

Random Perfect Graphs

Colin McDiarmid
 cmcd@stats.ox.ac.uk
 Department of Statistics
 University of Oxford

Nikola YOLOV
 nikola.yolov@cs.ox.ac.uk
 Department of Computer Science
 University of Oxford

April 5, 2016

Abstract

We investigate the asymptotic structure of a random perfect graph P_n sampled uniformly from the perfect graphs on vertex set $\{1, \dots, n\}$. Our approach is based on the result of Prömel and Steger that almost all perfect graphs are generalised split graphs, together with a method to generate such graphs almost uniformly.

We show that the distribution of the maximum of the stability number $\alpha(P_n)$ and clique number $\omega(P_n)$ is close to a concentrated distribution $L(n)$ which plays an important role in our generation method. We also prove that the probability that P_n contains any given graph H as an induced subgraph is asymptotically 0 or $\frac{1}{2}$ or 1. Further we show that almost all perfect graphs are 2-clique-colourable, improving a result of Bacsó et al from 2004; they are almost all Hamiltonian; they almost all have connectivity $\kappa(P_n)$ equal to their minimum degree; they are almost all in class one (edge-colourable using Δ colours, where Δ is the maximum degree); and a sequence of independently and uniformly sampled perfect graphs of increasing size converges almost surely to the graphon $W_P(x, y) = \frac{1}{2}(\mathbf{1}[x \leq 1/2] + \mathbf{1}[y \leq 1/2])$.

Keywords: Perfect graphs, Edge-colouring, Clique-colouring, Hamiltonian, Graph limits

1 Introduction

A graph is *perfect* if the chromatic number equals the clique number in each of its induced subgraphs. Perfect graphs have formed a central field in graph theory for some decades, partly because of the challenging open problems and partly because of the connections to polyhedral combinatorics, linear optimisation and computational complexity, see for example [RAR01]. Fundamental classes of graphs that are all perfect include bipartite graphs, chordal graphs, comparability graphs and interval graphs.

Generalised split graphs form another important class of perfect graphs. A graph G is *unipolar* if for some $k \geq 0$ its vertex set $V(G)$ can be partitioned

into $k + 1$ cliques C_0, C_1, \dots, C_k , so that there are no edges between C_i and C_j for $1 \leq i < j \leq k$. We call C_0 the *central clique*, and the C_i for $i \geq 1$ the *side cliques*; and we call the pair (G, C_0) a *unipolar arrangement of order $v(G)$* . A graph G is *co-unipolar* if its complement \overline{G} is unipolar; and it is a *generalised split graph* if it is unipolar or co-unipolar. We denote the classes of perfect, unipolar, co-unipolar and generalised split graphs by \mathcal{P} , \mathcal{GS}^+ , \mathcal{GS}^- and \mathcal{GS} respectively. Given a class \mathcal{G} of graphs, we let \mathcal{G}_n denote the set of graphs in \mathcal{G} on vertex set $[n] := \{1, \dots, n\}$. We say that a sequence of events $(A_n)_{n \geq 1}$ holds *with high probability (or whp)* if $\mathbb{P}(A_n) = 1 - o(1)$, and that it holds *with very high probability (or vwhp)* if $\mathbb{P}(A_n) = 1 - e^{-\Omega(n)}$.

Our goal is to describe the asymptotic properties of perfect graphs. We rely on the key theorem of Prömel and Steger [PS92] that almost all perfect graphs are generalised split graphs. We present and analyse an almost uniform generation process for the graphs in \mathcal{GS}_n . This generation process, together with the Prömel-Steger theorem, yields a powerful method for working with uniformly sampled perfect graphs. The usefulness of the method will be seen when we prove a range of results about the asymptotic behaviour of perfect graphs, described in the next subsection.

1.1 Plan of the paper

In § 2 we present our method for analysing random perfect graphs. The main result is that the total variation distance between a naturally generated random generalised split graph in \mathcal{GS}_n and the uniformly sampled perfect graph P_n is $e^{-\Theta(n)}$. We introduce an important family of concentrated distributions $L(n)$, and give a number of results essential for our analysis, including surveying relevant results about random partitions, but we defer the detailed proofs concerning the generation process to § 9 near the end of the paper.

In § 3 we study the stability number $\alpha(P_n)$ and clique number $\omega(P_n)$. We show that the minimum of the two numbers is asymptotically normally distributed (with specified mean and variance), and the distribution of the maximum is very close to the distribution $L(n)$ mentioned above.

In § 4 we discuss which induced subgraphs are likely to be found in a random perfect graph, and we find a trichotomy. Depending on whether (i) $H \in \mathcal{P} \setminus \mathcal{GS}$, (ii) $H \in \mathcal{GS} \setminus \mathcal{GS}^-$ or $H \in \mathcal{GS} \setminus \mathcal{GS}^+$ and (iii) $H \in \mathcal{GS}^+ \cap \mathcal{GS}^-$, the probability that P_n contains an induced copy of H is respectively (i) $e^{-\Theta(n)}$, (ii) $1/2 \pm e^{-\Omega(n)}$ and (iii) $1 - e^{-\Omega(n)}$. For example, the complete bipartite graph $K_{2,3} \in \mathcal{GS}^- \setminus \mathcal{GS}^+$, so the probability that P_n contains an induced $K_{2,3}$ is well-estimated by $\frac{1}{2}$.

In § 5 we show that almost all perfect graphs are 2-clique colourable, improving the result of Bacsó et al [BGG⁺04] from 2004 that almost all perfect graphs are 3-clique colourable.

In § 6 we discuss Hamilton cycles, and find that the distribution $L(n)$ appears again. We show that, if $X \sim L(n)$, the probability that P_n is Hamiltonian is

$$1 - \frac{1}{2} \mathbb{P}(X > n/2) \pm e^{-\Omega(n)} = 1 - 2^{-\frac{1}{4}(1+o(1)) \log^2 n} = 1 - o(1). \quad (1)$$

In § 7 we discuss connectivity and the chromatic index of perfect graphs. We show that for a random perfect graph P_n the connectivity $\kappa(P_n)$ equals the minimum degree wwhp, and the chromatic index equals the maximum degree wwhp.

In § 8 we study the limit of a sequence of independently sampled perfect graphs P_n . In particular, we show that

$$\begin{aligned}\mathbb{P}(\delta_{\square}(P_n, W_P) \leq n^{-1/2}) &= 1 - e^{-\Omega(\sqrt{n} \log n)} \text{ and} \\ \mathbb{P}(\delta_{\square}(P_n, W_P) \leq (\log n)^{-2}) &= 1 - e^{-\Theta(n)},\end{aligned}$$

where δ_{\square} is the cut distance, and $W_P(x, y)$ is the graphon $\frac{1}{2}(\mathbb{1}[x \leq 1/2] + \mathbb{1}[y \leq 1/2])$. This implies that a sequence of uniformly and independently sampled perfect graphs of increasing size converges almost surely to W_P and provides an analytic tool for estimating subgraph densities of P_n .

In § 9 we give the proofs concerning the generation process which were deferred from § 2; and finally in § 10 we make a few concluding remarks and mention some open problems.

2 Generating perfect graphs

Before we go any further, let us check that generalised split graphs are perfect. The weak perfect graph theorem states that the complement of a perfect graph is perfect [Lov72a, Lov72b], so it suffices to consider a unipolar graph G , and show it is perfect. Also, each induced subgraph of a unipolar graph is unipolar, so it suffices to show that $\chi(G) = \omega(G)$.

Consider a unipolar arrangement (G, C_0) , where C_0 is the central clique, and let C_1, \dots, C_k be the side cliques. Let G_j denote the induced subgraph $G[C_0 \cup C_j]$. Bipartite graphs are perfect, so co-bipartite graphs are perfect, and thus $\chi(G_j) = \omega(G_j)$ for each j . Therefore we can colour each G_j with $\max_j \omega(G_j)$ colours, and then make sure that the different colourings agree on the central clique C_0 . It is easy to see that $\omega(G) = \max_j \omega(G_j)$ and it follows that $\chi(G) = \omega(G)$, as required.

We next discuss some probability distributions on the set \mathcal{G}_n of all graphs on $[n]$. Probably the most well-known is that of the binomial random graph $G(n, p)$, which assigns probability $p^{e(G)}(1-p)^{e(\overline{G})}$ to each graph G in \mathcal{G}_n . We are predominantly interested in the following probability distribution on \mathcal{G}_n :

$$\mathbb{P}_{\mathcal{P}_n}(G) = \begin{cases} 1/|\mathcal{P}_n| & \text{if } G \in \mathcal{P}_n \\ 0 & \text{otherwise.} \end{cases}$$

Here is another distribution on \mathcal{G}_n :

$$\mathbb{P}_{\mathcal{GS}_n}(G) = \begin{cases} 1/|\mathcal{GS}_n| & \text{if } G \in \mathcal{GS}_n \\ 0 & \text{otherwise.} \end{cases}$$

flexible We rely heavily on Theorem 2.4 from [PS92], which for our purposes can be reformulated as follows:

$$|\mathcal{GS}_n| = (1 - e^{-\Omega(n)})|\mathcal{P}_n|. \quad (2)$$

In other words, the class of \mathcal{GS} -graphs is a very good asymptotic approximation to the class of perfect graphs, as all \mathcal{GS} -graphs are perfect and almost all perfect graphs are \mathcal{GS} -graphs. The bound in (2) is sharp. Fix any graph $U \in \mathcal{P} \setminus \mathcal{GS}$, for example take the disjoint union of $K_{2,3}$ and K_3 , and consider the family \mathcal{F}^U of the graphs that are a disjoint union of U and a perfect graph. We have $\mathcal{F}^U \subseteq \mathcal{P} \setminus \mathcal{GS}$ and $|\mathcal{F}_n^U| \geq |\mathcal{P}_{n-v(U)}|$, which combined with estimates on the size of \mathcal{P}_n , e.g. $|\mathcal{P}_n|/|\mathcal{P}_{n-1}| = 2^{(1+o(1))n/2}$, implies

$$|\mathcal{GS}_n| = (1 - e^{-\Theta(n)})|\mathcal{P}_n|. \quad (3)$$

We can also express (3) using total variation distance:

$$\begin{aligned} d_{TV}(\mathbb{P}_{\mathcal{GS}_n}, \mathbb{P}_{\mathcal{P}_n}) &= \sup_{A \subseteq \mathcal{G}_n} |\mathbb{P}_{\mathcal{GS}_n}(A) - \mathbb{P}_{\mathcal{P}_n}(A)| = \mathbb{P}_{\mathcal{P}_n}(\mathcal{P}_n \setminus \mathcal{GS}_n) \\ &= \frac{|\mathcal{P}_n \setminus \mathcal{GS}_n|}{|\mathcal{P}_n|} = e^{-\Theta(n)}. \end{aligned} \quad (4)$$

How many n -vertex perfect graphs are there? The number is very close to the number of n -vertex \mathcal{GS} -graphs, which is approximately twice the number of n -vertex unipolar graphs. The last class is easy to estimate: for each k between 0 and n there are $\binom{n}{k}$ ways to choose which k vertices to include in the central clique, B_{n-k} ways to specify the partition of the remaining vertices and $2^{k(n-k)}$ ways to specify the edges between the central clique and the remaining vertices. Here B_n stands for the n -th Bell number, which is the number of ways to partition an n -element set. Note that the last calculation is not precise, since unipolar graphs may be part of more than one unipolar arrangement. If we denote $\binom{n}{k} 2^{k(n-k)} B_{n-k}$ by $\ell_{n,k}$, we would expect that the number of unipolar graphs is close to $\mathcal{L}_n := \sum_{k=0}^n \ell_{n,k}$. This turns out to be a very precise estimate and we show in § 9 that

$$|\mathcal{GS}_n| = 2\mathcal{L}_n(1 - e^{-\Omega(n)}). \quad (5)$$

Definition 2.1. For each n let $L(n)$ be the discrete integer-valued distribution with probability mass function

$$p_{L(n)}(x) = \ell_{n,x}/\mathcal{L}_n \quad \text{for } x = 0, \dots, n.$$

2.1 The generation process

Now we describe our method to sample almost uniformly from the set of generalised split graphs on $[n]$.

Definition 2.2. For each n let $Gen(n)$ be a random quadruple $(B, E, (k, \sigma), \pi)$, such that

1. flexible B is a $\{-1, 1\}$ -valued random variable, taking each value with probability half,
2. E is the set of edges of an Erdős-Rényi random graph $G(n, \frac{1}{2})$,
3. (k, σ) is the result of the following random experiment: choose a number $k \in [0, n]$ with distribution $L(n)$, set $W = \{k+1, \dots, n\}$, and then sample σ uniformly at random from $\Pi(W)$, the set of all partitions of W ,
4. π is a uniformly sampled permutation from $\text{Sym}([n])$,
5. the components of the quadruple are independent.

Let ρ be the (deterministic) map from such quadruples to \mathcal{G}_n defined as follows. Let $U = (B, E, (k, \sigma), \pi)$ be an outcome of $\text{Gen}(n)$. Start with an n -vertex empty graph $G_0 = ([n], \emptyset)$. It is convenient to write $C = [k]$ and $S = V(G) \setminus C$. Create G_1 by adding edges to G_0 , so that C induces a clique, and S induces a disjoint union of cliques corresponding to σ , i.e. $ij \in E(G_1)$ for $1 \leq i < j \leq k$; and $ij \in E(G_1)$ whenever i and j are contained in the same part of σ for $k < i < j \leq n$. Let G_2 be obtained from G_1 by adding the edges $E \cap (C \times S)$, and let G_3 be G_2 if $B = 1$ and $\overline{G_2}$ otherwise. Finally let G_4 be obtained by permuting the vertices of G_3 according to π , i.e. $\pi(i)\pi(j) \in E(G_4)$ iff $ij \in E(G_3)$. We define $\rho(U) = G_4$ for each such vector U .

Consider $U = (B, E, (k, \sigma), \pi) \sim \text{Gen}(n)$ and the corresponding random graph $G = \rho(U)$. It is easy to see that G is unipolar if $B = 1$ and co-unipolar if $B = -1$. Indeed, $(G, \pi(C))$ is a unipolar arrangement if $B = 1$ and $(\overline{G}, \pi(C))$ is a unipolar arrangement if $B = 1$ otherwise. We often refer to the appropriate pair as the *induced arrangement* of $\rho(U)$ by the generation. Whenever we want to test if G has some isomorphism-closed graph property, π does not affect the membership, so we can assume that π is the identity permutation. All graph properties in which we are interested are isomorphism-closed, so π will largely be ignored.

We define our last probability distribution, $\mathbb{P}_{\text{Gen}(n)}$ on \mathcal{G}_n , as follows:

$$\mathbb{P}_{\text{Gen}(n)}(G) = \mathbb{P}(\rho(\text{Gen}(n)) = G).$$

By the above, $\mathbb{P}_{\text{Gen}(n)}(G) > 0$ if and only if $G \in \mathcal{GS}_n$. We shall show that

$$d_{TV}(\mathbb{P}_{\mathcal{GS}_n}, \mathbb{P}_{\text{Gen}(n)}) = e^{-\Omega(n)}. \quad (6)$$

It is not hard to prove (6) from (5). However, both equations have technical proofs, so they are postponed to § 9.

Theorem 2.3. *The uniform measure $\mathbb{P}_{\mathcal{P}_n}$ and the generated measure $\mathbb{P}_{\text{Gen}(n)}$ satisfy*

$$d_{TV}(\mathbb{P}_{\mathcal{P}_n}, \mathbb{P}_{\text{Gen}(n)}) = e^{-\Theta(n)}.$$

Proof. From the triangle inequality for total variation distance, (4) and (6), we have

$$d_{TV}(\mathbb{P}_{\mathcal{P}_n}, \mathbb{P}_{Gen(n)}) \leq d_{TV}(\mathbb{P}_{\mathcal{P}_n}, \mathbb{P}_{\mathcal{G}_{S_n}}) + d_{TV}(\mathbb{P}_{\mathcal{G}_{S_n}}, \mathbb{P}_{Gen(n)}) = e^{-\Theta(n)}.$$

A lowerbound follows from arguments similar to these for (3). \square

Theorem 2.3 describes the precision of our almost-uniform sampling. Note that any approach relying on (3) cannot hope to achieve better precision. The properties of the generated graph $\rho((B, E, P, \pi))$ depend a lot on the value of B . It is often the case that we break the analysis into different cases depending on the value of B . We introduce two new random 4-tuples - $Gen^+(n)$ and $Gen^-(n)$, which are the same as $Gen(n)$, but with overwritten values for B :

Definition 2.4. For each n define $Gen^+(n)$ to be $(1, E, (k, \sigma), \pi)$ and $Gen^-(n)$ to be $(-1, E, (k, \sigma), \pi)$, where $E, (k, \sigma)$ and π are the same as in the definition of $Gen(n)$.

The following observation is easily verified: Let Q be an arbitrary graph property. If $G \sim \rho(Gen(n))$, $G^+ \sim \rho(Gen^+(n))$ and $G^- \sim \rho(Gen^-(n))$, then

$$\mathbb{P}(G \in Q) = \frac{1}{2}\mathbb{P}(G^+ \in Q) + \frac{1}{2}\mathbb{P}(G^- \in Q). \quad (7)$$

2.2 On the parts of the generation process

Theorem 2.3 will prove to be very useful. Let $P_n \in_u \mathcal{P}_n$. Whenever we want to estimate $q_n = \mathbb{P}_{\mathcal{P}_n}(Q) = \mathbb{P}(P_n \in Q)$ for some graph property Q , we can instead estimate $q'_n = \mathbb{P}_{Gen(n)}(Q) = \mathbb{P}(\rho(Gen(n)) \in Q)$, and use the fact that $|q_n - q'_n| = e^{-\Omega(n)}$. The approximate sampling, $Gen(n)$, is a convenient method for estimating probability, because it decomposes the resulting graph into three independent parts – a central clique, a partition of the remaining vertices and a bipartite part in between.

2.2.1 Central clique

An essential part of the generation process involves determining the size of the central clique. For this purpose we introduced the distribution $L(n)$. Recall that if $X \sim L(n)$ and x is an integer between 0 and n , then $\mathbb{P}(X = x) = \ell_{n,x}/\mathcal{L}_n$, where $\ell_{n,x} = \binom{n}{x} 2^{x(n-x)} B_{n-x}$ and $\mathcal{L}_n = \sum_{k=0}^n \ell_{n,k}$. The term $2^{x(n-x)}$ dwarfs the other two terms and heavily concentrates the distribution. The term B_{n-x} breaks the symmetry around $n/2$ and tilts the mean towards zero by about $\frac{1}{2} \log n$.

Theorem 2.5. Suppose $X \sim L(n)$. For sufficiently large n we have

$$2^{-(x+1)^2-1} \leq \mathbb{P}(|X - \mu| \geq x) \leq 2^{-(x-2)^2+2} + n^{-n},$$

for each $x > 1$, where

$$\mu = \mu(n) = \frac{n - \log n + \log \ln n}{2}.$$

The upper bound in the theorem above is essential for handling the generation process. Indeed, Theorem 2.3 will always be used in combination with Theorem 2.5. Most often we take $x = \Omega(\sqrt{n})$ in order to have an exponential decay. However, sometimes there is phase transition when the central clique becomes bigger than the remaining graph, most notably when we look at Hamiltonian cycles in co-unipolar graphs, and then we take x to be $n/2 - \mu \sim \frac{1}{2} \log n$. The proof of Theorem 2.5 is postponed to § 9.3. In the proof we also show that if $X \sim L(n)$, then $|\mathbb{E}X - \mu(n)| < 1$.

2.2.2 Bipartite part

Let $U \sim \text{Gen}(n)$ and let (G, C) be the arrangement induced by the generation process. The edges between C and \overline{C} are best seen as a random bipartite graph with colour classes C and \overline{C} and edges in between, each present with probability $1/2$ and independently of the others. The sizes of C and \overline{C} are not necessarily identical, but Theorem 2.5 shows that their difference is $O(\sqrt{n})$ wvhp. Random bipartite graphs share many properties with binomial random graphs $G(n, p)$, which have been exhaustively studied.

2.2.3 Partition part

Let Π_n be the set of all partitions of $[n]$. The elements of each partition will be referred to as *parts*. We turn Π_n into a probability space by taking each partition with equal probability, namely $1/|\Pi_n|$. Let $\sigma = \sigma(n)$ be an outcome from this probability space, i.e. $\sigma \in_u \Pi_n$. The structure of σ can vary from having a single part with n elements to having n parts with one element each. However, σ has a well-defined shape with high probability.

Let $r = r(n)$ be the unique root of the equation $re^r = n$ and let $B_n = |\Pi_n|$ be the n th Bell number. It can be found in [DB70] that

$$r(n) = \ln n - (1 + o(1)) \ln \ln n, \quad \text{and} \quad (8)$$

$$B_n = \frac{1 + o(1)}{\sqrt{r}} e^{n(r-1+1/r)-1}. \quad (9)$$

There is a sizeable literature for random partitions, see in particular [Pit97]. Let $|\pi|$ be the number of parts in a partition π . If X is a random variable we will use \widehat{X} to denote $(X - \mathbb{E}X)/\sqrt{\text{Var}X}$. One of the earliest results on random partitions is by Harper:

Theorem 2.6. (Harper [Har66]) For $\sigma \in_u \Pi_n$

1. $|\widehat{\sigma}| \xrightarrow{d} N(0, 1)$,
2. $\mathbb{E}|\sigma| = \frac{B_{n+1}}{B_n} - 1 = (1 + o(1)) \frac{n}{r} = (1 + o(1)) \frac{n}{\ln n}$,
3. $\text{Var}|\sigma| = \frac{B_{n+2}}{B_n} - \left(\frac{B_{n+1}}{B_n}\right)^2 - 1 = \frac{n}{r(r+1)} - 1 + o(1) = (1 + o(1)) \frac{n}{(\ln n)^2}$.

Let $Y_t(\pi)$ be the number of parts of size t of an arbitrary partition $\pi \in \Pi_n$, so that $|\pi| = \sum_{t=1}^n Y_t(\pi)$ and $n = \sum_{t=1}^n tY_t(\pi)$. Suppose that $\sigma \in_u \Pi_n$ from now on.

Theorem 2.7. (*Pittel [Pit97]*) *Given a partition π and an interval $I = [k_1, k_2]$ where $1 \leq k_1 \leq k_2 \leq n$, let*

$$Y_I(\pi) = \sum_{j \in I} Y_j(\pi), \quad \lambda_I = \sum_{j \in I} \frac{r^j}{j!}.$$

Then, uniformly for $I = [k_1, k_2]$ and $\xi \geq 0$,

$$\mathbb{P}(|Y_I(\sigma) - \lambda_I| \geq \xi) = O(e^{-\min(\frac{\xi^2}{4\lambda_I}, \frac{\xi}{2})}).$$

The last theorem is useful for small deviations, but rather crude for deviations proportional to the mean. A counting argument yields a better bound:

Lemma 2.8. *Let $\epsilon > 0$ be fixed and let $\lambda = n/r = e^r$. Then for large n*

$$\mathbb{P}(|\sigma| - \lambda| \geq \epsilon\lambda) < ne^{-n(\epsilon - \ln(\epsilon+1))}.$$

In particular, $|\sigma| - \lambda < \epsilon\lambda$ wvhp.

Proof. For each $0 \leq k \leq n$ we have

$$\mathbb{P}(|\sigma| = k)B_n \leq \frac{k^n}{k!} \leq k^n \left(\frac{k}{e}\right)^{-k} = \exp(n \ln k - k \ln k + k).$$

Suppose that $k = (1+c)\lambda$ for some $c = c(n) > -1$. Substituting in the inequality above we get:

$$\begin{aligned} \mathbb{P}(|\sigma| = (1+c)\lambda) &\leq B_n^{-1} \exp(n \ln(1+c)\lambda - (1+c)\lambda \ln(1+c)\lambda + (1+c)\lambda) \\ &= B_n^{-1} \exp\{ n \ln \lambda - \lambda \ln \lambda + \lambda \\ &\quad + n \ln(1+c) - c\lambda \ln \lambda \\ &\quad + \lambda(c - \ln(1+c) - c \ln(1+c)) \} \\ &\leq \exp(n\{\ln(1+c) - c\} + \lambda\{c - \ln(1+c) - c \ln(1+c)\}) \end{aligned}$$

for sufficiently large n . We have used $\lambda e^\lambda = n$ and that (9) implies $B_n^{-1} \exp(n(r - 1 + 1/r)) < 1$ for sufficiently large n . Note that $c - \ln(1+c) - c \ln(1+c) \leq 0$ for all $c > -1$, so we can write

$$\mathbb{P}(|\sigma| = (1+c)\lambda) \leq \exp(n\{\ln(1+c) - c\})$$

for large n . Let $f(x) : (-1, +\infty) \rightarrow \mathbb{R}$ be defined as $f(x) = \ln(1+x) - x$, so that

$$\mathbb{P}(|\sigma| = (1+c)\lambda) \leq \exp(nf(c)).$$

The following properties are easily verified by taking derivatives: $f(x) \leq f(|x|)$; $f(0) = 0$; $f(x) < 0$ for $x \neq 0$; and f is strictly increasing in $(-1, 0]$ and strictly decreasing in $[0, \infty)$. We deduce that for every $I \subseteq (-1, \infty]$ we have $\sup_{x \in I} f(x) \leq f(\inf_{x \in I} |x|)$. Let $I_\delta = [0, n] \setminus [[(1 - \delta)\lambda], \lceil(1 + \delta)\lambda\rceil]$, for $\delta > 0$. We deduce

$$\mathbb{P}(|\sigma| - \lambda \geq \epsilon\lambda) = \mathbb{P}(|\sigma| \in I_\epsilon) \leq \max_{k \in I_\epsilon} \mathbb{P}(|\sigma| = k) < n \exp\{f(\epsilon)n\}. \quad \square$$

For $\pi \in \Pi_n$ let $L(\pi)$ be the maximum size of a part.

Lemma 2.9. *Suppose $0 \leq x \leq n/2$ and let $\sigma \in_u \Pi_n$. We have*

$$\mathbb{P}(L(\sigma) \geq x) < (1 + o(1))e^{-x(\ln x - \ln r - 2) + \ln n}.$$

Proof. We may assume that $x \geq e^2r$ (since otherwise the RHS is ≥ 1). Let E_x be the event that σ contains a part of size $\lfloor x \rfloor$. Then by (9) and noting that $r_{n-x} - 1 + \frac{1}{r_{n-x}} \leq r_n - 1 + \frac{1}{r_n}$ (this holds because $f(t) = t - 1 + t^{-1}$ and r_t are increasing in $t \in [1, \infty)$) we get

$$\begin{aligned} \mathbb{P}(E_x) &\leq \binom{n}{\lfloor x \rfloor} \frac{B_{n-\lfloor x \rfloor}}{B_n} \leq \left(\frac{en}{x}\right)^x (1 + o(1)) \sqrt{\frac{r_n}{r_{n-x}}} e^{-x(r_n - 1 + 1/r_n)} \\ &\leq (1 + o(1)) \left(\frac{en}{x}\right)^x \left(\frac{n}{er_n}\right)^{-x} \\ &= (1 + o(1))e^{-x(\ln x - \ln r - 2)}. \end{aligned}$$

Thus

$$\begin{aligned} \mathbb{P}(L(\sigma) \geq x) &= \mathbb{P}\left(\bigcup_{z \geq x} E_z\right) \leq \sum_{z \geq x} \mathbb{P}(E_z) < n \max_{z \geq x} \mathbb{P}(E_z) \\ &\leq (1 + o(1))e^{-x(\ln x - \ln r - 2) + \ln n}, \end{aligned}$$

since $\ln x - \ln r - 2 \geq 0$. \square

Lemma 2.10. *Suppose $a = a(n) = o(n)$ is integral for all n . Then*

$$\frac{B_{n-a}}{B_n} = \exp(-ar_{n-a} + o(n)).$$

Proof. We have

$$\begin{aligned} \frac{B_{n-a}}{B_n} &= (1 + o(1)) \exp\left((n-a) \left(r_{n-a} - 1 + \frac{1}{r_{n-a}}\right) - n \left(r_n - 1 + \frac{1}{r_n}\right)\right) \\ &= \exp((1 + o(1))n \ln(1 - a/n) - a(r_{n-a} - 1) + o(n)) \\ &= \exp((1 + o(1))(-a) - a(r_{n-a} - 1) + o(n)) \\ &= \exp(-ar_{n-a} + o(n)). \end{aligned} \quad \square$$

Corollary 2.11. *We have*

$$\mathbb{P}(L(\sigma) \geq n/\log n) = e^{-\Theta(n)}. \quad (10)$$

and further, if $\epsilon(n) = o(1)$, then

$$\mathbb{P}(L(\sigma) \geq \epsilon(n)n/\log n) = e^{-o(n)}. \quad (11)$$

Proof. Let $a = \lceil n/\log n \rceil$. Then $\mathbb{P}(L(\sigma) \geq n/\log n)$ is at least $\frac{B_n - a}{B_n}$; and is at most $\binom{n}{a} n \frac{B_n - a}{B_n}$ (where the factor n is to account for choosing a part (or the empty set) to add to the chosen part of size a). Now we can use the last lemma. \square

3 Stability and Clique number

In this section we discuss the stability number $\alpha(P_n)$ and clique number $\omega(P_n)$ of the random perfect graph P_n . Our main result is Theorem 3.7. Of course everything we say also applies to the clique covering number and chromatic number, as they equal the former two parameters for perfect graphs.

Lemma 3.1. *Let $U = (1, E, (k, \sigma), \pi)$ be an arbitrary outcome of $\text{Gen}^+(n)$ and let $G^+ = \rho(U)$. Then $0 \leq \alpha(G^+) - |\sigma| \leq 1$.*

Proof. Let (G^+, C) be the induced unipolar arrangement from the generation. We know that σ is a partition of \overline{C} into disjoint cliques, and therefore by picking a representative member from each part we obtain an independent set of size $|\sigma|$, hence $\alpha(G^+) \geq |\sigma|$. For the other direction, observe C together with σ gives a partition of $V(G^+)$ into $|\sigma| + 1$ cliques. Every stable set can intersect each of these cliques at most once, hence $\alpha(G^+) \leq |\sigma| + 1$. \square

The last lemma pins down $\alpha(G^+)$ to only two values, namely $|\sigma|$ and $|\sigma| + 1$. We next show that the second outcome is very unlikely.

Lemma 3.2. *Let $U = (1, E, (k, \sigma), \pi) \sim \text{Gen}^+(n)$, let $G^+ = \rho(U)$, and let $0 < \delta < 1 - \ln 2 \approx 0.3$. Then $\mathbb{P}(\alpha(G^+) = |\sigma| + 1) = O(e^{-n^\delta})$.*

We give an informal proof first. Condition on the induced unipolar arrangement (G^+, C) and on the partition σ . We have $\alpha(G^+) = |\sigma| + 1$ if and only if there is a vertex from C whose neighbourhood does not contain any side clique in \overline{C} . Let A_v be the event that v is such a vertex, and hence $\{\alpha(G^+) = |\sigma| + 1\} \equiv \bigcup_v A_v$. For a fixed vertex $u \in C$ and a fixed side clique R we have $\mathbb{P}(R \not\subseteq N(u)) = 1 - 2^{-|R|}$. Therefore

$$\mathbb{P}(A_v) = \prod_R \left(1 - 2^{-|R|}\right),$$

where that product is taken over the side cliques in \overline{C} . Up until now everything is rigorous. Observe from Theorem 2.7 that whp the number of side cliques of

size i is close to $r^i/i!$, where r is the unique root of the equation $n - k = re^r$. For a moment assume that the number of side cliques of size i equals this value, for each i . Then

$$\begin{aligned}
\mathbb{P}(A_v) &= \prod_{i=1}^n \prod_{R:|R|=i} (1 - 2^{-i}) = \prod_{i=1}^n (1 - 2^{-i})^{r^i/i!} \\
&\leq \prod_{i=1}^n \left(e^{-2^{-i}} \right)^{r^i/i!} \quad \text{since } 1 + x \leq e^x \\
&= \exp \left(- \sum_{i=1}^n (r/2)^i / i! \right) \leq \exp \left(-e^{r/2} + 2 \right) \\
&= \exp \left(-\sqrt{(n-k)/r} + 2 \right) = \exp \left(-(1 + o(1))\sqrt{n/(2 \ln n)} \right).
\end{aligned}$$

For the last step we assumed that $k \sim n/2$, which is easily justifiable from the concentration theorem, Theorem 2.5. We use the union bound to finalise our arguments:

$$\begin{aligned}
\mathbb{P}(\alpha(G^+) = |\sigma| + 1) &\leq \sum_v \mathbb{P}(A_v) \leq \exp \left(-(1 + o(1))\sqrt{n/(2 \ln n)} \right) \\
&= O(\exp(-n^{1/2-\epsilon})).
\end{aligned}$$

We observe that δ from the statement is $1/2 - \epsilon$ in this case, which may be the best possible.

Proof of Lemma 3.2. Let $\mathcal{J} = [n/2 - 2\sqrt{n}, n/2 + 1 + 2\sqrt{n}]$. We have $\mathbb{P}(k \in \mathcal{J}) \geq 1 - 2^{-n}$ from Theorem 2.5. Condition σ on $k = k_0$ for some $k_0 \in \mathcal{J}$. We use Theorem 2.7 for σ with $I = [0, r + r^{3/4}]$, $\lambda_I = e^r \mathbb{P}[Po(r) \leq r + r^{3/4}] \leq e^r$ and $\xi = 2(e^r)^{3/4}$, where r is the unique root of $n - k_0 = re^r$. The theorem says that

$$\mathbb{P}(Y_I(\sigma) \geq \lambda_I - \xi | k = k_0) \geq 1 - c \exp(-\sqrt{e^r}),$$

for some constant c . In order to give a lower bound on λ_I , we use a standard Chernoff bound for a Poisson random variable: $\mathbb{P}(Po(r) \geq (1 + \eta)r) < \exp(-r\eta^2/4)$, for $0 < \eta \leq 1$. Taking $\eta = r^{-1/4}$ we obtain:

$$\lambda_I - \xi > e^r(1 - \exp(-\sqrt{r}/4)) - 2(e^r)^{3/4} = e^r(1 - o(1)).$$

Define the events A_v as in the preceding informal discussion. The event $\{Y_I(\sigma) \geq \lambda_I - \xi\}$ implies that in σ there are at least $e^r(1 + o(1))$ parts of size at most

$r(1 + o(1))$, hence

$$\begin{aligned}
& \mathbb{P}(A_v | Y_I(\sigma) \geq \lambda_I - \xi \cap k = k_0) \\
& \leq (1 - 2^{-r(1+o(1))})e^{r(1+o(1))} \\
& \leq \exp\{-2^{-r(1+o(1))}e^r(1+o(1))\} \\
& = \exp\{-(e/2)^{r(1+o(1))}\} \\
& = \exp\{-\exp[r \cdot (1 - \ln 2)(1 + o(1))]\} \\
& = \exp\{((n - k_0)/r)^{(1 - \ln 2)(1+o(1))}\} \quad \text{since } e^r = (n - k_0)/r \\
& = O(\exp(-n^\delta)).
\end{aligned}$$

From the union bound we get

$$\begin{aligned}
\mathbb{P}(\alpha(G^+) = |\sigma| + 1 | k = k_0) & \leq \mathbb{P}(Y_I(\sigma) < \lambda_I - \xi | k = k_0) \\
& + \sum_v \mathbb{P}(A_v | (Y_I(\sigma) \geq \lambda_I - \xi) \cap (k = k_0)) \\
& = O(\exp(-n^\delta)).
\end{aligned}$$

Hence

$$\begin{aligned}
\mathbb{P}(\alpha(G^+) = |\sigma| + 1) & \leq \mathbb{P}(k \notin \mathcal{J}) + \max_{k_0 \in \mathcal{J}} \mathbb{P}(\alpha(G^+) = |\sigma| | k = k_0) \\
& = O(\exp(-n^\delta)). \quad \square
\end{aligned}$$

Note that the bound for δ in the last proof is $1 - \ln 2 \approx 0.3$, which is slightly worse than the constant in the informal discussion.

We shorten $G^+ \sim \rho(\text{Gen}^+(n))$ to $G^+ \sim \text{Gen}^+(n)$.

Proposition 3.3. *Let $G^+ \sim \text{Gen}^+(n)$, and denote $\alpha(G^+)$ by X . Then, as $n \rightarrow \infty$, $\widehat{X} \xrightarrow{d} N(0, 1)$, $\mathbb{E}[X] \sim \frac{n}{2 \ln n}$, and $\text{Var}[X] \sim \frac{n}{2 \ln^2 n}$.*

We write Y for $|\sigma|$ and Y_m for $|\sigma_m|$ when $\sigma_m \in_u \Pi_m$. To prove the last result we will use the following claim.

Claim 3.4. $\mathbb{E}Y_{n+1} = \mathbb{E}Y_n + \frac{1+o(1)}{\ln n}$.

Proof. From Theorem 2.6, $\mathbb{E}Y_n = \frac{B_{n+1}}{B_n} - 1$ and $\text{Var}Y_n = \frac{B_{n+2}}{B_n} - \left(\frac{B_{n+1}}{B_n}\right)^2 - 1$. Hence

$$\text{Var}Y_n = \frac{B_{n+2}}{B_{n+1}} \frac{B_{n+1}}{B_n} - \left(\frac{B_{n+1}}{B_n}\right)^2 - 1 = (\mathbb{E}Y_{n+1} + 1)(\mathbb{E}Y_n + 1) - (\mathbb{E}Y_n + 1)^2 - 1.$$

Therefore

$$(\mathbb{E}Y_{n+1} + 1) - (\mathbb{E}Y_n + 1) = \frac{\text{Var}Y_n + 1}{\mathbb{E}Y_n + 1} = \frac{1 + o(1)}{\ln n}. \quad \square$$

Proof of Proposition 3.3. Recall that σ is a random partition created by first choosing a value for $k \sim L(n)$, and then taking $\sigma \in_u \Pi_{n-k}$. It suffices to show that $\widehat{Y} \xrightarrow{d} N(0, 1)$. Indeed, if this were the case we get $Y \leq X \leq Y+1$, $\text{var}(Y) \rightarrow \infty$, $\widehat{Y} \xrightarrow{d} N(0, 1)$, and hence $\widehat{X} \xrightarrow{d} N(0, 1)$.

Let $\Phi(x)$ be the CDF of $N(0, 1)$. For an arbitrary fixed x and $\epsilon > 0$ we show that for large n

$$|\mathbb{P}(\widehat{Y} < x) - \Phi(x)| < \epsilon.$$

Find $\ell = \ell(\epsilon)$ such that $\mathbb{P}(k \notin \mathcal{I}) < \epsilon/2 + n^{-n}$ for $\mathcal{I} = \mathcal{I}(n) = [\lfloor \mathbb{E}k - \ell \rfloor, \lceil \mathbb{E}k + \ell \rceil]$. Such ℓ exists by Theorem 2.5. We emphasise that the length of \mathcal{I} is at most $2\ell + 1$, which is independent of n .

$$\begin{aligned} \left| \mathbb{P}(\widehat{Y} < x) - \Phi(x) \right| &= \sum_{c=0}^n \left| (\mathbb{P}(\widehat{Y} < x | k = c) - \Phi(x)) \mathbb{P}(k = c) \right| \\ &< \sum_{c \in \mathcal{I}} \left| (\mathbb{P}(\widehat{Y} < x | k = c) - \Phi(x)) \right| \mathbb{P}(k = c) + \epsilon/2 + n^{-n} \\ &\leq \max_{c \in \mathcal{I}} \left| \mathbb{P}(\widehat{Y} < x | k = c) - \Phi(x) \right| + \epsilon/2 + n^{-n}. \end{aligned}$$

Fix any integer valued function $q(n)$ with $q(n) \in \mathcal{I}(n)$.

$$\begin{aligned} |\mathbb{E}Y - \mathbb{E}Y_{n-q}| &= |\mathbb{E} \{ \mathbb{E}[Y | k] \} - \mathbb{E}Y_{n-q}| \\ &\leq \sum_{i=0}^n |\mathbb{E}Y_{n-i} - \mathbb{E}Y_{n-q}| \mathbb{P}(k = i) \\ &\leq \sum_{i \in \mathcal{I}} |\mathbb{E}Y_{n-i} - \mathbb{E}Y_{n-q}| \mathbb{P}(k = i) + \epsilon/2 + n^{-n} \\ &\leq \max_{i \in \mathcal{I}} |\mathbb{E}Y_{n-i} - \mathbb{E}Y_{n-q}| + \epsilon/2 + n^{-n} \\ &\leq (1 + o(1)) \frac{2\ell + 1}{2 \ln n} + \epsilon/2 + n^{-n} \\ &= \epsilon/2 + o(1). \end{aligned} \tag{12}$$

We proceed with the variance in a similar fashion.

$$|\text{Var}Y - \text{Var}Y_{n-q}| \leq \max_{i \in \mathcal{I}} |\text{Var}Y_{n-i} - \text{Var}Y_{n-q}| + \epsilon/2 + n^{-n} = o(\text{Var}Y_{n-q}).$$

For the last equality we have used Theorem 2.6 part 3. We conclude

$$\text{Var}Y = (1 + o(1)) \text{Var}Y_{n-q}. \tag{13}$$

Write $y = (x(\text{Var}Y)^{1/2} + \mathbb{E}Y - \mathbb{E}Y_{n-q})(\text{Var}Y_{n-q})^{-1/2}$, so that $\mathbb{P}(\widehat{Y} < x | k = q) = \mathbb{P}(\widehat{Y}_{n-q} < y)$. From (12), (13) and Theorem 2.6 we have $\lim_{n \rightarrow \infty} y = x$. Find $\delta > 0$, so that $\Phi(x) \leq \Phi(x - \delta) + \epsilon/6$. From Theorem 2.6 we have $\mathbb{P}(\widehat{Y}_{n-q} < x - \delta) \rightarrow \Phi(x - \delta)$ if $n - q \rightarrow \infty$. For large n

$$\mathbb{P}(\widehat{Y}_{n-q} < y) \geq \mathbb{P}(\widehat{Y}_{n-q} < x - \delta) \geq \Phi(x - \delta) - \epsilon/6 \geq \Phi(x) - \epsilon/3.$$

We can do the same in the other direction and deduce that for large n

$$|\mathbb{P}(\widehat{Y} < x | k = q) - \Phi(x)| < \epsilon/3.$$

But q was an arbitrary function, hence for large n

$$\max_{c \in \mathcal{I}} \left| \mathbb{P}(\widehat{Y} < x | k = c) - \Phi(x) \right| < \epsilon/3,$$

and the proposition is proven. \square

We continue with $\omega(G^+)$. Usually there is a unique clique of maximum size in G^+ – the central clique. The vertices in the central clique have degree $(1+o(1))3n/4$ whp, while the vertices outside have degree $(1+o(1))n/4$ whp, so the central clique C from the generation is easily recognisable in the resulting graph (and in the 4-tuple U , as well as k and B , we know σ up to relabelling of \overline{C} , we know a little about E and we know $\pi([k]) = C$).

Lemma 3.5. *Suppose $G^+ = \rho(U)$ where $U \sim \text{Gen}^+(n)$. Then $\mathbb{P}(\omega(G^+) \neq k) = 2^{-(\frac{1}{2}+o(1))n}$.*

Proof. As usual, let $\mathcal{I} = [n/2 - \sqrt{n}, n/2 + \sqrt{n}]$, and recall that $k \in \mathcal{I}$ whp by Theorem 2.5. Let $k_0 \in \mathcal{I}$, and condition on $k = k_0$.

We first give a lower bound for $\mathbb{P}(\omega(G^+) \neq k_0)$. With probability at least 2^{-k_0} there is a vertex outside the central clique which is adjacent to all vertices from the central clique, and hence $\mathbb{P}(\omega(G^+) > k_0) \geq 2^{-(\frac{1}{2}+o(1))n}$.

We now prove an upper bound. Let (G^+, C) be the unipolar arrangement induced by U . There are two different cases in which G^+ contains a clique M of size more than k_0 . The first case is when $|\overline{C} \cap M| = 1$, when also $M \supseteq C$. The probability of that happening is at most $(n - k_0)2^{-k_0} = 2^{-(\frac{1}{2}+o(1))n}$.

Let us focus on the second case; $|\overline{C} \cap M| > 1$. This implies that there are at least two vertices from \overline{C} with degree at least k_0 . If we substitute $n/\sqrt{\log n}$ for x in Lemma 2.9, we get that the probability that there is a part in σ of size more than $n/\sqrt{\log n}$ is less than e^{-n} for large n . Assume there is no such part in σ .

For every $v \in \overline{C}$ write $d(v) = d_C(v) + d_{\overline{C}}(v)$. We have $d_C(v) \sim \text{Bin}(k_0, 1/2)$ and $d_{\overline{C}}(v) \leq n/\sqrt{\log n}$. Let E_v be the event $\{d_C(v) \geq k_0 - 1 - n/\sqrt{\log n}\}$. Next we use the Chernoff bound in the form: $\mathbb{P}(X \geq mp + t) \leq e^{-2t^2/m}$, for $X \sim \text{Bin}(m, p)$ and $t \geq 0$. We obtain

$$\mathbb{P}(E_v) \leq e^{-(1+o(1))k_0/2}.$$

The events E_v and E_u are independent for $u \neq v$, hence $\mathbb{P}(E_v \cap E_u) \leq e^{-(1+o(1))k_0}$ for every pair $u \neq v$. We deduce that the probability that there are two vertices from \overline{C} with degree at least k_0 is at most $n^2 e^{-(1+o(1))k_0} + e^{-n} = e^{-(\frac{1}{2}+o(1))n}$.

Combining the two cases we conclude that $\mathbb{P}(\omega(G^+) > k_0) \leq 2^{-(\frac{1}{2}+o(1))n}$. \square

Lemma 3.6. For $G^+ \sim \text{Gen}^+(n)$ and a fixed $\epsilon > 0$ we have whp

$$(1 - \epsilon) \frac{n}{2 \ln n} < \alpha(G^+) < (1 + \epsilon) \frac{n}{2 \ln n}, \text{ and}$$

$$\frac{n}{2} - \epsilon \sqrt{n} < \omega(G^+) < \frac{n}{2} + \epsilon \sqrt{n}.$$

In particular, $\alpha(G^+) < \omega(G^+)$ whp.

Proof. The first line follows from Lemma 3.1, Lemma 2.8 and Theorem 2.5; the second line follows from Theorem 2.5 and Lemma 3.5. \square

Observe that a proof of Theorem 2.5 will suffice to complete the proof of Lemma 3.6: we shall use this later. For every graph G we have $\omega(G) = \alpha(\overline{G})$ and therefore everything we showed for $\alpha(G^+)$ and $\omega(G^+)$ directly translates to $\omega(G^-)$ and $\alpha(G^-)$. Let $h(G) = \min\{\alpha(G), \omega(G)\}$ and $H(G) = \max\{\alpha(G), \omega(G)\}$ for every graph G . (Here h and H come from homogeneous.)

Theorem 3.7. For $P_n \in_u \mathcal{P}_n$,

1. $h(P_n)$ is asymptotically normal with mean $\sim \frac{n}{2 \ln n}$ and variance $\sim \frac{n}{2 \ln^2 n}$,
2. $d_{TV}(H(P_n), L(n)) = e^{-\Omega(n)}$.

Proof. Let $U = (B, E, (k, \sigma), \pi) \sim \text{Gen}(n)$. We couple G_n, G_n^+ and G_n^- by writing $U^+ = (1, E, (k, \sigma), \pi)$, $U^- = (-1, E, (k, \sigma), \pi)$, $G_n = \rho(U)$, $G_n^+ = \rho(U^+)$ and $G_n^- = \rho(U^-)$. We have

$$\begin{aligned} d_{TV}(h(P_n), \alpha(G_n^+)) &\leq d_{TV}(h(P_n), h(G_n)) + d_{TV}(h(G_n), \alpha(G_n^+)) \\ &\leq d_{TV}(\mathbb{P}_{\mathcal{P}_n}, \mathbb{P}_{\text{Gen}(n)}) + \mathbb{P}(h(G_n) \neq \alpha(G_n^+)). \end{aligned}$$

But $d_{TV}(\mathbb{P}_{\mathcal{P}_n}, \mathbb{P}_{\text{Gen}(n)}) = e^{-\Omega(n)}$ from Theorem 2.3; and, since $h(G_n) = h(G_n^+)$,

$$\mathbb{P}(h(G_n) \neq \alpha(G_n^+)) = \mathbb{P}(\omega(G_n^+) < \alpha(G_n^+)) = e^{-\Omega(n)}$$

from Lemma 3.6. Thus $d_{TV}(h(P_n), \alpha(G^+)) = e^{-\Omega(n)}$. The first part in the statement now follows from Proposition 3.3.

Similar arguments establish the second part of the theorem. Much as above we may write

$$d_{TV}(H(P_n), L(n)) \leq d_{TV}(H(P_n), H(G_n)) + d_{TV}(H(G_n), \omega(G_n^+)) + d_{TV}(\omega(G_n^+), L(n)).$$

The first and third terms in the upper bound here are $e^{-\Omega(n)}$ by Theorem 2.3 and Lemma 3.5. Since $H(G_n) = H(G_n^+)$, the second term is at most $\mathbb{P}(\alpha(G_n^+) > \omega(G_n^+))$, which is $e^{-\Omega(n)}$ by Lemma 3.6. Thus $d_{TV}(H(P_n), L(n)) = e^{-\Omega(n)}$, as required. \square

The last result shows that the distribution $L(n)$ is not only needed for the approximate generation process, but also corresponds to an important property of perfect graphs.

4 Induced Subgraphs

Suppose $P_n \in_u \mathcal{P}_n$. From Lemma 3.6 and Theorem 2.3 we have $\omega(P_n) \geq \frac{(1-\epsilon)n}{2 \ln n}$ wvhp, so we can of course find any fixed graph H as a subgraph of P_n wvhp.

We use $H \subseteq_i G$ to denote that H is an induced subgraph of G . We observed already that the generated graphs, $\rho(\text{Gen}(n))$, are symmetric with respect to taking complements. The same is true for perfect graphs, that is $P_n \sim \overline{P_n}$, by the weak perfect graph theorem [Lov72a, Lov72b]. Thus we see that, if H is fixed and $P_n \in_u \mathcal{P}_n$, then

$$\mathbb{P}(H \subseteq_i P_n) = \mathbb{P}(\overline{H} \subseteq_i P_n). \quad (14)$$

Lemma 4.1. *Suppose that $G_n^+ \sim \text{Gen}^+(n)$ and that H is fixed. Then*

$$\mathbb{P}(H \subseteq_i G_n^+) = \begin{cases} 0 & \text{if } H \notin \mathcal{GS}^+, \\ 1 - e^{-\Omega(n \ln n)} & \text{if } H \in \mathcal{GS}^+. \end{cases}$$

To prove this lemma, we first establish two preliminary claims.

Claim 4.2. *Suppose $\sigma \in_u \Pi_n$, and let $a = a(n) \leq n^\delta$ for some $0 < \delta < 1$. The probability that σ contains at most a parts is at most $\exp\{-(1-\delta+o(1))n \ln n\}$.*

Proof. The probability of the event in the statement is at most a^n/B_n , and the bound follows, since $\ln B_n = (1+o(1))n \ln n$. \square

Claim 4.3. *Suppose $\sigma \in_u \Pi_n$, and let $l \geq 1$ be a fixed integer. The probability that all parts of σ are of size at most l is at most $\exp\{-(\frac{1}{l}+o(1))n \ln n\}$.*

Proof. We call a finite sequence $(a_i)_{i=1}^k$ *type* if $\sum_{i=1}^k a_i = n$. We say that σ has type (a_i) if (a_i) corresponds to the sizes of the parts of σ taken in any order. The number of partitions with type (a_i) with $a_i \leq l$ for all i is

$$\frac{n!}{k! \prod_{i=1}^k a_i!} \leq \frac{n!}{k!} \leq \frac{e\sqrt{n}}{\sqrt{2\pi k}} \cdot \frac{n^n e^{-n}}{k^k e^{-k}} = \frac{n^n}{(\frac{n}{l})^{\frac{n}{l}}} 2^{O(n)} = \exp\left\{\left(1 - \frac{1}{l}\right)n \ln n + O(n)\right\}.$$

The number of types (a_i) with $a_i \leq l$ for all i is at most $l^n = \exp(O(n))$, and hence the number of partitions with parts at most l is $\exp\{(1 - \frac{1}{l})n \ln n + O(n)\}$. The statement in the lemma follows after dividing by B_n . \square

Corollary 4.4. *For fixed $\delta > 0$ and $l \geq 1$, the probability that a random partition $\sigma \in_u \Pi_n$ contains at most n^δ parts of size more than l is $\exp(-\Theta(n \ln n))$.*

Proof. Let Π'_Q and Π''_Q be the classes of partitions of Q with at most n^δ parts and with parts of size at most l respectively. The number of partitions described in the statement can be bounded by $\sum_{S \subseteq [n]} |\Pi'_S| |\Pi''_S| = \exp(-O(n \ln n))$ by the previous two claims. For a lowerbound consider the partition π with n parts; $\mathbb{P}(\sigma = \pi) = B_n^{-1} = \exp(-\Theta(n \ln n))$. \square

Proof of lemma 4.1. Since \mathcal{GS}^+ is a hereditary class of graphs and G_n^+ takes as values only \mathcal{GS}^+ -graphs with non-zero probability, we get that if $H \notin \mathcal{GS}^+$, then $\mathbb{P}(H \subseteq_i G_n^+) = 0$ for any n .

Now suppose that $H \in \mathcal{GS}^+$. Fix a unipolar arrangement (H, A) . Suppose that $\pi \in \Pi_{\overline{A}}$ corresponds to $H[\overline{A}]$, and let $l = L(\pi)$ be the maximum size of a part in π . Let I_n be the interval $[n/2 - n^{2/3}, n/2 + n^{2/3}]$. By Theorem 2.5, $\mathbb{P}(k \notin I_n) \leq 2n^{-n}$ for n sufficiently large; so we may condition on $k \in I_n$. Let (G_n^+, C) be the unipolar arrangement induced by the generation process.

By Corollary 4.4 there are at least $n^{1/2}$ maximal cliques in \overline{C} of size at least l with probability $1 - \exp(-\Theta(n \ln n))$, so we may condition on this event. Hence we can find a collection of at least $t_1 = \lceil n^{1/2}/(3|\pi|) \rceil$ disjoint sets $T_j \subseteq \overline{C}$ such that $G_n^+[T_j] \cong H[\overline{A}]$. Pick $t_0 = \lceil n/(3|A|) \rceil$ disjoint subsets S_i of C . Then $G_n^+[S_i] \cong H[A]$ for every i , and hence for $1 \leq i \leq t_0$ and $1 \leq j \leq t_1$ we have $p_0 = \mathbb{P}(G_n^+[S_i \cup T_j] \cong H) = 2^{-|A||\overline{A}|}$. The probability that no such pair induces H is $(1 - p_0)^{t_0 t_1} = \exp(-\Theta(n^{3/2}))$. The failure probability is dominated by the term $\exp(-\Theta(n \ln n))$. \square

The bound is the best possible. For suppose H is the 4-edge path: then

$$\mathbb{P}(H \not\subseteq_i G_n) \geq \mathbb{P}(\sigma \text{ has one part}) \geq 1/B_n = e^{-O(n \log n)}.$$

Theorem 4.5. *Let H be any (fixed) graph, and let $P_n \in_u \mathcal{P}_n$. Then*

$$\mathbb{P}(H \subseteq_i P_n) = \begin{cases} e^{-\Omega(n)} & \text{if } H \notin \mathcal{GS}, \\ 1/2 \pm e^{-\Omega(n)} & \text{if } H \in \mathcal{GS} \setminus (\mathcal{GS}^+ \cap \mathcal{GS}^-), \\ 1 - e^{-\Omega(n)} & \text{if } H \in (\mathcal{GS}^+ \cap \mathcal{GS}^-). \end{cases}$$

Proof. Consider $Q_n = \{G \in \mathcal{G}_n : H \subseteq_i G\}$ and $\mathbb{P}_{\mathcal{P}_n}(Q_n)$. Now the result follows from Lemma 4.1, equation (7) and Theorem 2.3. \square

Note that Theorem 4.5 accords with (14). It also follows from Theorem 4.5 that for every graph H :

$$\mathbb{P}\left(\left((H \subseteq_i P_n) \cup (\overline{H} \subseteq_i P_n)\right)\right) = \begin{cases} e^{-\Omega(n)} & \text{if } H \notin \mathcal{GS}, \\ 1 - e^{-\Omega(n)} & \text{if } H \in \mathcal{GS}. \end{cases}$$

and

$$\mathbb{P}\left(\left((H \subseteq_i P_n) \cap (\overline{H} \subseteq_i P_n)\right)\right) = \begin{cases} e^{-\Omega(n)} & \text{if } H \notin \mathcal{GS}^+ \cap \mathcal{GS}^-, \\ 1 - e^{-\Omega(n)} & \text{if } H \in \mathcal{GS}^+ \cap \mathcal{GS}^-. \end{cases}$$

5 Clique Colouring

A j -clique colouring of a graph is a colouring of the vertices with j colours so that no maximal clique is monochromatic (ignoring any isolated vertices). We improve Corollary 6 of [BGG⁺04], which states that almost all perfect graphs are 3-clique-colourable.

Theorem 5.1. *Almost all perfect graphs are 2-clique-colourable.*

Corollary 6 of [BGG⁺04] is proved by showing that all generalised split graphs are 3-clique-colourable, and then using the theorem of Prömel and Steger discussed earlier. In the same article it is shown that there are generalised split graphs which are not 2-clique-colourable. There are several other subclasses of perfect graphs for which the clique-chromatic number is known to be at most 3, and it was conjectured in [DSSW91] that perfect graphs have bounded clique-chromatic number. This was recently disproved in [CPTT16]. See [MMP16] for recent work on clique-colourings of binomial random graphs and of geometric graphs. We prove Theorem 5.1 using Theorem 2.3. We consider unipolar and co-unipolar graphs separately.

5.1 Clique colouring of unipolar graphs

We start with a deterministic lemma. Let us say that a vertex *sees* a set of vertices if it is adjacent to some vertex in the set.

Lemma 5.2. *Let $n \geq 2$ and let $G \in \mathcal{GS}_n^+$, with a central clique C . Suppose that some vertex $x \in C$ sees each side clique Q which is a maximal clique in G . Then G is 2-clique-colourable.*

Proof. Choose a vertex $y_Q \in N(x) \cap Q$ for each side clique Q with $N(x) \cap Q \neq \emptyset$. Define a colouring $c : V(G) \rightarrow \{1, 2\}$ in the following way:

$$c(v) = \begin{cases} 1 & \text{if } v = x, \\ 2 & \text{if } v \in C - x, \\ 2 & \text{if } v = y_Q \text{ for some side clique } Q, \\ 1 & \text{otherwise.} \end{cases}$$

We claim that c is a proper clique colouring. Note first that $N(x) \neq \emptyset$: this is trivial if $|C| \geq 2$; and if $C = \{x\}$ then there is a side clique Q , and Q must contain a neighbour of x (whether it is a maximal clique or not).

Suppose for a contradiction that M is a monochromatic maximal clique. Since M is a clique, it must lie in the co-bipartite graph induced by C and a side clique Q . Further $M \not\subseteq C$: for if so then $M = C$, and C is not monochromatic if $|C| \geq 2$ and C is not a maximal clique if $C = \{x\}$.

If $N(x) \cap Q$ is empty then $M \subseteq Q$ and Q is not a maximal clique, so M is not a maximal clique, a contradiction. Thus we may assume that $N(x) \cap Q$ is nonempty, and so contains y_Q . Then either (1) $M \subseteq \{x\} \cup (Q \setminus \{y_Q\})$ or (2) $M \subseteq (C \setminus \{x\}) \cup \{y_Q\}$. In case (1), we could add y_Q to M , and in case (2) we could add x . Thus again M is not maximal, and we have our final contradiction. \square

Let $t = t(n) = \log n - 2 \log \log n$ (t is for threshold). Given a unipolar arrangement (G, C) of order n , we use t to split the side cliques of (G, C) into two categories - the *big* side cliques with at least t vertices, and the *small* side

cliques with less than t vertices. We shall see in the next two lemmas that, for $G^+ \sim \text{Gen}^+(n)$, whp the big side cliques can all be seen by some $x \in C$ and the small side cliques are not maximal, so we can apply the above deterministic lemma to deduce that whp G^+ is 2-clique-colourable.

Lemma 5.3. *Let $G^+ \sim \text{Gen}^+(n)$, with induced unipolar arrangement (G, C) . Then with probability $1 - 2^{-\Omega(\log^2 n)}$ there is a vertex $x \in C$ which sees each big side clique.*

Proof. Condition on $k = k_0 \in \mathcal{I} = [n/2 - \sqrt{n}, n/2 + \sqrt{n}]$, as usual. Let \mathcal{B} be the set of all big side cliques, so $|\mathcal{B}| \leq (n - k_0)/t$. Condition also on $|\mathcal{B}|$.

For $v \in C$ and $Q \in \mathcal{B}$, the probability that v does not see Q is

$$2^{-|Q|} \leq 2^{-t} = \frac{\log^2 n}{n}.$$

Thus the probability that v sees each $Q \in \mathcal{B}$ is at least

$$\begin{aligned} \left(1 - \frac{\log^2 n}{n}\right)^{\frac{n-k_0}{t}} &= \exp\left(-\frac{\log^2 n}{n} \frac{n-k_0}{t} + O\left(\frac{\log^3 n}{n}\right)\right) \\ &= \exp\left(-\left(\frac{1}{2} + o(1)\right) \log n\right) = n^{-\alpha+o(1)} \end{aligned}$$

where $\alpha = 1/(2 \log e) \approx 0.35$. Then, the probability that for each vertex $v \in C$ there is some $Q \in \mathcal{B}$ which v fails to see is at most

$$(1 - n^{-\alpha+o(1)})^{k_0} \leq e^{-n^{\frac{1}{2}-\alpha+o(1)}},$$

which is easily at most $2^{-\Omega(\log^2 n)}$. \square

Lemma 5.4. *Let $G^+ \sim \text{Gen}^+(n)$, with induced unipolar arrangement (G, C) . Then with probability $1 - 2^{-\Omega(\log^2 n)}$ no small side clique is a maximal clique.*

Proof. Condition on $k = k_0 \in \mathcal{I} = [n/2 - \sqrt{n}, n/2 + \sqrt{n}]$, as usual. Let us say that a vertex $v \in C$ extends a side clique Q if $Q \subseteq N(v)$. For $v \in C$ and $Q \in \mathcal{B}$, the probability that v does not extend Q is

$$1 - 2^{-|Q|} \leq 1 - 2^{-t} = 1 - \frac{\log^2 n}{n} \leq e^{-\frac{\log^2 n}{n}}.$$

Thus the probability that no $v \in C$ extends Q is at most

$$\left(e^{-\frac{\log^2 n}{n}}\right)^{k_0} = e^{-(\frac{1}{2}+o(1)) \log^2 n}.$$

But there are at most n small side cliques, so the same upper bound holds for the probability that some small side clique is a maximal clique. \square

Lemma 5.5. *$G^+ \sim \text{Gen}^+(n)$ is 2-clique-colourable with probability $1 - 2^{-\Omega(\log^2 n)}$.*

Proof. This follows from the last three lemmas. \square

5.2 Clique colouring of co-unipolar and perfect graphs

Co-unipolar graphs are easier to clique-colour than unipolar graphs.

Lemma 5.6. *The random graph $G^- \sim \text{Gen}^-(n)$ is 2-clique-colourable wvhp.*

Proof. We first prove an auxiliary claim.

Claim 5.7. *Let G be a co-unipolar graph, with a unipolar arrangement (\overline{G}, C) for \overline{G} . Suppose that C is non-empty, and every vertex of C has neighbours in at least two side independent sets of \overline{C} . Then G is 2-clique-colourable.*

Proof. Since C is non-empty, there are at least two side independent sets in \overline{C} . Let Q be a side independent set. Colour the vertices in $C \cup Q$ with 1 and the remaining vertices with 2. A maximal clique contained completely in \overline{C} contains a vertex from each side independent set and therefore is not monochromatic. There are no maximal cliques contained completely in C . A monochromatic clique which intersects both C and \overline{C} must contain precisely two vertices – $c \in C$ and $q \in Q$. However, such a clique can always be extended with a neighbour of c from a different side independent set. \square

Let us check that G^- satisfies the condition of the claim wvhp. By Theorem 2.5 we may assume $k \in \mathcal{I} = [n/2 - \sqrt{n}, n/2 + \sqrt{n}]$. Let (G^-, C) be the co-unipolar arrangement induced by the generation. By Lemma 2.9 wvhp the maximum size of an independent set of \overline{C} is at most $n/\sqrt{\ln n} = o(n)$. On the other hand every vertex of C has degree at least $n/5$ wvhp from Chernoff's inequality and the union bound. If the minimum degree of the vertices in C is bigger than the maximum size of a side independent set \overline{C} , then every vertex of C must see at least two side independent sets. \square

Theorem 5.1 now follows from equation (7), Lemma 5.5, Lemma 5.6 and Theorem 2.3. Indeed we see that the probability that P_n fails to be 2-clique-colourable is $2^{-\Omega(\log^2 n)}$.

6 Hamilton cycles in the random perfect graph P_n

A *Hamilton* cycle in a graph is a cycle visiting every vertex exactly once. Recall the distribution $L(n)$ in Definition 2.1. In this section we prove the following theorem.

Theorem 6.1. *Almost all perfect graphs are Hamiltonian. Indeed, for $P_n \in_u \mathcal{P}_n$ and $X \sim L(n)$,*

$$\mathbb{P}(P_n \text{ is Hamiltonian}) = 1 - \frac{1}{2} \mathbb{P}(X > n/2) \pm e^{-\Omega(n)}.$$

It follows from this result and Theorem 2.5 that

$$\mathbb{P}(P_n \text{ is Hamiltonian}) = 1 - 2^{-(\frac{1}{4} + o(1)) \log^2 n} = 1 - o(1),$$

which will complete the proof of (1). To prove Theorem 6.1 we consider unipolar and co-unipolar graphs separately, and in order to handle co-unipolar graphs we consider random bipartite graphs.

6.1 Hamilton cycles in random unipolar graphs

Lemma 6.2. *Suppose that (G, C) is a unipolar arrangement. Let $T_1, T_2 \subseteq \overline{C}$ each contain exactly one vertex from each side clique and be such that $T_1 \cap T_2$ consists of the vertices of \overline{C} forming side cliques of size one. Let A_1 and A_2 be disjoint subsets of C . Suppose that in G there are complete matchings M_1 from T_1 into A_1 and M_2 from T_2 into A_2 . Then G is Hamiltonian.*

Proof. Fix some ordering of the side cliques of \overline{C} . There is a Hamilton cycle which enters each clique from an edge of M_1 , visits all vertices inside and then leaves the clique from an edge of M_2 . After all side cliques are visited, the Hamilton cycle visits any remaining vertices in C and finishes at the initial vertex. \square

Definition 6.3. *Let $G_{n,m,\frac{1}{2}}$ denote a random $n \times m$ bipartite graph with the edges present independently with probability $1/2$.*

Lemma 6.4. *Suppose that $1 \leq n \leq m$. Then $G = (V_1, V_2, E) \sim G_{n,m,\frac{1}{2}}$ contains a complete matching from V_1 to V_2 with probability $1 - 2^{-m(1+o(1))}$.*

Proof. We can find an isolated vertex in V_1 with probability at least 2^{-m} , hence 2^{-m} is a lower bound on the probability of failing to have a complete matching (from V_1 to V_2). Hall's marriage theorem states that G contains a complete matching if and only if $|\Gamma(S)| \geq |S|$ for every $S \subseteq V_1$. Let E_S be the event that $|\Gamma(S)| < |S|$. We have $\mathbb{P}(E_S) \leq \binom{m}{m-s+1} 2^{-s(m-s+1)}$, where $s = |S|$. Thus

$$\begin{aligned} \mathbb{P}(\cup_S E_S) &\leq \sum_S \mathbb{P}(E_S) = \sum_{\substack{|S| \leq 2 \text{ or} \\ |S| \geq m-1}} \mathbb{P}(E_S) + \sum_{2 < |S| < m-1} \mathbb{P}(E_S) \\ &\leq (mn)^2 2^{-m} + 2^{n+m} 2^{-3(m-3)} = 2^{-m(1+o(1))}. \end{aligned} \quad \square$$

Lemma 6.5. *The random graph $G^+ \sim \text{Gen}^+(n)$ is Hamiltonian wvhp.*

Proof. Condition on $|k - n/2| < \sqrt{n}$, which happens with wvhp by Theorem 2.5. By Lemma 2.8, the number of side cliques is at most $n/16$ wvhp. Condition on this event and on the arrangement (G^+, C) induced by the generation process. Partition C into two near equal parts $A, B \subseteq C$ with $||A| - |B|| \leq 1$, so that $|A| \sim |B| \sim n/4$. Find $T_A, T_B \subset \overline{C}$ as in Lemma 6.2. By Lemma 6.4, we can find complete matchings between T_A and A and between T_B and B wvhp. We now see that G^+ satisfies the conditions of Lemma 6.2 wvhp and therefore G^+ is Hamiltonian wvhp. \square

6.2 Hamilton cycles in random bipartite graphs

Hamilton cycles in random bipartite graphs have been studied in Frieze [Fri85] and in Bollobás and Kohayakawa [BK91]. However, we are interested in the dense case and seek an exponentially small failure probability, which does not appear to have been done before. We shall show:

Theorem 6.6. *If $G \sim G_{n,n,\frac{1}{2}}$ then G is Hamiltonian with probability $1 - 2^{-(1+o(1))n}$.*

This theorem will follow from the deterministic Theorem 6.9 below, based on ‘Posa flips’, but first we need some preparation.

Definition 6.7. *Suppose G is a bipartite graph. We define $\tilde{\alpha}(G)$ to be the maximum integer k such that we can find k -sets $S_1 \subseteq V_1$ and $S_2 \subseteq V_2$ with $E(S_1, S_2) = \emptyset$ for a bipartition (V_1, V_2) of G . We call the pair (S_1, S_2) a bipartite hole of G .*

We denote the minimum degree of a graph G by $\delta(G)$.

Lemma 6.8. *Suppose G is a bipartite graph such that $\delta(G) > \tilde{\alpha}(G)$. Then G is $(\delta(G) - \tilde{\alpha}(G))$ -connected.*

Proof. Let $S \subseteq V(G)$ be a set of less than $\delta(G) - \tilde{\alpha}(G)$ vertices. We must show that $G' = G[V(G) \setminus S]$ is connected. Let (V_1, V_2) be a bipartition of G .

Let $v_1 \in V_1 \setminus S$ and $v_2 \in V_2 \setminus S$. Then $|N(v_1) \setminus S| \geq \tilde{\alpha}(G) + 1$ and $|N(v_2) \setminus S| \geq \tilde{\alpha}(G) + 1$; and hence either v_1 and v_2 are adjacent, or a neighbour of v_1 is adjacent to a neighbour of v_2 in G' . Thus v_1 and v_2 are connected in G' by a path of length at most 3. If $v_1, v_2 \in V_1 \setminus S$, then v_1 has a neighbour $u \in V_2 \setminus S$, and we have already seen that v_2 and u must be connected in G' , so v_1 and v_2 are connected in G' . The same conclusion holds if $v_1, v_2 \in V_2 \setminus S$. \square

Theorem 6.9. *An $n \times n$ bipartite graph G with $n \geq 2$ is Hamiltonian if $\delta(G) \geq 2\tilde{\alpha}(G) + 1$.*

Proof. All paths in this proof have their terminal vertices labelled as *start* and *end*. Suppose $P = \{v_1, \dots, v_l\}$ is a path with start v_1 and end v_l , and suppose v_l is adjacent to v_i . A *flip of P around v_i* is the path $\{v_1, \dots, v_i, v_l, \dots, v_{i+1}\}$ over $V(P)$, starting at v_1 and ending at v_{i+1} . (If $i = l - 1$ the flip does nothing.)

Let $P = \{v_1, \dots, v_l\}$ be a path starting at v_1 , ending at v_l , and having maximum length of any path in G . For $k = 0, 1, 2$ let W_k be the set of ends of paths obtained from P by at most k flips. Clearly $W_0 = \{v_l\}$. Since P is maximal, v_l is adjacent to vertices of P only, hence $W_1 = \{v_{i+1} : v_i v_l \in E(G)\}$ satisfies $|W_1| \geq \delta(G)$.

Assume that $\delta(G) \geq 2\tilde{\alpha}(G) + 1$, and let (V_1, V_2) be a bipartition of G . Without loss of generality suppose that $v_l \in V_1$. Thus each set $W_k \subseteq V_1$. To complete the proof it is sufficient to show that every vertex of V_2 is adjacent to a vertex of W_2 . Indeed, suppose this is the case. If $v_1 \in V_2$, then P can be closed to a cycle, which must be Hamiltonian since G is connected by Lemma 6.8 and P

has maximum length. If $v_1 \in V_1$, then there is a vertex $u \in V_2 \setminus V(P)$. But this contradicts the maximality of P , because there is a path Q with $V(Q) = V(P)$ ending at a neighbour of u obtained from P by two flips, and Q can be extended to contain u , yielding a longer path.

We claim that

$$|W_2| \geq n - 2\tilde{\alpha}(G). \quad (15)$$

Assuming this, and recalling that $\delta(G) \geq 2\tilde{\alpha}(G) + 1$, we see that every vertex of V_2 must be adjacent to a vertex of W_2 , and we are done. So it remains to establish the claim (15).

Suppose $v_{i+1} \in W_1$ is adjacent to $v_j \neq v_i$. Then

$$v_{j+1} \in W_2 \quad \text{if } j < i; \quad (16)$$

$$v_{j-1} \in W_2 \quad \text{if } j > i. \quad (17)$$

More precisely, the path obtained from P by first flipping around v_i and then v_j ends with v_{j+1} or v_{j-1} depending on whether v_j comes before v_i in P .

Since $|W_1| \geq 2\tilde{\alpha}(G) + 1$, we can find an integer t such that $v_{t+1} \in W_1$ and if we write

$$F_1 = \{v_i \in V(G) : i > t + 1\} \text{ and} \\ F_2 = \{v_i \in V(G) : i < t + 1\},$$

then $|F_1 \cap W_1| \geq \tilde{\alpha}(G)$ and $|F_2 \cap W_1| \geq \tilde{\alpha}(G)$. Let $M_1 = (F_1 \cap W_1) \cup \{v_{t+1}\}$ and $M_2 = (F_2 \cap W_1) \cup \{v_{t+1}\}$. It follows from (17) that if $v_j \in F_1$ has a neighbour in M_2 , then its predecessor, v_{j-1} , is contained in W_2 ; and from (16) it follows that if $v_j \in F_2$, $v_j \neq v_t$ has a neighbour in M_1 , then its successor, v_{j+1} , is contained in W_2 . But $v_{t+1} \in W_1$, and hence $v_{t+1} \in W_2$, so no special care is required for the case $v_j = v_t$. For $k = 1, 2$ let $N_k = \Gamma[M_{3-k}] \cap F_k$ and $N_k^c = (F_k \cap V_2) \setminus N_k$ (and hence $F_k \cap V_2 = N_k \cup N_k^c$); and let N_1^- be the set of predecessors in P of vertices in N_1 , and let N_2^+ be the set of successors of N_2 . We see that

1. $N_1^- \cup N_2^+ \subseteq W_2 \setminus \{v_t\}$;
2. $|N_1^-| = |N_1|$, $|N_2^+| = |N_2|$;
3. $N_1^- \cap N_2^+ = \{v_{t+1}\}$;
4. $v_t \in W_2$.

Since $F_k \cap V_2 = N_k \cup N_k^c$ and $v_{t+1} \in V_1$, we observe that $n = |V_2| = |N_1| + |N_1^c| + |N_2| + |N_2^c|$, and therefore

$$|W_2| \geq |N_1^- \cup N_2^+| + 1 = (|N_1^-| + |N_2^+| - 1) + 1 \\ = |N_1| + |N_2| = n - (|N_1^c| + |N_2^c|).$$

However, $E(M_{3-k}, N_k^c) = \emptyset$ (by the definition of N_k) and $|M_k| > \tilde{\alpha}(G)$, so $|N_k^c| \leq \tilde{\alpha}(G)$, and hence $|W_2| \geq n - 2\tilde{\alpha}(G)$. Thus we have proved (15), and the proof is complete. \square

Theorem 6.6 now follows because if $G \sim G_{n,n,\frac{1}{2}}$ and $t = \sqrt{n}$, then the events $\{\delta(G) < 2t + 1\}$ and $\{\tilde{\alpha}(G) > t\}$ both have probability $2^{-(1+o(1))n}$. As a lower bound consider the event that G contains an isolated vertex. For similar results, concerning non-bipartite graphs, we refer to [MY15]. The paper [MY15] explains how Theorem 6.6 could be extended to state that we can find n^δ , $0 < \delta < \frac{1}{2}$, edge-disjoint Hamilton cycles with probability $1 - 2^{-(1+o(1))n}$.

6.3 Hamilton cycles in random co-unipolar graphs

Lemma 6.10. *Suppose that $G^- \sim \text{Gen}^-(n)$ and $X \sim L(n)$. Then G^- is Hamiltonian with probability $\mathbb{P}(L(n) \leq n/2) - e^{-\Omega(n)}$.*

Proof. Recall that $\text{Gen}^-(n)$ is a random quadruple $(-1, E, (k, \sigma), \pi)$. If $k > n/2$, then $\alpha(G^-) > n/2$, which makes it impossible to contain a Hamilton cycle. We show that conditional on $k \leq n/2$, G^- is Hamiltonian wvhp, which will complete the proof.

From Theorem 2.5 and Lemma 2.8 we may condition on the event that $k \geq n/2 - \sqrt{n}$ and the event that the number of parts in σ is at least $n/(3 \ln n)$ (since both hold wvhp). Condition also on (G^-, C) being the co-unipolar arrangement induced by the generation. Partition \overline{C} as $Q \cup T$, so that $|C| = |Q| = k$, $|T| = n - 2k \leq 2\sqrt{n}$, and T contains vertices only from different parts of σ . By Corollary 6.6, there is a Hamilton cycle in the induced subgraph $G^-[C \cup Q]$ wvhp. Condition on this event and on H being a Hamiltonian cycle. If T is empty we are done, so assume not.

Now we have a Hamilton cycle H in $G^-[C \cup Q]$, and a non-empty set T of vertices inducing a clique. Let u, v be vertices in T , and note that there is a u - v Hamilton path in $G^-[T]$ (where $u = v$ if $|T| = 1$). We may assume that the maximum size of a side stable set is at most $n/\log n$ by Lemma 2.9. Let S_u, S_v be the side stable sets containing u, v respectively, and let R be the set of vertices in $Q \setminus S_u$. Then u is adjacent to each vertex in R , and $|R| \geq k - |S_u| \geq n/3 + |S_v|$ for n sufficiently large.

The vertices in R have at least $|R| - |S_v| \geq n/3$ neighbours along H which are not in S_v , and if v is adjacent to any of these neighbours then we have a Hamilton cycle in the whole graph. But the probability that this fails is at most $2^{-n/3}$, since any neighbour in \overline{C} must be adjacent to v and so we may assume there are none. \square

7 Connectivity and Chromatic index

7.1 Connectivity

A set of vertices in a graph G is a *cutset* if removing these vertices leaves a disconnected graph. The (*vertex*) *connectivity* of G is the minimum size of a cutset, except that, if G is complete, then by convention $\kappa(G) = |V(G)| - 1$.

We use $\delta(G)$ to denote the minimum vertex degree. Observe that $\delta(G)$ is natural upper bound for $\kappa(G)$. We prove the following theorem:

Theorem 7.1. *The connectivity of almost all perfect graphs equals their minimum degree. Indeed, wvhp $\kappa(P_n) = \delta(P_n)$, where $P_n \in_u \mathcal{P}_n$.*

Proof. We first consider unipolar graphs. Let (G^+, C) be the unipolar arrangement induced by the generation of $Gen^+(n)$. Then, for each fixed ϵ with $0 < \epsilon < \frac{1}{20}$, G^+ satisfies each of the following properties wvhp:

- (a1) $\forall u \in \overline{C}, d_C(u) \leq (1 + \epsilon)\frac{1}{4}n.$
- (a2) $\forall u \in \overline{C}, d_{\overline{C}}(u) \leq \epsilon n.$
- (a3) $\forall u, v \in \overline{C}, |(N(v) \cup N(u)) \cap C| \geq (1 - \epsilon)\frac{3}{8}n.$
- (a4) $|C| > (1 - \epsilon)\frac{n}{2}.$

Chernoff's inequality is sufficient to prove (a1) and (a3); for (a2) we can use Lemma 2.9, and (a4) follows from the concentration theorem, Theorem 2.5.

These conditions are sufficient to ensure $\kappa(G^+) = \delta(G^+)$. Note first that if $\overline{C} = \emptyset$ then G^+ is a clique and the result holds; so we may assume that $\overline{C} \neq \emptyset$. Suppose for contradiction that $Q \subset V = V(G^+)$ is a separator of G^+ with $|Q| < \delta(G^+)$. Note that $\delta(G^+) \leq (1 + 5\epsilon)\frac{n}{4} \leq |C|$, and so $C \not\subseteq Q$. We claim that every vertex of $V \setminus Q$ is connected to a vertex in $C \setminus Q$, which will yield our contradiction. This is obvious for vertices $u \in C \setminus Q$. If $u \in \overline{C}$, then since $|Q| < \delta(G^+)$, u must have a neighbour v outside Q . If $v \in C \setminus Q$ we are done, so suppose that $v \in N(u) \cap \overline{C}$. From (a3), u and v see together at least $(1 - \epsilon)\frac{3}{8}n$ vertices of C , while $|Q|$ is at most $(1 + 5\epsilon)\frac{n}{4}$, so there must be a vertex from $C \setminus Q$ connected to either u or v .

The situation is similar for co-unipolar graphs. Let (G^-, C) be the co-unipolar arrangement induced by the generation of $Gen^-(n)$. Then, for each fixed ϵ with $0 < \epsilon < \frac{1}{20}$, G^- satisfies the following conditions wvhp.

- (b1) $\forall u \in C, d(u) \leq (1 + \epsilon)\frac{1}{4}n.$
- (b2) Each side independent set in \overline{C} has size at most $\epsilon n.$
- (b3) $|\overline{C}| > (1 - \epsilon)\frac{n}{2}.$

To prove these, we can use a Chernoff bound for (b1), Lemma 2.9 for (b2), and Theorem 2.5 for (b3).

These conditions are sufficient to ensure $\kappa(G^-) = \delta(G^-)$. Note first that if $C = \emptyset$ then G^- is a complete multipartite graph and the result holds; so we may assume that $C \neq \emptyset$. Thus $\delta(G^-) \leq (1 + \epsilon)\frac{1}{4}n$. Suppose that $Q \subset V = V(G^-)$ is a cutset of G^- and $|Q| < \delta(G^-)$. Fix some vertex $w \in V \setminus Q$. We claim that every vertex of $V \setminus Q$ is connected to w .

First consider $u \in \overline{C}$. Then w and u are not neighbours (if and) only if they are contained in the same side independent set I . But then there is a vertex $x \in \overline{C} \setminus Q$ in a different side independent set, since $|\overline{C} \setminus Q| > (1 - 3\epsilon)\frac{n}{4} \geq |I|$; and the path u, x, w connects u to w . Now consider $u \in C$. Since $|Q| < \delta(G^-)$, u must

have a neighbour in \overline{C} , and we just seen that all vertices in \overline{C} are connected to w .

We have now shown that $\mathbb{P}_{Gen(n)}(\kappa(G) = \delta(G)) = 1 - e^{-\Omega(n)}$, and Theorem 2.3 completes the proof. \square

7.2 Chromatic Index

The *chromatic index* of a graph G , $\chi'(G)$, is the minimum number of colours required to colour the edges, so that no edges with the same colour share a vertex. Vizing proved in [Viz64] that $\chi'(G) = \Delta(G)$ or $\chi'(G) = \Delta(G) + 1$. If $\chi'(G) = \Delta(G)$, then G is said to be class one, otherwise it is said to be class two.

In this subsection we prove

Theorem 7.2. *Almost all perfect graphs are class one graphs.*

Class one graphs have a natural connection with perfect graphs. Denote the line graph of a graph G by $L(G)$. Then, for graphs G with $\Delta(G) \geq 3$, $\Delta(G) = \omega(L(G))$ and $\chi'(G) = \chi(L(G))$; and hence G is a class one graph iff $\omega(L(G)) = \chi(L(G))$. Vizing showed in [Viz65] that each graph of class two has at least 3 vertices of maximum degree (and indeed the vertices of maximum degree induce a subgraph with a cycle). Erdős and Wilson proved in [EW77] that almost all graphs have a unique vertex of maximum degree, and thus are class one, see also [FJMR88]. We shall show that almost all perfect graphs have a unique vertex of maximum degree, and thus are class one.

Let $d_j = d_j(G)$ be the j th largest degree in the list of all $v(G)$ degrees. Theorem 15 in Chapter III (The degree sequence) of the book [Bol01] concerns the first few gaps $d_i - d_{i+1}$ for a random graph $G_{n,p}$. The next lemma is a much easier result that may be proved along similar lines; we do not spell a proof here.

Lemma 7.3. *Whp $B := G_{n,m,1/2}$ with $m/n \rightarrow 1$ is such that*

$$d_1(B) - d_2(B) \geq \frac{n^{1/2}}{\log n}.$$

where $d_1(B)$ and $d_2(B)$ are the largest and second largest degrees of vertices in the first colour class.

Let (G^+, C) be the unipolar arrangement induced by the generation of $Gen^+(n)$, and let B denote the corresponding bipartite graph with parts C and \overline{C} . With high probability, $d(u) \geq (1 - \epsilon)\frac{3}{4}n$ for all u in C , and $d(v) \leq (1 + \epsilon)\frac{1}{4}n$ for all v in \overline{C} . Hence, by Lemma 7.3, whp

$$d_1(G^+) - d_2(G^+) = d_1(B) - d_2(B) \geq \frac{n^{1/2}}{2 \log n},$$

and so whp G^+ has a unique vertex of maximum degree.

Let (G^-, C) be the co-unipolar arrangement induced by the generation of $Gen^-(n)$, and let B denote the corresponding bipartite graph with parts \overline{C} and C (in this order). With high probability, $d(u) \geq (1 - \epsilon)\frac{3}{4}n$ for all u in \overline{C} , and $d(v) \leq (1 + \epsilon)\frac{1}{4}n$ for all v in C . Also, by Lemma 2.9, with high probability $|d_{\overline{C}}(u) - d_{\overline{C}}(v)| < 10 \ln n$ for each $u, v \in \overline{C}$, and therefore

$$\begin{aligned} d_1(G^+) - d_2(G^+) &\geq d_1(B) - d_2(B) - 10 \ln n \\ &\geq (1 - \epsilon) \frac{n^{1/2}}{2 \log n} - 10 \ln n, \end{aligned}$$

and so whp G^- has a unique vertex of maximum degree.

We have now seen that whp both G^+ and G^- have a unique vertex of maximum degree, which implies they are class one; Theorem 7.2 now follows from Theorem 2.3.

8 The limit of a sequence of random perfect graphs

We show that a sequence of uniformly and independently sampled n -vertex perfect graphs converges with probability one to the graphon

$$W_P(x, y) = \frac{\mathbf{1}[x \leq 1/2] + \mathbf{1}[y \leq 1/2]}{2}$$

and show how to use this for estimates on the subgraph densities.

8.1 Notation

We use the notation for left limits from [Lov12]. Suppose throughout (as usual) that P_n is a uniformly sampled perfect graph on vertex set $[n]$. Let λ be the Lebesgue measure on \mathbb{R}^2 . A *kernel* $W : [0, 1]^2 \rightarrow \mathbb{R}$ is an a.e. bounded symmetric measurable function. A *graphon* W is a kernel with $0 \leq W \leq 1$ a.s. The *cut norm*, $\|\cdot\|_{\square}$, is a norm on the vector space of kernels defined by

$$\|W\|_{\square} = \sup_{S, T \subseteq [0, 1]} \left| \int_{S \times T} W d\lambda \right|.$$

The supremum is taken over all measurable $S, T \subseteq [0, 1]$. Let $S_{[0, 1]}$ denote the set of invertible measure preserving maps on $[0, 1]$. Define $W^{\varphi}(x, y) := W(\varphi(x), \varphi(y))$ for $\varphi \in S_{[0, 1]}$. The *cut distance*, δ_{\square} , is defined by

$$\delta_{\square}(U, W) = \inf_{\varphi \in S_{[0, 1]}} \|U - W^{\varphi}\|_{\square}$$

for kernels U, W . For a graph G we define the graphon W_G by partitioning $[0, 1] = S_1 \sqcup \dots \sqcup S_{v(G)}$, $\lambda(S_i) = 1/v(G)$, and writing $W(x, y) = \mathbf{1}_{ij \in E(G)}$

if $(x, y) \in S_i \times S_j$. Let $\text{hom}(F, G)$ and $\text{hom}_{\text{inj}}(F, G)$ be the number of homomorphisms (edge preserving maps), injective homomorphisms respectively, from graph F to graph G . Let $t(F, G) = \frac{\text{hom}(F, G)}{n^k}$ and let $t_{\text{inj}}(F, G) = \frac{\text{hom}_{\text{inj}}(F, G)}{n(n-1)\dots(n-k+1)}$, where $k = v(F)$ and $n = v(G)$. For a graph F and kernel W define

$$t(F, W) = \int_{[0,1]^{v(F)}} \prod_{ij \in E(G)} W(x_i, x_j) \prod_{i \in V(G)} dx_i.$$

We have $t(F, G) = t(F, W_G^\varphi)$ for each pair of graphs F and G and each $\varphi \in S_{[0,1]}$.

8.2 Convergence results

Theorem 8.1. *We have*

$$\delta_{\square}(W_{P_n}, W_P) \leq n^{-1/2} \text{ with probability } 1 - e^{-\Omega(\sqrt{n} \log n)}, \quad (18)$$

$$\delta_{\square}(W_{P_n}, W_P) \leq (\log n)^{-2} \text{ with probability } 1 - e^{-\Omega(n)}. \quad (19)$$

Before we prove Theorem 8.1 we discuss a few corollaries. Suppose each P_n is sampled independently.

Corollary 8.2. *With probability one we have $P_n \rightarrow W_P$ with respect to δ_{\square} .*

Proof. This follows from Theorem 8.1 (either part), a Borel-Cantelli lemma, and Theorem 11.22 of [Lov12]. \square

Corollary 8.3. *For a fixed graph F*

$$|t_{\text{inj}}(F, P_n) - t(F, W_P)| \leq \binom{v(F)}{2} n^{-1} + e(F) n^{-1/2}$$

with probability at least $1 - e^{-\Omega(\sqrt{n} \log n)}$.

Proof. It is easy to see that for every graph G we have $|t_{\text{inj}}(F, G) - t(F, G)| \leq \binom{v(F)}{2} \frac{1}{v(G)}$. Theorem 10.23 in [Lov12] states that $|t(F, W_1) - t(F, W_2)| \leq e(F) \cdot \delta_{\square}(W_1, W_2)$ for every two graphons W_1 and W_2 . The rest follows from part 1 of Theorem 8.1. \square

The notation in the second corollary may be misleading, in the sense that the statement has nothing to do with graph limits. Indeed, the value $t(F, W_P)$ can be calculated explicitly and then used as an estimate on the number of subgraphs F of P_n . For example, if F is the single edge K_2 then $t(F, W_P) = \frac{1}{2}$ so the random perfect graph P_n contains $(1 + O(n^{-1/2})) \frac{n^2}{4}$ edges with probability at least $1 - e^{-\Omega(\sqrt{n} \log n)}$; and similarly P_n contains $(1 + O(n^{-1/2})) \frac{5n^3}{96}$ triangles with probability at least $1 - e^{-\Omega(\sqrt{n} \log n)}$. Using an inclusion-exclusion argument we can estimate the densities of induced graphs and in particular Corollary 8.3 implies that the split graphs are the only graphs with induced subgraph density bounded away from zero.

We can be more precise for example concerning the number $e(P_n)$ of edges of P_n . Note that the expected value is exactly $\frac{1}{2}\binom{n}{2}$. Also, whp $e(G_n^+) = \frac{1}{4}n^2 + (\frac{1}{4} + o(1))n \ln n$ and so whp $e(G_n^-) = \frac{1}{4}n^2 - (\frac{1}{4} + o(1))n \ln n$; and hence whp $e(P_n) = \frac{1}{4}n^2 + O(n \log n)$, with a bimodal distribution.

8.3 Proof of Theorem 8.1

We consider the unipolar case $G^+ \sim \text{Gen}^+(n)$ and the co-unipolar case $G^- \sim \text{Gen}^-(n)$ separately, and start with the former. Suppose $\text{Gen}^+(n) = (1, E, (k, \sigma), \pi)$. Let $\varphi \in S_{[0,1]}$ be a measure preserving map, mapping each vertex of G^+ to an interval of the form $V_i = [\frac{i-1}{n}, \frac{i}{n})$ where $i \in [n]$, and in addition let φ map the central clique of W_{G^+} to $[0, k/n)$ and the side cliques to $[k/n, 1)$. Pick arbitrary measurable $S, T \subseteq [0, 1]$. To prove that $\delta_{\square}(G^+, W_P) \leq \epsilon$ it is sufficient to show that

$$\left| \int_{S \times T} (W_{G^+}^{\varphi} - W_P) \right| \leq \epsilon.$$

Claim 8.4. *We can find $S', T' \subseteq [0, 1]$ of the form $S' = \cup_{i \in I_s} V_i$ and $T' = \cup_{i \in I_t} V_i$ such that*

$$\left| \int_{S \times T} (W_{G^+}^{\varphi} - W_P) \right| \leq \left| \int_{S' \times T'} (W_{G^+}^{\varphi} - W_P) \right|.$$

Proof. Initially set $S' = S$ and $T' = T$. For $i = 1, \dots, n$ in turn, check if $\int_{T'} (W_{G^+}^{\varphi} - W_P)(\frac{i-1}{n}, y) dy$ has the same sign as $\int_{S' \times T'} W_{G^+}^{\varphi} - W_P d\lambda$, and if so, add the entire interval V_i to S' , or remove V_i from S' otherwise. These operations can only increase the absolute value of the integral. Now repeat the same procedure for T' . It is clear that in the end we obtain S' and T' with the desired properties. \square

Every such pair (S', T') is measurable. We may assume that initially we are given S and T of this form.

Let $A_1 = [0, 1/2)$, $B_1 = [1/2, 1)$, $A_2 = [0, k/n)$, $B_2 = [k/n, 1)$, $A = A_1 \cap A_2$, $B = B_1 \cap B_2$ and $I = A \cup B$. By Theorem 2.5, wvhp both $\lambda(A)$ and $\lambda(B)$ are contained in $[\frac{1}{2} - \frac{n^{-1/2}}{24}, \frac{1}{2}]$. Condition on this event, and now we have

$$\left| \int_{S \times T} (W_{G^+}^{\varphi} - W_P) \right| \leq \left| \int_{(S \times T) \cap (I \times I)} (W_{G^+}^{\varphi} - W_P) \right| + \frac{n^{-1/2}}{12}.$$

Let $O_1 = (S \times T) \cap (A \times A)$, $O_2 = (S \times T) \cap (A \times B)$, $O_3 = (S \times T) \cap (B \times A)$, $O_4 = (S \times T) \cap (B \times B)$, and $w_i = \left| \int_{O_i} (W_{G^+}^{\varphi} - W_P) \right|$.

We have $W_{G^+}^{\varphi} = W_P = 1$ over O_1 , so clearly $w_1 = 0$. The restriction of $W_{G^+}^{\varphi}$ over O_2 is a step function corresponding to a random bipartite graph with density $1/2$, while W_P is uniformly $1/2$. This classical example motivated the study of the cut distance, but we give full details here.

Claim 8.5. *Suppose M is a random $\{-1, 1\}$ -valued $m \times m$ matrix, where each entry has expected value 0 and is independent from the others. Then the maximum absolute value of a rectangular sum in M is at most $\frac{5}{3}m^{3/2}$ wvhp.*

Proof. Let $R \subseteq [m]$ be a subset of the rows of M , and let $Q \subseteq [m]$ be a subset of the columns of M . Clearly $M_{Q,R} := \sum_{r \in R, q \in Q} M_{r,q} \sim 2X - |Q||R|$, where $X \sim \text{Bin}(|Q||R|, 1/2)$; so by Chernoff's inequality

$$\mathbb{P}(|M_{Q,R}| \geq x) = \mathbb{P}\left(\left|X - \frac{|Q||R|}{2}\right| \geq \frac{x}{2}\right) \leq 2e^{-\frac{x^2}{2m^2}}.$$

Taking $x = \frac{5}{3}m^{3/2}$, and the union bound over all 2^{2m} possible choices for R, Q completes the proof. \square

Setting $m = \lceil n/2 \rceil$ and rescaling in Claim 8.5 shows that w_2 and w_3 are bounded by $\frac{5}{12}n^{-1/2}$ wvhp.

Finally, w_4 is bounded by twice the number of edges in the partition part of G^+ divided by n^2 . Lemma 2.9 states that the maximum size of a part in a random partition is at most x with probability at least $1 - (1 + o(1))e^{-x(\ln x - \ln \ln n - 2) + \ln n}$. The maximum size of a part bounds the maximum degree. Here we have to make a compromise between low cut distance and high probability. To favour the former, use Lemma 2.9 with $x = \frac{1}{12}\sqrt{n}$ and deduce that

$$\left| \int_{S \times T} (W_{G^+}^\varphi - W_P) \right| \leq \frac{1}{\sqrt{n}}$$

with probability $1 - e^{-\Omega(\sqrt{n} \log n)}$.

Now let us aim for a bound that holds wvhp. Let $\log n \leq x \leq n/2$. Observe that if x vertices are contained in parts of σ each of size at least s , then x vertices are contained in at most $\lceil x/s \rceil$ parts. Thus, assuming $s \leq x$ and so $\lceil x/s \rceil \leq 2x/s$, the probability that some set of x points is contained in parts of σ each of size at least s is at most

$$\begin{aligned} & \binom{n}{x} \left(\frac{\lceil x \rceil}{s}\right)^x n \frac{B_{n-x}}{B_n} \\ & \leq \left(\frac{en}{x} \cdot \frac{2x}{s}\right)^x (1 + o(1)) \left(\frac{r_n}{r_{n-x}}\right)^{\frac{1}{2}} n \exp\{-x(r_n - 1 + 1/r_n)\} \\ & = \left(\frac{2en}{s}\right)^x \exp\{-x(\ln n - (1 + o(1)) \log \log n)\} \\ & = \exp\{-x(\ln s - (1 + o(1)) \log \log n)\}. \end{aligned}$$

The factor n in the first bound above arises since the 'last part' for the x -set may need to be amalgamated with some part for the $(n-x)$ -set. Set $x = \lceil n/(3 \log n) \rceil$ and $s = n/\log^2 n$, to see that wvhp at most $n/(3 \log n)$ points are contained in parts of σ each of size at least $n/\log^2 n$. Since also wvhp no part has size

greater than $n/\log n$ by (10), we see that wvhp the number of edges in $G^+[\bar{S}]$ is at most

$$\left(\frac{1}{2} + o(1)\right)n \cdot \frac{n}{\log^2 n} + \frac{n}{3 \log n} \frac{n}{\log n}$$

so wvhp $w_4 \leq \frac{5}{6}(\log n)^{-2}$. Hence wvhp $\delta_{\square}(G^+, W_P) \leq (\log n)^{-2}$.

Let J be the graphon with $J(x, y) = 1$ for each $x, y \in [0, 1]$; and note that $J - W_P$ may be written as W_P^{ψ} for some $\psi \in S_{[0,1]}$. Then

$$\int_{S \times T} (W_{G^-}^{\varphi} - W_{P^{\psi}}) = - \int_{S \times T} (W_{G^+}^{\varphi} - W_P).$$

Hence, the bounds for G^+ transfer to G^- . Theorem 2.3 completes the proof.

Using (11) we may see that the bound in (19) is best possible, in the sense that, if $\epsilon(n) = o(1)$, then $\delta_{\square}(W_{P_n}, W_P) \geq \epsilon(n)(\log n)^{-2}$ holds with probability $e^{-o(n)}$.

9 Proofs for the generation model

We conclude this paper with the proofs of Theorems 2.3 and 2.5, which were deferred to here.

9.1 Proof of Theorem 2.3

We shall prove (5) and (6) and thus complete the proof of Theorem 2.3. We denote the \mathcal{L}_p norm of X in the probability space $(\mathcal{G}_n, 2^{\mathcal{G}_n}, \mu)$ by $\|X\|_p^{\mu}$. Recall that $\|X\|_p^{\mu}$ is non-decreasing in p . Let ν and μ be discrete probability measures on (all subsets of) \mathcal{G}_n such that $\nu(G) > 0$ implies $\mu(G) > 0$. In this case ν is said to be *absolutely continuous* with respect to μ , written $\nu \ll \mu$. The Radon–Nikodym derivative $\frac{d\nu}{d\mu}$ is given by $\frac{d\nu}{d\mu}(G) = \frac{\nu(G)}{\mu(G)} \mathbf{1}_{\mu(G) > 0}$. We may express the total variation distance between ν and μ in terms of $\frac{d\nu}{d\mu}$ and the \mathcal{L}_1 norm $\|\cdot\|_1^{\mu}$: we have

$$2d_{TV}(\nu, \mu) = \sum_G |\nu(G) - \mu(G)| = \sum_G \left| \frac{d\nu}{d\mu}(G) - 1 \right| \mu(G) = \left\| \frac{d\nu}{d\mu} - 1 \right\|_1^{\mu}.$$

(The same hold for general, not necessarily discrete, probability measures, provided that $\nu \ll \mu$.) Also note that for each G

$$\frac{d\mathbb{P}_{Gen(n)}}{d\mathbb{P}_{\mathcal{P}_n}}(G) = \mathbb{P}(\rho(Gen(n)) = G) | \mathcal{P}_n|$$

The main result in this section is stronger than Theorem 2.3.

Theorem 9.1. *For each $p \geq 1$ we have*

$$2d_{TV}(\mathbb{P}_{Gen(n)}, \mathbb{P}_{\mathcal{P}_n}) = \left\| \frac{d\mathbb{P}_{Gen(n)}}{d\mathbb{P}_{\mathcal{P}_n}} - 1 \right\|_1^{\mathbb{P}_{\mathcal{P}_n}} \leq \left\| \frac{d\mathbb{P}_{Gen(n)}}{d\mathbb{P}_{\mathcal{P}_n}} - 1 \right\|_p^{\mathbb{P}_{\mathcal{P}_n}} = e^{-\Theta(n)}.$$

We have just noted the first two (in)equalities in the statement of the theorem: the non-trivial part is the final equality. To prove it we start from the triangle inequality

$$\left\| \frac{d\mathbb{P}_{Gen(n)}}{d\mathbb{P}_{\mathcal{P}_n}} - 1 \right\|_p^{\mathbb{P}_{\mathcal{P}_n}} \leq \left\| \frac{d\mathbb{P}_{Gen(n)}}{d\mathbb{P}_{\mathcal{P}_n}} - \frac{d\mathbb{P}_{\mathcal{G}\mathcal{S}_n}}{d\mathbb{P}_{\mathcal{P}_n}} \right\|_p^{\mathbb{P}_{\mathcal{P}_n}} + \left\| \frac{d\mathbb{P}_{\mathcal{G}\mathcal{S}_n}}{d\mathbb{P}_{\mathcal{P}_n}} - 1 \right\|_p^{\mathbb{P}_{\mathcal{P}_n}}.$$

It is straightforward to show from (3) that the second term on the right is $e^{-\Omega(n)}$. For the first term on the right, routine manipulations yield

$$\left\| \frac{d\mathbb{P}_{Gen(n)}}{d\mathbb{P}_{\mathcal{P}_n}} - \frac{d\mathbb{P}_{\mathcal{G}\mathcal{S}_n}}{d\mathbb{P}_{\mathcal{P}_n}} \right\|_p^{\mathbb{P}_{\mathcal{P}_n}} = \left\| \frac{d\mathbb{P}_{Gen(n)}}{d\mathbb{P}_{\mathcal{G}\mathcal{S}_n}} - 1 \right\|_p^{\mathbb{P}_{\mathcal{G}\mathcal{S}_n}} \left(\frac{|\mathcal{P}_n|}{|\mathcal{G}\mathcal{S}_n|} \right)^{1-1/p}.$$

Since $\frac{|\mathcal{P}_n|}{|\mathcal{G}\mathcal{S}_n|} = 1 + e^{-\Omega(n)}$ by (3), it follows that it is sufficient to show that

$$\left\| \frac{d\mathbb{P}_{Gen(n)}}{d\mathbb{P}_{\mathcal{G}\mathcal{S}_n}} - 1 \right\|_p^{\mathbb{P}_{\mathcal{G}\mathcal{S}_n}} = e^{-\Omega(n)}$$
 to complete the proof of Theorem 9.1.

We need a few definitions to continue.

Definition 9.2. Let $\mathcal{CG}\mathcal{S}_n^+$ (from coloured generalised split graphs) be the set of all unipolar arrangements of order n and let $\mathcal{CG}\mathcal{S}_n = \mathcal{CG}\mathcal{S}_n^+ \times \{-1, 1\}$.

Recall that \mathcal{L}_n was introduced just before Definition 2.1. We see that $|\mathcal{CG}\mathcal{S}_n^+| = \mathcal{L}_n$ and $|\mathcal{CG}\mathcal{S}_n| = 2\mathcal{L}_n$. A closer look at the definitions reveals that if $((G, C), B) \in_u \mathcal{CG}\mathcal{S}_n$ is uniformly selected, if we set $H_1 = G$ if $B = 1$ and $H_1 = \bar{G}$ otherwise, and set $H_2 \sim \rho(Gen(n))$, then H_1 and H_2 are equal in distribution. Therefore

$$\mathbb{P}_{Gen(n)}(G) = \frac{1}{|\mathcal{CG}\mathcal{S}_n|} \sum_{C \subseteq [n]} \{ \mathbb{1}[(G, C) \in \mathcal{CG}\mathcal{S}_n^+] + \mathbb{1}[(\bar{G}, C) \in \mathcal{CG}\mathcal{S}_n^+] \}.$$

To simplify the notation we define $R_+(G) = \sum_{C \subseteq [n]} \mathbb{1}[(G, C) \in \mathcal{CG}\mathcal{S}_n^+]$, $R_-(G) = R_+(\bar{G})$ and $R(G) = R_+(G) + R_-(G)$, so that $\sum_G R(G) = |\mathcal{CG}\mathcal{S}_n|$ and

$$\mathbb{P}_{Gen(n)}(G) = \frac{R(G)}{|\mathcal{CG}\mathcal{S}_n|}.$$

Note that $R(G) > 0$ iff $G \in \mathcal{G}\mathcal{S}_n$. Now we see that (for each G)

$$\frac{d\mathbb{P}_{Gen(n)}}{d\mathbb{P}_{\mathcal{G}\mathcal{S}_n}} = \frac{R}{\frac{1}{|\mathcal{G}\mathcal{S}_n|}} = R \frac{|\mathcal{G}\mathcal{S}_n|}{|\mathcal{CG}\mathcal{S}_n|}.$$

It is clear from the definitions that $\|R - 1\|_1^{\mathbb{P}_{\mathcal{G}\mathcal{S}_n}} = \frac{|\mathcal{CG}\mathcal{S}_n|}{|\mathcal{G}\mathcal{S}_n|} - 1$. We have

$$\begin{aligned} \left\| \frac{d\mathbb{P}_{Gen(n)}}{d\mathbb{P}_{\mathcal{G}\mathcal{S}_n}} - 1 \right\|_p^{\mathbb{P}_{\mathcal{G}\mathcal{S}_n}} &= \left\| R \frac{|\mathcal{G}\mathcal{S}_n|}{|\mathcal{CG}\mathcal{S}_n|} - 1 \right\|_p^{\mathbb{P}_{\mathcal{G}\mathcal{S}_n}} = \left\| R - \frac{|\mathcal{CG}\mathcal{S}_n|}{|\mathcal{G}\mathcal{S}_n|} \right\|_p^{\mathbb{P}_{\mathcal{G}\mathcal{S}_n}} \frac{|\mathcal{G}\mathcal{S}_n|}{|\mathcal{CG}\mathcal{S}_n|} \\ &\leq \left\| R - 1 + 1 - \frac{|\mathcal{CG}\mathcal{S}_n|}{|\mathcal{G}\mathcal{S}_n|} \right\|_p^{\mathbb{P}_{\mathcal{G}\mathcal{S}_n}} \leq \|R - 1\|_p^{\mathbb{P}_{\mathcal{G}\mathcal{S}_n}} + \left| 1 - \frac{|\mathcal{CG}\mathcal{S}_n|}{|\mathcal{G}\mathcal{S}_n|} \right| \\ &= \|R - 1\|_p^{\mathbb{P}_{\mathcal{G}\mathcal{S}_n}} + \|R - 1\|_1^{\mathbb{P}_{\mathcal{G}\mathcal{S}_n}} \leq 2 \|R - 1\|_p^{\mathbb{P}_{\mathcal{G}\mathcal{S}_n}}. \end{aligned}$$

Further simplifications lead us to

$$\begin{aligned}
\|R - 1\|_p^{\mathbb{P}^{\mathcal{GS}_n}} &= \|(R_+ - \mathbb{1}_{\mathcal{GS}_n^+}) + (R_- - \mathbb{1}_{\mathcal{GS}_n^-}) + (\mathbb{1}_{\mathcal{GS}_n^+} + \mathbb{1}_{\mathcal{GS}_n^-} - 1)\|_p^{\mathbb{P}^{\mathcal{GS}_n}} \\
&\leq 2\|R_+ - \mathbb{1}_{\mathcal{GS}_n^+}\|_p^{\mathbb{P}^{\mathcal{GS}_n}} + \|\mathbb{1}_{\mathcal{GS}_n^+} \mathbb{1}_{\mathcal{GS}_n^-}\|_p^{\mathbb{P}^{\mathcal{GS}_n}} \\
&= 2\|R_+ - 1\|_p^{\mathbb{P}^{\mathcal{GS}_n^+}} \frac{|\mathcal{GS}_n^+|}{|\mathcal{GS}_n|} + \|\mathbb{1}_{\mathcal{GS}_n^+} \mathbb{1}_{\mathcal{GS}_n^-}\|_p^{\mathbb{P}^{\mathcal{GS}_n}} \\
&\leq 2\|R_+ - 1\|_p^{\mathbb{P}^{\mathcal{GS}_n^+}} + \|\mathbb{1}_{\mathcal{GS}_n^+} \mathbb{1}_{\mathcal{GS}_n^-}\|_1^{\mathbb{P}^{\mathcal{GS}_n}}. \tag{20}
\end{aligned}$$

The only combinatorial argument we need to complete this proof is phrased as a lemma with a proof in the next section:

Lemma 9.3. *For every integral $j \geq 1$ we have*

$$\|R_+^j\|_1^{\mathbb{P}^{\mathcal{GS}_n^+}} \leq \|R_+\|_1^{\mathbb{P}^{\mathcal{GS}_n^+}} (1 + 2^{-n/2+o(n)}).$$

From now on we drop the superscript $\mathbb{P}^{\mathcal{GS}_n^+}$ for legibility. By combining the lemma for $j = 2$ with Jensen's inequality we get

$$\|R_+\|_1^2 \leq \|R_+^2\|_1 \leq \|R_+\|_1 (1 + e^{-\Omega(n)}),$$

and so $\|R_+\|_1 \leq 1 + e^{-\Omega(n)}$. Hence, by the last lemma, we have $\|R_+^j\|_1 \leq 1 + e^{-\Omega(n)}$. Now $1 \leq R_+^j \leq R_+^{j+1}$ a.s., and therefore

$$\|R_+^{j+1} - R_+^j\|_1 = \|R_+^{j+1}\|_1 - \|R_+^j\|_1 = e^{-\Omega(n)}.$$

Finally, we see that for each $j \geq 1$

$$\begin{aligned}
\|(R_+ - 1)^{j+1}\|_1 &= \left\| (R_+ - 1) \sum_{i=0}^j \binom{j}{i} R_+^i (-1)^{j-i} \right\|_1 \leq \sum_{i=0}^j \binom{j}{i} \|(R_+ - 1) R_+^i\|_1 \\
&\leq 2^j \|(R_+ - 1) R_+^j\|_1 \leq e^{-\Omega(n)}.
\end{aligned}$$

Therefore

$$\|R_+ - 1\|_p^{\mathbb{P}^{\mathcal{GS}_n^+}} = e^{-\Omega(n)} \tag{21}$$

for each real $p \geq 1$.

Thus we have filled the missing bits in the proof of Lemma 3.6 and now we can use it. We have $\mathbb{P}_{\mathcal{GS}_n^+}(\alpha(G) \geq \omega(G)) = e^{-\Omega(n)}$ and $\mathbb{P}_{\mathcal{GS}_n^-}(\alpha(G) \leq \omega(G)) = e^{-\Omega(n)}$. Partition the set of graphs which are both unipolar and co-unipolar into $U_1 = \{G \in \mathcal{GS}_n^+ \cap \mathcal{GS}_n^- : \alpha(G) > \omega(G)\}$ and $U_2 = \{G \in \mathcal{GS}_n^+ \cap \mathcal{GS}_n^- : \alpha(G) \leq \omega(G)\}$. We see from Lemma 3.6 that U_1 consists of rare unipolar graphs, and that U_2 consists of rare co-unipolar graphs, or more formally, $|U_i| = e^{-\Omega(n)} |\mathcal{GS}_n|$, hence $|\mathcal{GS}_n^+ \cap \mathcal{GS}_n^-| / |\mathcal{GS}_n^+| = e^{-\Omega(n)}$, and therefore

$$\|\mathbb{1}_{\mathcal{GS}_n^+} \mathbb{1}_{\mathcal{GS}_n^-}\|_1^{\mathbb{P}^{\mathcal{GS}_n}} = e^{-\Omega(n)}. \tag{22}$$

Now (20), (21) and (22) imply that $\|R - 1\|_p^{\mathbb{P}^{\mathcal{GS}_n}} = e^{-\Omega(n)}$, which completes the proof of Theorem 9.1.

9.2 Proving Lemma 9.3

Fix $j \geq 1$. Suppose throughout this section that $G \in \mathcal{GS}_n^+$, $C_1, \dots, C_j \subseteq V(G)$ and (G, C_i) is a unipolar arrangement of order n for each $i \in [j]$. (For this given G , the number of choices for C_1, \dots, C_j is $R^+(G)^j$, which we want to upper bound.) Further, let $C = \cap_{i=1}^j C_i$, $\tilde{C} = \cap_{i=1}^j \overline{C_i}$, and let $l = n - |C| - |\tilde{C}|$. The following claim is easy to verify.

Claim 9.4. *The pair $(G[C \cup \tilde{C}], C)$ is a unipolar arrangement of order $|C \cup \tilde{C}|$.*

Now let $C'_i = C_i \setminus C$ and $Q = V(G) \setminus (C \cup \tilde{C})$, so that $Q = \cup_{i=1}^j C'_i$.

Lemma 9.5. *Let $i \in [j]$. Then the pair $(G[Q], C'_i)$ is a unipolar arrangement of order l with at most $j - 1$ side cliques.*

Proof. Without loss of generality suppose that $i = 1$. It is easy to see that $G[Q]$ is a generalised split graph with a central clique C'_1 , since $C'_1 \subseteq C_1$, $Q \setminus C'_1 \subseteq \tilde{C}_1$ and (G, C_1) is a unipolar arrangement. Further, $Q = \cup_{i=1}^j C'_i$, so every vertex in $Q \setminus C'_1$ is covered by a clique C'_i , where $2 \leq i \leq j$. Hence $G[Q \setminus C'_1]$ is a disjoint union of cliques, covered by $j - 1$ cliques, so the unipolar arrangement $(G[Q], C'_1)$ contains at most $j - 1$ side cliques. \square

Since each C'_i is complete to C and $Q = \cup_{i=1}^j C'_i$, it follows that Q is complete to C .

We now focus on C'_1 . Observe that, for each vertex $v \in C'_1$, since $v \notin C$ we can find C'_i such that $v \notin C'_i$. Thus C'_1 can be expressed as a disjoint union $C''_2 \cup \dots \cup C''_j$ where $C''_i \cap C'_i = \emptyset$ for each $i = 2, \dots, j$. Let \mathcal{C} denote the family of sets $\{C'_i \setminus C'_1\}_{i=2}^j \cup \{C''_i\}_{i=2}^j$. Note that the union of these sets is Q .

Lemma 9.6. *For each $S \in \mathcal{C}$, either S is complete to a side clique of $(G[C \cup \tilde{C}], C)$ and non-adjacent to all other side cliques, or S is non-adjacent to \tilde{C} .*

Proof. In the unipolar arrangement (G, C_1) , each $C'_i \setminus C'_1$ is contained in a side clique, so it is either complete to a side clique of $(G[C \cup \tilde{C}], C)$, and non-adjacent to the others, or non-adjacent to any vertex of \tilde{C} . Similarly, in the unipolar arrangement (G, C_i) , C''_i is contained in a side clique, hence it has the same property. \square

Let f be the function mapping each set S in \mathcal{C} to the side clique S' of $(G[C \cup \tilde{C}], C)$ such that S is complete to S' , or to \emptyset if no vertex in S is adjacent to \tilde{C} .

Lemma 9.7. *The edge set $E(Q, \tilde{C})$ can be reconstructed from f .*

Proof. For each $v \in Q$ either $v \in C'_1$ and then v lies in some set $S \in \{C''_i\}_{i=2}^j$, or v lies in some set $S \in \{C'_i \setminus C'_1\}_{i=2}^j$. In either case, S is in \mathcal{C} and the set of neighbours of v in \tilde{C} is precisely the set $f(S)$. \square

The edges of G can be reconstructed from $G[C \cup \tilde{C}]$ – a unipolar graph, $G[Q]$ – a unipolar graph with at most $k - 1$ side cliques and $E(Q, C \cup \tilde{C})$, which in turn can be reconstructed from f , since Q is complete to C . It will turn out that the number of choices for $G[C \cup \tilde{C}]$ dominates the rest, and this number is maximised whenever $Q = \emptyset$. However, the vertex labels are a nuisance, and additional information is required to correctly recover G . The precise statement is given below.

Lemma 9.8. *The $j+1$ tuple (G, C_1, \dots, C_j) can be uniquely reconstructed from:*

1. *the pair $(G[C \cup \tilde{C}], C)$ seen as a unipolar $(n - l)$ -arrangement over the vertex set $[n - l]$ so that the order of the labels of the vertices is preserved;*
2. *the pair $(G[Q], C'_1)$ seen as a unipolar l -arrangement over the vertex set $[l]$ with at most $j - 1$ side cliques so that the order of the labels of the vertices is preserved;*
3. *the sets $C'_2, \dots, C'_j \subseteq [l]$;*
4. *the l -subset of $V(G)$ specifying the original labels of Q ;*
5. *the function f .*

Observe that there are at most $j^l 2^{\frac{l^2}{4}}$ choices for the pair $(G[Q], C'_1)$ above.

Recall that $\mathcal{L}_n = |\mathcal{CGS}_n^+|$. Let $T_n = 2^{\frac{n^2}{4} + \frac{n}{2} \log n - \frac{n}{2} \log \ln n - \frac{n}{2} \log \frac{e}{2}}$ for $n \geq 3$, and let $T_n = 1$ for $n = 0, 1, 2$. Prömel and Steger show in [PS92] that $\mathcal{L}_n \leq T_n \times 2^{O(\frac{n}{\ln \ln(n+3)})}$, hence for some c we have $|\mathcal{CGS}_n^+| \leq T_n \times 2^{\frac{cn}{\ln \ln(n+3)}}$. It follows that

$$U_{n,l} := \left(T_{n-l} \times 2^{\frac{cn}{\ln \ln(n+3)}} \right) \times \left(j^l 2^{\frac{l^2}{4}} \right) \times \left(2^{(j-1)l} \right) \times \binom{n}{l} \times \left((n+1)^{2(j-1)} \right)$$

is an upper bound on the number of choices for (G, C_1, \dots, C_j) when $|Q| = l$, and $U_n = \sum_{l=1}^n U_{n,l}$ is an upper bound on the number of choices for (G, C_1, \dots, C_j) when $Q \neq \emptyset$.

Lemma 9.9. *The number $U_{n,l}$ is maximised subject to $1 \leq l \leq n$ when $l = 1$.*

Proof. We may prove this by thinking of l as a continuous variable and taking the derivative of $U_{n,l}$ with respect to l . The analysis is straightforward but unpleasant: we leave the details to the reader. \square

Corollary 9.10. *We have*

$$U_n = |\mathcal{CGS}_n^+| 2^{-n/2+o(n)}.$$

Proof. Since $T_{n-1} = T_n 2^{-n/2+o(n)}$, we see that $\frac{U_{n,1}}{|\mathcal{CGS}_n^+|} = 2^{-n/2+o(n)}$, and hence

$$|\mathcal{CGS}_n^+| 2^{-n/2+o(n)} = U_{n,1} \leq U_n \leq nU_{n,1} = |\mathcal{CGS}_n^+| 2^{-n/2+o(n)}. \quad \square$$

Finally, the number of choices for the $j+1$ tuple (G, C_1, \dots, C_j) when $Q = \emptyset$ is clearly $|\mathcal{CGS}_n^+|$, so we have

$$\begin{aligned} \|R_+^j\|_1^{\mathbb{P}^{\mathcal{GS}_n^+}} &= \sum_{G \in \mathcal{GS}_n^+} R_+^j(G) \frac{1}{|\mathcal{GS}_n^+|} \leq (U_n + |\mathcal{CGS}_n^+|) \frac{1}{|\mathcal{GS}_n^+|} \\ &= |\mathcal{CGS}_n^+| (1 + 2^{-n/2+o(n)}) \frac{1}{|\mathcal{GS}_n^+|} = \|R_+\|_1^{\mathbb{P}^{\mathcal{GS}_n^+}} (1 + 2^{-n/2+o(n)}). \end{aligned}$$

This completes the proof of Lemma 9.3, Theorem 9.1 and Theorem 2.3.

9.3 Concentration: Proof of Theorem 2.5

Let $\mu = (n - \log n + \log \ln n)/2$ as in the statement of the theorem (for $n \geq 3$), let $\hat{\mu} = \lceil \mu \rceil$ and let $\check{\mu} = \hat{\mu} - 1$. As before, let $\ell_{n,k} = \binom{n}{k} 2^{k(n-k)} B_{n-k}$. We fix n and drop it in the subscript for legibility, i.e. write ℓ_k instead of $\ell_{n,k}$.

Lemma 9.11. *Using the notation above, there is an n_0 such that $\ell_k/\ell_{\hat{\mu}} \leq n^{-n-1}$ as long as $|k - \hat{\mu}| \geq n^{2/3}$ for $n \geq n_0$.*

Proof. Let $x = k - \hat{\mu}$, we have

$$\frac{\ell_k}{\ell_{\hat{\mu}}} = \frac{\binom{n}{k} 2^{k(n-k)} B_{n-k}}{\binom{n}{\hat{\mu}} 2^{\hat{\mu}(n-\hat{\mu})} B_{n-\hat{\mu}}} = \frac{(n - \hat{\mu})!}{(n - \hat{\mu} - x)!} \left(\frac{(\hat{\mu} + x)!}{\hat{\mu}!} \right)^{-1} 2^{x(n-2\hat{\mu})} 2^{-x^2} \frac{B_{n-\hat{\mu}-x}}{B_{n-\hat{\mu}}}.$$

We can bound each of the terms $\frac{(n-\hat{\mu})!}{(n-\hat{\mu}-x)!}$, $\left(\frac{(\hat{\mu}+x)!}{\hat{\mu}!} \right)^{-1}$, $2^{x(n-2\hat{\mu})}$ and $\frac{B_{n-\hat{\mu}-x}}{B_{n-\hat{\mu}}}$ with $n^{|x|}$, so we get

$$\frac{\ell_k}{\ell_{\hat{\mu}}} < n^{|x|} n^{|x|} 2^{-|x|^2} n^{|x|} = 2^{3|x|} \log n^{-|x|^2} \leq 2^{3n^{2/3}} \log n^{-n^{4/3}} \leq n^{-n-1}$$

for large n . □

The bound is rather crude, and it can be improved, but we do not need a better bound for our purposes. From now we focus on $k = \hat{\mu} + x$ with $|x| \leq n^{2/3}$. This implies $k \sim n/2$, which allows us to use asymptotics for k as n grows, for instance $(n-k)/(k+1) \rightarrow 1$ and more importantly

$$\frac{B_{n-k-1}}{B_{n-k}} = (1 + o_n(1)) \frac{\ln(n-k)}{n-k} = (1 + o_n(1)) \frac{2 \ln n}{n},$$

where the error term is uniform over k , provided that $|x| \leq n^{2/3}$.

Lemma 9.12. *There is an n_0 such that the following holds for each $n \geq n_0$. Suppose $0 \leq x \leq n^{2/3}$ is integral and $r \in \left\{ \frac{\ell_{\hat{\mu}-x}}{\ell_{\hat{\mu}}}, \frac{\ell_{\hat{\mu}+x}}{\ell_{\hat{\mu}}} \right\}$. Then $2^{-x^2-2x} \leq r \leq 2^{-x^2+2x}$.*

Proof. Define the continuous, strictly decreasing function $d(y) = 2^{n-2y} \frac{\ln n}{n}$, so that for $k \sim n/2$

$$\frac{\ell_{k+1}}{\ell_k} = \frac{n-k}{k+1} 2^{n-2k-1} (1 + o_n(1)) \frac{2 \ln n}{n} = (1 + o_n(1)) d(k).$$

Observe that $d(k) = 1 \iff 2k = n + \log \ln n - \log n \iff k = \mu$. We have $d(\mu + x) = d(\mu) \cdot 2^{-2x} = 2^{-2x}$, and

$$\ell_{\hat{\mu}+x} = \ell_{\hat{\mu}} \prod_{i=0}^{x-1} \frac{\ell_{\hat{\mu}+i+1}}{\ell_{\hat{\mu}+i}} = \ell_{\hat{\mu}} \prod_{i=0}^{x-1} (1 + o_n(1)) d(\hat{\mu} + i).$$

For each i we use the bounds

$$2^{-2i-2} = d(\mu + 1 + i) \leq d(\hat{\mu} + i) \leq d(\mu + i) = 2^{-2i},$$

while the error terms, $1 + o_n(1)$, we bound uniformly by $\frac{1}{2}$ and 2, provided $n \geq n_0$. By multiplying the terms we get $-2\binom{x}{2} - 3x$ and $-2\binom{x}{2} + x$ in the exponent for the lower and upper bounds respectively, which matches the statement of the lemma for the case $r = \ell_{\hat{\mu}+x}/\ell_{\hat{\mu}}$. The second case is analogous. \square

Claim 9.13. For $a > 1$ we have $\sum_{x=a}^{\infty} 2^{-x^2} < \frac{2^{-(a-1)^2}}{2(a-1) \ln 2}$.

Proof. Since $-x^2$ is decreasing, we have

$$\begin{aligned} \sum_{x=a+1}^{\infty} 2^{-x^2} &< \int_a^{\infty} 2^{-x^2} dx \leq \frac{1}{a} \int_a^{\infty} x 2^{-x^2} dx \\ &= \frac{1}{-2a \ln 2} \left(2^{-x^2} \Big|_{x=a}^{x=\infty} \right) = \frac{2^{-a^2}}{2a \ln 2}. \end{aligned} \quad \square$$

In the claim below and its proof, for simplicity we have written $n^{2/3}$ rather than $\lfloor n^{2/3} \rfloor$.

Claim 9.14. For each integer $k \geq 2$ we have

$$\sum_{i=k}^{n^{2/3}} \ell_{\hat{\mu}-i} \leq \ell_{\hat{\mu}} 2^{-(k-1)^2} \left(2 + \frac{1}{(k-1) \ln 2} \right) \quad \text{and} \quad \sum_{i=k}^{n^{2/3}} \ell_{\hat{\mu}+i} \leq \ell_{\hat{\mu}} 2^{-(k-1)^2} \left(2 + \frac{1}{(k-1) \ln 2} \right).$$

Proof. Using Lemma 9.12 we see that

$$\begin{aligned} \sum_{i=k}^{n^{2/3}} \ell_{\hat{\mu}-i} &\leq \ell_{\hat{\mu}} \sum_{i=k}^{n^{2/3}} 2^{-i^2+2i} = 2\ell_{\hat{\mu}} \sum_{i=k}^{n^{2/3}} 2^{-(i-1)^2} \leq 2\ell_{\hat{\mu}} \left(2^{-(k-1)^2} + \sum_{i=k}^{\infty} 2^{-i^2} \right) \\ &= 2\ell_{\hat{\mu}} \left(2^{-(k-1)^2} + \frac{2^{-(k-1)^2}}{2(k-1) \ln 2} \right). \end{aligned}$$

The last inequality follows from Claim 9.13. The second part of the statement is similar. \square

Now we put everything together to complete the proof of Theorem 2.5. Suppose $n \geq 3$, $X \sim L(n)$ and $x > 2$, then (noting the trivial bound $\mathcal{L}_n \geq \ell_{\hat{\mu}} + \ell_{\check{\mu}}$)

$$\begin{aligned}
\mathbb{P}(|X - \mu| \geq x) &\leq \mathbb{P}(\check{\mu} - n^{2/3} \leq X \leq \check{\mu} - x + 1) \\
&\quad + \mathbb{P}(\hat{\mu} + x - 1 \leq X \leq \hat{\mu} + n^{2/3}) \\
&\quad + \mathbb{P}(|X - \mu| \geq n^{2/3}) \\
&\leq \frac{\ell_{\hat{\mu}}}{\ell_{\hat{\mu}} + \ell_{\check{\mu}}} \cdot 2^{-(x-2)^2} \cdot 4 + \frac{\ell_{\check{\mu}}}{\ell_{\hat{\mu}} + \ell_{\check{\mu}}} \cdot 2^{-(x-2)^2} \cdot 4 + n^{-n} \\
&= 2^{-(x-2)^2+2} + n^{-n}, \text{ and} \\
\mathbb{P}(|X - \mu| \geq x) &\geq \mathbb{P}(X = \hat{\mu} + x) + \mathbb{P}(X = \check{\mu} - x) \\
&\geq \frac{2^{-x^2-2x}(\ell_{\hat{\mu}} + \ell_{\check{\mu}})}{(3 + \frac{1}{2 \ln 2})(\ell_{\hat{\mu}} + \ell_{\check{\mu}}) + n^{-n}\hat{\mu}} \\
&\geq 2^{-x^2-2x-2}.
\end{aligned}$$

10 Concluding remarks

In this paper we presented a generation model for perfect graphs which yields such graphs almost uniformly, with error $e^{-\Omega(n)}$ (in terms of total variation distance); and we used this approach to investigate several questions about such graphs.

We aimed to investigate the most natural questions about random perfect graphs, but of course there are further natural open problems. For example, concerning induced subgraphs H , we found the limiting probability that P_n has an induced subgraph H , but we did not discuss the distribution of the number of such induced subgraphs.

Our methods do not let us approach the following question about automorphisms. It is well known that almost all graphs G_n on vertex set $[n]$ have no non-trivial automorphisms, and easy to see that in fact the probability that G_n has a non-trivial automorphism is $2^{-(1+o(1))n}$. It is also not hard to see that, if $R_n \sim \text{Gen}(n)$, then the probability that R_n has a non-trivial automorphism is $2^{-(\frac{1}{2}+o(1))n}$. It is natural to expect that the probability for P_n is similar to that for R_n , but we do not know: we would need a finer result than (3). Similarly, if H is a fixed graph in $\mathcal{GS}^+ \cap \mathcal{GS}^-$ and $R_n \sim \text{Gen}(n)$, then by Lemma 4.1 that the probability that R_n fails to have an induced copy of H is $e^{-\Omega(n \log n)}$. Does such a result hold for P_n ?

A long standing open problem is to describe the structure of a random perfect graph with a prescribed number of edges, say $e(G) \sim v(G) \log v(G)$. This problem has been discussed before [PS92], and even directly approached [BTW12], but to the best of our knowledge a complete answer is a distant possibility.

Let C_5 -free denote the class of graphs with no induced subgraph C_5 . Then $\mathcal{GS} \subset \mathcal{P} \subset C_5$ -free, and the key result (2) from Theorem 2.4 of Prömel and

Steger [PS92] in fact extends as follows:

$$|\mathcal{GS}_n| = (1 - e^{-\Omega(n)})|(C_5\text{-free})_n|. \quad (23)$$

Thus our results refer also to random C_5 -free graphs. For related general results see [ABBM11, BB11] and references there.

References

- [ABBM11] N Alon, J Balogh, B Bollobás, and R Morris. The structure of almost all graphs in a hereditary property. *Journal of Combinatorial Theory, Series B*, 101(2):85–110, 2011.
- [BB11] J Balogh and J Butterfield. Excluding induced subgraphs: critical graphs. *Random Structures & Algorithms*, 38(1-2):100–120, 2011.
- [BGG⁺04] G Bacsó, S Gravier, A Gyárfás, M Preissmann, and A Sebo. Coloring the maximal cliques of graphs. *SIAM Journal on Discrete Mathematics*, 17(3):361–376, 2004.
- [BK91] B Bollobás and Y Kohayakawa. The hitting time of hamilton cycles in random bipartite graphs. *Graph Theory, Combinatorics, Algorithms, and Applications (Philadelphia)*(Y. Alavi, F. Chung, R. Graham, and D. Hsu, eds.), SIAM, pages 26–41, 1991.
- [Bol01] Béla Bollobás. *Random Graphs*, volume 73. Cambridge university press, 2001.
- [BTW12] J Böttcher, A Taraz, and A Würfl. Perfect graphs of fixed density: Counting and homogeneous sets. *Combinatorics Probability and Computing*, 21(5):661, 2012.
- [CPTT16] P Charbit, I Penev, S Thomassé, and N Trotignon. Perfect graphs of arbitrarily large clique-chromatic number. *Journal of Combinatorial Theory, Series B*, 116:456–464, 2016.
- [DB70] Nicolaas Govert De Bruijn. *Asymptotic methods in analysis*, volume 4. North-Holland Publishing, 1970.
- [DSSW91] D Duffus, B Sands, N Sauer, and R Woodrow. Two-colouring all two-element maximal antichains. *Journal of Combinatorial Theory, Series A*, 57(1):109–116, 1991.
- [EW77] P Erdős and R Wilson. On the chromatic index of almost all graphs. *Journal of Combinatorial Theory, Series B*, 23(2):255–257, 1977.
- [FJMR88] A Frieze, B Jackson, C McDiarmid, and B Reed. Edge-colouring random graphs. *Journal of Combinatorial Theory, Series B*, 45(2):135–149, 1988.

- [Fri85] Alan Frieze. Limit distribution for the existence of hamiltonian cycles in random bipartite graphs. *European Journal of Combinatorics*, 6:327 – 334, 1985.
- [Har66] L Harper. Stirling behavior is asymptotically normal. *Ann. Math. Stat.*, 38:410–414, 1966.
- [Lov72a] László Lovász. A characterization of perfect graphs. *Journal of Combinatorial Theory, Series B*, 13(2):95–98, 1972.
- [Lov72b] László Lovász. Normal hypergraphs and the perfect graph conjecture. *Discrete Mathematics*, 2(3):253–267, 1972.
- [Lov12] László Lovász. *Large networks and graph limits*, volume 60. American Mathematical Soc., 2012.
- [MMP16] C McDiarmid, D Mitsche, and P Pralat. Clique colourings of random graphs. 2016.
- [MY15] C McDiarmid and N YOLOV. Hamilton cycles in dense graphs. 2015.
- [Pit97] Boris Pittel. Random set partitions: Asymptotics of subset counts. *Journal of Combinatorial Theory, Series A*, 79(2):326 – 359, 1997.
- [PS92] H Prömel and A Steger. Almost all Berge graphs are perfect. *Combinatorics, Probability and Computing*, 1(01):53–79, 1992.
- [RAR01] J Ramírez-Alfonsín and B Reed. *Perfect graphs*, volume 44. Wiley, 2001.
- [Viz64] Vadim Vizing. On an estimate of the chromatic class of a p-graph. *Diskret. Analiz*, 3(7):25–30, 1964.
- [Viz65] Vadim Vizing. Critical graphs with given chromatic class. *Diskret. Analiz*, 5(9):17, 1965.