Topics in random graphs

Richard Montgomery

Lent 2018

Abstract

These notes accompanied a graduate course given in Cambridge in the Lent term of 2018. The course aimed to cover a range of different techniques relevant to random graphs. Any undefined notation is standard and can be found in, for example, [7, 13, 16]. Any comments or corrections are welcome.

Contents

1 Introduction

Consider graphs G which are finite and simple, with vertex set $V(G)$ and edge set $E(G)$, and write $|G| = |V(G)|$ and $e(G) = |E(G)|$. The study of random graphs began in full with the work of Erdős and Rényi [10]. We will study the following random graph model.

Definition. The *binomial random graph* or *Erdős-Rényi random graph*, $G = G(n,p)$, has $V(G) = [n] = \{1,\ldots,n\}$ and each possible edge is included independently at random with probability p.

Definition. A *graph property* $\mathcal{P}$ is a collection of graphs.

Example.

$\mathcal{P}_1 = \{G : e(G) > 0\}$ $\qquad$ $\mathcal{P}_4 = \{G : \chi(G) > 3\}$

$\mathcal{P}_2 = \{G : G \text{ contains a triangle}\}$ $\qquad$ $\mathcal{P}_5 = \{G : \chi(G) \neq 3\}$

$\mathcal{P}_3 = \{G : e(G) \text{ is odd}\}$

Definition. $\mathcal{P}$ is *non-trivial* if every $G \in \mathcal{P}$ has $e(G) > 0$ and, for sufficiently large n, $\exists G \in \mathcal{P}$ with $V(G) = [n]$.

Definition. $\mathcal{P}$ is *monotone* if $G \in \mathcal{P}$, $G \subset H$ and $V(H) = V(G) \implies H \in \mathcal{P}$. I.e., we cannot leave $\mathcal{P}$ by adding edges.

Example. $\mathcal{P}_1, \mathcal{P}_2, \mathcal{P}_4$ above are monotone, while $\mathcal{P}_3, \mathcal{P}_5$ are not.

Key meta-question. Is $G(n,p)$ typically in $\mathcal{P}$?

Definition. When $p = p(n)$, say $G(n,p) \in \mathcal{P}$ *almost surely* (a.s.) or *with high probability* (w.h.p.) if $\mathbb{P}(G(n,p) \in \mathcal{P}) \to 1$ as $n \to \infty$.

Asymptotic notation: For f, g non-zero functions of n, if $f(n)/g(n) \to \infty$ as $n \to \infty$, then we say $f = \omega(g)$ and $g = o(f)$. If $\exists C > 0$ such that, for all n, $|f(n)| \leq C|g(n)|$, then we say $f = O(g)$ and $g = \Omega(f)$. Where the implicit constant C depends on another variable, we will indicate this in the subscript.

2 Thresholds

Definition. A function $\hat{p}$ is a *threshold* for a property $\mathcal{P}$ if

- $p = o(\hat{p}) \implies G(n,p) \notin \mathcal{P}$ a.s., and
- $p = \omega(\hat{p}) \implies G(n,p) \in \mathcal{P}$ a.s.

Exercise. If $\hat{p}_0$, $\hat{p}_1$ are thresholds for $\mathcal{P}$ then $\exists c, C > 0$ such that $c\hat{p}_0 < \hat{p}_1 < C\hat{p}_0$.

Proposition 1. *The function $\hat{p} = 1/n^2$ is a threshold for $\mathcal{P}_1 = \{G : e(G) > 0\}$.*

Proof. If $p = \omega(\hat{p})$, then $\mathbb{P}(G(n,p) \notin \mathcal{P}) = (1-p)^{\binom{n}{2}} \leq e^{-p\binom{n}{2}} = o(1)$. On the other hand, if $p = o(\hat{p})$, then $\mathbb{P}(G(n,p) \in \mathcal{P}) \leq p\binom{n}{2} = o(1)$. □

The following lemma was first proved by Bollobás and Thomason [8].

Lemma 2. *All non-trivial properties have a threshold.*

Idea. We can couple probability models, i.e. 'reveal edges in batches'.

Proposition 3. *If $G_0 \sim G(n,p)$ and $G_1 \sim G(n,q)$ are independent, then $G_0 \cup G_1 \sim G(n, p + q - pq)$.*

Proof. The events $\{e \in E(G_0 \cup G_1)\}$, $e \in [n]^{(2)}$, are independent and, $\forall e \in [n]^{(2)}$,

$$\mathbb{P}(e \in E(G_0 \cup G_1)) = 1 - \mathbb{P}(e \notin E(G_0)) \cdot \mathbb{P}(e \notin E(G_1)) = p + q - pq. \qquad \square$$

Corollary 4. *If $\mathcal{P}$ is monotone, then $\forall p, q \in [0,1]$,*

$$\mathbb{P}(G(n, p+q) \notin \mathcal{P}) \leq \mathbb{P}(G(n,p) \notin \mathcal{P}) \cdot \mathbb{P}(G(n,q) \notin \mathcal{P}). \qquad \square$$

Proof of Lemma 2. For each (sufficiently large) n, let $\hat{p}(n)$ be such that $\mathbb{P}(G(n,\hat{p}) \in \mathcal{P}) = \frac{1}{2}$. Note that, as $G \in \mathcal{P} \implies e(G) > 0$, $\mathbb{P}(G(n,0) \in \mathcal{P}) = 0$. As there is some $G \in \mathcal{P}$ with $V(G) = [n]$, and $\mathcal{P}$ is monotone, the complete graph on $[n]$ is in $\mathcal{P}$. Thus, $\mathbb{P}(G(n,1) \in \mathcal{P}) = 1$, and such a $\hat{p}$ thus exists.

If $p = \omega(\hat{p})$, then, applying Corollary 4 $\lfloor p/\hat{p} \rfloor$ times, we have

$$\mathbb{P}(G(n,p) \notin \mathcal{P}) \leq \big(\mathbb{P}(G(n,\hat{p}) \notin \mathcal{P}\big)^{\lfloor p/\hat{p} \rfloor} = \left(\frac{1}{2}\right)^{\lfloor p/\hat{p} \rfloor} \to 0 \text{ as } n \to \infty.$$

If $p = o(\hat{p})$, then, similarly,

$$\frac{1}{2} = \mathbb{P}(G(n,\hat{p}) \notin \mathcal{P}) \leq \big(\mathbb{P}(G(n,p) \notin \mathcal{P})\big)^{\lfloor \hat{p}/p \rfloor}.$$

Thus, as $\hat{p} = \omega(p)$, $\mathbb{P}(G(n,p) \notin \mathcal{P}) = 1 - o(1)$. $\square$

Exercise. Prove that $\hat{p} = 1/n$ is a threshold for $G(n,p)$ to contain a triangle.

3 Sharp Thresholds

Definition. A function $\hat{p}$ is a *sharp threshold* for a property $\mathcal{P}$ if, $\forall \varepsilon > 0$, as $n \to \infty$,

- $\mathbb{P}(G(n, (1-\varepsilon)\hat{p}) \in \mathcal{P}) \to 0$, and
- $\mathbb{P}(G(n, (1+\varepsilon)\hat{p}) \in \mathcal{P}) \to 1$.

Exercise. If $\hat{p}_0$, $\hat{p}_1$ are sharp thresholds for $\mathcal{P}$, then $\hat{p}_0 - \hat{p}_1 \to 0$.

Definition. A threshold for $\mathcal{P}$ is *coarse* if $\mathcal{P}$ has no sharp threshold.

Example. $\mathcal{P}_1 = \{G : e(G) > 0\}$ has a coarse threshold. Indeed, if $p = C/n^2$, with C constant, then $\mathbb{P}(G(n,p) \in \mathcal{P}) = 1 - (1-p)^{\binom{n}{2}} \to 1 - e^{-C/2}$ as $n \to \infty$. Thus, if $\hat{p}$ was a sharp threshold for $\mathcal{P}_1$, then, as $G(n, 2\hat{p})$ is almost surely in $\mathcal{P}$, we have $\hat{p} = \omega(1/n^2)$, and as $G(n, \hat{p}/2)$ is almost surely not in $\mathcal{P}$, we have $\hat{p} = o(1/n^2)$, a contradiction.

Exercise. $\mathcal{P} = \{G : K_3 \subset G\}$ has a coarse threshold.

Rule of thumb. Properties that are 'locally checkable' have a coarse threshold. E.g. that G contains a triangle can be confirmed with only three edges, but to confirm that $\delta(G) > 0$ requires at least $n/2$ edges. This notion was formalised by Friedgut [12].

Example. $\mathcal{P} = \{G : \delta(G) \geq 2\}$ has a (very!) sharp threshold, as follows.

Lemma 5. *a) If $p = \frac{\log n + \log\log n + \omega(1)}{n}$, then $\mathbb{P}(\delta(G(n,p)) \geq 2) \to 1$ as $n \to \infty$.*

b) If $p = \frac{\log n + \log\log n - \omega(1)}{n}$*, then* $\mathbb{P}(\delta(G(n,p)) \leq 1) \to 1$ *as* $n \to \infty$.

Exercise. Show that, for each $p \in (0,1)$, $e^{-\frac{p}{1-p}} \leq 1 - p \leq e^{-p}$.

Proof of Lemma 5. Note that we can assume that $\frac{\log n}{2n} \leq p \leq \frac{2\log n}{n}$. Let X be the number of vertices with degree at most 1 in $G = G(n,p)$. Then,

$$\mathbb{E}X = n((1-p)^{n-1} + (n-1)p(1-p)^{n-2}) = (1+o(1))n^2p(1-p)^n. \tag{1}$$

a) *First moment method.* From (1), and as $p \leq \frac{2\log n}{n}$,

$$\mathbb{E}X \leq (2+o(1))n\log n \cdot e^{-pn} = o(1).$$

Thus, by Markov's inequality, $\mathbb{P}(X \geq 1) \leq \mathbb{E}X = o(1)$.

b) *Second moment method.* From (1), and as $p \geq \frac{\log n}{2n}$, we have

$$\mathbb{E}X \geq \Big(\frac{1}{2} + o(1)\Big)n\log n \cdot (1-p)^n \geq \Big(\frac{1}{2} + o(1)\Big)n\log n \cdot e^{\frac{pn}{1-p}} = \omega(1).$$

By Markov's inequality (or Chebychev's inequality),

$$\mathbb{P}(X = 0) \leq \mathbb{P}\big((X - \mathbb{E}X)^2 \geq (\mathbb{E}X)^2\big) \leq \frac{\mathbb{E}((X-\mathbb{E}X)^2)}{(\mathbb{E}X)^2} = \frac{\mathbb{E}(X^2)}{(\mathbb{E}X)^2} - 1.$$

Then,

$$\begin{aligned}
\mathbb{E}(X^2) &= \mathbb{E}X + \sum_{u \neq v} \mathbb{P}(d(u), d(v) \leq 1) \\
&\leq \mathbb{E}X + n^2((1-p)^{2n-3} + (2n-3)p(1-p)^{2n-4} + (n-2)^2p^2(1-p)^{2n-5}) \\
&= \mathbb{E}X + (1+o(1))n^4p^2(1-p)^{2n} = \mathbb{E}X + (1+o(1)) \cdot (\mathbb{E}X)^2.
\end{aligned}$$

Thus, $\mathbb{P}(X = 0) \leq 1/\mathbb{E}X + (1+o(1)) - 1 = o(1)$. □

4 Long paths and cycles

We will study how long a path we can typically find in $G(n, \frac{C}{n})$, with C a large constant. In order to cover different techniques, we will prove 5 increasingly good bounds on the size of a typical longest path in $G(n, \frac{C}{n})$. We will then show that if we start with the empty graph on n vertices and add random edges one-by-one, then the first graph with minimum degree 2 almost surely contains a Hamilton cycle (here, one with n vertices).

Definition. For each $u \in V(G)$, $N(u) = \{v : uv \in E(G)\}$, and, for each $U \subset V(G)$, $N(U) = (\cup_{u \in U} N(u)) \setminus U$.

4.1 Long paths in $G(n, \frac{C}{n})$ I: Depth First Search

Theorem 6. *$G(n, \frac{C}{n})$ almost surely has a path with length at least* $\big(1 - O_C(\frac{\log C}{C})\big)n$.

The following nice analysis of the Depth First Search algorithm and its use to find long paths (as part of a more general result) is due to Krivelevich, Lee and Sudakov [19]. The result was originally shown by Ajtai, Komlós and Szemerédi [1] and de la Vega [9].

Depth First Search (DFS) algorithm. Run the following algorithm on a graph G, governed by the following updating variables:
U – the set of unexplored vertices,
D – the set of dead vertices,
r – the number of active vertices, and
$A = (a_1, \dots, a_r)$ – the active vertices, which form a path in that order.

START: Pick some $v \in V(G)$, and let

$$U = V(G) \setminus \{v\}, \quad D = \emptyset, \quad r = 1, \quad \text{and} \quad A = (a_1) = (v).$$

STEP 1:
If possible, pick $w \in N(a_r) \cap U$, update

$$a_{r+1} \to w, \quad A \to (a_1, \dots, a_r, a_{r+1}), \quad U \to U \setminus \{w\}, \quad \text{and } r \to r+1,$$

and repeat STEP 1.

If not possible, go to STEP 2.

STEP 2:
The neighbours of a_r are fully explored. Update

$$D \to D \cup \{a_r\}, \quad A \to (a_1, \dots, a_{r-1}), \quad \text{and} \quad r \to r-1.$$

If $r \geq 1$, go to STEP 1.

Otherwise ($r = 0$), go to STEP 3.

STEP 3:
If $U \neq \emptyset$, pick $v \in U$, update

$$r \to 1, \quad a_1 \to v, \quad U \to U \setminus \{v\}, \quad \text{and} \quad A \to (a_1),$$

and go to STEP 1.

Otherwise, stop the algorithm.

OBSERVE: a) There are never any edges from D to U.
b) After each step, $r + |U| + |D| = |G|$.
c) During each step, either $|U|$ decreases by 1 and $|D|$ is unchanged, or $|D|$ increases by 1 and $|U|$ is unchanged.
d) The algorithm finishes when $U = \emptyset$ and $r = 0$, and thus $D = V(G)$.

Definition. Say G is m-joined if there is an edge in G between any 2 disjoint vertex sets with size m.

Lemma 7. *If G is m-joined, then it has a path with more than $|G| - 2m$ vertices.*

Proof. Run DFS on G. By c) and d) above, after some step we must have $|U| = |D| = s$, for some s. By a), and as G is m-joined, $s < m$. Thus, by b), at this point the active vertices form a path with more than $|G| - 2m$ vertices. □

Proposition 8. *For sufficiently large constant C, $G(n, \frac{C}{n})$ is almost surely m-joined for each $m \geq 3n \log C/C$.*

Proof. Let $m = 3n\log C/C$, $p = \frac{C}{n}$ and $G = G(n,p)$. Then, for sufficiently large C,

$$\begin{aligned}\mathbb{P}(\exists \text{ disjoint } A, B \subset V(G) \text{ with } |A|, |B| = m, e_G(A,B) = 0) &\leq \binom{n}{m}^2 (1-p)^{m^2} \\ &\leq \left(\frac{en}{m}\right)^{2m} e^{-pm^2} \\ &\leq (C^2 e^{-3\log C})^m \to 0. \qquad \square\end{aligned}$$

Thus, Lemma 7 and Proposition 8 imply Theorem 6.

Exercise. Prove $G(n, \frac{C}{n})$ almost surely has a cycle with length at least $\big(1 - O_C(\frac{\log C}{C})\big)n$.

Exercise. Changing the algorithm or the analysis, show that $G(n, \frac{C}{n})$ almost surely has a path with length at least $(1 - O_C(\frac{1}{C}))n$.

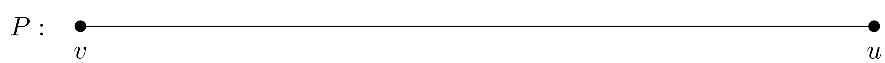
Exercise. Let G_p be a random subgraph of G, with each edge chosen independently at random with probability p. Show that, if $pk \to \infty$ and $\delta(G) \geq k$, then G_p almost surely (as $k \to \infty$) has a path with length $(1 - o_k(1))k$.

4.2 Long paths in $G(n, \frac{C}{n})$ II: Pósa rotation

To improve Theorem 6, we will use Pósa rotation, first introduced by Pósa [24] when studying the appearance of Hamilton cycles in the binomial random graph.

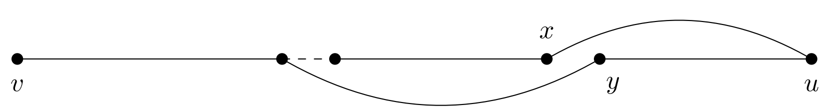
Theorem 9. *$G(n, \frac{C}{n})$ almost surely has a path with length at least $(1 - O_C(\frac{1}{C}))n$.*

Pósa rotation: Consider a path P, a longest path containing $V(P)$ in a graph G with v as one endpoint. Let u be the other endpoint and note that all the neighbours of u lie in P.



If $x \in N(u)$, and y is the neighbour of x furthest from v in the path P, then we can *break* xy and *rotate* P with v *fixed* to get the new path $P - xy + ux$ with endpoint y.

We find a new vertex y, whose neighbours must also all lie in P (by the maximality of P). We can then use edges from y to rotate again:

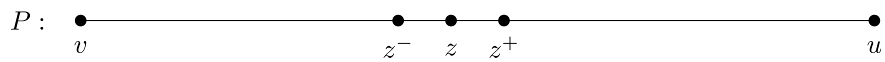


Definition. Let $F(P, v)$ be the set of endpoints not equal to v which can be achieved by iteratively rotating P with v fixed.

Observe: a) The maximality of P implies that, for each $u \in F(P,v)$, $N(u) \subset V(P)$.

b) If $xy \in E(P)$ is broken for a rotation, then x or y becomes an endpoint.

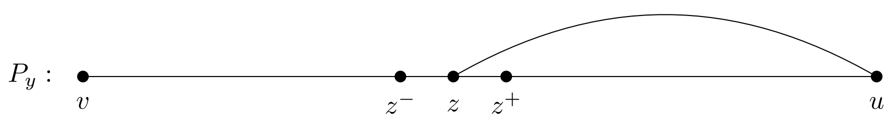
Now, label, for each $z \in V(P)$, the neighbours of z in P as



We then get the following claim.

Claim. If $z \in N(F(P,v))$, then either z^- or $z^+ \in F(P,v)$.

Proof. Let $z \in N(F(P,v))$ and assume $z^-, z^+ \notin F(P,v)$. Pick $y \in F(P,v)$ with $zy \in E(G)$, and let P_y be a path with endpoint y reached by rotating P with v fixed. As $z^-, z, z^+ \notin F(P,v)$, by b) above, both z^-z and zz^+ are not broken in rotating P to get P_y. Thus, either P_y is



and thus $z^+ \in F(P,v)$, a contradiction, or P_y is

and thus $z^- \in F(P,v)$, which is also a contradiction. □

Every vertex in $N(F(P,v))$ thus has a neighbour in $F(P,v)$ in the graph P. As $F(P,v)$ has fewer than $2|F(P,v)|$ neighbours within P (using that u has only 1 neighbour in P), we get the following.

Lemma 10. *(Pósa's Lemma) If P, a longest path containing $V(P)$ in G has v as an endpoint, then $|N(F(P,v))| < 2|F(P,v)|$.*

Definition. G is a (d,m)-*expander* if, for all $A \subset V(G)$ with $|A| \le m$, $|N(A)| \ge d|A|$.

That is, Pósa's lemma implies that if G is a $(2,m)$-expander then $|F(P,v)| > m$.

Lemma 11. *Let $m, n \in \mathbb{N}$ satisfy $m \le n/20$. Suppose $|G| = n$ and, for each $A \subset V(G)$ with $|A| = m$, $|N(A)| \ge 4n/5$. Then, G has a path with length at least $n - 2m$.*

Note. Compare this to Lemma 7, where G was required to be m-joined.

Proof of Lemma 11. We will find a large subgraph $G' \subset G$ which is a $(2, n/5)$-expander by removing a largest set that does not expand. Pick $B \subset V(G)$, with $|N(B)| < 2|B|$ and $|B| < m$, or $B = \emptyset$, to maximise $|B|$.

Claim. For all $U \subset V(G) \setminus B$ with $|U| \le n/5$, $|N(U) \setminus B| \ge 2|U|$.

Proof of claim. For contradiction, say there is some $U \subset V(G) \setminus B$ with $0 < |U| \le n/5$ and $|N(U) \setminus B| < 2|U|$. Then,

$$|N(U \cup B)| \le |N(U) \setminus B| + |N(B)| < 2|U| + 2|B| = 2|U \cup B|.$$

By the choice of B, then, $|U \cup B| \geq m$. Pick $A \subset U \cup B$ with $|A| = m$, so that

$$|N(U \cup B)| \geq |N(A)| - |U \cup B| \geq \frac{4n}{5} - \Big(\frac{n}{5} + m\Big) \geq \frac{2n}{5} + 2m \geq 2|U \cup B|,$$

a contradiction. □

Let $G' = G - B = G[V(G) \setminus B]$, which, by the claim, is a $(2, n/5)$-expander. Pick a longest path P in G', with an endpoint v, say. By Lemma 10, $|F_{G'}(P, v)| > n/5$. As P is a longest path in G', there are no edges between $F_{G'}(P, v)$ and $V(G') \setminus V(P) = V(G) \setminus (B \cup V(P))$, so that, by the condition in the lemma, $|V(G) \setminus (B \cup V(P))| < m$. Thus, $|P| > n - |B| - m > n - 2m$. □

To find the condition for Lemma 11 almost surely in $G(n, p)$, we use the following concentration inequality.

Lemma 12. *(Chernoff's inequality – see, for example [16, Corollary 2.3]) If X is a binomial variable and $0 < \varepsilon \leq 3/2$, then*

$$\mathbb{P}(|X - \mathbb{E}X| \geq \varepsilon \mathbb{E}X) \leq 2 \exp\Big(-\frac{\varepsilon^2 \mathbb{E}X}{3}\Big).$$

Proposition 13. *For sufficiently large C, in $G = G(n, \frac{C}{n})$ almost surely any $A \subset V(G)$ with $|A| = m := 100n/C$ satisfies $|N(A)| \geq 9n/10$.*

Proof. Let $p = C/n$. For each $A \subset V(G)$ with $|A| = m$, and $v \in V(G) \setminus A$,

$$\mathbb{P}(v \notin N(A)) = (1 - p)^m \leq e^{-pm} = e^{-100}.$$

For large n, then, $\mathbb{E}|N(A)| \geq 99n/100$. As $|N(A)| \sim Bin(n - m, (1 - p)^m)$, by Lemma 12 with $\varepsilon = 1/20$,

$$\mathbb{P}(|N(A)| < 9n/10) \leq \mathbb{P}\Big(|N(A)| < \Big(1 - \frac{1}{20}\Big)\mathbb{E}|N(A)|\Big) \leq 2 \exp\Big(-\frac{n}{10^4}\Big).$$

As $|V(G)^{(m)}| = \binom{n}{m} \leq (\frac{en}{m})^m \leq C^{100n/C} \leq \exp(n/10^5)$, for large C, by a union bound, we have

$$\mathbb{P}(\exists A \subset V(G) \text{ with } |A| = m \text{ and } |N(A)| < 9n/10) \leq 2 \exp\Big(-\frac{n}{10^4}\Big) \cdot \exp\Big(\frac{n}{10^5}\Big) = o(1).$$ □

Then, Lemma 11 and Proposition 13 imply Theorem 9.

4.3 Long paths in $G(n, \frac{C}{n})$ III: Revealing edges as needed

We will now improve Theorem 9, using an ad hoc method, which demonstrates how an algorithm can be combined with revealing whether each edge is present or not only when it is relevant to the algorithm.

We have used Pósa rotation on a maximal path P with an endpoint v in a $(2, \frac{n}{5})$-expander to get a linear-sized set $F(P, v)$ of possible endpoints. If we reveal more edges between a vertex $w \notin V(P)$ and $V(P)$ with probability $\frac{C}{n}$, then w has an neighbour in $F(P, v)$ with probability

$$1 - \Big(1 - \frac{C}{n}\Big)^{|F(P,v)|} \geq 1 - e^{-\frac{C}{5}},$$

and using such a neighbour we could find a path with vertex set $V(P) \cup \{w\}$. By repeating (something like) this, we can iteratively extend the path, suceeding in attaching each other vertex to the path with probability $1 - \exp(-\Omega(C))$. This will allow us to prove the following.

Theorem 14. *For sufficiently large C, $G(n, C/n)$ almost surely contains a path with length at least $(1 - e^{-C/100})n$.*

Definition. A sequence of random variables $X_1, X_2, \dots$ is a *martingale* (respectively, *supermartingale* or *submartingale*) if, for each i, $\mathbb{E}|X_i| < \infty$ and $\mathbb{E}(X_i | X_1, \dots, X_{i-1}) = X_{i-1}$ (respectively $\leq X_{i-1}$ or $\geq X_{i-1}$).

Theorem 15. *(Azuma-Hoeffding inequality) Let $(X_i)_{i=0}^n$ be a submartingale with X_0 constant, and, for each $1 \leq i \leq n$, $|X_i - X_{i-1}| \leq C_i$. Then, for all $t > 0$,*

$$\mathbb{P}(X_n \leq \mathbb{E}X_n - t) \leq \exp\left(-\frac{t^2}{2\sum_{i=1}^n C_i^2}\right).$$

Proof of Theorem 14. Consider independent random graphs $G_0, G_1 \sim G(n, C/2n)$. By Proposition 3, it is sufficient to show that $G_0 \cup G_1$ almost surely contains a path with length at least $(1 - e^{-C/100})n$.

Exercise. G_0 almost surely contains a $(4, n/10)$-expander $G_0' \subset G_0$ with at least $(1 - 1/100)n$ vertices. (See proof of Lemma 11.)

We will run an algorithm governed by the following updating variables:
- H - a large $(2, n/10)$-expander.
- P - a path which increases in length throughout the algorithm.
- D - the set of vertices we have failed to add to the path.
- W - a utility set to ensure expansion is kept in H as we add vertices and edges.

START: Let $H = G_0'$ and let P be a longest path in H, with an endpoint v, say. Let $W = D = \emptyset$. Carry out the following stage for each $1 \leq i \leq n$.

STAGE i:

If $V(P) \cup D = [n]$, then let $X_i = 1$.

Otherwise, pick $w \in [n] \setminus (V(P) \cup D)$ and reveal edges in G_1 between w and $V(P)$ - say this gives w the neighbourhood A in $V(P)$.

If $|A \cap (F_H(P, v) \setminus W)| \geq 6$, then let $X_i = 1$, and do the following.

- If $w \notin V(G_0')$, then add 6 vertices from $A \setminus W$ to W.
- Add w and the edges between w and A to H.
- Update P to be a maximal path in H containing $V(P)$ and w, with v an endpoint.

Otherwise, update $D \to D \cup \{w\}$, and let $X_i = 0$.

OBSERVE: a) Each edge in G_1 is revealed at most once.

b) At the end, $|P| \geq n - |D| = n - |\{i : X_i = 0\}| = \sum_{i=1}^n X_i$.

c) Always: for each $U \subset V(H) \setminus V(G_0')$, $|N_H(U)| \geq |N_H(U, W)| \geq 6|U| - |U| = 5|U|$.

d) Always: H is a $(2, n/10)$-expander, as follows.

Proof of d). Let $U \subset V(H)$, with $1 \leq |U| \leq n/10$, and set $U_1 = U \cap V(G_0')$ and $U_2 = U \setminus V(G_0')$.
Either $|U_1| \geq |U|/2$, whence $|N_H(U)| \geq |N_{G_0'}(U_1)| \geq 4|U_1| \geq 2|U|$,
or $|U_2| \geq |U|/2$, whence $|N_H(U)| \geq |N_H(U_2)| - |U_1| \geq 5|U_2| - |U_2| = 4|U_2| \geq 2|U|$. □

e) Always: $|W| \leq 6|V(G) \setminus V(G_0')| \leq 3n/50$, so, by Lemma 10 and d), $|F_H(P,v) \setminus W| \geq n/10 - 3n/50 = n/25$.

f) For each i, as at stage i we have, by e), $|F_H(P,v) \setminus W| \geq n/25$,

$$\begin{aligned}\mathbb{P}(X_i = 0 | X_1, \ldots, X_{i-1}) &\leq \mathbb{P}(Bin(n/25, C/2n) < 6) \\ &\leq \sum_{j=0}^{5} \left(\frac{n}{25}\right)^j \left(\frac{C}{2n}\right)^j \left(1 - \frac{C}{2n}\right)^{\frac{n}{25}-j} \\ &\leq 6C^5 e^{C/75} \\ &\leq e^{-C/100}/3,\end{aligned}$$

for large C.

Now, let $\varepsilon = e^{-C/100}/3$ and, for each $1 \leq j \leq m$, let $Y_j = \sum_{i=1}^{j}(X_i - (1-\varepsilon)n)$. As $\mathbb{E}(Y_j | Y_1, \ldots, Y_{j-1}) \geq 0$ for each j, $(Y_j)_{j=1}^{n}$ is a submartingale. As $|Y_j - Y_{j-1}| \leq 1$, for each j, by Theorem 15 we have

$$\mathbb{P}(Y_n < -\varepsilon n) \leq \exp\left(-\frac{(\varepsilon n)^2}{2n}\right) = o(1).$$

Note that $Y_n < -\varepsilon n$ implies that $(\sum_{i=1}^{n} X_i) - (1-\varepsilon)n < -\varepsilon n$. Thus, by a), at the end we almost surely have that $|P| > n - 2\varepsilon n$. $\square$

4.4 Long paths in $G(n, \frac{C}{n})$ IV: Pósa rotation and extension, and edge sprinkling

To improve Theorem 14, we will now use Pósa rotation and extension and edge sprinkling, a method developed in the literature beginning with the pioneering work of Pósa [24].

Suppose G is a $(2, \frac{n}{5})$-expander and $P \subset G$ is a maximal length path. By rotating *both* ends of P we can find $\Omega(n^2)$ paths with different endvertex pairs. Adding such an endvertex pair creates a cycle, and (often) a longer path.

Definition. Let $\ell(G)$ be the length of a longest path in G.

Definition. $e \in V(G)^{(2)}$ is a *booster for* G if $\ell(G+e) > \ell(G)$ or $G+e$ is Hamiltonian.

Observe: Iteratively adding $|G|$ boosters to G gives a Hamiltonian graph.

Exercise. Show that any $(2, n/5)$-expander is connected.

Lemma 16. *A $(2, n/5)$-expander G has at least $n^2/50$ boosters.*

Proof. Let P be a longest path in G, with v an endpoint. For each $x \in F(P,v)$, let P_x be a v,x-path with $V(P_x) = V(P)$. Let

$$E = \{xy : x \in F(P,v), y \in F(P_x, x)\}.$$

Claim A: $|E| \geq n^2/50$.

Proof of Claim A. By Lemma 10, $|F(P,v)| \geq \frac{n}{5}$ and, for each $x \in F(P,v)$, $|F(P_x,x)| \geq \frac{n}{5}$. Thus, $|E| \geq (\frac{n}{5})^2/2$. $\square$

Claim B: Each $e \in E$ is a booster.

Proof of Claim B. For each $e \in E$, $G + e$ contains a cycle with vertex set $V(P)$, C_e say. If $V(P) = V(G)$, $G + e$ is thus Hamiltonian. If not, then, as G is connected, there is some $w \in V(G) \setminus V(P)$ and $z \in V(P)$ with $wz \in E(G)$, so that $C_e + wz$ has a path with vertex set $V(P) \cup \{w\}$. Thus, $\ell(G + e) \geq \ell(C_e + wz) > \ell(G)$. □ □

Lemma 17. *If G_0 is a $(2, \frac{n}{5})$-expander with $V(G_0) = [n]$, and $G = G(n, \frac{1000}{n})$, then $G_0 \cup G$ is almost surely Hamiltonian.*

Proof. Let $m = 10n$ and $p = \frac{100}{n^2}$. Let $G_i \sim G(n, p)$, $1 \leq i \leq m$, be independent. By Proposition 3,

$$\mathbb{P}(G_0 \cup G \text{ is Hamiltonian}) \geq \mathbb{P}\big(\cup_{i=0}^{m} G_i \text{ is Hamiltonian}\big).$$

For each $1 \leq i \leq m$, let

$$X_i = \begin{cases} 1 & G_i \text{ contains a booster for } \cup_{j=0}^{i-1} G_j, \\ 0 & \text{otherwise.} \end{cases}$$

Observe: $\sum_{i=1}^{m} X_i \geq n \implies \cup_{i=0}^{m} G_i$ is Hamiltonian.

For each $1 \leq i \leq m$, $\cup_{j=0}^{i-1} G_j$ is a $(2, \frac{n}{5})$-expander as it contains G_0. Thus, by Lemma 16, $\cup_{j=0}^{i-1} G_j$ has at least $n^2/50$ boosters, and hence

$$\mathbb{P}(X_i = 0) \leq (1 - p)^{n^2/50} \leq e^{-2} \leq \frac{1}{2}.$$

Thus, $\left(\sum_{j=1}^{i} (X_j - \frac{1}{2})\right)_{i=1}^{m}$ is a submartingale, with differences at most 1. By Theorem 15, $\mathbb{P}(\sum_{j=1}^{m} (X_j - \frac{1}{2}) < -n) \leq \exp(-\frac{n^2}{2m}) = o(1)$. Thus, almost surely, $\sum_{j=1}^{m} X_j \geq m/2 - n \geq n$ and hence $\cup_{j=0}^{m} G_j$ is Hamiltonian. □

To apply this we need to find large expanding subgraphs almost surely in $G(n, \frac{C}{n})$. The expansion will follow from a few simple principles:

- Typically, there are few vertices with small degree.
- Vertices of small degree are mostly spaced apart.
- For sets U of non-small degree vertices, $U \cup N(U)$ contains many edges, so typically must be much bigger than U.

Definition. For any $A \subset V(G)$, let $e_G(A) = e(G[A])$.

Definition. Let $S \subset V(G)$. An S-path in G is a path with length at most 4 between two vertices in S. An S-cycle is a cycle with length at most 4 containing some vertex in S.

Lemma 18. *Let $m, D \geq 4$ be integers. Let G be an n-vertex graph and $S = \{v : d(v) < D\}$. Let x be the number of vertices in G with degree at most 1. Let y be the number of S-paths and S-cycles. Suppose that the following hold.*

(1) $x + 7y \leq n/5$ and $y \leq m$.

(2) Every set $A \subset V(G)$ with $|A| = m$ satisfies $|N(A)| \geq 4n/5$.

(3) There are no sets $A \subset V(G)$ with $|A| \leq 10m$ and $e_G(A) \geq D|A|/100$.

Then, G contains a $(2, n/5)$-expander with at least $n - x - 7y$ vertices.

Proof. Let B_0 be the set of vertices with degree at most 1 in G or vertices in an S-path or S-cycle. Note that $|B_0| \leq x + 5y$ and $|B_0 \setminus S| \leq 3y$. Let $B_1 \subset V(G) \setminus (B_0 \cup S)$ be a largest set subject to either $B_1 = \emptyset$ or $|B_1| \leq y$ and $e_G((B_0 \cup B_1) \setminus S) \geq D|B_1|/2$.

Claim A. $|B_1| < y$.

Proof of Claim A. Suppose to the contrary that $|B_1| = y$. Then, as $|B_0 \setminus S| \leq 3y$,

$$e_G((B_0 \cup B_1) \setminus S) \geq D|B_1|/2 \geq D|(B_0 \cup B_1) \setminus S|/8.$$

But, as $|(B_0 \cup B_1) \setminus S| \leq 4y \leq 10m$, this contradicts (3). □

Let $B_2 = N(B_1) \cap S$. Each vertex in B_1 is not in B_0, so has at most one neighbour in S. Thus, $|B_2| \leq |B_1| \leq y$. Let $H = G - (B_0 \cup B_1 \cup B_2)$, and note that $|H| \geq n - x - 7y$.

Claim B. For each $v \in V(H) \setminus S$, $|N_H(v) \setminus S| \geq D/4$.

Proof of Claim B. As $v \notin S$, $d_G(v) \geq D$. As $v \notin B_0$, v can have at most 1 neighbour in S. If $|N_G(v) \cap (B_0 \cup B_1)| \geq D/2$, then $e_G((B_0 \cup B_1 \cup \{v\}) \setminus S) \geq D(|B_1| + 1)/2$, contradicting the choice of B_1. Thus, $|N_H(v) \setminus S| \geq D - 1 - D/2 \geq D/4$. □

Claim C. H is a $(2, n/5)$-expander.

Proof of Claim C. Let $U \subset V(H)$ with $0 < |U| \leq n/5$. If $|U| \geq m$, then, picking a set $U' \subset U$ with $|U'| = m$ and considering $N(U')$, we have, using (1) and (2),

$$|N_H(U)| \geq 4n/5 - (|G| - |H|) - |U| \geq 2|U| + (n/5 - x - 7y) \geq 2|U|.$$

Suppose then that $|U| < m$. Let $U_1 = U \cap S$ and $U_2 = U \setminus S$. Each vertex in $U_1 \subset S$ has no neighbours in S, else we removed it. If a vertex in $U_1 \subset S$ has a neighbour in G on an S-path or S-cycle, or a neighbour in $S \setminus \{v\}$ or $N(S \setminus \{v\})$, then v is on an S-path or S-cycle, and so is in B_0, a contradiction. Thus, every vertex in U_1 has non neighbours in $V(G) \setminus V(H)$. Furthermore, no two neighbours in U_1 share any neighbours, as the vertices would then have been in B_0. Thus, $|N_H(U_1) \setminus S| \geq 2|U_1|$.

Furthermore, by Claim B, $e((U_2 \cup N_H(U_2)) \setminus S) \geq D|U_2|/8$, and therefore, by (3) (and as $|U_2| < m$), $|(U_2 \cup N_H(U_2)) \setminus S| > 10|U_2|$. Now, each vertex $v \in U_2$ has at most one neighbour in $N_H(U_1)$, for otherwise v is in an S-path or S-cycle. Thus, $|N_H(U_1) \cap N_H(U_2)| \leq |U_2|$. Thus,

$$|N_H(U)| \geq |N_H(U_1) \setminus S| - |U_2| + |N_H(U_2) \setminus S| - |N_H(U_1) \cap N_H(U_2)| \geq 2|U_1| + 7|U_2| \geq 2|U|. \quad \square \quad \square$$

Proposition 19. *Let C be a sufficiently large constant, $G = G(n, C/n)$ and $S = \{v : d(v) < C/100\}$. Then, almost surely, we have the following properties.*

a) There are at most $(1 + o_C(1))Ce^{-C}n$ vertices with degree at most 1.

b) There are at most $e^{-3C/2}n$ S-paths and S-cycles.

Proof. For illustration, we will show that the expected number of such vertices or paths and cycles satisfy a) and b), and leave their proofs by the second moment method to an exercise.

a) Let $p = C/n$ and let X be the number of vertices with degree at most 1. Then,

$$\mathbb{E}X = n \cdot ((1-p)^{n-1} + (n-1)p(1-p)^{n-2}) = (1+o(1))pn(1-p)^n = (1+o(1))Ce^{-C}n.$$

b) Let Y be the number of cycles with length at most 4. Then,

$$\mathbb{E}Y \leq \sum_{k=3}^{4} n^k p^k \leq 2C^4.$$

By Markov's inequality, we have $\mathbb{P}(Y \geq \log n) = o(1)$.

Let $\delta = 1/100$ and let Z be the number of S-paths. Then,

$$\begin{aligned}
\mathbb{E}Z &\leq \sum_{\ell=1}^{4} n^{\ell+1}p^{\ell} \cdot \left(\sum_{i=0}^{2\delta C} \binom{2n}{i} p^i (1-p)^{2n-5-i}\right) \\
&\leq 4nC^4 \cdot \sum_{i=0}^{2\delta C} \left(\frac{2enp}{i}\right)^i e^{-2pn(1-o(1))}. \\
&\leq 8\delta nC^5 \cdot \left(\frac{enp}{\delta C}\right)^{2\delta C} e^{-2C(1-o(1))}. \\
&\leq nC^5 \cdot \left(\frac{e}{\delta}\right)^{2\delta C} e^{-2C(1-o(1))}. \\
&\leq n \cdot C^5 \exp(-2(1-o(1)-2\delta \log(e/\delta))C) = o_C(e^{-3C/2}) \cdot n.
\end{aligned}$$

Exercise. Using the second moment method, prove a) and b). □

Proposition 20. *Let $p = C/n$ with C sufficiently large and $C \leq 2\log n$. Let $G = G(n,p)$ and $m = n/10^{16}$. Then, almost surely, we have the following properties.*

a) Every subset $A \subset V(G)$ with $|A| = m$ satisfies $|N(A)| \geq 4n/5$.

b) There is no set $A \subset V(G)$ with $|A| \leq 10m$ and $e_G(A) \geq C|A|/10^7$.

Proof. a) follows directly from Proposition 13.

b) Let $\delta = 10^{-7}$. For each $1 \leq t \leq 10m$, let p_t be the probability there is no set $A \subset V(G)$ with $|A| = t$ and $e_G(A) \geq \delta C|A|$. Then,

$$\begin{aligned}
p_t &\leq \binom{n}{t}\binom{t^2/2}{\delta Ct} p^{\delta Ct} \\
&\leq \left(\frac{en}{t}\right)^t \left(\frac{etp}{2\delta C}\right)^{\delta Ct} \\
&\leq \left(\frac{e^2tpn}{2\delta Ct}\right)^t \left(\frac{etp}{2\delta C}\right)^{\delta Ct - t} \\
&= \left(\frac{e^2}{2\delta}\right)^t \left(\frac{et}{2\delta n}\right)^{\delta Ct - t} \leq \left(\frac{e^3 t}{4\delta^2 n}\right)^{\delta Ct/2}.
\end{aligned}$$

If $t \leq \sqrt{n}$, then, as $\delta Ct \geq 10$, $p_t = o(n^{-1})$. If $t \geq \sqrt{n}$, then, as $e^3t/4\delta^2 n \leq 3/4$, $p_t = O((3/4)^t) = o(n^{-1})$. Thus, b) holds with probability at least $1 - \sum_{t=1}^{10m} p_t = 1 - o(1)$. □

Theorem 21. *$G(n, \frac{C}{n})$ almost surely has a cycle with at least $(1 - e^{-(1-o(1))C})n$ vertices.*

Proof. Let $\lambda = C - 1000$ and $G = G(n, \lambda/n)$. We will show that, almost surely, G contains a $(2, n/5)$-expander on at least $e^{-(1-o_C(1))C}n$ vertices. The result will then follow by Lemma 17 and Proposition 3.

Let $S = \{v : d(v) < \lambda/100\}$, $X = |\{v : d(v) \leq 1\}|$ and let Y be the number of S-paths and S-cycles. Almost surely, by Proposition 19, $X + 7Y \leq (1+o(1))\lambda e^{-\lambda}n \leq e^{-(1-o_C(1))C}n$. Almost surely, a) and b) in 20 hold in G with C replaced by λ. Then, by Lemma 18 with $m = n/10^{16}$ and $D = \lambda/100$, G is a $(2, n/5)$-expander. □

We can also use Lemma 17 to find a (very) sharp threshold for the appearance of Hamilton cycles in the binomial random graph, for which we need the following.

Proposition 22. *Let $\log n \leq \lambda \leq 2\log n$, $G = G(n, \lambda/n)$ and $S = \{v : d(v) \leq \lambda/100\}$. Then, almost surely, G has no S-paths or S-cycles.*

Proof. Let $\delta = 1/100$ and $p = \lambda/n$. Let X be the number of S-paths. Then,

$$\begin{aligned}
\mathbb{E}X &\leq \binom{n}{2}\sum_{k=0}^{2} p^{k+1}n^k \cdot \sum_{i=0}^{2\delta\lambda}\binom{2n}{i}p^i(1-p)^{2n-5-i} \\
&\leq 3n^2p\lambda^2 \cdot (2\delta\lambda + 1)\left(\frac{2enp}{2\delta\lambda}\right)^{2\delta\lambda} e^{-(2+o(1))np} \\
&\leq (n\log^4 n)\cdot \exp(2\delta\lambda\log(1/\delta) - (2+o(1))\lambda) = o(1).
\end{aligned}$$

Then, almost surely, there are no paths with length at most 4 and endpoints in S.

Let Y be the number of S-cycles. Then, similarly,

$$\begin{aligned}
\mathbb{E}Y &\leq n \cdot \sum_{k=2}^{3} p^{k+1}n^k \cdot \sum_{i=0}^{\delta\lambda}\binom{n-3}{i}p^i(1-p)^{n-3-i} \\
&\leq 2\lambda^4 \cdot (\delta\lambda+1)\cdot\left(\frac{enp}{\delta\lambda}\right)^{\delta\lambda} e^{-(1+o(1))np} \leq \lambda^5 \cdot \left(\frac{e}{\delta}\right)^{\delta\lambda} e^{-(1+o(1))\lambda} = o(1).
\end{aligned}$$

Thus, almost surely, there are no such cycles in G. □

This gives the following theorem, due originally to Bollobás [5] and Komlós and Szemerédi [18].

Theorem 23. *If $p = \frac{\log n + \log\log n + \omega(1)}{n}$, then $G(n,p)$ is almost surely Hamiltonian.*

Proof. Let $q = p - 1000/n$ and $G = G(n,q)$. By Lemma 17 and Proposition 3, it is sufficient to show that G is almost surely a $(2, n/5)$-expander.

Note that $q = \frac{\log n + \log\log n + \omega(1)}{n}$. Thus, by Lemma 5, $\delta(G) \geq 2$ almost surely. Let $\delta = 10^{-5}$, $m = n/10^{16}$ and $S = \{v : d(v) < \delta qn\}$. Almost surely, by Proposition 20, any subset $A \subset V(G)$ with $|A| = m$ satisfies $|N(A)| \geq 4n/5$ and there is no set $A \subset V(G)$ with $|A| \leq 10m$ and $e_G(A) \geq \delta qn/100$. Almost surely, by Proposition 22, there are no cycles with a vertex in S or paths with length at most 4 between vertices in S. Thus, by Lemma 18 with m and $D = \delta qn$, G is a $(2, n/5)$-expander, as required. □

4.5 Long paths in $G(n, \frac{C}{n})$ V: Conditional existence of boosters

We will improve on Theorem 21 to give our final result on long cycles in $G(n, \frac{C}{n})$ using a conditioning argument by Lee and Sudakov [21].

Given any fixed $(2, \frac{n}{5})$-expander $H \subset K_n$, H has many boosters, so that it is very likely one of them will appear in $G = G(n,p)$. Typically there are many such graphs H which will appear in G, but by restricting such graphs further we can find a large set of $(2, \frac{n}{5})$-expanders in G which all have boosters in G. The graphs H being sparse in comparison to G is sufficient for this to hold, as follows.

Lemma 24. *Let $p \geq 1/n$ and $G = G(n,p)$. Almost surely, any $(2, \frac{n}{5})$-expander $H \subset G$ with $e(H) \leq pn^2/10^4$ has a booster in G.*

Proof. Let $\delta = 10^{-4}$. Let $\mathcal{H}$ be the set of $(2, n/5)$-expander graphs H with $V(H) \subset [n]$ and $e(H) \leq \delta pn^2$. For each $H \in \mathcal{H}$, let B_H be the set of boosters in H which are not in $E(H)$, so that, by Lemma 10,

$$|B_H| \geq n^2/50 - \delta pn^2 \geq n^2/100.$$

For each $H \in \mathcal{H}$, B_H contains no edges in H, so that

$$\mathbb{P}(B_H \cap E(G) = \emptyset | H \subset G) = \mathbb{P}(B_H \cap E(G) = \emptyset) = (1-p)^{|B_H|} \leq e^{-pn^2/100}.$$

Let q be the probability that, for some $H \in \mathcal{H}$, $H \subset G$ and $e(G) \cap B_H = \emptyset$. Then,

$$q \leq \sum_{H \in \mathcal{H}} \mathbb{P}((H \subset G) \wedge (B_H \cap E(G) = \emptyset)) = \sum_{H \in \mathcal{H}} \mathbb{P}(B_H \cap E(G) = \emptyset | H \subset G) \cdot \mathbb{P}(H \subset G).$$

Therefore,

$$\begin{aligned}
\mathbb{P}(\exists H \in \mathcal{H} \text{ with } H \subset G \text{ and } B_H \cap E(G) = \emptyset) &\leq \sum_{H \in \mathcal{H}} \mathbb{P}((H \subset G) \wedge (B_H \cap E(G) = \emptyset)) \\
&= \sum_{H \in \mathcal{H}} \mathbb{P}(B_H \cap E(G) = \emptyset | H \subset G) \cdot \mathbb{P}(H \subset G) \\
&\leq e^{-pn^2/100} \sum_{H \subset K_n, e(H) \leq \delta pn^2} \mathbb{P}(H \subset G) \\
&\leq e^{-pn^2/100} \sum_{t=0}^{\delta pn^2} \binom{n^2}{t} p^t \\
&\leq e^{-pn^2/100} \sum_{t=0}^{\delta pn^2} \left(\frac{en^2 p}{t}\right)^t \\
&\leq e^{-pn^2/100} \cdot n^2 \cdot \left(\frac{e}{\delta}\right)^{\delta pn^2} \\
&\leq n^2 \cdot \exp((-1/100 + \delta \log(e/\delta)) pn^2) \\
&\leq n^2 \cdot \exp(-pn^2/1000) = o(1).
\end{aligned}$$

□

Corollary 25. *Let $p \geq 10^5/n$ and $G = G(n,p)$. Almost surely, for any $(2, \frac{n}{5})$-expander $H \subset G$ with $e(H) \leq 5pn^2/10^5$, $G[V(H)]$ is Hamiltonian.*

Proof. Let $\delta = 5/10^5$. Almost surely, G has the property from Lemma 24. Let then $H \subset G$ be a $(2, \frac{n}{5})$-expander with $e(H) \leq \delta pn^2$. Let $H_0 = H$ and let $\ell \leq n$ be the largest integer for which there is a sequence $H_0 \subset H_1 \subset \ldots \subset H_\ell \subset G$ such that, for each $1 \leq i \leq \ell$, H_i is formed by adding a booster to H_{i-1}.

Now, $e(H_\ell) \leq \delta pn^2 + \ell \leq pn^2/10^4$ and H_ℓ is a $(2, n/5)$-expander, so that, by the property from Lemma 24, H_ℓ has a booster in G. Thus, by the choice of ℓ, we must have $\ell = n$. Thus, H_ℓ, and hence $G[V(H)]$, is Hamiltonian. □

We can now prove two essentially best possible results about long cycles in $G(n, C/n)$ and Hamiltonicity in $G(n,p)$. The first of these was originally proved by Frieze [14], and the second by Ajtai, Komlós and Szemerédi [2].

Theorem 26. *For any sufficiently large constant C, $G(n, C/n)$ almost surely has a cycle on at least $(1 - (1 + o_C(1))Ce^{-C})n$ vertices.*

Theorem 27. *If $p = \frac{\log n + \log\log n + c_n}{n}$ and $c \in \mathbb{R}$, then*

$$\mathbb{P}(G(n,p) \text{ is Hamiltonian}) \to \begin{cases} 0 & \text{if } c_n \to -\infty \\ e^{-e^{-c}} & \text{if } c_n \to c \\ 1 & \text{if } c_n \to \infty. \end{cases}$$

These results follow from the following theorem.

Theorem 28. *There exists some sufficiently large C_0 such that the following holds for any $C_0 \leq \lambda \leq 2\log n$. Let $G = G(n, \lambda/n)$, $X = |\{v : d(v) \leq 1\}|$, $S = \{v : d(v) < \lambda/100\}$ and let Y be the number of S-paths and S-cycles in G. Then, almost surely, G has a cycle on at least $n - X - 7Y$ vertices.*

Proof. Almost surely, by Corollary 25, for any $(2, n/5)$-expander subgraph $H \subset G$ with $e(H) \leq 5\lambda n/10^5$, $G[V(H)]$ is Hamiltonian. Let $m = n/10^{16}$. Almost surely, by Proposition 20, the following holds.

a) There are no sets $A \subset V(G)$ with $|A| \leq 10m$ and $e_G(A) \geq \lambda|A|/10^7$.

Form $G_0 \subset G$, a graph with vertex set $V(G)$ by, for each $v \in V(G)\backslash S$, adding $\min\{d(v), \lambda/10^5\}$ edges adjacent to v. Note that $e(G_0) \leq \lambda n/10^5$.

Let $G_1 \subset G$ be a subset of G with each edge chosen independently at random with probability $1/10^5$. Noting that $G_1 \sim G(n, \lambda/10^5 n)$, by Lemma 12 we almost surely have $e(G_1) \leq \lambda n/10^5$ and, by Proposition 13, we almost surely have the following.

b) For all $A \subset V(G)$ with $|A| = m$, $|N(A)| \geq 4n/5$.

Now, $G_0 \cup G_1 \subset G$ has X vertices with degree at most 1 and, letting $S' = \{v : d_{G'}(v) < \lambda/10^7\} \subset S$, at most Y S'-paths and S'-cycles, and a) and b) hold. Thus, by Lemma 18 there is some $(2, n/5)$-expander $H \subset G_0 \cup G_1$ on at least $n - X - 7Y$ vertices. Hence, as $e(H) \leq e(G_0 \cup G_1) \leq 5\lambda/10^5$, $G[V(H)]$ is Hamiltonian. I.e., G contains a cycle with at least $n - X - 7Y$ vertices. □

Theorem 26 follows from Proposition 19 and Theorem 28.

Exercise. Prove Theorem 27 by studying $\mathbb{P}(\delta(G(n,p)) \geq 2)$.

4.6 Hamiltonicity in the random graph process

Let G_0 be the empty graph with vertex set $[n]$ and create the following sequence of random graphs $G_0 \subset G_1 \subset \ldots \subset G_{\binom{n}{2}}$, where, for each $1 \leq M \leq \binom{n}{2}$, G_M is formed from G_{M-1} by, uniformly at random, adding a non-edge. We call $\{G_M\}_{M\geq 0}$ the *random graph process*. Our aim here is to show that, almost surely, the very edge we add in this process which gives the graph minimum degree 2 also creates the first Hamilton cycle. That is, the *hitting time* for Hamiltonicity and the hitting time for minimum degree at least 2 almost surely coincide.

Definition. Let $G_{n,M}$ be the random graph with vertex set $[n]$ and M edges, with all such graphs equally likely.

Definition. Say an event *holds in almost every (a.e.) random graph process* if the event holds almost surely in the random graph process.

Switching between graph models. Note that in the n-vertex random graph process $\{G_M\}_{M\geq 0}$ each G_M is distributed as $G_{n,M}$. Often $G(n,p)$ is easier to work with due to the independence between edges. Fortunately, the random graph $G_{n,M}$ with $M = p\binom{n}{2}$ edges is closely related to $G(n,p)$. For example $G(n,p)$ almost surely has around $p\binom{n}{2}$ edges, as follows.

Proposition 29. *Let $p \geq 10/n$ satisfy $p = o(1)$. Then, almost surely, $-n \leq e(G) - p\binom{n}{2} \leq n$.*

Proof. Let $M = p\binom{n}{2}$ and $\varepsilon = n/M = 2/(n-1)p \leq 1$, so that $\varepsilon^2 M = n^2/M = \Omega(1/p) = \omega(1)$. By Lemma 12 with ε, we have

$$\mathbb{P}(|e(G) - M| \leq n) \leq 2\exp(-\varepsilon^2 M/3) = o(1). \qquad \square$$

This allows us to move simply between models, as follows.

Proposition 30. *If $\mathcal{P}$ is a graph property, then the following hold.*

a) Let $p \geq 10/n$ satisfy $p = o(1)$. Suppose $\mathcal{P}$ almost surely contains $G(n,p)$. Then, there is some M with $p\binom{n}{2} - n \leq M \leq p\binom{n}{2} + n$ for which, almost surely, $G_{n,M} \in \mathcal{P}$.

b) Suppose that in almost every graph process, every graph has property $\mathcal{P}$. Then, for any $p = p(n)$, $G(n,p)$ is almost surely in $\mathcal{P}$.

Proof. a) Let $\mathcal{M} = \{M : p\binom{n}{2} - n \leq M \leq p\binom{n}{2} + n\}$. Let $\varepsilon > 0$ and let n be sufficiently large that $\mathbb{P}(G(n,p) \in \mathcal{P}) \geq 1 - \varepsilon/2$ and, using Proposition 29, that $\mathbb{P}(e(G(n,p)) \in \mathcal{M}) \geq 1 - \varepsilon/2$. Then,

$$\mathbb{P}(e(G(n,p) \in \mathcal{M} \wedge G(n,p) \in \mathcal{P})) \geq 1 - \varepsilon,$$

so that

$$\sum_{M\in\mathcal{M}} \mathbb{P}(G_{n,M} \in \mathcal{P}) \cdot \mathbb{P}(e(G(n,p)) = M) = \mathbb{P}(e(G(n,p)) \in \mathcal{M} \wedge G(n,p) \in \mathcal{P}) \geq 1 - \varepsilon.$$

Thus, for some $M \in \mathcal{M}$, $\mathbb{P}(G_{n,M} \in \mathcal{P}) \geq 1 - \varepsilon$.

b) For each $\varepsilon > 0$, we must have, for sufficiently large n, that $\mathbb{P}(G_{n,M} \in \mathcal{P}) \geq 1 - \varepsilon$ for each $0 \leq M \leq \binom{n}{2}$. Thus, for any p,

$$\mathbb{P}(G(n,p) \in \mathcal{P}) = \sum_{M=0}^{\binom{n}{2}} \mathbb{P}(G(n,p) = M) \cdot \mathbb{P}(G_{n,M} \in \mathcal{P}) \geq \sum_{M=0}^{\binom{n}{2}} \mathbb{P}(G(n,p) = M) \cdot (1-\varepsilon) = 1-\varepsilon. \qquad \square$$

We will show the following theorem, which was originally proved by Bollobás [6]

Theorem 31. *In almost every random graph process, for each $0 \leq M \leq \binom{n}{2}$, $\delta(G_M) \geq 2 \implies G_M$ is Hamiltonian.*

Thus, we have by Proposition 30b) applied with $\mathcal{P} = \{G : \delta(G) \geq 2 \implies G \text{ is Hamiltonian}\}$, we have the following.

Corollary 32. *For any $p = p(n)$, we have*

$$\mathbb{P}(G(n,p) \text{ is Hamiltonian}) = \mathbb{P}(\delta(G(n,p)) \geq 2) + o(1).$$

Before we prove Theorem 31, we will record a couple of useful propositions, the first giving a simple bound on the likely maximum degree.

Proposition 33. *If $p \leq 2\log n/n$, then, almost surely, $\Delta(G(n,p)) \leq 10\log n$.*

Proof. Let X be the number of vertices with degree more than $10\log n$. Then,

$$\mathbb{P}(X \geq 1) \leq \mathbb{E}X \leq n \cdot \binom{n}{10\log n} p^{10\log n} \leq n \cdot \left(\frac{enp}{10\log n}\right)^{10\log n} \leq n \cdot (e/5)^{10\log n} = o(1). \qquad \square$$

Proposition 34. *Almost surely, $G = G(n, (1 - \frac{1}{100})\frac{\log n}{n})$ satisfies the following properties with $S = \{v : d(v) < \frac{\log n}{100}\}$ and $m = n/10^{16}$.*

A1 *$\delta(G) \leq 1$.*

A2 *There is no set $A \subset V(G)$ with $|A| \leq 10m$ and $e_G(A) \geq |A|\log n/10^7$.*

A3 *Every subset $A \subset V(G)$ with $|A| = m$ satisfies $|N(A)| \geq 4n/5$.*

A4 *There are no S-paths or S-cycles.*

A5 *$|S| \leq n^{1/4}$.*

Proof. **A1** follows from Lemma 5. **A2** and **A3** restate part of Proposition 20, while **A4** restates Proposition 22.

Proof of **A5**: Let $X = |S|$ and $\delta = 1/100$. Then,

$$\begin{aligned}\mathbb{E}X &\leq n \sum_{i=0}^{\delta\log n} \binom{n}{i} p^i (1-p)^{n-1-i} \leq n \sum_{i=0}^{\delta\log n} \left(\frac{enp}{i}\right)^i e^{-(1-o(1))pn} \\ &\leq n \cdot \log n \cdot (e/\delta)^{\delta\log n} e^{-(1-\delta-o(1))\log n} = o(n^{1/4}).\end{aligned}$$

Thus, by Markov's inequality, $\mathbb{P}(X \leq n^{1/4}) = o(n^{-1})$. $\square$

When combined with Proposition 30a), this gives the following corollary.

Corollary 35. *There exists some $M_0 = M_0(n) \leq (1 - \frac{1}{200})\frac{n\log n}{2}$ such that in almost every n-vertex random graph process $\{G_M\}_{M\geq 0}$, G_{M_0} satisfies **A1** – **A5**.*

We can now prove Theorem 31, using a method from the work of Krivelevich, Lubetzky and Sudakov [20].

Proof of Theorem 31. Let $\{G_M\}_{M\geq 0}$ be the n-vertex random graph process, and, for each M, let e_M be the edge added to G_{M-1} to create G_M. Let $\delta = 1/200$, and take $M_0 \leq (1-\delta)\frac{n\log n}{2}$ from Corollary 35. Let $M_1 = \frac{n\log n}{2}$ and $M_2 = (1+\delta)\frac{n\log n}{2}$.

Reveal edges $e_1, e_2, \dots, e_{M_0}$. By the choice of M_0, almost surely, **A1**–**A5** hold in G_{M_0}. Let $S = \{v : d_{G_{M_0}}(v) < \frac{\log n}{100}\}$.

Now, let $H_{M_0} = G_{M_0}$. For each $M_0 < i \leq M_2$, reveal whether e_i contains a vertex in S or not.

- If it does, let $X_M = 0$, reveal e_M (i.e., its exact location) and let $H_M = H_{M-1} + e_M$.
- If it does not, let $X_M = 1$ and $H_M = H_{M-1}$.

When finished, let $I = \{M : M_0 < M \leq M_1, X_M = 1\}$.

Claim A. Almost surely, $|I| \geq \delta n \log n/4$.

Proof of Claim A. As, by **A5**, $|S| \leq n^{1/4}$, we have, for each M, $M_0 < M \leq M_1$,

$$\mathbb{E}(X_M | X_{M_0}, \dots, X_{M-1}) \geq \left(1 - \frac{|S|\cdot n}{\binom{n}{2} - (M-1)}\right) \geq \frac{3}{4}.$$

Thus, $\left(\sum_{M=M_0+1}^{i}(X_M - 3/4)\right)_{i=M_0+1}^{M_1}$ is a submartingale. By Theorem 15, then, for $t = \delta \log n/8$,

$$\mathbb{P}\left(\sum_{M=M_0+1}^{M_1} X_M < 3(M_1 - M_0)/4 - t\right) \leq \exp\left(-\frac{t^2}{2(M_1 - M_0)}\right),$$

so that, as $|I| = \sum_{M=M_0+1}^{M_1} X_M$ and $M_1 - M_0 \geq \frac{\delta n \log n}{2}$,

$$\mathbb{P}\left(|I| < \frac{\delta n \log n}{4}\right) = o(1). \qquad \square$$

We know the edges e_M, $M \in I$, lie within $[n] \setminus S$, but nothing further about their location, other than they are not in H_{M_0}.

Claim B. Almost surely, for each $M_0 \leq M \leq M_2$, the following event E_M holds.

$$E_M: \text{ There are no } S\text{-paths or } S\text{-cycles in } H_M.$$

Proof of Claim B. Note that we know E_{M_0} holds by **A5**. If E_{M-1} holds then, for E_M not to hold e_M must lie within $S_M := S \cup N_{H_{M-1}}(S) \cup N_{H_{M-1}}(N_{H_{M-1}}(S))$. If, in addition, $\Delta(H_{M-1}) \leq 10 \log n$, then, by **A5**,

$$|S_M| \leq 2(10\log n)^2 |S| = o(n^{1/3}).$$

Thus, letting $\bar{E}_M$ be the event that E_M fails,

$$\mathbb{P}(\bar{E}_M | E_{M-1} \wedge (\Delta(H_{M-1}) \leq 10 \log n)) = o\left(\frac{n^{2/3}}{\binom{n}{2} - M}\right) = o(n^{-4/3}).$$

Now,

$$\begin{aligned}&\mathbb{P}(\bar{E}_M \text{ for some } M_0 \leq M \leq M_2)\\ &\qquad \leq \mathbb{P}(\Delta(H_{M_2}) > 10\log n) + \mathbb{P}((\bar{E}_M \text{ for some } M_0 \leq M \leq M_2) \wedge (\Delta(H_{M_2}) \leq 10 \log n)).\end{aligned}$$

By Propositions 33 and 30a), $\mathbb{P}(\Delta(H_{M_2}) > 10\log n) \leq \mathbb{P}(\Delta(G_{n,M_2}) > 10\log n) = o(1)$. Therefore,

$$\begin{aligned}
&\mathbb{P}(\bar{E}_M \text{ for some } M_0 \leq M \leq M_2) \\
&\quad \leq o(1) + \sum_{M=M_0+1}^{M_2} \mathbb{P}(\bar{E}_M \wedge E_{M-1} \wedge (\Delta(H_{M_2}) \leq 10\log n)) \\
&\quad \leq o(1) + \sum_{M=M_0+1}^{M_2} \mathbb{P}(\bar{E}_M \wedge E_{M-1} \wedge (\Delta(H_{M-1}) \leq 10\log n)) \\
&\quad \leq o(1) + \sum_{M=M_0+1}^{M_2} \mathbb{P}(\bar{E}_M | E_{M-1} \wedge (\Delta(H_{M-1}) \leq 10\log n)) \\
&\quad \leq o(1) + M_2 \cdot n^{-4/3} \leq o(1) + (n\log n) \cdot n^{-4/3} = o(1). \qquad \square
\end{aligned}$$

Note that, for each $M_0 \leq M \leq M_2$, $\delta(H_M) \geq 2$ exactly when $\delta(G_M) \geq 2$. Thus, by Lemma 5 and Proposition 30, we almost surely have $\delta(H_{M_2}) \geq 2$. Let N be the smallest $N \leq M_2$ such that $\delta(H_N) \geq 2$. By **A1**, $N > M_0$, so that G_N is the first graph in the sequence $\{G_M\}_{M\geq 0}$ with minimum degree 2. Let $L_0 = H_N$. By **A2**–**A3**, Claim B and Lemma 18, and as $\delta(L_0) \geq 2$, L_0 is a $(2, n/5)$-expander.

Let $m = |I| \geq \delta n \log n/4$ (by Claim A) and label $I = \{j_i : i \in [m]\}$ so that $j_1 < \ldots < j_m$. For each $1 \leq M \leq m$, let $L_M = L_0 + (\sum_{i=1}^{M} e_{j_i})$. For each $1 \leq M \leq m$, let Y_M be the indicator function for the event that e_{j_M} is a booster for L_{M-1}.

For each $1 \leq M \leq m$, As L_0, and thus L_{M-1}, is a $(2, \frac{n}{5})$-expander, by Lemma 10, the number of boosters for L_{M-1} within $[n] \setminus S$ but not in $E(L_{M-1})$ is at least $\frac{n^2}{50} - n \cdot |S| - M \geq n^2/100$. Thus, for each $1 \leq M \leq m$,

$$\mathbb{E}(Y_M | Y_1, \ldots, Y_{M-1}) \geq \left(\frac{n^2}{100}\right) \Big/ \binom{n}{2} \geq \frac{1}{50}.$$

By applying Theorem 15 to the submartingale $(\sum_{M=1}^{i}(Y_M - 1/50))_{i=1}^{m}$, we have that, almost surely, $\sum_{M=1}^{m} Y_M \geq m/100 \geq n$. Thus, L_m is formed from L_0 by adding iteratively n boosters (and some other edges), and thus is Hamiltonian. Therefore, $G_N = L_m$, the first graph in the sequence with minimum degree 2 is Hamiltonian. $\square$

5 Random Directed Graphs

Orienting the edges of a graph can often create an interesting new problem. For example, giving each edge $uv = \{u, v\}$ an orientation and looking for a long directed cycle — where the edges have the same direction around the cycle — nullifies our previous approach using Pósa rotation, as rotation does not maintain a directed path.

Definition. An *oriented graph* D has vertex set $V(D)$ and a set $E(D) \subset \{\vec{uv} : u, v \in V(D), u \neq v\}$ of directed edges, where at most one of $\vec{uv}$ and $\vec{vu}$ are in $E(D)$.

Definition. A *directed graph (digraph)* D has vertex set $V(D)$ and a set $E(D) \subset \{\vec{uv} : u, v \in V(D), u \neq v\}$ of directed edges, where $\vec{uv}$ and $\vec{vu}$ may both be in $E(D)$.

We will work in random digraphs, though in practice we will typically have few $u, v \in V(D)$ with $\vec{uv}, \vec{vu} \in E(D)$

Definition. Let $D(n,p)$ be a random digraph with vertex set $[n]$, where each of the possible $n(n-1)$ edges occurs independently at random with probability p.

An elegant coupling argument of McDiarmid [22] allows us to give a good bound on the appearance of oriented subgraphs in $D(n,p)$, if we have a good bound on the appearence of the underlying subgraph in $G(n,p)$.

Theorem 36. *Let F be a digraph created from a graph H by orienting edges. Then, for any $n \in \mathbb{N}$ and $p \in (0,1)$,*

$$\mathbb{P}(D(n,p) \text{ contains a copy of } F) \geq \mathbb{P}(G(n,p) \text{ contains a copy of } H).$$

Proof. We will interpolate between $D_0 \sim G(n,p)$ and $D_N \sim D(n,p)$, with $N = \binom{n}{2}$, using a sequence of random digraphs $D_0, D_1, \ldots, D_N$ defined as follows using an arbitrary enumeration $e_1, \ldots, e_N$ of the edges of K_n, with $e_i = \{v_i, w_i\}$.

For each $0 \leq i \leq N$, let D_i be the random digraph where, for each $j \in [N]$,

- if $j \leq i$, then $\vec{v_j w_j}$ and $\vec{w_j v_j}$ are included independently *of each other* with probability p, and
- if $j > i$, then $\vec{v_j w_j}$ and $\vec{w_j v_j}$ are included *together* with probability p, and otherwise omitted.

Note that $D_N \sim D(n,p)$, and D_0 can be formed by replacing each $uv \in E(G(n,p))$ with $\vec{uv}$ and $\vec{vu}$. Considering the latter, in such a coupling any copy of H in $G(n,p)$ corresponds to some copy of F in D_0. Thus, using $\precsim$ to denote 'contains a copy of',

$$\mathbb{P}(F \precsim D_0) = \mathbb{P}(H \precsim G(n,p)).$$

Now, let $i \in [N]$ and note that D_i and D_{i-1} differ only on $\{\vec{v_i w_i}, \vec{w_i v_i}\}$. Let $\tilde{D}_i = D_i - \{\vec{v_i w_i}, \vec{w_i v_i}\} = D_{i-1} - \{\vec{v_i w_i}, \vec{w_i v_i}\}$.

EITHER a) $F \precsim \tilde{D}_i$, so that $\mathbb{P}(F \precsim D_i | \tilde{D}_i) = \mathbb{P}(F \precsim D_{i-1} | \tilde{D}_i) = 1$.

OR b) $F \not\precsim \tilde{D}_i$, but there is some $e \in \{\vec{v_i w_i}, \vec{w_i v_i}\}$ such that $F \precsim \tilde{D}_i + e$, so that $\mathbb{P}(F \precsim D_i | \tilde{D}_i) \geq p = \mathbb{P}(F \precsim D_{i-1} | \tilde{D}_i)$.

OR c) $F \not\precsim \tilde{D}_i + \{\vec{v_i w_i}, \vec{w_i v_i}\}$, so that $\mathbb{P}(F \precsim D_i | \tilde{D}_i) = \mathbb{P}(F \precsim D_{i-1} | \tilde{D}_i) = 0$.

Note that in this division we have used that no copy of F contains both $\vec{v_i w_i}$ and $\vec{w_i v_i}$.

Thus, in each case

$$\mathbb{P}(F \precsim D_i | \tilde{D}_i) \geq \mathbb{P}(F \precsim D_{i-1} | \tilde{D}_i),$$

so that $\mathbb{P}(F \precsim D_i) \geq \mathbb{P}(F \precsim D_{i-1})$.

Therefore,

$$\mathbb{P}(F \precsim D(n,p)) = \mathbb{P}(F \precsim D_N) \geq \mathbb{P}(F \precsim D_0) = \mathbb{P}(H \precsim G(n,p)). \qquad \square$$

Remark. Nothing of the specific structure of graphs is used in the proof of Theorem 36. We equally could have proved the following

Theorem 36'. *Let $[n]_p$ and $([n] \times [2])_p$ be subsets of $[n]$ and $[n] \times [2]$ with each element chosen independently at random with probability p. Let $\pi : [n] \times [2] \to [n]$ be the projection onto the first co-ordinate. Let $\mathcal{A} \subset \mathcal{P}([n])$ and let $\mathcal{B} \subset \mathcal{P}([n] \times [2])$ be such that for all $B \in \mathcal{B}$ and $x \in [n]$, $\{x\} \times [2] \not\subset B$, and $\mathcal{A} \subset \pi(\mathcal{B})$. Then,*

$$\mathbb{P}(\exists B \in \mathcal{B} : B \subset ([n] \times [2])_p) \geq \mathbb{P}(\exists A \in \mathcal{A} : A \subset [n]_p).$$

Definition. A digraph D with n vertices is *Hamiltonian* if it contains a spanning directed cycle (i.e. one with vertex set $V(D)$).

Combining Theorem 36 and Theorem 23, gives the following.

Corollary 37. *If* $p = \frac{\log n + \log\log n + \omega(1)}{n}$*, then* $D(n,p)$ *is almost surely Hamiltonian.*

Note. Corollary 37 was subsequently improved by Frieze [15].

Corollary 38. *For each constant* $C > 0$*,* $D(n, \frac{C}{n})$ *almost surely contains a directed cycle with at least* $(1-(1+o_C(1))Ce^{-C})n$ *vertices.*

Are these corollaries best possible? Note that the best degree condition that holds in every Hamiltonian digraph D is that $\delta^{\pm}(D) \geq 1$.

Exercise. Show, for each $v \in [n]$, if $p = \frac{\log n + \omega(1)}{n}$, $D = D(n,p)$, $G = G(n,p)$ and $j \in \{+,-\}$, then

$$\mathbb{P}(d_D^j(v) \geq 1) = \mathbb{P}(d_G(v) \geq 1) = 1 - o(n^{-1}),$$

so that

$$\mathbb{P}(\delta^{\pm}(D) \geq 1) \geq 1 - 2n \cdot o(n^{-1}) = 1 - o(1).$$

Exercise^{++}. Show that if $p = \frac{\log n + \omega(1)}{n}$, then $D(n,p)$ is almost surely Hamiltonian.

Exercise^{++}. Show that $D(n, \frac{C}{n})$ almost surely contains a directed cycle with at least $(1-(1+o_C(1))e^{-C})n$ vertices.

6 The chromatic number of dense random graphs

One simple restriction on $\chi(G)$, the chromatic number of G, follows from the size of the largest independent set, $\alpha(G)$. That is, as $\alpha(G)$ is an upper bound on the size of any colour class, $\chi(G) \geq |G|/\alpha(G)$.

Lemma 39. *Let* $p \in (0,1)$ *be constant, and* $b = \frac{1}{1-p}$*. Then, almost surely* $\alpha(G(n,p)) \leq \lceil 2\log_b n \rceil$.

Proof. Let $k = \lceil 2\log_b n \rceil$ and let X be the number of independent sets with size k. Then,

$$\mathbb{P}(X > 0) \leq \mathbb{E}X = \binom{n}{k}(1-p)^{\binom{k}{2}} \leq \left(\frac{en(1-p)^{(k-1)/2}}{k}\right)^k \leq \left(\frac{e}{k(1-p)^{1/2}}\right)^k = o(1). \qquad \square$$

Corollary 40. *Let* $p \in (0,1)$ *be constant, and* $b = \frac{1}{1-p}$*. Then, almost surely,* $\chi(G(n,p)) \geq \frac{n}{2\log_b n}$.

Idea. If we iteratively remove maximal independent sets, we can (almost surely) find a colouring with $(1+o(1))\frac{n}{2\log_b n}$ colours.

To show such large independent sets exist, we will use the following definition and theorem.

Definition. For any graphs G and H, $G \cap H$ is the graph with vertex set $V(G) \cap V(H)$ and edge set $E(G) \cap E(H)$.

We will use the following version of Janson's inequality, which, for example, follows directly from Theorems 8.1.1 and 8.1.2 in [3].

Theorem 41. *(Janson's inequality) Let $p \in (0,1)$ and let $\{H_i\}_{i\in\mathcal{I}}$ be a family of subgraphs of the complete graph on the vertex set $[n]$. For each $i \in \mathcal{I}$, let X_i denote the indicator random variable for the event that $H_i \subset G(n,p)$ and, for each ordered pair $(i,j) \in \mathcal{I}\times\mathcal{I}$, with $i \neq j$, write $H_i \sim H_j$ if $E(H_i)\cap E(H_j) \neq \emptyset$. Then, for*

$$X = \sum_{i\in\mathcal{I}} X_i, \qquad \mu = \mathbb{E}X = \sum_{i\in\mathcal{I}} p^{e(H_i)},$$

$$\delta = \sum_{(i,j)\in\mathcal{I}\times\mathcal{I}, H_i\sim H_j} \mathbb{E}[X_iX_j] = \sum_{(i,j)\in\mathcal{I}\times\mathcal{I}, H_i\sim H_j} p^{e(H_i)+e(H_j)-e(H_i\cap H_j)}$$

and any $0 < \gamma < 1$, we have

$$\mathbb{P}[X < (1-\gamma)\mu] \leq e^{-\frac{\gamma^2\mu^2}{2(\mu+\delta)}}.$$

6.1 Diversion: Upper and lower tails

Let $\frac{\log n}{n} \leq p \leq n^{-2/3}$ and let X be the number of copies of C_4 in $G(n,p)$. Let H_i, $i \in \mathcal{I}$, be the copies of C_4 in K_n and H some specific such copy of C_4 in K_n. Note that if $H_i \sim H$ then H_i must share either 1 or 2 edges with H. Using the notation in Theorem 41, we have $\mu = 3\binom{n}{4}p^4$ and, by symmetry,

$$\delta = 3\binom{n}{4}\sum_{i\in\mathcal{I}:H_i\sim H} p^{8-e(H_i\cap H)} = \mu\cdot p^4\cdot O(n^2p^{-1}+np^{-2}) = \mu\cdot O(n^2p^3+np^2) = \mu\cdot O(n^2p^3) = O(\mu).$$

Thus, by Theorem 41, we have, for each $\gamma \in (0,1)$, that

$$\mathbb{P}(X < (1-\gamma)\mu) \leq e^{-\Omega(\gamma^2\mu)}. \tag{2}$$

That is, the *lower tail* of the distribution of X is exponentially small in μ.

The *upper tail* is harder to analyse (but has been done by Vu [27]), and does not decrease as fast. To see this, keeping the same notation, let $\lambda = 2\sqrt[4]{\mu} = o(n)$ and pick disjoint sets $A, B \subset [n]$ with $|A| = |B| = \lambda$. The complete bipartite graph with vertex sets A and B, $K_n[A,B]$, contains $\lambda^4/4 = 4\mu$ copies of C_4. Thus,

$$\mathbb{P}(X \geq 4\mu) \geq \mathbb{P}(K_n[A,B] \subset G(n,p)) = p^{\lambda^2} = e^{-\sqrt{\mu}\log(1/p)} = e^{-o(\mu)},$$

where we have used that $p \geq \log n/n$. Thus, (2) cannot hold for the upper tail.

6.2 The chromatic number of $G(n,p)$.

Lemma 42. *Let $p \in (0,1)$ be constant and $b = \frac{1}{1-p}$. Let $k \in \mathbb{N}$ satisfy $k = 2\log_b n - \omega((\log\log n)^2)$. Then, with probability at least $1 - \exp(-\omega(n\log^3 n)$, $G(n,p)$ has an independent set with size k.*

Proof. First note that

$$\frac{n}{b^{k/2}} = \exp(\omega((\log\log n)^2)) = \omega(\log^{10} n),$$

and that we can assume that $k = \omega(\log n/\log\log n)$. Let H_i, $i \in \mathcal{I}$, be the independent k-sets in $\bar{K}_n$, say H is one, and let us use the notation in Theorem 41. Then,

$$\mu = \binom{n}{k}(1-p)^{\binom{k}{2}} \geq \left(\frac{n(1-p)^{\frac{k-1}{2}}}{k}\right)^k = \omega(\log^{8k} n) = \omega(n\log^3 n). \tag{3}$$

Furthermore,

$$
\begin{aligned}
\delta &= \binom{n}{k} \sum_{H_i \sim H} (1-p)^{k(k-1)-e(H_i \cap H)} \\
&\leq \frac{\mu^2}{\binom{n}{k}} \cdot \sum_{i=2}^{k-1} \binom{k}{i} \binom{n}{k-i} (1-p)^{-\binom{i}{2}} \\
&\leq \mu^2 \cdot \sum_{i=2}^{k-1} k^i \cdot \left(\frac{2k}{n}\right)^i (1-p)^{-\binom{i}{2}} \\
&\leq \mu^2 \cdot \frac{2k^2}{n} \cdot \sum_{i=2}^{k-1} \left(\frac{2k^2(1-p)^{-i/2}}{n}\right)^{i-1} \\
&= \mu^2 \cdot \frac{2k^2}{n} \cdot o\left(\sum_{i=2}^{k-1} \left(\log^{-6} n\right)^{i-1}\right) = o\left(\frac{\mu^2}{n \log^3 n}\right). \qquad (4)
\end{aligned}
$$

Thus, by (3), (4), and Theorem 41, we have $\mathbb{P}(X=0) \leq e^{-\omega(n \log^3 n)}$. □

Theorem 43. *Let $p \in (0,1)$ be constant and $b = \frac{1}{1-p}$. Almost surely,*

$$\chi(G(n,p)) = (1+o(1)) \cdot \frac{n}{2 \log_b n}. \qquad (5)$$

Note. This theorem was first proved by Bollobás, using a different technique as the result predated Theorem 41.

Proof of Theorem 43. Let $\ell = n/\log^2 n$ and $k = 2\log_b n - (\log\log n)^3 = 2\log_b \ell - \omega((\log\log \ell)^2)$. By Lemma 42, the probability there is some set with size ℓ in $G = G(n,p)$ with no independent set with size k is at most

$$2^n \cdot \exp(-\omega(\ell \log^3 \ell)) = o(1).$$

Then, iteratively remove maximal independent sets until fewer than ℓ vertices remain. This gives a colouring using at most

$$\frac{n}{k} + \ell = (1+o(1)) \cdot \frac{n}{2\log_b n}$$

colours. When combined with Corollary 40, this gives (5). □

7 Squared Hamilton cycles in random graphs

Definition. The *kth power of a graph H*, H^k is the graph with vertex set $V(H)$ and edge set $E(H) = \{uv : d_H(u,v) \leq k\}$, where $d_H(u,v)$ is the length of a shortest u,v-path in H if one exists.

It is known (as follows from a general theorem of Riordan [25]) that, for all $k \geq 3$, if $p = \omega(n^{-1/k})$, then, almost surely, $C_n^k \subsetneq G(n,p)$. We expect $\frac{n!}{2n} \cdot p^{2n}$ copies of C_n^2 in $G(n,p)$, so the threshold for $C_n^2 \subsetneq G(n,p)$ must be at least $n^{-1/2}$. This is its conjectured value, which remains open, while the best current bound is due to Nenadov and Škorić [23]. Using improved path connection techniques, we will show the following.

Theorem 44. *If $p = \omega\left(\frac{\log^2 n}{n}\right)$, then $G(n,p)$ almost surely contains a squared Hamilton cycle.*

7.1 Overview of the proof of Theorem 44

We will prove Theorem 44 using absorption and path connection techniques. Absorption was introduced as a general concept by Rödl, Ruciński and Szemerédi [26]. We will build up the squared path, initially with a special structure that allows us to *absorb* vertices to lengthen the path without altering its endpoints, as specified by the following definition.

Definition. A squared (a,b,c,d)-path $P \subset G$ is an *absorber for a vertex set* $X \subset V(G) \setminus V(P)$ *in* G if, for any $X' \subset X$ there is a squared (a,b,c,d)-path in G with vertex set $V(P) \cup X'$.

The set X is known as the reservoir. Using vertices from the reservoir we can then complete the squared cycle while covering all the unused vertices not in the reservoir. The unused vertices in the reservoir are then absorbed into the cycle to complete the squared Hamilton cycle.

We will seperate the absorbing and covering stage into the following two lemmas.

Lemma 45. *Let* $p = \omega\left(\frac{\log n}{\sqrt{n}}\right)$*. Almost surely,* $G = G(n,p)$ *contains a set* X *with size at least* $\frac{n}{100\log^2 n}$ *and a squared path* P *with* $V(P) \subset V(G) \setminus X$ *which is an absorber for* X *in* G.

For the second lemma, we use the following definition.

Definition. Let $u,v,x,y \in [n]$ be distinct vertices. We say a graph P is a (u,v,x,y)*-connector* with length $k+3$, if we can label its vertices as $\{u,v,w_1,\dots,w_k,x,y\}$ so that P is the square of the path $uvw_1\dots w_kxy$ with the edges uv and xy removed. We call $w_1,\dots,w_k$ the *interior vertices* of P.

Lemma 46. *Let* $X \subset [n]$ *satisfy* $|X| \geq n/100\log^2 n$*, and let* $a,b,c,d \in [n] \setminus X$ *be distinct. If* $p = \omega\left(\frac{\log^2 n}{n}\right)$*, then, almost surely,* $G(n,p)$ *contains an* (a,b,c,d)*-connector that covers* $[n] \setminus X$.

Proof of Theorem 44. Reveal edges in G within $[n]$ with probability $p/2$. Almost surely, by Lemma 45, G contains a set X with size at least $\frac{n}{100\log^2 n}$ and a squared path P with $V(P) \subset V(G) \setminus X$ which is an absorber for X in G. Say P is a squared (a,b,c,d)-path.

Reveal more edges in G with probability $p/2$. By Lemma 46, $G - (V(P) \setminus \{a,b,c,d\})$ contains a (b,a,d,c)-connector, Q say, which covers $[n] \setminus (V(P) \cup X)$. Let $X' = X \setminus V(Q)$, and let P' be a squared (a,b,c,d)-path with vertex set $V(P) \cup X'$ (which exists by the definition of an absorber). Combining Q and P' gives a squared Hamilton cycle. $\square$

For both the proof of Lemma 45 and Lemma 46, we will find connectors.

7.2 Finding connectors

We wish to find connectors with length around $\log n$ between $\Omega(n/\log n)$ many distinct pairs of vertex pairs. We will show that some connector is very likely to exist using Janson's inequality.

Lemma 47. *Let* $\ell = n/10$ *and let* u_i, v_i, x_i, y_i*,* $i \in [\ell]$*, be distinct vertices in* $[n]$*. Let* $p = \omega\left(\frac{\log n}{\sqrt{n}}\right)$*,* $G = G(n,p)$ *and* $k = \log n$*. Then, with probability at least* $1 - e^{-\omega(n\log n)}$*, there is for some* $i \in [\ell]$ *a* (u_i,v_i,x_i,y_i)*-connector in* G *with length* $k+3$ *and interior vertices in* $V := [n] \setminus \{u_i,v_i,x_i,y_i : i \in [\ell]\}$*.*

Proof. Note that we can assume that $p = o(\frac{\log^2 n}{\sqrt{n}})$. Let P be a (u_1,v_1,x_1,y_1)-connector with length $k+3$ and interior vertices in V. Let $U = \{u_1,v_1,x_1,y_1\}$. Let $\mathcal{H}$ be the set of copies of P

in K_n with interior vertices in V making a (u_i, v_i, x_i, y_i)-connector for some $i \in [\ell]$. Note that

$$n^{k+1} \geq |\mathcal{H}| = \ell \cdot |V|(|V|-1)\dots(|V|-k+1) \geq \ell \cdot |V|^k \cdot \left(1 - \frac{2}{n}\right)^k = \Omega(n^{k+1}). \tag{6}$$

We will use Theorem 41 to bound the probability no graph in $\mathcal{H}$ appears in $G(n,p)$. Let X be the number of graphs in $\mathcal{H}$ appearing in $G(n,p)$ and let $\mu = \mathbb{E}X$. Then, by (6),

$$\mu = \Omega(|\mathcal{H}| \cdot p^{e(P)}) = \Omega(n^{k+1}p^{2k+3}) = \Omega(p \cdot (np^2)^{k+1}) = \omega(n \log n). \tag{7}$$

For each $H, H' \in \mathcal{H}$, write $H \sim H'$ if $e(H \cap H) > 0$, and let

$$\delta = \sum_{H,H' \in \mathcal{H}: H \sim H'} p^{e(H)+e(H')-e(H \cap H')} = |\mathcal{H}| \sum_{H \in \mathcal{H}: H \sim P} p^{2e(P)-e(H \cap P)},$$

where we have used symmetry.

Using some subsequent claims, we will show that $\delta = o\left(\frac{\mu^2}{n \log n}\right)$. Whence, with (7) and Theorem 41, the claim in the lemma will follow.

In order to bound δ, we divide into cases depending on whether $H \in \mathcal{H}$ with $H \sim P$ is a (u_1, v_1, x_1, y_1)-connector (and so, like P, contains U) or not. For this, let

$$\mathcal{J}_1 = \{P \cap H : H \in \mathcal{H}, e(P \cap H) > 0, U \cap V(J) = \emptyset\}$$

and

$$\mathcal{J}_2 = \{P \cap H : H \in \mathcal{H}, e(P \cap H) > 0, U \subset V(J)\}.$$

First we shall calculate, for each possible intersection J, a bound on how many graphs in $\mathcal{H}$ intersect with P in J.

Claim A1. For each $J \in \mathcal{J}_1$, $|\{H \in \mathcal{H} : H \cap P = J\}| \leq 4k^{|J|-1}n^{k+1-|J|}$.

Proof of Claim A1. To choose $H \in \mathcal{H}$ with $H \cap P = J$, we can choose where the copy of J appears in H, choose $i \in [\ell]$ ($\ell \leq n$ choices), and then choose the remaining vertices in H (at most $n^{k-|J|}$ choices).

Recall that $e(J) \geq 1$, choosing where J appears in H can be done by choosing the location of some edge (at most $e(P - U) \leq 4k$ choices), and then the location of the other vertices in J (at most $k^{|J|-2}$ choices). Thus, $|\{H \in \mathcal{H} : H \cap P = J\}| \leq 4k \cdot k^{|J|-2} \cdot \ell \cdot n^{k-|J|}$. □

Claim A2. For each $J \in \mathcal{J}_2$, $|\{H \in \mathcal{H} : H \cap P = J\}| \leq 4k^{|J-U|-1}n^{k-|J-U|}$.

Proof of Claim A2. Note that if $e(J - U) > 0$, then the reasoning for Claim A1, noting that the choice of $i \in [\ell]$ is fixed at $i = 1$, implies the result.

If $e(J - U) = 0$, then J must contain some edge incident to U, and therefore there are at most 2 choices for its other vertex in the copy of H. Thus, there are at most $2k^{|J-U|-1}$ ways J could be embedded in H. Along with at most $n^{k-|J-U|}$ choices for the other vertices of H, this gives the result. □

Next, we shall prove bounds on the number of edges in the intersection of graphs with P.

Claim B1. For each $J \in \mathcal{J}_1$, $e(J) \leq 2(|J| - 1) - 1$, so that

$$p^{-e(J)} \leq p \cdot \left(\frac{1}{p^2}\right)^{|J|-1}.$$

Proof of Claim B1. Consider the natural ordering of P beginning a, b and ending c, d. Count the edges in J going to the left from each vertex in J using this order. The leftmost two vertices between them have at most 1 edge going to the left, while every other vertex has at most 2 edges going to the left. Thus, $e(J) \leq 2(|J|-2)+1 = 2(|J|-1)-1$, as required. □

Claim B2. For each $J \in \mathcal{J}_2$,

$$p^{-e(J-U)} \leq \left(\frac{10}{p^2}\right)^{|J-U|}. \tag{8}$$

Proof of Claim B2. If two consecutive vertices in P are not in J, then counting leftwards edges and rightwards edges from that point, we have $e(J-U) \leq 2|J-U|$, so that (8) holds.

If this does not happen and $|J-U| \leq k-3$, then counting along leftward edges we have $e(J-U) \leq 2|J-U|$, so that (8) holds.

Finally, if $|J-U| \geq k-2$, then counting along leftward edges we have $e(J-U) \leq 2|J-U|+3$. As $10^{k-2} = \omega(n^2) = \omega(p^{-3})$, so that (8) holds. □

Finally, we will prove a simple bound on the number of possible intersections $J \in \mathcal{J}_1$ and $J \in \mathcal{J}_2$ of each order.

Claim C1. For each $2 \leq i \leq k$, $|\{J \in \mathcal{J}_1 : |J| = i\}| \leq 4k^{i-1}4^i$.

Proof of Claim C1. Pick $V \subset V(P)$ with $|V| = i$ with $e(P[V]) > 0$ by choosing an edge (in at most $2k$ ways) and then choosing the remaining $i-2$ vertices (in at most k^{i-2} ways). As $P[V]$ has at most $2i$ edges, there can be at most 2^{2i} subgraphs of $P[V]$ with vertex set V. Thus, $|\{J \in \mathcal{J}_1 : |J| = i\}| \leq 4k^{i-1}4^i$, as required. □

Claim C2. For each $1 \leq i \leq k$, $|\{J \in \mathcal{J}_2 : |J-U| = i\}| \leq 8k^{i-1}4^i$.

Proof of Claim C2. For each $1 \leq i \leq k$, $|\{J \in \mathcal{J}_2 : |J-U| = i, e(J-U) > 0\}| \leq 4k^{i-1}4^i$, similarly to the proof of Claim C1. For each $J \in \mathcal{J}_2$ with $e(J-U) = 0$, as $e(J) > 0$, $V(J)$ must contains some vertex in $N_P(U)$. Thus, $|\{J \in \mathcal{J}_2 : |J-U| = i, e(J-U) = 0\}| \leq 4k^{i-1}2^{2i}$, and the claim holds. □

We can now bound δ, where we use, for example, A to indicate where the Claims A1 and A2 are applied. This gives us

$$\begin{aligned}
\delta &= |\mathcal{H}| \sum_{H \in \mathcal{H} : E(H \cap P) \neq \emptyset} p^{2e(P)-e(H\cap P)} \\
&= \mu p^{e(P)} \cdot \left(\sum_{J \in \mathcal{J}_1} \sum_{H \in \mathcal{H} : H \cap P = J} p^{-e(J)} + \sum_{J \in \mathcal{J}_2} \sum_{H \in \mathcal{H} : H \cap P = J} p^{-e(J)} \right) \\
&\overset{A}{\leq} \mu p^{e(P)} \cdot \left(\sum_{J \in \mathcal{J}_1} n^{k+1-|J|} \cdot 4k^{|J|-1} \cdot p^{-e(J)} + \sum_{J \in \mathcal{J}_2} n^{k-|J-U|} \cdot 4k^{|J-U|-1} \cdot p^{-e(J)} \right) \\
&= O\left(\frac{\mu^2}{n} \cdot \left(\sum_{J \in \mathcal{J}_1} n^{1-|J|} \cdot k^{|J|-1} \cdot p^{-e(J)} + \frac{1}{k} \sum_{J \in \mathcal{J}_2} n^{-|J-U|} \cdot k^{|J-U|} \cdot p^{-e(J)} \right) \right) \\
&\overset{B}{=} O\left(\frac{\mu^2}{n} \cdot \left(\sum_{J \in \mathcal{J}_1} \left(\frac{k}{np^2}\right)^{|J|-1} \cdot p + \frac{1}{k} \sum_{J \in \mathcal{J}_2} \left(\frac{10k}{np^2}\right)^{|J-U|} \right) \right)
\end{aligned}$$

$$\overset{C}{=} O\left(\frac{\mu^2}{n}\cdot\left(p\cdot\sum_{i=2}^{k}\left(\frac{k}{np^2}\right)^{i-1}\cdot 4k^{i-1}\cdot 4^i+\frac{1}{k}\cdot\sum_{i=1}^{k}\left(\frac{10k}{np^2}\right)^{i}\cdot 8k^{i-1}\cdot 4^i\right)\right)$$
$$= O\left(\frac{\mu^2}{n}\cdot\left(p\cdot\sum_{i=2}^{k}\left(\frac{4k^2}{np^2}\right)^{i-1}+\frac{1}{k}\sum_{i=1}^{k}\left(\frac{40k^2}{np^2}\right)^{i}\right)\right).$$

Thus, as $np^2=\omega(k^2)$,

$$\delta = o\left(\frac{\mu^2}{n}\cdot\left(p+\frac{1}{k}\right)\right)=o\left(\frac{\mu^2}{n\log n}\right).$$

Therefore, by this, (7) and Theorem 41, the probability some graph in $\mathcal{H}$ exists in $G(n,p)$ is at least $1-\exp(-\omega(n\log n))$. □

In $G(n,p)$ with $p=\omega\left(\frac{\log n}{\sqrt{n}}\right)$, we almost surely have the following property. Given any linearly sized subset of 'unused' vertices, revealing more edges is likely to connect any given pair of vertex pairs. That is, we have the following.

Corollary 48. *If $p=\omega\left(\frac{\log n}{\sqrt{n}}\right)$ and $k=\log n$, then the following is almost surely true in $G=G(n,p)$ for any $A\subset[n]$ with $|A|\geq n/2$.*

- ***P*** *Given any distinct vertices $u,v,x,y\in[n]\setminus A$, if more edges are revealed between $\{u,v,x,y\}$ and A with probability p, forming the edge set E, then, with probability at least $1-n^{-2}$, there is a (u,v,x,y)-connector in $G+E$ with length $k+3$ and interior vertices in A.*

Proof. We will show that, for large n and any set $A\subset[n]$ with $|A|\geq n/2$, **P** does not hold with probability $\exp(-\omega(n\log n))$. As there are certainly at most 2^n such sets A, the lemma follows by a union bound.

Fix $A\subset[n]$ with $|A|\geq n/2$ and let q be the probability that **P** does not hold for A. Reveal edges in G with probability p, where $V(G)=[n]$, and suppose that **P** does not hold.

Let $\ell=n/10$ and take disjoint vertices u_i,v_i,x_i,y_i, $i\in[\ell]$ in $[n]\setminus A$. Reveal edges between A and $[n]\setminus A$ with probability p. As **P** does not hold, for each $i\in[\ell]$ the probability there is no (u_i,v_i,x_i,y_i)-connector with length $k+3$ and interior vertices in A is at least n^{-2}. Furthermore the that this happens is independent for each $i\in[\ell]$, and each possible edge has been revealed with probability at least p.

Therefore, in $G(n,p)$, the probability that there is, for each $i\in[\ell]$, no (u_i,v_i,x_i,y_i)-connector with length $k+3$ and interior vertices in A is at least $q\cdot n^{-2\ell}=q\cdot\exp(-O(n\log n))$. But, by Lemma 47, this probability is at most $\exp(-\omega(n\log n))$, so that we must therefore have $q=\exp(-\omega(n\log n))$, as required. □

We can use Corollary 48 to reveal edges and apply **P** iteratively to find many connectors.

Corollary 49. *Let $\ell=n/5\log n$, and let u_i,v_i,x_i,y_i, $i\in[\ell]$, be distinct vertices in $[n]$. Let $k=\log n$ and $p=\omega\left(\frac{\log n}{\sqrt{n}}\right)$. Then, almost surely, in $G(n,2p)$ there are a set of disjoint (u_i,v_i,x_i,y_i)-connectors, $i\in[\ell]$, with length $k+3$.*

Proof. Let $G_0=G(n,p)$ and $V:=[n]\setminus\{u_i,v_i,x_i,y_i : i\in[\ell]\}$. Almost surely, by Corollary 48, each $A\subset[n]$ with $|A|\geq n/2$ satisfies **P**. We will, almost surely, iteratively find disjoint (u_i,v_i,x_i,y_i)-connectors P_i, $i\in[\ell]$, with length $k+3$ and interior vertices in $V\setminus(\cup_{j<i}V(P_j))$.

For each i, $1\leq i\leq\ell$, reveal edges with probability p between $\{u_i,v_i,x_i,y_i\}$ and V and add them to G_{i-1} to get G_i. By **P**, and as $|V\setminus(\cup_{j<i}V(P_j))|\geq n-(k+4)\ell\geq n/2$, with probability

at least $1 - n^{-2}$, we can find a (u_i, v_i, x_i, y_i)-connector P_i with length $k+3$ and interior vertices in $V \setminus (\cup_{j<i} V(P_j))$. Thus, with probability at least $1 - \ell n^{-2} = 1 - o(1)$ we will find a disjoint collection of connectors as desired.

Note that we have revealed each possible edge with probability at most $2p$, so that we may compare favourably the resulting random graph with $G(n, 2p)$. □

7.3 Finding a covering connector

To prove Lemma 46, we first find a square path which covers most of the vertices in $[n] \setminus (X \cup \{a, b, c, d\})$, using the following result.

Lemma 50. *For all $p = p(n)$, $G(n, p)$ almost surely contains a squared path covering all but at most $\frac{10 \log n}{p^2}$ vertices.*

Proof. Let $\ell = \frac{10 \log n}{p^2}$ and note we can assume $p \geq 1/\sqrt{n}$. Pick a vertex $v_0 \in [n]$ and reveal the edges between v_0 and $[n] \setminus \{v_0\}$ with probability $p/2$ to get the graph G. Almost surely, v_0 has some neighbour, v_1 say. Then, for each $2 \leq i \leq n - \ell$, reveal more edges in G between $\{v_{i-1}, v_{i-2}\}$ and $[n] \setminus \{v_0, v_1, \dots, v_{i-1}\}$ with probability $p/2$. With probability at least $1 - (1 - (p/2)^2)^\ell \geq 1 - e^{-p^2 \ell/4} = 1 - o(n^{-1})$, there is some neighbour, v_i say, of v_{i-1} and v_{i-2} in $[n] \setminus \{v_0, v_1, \dots, v_{i-1}\}$. Thus, this almost surely finds a squared path with at least $n - \ell$ vertices in G. As each possible edge within $[n]$ has been revealed with probability at most p, this gives the result of the lemma. □

Proof of Lemma 46. By removing vertices from X if necessary, assume that $|X| \leq n/2$. Let $G_0 = G(n, p/4)$. Almost surely, by Lemma 50, there is a squared path in $G_0 - (X \cup \{a, b, c, d\})$ covering all but at most $250 \log n/p^2 = o(n/\log^3 n)$ vertices, say this path is P_0 and is a squared (x, y, u_0, v_0)-path. Let $V = [n] \setminus (X \cup V(P_0))$ and $\ell = |V| \leq= o(n/\log^3 n)$.

Reveal more edges with probability $p/4$ and find a matching from V into X (using, for example, an iterative argument), and label its edges $u_i v_i$, $i \in [\ell]$.

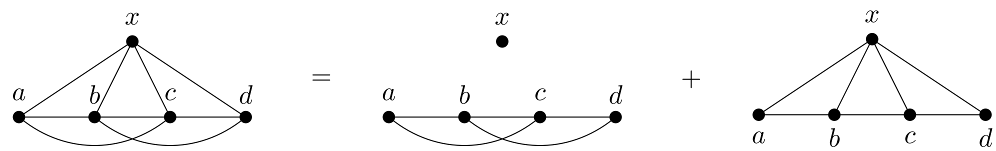
Let $X' = X \setminus \{u_i, v_i : i \in [\ell]\}$, so that $|X'| = \omega(\ell \log n)$ and $p = \omega(\frac{\log n}{\sqrt{|X'|}})$. Revealing more edges in two rounds with probability $p/4$, and applying Corollary 49 to vertex-disjoint pairs of vertex pairs, we can find the following connectors which are internally vertex disjoint:

- an (a, b, x, y)-connector,
- for each $i \in [\ell]$, a $(u_{i-1}, v_{i-1}, u_i, v_i)$-connector, and
- a (u_ℓ, v_ℓ, c, d)-connector.

Combining these connectors with the squared (x, y, u_0, v_0)-path P_0 and the edges $u_i v_i$, $i \in [\ell]$, we get an (a, b, c, d)-connector covering $[n] \setminus X$. □

7.4 Diversion: Realistic ambitions for constructing absorbers

We will create an absorber for a single vertex, before joining together many such absorbers with connectors to create an absorber for a large set. The following (a, b, c, d)-path is an absorber for x:

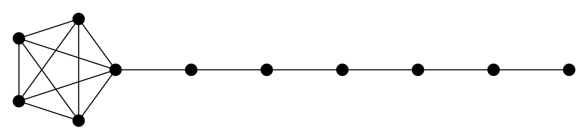
$H :=$ 

The graph H has 9 edges, so $\mathbb{E}|\{\text{copies of } H \text{ in } G(n,p)\}| \le n^5p^9$. Thus, if, for example $p = \frac{\log^2 n}{\sqrt{n}}$, then we can expect only $o(n^{2/3})$ copies of H in $G(n,p)$, which would give rise to a reservoir set X with size $o(n^{2/3})$.

We want instead an absorber for which we can find disjoint copies covering $\Omega(n)$ vertices. This is the same as finding a copy of the graph created by the disjoint union of different copies of the absorber. When in general is there likely to be a copy of a graph H in $G(n,p)$? As before, we can look to the expected number of copies of H, and deduce that certainly we will need

$$n^{|H|}p^{e(H)} = \Omega(1).$$

Even for graphs H with fixed size, this is not the whole story. For example, the expected number of copies of the following graph in $G\big(n, \frac{1}{n^{2/3}}\big)$ is $\omega(1)$:

$K :$ 

However, $G\big(n, \frac{1}{n^{2/3}}\big)$ almost surely contains no copy of K_5, so in fact almost surely contains no such copy of K. The problem is that K is not *balanced*, as it has a subgraph denser than the overall graph, so we need to consider the following quantity:

$$m_0(H) = \max\Big\{\frac{e(H')}{|H'|} : H' \subset H\Big\}.$$

Exercise. For any fixed graph H with at least 1 edge, prove that $p = n^{-1/m_0(H)}$ is a threshold for $G(n,p)$ to contain a copy of H. (This was originally shown for some graphs by Erdős and Rényi [11] and all graphs by Bollobás [4])

When finding absorbers we wish to find vertex disjoint copies of some graph H in $G(n,p)$. An *H-factor* is $\lfloor n/|H| \rfloor$ vertex disjoint copies of H in $G(n,p)$. We wish to find an *almost H-factor*, one covering $\Omega(n)$ vertices, so we will certainly need $\mathbb{E}|\{\text{copies of } H' \text{ in } G(n,p)\}| = \Omega(n)$ for each $H' \subset H$. Define then the following quantity:

$$m_1(H) = \max\Big\{\frac{e(H')}{|H'|-1} : H' \subset H, |H'| \ge 2\Big\}.$$

Exercise. For any fixed graph H with at least 1 edge, prove that $p = n^{-1/m_1(H)}$ is a threshold for $G(n,p)$ to contain vertex disjoint copies of H covering at least $n/2$ vertices.

As we have seen (for the Hamilton cycle), looking at the expectation of different subgraphs does not tell the whole story. Many general H-factors also demonstrate this behaviour. For example, in a remarkable paper, Johansson, Kahn and Vu [17] proved that for graphs H where

$e(H')/(|H'|-1)$ obtains a unique maximum over all $H' \subset H$ and $|H'| \geq 2$ when $H' = H$, the threshold for an H-factor in $G(n,p)$ is the same as the threshold in $G(n,p)$ for each vertex to be contained in some copy of H.

Exercise. For each graph H, find the threshold in $G(n,p)$ for each vertex to be contained in some copy of H.

Exercise. Give an example of a graph H in which the threshold in $G(n,p)$ for each vertex to be contained in some copy of H is not a threshold in $G(n,p)$ for an H-factor.

Ideal vertex absorbers. Say we have an absorber P for x and P_x is a squared path with same endpoint pairs as P and vertex set $V(P) \cup \{x\}$. We wish to find such a P with as few vertices as possible so that we may still cover $\Omega(n)$ vertices with disjoint copies of P. We will find such a P with $O(\log^2 n)$ vertices, so to see whether we can find a smaller such P suppose that $|P| = O(\log^2 n)$ and we want to find $\Theta(n/\log^2 n)$ disjoint copies of $P \cup P_x$.

Suppose $p = \frac{\log^2 n}{\sqrt{n}}$, i.e., a fairly low edge probability that we still need to consider for Lemma 45. For each $H \subset P \cup P_x$ with $e(H) \geq 1$ we need

$$n^{|H|}p^{e(H)} = \Omega(n/\log^2 n)$$

so that, as $p = n^{-1/2+O(\log\log n/\log n)}$,

$$n^{|H|-1}n^{-(1/2-O(\log\log n/\log n))e(H)} = \Omega(1).$$

Thus we need $m_1(P \cup P_x) \leq 1/(1/2 - O(\log\log n/\log n) = 2 + O(\log\log n/\log n)$.

Now, to simplify the argument, let us work with paths rather than squares of paths. Here, we consider an (a,b)-path P for a vertex x and an a,b-path P_x with vertex set $V(P) \cup \{x\}$, and wish to have $m_1(P \cup P_x) \leq 1 + O(\log\log n/\log n)$ (so that it is plausible we can find $n/2|P|$ disjoint copies of $P \cup P_x$ in $G\big(n, \frac{\log^2 n}{n}\big)$).

Let H be the graph with edge set $E(P)\triangle E(P_x)$ and no isolated vertices. As $x \notin V(P)$, $x \in V(H)$ and $d_H(x) = 2$. Note that every vertex in H must in fact have at least 2 neighbours in H. Thus, $e(H) \geq |H|$, so that $m_1(P \cup P_x) \geq e(H)/(|H|-1) \geq |H|/(|H|-1) = 1 + O(1/|H|)$. Hence, $e(H) \geq |H| = \Omega(\log n/\log\log n)$.

Thus, $e(P \cup P_x) \geq |P| + |E(P)\triangle E(P_x)|/2 = |P| + \Omega(\log n/\log\log n)$, so that

$$m_1(P \cup P_x) \geq \frac{e(P \cup P_x)}{|P|-1} \geq 1 + \Omega\Big(\frac{\log n}{|P| \cdot \log\log n}\Big).$$

Thus, we must have $|P| = \Omega(\log^2 n/(\log\log n)^2)$.

Similar, but more involved arguments, show that for squared paths the best we can do is find an absorber for a vertex with $\Omega(\log^2 n/(\log\log n)^2)$ vertices. We will get close to this, constructing an absorber with $O(\log^2 n)$ vertices – the difference being for ease of calculation rather than a fundamental restriction of the construction.

7.5 Constructing absorbers

To find an absorber for a large set X we will construct an absorber for each $x \in X$ before connecting them using Corollary 49. Each absorber will be constructed from a subgraph we call a gadget, many copies of which will exist in our random graph, as shown by Janson's inequality.

Definition. As depicted in Figure 1, a *k-gadget for x in G* is a $2k$-vertex subgraph H of $G - \{x\}$, with vertex set $\{u_i, v_i : i \in [k]\}$ say, so that

- for each $1 \leq i \leq k-1$, $u_i v_{i+1}, v_i u_{i+1} \in E(H)$,
- for each $1 \leq i \leq k$, $u_i v_i \in E(H)$,
- $u_1 u_k \in E(H)$ and $xu_1, xv_1, xu_k, xv_k \in E(G)$.

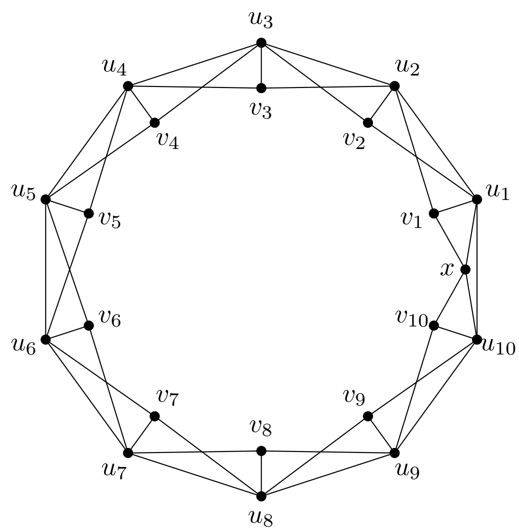


Figure 1: A k-gadget for x, with $k = 10$.

Exercise. Show that if $k = \log n$ and H is a k-gadget, then $m_1(H) = 2 + O(\log\log n/\log n)$.

Lemma 51. *Let $k = \log n$ and $\ell = n/10k$. If $p = \omega(\frac{\log n}{n})$, then, almost surely, $G = G(n,p)$ contains, disjointly, vertices x_i, $i \in [\ell]$, and subgraphs H_i, $i \in [\ell]$, such that, for each $i \in [\ell]$, H_i is a $(4k+2)$-gadget for x_i in G.*

Exercise. Prove Lemma 51 using an application of Janson's inequality (see also the proof of Lemma 47).

Lemma 52. *Suppose H with vertex set $\{u_i, v_i : i \in [4k+2]\}$ is a $(4k+2)$*-gadget for x in *G as depicted in Figure 1. Suppose P_i, $1 \leq i \leq 2k$, is a disjoint collection of connectors in $G - x$ so that, if i is odd, P_i is a $(u_{i+2k+1}, v_{i+2k+1}, v_i, u_i)$-connector and, if i is even, P_i is a $(u_i, v_i, v_{i+2k+1}, u_{i+2k+1})$-connector. Then, the following squared $(v_{2k+1}, u_{2k+1}, u_{4k+2}, v_{4k+2})$-path P, as depicted in Figure 2, is an absorber for x.*

$$v_{2k+1}u_{2k+1}P_1P_2P_3P_4\dots P_{2k-1}P_{2k}u_{4k+2}v_{4k+2}.$$

Proof. As depicted in Figure 3, the following is a squared $(v_{2k+1}, u_{2k+1}, u_{4k+2}, v_{4k+2})$-path with vertex set $V(P) \cup \{x\}$.

$$v_{2k+1}u_{2k+1}P_{2k}P_{2k-1}P_{2k-2}P_{2k-3}\dots P_2P_1xu_{2k}v_{2k}. \qquad \square$$

We can now find disjoint gadgets and connect them together as in Lemma 52 to create a squared path capable of absorbing any subset of a large set X, which proves Lemma 45.

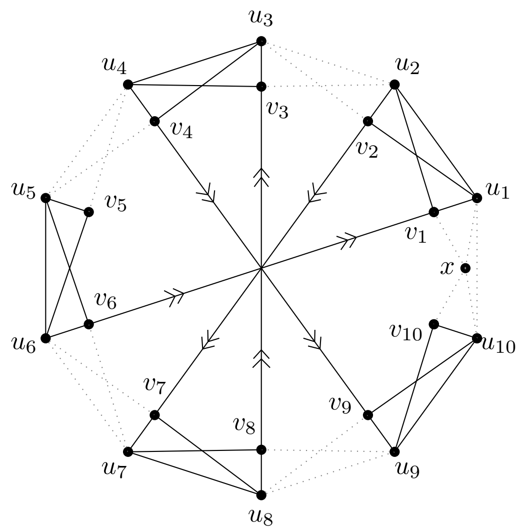


Figure 2: A $(v_5, u_5, u_{10}, v_{10})$-squared path P which is an absorber for $\{x\}$. The dotted edges exist in the parent graph to allow the path to absorb x.

Proof of Lemma 45. Let $\ell = n/100\log^2 n$ and $k = \log n$. By Lemma 51, $G_0 = G(n, p/2)$ almost surely contains disjointly vertices x_i, $i \in [\ell]$, and subgraphs H_i, $i \in [\ell]$, such that H_i is a $(4k+2)$-gadget for x_i in G_0. For each $i \in [\ell]$, label $V(H_i)$ as $\{u_{i,1}, v_{i,1}, \ldots, u_{i,4k+2}, v_{i,4k+2}\}$, in the manner of Figure 1. Let $X = \{x_1, \ldots, x_\ell\}$.

Note that $\ell(2k+1) \leq n/10k$. Let $G_1 = G(n, p/2)$ and $G = G_0 \cup G_1$. We will find the required absorbing squared path almost surely in G. Almost surely, by Corollary 49, $G_1 - X$ contains disjointly the following connectors.

- For each $i \in [\ell]$ and $j \in [2k]$ a $(u_{i,j+2k+1}, v_{i,j+2k+1}, v_{i,j}, u_{i,j})$-connector, $Q_{i,j}$ say, if j is odd, and a $(u_{i,j}, v_{i,j}, v_{i,j+2k+1}, u_{i,j+2k+1})$-connector, $Q_{i,j}$ say, if j is even, and
- for each $i \in [\ell - 1]$ a $(u_{i,4k+2}, v_{i,4k+2}, v_{i+1,2k+1}, u_{i+1,2k+1})$-connector P_i.

By Lemma 52, for each $i \in [\ell]$, the following squared $(v_{i,2k+1}, u_{i,2k+1}, u_{i,4k+2}, v_{i,4k+2})$-path Q_i, is an absorber for x_i.

$$v_{i,2k+1}u_{i,2k+1}Q_{i,1}Q_{i,2}\ldots Q_{i,2k-1}Q_{i,2k}u_{i,4k+2}v_{i,4k+2}$$

For each $i \in [\ell]$, let $\hat{Q}_i$ be a squared $(v_{i,2k+1}, u_{i,2k+1}, u_{i,4k+2}, v_{i,4k+2})$-path with vertex set $V(Q_i) \cup \{x_i\}$, which exists as Q_i is an absorber for x_i.

Then the following squared $(v_{1,2k+1}, u_{1,2k+1}, u_{\ell,4k+2}, v_{\ell,4k+2})$-path is an absorber for X.

$$Q_1P_1Q_2P_2\ldots Q_{\ell-1}P_{\ell-1}Q_\ell$$

Indeed, for any $X' \subset X$, for each $i \in [\ell]$, let $Q_i' = Q_i$ if $x_i \notin X'$ and $Q_i' = \hat{Q}_i$ if $x_i \in X'$, so that

$$Q_1'P_1Q_2'P_2\ldots Q_{\ell-1}'P_{\ell-1}Q_\ell'$$

is a squared $(v_{1,2k+1}, u_{1,2k+1}, u_{\ell,4k+2}, v_{\ell,4k+2})$-path with vertex set $V(P) \cup X'$. □

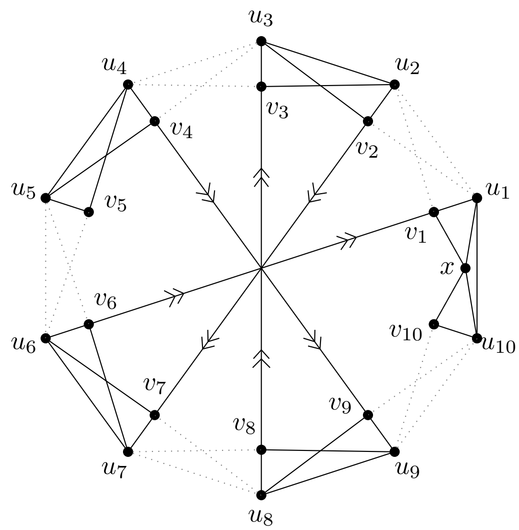


Figure 3: The $(v_5, u_5, u_{10}, v_{10})$-squared path P with x absorbed.

References

[1] M. Ajtai, J. Komlós, and E. Szemerédi. The longest path in a random graph. *Combinatorica*, 1(1):1–12, 1981.

[2] M. Ajtai, J. Komlós, and E. Szemerédi. First occurrence of Hamilton cycles in random graphs. *North-Holland Mathematics Studies*, 115(C):173–178, 1985.

[3] N. Alon and J.H. Spencer. **The probabilistic method**. John Wiley & Sons, 2004.

[4] B. Bollobás. Random graphs. *Proceedings of the Eigth Combinatorial Conference, University College*, pages 41–52, 1981.

[5] B. Bollobás. The evolution of random graphs. *Transactions of the American Mathematical Society*, 286(1):257–274, 1984.

[6] B. Bollobás. The evolution of sparse graphs, Graph Theory and Combinatorics (Cambridge 1983), 35-57, 1984.

[7] B. Bollobás. **Random graphs**. Cambridge University Press, 2001.

[8] B. Bollobás and A.G. Thomason. Threshold functions. *Combinatorica*, 7(1):35–38, 1987.

[9] W.F. de la Vega. Long paths in random graphs. *Studia Sci. Math. Hungar*, 14(4):335–340, 1979.

[10] P. Erdős and A. Rényi. On random graphs I. *Publicationes Mathematicae (Debrecen)*, 6:290–297, 1959.

[11] P. Erdős and A. Rényi. On the evolution of random graphs. *Publ. Math. Inst. Hung. Acad. Sci*, 5:17–61, 1960.

[12] E. Friedgut. Sharp thresholds of graph properties, and the k-sat problem. *Journal of the American mathematical Society*, 12(4):1017–1054, 1999.

[13] A. Frieze and M. Karoński. **Introduction to random graphs**. Cambridge University Press, 2015.

[14] A.M. Frieze. On large matchings and cycles in sparse random graphs. *Discrete Mathematics*, 59(3):243–256, 1986.

[15] A.M. Frieze. An algorithm for finding Hamilton cycles in random directed graphs. *Journal of Algorithms*, 9(2):181–204, 1988.

[16] S. Janson, T. Luczak, and A. Rucinski. **Random graphs**. John Wiley & Sons, 2011.

[17] A. Johansson, J. Kahn, and V. Vu. Factors in random graphs. *Random Structures & Algorithms*, 33(1):1–28, 2008.

[18] J. Komlós and E. Szemerédi. Limit distribution for the existence of Hamiltonian cycles in a random graph. *Discrete Mathematics*, 43(1):55–63, 1983.

[19] M. Krivelevich, C. Lee, and B. Sudakov. Long paths and cycles in random subgraphs of graphs with large minimum degree. *Random Structures & Algorithms*, 46(2):320–345, 2015.

[20] M. Krivelevich, E. Lubetzky, and B. Sudakov. Cores of random graphs are born Hamiltonian. *Proceedings of the London Mathematical Society*, 2014.

[21] C. Lee and B. Sudakov. Dirac's theorem for random graphs. *Random Structures & Algorithms*, 41(3):293–305, 2012.

[22] C. McDiarmid. Clutter percolation and random graphs. In *Combinatorial Optimization II*, pages 17–25. Springer, 1980.

[23] R. Nenadov and N. Škoric. Powers of cycles in random graphs and hypergraphs. *arXiv preprint arxiv:1601.04034*, 2016.

[24] L. Pósa. Hamiltonian circuits in random graphs. *Discrete Mathematics*, 14(4):359–364, 1976.

[25] O. Riordan. Spanning subgraphs of random graphs. *Combinatorics, Probability and Computing*, 9(2):125–148, 2000.

[26] V. Rödl, A. Ruciński, and E. Szemerédi. A Dirac-type theorem for 3-uniform hypergraphs. *Combinatorics, Probability and Computing*, 15(1-2):229–251, 2006.

[27] V.H. Vu. A large deviation result on the number of small subgraphs of a random graph. *Combinatorics, Probability and Computing*, 10(1):79–94, 2001.