Combinatorics, Probability and Computing (2009) 18, 271–300. © 2009 Cambridge University Press
doi:10.1017/S0963548308009632 Printed in the United Kingdom

Online Ramsey Games in Random Graphs

MARTIN MARCINISZYN, RETO SPÖHEL[†] and ANGELIKA STEGER[‡]

Institute of Theoretical Computer Science, ETH Zürich, 8092 Zürich, Switzerland
(e-mail: {mmarcini|rspoehel|steger}@inf.ethz.ch)

Received 14 March 2006; revised 13 October 2008; first published online 21 January 2009

Consider the following one-player game. Starting with the empty graph on n vertices, in every step a new edge is drawn uniformly at random and inserted into the current graph. This edge has to be coloured immediately with one of r available colours. The player's goal is to avoid creating a monochromatic copy of some fixed graph F for as long as possible. We prove a lower bound of $n^{\beta(F,r)}$ on the typical duration of this game, where $\beta(F,r)$ is a function that is strictly increasing in r and satisfies $\lim_{r\to\infty}\beta(F,r) = 2 - 1/m_2(F)$, where $n^{2-1/m_2(F)}$ is the threshold of the corresponding offline colouring problem.

1. Introduction

A well-studied problem in deterministic graph theory is finding edge-colourings of a given graph G that avoid a monochromatic copy of a given fixed graph F. It follows from Ramsey's celebrated result [12] that *every* two-colouring of the edges of a complete graph on n vertices contains a monochromatic clique on $\Theta(\log n)$ vertices and thus, for n sufficiently large, also a monochromatic copy of F. While that seems to rely on the fact that the K_n is a very dense graph, Folkman [4] and, in a more general setting, Nešetřil and Rödl [11] showed that there also exist locally sparse graphs $G = G(F)$ with the property that every two-colouring of the edges of G contains a monochromatic copy of F. By transferring the problem into a random setting, Rödl and Ruciński showed that in fact such graphs G are quite frequent. More precisely, they proved that the existence of a two-colouring without a monochromatic copy of F a.a.s. depends only on the density p of the random graph $G_{n,p}$.

Theorem 1.1 ([7, 13, 14]). *Let $r \geqslant 2$ and F be a graph that is not a star forest and, in the case $r = 2$, not a forest of stars and P_3s. Moreover, let $\mathcal{P}$ = 'every r-edge-colouring of G contains a*

[†] Supported by the Swiss National Science Foundation, grant 200021-108158.
[‡] A preliminary version of this paper appeared in the proceedings of the 19th International Workshop on Randomized Techniques in Computation (RANDOM'05) [9].

https://doi.org/10.1017/S0963548308009632 Published online by Cambridge University Press

monochromatic copy of F'. Then there exist positive constants $c = c(F,r)$ and $C = C(F,r)$ such that

$$\lim_{n\to\infty} \mathbb{P}[G_{n,m} \in \mathcal{P}] = \begin{cases} 1 & \text{if } m > Cn^{2-1/m_2(F)}, \\ 0 & \text{if } m < cn^{2-1/m_2(F)}, \end{cases}$$

where

$$m_2(F) := \max_{H\subseteq F} \frac{e_H - 1}{v_H - 2}.$$

While the above results consider a static or 'offline' situation, Friedgut, Kohayakawa, Rödl, Ruciński and Tetali [6] transferred these problems into an algorithmic or 'online' setting. They studied the following one-player game. The board is a graph with n vertices, which initially contains no edges. The edges are presented to the player, henceforth called Painter, one by one in an order chosen uniformly at random among all permutations of the underlying complete graph. Painter must assign one of r available colours to each edge immediately. Her objective is to colour as many edges as possible without creating a monochromatic copy of some fixed graph F. The game ends as soon as the first monochromatic copy of F is closed. We refer to this as the online F-avoidance game with r colours and to the number of properly coloured edges as its *duration*. Friedgut, Kohayakawa, Rödl, Ruciński and Tetali [6] showed that for the triangle-avoidance game with two colours, there is a threshold that differs dramatically from the one in the offline case. While in the offline setting for a small $c > 0$ a.a.s. every random graph with $cn^{3/2}$ edges can be properly coloured [7], the online case will a.a.s. stop whenever Painter has seen substantially more than $n^{4/3}$ edges.

We say that $N_0 = N_0(F,r,n)$ is a threshold for the game if there exists a strategy such that Painter a.a.s. survives with this strategy for any $N \ll N_0$ edges, and if, moreover, Painter a.a.s. loses the game within any $N \gg N_0$ edges, regardless of her strategy. Note that the existence of such a threshold does not immediately follow from the Bollobás–Thomason result about monotone graph properties [2]. However, a variant of the well-known multi-round exposure proof of their result yields that indeed a threshold exists for all F and r (*cf.* Lemma 2.1 below).

In this paper, we prove that for every graph F and every integer $r \geqslant 1$, the threshold for the online F-avoidance game with r colours exists and satisfies

$$n^{\beta(F,r)} \leqslant N_0(F,r,n) \leqslant n^{2-1/m_2(F)},$$

where $\beta(F,r)$ is a function satisfying

$$\lim_{r\to\infty} \beta(F,r) = 2 - 1/m_2(F). \tag{1.1}$$

Note that $n^{2-1/m_2(F)}$ is an *a priori* upper bound on the duration of the game with any fixed number r of colours due to Theorem 1.1.

The function $\beta(F,r)$ we prove is different for forests and non-forests. If F contains at least one cycle, it is given by the density measures $\overline{m}_2^r(F)$, which are defined inductively as follows:

$$\overline{m}_2^r(F) := \begin{cases} \max\limits_{H\subseteq F} \dfrac{e_H}{v_H} & \text{if } r = 1, \\ \max\limits_{H\subseteq F} \dfrac{e_H}{v_H - 2 + 1/\overline{m}_2^{r-1}(F)} & \text{if } r \geqslant 2. \end{cases} \tag{1.2}$$

https://doi.org/10.1017/S0963548308009632 Published online by Cambridge University Press

Theorem 1.2 (Main result). *Let F be a graph that is not a forest, and let $r \geqslant 1$. Then the online F-avoidance edge-colouring game with r colours has a threshold $N_0(F,r,n)$ that satisfies*

$$N_0(F,r,n) \geqslant n^{2-1/\overline{m}_2^r(F)}.$$

Note that the parameter $\overline{m}_2^r(F)$ is strictly increasing in r (*cf.* Lemma 2.2 below), and satisfies

$$\lim_{r\to\infty} \overline{m}_2^r(F) = m_2(F). \tag{1.3}$$

The lower bound given by Theorem 1.2 thus depends on the number of colours r, in contrast to Theorem 1.1, and approaches the offline threshold as the number of colours grows. In the special case when F is 2-balanced (F is called 2-balanced if $m_2(F) = (e_F - 1)/(v_F - 2)$), the convergence (1.3) takes the form of a geometric series and yields a more explicit bound.

Corollary 1.3. *Let F be a 2-balanced graph that is not a forest, and let $r \geqslant 1$. Then the online F-avoidance edge-colouring game with r colours has a threshold $N_0(F,r,n)$ that satisfies*

$$N_0(F,r,n) \geqslant n^{\left(2-1/m_2(F)\right)\left(1-e_F^{-r}\right)}.$$

We believe that for a large class of graphs including cliques and cycles of arbitrary size, the lower bound given in Theorem 1.2 is in fact the threshold of the game. In [10] we prove this when $r = 2$. For arbitrary r, we believe that in particular the following is true.

Conjecture 1.4. *For all $\ell \geqslant 2$ and $r \geqslant 1$, the threshold for the online K_ℓ-avoidance edge-colouring game with r colours is*

$$N_0(\ell,r,n) = n^{\left(2-\frac{2}{\ell+1}\right)\left(1-\binom{\ell}{2}^{-r}\right)}.$$

In the proof of Theorem 1.2, we show that Painter can survive with the following strategy. For a given non-empty graph F, define an infinite sequence of graphs as follows:

$$H_r = H_r(F) := \begin{cases} \operatorname{argmax}_{H\subseteq F} \dfrac{e_H}{v_H} & \text{if } r = 1, \\ \operatorname{argmax}_{H\subseteq F} \dfrac{e_H}{v_H - 2 + 1/\overline{m}_2^{r-1}(F)} & \text{if } r \geqslant 2. \end{cases} \tag{1.4}$$

Label the available colours with the numbers $1,\dots,r$. Painter assigns colour r to all edges unless this closes a monochromatic copy of H_r. In that case, she uses colour $r-1$, unless this closes a monochromatic copy of H_{r-1}, and so on. She stops playing as soon as she cannot avoid creating a copy of H_1 in colour 1 since all colours are exhausted. (Of course she could continue playing until she is forced to create a copy of F in colour 1, but it turns out that this makes no difference asymptotically.) Corollary 1.3 corresponds to the special case when $H_1 = \cdots = H_r = F$.

This maximization over r potentially different subgraphs in (1.2) gives rise to the curious phenomenon that a disconnected graph F may have a higher threshold than each of his components. For an example, consider $r = 2$ and F the disjoint union of the complete graph K_4 on four vertices and the wheel W_4 with four spokes. Together with results from [10], Theorem 1.2 yields that the individual components have thresholds $N_0(K_4,2,n) = n^{14/9} \approx n^{1.556}$ and $N_0(W_4,2,n) = n^{99/64} \approx n^{1.547}$ respectively, whereas Theorem 1.2 yields that $N_0(F,2,n)$ is at least $n^{25/16} \approx n^{1.563}$.

https://doi.org/10.1017/S0963548308009632 Published online by Cambridge University Press

While we do believe that the lower bound from Theorem 1.2 provides the correct threshold for cliques, cycles, and many other graphs, this is unfortunately not true for all graphs. Here one presumably needs to define a more complex colouring strategy than our greedy strategy formulated above. For an example, consider $r = 2$ and F the graph formed by two triangles overlapping in exactly one vertex. Our proof of Theorem 1.2 shows that greedily avoiding triangles (say, greedily using red, and using blue to avoid red triangles) guarantees survival for any $N \ll n^{2-1/\overline{m}_2^2(F)} = n^{25/18} \approx n^{1.389}$ edges. It turns out that switching back to red when this strategy would complete a blue copy of F improves the lower bound to $n^{17/12} \approx n^{1.417}$.

If F is a forest, things behave somewhat differently. Again a lower bound $n^{\beta(F,r)}$ satisfying (1.1) can be achieved by successively avoiding a suitably chosen subgraph in each colour. However, to obtain the best-possible lower bound within this framework, these subgraphs need to be chosen according to a different criterion to (1.4), and the resulting lower bound is strictly less than $n^{2-1/\overline{m}_2^r(F)}$. For the precise statement, we refer to Theorem 5.1. We were able to verify by *ad hoc* arguments that for $r = 2$ and all forests on up to five vertices, our lower bound is in fact the threshold of the game. We leave it as an intriguing open problem whether this is true in general.

Note that both in Theorem 1.2 and in Theorem 5.1, the case $r = 1$ is a mere reformulation of the well-known theorem about appearance of small subgraphs due to Bollobás.

Theorem 1.5 ([1]). *Let F be a non-empty graph, and let $\mathcal{P} =$ 'G contains a copy of F'. Then*

$$\lim_{n\to\infty} \mathbb{P}[G_{n,m} \in \mathcal{P}] = \begin{cases} 1 & \text{if } m \gg n^{2-1/m(F)}, \\ 0 & \text{if } m \ll n^{2-1/m(F)}, \end{cases}$$

where

$$m(F) := \max_{H\subseteq F} \frac{e_H}{v_H}.$$

As already pointed out in [6] for the case of triangles, in all cases where a threshold is known, the proofs show that it cannot be strengthened to a 'semi-sharp' threshold as in Theorem 1.1. This suggests that online colourability is essentially determined by local properties, and in some sense more closely related to Theorem 1.5 than to Theorem 1.1 (see [5] for a discussion of sharp thresholds and global versus local graph properties).

1.1. Vertex-colourings

In [8], the analogous question for vertex-colourings instead of edge-colourings was studied. It was shown that densities $\overline{m}_1^r(F)$ defined similarly to (1.2) determine threshold functions for the online F-avoidance vertex-colouring game with r colours for all $r \geqslant 1$ and a large class of graphs F including cliques and cycles. Analogously to (1.3), the density measures $\overline{m}_1^r(F)$ are strictly increasing in r and satisfy

$$\lim_{r\to\infty} \overline{m}_1^r(F) = m_1(F),$$

where $m_1(F) = \max_{H\subseteq F} e_H/(v_H - 1)$ determines the offline threshold for the existence of a vertex-colouring without monochromatic copies of F [7].

https://doi.org/10.1017/S0963548308009632 Published online by Cambridge University Press

These results yield in particular the following vertex-colouring analogue of Conjecture 1.4.

Theorem 1.6 ([8]). *For all $\ell \geqslant 2$ and $r \geqslant 1$, the threshold for the online K_ℓ-avoidance vertex-colouring game with r colours is*

$$p_0(\ell, r, n) = n^{-\frac{2}{\ell(1-\ell^{-r})}}.$$

1.2. Organization of this paper

We explain our notation and prove some auxiliary results in Section 2. In Section 3, we give the main part of the proof of Theorem 1.2. The argument is simple, but relies on a quite involved combinatorial statement, which is proved in Section 4. We investigate the case of forests in Section 5.

2. Preliminaries and notation

We consider the random graph process $(G(n,N))_{0\leqslant N\leqslant\binom{n}{2}}$, where the edges appear uniformly at random one after the other, *i.e.*, in one of $\binom{n}{2}!$ possible permutations. It is easily seen that $G(n,N)$ is uniformly distributed over all graphs on n vertices with exactly N edges. We let $G_{n,m}$ denote a graph chosen uniformly at random from all graphs on n vertices with exactly $m = m(n)$ edges.

Lemma 2.1. *Let F be a non-empty graph, and let $r \geqslant 1$. Then the online F-avoidance edge-colouring game with r colours has a threshold $N_0(F,r,n)$.*

Proof. For the proof, consider F and r fixed. Let Π denote the space of all possible strategies, *i.e.*, functions mapping all (countably infinitely many) game positions Painter may face to one of the r available colours. For a fixed strategy $\pi \in \Pi$, let the random variable $D_\pi(n)$ denote the duration of the online F-avoidance game with r colours on n vertices when played according to π.

Note that for all $a, b, n \in \mathbb{N}$ we have

$$\sup_{\pi\in\Pi} \mathbb{P}[D_\pi(n) \geqslant a+b] \leqslant \sup_{\pi\in\Pi} \mathbb{P}[D_\pi(n) \geqslant a] \cdot \sup_{\pi\in\Pi} \mathbb{P}[D_\pi(n) \geqslant b], \tag{2.1}$$

since the right-hand side is the maximum probability that Painter survives for $a+b$ moves in a relaxed game where the board is erased and the game restarted after the first a moves.

With this observation at hand, it is easy to prove that

$$N_0(n) := \max\Big\{ N : \sup_{\pi\in\Pi} \mathbb{P}[D_\pi(n) \geqslant N] \geqslant 1/2 \Big\} \tag{2.2}$$

is a threshold function. On one hand, there exists no function $N_1(n) \ll N_0(n)$ with

$$\liminf_{n\to\infty} \sup_{\pi\in\Pi} \mathbb{P}[D_\pi(n) \geqslant N_1(n)] =: c_1 < 1, \tag{2.3}$$

https://doi.org/10.1017/S0963548308009632 Published online by Cambridge University Press

since otherwise we would obtain for $k := \lceil \log_{c_1}(1/4) \rceil$ (or $k := 1$ if $c_1 = 0$) and infinitely many n that

$$\begin{aligned} 1/2 &\overset{(2.2)}{\leqslant} \sup_{\pi\in\Pi} \mathbb{P}[D_\pi(n) \geqslant N_0(n)] \\ &\leqslant \sup_{\pi\in\Pi} \mathbb{P}[D_\pi(n) \geqslant kN_1(n)] \\ &\overset{(2.1)}{\leqslant} \sup_{\pi\in\Pi} \mathbb{P}[D_\pi(n) \geqslant N_1(n)]^k \\ &\overset{(2.3)}{=} (c_1 + o(1))^k \\ &\leqslant 1/4 + o(1), \end{aligned}$$

an obvious contradiction. On the other hand, there exists no function $N_2(n) \gg N_0(n)$ with

$$\limsup_{n\to\infty} \sup_{\pi\in\Pi} \mathbb{P}[D_\pi(n) \geqslant N_2(n)] =: c_2 > 0, \tag{2.4}$$

since for $k := \lceil -\log_2(c_2/2) \rceil$ and infinitely many n, we would obtain

$$\begin{aligned} c_2 + o(1) &\overset{(2.4)}{=} \sup_{\pi\in\Pi} \mathbb{P}[D_\pi(n) \geqslant N_2(n)] \\ &\leqslant \sup_{\pi\in\Pi} \mathbb{P}[D_\pi(n) \geqslant k(N_0(n)+1)] \\ &\overset{(2.1)}{\leqslant} \sup_{\pi\in\Pi} \mathbb{P}[D_\pi(n) \geqslant N_0(n)+1]^k \\ &\overset{(2.2)}{<} (1/2)^k \\ &\leqslant c_2/2, \end{aligned}$$

which is again a contradiction. □

All graphs are simple and undirected. The number of vertices of a graph G is denoted by v_G or $v(G)$, and similarly the number of edges by e_G or $e(G)$. We denote a clique on ℓ vertices by K_ℓ, a cycle on ℓ vertices by C_ℓ, and a path with $\ell + 1$ vertices by P_ℓ.

The standard density measure for graphs is $d(G) := e_G/v_G$, which is exactly half of the average degree. Besides $d(G)$, we also use the so-called 2-density $d_2(G) := (e_G - 1)/(v_G - 2)$. For the sake of completeness, we also define $d_2(K_2) := 1/2$ and $d(G) = d_2(G) := 0$ if G is empty. For a given density function d_i, we let $m_i(G) := \max_{H\subseteq G} d_i(H)$. We say that G is *balanced with respect to* d_i if $m_i(G) = d_i(G)$. We simply write *balanced* for balancedness w.r.t. d, and 2-*balanced* for balancedness w.r.t. d_2.

For non-empty graphs F and G and any integer $r \geqslant 2$, we define

$$\overline{d}_2^r(F,G) := \frac{e_G}{v_G - 2 + 1/\overline{m}_2^{r-1}(F)},$$

where $\overline{m}_2^{r-1}(F)$ is defined as in (1.2). We set $\overline{d}_2^r(F,G) := 0$ if F or G is empty. Note that

$$\overline{m}_2^r(F) = \max_{H\subseteq F} \overline{d}_2^r(F,H)$$

https://doi.org/10.1017/S0963548308009632 Published online by Cambridge University Press

for all $r \geqslant 2$, and recall that $\overline{m}_2^1(F) = m(F)$ by definition. We say that F is balanced w.r.t. $\overline{d}_2^r$ if $\overline{m}_2^r(F) = \overline{d}_2^r(F, F)$.

The maximum density measures m and m_2 are well known and motivated by Theorems 1.5 and 1.1 respectively. It is also well known that every non-empty graph that is not a matching satisfies $m(F) < m_2(F)$, and that every 2-balanced graph is balanced. In contrast, the maximum densities $\overline{m}_2^r(F)$ seem not to have been studied before. The next lemma shows that they interpolate between m and m_2 in some sense.

Lemma 2.2. *Let F be a graph that is not a matching.*

(i) *We have*

$$m(F) = \overline{m}_2^1(F) < \overline{m}_2^2(F) < \cdots < \overline{m}_2^r(F) < \cdots < m_2(F).$$

(ii) *We have*

$$\lim_{r\to\infty} \overline{m}_2^r(F) = m_2(F).$$

(iii) *If F is 2-balanced, it is is balanced with respect to $\overline{d}_2^r$ for all $r \geqslant 1$.*

(iv) *For all $r \geqslant 2$, if F is balanced with respect to $\overline{d}_2^r$, it is balanced with respect to $\overline{d}_2^{r-1}$.*

For the proof we use the following observation, which we state separately for further reference.

Proposition 2.3. *For $a, c, C \in \mathbb{R}$ and $b > d > 0$, we have*

$$\frac{a}{b} \geqslant C \wedge \frac{c}{d} \leqslant C \quad \Longrightarrow \quad \frac{a-c}{b-d} \geqslant C$$

and

$$\frac{a-c}{b-d} \geqslant \frac{a}{b} \quad \Longleftrightarrow \quad \frac{c}{d} \leqslant \frac{a}{b}.$$

□

Proof of Lemma 2.2. We prove (i) by induction on r, showing that $\overline{m}_2^r(F) < \overline{m}_2^{r+1}(F)$ and $\overline{m}_2^r(F) < m_2(F)$. In the base case $r = 1$, these statements follow from the inequalities $d(H) < \overline{d}_2^2(F, H)$ and $d(H) < d_2(H)$ for all subgraphs $H \subseteq F$ that are not matchings. For $r \geqslant 2$, the first inequality follows from

$$\overline{m}_2^r(F) = \max_{H\subseteq F} \frac{e_H}{v_H - 2 + 1/\overline{m}_2^{r-1}(F)} \overset{\text{ind.}}{<} \max_{H\subseteq F} \frac{e_H}{v_H - 2 + 1/\overline{m}_2^{r}(F)} = \overline{m}_2^{r+1}(F).$$

The second inequality follows with $(e_H - 1)/(v_H - 2) \leqslant m_2(F)$ from

$$\overline{m}_2^r(F) = \max_{H\subseteq F} \frac{e_H}{v_H - 2 + 1/\overline{m}_2^{r-1}(F)} \overset{\text{ind.}}{<} \max_{H\subseteq F} \frac{e_H}{(e_H - 1)/m_2(F) + 1/m_2(F)} = m_2(F).$$

This concludes the proof of (i).

Clearly, (i) implies that $\overline{m}_2^\infty(F) := \lim_{r\to\infty} \overline{m}_2^r(F)$ exists and is at most $m_2(F)$. On the other hand, letting $r \to \infty$ in (1.2) yields that $\frac{e_H}{v_H - 2 + 1/\overline{m}_2^\infty(F)} \leqslant \overline{m}_2^\infty(F)$ for all $H \subseteq F$, which by elementary calculations is equivalent to $\overline{m}_2^\infty(F) \geqslant m_2(F)$. This proves (ii).

https://doi.org/10.1017/S0963548308009632 Published online by Cambridge University Press

An easy calculation shows that $\overline{d}_2^r(F,F) \leqslant d_2(F)$ is equivalent to $\overline{m}_2^{r-1}(F) \leqslant d_2(F)$. Hence, F satisfies $\overline{d}_2^r(F,F) \leqslant d_2(F)$ by (i). If F is 2-balanced, we obtain for every subgraph $H \subsetneq F$ with $3 \leqslant v_H < v_F$ that

$$\frac{e_F - e_H}{(v_F - 2 + 1/\overline{m}_2^{r-1}(F)) - (v_H - 2 + 1/\overline{m}_2^{r-1}(F))} = \frac{(e_F - 1) - (e_H - 1)}{(v_F - 2) - (v_H - 2)} \overset{\text{Prop. 2.3}}{\geqslant} d_2(F) \geqslant \overline{d}_2^r(F,F),$$

which again by Proposition 2.3 implies that $\overline{d}_2^r(F,H) \leqslant \overline{d}_2^r(F,F)$, concluding the proof of (iii).

Claim (iv) follows analogously from

$$\frac{e_F - e_H}{(v_F - 2 + 1/\overline{m}_2^{r-1}(F)) - (v_H - 2 + 1/\overline{m}_2^{r-1}(F))} = \frac{e_F - e_H}{(v_F - 2 + 1/\overline{m}_2^{r}(F)) - (v_H - 2 + 1/\overline{m}_2^{r}(F))} \overset{\text{Prop. 2.3}}{\geqslant} \overline{d}_2^r(F,F) \geqslant \overline{d}_2^{r-1}(F,F)$$

for every subgraph $H \subsetneq F$ with $2 \leqslant v_H < v_F$. □

Clearly, if F is a forest, the graph $H_r = H_r(F)$ defined by (1.4) is a forest. The next lemma states that the converse is also true.

Lemma 2.4. *Let F be a graph that is not a forest. For all $r \geqslant 2$, a graph H with $\overline{d}_2^r(F,H) = \overline{m}_2^r(F)$ satisfies $e_H \geqslant v_H$. In particular, $H_r = H_r(F)$ defined by* (1.4) *is not a forest.*

Proof. If $e_H \leqslant v_H - 1$, we obtain with $\overline{m}_2^{r-1}(F) \geqslant m(F) \geqslant 1$ that

$$\overline{m}_2^r(F) = \overline{d}_2^r(F,H) \leqslant \frac{v_H - 1}{v_H + 1/\overline{m}_2^{r-1}(F)} = 1 + \frac{1 - 1/\overline{m}_2^{r-1}(F)}{v_H - 2 + 1/\overline{m}_2^{r-1}(F)} \overset{v_H \geqslant 2}{\leqslant} \overline{m}_2^{r-1}(F),$$

contradicting Lemma 2.2. □

The inductive definition of $\overline{m}_2^r(F)$ can also be written in an explicit form, which simplifies to a nice expression if F is 2-balanced.

Lemma 2.5. *For all non-empty graphs F and $r \geqslant 1$, we have*

$$\overline{m}_2^r(F) = \max_{H_1,\dots,H_r \subseteq F} \frac{\prod_{i=1}^r e(H_i)}{2 + \sum_{i=1}^r (v(H_i) - 2) \prod_{j=1}^{i-1} e(H_j)}. \tag{2.5}$$

If F is 2-balanced, we have

$$2 - 1/\overline{m}_2^r(F) = \big(2 - 1/m_2(F)\big)\big(1 - e_F^{-r}\big).$$

https://doi.org/10.1017/S0963548308009632 Published online by Cambridge University Press

Proof. To prove the first statement, we apply induction on r. For $r = 1$, the claim follows directly from (1.2). For $r \geqslant 2$, we have

$$\begin{aligned}\overline{m}_2^r(F) &= \max_{H_r \subseteq F} \frac{e(H_r)}{v(H_r) - 2 + 1/\overline{m}_2^{r-1}(F)} \\ &\overset{\text{ind.}}{=} \max_{H_1,\dots,H_r \subseteq F} \frac{e(H_r)}{v(H_r) - 2 + \frac{2+\sum_{i=1}^{r-1}(v(H_i)-2)\prod_{j=1}^{i-1} e(H_j)}{\prod_{i=1}^{r-1} e(H_i)}} \\ &= \max_{H_1,\dots,H_r \subseteq F} \frac{\prod_{i=1}^{r} e(H_i)}{2 + \sum_{i=1}^{r}(v(H_i) - 2)\prod_{j=1}^{i-1} e(H_j)}.\end{aligned}$$

The above calculation shows that the graphs $H_1, \dots, H_r$ maximizing (2.5) are the ones defined in (1.4). Moreover, if F is 2-balanced, by Lemma 2.2 it is also balanced w.r.t. $\overline{d}_2^r$, $r \geqslant 1$. Plugging $H_1 = \cdots = H_r = F$ into (2.5), we thus obtain

$$\begin{aligned}2 - \frac{1}{\overline{m}_2^r(F)} &= 2 - \frac{2 + (v_F - 2)\sum_{j=0}^{r-1} e_F^j}{e_F^r} \\ &= 2 - \frac{2 + (v_F - 2)\frac{e_F^r - 1}{e_F - 1}}{e_F^r} \\ &= \left(2 - \frac{v_F - 2}{e_F - 1}\right)\left(1 - e_F^{-r}\right).\end{aligned}$$

□

Corollary 1.3 now follows immediately from Theorem 1.2 and Lemma 2.5.

3. Proof of Theorem 1.2

Consider the game with two colours, say red and blue, and assume that Painter uses one colour, say red, in every move if this does not create an entirely red copy of F. Clearly, the game will end with a blue copy of F, which was forced by a surrounding red structure. More precisely, when the game is over, $G(n, N)$ contains a blue copy of F, each edge of which completes a red subgraph to a copy of F. We say that the graph formed by these copies is a dangerous graph for Painter. Figure 1 shows two such dangerous graphs for the case $F = K_4$.

This simple greedy strategy yields the claimed lower bound if F is balanced w.r.t. $\overline{d}_2^2$. For general graphs, it may be smarter to play the greedy strategy with respect to an appropriately chosen subgraph H of F. For an example, consider the graph F consisting of a triangle with one edge attached to it. Here it turns out that greedily avoiding triangles and forgetting about the additional edge guarantees longer survival than greedily avoiding F itself. If Painter follows this greedy strategy for a fixed $H \subseteq F$, the game ends with a blue copy of F, each edge of which completes a red subgraph to a copy of H.

For arbitrary non-empty graphs H_1 and H_2, let $\mathcal{F}(H_1, H_2)$ denote the class of all graphs that have an 'inner' (blue) copy of H_1, each edge of which also completes an 'outer' (red) copy of H_2. Here the colours should only provide the intuitive connection to the greedy strategy: the members of the family $\mathcal{F}(H_1, H_2)$ are not associated with a colouring. We say that the inner copy of H_1 is formed by inner vertices and edges, and refer to the surrounding elements as outer vertices and edges respectively. Formally, we define the family of graphs $\mathcal{F}(H_1, H_2)$ as follows.

https://doi.org/10.1017/S0963548308009632 Published online by Cambridge University Press

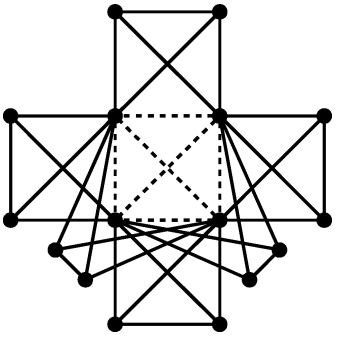


Figure 1. Two graphs from the class $\mathcal{F}(K_4, K_4)$, which are dangerous to a greedy Painter in the K_4-avoidance game with two colours. The graph on the left-hand side is the unique graph K_4^*. In both graphs, the inner edges are drawn dashed.

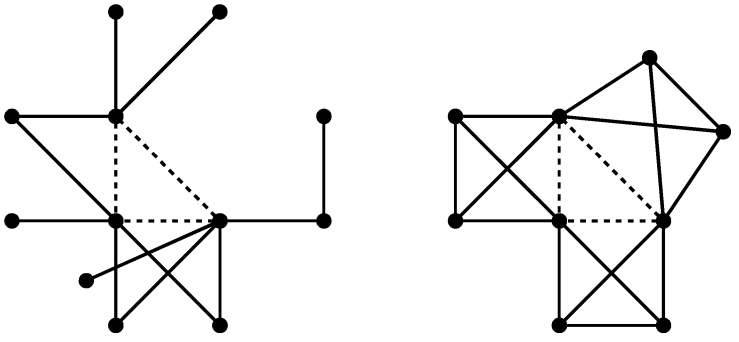


Figure 2. Two graphs from the class $\mathcal{F}(C_4, C_4)$. Note that the inner vertex v in the right graph also serves as an outer vertex in the outer copy associated with the edge f.

Definition 1. For all graphs $H_1 = (V, E)$ and H_2, let

$$\mathcal{F}(H_1, H_2) := \Big\{ F' = (V \mathbin{\dot\cup} U, E \mathbin{\dot\cup} D) : F' \text{ is a } \textit{minimal} \text{ graph such that for all } f \in E \text{ there are sets } U(f) \subseteq V \mathbin{\dot\cup} U \text{ and } D(f) \subseteq D \text{ with } \big(f \mathbin{\dot\cup} U(f), \{f\} \mathbin{\dot\cup} D(f)\big) \cong H_2 \Big\}.$$

The inner vertices V and edges E form the inner copy of H_1. Every edge $f \in E$ together with $U(f)$ and $D(f)$ forms a copy of H_2. Hence, $|U(f)| = v(H_2) - 2$ and $|D(f)| = e(H_2) - 1$. Note that inner vertices $v \in V$ may also serve as outer vertices $U(f)$ for a non-incident inner edge f, as illustrated in Figure 2. Therefore, every outer copy contains exactly one inner edge and two or more inner vertices. We take F' as a minimal element with respect to subgraph inclusion, *i.e.*, F' does not have a subgraph which satisfies the same properties. This ensures in particular that $\mathcal{F}(H_1, H_2)$ is finite.

In the game with three colours, say, with colours yellow (3), red (2) and blue (1), a greedy Painter first avoids some subgraph H_3 in yellow, H_2 in red, and eventually H_1 in blue. We call this strategy the greedy $\langle H_3, H_2, H_1 \rangle$-avoidance strategy. By the same argument as before, when

https://doi.org/10.1017/S0963548308009632 Published online by Cambridge University Press

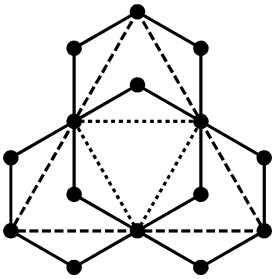


Figure 3. The unique graph from the class $\mathcal{F}(K_3, K_3, K_3)$ in which no outer copies overlap.

the game is over, the board contains a red–blue copy of a member from the family $\mathcal{F}(H_1, H_2)$, each edge of which completes an entirely yellow copy of H_3. We denote the class of all such graphs by $\mathcal{F}(H_1, H_2, H_3)$. Figure 3 depicts a graph from the class $\mathcal{F}(K_3, K_3, K_3)$.

This motivates the following inductive definition for general r.

Definition 2. For any graph H_1, let $\mathcal{F}(H_1) := \{H_1\}$. For $r \geqslant 2$ and any graphs $H_1, \dots, H_r$, let

$$\mathcal{F}(H_1, \dots, H_r) := \{F^r \in \mathcal{F}(F^{r-1}, H_r) : F^{r-1} \in \mathcal{F}(H_1, \dots, H_{r-1})\}.$$

By the same argument as before, if Painter uses the greedy $\langle H_r, \dots, H_1 \rangle$-avoidance strategy in the game with r colours, at the end of the game the board will contain a copy of a graph from $\mathcal{F}(H_1, \dots, H_r)$. With this observation at hand, Theorem 1.2 is an immediate consequence of the next, purely deterministic lemma. Its proof is postponed to Section 4.

Lemma 3.1. *Let F be a graph that is not a forest, and let $r \geqslant 1$. If the subgraphs $H_1, \dots, H_r \subseteq F$ are chosen according to* (1.4)*, all graphs $F^r \in \mathcal{F}(H_1, \dots, H_r)$ satisfy*

$$m(F^r) \geqslant \overline{m}_2^r(F).$$

Proof of Theorem 1.2. Since the class $\mathcal{F}(H_1, \dots, H_r)$ is finite, Lemma 3.1 implies by Theorem 1.5 that $G(n, N)$ a.a.s. contains no graph from this class as long as $N \ll n^{2-1/\overline{m}_2^r(F)}$. Therefore, Painter will a.a.s. not lose the game if she follows the greedy $\langle H_r, \dots, H_1 \rangle$-avoidance strategy. □

We conclude this section by giving some intuition for Lemma 3.1. Consider again the case $r = 2$. Among the graphs $F' \in \mathcal{F}(H_1, H_2)$, there are some distinguished ones F^* in which no outer copies overlap. Such graphs are depicted in Figures 1 and 2 respectively on the left-hand side. Clearly, these graphs have exactly $e(H_1)(v(H_2) - 2) + v(H_1)$ vertices and $e(H_1)(e(H_2) - 1) + e(H_1) = e(H_1)e(H_2)$ edges. If H_1 and H_2 are chosen according to (1.4), this yields

$$d(F^*) = \frac{e(H_1)e(H_2)}{e(H_1)(v(H_2) - 2) + v(H_1)} = \frac{e(H_2)}{v(H_2) - 2 + 1/d(H_1)} = \overline{m}_2^2(F).$$

More generally, the density of a 'nice' member of the class $\mathcal{F}(H_1, \dots, H_r)$ with no overlapping outer copies is given by the fraction on the right-hand side of (2.5) (see Figure 3 for an example). Thus the statement of Lemma 3.1 is essentially that members of the family $\mathcal{F}(H_1, \dots, H_r)$ that

https://doi.org/10.1017/S0963548308009632 Published online by Cambridge University Press

contain overlapping substructures are at least as dense as the 'nice' members in which no outer copies overlap. While this seems intuitively clear in many cases, the proof of Lemma 3.1 is surprisingly technical.

4. Proof of Lemma 3.1

We shall prove Lemma 3.1 by induction on r using Lemma 4.1.

Lemma 4.1. *Let F be a graph that is not a forest, and let $r \geqslant 2$. Let S be a graph satisfying $m(S) \geqslant \overline{m}_2^{r-1}(F)$ if $r \geqslant 3$, and $m(S) = m(F)$ if $r = 2$. If the subgraph $H = H_r \subseteq F$ is chosen according to* (1.4)*, all graphs $S' \in \mathcal{F}(S, H)$ satisfy*

$$m(S') \geqslant \overline{m}_2^r(F).$$

With Lemma 4.1 at hand, Lemma 3.1 follows easily.

Proof of Lemma 3.1. We proceed by induction on r. For $r = 1$, we have $\mathcal{F}(F) = \{F\}$ and $m(F) = \overline{m}_2^1(F)$ by definitions. For $r = 2$, we have by choice of H_1 that $m(H_1) = m(F)$, and Lemma 4.1 implies Lemma 3.1 with $S \leftarrow H_1$ and $H \leftarrow H_2$.

Now suppose we have $r \geqslant 3$, and let F^r be any graph from $\mathcal{F}(H_1, H_2, \ldots, H_r)$. Then by definition there is a graph $F^{r-1} \in \mathcal{F}(H_1, H_2, \ldots, H_{r-1})$ such that $F^r \in \mathcal{F}(F^{r-1}, H_r)$, and we have

$$m(F^{r-1}) \geqslant \overline{m}_2^{r-1}(F)$$

by the induction hypothesis. Applying Lemma 4.1 with $S \leftarrow F^{r-1}$ and $H \leftarrow H_r$ yields the statement of Lemma 3.1.

For the proof of Lemma 4.1, we need the following technical lemma. Unfortunately, there are two exceptional cases where its statement does not hold and which we will have to consider separately.

Lemma 4.2. *Let F be a graph that is not a forest, and let $r \geqslant 2$. Let S be a graph satisfying $\overline{m}_2^{r-1}(F) \leqslant d(S) \leqslant \overline{m}_2^r(F)$ if $r \geqslant 3$, and $d(S) = m(F)$ if $r = 2$. Then we have*

$$\overline{m}_2^r(F) - \lfloor 2d(S) \rfloor \left(\frac{\overline{m}_2^r(F)}{\overline{m}_2^{r-1}(F)} - 1 \right) - 1 \geqslant 0, \tag{4.1}$$

unless one of the following exceptional cases occurs:

(a) *$r = 2$, and* (1.4) *yields $H_2 = H_2(F) = K_4$,*
(b) *$d(S) < 3/2$, and* (1.4) *yields a graph $H_r = H_r(F)$ with $e(H_r) \leqslant v(H_r) + 1$.*

Proof of Lemma 4.2. Throughout, H denotes the graph H_r given by (1.4). By Lemma 2.4 we have $e_H \geqslant v_H \geqslant 3$.

Observe that $d(S) \geqslant \overline{m}_2^{r-1}(F) \geqslant m(F) \geqslant 1$. We distinguish the three cases:

(i) $d(S) \geqslant 3/2$ and $r \geqslant 3$,
(ii) $d(S) \geqslant 3/2$ and $r = 2$, and
(iii) $1 \leqslant d(S) < 3/2$.

https://doi.org/10.1017/S0963548308009632 Published online by Cambridge University Press

(i) Suppose $d(S) \geqslant 3/2$ and $r \geqslant 3$. Using that $d(S) \leqslant \overline{m}_2^r(F)$ and dropping the Gauss bracket, we see that (4.1) holds if

$$\overline{m}_2^r(F)\left(3 - 2\frac{\overline{m}_2^r(F)}{\overline{m}_2^{r-1}(F)}\right) - 1 \geqslant 0. \tag{4.2}$$

For $e_H \geqslant v_H \geqslant 4$, we have due to $\overline{m}_2^{r-1}(F) \geqslant \overline{m}_2^2(F) \geqslant \overline{d}_2^2(F,H)$ that

$$\begin{aligned}
\frac{\overline{m}_2^r(F)}{\overline{m}_2^{r-1}(F)} &= \frac{e_H}{(v_H-2)\overline{m}_2^{r-1}(F)+1} \leqslant \frac{e_H}{(v_H-2)\overline{d}_2^2(F,H)+1} \\
&= \frac{e_H}{(v_H-2)\frac{e_H m(F)}{(v_H-2)m(F)+1}+1} \\
&= 1 + \frac{e_H-(v_H-2)m(F)-1}{(e_H+1)(v_H-2)m(F)+1} \\
&\leqslant 1 + \frac{e_H-(v_H-2)\frac{e_H}{v_H}-1}{(e_H+1)(v_H-2)\frac{e_H}{v_H}+1} \\
&= 1 + \frac{2e_H-v_H}{e_H(e_H+1)(v_H-2)+v_H} \leqslant 7/6.
\end{aligned}$$

Thus, the parenthesis in (4.2) is at least 2/3, and (4.2) holds due to $\overline{m}_2^r(F) \geqslant d(S) \geqslant 3/2$. For $H = K_3$, plugging $\overline{m}_2^r(F) = 3/(1 + 1/\overline{m}_2^{r-1}(F))$ into (4.2) yields the condition

$$8\overline{m}_2^{r-1}(F)^2 - 11\overline{m}_2^{r-1}(F) - 1 \geqslant 0,$$

which is easily checked using that $\overline{m}_2^{r-1}(F) \geqslant \overline{m}_2^2(H) = 3/2$ due to $H \subseteq F$.

(ii) Suppose $d(S) \geqslant 3/2$ and $r = 2$. Using that $d(S) = m(F)$ by assumption and dropping the Gauss bracket, we see that (4.1) holds if

$$\overline{m}_2^2(F) \leqslant 2m(F) - 1. \tag{4.3}$$

Using

$$\overline{m}_2^2(F) = \frac{e_H}{v_H - 2 + 1/m(F)} \leqslant \frac{v_H m(F)}{v_H - 2 + 1/m(F)},$$

we deduce that (4.3) holds if

$$v_H \geqslant 4 + \frac{1}{m(F)(m(F)-1)}, \tag{4.4}$$

which due to $m(F) \geqslant 3/2$ is satisfied for $v_H \geqslant 6$. Using that $\overline{m}_2^2(F) = e_H/(v_H - 2 + 1/m(F))$ and expanding (4.3) yields

$$e_H \leqslant 2m(F)(v_H - 2) - 1/m(F) + 4 - v_H.$$

Since $m(F) \geqslant \max\{3/2, e_H/v_H\}$, the claim follows if either

$$e_H \leqslant 2v_H - 8/3$$

or

$$e_H - 4e_H/v_H - v_H/e_H + 4 - v_H \geqslant 0$$

https://doi.org/10.1017/S0963548308009632 Published online by Cambridge University Press

is satisfied. All graphs with $v_H \leqslant 5$ and $e_H \geqslant v_H$ satisfy one of these equations, except H with $v_H = 5$, $e_H = 8$ and $H = K_4$. In the first case, if $m(F) \geqslant 2$, the claim follows from (4.4), and otherwise with $m(F) \geqslant e_H/v_H = 8/5$ and $\lfloor 2d(S) \rfloor = \lfloor 2m(F) \rfloor = 3$ directly from (4.1) with elementary calculations. The second case is listed as exceptional case (a).

(iii) Suppose $1 \leqslant d(S) < 3/2$**.** Observing that $\lfloor 2d(S) \rfloor = 2$, we see that (4.1) holds if

$$\overline{m}_2^r(F) \leqslant \frac{\overline{m}_2^{r-1}(F)}{2 - \overline{m}_2^{r-1}(F)}.$$

Plugging in that $\overline{m}_2^r(F) = e_H/(v_H - 2 + 1/\overline{m}_2^{r-1}(F))$, this reduces to

$$\overline{m}_2^{r-1}(F) \geqslant \frac{2e_H - 1}{e_H + v_H - 2},$$

Since $\overline{m}_2^{r-1}(F) \geqslant e_H/v_H$, the last inequality holds if

$$\frac{e_H}{v_H} \geqslant \frac{2e_H - 1}{e_H + v_H - 2}$$

or, equivalently,

$$e_H(e_H - v_H - 2) + v_H \geqslant 0,$$

which holds if $e_H \geqslant v_H + 2$. The remaining possibilities are listed as exceptional case (b). □

With Lemma 4.2 at hand, we are ready to prove Lemma 4.1.

Proof of Lemma 4.1. Note that we may assume that $m(S) \leqslant \overline{m}_2^r(F)$, as otherwise the claim follows trivially from $S \subseteq S'$. Next, we argue that it suffices to prove Lemma 4.1 for the case when S is balanced. Suppose it is not, and let $G \subset S$ be a (balanced) subgraph satisfying $d(G) = m(S)$. If S satisfies the assumption $m(S) \geqslant \overline{m}_2^{r-1}(F)$, so does G. Moreover, every graph $S' \in \mathcal{F}(S, H)$ contains a subgraph $G' \in \mathcal{F}(G, H)$. Therefore, if Lemma 4.1 holds for balanced graphs, it follows that

$$m(S') \geqslant m(G') \overset{\text{Lem. 4.1}}{\geqslant} \overline{m}_2^r(F).$$

For the rest of the proof, we assume that S is balanced with $\overline{m}_2^{r-1}(F) \leqslant d(S) \leqslant \overline{m}_2^r(F)$. Note that the preconditions of Lemma 4.2 reflect these assumptions. In order to prove $m(S') \geqslant \overline{m}_2^r(F)$ for all $S' \in \mathcal{F}(S, H)$, we show the slightly stronger assertion $d(S') \geqslant \overline{m}_2^r(F)$.

We now give a brief outline of the main ideas of our proof. Let S' be any graph from the family $\mathcal{F}(S, H)$. Then there exists a graph S^* from the same family that has non-overlapping outer copies of H (as on the left-hand side of Figures 1 and 2) and can be transformed into S' by merging those outer copies in $e(S)$ steps as follows. We fix some linear ordering on the set of inner edges $E = E(S)$. For every edge $f \in E$, we merge the outer copy $\widehat{H}(f)$ of H associated with f to outer copies associated with those edges preceding f. Throughout the entire process, we keep track of the number of edges $\Delta_e(f)$ and vertices $\Delta_v(f)$ vanishing in each step. If one could show that $d(S') \geqslant d(S^*)$ by proving that the density of the graph is non-decreasing in every step of the process, the proof would be finished, since we have

$$d(S^*) = \frac{e_S e_H}{e_S(v_H - 2) + v_S} = \frac{e_H}{v_H - 2 + 1/d(S)} \geqslant \frac{e_H}{v_H - 2 + 1/\overline{m}_2^{r-1}(F)} = \overline{m}_2^r(F) \qquad (4.5)$$

https://doi.org/10.1017/S0963548308009632 Published online by Cambridge University Press

by our assumptions on S. To show that the density of S^* is indeed non-decreasing in this process, it would suffice to show that, in each step, the ratio of vanishing edges to vanishing vertices is bounded by $d(S^*)$, *i.e.*, that

$$\frac{\Delta_e(f)}{\Delta_v(f)} \leqslant d(S^*)$$

for all $f \in E$ (*cf.* Proposition 2.3). Unfortunately, this monotonicity does not hold in general. However, we are able to rescue this proof strategy by showing that there are sufficiently many 'good' steps in this process that can amortize all 'bad' ones which may arise. Essentially, we group the edges $f \in E$ into suitable 'phases' such that the ratio of vanished edges to vanished vertices *in each phase* is bounded by $d(S^*)$.

Additional technical difficulties arise from the fact that outer vertices may be glued to *inner* vertices, as depicted in Figure 2. Indeed, this complication might spoil the amortization inside certain phases, and we have to compensate such 'irregular' phases differently, amortizing also between phases. In the following we make these ideas precise.

Recall that in every graph $S' = (V \mathbin{\dot\cup} U, E \mathbin{\dot\cup} D)$, the inner copy (V, E) is isomorphic to S. By definition, for each inner edge $f \in E$, we can identify sets of outer vertices $U(f) \subseteq V \mathbin{\dot\cup} U$ and outer edges $D(f) \subseteq D$ such that $\widehat{H}(f) := (f \mathbin{\dot\cup} U(f), \{f\} \mathbin{\dot\cup} D(f))$ is isomorphic to H. While these sets are not necessarily unique, for the rest of the proof we fix one choice of appropriate sets $U(f)$ and $D(f)$. Note that by the minimality condition in Definition 1, every vertex and edge of S' is included in at least one outer copy. Let $\widehat{H}_-(f) := (f \mathbin{\dot\cup} U(f), D(f))$ denote the subgraph that is obtained by removing f from $\widehat{H}(f)$. For every vertex $u \in V \mathbin{\dot\cup} U$ and for every outer edge $d \in D$ of S', the sets

$$E(u) := \{f \in E : u \in U(f)\}$$

and

$$E(d) := \{f \in E : d \in D(f)\}$$

indicate in which outer copies it participates.

Note that

$$\sum_{d \in D} |E(d)| = \sum_{d \in D} \sum_{\substack{f \in E: \\ d \in D(f)}} 1 = \sum_{f \in E} \sum_{\substack{d \in D: \\ d \in D(f)}} 1 = \sum_{f \in E} |D(f)| = e_S(e_H - 1)$$

and analogously

$$\sum_{u \in U \dot\cup V} |E(u)| = e_S(v_H - 2).$$

Recall that S^* has $e_S e_H$ edges and $v_S + e_S(v_H - 2)$ vertices. Due to

$$\begin{aligned} e(S') - |D| = e_S &= e(S^*) - e_S(e_H - 1) \\ &= e(S^*) - \sum_{d \in D} |E(d)|, \end{aligned}$$

https://doi.org/10.1017/S0963548308009632 Published online by Cambridge University Press

we have

$$e(S') = e(S^*) - \sum_{d \in D} \big(|E(d)| - 1\big). \tag{4.6}$$

Analogously,

$$\begin{aligned} v(S') - |U| = v_S &= v(S^*) - e_S(v_H - 2) \\ &= v(S^*) - \sum_{u \in U \dot\cup V} |E(u)| \end{aligned}$$

yields

$$v(S') = v(S^*) - \sum_{u \in U} \big(|E(u)| - 1\big) - \sum_{v \in V} |E(v)|. \tag{4.7}$$

Next, we impose a linear order on the vertices and edges of S. For ease of notation we will use the abbreviation $\ell := v_S$ in the remainder of the proof. Using the averaging principle, it is easy to see that for every balanced graph S there exists an ordering $[v_1, \dots, v_\ell]$ of its vertices such that for all $2 \leqslant i \leqslant \ell$, v_i has at most $\lfloor 2d(S) \rfloor$ neighbours among $\{v_1, \dots, v_{i-1}\}$ in S. Without loss of generality, we assume that the inner vertices V are ordered in this way. This ordering induces a mapping $p : E \to \{2, \dots, \ell\}$, which assigns every inner edge to the greater label of its two vertices. We call $p(f)$ the phase of edge f. This mapping induces a partial order on E, which can be extended to a total order $\prec$ by choosing an arbitrary order on edges belonging to the same phase.

For $f \in E$, we define

$$\Delta_E(f) := D(f) \cap \left(\bigcup_{f' \prec f} D(f') \right) \tag{4.8}$$

$$\Delta_V(f) := U(f) \cap \left(\bigcup_{f' \prec f} U(f') \cup \{v_1, \dots, v_{p(f)}\} \right) \tag{4.9}$$

$$\Phi(f) := U(f) \cap \left(\{v_{p(f)+1}, \dots, v_\ell\} \setminus \bigcup_{f' \prec f} U(f') \right)$$

and $\Delta_e(f) := |\Delta_E(f)|$ and $\Delta_v(f) := |\Delta_V(f)|$ for the cardinalities of those sets. Intuitively, $\Delta_e(f)$ is the number of edges vanishing when $\widehat{H}_-(f)$ is merged with preceding outer copies and possibly inner vertices. Analogously, $\Delta_v(f) + |\Phi(f)|$ is the number of vertices vanishing in this merge operation. We introduce the sets $\Phi(f)$ to deal with the problem that vertices of outer copies can be merged onto inner vertices. $\Phi(f)$ is the set of inner vertices v_j with $j > p(f)$ which are merged with an outer vertex for the *first* time in step f.

$\Delta_E(f)$ contains all edges $d \in D(f)$ that also belong to $D(f')$ for some edge $f' \prec f$. By definition, both f and f' are in $E(d)$. Therefore, we have

$$\sum_{f \in E} \Delta_e(f) = \sum_{f \in E} \sum_{\substack{d \in D: \\ d \in D(f) \\ f \neq \min E(d)}} 1 = \sum_{d \in D} \sum_{\substack{f \in E: \\ d \in D(f) \\ f \neq \min E(d)}} 1 = \sum_{d \in D} |E(d) - 1|,$$

https://doi.org/10.1017/S0963548308009632 Published online by Cambridge University Press

and by (4.6),

$$e(S') = e(S^*) - \sum_{f \in E} \Delta_e(f). \tag{4.10}$$

Similarly, it follows from

$$\Delta_v(f) = \sum_{\substack{u \in U: \\ u \in U(f) \\ f \neq \min E(u)}} 1 + \sum_{\substack{v \in \{v_{p(f)+1}, \dots, v_\ell\}: \\ v \in U(f) \\ f \neq \min E(v)}} 1 + \sum_{\substack{v \in \{v_1, \dots, v_{p(f)}\}: \\ v \in U(f)}} 1$$

and

$$|\Phi(f)| = \sum_{\substack{v \in \{v_{p(f)+1}, \dots, v_\ell\}: \\ v \in U(f) \\ f = \min E(v)}} 1$$

that

$$\Delta_v(f) + |\Phi(f)| = \sum_{\substack{u \in U: \\ u \in U(f) \\ f \neq \min E(u)}} 1 + \sum_{\substack{v \in V: \\ v \in U(f)}} 1$$

and thus

$$\sum_{f \in E} \Delta_v(f) + \sum_{f \in E} |\Phi(f)| = \sum_{u \in U} |E(u) - 1| + \sum_{v \in V} |E(u)|.$$

By (4.7), we obtain

$$v(S') = v(S^*) - \sum_{f \in E} \Delta_v(f) - \sum_{f \in E} |\Phi(f)|. \tag{4.11}$$

Note that the sets $\Phi(f) \subseteq V$ are disjoint because $v_i \in \Phi(f)$ implies that $v_i \in U(f)$, which prevents v_i from being contained in $\Phi(g)$ for all $g \succ f$. Recall also that $\Phi(f)$ contains those vertices v_j with $j > p(f)$ that are merged with an outer vertex for the first time in step f. Their union

$$\Phi := \dot{\bigcup}_{f \in E} \Phi(f)$$

contains all vertices $v_j \in V$ that form part of some outer copy of an edge f with $p(f) < j$, *i.e.*,

$$\Phi = \left\{ v_j \in V : v_j \in \bigcup_{\substack{f \in E: \\ p(f) < j}} U(f) \right\}.$$

In order to calculate the density of S', we introduce the following quantities. For every phase i, we define

$$\Delta_e^i := \sum_{\substack{f \in E: \\ p(f) = i}} \Delta_e(f),$$

https://doi.org/10.1017/S0963548308009632 Published online by Cambridge University Press

and analogously

$$\Delta_v^i := \sum_{\substack{f \in E: \\ p(f)=i}} \Delta_v(f).$$

Due to (4.10) and (4.11), we can express the density of S' simply as

$$\frac{e(S')}{v(S')} = \frac{e(S^*) - \Delta_e^2 - \cdots - \Delta_e^\ell}{v(S^*) - \Delta_v^2 - \cdots - \Delta_v^\ell - |\Phi|}. \tag{4.12}$$

We will distinguish between regular and irregular phases. A phase i is regular if $v_i \notin \Phi$. Let $\overline{\Phi} := V \setminus \Phi$. We can rewrite (4.12) by grouping into regular and irregular phases:

$$\frac{e(S')}{v(S')} = \frac{e(S^*) - \sum_{v_i \in \overline{\Phi}} \Delta_e^i - \sum_{v_i \in \Phi} \Delta_e^i}{v(S^*) - \sum_{v_i \in \overline{\Phi}} \Delta_v^i - \sum_{v_i \in \Phi} (\Delta_v^i + 1)}. \tag{4.13}$$

We call phase i *trivial* if $\Delta_v^i = 0$, which implies $\Delta_e^i = 0$. Applying Proposition 2.3 to (4.13), it follows with (4.5) that

$$d(S') \geqslant \overline{m}_2^r(F)$$

if the next two claims hold.

Claim 4.3. *In every non-trivial, regular phase i, we have*

$$\frac{\Delta_e^i}{\Delta_v^i} \leqslant \overline{m}_2^r(F).$$

Claim 4.4. *In every non-trivial, irregular phase i, we have*

$$\frac{\Delta_e^i}{\Delta_v^i + 1} \leqslant \overline{m}_2^r(F).$$

In order to prove these claims, we derive bounds on $\Delta_e(f)$ by carefully examining appropriate graphs with edge set $\Delta_E(f)$. For every edge $f \in E$, let

$$T(f) := (f \mathbin{\dot\cup} \Delta_V(f), \Delta_E(f)).$$

This is a well-defined graph since the edges $\Delta_E(f)$ have endpoints in

$$\begin{aligned}
&\big(f \mathbin{\dot\cup} U(f)\big) \cap \left(\bigcup_{f' \prec f} \big(f' \mathbin{\dot\cup} U(f')\big) \right) \\
&\quad \subseteq \big(f \mathbin{\dot\cup} U(f)\big) \cap \left(\bigcup_{f' \prec f} U(f') \cup \{v_1, \ldots, v_{p(f)}\} \right) \\
&\quad = f \mathbin{\dot\cup} \Delta_V(f)
\end{aligned}$$

(*cf.* (4.8) and (4.9)).

We shall use that $T(f)$ is a subgraph of $\widehat{H}(f) \cong H$ to derive an easy upper bound on $\Delta_e(f)$. Since the graph $T^+(f)$ obtained by adding the edge f to $T(f)$ is still a subgraph of $\widehat{H}(f) \cong H \subseteq$

https://doi.org/10.1017/S0963548308009632 Published online by Cambridge University Press

F, we use that $\overline{d}_2^r(F, T^+(f)) \leqslant \overline{m}_2^r(F)$ to obtain that

$$\begin{aligned}\Delta_e(f) &= e(T^+(f)) - 1\\ &\leqslant \overline{m}_2^r(F)\big(v(T^+(f)) - 2 + 1/\overline{m}_2^{r-1}(F)\big) - 1\\ &= \overline{m}_2^r(F)\Delta_v(f) + \overline{m}_2^r(F)/\overline{m}_2^{r-1}(F) - 1.\end{aligned} \tag{4.14}$$

This simple bound already suffices to prove Claim 4.4.

Proof of Claim 4.4. Recall that phase i consists of at most $\lfloor 2d(S)\rfloor$ merge operations. Using this, we derive from (4.14) that

$$\begin{aligned}\Delta_e^i &= \sum_{f\in E, p(f)=i} \Delta_e(f)\\ &\leqslant \sum_{f\in E, p(f)=i} \left(\overline{m}_2^r(F)\Delta_v(f) + \frac{\overline{m}_2^r(F)}{\overline{m}_2^{r-1}(F)} - 1\right)\\ &= \overline{m}_2^r(F)\Delta_v^i + |\{f \in E, p(f) = i\}|\left(\frac{\overline{m}_2^r(F)}{\overline{m}_2^{r-1}(F)} - 1\right)\\ &\leqslant \overline{m}_2^r(F)(\Delta_v^i + 1) + \lfloor 2d(S)\rfloor\left(\frac{\overline{m}_2^r(F)}{\overline{m}_2^{r-1}(F)} - 1\right) - \overline{m}_2^r(F).\end{aligned}$$

Claim 4.4 now follows if

$$\overline{m}_2^r(F) - \lfloor 2d(S)\rfloor\left(\frac{\overline{m}_2^r(F)}{\overline{m}_2^{r-1}(F)} - 1\right) \geqslant 0, \tag{4.15}$$

which is slightly weaker than the statement of Lemma 4.2. Therefore, the claim follows unless one of the exceptional cases in Lemma 4.2 occurs.

If exceptional case (a) occurs, *i.e.*, $r = 2$ and $H = K_4$, using $d(S) = m(F)$ and dropping the Gauss bracket, (4.15) reduces to $\overline{m}_2^2(F) \leqslant 2m(F)$, which is satisfied due to $\overline{m}_2^2(F) = 6/(2 + 1/m(F)) \leqslant 3 \leqslant 2m(F)$.

If exceptional case (b) occurs, *i.e.*, $d(S) < 3/2$ and $e_H \leqslant v_H + 1$, we proceed similarly to the calculations in the proof of case (iii) in Lemma 4.2. Observing that $\lfloor 2d(S)\rfloor = 2$, we see that (4.1) holds if

$$\overline{m}_2^r(F) \leqslant \frac{2\overline{m}_2^{r-1}(F)}{2 - \overline{m}_2^{r-1}(F)}.$$

Plugging in that $\overline{m}_2^r(F) = e_H/(v_H - 2 + 1/\overline{m}_2^{r-1}(F))$, this reduces to

$$\overline{m}_2^{r-1}(F) \geqslant \frac{2e_H - 2}{e_H + 2v_H - 4}.$$

Since $\overline{m}_2^{r-1}(F) \geqslant e_H/v_H$, the last inequality holds if

$$\frac{e_H}{v_H} \geqslant \frac{2e_H - 2}{e_H + 2v_H - 4}$$

or, equivalently,

https://doi.org/10.1017/S0963548308009632 Published online by Cambridge University Press

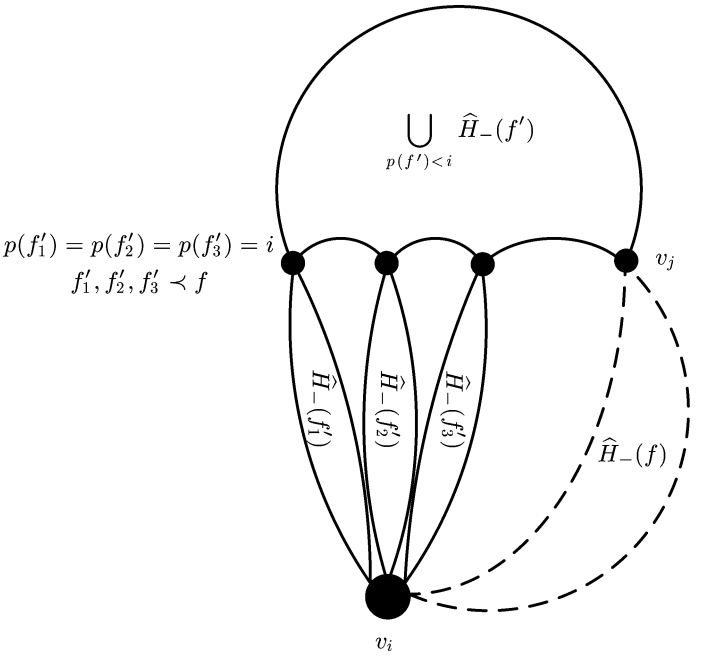


Figure 4. The structure of $G(f)$ in the case where $f = \{v_j, v_i\}$ is the first edge in phase i with $\Delta_v(f) > 0$. The outer copy $\widehat{H}_-(f)$ is drawn dashed because it is not part of $G(f)$. Intuitively, the graph $T(f)$ depicted in the next figure indicates how it is merged onto $G(f)$. Here it is shown in its non-merged position.

$$e_H(e_H - 4) + 2v_H \geqslant 0,$$

which holds for all graphs H. This concludes the proof of Claim 4.4. □

Proof of Claim 4.3. To deal with regular phases, we derive a better bound on $\Delta_e(f)$ than (4.14) for a suitably chosen edge f in phase i. Note that for $g \in E$, $\Delta_v(g) = 0$ implies $\Delta_e(g) = 0$. In the following, let $f = \{v_j, v_i\}$, $j < i$, be the first edge in phase i according to the ordering $\prec$ satisfying $\Delta_v(f) > 0$. That is, no merging took place in the previous steps of this phase, and therefore the outer copies of H associated with the edges preceding f in phase i are still untouched. Then

$$G(f) := (\{v_1, \ldots, v_i\}, \emptyset) \cup \bigcup_{f' \prec f} \widehat{H}_-(f'),$$

the union of all outer copies associated with edges preceding f, has the structure shown in Figure 4. This structure enables us to obtain a stronger bound on $\Delta_e(f)$ than (4.14).

Due to (4.8) and (4.9), we have $T(f) = \big(f \,\dot\cup\, \Delta_V(f), \Delta_E(f)\big) = \widehat{H}_-(f) \cap G(f)$. In particular, $T(f)$ is a subgraph of $G(f)$ and inherits the structure of $G(f)$. Therefore, $T(f)$ looks as sketched in Figure 5. This observation gives rise to the following decomposition. To simplify notation, we write T for $T(f)$ in the remainder of the proof. By definition, T contains the inner vertex v_i. We consider all edges preceding f in phase i, distinguishing them according to whether their other inner vertex is also contained in T or not. Let

$$\mathcal{A} := \{f' \prec f : p(f') = i \wedge f' \notin V(T)\}$$

https://doi.org/10.1017/S0963548308009632 Published online by Cambridge University Press

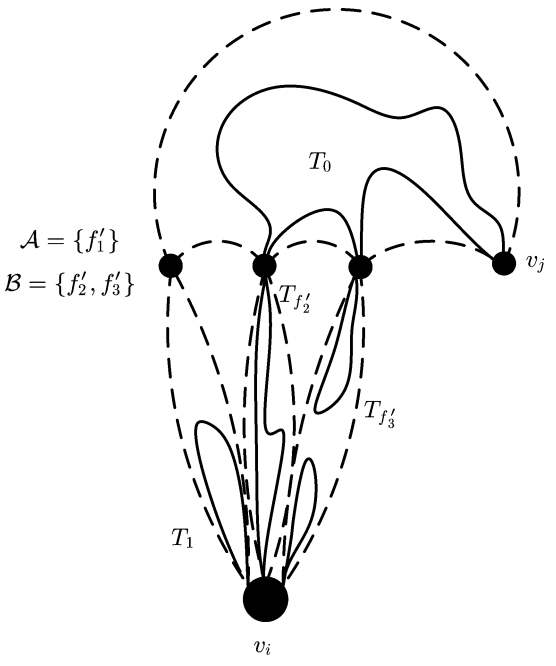


Figure 5. The decomposition of T into T_0, T_1, and $T_{f'}$, $f' \in \mathcal{B}$. The dashed lines indicate $G(f)$ (*cf.* Figure 4).

and

$$\mathcal{B} := \{f' \prec f : p(f') = i \wedge f' \subseteq V(T)\}.$$

We will show that

$$\Delta_e(f) \leqslant \overline{m}_2^r(F)\big(\Delta_v(f) - 1\big) + |\mathcal{B}|\left(\frac{\overline{m}_2^r(F)}{\overline{m}_2^{r-1}(F)} - 1\right) + \frac{\overline{m}_2^r(F)}{\overline{m}_2^{r-1}(F)}. \tag{4.16}$$

As $|\mathcal{B}|$ grows, this bound becomes worse, but at the same time, we have fewer edges that follow f in phase i. We will see that these two effects cancel out in our framework.

Recall that $f = \{v_j, v_i\}$ with $j < i$. We define the following subgraphs of T:

$$T_0 := \begin{cases} (\{v_j\}, \emptyset) & \text{if } \{f' \in E : p(f') < i\} = \emptyset, \\ T \cap \big[\bigcup_{p(f')<i} \widehat{H}_-(f')\big] & \text{otherwise,} \end{cases}$$

$$T_1 := \begin{cases} (\{v_i\}, \emptyset) & \text{if } \mathcal{A} = \emptyset, \\ T \cap \big[\bigcup_{f' \in \mathcal{A}} \widehat{H}_-(f')\big] & \text{otherwise,} \end{cases}$$

$$T_{f'} := T \cap \widehat{H}_-(f') \qquad \text{for all } f' \in \mathcal{B}.$$

Clearly, we have

$$T = T_0 \cup T_1 \cup \bigcup_{f' \in \mathcal{B}} T_{f'}.$$

https://doi.org/10.1017/S0963548308009632 Published online by Cambridge University Press

These definitions are illustrated in Figure 5. Since, by choice of f, all edges in $\mathcal{A}$ and $\mathcal{B}$ correspond to still untouched outer copies of H, the graphs T_0, T_1 and $T_{f'}$ are edge-disjoint. For the vertex sets, we have the following relations for all $f' = \{v_i, v_{j'}\}$ and $f'' = \{v_i, v_{j''}\}$ in $\mathcal{B}$:

$$\begin{aligned} V(T_0) \cap V(T_1) &= \emptyset, \\ V(T_0) \cap V(T_{f'}) &= \{v_{j'}\}, \\ V(T_1) \cap V(T_{f'}) &= \{v_i\}, \\ V(T_{f'}) \cap V(T_{f''}) &= \{v_i\}. \end{aligned} \tag{4.17}$$

These equations hold since we assumed that phase i is regular, which translates to $v_i \notin V(T_0)$. Since all graphs T_0, T_1, and $T_{f'}$, $f' \in \mathcal{B}$, are pairwise edge-disjoint, we conclude that

$$\Delta_e(f) = e(T) = e(T_0) + e(T_1) + \sum_{f' \in \mathcal{B}} e(T_{f'}). \tag{4.18}$$

Similarly, (4.17) implies that

$$v(T_0) + v(T_1) + \sum_{f' \in \mathcal{B}} v(T_{f'}) = v(T) + 2|\mathcal{B}| \tag{4.19}$$

since on the left-hand side each of the $|\mathcal{B}|$ vertices $v_{j'} \in \dot{\bigcup}_{f' \in \mathcal{B}} (f' \setminus \{v_i\})$ is counted exactly twice, v_i is counted exactly $|\mathcal{B}| + 1$ times, and all other vertices of T are counted only once.

Since T_0 and T_1 are subgraphs of $T \subseteq H \subseteq F$, it follows for $j = 0, 1$ from $\overline{d}_2^r(F, T_j) \leqslant \overline{m}_2^r(F)$ that we have

$$e(T_j) \leqslant \overline{m}_2^r(F)\big(v(T_j) - 2 + 1/\overline{m}_2^{r-1}(F)\big), \tag{4.20}$$

provided that T_j contains at least two vertices. Otherwise, we observe that both T_0 and T_1 contain at least one vertex by definition. Using that $\overline{m}_2^{r-1}(F) \geqslant m(F) \geqslant 1$, we deduce that

$$e(T_j) \leqslant \overline{m}_2^r(F)\big(v(T_j) - 1\big) \tag{4.21}$$

in all cases. Note that this bound is weaker than (4.20). Similarly, we obtain a bound on $e(T_{f'})$ for any fixed edge $f' \in \mathcal{B}$. Consider the graph $T_{f'}^+$ in which f' is added to $T_{f'}$. As $T_{f'}^+$ is a subgraph of $\widehat{H}(f') \cong H \subseteq F$, we obtain with $\overline{d}_2^r(F, T_{f'}^+) \leqslant \overline{m}_2^r(F)$ that

$$e(T_{f'}) \ = e(T_{f'}^+) - 1 \leqslant \overline{m}_2^r(F)\big(v(T_{f'}) - 2 + 1/\overline{m}_2^{r-1}(F)\big) - 1. \tag{4.22}$$

If $v(T_0) = 1$, we have $\mathcal{B} = \emptyset$, and if moreover $v(T_1) = 1$, the graph $T = T_0 \cup T_1 \cup \bigcup_{f' \in \mathcal{B}} T_{f'}$ contains no edges at all, and (4.16) is trivially true. We may therefore assume w.l.o.g. that either T_0 or T_1 contains at least two vertices. Denote that graph by T_j, $j \in \{0, 1\}$. Thus, combining (4.20), (4.21), and (4.22), we continue (4.18) and derive that

$$\begin{aligned} \Delta_e(f) \ &= \ e(T_0) + e(T_1) + \sum_{f' \in \mathcal{B}} e(T_{f'}) \\ &\leqslant \ \overline{m}_2^r(F)\big(v(T_j) - 2 + 1/\overline{m}_2^{r-1}(F)\big) + \overline{m}_2^r(F)\big(v(T_{1-j}) - 1\big) \\ &\quad + \sum_{f' \in \mathcal{B}} \big(\overline{m}_2^r(F)\big(v(T_{f'}^+) - 2 + 1/\overline{m}_2^{r-1}(F)\big) - 1\big) \end{aligned}$$

https://doi.org/10.1017/S0963548308009632 Published online by Cambridge University Press

$$
\begin{aligned}
&= \overline{m}_2^r(F)\left(v(T_0)+v(T_1)+\sum_{f'\in\mathcal{B}} v(T_{f'}) - 2|\mathcal{B}| - 3 + (|\mathcal{B}|+1)/\overline{m}_2^{r-1}(F)\right) - |\mathcal{B}| \\
&\overset{(4.19)}{=} \overline{m}_2^r(F)\big(v(T) - 3 + (|\mathcal{B}|+1)/\overline{m}_2^{r-1}(F)\big) - |\mathcal{B}| \\
&= \overline{m}_2^r(F)\big(\Delta_v(f) - 1 + (|\mathcal{B}|+1)/\overline{m}_2^{r-1}(F)\big) - |\mathcal{B}| \\
&= \overline{m}_2^r(F)\big(\Delta_v(f)-1\big) + |\mathcal{B}|\big(\overline{m}_2^r(F)/\overline{m}_2^{r-1}(F) - 1\big) + \overline{m}_2^r(F)/\overline{m}_2^{r-1}(F),
\end{aligned}
$$

as claimed in (4.16).

Claim 4.3 now follows with a similar calculation as in the proof of Claim 4.4. Recall that each phase consists of at most $\lfloor 2d(S)\rfloor$ merge operations. Let $\mathcal{G} := \{g \in E : g \succ f \wedge p(g) = i\}$, and note that $|\mathcal{B}| + |\mathcal{G}| \leqslant \lfloor 2d(S)\rfloor - 1$ since f belongs neither to $\mathcal{B}$ nor to $\mathcal{G}$. Using (4.14) and (4.16), we compute

$$
\begin{aligned}
\Delta_e^i &= \Delta_e(f) + \sum_{g\in\mathcal{G}} \Delta_e(g) \\
&\leqslant \overline{m}_2^r(F)\big(\Delta_v(f)-1\big) + |\mathcal{B}|\left(\frac{\overline{m}_2^r(F)}{\overline{m}_2^{r-1}(F)} - 1\right) + \frac{\overline{m}_2^r(F)}{\overline{m}_2^{r-1}(F)} \\
&\quad + \sum_{g\in\mathcal{G}}\left(\overline{m}_2^r(F)\Delta_v(g) + \frac{\overline{m}_2^r(F)}{\overline{m}_2^{r-1}(F)} - 1\right) \\
&= \overline{m}_2^r(F)\big(\Delta_v^i - 1\big) + (|\mathcal{B}|+|\mathcal{G}|)\left(\frac{\overline{m}_2^r(F)}{\overline{m}_2^{r-1}(F)} - 1\right) + \frac{\overline{m}_2^r(F)}{\overline{m}_2^{r-1}(F)} \\
&\leqslant \overline{m}_2^r(F)\big(\Delta_v^i - 1\big) + \big(\lfloor 2d(S)\rfloor - 1\big)\left(\frac{\overline{m}_2^r(F)}{\overline{m}_2^{r-1}(F)} - 1\right) + \frac{\overline{m}_2^r(F)}{\overline{m}_2^{r-1}(F)} \\
&= \overline{m}_2^r(F)\Delta_v^i - \left(\overline{m}_2^r(F) - \lfloor 2d(S)\rfloor\left(\frac{\overline{m}_2^r(F)}{\overline{m}_2^{r-1}(F)} - 1\right) - 1\right).
\end{aligned}
$$

Claim 4.3 now follows from Lemma 4.2, unless one of the exceptional cases occurs.

If exceptional case (a) occurs, *i.e.*, $r = 2$ and $H = K_4$, we prove Claim 4.3 directly. Since $K_4 = H \subseteq F$, Lemma 2.5 implies that $\overline{m}_2^2(F) \geqslant \overline{m}_2^2(K_4) = 9/4 = 2.25$. This yields with easy case checking that the only edges $g \in E$ with $\Delta_e(g)/\Delta_v(g) > \overline{m}_2^2(F)$ are those with $\Delta_v(g) = 2$ and $\Delta_e(g) = 5$. We call these edges *bad* and all others *good*.

Recall that $f = \{v_j, v_i\}$, $j < i$, denotes the first edge in phase i with $\Delta_v(f) > 0$. The key observation is that all vertices in T_0 at distance 2 from v_i are in the set

$$
\dot{\bigcup}_{f'\in\mathcal{B}} (f' \setminus \{v_i\}),
$$

since this set separates v_i from v_j in $T(f)$, and the edges $f' \in \mathcal{B}$ themselves are not in $T(f)$. In particular, v_i and v_j are not connected by a path of length two in $T(f)$, and therefore for every vertex $u \in \Delta_V(f)$ at most one of the two edges $\{u, v_i\}$ and $\{u, v_j\}$ can be contained in $\Delta_E(f) = E(T(f))$. Consequently, $\Delta_v(f) = 1$ implies $\Delta_e(f) \leqslant 1$, and $\Delta_v(f) = 2$ implies $\Delta_e(f) \leqslant 3$. This shows in particular that f is good. In fact, our argument is that it is 'good enough' to compensate for all bad edges which may occur later in phase i. Formally, we show that

$$
\frac{\Delta_e(f) + \sum_{g\in E, p(g)=i, g \text{ bad}} \Delta_e(g)}{\Delta_v(f) + \sum_{g\in E, p(g)=i, g \text{ bad}} \Delta_v(g)} \leqslant \overline{m}_2^2(F), \tag{4.23}
$$

https://doi.org/10.1017/S0963548308009632 Published online by Cambridge University Press

and infer Claim 4.3 with Proposition 2.3 from $\Delta_e(g)/\Delta_v(g) \leqslant \overline{m}_2^2(F)$ for all good edges of phase i.

To prove (4.23), recall that we have $m(F) < \overline{m}_2^2(F) = 6/(2 + 1/m(F))$ by Lemma 2.2. Solving this for $m(F)$ yields $m(F) < 5/2$. Moreover, we have $m(F) \geqslant d(K_4) = 3/2$. We distinguish the cases $3/2 \leqslant m(F) < 2$ and $2 \leqslant m(F) < 5/2$.

If $3/2 \leqslant m(F) < 2$, there are at most $\lfloor 2d(S) \rfloor = \lfloor 2m(F) \rfloor \leqslant 3$ edges in every phase, and since f is good, there are at most two bad edges. Recall that $\overline{m}_2^2(F) \geqslant 2.25$. Letting x denote the number of bad copies in phase i, the left-hand side of (4.23) is either

$$\frac{1 + x \cdot 5}{1 + x \cdot 2} \overset{x \leqslant 2}{=} 11/5 = 2.2 < \overline{m}_2^2(F)$$

or

$$\frac{3 + x \cdot 5}{2 + x \cdot 2} \overset{x \leqslant 2}{=} 13/6 = 2.166\cdots < \overline{m}_2^2(F),$$

depending on whether $\Delta_v(f) = 1$ or $\Delta_v(f) = 2$.

If $2 \leqslant m(F) < 5/2$, we obtain that $\overline{m}_2^2(F) = 6/(2 + 1/m(F) \geqslant 12/5 = 2.4$, and that every phase consists of at most four edges. The claim follows analogously from

$$\frac{1 + x \cdot 5}{1 + x \cdot 2} \overset{x \leqslant 3}{=} 16/7 = 2.285\cdots < \overline{m}_2^2(F)$$

or

$$\frac{3 + x \cdot 5}{2 + x \cdot 2} \overset{x \leqslant 3}{=} 19/8 = 2.375 < \overline{m}_2^2(F).$$

This settles exceptional case (a).

It remains to consider exceptional case (b), *i.e.*, the case $d(S) < 3/2$ and $e_H \leqslant v_H + 1$. By Lemma 2.4 we have $e_H \geqslant v_H \geqslant 3$. We will solve the case $e_H = v_H$ by a simple combinatorial argument, and the case $e_H = v_H + 1$ by redoing calculations (4.14) and (4.16) taking into account integrality constraints. In both cases, phase i consists of at most $\lfloor 2d(S) \rfloor \leqslant 2$ edges due to $d(S) < 3/2$.

We first deal with the easier case $e_H = v_H$. It follows from (1.4) that H is a cycle (in fact, the shortest cycle in F). For any edge $g \in E$, $T(g)$ is a subgraph of the path $\widehat{H}_-(g)$, of which it contains both endpoints. Hence, we have $\Delta_e(g) = e(T(g)) \leqslant v(T(g)) - 2 = \Delta_v(g)$ unless $T(g) = \widehat{H}_-(g)$, *i.e.*, $\Delta_v(g) = v_H - 2$. However, the latter is only possible if there exists a path of length $e_H - 1$ connecting the endpoints of g in $G(g)$. This can only be the case if there has been another merge operation earlier in phase i, *i.e.*, only for the second of the at most $\lfloor 2d(S) \rfloor \leqslant 2$ edges of phase i.

If phase i consists of only one edge with $\Delta_v(g) > 0$, we have $\Delta_e^i/\Delta_v^i = \Delta_e(g)/\Delta_v(g) \leqslant 1 \leqslant \overline{m}_2^r(F)$. Otherwise, denoting the first edge by f and the second one by g, it follows from the above that

$$\frac{\Delta_e^i}{\Delta_v^i} \leqslant \frac{\Delta_e(f) + \Delta_e(g)}{\Delta_v(f) + \Delta_v(g)} = 1 \leqslant \overline{m}_2^r(F)$$

https://doi.org/10.1017/S0963548308009632 Published online by Cambridge University Press

unless $\Delta_v(g) = v_H - 2$, in which case we have

$$\begin{aligned}\frac{\Delta_e^i}{\Delta_v^i} &\leqslant \frac{\Delta_v(f) + \Delta_v(g) + 1}{\Delta_v(f) + \Delta_v(g)} \\ &\leqslant \frac{v_H}{v_H - 1}, \qquad \text{as } \Delta_v(f) \geqslant 1 \text{ and } \Delta_v(g) = v_H - 2 \\ &\leqslant \frac{v_H}{v_H - 2 + 1/\overline{m}_2^{r-1}(F)} = \overline{m}_2^r(F), \qquad \text{as } \overline{m}_2^{r-1}(F) \geqslant 1.\end{aligned}$$

This settles the case $e_H = v_H$.

Now suppose we have $e_H = v_H + 1$. Recall that phase i consists of at most two edges. If there is only one edge with $\Delta_v(g) > 0$, according to (4.16) we have

$$\begin{aligned}\Delta_e(g) &\leqslant \overline{m}_2^r(F)\big(\Delta_v(g) - 1\big) + 2\frac{\overline{m}_2^r(F)}{\overline{m}_2^{r-1}(F)} - 1 \\ &= \frac{e_H\big(\Delta_v(g) - 1 + 2/\overline{m}_2^{r-1}(F)\big)}{v_H - 2 + 1/\overline{m}_2^{r-1}(F)} - 1 \\ &\overset{\{e_H = v_H + 1\}}{=} \Delta_v(g) - 1 + \frac{\big(3 - 1/\overline{m}_2^{r-1}(F)\big)\Delta_v(g) - (v_H + 1)(1 - 2/\overline{m}_2^{r-1}(F))}{v_H - 2 + 1/\overline{m}_2^{r-1}(F)}.\end{aligned} \tag{4.24}$$

If $\Delta_v(g) \leqslant v_H/2 - 1$ this implies

$$\begin{aligned}\Delta_e(g) &\leqslant \Delta_v(g) + 1 \\ &\quad - \frac{2(v_H - 2 + 1/\overline{m}_2^{r-1}(F)) - (3 - 1/\overline{m}_2^{r-1}(F))(v_H/2 - 1) + (v_H + 1)(1 - 2/\overline{m}_2^{r-1}(F))}{v_H - 2 + 1/\overline{m}_2^{r-1}(F)} \\ &= \Delta_v(g) + 1 - \frac{\frac{3}{2}v_H\overline{m}_2^{r-1}(F) - \frac{3}{2}v_H - 1}{\overline{m}_2^{r-1}(F) \cdot (v_H - 2 + 1/\overline{m}_2^{r-1}(F))}.\end{aligned}$$

Using that $\overline{m}_2^{r-1}(F) \geqslant e_H/v_H = (v_H + 1)/v_H$, we deduce that both numerator and denominator of the quotient are positive. Since $\Delta_e(g)$ is an integer, this implies that $\Delta_e(g) \leqslant \Delta_v(g)$ and thus

$$\frac{\Delta_e^i}{\Delta_v^i} = \frac{\Delta_e(g)}{\Delta_v(g)} = 1 \leqslant \overline{m}_2^r(F),$$

as desired. If $\Delta_v(g) > v_H/2 - 1$, integrality implies that $\Delta_v(g) \geqslant (v_H - 1)/2$. Since we trivially also have that $\Delta_v(g) \leqslant v_H - 2$, we deduce from (4.24)

$$\begin{aligned}\Delta_e(g) &\leqslant \Delta_v(g) + 2 \\ &\quad - \frac{3(v_H - 2 + 1/\overline{m}_2^{r-1}(F)) - (3 - 1/\overline{m}_2^{r-1}(F))(v_H - 2) + (v_H + 1)(1 - 2/\overline{m}_2^{r-1}(F))}{v_H - 2 + 1/\overline{m}_2^{r-1}(F)} \\ &= \Delta_v(g) + 1 - \frac{(v_H + 1)(1 - /\overline{m}_2^{r-1}(F))}{v_H - 2 + 1/\overline{m}_2^{r-1}(F)},\end{aligned}$$

https://doi.org/10.1017/S0963548308009632 Published online by Cambridge University Press

which implies by integrality of $\Delta_e(g)$ that $\Delta_e(g) \leqslant \Delta_v(g) + 1$. Thus, we have

$$\begin{aligned}\frac{\Delta_e^i}{\Delta_v^i} &\leqslant \frac{\Delta_v(g)+1}{\Delta_v(g)} \\ &\leqslant \frac{(v_H+1)/2}{(v_H-1)/2}, \qquad \text{as } \Delta_v(g) \geqslant (v_H-1)/2 \\ &\leqslant \frac{v_H+1}{v_H-2+1/\overline{m}_2^{r-1}(F)} = \overline{m}_2^r(F), \qquad \text{as } \overline{m}_2^{r-1}(F) \geqslant 1.\end{aligned}$$

This proves the claim for the case when only one edge satisfies $\Delta_v(g) > 0$.

Suppose now that there are two edges in phase i with $\Delta_v(\cdot) > 0$, and denote these by f and g, $f \prec g$. Applying (4.16) to f and (4.14) to g we get

$$\Delta_e(f) \leqslant \overline{m}_2^r(F)\big(\Delta_v(f)-1\big) + \frac{\overline{m}_2^r(F)}{\overline{m}_2^{r-1}(F)}$$

and

$$\Delta_e(g) \leqslant \overline{m}_2^r(F)\Delta_v(g) + \frac{\overline{m}_2^r(F)}{\overline{m}_2^{r-1}(F)} - 1.$$

Note that these equations are similar but not quite identical to (4.24). Nevertheless similar case distinctions as above yield

$$\Delta_e(f) \leqslant \begin{cases}\Delta_v(f)+1 & \text{if } v_H/2 \leqslant \Delta_v(f) \leqslant v_H-2,\\ \Delta_v(f) & \text{if } 1 \leqslant \Delta_v(f) \leqslant (v_H-1)/2,\end{cases}$$

and

$$\Delta_e(g) \leqslant \begin{cases}\Delta_v(g)+2 & \text{if } \Delta_v(g) = v_H-2,\\ \Delta_v(g)+1 & \text{if } v_H/2-1 \leqslant \Delta_v(g) < v_H-2,\\ \Delta_v(g) & \text{if } 1 \leqslant \Delta_v(g) \leqslant (v_H-3)/2.\end{cases}$$

With these inequalities at hand, the claim follows for all possible values of $\Delta_v(f)$ and $\Delta_v(g)$ from one of the following calculations.

In the first case we have $\Delta_v(f) \geqslant v_H/2$ and $\Delta_v(g) = v_H - 2$. Using that $\overline{m}_2^{r-1}(F) \geqslant d(H) = (v_H+1)/v_H$ and with some elementary calculations, this yields

$$\begin{aligned}\frac{\Delta_e^i}{\Delta_v^i} &\leqslant \frac{\Delta_v(f)+\Delta_v(g)+3}{\Delta_v(f)+\Delta_v(g)} \\ &\leqslant \frac{3v_H/2+1}{3v_H/2-2} \\ &\leqslant \frac{v_H+1}{v_H-2+v_H/(v_H+1)} \\ &\leqslant \frac{v_H+1}{v_H-2+1/\overline{m}_2^{r-1}(F)} = \overline{m}_2^r(F).\end{aligned}$$

https://doi.org/10.1017/S0963548308009632 Published online by Cambridge University Press

In the second case we suppose that either $\Delta_v(f) \geqslant v_H/2$ and $tv_H - 2 > \Delta_v(g) \geqslant v_H/2 - 1$ or $(v_H - 1)/2 \geqslant \Delta_v(f) \geqslant 1$ and $\Delta_v(g) = v_H - 2$. We obtain similarly

$$\begin{aligned}\frac{\Delta_e^i}{\Delta_v^i} &\leqslant \frac{\Delta_v(f) + \Delta_v(g) + 2}{\Delta_v(f) + \Delta_v(g)} \\ &\leqslant \frac{v_H + 1}{v_H - 1} \leqslant \frac{v_H + 1}{v_H - 2 + 1/\overline{m}_2^{r-1}(F)} = \overline{m}_2^r(F).\end{aligned}$$

In the third case we assume that either $(v_H - 1)/2 \geqslant \Delta_v(f) \geqslant 1$ and $v_H - 2 > \Delta_v(g) \geqslant v_H/2 - 1$ or $\Delta_v(f) \geqslant v_H/2$ and $(v_H - 3)/2 \geqslant \Delta_v(g) \geqslant 1$. This yields

$$\begin{aligned}\frac{\Delta_e^i}{\Delta_v^i} &\leqslant \frac{\Delta_v(f) + \Delta_v(g) + 1}{\Delta_v(f) + \Delta_v(g)} \\ &\leqslant \frac{v_H/2 + 1}{v_H/2} \leqslant \frac{v_H + 1}{v_H - 1} \leqslant \frac{v_H + 1}{v_H - 2 + 1/\overline{m}_2^{r-1}(F)} = \overline{m}_2^r(F).\end{aligned}$$

In the last case when $(v_H - 1)/2 \geqslant \Delta_v(f) \geqslant 1$ and $(v_H - 3)/2 \geqslant \Delta_v(g) \geqslant 1$, we readily have

$$\frac{\Delta_e^i}{\Delta_v^i} \leqslant \frac{\Delta_v(f) + \Delta_v(g)}{\Delta_v(f) + \Delta_v(g)} = 1 \leqslant \overline{m}_2^r(F).$$

This settles exceptional case (b), and concludes the proof of Claim 4.3. □

As explained before, Claims 4.3 and 4.4 imply Lemma 4.1 by (4.13) and Proposition 2.3. □

5. Forests

To state our results for forests, we need to introduce some notation. For a given forest F, let $e_{\max} = e_{\max}(F)$ denote the number of edges in a largest connected component, let $k = k(F)$ be the maximal size of a minimal vertex cover of a component of F, and let $e_{\mathrm{mvc}} = e_{\mathrm{mvc}}(F)$ be the number of edges in a largest connected component with a minimal vertex cover of size k (thus, $e_{\mathrm{mvc}} \leqslant e_{\max}$).

With this notation, define for any forest F and $r \geqslant 1$,

$$\widetilde{e}^r(F) := 1 + (e_{\mathrm{mvc}} - 1)\sum_{i=0}^{r-2} k^i + (e_{\max} - 1)k^{r-1}.$$

Theorem 5.1. *Let F be a forest, and let $r \geqslant 1$. Then the online F-avoidance edge-colouring game with r colours has a threshold $N_0(F, r, n)$ that satisfies*

$$N_0(F, r, n) \geqslant n^{1 - 1/\widetilde{e}^r(F)}.$$

Note that $k = 1$ if and only if F is a star forest, and that therefore

$$\widetilde{e}^r(F) = \begin{cases} 1 + (e_{\max} - 1)r & \text{if } F \text{ is a star forest,} \\ 1 + (e_{\mathrm{mvc}} - 1)\frac{k^{r-1}-1}{k-1} + (e_{\max} - 1)k^{r-1} & \text{otherwise.} \end{cases}$$

https://doi.org/10.1017/S0963548308009632 Published online by Cambridge University Press

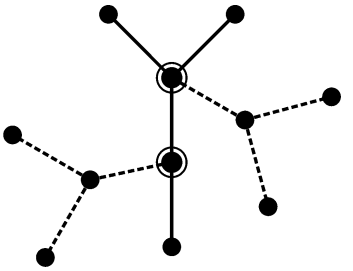


Figure 6. A smallest tree in $\mathcal{F}(T,T)$ for T the only tree with four edges which is neither a star nor P_4. The circled dots form a vertex cover of the inner copy of T.

If $F = T$ is a tree, we have $e_{\max} = e_{\mathrm{mvc}} = e_T$, which yields that

$$\widetilde{e}^r(T) := \begin{cases} 1 + (e_T - 1)r & \text{if } T \text{ is a star,} \\ 1 + (e_T - 1)\frac{k^r - 1}{k-1} & \text{otherwise.} \end{cases}$$

As in Theorem 1.2, setting $r = 1$ yields the statement of Theorem 1.5, and for $r \to \infty$ the exponent tends to $1 = 2 - 1/m_2(F)$, which by Theorem 1.1 is the exponent in the offline case (unless F is a star forest). In fact, for F a star forest the lower bound stated in Theorem 5.1 coincides with the well-known offline threshold given by the pigeon-hole principle, which immediately yields the same threshold for the online game.

We do not know whether $\widetilde{e}^r(F)$ determines a general threshold. As mentioned in the Introduction, $n^{1-1/\widetilde{e}^r(F)}$ is indeed the threshold of the game for $r = 2$ and all forests on up to five vertices. In particular, we have that $N_0(P_3, 2, n) = n^{6/7}$ and $N_0(P_4, 2, n) = n^{9/10}$.

The proof of Theorem 5.1 relies on a well-known theorem due to König. A proof can be found, *e.g.*, in [3].

Theorem 5.2 (König 1931). *In every bipartite graph, the maximum cardinality of a matching is equal to the minimum cardinality of a vertex cover.*

For any graph F, let $\mathrm{mvc}(F)$ denote the size of a minimum vertex cover of F. Similarly to Lemmas 3.1 and 4.1, we will first prove a statement which essentially covers the case $r = 2$ and then derive a more general statement by induction. Figure 6 illustrates the statement of Lemma 5.3 for a simple example.

Lemma 5.3. *Let S and T be trees. Then every tree $S' \in \mathcal{F}(S,T)$ has at least*

$$e_S + \mathrm{mvc}(S)(e_T - 1)$$

edges, and the size of a minimal vertex cover is at least

$$\mathrm{mvc}(S)\,\mathrm{mvc}(T).$$

Proof of Lemma 5.3. Overlapping outer copies belonging to non-adjacent edges of S would close a cycle and thus create a graph that is not a tree. Hence the edges of every matching $M \subseteq S$ need their own outer copy, and every tree in $\mathcal{F}(S,T)$ has at least $e_S + e_M(e_T - 1)$ edges. By

https://doi.org/10.1017/S0963548308009632 Published online by Cambridge University Press

Theorem 5.2, the cardinality e_M of the largest matching in S is $\mathrm{mvc}(S)$, so every tree in $\mathcal{F}(S,T)$ has at least $e_S + \mathrm{mvc}(S)(e_T - 1)$ edges. By the same argument, there are $\mathrm{mvc}(S)$ disjoint copies of T in S', which immediately implies the second inequality. □

Lemma 5.4. *Let $T_1,\dots,T_r$ be trees. Then every tree $S^r \in \mathcal{F}(T_1,\dots,T_r)$ has at least*

$$1 + \sum_{i=1}^{r} (e(T_i) - 1) \prod_{j=1}^{i-1} \mathrm{mvc}(T_j)$$

edges, and the size of a minimal vertex cover is at least

$$\prod_{j=1}^{r} \mathrm{mvc}(T_j).$$

Proof of Lemma 5.4. We proceed by induction on r. For $r = 1$ the claim is immediate. For $r \geqslant 2$, by definition there exists a tree $S^{r-1} \in \mathcal{F}(T_1,\dots,T_{r-1})$ such that $S^r \in \mathcal{F}(S^{r-1}, T_r)$. Using the induction hypothesis, the claim follows from Lemma 5.3 with $S \leftarrow S^{r-1}$ and $T \leftarrow T_r$. □

Proof of Theorem 5.1. Let $T_{\max}$ denote a largest component of F, and T_{mvc} a largest component with a minimal vertex cover of size k, such that $e_{\max} = e(T_{\max})$ and $e_{\mathrm{mvc}} = e(T_{\mathrm{mvc}})$. By the arguments given in Section 3, if Painter plays the greedy $\langle T_{\max}, T_{\mathrm{mvc}},\dots,T_{\mathrm{mvc}}\rangle$-avoidance strategy, she will not lose the game if contains no graph from $\mathcal{F}(T_{\mathrm{mvc}},\dots,T_{\mathrm{mvc}},T_{\max})$. Clearly, all graphs in this class that are not trees have density at least 1 and a.a.s. do not appear in $G(n,N)$ with $N \ll n$ due to Theorem 1.5. Similarly, every tree in $\mathcal{F}(T_{\mathrm{mvc}},\dots,T_{\mathrm{mvc}},T_{\max})$ has at least $\widetilde{e}^r(F)$ edges by Lemma 5.4, and a.a.s. does not appear in $G(n,N)$ as long as $N \ll n^{1-1/\widetilde{e}^r(F)}$. □

It is easy to see that the bounds in Lemma 5.3 and Lemma 5.4 are tight (*cf.* Figure 6), and that the choice of $T_1,\dots,T_r$ in the proof of Theorem 5.1 is optimal. Therefore, Theorem 5.1 states the best bound obtainable in our framework.

References

[1] Bollobás, B. (1981) Threshold functions for small subgraphs. *Math. Proc. Cambridge Philos. Soc.* **90** 197–206.

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

[3] Diestel, R. (2005) *Graph Theory*, Vol. 173 of *Graduate Texts in Mathematics*, 3rd edn, Springer, Berlin.

[4] Folkman, J. (1970) Graphs with monochromatic complete subgraphs in every edge coloring. *SIAM J. Appl. Math.* **18** 19–24.

[5] Friedgut, E. (2005) Hunting for sharp thresholds. *Random Struct. Alg.* **26** 37–51.

[6] Friedgut, E., Kohayakawa, Y., Rödl, V., Ruciński, A. and Tetali, P. (2003) Ramsey games against a one-armed bandit. *Combin. Probab. Comput.* **12** 515–545.

[7] Łuczak, T., Ruciński, A. and Voigt, B. (1992) Ramsey properties of random graphs. *J. Combin. Theory Ser. B* **56** 55–68.

[8] Marciniszyn, M. and Spöhel, R. (2007) Online vertex colorings of random graphs without monochromatic subgraphs. In *Proc. 18th annual ACM–SIAM Symposium on Discrete Algorithms* (SODA 2007), pp. 486–493. Journal version to appear in *Combinatorica*.

https://doi.org/10.1017/S0963548308009632 Published online by Cambridge University Press

[9] Marciniszyn, M., Spöhel, R. and Steger, A. (2005) The online clique avoidance game on random graphs. In *Approximation, Randomization and Combinatorial Optimization*, Vol. 3624 of *Lecture Notes in Computer Science*, Springer, pp. 390–401.
[10] Marciniszyn, M., Spöhel, R. and Steger, A. (2009) Upper bounds for online Ramsey games in random graphs. *Combin. Probab. Comput.*, this issue.
[11] Nešetřil, J. and Rödl, V. (1976) The Ramsey property for graphs with forbidden complete subgraphs. *J. Combin. Theory Ser. B* **20** 243–249.
[12] Ramsey, F. P. (1930) On a problem of formal logic. In *Proc. London Math. Soc.* **30** 264–286.
[13] Rödl, V. and Ruciński, A. (1993) Lower bounds on probability thresholds for Ramsey properties. In *Combinatorics: Paul Erdős is Eighty*, Vol. 1 of *Bolyai Society Mathematical Studies*, pp. 317–346.
[14] Rödl, V. and Ruciński, A. (1995) Threshold functions for Ramsey properties. *J. Amer. Math. Soc.* **8** 917–942.

https://doi.org/10.1017/S0963548308009632 Published online by Cambridge University Press