

## The early evolution of the $H$ -free process

Tom Bohman · Peter Keevash

Received: 9 June 2009 / Accepted: 22 March 2010 / Published online: 4 May 2010  
© Springer-Verlag 2010

**Abstract** The  $H$ -free process, for some fixed graph  $H$ , is the random graph process defined by starting with an empty graph on  $n$  vertices and then adding edges one at a time, chosen uniformly at random subject to the constraint that no  $H$  subgraph is formed. Let  $G$  be the random maximal  $H$ -free graph obtained at the end of the process. When  $H$  is strictly 2-balanced, we show that for some  $c > 0$ , with high probability as  $n \rightarrow \infty$ , the minimum degree in  $G$  is at least  $cn^{1-(v_H-2)/(e_H-1)}(\log n)^{1/(e_H-1)}$ . This gives new lower bounds for the Turán numbers of certain bipartite graphs, such as the complete bipartite graphs  $K_{r,r}$  with  $r \geq 5$ . When  $H$  is a complete graph  $K_s$  with  $s \geq 5$  we show that for some  $C > 0$ , with high probability the independence number of  $G$  is at most  $Cn^{2/(s+1)}(\log n)^{1-1/(e_H-1)}$ . This gives new lower bounds for Ramsey numbers  $R(s, t)$  for fixed  $s \geq 5$  and  $t$  large. We also obtain new bounds for the independence number of  $G$  for other graphs  $H$ , including the case when  $H$  is a cycle. Our proofs use the differential equations method for random graph processes to analyse the evolution of the process, and give further information about the structure of the graphs obtained, including asymptotic formulae for a broad class of subgraph extension variables.

---

T. Bohman was supported in part by NSF grant DMS-0701183.

P. Keevash was supported in part by NSF grant DMS-0555755.

---

T. Bohman (✉)

Department of Mathematical Sciences, Carnegie Mellon University, Pittsburgh, USA  
e-mail: [tbohman@math.cmu.edu](mailto:tbohman@math.cmu.edu)

P. Keevash

School of Mathematical Sciences, Queen Mary, University of London, Mile End Road,  
London E1 4NS, UK  
e-mail: [p.keevash@qmul.ac.uk](mailto:p.keevash@qmul.ac.uk)

## 1 Introduction

Random graph processes provide a natural context for modeling a complex network that evolves over time. While there has been considerable recent interest in using such processes to model networks that arise in applications (see [11] and the references therein), random graphs have long been an important component in the construction of sophisticated combinatorial objects (see [2]). In the classical Erdős-Rényi random graph model  $G(n, p)$  each pair of vertices appears as an edge with probability  $p = p(n)$  and these choices are mutually independent. The closely related random graph  $G(n, i)$  is chosen uniformly at random from the collection of all graphs with  $n$  vertices and  $i$  edges. These models are well understood, but distributions on graphs given by random processes in which there is significant dependence among the choices made in different rounds are typically much more difficult to analyse. For many such processes even the most basic quantities, such as the number of edges in the final graph, are not known (see [21], for example).

In this paper we analyse a significant portion of the initial evolution of the  $H$ -free process, for some fixed graph  $H$ , defined by starting with an empty graph on  $n$  vertices and then adding edges one at a time, chosen uniformly at random subject to the constraint that no  $H$  subgraph is formed. More formally, we begin with the graph on  $n$  vertices with no edges, which we denote  $G(0)$ . Now suppose  $i > 0$  and we have some graph  $G(i - 1)$ . We say that a pair  $uv$  of vertices is *open* in  $G(i - 1)$  if  $uv$  is not an edge of  $G(i - 1)$  and  $G(i - 1) \cup \{uv\}$  does not contain  $H$  as a subgraph. We choose  $uv$  uniformly at random among all open pairs in  $G(i - 1)$  and then  $G(i)$  is obtained from  $G(i - 1)$  by adding the edge  $e_i = uv$ . The process terminates when there are no open pairs, with some graph  $G$  on  $n$  vertices that is a maximal  $H$ -free graph. Beside being of interest in its own right, our analysis of this process produces new results in Ramsey theory and the theory of Turán problems.

Erdős, Suen and Winkler [17] suggested this process as a means to generate an interesting probability distribution on the collection of maximal  $H$ -free graphs, or more generally maximal graphs with any fixed graph property.<sup>1</sup> They obtained results on the triangle-free process and the bipartite process, using a differential equations method that had been previously applied by Ruciński and Wormald [28] to analyse the ‘maximum degree  $d$ ’ process. Another motivation for their work was that their analysis of the triangle-free process led to the best lower bound on the Ramsey number  $R(3, t)$  known at that time.

Ramsey theory encompasses a variety of results expressing the informal principle that all large systems have some structure. It is a source of many

---

<sup>1</sup>Bollobás (personal communication) informs us that such processes were considered earlier, if not in print.

challenging unsolved combinatorial problems and has applications throughout mathematics. We refer the reader to [22] for an introduction to the subject. The Ramsey number  $R(s, t)$  is the least number  $n$  such that any graph on  $n$  vertices contains a complete graph with  $s$  vertices or an independent set with  $t$  vertices. In general, very little is known about these numbers, even approximately. The upper bound  $R(3, t) = O(t^2/\log t)$  was obtained by Ajtai et al. [1], but for many years the best known lower bound, due to Erdős [12], was  $\Omega(t^2/\log^2 t)$ . Spencer [33] conjectured that the triangle-free process is likely to produce a graph that establishes a good lower bound on  $R(3, t)$  for  $t$  large; the idea being that the triangle-free process admits enough random edges to bring the independence number close to the smallest possible for a triangle-free graph. Finally, Kim [23] determined the order of magnitude, showing that  $R(3, t) = \Theta(t^2/\log t)$ . His proof made use of a semi-random construction that is motivated (even guided) by the triangle-free process, but the question remained open as to whether the triangle-free process itself gives such a good construction. This was answered by Bohman [7], who showed that with high probability, the graph produced by the triangle-free process has independence number bounded above by  $O(n^{1/2} \log^{1/2} n)$  and minimum degree bounded below by  $\Omega(n^{1/2} \log^{1/2} n)$ . He went on to analyse the  $K_4$ -free process, improving the best known lower bound on  $R(4, t)$  to  $R(4, t) > \Omega(t^{5/2}/\log^2 t)$ .

The general  $H$ -free process was independently studied by Osthus and Taraz [26] and by Bollobás and Riordan [8]. Say that a graph  $H$  is *strictly 2-balanced* if the number of vertices  $v_H$  and edges  $e_H$  in  $H$  are both at least 3 and

$$\frac{e_H - 1}{v_H - 2} > \frac{e_K - 1}{v_K - 2}$$

for all proper subgraphs  $K$  of  $H$  with  $v_K \geq 3$ . Osthus and Taraz showed that if  $H$  is strictly 2-balanced then for some  $c, C > 0$  with high probability, for the  $H$ -free process  $G$  has average degree at least  $cn^{1-(v_H-2)/(e_H-1)}$  and maximum degree at most  $Cn^{1-(v_H-2)/(e_H-1)}(\log n)^{1/(\Delta(H)-1)}$ . (In fact they proved the average degree bound under a similar but weaker condition on  $H$ .) Wolfowitz [36] showed that if  $H$  is strictly 2-balanced and regular then the expected number of edges in  $G$  is at least  $cn^{2-(v_H-2)/(e_H-1)}(\log \log n)^{1/(e_H-1)}$ . An immediate consequence is an improved lower bound for Turán numbers, which leads us to another motivation for studying the  $H$ -free process.

The Turán number  $\text{ex}(n, H)$  is the maximum possible number of edges in a graph on  $n$  vertices that does not contain an  $H$  subgraph. More generally, the theory of Turán problems concerns the study of combinatorial structures that have maximum size subject to not containing some fixed structure. We refer the reader to [18] for a survey of this subject. Turán [35] determined the value of  $\text{ex}(n, H)$  when  $H = K_r$  is complete: the unique largest graph

on  $n$  vertices with no  $K_r$  subgraph is complete  $(r - 1)$ -partite with part sizes as equal as possible. For general  $H$ , the Erdős-Stone-Simonovits theorem [14, 16] gives the estimate  $\text{ex}(n, H) = \text{ex}(n, K_r) + o(n^2)$ , where  $r = \chi(H)$  is the chromatic number of  $H$ . This gives an asymptotic formula for the Turán number when  $H$  is not bipartite. However, when  $H$  is bipartite it is an open problem in general to determine even the order of magnitude of  $\text{ex}(n, H)$ . For example, when  $H = K_{r,r}$  is complete bipartite with  $r \geq 5$ , for many years the best known lower bound was  $\text{ex}(n, K_{r,r}) = \Omega(n^{2-2/(r+1)})$ , a result of Erdős and Spencer [15] proved via a simple application of the probabilistic method. Wolfowitz’s analysis of the  $H$ -free process improved this to  $\text{ex}(n, K_{r,r}) = \Omega(n^{2-2/(r+1)}(\log \log n)^{1/(r^2-1)})$ .

### 1.1 Results I: Ramsey and Turán bounds

In this paper we extend the methods from [7] to an analysis of the  $H$ -free process when  $H$  is strictly 2-balanced, leading to new lower bounds for Ramsey and Turán numbers. We also investigate other properties of the process, viewing it as a model of interest in its own right, and give certain extension counting formulae that address a question of Spencer. In particular, we show that the graph produced by the  $H$ -free process is very similar to the corresponding random graph  $G(n, i)$  with respect to small subgraph counts, with the exception that the  $H$ -free process produces no copies of graphs containing  $H$ . We begin with the Turán and Ramsey results.

Our first theorem gives a new lower bound for the number of edges in  $G$ . In fact we have a new lower bound for the minimum degree, and it holds with high probability, not just in expectation. An immediate consequence is a lower bound for the Turán number  $\text{ex}(n, H)$ .

**Theorem 1.1** *Suppose that  $H$  is a strictly 2-balanced graph with  $v_H$  vertices and  $e_H$  edges. Then for some  $c > 0$  with high probability the minimum degree in the final graph of the  $H$ -free process is at least  $cn^{1-(v_H-2)/(e_H-1)}(\log n)^{1/(e_H-1)}$ . In particular, the Turán number satisfies*

$$\text{ex}(n, H) = \Omega(n^{2-(v_H-2)/(e_H-1)}(\log n)^{1/(e_H-1)}).$$

Note that it follows immediately from Theorem 1.1 that we have

$$\text{ex}(n, K_{r,r}) = \Omega(n^{2-2/(r+1)}(\log n)^{1/(r^2-1)}).$$

For general complete bipartite graphs  $K_{r,s}$  with  $r \leq s$ , the ‘Zarankiewicz problem’ of estimating  $\text{ex}(n, K_{r,s})$  is a subject of special interest in extremal graph theory. A general upper bound of order  $n^{2-1/r}$  was given by Kővári, Sós and Turán [24]. The only known asymptotic results are  $\text{ex}(n, K_{2,r}) \sim$

$\frac{1}{2}(r - 1)^{1/2}n^{3/2}$  (see [19]) and  $\text{ex}(n, K_{3,3}) \sim \frac{1}{2}n^{5/3}$  (see [9] and [20]). Note that the lower bound construction for  $K_{3,3}$  also gives the best known lower bound for  $K_{4,4}$ . The only other case when the upper bound is known to be of the correct order of magnitude is when  $s > (r - 1)!$  (see [3]). The known constructions are based on algebraic and geometric structures that may not exist for other values of the parameters  $r$  and  $s$ . However, it is widely believed that  $\text{ex}(n, K_{r,s})$  for general  $r \leq s$  is on the order of  $n^{2-1/r}$ .

For Ramsey numbers, we obtain the following new lower bounds.

**Theorem 1.2** *For fixed  $s \geq 5$  and  $t \rightarrow \infty$ , the Ramsey number satisfies*

$$R(s, t) = \Omega\left(t^{\frac{s+1}{2}} (\log t)^{\frac{1}{s-2} - \frac{s+1}{2}}\right).$$

The previously best known lower bound on  $R(s, t)$  when  $s$  is fixed and  $t$  is large was  $R(s, t) = \Omega((t/\log t)^{\frac{s+1}{2}})$ , established by Spencer [32] using the Lovász Local lemma. Theorem 1.2 improves this by a multiplicative factor of  $(\log t)^{1/(s-2)}$ . There is no particular reason to believe that our lower bound is anywhere near optimal, since the best known general upper bound is essentially  $t^{s-1}$  (up to a polylogarithmic factor in  $t$ ). On the other hand, as Theorem 1.2 can be viewed as the natural generalisation of the construction that gives the correct order of magnitude for  $R(3, t)$ , it would be interesting to see a significant improvement on the bound in Theorem 1.2 for  $s \geq 4$ .

We also obtain new lower bounds for cycle-complete Ramsey numbers. Given graphs  $H_1, H_2$ , the graph Ramsey number  $R(H_1, H_2)$  is the least number  $n$  such that for any 2-colouring of the edges of  $K_n$  there is a monochromatic copy of  $H_1$  or  $H_2$ . Note that  $R(C_\ell, K_t) \geq n$  if and only if there is a  $C_\ell$ -free graph on  $n$  vertices with no independent set of size  $t$ . We prove the following bound.

**Theorem 1.3** *For fixed  $\ell \geq 4$  and  $t \rightarrow \infty$  the cycle-complete Ramsey number satisfies*

$$R(C_\ell, K_t) = \Omega\left(t^{\frac{\ell-1}{\ell-2}} / \log t\right).$$

Again this is quite far from the best known upper bound (see [10, 25, 34]). For example, Erdős [13] conjectured that  $R(C_4, K_t) = O(t^{2-\epsilon})$  for some absolute constant  $\epsilon > 0$ , but this is still open.

In fact, we establish more general properties of the  $H$ -free process from which these theorems follow. In order to show that the process continues to run for a certain number of steps, we will establish asymptotic formulae for various graph parameters at any given time in the process, including the degree of any vertex, but also more general extension parameters. To state these formulae we need some terminology and notation.

### 1.2 Terminology and notation I

We write  $[n] = \{1, \dots, n\}$  for the vertex set of the process. At step  $i$  of the process let  $E(i)$  be the edges of the graph  $G(i)$ , let  $O(i)$  be the pairs of vertices that are open (as defined above), and let  $C(i)$  be the pairs of vertices that are neither edges nor open, which we refer to as *closed*.

We fix some strictly 2-balanced graph  $H$  throughout the paper and write

$$p = n^{-\frac{v_H-2}{e_H-1}}.$$

For any graph  $\Gamma$  we write  $V_\Gamma$  for the vertex set of  $\Gamma$ ,  $E_\Gamma$  for the edge set of  $\Gamma$ ,  $v_\Gamma = |V_\Gamma|$  and  $e_\Gamma = |E_\Gamma|$ . For  $A \subseteq V_\Gamma$  we write

$$S_\Gamma = p^{e_\Gamma} n^{v_\Gamma} \quad \text{and} \quad S_{A,\Gamma} = p^{e_\Gamma - e_{\Gamma[A]}} n^{v_\Gamma - |A|}.$$

We say that such a pair  $(A, \Gamma)$  is *strictly balanced* if  $S_{A,\Gamma[B]} > S_{A,\Gamma}$  for every  $A \subsetneq B \subsetneq V_\Gamma$  and *strictly dense* if  $S_{A,\Gamma[B]} > 1$  for every  $A \subsetneq B \subseteq V_\Gamma$ .

A key element of our analysis of the  $H$ -free process is closely tracking the number of extensions from fixed sets of vertices to fixed subgraphs of  $G(i)$ . Intuitively, the graph  $G(i)$  produced by the  $H$ -free process should be roughly equal to the random graph  $G(n, i)$ , the graph chosen uniformly at random from the collection of graphs with  $n$  vertices and  $i$  edges, up until the number of copies of  $H$  in  $G(n, i)$  is roughly equal to the number of edges. This occurs when  $i$  is roughly  $pn^2$ , with  $p$  as defined above. We expect the more interesting part of the evolution of the  $H$ -free process to be at and beyond this range of  $i$ . Considering  $G(n, p)$ , which is very similar to  $G(n, i)$  here, we note that  $S_\Gamma$  is roughly the expected number of labeled copies of  $\Gamma$ , and  $S_{A,\Gamma}$  is roughly the expected number of labeled extensions to  $\Gamma$  from a fixed set of vertices playing the role of  $A$ . Thus we can think of these quantities as anticipated *scalings* by which we should measure the same parameters in the  $H$ -free process.

In order to track extensions, we track all ‘open routes’ to such extensions. Suppose  $\Gamma$  is a graph and  $J$  is a spanning subgraph of  $\Gamma$ . Suppose also that  $A \subseteq V_\Gamma$  is an independent set in  $\Gamma$  and  $\phi : A \rightarrow [n]$  is an injective mapping. We define the *extension variables*  $X_{\phi,J,\Gamma}(i)$  to be the number of injective maps  $f : V_\Gamma \rightarrow [n]$  such that

- (i)  $f(e) \in O(i)$  for every  $e \in E_\Gamma \setminus E_J$ ,
- (ii)  $f(e) \in E(i)$  for every  $e \in E_J$ , and
- (iii)  $f$  restricts to  $\phi$  on  $A$ .

We say that the random variable  $X_{\phi,J,\Gamma}(i)$  is *trackable* if one of the following two conditions holds:

- (a)  $(A, \Gamma)$  is strictly dense and  $\Gamma$  does not contain  $H$  as a subgraph, or
- (b)  $S_{A,\Gamma} = 1$ ,  $(A, \Gamma)$  is strictly balanced,  $E_J \subsetneq E_\Gamma$ , and  $H$  is not a subgraph of the graph  $\Gamma'$  obtained from  $\Gamma$  by adding the edges  $ab$  for all  $a, b \in A$  with  $\phi(a)\phi(b) \in E(i)$ .

It follows easily from the definitions that for any trackable extension variable  $X_{\phi,J,\Gamma}(i)$  the pair  $(A, J)$  is strictly dense. Note further that condition (b) includes the case where  $\Gamma = H \setminus ab$  for some  $ab \in E_H$ ,  $e_J \leq e_H - 2$ ,  $A = \{a, b\}$  and  $\phi(ab) \notin E(i)$ . These extensions comprise the set of *open routes* to a copy of  $H$  less an edge, where  $\phi(ab)$  plays the role of the missing edge. As the appearance of such an extension is the mechanism whereby the pair  $\phi(ab)$  becomes closed, these particular extension variables play a central role in our analysis of the  $H$ -free process.

We fix constants  $V, W, \epsilon, \mu$  throughout the paper which satisfy  $0 < \mu \ll \epsilon \ll 1/W \ll 1/V \ll 1/e_H$ . (The notation  $0 < \alpha \ll \beta$  means that there is an increasing function  $f(x)$  so that the following argument is valid for  $0 < \alpha < f(\beta)$ .) We introduce a continuous time variable  $t$ , using the scaling  $t = t(i) = i/s$  with  $s = pn^2$ , and analyse the process up to time  $t_{\max} = \mu(\log n)^{1/(e_H-1)}$ , which corresponds to

$$m = \mu(\log n)^{1/(e_H-1)} pn^2$$

edges. Let  $\mathcal{T}$  be the set of all triples  $(A, J, \Gamma)$  where  $J$  is a spanning subgraph of a graph  $\Gamma$  with  $v_\Gamma, e_\Gamma < V$ ,  $A$  is an independent set in  $\Gamma$ , and the variables  $X_{\phi,J,\Gamma}(0)$  are trackable. Write  $aut(H)$  for the number of automorphisms of  $H$  and define

$$\begin{aligned} q(t) &= e^{-2e_H aut(H)^{-1}(2t)^{e_H-1}}, & a_H &= 4e_H(e_H - 1)/aut(H), \\ c(t) &= a_H(2t)^{e_H-2}q(t), & P(t) &= W(t^{e_H-1} + t), \\ e(t) &= e^{P(t)} - 1 & \text{and} & \quad s_e = n^{1/2e_H-\epsilon}. \end{aligned}$$

We also define  $\gamma(t)$  to be any smooth increasing function such that  $\gamma(t) = 40Ve^{40V}t$  for  $0 \leq t \leq 40V/W$ ,  $\gamma'(t) > 20V$  for  $40V/W < t \leq 1/(50V)$ , and  $\gamma(t) < 1/2$ ,  $\gamma'(t) < W$  for all  $t \geq 0$ . Then we set  $\theta(t) = 1/2 + \gamma(t)$ , so that  $1/2 \leq \theta(t) < 1$  for all  $t \geq 0$ .

### 1.3 Results II: the $H$ -free process

Our first main theorem gives asymptotic formulae for trackable extension variables throughout the process.

**Theorem 1.4** *With high probability, for every  $i \leq m$  and trackable extension variable  $X_{\phi,J,\Gamma}(i)$  corresponding to a triple in  $\mathcal{T}$ , we have*

$$X_{\phi,J,\Gamma}(i) = (1 \pm e(t)/s_e)(x_{A,J,\Gamma}(t) \pm 1/s_e)S_{A,J},$$

where

$$x_{A,J,\Gamma}(t) = (2t)^{e_J} q(t)^{e_{\Gamma} - e_J}.$$

(For this theorem to be useful we choose  $\epsilon < \epsilon(V)$  sufficiently small and then  $\mu < \mu(\epsilon)$  sufficiently small so that  $e(t)$  and  $q(t)^{-V}$  are both at most  $n^\epsilon$  for  $t \leq t_{\max}$ .) Note, for example, that there is a trackable extension variable describing the number of common neighbours of a set of size  $d$  whenever  $p^d n > 1$ , so we have the following corollary.

**Corollary 1.5** *With high probability, for every  $d$  with  $p^d n > 1$ , set  $A$  of  $d$  vertices and  $n^2 p \leq i \leq m$ , the number of common neighbours of  $A$  in  $G(i)$  is  $(1 + o(1))(2i/n^2)^d n$ .*

A remarkable consequence of Theorem 1.4 is that the graph  $G(i)$  for  $i \leq m$  is similar to the uniform random graph  $G(n, i)$  with respect to small subgraph counts, with the notable exception that there are no copies of graphs containing  $H$  in  $G(i)$ . The possibility of this intriguing behavior was first suggested by Joel Spencer. The following theorem gives the correct asymptotic counts for labeled copies of a graph  $\Gamma$  in the ‘subcritical’ case (i) and the ‘supercritical’ case (ii). For the sake of brevity we just establish existence of at least one copy in the ‘critical’ case (iii), although our discussion in Sect. 10 points the way towards better results in this case.

**Theorem 1.6** *Suppose  $\Gamma$  is an  $H$ -free graph and write  $X_\Gamma(i)$  for the number of labelled copies of  $\Gamma$  in  $G(i)$ . Then with high probability*

- (i) *If there exists  $B \subseteq V_\Gamma$  with  $S_{\Gamma[B]} < 1$  then  $X_\Gamma(m) = 0$ .*
- (ii) *If  $S_{\Gamma[B]} > 1$  for all non-empty  $B \subseteq V_\Gamma$  then there exists a constant  $\xi > 0$  such that  $X_\Gamma(i) \sim (2i/n^2)^{e_\Gamma} n^{v_\Gamma}$  for  $n^2 p \leq i \leq \xi m$ .*
- (iii) *If  $S_{\Gamma[B]} \geq 1$  for all  $B \subseteq V_\Gamma$  then  $X_\Gamma(m) > 0$ .*

While Theorem 1.4 alone is enough to establish the Turán bounds stated above, our results on the Ramsey numbers require an upper bound on the independence number of  $G(m)$ . Theorem 1.2 follows easily from Theorem 1.8 below. This in turn follows from the following more general result for  $s \geq 6$ . (Then we will need to modify the proof slightly to deal with the case  $s = 5$ .)

**Theorem 1.7** *Suppose that  $H$  is strictly 2-balanced and that for any two edges  $uv, xy$  of  $H$  and  $\{x, y\} \subsetneq B \subsetneq V_H$  we have  $S_{B, H \setminus uv} < 1$ . Then there is  $C > 0$  such that with high probability the final graph of the  $H$ -free process has independence number at most  $Cn^{(v_H - 2)/(e_H - 1)} (\log n)^{1 - 1/(e_H - 1)}$ .*



**Theorem 1.8** *For any  $s \geq 5$  there is  $C > 0$  such that with high probability the final graph of the  $K_s$ -free process has independence number at most  $Cn^{\frac{2}{s+1}}(\log n)^{1-\binom{s}{2}^{-1}}$ .*

Alon, Ben-Shimon and Krivelevich [4] recently proposed a construction that takes a nearly regular  $K_s$ -free graph  $G$  and produces a regular  $K_s$ -free graph with roughly the same independence number as the original graph. It follows from Corollary 1.5 that the graph produced after  $m$  steps of the  $K_s$ -free process is a suitable input for this construction. This suggests that the bound on  $R(s, t)$  given in Theorem 1.2 can be achieved by a regular graph. (A formal proof would need to provide some details missing from the sketch given in [4].)

We also obtain the following bound when  $H$  is a cycle, which implies Theorem 1.3.

**Theorem 1.9** *For any  $\ell \geq 3$  there is  $C > 0$  such that with high probability the final graph of the  $C_\ell$ -free process has independence number at most  $C(n \log n)^{(\ell-2)/(\ell-1)}$ .*

#### 1.4 Organisation of the paper

In the next section we give a heuristic explanation for the differential equations leading to the formulae in Theorem 1.4. In Sect. 3 we develop some theory of strictly 2-balanced graphs and balanced extensions. Over the following three sections we collect various properties that hold with high probability on the ‘good’ event at a given time that the process has followed the trajectory of the differential equations so far. Section 4 contains various union bound arguments, Sect. 5 gives upper bounds on the extension variables and Sect. 6 provides a means to approximate the number of pairs that become closed when some particular pair is added as an edge. In Sect. 7 we formulate our framework for showing that the process follows the differential equations, which is based to some extent on that given by Wormald [37], but also incorporates martingale estimates from [7]. Section 8 concerns trackable random variables: we obtain bounds on the one-step changes of trackable random variables sufficient to apply the differential equations method. Then we apply the differential equation method in Sect. 9 to prove Theorem 1.4, from which Theorem 1.1 immediately follows. We also apply Theorem 1.4 to prove Theorem 1.6 in Sect. 10. Next we turn our attention to the independence number. In Sect. 11 we formulate a general property, which we call ‘smooth independence’, and bound the independence number under the assumption that  $H$  has this property. Then in Sect. 12 we show that cycles and complete graphs  $K_s$ ,  $s \geq 5$  have smooth independence, from which Theorems 1.9 and 1.2 follow. We also prove Theorem 1.7 in this section. The final section contains some concluding remarks.

### 1.5 Terminology and notation II

We write  $\mathcal{G}_j$  for the *good* event that for every  $0 \leq i \leq j$  and trackable extension variable  $X_{\phi, J, \Gamma}(i)$  corresponding to a triple in  $\mathcal{T}$ , we have

$$X_{\phi, J, \Gamma}(i) = (1 \pm e(t)/s_e)(x_{A, J, \Gamma}(t) \pm \theta(t)/s_e)S_{A, J}.$$

Note that this implies the formulae in the statement of Theorem 1.4, since  $\theta(t) < 1$  for all  $t \geq 0$ .

When we count extensions it is convenient to work with labeled graphs, and we will often write  $uv$  for the ordered pair  $(u, v)$  as well as the edge  $\{u, v\}$ . The prime symbol  $'$  is occasionally used to denote differentiation with respect to the time variable  $t$ : this will be clear from the context.

Statements containing the symbols  $\pm$  and/or  $\mp$  are shorthand for two separate statements: one with every  $\pm$  replaced by  $+$  and every  $\mp$  by  $-$ , the other with  $\pm$  replaced by  $-$  and  $\mp$  by  $+$ . We also use the notation  $a = b \pm c$  to mean  $b - c < a < b + c$ . Where there is possibility for confusion we label the symbols as  $\pm_1$  and  $\pm_2$ , e.g.  $a^{\pm_1 \pm_2} = b^{\pm_1} \pm c^{\mp_2}$  is shorthand for 4 separate statements, one of which is  $a^{++} = b^+ \pm c^-$ .

The parameter  $n$  will always be sufficiently large compared to all other parameters, and we use the phrase ‘with high probability’ to refer to an event that has probability  $1 - o_n(1)$ , i.e. the probability tends to 1 as  $n$  tends to infinity. In fact we can arrange that our high probability events fail with probability at most  $\exp(-n^\epsilon)$ .

We say that a graph  $W$  is a *join* of two graphs  $W_1$  and  $W_2$  if it has subgraphs  $J_1$  isomorphic to  $W_1$  and  $J_2$  isomorphic to  $W_2$  such that  $V_W = V_{J_1} \cup V_{J_2}$  and  $E_W = E_{J_1} \cup E_{J_2}$ . For convenient notation we use names for vertices in  $J_1$  interchangeably with their corresponding vertices in  $W_1$ , and similarly for  $J_2$  and  $W_2$ .

If  $X$  is a set and  $k$  is a non-negative integer then we write  $\binom{X}{k}$  for the set of subsets of  $X$  of size  $k$ .

We will not often refer explicitly to the underlying probability space for the  $H$ -free process, but we note here the following natural construction. Let  $\Omega = \Omega_n$  be the set of all maximal sequences in  $\binom{[n]}{2}$  with distinct entries and the property that each initial sequence gives an  $H$ -free graph on vertex set  $[n]$ . We stress that our measure is not uniform: it is the measure given by the uniform random choice at each step. We always work with the natural filtration  $\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \dots$  given by the process. Two elements  $x, y$  of  $\Omega$  are in the same atom (i.e. part of the generating partition) of  $\mathcal{F}_j$  exactly when the first  $j$  entries of  $x$  and  $y$  agree.

## 2 Trajectory equations

We start by giving a heuristic explanation of the equations describing the evolution of the  $H$ -free process. We will then prove the validity of these equations in subsequent sections. Recall that  $G(i)$  denotes the graph on  $[n]$  obtained after  $i$  steps of the  $H$ -free process: its edge set  $E(i)$  contains  $i$  edges. We partition the non-edges  $\binom{[n]}{2} \setminus E(i)$  into two sets  $O(i)$  and  $C(i)$ , which we call *open* pairs and *closed* pairs, respectively. We say that a pair  $uv$  is open if  $G(i) \cup uv$  does not contain a copy of  $H$ , i.e.  $uv$  is a possible choice for the next edge in the process.

*Notation* We consider the following random variables. Suppose  $\Gamma$  is a graph and  $J$  is a spanning subgraph of  $\Gamma$  (i.e.  $V_J = V_\Gamma$ ). Suppose also that  $A \subseteq V_J$  is an independent set (i.e. does not span any edges) in  $\Gamma$  and  $\phi : A \rightarrow [n]$  is an injective mapping. *Throughout this paper we assume that  $\Gamma, J, A, \phi$  satisfy these conditions, even if this is not explicitly stated.* We define the *extension set*  $\Xi_{\phi, J, \Gamma}(i)$  to be the set of injective maps  $f : V_\Gamma \rightarrow [n]$  such that (i)  $f(e) \in O(i)$  for every  $e \in E_\Gamma \setminus E_J$ , (ii)  $f(e) \in E(i)$  for every  $e \in E_J$ , and (iii)  $f$  restricts to  $\phi$  on  $A$ . Then we define the *extension variables* by  $X_{\phi, J, \Gamma}(i) = |\Xi_{\phi, J, \Gamma}(i)|$ . In words, we are counting labeled copies (not necessarily induced) of a graph  $J$  in  $G(i)$  that extend a particular embedding  $\phi : A \rightarrow [n]$ , with the extra condition that some extra pairs (i.e. the edges of  $\Gamma \setminus J$ ) are open. Actually we will be interested in the number of copies up to isomorphism, but the equations for labeled copies are easier to work with.

*Examples* One special case of this definition is the number of labeled copies of a graph  $\Gamma$  in  $G(i)$ : this can be written as  $X_{\phi_0, \Gamma, \Gamma}(i)$ , where we write  $\phi_0$  for the unique function  $\phi_0 : \emptyset \rightarrow [n]$ . To count edges and open pairs with this notation we write  $e$  and  $\bar{e}$  for the two graphs on two vertices, say  $\{a, b\}$ , with one edge and no edges respectively. Then  $X_{\phi_0, \bar{e}, e}(i) = 2|O(i)|$  and  $X_{\phi_0, e, e}(i) = 2|E(i)|$ . We can also express the degree  $d_{G(i)}(v)$  of a vertex  $v$  in  $G(i)$  as  $X_{\phi_v, e, e}(i)$ , where again  $e$  is the edge  $ab$  and we write  $\phi_v$  for the function  $\phi : \{a\} \rightarrow [n]$  defined by  $\phi(a) = v$ .

We write  $Q(i) = 2|O(i)|$  for the number of ordered pairs that are open. For an ordered pair  $uv \in O(i)$ , write  $C_{uv}(i)$  for the set of ordered pairs  $xy \in O(i)$  that would become closed, i.e. belong to  $C(i + 1)$ , if at time  $i + 1$  the process chooses  $uv$  as the edge  $e_{i+1}$ . By the definition of  $C(i + 1)$  this means that adding  $uv$  and  $xy$  to  $G(i)$  would create a copy of  $H$ . Another way to say this is that there is a subgraph  $J$  obtained by deleting two edges  $ab$  and  $cd$  from  $H$  and an injective map  $f : V_H \rightarrow [n]$  such that  $f(a) = u, f(b) = v, f(c) = x, f(d) = y$  and  $f(e) \in E(i)$  for every edge of  $J$ . We have  $f \in \Xi_{\phi_T, J_T, \Gamma_T}(i)$ , where given such a quadruple  $T = (a, b, c, d)$ , we write  $\Gamma_T = H \setminus ab, J_T =$

$H \setminus \{ab, cd\}$  and define  $\phi_T$  by  $\phi_T(a) = u$  and  $\phi_T(b) = v$ . In principle there could be many embeddings  $f$  giving the same pair  $xy$ , but we will show in Lemma 6.1 that this is very unlikely: for most  $xy \in C_{uv}(i)$  there will be exactly one such embedding  $f$ , up to an automorphism of  $H$ . We will see that  $C_{uv}(i) \sim \text{aut}(H)^{-1} \sum_T X_{\phi_T, J_T, \Gamma_T}(i)$ , where the sum is over quadruples  $T = (a, b, c, d)$  such that  $ab$  and  $cd$  are distinct (but not necessarily disjoint) edges of  $H$ .

To approximate the extension variables we introduce a continuous time variable  $t$ , using the scaling  $t = t(i) = i/s$  with  $s = pn^2$ , where we recall that  $p = n^{-(v_H-2)/(e_H-1)}$ . We noted above that this is the point at which the number of copies of  $H$  in the random graph  $G(n, s)$  is comparable to the number of edges  $s$ , so it is natural to anticipate the interesting behaviour to occur at this scale. We analyse the process up to time  $t_{\max} = \mu(\log n)^{1/(e_H-1)}$ , for some small constant  $\mu > 0$ , which corresponds to  $m = \mu(\log n)^{1/(e_H-1)}pn^2$  edges. For the variable  $X_{\phi, J, \Gamma}(i)$  with  $\phi : A \rightarrow [n]$  we use the scaling  $S_{A, J} = p^{e_J}n^{v_J-|A|}$ . Again, we noted above that the count of these extensions in  $G(n, s)$  suggests the use of this scaling. Our eventual aim is to prove that with high probability, for every  $i \leq m$  and for every trackable extension variable  $X_{\phi, J, \Gamma}(i)$  corresponding to a triple in  $\mathcal{T}$ , we have the asymptotic formula

$$X_{\phi, J, \Gamma}(i) = (1 \pm e(t)/s_e)(x_{A, J, \Gamma}(t) \pm \theta(t)/s_e)S_{A, J},$$

where  $x_{A, J, \Gamma}(t) = (2t)^{e_J}q(t)e^{\Gamma-e_J}$  and  $q(t), e(t), \theta(t), s_e$  are as defined above.

Note that  $x_{\phi_0, \bar{e}, e}(t) = q(t)$ , so the good event pertaining to  $Q(i)$  is  $Q(i) = (1 \pm e(t)/s_e)(q(t) \pm \theta(t)/s_e)n^2$ . We also write  $c(t) = \text{aut}(H)^{-1} \times \sum_T x_{\phi_T, J_T, \Gamma_T}(t)$ , where as above the sum is over quadruples  $T = (a, b, c, d)$  such that  $ab$  and  $cd$  are distinct edges of  $H$ .

Now we give an informal derivation of the differential equations satisfied by the functions  $x_{A, J, \Gamma}(t)$ , which describe the main terms for the behaviour of the variables  $X_{\phi, J, \Gamma}$ . We stress that this discussion does not constitute a proof of Theorem 1.4; rather, it motivates the functions  $x_{A, J, \Gamma}(t)$  defined above, and presages the proper proof given below, in which the calculations we make here will play a central role. For the sake of the discussion we ignore the error terms described by  $e(t)$  and  $s_e$ , and use the approximations  $X_{\phi, J, \Gamma}(i) \approx x_{A, J, \Gamma}(t)S_{A, J}$ , so  $Q(i) \approx q(t)n^2$  and  $C_{uv}(i) \approx c(t)p^{e_H-2}n^{v_H-2} = c(t)p^{-1}$ . The system of differential equations will follow from the approximation  $x_{A, J, \Gamma}(t + s^{-1}) \approx x_{A, J, \Gamma}(t) + s^{-1}x'_{A, J, \Gamma}(t)$  and replacing changes  $X_{\phi, J, \Gamma}(i + 1) - X_{\phi, J, \Gamma}(i)$  by their expected value given  $\mathcal{G}_i$ . Intuitively, although the change in a single step may be far from its expected value, over many steps a ‘law of large numbers’ will apply to the accumulated changes. We also ignore two ‘pathological’ behaviours that will need to be dealt with in Sect. 8.

As an illustrative case, we start by counting open edges  $|O(i)| = Q(i)/2$ . When we choose the edge  $e_{i+1}$  we have

$$Q(i + 1) = Q(i) - 1 - C_{e_{i+1}}(i) \approx q(t)n^2 - c(t)p^{-1}.$$

Since

$$Q(i + 1) \approx q(t + 1/s)n^2 \approx (q(t) + s^{-1}q'(t))n^2 = q(t)n^2 + p^{-1}q'(t)$$

we have the equation  $q'(t) = -c(t)$ .

To derive the differential equation for the general extension variable  $x_{A,J,\Gamma}(t)$ , we write

$$X_{\phi,J,\Gamma}(i + 1) - X_{\phi,J,\Gamma}(i) = Y_{\phi,J,\Gamma}^+(i) - Y_{\phi,J,\Gamma}^-(i),$$

where  $Y_{\phi,J,\Gamma}^+(i) \geq 0$  is the number of functions  $f : V_{\Gamma} \rightarrow [n]$  in  $\Xi_{\phi,J,\Gamma}(i + 1) \setminus \Xi_{\phi,J,\Gamma}(i)$ , and  $Y_{\phi,J,\Gamma}^-(i) \geq 0$  is the number of functions  $f : V_{\Gamma} \rightarrow [n]$  in  $\Xi_{\phi,J,\Gamma}(i) \setminus \Xi_{\phi,J,\Gamma}(i + 1)$ . The term  $Y_{\phi,J,\Gamma}^+(i)$  has contributions corresponding to each edge  $e$  of  $J$ . A function  $f$  in  $\Xi_{\phi,J \setminus e,\Gamma}(i)$  will be counted by  $Y_{\phi,J,\Gamma}^+(i)$  if the process chooses the edge  $e_{i+1}$  equal to  $f(e)$ . Since  $e_{i+1}$  is chosen uniformly at random among  $Q(i)/2$  open edges, we can estimate

$$\mathbb{E}(Y_{\phi,J,\Gamma}^+(i)|\mathcal{G}_i) \approx \frac{2}{Q(i)} \sum_{e \in J} X_{\phi,J \setminus e,\Gamma}(i) \approx \frac{2p^{-1}S_{A,J}}{q(t)n^2} \cdot \sum_{e \in J} x_{A,J \setminus e,\Gamma}(t).$$

The term  $Y_{\phi,J,\Gamma}^-(i)$  has contributions corresponding to each edge  $e$  of  $\Gamma \setminus J$ . A function  $f$  in  $\Xi_{\phi,J,\Gamma}(i)$  will be counted by  $Y_{\phi,J,\Gamma}^-(i)$  if the process either chooses the edge  $e_{i+1}$  equal to  $f(e)$  or  $f(e)$  becomes closed, i.e.  $f(e) \in C(i + 1)$ . Thinking of  $e_{i+1}$  as an ordered pair, the number of choices is  $2 + C_{f(e)}(i)$ , each occurring with probability  $Q(i)^{-1}$ . Therefore

$$\begin{aligned} \mathbb{E}(Y_{\phi,J,\Gamma}^-(i)|\mathcal{G}_i) &= \frac{1}{Q(i)} \sum_{e \in \Gamma \setminus J} \sum_{f \in \Xi_{\phi,J,\Gamma}(i)} (2 + C_{f(e)}(i)) \\ &\approx (e_{\Gamma} - e_J) \frac{c(t)p^{-1}x_{A,J,\Gamma}(t)S_{A,J}}{q(t)n^2}. \end{aligned}$$

On the other hand, we have

$$\begin{aligned} Y_{\phi,J,\Gamma}^+(i) - Y_{\phi,J,\Gamma}^-(i) &= X_{\phi,J,\Gamma}(i + 1) - X_{\phi,J,\Gamma}(i) \\ &\approx (x_{A,J,\Gamma}(t + s^{-1}) - x_{A,J,\Gamma}(t))S_{A,J} \\ &\approx s^{-1}x'_{A,J,\Gamma}(t)S_{A,J} \end{aligned}$$

so we have the equation

$$q(t)x'_{A,J,\Gamma}(t) = 2 \sum_{e \in J} x_{A,J \setminus e,\Gamma}(t) - (e_\Gamma - e_J)c(t)x_{A,J,\Gamma}(t). \tag{1}$$

Note that the equation  $q'(t) = -c(t)$  derived above is simply a special case of (1).

To solve these equations we use the substitution  $x_{A,J,\Gamma}(t) = q(t)^{e_\Gamma - e_J} \times z_\ell(t)$ , where we will see that the functions  $z_\ell(t)$  can be parameterised by a single number  $\ell = e_J$ . Then, since  $q'(t) = -c(t)$ , we have  $q(t)x'_{A,J,\Gamma}(t) = q(t)^{e_\Gamma - e_J + 1}z'_\ell(t) - c(t)(e_\Gamma - e_J)q(t)^{e_\Gamma - e_J}z_\ell(t)$ , which also equals

$$\begin{aligned} & 2 \sum_{e \in J} x_{A,J \setminus e,\Gamma}(t) - (e_\Gamma - e_J)c(t)x_{A,J,\Gamma}(t) \\ &= 2\ell q(t)^{e_\Gamma - e_J + 1}z_{\ell-1}(t) - (e_\Gamma - e_J)c(t)q(t)^{e_\Gamma - e_J}z_\ell(t). \end{aligned}$$

We deduce that  $z'_\ell(t) = 2\ell z_{\ell-1}(t)$ . Now we use the initial conditions that  $x_{A,J,\Gamma}(0)$  is equal to 1 if  $e_J = 0$ , otherwise 0 (e.g.  $q(0) = 1$ ). So  $z_0(0) = 1$  and  $z_\ell(0) = 0$  for  $\ell > 0$ . We obtain the solution  $z_\ell(t) = (2t)^\ell$ . Also  $q'(t) = -c(t) = -\text{aut}(H)^{-1} \sum_T x_{\phi_T, J_T, \Gamma_T}(t) = -\text{aut}(H)^{-1} 4e_H(e_H - 1) \times q(t)(2t)^{e_H - 2}$ . Integrating and substituting we conclude that

$$\begin{aligned} q(t) &= e^{-2e_H \text{aut}(H)^{-1} (2t)^{e_H - 1}}, \\ x_{A,J,\Gamma}(t) &= (2t)^{e_J} e^{-2(e_\Gamma - e_J)e_H \text{aut}(H)^{-1} (2t)^{e_H - 1}} = (2t)^{e_J} q(t)^{e_\Gamma - e_J}. \end{aligned}$$

*Remark* As discussed above, we expect these random variables to evolve as they do in the unconstrained random graph  $G(n, i)$ . Thus it is natural to compare the process  $G(i)$  at step  $i$  to the random graph  $G(n, \rho)$ , where  $\rho n^2 / 2 = i = tpn^2$ , i.e.  $\rho = 2tp$ . In  $G(n, \rho)$  we can define open/closed pairs and the variables  $X_{\phi, J, \Gamma}(i)$ . For any ordered pair  $uv$  in  $[n]$ , edge  $ab$  of  $H$  and function  $f : V_H \rightarrow [n]$  with  $f(a) = u, f(b) = v$  the edges of  $f(H \setminus ab)$  will all be present in  $G(n, \rho)$  with probability  $\rho^{e_H - 1}$ . (For the purpose of this discussion we ignore the negligible contributions from functions  $f$  that are not injective.) Given  $uv$ , there are  $2e_H n^{v_H - 2}$  such functions  $f : V_H \rightarrow [n]$ , corresponding to  $2e_H \text{aut}(H)^{-1} n^{v_H - 2}$  distinct sets of edges. The probability that  $uv$  is open should be approximately

$$\begin{aligned} (1 - \rho^{e_H - 1})^{2e_H \text{aut}(H)^{-1} n^{v_H - 2}} &\approx \exp(-(2tp)^{e_H - 1} 2e_H \text{aut}(H)^{-1} n^{v_H - 2}) \\ &= q(t). \end{aligned}$$

Similar reasoning applies to general extension variables, and the equations we derived above agree with the corresponding equations for  $G(n, \rho)$ . (See

Spencer [31] for results on extension variables in this model.) We could use this correspondence as the starting point of our discussion and as a heuristic for the trajectories our variables follow, but this would not provide any insight into how to prove that our random variables actually follow the given trajectories. As we noted above, the calculations in this section play a central role in the proof of Theorem 1.4.

### 3 Strictly balanced graphs and balanced extensions

In this section we obtain some basic properties of our fixed strictly 2-balanced graph  $H$ . We also introduce a more general concept of strictly balanced extensions, and discuss the manner by which arbitrary extensions can be decomposed into a series of such extensions. First we recall the relevant definitions. We suppose that  $H$  is *strictly 2-balanced*, in the sense that  $v_H, e_H \geq 3$  and  $\frac{e_H-1}{v_H-2} > \frac{e_K-1}{v_K-2}$  for all proper subgraphs  $K$  of  $H$  with  $v_K \geq 3$ . We also fix the parameter

$$p = n^{-\frac{v_H-2}{e_H-1}}.$$

For any graph  $\Gamma$  we define the *scaling* of  $\Gamma$  to be  $S_\Gamma = n^{v_\Gamma} p^{e_\Gamma}$ . The condition that  $H$  is strictly 2-balanced can be also be written as  $S_K > S_H$  for all subgraphs  $K$  of  $H$  with  $2 < v_K < v_H$ , since  $S_H = n^{v_H} p^{e_H} = pn^2$  and

$$S_K/S_H = n^{v_K-2} p^{e_K-1} = n^{(e_K-1)(\frac{v_K-2}{e_K-1} - \frac{v_H-2}{e_H-1})} > 1.$$

Note that the scaling  $S_\Gamma$  is always an integer power of  $n^{1/(e_H-1)}$ . It follows that the inequality  $S_\Gamma > 1$  actually implies  $S_\Gamma \geq n^{1/(e_H-1)}$  and similarly that  $S_\Gamma < 1$  implies  $S_\Gamma \leq n^{-1/(e_H-1)}$ .

The following lemma collects some simple properties of  $H$  and  $p$ .

#### Lemma 3.1

- (i) If  $d$  is the largest integer for which  $np^{d-1} > 1$  then  $H$  has minimum degree at least  $d$ .
- (ii) We have  $p > 1/n$ , and so  $H$  has minimum degree at least 2.
- (iii)  $H$  is a 2-connected graph, and if  $\{x, y\}$  is a cutset then  $xy \notin E_H$ .

*Proof* First note that  $H$  cannot have a vertex  $v$  of degree at most  $d - 1$ : otherwise  $S_H/S_{H \setminus v} = np^{d(v)} > 1$ , which contradicts the fact that  $H$  is strictly 2-balanced. We deduce that  $H$  has minimum degree at least 1. Next, suppose for a contradiction that  $p \leq 1/n$ . Then  $e_H \leq v_H - 1$ . However, for every connected subgraph  $K$  of  $H$  we have  $e_K \geq v_K - 1$ , so  $\frac{e_K-1}{v_K-2} \geq 1 \geq \frac{e_H-1}{v_H-2}$ , which contradicts the definition of  $H$  being strictly 2-balanced. Therefore

$p > 1/n$ . Now suppose for a contradiction that  $H$  is not 2-connected. Then we can write  $V_H = X \cup Y$  so that  $E_H = E_{H[X]} \cup E_{H[Y]}$  and  $|X \cap Y| = 1$ . Then  $S_{H[X]}S_{H[Y]} = nS_H$ , so without loss of generality we have  $S_{H[X]} \leq (nS_H)^{1/2}$ , and since  $S_H = pn^2$  we have  $S_{H[X]}/S_H \leq (n/S_H)^{1/2} = (1/pn)^{1/2} < 1$ . This contradicts  $H$  being strictly 2-balanced, so  $H$  is 2-connected. Finally, suppose that  $\{x, y\}$  is a cutset, but that  $xy \in E_H$ . Write  $V_H = X \cup Y$  so that  $E_H = E_{H[X]} \cup E_{H[Y]}$  and  $X \cap Y = \{x, y\}$ . Then  $S_{H[X]}S_{H[Y]} = pn^2S_H = (pn^2)^2$ , so without loss of generality  $S_{H[X]} \leq pn^2 = S_H$ . But this contradicts  $H$  being strictly 2-balanced, so  $xy \notin E_H$ .  $\square$

Recall that if  $\Gamma$  is a graph and  $A \subseteq V_\Gamma$  we define the *scaling* of the pair  $(A, \Gamma)$  to be

$$S_{A,\Gamma} = p^{e_\Gamma - e_{\Gamma[A]}} n^{v_\Gamma - |A|}.$$

Note that  $S_{A,\Gamma} = S_\Gamma/S_{\Gamma[A]}$ . Also, for any  $A \subseteq B \subseteq V_\Gamma$  we have  $S_{B,\Gamma} = S_\Gamma/S_{\Gamma[B]} = S_\Gamma/S_{\Gamma[A]} \cdot S_{\Gamma[A]}/S_{\Gamma[B]} = S_{A,\Gamma}/S_{A,\Gamma[B]}$ . We say that  $(A, \Gamma)$  is *strictly balanced* if for any  $A \subsetneq B \subsetneq V_\Gamma$  we have  $S_{A,\Gamma} < S_{A,\Gamma[B]}$ , or equivalently  $S_{B,\Gamma} < 1$ . For example, we can again rephrase our assumption that  $H$  is strictly 2-balanced to say that for any edge  $e = ab$  of  $H$ , with  $A = \{a, b\}$  the pair  $(A, H)$  is strictly balanced. Indeed,  $S_{A,H} = p^{e_H - 1} n^{v_H - 2} = 1$ , and for  $A \subsetneq B \subsetneq V_H$  we have  $S_{B,H} = S_H/S_{H[B]} < 1$ .

We will apply results on strictly balanced extensions to arbitrary pairs  $(A, \Gamma)$  using the *extension series*  $A = B_0 \subsetneq B_1 \subsetneq \dots \subsetneq B_d = V_\Gamma$  of  $(A, \Gamma)$ , which we construct by the following rule. If  $(B_i, \Gamma)$  is not strictly balanced then  $B_{i+1}$  is chosen to be a minimal set  $C$  with  $B_i \subsetneq C \subsetneq V_\Gamma$  that minimises  $S_{B_i,\Gamma[C]} = n^{|C|-|B_i|} p^{e_{\Gamma[C]} - e_{\Gamma[B_i]}}$ , otherwise we choose  $B_d = B_{i+1} = V_\Gamma$ . For more compact notation we also write  $S_i^A(\Gamma) = S_{B_i,\Gamma[B_{i+1}]}$ . We note the following properties of extension series.

- $(B_i, \Gamma[B_{i+1}])$  is strictly balanced.
- For  $i \geq 1$  we have  $S_i^A(\Gamma) = S_{B_i,\Gamma[B_{i+1}]} = S_{B_{i-1},\Gamma[B_{i+1}]} / S_{B_{i-1},\Gamma[B_i]} \geq 1$ .  
Therefore the sequence  $S_{A,\Gamma[B_i]} = \prod_{j=0}^{i-1} S_j^A(\Gamma)$  is non-decreasing. However, it is not necessarily true that the sequence of successive factors  $S_i^A(\Gamma)$  is non-decreasing. For example, consider the  $K_7$ -free process, where  $p = n^{-1/4}$ , and let  $\Gamma = K_4$ . Choosing  $A$  of size 2 we have  $\Gamma[B_0] = K_2$ ,  $\Gamma[B_1] = K_3$ ,  $\Gamma[B_2] = K_4$  with  $S_0^A(\Gamma) = np^2 = n^{1/2}$  and  $S_1^A(\Gamma) = np^3 = n^{1/4}$ .
- It is possible that  $S_{A,\Gamma} < 1$  but some factors  $S_i^A(\Gamma)$  are greater than 1. For example, consider the  $C_5$ -free process, where  $p = n^{-3/4}$ , and let  $\Gamma$  be the graph consisting of  $K_4$  plus an isolated vertex. Choosing  $A$  to be 2 vertices of the  $K_4$  we have  $\Gamma[B_0] = K_2$ ,  $\Gamma[B_1] = K_4$ ,  $\Gamma[B_2] = \Gamma$ , so  $S_0^A(\Gamma) = n^2 p^5 = n^{-7/4}$ ,  $S_1^A(\Gamma) = n$  and  $S_{A,\Gamma} = n^{-3/4}$ .



### 4 Union bounds

In this section we collect some useful properties of the  $H$ -free process, assuming that the good events  $\mathcal{G}_i$  hold. Recall that on  $\mathcal{G}_i$  we have  $Q(i) = (1 \pm e(t)/s_e)(q(t) \pm \theta(t)/s_e)n^2$ , and  $q(t) = \exp(-\Theta(t^{e_H-1}))$ , where the constant in the  $\Theta$ -notation depends only on  $H$ . We analyse the process up to time  $t_{\max} = m/s = \mu(\log n)^{1/(e_H-1)}$ , and choose  $\mu > 0$  sufficiently small so that  $e(t), q(t)^{-V} < n^\epsilon$ . Since  $s_e = n^{1/2e_H-\epsilon}$  we have  $Q(i) > n^{2-\epsilon}$  (say) for  $i \leq m$ . The following lemmas use this lower bound for  $Q(i)$  and union bound estimates. We will state the bounds at time  $m$ , but they also hold at any time  $i \leq m$  by monotonicity. Our first lemma bounds the probability that  $G(m)$  contains some fixed graph  $F$ .

**Lemma 4.1** *For any fixed graph  $F$  on  $[n]$ , the probability that  $\mathcal{G}_m$  holds and  $G(m)$  contains  $F$  is at most  $p^{e_F} n^{2e_F\epsilon}$ .*

*Proof* We take a union bound over all choices of steps  $1 \leq i_1 \leq \dots \leq i_{e_F} \leq m$  where the  $j$ th edge of  $F$  is chosen as the edge  $e_{i_j}$  added to form  $G(i_j)$  from  $G(i_j - 1)$ . Since edges are chosen uniformly at random from at least  $n^{2-\epsilon}$  options, each choice has probability at most  $n^{-(2-\epsilon)}$  conditional on the history of the process. Therefore  $\mathbb{P}(F \subseteq G(m)) \leq m^{e_F} n^{-(2-\epsilon)e_F} < p^{e_F} n^{2e_F\epsilon}$ , say, since  $m = \mu(\log n)^{1/(e_H-1)} pn^2$ . □

Given sets  $A, B \subseteq [n]$ , write  $e(A, B)$  for the number of edges in  $G(m)$  that have one endpoint in  $A$  and the other in  $B$ . Our next lemma gives a bound for  $e(A, B)$  holding with high probability for all choices of  $A, B$  of specified size.

**Lemma 4.2** *For any  $a, b \geq 1$ , the probability  $p_{a,b}$  that  $\mathcal{G}_m$  holds and there exist sets  $A, B \subseteq [n]$  such that  $|A| = a, |B| = b$  and  $e(A, B) \geq \max\{4\epsilon^{-1}(a + b), pabn^{2\epsilon}\}$  satisfies  $p_{a,b} < n^{-(a+b)}$ .*

*Proof* Write  $x = \max\{4\epsilon^{-1}(a + b), pabn^{2\epsilon}\}$ . We take a union bound over  $\binom{n}{a}$  choices for  $A, \binom{n}{b}$  choices for  $B$ , at most  $\binom{ab}{x}$  ways to choose  $x$  pairs with one endpoint in  $A$  and the other in  $B$ , and less than  $m^x$  choices of steps  $1 \leq i_1 < \dots < i_x \leq m$  in which to choose these pairs as edges of the process. Since edges are chosen uniformly at random from at least  $n^{2-\epsilon}$  options, each choice has probability at most  $n^{-(2-\epsilon)}$  conditional on the history of the process. Therefore we can estimate the probability by  $p_{a,b} < \binom{n}{a} \binom{n}{b} \binom{ab}{x} m^x n^{-(2-\epsilon)x}$ . Since  $m = \mu(\log n)^{1/(e_H-1)} pn^2$ , we have

$$\begin{aligned} \log p_{a,b} &< a(\log(n/a) + 1) + b(\log(n/b) + 1) \\ &\quad + x(\log(ab/x) + 1 + \log(pn^\epsilon)) + \log \mu + (e_H - 1)^{-1} \log \log n \\ &< (a + b - \epsilon x/2) \log n, \end{aligned}$$

since  $x \geq pabn^{2\epsilon}$  and  $n$  large imply that  $-(\log(ab/x) + \log(pn^\epsilon)) \geq \epsilon \log n \gg \log \log n$ . Since  $x \geq 4\epsilon^{-1}(a + b)$  the stated bound follows.  $\square$

For  $A \subseteq [n]$  let  $D_{A,d}$  be the set of vertices  $v$  such that  $|N_{G(m)}(v) \cap A| \geq d$ , i.e. in  $G(m)$ ,  $v$  has at least  $d$  neighbours in  $A$ . We conclude this section by applying the previous lemma to give an upper bound for  $D_{A,d}$ .

**Lemma 4.3** *For any  $8\epsilon^{-1} \leq d \leq a \leq dp^{-1}n^{-2\epsilon}$ , the probability that  $\mathcal{G}_m$  holds and there exists  $A \subseteq [n]$  with  $|A| = a$  and  $|D_{A,d}| \geq 8\epsilon^{-1}d^{-1}a$  is at most  $n^{-a}$ .*

*Proof* Set  $B = D_{A,d}$ ,  $b = |B|$  and consider the event that  $b \geq 8\epsilon^{-1}d^{-1}a$ . Since  $e(A, B) \geq db$  and  $d \geq 8\epsilon^{-1}$  we have  $e(A, B) - 4\epsilon^{-1}b \geq db/2 \geq 4\epsilon^{-1}a$ . Also, the bound  $a \leq dp^{-1}n^{-2\epsilon}$  implies that  $e(A, B) \geq db \geq pabn^{2\epsilon}$ . By Lemma 4.2 this event has probability at most  $n^{-(a+b)} \leq n^{-a}$ .  $\square$

### 5 Counting extensions

In this section we see how to obtain general upper bounds on extension variables, assuming that the good events  $\mathcal{G}_i$  hold. We will state the bounds at time  $m$ , but they also hold at any time  $i \leq m$  by monotonicity. Let  $N_{\phi,J} = X_{\phi,J,J}(m)$ : the number of extensions of a fixed embedding  $\phi : A \rightarrow [n]$  to an embedding  $f : J \rightarrow G(m)$ , where  $A \subseteq V_J$  is independent. Note that this is an upper bound for  $X_{\phi,J,\Gamma}(m)$ . The following lemma gives a good estimate on  $N_{\phi,J}$  when the extension is strictly balanced.

**Lemma 5.1** *Suppose  $(A, J)$  is strictly balanced and  $\phi : A \rightarrow [n]$  is an injective map. Let  $\omega(n)$  be any function such that  $\omega(n) \rightarrow \infty$  as  $n \rightarrow \infty$ . On  $\mathcal{G}_m$ , with high probability we have  $N_{\phi,J} < S_{A,J}n^{4e_J\epsilon}$  if  $S_{A,J} \geq 1$  and  $N_{\phi,J} < \omega(n)$  if  $S_{A,J} < 1$ .*

*Proof* We start by estimating the maximum number of vertex-disjoint extensions of  $\phi$  to an embedding of  $J$ . Let  $N'_{\phi,J}$  be the maximum number  $s$  such that there are embeddings  $f_1, \dots, f_s$  of  $J$  in  $G(m)$ , all restricting to  $\phi$  on  $A$ , with  $f_i(V_J \setminus A)$  and  $f_j(V_J \setminus A)$  disjoint for all  $1 \leq i < j \leq s$ . We can estimate  $\mathbb{P}(N'_{\phi,J} \geq s)$  by a union bound over at most  $s!^{-1}(n^{v_J - |A|})^s$  possible functions  $f_1, \dots, f_s$ , where for each choice of functions, we can apply

Lemma 4.1 to obtain an upper bound  $p^{se_J} n^{2se_J\epsilon}$  on the probability that the graph  $F = \bigcup_{i=1}^s f_i(J)$  is a subgraph of  $G(m)$ . Therefore

$$\mathbb{P}(N'_{\phi,J} \geq s) \leq s!^{-1} (n^{v_J - |A|})^s p^{se_J} n^{2se_J\epsilon} < (3s^{-1} S_{A,J} n^{2e_J\epsilon})^s.$$

If  $S_{A,J} \geq 1$  then we can set  $s = S_{A,J} n^{3e_J\epsilon}$  to get a bound holding with failure probability much less than  $\exp(-n^\epsilon)$ . On the other hand, if  $S_{A,J} = p^{e_J} n^{v_J - |A|} < 1$  then, since  $p = n^{-\frac{v_H - 2}{e_H - 1}}$ , we in fact have  $S_{A,J} \leq n^{-1/(e_H - 1)}$ . Assuming that  $\epsilon < (2e_J e_H)^{-1}$  we then have  $S_{A,J} n^{2e_J\epsilon} < 1$ , and we can set  $s = \omega'(n)$  for any function  $\omega'(n) \rightarrow \infty$  as  $n \rightarrow \infty$  to get a bound holding with failure probability much less than  $n^{-C}$  for any constant  $C > 0$ .

Now we argue by induction on  $v_J - |A|$  to show the following bounds on  $N_{\phi,J}$ : if  $S_{A,J} \geq 1$  then  $N_{\phi,J} < S_{A,J} n^{3e_J\epsilon} \omega'(n)^{2(v_J - |A|)}$  and if  $S_{A,J} < 1$  then  $N_{\phi,J} < \omega'(n)^{2(v_J - |A|)}$ . Then we can choose  $\omega'(n)^{2(v_J - |A|)} < \omega(n) < n^\epsilon$  to obtain the bounds required for the theorem. Our base case is  $v_J - |A| = 1$ , when we have  $N_{\phi,J} = N'_{\phi,J}$ , and we can apply the bounds just shown for  $N'_{\phi,J}$ .

Next suppose  $v_J - |A| > 1$ . We claim that for any embedding  $f$  counted by  $N_{\phi,J}$  there are at most  $\omega'(n)^{2(v_J - |A|) - 1}$  embeddings  $f'$  counted by  $N_{\phi,J}$  with  $f'(V_J \setminus A) \cap f(V_J \setminus A) \neq \emptyset$ . To see this, consider any such  $f'$  and let  $B = \{b \in V_J : f'(b) \in f(V_J)\}$ , so that  $A \subsetneq B \subsetneq V_J$ . Let  $\phi'$  be the restriction of  $f'$  to  $B$  and let  $J' = J \setminus E_{J[B]}$  be the graph obtained from  $J$  by deleting all edges inside  $B$ . Then, as noted above,  $S_{B,J'} = S_{A,J} / S_{A,J[B]}$ , and since  $(A, J)$  is strictly balanced we have  $S_{B,J'} < 1$ . By induction hypothesis we have  $N_{\phi',J'} < \omega'(n)^{2(v_J - |B|)}$ . Also, there are at most  $v_J^{|B| - |A|} < v_J^{v_J}$  choices for  $\phi'$ , so at most  $v_J^{v_J} \omega'(n)^{2(v_J - |B|)}$  embeddings  $f'$  corresponding to this set  $B$ . Summing over all  $A \subsetneq B \subsetneq V_J$  we obtain at most  $\omega'(n)^{2(v_J - |A|) - 1}$  (say) such embeddings  $f'$ .

Finally, we can estimate  $N_{\phi,J}$  by means of a maximum collection  $F = \{f_1, \dots, f_s\}$  of vertex-disjoint extensions of  $\phi$  (so  $|F| = N'_{\phi,J}$ ). Any extension  $f$  counted by  $N_{\phi,J}$  has a common image with some  $f_i \in F$  outside of  $A$ , and for each  $f_i \in F$  we have at most  $\omega'(n)^{2(v_J - |A|) - 1}$  such embeddings  $f$ . Therefore  $N_{\phi,J} \leq N'_{\phi,J} \omega'(n)^{2(v_J - |A|) - 1}$ . If  $S_{A,J} \geq 1$  then  $N'_{\phi,J} < S_{A,J} n^{3e_J\epsilon}$  and so  $N_{\phi,J} < S_{A,J} n^{3e_J\epsilon} \omega'(n)^{2(v_J - |A|)}$ . On the other hand, if  $S_{A,J} < 1$  then  $N'_{\phi,J} < \omega'(n)$  and so  $N_{\phi,J} < \omega'(n)^{2(v_J - |A|)}$ . This completes the proof.  $\square$

For general extensions  $N_{\phi,J}$  may be considerably larger than  $S_{A,J}$ , but the following lemma gives a useful bound.

**Lemma 5.2** *On  $\mathcal{G}_m$ , with high probability we have  $N_{\phi,J} < n^{4e_J\epsilon} \times \max_{A \subsetneq B \subsetneq V_J} S_{B,J}$ .*

*Proof* Consider the extension series  $A = B_0 \subsetneq B_1 \subsetneq \dots \subsetneq B_d = V_J$ . We repeatedly apply Lemma 5.1 to bound the number of extensions in each step of the series. At the first step we either have  $S_0^A(J) < 1$  and so  $N_{\phi, J[B_1]} < \omega(n)$  or  $S_0^A(J) \geq 1$  and so  $N_{\phi, J[B_1]} < S_0^A(J)n^{4e_{J[B_1]}\epsilon}$ . At subsequent steps  $i \geq 1$  we have  $S_i^A(J) \geq 1$ , so for each injection  $\phi' : B_i \rightarrow [n]$  we have  $N_{\phi', J_i[B_{i+1}]} < S_i^A(J)n^{4(e_{J[B_{i+1}]} - e_{J[B_i]})\epsilon}$ . Multiplying these bounds and using  $S_{A, J} = \prod_{i=0}^{d-1} S_i^A(J)$  gives a bound equal to either  $n^{4e_{J}\epsilon} S_{A, J}$  when  $S_0^A(J) \geq 1$  or  $\omega(n)n^{4(e_{J[B_1]})\epsilon} S_{B_1, J}$  when  $S_0^A(J) < 1$ . By definition of the extension series,  $\max_{A \subseteq B \subseteq V_J} S_{B, J}$  is either  $S_{A, J}$  when  $S_0^A(J) \geq 1$  or  $S_{B_1, J}$  when  $S_0^A(J) < 1$ . Also, we may assume that  $e_{J[B_1]} \geq 1$  (otherwise  $E_J$  is empty), so we can choose  $\omega(n) < n^\epsilon$  to obtain the required bound.  $\square$

*Remark* In both of the preceding lemmas we can choose  $\omega(n) = n^{c\epsilon}$  for some constant  $c > 0$  to make the failure probability exponentially small.

We say that the pair  $(A, J)$  is *dense* if  $S_0^A(J) = S_{A, J[B_1]} \geq 1$  and *strictly dense* if  $S_0^A(J) > 1$  (and so  $S_0^A(J) \geq n^{1/(e_H - 1)}$ ). Since  $S_i^A(J) \geq 1$  for  $i \geq 1$ , for a dense pair we have  $\max_{A \subseteq B \subseteq V_J} S_{B, J} = S_{A, J}$ , so the previous lemma gives an approximate upper bound of  $S_{A, J}$  for  $N_{\phi, J}$ . Note that if  $(A, J)$  is strictly dense then so is  $(A, J')$  for any subgraph  $J'$  of  $J$ , since we have  $S_{A, J'[B]} \geq S_{A, J[B]} > 1$  for any  $B$  with  $A \subsetneq B \subseteq V_J$ . The same argument shows that if  $J$  is a subgraph of  $H$  with  $e_J \leq e_H - 2$  and  $A = \{u, v\}$ , where  $uv \in E_H \setminus E_J$ , then  $(A, J)$  is strictly dense.

We conclude this section by showing that adding an edge to a strictly dense pair gives a significant improvement on the bound for  $N_{\phi, J}$ .

**Lemma 5.3** *Suppose that  $(A, J)$  is a strictly dense pair,  $a, b$  are vertices of  $J$  with  $ab \notin E_J$  and  $\{a, b\} \not\subseteq A$ , and  $J' = J \cup \{ab\}$  is obtained by adding the edge  $ab$  to  $J$ . Then  $\max_{A \subseteq B \subseteq V_{J'}} S_{B, J'} < S_{A, J}$ , and so on  $\mathcal{G}_m$ , with high probability we have  $N_{\phi, J'} < n^{-1/(e_H - 1) + 4e_{J'}\epsilon} S_{A, J}$ .*

*Proof* Choose  $B$  with  $A \subseteq B \subseteq V_J$  maximising  $S_{B, J'}$ . If  $B = A$  we have  $S_{B, J'} = pS_{A, J}$ , whereas if  $B \neq A$  we have  $S_{B, J'} \leq S_{B, J} = S_{A, J}/S_{A, J[B]} < S_{A, J}$ , as  $(A, J)$  is strictly dense. Either way we have  $S_{B, J'} \leq n^{-1/(e_H - 1)} S_{A, J}$ , since it is an integer power of  $n^{1/(e_H - 1)}$ , so the bound on  $N_{\phi, J'}$  follows from Lemma 5.2.  $\square$

### 6 Closure fidelity

Recall that for an ordered pair  $uv \in O(i)$ , we write  $C_{uv}(i)$  for the set of ordered pairs  $xy \in O(i)$  that would become closed, i.e. belong to  $C(i + 1)$ ,

if at time  $i + 1$  the process chooses  $uv$  as the edge  $e_{i+1}$ . By definition of  $C(i + 1)$  this means that adding  $uv$  and  $xy$  to  $G(i)$  would create a copy of  $H$ . Also, since  $uv$  and  $xy$  are open, any such copy of  $H$  must use both  $uv$  and  $xy$ . In principle there could be many such copies of  $H$ , but we will show in this section that in fact this is not the case, and moreover, by counting these copies of  $H$  we obtain an accurate estimate for the number of pairs closed by  $uv$ .

We frequently need to estimate the number of overlapping extensions of two pairs  $(A_1, J_1)$  and  $(A_2, J_2)$ , so we will introduce some notation for this situation. Recall that a graph  $W$  is a *join* of two graphs  $W_1$  and  $W_2$  if it has subgraphs  $J_1$  isomorphic to  $W_1$  and  $J_2$  isomorphic to  $W_2$  such that  $V_W = V_{J_1} \cup V_{J_2}$  and  $E_W = E_{J_1} \cup E_{J_2}$ . For convenient notation we use names for vertices in  $J_1$  interchangeably with their corresponding vertices in  $W_1$ , and similarly for  $J_2$  and  $W_2$ . Whenever we use this notation the sets  $A_1$  and  $A_2$  will be independent and we will write  $C = V_{J_1} \cap V_{J_2}$ .

We need some further notation for describing the possibilities by which a pair  $uv$  can close a pair  $xy$ . There must be a subgraph  $J$  obtained by deleting two edges  $ab$  and  $cd$  from  $H$  and an injective map  $f : V_H \rightarrow [n]$  such that  $f(a) = u, f(b) = v, f(c) = x, f(d) = y$  and  $f(e) \in E(i)$  for every edge of  $J$ . The map  $f$  is counted by  $X_{\phi_T, J_T, \Gamma_T}(i)$ , where given such a quadruple  $T = (a, b, c, d)$ , we write  $\Gamma_T = H \setminus ab, J_T = H \setminus \{ab, cd\}$  and define  $\phi_T$  by  $\phi_T(a) = u$  and  $\phi_T(b) = v$ .

For the sake of an argument needed in the proof of Lemma 11.1 we extend the definition of  $C_{uv}(i)$  to allow the case when  $uv \in C(i)$  is a closed pair: we define it as the number of pairs  $xy$  such that adding  $uv$  and  $xy$  to  $G(i)$  creates a copy of  $H$  containing both  $uv$  and  $xy$ .

**Lemma 6.1** *With high probability, for every  $1 \leq i \leq m$  and ordered pair  $uv \in O(i) \cup C(i)$ , assuming  $\mathcal{G}_i$ , we have  $|C_{uv}(i)| = \text{aut}(H)^{-1} \sum_T X_{\phi_T, J_T, \Gamma_T}(i) \pm n^{-1/e_H} p^{-1}$ , where the sum is over quadruples  $T = (a, b, c, d)$  such that  $ab$  and  $cd$  are distinct (but not necessarily disjoint) edges of  $H$ .*

*Proof* Let  $P$  be the set of ordered pairs  $xy$  for which there exist (at least) two embeddings  $f_1, f_2$  of  $H$  in  $G(i) \cup \{uv, xy\}$  with  $f_1(E_H) \neq f_2(E_H)$  such that both embedded copies  $f_1(H)$  and  $f_2(H)$  use the edges  $uv$  and  $xy$ . Given any  $xy \in P$  we fix any two such embeddings  $f_1$  and  $f_2$ . Let  $W$  be a graph isomorphic to  $(f_1(H) \cup f_2(H)) \setminus \{uv, xy\}$  and write  $a, b, c, d$  for the vertices in  $W$  corresponding to  $u, v, x, y$  respectively. Note that these are not necessarily distinct, but there are at least 3 distinct vertices in the list, since  $\{u, v\} \neq \{x, y\}$ . Let  $\phi$  be the function defined by  $\phi(a) = u$  and  $\phi(b) = v$ . We bound  $P$  by estimating, for all such  $W$ , the number  $N_{\phi, W}$  of embeddings of  $W$  in  $G(i)$  where  $a$  is mapped to  $u$  and  $b$  to  $v$ .

There are two cases, according to whether or not we have  $f_1(V_H) = f_2(V_H)$ . If  $f_1(V_H) = f_2(V_H)$  then, since  $f_1(E_H) \neq f_2(E_H)$ ,  $W$  is obtained from a subgraph  $J = H \setminus \{ab, cd\}$  of  $H$  by adding at least one edge. As noted above,  $(ab, J)$  is strictly dense, and so by Lemma 5.3 we have  $N_{\phi, W} < n^{-1/(e_H-1)+4e_W\epsilon} p^{-1}$ . Now suppose that  $f_1(V_H) \neq f_2(V_H)$ . We need to estimate  $N_{\phi, W}$  where  $W$  is the join of  $J_1 = f_1(H) \setminus \{uv, xy\}$  and  $J_2 = f_2(H) \setminus \{uv, xy\}$ . With the above notation we have  $A_1 = A_2 = \{a, b\}$ , and  $C = V_{J_1} \cap V_{J_2}$  contains  $\{a, b\}$  and  $\{c, d\}$ , so  $C \setminus A_1$  and  $C \setminus A_2$  are non-empty. Choose  $B$  with  $A_1 \cup A_2 \subseteq B \subseteq V_W$  maximising  $S_{B, W}$  and write  $B_1 = B \cap V_{J_1}$ ,  $B_2 = B \cap V_{J_2}$ . We consider three subcases according to  $B_1$  and  $B_2$ . The first subcase is  $B_1 \cup C \neq V_{J_1}$ . Then we have  $S_{B_1 \cup C, J_1} = S_{B_1 \cup C, H} < 1$ , as  $\{c, d\} \subseteq C$  and  $H$  is strictly 2-balanced. Also  $S_{B_2, J_2} \leq S_{A_2, J_2}$ , since  $(A_2, J_2)$  is (strictly) dense, so  $S_{B, W} \leq S_{B_2, J_2} S_{B_1 \cup C, J_1} < S_{A_2, J_2} = p^{-1}$ . The second subcase is  $B_2 \cup C \neq V_{J_2}$ , when a similar argument gives  $S_{B, W} = S_{B_1, J_1} S_{B_2 \cup C, J_2} < S_{A_1, J_1} = p^{-1}$ . Finally, the third subcase is  $B_1 \cup C = V_{J_1}$  and  $B_2 \cup C = V_{J_2}$ . Then  $V_{J_1} \setminus (A_1 \cup C)$  and  $V_{J_2} \setminus (A_2 \cup C)$  are non-empty, since  $f_1(V_H) \neq f_2(V_H)$ , and thus  $B_1 \neq A_1$  and  $B_2 \neq A_2$ . Since  $(A_1, J_1)$  is strictly dense we have  $S_{B_1, J_1} < S_{A_1, J_1} = p^{-1}$ , so  $S_{B, W} = S_{B_1, J_1} S_{B_2 \cup C, J_2} = S_{B_1, J_1} < p^{-1}$ . In all cases we have  $S_{B, W} < p^{-1}$ , so  $S_{B, W} \leq n^{-1/(e_H-1)} p^{-1}$ , since it is an integer power of  $n^{-1/(e_H-1)}$ . Now Lemma 5.2 gives  $N_{\phi, W} < n^{-1/(e_H-1)+4e_W\epsilon} p^{-1}$ . Summing over less than  $|V_H|^{2|V_H|}$  (say) choices of  $W$  we obtain a bound  $|P| \leq n^{-1/(e_H-1/2)} p^{-1}$ , say.

To finish the proof we calculate the number of ordered pairs  $xy \notin P$  counted by  $C_{uv}(i)$ . For each such pair  $xy$  there is a unique copy  $H^c$  of  $H$  in  $G(i) \cup \{uv, xy\}$ . For each quadruple  $T = (a, b, c, d)$  in  $H$  such that there is an isomorphism  $f : H \rightarrow H^c$  with  $f(a) = u, f(b) = v, f(c) = x, f(d) = y$  we count  $xy$  by  $X_{\phi_T, J_T, \Gamma_T}(i)$ . Also, any other such quadruple  $T' = (a', b', c', d')$  and isomorphism  $f' : H \rightarrow H^c$  with  $f'(a') = u, f'(b') = v, f'(c') = x, f'(d') = y$  corresponds to the automorphism  $f^{-1}f'$  of  $H$ , and this is a one-to-one correspondence. Therefore we can estimate the number of ordered pairs  $xy \notin P$  that close  $uv$  by  $aut(H)^{-1} \sum_T (X_{\phi_T, J_T, \Gamma_T}(i) \pm |P|)$ . Including the pairs in  $P$ , we can estimate  $|C_{uv}(i)|$  by  $aut(H)^{-1} \sum_T X_{\phi_T, J_T, \Gamma_T}(i) \pm n^{-1/e_H} p^{-1}$ , say. This completes the proof.  $\square$

Note that the extension variables which appear in Lemma 6.1 are trackable: they satisfy condition (b) in the definition, since  $uv \notin E(i)$ . Substituting the formulae  $X_{\phi_T, J_T, \Gamma_T}(i) = (1 \pm e(t)/s_e)((2t)^{e_H-2}q(t) \pm \theta(t)/s_e)p^{-1}$  and recalling that  $s_e = n^{1/2e_H-\epsilon} \ll n^{1/e_H}$  we obtain the following estimate.

**Corollary 6.2** *With high probability, for every  $1 \leq i \leq m$  and ordered pair  $uv \in O(i) \cup C(i)$ , assuming  $\mathcal{G}_i$ , we have*

$$|C_{uv}(i)| = (1 \pm e(t)/s_e)((2t)^{e_H-2}q(t) \pm 2\theta(t)/s_e)a_H p^{-1},$$

where  $a_H = 4e_H(e_H - 1)/aut(H)$ .

### 7 Martingale estimates: the differential equations method

Our main tool for establishing concentration of random variables will be the following versions of the Azuma-Hoeffding inequality, Lemmas 6 and 7 from [7]. First we need some definitions. Suppose we have a sequence of random variables  $X_0, X_1, \dots$  and a filtration  $\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \dots$  (which will always be the natural filtration given by the process). We say that the sequence  $X_0, X_1, \dots$  is a *martingale* if  $\mathbb{E}(X_{i+1}|\mathcal{F}_i) = X_i$  for  $i \geq 0$ . We say it is a *submartingale* if  $\mathbb{E}(X_{i+1}|\mathcal{F}_i) \geq X_i$  for  $i \geq 0$  or a *supermartingale* if  $\mathbb{E}(X_{i+1}|\mathcal{F}_i) \leq X_i$  for  $i \geq 0$ . We say that a sequence of random variables  $X_0, X_1, \dots$  is  $(\eta, N)$ -*bounded*, for some  $\eta, N > 0$ , if  $X_i - \eta \leq X_{i+1} \leq X_i + N$  for all  $i \geq 0$ . In our application below we consider sequences of random variables  $A_0, A_1, \dots$  where the difference sequence  $D_i = A_{i+1} - A_i$  satisfies  $0 \leq D_i \leq N$  and  $\mathbb{E}D_i = (1 \pm e_i)d_i$  for some  $d_i \leq \eta/2$  and a small error term  $0 < e_i < 1$ . We will define  $A_i^+ = \sum_{j < i} (D_j - (1 - e_j)d_j)$ , and  $A_i^- = \sum_{j < i} (D_j - (1 + e_j)d_j)$ . Then each of  $A_i^\pm$  is  $(\eta, N)$ -bounded,  $A_i^+$  is a submartingale and  $A_i^-$  is a supermartingale. We refer to  $A_i^\pm$  as a *martingale pair with parameters*  $(\eta, N)$ .

**Lemma 7.1** *Suppose  $\eta \leq N/10, m \geq 1, a > 0$  and  $A_0, A_1, \dots$  is an  $(\eta, N)$ -bounded submartingale. Then  $\mathbb{P}(A_m \leq A_0 - a) \leq e^{-a^2/3\eta m N}$ .*

**Lemma 7.2** *Suppose  $\eta \leq N/10, m \geq 1, 0 < a \leq \eta m/10$  and  $A_0, A_1, \dots$  is an  $(\eta, N)$ -bounded supermartingale. Then  $\mathbb{P}(A_m \geq A_0 + a) \leq e^{-a^2/3\eta m N}$ .*

We now come to the formulation of the differential equations method. Although it is technically involved, the idea behind it is quite simple. We have a collection of sequences of random variables, and would like to prove that certain asymptotic approximations hold with high probability at each step of each sequence. The asymptotic formulae are heuristically derived by considering the one-step expected changes in these variables. We let  $\mathcal{G}_i$  be the event all formulae hold up to step  $i$ . If, conditional on  $\mathcal{G}_i$ , the expected change of a random variable from step  $i$  to step  $i + 1$  is close to what it should be for these formulae to hold, and we also have a useful absolute bound for these one-step changes, then we can apply martingale estimates to show that the event  $\mathcal{G}_i$  indeed holds with high probability. We recommend the survey of Wormald [37] for an introduction to this method, and a comparison of Lemma 7.3 below with Theorem 5.1 in Wormald [37] may be helpful. We also note that Seierstad [29, 30] has recently given improved large deviation bounds and a central limit theorem for the method under certain general criteria. One difference in our theorem is that we phrase our result in terms of a known smooth solution to a system of differential equations, and thus side-step the issue of the existence of a solution. However, the important difference is in the hypothesis for

the bounds on the one-step changes of the variables: by using Lemmas 7.1 and 7.2 we can make do with much weaker estimates than those needed to apply the general result from [37].

*Set-up for Lemma 7.3* Suppose we have a stochastic graph process defined on the vertex set  $[n]$ , where  $n$  is large. Let  $r$  be a fixed positive integer, and for each  $j \in [r]$  let  $k_j, S_j$  be parameters (which can depend on  $n$ ). Suppose that for each  $j \in [r]$  and  $A \in \binom{[n]}{k_j}$  there is a sequence of random variables  $X_{j,A}(i)$ , defined for  $i = 0, \dots, m$  and measurable with respect to the underlying graph process. We suppose further that

$$X_{j,A}(i + 1) - X_{j,A}(i) = Y_{j,A}^+(i) - Y_{j,A}^-(i),$$

where  $Y_{j,A}^+(i), Y_{j,A}^-(i) \geq 0$ . We relate these sequences of random variables to functions on  $[0, \infty)$  by introducing  $t = i/s$  for some function  $s = s(n)$  that goes to infinity. We hope to find a collection  $x_j(t)$  of continuous functions such that

$$X_{j,A}(i) \approx x_j(t)S_j$$

for all  $j \in [r], A \in \binom{[n]}{k_j}$  and  $i = 0, \dots, m$ . Note that in our application  $i$  will be the number of edges that have been added, and we can think of  $s$  as the time-scaling for the underlying process. We can think of  $1 \leq j \leq r$  as the ‘type’ of a random variable and the set  $A$  as giving its ‘position’ in the graph. The parameter  $S_j$  is the size-scaling for the  $j$ th type of random variable.

Now we will formally state our lemma. Note that for technical reasons we also allow the introduction of an additional sequence  $\mathcal{H}_i$  of high probability events.

**Lemma 7.3** *Let  $0 < \epsilon < 1$  and  $c, C > 0$  be constants, and suppose that for each  $j \in [r]$  we have a parameter  $s_j = s_j(n)$ , and functions  $x_j(t), e_j(t), \theta_j(t), \gamma_j(t)$  that are smooth and non-negative for  $t \geq 0$ . For  $i^* = 1, \dots, m$  let  $\mathcal{G}_{i^*}$  be the event that*

$$X_{j,A}(i) = \left(1 \pm \frac{e_j(t)}{s_j}\right) \left(x_j(t) \pm \frac{\theta_j(t)}{s_j}\right) S_j$$

*for all  $1 \leq i \leq i^*, 1 \leq j \leq r$  and  $A \in \binom{[n]}{k_j}$ . Suppose that also there is a decreasing sequence of events  $\mathcal{H}_i, 1 \leq i \leq m$  such that  $\mathbb{P}(\mathcal{H}_m \mid \mathcal{G}_m) \rightarrow 1$  as  $n \rightarrow \infty$ , and that the following conditions hold:*



1. (trend hypothesis) When conditioning on  $\mathcal{G}_i \wedge \mathcal{H}_i$  we have

$$\mathbb{E}Y_{j,A}^\pm(i) = \left( y_j^\pm(t) \pm \frac{h_j(t)}{4s_j} \right) \frac{S_j}{s},$$

for all  $j \in [r]$  and  $A \in \binom{[n]}{k_j}$ , where  $y_j^\pm(t)$  and  $h_j(t)$  are smooth non-negative functions such that

$$x'_j(t) = y_j^+(t) - y_j^-(t) \quad \text{and} \quad h_j(t) = (e_j x_j + \gamma_j)'(t);$$

2. (boundedness hypothesis) For each  $j \in [r]$ , conditional on  $\mathcal{G}_i \wedge \mathcal{H}_i$  we have

$$Y_{j,A}^\pm(i) < \frac{S_j}{s_j^2 k_j n^\epsilon};$$

3. (initial condition) For all  $j \in [r]$  we have  $e_j(0) = \gamma_j(0) = 0$ ; and  $X_{j,A}(0) = S_j x_j(0)$  for all  $A \in \binom{[n]}{k_j}$ ;

4. We have  $n^{3\epsilon} < s < m < n^2$ ,  $m < n^{\epsilon/2} s$ ,  $s \geq 40Cs_j^2 k_j n^\epsilon$ ,  $n^{2\epsilon} \leq s_j < n^{-\epsilon} s$ ,

$$\inf_{t \geq 0} \theta_j(t) + e_j(t)x_j(t)/2 - \gamma_j(t)/2 > c,$$

$$\sup_{t \geq 0} |y_j^\pm(t)| < C, \quad \sup_{t \geq 0} |x'_j(t)| < C, \quad \int_0^\infty |x''_j(t)| dt < C,$$

$$\sup_{0 \leq t \leq m/s} |h_j(t)| < n^\epsilon, \quad \int_0^{m/s} |h'_j(t)| dt < n^\epsilon.$$

Then  $\mathbb{P}(\mathcal{G}_m \wedge \mathcal{H}_m) \rightarrow 1$  as  $n \rightarrow \infty$ .

*Proof* On the event  $\mathcal{G}_i \wedge \mathcal{H}_i$  we define

$$Y_{j,A}^{\pm_1 \pm_2}(i) = Y_{j,A}^{\pm_1}(i) - (y_j^{\pm_1}(t) \mp_2 h_j(t)/4s_j)S_j/s.$$

(Recall our convention that this is shorthand for 4 separate sequences of variables, one for each way of choosing signs for  $\pm_1$  and for  $\pm_2$ .) If any event  $\mathcal{G}_i$  or  $\mathcal{H}_i$  fails we define all  $Y_{j,A}^{\pm_1 \pm_2}(i')$  to be 0 for  $i' > i$ . Define

$$Z_{j,A}^{\pm_1 \pm_2}(i) = \sum_{i'=0}^{i-1} Y_{j,A}^{\pm_1 \pm_2}(i'), \quad N_j = \frac{S_j}{s_j^2 k_j n^\epsilon} \quad \text{and} \quad \eta_j = 4CS_j/s.$$

Using the bounds  $|h_j(t)| < n^\epsilon$ ,  $s_j > n^{2\epsilon}$ ,  $|y_j^\pm(t)| < C$  we see that  $Z_{j,A}^{+\pm}(i)$  and  $Z_{j,A}^{-\pm}(i)$  are martingale pairs with parameters  $(\eta_j, N_j + \eta_j)$ . For example

$Z_{j,A}^{++}(i + 1) - Z_{j,A}^{++}(i) = Y_{j,A}^{++}(i) = Y_{j,A}^+ - (y_j^+(t) - h_j(t)/4s_j)S_j/s$  is a submartingale by the trend hypothesis, is bounded above by  $N_j + n^{-\epsilon}S_j/4s < N_j + \eta_j$  by the boundedness hypothesis and below by  $-CS_j/s > -\eta_j$ . (The other cases are similar.)

Next we need the Euler-Maclaurin summation formula (see [5]), which is as follows. Suppose  $f(t)$  is a smooth function and  $a$  is a natural number. Then  $I = \int_0^a f(i) di$  can be approximated by  $S = \frac{1}{2}f(0) + f(1) + \dots + f(a - 1) + \frac{1}{2}f(a)$  with error  $|S - I| < \int_0^a |f'(i)| di$ . We apply the formula to  $f(i) = x'_j(t(i))$  for any  $j \in [r]$  and  $a = i^*$  with  $1 \leq i^* \leq m$ . Write  $t^* = i^*/s$ . Then

$$I = \int_0^{i^*} x'_j(t(i)) di = \int_0^{t^*} x'_j(\tau)s d\tau = s(x_j(t^*) - x_j(0))$$

and

$$|S - I| < \frac{1}{s} \int_0^{i^*} |x''_j(t(i))| di = \int_0^{t^*} |x''_j(\tau)| d\tau < C,$$

so

$$\begin{aligned} & \left| x_j(t^*) - x_j(0) - \frac{1}{s} \sum_{i=0}^{i^*-1} x'_j(t(i)) \right| \\ & < \frac{1}{s} \left( \left| \frac{x'_j(0)}{2} \right| + \left| \frac{x'_j(t^*)}{2} \right| + \int_0^{t^*} |x''_j(\tau)| d\tau \right) < \frac{3C}{s}. \end{aligned}$$

We can rewrite this as

$$\frac{1}{s} \sum_{i=0}^{i^*-1} x'(t(i))S_j = \left( x_j(t^*) - x_j(0) \pm \frac{3C}{s} \right) S_j. \tag{2}$$

Similarly, our assumptions on  $h_j$  and the initial conditions  $e_j(0) = \gamma_j(0) = 0$  give  $|e_j(t^*)x_j(t^*) + \gamma_j(t^*) - \sum_{i=0}^{i^*-1} h_j(t(i))/s| < 3n^\epsilon/s$ , which we can rewrite as

$$\sum_{i=0}^{i^*-1} h_j(t(i))/4s_j \cdot S_j/s = (e_j(t^*)x_j(t^*) + \gamma_j(t^*) \pm 3n^\epsilon/s)S_j/4s_j. \tag{3}$$

Now we will estimate the probability that any event  $\mathcal{G}_i$  fails. We can restrict attention to events where all  $\mathcal{H}_i$  hold, as by assumption they all hold with high probability. Fix  $1 \leq j \leq k$ ,  $A \in \binom{[n]}{k_j}$ ,  $1 \leq i^* \leq m$ ,  $t^* = i^*/s$ . Consider the event that  $i^*$  is the first step at which  $\mathcal{H}_{i^*}$  holds but  $\mathcal{G}_{i^*}$  fails and that it fails for the variable  $X_{j,A}(i^*)$ . One possibility is that  $X_{j,A}(i^*) >$

$(1 + e_j(t^*)/s_j)(x_j(t^*) + \theta_j(t^*)/s_j)S_j$ . By definition (using the convention  $t = t(i) = i/s$ )

$$\begin{aligned} X_{j,A}(i^*) - X_{j,A}(0) &= \sum_{i=0}^{i^*-1} x'(t)S_j/s \\ &= \sum_{i=0}^{i^*-1} (Y_{j,A}^+(i) - y_j^+(t)S_j/s - Y_{j,A}^-(i) + y_j^-(t)S_j/s) \\ &= Z_{j,A}^{+-}(i^*) - Z_{j,A}^{-+}(i^*) + 2 \sum_{i=0}^{i^*-1} h_j(t)/4s_j \cdot S_j/s. \end{aligned}$$

Applying (2) and the assumed lower bound for  $X_{j,A}(i^*)$  gives

$$\begin{aligned} Z_{j,A}^{+-}(i^*) - Z_{j,A}^{-+}(i^*) + 2 \sum_{i=0}^{i^*-1} h_j(t)/4s_j \cdot S_j/s \\ > (e_j(t^*)x_j(t^*) + \theta_j(t^*) + \theta_j(t^*)e_j(t^*)/s_j - 3Cs_j/s)S_j/s_j. \end{aligned}$$

Then (3),  $n^{2\epsilon} < s_j < n^{-\epsilon}s$  and  $\theta_j(t^*) + e_j(t^*)x_j(t^*)/2 - \gamma_j(t^*)/2 > c$  give

$$\begin{aligned} Z_{j,A}^{+-}(i^*) - Z_{j,A}^{-+}(i^*) \\ > (e_j(t^*)x_j(t^*)/2 - \gamma_j(t^*)/2 + \theta_j(t^*) - (n^\epsilon + 3Cs_j/s)S_j/s_j) \\ > cS_j/2s_j. \end{aligned}$$

We deduce that  $Z_{j,A}^{+-}(i^*) > cS_j/4s_j$  or  $Z_{j,A}^{-+}(i^*) < -cS_j/4s_j$ . Since  $Z_{j,A}^{\pm\pm}$  does not change once  $\mathcal{G}_{i^*}$  fails, it follows that  $Z_{j,A}^{+-}(m) > cS_j/4s_j$  or  $Z_{j,A}^{-+}(m) < -cS_j/4s_j$ . Now we apply Lemmas 7.1 and 7.2 with  $a = cS_j/4s_j$ , which is valid using our assumptions  $s \geq 40Cs_j^2k_jn^\epsilon$ ,  $s_j > n^{2\epsilon}$  and  $m > s$  which give  $\eta_j < N_j/10$  and  $a < \eta_jm/10$ . We deduce that these events have probability at most

$$\exp(-(cS_j/4s_j)^2/3\eta_jm(N_j + \eta_j)) < \exp(-5k_j \log n) \ll \left( \binom{[n]}{k_j} \right)^{-1} n^{-3k_j},$$

say. A similar bound holds for the probability that  $X_{j,A}(i^*) < (1 - e_j(t^*)/s_j)(x_j(t^*) - c/s_j)S_j$ , when we have  $Z_{j,A}^{-+}(i^*) > cS_j/4s_j$  or  $Z_{j,A}^{+-}(i^*) < -cS_j/4s_j$ . Taking a union bound over  $1 \leq j \leq r$  and  $A \in \binom{[n]}{k_j}$  completes the proof.  $\square$

### 8 Trackable variables

To apply Lemma 7.3 to the extension variables  $X_{\phi,J,\Gamma}(i)$ , we need to estimate the expected and maximum number of extensions that may be created or destroyed in each step of the process. In this section we establish a bound on the maximum number of extensions created or destroyed; in other words, we verify the boundedness hypothesis. Also, in anticipation of the expected change calculations needed for the trend hypothesis, we show that two types of pathological subgraph configurations that could potentially spoil these calculations are suitably rare. More specifically, we show that, on the event  $\mathcal{G}_i$ , there are very few extensions in  $\Xi_{\phi,J,\Gamma}$  that contain a pair of open pairs  $e, f$  such that the inclusion of one as an edge causes the other to become closed, and very few extensions in  $\Xi_{\phi,J,\Gamma}$  for which there are two edges in  $\phi(E_\Gamma \setminus E_J)$  that can both be closed by the addition of the same edge  $e_{i+1}$ . We stress that we obtain these bounds whenever the variable is *trackable* (as defined in Sect. 1.2). In particular, this condition holds for the extension variables that track the open routes to  $H$  less an edge, the central variables in the proof of Theorem 1.4.

We begin with a technical lemma that amounts to showing that if  $X_{\phi,J,\Gamma}$  is trackable then there are no ‘implicitly’ closed edges in  $E_\Gamma \setminus E_J$ .

**Lemma 8.1** *If  $X_{\phi,J,\Gamma}(i)$  is a trackable variable and  $uv \in E_\Gamma \setminus E_J$  then there does not exist  $C \subseteq V_H$  with an injective embedding  $\psi : C \rightarrow V_\Gamma$  such that*

1.  $\psi(H[C])$  is a subgraph of the graph  $\Gamma' = \Gamma \cup (\phi^{-1}(E(i)) \cap \binom{A}{2})$  obtained from  $\Gamma$  by adding the edges  $ab$  for all  $a, b \in A$  with  $\phi(a)\phi(b) \in E(i)$ ,
2. for any vertex  $c \in C$  with  $\psi(c) \notin A$ , every neighbour of  $c$  in  $H$  belongs to  $C$ , and
3. there is some edge  $e$  in  $H[C]$  with  $\psi(e) = uv$ .

*Proof* Assume for a contradiction that  $\psi$  is an embedding satisfying conditions (1–3) of the lemma. Define  $A' = \{c \in C : \psi(c) \in A\}$ . We claim that  $|A'| \geq 2$ . This is clear if  $H$  contains an edge  $e$  with  $\psi(e) \subseteq A$ . Otherwise, condition (1) implies that  $C \neq V_H$ , as  $H$  is not a subgraph of  $\Gamma$  by definition of trackability. Then, since condition (3) implies  $\psi(C) \not\subseteq A$ , condition (2) implies that  $A'$  disconnects  $H$ , and since  $H$  is 2-connected we deduce that  $|A'| \geq 2$ .

Let  $K$  be the graph obtained from  $H[C]$  by deleting all edges inside  $A'$ . Then  $K$  is isomorphic to a subgraph of  $\Gamma$  by condition (1), so  $S_{A,\Gamma[A \cup \psi(C)]} \leq S_{A',K}$ . Also,  $S_{A',K}$  is equal to  $S_{(V_H \setminus C) \cup A',H}$  by condition (2). This in turn is at most 1, as  $H$  is strictly balanced. We deduce that  $S_{A,\Gamma[A \cup \psi(C)]} \leq 1$ .

Now  $(A, \Gamma)$  cannot be strictly dense, since  $\psi(C) \not\subseteq A$ , so it remains to consider possibility (b) in the definition of trackability. In this case we must have  $S_{A,\Gamma[A \cup \psi(C)]} = 1$ , and so  $S_{(V_H \setminus C) \cup A',H} = 1$ . Then the fact that  $H$  is strictly balanced implies that  $C = V_H$ ,  $|A'| = 2$  and  $A' \in E_H$ . (We cannot

have  $(V_H \setminus C) \cup A' = V_H$  as  $A' \subsetneq C$ .) However, the existence of such an embedding of  $H$  in  $\Gamma'$  is specifically ruled out by the definition of trackability, so we have the required contradiction.  $\square$

Now we are ready to verify the boundedness hypothesis. Following the notation of Lemma 7.3 we write  $X_{\phi,J,\Gamma}(i + 1) - X_{\phi,J,\Gamma}(i) = Y_{\phi,J,\Gamma}^+(i) - Y_{\phi,J,\Gamma}^-(i)$ , where  $Y_{\phi,J,\Gamma}^+(i) \geq 0$  is the number of maps  $f$  in  $\Xi_{\phi,J,\Gamma}(i + 1) \setminus \Xi_{\phi,J,\Gamma}(i)$  and  $Y_{\phi,J,\Gamma}^-(i) \geq 0$  is the number of maps  $f$  in  $\Xi_{\phi,J,\Gamma}(i) \setminus \Xi_{\phi,J,\Gamma}(i + 1)$ . Recall that  $f : V_\Gamma \rightarrow [n]$  is counted by  $X_{\phi,J,\Gamma}(i)$  if  $f(e) \in O(i)$  for every  $e \in E_\Gamma \setminus E_J$ ,  $f(e) \in E(i)$  for every  $e \in E_J$ , and  $f$  restricts to  $\phi$  on  $A$ . Then  $f$  will be counted by  $Y_{\phi,J,\Gamma}^-(i)$  if there is at least one  $e \in E_\Gamma \setminus E_J$  such that  $f(e)$  either becomes closed at step  $i + 1$  or is the edge  $e_{i+1}$  chosen by the process at step  $i + 1$ . Also, for each edge  $e$  of  $J$  and  $f$  counted by  $X_{\phi,J \setminus e,\Gamma}(i)$ ,  $f$  might be counted by  $Y_{\phi,J,\Gamma}^+(i)$  if  $e_{i+1} = f(e)$ . (We will see below that  $f$  may not actually be counted, but for the purpose of an upper bound we do not need to take this into account here.)

**Lemma 8.2** (Boundedness hypothesis) *With high probability, for every  $1 \leq i \leq m$ , assuming  $\mathcal{G}_i$  and that  $X_{\phi,J,\Gamma}(i)$  is trackable, we have  $Y_{\phi,J,\Gamma}^+(i) \leq n^{-1/e_H} S_{A,J}$  and  $Y_{\phi,J,\Gamma}^-(i) \leq n^{-1/e_H} S_{A,J}$ .*

*Proof* We start with the variable  $Y_{\phi,J,\Gamma}^+(i)$ . Fix an edge  $e = ab$  of  $J$  and suppose the process chooses the edge  $e_{i+1} = uv$  in step  $i + 1$ . Let  $A' = A \cup \{a, b\}$ ,  $J' = J \setminus E_{J[A']}$  and define  $\phi' : A' \rightarrow [n]$  agreeing with  $\phi$  on  $A$  and satisfying  $\phi'(a) = u$ ,  $\phi'(b) = v$ . Note that one of  $a$  or  $b$  may belong to  $A$ , but not both, as  $A$  is independent in  $J$ . Any  $f$  counted by  $Y_{\phi,J,\Gamma}^+(i)$  with  $f(a) = u$  and  $f(b) = v$  is counted by  $X_{\phi',J',\Gamma}(i)$ ; we can bound this by  $N_{\phi',J'}$ , which by Lemma 5.2 satisfies  $N_{\phi',J'} < n^{4e_{J'}} \max_{A' \subseteq B \subseteq V_{J'}} S_{B,J'}$ . Since  $A \subsetneq A'$  and  $(A, J)$  is strictly dense we have  $\max_{A' \subseteq B \subseteq V_{J'}} S_{B,J'} \leq n^{-1/(e_H - 1)} S_{A,J}$ . Summing over all edges  $e$  of  $J$  we estimate  $Y_{\phi,J,\Gamma}^+(i) < n^{-1/e_H} S_{A,J}$ .

Now consider the variable  $Y_{\phi,J,\Gamma}^-(i)$ . Suppose the process chooses the edge  $e_{i+1} = uv$  in step  $i + 1$ . Fix an edge  $e$  of  $\Gamma \setminus J$ . We want to estimate the number of embeddings  $f$  in  $\Xi_{\phi,J,\Gamma}(i)$  for which  $f(e)$  is either equal to  $e_{i+1}$  or becomes closed in step  $i + 1$ . Since  $(A, J)$  is strictly dense, Lemma 5.3 gives an upper bound of  $n^{-1/(e_H - 1) + 4(e_J + 1)\epsilon} S_{A,J}$  on the number of embeddings  $f$  with  $f(e) = e_{i+1}$ .

Next consider an embedding  $f$  where  $f(e) = xy$  becomes closed in step  $i + 1$ . Then there is an embedding  $f_2$  of  $H$  in  $G(i) \cup \{uv, xy\}$ . Write  $C' = f(V_J) \cap f_2(V_H)$  and identify the sets  $f^{-1}(C')$  and  $f_2^{-1}(C')$  as a set  $C$  on which  $f$  and  $f_2$  agree. Then we have  $f_2(a) = u$ ,  $f_2(b) = v$  for some  $a, b \in V_H$ , and we have some  $c, d \in C$  with  $f(c) = f_2(c) = x$ ,

$f(d) = f_2(d) = y$ , where  $\{c, d\} \neq \{a, b\}$  and  $\{c, d\} \not\subseteq A$  (since  $A$  is independent in  $\Gamma$ ). Write  $H' = H \setminus \{ab, cd\}$  and let  $W$  be the join of  $J_1 = J$  and  $J_2 = H'$  formed by identifying vertices in  $C$  and removing any edges within  $A' = A \cup \{a, b\}$ .

We claim that  $S_{B,W} \leq n^{-1/(e_H-1)} S_{A,J}$  for all  $A' \subseteq B \subseteq V_W$ . Fix such a set  $B$  and write  $B_1 = B \cap V_{J_1}$  and  $B_2 = B \cap V_{J_2}$ . We have

$$S_{B,W} = S_{B_1,J} \cdot S_{B_2 \cup C, H} \cdot p^\beta$$

where  $\beta$  is the number of edges in  $J_2$  joining  $B_2 \setminus C$  and  $C \setminus B_2$ . Since  $(A, J)$  is strictly dense we have  $S_{B_1,J} \leq S_{A,J}$ , with equality only if  $B_1 = A$ . Furthermore, since  $\{a, b\} \cup C$  has at least 3 vertices, we have  $S_{C \cup B_2, H} \leq 1$ , with equality only if  $C \cup B_2 = V_H$ . Thus we can restrict our attention to the situation where  $B_1 = A$ ,  $B_2 \supset V_{J_2} \setminus V_{J_1}$  and  $\beta = 0$ . In this case we will use Lemma 8.1 to obtain a contradiction. We view  $C$  as a subset of  $V_H$  and let  $\psi$  be the identification of  $C$  with the subset of  $V_\Gamma$  which is also called  $C$ . We can assume that condition (1) is satisfied, as otherwise  $f$  is an extension of  $\phi$  to an embedding of a supergraph of  $J$  and then we have the required estimate on  $S_{B,W}$  by Lemma 5.3. Also,  $\beta = 0$  gives condition (2), and  $f_2(cd) = xy = f(e)$  with  $e \in E_\Gamma \setminus E_J$  and  $c, d \in C$ , which gives condition (3). Thus Lemma 8.1 shows that this case does not actually arise. We deduce that  $S_{B,W} \leq n^{-1/(e_H-1)} S_{A,J}$ .

Now applying Lemma 5.2 and summing over all possibilities for  $e$  and  $W$  gives the required bound  $Y_{\phi,J,\Gamma}^-(i) < n^{-1/e_H} S_{A,J}$ . □

Now we turn to two technical issues regarding the expected values of  $Y_{\phi,J,\Gamma}^+(i)$  and  $Y_{\phi,J,\Gamma}^-(i)$ . We would like to approximate these using our estimates for extension variables. In the case of  $Y_{\phi,J,\Gamma}^+(i)$ , our first approximation is that for each edge  $e$  of  $J$ , an embedding  $f$  counted by  $X_{\phi,J \setminus e,\Gamma}(i)$  should be counted by  $Y_{\phi,J,\Gamma}^+(i)$  if  $e_{i+1} = f(e)$ . However, we need to account for the possibility that the addition of the edge  $e_{i+1} = f(e)$  closes some edge  $f(e')$  where  $e' \in E_\Gamma \setminus E_J$ . In the case of  $Y_{\phi,J,\Gamma}^-$ , we sum  $|C_{f(uv)}(i)|$  over  $uv \in E_\Gamma \setminus E_J$  to estimate the number of open edges  $xy$  such that choosing  $e_{i+1} = xy$  causes a given embedding  $f$  in  $\Xi_{\phi,J,\Gamma}$  to leave this set. However, we need to account for the possibility that there could be edges  $uv, u'v' \in E_\Gamma \setminus E_J$  such that  $C_{f(uv)}(i)$  and  $C_{f(u'v')}(i)$  have large intersection. We now establish two lemmas showing that these two ‘pathological’ possibilities have a negligible impact.

**Lemma 8.3** (Creation fidelity) *If  $X_{\phi,J,\Gamma}$  is a trackable variable then, with high probability on the event  $\mathcal{G}_i$ , the number of extensions  $f \in \Xi_{\phi,J,\Gamma}$  with the property that there are distinct  $uv, xy \in E_\Gamma \setminus E_J$  such that  $G(i) \cup \{f(uv), f(xy)\}$  contains a copy of  $H$  is at most  $n^{-1/e_H} S_{A,J}$ .*

*Proof* Let  $uv, xy \in E_\Gamma \setminus E_J$  be distinct and fixed. Consider any graph  $W$  given by the join  $J$  and a copy of  $H$  less two edges, where  $uv$  and  $xy$  are identified with these missing edges. As in Lemma 8.2 it suffices to show that  $S_{B,W} \leq n^{-1/(e_H-1)} S_{A,J}$  for all  $A \subseteq B \subseteq V_W$ . The argument is almost identical to that in Lemma 8.2. With the same notation we again have  $S_{B,W} = S_{B_1,J} \cdot S_{B_2 \cup C, H} \cdot p^\beta$ . We again have  $S_{B_1,J} \leq S_{A,J}$ , with equality only if  $B_1 = A$ . Furthermore, in the current lemma we have  $u, v, x, y \in C$ , so  $|C| \geq 3$ , and  $S_{C \cup B_2, H} \leq 1$ , with equality only if  $C \cup B_2 = V_H$ . Then Lemma 8.1 applies as before to complete the proof.  $\square$

**Lemma 8.4** (Destruction fidelity) *If  $uv, u'v' \in O(i)$  are distinct then, on  $\mathcal{G}_i$ , we have  $|C_{uv}(i) \cap C_{u'v'}(i)| \leq n^{-1/e_H} p^{-1}$  with high probability.*

*Proof* Let  $ab$  and  $cd$  be distinct edges of  $H$  and set  $H_1 = H \setminus \{ab, cd\}$ . Similarly, let  $a'b'$  and  $c'd'$  be distinct edges of  $H$  and set  $H_2 = H \setminus \{a'b', c'd'\}$ . Now let  $W$  be any join of  $H_1$  and  $H_2$  where  $c = c'$  and  $d = d'$  but  $ab \neq a'b'$ . Set  $A = \{a, b\} \cup \{a', b'\}$ . Then  $|A| \geq 3$ . Appealing to Lemma 5.2, it suffices to show  $S_{B,W} < p^{-1}$  for all  $A \subseteq B \subseteq V_W$ . Fix such a set  $B$ . Similarly to before we have  $S_{B,W} \leq S_{B_1, H_1} S_{C \cup B_2, H_2} p^{\beta_2}$ , where  $B_1 = B \cap V_{H_1}$ ,  $B_2 = B \cap V_{H_2}$ , and  $C = V_{H_1} \cap V_{H_2}$  and  $\beta_2$  is the number of edges in  $H_2$  joining  $B_2 \setminus C$  and  $C \setminus B_2$ .

Note that  $c, d \in C$ , so  $S_{C \cup B_2, H_2} = S_{C \cup B_2, H} \leq 1$ , with equality only when  $C \cup B_2 = V_H$ . Also, since  $H_1$  is strictly dense we have  $S_{B_1, H_1} \leq 1/p$ , with equality only when  $B_1 = \{a, b\}$ . Thus we obtain the desired inequality  $S_{B,W} < p^{-1}$ , except possibly in the case when  $C \cup B_2 = V_H$ ,  $B_1 = \{a, b\}$  and  $\beta_2 = 0$ . Also, the same argument reversing the roles of  $H_1$  and  $H_2$  shows that we obtain the desired inequality, except possibly in the case when  $C \cup B_1 = V_H$ ,  $B_2 = \{a', b'\}$  and  $\beta_1 = 0$ , where  $\beta_1$  is the number of edges in  $H_1$  joining  $B_1 \setminus C$  and  $C \setminus B_1$ . Since  $H$  is 2-connected, the only remaining possibility is when  $V_{H_1} = V_{H_2}$ . But then  $S_{B,W} \leq S_{A, H_1} < 1/p$ , as  $H_1$  is strictly dense and  $|A| \geq 3$ . Thus in all cases we have the desired inequality.  $\square$

### 9 Trajectory verification and Turán bounds

Now we use the above bounds and Lemma 7.3 to prove Theorem 1.4, which shows that trackable extension variables are well described by the differential equations given earlier in the paper. It will then follow that the process does indeed continue until at least time  $t = t_{\max} = \mu(\log n)^{1/(e_H-1)}$ , i.e.  $m = \mu(\log n)^{1/(e_H-1)} pn^2$  edges. In particular, it will follow that variables counting common neighbours of  $d$ -sets with  $p^d n > 1$  and variables counting extensions from non-edge pairs to subgraphs of  $H$  with at most  $e_H - 2$  edges satisfy these equations. Then Corollary 1.5 is an immediate consequence of

the formulae for common neighbours. In particular, when  $d = 1$  we deduce the minimum degree statement needed to prove Theorem 1.1. To prove Theorem 1.4 we will show that the good event  $\mathcal{G}_m$  holds with high probability, i.e. for every  $i \leq m$  and trackable extension variable  $X_{\phi,J,\Gamma}(i)$  corresponding to a triple in  $\mathcal{T}$ , we have

$$X_{\phi,J,\Gamma}(i) = (1 \pm e(t)/s_e)(x_{A,J,\Gamma}(t) \pm \theta(t)/s_e)S_{A,J},$$

where  $x_{A,J,\Gamma}(t) = q(t)^{e_{\Gamma}-e_J} (2t)^{e_J}$  and  $t, s_e, S_{A,J}, q(t), e(t), \theta(t)$  are as defined in Sect. 1.2.

*Proof of Theorem 1.4* To apply Lemma 7.3 we arbitrarily number the triples in  $\mathcal{T}$  by  $1 \leq j \leq r$  and identify the extension variables  $X_{\phi,J,\Gamma}(i)$  with the variables  $X_{j,A}(i)$  appearing in the statement of the lemma. We take  $e_j(t) = e(t)$  and  $\theta_j(t) = \theta(t)$  and  $s_j = s_e$  for all  $1 \leq j \leq r$ . The event  $\mathcal{H}_i$  is the event that the estimates given in Lemmas 6.1, 8.2, 8.3 and 8.4 hold up to step  $i$ . We will give values for the other parameters of the lemma later in this proof.

We start with the main step, which is checking the trend hypothesis. For the expected one-step changes  $\mathbb{E}[Y_{\phi,J,\Gamma}^{\pm}(i) | \mathcal{G}_i \wedge \mathcal{H}_i]$  we analyse the error terms in our earlier heuristic derivation. We start with the variable  $Q(i)$ , which counts the number of ordered pairs that are open at step  $i$ . Write  $Q(i+1) - Q(i) = Q^+(i) - Q^-(i)$  with  $Q^+(i), Q^-(i) \geq 0$ . Since  $Q(i+1) = Q(i) - 1 - |C_{e_{i+1}}(i)|$  we have  $Q^+(i) = 0$  and  $Q^-(i) = 1 + |C_{e_{i+1}}(i)|$ . Then Corollary 6.2 gives

$$Q^-(i) = 1 + (1 \pm e(t)/s_e)((2t)^{e_H-2}q(t) \pm 2\theta(t)/s_e)a_H p^{-1}.$$

We have  $q'(t) = y_q^+(t) - y_q^-(t)$ , where  $y_q^+(t) = 0$  for all  $t$  and  $y_q^-(t) = c(t) = a_H(2t)^{e_H-2}q(t)$ . Recall that  $h_q(t) = (eq + \gamma)'(t)$ . Now  $e'(t) = P'(t)e^{P(t)} > W(t^{e_H-2} + 1)e^{P(t)}$ , so since  $W \gg V \gg e_H$  we have  $h_q(t)/y_q^-(t) > (V + Wa_H^{-1}(2t)^{-(e_H-2)})e^{P(t)}$  for  $t > 0$ . Since  $s = pn^2$  and  $\theta(t) < 1$  we easily have the required condition for  $Q^-(i)$ , namely

$$Q^-(i) = (y_q^-(t) \pm h_q(t)/4s_e)n^2/s.$$

(We only need this estimate for  $\mathbb{E}(Q^-(i) | \mathcal{G}_i \wedge \mathcal{H}_i)$ , but actually it always holds on the event  $\mathcal{G}_i$ .)

Now we check the trend hypothesis in the general case. We write  $X_{\phi,J,\Gamma}(i+1) - X_{\phi,J,\Gamma}(i) = Y_{\phi,J,\Gamma}^+(i) - Y_{\phi,J,\Gamma}^-(i)$ . The term  $Y_{\phi,J,\Gamma}^+(i)$  has contributions corresponding to each edge  $e$  of  $J$ . A function  $f$  in  $\Xi_{\phi,J \setminus e,\Gamma}(i)$  will be counted by  $Y_{\phi,J,\Gamma}^+(i+1)$  if the process chooses the edge  $e_{i+1}$  equal to  $f(e)$  and this choice of  $e_{i+1}$  does not close any edge in  $f(E_{\Gamma} \setminus E_J)$ . Now



$e_{i+1}$  is chosen uniformly at random among  $Q(i)/2$  open edges, so appealing to Lemma 8.3 we can estimate

$$\mathbb{E}(Y_{\phi, J, \Gamma}^+(i) | \mathcal{G}_i \wedge \mathcal{H}_i) = 2Q(i)^{-1} \sum_{e \in J} (X_{\phi, J \setminus e, \Gamma}(i) \pm n^{-1/e_H} S_{A, J \setminus e}).$$

Now  $X_{\phi, J \setminus e, \Gamma}(i) = (1 \pm e(t)/s_e)(x_{A, J \setminus e, \Gamma}(t) \pm \theta(t)/s_e)S_{A, J \setminus e}$ . Since  $S_{A, J \setminus e} = p^{-1}S_{A, J}$ ,  $n^{-1/e_H} \ll 1/s_e$  and  $\theta(t) \geq 1/2$  for  $t \geq 0$  we estimate  $\mathbb{E}(Y_{\phi, J, \Gamma}^+(i) | \mathcal{G}_i \wedge \mathcal{H}_i)$  as

$$2((1 \pm e(t)/s_e)(q(t) \pm \theta(t)/s_e)n^2)^{-1} \times e_J \cdot (1 \pm e(t)/s_e)(q(t)^{e_{\Gamma} - e_J + 1}(2t)^{e_J - 1} \pm 2\theta(t)/s_e)p^{-1}S_{A, J}.$$

We have  $x'_{A, J, \Gamma}(t) = y_{A, J, \Gamma}^+(t) - y_{A, J, \Gamma}^-(t)$ , where  $y_{A, J, \Gamma}^+(t) = 2e_J q(t)^{e_{\Gamma} - e_J} \times (2t)^{e_J - 1}$  and  $y_{A, J, \Gamma}^-(t) = a_H(e_{\Gamma} - e_J)q(t)^{e_{\Gamma} - e_J}(2t)^{e_J + e_H - 2}$ . We also have  $h_{A, J, \Gamma}(t) = (ex_{A, J, \Gamma} + \gamma)'(t)$ . To establish the required bound, i.e.

$$\mathbb{E}(Y_{\phi, J, \Gamma}^+(i) | \mathcal{G}_i \wedge \mathcal{H}_i) = (y_{A, J, \Gamma}^+(t) \pm h_{A, J, \Gamma}(t)/4s_e)S_{A, J}/s,$$

it suffices to show that

$$(1 \pm 4e(t)/s_e)(1 \pm 2\theta(t)q(t)^{-1}/s_e)(1 \pm 2\theta(t)(q(t)^{e_{\Gamma} - e_J + 1}(2t)^{e_J - 1})^{-1}/s_e) \subseteq 1 \pm (2e_J q(t)^{e_{\Gamma} - e_J}(2t)^{e_J - 1})^{-1}h_{A, J, \Gamma}(t)/4s_e. \tag{4}$$

Setting  $x(t) = x_{A, J, \Gamma}(t) = (2t)^{e_J} q(t)^{e_{\Gamma} - e_J}$  we see that it is necessary to establish that

$$\frac{4e_J e(t)x(t)}{t} + \frac{2e_J \theta(t)x(t)}{tq(t)} + \frac{4e_J \theta(t)}{q(t)} \tag{5}$$

is bounded above by

$$\begin{aligned} & \frac{1}{4} (x(t)e'(t) + x'(t)e(t) + \gamma'(t)) \\ & > \frac{e^{P(t)}x(t)}{4} \left( Wt^{e_H - 2} + W + (e_J/t - (e_{\Gamma} - e_J)a_H(2t)^{e_H - 2}) \frac{e^{P(t)} - 1}{e^{P(t)}} \right) \\ & \quad + \frac{\gamma'(t)}{4} \\ & > \frac{e^{P(t)}x(t)}{4} \left( \frac{W}{2}t^{e_H - 2} + W \right) + \frac{\gamma'(t)}{4}. \end{aligned}$$

Note that establishing this bound is in fact sufficient. This follows from the inequality  $|h_{A, J, \Gamma}(t)| < n^\epsilon$  (explained later in the proof), which implies that

the terms omitted in (4) are bounded above by  $n^\epsilon / (x(t)s_e^2) + n^\epsilon / s_e^2$ , and so do not cause the inequality to be violated when  $n$  is sufficiently large. Note also that we can assume that  $e_J > 0$ , as otherwise  $Y_{\phi, J, \Gamma}^+ = y_{A, J, \Gamma}^+ = 0$ . To verify the bound for  $t < 40V/W$  we note that  $x(t) \leq 9t/4$ , as  $e_J > 0$ , and therefore (5) is at most  $9Ve(t) + 15V < 10Ve^{40V} = \gamma'(t)/4$ . On the other hand, for  $t > 40V/W$  we note that the first two terms in (5) can each be bounded by  $We^{P(t)}x(t)/10$ ; the remaining term is bounded by  $\gamma'(t)/4 > 5V$  for  $40V/W < t < 1/(50V)$  and by  $e^{P(t)}x(t)$  for larger  $t$ .

Next consider the term  $Y_{\phi, J, \Gamma}^-(i)$ , which has contributions corresponding to each edge  $e$  of  $\Gamma \setminus J$ . A function  $f$  in  $\Xi_{\phi, J, \Gamma}(i)$  will be counted by  $Y_{\phi, J, \Gamma}^-(i + 1)$  if the process either chooses the edge  $e_{i+1}$  equal to  $f(e)$  or  $f(e)$  becomes closed, i.e.  $f(e) \in C(i + 1)$ . Thinking of  $e_{i+1}$  as an ordered pair, the number of choices is  $2 + |C_{f(e)}(i)|$ , each occurring with probability  $Q(i)^{-1}$ . Therefore, appealing to Lemma 8.4, we have

$$\begin{aligned} &\mathbb{E}(Y_{\phi, J, \Gamma}^-(i) | \mathcal{G}_i \wedge \mathcal{H}_i) \\ &= Q(i)^{-1} \sum_{f \in \Xi_{\phi, J, \Gamma}(i)} \sum_{e \in \Gamma \setminus J} (2 + |C_{f(e)}(i)| \pm n^{-1/e_H} p^{-1}). \end{aligned}$$

We can estimate  $|C_{f(e)}(i)|$  by Corollary 6.2, so we estimate  $\mathbb{E}(Y_{\phi, J, \Gamma}^-(i) | \mathcal{G}_i \wedge \mathcal{H}_i)$  as

$$\begin{aligned} &((1 \pm e(t)/s_e)(q(t) \pm \theta(t)/s_e)n^2)^{-1} \\ &\quad \times (e_\Gamma - e_J) \cdot (1 \pm e(t)/s_e)(q(t)^{e_\Gamma - e_J} (2t)^{e_J} \pm \theta(t)/s_e)S_{A, J} \\ &\quad \times (1 \pm e(t)/s_e)((2t)^{e_H - 2}q(t) \pm 2\theta(t)/s_e \pm n^{-1/e_H})a_H p^{-1}. \end{aligned}$$

Now to establish the required bound, i.e.

$$\mathbb{E}(Y_{\phi, J, \Gamma}^-(i) | \mathcal{G}_i \wedge \mathcal{H}_i) = (y_{A, J, \Gamma}^-(t) \pm h_{A, J, \Gamma}(t)/4s_e)S_{A, J}/s,$$

it suffices to show that

$$\begin{aligned} &(1 \pm 4e(t)/s_e)(1 \pm 2\theta(t)q(t)^{-1}/s_e)(1 \pm \theta(t)(q(t)^{e_\Gamma - e_J} (2t)^{e_J})^{-1}/s_e) \\ &\quad \times (1 \pm 2\theta(t)((2t)^{e_H - 2}q(t))^{-1}/s_e) \\ &\subseteq 1 \pm (a_H(e_\Gamma - e_J)q(t)^{e_\Gamma - e_J} (2t)^{e_J + e_H - 2})^{-1}h_{A, J, \Gamma}(t)/4s_e. \end{aligned}$$

And this reduces to showing that

$$\begin{aligned} &4a_H(e_\Gamma - e_J)(2t)^{e_H - 2}x(t)e(t) + \frac{2a_H(e_\Gamma - e_J)(2t)^{e_H - 2}x(t)\theta(t)}{q(t)} \\ &+ a_H(e_\Gamma - e_J)(2t)^{e_H - 2}\theta(t) + \frac{2a_H(e_\Gamma - e_J)x(t)\theta(t)}{q(t)} \end{aligned}$$

is bounded above by

$$\frac{e^{P(t)}x(t)}{4} \left( \frac{W}{2}t^{e_H-2} + W \right) + \frac{\gamma'(t)}{4}.$$

This follows by estimates very similar to those given above for  $Y_{\phi,J,\Gamma}^+(i)$ . We omit the details, except for remarking that is helpful to observe that the term  $a_H(e_\Gamma - e_J)(2t)^{e_H-2}\theta(t)$  is bounded by  $\gamma'(t)/4$  for  $t < 1/(50V)$ .

This verifies the trend hypothesis of Lemma 7.3. To finish the proof we check the remaining conditions. The boundedness hypothesis follows from Lemma 8.2 as we have  $n^{1/e_H} \gg n^{1/e_H-\epsilon} = s_e^2 n^\epsilon$ . We have  $|T| = r < V^{3V}$ ,  $n^{2\epsilon} < s_e^2 n^\epsilon < n < pn^2 = s < m < n^2$ ,  $s_e = n^{1/2e_H-\epsilon} > n^{2\epsilon}$  and  $m < sn^\epsilon$  for  $n$  sufficiently large. The functions  $x_{A,J,\Gamma}(t)$  and  $y_{A,J,\Gamma}^\pm(t)$  all have the form  $F(t)e^{-Kt^{e_H-1}}$ , where  $F$  is a polynomial of degree at most  $V + e_H$ , and  $K$  and all coefficients in  $F$  are non-negative and bounded above by  $W$ , say. Here we can use

$$\int_0^\infty t^a e^{-t} = a! \quad \text{and} \quad \sup_{t \geq 0} t^a e^{-t} = (a/e)^a \quad \text{for } a \in \mathbb{N}$$

to see that  $\sup_{t \geq 0} |y_{A,J,\Gamma}^\pm(t)|$ ,  $\sup_{t \geq 0} |x'_{A,J,\Gamma}(t)|$  and  $\int_0^\infty |x''_{A,J,\Gamma}(t)| dt$  are all bounded by some constant  $C$  depending only on  $W$ . Also, recall that  $e(t) = e^{P(t)} - 1$  with  $P(t) = W(t^{e_H-1} + t)$ ,  $h_{A,J,\Gamma}(t) = (ex_{A,J,\Gamma} + \gamma)'(t)$ , and  $\gamma(t)$  is a smooth increasing function such that  $\gamma(t)$  and  $\gamma'(t)$  are bounded by absolute constants. The initial conditions  $e(0) = \gamma(0) = 0$  hold. Since  $t < t^* = \mu(\log n)^{1/(e_H-1)}$ , by choosing  $\mu$  sufficiently small we can ensure that  $\sup_{0 \leq t \leq m/s} |h_{A,J,\Gamma}(t)| < n^\epsilon$  and  $\int_0^{m/s} |h'_{A,J,\Gamma}(t)| dt < n^\epsilon$ . Finally, we can choose  $c = 1/2$ , since  $\theta(t) = 1/2 + \gamma(t)$ , so  $\theta(t) + e(t)x(t)/2 - \gamma(t)/2 > 1/2$ . □

### 10 Counting small subgraphs

In this short section we apply our results to count small subgraphs in the  $H$ -free process and compare these counts to those known for the  $G(n, p)$  model. A rough summary is that the  $H$ -free process looks very much like  $G(n, p)$  from this perspective, except that it does not contain any graphs that contain  $H$ . A more precise description is given by Theorem 1.6, which we now prove.

*Proof of Theorem 1.6* Statement (i) follows from Lemma 4.1, as  $\Gamma[B]$  does not appear in  $G(i)$  with high probability, and therefore  $\Gamma$  itself does not appear with high probability (note that the failure probability here decays

polynomially in  $n$ , not exponentially). Statement (ii) follows from Theorem 1.4 applied to the trackable variable  $X_\Gamma(i) = X_{\emptyset, \Gamma}(i)$ ; for any fixed  $\Gamma$  we can reapply the theorem with a new parameter  $V' > v_\Gamma, e_\Gamma$ , and track the variable until step  $m' = \xi m$  for some  $\xi = \xi(\Gamma) > 0$ . It remains to consider the case when  $S_{\Gamma[B]} \geq 1$  for all  $B \subseteq V_\Gamma$ . For convenience we assume that  $V_\Gamma, e_\Gamma < V$ , although this assumption can be eliminated by restricting time to  $t \leq \xi m$  for some  $\xi = \xi(\Gamma) > 0$ , as in (ii). Form the extension series  $\emptyset = B_0 \subsetneq B_1 \subsetneq \cdots \subsetneq B_d = V_\Gamma$ , as defined in Sect. 3. We divide the  $m$  steps of the process into  $d$  equal intervals, and in the  $j$ th interval we show that with high probability there is an extension from a *fixed* copy of  $\Gamma[B_{j-1}]$  (found in the previous interval) to a copy of  $\Gamma[B_j]$ . By construction every step of the extension series is strictly balanced, and our assumption in this case implies that the scalings in each step satisfy  $S_{B_{j-1}, \Gamma[B_j]} \geq 1$ . Suppose that  $\phi : B_{j-1} \rightarrow [n]$  is an embedding of  $\Gamma[B_{j-1}]$  in  $G((j-1)m/d)$ . If  $S_{B_{j-1}, \Gamma[B_j]} > 1$  then the variable  $X_{\phi, \Gamma[B_j], \Gamma[B_j]}(i)$  is trackable, so the required extension exists by Theorem 1.4 (in fact there are many such extensions). On the other hand, if  $S_{B_{j-1}, \Gamma[B_j]} = 1$  we can apply Theorem 1.4 to the trackable variables  $X_{\phi, \Gamma[B_j] \setminus e, \Gamma[B_j]}(i)$  with  $e \in E_{\Gamma[B_j]} \setminus E_{\Gamma[B_{j-1}]}$ . Writing  $a_j = e_{\Gamma[B_{j+1}]} - e_{\Gamma[B_j]}$  we can estimate the probability that in step  $i$  the edge  $e_i$  completes some embedding of  $\Gamma[B_j] \setminus e$  for some  $e$  to an embedding of  $\Gamma[B_j]$  by  $Q(i)^{-1} \sum_e X_{\phi, \Gamma[B_j] \setminus e, \Gamma[B_j]}(i) \sim a_j (2t)^{a_j-1} / (pn^2)$ . Since the length of each interval is  $m/d \gg s = pn^2$  and  $t \gg 1$  (ignoring the first half of the first interval, say) we see that the required extension appears with high probability.  $\square$

*Remark* Our results for counting labelled copies of  $\Gamma$  in the  $H$ -free process mirror those obtained for the analogous counts in  $G(n, p)$ . However, rather more is known in the  $G(n, p)$  model, some of which is surveyed in Sect. VII of [27]. In the supercritical case Barbour, Karoński and Ruciński [6] gave a central limit theorem with estimates on the rate of convergence for the appropriately normalised count. Spencer [31] analysed the critical case: one of his results concerns the case when  $\Gamma$  is strictly balanced, when he obtains the asymptotic probability for  $\Gamma$  to appear when  $p$  is near the threshold. It seems plausible that similar results may hold for the  $H$ -free process: in the supercritical case one would need to extract distributional information from the differential equations method (along the lines of [29]), and in the critical case one would need a more accurate analysis of the above proof (which seems to suggest a Poisson approximation). For the sake of brevity we do not pursue these possibilities here.

### 11 Smooth independence

We have now shown that the  $H$ -free process continues until at least the time  $t_{\max} = \mu(\log n)^{1/(e_H-1)}$ , when it has  $m = \mu(\log n)^{1/(e_H-1)}pn^2$  edges. In this section we describe an additional assumption (‘smooth independence’) on  $H$ , under which we show that the independence number of the resulting graph is at most

$$\alpha = 3\mu^{-1}(\log n)^{1-1/(e_H-1)}p^{-1}.$$

Since the independence number cannot increase when more edges are added, we also have the same upper bound for the terminal graph of the process. The main step of our proof will be to show that, for any set  $I$  of size  $\alpha$ , with high probability we can track the number of open pairs contained within  $I$ : at time  $t$  there will be roughly  $q(t)|I|^2$  open ordered pairs in  $I$ . Then a simple union bound calculation will show that with high probability  $I$  is not independent at time  $t_{\max}$ .

To track the open pairs within a set  $I$  we use Lemma 7.3, but we cannot simply apply the lemma directly, due to the possibility of closing a large number of pairs in  $I$  in a single step of the process. Note that in this application of Lemma 7.3 we will take  $r = 1$ ,  $k_1 = \alpha$  and  $S_1 = \alpha^2$ . So we will not be able to achieve the boundedness hypothesis in a useful way if we allow our process to close  $\alpha$  edges in the set  $I$  in a single step (and this certainly is a possibility for many choices of  $H$ ). To deal with this, we say that an open pair  $uv \in O(i - 1)$  is  $I$ -good at step  $i$  if

$$|C_{uv}(i - 1) \cap I^2| < n^{-5\epsilon}p^{-1},$$

that is, if the addition of the pair as an edge would close at most  $n^{-5\epsilon}p^{-1}$  ordered pairs in  $I$ , otherwise the pair  $uv$  is  $I$ -bad at step  $i$ . Then we say that a pair  $xy$  in  $I$  is  $I$ -closed at step  $i$  if there is some step  $i' \leq i$  such that  $e_{i'}$  is  $I$ -good at step  $i$  and  $G(i') \cup \{xy\}$  contains a copy of  $H$  using  $xy$  and  $e_{i'}$ . If  $xy$  in  $I$  is not in  $E(i)$  and not  $I$ -closed we say that it is  $I$ -open at step  $i$ . Note that an  $I$ -closed pair is closed, but an  $I$ -open pair could be open or closed (but not an edge). Let  $Q_I(i)$  be the number of open ordered pairs in  $I$  at step  $i$  and  $X_I(i)$  be the number of  $I$ -open ordered pairs in  $I$  at step  $i$ . We write  $B_I(i)$  for the set of pairs  $uv \in O(i)$  that are  $I$ -bad at step  $i + 1$ . Also, write  $P_I$  for the set of  $I$ -bad edges chosen through  $m$  steps of the process. Thus  $P_I = \{e_i : e_i \text{ is } I\text{-bad at step } i\}$ . We say that  $H$  has *smooth independence* if with high probability, for every set  $I$  of size  $\alpha$ ,

$$|B_I(i)| < n^{-6\epsilon}p^{-2} \text{ for all } i, \text{ and } |P_I| < n^{-5\epsilon}p^{-1}.$$

Our first step is to apply Lemma 7.3 to track the number of  $I$ -open pairs in  $I$ .

**Lemma 11.1** *If  $H$  has smooth independence, then with high probability, for any set  $I$  of size  $\alpha$ , the number of  $I$ -open ordered pairs in  $I$  at step  $i$  is  $X_I(i) = (1 \pm e(t)n^{-2\epsilon})(q(t) \pm n^{-2\epsilon})\alpha^2$ .*

*Proof* We apply Lemma 7.3 with  $r = 1, k_1 = \alpha, X_{1,I}(i) = X_I(i)$  for  $I \in \binom{[n]}{\alpha}$ ,  $x_1(t) = q(t), e_1(t) = e(t), \gamma_1(t) = \gamma(t), \theta_1(t) = \theta(t), s_1 = n^{2\epsilon}$  and  $S_1 = \alpha^2$ . We let  $\mathcal{H}_i$  be the event that the estimates given by Theorem 1.4 and  $|B_I(i)| \leq p^{-2}n^{-6\epsilon}$  hold up to step  $i$  for every set  $I$  of size  $\alpha$ .

The main step is verifying the trend hypothesis of Lemma 7.3. Note that adding an edge cannot create any new  $I$ -open pairs, so we always have  $Y_{1,I}^+(i) = 0$ . Now we calculate the expected one-step change  $\mathbb{E}(Y_{1,I}^- | \mathcal{G}_i \wedge \mathcal{H}_i)$ . Recall that a pair  $e$  becomes closed at step  $i + 1$  if the process chooses the edge  $e_{i+1}$  in  $C_e(i)$  so a pair  $e$  in  $I$  becomes  $I$ -closed if is  $I$ -open and  $e_{i+1}$  is chosen in  $C_e(i) \setminus B_I(i)$ . Also, if  $e$  in  $I$  is open as well as  $I$ -open it may become an edge if the process chooses  $e_{i+1} = e$ . Now  $e_{i+1}$  is chosen uniformly among  $Q(i)$  open ordered pairs at step  $i$ , so

$$\mathbb{E}(Y_{1,I}^- | \mathcal{G}_i \wedge \mathcal{H}_i) = Q(i)^{-1} \left[ \left( \sum_{e \in X_I(i)} (|C_e(i)| \pm 1) \right) \pm |B_I(i)| p^{-1} n^{\epsilon/2} \right].$$

Here we also wrote  $X_I(i)$  for the set of  $I$ -open pairs in  $I$ , and used the estimate that each pair in  $B_I(i)$  can close at most  $p^{-1}n^{\epsilon/2}$  pairs by Corollary 6.2. Temporarily ignoring the error terms, this suggests the equation  $x_1'(t) = -q(t)^{-1}x_1(t)c(t)$ , which has  $q(t)$  as a solution, explaining our choice of  $x_1(t)$  above. To account for the error terms, we estimate  $Q(i)$  by Theorem 1.4,  $C_e(i)$  by Corollary 6.2,  $X_I(i)$  by the fact that we are conditioning on  $\mathcal{G}_i$  (interpreted for the current application of Lemma 7.3) and  $B_I(i)$  by definition of the event  $\mathcal{H}_i$ . Thus we estimate  $\mathbb{E}(Y_{1,I}^- | \mathcal{G}_i \wedge \mathcal{H}_i)$  as

$$\begin{aligned} & ((1 \pm e(t)/s_e)(q(t) \pm \theta(t)/s_e)n^2)^{-1} \cdot (1 \pm e(t)/s_1)(q(t) \pm \theta(t)/s_1)\alpha^2 \\ & \times (1 \pm e(t)/s_e)((2t)^{e_H-2}q(t) \pm 2\theta(t)/s_e \pm n^{-5\epsilon})a_H p^{-1}. \end{aligned}$$

Recalling that  $s = pn^2, y_q^-(t) = a_H(2t)^{e_H-2}q(t)$  and  $h_q(t) = (eq + \gamma)'(t)$  we see that we have the required condition

$$\mathbb{E}(Y_{1,I}^- | \mathcal{G}_i \wedge \mathcal{H}_i) = (y_q^-(t) \pm h_q(t)/4s_1)\alpha^2/s.$$

The boundedness hypothesis follows immediately from the definition of  $I$ -open pairs. Note that we can arrange for  $s_1^2 k_1 n^\epsilon = \alpha n^{5\epsilon} < n$ , since  $\epsilon$  is small. The remaining conditions of Lemma 7.3 follow by similar calculations as in the proof of Theorem 1.4. □

Next we show that a similar estimate holds for the number of open pairs in  $I$ .

**Lemma 11.2** *If  $H$  has smooth independence, then with high probability, for every set  $I$  of size  $\alpha$ , the number of open ordered pairs in  $I$  at step  $i$  is  $Q_I(i) = (1 \pm e(t)n^{-2\epsilon})(q(t) \pm 2n^{-2\epsilon})\alpha^2$ .*

*Proof* We need to estimate the number of ordered pairs in  $I$  that are  $I$ -open but not open. By Corollary 6.2 we can bound the number of pairs closed by any edge by  $p^{-1} \log n$  (say). By smooth independence we can assume that  $|P_I| < n^{-5\epsilon} p^{-1}$ , so at most  $n^{-5\epsilon} p^{-1} \cdot p^{-1} \log n$  pairs in  $I$  are closed but  $I$ -open. The required bound follows from these estimates and Lemma 11.1.  $\square$

Finally, we can show that the independence number of the process at time  $m$  is at most  $\alpha$ .

**Lemma 11.3** *If  $H$  has smooth independence, then with high probability, at time  $m$  every set  $I$  of size  $\alpha$  contains at least one edge.*

*Proof* At step  $i + 1$  the process chooses an edge uniformly at random from one of the  $Q(i)$  open ordered pairs. Since  $Q_I(i)$  of these belong to  $I$ , it fails to choose an edge in  $I$  with probability  $1 - Q_I(i)/Q(i)$ . Multiplying these probabilities and taking a union bound over  $I$  we can bound the probability that there is an independent set  $I$  of size  $\alpha$  by  $p_\alpha = \binom{n}{\alpha} \max_I \prod_{i=1}^m (1 - Q_I(i)/Q(i))$ . By Theorem 1.4 and Lemma 11.2 we have

$$Q_I(i)/Q(i) = ((1 \pm e(t)/s_e)(q(t) \pm \theta(t)/s_e)n^2)^{-1} \times (1 \pm e(t)n^{-2\epsilon})(q(t) \pm 2n^{-2\epsilon})\alpha^2.$$

Recalling that  $s_e = n^{1/2e_H - \epsilon}$  and  $\mu$  is chosen small enough that  $q(t)^{-1}$  and  $e(t)$  are at most  $n^\epsilon$  for  $t \leq t_{\max}$  we can estimate  $Q_I(i)/Q(i) = (1 \pm 10n^{-\epsilon})(\alpha/n)^2$ . Therefore

$$\log p_\alpha = \alpha(\log n - \log \alpha + 1 + O(1/n)) - m((1 \pm 10n^{-\epsilon})(\alpha/n)^2 \pm 2(\alpha/n)^4).$$

Also, since  $m = \mu(\log n)^{1/(e_H - 1)} pn^2$  and  $\alpha = 3\mu^{-1}(\log n)^{1 - 1/(e_H - 1)} p^{-1}$  we have  $m(\alpha/n)^2 = 3\alpha \log n$ . Thus we obtain

$$\log p_\alpha < -\alpha \log n = -3\mu^{-1}(\log n)^{2 - 1/(e_H - 1)} p^{-1},$$

so  $p_\alpha < \exp(-n^{1/e_H})$  (say), as required.  $\square$

## 12 Independence number and Ramsey bounds

In this section we show that cliques and cycles both have the smooth independence property. By Lemma 11.3, this is enough to prove Theorems 1.8

and 1.9, and then Theorem 1.2 follows immediately from Theorem 1.8. We will also show that a graph  $H$  satisfying the hypothesis of Theorem 1.7 has smooth independence, which is enough to prove that theorem.

We start with cycles, where we deduce smooth independence from a path-counting argument.

**Lemma 12.1** *The  $\ell$ -cycle  $C_\ell$  has smooth independence for  $\ell \geq 4$ .*

*Proof* Suppose  $I \subseteq [n]$  is a set of  $\alpha$  vertices and let  $B_I(i) \subseteq O(i - 1)$  be the ordered pairs at step  $i$  that are  $I$ -bad. We need to show that with high probability  $|B_I(i)| < n^{-6\epsilon} p^{-2}$  for all such  $I$  and all  $i \leq m$ . Consider the contrary event that  $|B_I(i)| \geq n^{-6\epsilon} p^{-2}$ , i.e. there are at least  $n^{-6\epsilon} p^{-2}$  ordered pairs that each close at least  $n^{-5\epsilon} p^{-1}$  ordered pairs in  $I$ . Then there is some ordered pair of edges  $uv, xy$  of  $C_\ell$  and  $B'_I(i) \subseteq B_I(i)$  with  $|B'_I(i)| \geq \ell^{-1} n^{-6\epsilon} p^{-2}$  such that for every pair  $cd$  in  $B'_I(i)$  there are at least  $\ell^{-1} n^{-6\epsilon} p^{-1}$  embeddings  $f$  of  $C_\ell \setminus \{uv, xy\}$  with  $f(x) = c, f(y) = d$  and  $f(u), f(v) \in I$ .

Set  $I_0 = I$  and for  $1 \leq j \leq \ell - 2$  define

$$I_j = \{v : |N_{G(i)}(v) \cap I_{j-1}| > n^{-10\epsilon} pn\}.$$

By Theorem 1.4 the degree of any vertex at time  $t$  is  $(1 \pm e(t)/s_e)(2t \pm 1/s_e)pn$ . Now  $p = n^{-(\ell-2)/(\ell-1)}$  and  $t \leq t_{\max} = \mu(\log n)^{1/(\ell-1)}$ , so  $pn = n^{1/(\ell-1)}$  and we can bound all degrees by  $(n \log n)^{1/(\ell-1)}$ . It follows that there are at most  $(n \log n)^{j/(\ell-1)}$  paths of length  $j$  starting at any given vertex, for any  $j$ . Also, if  $v \notin I_j$  we can improve on this estimate when counting paths of length  $j$  that start at  $v$  and end in  $I$ . To see this, consider choosing the vertex sequence of such a path starting at  $v$ , say  $v = v_{j-1}, \dots, v_0 \in I$ . At each step we have at most  $(n \log n)^{1/(\ell-1)}$  choices, and there must be some  $j - 1 \geq j' \geq 1$  where  $v_{j'} \notin I_{j'}$  but  $v_{j'-1} \in I_{j'-1}$ , when by definition we have at most  $n^{-10\epsilon} pn$  choices. This gives at most  $j(n^{-10\epsilon} pn)((n \log n)^{(j-1)/(\ell-1)}) < n^{-9\epsilon} n^{j/(\ell-1)}$  paths of length  $j$  that start at  $v$  and end in  $I$ .

Suppose without loss of generality that removing  $uv$  and  $xy$  from the cycle leaves a path of length  $\ell_1$  joining  $u$  to  $x$  and a path of length  $\ell_2$  joining  $v$  to  $y$ , with  $\ell_1 + \ell_2 = \ell - 2$  and  $\ell_1 > 0$  (we might have  $\ell_2 = 0$ , i.e.  $v = y$ ). We claim that for any pair  $cd$  in  $B'_I(i)$  we must have  $c \in I_{\ell_1}$  and  $d \in I_{\ell_2}$ . For suppose that  $c \notin I_{\ell_1}$ . Then there are at most  $n^{-9\epsilon} n^{\ell_1/(\ell-1)}$  paths of length  $\ell_1$  that start at  $c$  and end in  $I$ . Also, there are at most  $(n \log n)^{\ell_2/(\ell-1)}$  paths of length  $\ell_2$  that start at  $d$  and end in  $I$ . Thus we bound the number of embeddings  $f$  of  $C_\ell \setminus uv$  with  $f(x) = c, f(y) = d$  and  $f(u), f(v) \in I$  by  $n^{-9\epsilon} n^{\ell_1/(\ell-1)} \times (n \log n)^{\ell_2/(\ell-1)} < \ell^{-1} n^{-5\epsilon} p^{-1}$ , contradiction. Thus we have  $c \in I_{\ell_1}$ , and the same argument gives  $d \in I_{\ell_2}$ .



Now by Lemma 4.3, with high probability we have  $|I_j| \leq \alpha(8^{-1}\epsilon n^{-10\epsilon} pn)^{-j} < n^{1-(j+1)/(\ell-1)+11j\epsilon}$  for  $1 \leq j \leq \ell - 2$  and every  $I$  of size  $\alpha$ . Therefore,

$$\begin{aligned} |B'_I(i)| &\leq |I_{\ell_1}| \cdot |I_{\ell_2}| \leq n^{2-(\ell_1+\ell_2+2)/(\ell-1)+11(\ell_1+\ell_2)\epsilon} \\ &= n^{2-\ell/(\ell-1)+11(\ell-2)\epsilon} < n < n^{-6\epsilon} p^{-2}. \end{aligned}$$

This contradicts our assumption on  $B'_I(i)$ .

It remains to bound  $|P_I|$ . Now suppose  $|P_I| \geq n^{-5\epsilon} p^{-1}$ . Under this assumption there are at least  $n^{-5\epsilon} p^{-1}$  edges that are chosen by the process that each close at least  $n^{-5\epsilon} p^{-1}$  ordered pairs in  $I$ . Arguing as above, we see that there is some ordered pair of edges  $uv, xy$  of  $C_\ell$  and  $P'_I \subseteq P_I$  with  $|P'_I| \geq \ell^{-1} n^{-5\epsilon} p^{-1}$  such that for every edge  $cd$  in  $P'_I$  there are at least  $\ell^{-1} n^{-5\epsilon} p^{-1}$  embeddings  $f$  of  $C_\ell \setminus uv$  with  $f(x) = c, f(y) = d$  and  $f(u), f(v) \in I$ . Following the argument above, each such pair  $cd$  must have  $c \in I_{\ell_1}$  and  $d \in I_{\ell_2}$ . And here we have the added condition  $cd \in E(m)$ . By Lemma 4.2, with high probability we have

$$e(I_{\ell_1}, I_{\ell_2}) < \max\{4\epsilon^{-1}(|I_{\ell_1}| + |I_{\ell_2}|), p|I_{\ell_1}||I_{\ell_2}|n^{2\epsilon}\}.$$

This is less than  $n^{-1/\ell} p^{-1}$  unless  $\ell_2 = 0$ . Also, if  $\ell_2 = 0$  then  $\ell_1 = \ell - 2$ , so  $|I_{\ell_1}| < n^{11\ell\epsilon}$  and we can bound the number of edges incident to  $I_{\ell_1}$  by  $|I_{\ell_1}|(n \log n)^{1/(\ell-1)} < n^{1/(\ell-1)+12\ell\epsilon}$ . Either way we have  $e(I_{\ell_1}, I_{\ell_2}) < \ell^{-1} n^{-5\epsilon} p^{-1} \leq |P'_I|$ , by our earlier assumption, which contradicts the fact any edge  $cd$  in  $P'_I$  has  $c \in I_{\ell_1}$  and  $d \in I_{\ell_2}$ . Therefore with high probability we have  $|P_I| < n^{-5\epsilon} p^{-1}$  for all  $I$ , i.e.  $H$  has the smooth independence property.  $\square$

For cliques, we first consider the case  $H = K_s$  for some  $s \geq 6$ . Then  $p = n^{-2/(s+1)}$ . Consider any two edges  $uv, xy$  of  $H$  and let  $H^- = H \setminus uv$ . We have  $S_{xy, H^-} = p^{-1}$  and for  $s \geq 6$  we have  $S_{xy, H^-[B]} \geq p^2 n > p^{-1}$  for any  $B$  with  $xy \subsetneq B \subsetneq V_H$ , i.e.  $(xy, H^-)$  is strictly balanced. We show that this more general property suffices for smooth independence. Note that if  $H$  is any graph such that  $(xy, H^-)$  is strictly balanced for all  $xy, uv \in E_H$  then  $H$  has minimum degree at least 3. (To see this, assume for a contradiction that  $d_H(u) = 2$  and consider an extension  $(xy, H^-)$  where  $u \notin xy$ .)

**Lemma 12.2** *Suppose that  $(xy, H \setminus uv)$  is strictly balanced for any two edges  $uv, xy$  of  $H$ . Then  $H$  has smooth independence.*

*Proof* Suppose  $I \subseteq [n]$  is a set of  $\alpha$  vertices, and let  $P_I(i)$  be the set of ordered pairs of vertices  $cd$  with the property that there at least  $n^{-5\epsilon} p^{-1}$  ordered pairs  $ab \in I^2$  such that  $G(i - 1) + \{ab, cd\}$  contains a copy of  $H$  that

includes both  $ab$  and  $cd$ . Note that  $P_I(i)$  contains  $B_I(i)$  as well as all of the  $I$ -bad edges that have been chosen by the process so far (i.e.  $P_I \cap E(i)$ ). We show that with high probability  $|P_I(i)| < n^{-5\epsilon} p^{-1}$  for all such  $I$  and all  $i \leq m$ . Consider the contrary event that  $|P_I(i)| \geq n^{-5\epsilon} p^{-1}$ , i.e. there are at least  $n^{-5\epsilon} p^{-1}$  ordered pairs that each close at least  $n^{-5\epsilon} p^{-1}$  ordered pairs in  $I$ . Then there is some ordered pair of ordered edges  $uv, xy$  of  $H$  with  $u \notin \{x, y\}$  and  $P'_I(i) \subseteq P_I(i)$  with  $|P'_I(i)| \geq (2e_H)^{-1} n^{-5\epsilon} p^{-1}$  such that for every pair  $cd$  in  $P'_I(i)$  there are at least  $(2e_H)^{-1} n^{-5\epsilon} p^{-1}$  embeddings  $f$  of  $H^- = H \setminus \{uv, xy\}$  with  $f(x) = c, f(y) = d$  and  $f(u), f(v) \in I$ .

Write  $H^- = H \setminus \{uv, xy\}$ . Since  $(xy, H^-)$  is strictly balanced we have  $S_{B, H^-} < 1$  for any  $B$  with  $xyu \subseteq B \subsetneq V_H$ . Applying Lemma 5.2, we see that for any  $a, c, d \in [n]$  there are at most  $n^{4e_H\epsilon}$  embeddings  $f$  of  $H^- = H \setminus \{uv, xy\}$  with  $f(x) = c, f(y) = d$  and  $f(u) = a$ . For each pair  $cd \in P'_I(i)$  let  $U_{cd}$  be the set of vertices  $a \in I$  such that there is at least one embedding  $f$  of  $H^- = H \setminus \{uv, xy\}$  with  $f(x) = c, f(y) = d$  and  $f(u) = a$ . By definition of  $P'_I(i)$  we must have

$$|U_{cd}| > (2e_H)^{-1} n^{-5\epsilon} p^{-1} / n^{4e_H\epsilon} > n^{-10e_H\epsilon} p^{-1}$$

(say) for every pair  $cd \in P'_I(i)$ . Next we need the following claim.

**Claim 1**  $|U_{cd} \cap U_{c'd'}| < n^{-1/e_H} p^{-1}$  for any two pairs  $cd, c'd' \in P'_I(i)$ .

*Proof* Consider two embeddings  $f_1, f_2$  of  $H^-$  such that  $f_1(x) = c, f_1(y) = d, f_2(x) = c', f_2(y) = d'$  and  $f_1(u) = f_2(u) = a$ . Let  $C' = f_1(V_H) \cap f_2(V_H)$ . Let  $W$  be the join of  $J_1 = H^-$  and  $J_2 = H^-$  formed by identifying the sets  $f_1^{-1}(C')$  and  $f_2^{-1}(C')$  as a single set  $C$  on which  $f_1$  and  $f_2$  agree. Note that we have  $u \in C$ . For ease of notation we let  $x, y$  denote the copies of  $x, y$  in  $J_1$  and  $x', y'$  the copies of  $x, y$  in  $J_2$ . Let  $A = \{x, y\} \cup \{x', y'\}$ . Since  $cd \neq c'd'$  we have  $|A| \geq 3$ . Define  $\phi : A \rightarrow [n]$  by  $\phi(x) = c, \phi(y) = d, \phi(x') = c', \phi(y') = d'$ . We want to estimate  $N_{\phi, W}$ . The argument is very similar to that in Lemma 6.1. Choose  $B$  with  $A \subseteq B \subseteq V_W$  maximising  $S_{B, W}$ . We have cases depending on how  $V_{J_1}$  and  $V_{J_2}$  intersect. If  $f_1(V_H) = f_2(V_H)$ , i.e.  $V_{J_1} = V_{J_2}$ , then we have  $S_{B, W} \leq S_{B, H^-} \leq 1 < p^{-1}$ , since  $(xy, H^-)$  is strictly balanced and  $|A| \geq 3$ . We henceforth suppose that  $f_1(V_H) \neq f_2(V_H)$ . Define  $B_1 = B \cap V_{J_1}$  and  $B_2 = B \cap V_{J_2}$ . Next we consider the case  $V_{J_1} \subseteq V_{J_2} \cup A$ . If  $B_2 \neq \{x', y'\}$  then we have  $S_{B, W} \leq S_{B_2, J_2} \leq 1 < 1/p$  because  $(x'y', J_2)$  is strictly balanced. If  $B_2 = \{x', y'\}$  then we note that, since  $H$  has minimum degree at least 3 and  $V_{J_1} \setminus V_{J_2} \neq \emptyset$ , we have  $S_{B, W} \leq p S_{B_2, J_2} \leq 1 < 1/p$ . The analogous argument handles the case  $V_{J_2} \subseteq V_{J_1} \cup A$ .

Now suppose that  $V_{J_1} \setminus (V_{J_2} \cup A)$  and  $V_{J_2} \setminus (V_{J_1} \cup A)$  are non-empty. We consider subcases according to  $B_1$  and  $B_2$ . The first subcase is  $B_1 \cup C \neq V_{J_1}$ . Then we have  $S_{B_1 \cup C, J_1} = S_{B_1 \cup C, H^-} < 1$ , since  $u \in C$  and  $(xy, H^-)$  is

strictly balanced. Also  $S_{B_2, J_2} \leq S_{x'y', H^-} = p^{-1}$ , so  $S_{B, W} = S_{B_2, J_2} S_{B_1 \cup C, J_1} < p^{-1}$ . The second subcase is  $B_2 \cup C \neq V_{J_2}$ , when a similar argument gives  $S_{B, W} = S_{B_1, J_1} S_{B_2 \cup C, J_2} < p^{-1}$ . Finally, the third subcase is  $B_1 \cup C = V_{J_1}$  and  $B_2 \cup C = V_{J_2}$ . Then  $B_1$  contains  $V_{J_1} \setminus (A_1 \cup C)$  and  $B_2$  contains  $V_{J_2} \setminus (A_2 \cup C)$ , which are both non-empty. Since  $(xy, H^-)$  is strictly balanced we have  $S_{B_1, J_1} \leq 1$  and  $S_{B_2 \cup C, J_2} \leq 1$ , and so  $S_{B, W} = S_{B_1, J_1} S_{B_2 \cup C, J_2} \leq 1$ . In all cases we have  $S_{B, W} < p^{-1}$ , so  $S_{B, W} \leq n^{-1/(e_H-1)} p^{-1}$ , since it is an integer power of  $n^{-1/(e_H-1)}$ . Now Lemma 5.2 gives  $N_{\phi, W} < n^{4e_W \epsilon - 1/(e_H-1)} p^{-1}$ . Summing over all possible joins  $W$  we estimate  $|U_{cd} \cap U_{c'd'}| < n^{-1/e_H} p^{-1}$ , which proves the claim.  $\square$

Returning to the proof of the lemma, we now set  $\omega = n^{11e_H \epsilon}$  and choose  $\omega$  pairs of  $P'_I(i)$ , say  $c_1 d_1, \dots, c_\omega d_\omega$ . Recall that  $|U_{cd}| > n^{-10e_H \epsilon} p^{-1}$  for every  $cd \in P'_I(i)$ . Then  $|U_{c_i d_i} \setminus \cup_{j < i} U_{c_j d_j}| > n^{-10e_H \epsilon} p^{-1} - i n^{-1/e_H} p^{-1}$  for  $1 \leq i \leq \omega$  by the claim. This gives

$$\left| \bigcup_{i=1}^{\omega} U_{c_i d_i} \right| > \omega n^{-10e_H \epsilon} p^{-1} - \frac{1}{2} \omega^2 n^{-1/e_H} p^{-1} > n^\epsilon p^{-1},$$

say. But by definition the sets  $U_{c_i d_i}$  are contained in  $I$ , for which  $|I| = \alpha = 3\mu^{-1}(\log n)^{1-1/(e_H-1)} p^{-1}$  is too small. This contradiction shows that we cannot have  $|P_I(i)| \geq n^{-5\epsilon} p^{-1}$  for some  $I$  holding together with the bounds used from Lemma 5.2. These bounds hold with high probability, so with high probability we have  $|P_I(i)| < n^{-5\epsilon} p^{-1}$  for all  $I$ , i.e.  $H$  has the smooth independence property.  $\square$

The two arguments above can be generalised to prove smooth independence for a wider class of graphs  $H$ . However, for the sake of brevity and clarity, we restrict our attention to these simple cases here. We complete the discussion of cliques by showing that  $K_5$  has smooth independence. (The independence numbers for the  $K_3$ -free and  $K_4$ -free processes have already been obtained in [7].)

**Lemma 12.3**  $K_5$  has smooth independence.

*Proof* Write  $H = K_5$ . We argue as in the proof of Lemma 12.2. Consider  $I, P_I(i), uv, xy, P'_I(i), U_{cd}$  as defined in that proof. Now  $(xy, H^-)$  is not strictly balanced, but we do have  $S_{B, H^-} \leq 1$  for any  $B$  with  $xyu \subseteq B \subseteq V_H$ , so for every  $cd \in P'_I(i)$  we still obtain the bound  $|U_{cd}| > n^{-10e_H \epsilon} p^{-1}$ . Following that proof, our next step is to show that  $|U_{cd} \cap U_{c'd'}| < n^{-1/e_H} p^{-1}$  for any two pairs  $cd, c'd' \in P'_I(i)$ . In fact we will obtain a much stronger bound. Consider two embeddings  $f_1, f_2$  of  $H$  such that  $f_1(x) = c, f_1(y) = d,$

$f_2(x) = c', f_2(y) = d'$  and  $f_1(u) = f_2(u) = a$ . Define  $C', H^-, J_1, J_2, W, x', y', A$  and  $\phi$  as before. Choose  $B$  with  $A \subseteq B \subseteq V_W$  maximising  $S_{B,W}$ . Note that for any  $K$  with  $xy \not\subseteq K \subseteq V_H$  we have  $S_{K,H^-} \leq 1$ . So if  $V_{J_1} = V_{J_2}$  we have  $S_{B,W} \leq 1$ . Otherwise we consider cases according to  $B_1 = B \cap V_{J_1}$  and  $B_2 = B \cap V_{J_2}$ . Since  $S_{B,W} = S_{B_1,J_1} S_{B_2 \cup C, J_2}$ ,  $S_{B,W} = S_{B_2,J_2} S_{B_1 \cup C, J_1}$  and  $|B_1 \cup C|, |B_2 \cup C| \geq 3$  we see that  $S_{B,W} \leq 1$ , except possibly in the case  $B_1 = \{x, y\}$  and  $B_2 = \{x', y'\}$ . In this case we note that there is an edge from  $a \in C$  to  $B_2$  that is not contained in  $J_1$ , so  $S_{B,W} \leq p S_{B_1,J_1} = 1$ . In all cases we have  $S_{B,W} \leq 1$  and so  $N_{\phi,W} < n^{4\epsilon_W}$ . Summing over all possible joins  $W$  we estimate  $|U_{cd} \cap U_{c'd'}| < n^{5\epsilon_W}$ , say. Now the remainder of the proof follows as in Lemma 12.2. □

### 13 Concluding remarks

We have restricted our attention in this paper to those aspects of the  $H$ -free process needed for our applications to Ramsey and Turán bounds. However, we also view this work as the first stage in the study of this process as a model of independent interest. In the course of our arguments we have already described some properties of the model via our asymptotic formulae for trackable extension variables; for example, we have shown that for fixed graphs  $\Gamma$  that do not contain  $H$  as a subgraph, excluding ‘critical’ cases, the number of copies of  $\Gamma$  in  $G(i)$  is roughly the same as the number of copies of  $\Gamma$  in the unconstrained random graph  $G(n, i)$ . In principle, one may ask for analogues in the graph  $G(i)$  produced by the  $H$ -free process of any property known to hold in  $G(n, i)$ . But the most natural next steps are continued investigation of the independence number and development of upper bounds on the number of steps in the  $H$ -free process. For independent sets, there are other classes of graphs covered by our methods, but for clarity we have restricted our attention to certain concrete settings rather than stating a complicated general theorem. One might hope that any strictly 2-balanced graph can be analysed by these methods. With respect to upper bounds, we believe that the number of steps in the  $H$ -free process is at most a constant times the lower bound we establish here for any strictly 2-balanced  $H$ . In fact, we are even prepared to make this conjecture for the degree of each vertex.

**Conjecture 13.1** *For any strictly 2-balanced graph  $H$  there is a constant  $C$  so that with high probability the maximal  $H$ -free graph  $G$  on  $n$  vertices produced by the  $H$ -free process has maximum degree*

$$\Delta(G) < C n^{1-(v_H-2)/(\epsilon_H-1)} (\log n)^{1/(\epsilon_H-1)}.$$

For the triangle-free process this follows from the bound on the independence number (see [7]), but in general it is a separate question. The later

evolution of the process, where Theorem 1.4 no longer applies, is also an intriguing topic for further study.

## References

1. Ajtai, M., Komlós, J., Szemerédi, E.: A note on Ramsey numbers. *J. Comb. Theory, Ser. A* **29**, 354–360 (1980)
2. Alon, N., Spencer, J.: *The Probabilistic Method*, second edn. Wiley, New York (2000)
3. Alon, N., Rónyai, L., Szabó, T.: Norm-graphs: variations and applications. *J. Comb. Theory, Ser. B* **76**, 280–290 (1999)
4. Alon, N., Ben-Shimon, S., Krivelevich, M.: A note on regular Ramsey graphs. *J. Graph Theory* (to appear)
5. Apostol, T.M.: An elementary view of Euler’s summation formula. *Am. Math. Mon.* **106**, 409–418 (1999)
6. Barbour, A.D., Karoński, M., Ruciński, A.: A central limit theorem for decomposable random variables, with applications to random graphs. *J. Comb. Theory, Ser. B* **47**, 125–145 (1989)
7. Bohman, T.: The triangle-free process. *Adv. Math.* **221**, 1653–1677 (2009)
8. Bollobás, B., Riordan, O.: Constrained graph processes. *Electron. J. Comb.* **7**, R18 (2000)
9. Brown, W.G.: On graphs that do not contain a Thomsen graph. *Can. Math. Bull.* **9**, 281–289 (1966)
10. Caro, Y., Li, Y., Rousseau, C.C., Zhang, Y.: Asymptotic bounds for some bipartite graph—complete graph Ramsey numbers. *Discrete Math.* **220**, 51–56 (2000)
11. Durrett, R.: *Random Graph Dynamics*. Cambridge Univ. Press, Cambridge (2007)
12. Erdős, P.: Graph theory and probability, II. *Can. J. Math.* **13**, 346–352 (1961)
13. Erdős, P.: Extremal problems in number theory, combinatorics and geometry. In: *Proceedings of the ICM*, pp. 51–70. PWN, Warsaw (1984)
14. Erdős, P., Simonovits, M.: Some extremal problems in graph theory. *Colloq. Math. Soc. János Bolyai* **4**, 377–390 (1969)
15. Erdős, P., Spencer, J.H.: *Probabilistic Methods in Combinatorics*. Academic Press, San Diego (1974)
16. Erdős, P., Stone, A.H.: On the structure of linear graphs. *Bull. Am. Math. Soc.* **52**, 1087–1091 (1946)
17. Erdős, P., Suen, S., Winkler, P.: On the size of a random maximal graph. *Random Struct. Algorithms* **6**, 309–318 (1995)
18. Füredi, Z.: Turán type problems. In: *Surveys in Combinatorics*. London Math. Soc. Lecture Note Ser., vol. 166, pp. 253–300. Cambridge Univ. Press, Cambridge (1991)
19. Füredi, Z.: New asymptotics for bipartite Turán numbers. *J. Comb. Theory, Ser. A* **75**, 141–144 (1996)
20. Füredi, Z.: An upper bound on Zarankiewicz’ problem. *Comb. Probab. Comput.* **5**, 29–33 (1996)
21. Grable, D.: On random greedy triangle packing. *Electron. J. Comb.* **4**, R11 (1997)
22. Graham, R.L., Rothschild, B.L., Spencer, J.H.: *Ramsey Theory*. Wiley, New York (1990)
23. Kim, J.H.: The Ramsey number  $R(3, t)$  has order of magnitude  $t^2/\log t$ . *Random Struct. Algorithms* **7**, 173–207 (1995)
24. Kövari, T., Sós, V.T., Turán, P.: On a problem of K. Zarankiewicz. *Colloq. Math.* **3**, 50–57 (1954)
25. Li, Y., Zang, W.: The independence number of graphs with a forbidden cycle and Ramsey numbers. *J. Comb. Optim.* **7**, 353–359 (2003)
26. Osthus, D., Taraz, A.: Random maximal  $H$ -free graphs. *Random Struct. Algorithms* **18**, 61–82 (2001)

27. Ruciński, A.: Recent developments in random graphs. In: Proceedings of the International Summer School on Probability and Statistics, Varna. Online at: <http://www.staff.amu.edu.pl/~rucinski/papers/43.pdf> (1994)
28. Ruciński, A., Wormald, N.: Random graph processes with degree restrictions. *Comb. Probab. Comput.* **1**, 169–180 (1992)
29. Seierstad, T.G.: A central limit theorem via differential equations. *Ann. Appl. Probab.* **19**, 661–675 (2009)
30. Seierstad, T.G.: Stronger large deviation bounds for Wormald’s differential equation method (submitted)
31. Spencer, J.: Counting extensions. *J. Comb. Theory, Ser. A* **55**, 247–255 (1990)
32. Spencer, J.: Asymptotic lower bounds for Ramsey functions. *Discrete Math.* **20**, 69–76 (1997)
33. Spencer, J.: Maximal trianglefree graphs and Ramsey  $R(3, k)$ . Available online at: <http://www.cs.nyu.edu/spencer/papers/ramsey3k.pdf> (unpublished manuscript)
34. Sudakov, B.: A note on odd cycle-complete graph Ramsey numbers. *Electron. J. Comb.* **9**, N1 (2002)
35. Turán, P.: Eine Extremalaufgabe aus der Graphentheorie. *Mat. Fiz. Lapok* **48**, 436–452 (1941)
36. Wolfowitz, G.: Lower bounds for the size of random maximal  $H$ -free graphs. *Electron. J. Comb.* **16**, R4 (2009)
37. Wormald, N.C.: The differential equation method for random graph processes and greedy algorithms. In: *Lectures on Approximation and Randomized Algorithms*, pp. 73–155. PWN, Warsaw (1999)