogonal decomposition method of [3], summarized in [4, Section 6.4]. For convenience, we repeat the main definitions and results here, referring to [3] for proofs. We begin by defining an orthogonal family of functionals of $G_{n,p}$.

Let H be a graph. Consider the $(n)_{v_H}$ injective mappings from the vertex set of H into $\{1,\dots,n\}$. Each such mapping φ maps H onto a copy $\varphi(H)$ of H in K_n, and we define

$$S_{n,p}(H) := \sum_{\varphi}\prod_{e\in\varphi(H)}(I_e - p), \tag{3.1}$$

where $I_e = \mathbf{1}[e \in G_{n,p}]$ is the indicator that the edge e is present. In other words, we sum $\prod_{e\in H'}(I_e-p)$ over all copies of H in $G_{n,p}$, counted with multiplicities $\operatorname{aut}(H)$. Note that if $X_H(G)$ denotes the number of copies of H in G, each counted with multiplicity $\operatorname{aut}(H)$, we have the similar formula

$$X_H(G_{n,p}) = \sum_{\varphi}\prod_{e\in\varphi(H)} I_e, \tag{3.2}$$

where, however, the terms in the sum not are orthogonal.

$S_{n,p}(H)$ depends on H only up to isomorphism. Hence we may regard H as an unlabelled graph.

Let $\mathcal{U}^0$ denote the set of unlabelled graphs without isolated vertices. Then the random variables $\{S_{n,p}(H)\}_{H\in\mathcal{U}^0}$ are orthogonal, and each functional of $G_{n,p}$ that depends only on the isomorphism type is a linear combination of these variables. In particular,

$$W_k(G_{n,p}) = \sum_{H\in\mathcal{U}^0} \hat{w}_k(n,p;H)S_{n,p}(H) \tag{3.3}$$

for some coefficients $\hat{w}_k$.

We allow here H to be the empty graph $\emptyset$ with $v(\emptyset) = e(\emptyset) = 0$; then $S_{n,p}(\emptyset) = 1$. Since $\mathbb{E}\, S_{n,p}(H) = 0$ when $H \neq \emptyset$, we have

$$\hat{w}_k(n,p;\emptyset) = \mathbb{E}\, W_k(G_{n,p}). \tag{3.4}$$

We can find the decomposition (3.3) as follows. A closed walk of length k may have a finite number (depending on k) different shapes, since one or several vertices may be repeated. Hence W_k can be written as a linear combination of different subgraph counts X_H. For example, with $k = 4$ we can have a 4-cycle, a path of length 2 with each edge traversed twice, or a single edge traversed four times, and we find

$$W_4 = X_{C_4} + 2X_{P_2} + X_{K_2}.$$

(P_l denotes the path with l edges and thus $l+1$ vertices.)

Next, substituting $I_e = (I_e - p) + p$ in (3.2) and expanding, each X_H becomes a linear combination of $S_{n,p}(K)$ for $K \subseteq H$. For example, straightforward calculations yield, with $(n)_k = n(n-1)\cdots(n-k+1)$,

$$\begin{aligned}
X_{K_2}(G_{n,p}) &= S_{n,p}(K_2) + (n)_2 p\\
X_{P_2}(G_{n,p}) &= S_{n,p}(P_2) + 2(n-2)pS_{n,p}(K_2) + (n)_3 p^2\\
X_{C_4}(G_{n,p}) &= S_{n,p}(C_4) + 4pS_{n,p}(P_3) + 4(n-3)p^2 S_{n,p}(P_2) + 2p^2 S_{n,p}(2K_2)\\
&\quad + 4(n-2)(n-3)p^3 S_{n,p}(K_2) + (n)_4 p^4.
\end{aligned}$$

In this way, we can obtain a decomposition (3.3) for any k explicitly (but the amount of work increases rapidly with k). Note that only terms with $e(H) \le k$ appears.

For $H \in \mathcal{U}^0$,

$$S_{n,p}(H) = O_p\bigl(n^{v(H)/2} p^{e(H)/2}\bigr). \tag{3.5}$$

Hence we also define

$$S^*_{n,p}(H) := n^{-v(H)/2} p^{-e(H)/2} S_{n,p}(H), \tag{3.6}$$

$$\hat{w}^*_k(n,p;H) := n^{v(H)/2} p^{e(H)/2} \hat{w}_k(n,p;H); \tag{3.7}$$

thus (3.3) can be rewritten

$$W_k(G_{n,p}) = \sum_{H\in\mathcal{U}^0} \hat{w}^*_k(n,p;H) S^*_{n,p}(H), \tag{3.8}$$

where by (3.5), for every H,

$$S^*_{n,p}(H) = O_p(1). \tag{3.9}$$

If further $H \neq \emptyset$ and H is connected, we have the limit result [3, Theorem 1], [4, Theorem 6.43] that if $n \to \infty$, $p \to p_0 \in [0,1]$ and $np^{m(H)} \to \infty$, where $m(H) := \max\{e(F)/v(F) : F \subseteq H, v(F) > 0\}$, then, for some random variables $U(H)$,

$$S^*_{n,p}(H) \overset{\mathrm{d}}{\to} U(H) \sim N\bigl(0, \operatorname{aut}(H)(1-p_0)^{e(H)}\bigr). \tag{3.10}$$

To prove Lemma 2.3, it is now sufficient to verify, for $k \geq 6/\delta$,

$$\mathbb{E}\, W_k(G_{n,p}) = \alpha_{n,p}^k\bigl(1 + o(n^{-1}p^{-1/2})\bigr) \tag{3.11}$$
$$\hat{w}_k^*(n,p;K_2) = \alpha_{n,p}^k n^{-1}p^{-1/2}\bigl(k + o(1)\bigr) \tag{3.12}$$
$$\hat{w}_k^*(n,p;H) = o(\alpha_{n,p}^k n^{-1}p^{-1/2}), \qquad H \in \mathcal{U}^0, v(H) \geq 3, \tag{3.13}$$

because then (3.8) yields by (3.4), (3.9)

$$\begin{aligned}\alpha_{n,p}^{-k}W_k(G_{n,p}) &= \alpha_{n,p}^{-k}\,\mathbb{E}\, W_k(G_{n,p}) + \alpha_{n,p}^{-k}\hat{w}_k^*(n,p;K_2)S_{n,p}^*(K_2) + o_p(n^{-1}p^{-1/2})\\ &= 1 + n^{-1}p^{-1/2}kS_{n,p}^*(K_2) + o_p(n^{-1}p^{-1/2}),\end{aligned}$$

and Lemma 2.3 follows by (3.10), with $Y = U(K_2)$.

To prove Lemma 2.4, we define $p := m/\binom{n}{2}$ and note that

$$\alpha_{n,p} = \alpha_{n,m} + O(1/n) = \alpha_{n,m}\bigl(1 + O(n^{-2}p^{-1})\bigr) = \alpha_{n,m}\bigl(1 + o(n^{-3/2}p^{-1})\bigr).$$

For Lemma 2.4, we now need, for $k \geq 6/\delta$,

$$\mathbb{E}\, W_k(G_{n,p}) = \alpha_{n,p}^k\bigl(1 + o(n^{-3/2}p^{-1})\bigr) \tag{3.14}$$
$$\hat{w}_k^*(n,p;P_2) = \alpha_{n,p}^k n^{-3/2}p^{-1}\bigl(k + o(1)\bigr) \tag{3.15}$$
$$\hat{w}_k^*(n,p;H) = o(\alpha_{n,p}^k n^{-3/2}p^{-1}), \qquad H \in \mathcal{U}^0, H \neq \emptyset, K_2, P_2. \tag{3.16}$$

(No condition on $\hat{w}_k^*(n,p;K_2)$ is needed.) Indeed, using these estimates, [3, Theorem 7] or [4, Theorem 6.54], with $\beta_n := n^{-3/2}p^{-1}\alpha_{n,m}^k$, shows that

$$n^{3/2}p\left(\frac{W_k(G_{n,m})}{\alpha_{n,m}^k} - 1\right) \xrightarrow{\mathrm{d}} kU(P_2)$$

(again jointly for different k), which yields Lemma 2.4 and thus Theorem 1.2.

It is important to note that we here draw a conclusion for $G_{n,m}$ from the estimates (3.14)–(3.16) for $G_{n,p}$. In the remainder of the paper, we thus consider $G_{n,p}$ only.

It remains to prove the estimates (3.11)–(3.13) and (3.14)–(3.16). Using $\alpha_{n,p} \sim np$ and changing J, we restate (and partly improve) them slightly as the following lemmas, which thus contain the combinatorial part of the proof of Theorems 1.1 and 1.2. (We treat $\hat{w}_k(n,p;\emptyset) = \mathbb{E}\, W_k(G_{n,p})$ separately because a much smaller relative error is required.)

Lemma 3.1. *Let $\delta > 0$ and let k and J be fixed integers with $J \geq 1/\delta$ and $k \geq 6/\delta$. If $n \to \infty$ and $np/n^\delta \to \infty$, then*

$$\mathbb{E}\, W_k(G_{n,p}) = (np)^k\Bigl(1 - \frac{2}{n} + \sum_{j=1}^{J} a_j(np)^{-j} + O(n^{-2}p^{-1})\Bigr)^k.$$

Lemma 3.2. *Let $\delta > 0$ and $k \geq 4/\delta$ be fixed, and suppose that $H \in \mathcal{U}^0$ with $H \neq \emptyset$. If $n \to \infty$ and $np/n^\delta \to \infty$, then*

$$(np)^{-k}\hat{w}_k(n,p;H) = \begin{cases} kn^{-2}p^{-1} + o(n^{-2}p^{-1}), & H = K_2,\\ kn^{-3}p^{-2} + o(n^{-3}p^{-2}), & H = P_2,\\ o\Bigl(n^{-v(H)/2-3/2}p^{-e(H)/2-1}\Bigr), & H \neq \emptyset, K_2, P_2.\end{cases}$$

4. Proof of Lemma 3.1

We begin by giving an explicit, although rather opaque, definition of the numbers a_j in Theorems 1.1 and 1.2.

For a tree T, let $b_k(T)$ be the number of (not necessarily closed) walks of length k on T that traverse every edge at least twice. Let $\mathcal{T}_n$ be the set of the n^{n-2} trees on $\{1,\dots,n\}$, and let $\mathcal{T} := \bigcup_{n=1}^{\infty} \mathcal{T}_n$, and define the formal power series

$$\Psi(\varepsilon, z) := \sum_{k=0}^{\infty}\sum_{n=1}^{\infty}\frac{1}{n!}\sum_{T\in\mathcal{T}_n} b_k(T)\varepsilon^{k-e(T)}z^k = \sum_{k=0}^{\infty}\sum_{T\in\mathcal{T}}\frac{1}{v(T)!}b_k(T)\varepsilon^{k-e(T)}z^k.$$

By symmetry, we can eliminate the factor $1/n!$ by only considering walks on $T \in \mathcal{T}_n$ such that the first visits to the vertices come in order $1, 2, \dots, n$. Thus Ψ has integer coefficients.

If a term $\varepsilon^j z^k$ appears in $\Psi(\varepsilon, z)$ with non-zero coefficient, then $j = k - e(T)$ for some tree with a walk of length k that uses every edge at least twice. Thus $k \ge 2e(T)$, so $k/2 \le j \le k$. We can thus regard $\Psi(\varepsilon, z)$ as a power series in ε, with coefficients that are polynomials in z with integer coefficients. Note also that the constant term $\Psi(0, z) = 1$. It follows that there exists a unique power series $Z(\varepsilon)$ such that

$$Z(\varepsilon)\Psi\big(\varepsilon, Z(\varepsilon)\big) = 1. \tag{4.1}$$

Z has integer coefficients and $Z(0) = 1$. Finally, define the formal power series

$$A(\varepsilon) = \sum_{k=0}^{\infty} a_k\varepsilon^k := \frac{1}{Z(\varepsilon)}. \tag{4.2}$$

Note that each a_k is an integer and $a_0 = 1$.

Proof of Lemma 3.1. A closed walk with k steps defines a connected graph F consisting of all vertices and edges in the walk. Since $e(F) \le k$, there is only a finite number of possible F (regarded as unlabelled graphs). The contribution to $(np)^{-k}\,\mathbb{E}\,W_k(G_{n,p})$ for a given unlabelled F is clearly

$$(np)^{-k}O\big(n^{v(F)}p^{e(F)}\big) = O\big(n^{v(F)-k}p^{e(F)-k}\big). \tag{4.3}$$

We consider three cases separately.

Case 1: F is a tree, $v(F) = e(F) + 1$.

Since a closed walk on a tree has to traverse each edge at least twice, we have $2e(F) \le k$ and thus the contribution is, by (4.3),

$$O\big(n^{v(F)-k}p^{e(F)-k}\big) = O\big(n(np)^{e(F)-k}\big) = O\big(n(np)^{-k/2}\big) = O\big(n^{-2}\big) \tag{4.4}$$

because $(np)^{k/2} \ge (np)^{3/\delta} \gg n^3$.

Case 2: F has more that one cycle, $v(F) < e(F)$.

The contribution from F is by (4.3)

$$O\big(n^{-(e(F)-v(F))}(np)^{e(F)-k}\big)$$

which is $O\big(n^{-2}p^{-1}\big)$ except when $v(F) = e(F) - 1$ and $e(F) = k$. The latter case means that the edges of the walk are distinct but one vertex is repeated. Labelling the vertices $v_1, \dots, v_k$, we thus have $v_i = v_j$ for two indices i and j, while the v_i's otherwise are distinct. Moreover, $3 \le |i - j| \le k - 3$ since each of the two cycles in F has at least 3 vertices. The indices i and j may thus be chosen in $k(k-5)/2$ ways, and thus the contribution from such walks is

$$(np)^{-k}\frac{k(k-5)}{2}(n)_{k-1}p^k = \frac{k(k-5)}{2}n^{-1} + O(n^{-2}).$$

The total contribution from F with $v(F) < e(F)$ is thus

$$\frac{k(k-5)}{2}n^{-1} + O(n^{-2}p^{-1}). \tag{4.5}$$

Case 3: F is unicyclic, $v(F) = e(F)$.

Then F consists of a cycle with attached trees. Given a closed walk on F traversing all edges, colour all edges of F that are traversed at least twice red and colour the remaining edges green. Each edge in the attached trees is red, while the edges in the cycle may be either red or green. Let $l \geq 0$ be the number of green edges.

If there are $l \geq 1$ green edges, the removal of them from F leaves l red components $T_1, \dots, T_l$. Each T_l is a tree (possibly a single vertex only) and $v(F) = \sum_{i=1}^{l} v(T_i)$; moreover, the green edges join the red components into a cycle.

Fix $l \geq 3$ and trees $T_1, \dots, T_l$ (regarded as disjoint subgraphs of K_n), and consider together all F that are obtained by joining the trees by l edges, one from each T_i to T_{i+1} (and from T_l to T_1). A closed walk on one of these F with red subtrees $T_1, \dots, T_l$, that starts with the green edge leading from T_l to T_1, is called *special*. A special closed walk thus consists of a walk in each T_i that traverses each edge at least twice, together with single (green) steps linking the walks. The green links are determined by the walks in the trees, and thus the number of special walks with k_i steps inside T_i, $i = 1, \dots, l$, is $\prod_{i=1}^{l} b_{k_i}(T_i)$; summing we find that the number of special walks with length k, for given $T_1, \dots, T_l$, is, with $B(x;T) := \sum_{k=0}^{\infty} x^k b_k(T)$ and using $[x^j]f(x)$ to denote the coefficient of x^j in a power series $f(x)$,

$$\sum_{k_1+\cdots+k_l=k-l}\prod_{i=1}^{l} b_{k_i}(T_i) = [x^{k-l}]B(x;T_1)\cdots B(x;T_l). \tag{4.6}$$

Each of these walks uses $\sum_1^l v(T_i)$ edges, so to get the contribution to $(np)^{-k}\,\mathbb{E}\,W_k(G_{n,p})$ we multiply by $(np)^{-k}p^{\sum v(T_i)}$.

Summing first over all choices of $T_1, \dots, T_l$ with given vertex sets and then over all ways to choose these vertex sets in $\{1, \dots, n\}$ we obtain

$$\begin{aligned}(np)^{-k}\sum_{n_1,\dots,n_l\geq 1}\binom{n}{n_1,\dots,n_l}\sum_{T_i\in\mathcal{T}_{n_i}}[x^{k-l}]B(x;T_1)\cdots B(x;T_l)p^{\sum_i v(T_i)}\\ = (np)^{-k}[x^{k-l}]\Bigl(\sum_{T\in\mathcal{T}}\frac{B(x;T)}{v(T)!}(np)^{v(T)}\Bigr)^l\Bigl(1+O\Bigl(\frac{1}{n}\Bigr)\Bigr).\end{aligned} \tag{4.7}$$

This is, for a given $l \geq 3$, the contribution from the walks that generate a unicyclic F with l red subgraphs, and that begin with a green edge. A walk generating such an F may be shifted (cyclically) in k ways by changing the starting point, and l of these shifts begin with a green edge; hence, the contribution from the walks that begin with a green edge is l/k times the total contribution for this F. Consequently, the contribution from all walks that generate a unicyclic F with l red subgraphs (for given $l \geq 3$) is k/l times the value in (4.7).

If $l < k$, then $v(T_i) > 1$ for some i so F contains a red edge. This means that $k > e(F) = v(F) = \sum_i v(T_i)$. Since each term in the sum in (4.7) then is

$$O\big((np)^{\sum_i v(T_i)-k}\big) = O\big((np)^{-1}\big),$$

the contribution of the term $O(1/n)$ in (4.7) then is $O\big((n^2p)^{-1}\big)$. Moreover,

$$\sum_{T\in\mathcal{T}} \frac{B(x;T)}{v(T)!}(np)^{v(T)} = \sum_{k=1}^{\infty}\sum_{T\in\mathcal{T}} \frac{b_k(T)x^k}{v(T)!}(np)^{e(T)+1} = np\Psi\Big(\frac{1}{np}, npx\Big).$$

Hence we find from (4.7) that the contribution to $(np)^{-k}\,\mathbb{E}\,W_k(G_{n,p})$ from all walks that generate a unicyclic F with l red subtrees is, for $3 \le l < k$,

$$\frac{k}{l}(np)^{-k}[x^{k-l}](np)^l\Psi\Big(\frac{1}{np}, npx\Big)^l + O\big((n^2p)^{-1}\big) = \frac{k}{l}[x^{k-l}]\Psi\Big(\frac{1}{np}, x\Big)^l + O\big((n^2p)^{-1}\big). \tag{4.8}$$

For $l = k$ we are considering walks without repeated edges, i.e. cycles. Clearly, the contribution from them is

$$(np)^{-k}(n)_k p^k = 1 - \binom{k}{2}\frac{1}{n} + O\big(n^{-2}\big) = [x^0]\Psi\Big(\frac{1}{np}, x\Big)^k - \binom{k}{2}\frac{1}{n} + O\big(n^{-2}\big). \tag{4.9}$$

For $l \le 2$, the formulas above are not quite correct. However, with $l \ge 0$ green edges and thus $e(F) - l$ red edges, we have $k \ge l + 2(e(F) - l)$ and thus $e(F) \le (k + l)/2$. If $l \le 2$ we thus have $e(F) \le 1 + k/2$, and by (4.3), the contribution from such F is, since $k/2 \ge 3/\delta$,

$$O\big((np)^{e(F)-k}\big) = O\big((np)^{1-k/2}\big) = O\big(n(np)^{-3/\delta}\big) = O\big(n^{-2}\big). \tag{4.10}$$

Summing (4.4), (4.5), (4.8) for $3 \le l < k$ and (4.9), (4.10) we find

$$(np)^{-k}\,\mathbb{E}\,W_k(G_{n,p}) = \sum_{l=3}^{k} \frac{k}{l}[x^{k-l}]\Psi\Big(\frac{1}{np}, x\Big)^l - 2\frac{k}{n} + O\big(n^{-2}p^{-1}\big).$$

Lemma 3.1 now follows from the following algebraic lemma. □

Lemma 4.1. *If $J \ge 0$ and $k \ge m \ge 2J$, then*

$$\sum_{l=k-m}^{k} \frac{k}{l}[z^{k-l}]\Psi(\varepsilon, z)^l = \left(\sum_{j=0}^{J} a_j\varepsilon^j\right)^k + O\left(\varepsilon^{J+1}\right).$$

Here and in the proof, $O(\varepsilon^a)$, with a real, denotes a polynomial or power series in ε containing only powers ε^j with $j \ge a$.

Proof. Define Φ_ε as the power series that solves the equation

$$\Psi(\varepsilon, z) = \Phi_\varepsilon\big(z\Psi(\varepsilon, z)\big). \tag{4.11}$$

Since $\Psi(\varepsilon, 0) = 1$, it is easily seen that Φ_ε exists and is unique; moreover, by an easy induction, each coefficient $[z^k]\Phi_\varepsilon(z)$ is a polynomial in ε with nonzero terms $c_j\varepsilon^j$ for $k/2 \le j \le k$ only, because Ψ is of this type. The same then is true for any power of $\Phi_\varepsilon(z)$.

By Lagrange's inversion formula [7, Theorem 5.4.2], for $1 \le l \le k$,

$$\frac{k}{l}[z^{k-l}]\Psi(\varepsilon, z)^l = \frac{k}{l}[z^k]\big(z\Psi(\varepsilon, z)\big)^l = [u^{k-l}]\Phi_\varepsilon(u)^k.$$

This is a polynomial in ε and is $O(\varepsilon^{(k-l)/2})$. Hence,

$$\sum_{l=k-m}^{k} \frac{k}{l}[z^{k-l}]\Psi(\varepsilon,z)^l = \sum_{j=0}^{m}[u^j]\Phi_\varepsilon(u)^k = \sum_{j=0}^{\infty}[u^j]\Phi_\varepsilon^k(u) + O\left(\varepsilon^{(m+1)/2}\right),$$

where the infinite sum is well defined as a power series in ε. This sum of all coefficients of Φ_ε^k is

$$\Phi_\varepsilon^k(1) = \Phi_\varepsilon(1)^k$$

and, substituting (4.1) in (4.11) and using (4.1) and (4.2),

$$\Phi_\varepsilon(1) = \Phi_\varepsilon\big(Z(\varepsilon)\Psi\big(\varepsilon, Z(\varepsilon)\big)\big) = \Psi\big(\varepsilon, Z(\varepsilon)\big) = \frac{1}{Z(\varepsilon)} = A(\varepsilon).$$

(These manipulations are easily justified modulo ε^N for any fixed N.) The lemma follows. □

5. Proof of Lemma 3.2

It is easily seen from the discussion if Section 3 that $\hat{w}_k(n,p;H)$ can be computed as follows. Fix a copy H_0 of H in K_n and consider the set $\mathcal{W}$ of closed walks of length k in K_n that use every edge in H_0 at least once. If $\gamma \in \mathcal{W}$, let $\bar{\gamma}$ denote its trace, i.e. the subgraph of K_n consisting of the edges and vertices in γ. Then

$$\hat{w}_k(n,p;H) = \frac{1}{\operatorname{aut}(H)} \sum_{\gamma\in\mathcal{W}} p^{e(\bar{\gamma})-e(H)}. \tag{5.1}$$

Let $c = c(H)$ be the number of components of H, and note that $v(H) \le c + e(H)$.

Fix $j \ge 0$ and consider the closed walks γ in this sum that pass through j vertices outside H_0. Clearly, the number of such γ is $O(n^j)$.

Since $\bar{\gamma}$ connects the j vertices outside H_0 and the c components of H_0, it has at least $j+c-1$ edges outside H_0, i.e.

$$e(\bar{\gamma}) - e(H) \ge j + c - 1.$$

Case 1: $e(\bar{\gamma}) - e(H) = j + c - 1$.

In this case, if we collapse each component of H_0 to a single point, $\bar{\gamma}$ becomes a connected graph with $j+c$ vertices and $j+c-1$ edges, i.e. a tree. The closed walk γ has to traverse each edge in this tree an even number of times, and thus

$$k \ge 2(j+c-1) + e(H).$$

The contribution to $(np)^{-k}\hat{w}_k(n,p;H)$ from all γ in Case 1 is thus, using (5.1),

$$\begin{aligned} O\left((np)^{-k}n^j p^{j+c-1}\right) &= O\left((np)^{-k/2} n^{1-c-e(H)/2} p^{-e(H)/2}\right) \\ &= o\left(n^{-k\delta/2+1-c/2-v(H)/2} p^{-e(H)/2}\right) \\ &= o\left(n^{-3/2-v(H)/2} p^{-e(H)/2}\right). \end{aligned}$$

This is covered by the o term in the lemma.

Case 2: $e(\bar{\gamma}) - e(H) \ge j + c$.

Then $k \geq e(\bar{\gamma}) \geq j + c + e(H)$. The contribution to $(np)^{-k}\hat{w}_k(n,p;H)$ is, using (5.1),

$$\begin{aligned} O\left((np)^{-k}n^j p^{j+c}\right) &= O\left(n^{-c-e(H)}p^{-e(H)}\right) \\ &= O\left((np)^{-e(H)/2}n^{-c/2-v(H)/2}p^{-e(H)/2}\right). \end{aligned} \tag{5.2}$$

If $e(H) \geq 3$, or if $e(H) = 2$ and $c > 1$, this is $o\left(n^{-3/2}p^{-1}n^{-v(H)/2}p^{-e(H)/2}\right)$, which verifies the lemma for these H, i.e. all H except K_2 and P_2.

For $H = P_2$, the calculation in (5.2) yields $O\left(n^{-3}p^{-2}\right)$, and $o\left(n^{-3}p^{-2}\right)$ unless $k = e(\bar{\gamma}) = j + c + e(H) = j + 3$. We thus only have to consider γ that go through k different vertices, i.e. cycles of length k. The number of such cycles passing through H_0 is $2k(n)_k/(n)_3$, since there are $(n)_k/(n)_3$ choices of the cycle $\bar{\gamma}$, and for each $\bar{\gamma}$, γ may start at k places and in 2 directions. Thus, (5.1) yields

$$(np)^{-k}\hat{w}_k(n,p;P_2) = (np)^{-k}\frac{k(n)_k}{(n)_3}p^{k-2} + o\left(n^{-3}p^{-2}\right) = kn^{-3}p^{-2} + o\left(n^{-3}p^{-2}\right).$$

Finally, for $H = K_2$, (5.2) yields $O\left(n^{-2}p^{-1}\right)$, and again we have o unless $k = e(\bar{\gamma}) = j + c + e(H) = j + v(H)$. Thus, again, we only have to consider γ that go through k different vertices, i.e. cycles of length k. Arguing as for P_2 we find that the number of such cycles passing through H_0 is $2k(n)_k/(n)_2$, and

$$(np)^{-k}\hat{w}_k(n,p;K_2) = (np)^{-k}\frac{k(n)_k}{(n)_2}p^{k-1} + o\left(n^{-2}p^{-1}\right) = kn^{-2}p^{-1} + o\left(n^{-2}p^{-1}\right).$$

□

References

[1] N. Alon, M. Krivelevich & V. H. Vu, On the concentration of eigenvalues of random symmetric matrices. *Israel J. Math.* **131** (2002), 259–267.

[2] Z. Füredi & J. Komlós, The eigenvalues of random symmetric matrices. *Combinatorica* **1** (1981), no. 3, 233–241.

[3] S. Janson, *Orthogonal Decompositions and Functional Limit Theorems for Random Graph Statistics*. Mem. Amer. Math. Soc., vol. 111, no. 534, American Mathematical Society, Providence, R.I., 1994.

[4] S. Janson, T. Łuczak, A. Ruciński, *Random Graphs*. John Wiley & Sons, New York, 2000.

[5] O. Kallenberg, *Foundations of Modern Probability*, 2nd ed., Springer, New York, 2002.

[6] M. Krivelevich & B. Sudakov, The largest eigenvalue of sparse random graphs. *Combin. Probab. Comput.* **12** (2003), 61–72.

[7] R.P. Stanley, *Enumerative Combinatorics, Volume 2*. Cambridge Univ. Press, 1999.

Department of Mathematics, Uppsala University, PO Box 480, SE-751 06 Uppsala, Sweden

E-mail address: svante.janson@math.uu.se

URL: http://www.math.uu.se/~svante/