

MINORS IN RANDOM REGULAR GRAPHS

NIKOLAOS FOUNTOULAKIS, DANIELA KÜHN AND DERYK OSTHUS

ABSTRACT. We show that there is a constant c so that for fixed $r \geq 3$ a.a.s. an r -regular graph on n vertices contains a complete graph on $c\sqrt{n}$ vertices as a minor. This confirms a conjecture of Markström [17]. Since any minor of an r -regular graph on n vertices has at most $rn/2$ edges, our bound is clearly best possible up to the value of the constant c . As a corollary, we also obtain the likely order of magnitude of the largest complete minor in a random graph $G_{n,p}$ during the phase transition (i.e. when $pn \rightarrow 1$).

1. INTRODUCTION

We say that a graph G contains a complete graph on k vertices (denoted by K_k) as a minor if we can obtain a copy of K_k after a series of contractions of the edges and deletions of vertices or edges of G . We write $K_k \prec G$ in this case. Equivalently, G has a K_k minor if there are k pairwise disjoint non-empty subsets of $V(G)$ (which we call *branch sets*) such that each of them is connected and any two of them are joined by an edge. The *contraction clique number* $\text{ccl}(G)$ of G is the largest integer k such that G has a K_k minor.

Originally, the study of the order of the largest complete minor in a random graph was motivated by Hadwiger's conjecture which states that $\text{ccl}(G) \geq \chi(G)$ for any graph G . Bollobás, Erdős and Catlin [7] showed that the proportion of graphs on n vertices that satisfy Hadwiger's conjecture tends to 1 as n tends to infinity. For this, they determined the likely value of $\text{ccl}(G_{n,p})$ for the random graph $G_{n,p}$ with constant edge probability p and compared this with known results on $\chi(G_{n,p})$. Krivelevich and Sudakov [13] investigated $\text{ccl}(G)$ for expanding graphs G and derived the order of magnitude of $\text{ccl}(G_{n,p})$ from their results when $p = n^{-c}$, with $c < 1$. In [9], we extended these results to any p with $pn \geq c$ for some constant $c > 1$, which answered a question from [13]. In particular, we showed that if $pn = c$ for some fixed $c > 1$ then a.a.s.

$$(1) \quad \text{ccl}(G_{n,p}) = \Theta(\sqrt{n}).$$

The upper bound is immediate, as for such p a.a.s. the random graph $G_{n,p}$ has $\Theta(n)$ edges and no minor of a graph G can contain more edges than G itself. Here we write that an event regarding a graph on n vertices holds a.a.s. if the probability of this event tends to 1 as n tends to infinity.

Markström [17] had earlier conjectured a similar phenomenon as in (1) for the case of random regular graphs. For any $r \geq 3$ and $n \geq 4$ such that rn is even, we denote by $G(n, r)$ a graph chosen uniformly at random from the set of r -regular simple graphs on n vertices. Throughout, we consider the case where r is fixed. The number of edges of $G(n, r)$ is $rn/2$ and so the same argument as above shows that $\text{ccl}(G(n, r)) \leq 2\sqrt{rn}$. However the lower bound in (1) does not imply that a random r -regular graph satisfies $\text{ccl}(G(n, r)) = \Omega(\sqrt{n})$ as the asymptotic structure of $G(n, r)$ is quite different from that of $G_{n,r/n}$ (see for example [20] or Chapter 9 in [11]). Markström [17] proved that $G(n, 3)$ a.a.s. contains a complete minor

Date: October 23, 2008.

N. Fountoulakis and D. Kühn were supported by the EPSRC, grant no. EP/D50564X/1.

of order k , for any integer $k \geq 3$ and conjectured that $G(n, 3)$ contains a complete minor of order $\Omega(\sqrt{n})$. In this paper, we verify this conjecture for any $r \geq 3$:

Theorem 1. *There exists an absolute constant $c > 0$ such that for every fixed $r \geq 3$ a.a.s. $c\sqrt{n} \leq \text{ccl}(G(n, r)) \leq 2\sqrt{rn}$.*

The bound on c that is obtained from our calculations is very small; approximately 10^{-40} . We have made no effort to optimize this constant. This result can be combined with results of Łuczak [15] to determine the likely order of magnitude of $\text{ccl}(G_{n,p})$ during the phase transition, i.e. when $pn \rightarrow 1$ (see Section 3 for details).

Corollary 2. *There exists an absolute constant $c > 0$ such that whenever $np = 1 + \lambda n^{-1/3}$, where $\lambda = \lambda(n) \rightarrow \infty$ but $\lambda = o(n^{1/3})$, then a.a.s. $c\lambda^{3/2} \leq \text{ccl}(G_{n,p}) \leq 4\lambda^{3/2}$.*

Łuczak, Pittel and Wierman [16] previously showed that a.a.s. $\text{ccl}(G_{n,p})$ is unbounded for p as in Corollary 2. For smaller p (i.e. when $np \leq 1 + \lambda n^{-1/3}$ for some constant λ) they showed that $\text{ccl}(G_{n,p})$ is bounded in probability, i.e. for every $\delta > 0$ there exists $C = C(\delta)$ such that $\mathbb{P}(\text{ccl}(G_{n,p}) > C) < \delta$. As described earlier, values of p which are larger than those allowed for in Corollary 2 but bounded away from 1 are covered in [9]. So altogether, all these results determine the likely order of magnitude of $\text{ccl}(G_{n,p})$ for any p which is bounded away from 1.

The results in [16] were proved in connection with the following result on the limiting probability $g(\lambda)$ that $G_{n,p}$ is planar in the above range. The authors proved that if λ is bounded, then $g(\lambda)$ is bounded away from 0 and 1. If $\lambda \rightarrow -\infty$, then $g(\lambda) \rightarrow 1$, whereas if $\lambda \rightarrow \infty$, then $g(\lambda) \rightarrow 0$.

2. OPEN QUESTIONS

The result in (1) and Theorem 1 raise the question of whether one can extend these results to other (not necessarily random) graphs. A natural class to consider are expanding graphs: A graph G on n vertices is an (t, α) -expander if any $X \subseteq V(G)$ with $|X| \leq \alpha n/t$ satisfies $|N(X)| \geq t|X|$, where $N(X)$ denotes the external neighbourhood of X .

Problem 3. *Is there a constant $c > 0$ such that every $(10, 1/20)$ -expander satisfies $\text{ccl}(G) \geq c\sqrt{n}$?*

An answer to the problem would indicate whether expansion alone is sufficient when trying to force a complete minor of the largest possible order in a sparse graph, or whether other parameters are also relevant. Krivelevich and Sudakov [13] showed that we do have $\text{ccl}(G) \geq c\sqrt{n/\log n}$. (They also considered the case when t is not bounded but grows with n .) As observed in [13], this bound can also be deduced from a result of Plotkin, Rao and Smith [19] on separators in graphs without a large complete minor. Kleinberg and Rubinfeld [12] also considered the same problem but with a weaker definition of expansion.

Secondly, one can ask similar questions as above for topological minors. Topological minors in random graphs were investigated initially by Ajtai, Komlós and Szemerédi [1] and subsequently by Bollobás, Catlin and Erdős [7] and Bollobás and Catlin [6]. An analogue of Problem 3 would be to ask for which values of α, t, r an r -regular (t, α) -expander on n vertices contains a subdivision of a K_{r+1} . We expect that this might not be difficult to prove for fixed r but harder if r is no longer very small compared to n .

Thirdly, the question arises whether one could generalize Theorem 1 as follows: is there a constant c so that almost surely one can find *any* graph H which has at most cn edges as a minor? However, this is easily seen to be false. Given any constant c , let H be the graph consisting of $cn/3$ triangles. If $G(n, 3)$ contains H as a minor then $G(n, 3)$ would need to

have $cn/6$ cycles of length at most $6/c$. On the other hand, it is well known that with high probability $G(n, 3)$ contains only a bounded number of cycles of length at most $6/c$. Note that this argument works for $G_{n,p}$ with pn constant too.

Finally, it is not clear what happens if we consider minors in r -regular graphs where r is not a constant.

3. PROOF OF COROLLARY 2

The upper bound in Corollary 2 will follow from basic facts about minors as well as the structure of $G_{n,p}$. Bollobás [5] (see also [3] or [11]) proved that a.a.s. all the components of $G_{n,p}$, except from the largest one, are either trees or unicyclic. Therefore none of them contains a K_4 minor. Let $L_1(G_{n,p})$ denote the largest component of $G_{n,p}$. Given a graph G , we define its *excess* as $\text{exc}(G) := e(G) - |G| + 1$. ($\text{exc}(G)$ is also called the *cyclomatic number* of G .) Observe that if H and G are connected graphs and $H \prec G$ then $\text{exc}(H) \leq \text{exc}(G)$. Since $\text{exc}(K_r) = \binom{r}{2} - r + 1 \geq r^2/16$ for $r \geq 4$ this implies that if $K_r \prec L_1(G_{n,p})$ for some $r \geq 4$ then $r^2/16 \leq \text{exc}(L_1(G_{n,p}))$, or equivalently

$$(2) \quad r \leq 4\sqrt{\text{exc}(L_1(G_{n,p}))}.$$

Łuczak [14] gave a tight estimate on $\text{exc}(L_1(G_{n,m}))$, where $G_{n,m}$ is a random graph with n vertices and m edges (i.e. $G_{n,m}$ is chosen uniformly at random among all such graphs). He proved that if $m = n/2 + \bar{\lambda}n^{2/3}$, where $\bar{\lambda} = \bar{\lambda}(n) \rightarrow \infty$ but $\bar{\lambda} = o(n^{1/3})$, then a.a.s. $\text{exc}(L_1(G_{n,m})) = (1 + o(1))16\bar{\lambda}^3/3$. This trivially implies that if $m = n/2 + \bar{\lambda}n^{2/3} + O(\sqrt{n})$ then a.a.s. $\text{exc}(L_1(G_{n,m})) \leq 8\bar{\lambda}^3$. Together with the fact that $\binom{n}{2}p = n/2 + \lambda n^{2/3}/2 + O(1)$ and Proposition 1.12 in [11] this implies that a.a.s. $\text{exc}(L_1(G_{n,p})) \leq \lambda^3$. But if the latter holds and $K_r \prec L_1(G_{n,p})$ for some $r \geq 4$ then (2) gives $r \leq 4\lambda^{3/2}$. Thus a.a.s. $\text{ccl}(G_{n,p}) \leq 4\lambda^{3/2}$.

For the lower bound in Corollary 2 we will use the following result of Łuczak which is contained in the proof of Theorem 5* in [15].

Theorem 4. *Suppose that $m = n/2 + \lambda n^{2/3}$, where $\lambda \rightarrow \infty$ and $\lambda = o(n^{1/3})$. Then there is a procedure which in any given graph G with n vertices and m edges finds a subdivision of a (possibly empty) 3-regular graph $C = C(G)$ such that a.a.s. $|C(G_{n,m})| = (32/3 + o(1))\lambda^3$ and conditional on $|C(G_{n,m})| = s$ in this range the distribution of $C(G_{n,m})$ is the same as $G(s, 3)$.*

Loosely speaking, Theorem 4 implies that a.a.s. $L_1(G_{n,m})$ contains a subdivision of a random 3-regular graph $G(s, 3)$ where $s = (32/3 + o(1))\lambda^3$. Together with Theorem 1 this implies that a.a.s.

$$\text{ccl}(G_{n,m}) \geq \text{ccl}(C(G_{n,m})) \geq c\lambda^{3/2}.$$

Again Proposition 1.12 from [11] now yields the lower bound of Corollary 2.

4. SKETCH OF PROOF OF THEOREM 1

We will use a result of Janson [10] which implies that it suffices to find a complete minor in the union of a random Hamilton cycle and a random perfect matching. We split the Hamilton cycle into paths P_1 and P_2 of equal length. We further split P_1 into k connected candidate branch sets B_i , where k is close to \sqrt{n} . Each of these candidate branch sets has size roughly \sqrt{n} . We now split P_2 into sets \mathcal{P}_i of disjoint paths. The lengths of the paths in \mathcal{P}_i is roughly 3^i , whereas the number of paths in \mathcal{P}_i is roughly $n/9^i$. For each pair (B, B') of candidate branch sets we aim to find a path P in some \mathcal{P}_i such that both B and B' are

joined to P by an edge of the random perfect matching. We let \mathcal{U}_{i-1} denote the set of pairs of candidate branch sets for which we were not able to find such a path P in $\bigcup_{j < i} \mathcal{P}_j$. We will show inductively that $|\mathcal{U}_i| \leq |\mathcal{U}_{i-1}|/27$ (with sufficiently high probability). By continuing this for $(\log_3 n)/6$ stages and discarding a few atypical branch sets, we eventually obtain the desired minor. This strategy is similar to that of [9]. However, the proof that it works is very different: the argument in [9] was based on a greedy matching algorithm whose analysis crucially relied on the independence of certain events. In the current setting, this no longer works. So instead, in each stage we use Hall's theorem to find a large matching in the bipartite auxiliary graph whose vertex classes are \mathcal{U}_{i-1} and \mathcal{P}_i and where a pair (B, B') is adjacent to a path $P \in \mathcal{P}_i$ if P can be used to join B and B' as above. (Actually, it turns out that we need to consider suitable subsets $\mathcal{U}'_{i-1} \subseteq \mathcal{U}_{i-1}$ and $\mathcal{S} \subseteq \mathcal{P}_i$ for the argument to work.) Though the number $|\mathcal{P}_i|$ of paths decreases in each stage, the increasing path length means that the average degree of a pair (B, B') in this auxiliary graph remains large (but bounded) in each stage and so we can indeed expect to find a large matching. On the other hand, one can show that there might be a significant number of pairs from \mathcal{U}_{i-1} which are isolated in the auxiliary graph. So we cannot hope to get away with just a single stage.

5. MODELS OF RANDOM r -REGULAR GRAPHS

The aim of this section is to show that it suffices to find our complete minor in the union of a random Hamilton cycle and a random perfect matching. For $n \geq 1$ let $V_n := \{1, \dots, n\}$. Also for those n for which rn is even, we let $P := V_n \times [r]$. A *configuration* is a perfect matching on P (configurations were introduced by Bender and Canfield [2] and independently by Bollobás [4]). If we project a configuration onto V_n , then we obtain an r -regular multigraph on V_n . Let $G^*(n, r)$ denote the random multigraph that is the projection of a configuration on P which is chosen uniformly at random. It can be shown (see e.g. [11, p. 236]) that if we condition on $G^*(n, r)$ being simple (i.e. it does not have loops or multiple edges), then this has the same distribution as $G(r, n)$. Moreover, it is well known that $\lim_{n \rightarrow \infty} \mathbb{P}(G^*(n, r) \text{ is simple}) > 0$ (see Corollary 9.7 in [11]).

We let $G'(n, r)$ denote a random multigraph whose distribution is that of $G^*(n, r)$ conditional on having no loops. Let A_n be a subset of the set of r -regular multigraphs on V_n . It is easy to see that the above facts imply that if $\mathbb{P}(G'(n, r) \in A_n) \rightarrow 0$ as $n \rightarrow \infty$ then $\mathbb{P}(G(n, r) \in A_n) \rightarrow 0$. This allows us to work with $G'(n, r)$ instead of $G(n, r)$ itself.

If (μ_n) and (ν_n) are two sequences of probability measures such that for each n , μ_n and ν_n are measures on the same measurable space Ω_n , then we say that they are *contiguous* if for every sequence of measurable sets (A_n) with $A_n \in \Omega_n$ we have $\lim_{n \rightarrow \infty} \mu_n(A_n) = 0$ if and only if $\lim_{n \rightarrow \infty} \nu_n(A_n) = 0$. Now let $H(n) + G(n, 1)$ denote the random multigraph on V_n that is obtained from a Hamilton cycle on V_n chosen uniformly at random by adding a random perfect matching on V_n chosen independently from the Hamilton cycle. Janson [10] (see also Theorem 9.30 in [11]) proved that $H(n) + G(n, 1)$ is contiguous to $G'(n, 3)$. So the following result implies the case when $r = 3$ of Theorem 1.

Theorem 5. *There exists an absolute constant $c' > 0$ such that a.a.s. the random multigraph $H(n) + G(n, 1)$ contains a complete minor of order at least $c'\sqrt{n}$.*

The lower bound for $r > 3$ follows from Theorem 9.36 in [11] which states that for each $s \geq 3$ an increasing property that holds a.a.s. for $G(n, s)$ also holds a.a.s. for $G(n, s + 1)$.

6. NOTATION AND LARGE DEVIATION INEQUALITIES

6.1. Notation. Given a graph G and two disjoint sets A and B of vertices, we say that an edge of G is an *A - B edge* if it joins a vertex in A to a vertex in B . Given disjoint

subgraphs H and H' of G , we define H - H' edges of G similarly. Given $a, b \in \mathbb{R}$ we write $[a \pm b]$ for the interval $[a - b, a + b]$. We will write $\ln^2 n$ for $(\ln n)^2$. Also, if x is real number and r a positive integer, we write $(x)_r := x(x-1) \cdots (x-r+1)$. If K is a graph or a set of vertices, then in order to simplify notation we sometimes also write $|K|_r$ for $(|K|)_r$. We omit floors and ceilings whenever this does not affect the argument.

6.2. A concentration inequality. In this subsection, we will state a concentration inequality which we will use several times during the proof of Theorem 5. This is Theorem 7.4 in [18]. We first describe the more general setting to which this theorem applies.

Let W be a finite probability space that is also a metric space with its metric denoted by d . Suppose that F_0, \dots, F_s is a sequence of partitions of W such that F_{j+1} refines F_j , F_0 is the partition consisting of only one part (i.e. $F_0 = \{W\}$) and F_s is the partition where each part is a single element of W . Suppose that whenever $A, B \in F_{j+1}$ and $C \in F_j$ are such that $A, B \subseteq C$, then there is a bijection $\phi : A \rightarrow B$ such that $d(x, \phi(x)) \leq c$. Now, let $w \in W$ be chosen uniformly at random and let $f : W \rightarrow \mathbb{R}$ be a function on W satisfying $|f(x) - f(y)| \leq d(x, y)$. Then for all $a > 0$

$$(3) \quad \mathbb{P}(|f - \mathbb{E}(f)| > a) \leq 2 \exp\left(\frac{-2a^2}{sc^2}\right).$$

6.3. The hypergeometric distribution. Let Z be a non-empty finite set and $Z' \subseteq Z$. Assume that we sample a set Y uniformly at random among all subsets of Z having size y . Recall that the size of $Y \cap Z'$ is a random variable whose distribution is *hypergeometric* and whose expected value is $\lambda := y|Z'|/|Z|$. We will often use the following concentration inequality that follows e.g. from Theorem 2.10 and Inequalities (2.5) and (2.6) in [11]:

$$(4) \quad \mathbb{P}(|Y \cap Z'| - \lambda| \geq a) \leq 2 \exp\left(-\frac{a^2}{2(\lambda + a/3)}\right)$$

for all $a \geq 0$.

7. PROOF OF THEOREM 5

7.1. Setup. Let V_n be a set of n vertices. We will expose the random multigraph $H(n) + G(n, 1)$ on V_n in stages starting with the Hamilton cycle $H(n)$. We split $H(n)$ into two paths P_1, P_2 of equal lengths each having $n/2$ vertices. As described in Section 4, the (candidate) branch sets for our minor will be subpaths of P_1 and we will use the edges of the random perfect matching $G(n, 1)$ as well as subpaths of P_2 to join them. Let us now turn to $G(n, 1)$. So consider a perfect matching M^* on V_n chosen uniformly at random. Our first aim is to estimate the number of P_1 - P_2 edges of M^* .

Lemma 6. *With probability $1 - O(1/\ln^2 n)$ the number of P_1 - P_2 edges of M^* lies in the interval $[n/4 \pm \sqrt{n} \ln n]$.*

Proof. This is a simple application of Chebyshev's inequality. For each vertex $v \in V(P_1)$ set $X_v := 1$ if M^* matches v to a vertex of P_2 and set $X_v := 0$ otherwise. Then $X := \sum_{v \in V(P_1)} X_v$ is the number of P_1 - P_2 edges of M^* . Note that for every v we have

$$\mathbb{P}(X_v = 1) = \frac{n/2}{n-1} = 1/2 + O(1/n).$$

So $\mathbb{E}X = (n/2)(1/2 + O(1/n)) = n/4 + O(1)$. Also, for distinct $v, w \in V(P_1)$ we have

$$\mathbb{P}(X_v = 1 \mid X_w = 1) = \frac{n/2 - 1}{n - 3} = 1/2 + O(1/n).$$

This implies that

$$\begin{aligned}\mathbb{E}(X^2) &= \sum_{v \in V(P_1)} \mathbb{P}(X_v = 1) + \sum_{v \neq w \in V(P_1)} \mathbb{P}(X_v = X_w = 1) \\ &= \mathbb{E}X + \frac{n}{2} \left(\frac{n}{2} - 1 \right) \left(\frac{1}{4} + O(1/n) \right) = \frac{n^2}{16} + O(n) = (\mathbb{E}X)^2(1 + O(1/n)).\end{aligned}$$

So Chebyshev's inequality implies that

$$\mathbb{P}(|X - n/4| \geq \sqrt{n} \ln n) \leq \mathbb{P}(|X - \mathbb{E}X| \geq (\sqrt{n} \ln n)/2) = \frac{O(1/n)(\mathbb{E}X)^2}{n \ln^2 n} = O(1/\ln^2 n),$$

as required. \square

Fix a positive constant ε . Throughout the proof we will assume that ε is sufficiently small for our estimates to hold. (All conditions on ε will involve only absolute constants, i.e. will be independent of n .) Suppose that n is sufficiently large compared to $1/\varepsilon$. Let k and t be integers such that

$$(5) \quad \binom{k}{2} = \varepsilon^4 n \quad \text{and} \quad t = \frac{\sqrt{n}}{\varepsilon}.$$

So $k = (1 + o(1))\varepsilon^2 \sqrt{2n}$. Consider any $X_1 \subseteq V(P_1)$ and $X_2 \subseteq V(P_2)$ such that $|X_1| = |X_2| \in [n/4 \pm \sqrt{n} \ln n]$. Let $X'_1 \subseteq X_1$ be the set of the first kt vertices on P_1 in X_1 . Let $X'_2 \subseteq X_2$ be any subset of size kt . Let \mathcal{X} denote the event that X_1 and X_2 are the set of endvertices of the P_1 - P_2 edges in our random perfect matching M^* on V_n . Similarly, let \mathcal{X}' be the event that M^* matches X'_1 to X'_2 . In what follows, we will condition on both \mathcal{X} and \mathcal{X}' . All our probability bounds will hold regardless of what the sets X_1, X_2, X'_1, X'_2 actually are (provided that $|X_1| = |X_2|$ is in the specified range).

Pick k consecutive disjoint subpaths B_1, \dots, B_k of P_1 such that $|V(B_i) \cap X_1| = |V(B_i) \cap X'_1| = t$ for all $i = 1, \dots, k$. The B_i 's will be called *candidate branch sets* and the vertices in $V(B_i) \cap X'_1$ will be called the *effective vertices of B_i* . We will show that a.a.s. there is a complete minor whose branch sets are almost all the B_i 's. Set

$$(6) \quad i_0 := (\log_3 n)/6.$$

Choose consecutive disjoint subpaths Q_1, \dots, Q_{i_0} of P_2 such that

$$(7) \quad |Q_i|_{\text{eff}} := |V(Q_i) \cap X'_2| = \frac{|X'_2|}{3^i} = \frac{kt}{3^i} = \frac{(1 + o(1))\sqrt{2}\varepsilon n}{3^i}.$$

The vertices in $V(Q_i) \cap X'_2$ are the *effective vertices of Q_i* and $|Q_i|_{\text{eff}}$ is the *effective length of Q_i* . We further divide each Q_i into a set \mathcal{P}_i of consecutive disjoint subpaths, each of effective length

$$(8) \quad \ell_i := 100 \cdot 3^{i-1}.$$

(So each of these subpaths meets X'_2 in precisely ℓ_i vertices.) Note that

$$(9) \quad \ell_{i_0} \leq 100 \cdot 3^{i_0} = 100n^{1/6}$$

and

$$(10) \quad |\mathcal{P}_i| = \frac{|Q_i|_{\text{eff}}}{\ell_i} = \frac{kt}{300 \cdot 9^{i-1}}.$$

Thus $|\mathcal{P}_{i_0}| = \Theta(n^{2/3})$. The strategy of our proof is to expose the neighbours of the (effective) vertices from X'_2 in our random perfect matching M^* in i_0 stages. More precisely, during the i th stage we will expose the neighbours of the effective vertices in Q_i (for every $1 \leq i \leq i_0$).

We will show that with high probability during each stage we can use the paths in \mathcal{P}_i to join a large proportion of all those pairs of candidate branch sets that are still unjoined after the previous stages. More precisely, an unjoined pair (B, B') of candidate branch sets can be joined through $P \in \mathcal{P}_i$ if our random perfect matching M^* contains both a B - P edge and a B' - P edge. In this case we will say that P can be used to join the pair (B, B') . Of course, if we use P to join (B, B') then P cannot be used to join another unjoined pair of candidate branch sets.

Let us make the above more precise. Given $1 \leq i \leq i_0$, let \mathcal{U}_{i-1} denote the set of pairs of candidate branch sets that are still unjoined after the $(i-1)$ th stage. So \mathcal{U}_0 is the set of all pairs of candidate branch sets. Note that

$$(11) \quad U_0 := |\mathcal{U}_0| = \binom{k}{2} = \varepsilon^4 n.$$

We will show that with high probability during the i th stage we can join $26|\mathcal{U}_{i-1}|/27$ pairs in \mathcal{U}_{i-1} using the paths belonging to \mathcal{P}_i . So inductively we will prove that with high probability

$$(12) \quad U_i := |\mathcal{U}_i| = \frac{U_0}{27^i} = \frac{\varepsilon^4 n}{27^i}.$$

Suppose that (12) holds for all $j < i$ and that we now wish to analyze the i th stage. It will turn out that the pairs in \mathcal{U}_{i-1} which contain candidate branch sets lying in too many other pairs from \mathcal{U}_{i-1} create problems. So we will ignore these pairs. More precisely, let \mathcal{B}_{i-1} be the set of all those candidate branch sets that belong to more than

$$(13) \quad \Delta_{i-1} := \frac{(3/2)^{i-1} U_{i-1}}{\varepsilon^{1/8} k} = \frac{U_0}{\varepsilon^{1/8} (2 \cdot 9)^{i-1} k}$$

pairs in \mathcal{U}_{i-1} . Note that since $|\mathcal{B}_{i-1}| \Delta_{i-1} \leq 2U_{i-1}$ we have

$$(14) \quad |\mathcal{B}_{i-1}| \leq \frac{2U_{i-1}}{\Delta_{i-1}} \leq \frac{2\varepsilon^{1/8} k}{(3/2)^{i-1}}.$$

Let \mathcal{U}_{i-1}^* be the set of all those pairs in \mathcal{U}_{i-1} having at least one branch set in \mathcal{B}_{i-1} . Call these pairs *bad*. If $|\mathcal{U}_{i-1}^*| \geq 26U_{i-1}/27$, delete precisely $26U_{i-1}/27$ bad pairs from \mathcal{U}_{i-1} to obtain \mathcal{U}_i . If $|\mathcal{U}_{i-1}^*| < 26U_{i-1}/27$ we let $\mathcal{U}'_{i-1} := \mathcal{U}_{i-1} \setminus \mathcal{U}_{i-1}^*$. We will show that during the i th stage with high probability we can join all but $U_{i-1}/27$ pairs in \mathcal{U}'_{i-1} . We let \mathcal{U}_i be the set of the remaining unjoined pairs in \mathcal{U}'_{i-1} . Thus in both cases $U_i = |\mathcal{U}_i|$ satisfies (12) with high probability.

After the end of stage i_0 will delete all the candidate branch sets in $\mathcal{B}_0 \cup \dots \cup \mathcal{B}_{i_0-1}$ (see Section 7.5). The number of these candidate branch sets is

$$(15) \quad \sum_{i=1}^{i_0} |\mathcal{B}_{i-1}| \stackrel{(14)}{\leq} \sum_{i \geq 1} \frac{2\varepsilon^{1/8} k}{(3/2)^{i-1}} = 6\varepsilon^{1/8} k.$$

7.2. Bounds on the number of effective vertices still available. We will now estimate the number of all those effective vertices in each candidate branch set that are joined to a path in $\mathcal{P}_1 \cup \dots \cup \mathcal{P}_{i-1}$, i.e. that are matched after the first $i-1$ stages. The total number of effective vertices in the candidate branch sets that are matched after the first $i-1$ stages is

$$(16) \quad \sum_{j \leq i-1} |Q_j|_{\text{eff}} \stackrel{(7)}{=} kt \sum_{j=1}^{i-1} \frac{1}{3^j} = \frac{kt}{2} \left(1 - 3^{-(i-1)}\right) =: x_{i-1}.$$

Each x_{i-1} -subset of the union X'_1 of all the effective vertices in the candidate branch sets is equally likely to be the set of these matched vertices. Thus for every candidate branch set B the distribution of the number $\text{eff}'_i(B)$ of all those effective vertices in B which are matched to (effective vertices on) paths in $\mathcal{P}_1 \cup \dots \cup \mathcal{P}_{i-1}$ is hypergeometric. Since in total B contains t effective vertices and $|X'_1| = kt$ we can now use (4) to see that

$$\mathbb{P} \left(|\text{eff}'_i(B) - x_{i-1}t/kt| \geq n^{1/4} \ln n \mid \mathcal{X}, \mathcal{X}' \right) \leq \exp(-\Omega(\ln^2 n)).$$

Thus,

$$\text{eff}'_i(B) \in \left[\frac{x_{i-1}}{k} \pm n^{1/4} \ln n \right] \subseteq \left[\frac{t}{2}(1 - 3^{-(i-1)}) \pm n^{1/3} \right],$$

with (conditional) probability $1 - \exp(-\Omega(\ln^2 n))$. Now, let $\text{eff}_i(B) := t - \text{eff}'_i(B)$ be the number of all those effective vertices in B that are still unmatched after the first $i - 1$ stages and let $\text{Eff}_i(B)$ denote the set of all those effective vertices. Thus with (conditional) probability $1 - k \exp(-\Omega(\ln^2 n)) = 1 - \exp(-\Omega(\ln^2 n))$ we have

$$(17) \quad \text{eff}_i(B) \in \left[\frac{t}{2}(1 + 3^{-(i-1)}) \pm n^{1/3} \right]$$

for all candidate branch sets B .

Let M_{i-1}^* be any matching which matches the set $\text{Eff}(Q_1) \cup \dots \cup \text{Eff}(Q_{i-1})$ of effective vertices on the paths Q_1, \dots, Q_{i-1} (equivalently the set of effective vertices on the paths in $\mathcal{P}_1 \cup \dots \cup \mathcal{P}_{i-1}$) into the set of effective vertices in the candidate branch sets. Suppose that M_{i-1}^* is the submatching of our random matching M^* exposed after the first $i - 1$ stages. Then M_{i-1}^* determines $\text{Eff}_i(B)$ for every candidate branch set B . Moreover, by considering a fixed ordering of all the pairs in \mathcal{U}_0 , we may assume that M_{i-1}^* also determines \mathcal{U}_{i-1} . Call M_{i-1}^* *good* if (17) holds for all candidate branch sets B and if (12) holds for $i - 1$. Consider any good M_{i-1}^* and let \mathcal{M}_{i-1}^* denote the event that M_{i-1}^* is the submatching of our random matching M^* exposed after the first $i - 1$ stages. From now on we will condition on \mathcal{X} , \mathcal{X}' and \mathcal{M}_{i-1}^* and we let $\mathbb{P}_i(\cdot)$ denote the corresponding conditional probability measure that arises from choosing a random matching from the set $\text{Eff}(Q_i)$ of effective vertices on Q_i into the set $\bigcup_{j=1}^k \text{Eff}_i(B_j)$ of all those effective vertices in the candidate branch sets which are not already endvertices of edges in M_{i-1}^* (i.e. into the set of all those effective vertices in the candidate branch sets that are still unmatched after the first $i - 1$ stages).

Given $\mathcal{S} \subseteq \mathcal{P}_i$ and a candidate branch set B , we let $\text{Eff}_{\mathcal{S}}(B)$ denote the set of all those effective vertices in B that are matched to some (effective) vertex on a path in \mathcal{S} (in our random matching M^*). Assume that $|\mathcal{S}| = \alpha|\mathcal{P}_i|$ where $1/2 \leq \alpha \leq 1$. Let

$$(18) \quad I(\alpha) := \left[\frac{\alpha t}{3^i} \left(1 \pm \frac{1}{4} \right) \right]$$

and let $\mathcal{E}_{\mathcal{S}}$ denote the event that $|\text{Eff}_{\mathcal{S}}(B)| \in I(\alpha)$ for every candidate branch set B . Let $\overline{\mathcal{E}_{\mathcal{S}}}$ denote the complement of $\mathcal{E}_{\mathcal{S}}$.

Lemma 7. $\mathbb{P}_i(\overline{\mathcal{E}_{\mathcal{S}}}) = \exp(-\Omega(\ln^2 n))$.

Proof. Consider any candidate branch set B . Note that $|\text{Eff}_{\mathcal{S}}(B)|$ is hypergeometrically distributed with mean $\lambda := \text{eff}_i(B)|\mathcal{S}|l_i/(kt - x_{i-1})$. But $|\mathcal{S}|l_i = \alpha|\mathcal{P}_i|l_i = \alpha|Q_i|_{\text{eff}}$ and $kt - x_{i-1} = \frac{kt}{2}(1 + 3^{-(i-1)})$ by (16). So (17) implies that

$$\frac{\text{eff}_i(B)}{kt - x_{i-1}} \in \left[\frac{\frac{t}{2}(1 + 3^{-(i-1)}) \pm n^{1/3}}{\frac{kt}{2}(1 + 3^{-(i-1)})} \right] \subseteq \left[\frac{1}{k} \left(1 \pm \frac{1}{5} \right) \right]$$

and thus

$$\lambda \in \left[\alpha |Q_i|_{\text{eff}} \left(\frac{1}{k} \left(1 \pm \frac{1}{5} \right) \right) \right] \stackrel{(7)}{=} \left[\frac{\alpha t}{3^i} \left(1 \pm \frac{1}{5} \right) \right].$$

In particular, together with (4) this implies that

$$\mathbb{P}_i \left(\left| |\text{Eff}_{\mathcal{S}}(B)| - \lambda \right| \geq n^{1/4} \ln n \right) = \exp \left(-\Omega(\ln^2 n) \right).$$

So with probability at most $k \exp \left(-\Omega(\ln^2 n) \right) = \exp \left(-\Omega(\ln^2 n) \right)$ there is a candidate branch set B with

$$|\text{Eff}_{\mathcal{S}}(B)| \notin \left[\frac{\alpha t}{3^i} \left(1 \pm \frac{1}{5} \right) \pm n^{1/4} \ln n \right] \stackrel{(5),(18)}{\subseteq} I(\alpha),$$

as required. \square

7.3. A lower bound for the degrees of the pairs of candidate branch sets in \mathcal{G}_i .

Recall that, as described in the paragraph after (14), when analyzing the i th stage, we may assume that $|\mathcal{U}_{i-1}^*| < 26U_{i-1}/27$ and thus \mathcal{U}'_{i-1} is well defined. Given a candidate branch set B and path $P \in \mathcal{P}_i$, we write $P \sim B$ if some effective vertex on P is matched to some vertex in $\text{Eff}_i(B)$ (in our random matching M^*). Consider an auxiliary bipartite graph \mathcal{G}_i whose vertex classes are \mathcal{U}'_{i-1} and \mathcal{P}_i and in which a pair $(B, B') \in \mathcal{U}'_{i-1}$ is adjacent to $P \in \mathcal{P}_i$ if P can be used to join (B, B') , i.e. if $P \sim B$ and $P \sim B'$. We will now estimate the degrees of the vertices in \mathcal{U}'_{i-1} in \mathcal{G}_i . Given $\mathcal{S} \subseteq \mathcal{P}_i$, we let $d_{\mathcal{G}_i}(L, \mathcal{S})$ denote the degree of a vertex/pair $L \in \mathcal{U}'_{i-1}$ into the set \mathcal{S} (in \mathcal{G}_i).

Lemma 8. *Suppose that $1/2 \leq \alpha \leq 1$ (where α may depend on n). Fix $L \in \mathcal{U}'_{i-1}$ and $\mathcal{S} \subseteq \mathcal{P}_i$ with $|\mathcal{S}| = \alpha |\mathcal{P}_i|$. Then $\mathbb{P}_i(d_{\mathcal{G}_i}(L, \mathcal{S}) \leq 1/(2\varepsilon^3)) < 3\varepsilon$.*

Proof. Let $L = (B, B')$. Our aim is to show that

$$(19) \quad \mathbb{P}_i(d_{\mathcal{G}_i}(L, \mathcal{S}) \leq 1/(2\varepsilon^3) \mid \mathcal{E}_{\mathcal{S}}) < 2\varepsilon.$$

This implies the lemma since

$$\begin{aligned} \mathbb{P}_i(d_{\mathcal{G}_i}(L, \mathcal{S}) \leq 1/(2\varepsilon^3)) &\leq \mathbb{P}_i(d_{\mathcal{G}_i}(L, \mathcal{S}) \leq 1/(2\varepsilon^3) \mid \mathcal{E}_{\mathcal{S}}) + \mathbb{P}_i(\overline{\mathcal{E}_{\mathcal{S}}}) \\ &\stackrel{(19), \text{Lemma 7}}{\leq} 2\varepsilon + \exp(-\Omega(\ln^2 n)) < 3\varepsilon. \end{aligned}$$

To estimate the number of all those paths in \mathcal{S} that are neighbours of both B and B' in the auxiliary graph \mathcal{G}_i , we will first bound the number of paths in \mathcal{S} that are neighbours of B and then we will estimate how many of them are neighbours of B' . More precisely, we will first show that most of the paths $P \in \mathcal{S}$ with $P \sim B$ are joined to B by exactly one (matching) edge. Let us condition first on a particular realization E_B of $\text{Eff}_{\mathcal{S}}(B)$ with $|E_B| \in I(\alpha)$. Denote the corresponding probability subspace of \mathbb{P}_i (where we condition on the event that $\text{Eff}_{\mathcal{S}}(B) = E_B$ and on $\mathcal{E}_{\mathcal{S}}$) by $\mathbb{P}_{i, \mathcal{E}_{\mathcal{S}}, E_B}$. Assuming an arbitrary ordering of the vertices in E_B , we expose their neighbours (in the random matching) on the paths in \mathcal{S} one by one according to this ordering. We say that the j th vertex *fails* if its neighbour lies in a path from \mathcal{S} that already contains a neighbour of the previously exposed vertices. Note that the number of paths in \mathcal{S} containing more than one neighbour of E_B is bounded above by the number of failures that occur during the exposure of the neighbours of E_B . Suppose we have exposed the neighbours of the first $j-1$ vertices in E_B . Let the corresponding event be \mathcal{C}_{j-1} . To estimate the probability that the j th vertex fails, observe that the number of all those paths in \mathcal{S} that already have a neighbour in E_B is less than j and each of them

contains less than ℓ_i effective vertices which are still available. Note that this holds regardless of what \mathcal{C}_{j-1} is. Thus

$$\mathbb{P}_{i,\mathcal{E}_S,E_B}(\text{the } j\text{th vertex fails} \mid \mathcal{C}_{j-1}) < \frac{j\ell_i}{\alpha|\mathcal{P}_i|\ell_i - (j-1)} \leq \frac{|E_B|\ell_i}{\alpha|\mathcal{P}_i|\ell_i - |E_B|} \stackrel{(6),(10)}{\leq} \frac{2|E_B|}{\alpha|\mathcal{P}_i|} =: \tilde{p}.$$

In particular, let \mathcal{D}_{j-1} be any event which depends only on the neighbours of the first $j-1$ vertices in E_B . Then

$$(20) \quad \mathbb{P}_{i,\mathcal{E}_S,E_B}(\text{the } j\text{th vertex fails} \mid \mathcal{D}_{j-1}) \leq \tilde{p}.$$

Now let $A = \{a_1, \dots, a_r\}$ be any set of vertices in E_B (where a_q precedes a_{q+1} in the ordering of E_B) and let Fail_A denote the event that the set of failure vertices equals A . Then

$$\mathbb{P}_{i,\mathcal{E}_S,E_B}(\text{Fail}_A) \leq \prod_{q=1}^r \mathbb{P}_{i,\mathcal{E}_S,E_B}(a_q \text{ fails} \mid a_1, \dots, a_{q-1} \text{ fail}) \stackrel{(20)}{\leq} \tilde{p}^r.$$

This in turn implies that

$$\mathbb{P}_{i,\mathcal{E}_S,E_B}(\geq f \text{ failures}) \leq \sum_{r=f}^{|E_B|} \sum_{\substack{A \subseteq E_B \\ |A|=r}} \mathbb{P}_{i,\mathcal{E}_S,E_B}(\text{Fail}_A) \leq \sum_{r=f}^{|E_B|} \binom{|E_B|}{r} \tilde{p}^r \leq \sum_{r=f}^{|E_B|} \left(\frac{e|E_B|}{r} \frac{2|E_B|}{\alpha|\mathcal{P}_i|} \right)^r.$$

Since $|E_B| \in I(\alpha)$ we have

$$(21) \quad \frac{|E_B|}{\alpha|\mathcal{P}_i|} \stackrel{(10),(18)}{\in} \left[(1 \pm 1/4)^2 \frac{\alpha^2 t^2}{9^i} \frac{1}{\alpha} \frac{300 \cdot 9^i}{9kt} \right] \stackrel{(5)}{\subseteq} \left[(1 \pm 2/3) \frac{100\alpha}{3\sqrt{2}\varepsilon^3} \right].$$

Let \mathcal{F}_B denote the event that at least $1000/\varepsilon^3$ failures occur when we expose the neighbours of the vertices in E_B . Thus by setting $f := 1000/\varepsilon^3$, we obtain

$$(22) \quad \mathbb{P}_{i,\mathcal{E}_S,E_B}(\mathcal{F}_B) = \mathbb{P}_{i,\mathcal{E}_S,E_B}(\geq f \text{ failures}) \leq \sum_{r \geq f} \left(\frac{5\sqrt{2} \cdot 100e\alpha}{9\varepsilon^3 r} \right)^r \leq \sum_{r \geq f} (1/2)^r \leq \varepsilon.$$

Let $\overline{\mathcal{F}_B}$ denote the complement of \mathcal{F}_B . Note that if $\overline{\mathcal{F}_B}$ occurs, then there are at least $|E_B| - 1000/\varepsilon^3$ paths in \mathcal{S} that are joined to $E_B \subseteq B$ by exactly one (matching) edge. Let $\mathcal{S}(B)$ denote the set of these paths. So

$$(23) \quad |E_B| - 1000/\varepsilon^3 \leq |\mathcal{S}(B)| \leq |E_B|$$

(for the second inequality we need not assume that $\overline{\mathcal{F}_B}$ holds). Now we additionally condition on a specific realization $E_{B'}$ of $\text{Eff}_S(B')$ with $|E_{B'}| \in I(\alpha)$. As above, we fix an arbitrary ordering on the vertices in $E_{B'}$ according to which we expose their neighbours in \mathcal{S} . We say that the j th vertex of $E_{B'}$ is *useful* if it is adjacent to a vertex lying on a path from $\mathcal{S}(B)$ such that none of the previous vertices in $E_{B'}$ is joined to this path. Note that if $U(B')$ denotes the set of vertices in $E_{B'}$ that are useful, then

$$(24) \quad |U(B')| \leq d_{\mathcal{G}_i}(L, \mathcal{S}).$$

Given $\overline{\mathcal{F}_B}$, we will show that with high probability $|U(B')| \geq 1/(2\varepsilon^3)$.

Note that there are exactly $R := \alpha|\mathcal{P}_i|\ell_i - |E_B|$ effective vertices on the paths in \mathcal{S} that are still available to be matched to the vertices of $E_{B'}$. Put $s := |\mathcal{S}(B)|(\ell_i - 1)$ and let C be any subset of $E_{B'}$ with $c := |C| \leq 1/(2\varepsilon^3)$. Suppose that C is the set of useful vertices in $E_{B'}$. Then the vertices in C are matched to effective vertices on different paths in $\mathcal{S}(B)$. So there are $|\mathcal{S}(B)|_c(\ell_i - 1)^c \leq s^c$ choices for the neighbours of C . Moreover, each vertex $x \in E_{B'} \setminus C$ is either matched to an effective vertex on a path in $\mathcal{S}(B)$ which already contains a neighbour of C or x is matched to an effective vertex on a path in $\mathcal{S} \setminus \mathcal{S}(B)$. There are

less than $c\ell_i$ choices for a neighbour of x having the first property and $R - s$ choices for a neighbour of x having the second property. Thus in total the number of choices for the neighbours of $E_{B'} \setminus C$ is at most

$$\begin{aligned} \sum_{q=0}^{|E_{B'}|-c} (|E_{B'}| - c)_q (c\ell_i)^q (R - s)_{|E_{B'}|-c-q} &\leq (R - s)_{|E_{B'}|-c} \sum_{q=0}^{|E_{B'}|-c} \left(\frac{c\ell_i |E_{B'}|}{R/2} \right)^q \\ &\leq (R - s)_{|E_{B'}|-c} \sum_{q \geq 0} \left(\frac{1}{2} \right)^q = 2(R - s)_{|E_{B'}|-c}. \end{aligned}$$

(Here we used that $|E_{B'}| = O(\sqrt{n}) = o(|\mathcal{P}_i|)$ and so $s = o(R)$ as well as $|E_{B'}|\ell_i = o(R)$.) Setting $p := s/R$ we obtain

$$\begin{aligned} \mathbb{P}_{i, \mathcal{E}_S, E_B}(U(B') = C \mid \overline{\mathcal{F}_B}, \text{Eff}_S(B') = E_{B'}) &\leq \frac{2s^c (R - s)_{|E_{B'}|-c}}{(R)_{|E_{B'}|}} \\ &\leq 2s^c \left(\frac{R - s}{R} \right)^{|E_{B'}|-c} \left(\frac{1}{R - |E_{B'}|} \right)^c = 2 \left(\frac{s}{R - |E_{B'}|} \right)^c \left(1 - \frac{s}{R} \right)^{|E_{B'}|-c} \\ &\leq 2 \left(\frac{10}{9} \right)^c \left(\frac{s}{R} \right)^c \left(1 - \frac{s}{R} \right)^{|E_{B'}|-c} = 2 \left(\frac{10}{9} \right)^c p^c (1 - p)^{|E_{B'}|-c}. \end{aligned}$$

(In the second inequality we used that $\frac{a-j}{b-j} < \frac{a}{b}$, for $0 < j < a < b$ and in the last inequality we again used that $|E_{B'}| = o(R)$.) Thus

$$(25) \quad \begin{aligned} \mathbb{P}_{i, \mathcal{E}_S, E_B}(|U(B')| \leq 1/(2\varepsilon^3) \mid \overline{\mathcal{F}_B}, \text{Eff}_S(B') = E_{B'}) \\ \leq 2(10/9)^{1/(2\varepsilon^3)} \sum_{c \leq 1/(2\varepsilon^3)} \binom{|E_{B'}|}{c} p^c (1 - p)^{|E_{B'}|-c}. \end{aligned}$$

Observe that the sum on the right-hand side is the probability that a binomial random variable Y with parameters $|E_{B'}|, p$ is at most $1/(2\varepsilon^3)$. To bound this probability from above, we will use the following Chernoff bound (see e.g. Inequality (2.9) in [11]):

$$(26) \quad \mathbb{P}(Y \leq \mathbb{E}Y/2) \leq 2 \exp(-\mathbb{E}Y/12).$$

Note that by (23) and the definition of R , we have

$$p = \frac{s}{R} = \frac{|\mathcal{S}(B)|(\ell_i - 1)}{R} \geq \frac{(|E_B| - 1000/\varepsilon^3)(\ell_i - 1)}{\alpha|\mathcal{P}_i|\ell_i} > \frac{|E_B|}{\alpha|\mathcal{P}_i|} \frac{98}{100},$$

where the last inequality holds since $\ell_i \geq 100$. Moreover, the bound (21) also holds if we replace $|E_B|^2$ by $|E_B||E_{B'}|$. So altogether we have

$$\mathbb{E}Y = |E_{B'}|p \geq \frac{98}{100} \frac{|E_{B'}||E_B|}{\alpha|\mathcal{P}_i|} \stackrel{(21)}{\geq} \frac{2\alpha}{\varepsilon^3} \geq \frac{1}{\varepsilon^3}.$$

Thus (26) implies that

$$\sum_{c \leq 1/(2\varepsilon^3)} \binom{|E_{B'}|}{c} p^c (1 - p)^{|E_{B'}|-c} \leq 2 \exp(-1/(12\varepsilon^3)).$$

Substituting this bound into (25), we obtain

$$\mathbb{P}_{i, \mathcal{E}_S, E_B}(|U(B')| \leq 1/(2\varepsilon^3) \mid \overline{\mathcal{F}_B}, \text{Eff}_S(B') = E_{B'}) \leq 4 \left((10/9)^6 / e \right)^{1/(12\varepsilon^3)} \leq (9/10)^{1/(12\varepsilon^3)} \leq \varepsilon.$$

Since $E_{B'}$ was an arbitrary realization of $\text{Eff}_{\mathcal{S}}(B')$ with $|E_{B'}| \in I(\alpha)$, this implies that

$$\mathbb{P}_{i, \mathcal{E}_{\mathcal{S}}, E_B}(|U(B')| \leq 1/(2\varepsilon^3) \mid \overline{\mathcal{F}_B}) \leq \varepsilon$$

and thus

$$(27) \quad \begin{aligned} \mathbb{P}_{i, \mathcal{E}_{\mathcal{S}}, E_B}(|U(B')| \leq 1/(2\varepsilon^3)) &\leq \mathbb{P}_{i, \mathcal{E}_{\mathcal{S}}, E_B}(|U(B')| \leq 1/(2\varepsilon^3) \mid \overline{\mathcal{F}_B}) + \mathbb{P}_{i, \mathcal{E}_{\mathcal{S}}, E_B}(\mathcal{F}_B) \\ &\stackrel{(22)}{\leq} 2\varepsilon. \end{aligned}$$

Finally, since E_B was an arbitrary realization of $\text{Eff}_{\mathcal{S}}(B)$ with $|E_B| \in I(\alpha)$ it follows that

$$\mathbb{P}_i(d_{\mathcal{G}_i}(L, \mathcal{S}) \leq 1/(2\varepsilon^3) \mid \mathcal{E}_{\mathcal{S}}) \stackrel{(24)}{\leq} \mathbb{P}_i(|U(B')| \leq 1/(2\varepsilon^3) \mid \mathcal{E}_{\mathcal{S}}) \stackrel{(27)}{\leq} 2\varepsilon,$$

as required. \square

Now given $\mathcal{S} \subseteq \mathcal{P}_i$, we let $\mathcal{U}(\mathcal{S})$ be the set of all those pairs in \mathcal{U}'_{i-1} that have degree at most $1/(2\varepsilon^3)$ into \mathcal{S} (in our auxiliary graph \mathcal{G}_i). So if $1/2 \leq \alpha \leq 1$ and $|\mathcal{S}| = \alpha|\mathcal{P}_i|$ then Lemma 8 implies that

$$(28) \quad \mathbb{E}_i(|\mathcal{U}(\mathcal{S})|) \leq 3\varepsilon|\mathcal{U}'_{i-1}| \leq \frac{U_{i-1}}{2 \cdot 27},$$

where $\mathbb{E}_i(\cdot)$ denotes the expectation that arises from the probability measure $\mathbb{P}_i(\cdot)$.

Lemma 9. *Let $1/2 \leq \alpha \leq 1$. Then every $\mathcal{S} \subseteq \mathcal{P}_i$ with $|\mathcal{S}| = \alpha|\mathcal{P}_i|$ satisfies*

$$(29) \quad \mathbb{P}_i\left(|\mathcal{U}(\mathcal{S})| > \frac{U_{i-1}}{27}\right) \leq 2 \exp\left(-\frac{2\varepsilon^8 n}{(3^{i-1})^7}\right)$$

as well as

$$(30) \quad \mathbb{P}_i\left(|\mathcal{U}(\mathcal{S})| > \frac{U_{i-1}}{27}\right) \leq 2 \exp\left(-\varepsilon^4 3^{(i-1)/4}\right).$$

Proof. Our aim is to apply (3) to show that $|\mathcal{U}(\mathcal{S})|$ is concentrated around its expected value. We first prove (29). Here W will be the space of all those matchings which match the set $\text{Eff}(Q_i)$ of effective vertices on Q_i into the set $\bigcup_{j=1}^k \text{Eff}_i(B_j)$ of all those effective vertices in the candidate branch sets that are still unmatched after the first $i-1$ stages (equipped with the uniform distribution). (Recall that $\text{Eff}_i(B)$ is fixed since we condition on \mathcal{M}_{i-1} .) So each matching in W consists of $|Q_i|_{\text{eff}}$ edges. The metric d on W is defined by $d(M, M') := 2\ell_i|M \Delta M'|$ for all $M, M' \in W$. It is easy to see that this is indeed a metric.

So let us now define the partitions $F_0, \dots, F_{|Q_i|_{\text{eff}}}$. $F_0 := \{W\}$ and each part of $F_{|Q_i|_{\text{eff}}}$ will consist of a single matching in W . To define F_j for $1 \leq j < |Q_i|_{\text{eff}}$, fix a linear ordering on the vertices in $\text{Eff}(Q_i)$. Given a matching $M \in W$, the j -prefix of M is the set of all edges in M adjacent to the first j vertices in $\text{Eff}(Q_i)$. Each part of the partition F_j will consist of all those matchings in W having the same j -prefix. Clearly F_{j+1} refines F_j .

To define the bijection ϕ , consider any two parts $A \neq B$ of F_{j+1} and any part C of F_j such that $A, B \subseteq C$. So if $M \in A$ and $M' \in B$, then M and M' have the same j -prefix and they differ at the edge that is adjacent to the $(j+1)$ th vertex in $\text{Eff}(Q_i)$. Let v_A and v_B be the neighbours of the $(j+1)$ th vertex in M and M' , respectively. Note that v_A does not depend on the choice of $M \in A$ and similarly for v_B . We define $\phi : A \rightarrow B$ by saying that for all $M \in A$ the matching $\phi(M)$ is obtained from M as follows: the $(j+1)$ th vertex in $\text{Eff}(Q_i)$ is now matched to v_B and v_A is matched to the neighbour of v_B in M , every other edge of M remains unchanged. Thus the size of the symmetric difference of M and $\phi(M)$ is 4 and so $d(M, \phi(M)) \leq 8\ell_i$. So we can take $c := 8\ell_i$.

Now note that $|\mathcal{U}(\mathcal{S})|$ is a function whose value is determined by a matching from W chosen uniformly at random. So we take $f : W \rightarrow \mathbb{R}$ to be the function defined by setting $f(M)$ to be the value of $|\mathcal{U}(\mathcal{S})|$ on M (for all $M \in W$). We have to show that for any $M, M' \in W$ we have $|f(M) - f(M')| \leq d(M, M')$. To do so, we will construct a sequence M_0, M_1, \dots, M_q of matchings in W such that $M_0 := M$, $M_q := M'$ and such that M_j and M' agree on the first j vertices in $\text{Eff}(Q_i)$ (i.e. M_j and M' have the same prefix). Suppose that we have constructed M_j for some $j < q$ and that we now wish to construct M_{j+1} . Let v be the first vertex in $\text{Eff}(Q_i)$ on which M_j and M' differ. Let b be its neighbour in M' and let v' be the neighbour of b in M_j . Define M_{j+1} to be the matching obtained from M_j by swapping the neighbours of v and v' in M_j . So M_{j+1} now agrees with M' on v and all (the at least j) vertices preceding v in $\text{Eff}(Q_i)$. Note that $|f(M_j) - f(M_{j+1})| \leq 4\ell_i$ since swapping two edges can change $|\mathcal{U}(\mathcal{S})|$ by at most $4\ell_i$. Indeed, to see the latter, note that for each one of these two edges there are $\ell_i - 1$ other edges starting from the same path in \mathcal{P}_i , and therefore each of these two edges contributes to the degree of at most ℓ_i pairs in \mathcal{U}'_{i-1} . If we swap these edges, this might change the degree of at most $4\ell_i$ pairs. So $|\mathcal{U}(\mathcal{S})|$ can be increased or decreased by at most $4\ell_i$. Also observe that $q \leq |M \Delta M'|/2$ since initially the number of vertices in $\text{Eff}(Q_i)$ on which M and M' differ equals $|M \Delta M'|/2$ and in each step this number decreases by at least 1. Therefore,

$$(31) \quad |f(M) - f(M')| \leq \sum_{j=0}^{q-1} |f(M_j) - f(M_{j+1})| \leq 4q\ell_i \leq 2\ell_i |M \Delta M'| = d(M, M').$$

Now, we are ready to apply (3): if $|\mathcal{U}(\mathcal{S})| > U_{i-1}/27$, then by (28) we have $|\mathcal{U}(\mathcal{S})| - \mathbb{E}(|\mathcal{U}(\mathcal{S})|) > U_{i-1}/(2 \cdot 27)$ and (3) yields

$$\begin{aligned} \mathbb{P}_i \left(|\mathcal{U}(\mathcal{S})| > \frac{U_{i-1}}{27} \right) &\leq 2 \exp \left(-2 \frac{(U_{i-1}/(2 \cdot 27))^2}{|Q_i|_{\text{eff}} 8^2 \ell_i^2} \right) \\ &\stackrel{(7),(8),(12)}{=} 2 \exp \left(-\frac{1}{2^7 \cdot 27^2} \frac{\varepsilon^8 n^2}{(27^{i-1})^2} \frac{3^i}{(1 + o(1))\sqrt{2}\varepsilon n} \frac{1}{100^2 9^{i-1}} \right) \\ &\leq 2 \exp \left(-\frac{2\varepsilon^8 n}{(3^{i-1})^7} \right). \end{aligned}$$

Now we prove (30). In this case we can apply (3) with metric $d(M, M') := 2\Delta_{i-1}|M \Delta M'|$ and $c := 8\Delta_{i-1}$. Indeed, for each candidate branch set B and each $P \in \mathcal{P}_i$ the removal/addition of a B - P edge can only affect the degrees of those pairs in \mathcal{U}'_{i-1} which contain B . But there are at most Δ_{i-1} such pairs. Thus

$$\begin{aligned} \mathbb{P}_i \left(|\mathcal{U}(\mathcal{S})| > \frac{U_{i-1}}{27} \right) &\leq 2 \exp \left(-2 \frac{(U_{i-1}/(2 \cdot 27))^2}{|Q_i|_{\text{eff}} 8^2 \Delta_{i-1}^2} \right) \\ &\stackrel{(7),(13)}{\leq} 2 \exp \left(-\frac{U_{i-1}^2}{2^7 \cdot 27^2} \frac{3^i}{kt} \frac{\varepsilon^{1/4} k^2}{(3/2)^{2(i-1)} U_{i-1}^2} \right) \\ &\stackrel{(5)}{\leq} 2 \exp \left(-\frac{1}{2^7 \cdot 27^2} \frac{3\varepsilon^{1/4}(1 + o(1))\sqrt{2}\varepsilon^2}{1/\varepsilon} (4/3)^{i-1} \right) \\ &\leq 2 \exp \left(-\varepsilon^4 (4/3)^{i-1} \right) \leq 2 \exp \left(-\varepsilon^4 3^{(i-1)/4} \right), \end{aligned}$$

as required. \square

Define β by

$$(32) \quad \beta := \left(\frac{\varepsilon^8}{2(3^{i-1})^7} \right)^2.$$

Lemma 10. *For each i with $3^{i-1} \leq n^{2/33}$ let Y_i denote the number of all those subsets \mathcal{S} of \mathcal{P}_i with $|\mathcal{S}| = (1 - \beta)|\mathcal{P}_i|$ for which $|\mathcal{U}(\mathcal{S})| > U_{i-1}/27$. Then $\mathbb{P}_i(Y_i > 0) \leq n^{-1/34}$.*

Proof. Note that (10) and the restriction on i together imply that $\beta|\mathcal{P}_i| = \Omega(n/3^{16i}) = \Omega(n^{1/33})$ and so we may treat it as an integer. (29) implies that

$$(33) \quad \mathbb{E}_i(Y_i) \leq \binom{|\mathcal{P}_i|}{(1 - \beta)|\mathcal{P}_i|} 2 \exp\left(-\frac{2\varepsilon^8 n}{(3^{i-1})^7}\right).$$

Note that $|\mathcal{P}_i| \leq n$. So

$$\binom{|\mathcal{P}_i|}{(1 - \beta)|\mathcal{P}_i|} = \binom{|\mathcal{P}_i|}{\beta|\mathcal{P}_i|} \leq \left(\frac{e}{\beta}\right)^{\beta|\mathcal{P}_i|} \leq \beta^{-2\beta|\mathcal{P}_i|} \leq \beta^{-2\beta n}.$$

Now note that if $a > 0$ is sufficiently small then $a \ln(a^{-1}) \leq a^{1/2}$. Thus

$$\binom{|\mathcal{P}_i|}{(1 - \beta)|\mathcal{P}_i|} \leq e^{2\beta^{1/2}n} \stackrel{(32)}{=} \exp\left(\frac{\varepsilon^8 n}{(3^{i-1})^7}\right).$$

So if $3^{i-1} \leq n^{2/33}$ then

$$\mathbb{P}_i(Y_i > 0) \leq \mathbb{E}_i(Y_i) \stackrel{(33)}{\leq} 2 \exp\left(-\frac{\varepsilon^8 n}{(3^{i-1})^7}\right) = \exp\left(-\Omega\left(n^{19/33}\right)\right) \leq n^{-1/34},$$

as required. \square

7.4. An upper bound on the degrees of the paths in \mathcal{G}_i . Let $d := 10^6$. We now estimate the probability that a given path $P \in \mathcal{P}_i$ joins at least d unjoined pairs in \mathcal{U}'_{i-1} .

Lemma 11. *If $d = 10^6$, $i \leq i_0$ and i satisfies*

$$(34) \quad 3^{i-1} \geq 1/\varepsilon,$$

then for every fixed $P \in \mathcal{P}_i$ we have $\mathbb{P}_i(d_{\mathcal{G}_i}(P) \geq d) \leq \beta/3^i$.

Proof. Suppose that $C \subseteq \mathcal{U}'_{i-1}$ is a set of size d which lies in the neighbourhood of P in the auxiliary graph \mathcal{G}_i . Let $\mathcal{B}(C)$ denote the set of candidate branch sets involved in the pairs from C . Note that

$$\sqrt{2d} \leq |\mathcal{B}(C)| \leq 2d.$$

Moreover, $P \sim B$ for each candidate branch set $B \in \mathcal{B}(C)$. (Recall that this means that there is an effective vertex on P that is matched to some vertex in $\text{Eff}_i(B)$, where $\text{Eff}_i(B)$ was the set of all those effective vertices in B that are still available after the $(i-1)$ th stage.) Now let \mathbf{B} be the collection of all the sets \mathcal{B} of candidate branch sets such that for each $B \in \mathcal{B}$ there is a $B' \in \mathcal{B}$ with $(B, B') \in \mathcal{U}'_{i-1}$ and such that $b := |\mathcal{B}|$ satisfies

$$(35) \quad \sqrt{2d} \leq b \leq 2d.$$

Thus $\mathcal{B}(C) \in \mathbf{B}$ for any C as above and hence

$$(36) \quad \mathbb{P}_i(d_{\mathcal{G}_i}(P) \geq d) \leq \sum_{\mathcal{B} \in \mathbf{B}} \mathbb{P}_i(P \sim B \forall B \in \mathcal{B}).$$

To bound the latter probability, consider any $\mathcal{B} \in \mathbf{B}$, let $b := |\mathcal{B}|$ and $s := \sum_{B \in \mathcal{B}} \text{eff}_i(B) \leq bt$. Recall that x_{i-1} was the total number of all those effective vertices in the branch sets that are matched after the first $i-1$ stages. So

$$\begin{aligned} \mathbb{P}_i(P \sim B \forall B \in \mathcal{B}) &\leq \binom{\ell_i}{b} \frac{(s)_b}{(kt - x_{i-1})_b} \leq \left(\frac{e\ell_i}{b}\right)^b \left(\frac{s}{kt - x_{i-1}}\right)^b \\ &\leq \left(\frac{e\ell_i}{b} \frac{bt}{kt - x_{i-1}}\right)^b \stackrel{(16)}{\leq} \left(\frac{et\ell_i}{kt/2}\right)^b = \left(\frac{2e\ell_i}{k}\right)^b. \end{aligned}$$

In the second inequality, we used that $\frac{a-j}{b-j} < \frac{a}{b}$, for $0 < j < a < b$. To bound $|\mathbf{B}|$, consider an auxiliary graph \mathcal{A}_{i-1} whose vertex set is the set of candidate branch sets and whose edges correspond to the pairs in \mathcal{U}'_{i-1} . Since \mathcal{A}_{i-1} involves only edges/pairs from \mathcal{U}'_{i-1} its maximum degree is at most Δ_{i-1} . Consider any b as in (35). Note that each $\mathcal{B} \in \mathbf{B}$ with $|\mathcal{B}| = b$ corresponds to a subgraph F of \mathcal{A}_{i-1} which has order b and in which no vertex is isolated. We claim that for all $q \leq b/2$, the number of such subgraphs F having precisely q components is at most $U_{i-1}^q (b\Delta_{i-1})^{b-2q}$. To see this, note that each component of F has to contain at least one edge (this is also the reason why it makes sense only to consider $q \leq b/2$). So each subgraph F as above can be obtained as follows. First choose q (independent) edges of \mathcal{A}_{i-1} . The number of choices for this is at most U_{i-1}^q . Now successively add the remaining $b-2q$ vertices to the existing subgraph without creating new components. At each step there are at most b vertices y to which a new vertex z can be attached and once we have chosen y , there are at most $\Delta(\mathcal{A}_{i-1}) \leq \Delta_{i-1}$ choices for z , which proves the claim.

Let $\mathbf{B}_{b,q}$ be the set of all those $\mathcal{B} \in \mathbf{B}$ that have size b and induce q components in \mathcal{A}_{i-1} . Then

$$\begin{aligned} \sum_{\mathcal{B} \in \mathbf{B}_{b,q}} \mathbb{P}_i(P \sim B \forall B \in \mathcal{B}) &\stackrel{(35)}{\leq} U_{i-1}^q (2d\Delta_{i-1})^{b-2q} \left(\frac{2e\ell_i}{k}\right)^b \\ \stackrel{(8),(12),(13)}{\leq} &\left(\frac{U_0}{27^{i-1}} \frac{1}{4d^2} \frac{\varepsilon^{1/4} k^2 81^{i-1} 4^{i-1}}{U_0^2}\right)^q \left(2d \frac{U_0}{\varepsilon^{1/8} k 9^{i-1} 2^{i-1}} \frac{2e \cdot 100 \cdot 3^{i-1}}{k}\right)^b \\ \stackrel{(11)}{\leq} &(12^{i-1})^q \left(\frac{2000d}{\varepsilon^{1/8} 6^{i-1}}\right)^b \leq \left(\frac{4 \cdot 10^6 d^2}{\varepsilon^{1/4} 3^{i-1}}\right)^{b/2} \leq \left(\frac{1}{\varepsilon^{1/3} 3^{i-1}}\right)^{b/2} \stackrel{(34)}{\leq} \left(\frac{\varepsilon^{1/6}}{3^{(i-1)/2}}\right)^{b/2}. \end{aligned}$$

Since $b \geq \sqrt{2d} \geq 4 \cdot 48$ by (35) this implies

$$\mathbb{P}_i(d_{\mathcal{G}_i}(P) \geq d) \stackrel{(36)}{\leq} \sum_{\sqrt{2d} \leq b \leq 2d} \sum_{q=1}^{b/2} \left(\frac{\varepsilon^{1/6}}{3^{(i-1)/2}}\right)^{b/2} \leq 2d^2 \left(\frac{\varepsilon^{1/3}}{3^{i-1}}\right)^{\sqrt{2d}/4} \leq 2d^2 \left(\frac{\varepsilon^{1/3}}{3^{i-1}}\right)^{48} \leq \frac{\beta}{3^i},$$

as required. \square

Given $d = 10^6$ and i satisfying (34), let \mathcal{S}_i denote the set of paths in \mathcal{P}_i which have degree less than d in \mathcal{G}_i . We will now use Lemma 11 to show that with high probability \mathcal{S}_i is large.

Lemma 12. $\mathbb{P}_i(|\mathcal{S}_i| \geq (1-\beta)|\mathcal{P}_i|) \geq 1 - n^{-1/34}$ for every $i \leq i_0$ which satisfies (34).

Proof. Let $\overline{\mathcal{S}}_i := \mathcal{P}_i \setminus \mathcal{S}_i$. Note that Lemma 11 implies

$$\mathbb{E}_i(\overline{\mathcal{S}}_i) = \mathbb{P}_i(d_{\mathcal{G}_i}(P) \geq d) |\mathcal{P}_i| \leq \frac{\beta}{3^i} |\mathcal{P}_i|.$$

If $(\log_3 n)/34 \leq i \leq i_0$ then together with Markov's inequality this yields

$$\mathbb{P}_i(|\overline{\mathcal{S}}_i| > \beta|\mathcal{P}_i|) \leq \frac{1}{3^i} \leq \frac{1}{n^{1/34}}$$

and thus Lemma 12 holds for all such i .

If $i \leq (\log_3 n)/34$ we will use (3). As in the proof of Lemma 9, the underlying metric space will be the set of all those matchings which match the set $\text{Eff}(Q_i)$ of effective vertices on Q_i into the set $\bigcup_{j=1}^k \text{Eff}_i(B_j)$ of all those effective vertices in the candidate branch sets that are still unmatched after the first $i-1$ stages. The series of partitions and the bijections ϕ are also as defined there. However, the metric imposed on W now changes: for any $M, M' \in W$ we set $d(M, M') = |M \Delta M'|$. In particular this means that we can take $c := 4$. $f : W \rightarrow \mathbb{R}$ will be the function defined by taking $f(M)$ to be the value of $|\overline{\mathcal{S}}_i|$ on M (for all $M \in W$). Note that the analogue of (31) is satisfied, since if we switch the endpoints of two edges of a matching (as it is the case when we obtain M_{j+1} from M_j as in the proof of Lemma 9) $|\overline{\mathcal{S}}_i|$ changes by at most 2 as switching two edges only affects the degree of the (at most) two paths involved. Thus

$$|f(M) - f(M')| \leq \sum_{j=0}^{q-1} |f(M_j) - f(M_{j+1})| \leq 2q \leq |M \Delta M'| \leq d(M, M').$$

Hence applying (3) with $a := \beta|\mathcal{P}_i|/2$ we obtain

$$\mathbb{P}_i(|\overline{\mathcal{S}}_i| \geq \beta|\mathcal{P}_i|) \leq 2 \exp\left(-2 \frac{\beta^2 |\mathcal{P}_i|^2}{4 \cdot 16 |Q_i|_{\text{eff}}}\right).$$

So to complete the proof, it suffices to show that $\beta^2 |\mathcal{P}_i|^2 / |Q_i|_{\text{eff}} = \Omega(n^{3/34})$, as this gives an error bound of $\exp(-\Omega(n^{3/34})) \leq 1/n^{34}$. To prove the former, note that by (7), (10) and (32) we obtain

$$\frac{\beta^2 |\mathcal{P}_i|^2}{|Q_i|_{\text{eff}}} = \Theta\left(\frac{1}{(3^{i-1})^{28}} \left(\frac{n}{9^{i-1}}\right)^2 \frac{3^i}{n}\right) = \Theta\left(\frac{n}{3^{31(i-1)}}\right) = \Omega\left(\frac{n}{3^{31(\log_3 n)/34}}\right) = \Omega(n^{3/34}),$$

as required. \square

7.5. Finding a large matching of \mathcal{G}_i . The next lemma shows that with high probability we can join the required number of pairs from \mathcal{U}'_{i-1} during the i th stage.

Lemma 13. *For each $i \leq i_0$ we have $\mathbb{P}_i(|\mathcal{U}_i| > U_{i-1}/27) \leq 2n^{-1/34}$.*

Proof. Recall that \mathcal{U}'_{i-1} (defined after (14)) was obtained from \mathcal{U}_{i-1} by discarding all those pairs containing a candidate branch set from \mathcal{B}_i . By definition of \mathcal{U}_i we may assume that $|\mathcal{U}'_{i-1}| \geq U_{i-1}/27$ and it suffices to show that in \mathcal{G}_i we can find a matching which covers all but at most $U_{i-1}/27$ vertices/pairs in \mathcal{U}'_{i-1} .

Case 1: $3^{i-1} < 1/\varepsilon$.

In this case, we apply (29) with $\mathcal{S} := \mathcal{P}_i$ (i.e. $\alpha = 1$) to obtain that with probability at least $1 - 2 \exp(-2\varepsilon^8 n / (3^{i-1})^7) \geq 1 - 2n^{-1/34}$ we have the following: there is a set $\mathcal{W} \subseteq \mathcal{U}'_{i-1}$ with $|\mathcal{W}| = |\mathcal{U}'_{i-1}| - U_{i-1}/27$ so that every pair in \mathcal{W} has degree at least $1/(2\varepsilon^3)$ in \mathcal{G}_i . On the other hand, clearly every path in \mathcal{P}_i has degree at most $\ell_i^2 = 10^4 9^{i-1} < 10^4/\varepsilon^2$ in \mathcal{G}_i . This implies that the subgraph \mathcal{G}'_i of \mathcal{G}_i induced by \mathcal{W} and \mathcal{P}_i has a matching covering all of \mathcal{W} . To see this, consider any $\mathcal{W}' \subseteq \mathcal{W}$ and let $N(\mathcal{W}') \subseteq \mathcal{P}_i$ denote its neighbourhood in \mathcal{G}'_i . Then by counting edges between \mathcal{W}' and $N(\mathcal{W}')$ we obtain that $|\mathcal{W}'|/(2\varepsilon^3) \leq (10^4/\varepsilon^2)|N(\mathcal{W}')|$.

This in turn implies that $|N(\mathcal{W}')| \geq |\mathcal{W}'|$ and so Hall's condition is satisfied. But this means that we can take $\mathcal{U}_i := \mathcal{U}'_{i-1} \setminus \mathcal{W}$. Note that $U_i = |\mathcal{U}'_{i-1}| - |\mathcal{W}| = U_{i-1}/27$, as required.

Case 2: $1/\varepsilon \leq 3^{i-1} \leq n^{2/33}$.

In this case we first apply Lemma 12 to see that with probability at least $1 - n^{-1/34}$ we have $|\mathcal{S}_i| \geq (1 - \beta)|\mathcal{P}_i|$. By taking a subset we may assume that $|\mathcal{S}_i| = (1 - \beta)|\mathcal{P}_i|$. On the other hand, Lemma 10 implies that with probability at least $1 - n^{-1/34}$ any set \mathcal{S} of this size satisfies $|\mathcal{U}(\mathcal{S})| \leq U_{i-1}/27$. So with probability at least $1 - 2n^{-1/34}$ we have $|\mathcal{U}(\mathcal{S}_i)| \leq U_{i-1}/27$. But if this is the case then there is a set $\mathcal{W} \subseteq \mathcal{U}'_{i-1}$ with $|\mathcal{W}| = |\mathcal{U}'_{i-1}| - U_{i-1}/27$ so that every pair in \mathcal{W} has degree at least $1/(2\varepsilon^3)$ in the subgraph \mathcal{G}_i'' of \mathcal{G}_i induced by \mathcal{W} and \mathcal{S}_i . On the other hand, the definition of \mathcal{S}_i implies that in \mathcal{G}_i'' , the degree of every vertex in \mathcal{S}_i is at most $d = 10^6$. As in the previous case, this implies that \mathcal{G}_i'' has a matching covering all of \mathcal{W} . Indeed, to verify Hall's condition consider any $\mathcal{W}' \subseteq \mathcal{W}$ and let $N(\mathcal{W}') \subseteq \mathcal{S}_i$ denote its neighbourhood in \mathcal{G}_i'' . Then $|\mathcal{W}'|/(2\varepsilon^3) \leq 10^6|N(\mathcal{W}')|$. As before, we can take $\mathcal{U}_i := \mathcal{U}'_{i-1} \setminus \mathcal{W}$.

Case 3: $3^{i-1} \geq n^{2/33}$.

In this case, we apply (30) to $\mathcal{S} := \mathcal{P}_i$ in order to obtain that with probability at least

$$1 - 2 \exp\left(-\varepsilon^4 3^{(i-1)/4}\right) \geq 1 - 2 \exp\left(-\varepsilon^4 n^{1/66}\right) \geq 1 - n^{-1/34}$$

we have the following: there is a set $\mathcal{W} \subseteq \mathcal{U}'_{i-1}$ with $|\mathcal{W}| = |\mathcal{U}'_{i-1}| - U_{i-1}/27$ so that every pair in \mathcal{W} has degree at least $1/(2\varepsilon^3)$ in \mathcal{G}_i . On the other hand, Lemma 11 implies that the probability that \mathcal{P}_i does not contain a path of degree at least $d = 10^6$ in \mathcal{G}_i is at least

$$1 - \frac{\beta|\mathcal{P}_i|}{3^i} \stackrel{(10),(32)}{\geq} 1 - O(n/3^{17i}) = 1 - O(n^{-1/33}) \geq 1 - n^{-1/34}.$$

So we may assume that both events occur and we get a matching covering all of \mathcal{W} in the subgraph \mathcal{G}_i' of \mathcal{G}_i induced by \mathcal{W} and \mathcal{P}_i as before. So we again obtain a set \mathcal{U}_i of the desired size, with the required error bounds. \square

To complete the proof of Theorem 5 it remains to combine all the error probabilities for all the $i_0 = (\log_3 n)/6$ stages. Recall that when analyzing the i th stage we conditioned on \mathcal{M}_{i-1}^* (defined after (17)). However, all our probability bounds hold regardless of what the actual value of M_{i-1}^* is (as long as M_{i-1}^* is good). So suppose that $|\mathcal{U}_{i-1}| = U_0/27^{i-1}$ for some $i \leq i_0$. If $|\mathcal{U}_i| \neq U_0/27^i$ then either some candidate branch set violated (17) or we had the undesired event that $|\mathcal{U}_i| > U_{i-1}/27$ in Lemma 13. Thus

$$\mathbb{P}(|\mathcal{U}_i| = U_0/27^i \mid |\mathcal{U}_{i-1}| = U_0/27^{i-1}, \mathcal{X}, \mathcal{X}') \geq 1 - \exp(-\Omega(\ln^2 n)) - 2n^{-1/34} \geq 1 - 3n^{-1/34}$$

and so

$$\mathbb{P}(|\mathcal{U}_i| = U_0/27^i \forall i \leq i_0 \mid \mathcal{X}, \mathcal{X}') \geq 1 - 3i_0 n^{-1/34} \geq 1 - n^{-1/35}.$$

This bound holds regardless of what the choices of X_1, X_2, X'_1, X'_2 actually are (as long as $|X_1| = |X_2|$ is within the range determined in Lemma 6). The only other reason why $|\mathcal{U}_{i_0}| \neq U_0/27^{i_0}$ is that we had an undesired event in Lemma 6. This happens with probability $O(1/\ln^2 n)$. Altogether this shows that with probability $1 - n^{-1/35} - O(1/\ln^2 n) = 1 - o(1)$ after the i_0 th stage we are left with

$$(37) \quad U_{i_0} = \frac{U_0}{27^{i_0}} \stackrel{(6),(11)}{\geq} \varepsilon^4 n^{1/2}$$

unjoined pairs. We now discard a candidate branch set in each of these pairs as well as all the candidate branch sets in $\mathcal{B}_0 \cup \dots \cup \mathcal{B}_{i_0-1}$. By (15) and (37) this gives a complete minor on $k - 6\varepsilon^{1/8}k - \varepsilon^4 n^{1/2} \geq \varepsilon^2 n^{1/2}$ vertices, as required.

8. ACKNOWLEDGEMENTS

We are grateful to Tomasz Łuczak for helpful discussions on the phase transition of $G_{n,p}$.

REFERENCES

- [1] M. Ajtai, J. Komlós and E. Szemerédi, Topological complete subgraphs in random graphs, *Studia Sci. Math. Hungar.* **14** (1979), 293–297.
- [2] E.A. Bender and E.R. Canfield, The asymptotic number of labelled graphs with given degree sequences, *J. Combin. Theory A* **24** (1978), 296–307.
- [3] B. Bollobás, *Random Graphs*, 2nd. edition, Cambridge University Press, 2001.
- [4] B. Bollobás, A probabilistic proof of an asymptotic formula for the number of labelled regular graphs, *Europ. J. Combin.* **1** (1980), 311–316.
- [5] B. Bollobás, The evolution of random graphs, *Trans. Amer. Math. Soc* **286** (1984), 257–274.
- [6] B. Bollobás and P.A. Catlin, Topological cliques in random graphs, *J. Combin. Theory B* **30** (1981), 224–227.
- [7] B. Bollobás, P.A. Catlin and P. Erdős, Hadwiger’s conjecture is true for almost every graph, *Europ. J. Combin.* **1** (1980), 195–199.
- [8] P. Erdős and S. Fajtlowicz, On the conjecture of Hajós, *Combinatorica* **1** (1981), 141–143.
- [9] N. Fountoulakis, D. Kühn and D. Osthus, The order of the largest complete minor in a random graph, *Random Structures and Algorithms* **33** (2008), 127–141.
- [10] S. Janson, Random regular graphs: asymptotic distributions and contiguity, *Combinatorics, Probability & Computing* **4** (1995), 369–405.
- [11] S. Janson, T. Łuczak and A. Ruciński, *Random Graphs*, Wiley Interscience, 2000.
- [12] J. Kleinberg and R. Rubinfeld, Short paths in expander graphs, *Proc. 37th Symposium on Foundations of Computer Science (FOCS)*, IEEE Comput. Soc. Press (1996), 86–95.
- [13] M. Krivelevich and B. Sudakov, Minors in expanding graphs, *Geometric and Functional Analysis*, to appear.
- [14] T. Łuczak, Component behavior near the critical point of the random graph process, *Random Structures and Algorithms* **1** (1990), 287–310.
- [15] T. Łuczak, Cycles in a random graph near the critical point, *Random Structures and Algorithms* **2** (1991), 421–440.
- [16] T. Łuczak, B. Pittel and J.C. Wierman, The structure of a random graph at the point of the phase transition, *Trans. Amer. Math. Soc.* **341** (1994), 721–748.
- [17] K. Markström, Complete minors in cubic graphs with few short cycles and random cubic graphs, *Ars Combinatoria* **70** (2004), 289–295.
- [18] C.J.H. McDiarmid, On the method of bounded differences, In *Surveys in Combinatorics, 1989 (Norwich, 1989)*, London Math. Soc. Lecture Note Ser., 141, Cambridge Univ. Press, Cambridge, 1989, 148–188.
- [19] S. Plotkin, S. Rao and W. Smith, Shallow excluded minors and improved graph decompositions, *Proceedings 5th ACM-SIAM Symposium on Discrete Algorithms (SODA)*, ACM Press, New York (1994), 462–470.
- [20] N. Wormald, Models of random regular graphs, In *Surveys in Combinatorics, 1999 (Canterbury, 1999)*, London Math. Soc. Lecture Note Ser., 267, Cambridge Univ. Press, Cambridge, 1999, 239–298.

Nikolaos Fountoulakis, Daniela Kühn & Deryk Osthus

School of Mathematics

University of Birmingham

Edgbaston

Birmingham

B15 2TT

UK

E-mail addresses: {nikolaos,kuehn,osthus}@maths.bham.ac.uk