*Combinatorics, Probability and Computing* (2018) 27, 475–495. © Cambridge University Press 2018
doi:10.1017/S0963548318000159

# Packing Hamilton Cycles Online

JOSEPH BRIGGS[1], ALAN FRIEZE[1†], MICHAEL KRIVELEVICH[2‡], PO-SHEN LOH[1§] and BENNY SUDAKOV[3¶]

[1]Department of Mathematical Sciences, Carnegie Mellon University, Pittsburgh, PA 15213, USA
(e-mail: jbriggs@andrew.cmu.edu, alan@random.math.cmu.edu, ploh@cmu.edu)

[2]School of Mathematical Sciences, Raymond and Beverly Sackler Faculty of Exact Sciences, Tel Aviv University, Tel Aviv 6997801, Israel
(e-mail: krivelev@post.tau.ac.il)

[3]Department of Mathematics, ETH, 8092 Zurich, Switzerland
(e-mail: benjamin.sudakov@math.ethz.ch)

*Received 22 August 2016; revised 7 November 2017; first published online 22 March 2018*

It is known that w.h.p. the hitting time $\tau_{2\sigma}$ for the random graph process to have minimum degree $2\sigma$ coincides with the hitting time for $\sigma$ edge-disjoint Hamilton cycles [4, 9, 13]. In this paper we prove an online version of this property. We show that, for a fixed integer $\sigma \geqslant 2$, if random edges of $K_n$ are presented one by one then w.h.p. it is possible to colour the edges online with $\sigma$ colours so that at time $\tau_{2\sigma}$ each colour class is Hamiltonian.

2010 *Mathematics subject classification*: Primary 05C80
Secondary 05C15

## 1. Introduction

The celebrated *random graph process*, introduced by Erdős and Rényi [5] in the 1960s, begins with an empty graph on $n$ vertices, and at every step $t = 1, \ldots, \binom{n}{2}$ adds to the current graph a single new edge chosen uniformly at random out of all missing edges. Taking a snapshot of the random graph process after $m$ steps produces the distribution $G_{n,m}$. An equivalent 'static' way of defining $G_{n,m}$ would be to choose $m$ edges uniformly at random out of all $\binom{n}{2}$ possible ones. One

† Research supported in part by NSF grants DMS1362785, CCF1522984 and a grant (333329) from the Simons Foundation.
‡ Research supported in part by a USA–Israel BSF grant and by a grant from Israel Science Foundation.
§ Research supported in part by NSF grant DMS-1455125.
¶ Research supported in part by SNSF grant 200021-149111.

https://doi.org/10.1017/S0963548318000159 Published online by Cambridge University Press

advantage in studying the random graph process, rather than the static model, is that it allows for a higher resolution analysis of the appearance of monotone graph properties (a graph property is monotone if it is closed under edge addition).

A *Hamilton cycle* of a graph is a simple cycle that passes through every vertex of the graph, and a graph containing a Hamilton cycle is called *Hamiltonian*. Hamiltonicity is one of the most fundamental notions in graph theory, and has been intensively studied in various contexts, including random graphs. The earlier results on Hamiltonicity of random graphs were obtained by Pósa [15] and Korshunov [11]. Improving on these results, Bollobás [3] and Komlós and Szemerédi [10] proved that if $m' = \frac{1}{2}n\log n + \frac{1}{2}n\log\log n + \omega n$, then $G_{n,m'}$ is Hamiltonian w.h.p. Here $\omega$ is any function of $n$ tending to infinity together with $n$. One obvious necessary condition for the graph to be Hamiltonian is for the minimum degree to be at least 2, and the above result indicates that the events of being Hamiltonian and of having all degrees at least two are indeed bundled together closely. Bollobás [3], and independently Ajtai, Komlós and Szemerédi [1], further strengthened this by proving that w.h.p. the random graph process becomes Hamiltonian when the last vertex of degree one disappears. A more general property $\mathcal{H}_\sigma$ of having $\sigma$ edge-disjoint Hamilton cycles was studied by Bollobás and Frieze [4]. They showed that if $\sigma = O(1)$ then w.h.p. the random graph process satisfies $\mathcal{H}_\sigma$ when the minimum degree becomes $2\sigma$. It took quite a while, but this result was extended to the more difficult case of growing $\sigma$ in the $G_{n,m}$ context by Knox, Kühn and Osthus [9] and Krivelevich and Samotij [13].

Recently, quite a lot of attention and research effort has been devoted to controlled random graph processes. In processes of this type, an input graph or a graph process is usually generated fully randomly, but then an algorithm has access to this random input and can manipulate it in some well-defined way (say, by dropping some of the input edges, or by colouring them), aiming to achieve some preset goal. There is usually the so-called *online* version where the algorithm must decide on its course of action based only on the history of the process so far and without assuming any familiarity with future random edges. For example, in the so-called *Achlioptas process* the random edges arrive in batches of size $k$. An online algorithm chooses one of them and puts it into the graph. By doing this one can attempt to accelerate or to delay the appearance of some property. Hamiltonicity in Achlioptas processes was studied in [12]. Another online result on Hamiltonicity was proved in [14]. There, it was shown that one can orient the edges of the random graph process so that w.h.p. the resulting graph has a directed Hamilton cycle exactly at the time when the underlying graph has minimum degree two.

Here we consider a Ramsey-type version of controlled random processes. In this version, the incoming random edge, when it is exposed, is irrevocably coloured by an algorithm in one of $r$ colours, for a fixed $r \geqslant 2$. The goal of the algorithm is to achieve or to maintain a certain monotone graph property in all of the colours. For example, in [2] the authors considered the problem of creating a linear size (so-called *giant*) component in every colour.

The above-mentioned result of Bollobás and Frieze [4] gives rise to the following natural question. Can one typically construct $\sigma$ edge-disjoint Hamilton cycles in an online fashion by the time the minimum degree becomes $2\sigma$? We answer this question affirmatively in the case $\sigma = O(1)$.

**Theorem 1.1.** *For a fixed integer $\sigma \geqslant 2$, let $\tau_{2\sigma}$ denote the hitting time for the random graph process $G_i, i = 1,2,\ldots$ to have minimum degree $2\sigma$. Then w.h.p. we can colour the edges of*

https://doi.org/10.1017/S0963548318000159 Published online by Cambridge University Press

*$G_i, i = 1,2,\ldots$ online with $\sigma$ colours so that $G_{\tau_{2\sigma}}$ contains $\sigma$ Hamilton cycles $C_1, C_2, \ldots, C_\sigma$, where the edges of cycle $C_j$ all have colour $j$.*

## 2. Description of the colouring procedure

We describe our colouring procedure in terms of $q = 2\sigma$ colours; we aim to colour the edges so that each vertex has degree at least one in each colour. Think of colours 1 and $1+\sigma$ as being light red and dark red, say, and then that each vertex is incident with at least two red edges. This may appear cumbersome, but it does make some of the description of the analysis a little easier.

In the broadest terms, we construct two sets of edges $E^+$ and $E^*$. Let $\Gamma_c^*$ be the subgraph of $G_{\tau_{2\sigma}}$ induced by the edges of colour $c$ in $E^*$. We ensure that w.h.p. this has minimum degree at least one for all $c$. We then show that w.h.p. after merging colours $c$ and $c+\sigma$ for $c \in [\sigma]$ the subgraph $\Gamma_c^{**} = \Gamma_c^* \cup \Gamma_{c+\sigma}^*$ has sufficient expansion properties so that standard arguments using Pósa rotations can be applied. For every colour $c$, the edges of $E_c^*$ are used to help create a good expander, and produce a backbone for rotations. And the edges in $E_c^+$ are used to close cycles in this argument.

**Notation.** 'At time $t$' is taken to mean 'when $t$ edges have been revealed'.

**Notation.** Let $N^{(t)}(v)$ denote the set of neighbours of $v$ in $G_t$ and let $d_v^{(t)} = |N^{(t)}(v)|$.

For colour $c \in [q]$, write $d_c = d_{c,t}, N_c = N_{c,t}$ for the degrees and neighbourhoods of vertices and sets in $\Gamma_c$.

**Definition.** Let *Full* denote the set of vertices with degree at least $\varepsilon \log n/(1000q)$ in every colour at time

$$t_\varepsilon := \varepsilon n \log n,$$

where $\varepsilon$ is some sufficiently small constant depending only on the constant $q$. The actual value of $\varepsilon$ needed will depend on certain estimates below being valid, in particular equation (4.41). A vertex is *Full* if it lies in *Full*. Similarly, let $Full' \subseteq Full$ denote the set of vertices with degree at least $\varepsilon \log n/(1000q)$ in every colour at time $\frac{1}{2}\varepsilon n \log n$.

This definition only makes sense if $t_\varepsilon$ is an integer. Here and below we use the following convention. If we give an expression for an integer quantity that is not clearly an integer, then rounding the expression up or down will give a value that can be used to satisfy all requirements.

### 2.1. Colouring algorithm COL

We now describe our algorithm for colouring edges as we see them. At any time $t$, vertex $v$ has a list $C_v^{(t)} := \{c \in [q] : d_c^{(t)}(v) = 0\}$ of colours currently not present among edges incident to $v$; 'the colours that $v$ needs'. A vertex is *needy* at time $t$ if $C_v^{(t)} \neq \emptyset$. If the next edge to colour contains a needy vertex then we try to reduce the need of this vertex. Otherwise, we make choices to guarantee expansion in $E^*$, needed to generate many endpoints in the rotation phase, and to provide edges for $E^+$, which are used to close cycles, if needed.

https://doi.org/10.1017/S0963548318000159 Published online by Cambridge University Press

**FOR** $t = 1, 2, \ldots, \tau_q$ **DO**
**BEGIN**
Step 1. Let $e_t = uv$.
Step 2. If $C_v^{(t)} \cup C_u^{(t)} = \emptyset$, $t > t_\varepsilon$, and precisely one of $\{u, v\}$ (without loss of generality $u$) is *Full*, then give $uv$ the colour $c$ that minimizes $d_c(v)$ (breaking ties arbitrarily). Add $uv$ to $E_c^*$.
Step 3. If $C_v^{(t)} \cup C_u^{(t)} = \emptyset$, $t > t_\varepsilon$ and *both* $u, v \in$ *Full*, give $uv$ a colour $c$ uniformly at random from $[q]$. Then add this edge to $E_c^+$ or $E_c^*$, each with probability $1/2$.
Step 4. If $C_v^{(t)} \cup C_u^{(t)} = \emptyset$ but $t \leqslant t_\varepsilon$ or *both* $u, v \notin$ *Full*, then colour $uv$ with colour $c$ chosen uniformly at random from $[q]$. Add $uv$ to $E_c^*$.
Step 5. Otherwise, colour $uv$ with colour $c$ chosen uniformly at random from $C_u^{(t)} \cup C_v^{(t)}$. Add $uv$ to $E_c^*$.
**END**

Let

$$E^* = \bigcup_{c \in [q]} E_c^* \quad \text{and} \quad E^+ = \bigcup_{c \in [q]} E_c^+.$$

## 3. Structural properties

Let

$$p = \frac{\log n + (q-1)\log\log n - \omega}{n} \quad \text{and} \quad m = \binom{n}{2} p$$

where

$$\omega = \omega(n) \to \infty, \quad \omega = o(\log\log n).$$

We will use the following well-known properties relating $G_{n,p}$ and $G_{n,m}$: see *e.g.* [7, Chapter 1]. Let $\mathcal{P}$ be a graph property. It is monotone increasing if adding an edge preserves it, and it is monotone decreasing if deleting an edge preserves it. We have

$$\mathbb{P}(G_{n,m} \in \mathcal{P}) \leqslant 10 m^{1/2} \mathbb{P}(G_{n,p} \in \mathcal{P}), \tag{3.1}$$

$$\mathbb{P}(G_{n,m} \in \mathcal{P}) \leqslant 3 \mathbb{P}(G_{n,p} \in \mathcal{P}) \quad \text{if } \mathcal{P} \text{ is monotone.} \tag{3.2}$$

A vertex $v \in [n]$ is *small* if its degree $d(v)$ in $G_{n,m}$ satisfies $d(v) < \log n/(100q)$. It is *large* otherwise. The set of small vertices is denoted by *SMALL* and the set of large vertices is denoted by *LARGE*.

**Definition.** A subgraph $H$ of $G_{n,m}$ with a subset $S(H) \subset V(H)$ is called a *small structure* if

$$|E(H)| + |S(H)| - |V(H)| \geqslant 1.$$

We say that $G_{n,m}$ *contains* $H$ if there is an injective homomorphism $\phi : H \hookrightarrow G_{n,m}$ such that $\phi(S(H)) \subseteq$ *SMALL*. The important examples of $H$ include:

https://doi.org/10.1017/S0963548318000159 Published online by Cambridge University Press

- a single edge between two *small* vertices,
- a path of length at most five between two *small* vertices,
- a copy of $C_3$ or $C_4$ with at least one *small* vertex,
- two distinct triangles sharing at least one vertex.

**Lemma 3.1.** *For any fixed small structure $H$ of constant size,*

$$\mathbb{P}(G_{n,m} \text{ contains } H) = o(n^{-1/5}).$$

**Proof.** We will prove that

$$\mathbb{P}(G_{n,p} \text{ contains } H) = o(n^{-3/4}). \tag{3.3}$$

This along with (3.1) implies the lemma.

Let $h = |V(H)|, f = |E(H)|, s = |S(H)|$ so that $f + s \geqslant h + 1$. Then

$$\begin{aligned}
\mathbb{P}(G_{n,p} \text{ contains } H) &\leqslant \binom{n}{h} h! p^f \left( \sum_{i=0}^{\log n/(100q)} \binom{n-h}{i} p^i (1-p)^{n-h-i} \right)^s \\
&\lesssim n^h \left( \frac{\log n}{n} \right)^f \left( \sum_{i=0}^{\log n/(100q)} \left( \frac{(e+o(1))\log n}{i} \right)^i e^{-\log n - (q-1)\log\log n + \omega + o(1)} \right)^s \\
&\leqslant n^h \left( \frac{\log n}{n} \right)^f \left( \frac{(300q)^{\log n/(100q)}}{n(\log n)^{q-1-o(1)}} \right)^s \\
&= o(n^{h-f-s+1/4}) = o(n^{-3/4}).
\end{aligned}$$

(We used the notation $A \lesssim B$ in place of $A \leqslant (1+o(1))B$.) In the calculation above, in the first line we placed the vertices of $H$ and decided about the identity of $s$ vertices falling into *SMALL*, then required that all $f$ edges of $H$ are present in $G_{n,p}$, and finally required that for each of the $s$ vertices in *SMALL*, their degree outside the copy of $H$ is at most $\log n/(100q)$. □

**Lemma 3.2.** *With high probability, for every*

$$k \in \left[ q-1, \frac{\log n}{100q} \right],$$

*there are fewer than*

$$\nu_k = \frac{e^{2\omega}(\log n)^{k-q+1}}{(k-1)!}$$

*vertices of degree $k$ in $G_{n,m}$.*

**Remark 1.** $\nu_k$ is increasing in $k$ for this range, and for the largest $k = \log n/(100q)$ we have

$$\nu_k \lesssim n^{\log(100eq)/(100q)}.$$

https://doi.org/10.1017/S0963548318000159 Published online by Cambridge University Press

**Proof.** Fix $k$, and then we have

$$\begin{aligned}
&\mathbb{P}(G_{n,p} \text{ has at least } \nu_k \text{ vertices of degree at most } k)\\
&\leqslant \binom{n}{\nu_k}\left(\sum_{\ell=0}^{k}\binom{n-\nu_k}{\ell}p^{\ell}(1-p)^{n-\nu_k-\ell}\right)^{\nu_k}\\
&= \binom{n}{\nu_k}\left((1+o(1))\binom{n-\nu_k}{k}p^{k}(1-p)^{n-\nu_k-k}\right)^{\nu_k}\\
&\leqslant \left(\frac{ne}{\nu_k}\times\frac{n^k}{k!}\left(\frac{\log n+(q-1)\log\log n-\omega}{n}\right)^{k}e^{-\log n-(q-1)\log\log n+\omega+o(1)}\right)^{\nu_k}\\
&\leqslant \left(\frac{e^{\omega+O(1)}}{(\log n)^{q-1}}\frac{(\log n+q\log\log n)^k}{k!\nu_k}\right)^{\nu_k}\\
&= \left(\frac{e^{-\omega+O(1)}}{k}\left(1+\frac{q\log\log n}{\log n}\right)^{k}\right)^{\nu_k}\\
&\leqslant \left(e^{-\omega+O(1)}\frac{(\log n)^{kq/\log n}}{k}\right)^{\nu_k}.
\end{aligned}$$

The function

$$f(k)=\frac{(\log n)^{kq/\log n}}{k}$$

is log-convex, and so $f$ is maximized at the extreme values of $k$, specifically

$$f(q-1)=e^{O(1)}>f\left(\frac{\log n}{100q}\right)=o(1).$$

Hence,

$$\mathbb{P}(\exists k: G_{n,p} \text{ has at least } \nu_k \text{ vertices of degree } k)\leqslant\sum_{k=q-1}^{\log n/(100q)}e^{-\omega\nu_k/2}=o(1).$$

Applying (3.2), we see that

$$\mathbb{P}(\exists k: G_{n,m} \text{ has at least } \nu_k \text{ vertices of degree } k)=o(1),$$

which is stronger than required. □

**Lemma 3.3.** *With probability* $1-o(n^{-10})$, $G_{n,m}$ *has no vertices of degree* $\geqslant 20\log n$.

**Proof.** We will prove that w.h.p. $G_{n,p}$ has the stated property. We can then obtain the lemma by applying (3.2), that is,

$$\begin{aligned}
\mathbb{P}(\exists v: d(v)\geqslant 20\log n)&\leqslant n\binom{n-1}{20\log n}p^{20\log n}\\
&\leqslant n\left(\frac{en}{20\log n}\frac{2\log n}{n}\right)^{20\log n}
\end{aligned}$$

https://doi.org/10.1017/S0963548318000159 Published online by Cambridge University Press

$$\leqslant n\left(\frac{e}{10}\right)^{20\log n}$$
$$= o(n^{-10}).$$ □

## 4. Analysis of COL

Let $\Gamma = G_m$ and let $d(v)$ denote the degree of $v \in [n]$ in $\Gamma$. Let

$$\theta_v = \begin{cases} 0 & d(v) \geqslant q, \\ 1 & d(v) = q-1. \end{cases}$$

**Lemma 4.1.** *Suppose we run COL as described above. Then w.h.p. $|C_v^{(m)}| = \theta_v$ for all $v \in [n]$.*

In words, Lemma 4.1 guarantees that the algorithm COL typically performs so that at time $m$, each vertex of degree at least $q$ has all colours present at its incident edges, while each vertex of degree $q-1$ has exactly one colour missing. (It is well known that w.h.p. $\delta(G_m) = q-1$: see *e.g.* [7, Section 4.2].)

**Proof.** Fix $v$ and suppose $v$ has $k$ neighbours in *LARGE*, via edges $\{f_i = vu_i\}_{i=1}^k$. Then in general $d(v)-1 \leqslant k \leqslant d(v)$ as small vertices do not share a path of length two. Also, when $v$ is small, $k = d(v)$. Write $t(e)$ for the time $t \in [1,m]$ at which an edge $e$ appears in the random graph process, that is, $t(e_i) = i$. Let $t_i = t(f_i)$ and assume that $t_i < t_{i+1}$ for $i > 0$. We omit $i = 1$ in the next consideration since $v$ will always get a colour it needs by time $t_1$. (It may get a colour before $t_1$ through an edge $vw$ where $w$ is not in *LARGE*.) Every time an $f_i, i \geqslant 2$ appears while $u_i$ needs no additional colours, $v$ gets a colour it needs. So for $v$ to have $|C_v^{(m)}| > \theta_v$ at the end of the process, this must happen at most $q-2-\theta_v$ times, so there is certainly some set

$$S = \{i_1 < i_2 < \cdots < i_s\} \subseteq [2,k] \text{ of } s = k-q+1+\theta_v \text{ indices}$$

whose corresponding edges $\{f_i, i \in S\}$ incident with $v$ satisfy $C_{u_i}^{(t_i)} \neq \emptyset$. Let $\mathbf{T_S}$ denote $\{t_i : i \in S\}$ and let $\mathbf{U}$ denote the sequence $u_1, u_2, \ldots, u_k$. In the following we will sum over $S$ and condition on the choices for $\mathbf{T_S}$ and then estimate the probability that $C_{u_i}^{(t_i)} \neq \emptyset$ for $i \in S$. For a fixed $S$ there will be at least $\binom{m-k}{|S|+1}$ equally likely choices for the set $\{t_i, i \in \{1\} \cup S\}$. (We do not condition on $t_1$. The factor $t_{i_1} - 1$ in (4.2) below will allow for the number of choices for $t_1$.) Let $\mathcal{L}$ denote the occurrence of the bound of $20\log n$ on the degree of $v$ and its neighbours (see Lemma 3.3), and note that $\mathbb{P}(\mathcal{L}) = 1 - o(n^{-10})$.

Taking a union bound over $S$, and letting

$$A_i := \{C_{u_i}^{(t_i)} \neq \emptyset\},$$

https://doi.org/10.1017/S0963548318000159 Published online by Cambridge University Press

we have

$$\mathbb{P}(|C_v^{(m)}| > \theta_v \mid \mathcal{L}, \mathbf{U}) \leqslant \sum_{\substack{S\subset[2,k]\\|S|=s}} \sum_{t_i : i\in\{1\}\cup S} \frac{1}{\binom{m-k}{k-q+2+\theta_v}} \mathbb{P}\left( \bigwedge_{i\in S} A_i \;\middle|\; \mathbf{T_S}, \mathbf{U}, \mathcal{L} \right) \tag{4.1}$$

$$\approx \sum_{\substack{S\subset[2,k]\\|S|=s}} \sum_{t_i : i\in S} \frac{t_{i_1}-1}{\binom{m}{k-q+2+\theta_v}} \mathbb{P}\left( \bigwedge_{i\in S} A_i \;\middle|\; \mathbf{T_S}, \mathbf{U}, \mathcal{L} \right), \tag{4.2}$$

since there are $t_{i_1}-1$ choices for $t_1$ and $k^2 = o(m)$, implying

$$\binom{m-k}{k-q+2+\theta_v} \approx \binom{m}{k-q+2+\theta_v},$$

given $\mathcal{L}$. Next let

$$\begin{aligned} Y_i &= \{\text{edges of } u_i \text{ that appeared before } t_i \text{ excluding edges contained in } N^{(m)}(v)\},\\ d_r &= d(u_r) \quad \text{and} \quad Z_r := |Y_r| \quad \text{for } r = 1,2,\dots,s,\\ \mathbf{D_S} &= \{d_i : i \in S\}. \end{aligned}$$

Now fix $\mathbf{U}$ and $S$ and $\mathbf{T_S}$ and $\mathbf{D_S}$.

**Remark 2.** Going back to Algorithm COL, we observe that Step 5 implies that if $C_v^{(t)} \neq \emptyset$ then $uv$ is coloured with a colour in $C_v^{(t)}$ with probability at least $1/q$. This holds regardless of the previous history of the algorithm, and also holds conditional on $\mathbf{T_S}, \mathbf{U}, \mathbf{D_S}$. Indeed, the random bits used in Step 5 are independent of the history and are distinct from those used to generate the random graphs. The latter explains why we can condition on the future by fixing $\mathbf{T_S}, \mathbf{U}, \mathbf{D_S}$. We condition on $\mathcal{L}$ in order to control $s$ as $O(\log n)$.

Then,

$$\mathbb{P}(A_{i_1} \wedge \cdots \wedge A_{i_s} \mid \mathbf{T_S}, \mathbf{U}, \mathbf{D_S}, \mathcal{L}) \tag{4.3}$$

$$= \sum_{z_s} \underbrace{\mathbb{P}(A_{i_s} \mid A_{i_1}, \dots, A_{i_{s-1}}, Z_s = z_s, \mathbf{T_S}, \mathbf{U}, \mathbf{D_S}, \mathcal{L})}_{\leqslant \mathbb{P}(\mathrm{Bin}(z_s, q^{-1}) \leqslant q-1) \text{ by Remark 2}} \mathbb{P}(A_{i_1}, \dots, A_{i_{s-1}}, Z_s = z_s \mid \mathbf{T_S}, \mathbf{U}, \mathbf{D_S}, \mathcal{L}) \tag{4.4}$$

$$\leqslant \sum_{z_s} g(z_s) \sum_{z_{s-1}} \mathbb{P}(A_{i_{s-1}} \mid A_{i_1}, \dots, A_{i_{s-2}}, Z_{s-1} = z_{s-1}, Z_s = z_s, \mathbf{T_S}, \mathbf{U}, \mathbf{D_S}, \mathcal{L}) \tag{4.5}$$

$$\times \mathbb{P}(A_{i_1}, \dots, A_{i_{s-2}}, Z_{s-1} = z_{s-1}, Z_s = z_s \mid \mathbf{T_S}, \mathbf{U}, \mathbf{D_S}, \mathcal{L}) \tag{4.6}$$

$$\leqslant \sum_{z_s, z_{s-1}} g(z_s) g(z_{s-1}) \mathbb{P}(A_{i_1}, \dots, A_{i_{s-2}}, Z_{s-1} = z_{s-1}, Z_s = z_s \mid \mathbf{T_S}, \mathbf{U}, \mathbf{D_S}, \mathcal{L}) \tag{4.7}$$

$$\leqslant \sum_{z_s, \dots, z_1} g(z_s) \cdots g(z_2) \mathbb{P}(Z_r = z_r, r = 1, \dots, s \mid \mathbf{T_S}, \mathbf{U}, \mathbf{D_S}, \mathcal{L}) \quad \text{(by induction)}. \tag{4.8}$$

Here $g(z) := \mathbb{P}(\mathrm{Bin}(z, q^{-1}) \leqslant q-1)$ for any $z \geqslant 0$.

https://doi.org/10.1017/S0963548318000159 Published online by Cambridge University Press

**Claim 4.2.**

$$\mathbb{P}(Z_r = z_r, r = 1,2,\dots,s \mid \mathbf{T_S}, \mathbf{U}, \mathbf{D_S}, \mathcal{L}) \leqslant (1+\tilde{O}(n^{-1}))\prod_{r=1}^{s} \frac{\binom{t_r}{z_r}\binom{m-t_r}{d_r-z_r}}{\binom{m}{d_r}},$$

*where $\tilde{O}$ hides polylog factors.*

**Proof.** Fix

$$\frac{\log n}{100q} \leqslant d_1, d_2, \dots, d_s = O(\log n) \quad \text{and} \quad t_1, t_2, \dots, t_s.$$

Then, for every $1 \leqslant r \leqslant s$,

$$\mathbb{P}(Z_r = z_r \mid Z_{r-1} = z_{r-1}, \dots, Z_1 = z_1, \mathbf{T_S}, \mathbf{U}, \mathbf{D_S}, \mathcal{L}) \leqslant (1+o(n^{-10}))\frac{\binom{t_r}{z_r}\binom{m-t_r}{d_r-z_r}}{\binom{m-d_2-\cdots-d_{r-1}-s}{d_r}} \tag{4.9}$$

$$\leqslant (1+\tilde{O}(n^{-1}))\frac{\binom{t_r}{z_r}\binom{m-t_r}{d_r-z_r}}{\binom{m}{d_r}}. \tag{4.10}$$

**Explanation for (4.9).** The first binomial coefficient in the numerator in (4.9) bounds the number of choices for the $z_r$ positions in the sequence where an edge contributing $Y_r$ occurs. This holds regardless of $z_1, z_2, \dots, z_{r-1}$. The second binomial coefficient bounds the number of choices for the $d_r - z_r$ positions in the sequence where we choose an edge incident with $u_r$ after time $t_r$. Conversely, the denominator in (4.9) is a lower bound on the number of choices for the $d_r$ positions where we choose an edge incident with $u_r$, given $d_1, d_2, \dots, d_{r-1}$. We subtract the extra $s$ to (over)count for edges from $v$ to $u_{r+1}, \dots, u_s$. The factor $(1+o(n^{-10}))$ accounts for the conditioning on $\mathcal{L}$.

Expanding $\mathbb{P}(Z_r = z_r, r = 1, \dots, s \mid \mathbf{T_S}, \mathbf{U}, \mathbf{D_S}, \mathcal{L})$ as a product of $s = O(\log n)$ of these terms completes the proof of Claim 4.2. □

Going back to (4.8), we see that given $d_1, d_2, \dots, d_s$,

$$\mathbb{P}(A_{i_1} \wedge \cdots \wedge A_{i_s} \mid \mathbf{T_S}, \mathbf{U}, \mathbf{D_S}, \mathcal{L}) \tag{4.11}$$

$$\lesssim \prod_{r=1}^{s}\sum_{z_r=0}^{d_r}\left(\mathbb{P}(\mathrm{Bin}(z_r, q^{-1}) \leqslant q-1) \times \frac{\binom{t_r}{z_r}\binom{m-t_r}{d_r-z_r}}{\binom{m}{d_r}}\right) \tag{4.12}$$

$$\leqslant \prod_{r=1}^{s}\sum_{z_r=0}^{d_r}\left(C_1\binom{z_r}{\min\{z_r, q-1\}}\frac{1}{q^{q-1}}\left(1-\frac{1}{q}\right)^{z_r} \times \frac{\binom{t_r}{z_r}\binom{m-t_r}{d_r-z_r}}{\binom{m}{d_r}}\right) \tag{4.13}$$

$$\leqslant \prod_{r=1}^{s}\sum_{z_r=0}^{d_r}\left(C_1 \max\{1, z_r^{q-1}\}e^{-z_r/q} \times \frac{\binom{t_r}{z_r}\binom{m-t_r}{d_r-z_r}}{\binom{m}{d_r}}\right). \tag{4.14}$$

Here, $C_1 = C_1(q)$ depends only on $q$. We will use constants $C_2, C_3, \dots$ in a similar fashion without further comment.

https://doi.org/10.1017/S0963548318000159 Published online by Cambridge University Press

**Justification for (4.13).** If $z_r \leqslant q-1$ then $\mathbb{P}(\mathrm{Bin}(z_r, q^{-1}) \leqslant q-1) = 1$, and $C_1 = eq^q$ will suffice. If $q \leqslant z_r \leqslant 10q$ we use

$$\mathbb{P}(\mathrm{Bin}(z_r, q^{-1}) \leqslant q-1) \leqslant 1 \quad \text{and} \quad \binom{z_r}{q-1}\frac{1}{q^{q-1}}\left(1-\frac{1}{q}\right)^{z_r} \geqslant \frac{1}{q^{q-1}}\left(1-\frac{1}{q}\right)^{10q},$$

and $C_1 = e^{20}q^q$ will suffice in this case.

If $z_r > 10q$, then putting

$$a_i := \mathbb{P}(\mathrm{Bin}(z_r, q^{-1}) = i) = \binom{z_r}{i}\frac{1}{q^i}\left(1-\frac{1}{q}\right)^{z_r-i}$$

for $i \leqslant q-1$, we see that

$$\frac{a_i}{a_{i-1}} = \frac{z_r - i + 1}{i}\cdot\frac{1}{q-1} \geqslant \frac{z_r - q}{q^2} > \frac{z_r}{2q^2} \geqslant \frac{5}{q}.$$

So here

$$\begin{aligned}\mathbb{P}(\mathrm{Bin}(z_r, q^{-1}) \leqslant q-1) &= \sum_{i=0}^{q-1} a_i \leqslant a_{q-1}\left(1 + \frac{2q^2}{z_r} + \cdots + \left(\frac{2q^2}{z_r}\right)^{q-2}\right)\\ &\leqslant \left(1-\frac{1}{q}\right)^{1-q}\left(\binom{z_r}{q-1}\frac{1}{q^{q-1}}\left(1-\frac{1}{q}\right)^{z_r}\right)\frac{(q/5)^{q-1}-1}{q/5-1},\end{aligned}$$

and thus $C_1 = (5q)^q$ suffices. This completes the verification of (4.13).

Now, writing $(t)_z$ for the *falling factorial* $t!/(t-z)! = t(t-1)(t-2)\dots(t-z+1)$,

$$\frac{\binom{t_r}{z_r}\binom{m-t_r}{d_r-z_r}}{\binom{m}{d_r}} = \binom{d_r}{z_r}\frac{(t_r)_{z_r}(m-t_r)_{d_r-z_r}}{(m)_{d_r}} \tag{4.15}$$

$$= \binom{d_r}{z_r}\prod_{i=0}^{z_r-1}\frac{t_r - i}{m-(d_r-z_r)-i}\cdot\prod_{i=0}^{d_r-z_r-1}\frac{m-t_r-i}{m-i} \tag{4.16}$$

$$\leqslant \left(1+O\left(\frac{d_r^2}{m}\right)\right)\binom{d_r}{z_r}\left(\frac{t_r}{m}\right)^{z_r}\left(1-\frac{t_r}{m}\right)^{d_r-z_r}. \tag{4.17}$$

Observe next that if $z_r \geqslant q^2$ then

$$(z_r)_{q-1} = z_r^{q-1}\prod_{i=0}^{q-1}\left(1-\frac{i}{z_r}\right) \geqslant z_r^{q-1}\left(1-\frac{q^2}{2z_r}\right) \geqslant \frac{z_r^{q-1}}{2}. \tag{4.18}$$

It follows from (4.17) and (4.18) that

$$\sum_{z_r=q^2}^{d_r} C_1 z_r^{q-1}e^{-z_r/q} \times \frac{\binom{t_r}{z_r}\binom{m-t_r}{d_r-z_r}}{\binom{m}{d_r}} \tag{4.19}$$

$$\leqslant 2C_1\sum_{z_r=q-1}^{d_r}(z_r)_{q-1}\binom{d_r}{z_r}\left(\frac{t_r e^{-1/q}}{m}\right)^{z_r}\left(1-\frac{t_r}{m}\right)^{d_r-z_r} \tag{4.20}$$

$$\leqslant 2C_1(d_r)_{q-1}\left(\frac{t_r}{m}\right)^{q-1}\sum_{z_r=q-1}^{d_r}\binom{d_r-q+1}{z_r-q+1}\left(\frac{t_r e^{-1/q}}{m}\right)^{z_r-q+1}\left(1-\frac{t_r}{m}\right)^{d_r-z_r} \tag{4.21}$$

https://doi.org/10.1017/S0963548318000159 Published online by Cambridge University Press

$$\leqslant 2C_1\left(\frac{d_r t_r}{m}\right)^{q-1}\left(1-\frac{t_r}{m}(1-e^{-1/q})\right)^{d_r-q+1} \tag{4.22}$$

$$\leqslant 2C_1\left(\frac{d_r t_r}{m}\right)^{q-1}\exp\left\{-\frac{(d_r-q+1)t_r}{m}(1-e^{-1/q})\right\}. \tag{4.23}$$

Furthermore, we must not forget that

$$\sum_{z_r=0}^{q^2-1} C_1\max\{1,z_r^{q-1}\}e^{-z_r/q}\times\frac{\binom{t_r}{z_r}\binom{m-t_r}{d_r-z_r}}{\binom{m}{d_r}} \leqslant C_2\sum_{z_r=0}^{q^2-1}\frac{\binom{t_r}{z_r}\binom{m-t_r}{d_r-z_r}}{\binom{m}{d_r}} \tag{4.24}$$

$$\leqslant C_3\sum_{z_r=0}^{q^2-1} t_r^{z_r}\cdot\frac{(m-t_r)^{d_r-z_r}}{(d_r-z_r)!}\cdot\frac{d_r!}{m^{d_r}} \tag{4.25}$$

$$\leqslant C_3\sum_{z_r=0}^{q^2-1}\left(\frac{d_r t_r}{m}\right)^{z_r} e^{-(d_r-z_r)t_r/m} \tag{4.26}$$

$$\leqslant C_4\psi\left(\frac{d_r t_r}{m}\right), \tag{4.27}$$

where $\psi(x)=e^{-x}\sum_{z=0}^{q^2-1}x^z$. (Now $z_r\leqslant q^2$ and so the factor $e^{z_r t_r/m}\leqslant e^{q^2}$ can be absorbed into $C_4$.) Going back to (4.14), we have

$$\begin{aligned}&\mathbb{P}(A_{i_1}\wedge\cdots\wedge A_{i_s}\mid \mathbf{T_S},\mathbf{U},\mathbf{D_S},\mathcal{L})\\ &\quad\leqslant C_5^s\prod_{r=1}^{s}\left(\left(\frac{d_r t_r}{m}\right)^{q-1}\exp\left\{-\frac{d_r t_r}{m}(1-e^{-1/q})\right\}+\psi\left(\frac{d_r t_r}{m}\right)\right).\end{aligned} \tag{4.28}$$

It follows from (4.2) and (4.28) that

$$\begin{aligned}p_v &:= \mathbb{P}(|C_v^{(m)}|>\theta_v\mid \mathbf{T_S},\mathbf{U},\mathbf{D_S},\mathcal{L})\\ &\leqslant \sum_{\substack{S\subset[2,k]\\|S|=s}}\sum_{t_i:i\in S}\frac{t_{i_1}C_5^s}{\binom{m}{s+1}}\prod_{r=1}^{s}\left(\left(\frac{d_r t_r}{m}\right)^{q-1}\exp\left\{-\frac{d_r t_r}{m}(1-e^{-1/q})\right\}+\psi\left(\frac{d_r t_r}{m}\right)\right).\end{aligned}$$

Replacing a sum of products with a product of sums and dividing by $s!$ to account for repetitions, we get

$$p_v\leqslant\sum_{\substack{S\subset[2,k]\\|S|=s}}\frac{C_5^s}{\binom{m}{s+1}s!}\prod_{r=2}^{s}\left(\sum_{t=1}^{m}\left(\left(\frac{d_r t}{m}\right)^{q-1}\exp\left\{-\frac{d_r t}{m}(1-e^{-1/q})\right\}+\psi\left(\frac{d_r t}{m}\right)\right)\right) \tag{4.29}$$

$$\times\left(\sum_{t=1}^{m}\left(t\left(\frac{d_1 t}{m}\right)^{q-1}\exp\left\{-\frac{d_1 t}{m}(1-e^{-1/q})\right\}+t\psi\left(\frac{d_1 t}{m}\right)\right)\right). \tag{4.30}$$

https://doi.org/10.1017/S0963548318000159 Published online by Cambridge University Press

We now replace the sums with integrals. This is valid seeing as the summands have a bounded number of extrema, and we replace $C_5$ with $C_6$ to absorb any small error factors. Thus

$$p_v \leqslant \sum_{\substack{S\subset[2,k]\\|S|=s}} \frac{C_6^s}{\binom{m}{s+1}s!}\prod_{r=2}^{s}\left(\int_{t=0}^{\infty}\left[\left(\frac{d_r t}{m}\right)^{q-1}\exp\left\{-\frac{d_r t}{m}(1-e^{-1/q})\right\}+\psi\left(\frac{d_r t}{m}\right)\right]dt\right) \tag{4.31}$$

$$\times\left(\int_{t=0}^{\infty}\left(t\left(\frac{d_1 t}{m}\right)^{q-1}\exp\left\{-\frac{d_1 t}{m}(1-e^{-1/q})\right\}+\psi\left(\frac{d_1 t}{m}\right)\right)dt\right) \tag{4.32}$$

$$= \sum_{\substack{S\subset[2,k]\\|S|=s}} \frac{C_6^s}{\binom{m}{s+1}s!}\prod_{r=2}^{s}\left(\frac{m}{d_r}\int_{x=0}^{\infty}(x^{q-1}e^{-(1-e^{-1/q})x}+\psi(x))\,dx\right) \tag{4.33}$$

$$\times\frac{m^2}{d_1^2}\left(\int_{x=0}^{\infty}(x^q\exp\{-(1-e^{-1/q})x\})+x\psi(x)\,dx\right) \tag{4.34}$$

$$\leqslant \sum_{\substack{S\subset[2,k]\\|S|=s}} \frac{C_6^s}{\binom{m}{s+1}s!}\cdot\left(\frac{C_7 m}{\min_r\{d_r\}}\right)^{s+1} \tag{4.35}$$

$$\leqslant \frac{C_8^k}{(\log n)^{k-q+2+\theta_v}}. \tag{4.36}$$

Applying Lemmas 3.2 and 3.3 and removing the conditioning on $\mathbf{T_S}, \mathbf{U}, \mathbf{D_S}, \mathcal{L}$, we see that with $k_0 = \log n/(100q)$,

$$\begin{aligned}
&\mathbb{P}(\exists v : |C_v^{(m)}| > \theta_v)\\
&\quad\leqslant \mathbb{P}(\neg\mathcal{L}) + \sum_{k=q-1}^{k_0}\frac{e^{2\omega}(\log n)^{k-q+1}}{(k-1)!}\times\frac{C_8^k}{(\log n)^{k-q+2+\theta_v}} + n\sum_{k=k_0}^{20\log n}\frac{C_8^k}{(\log n)^{k-q+2}}\\
&\quad\leqslant o(1)+\frac{e^{2\omega}}{\log n}\sum_{k\geqslant q-1}\frac{C_8^k}{(k-1)!}+n\sum_{k=k_0}^{20\log n}\frac{C_8^k}{(\log n)^{k/2}}\\
&\quad\leqslant o(1)+\frac{C_9 e^{2\omega+C_9}}{\log n}\\
&\quad= o(1).
\end{aligned}$$

□

We show next that at time $m$, w.h.p. sets of size up to $\Omega(n)$ have large neighbourhoods in every colour.

We first prove that typically 'large-degree vertices have large degree in every colour': let $d_c^*(v)$ denote the number of edges incident with $v$ that COL colours $c$, except for those edges that are coloured in Step 3.

**Theorem 4.3.** *There exists $\varepsilon = \varepsilon(q) > 0$ such that w.h.p. on completion of COL every $v \in$ LARGE has*

$$d_c^*(v) \geqslant \frac{\varepsilon\log n}{1000q} \quad \textit{for all } c \in [q].$$

https://doi.org/10.1017/S0963548318000159 Published online by Cambridge University Press

Suppose we define a vertex to be *small*$_c$ if it has $d_c(v) \leqslant \varepsilon \log n/(1000q)$. Theorem 4.3 says w.h.p. the set of *small*$_c$ vertices $SMALL_c \subset SMALL$ so that by Lemma 3.1, $G$ does not contain any *small*$_c$ structures of constant size. Here a *small*$_c$ structure is a small structure made up of *small*$_c$ vertices.

The proof of Theorem 4.3 will follow from Lemmas 4.4, 4.5 and 4.6 below.

**Lemma 4.4.** *There exists $\delta = \delta(q) > 0$ such that the following holds w.h.p. Let $Full', Full$ be as in Definition 2. Then $|Full'| \geqslant n - 203qn/(\varepsilon \log n)$ and $|Full| \geqslant n - n^{1-\delta}$.*

**Proof.** We first note that for $v \in [n]$, if $t_\varepsilon = \varepsilon n \log n$ then

$$\mathbb{P}\left(d^{(t_\varepsilon/2)}(v) < \lambda_0 := \frac{\varepsilon \log n}{100}\right) \leqslant 3n^{-\varepsilon/2} < n^{-\varepsilon/3}. \tag{4.37}$$

Indeed, with $p_1 = (t_\varepsilon/2)/\binom{n}{2}$ we see that in the random graph model $G_{n,p_1}$,

$$\begin{aligned}\mathbb{P}(d(v) < \lambda_0) &= \sum_{i=0}^{\lambda_0 - 1} \binom{n}{i} p_1^i (1-p_1)^{n-i} \\ &\leqslant 2\binom{n}{\lambda_0} p_1^{\lambda_0} (1-p_1)^{n-\lambda_0} \\ &\leqslant 2\left(\frac{nep_1}{\lambda_0}\right)^{\lambda_0} n^{-\varepsilon + o(1)} \leqslant n^{-\varepsilon/2}.\end{aligned} \tag{4.38}$$

The first inequality follows from the fact that the ratio of successive summands in the sum is at least $(n-\lambda_0)p_1/\lambda_0 > 50$.

Equation (4.37) now follows from (3.2) (with $p$ replaced by $p_1$) and (4.38).

Thus the Markov inequality shows that with probability at least $1 - n^{-\varepsilon/3}$, at least $n - n^{1-\varepsilon/6}$ of the vertices $v$ have $d^{(t_\varepsilon/2)}(v) \geqslant \varepsilon \log n/100$. Now note that at most $qn$ of the first $t_\varepsilon/2$ edges were restricted in colour by being incident to at least one needy vertex. This is because each time a needy vertex gets an edge incident to it, the total number of needed colours decreases by at least one. Therefore at most $200qn/(\varepsilon \log n)$ of these vertices $v$ have fewer than $\varepsilon \log n/200$ of their $\geqslant \varepsilon \log n/100$ initial edges coloured completely at random, as in Step 4 of COL. Hence, there are at least $n - 201qn/(\varepsilon \log n)$ vertices $v$ which have $d^{(t_\varepsilon/2)}(v) \geqslant \varepsilon \log n/100$ and $\varepsilon \log n/200$ edges of fully random colour. For such a $v$, and any colour $c$,

$$\begin{aligned}\mathbb{P}\left(d_c^{(t_\varepsilon/2)}(v) < \frac{\varepsilon \log n}{1000q}\right) &\leqslant \mathbb{P}\left(\mathrm{Bin}\left(\frac{\varepsilon \log n}{200}, \frac{1}{q}\right) \leqslant \frac{\varepsilon \log n}{1000q}\right) \\ &\leqslant \exp\left\{-\frac{1}{2} \cdot \frac{16}{25} \cdot \frac{\varepsilon \log n}{200q}\right\} \leqslant n^{-\varepsilon/1000q}.\end{aligned} \tag{4.39}$$

So $\mathbb{P}(v \notin Full') \leqslant qn^{-\varepsilon/1000q}$, and the Markov inequality shows that w.h.p.

$$|Full'| \geqslant n - \frac{201qn}{\varepsilon \log n} - n^{1-\varepsilon/2000q} \geqslant n - \frac{202qn}{\varepsilon \log n}. \tag{4.40}$$

Now, for $v \notin Full'$, let $S(v) := Full' \setminus N^{(t_\varepsilon/2)}(v)$. Since $d^{(t_\varepsilon/2)}(v) \leqslant d(v) \leqslant 20 \log n$, we have $|S(v)| \geqslant n - 203qn/(\varepsilon \log n)$. Furthermore, every $w \in S(v)$ is no longer needy, and so among the next $t_\varepsilon/2$

https://doi.org/10.1017/S0963548318000159 Published online by Cambridge University Press

edges, at most $q$ of the edges between $v$ and $S(v)$ have their choice of colour restricted by $v$, and the rest are coloured randomly as in Step 4 of COL. Now

$$\mathbb{P}\left(|e(v,S(v))| < \frac{\varepsilon \log n}{100}\right) = O(n^{-\varepsilon/3})$$

by a similar calculation to (4.38). Conditioning on $|e(v,S(v))| \geqslant \varepsilon \log n/100$ we have $\mathbb{P}(v \notin \mathit{Full}) \leqslant qn^{-\varepsilon/1000q}$ by a similar calculation to (4.39) and so $|\mathit{Full}| \geqslant n - n^{1-\varepsilon/2000q}$ with probability $\geqslant 1 - O(n^{-\varepsilon/2000q})$ by the Markov inequality. □

We are working towards showing that vertices with low degree in some colour must have a low overall degree. The point is that all *Full* vertices no longer need additional colours later than $t_\varepsilon = \varepsilon n \log n$, so any new edge connecting *Full* to $V \setminus \mathit{Full}$ after time $t_\varepsilon$ has its colour determined by the vertex not in *Full*, as in Step 2 of COL. Indeed, suppose a vertex $v \notin \mathit{Full}$ has at least $\varepsilon \log n/400$ edges to *Full* after time $t_\varepsilon$. Then $v$ gets at least $\varepsilon \log n/(400q) > \varepsilon \log n/(1000q)$ edges of every colour incident with it.

**Lemma 4.5.** *With high probability there are no vertices $v \notin \mathit{Full}$ with at least $\varepsilon \log n/200$ edges after time $t_\varepsilon$, i.e., $d^{(m)}(v) - d^{(t_\varepsilon)}(v) \geqslant \varepsilon \log n/200$ but with at most $\varepsilon \log n/400$ of these edges to Full.*

**Proof.** Take any vertex $v \notin \mathit{Full}$ and consider the first $\varepsilon \log n/200$ edges incident to $v$ after time $t_\varepsilon$. We must estimate the probability that at least half of these edges are to vertices not in *Full*. We bound this by

$$\binom{\varepsilon \log n/200}{\varepsilon \log n/400}\left(\frac{n^{1-\delta}}{n - 20\log n}\right)^{\varepsilon \log n/400} = o(1/n).$$

We subtract off a bound of $20\log n$ on the number of edges from $v$ to *Full* in $E_{t_\varepsilon}$. Note that we do not need to multiply by the number of choices for *Full*, as *Full* is defined by the first $t_\varepsilon$ edges. There are at most $n$ choices for $v$ and so the lemma follows. □

**Lemma 4.6.** *There are no* large *vertices $v$ with $d^{(m)}(v) - d^{(t_\varepsilon)}(v) < \varepsilon \log n/200$.*

**Proof.** Any $v$ satisfying these conditions must have $d^{(t_\varepsilon)}(v) \geqslant \log n/(200q)$, if $\varepsilon \leqslant 1/q$ say. However, with $p_2 = t_\varepsilon / \binom{n}{2}$ we have that in the random graph model $G_{n,p_2}$,

$$\mathbb{P}\left(d(v) \geqslant \frac{\log n}{200q}\right) \leqslant \binom{n}{\log n/200q} p_2^{\log n/200q} \leqslant (400qe\varepsilon)^{\log n/200q} = o(n^{-2}), \tag{4.41}$$

for $\varepsilon$ sufficiently small.

The result follows by taking a union bound over choices of $v$ and using (3.2) (again noting $\binom{n}{2} p_2 = t_\varepsilon \to \infty$). □

**Proof Proof of Theorem 4.3.** It follows from Lemmas 4.5 and 4.6 that every large vertex has at least $\varepsilon \log n/400$ edges to *Full* that occur after time $t_\varepsilon$. These edges will provide all needed edges of all colours. □

https://doi.org/10.1017/S0963548318000159 Published online by Cambridge University Press

It is known that w.h.p. $m \leqslant \tau_q \leqslant m' = m + 2\omega n$: see Erdős and Rényi [6]. We have shown that at time $m$ all vertices, other than vertices of degree $q-1$, are incident with edges of all colours. Furthermore, vertices of degree $q-1$ are only missing one colour. As we add the at most $2\omega n$ edges needed to reach $\tau_q$ we find (see Claim 4.7 below) that w.h.p. the edges we add incident to a vertex $v$ of degree $q-1$ have their other end in *LARGE*. As such COL will now give vertex $v$ its missing colour.

**Claim 4.7.** *With high probability an edge of $E_{m'} \setminus E_m$ that meets a vertex of degree $q-1$ in $G_m$ has its other end in LARGE.*

**Proof.** It follows from Lemma 3.2 that at time $m$ and later there are w.h.p. at most $e^{2\omega}$ vertices of degree $q-1$ and at most

$$M = \sum_{k=q-1}^{\log n/100q} \frac{e^{2\omega}(\log n)^{k-q+1}}{(k-1)!} \leqslant 2e^{2\omega}\left(\frac{e\log n}{\log n/100q}\right)^{-q+1+\log n/100q} \leqslant n^{1/3}$$

vertices in *SMALL*. (The first inequality follows from the fact that the summands grow by a factor of at least $100q$.)

Thus the probability that there is an edge contradicting the claim is at most

$$2\omega n \times \frac{e^{2\omega} \times n^{1/3}}{\binom{n}{2} - m'} = o(1). \qquad \square$$

We remind the reader that $q = 2\sigma$ where we only use $\sigma$ colours. We apply the above analysis by identifying colours $\bmod \sigma$. We therefore have the following.

**Corollary 4.8.** *With high probability the algorithm COL applied to $G_{\tau_{2\sigma}}$ yields a colouring for which $d_c^*(v) \geqslant 2$ for all $v \in [n]$.*

**Proof.** We can see from the above that w.h.p. at time $\tau_{2\sigma}$ we have that $d_c(v) \geqslant 2$ for all $c \in [\sigma], v \in [n]$. Furthermore, by construction, for each $c \in [q], v \in [n]$ the first edge incident with $v$ that gets colour $c$ will be in $E_c^*$. (The only time we place an edge in $E_c^+$ is when it joins two full vertices.) $\square$

From now on we think in terms of $\sigma$ colours.

### 4.1. Expansion

For a set $S \subseteq [n]$ we let

$$N_c^*(S) = \{v \notin S : \exists u \in S \; s.t. \; uv \in E_c^*\} \subset N_c(S).$$

Let

$$\alpha = \frac{1}{10^6 q}.$$

**Lemma 4.9.** *Then w.h.p. $|N_c^*(S)| \geqslant 19|S|$ for all $S \subset LARGE$, $|S| \leqslant \alpha n$.*

https://doi.org/10.1017/S0963548318000159 Published online by Cambridge University Press

**Claim 4.10.** *At time m, for every $R \subset V(G)$ with $|R| \leqslant n/(\log n)^3$, there are w.h.p. at most $2|R|$ edges within R of colour c for every $c \in [q]$.*

**Proof.** [Proof of claim.] We will show that w.h.p. every such $R$ does not have this many edges irrespective of colour. Note that the desired property is monotone decreasing, so it suffices to use (3.2) and show this occurs w.h.p. in $G_{n,p}$:

$$\begin{aligned}
&\mathbb{P}\left(\exists |R| \leqslant \frac{n}{(\log n)^3} : |E(G[R])| > 2|R|\right) \\
&\quad\leqslant \sum_{r=4}^{n/(\log n)^3} \binom{n}{r}\binom{\binom{r}{2}}{2r} p^{2r} \\
&\quad\leqslant \sum_{r=4}^{n/(\log n)^3} \left(\frac{ne}{r}\left(\frac{re^{1+o(1)}\log n}{4n}\right)^2\right)^r \\
&\quad\leqslant \sum_{r=4}^{n/(\log n)^3} \left(\frac{r}{n}\cdot\frac{e^{3+o(1)}(\log n)^2}{16}\right)^r = o(n^{-3}).
\end{aligned}$$

□

**Proof.** [Proof of Lemma 4.9.]

*Case 1.* $|S| \leqslant n/(\log n)^4$.

We may assume that $S \cup N^*_c(S)$ is small enough for Claim 4.10 to apply (otherwise $|N^*_c(S)| \geqslant n/(\log n)^3 - n/(\log n)^4$ so that $S$ actually has logarithmic expansion in colour $c$). Then, using $e_c$ to denote the number of edges in colour $c$, and using Theorem 4.3,

$$\frac{\varepsilon \log n}{1000q}|S| \leqslant \sum_{v\in S} d^*_c(v) = 2e_c(S) + e_c(S, N^*_c(S)) \leqslant 4|S| + 2|N^*_c(S) \cup S|.$$

Hence,

$$|N^*_c(S)| \geqslant \frac{\varepsilon \log n}{2001q}|S| \geqslant 19|S|,$$

which verifies the truth of the lemma for this case.

*Case 2.* $n/(\log n)^4 \leqslant |S| \leqslant n/(50\log n)$.

Let

$$m_+ := \frac{n\log n}{8q}.$$

Let $E^+_c, E^*_c$ denote the edges of $E^+, E^*$ respectively, which are coloured $c$. We begin by proving the following claim.

**Claim 4.11.** $|E^+_c|, |E^*_c| \geqslant m_+$ *w.h.p.*

**Proof.** Once *Full* has been formed, it follows from Lemma 4.4 that at most

$$(n^{1-\delta}(n - n^{1-\delta})) + \binom{n^{1-\delta}}{2} < 2n^{-\delta}\binom{n}{2}$$

https://doi.org/10.1017/S0963548318000159 Published online by Cambridge University Press

spaces remain in $E(Full, V \setminus Full)$ or $E(V \setminus Full)$. For each of the $m - t_\varepsilon \sim (1/2 - \varepsilon) n \log n$ edges appearing thereafter, since $\lesssim n \log n < n^{-\delta}\binom{n}{2}$ edges have been placed already, each has a probability $\geqslant 1 - 4n^{-\delta}$ of having both ends in *Full*, independently of what has happened previously. Applying the Chernoff bounds (see *e.g.* [7, Chapter 21.4]) we see that the probability that fewer than $\frac{1}{3} n \log n$ of these $(1/2 - \varepsilon) n \log n$ edges were between vertices in *Full* is at most $e^{-\Omega(n \log n)}$. We remind the reader that every edge with both endpoints in *Full* is randomly coloured and placed in $E^+$ or $E^*$ in Step 3 of COL.

So, we may assume there are at least $(3q)^{-1} n \log n$ of these edges in $E^+ \cup E^*$ of colour $c$ in expectation and then the Chernoff bounds imply that there are at least $(8q)^{-1} n \log n = m^+$ w.h.p. in both $E^+$ and $E^*$. □

Suppose there exists $S$ as above with

$$|N_c^*(S)| < \frac{\log n}{1000q} |S|.$$

For $F := S \cap Full$, note that $|F| \geqslant |S| - n^{1-\delta} = |S|(1 - o(1))$. Therefore

$$|N_c^*(F) \cap Full| < \frac{\log n}{1000q} |S| \leqslant \frac{\log n}{999q} |F|.$$

We will show that w.h.p. there are no such $F \subseteq Full$.

We consider the graphs $H_1 = G_{|Full|, m_+} \setminus E_{t_\varepsilon}$ and the corresponding independent model $H_2 = G_{|Full|, p_+} \setminus E_{t_\varepsilon}$ where $p_+ \sim \log n/(4qn)$. We will show that w.h.p. $H_2$ contains no set $F$ of the postulated size and small neighbourhood. Together with (3.2) (and $\binom{|Full|}{2} p_+ \to \infty$) this implies that w.h.p. $H_1$ has no such set either. Note that by Lemma 3.3, we see that w.h.p. at most $20|F| \log n$ edges of $E_{t_\varepsilon}$ are incident with $F$. This calculation is relevant because $(E^* \setminus E_{t^*})$'s only dependence on $E_{t_\varepsilon}$ is that it is disjoint from it.

Hence, in $H_2$,

$$\begin{aligned}
\mathbb{P}(\exists F) &\leqslant \sum_{f=(n-o(n))/(\log n)^4}^{n/50\log n} \sum_{k=0}^{(\log n/(999q))f} \binom{|Full|}{f} \binom{|Full|}{k} f^k p_+^k (1-p_+)^{(|Full|-k)f - 20f\log n} \\
&\leqslant \sum_{f=(n-o(n))/(\log n)^4}^{n/50\log n} \sum_{k=0}^{(\log n/(999q))f} \underbrace{\left(\frac{ne}{f}\right)^f \left(\frac{nf}{k} \cdot \frac{\log n}{qn}\right)^k e^{-nfp_+(1-o(1))}}_{u_{f,k}}.
\end{aligned}$$

Here, the ratio

$$\frac{u_{f,k+1}}{u_{f,k}} = \frac{f \log n}{q(k+1)} \left(\frac{k}{k+1}\right)^k \geqslant 999/e.$$

Therefore,

$$\begin{aligned}
\mathbb{P}(\exists F) &\leqslant 2 \sum_{f \sim n/(\log n)^4}^{n/50\log n} \left(\frac{ne}{f} \cdot (999)^{\log n/(999q)} n^{-1/5q}\right)^f \\
&\leqslant 2n\left(3(\log n)^4 n^{-1/10q}\right)^{(1-o(1))n/(\log n)^4} = o(1).
\end{aligned}$$

*Case 3.* $n/(50\log n) \leqslant |S| \leqslant n/(10^6 q)$.

https://doi.org/10.1017/S0963548318000159 Published online by Cambridge University Press

Choose any $S_1 \subset S$ of size $n/(50\log n)$, then

$$|N_c^*(S)| \geqslant |N_c^*(S_1)| - |S| \geqslant \frac{\log n}{1000q} \cdot \frac{n}{50\log n} - \frac{n}{10^6 q} = 19\alpha n \geqslant 19|S|.$$

The following corollary applies to the subgraph of $G_{\tau_{2\sigma}}$ induced by $E_c^*$.

**Corollary 4.12.** *With high probability $|N_c^*(S)| \geqslant 2|S|$ for all $S \subset V(G)$ with $|S| \leqslant \alpha n$.*

**Proof.** We know from Corollary 4.8 that w.h.p. at time $m$ every vertex $v$ has $d_c^*(v) \geqslant 2$. Let $S_2 = S \cap LARGE$, $S_1 = S \setminus S_2$. Then

$$\begin{aligned} |N_c^*(S)| &= |N_c^*(S_1)| + |N_c^*(S_2)| - |N_c^*(S_1) \cap S_2| - |N_c^*(S_2) \cap S_1| - |N_c^*(S_1) \cap N_c^*(S_2)| \\ &\geqslant |N_c^*(S_1)| + |N_c^*(S_2)| - |S_2| - |N_c^*(S_2) \cap S_1| - |N_c^*(S_1) \cap N_c^*(S_2)|. \end{aligned}$$

Clearly, $|S_2| \leqslant |S| \leqslant \alpha n$, and so Lemma 4.9 gives $|N_c^*(S_2)| \geqslant 19|S_2|$, w.h.p. Also, recall from Lemma 3.1 that w.h.p. there are no small structures in $G_m$, and since $SMALL_c \subset SMALL$ w.h.p., this means there are no small-$c$-structures either. In particular,

- no $small_c$ vertices are adjacent and there is no path of length two between $small_c$ vertices, which implies that $|N_c^*(S_1)| \geqslant 2|S_1|$ and $|N_c^*(S_2) \cap S_1| \leqslant |S_2|$.
- in addition, there is no $C_4$ containing a $small_c$ vertex, and no path of length four between $small_c$ vertices. This means that $|N_c^*(S_1) \cap N_c^*(S_2)| \leqslant |S_2|$.

We deduce that $|N_c^*(S)| \geqslant 2|S_1| + 19|S_2| - 3|S_2| \geqslant 2|S|$. □

Recall that $\Gamma_c^*$ is the subgraph induced by edges of colour $c$ that are not in $E^+$.

**Corollary 4.13.** *With high probability $\Gamma_c^*$ is connected for every $c \in [q]$.*

**Proof.** If $[S, V \setminus S]$ is a cut in $\Gamma_c^*$ then Corollary 4.12 implies that $|S|, |V \setminus S| \geqslant \alpha n$. Let $F = S \cap Full$. Since $|V \setminus Full| \leqslant n^{1-\delta} < \alpha n/2$ we see that $|F|, |Full \setminus F| \geqslant \alpha n/2$. As in Case 2 of Lemma 4.9 we show that w.h.p. no such $F$ exists by doing the relevant computation in $H_2$ with $p_+ \sim \log n/(4qn)$:

$$\mathbb{P}(\exists F) \leqslant 2 \sum_{f=\alpha n/2}^{|Full|-\alpha n/2} \binom{|Full|}{f} (1-p_+)^{f(|Full|-f)-t_\varepsilon} \tag{4.42}$$

$$\leqslant n2^n (1-p_+)^{(\alpha n/2)(n-n^{1-\delta}-(\alpha/2)n)-o(n^2)} \tag{4.43}$$

$$\leqslant n2^n e^{-\alpha n \log n/10q} = o(1). \tag{4.44}$$

We subtract $t_\varepsilon$ from $f(|Full| - f)$ because we do not include the first $t_\varepsilon$ edges in this calculation. This is because *Full* depends on them. □

## 5. Rotations

We now use $E_c^+$ to build the Hamiltonian cycles for every colour $c$ using Pósa rotations. We let $G_c$ denote the graph induced by the edges of colour $c$. Given a path $P = (x_1, x_2, \ldots, x_k)$ and an

https://doi.org/10.1017/S0963548318000159 Published online by Cambridge University Press

edge $x_ix_k, 2 \leqslant i \leqslant k-2$, we say that the path $P' = (x_1, \dots, x_i, x_k, \dots, x_{i+1})$ is obtained from $P$ by a rotation with $x_1$ as the fixed endpoint.

For a path $P$ in $G_c$ with endpoint $a$, let $END(a)$ denote the set of all endpoints of paths obtainable from $P$ by a sequence of Pósa rotations with $a$ as the fixed endpoint. In this context, Pósa [15] shows that $|N_c(END(a))| < 2|END(a)|$. This is assuming that in the course of executing the rotations, no simple extension of our path is found. It follows from Corollary 4.12 that w.h.p. $|END(a)| \geqslant \alpha n$. For each $b \in END(a)$ there will be a path $P_b$ of the same length as $|P|$ with endpoints $a, b$. We let $END(b)$ denote the set of all endpoints of paths obtainable from $P_b$ by a sequence of Pósa rotations with $b$ as the fixed endpoint. It also follows from Corollary 4.12 that w.h.p. $|END(b)| \geqslant \alpha n$ for all $b \in END(a)$. Let $END(P) = \{a\} \cup END(a)$.

An edge $u = \{x, y\}$ of colour $c$ with $y \in END(x)$ is called a *booster*. Let $P_{x,y}$ be the path of length $|P|$ from $x$ to $y$ implied by $y \in END(x)$. Adding the edge $u$ to $P_{x,y}$ will either create a Hamilton cycle or imply the existence of a path of length $|P|+1$ in $G_c$, after using Corollary 4.13. Indeed, if the cycle $C$ created is not a Hamilton cycle, then the connectivity of $\Gamma_c^*$ implies that there is an edge $u = xy$ of colour $c$ with $x \in V(c)$ and $y \notin V(C)$. Then adding $u$ and removing an edge of $C$ incident to $x$ creates a path of length $|P|+1$.

We start with a longest path in $\Gamma_c^*$ and let $E_c^+ = \{f_1, f_2, \dots, f_\ell\}$ where w.h.p. $\ell \geqslant m_+ = n \log n/(8q)$: see Claim 4.11. A round consists of an attempt to find a longer path than the current one or to close a Hamilton path to a cycle. Suppose we start a round with a path $P$ of length $k$. We use rotations and construct many paths. If one of these paths has an endpoint with a neighbour outside the path then we add this neighbour to the current path and start a new round with a path of length $k+1$. Here we only use edges not in $E_c^+$. Failing this we compute $END(P)$ and look for a booster in $E_c^+$. In the search for boosters we start from $f_r$ assuming that we have already examined $f_1, f_2, \dots, f_{r-1}$ in previous rounds. Now $f_r$ is chosen uniformly from $(1-o(1))\binom{n}{2}$ pairs, and so the probability it is a booster is at least $\beta = (1-o(1))\alpha^2$. It is clear that at most $n$ boosters are needed to create a Hamilton cycle. Adding a booster increases the length of the current path by one, or creates a Hamilton cycle. So the probability we fail to find a Hamilton cycle of colour $c$ is at most $\mathbb{P}(\mathrm{Bin}(m_+, \beta) \leqslant n) = o(1)$. We can inflate this $o(1)$ by $\sigma$ to show that w.h.p. we find a Hamilton cycle in each colour, completing the proof of Theorem 1.1.

## 6. Concluding remarks

In this paper we studied a very natural variant of the classical problem of the appearance of $\sigma$ edge-disjoint Hamilton cycles in a random graph process. We showed that one can colour the edges of the process online so that every colour class has a Hamilton cycle exactly at the moment when the underlying graph has $\sigma$ edge-disjoint ones.

The paper [4] shows that at the hitting time $\tau_{2\sigma+1}$ there will w.h.p. be $\sigma$ edge-disjoint Hamilton cycles plus an edge-disjoint matching of size $\lfloor n/2 \rfloor$. It is straightforward to extend this result to the online situation. It should be clear that at time $\tau_{2\sigma+1}$ COL can be used to construct w.h.p. $E_c^*, E_c^+, c = 1, 2, \sigma+1$ such that $E_c^* \cup E_c^+$ induce Hamiltonian graphs for $1 \leqslant c \leqslant \sigma$ and $d_{\sigma+1}^*(v) \geqslant 1$ for $v \in [n]$. For colour $\sigma+1$, we replace the statement of Corollary 4.12 with

$$\text{w.h.p. } |N_{\sigma+1}^*(S)| \geqslant |S| \quad \text{for all } S \subset V(G) \text{ with } |S| \leqslant \alpha n. \tag{6.1}$$

https://doi.org/10.1017/S0963548318000159 Published online by Cambridge University Press

We then replace rotations with alternating paths, using $E^+_{\sigma+1}$ as boosters. The details are as described in Chapter 6 of [7]. In outline, let $G = (V,E)$ be a graph without a matching of size $\lfloor |V(G)|/2 \rfloor$. For $v \in V$ such that $v$ is isolated by some maximum matching, let

$$A(v) = \{w \in V : w \neq v \text{ and } \exists \text{ a maximum matching of } G \text{ that isolates } v \text{ and } w\}.$$

We use the following lemma.

**Lemma 6.1.** *Let $G$ be a graph without a matching of size $\lfloor |V(G)|/2 \rfloor$. Let $M$ be a maximum matching of $G$. If $v \in V$ and $A(v) \neq \emptyset$, then $|N_G(A(v))| < |A(v)|$.*

We start with a maximum matching $M$ of $\Gamma^*_{\sigma+1}$. Suppose that $v$ is not covered by $M$. Using (6.1), we see that w.h.p. $|A(v)| \geqslant \alpha n$. Further, if $u \in A(v)$ and $uv \in E^+_{\sigma+1}$, then adding this edge gives a larger matching. Also, because $u$ is isolated by a maximum matching, there is a corresponding set $A_u$ of size at least $\alpha n$ such that if $w \in A_u$ and $uw \in E^+_{\sigma+1}$ then we can find a larger matching. Therefore we have $\Omega(n^2)$ boosters and the proof is similar to that for Hamilton cycles.

There are several related problems which can likely be treated using our approach. One potential application for our technique is to show that for any fixed positive integer $k$ and any decomposition $k = k_1 + \cdots + k_s$ into the sum of $s$ positive integers, there is an online algorithm, colouring the edges of a random graph process in $s$ colours so that exactly at the hitting time $\tau_k$ the $i$th colour forms a $k_i$-connected spanning graph for $i = 1, \ldots, s$. In general, one can generate many more interesting problems by considering the online Ramsey version of other results in the theory of random graphs.

## Acknowledgement

We thank the referee(s) for their comments.

## References

[1] Ajtai, M., Komlós, J. and Szemerédi, E. (1985) First occurrence of Hamilton cycles in random graphs. In *Cycles in Graphs* (B. R. Alspach and C. D. Godsil, eds), Vol. 115 of North-Holland Mathematics Studies, North-Holland, pp. 173–178.

[2] Bohman, T., Frieze, A., Krivelevich, M., Loh, P. and Sudakov, B. (2011) Ramsey games with giants. *Random Struct. Alg.* **38** 68–99.

[3] Bollobás, B. (1984) The evolution of sparse graphs. In *Graph Theory and Combinatorics: Proceedings of the Cambridge Combinatorial Conference in Honour of Paul Erdős* (B. Bollobás, ed.), pp. 35–57.

[4] Bollobás, B. and Frieze, A. M. (1985) On matchings and Hamiltonian cycles in random graphs. *Ann. Discrete Math.* **28** 23–46.

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

[6] Erdős, P. and Rényi, A. (1961) On the strength of connectedness of a random graph. *Acta. Math. Acad. Sci. Hungar.* **8** 261–267.

[7] Frieze, A. M. and Karoński, M. (2016) *Introduction to Random Graphs*, Cambridge University Press.

[8] Hoeffding, W. (1963) Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.* **58** 13–30.

https://doi.org/10.1017/S0963548318000159 Published online by Cambridge University Press

[9] Knox, F., Kühn, D. and Osthus, D. (2015) Edge-disjoint Hamilton cycles in random graphs. *Random Struct. Alg.* **46** 397–445.
[10] Komlós, J. and Szemerédi, E. (1983) Limit distribution for the existence of Hamilton cycles in random graphs. *Discrete Math.* **43** 55–63.
[11] Korshunov, A. (1976) Solution of a problem of Erdős and Rényi on Hamilton cycles in non-oriented graphs. *Soviet Math. Dokl.* **17** 760–764.
[12] Krivelevich, M., Lubetzky, E. and Sudakov, B. (2010) Hamiltonicity thresholds in Achlioptas processes. *Random Struct. Alg.* **37** 1–24.
[13] Krivelevich, M. and Samotij, W. (2012) Optimal packings of Hamilton cycles in sparse random graphs. *SIAM J. Discrete Math.* **26** 964–982.
[14] Lee, C., Sudakov, B. and Vilenchik, D. (2012) Getting a directed Hamilton cycle two times faster. *Combin. Probab. Comput.* **21** 773–801.
[15] Pósa, L. (1976) Hamiltonian circuits in random graphs. *Discrete Math.* **14** 359–364.

https://doi.org/10.1017/S0963548318000159 Published online by Cambridge University Press