REPEATED PATTERNS IN PROPER COLORINGS*

DAVID CONLON[†] AND MYKHAYLO TYOMKYN[‡]

Abstract. For a fixed graph H , what is the smallest number of colors C such that there is a proper edge-coloring of the complete graph K_n with C colors containing no two vertex-disjoint color-isomorphic copies, or repeats, of H ? We study this function and its generalization to more than two copies using a variety of combinatorial, probabilistic, and algebraic techniques. For example, we show that for any tree T there exists a constant c such that any proper edge-coloring of K_n with at most cn^2 colors contains two repeats of T , while there are colorings with at most $c'n^{3/2}$ colors for some absolute constant c' containing no three repeats of any tree with at least two edges. We also show that for any graph H containing a cycle there exist k and c such that there is a proper edge-coloring of K_n with at most cn colors containing no k repeats of H , while for a tree T with m edges, a coloring with $o(n^{(m+1)/m})$ colors contains $\omega(1)$ repeats of T .

Key words. graphs, coloring, extremal problems

AMS subject classification. 05C35

DOI. 10.1137/21M1414103

1. Introduction. A considerable body of recent work in extremal combinatorics is devoted to the study of rainbow patterns in proper edge-colorings of complete graphs. To mention two such results (among many [3, 4, 11, 12, 14, 15, 16, 21, 22, 25, 26, 28, 29, 30]), there is the work of Alon, Pokrovskiy, and Sudakov [1] showing that any proper edge-coloring of K_n contains a rainbow path of length $n - o(n)$, and the work of Montgomery, Pokrovskiy, and Sudakov [27] and, independently, Keevash and Staden [20] resolving a celebrated conjecture of Ringel, one of whose statements involves finding a rainbow copy of any tree with n edges in a particular proper edge-coloring of K_{2n+1} . For the most part, this recent work has focused on finding large structures in proper edge-colorings. We instead study small structures, our aim being to understand when a proper edge-coloring contains two or more repeats of a particular graph H .

To be more precise, we say that two copies of a graph H in a coloring of K_n are *color isomorphic* if there exists an isomorphism between them preserving the colors. The following function is our main object of study.

DEFINITION 1.1. For $k, n \geq 2$ and a graph H , define $f_k(n, H)$ to be the smallest integer C such that there is a proper edge-coloring of K_n with C colors containing no k vertex-disjoint color-isomorphic copies (or “repeats”) of H .

We make several remarks about this definition. First, one could, in principle, ask the same question without the restriction to proper colorings. However, this changes the character of the question completely. Indeed, consider the coloring of the complete graph on vertex set $\{1, 2, \dots, n\}$, where we color the edge ij with $i < j$ by the color i .

*Received by the editors April 21, 2021; accepted for publication (in revised form) July 1, 2021; published electronically September 28, 2021.

<https://doi.org/10.1137/21M1414103>

Funding: The first author was supported by NSF award DMS-2054452. The second author was supported by ERC Synergy grant DYNASNET 810115, the H2020-MSCA-RISE project CoSP-GA 823748, and GACR grant 19-04113.

[†]Department of Mathematics, California Institute of Technology, Pasadena, CA 91125 USA (dconlon@caltech.edu).

[‡]Department of Applied Mathematics, Charles University, 11800 Prague, Czech Republic (tyomkyn@kam.mff.cuni.cz).

Then this is a coloring with $n - 1$ colors which does not even contain two disjoint edges of the same color. On the other hand, when we restrict ourselves to proper colorings, we have that $f_k(n, K_2) \geq \lceil \frac{1}{k-1} \binom{n}{2} \rceil$ by a straightforward application of the pigeonhole principle. For n sufficiently large in terms of k , it also follows from several well-known decomposition results, such as Gustavsson's theorem [18], that this bound is tight.

Our second remark collects together several simple observations that we will use throughout.

Remark 1.2. The quantity $f_k(n, H)$ is monotone increasing in n , but decreasing in k and in H (with respect to taking subgraphs). Moreover, since every proper coloring has at least $n - 1$ colors,

$$\binom{n}{2} \geq f_k(n, H) \geq n - 1.$$

Finally, although our definition contains no requirement that the copies of H should be rainbow, all of the results where we find repeats of a particular graph H remain true up to a constant factor if we insist that each copy is rainbow. This then brings our work more fully in line with the body of research discussed at the outset. With these preliminaries out of the way, we now describe our main results.

In the classical Turán problem, the growth rate of the extremal function $\text{ex}(n, H)$ is subject to a well-known trichotomy. Namely, nonbipartite graphs, bipartite graphs with a cycle, and forests satisfy $\text{ex}(n, H) = \Theta(n^2)$, $n^{1+\Omega(1)} \leq \text{ex}(n, H) \leq n^{2-\Omega(1)}$, and $\text{ex}(n, H) = \Theta(n)$, respectively. Our first theorem shows that something broadly similar holds for $f_2(n, H)$, although, unlike the extremal function, our function can, and usually does, degenerate for bipartite graphs with a cycle. Note that here and throughout, all terms in the O -notation are to be interpreted with respect to n , with all other variables treated as constants.

THEOREM 1.3. *The growth rate of $f_2(n, H)$ satisfies the following:*

- (i) $f_2(n, H) = \Theta(n^2)$ if H is a forest. Otherwise, $f_2(n, H) = O(n^{2-\Omega(1)})$.
- (ii) If H is nonbipartite, then $f_2(n, H) \leq n + 1$.
- (iii) If H is bipartite and $e(H) \geq 2|H| - 2$, then $f_2(n, H) = \Theta(n)$.
- (iv) There exist bipartite graphs H with $n^{1+\Omega(1)} \leq f_2(n, H) \leq n^{2-\Omega(1)}$.

For three or more repeats, the class of graphs for which we know that $f_k(n, H) = O(n)$ grows as k increases. In fact, for any graph H containing a cycle, we can show, by using a variant of Bukh's random algebraic method [5], that there exists k such that $f_k(n, H) = O(n)$.

THEOREM 1.4. *For any graph H containing a cycle, there exists k such that $f_k(n, H) = O(n)$.*

For trees, the situation is much more involved, as spelled out in the next theorem, whose proof relies on a mixture of novel combinatorial and algebraic methods.

THEOREM 1.5. *For any tree T with m edges and any $k \geq 3$, the following hold:*

- (i) $f_k(n, T) = \Omega(n^{k/(k-1)})$. Moreover, if T has at least two edges, then $f_3(n, T) = \Theta(n^{3/2})$.
- (ii) $f_k(n, T) = \Omega(n^{(m+1)/m})$ and there exists k' such that $f_{k'}(n, T) = O(n^{(m+1)/m})$.

The paper is structured as follows. In section 2, we collect some general upper and lower bounds for $f_k(n, H)$. We then proceed to study trees in section 3. In section 4, we study even cycles. We conclude, in section 5, by discussing a broad range of open

problems suggested by our work. Unless otherwise specified, the term “coloring” will refer to proper edge-colorings of a complete graph.

2. Preliminary observations. Since it is not entirely obvious, we first verify that finding $f_k(n, H)$ is indeed an extremal problem.

LEMMA 2.1. *For any k, n, H and $\binom{n}{2} \geq C > f_k(n, H)$, there exists a C -coloring of K_n which does not contain k repeats of H .*

Proof. Suppose for contradiction that the statement does not hold for some $C > f_k(n, H)$ and consider an $f_k(n, H)$ -coloring of K_n which avoids k repeats of H . In this coloring, take any $C - f_k(n, H)$ edges from color classes of size at least 2, leaving at least one edge in each color class and recolor each of these edges with an entirely new color. By assumption, the resulting C -coloring contains k repeats of H . Since each new color class has only one edge, the repeats must be in the old colors, a contradiction. $\square$

The following lemma effectively reduces the problem to connected graphs.

LEMMA 2.2. *Suppose that H is a graph with connected components $H_1, \dots, H_\ell$. Then, for all $k \geq 2$,*

$$f_k(n, H) = \Theta(\min_{i \in [\ell]} \{f_k(n, H_i)\}).$$

Proof. Since containing no k -fold repeat of one of the H_i implies that there is no k -fold repeat of H , we immediately have

$$(2.1) \quad f_k(n, H) \leq \min_{i \in [\ell]} \{f_k(n, H_i)\}.$$

For the other direction, let $k \geq 2$ be fixed. We will prove the result by induction on ℓ , noting first that the statement is obviously true for $\ell = 1$. Given $\ell \geq 2$, H , and n , we may assume, without loss of generality, that

$$f_k(n, H_\ell) = \min_{i \in [\ell]} \{f_k(n, H_i)\}$$

and let $H' := H \setminus H_\ell$. By the induction hypothesis, there exists a constant $c = c(k, \ell, H_1, \dots, H_{\ell-1})$ such that

$$f_k(n, H') \geq c \cdot \min_{i \in [\ell-1]} f_k(n, H_i) \geq c \cdot f_k(n, H_\ell).$$

Observe now that

$$(2.2) \quad f_k(n, H) \geq \min\{f_k(n, H_\ell), f_k(n - k|H_\ell|, H')\},$$

as otherwise we can find a k -repeat of H_ℓ and then a k -repeat of H' on the remaining vertices, resulting in a k -repeat of H . If $f_k(n, H_\ell) \leq f_k(n - k|H_\ell|, H')$, then, by (2.2) and (2.1), $f_k(n, H) = f_k(n, H_\ell)$ and we are done. So suppose now that $f_k(n, H_\ell) > f_k(n - k|H_\ell|, H')$, in which case (2.2) becomes

$$(2.3) \quad f_k(n, H) \geq f_k(n - k|H_\ell|, H').$$

The right-hand side of (2.3) is at least $f_k(n, H') - k|H_\ell|n$, as can be seen by extending a coloring of the complete graph on $n - k|H_\ell|$ vertices containing no k -repeat of H'

to a coloring of K_n with the same property by using a new unique color for each edge (for a total of at most $k|H_\ell|n$ new colors). Therefore,

$$(2.4) \quad f_k(n, H) \geq f_k(n - k|H_\ell|, H') \geq f_k(n, H') - k|H_\ell|n \geq c \cdot f_k(n, H_\ell) - k|H_\ell|n.$$

If $2(k/c)|H_\ell|n < f_k(n, H_\ell)$, then, by (2.4),

$$f_k(n, H) \geq cf_k(n, H_\ell) - \frac{c}{2}f_k(n, H_\ell) = \frac{c}{2}f_k(n, H_\ell).$$

On the other hand, by Remark 1.2, we have $f_k(n, H) \geq n - 1 \geq n/2$. Hence, if $2(k/c)|H_\ell|n \geq f_k(n, H_\ell)$, then

$$f_k(n, H) \geq \frac{n}{2} = \frac{2(k/c)|H_\ell|n}{4(k/c)|H_\ell|} \geq \frac{c}{4k|H_\ell|}f_k(n, H_\ell),$$

completing the induction step and the proof. $\square$

Next, we quickly resolve the case of nonbipartite graphs, proving Theorem 1.3(ii).

THEOREM 2.3. *If H contains an odd cycle, then $f_k(n, H) \leq n + 1$.*

Proof. It suffices to show this for $k = 2$ and $H = C_{2\ell+1}$. To this end, consider the following well-known coloring c . Suppose that $n = 2p + 1$ is an odd number, let $V(G) = \mathbb{Z}_n$, and put $c(a, b) := a + b \pmod n$. Clearly this is a proper coloring using n colors. Moreover, whenever there is a copy of $C_{2\ell+1}$ with vertices $v_0, \dots, v_{2\ell+1} = v_0$, labeled along the cycle, we have $c_i := c(v_i, v_{i+1}) = v_i + v_{i+1} \pmod n$ for each i . The vertices $v_0, \dots, v_{2\ell}$ are then uniquely determined by the colors $c_0, \dots, c_{2\ell}$, since $v_0 = (c_0 - c_1 + c_2 - \dots + c_{2\ell})/2 \pmod n$ and a similar identity holds for the other v_i . Hence, no two distinct, let alone vertex-disjoint, cycles will be color isomorphic. This proves that $f_k(n, H) \leq n$ when n is odd. For even n , monotonicity implies that $f_k(n, H) \leq f_k(n + 1, H) \leq n + 1$. $\square$

For odd n , the proof of Theorem 2.3 gives $f_k(n, H) \leq n$, which is best possible, as this is the minimum number of colors in any proper coloring when n is odd. For even n , the bounds in Remark 1.2 and Theorem 2.3 differ by 2. We refer the reader to section 5 for further discussion.

For upper-bound constructions, the following observation will be very useful. A not necessarily proper edge-coloring of K_n is called b -bounded if each color class is a graph of maximum degree at most b .

PROPOSITION 2.4. *For any graph H and integers $b, k, n \geq 2$, any b -bounded C -coloring of K_n without k repeats of any (not necessarily proper) coloring of H can be refined to a proper coloring with at most $(b + 1)C$ colors and no k -repeat of H .*

Proof. Since each color class is a graph of maximum degree b , Vizing's theorem implies that the edges in each color class can be recolored with at most $b + 1$ new colors so that the new coloring is proper. A k -repeat of H in the new colors would imply the same in the old colors and is therefore impossible. $\square$

The next theorem provides, via a simple probabilistic argument, a general upper bound for $f_k(n, H)$. In light of Theorem 2.3, it is only meaningful for bipartite H .

THEOREM 2.5. *For any graph H with v vertices and e edges,*

$$f_k(n, H) = O(\max\{n, n^{\frac{kv-2}{(k-1)e}}\}).$$

Proof. This is a standard application of the Lovász Local Lemma (see, for instance, [2]). We would like to find a probability $p = p(n)$ and assign each edge of K_n independently one of $1/p$ colors so that with positive probability the following will hold:

(P1) There is no k -fold repeat of any (not necessarily proper) coloring of H .

(P2) The coloring is $(kv - 2)$ -bounded.

For this, we define two collections of “bad” events, as follows. Given k disjoint copies of H in K_n , say $H_1, \dots, H_k$, let $A(H_1, \dots, H_k)$ be the event that $H_1, \dots, H_k$ are color isomorphic. We have

$$q_A := \mathbb{P}(A(H_1, \dots, H_k)) = \Theta(p^{(k-1)e}).$$

For every copy S of S_ℓ , the star with $\ell := kv - 1$ edges, we define $B(S)$ to be the event that all edges of S receive the same color. We have

$$q_B := \mathbb{P}(B(S)) = p^{\ell-1}.$$

For every fixed $H_1, \dots, H_k$, the event $A(H_1, \dots, H_k)$ is independent of any event where the subgraphs involved do not share an edge with any of $H_1, \dots, H_k$. Therefore, the number of dependent events is at most

$$\Delta = O(n^{kv-2} + n^{(\ell+1)-2}) = O(n^{\ell-1}).$$

Similarly, each $B(S)$ is independent from all but at most $\Delta = O(n^{\ell-1})$ other events. By the Local Lemma, in order to establish that $\mathbb{P}(\bigcap \overline{A} \cap \bigcap \overline{B}) > 0$, it suffices to show that $\max\{q_A, q_B\} < 1/e\Delta$. This will be the case when

$$p \leq \gamma \min\{n^{-1}, n^{-\frac{\ell-1}{(k-1)e}}\}$$

for some constant $\gamma = \gamma(k, v, e) > 0$. Therefore, there exists a $(kv - 2)$ -bounded coloring with $O(\max\{n, n^{\frac{kv-2}{(k-1)e}}\})$ colors satisfying (P1). By Proposition 2.4, it can be refined to a proper $O(\max\{n, n^{\frac{kv-2}{(k-1)e}}\})$ -coloring with no k -repeat of H , as desired. $\square$

We immediately deduce the following corollary, which includes Theorem 1.3(iii) as a special case.

COROLLARY 2.6. *If H is a bipartite graph with $e(H) \geq \frac{k}{k-1}|H| - \frac{2}{k-1}$, then*

$$f_k(n, H) = \Theta(n).$$

For $e(H) \geq |H| + 1$, this already implies that there exists $k = k(H)$ such that $f_k(n, H) = O(n)$. We will later show that this holds for all graphs containing a cycle, but using Theorem 2.5 gives an explicit upper bound for the smallest k such that $f_k(n, H) = O(n)$, namely, $k \leq \lceil \frac{e-2}{e-v} \rceil$, which our later approach does not provide.

3. Trees. Applied to trees, Theorem 2.5 reads as follows.

COROLLARY 3.1. *For any tree T with m edges,*

$$f_k(n, T) = O(n^{\frac{k(m+1)-2}{(k-1)m}}).$$

On the other hand, we have the following general lower bound for trees, showing that Corollary 3.1 is not far from the truth for large trees.

THEOREM 3.2. *For any tree T with m edges, any $k \geq 2$, and n sufficiently large,*

$$f_k(n, T) \geq n^{\frac{k}{k-1}}/q,$$

where $q = (4k(km+1)(k^2-k+1))^{\frac{1}{k-1}}$.

Proof. Fix k and T , let q be as above, and suppose that for some large n we have a coloring of $G = K_n$ with $C = n^{\frac{k}{k-1}}/q$ colors. Let F be the following auxiliary graph. Set $V(F) = \binom{V(G)}{k}$, that is, each vertex of F is a vertex subset of G of order k . Two vertices of F , corresponding to sets $U = \{u_1, \dots, u_k\}$ and $W = \{w_1, \dots, w_k\}$, are connected by an edge if $U \cap W = \emptyset$ and the bipartite graph $G[U, W]$ has a matching of size k which is monochromatic.

By the above definition, each monochromatic k -matching gives rise to 2^{k-1} edges of F , while each edge of F is accounted for by at most k different k -matchings. This implies that

$$e(F) \geq \frac{2^{k-1}}{k} E_k,$$

where E_k is the number of monochromatic k -matchings in G .

Note now that for the average degree of F , we have

$$d_{\text{avg}}(F) = \frac{2e(F)}{|F|} \geq \frac{2^k E_k}{k \binom{n}{k}}.$$

By the convexity of binomial coefficients,

$$E_k \geq C \binom{\frac{1}{C} \binom{n}{2}}{k}.$$

Thus,

$$d_{\text{avg}}(F) \geq \frac{2^k C \binom{\frac{1}{C} \binom{n}{2}}{k}}{k \binom{n}{k}} = (1 + o(1)) \frac{n^k}{k C^{k-1}} = (1 + o(1)) \frac{q^{k-1}}{k}.$$

By a folklore fact, F therefore contains a subgraph F' with

$$d := \delta(F') \geq \frac{d_{\text{avg}}(F)}{2} \geq (1 + o(1)) \frac{q^{k-1}}{2k}.$$

Let $v_0, \dots, v_m$ be the vertices of T ordered so that each v_i is a leaf attached to $T[v_0, \dots, v_{i-1}]$. We claim that there exists an embedding $\phi : T \rightarrow F'$ such that the k -vertex sets of G corresponding to $\phi(v_0), \dots, \phi(v_m)$ are disjoint.

We let $\phi(v_0)$ be an arbitrary vertex of F' and proceed inductively. Suppose that for some $1 \leq i \leq m$ we have defined $\phi(v_0), \dots, \phi(v_{i-1})$ and let v_j be the unique neighbor of v_i in $T[v_0, \dots, v_{i-1}]$. Take any d vertices in $N_{F'}(\phi(v_j))$ and consider the underlying k -sets in $V(G)$. This defines a k -uniform hypergraph H .

Observe that the maximum degree in H is at most k , since for every vertex $u \notin \phi(v_j)$ there are at most k monochromatic matchings in G connecting $\phi(v_j)$ and a set containing u . Thus, the line graph of H has maximum degree at most $k(k-1)$, so we can properly vertex color it greedily with at most $k^2 - k + 1$ colors. Take the largest color class: this is a matching $\mu \subseteq E(H)$ of size at least $d/(k^2 - k + 1)$. But

$$\frac{d}{k^2 - k + 1} \geq (1 + o(1)) \frac{q^{k-1}}{2k(k^2 - k + 1)} = (2 + o(1))(km + 1).$$

Therefore, when n is sufficiently large, we have $|\mu| \geq km + 1 \geq ki + 1$, so there exists $e \in \mu$ such that the underlying k -vertex set in G satisfies $e \cap \phi(v_\ell) = \emptyset$ for all $\ell = 0, \dots, i - 1$. We can therefore let $\phi(v_i) := e$, completing the induction step. The resulting copy of $\phi(T) \subseteq F'$ gives k vertex-disjoint color-isomorphic copies of T in the original coloring of K_n , completing the proof. $\square$

In particular, when $k = 2$, we have the following result.

COROLLARY 3.3. *For any tree T with m edges and n sufficiently large,*

$$f_2(n, T) \geq \frac{1}{24(2m+1)} n^2.$$

This also completes the proof of Theorem 1.3(i). Indeed, if H is a forest, then $f_2(n, H) = \Theta(n^2)$ by Corollary 3.3 and Lemma 2.2. Conversely, if H is not a forest, it has a connected component H' with $e(H') \geq |H'|$. By Theorem 2.5, $f_2(n, H') = O(n^{2-\Omega(1)})$, so the same then holds for H by Lemma 2.2.

We now observe that the bound in Corollary 3.3 is tight up to an absolute constant.

LEMMA 3.4. *For every tree T with $m \geq 2$ edges,*

$$f_2(n, T) \leq (1 + o(1)) \frac{n^2}{2m}.$$

Proof. By Wilson's theorem [31], there is an edge decomposition of all but $o(n^2)$ edges of K_n into at most $\binom{n}{2} / \binom{2m+1}{2}$ copies of K_{2m+1} . We color each of the $o(n^2)$ edges in a unique color and decompose each of the copies of K_{2m+1} into at most $2m+1$ (near-)perfect matchings, each of which we view as a separate color class. Since any two copies of K_{2m+1} share at most one vertex, there will be no repeat of T with colors from two different K_{2m+1} . On the other hand, since $2m+1 < 2(m+1)$ and T has $m+1$ vertices, there cannot be two vertex-disjoint copies of T inside any K_{2m+1} . $\square$

Next, we show that the bound in Theorem 3.2 is essentially tight when $k = 3$, thus completing the proof of Theorem 1.5(i).

THEOREM 3.5. *For any tree T with at least 2 edges, $f_3(n, T) = O(n^{3/2})$.*

Proof. Suppose that A is a vertex set of order $n \leq q^2 = (1 + o(1))n$ for some prime q , noting that such a prime exists by the prime number theorem. Our aim is to find a proper coloring of the complete graph on A with $O(n^{3/2})$ colors such that no three vertices are incident to the same two colors. In particular, this will imply that the coloring has no 3-fold repeat of the star with two edges and, hence, no 3-fold repeat of any tree T with more than one edge.

To achieve this, let A be indexed by $\mathbb{F}_q^2$, where $\mathbb{F}_q$ denotes the finite field of order q . We first deal with certain “degenerate” edges in $\binom{A}{2}$. For our purposes, these will be edges between two vertices $u = (a, b)$ and $v = (c, d)$, where either $a = c$, $a = 1$, or $c = 1$. Note that there are $O(q^3) = O(n^{3/2})$ such edges, so we may color each of them with a unique color. To each remaining edge $\{u, v\} \in \binom{A}{2}$ with $u = (a, b)$ and $v = (c, d)$, we assign the color $(x_1, x_2, x_3) \in \mathbb{F}_q^3$ (so that there are at most $q^3 = O(n^{3/2})$ additional colors) satisfying the equations $x_1 + x_2a + x_3a^2 = b$, $x_1 + x_2c + x_3c^2 = d$, and $x_1 + x_2 + x_3 = a + c$. This system of three equations has a unique solution (x_1, x_2, x_3) , since the underlying matrix is a Vandermonde matrix (and, since a, c , and 1 are all distinct, the matrix is nonsingular). Now, for a fixed $u = (a, b)$, how many nondegenerate edges uv with $v = (c, d)$ receive the color (x_1, x_2, x_3) ? Each such

edge has to satisfy $c = x_1 + x_2 + x_3 - a$, so c is fixed by the choice of u and the color (x_1, x_2, x_3) . Moreover, since $d = x_1 + x_2c + x_3c^2$, d is also fixed. Hence, there is at most one edge of each color adjacent to u . That is, we have a proper coloring.

Note now that each nondegenerate color (x_1, x_2, x_3) incident to a vertex (a, b) satisfies the equation $x_1 + x_2a + x_3a^2 = b$. Therefore, since there is at most one quadratic passing through any three points in $\mathbb{F}_q^2$, no three vertices can be incident to the same two colors. $\square$

Remark 3.6. If H is not a forest, $f_3(n, H)$ is always polynomially smaller than $n^{3/2}$. Indeed, if H is not a forest, it has a connected component H' with $e(H') \geq |H'|$. By Theorem 2.5, $f_3(n, H') = O(n^{3/2 - \Omega(1)})$, so the same then holds for H by Lemma 2.2.

We conclude this section by establishing the first part of Theorem 1.5(ii).

PROPOSITION 3.7. *For any tree T with m edges and any $k \geq 2$, $f_k(n, T) = \Omega(n^{\frac{m+1}{m}})$.*

Proof. For convenience of notation, we will instead prove the equivalent statement that a proper C -coloring of K_n with $C = o(n^{\frac{m+1}{m}})$ colors contains $\omega(1)$ repeats of T . The number of copies of T in K_n is $\Theta(n^{m+1}) = \omega(C^m)$. Thus, some m -tuple of colors will be repeated $\omega(1)$ times. Since the coloring is proper, every vertex appears in at most $|T| = m + 1$ copies of T with a fixed coloring. Hence, $\omega(1)$ of the above repeats will be vertex disjoint. $\square$

4. Cycles. We begin this section by proving two of our main results at once, Theorem 1.4 and the second part of Theorem 1.5(ii).

THEOREM 4.1. *For every bipartite graph H containing a cycle, there exists $k = k(H)$ such that*

$$f_k(n, H) = O(n).$$

For every m -edge tree T , there exists $k = k(T)$ such that

$$f_k(n, T) = O(n^{\frac{m+1}{m}}).$$

In the proof of this theorem, we will make use of Bukh's random algebraic method [5] (see also [6, 8]). It will be useful to briefly recall some basic terminology from algebraic geometry. Let $\overline{\mathbb{F}}_q$ stand for the algebraic closure of $\mathbb{F}_q$. A *variety* over $\overline{\mathbb{F}}_q$ is a set of the form

$$W = \{x \in \overline{\mathbb{F}}_q^t : f_1(x) = \cdots = f_s(x) = 0\}$$

for a collection of multivariate polynomials $f_1, \dots, f_s : \overline{\mathbb{F}}_q^t \rightarrow \overline{\mathbb{F}}_q$. We also write $W(\mathbb{F}_q) := W \cap \mathbb{F}_q^t$. The variety W is said to be *defined* over $\mathbb{F}_q$ if the coefficients of $f_1, \dots, f_s$ are in $\mathbb{F}_q$. Finally, we say that W has *complexity* at most M if s, t , and the degrees of the polynomials f_i are all at most M . We will repeatedly use the following lemma, a consequence of the well-known Lang–Weil bound [24].

LEMMA 4.2 ([6, Lemma 2.7]). *Suppose W and D are varieties over $\overline{\mathbb{F}}_q$ of complexity at most M which are defined over $\mathbb{F}_q$. Then one of the following holds for all q sufficiently large in terms of M :*

- $|W(\mathbb{F}_q) \setminus D(\mathbb{F}_q)| > q/2$ or
- $|W(\mathbb{F}_q) \setminus D(\mathbb{F}_q)| < c$, where $c = c_M$ depends only on M .

Proof of Theorem 4.1. By Remark 1.2, it suffices to prove that, for a cycle or a tree H with $(v, e) := (|V(H)|, |E(H)|)$, there exists k such that

$$f_k(n, H) = O(n^{v/e}).$$

Suppose first that H is a cycle, so $v = e$. Let $n \leq q = (1 + o(1))n$ be a prime, guaranteed for large n by the prime number theorem. We also let d and t be positive integers, with d chosen to be sufficiently large in terms of t and t sufficiently large in terms of v . We write $\mathcal{P}_d$ for the set of all two-variable polynomials over $\mathbb{F}_q$ of degree at most d and let g be a polynomial taken from $\mathcal{P}_d$ uniformly at random. Such a polynomial can be generated by selecting a coefficient $a_{d_1 d_2} \in \mathbb{F}_q$ independently at random for each pair of nonnegative integers (d_1, d_2) with $d_1 + d_2 \leq d$ and writing

$$g(X_1, X_2) := \sum a_{d_1 d_2} X_1^{d_1} X_2^{d_2}.$$

We will make repeated use of the following property of such random polynomials. Because we will need it below, we state the result for the more general case of t -variate random polynomials.

LEMMA 4.3 ([6, Lemma 2.3]). *Suppose that $q > \binom{m}{2}$ and $d \geq m - 1$. Then, if g is a random polynomial from $\mathcal{P}_d$ and $x_1, \dots, x_m$ are m distinct points in $\mathbb{F}_q^t$,*

$$\mathbb{P}[g(x_i) = 0 \text{ for all } i = 1, \dots, m] = 1/q^m.$$

Define a q -coloring of $G = K_{[n]}$ by assigning each edge ij with $i < j$ the color $g(i, j)$. We claim that with positive probability this coloring has the following properties:

- (Q1) Every (not necessarily proper) coloring of H occurs fewer than $k = k(H)$ times.
- (Q2) The coloring is $2d$ -bounded.

Given a coloring satisfying these conditions, we can apply Proposition 2.4 to obtain a proper coloring with at most $(2d + 1)q = O(n)$ colors and no k -repeats of H .

Suppose now that H has vertices $1, \dots, v$. We consider all copies of H in G of fixed order and color type. That is, given an ordering $(\tau_1, \dots, \tau_v)$ of $1, \dots, v$ and a sequence of e colors $(b_{ij} : ij \in E(H))$, we consider only those copies of H in G where the images $y_1, \dots, y_v$ of $1, \dots, v$ in $[n]$ satisfy $y_{\tau_1} < \dots < y_{\tau_v}$ and the edge between y_i and y_j is colored with color b_{ij} for all $ij \in E(H)$.

Let W be the variety defined by the system of equations

$$(4.1) \quad g(x_{i,j,1}, x_{i,j,2}) = b_{ij} \text{ for } ij \in E(H)$$

with v variables $x_1, \dots, x_v$, where each $(x_{i,j,1}, x_{i,j,2})$ equals (x_i, x_j) if $\tau_i < \tau_j$ and (x_j, x_i) otherwise. Let D be the variety defined by the equation

$$\prod_{1 \leq i \neq j \leq v} (x_i - x_j) = 0.$$

Note that the number of copies of H of given type is at most $S = |W(\mathbb{F}_q) \setminus D(\mathbb{F}_q)|$, i.e., the random variable counting the number of “nondegenerate” solutions to the system of equations (4.1).

By Lemma 4.3, for any $s \leq d + 1$ distinct points $y_1, \dots, y_s$ in $\mathbb{F}_q^2$ and any $b_1, \dots, b_s \in \mathbb{F}_q$, we have

$$\mathbb{P}[g(y_i) = b_i \text{ for all } i = 1, \dots, s] = 1/q^s.$$

Now consider S^t , observing that it counts ordered collections of t (potentially overlapping or identical) copies of H . Note that the graph H_t spanned by any collection of t copies of H will satisfy $e(H_t) \geq |H_t|$, since each connected component of H_t contains a cycle. Therefore, for d sufficiently large in terms of v and t ,

$$\mathbb{E}[S^t] \leq \sum_{s=v}^{vt} \frac{s^{vt} q^s}{q^s} < (vt)^{2vt},$$

where the factor s^{vt} is an upper bound for the number of ordered ways of placing t ordered copies of H on a fixed set of s vertices.

Since W and D are defined over $\mathbb{F}_q$ and their complexities are bounded by a function of d and v , Lemma 4.2 implies that there exists a constant $K = K(d, v)$ such that the random variable $S = |W(\mathbb{F}_q) \setminus D(\mathbb{F}_q)|$ satisfies either $S < K$ or $S > q/2$. Thus, by Markov's inequality,

$$\mathbb{P}[S \geq K] = \mathbb{P}[S > q/2] = \mathbb{P}[S^t > (q/2)^t] \leq \frac{\mathbb{E}[S^t]}{(q/2)^t} < (vt)^{2vt} (q/2)^{-t}.$$

This gives an upper bound on the probability of having at least K distinct copies of H for any given order and color type. Since there are at most $v!$ different order types and at most $q^v = q^v$ choices for the sequence of colors, the union bound implies that, for t sufficiently large in terms of v , there are with high probability fewer than K distinct copies of H for each order and color type. Therefore, writing $k = v!K$, we see that with high probability the coloring has fewer than k copies of H for each color type and so it satisfies (Q1) with probability $1 - o(1)$.

To see that it also satisfies (Q2), we have to show that with high probability, for each fixed i and a , the equations $g(x, i) = a$ and $g(i, y) = a$ have at most d solutions, i.e., that each of $g(x, i)$ and $g(i, y)$ has a nonconstant coefficient not equal to zero.

Consider $g(x, i)$ and note that, when viewing g as a polynomial in $\mathbb{F}_q[X_2][X_1]$, the coefficient of X_1^p is a random univariate polynomial $h_p(X_2)$ of degree $d - p$ with the coefficient for each different p being independent. Hence, the probability that $h_p(i)$ is 0 for all $p \geq 1$ is q^{-d} . Since there are at most q^2 choices for i and a , the union bound implies that with high probability $g(x, i) = a$ has at most d solutions for all i and a . Together with the analogous result for $g(i, y) = a$, this establishes (Q2) with probability $1 - o(1)$. By the union bound, with positive probability the coloring satisfies both properties (Q1) and (Q2), as claimed.

Suppose now that H is a tree with $m \geq 1$ edges. Let q be a prime with $n^{1/m} \leq q = (1 + o(1))n^{1/m}$. Let d and t again be positive integers, with d chosen to be sufficiently large in terms of t and t sufficiently large in terms of m , while $\mathcal{P}_d$ is now the set of all $2m$ -variable polynomials over $\mathbb{F}_q$ of degree at most d . Let $g_1, \dots, g_{m+1}$ be $m+1$ polynomials taken from $\mathcal{P}_d$ independently and uniformly at random.

Fix an ordering $\prec$ on $\mathbb{F}_q^m$ and associate $V(K_n)$ with a subset of $\mathbb{F}_q^m$. We define a q^{m+1} -coloring on $\binom{\mathbb{F}_q^m}{2}$ by assigning each edge ij with $i \prec j$ the color $(g_1(i, j), \dots, g_{m+1}(i, j))$, where we interpret (i, j) as an element of $\mathbb{F}_q^m \times \mathbb{F}_q^m \cong \mathbb{F}_q^{2m}$. As in the cycle case, we claim that with positive probability the following statements hold (in which case we are again done by Proposition 2.4):

(Q1) Every (not necessarily proper) coloring of H occurs fewer than $k = k(H)$ times.

(Q2') The coloring is k' -bounded for some $k' = k'(H)$.

Define order and color types as before, but with respect to the ordering $\prec$. For a fixed order type $(\tau_1, \dots, \tau_{m+1})$ and color type $(b_{ij} : ij \in E(H))$, let W be the variety

defined by the system of equations

$$g_\ell(x_{i,j,1}, x_{i,j,2}) = b_{ij}^{(\ell)} \text{ for } ij \in E(H) \text{ and } \ell \in [m+1]$$

with $m+1$ variables $x_1, \dots, x_{m+1} \in \mathbb{F}_q^m$, where each $(x_{i,j,1}, x_{i,j,2})$ equals (x_i, x_j) if $\tau_i < \tau_j$ and (x_j, x_i) otherwise and $b_{ij} = (b_{ij}^{(1)}, \dots, b_{ij}^{(m+1)})$. For each $\{i, j\} \in \binom{[m+1]}{2}$, let D_{ij} be the variety defined by the system of equations

$$x_i^{(\ell)} = x_j^{(\ell)} \text{ for } \ell \in [m],$$

where $x_i^{(\ell)}$ and $x_j^{(\ell)}$ are the ℓ th coordinates of x_i and $x_j \in \mathbb{F}_q^m$, respectively, and let

$$D = \bigcup_{1 \leq i \neq j \leq m+1} D_{ij}.$$

We again have that the number of copies of H of given type is bounded above by $S = |W(\mathbb{F}_q) \setminus D(\mathbb{F}_q)|$.

By Lemma 4.3 and the independence of the g_ℓ , for any $s \leq d+1$ distinct points $y_1, \dots, y_s$ in $\mathbb{F}_q^{2m}$ and any choice of $b_i^{(\ell)} \in \mathbb{F}_q$ for each $i \in [s]$ and $\ell \in [m+1]$, we have

$$\mathbb{P}[g_\ell(y_i) = b_i^{(\ell)} \text{ for all } i \in [s] \text{ and } \ell \in [m+1]] = 1/q^{s(m+1)}.$$

Now consider S^t , observing that it counts ordered collections of t (potentially overlapping or identical) copies of H . Note that the graph H_t spanned by any collection of t copies of H will satisfy $e(H_t) \geq \frac{m}{m+1}|H_t|$, since each connected component of H_t either is a tree with at least m edges or contains a cycle. Therefore, for d sufficiently large in terms of m and t ,

$$\mathbb{E}[S^t] \leq \sum_{s=m+1}^{(m+1)t} \frac{s^{(m+1)t} q^{ms}}{q^{\frac{m}{m+1}s(m+1)}} < (2mt)^{4mt}.$$

Since W and D are defined over $\mathbb{F}_q$ and their complexities are bounded by a function of d and m , Lemma 4.2 implies that there exists a constant $K = K(d, m)$ such that the random variable $S = |W(\mathbb{F}_q) \setminus D(\mathbb{F}_q)|$ satisfies either $S < K$ or $S > q/2$. Thus, by Markov's inequality,

$$\mathbb{P}[S \geq K] = \mathbb{P}[S > q/2] = \mathbb{P}[S^t > (q/2)^t] \leq \frac{\mathbb{E}[S^t]}{(q/2)^t} < (2mt)^{4mt} (q/2)^{-t}.$$

Hence, as in the case of cycles, writing $k = (m+1)!K$, we obtain that with high probability the coloring has fewer than k copies of H for each color type and so it satisfies (Q1) with probability $1 - o(1)$.

To see that it also satisfies (Q2'), we have to show that with high probability, for each fixed $i \in \mathbb{F}_q^m$ and $a = (a_\ell)_{\ell \in [m+1]} \in \mathbb{F}_q^{m+1}$, the systems of equations

$$g_\ell(x, i) = a_\ell \text{ for } \ell \in [m+1]$$

and

$$g_\ell(i, y) = a_\ell \text{ for } \ell \in [m+1]$$

have at most $k'/2$ solutions for some $k' = k'(d)$. For this, we again resort to the random algebraic method.

Fix $i \in \mathbb{F}_q^m$. For each $\ell \in [m+1]$, define $g'_\ell(x) := g_\ell(x, i)$. It is not hard to see that $g'_\ell(x)$ is a uniformly random m -variable polynomial of degree d . Note also that the set

$$(4.2) \quad \{x \in \mathbb{F}_q^m : g'_\ell(x) = a_\ell \text{ for all } \ell \in [m+1]\}$$

is of the form $W'(\mathbb{F}_q)$ for a variety W' of complexity bounded by a function of d and m . Moreover, by Lemma 4.3 and the independence of the g'_ℓ , for any $s \leq d+1$ distinct points $x_1, \dots, x_s \in \mathbb{F}_q^m$ and any choice of $a_j^{(\ell)} \in \mathbb{F}_q$ for each $j \in [s]$ and $\ell \in [m+1]$, we have

$$\mathbb{P}[g'_\ell(x_j) = a_j^{(\ell)} \text{ for all } j \in [s] \text{ and } \ell \in [m+1]] = 1/q^{s(m+1)}.$$

Consider the random variable $S' = |W'(\mathbb{F}_q)|$ and observe that S'^t counts ordered collections of t (not necessarily distinct) solutions to (4.2). Therefore, for d sufficiently large in terms of t ,

$$\mathbb{E}[S'^t] \leq \sum_{s=1}^t \frac{t^s q^{sm}}{q^{s(m+1)}} = \sum_{s=1}^t \left(\frac{t}{q}\right)^s < 1.$$

By Lemma 4.2, there exists a constant $K' = K'(d, m)$ such that S' satisfies either $S' < K'$ or $S' > q/2$. Thus, by Markov's inequality, provided t is sufficiently large in terms of m ,

$$\mathbb{P}[S' \geq K'] = \mathbb{P}[S' > q/2] = \mathbb{P}[S'^t > (q/2)^t] \leq \frac{\mathbb{E}[S'^t]}{(q/2)^t} < \frac{2^t}{q^t} < \frac{1}{q^{3m+1}}.$$

In other words, the system of equations $g_\ell(x, i) = a_\ell$ with $\ell \in [m+1]$ has with high probability at most K' solutions. Together with the analogous result for the system of equations $g_\ell(i, y) = a_\ell$ with $\ell \in [m+1]$, a union bound over all $i \in \mathbb{F}_q^m$ and $a = (a_\ell)_{\ell \in [m+1]} \in \mathbb{F}_q^{m+1}$ establishes (Q2') with high probability for $k' = 2K'$. By another simple application of the union bound, with positive probability the coloring satisfies both properties (Q1) and (Q2'), as claimed. $\square$

We finish this section by completing the proof of Theorem 1.3. Indeed, Theorem 2.5 implies that $f_2(n, C_6) = O(n^{5/3})$, while the next lemma says that $f_2(n, C_6) = \Omega(n^{4/3})$. Together, these results establish Theorem 1.3(iv) with $H = C_6$.

THEOREM 4.4. $f_2(n, C_6) = \Omega(n^{4/3})$.

Proof. Suppose that $C = \gamma n^{4/3}$, where γ is a small constant. Suppose also that n is taken sufficiently large and we have a C -coloring of $G = K_n$. For simplicity in our notation, we will also assume that n is even. Take an arbitrary equipartition of $V(G)$ into parts X and Y . Let $U := \binom{X}{2}$, $W := \binom{Y}{2}$ and let F be the following auxiliary graph, similar to the graph used in the proof of Theorem 3.2. The vertex set is $V(F) := U \cup W$, and $F = F[U, W]$ is bipartite with $uw \in E(F)$ for $u = \{x_1, x_2\}$, $w = \{y_1, y_2\}$ if and only if at least one of $\{x_1 y_1, x_2 y_2\}$, $\{x_1 y_2, x_2 y_1\}$ is a monochromatic matching in G . Therefore,

$$|F| = 2 \binom{n/2}{2} = (1 + o(1)) \frac{n^2}{4} \geq \frac{n^2}{5}$$

and $e(F) \geq E_2/2$, where E_2 is the number of monochromatic matchings of size 2 between X and Y , as every such matching gives rise to an edge of F and each edge is counted at most twice. By the convexity of binomial coefficients, we obtain

$$(4.3) \quad e(F) \geq \frac{E_2}{2} \geq \frac{C}{2} \binom{\frac{n^2}{5C}}{2} > \frac{n^4}{200C} = \frac{n^{8/3}}{200\gamma} > \frac{|F|^{4/3}}{200\gamma}.$$

The *theta graph* $\theta_{3,\ell}$ is the bipartite graph composed of ℓ internally disjoint paths of length 3 sharing the same pair of endpoints. For example, $\theta_{3,2} = C_6$. A theorem of Faudree and Simonovits [13] (see also the recent paper [7]) states that, for every $\ell \geq 2$, $\text{ex}(n, \theta_{3,\ell}) = O_\ell(n^{4/3})$. Hence, by this and (4.3), putting $\ell = 60$, there exists γ_0 such that if $\gamma < \gamma_0$, then F contains a copy of $\theta_{3,60}$, noting that, since F is bipartite, its endpoints cannot be in the same part.

Suppose now that $u \in U$ and $w \in W$ are the endpoints of our copy of $\theta_{3,60}$, the neighbors of u and w are $w_1, \dots, w_\ell$ and $u_1, \dots, u_\ell$, respectively, and uw_iu_iw is a path for each $i = 1, 2, \dots, \ell$. We claim that, after relabeling, there are three indices 1, 2, 3 such that the eight vertex pairs in G encoded by the vertices $u, w, u_1, u_2, u_3, w_1, w_2, w_3$ are disjoint. Since $u, u_1, u_2, u_3 \in U$ and $w, w_1, w_2, w_3 \in W$, we only need to make sure that the vertex pairs are disjoint in each part.

To this end, note that, by the same hypergraph coloring argument used in the proof of Theorem 3.2, any set of t vertices $w_{i_1}, \dots, w_{i_t} \in N_F(u)$ contains a subset of size at least $t/3$ in which all vertices correspond to disjoint pairs in G . The same holds for any set of t vertices in $N_F(w)$.

Thus, at least 15 of the vertices $w_1, \dots, w_\ell$ correspond to pairs in Y that are disjoint from each other and from w (since there are at least 20 of the w_i whose corresponding pairs are disjoint from each other and w overlaps with at most two of these). Relabeling, we can assume that these vertices are $w_1, \dots, w_{15}$. Applying the same argument to $u_1, \dots, u_{15}$ and relabeling if necessary, we obtain three vertices u_1, u_2, u_3 corresponding to pairs in X which are disjoint from each other and from u .

Let $u := \{x^1, x^2\}$, $w := \{y^1, y^2\}$, $u_i := \{x_i^1, x_i^2\}$, and $w_i := \{y_i^1, y_i^2\}$ for $i = 1, 2, 3$. By the definition of F , without loss of generality we may assume that we have monochromatic pairs $\{x^1y_i^1, x^2y_i^2\}$ and $\{y_i^1x_i^1, y_i^2x_i^2\}$ for $i = 1, 2, 3$. For each of the three edges $u_iw \in E(F)$, we have two options: either the matching $\{x_i^1y^1, x_i^2y^2\}$ is monochromatic or the matching $\{x_i^1y^2, x_i^2y^1\}$ is. By the pigeonhole principle, there will be two indices, say $i = 1, 2$, where the same situation occurs. In the case where $\{x_i^1y^1, x_i^2y^2\}$ is monochromatic for $i = 1, 2$, we have a 2-repeat of C_6 with copies $x^1y_1^1x_1^1y_1^2x_2^2y_2^2x^2$ and $x^2y_1^2x_1^2y_2^1x_2^1y_2^2x^1$. If instead $\{x_i^1y^2, x_i^2y^1\}$ is monochromatic for $i = 1, 2$, our 2-repeat of C_6 consists of the copies $x^1y_1^1x_1^1y_2^1x_2^2y_2^1x^2$ and $x^2y_1^2x_1^2y_2^2x_2^1y_2^2x^1$. $\square$

5. Open problems.

Trees. Proposition 3.7 and Theorem 4.1 together imply that for each m -edge tree T , there exists a constant $k_0 = k_0(T)$ such that, for all $k \geq k_0$,

$$f_k(n, T) = \Theta(n^{\frac{m+1}{m}}).$$

It would be interesting to prove a tight bound for k_0 —we suspect that the answer might be $m+1$ for any m -edge tree. When T is the star with 2 edges, our Theorem 3.5 verifies this suspicion. When T is the star with m edges, this problem seems closely related to the Zarankiewicz problem for $K_{m,m+1}$, suggesting that it may already be very difficult for $m \geq 3$. An easier problem might be to give an upper bound for $f_3(n, T)$ when T is an m -edge tree which matches the lower bound $\Omega(n^{3/2}/\sqrt{m})$ of Theorem 3.2 up to an absolute constant factor.

Bipartite graphs containing a cycle. Corollary 2.6 tells us that if $e(H) \geq 2|H| - 2$, then $f_2(n, H) = O(n)$. How sharp is this bound? Recalling that $\theta_{3,\ell}$ is the bipartite graph consisting of ℓ internally disjoint paths of length 3 sharing the same

pair of endpoints, an extension of Theorem 4.4 implies that, for any $\ell \geq 2$,

$$f_2(n, \theta_{3,\ell}) = \Omega(n^{4/3}).$$

Since $e(\theta_{3,\ell}) = \frac{3}{2}|\theta_{3,\ell}| - 3$, we see that Corollary 2.6 cannot be improved to say that $e(H) \geq \frac{3}{2}|H| - 3$ implies that $f_2(n, H) = O(n)$. An interesting test case for deciding whether the latter bound can be pushed closer to $2|H|$ might be to study $f_2(n, K'_t)$, where K'_t is the 1-subdivision of the complete graph K_t . The extremal properties of these graphs have received close attention in the recent literature [9, 10, 19], and some of the ideas developed in these papers might also prove useful in our context.

Even cycles. There is an abundance of open problems regarding even cycles, with the simplest being whether $f_2(n, C_4) = \Theta(n)$. We have only shown that $f_2(n, C_4) = O(n^{3/2})$, as a particular instance of Theorem 2.5. It would also be interesting to determine the correct exponent for C_6 . That is, determine the exponent α , if it exists, such that $f_2(n, C_6) = \Theta(n^\alpha)$. Our results only show that $4/3 \leq \alpha \leq 5/3$. Finally, for longer cycles, we suspect that the exponent approaches 2. That is, is it true that for every $\varepsilon > 0$, there exists $m_0 = m_0(\varepsilon)$ such that, for all $m \geq m_0$, $f_2(n, C_{2m}) = \Omega(n^{2-\varepsilon})$? This problem seems to bear some relation to the even cycle case of Conjecture 3.1 from [17].

Nonbipartite graphs. When H is nonbipartite, we know, by the remark after Theorem 2.3, that $f_2(n, H) = n$ for n odd. On the other hand, when n is even, we have only shown that $f_2(n, H) \leq n + 1$. However, in many cases, this bound can be improved to $f_2(n, H) = n - 1$. To see this, we start with the coloring of K_{n-1} with $n - 1$ colors used in the proof of Theorem 2.3 and then consider the unique extension of this coloring to K_n . The resulting coloring does not contain repeats of “most” nonbipartite H , such as graphs containing two disjoint odd cycles, so that $f_2(n, H) = n - 1$ for these graphs. Similarly, this coloring allows us to conclude that $f_3(n, H) = n - 1$ for n even and any nonbipartite H . However, it remains to determine $f_2(n, H)$ exactly in the natural case where n is even and H is an odd cycle. Curiously, when H is a triangle, each 1-factorization of K_n has $\binom{n-1}{3}$ triples of possible colors, which is less than $\binom{n}{3}$, the total number of triangles. Hence, by the pigeonhole principle, there will be two color-isomorphic triangles, but they may not be vertex disjoint.

Repeats of other patterns. Our results all yield repeated rainbow copies, where the edges in each copy all receive a different color, but one could also ask for repeats of other patterns. For instance, what is the smallest number of colors in a proper edge-coloring which does not contain a 2-repeat of the properly 2-colored path of length 3? From the construction in Theorem 3.5, it is not hard to deduce an upper bound of $O(n^{3/2})$, whereas a simple counting argument yields a lower bound of $\Omega(n^{4/3})$. We believe the latter to be tight.

A problem in Ramsey theory. Our original motivation for studying repeats came from a problem in generalized Ramsey theory raised by Krueger [23, Problem 1.2]. His question asks for the minimum number of colors in an edge-coloring of K_n such that every copy of P_t , the path with t vertices, where t is odd, contains at least $\frac{t+1}{2}$ colors. In our context, the most closely related problem is to study lower bounds for the appearance, in proper edge-colorings, of the pattern consisting of two repeats of the path $P_{(t+1)/2}$ sharing an endpoint. At present, the best lower bound for Krueger’s problem is $\Omega(n^{3/2})$ for all $t \geq 9$. It would be interesting to determine whether the exponent tends to 2 as t tends to infinity both for this problem and ours.

Hypergraphs. Finally, we note that there are several ways to generalize our problem to the setting of r -uniform hypergraphs, depending on how we define a proper coloring. Indeed, for any $1 \leq t < r$, one could ask that any two hyperedges sharing at least t vertices receive distinct colors, resulting in a family of problems. It may also be interesting to study proper colorings of Steiner Triple Systems, looking for repeats of fixed linear 3-graphs.

Acknowledgment. We are extremely grateful to Sean English and Bob Krueger for spotting an error in an earlier version of this paper and suggesting a fix.

REFERENCES

- [1] N. ALON, A. POKROVSKIY, AND B. SUDAKOV, *Random subgraphs of properly edge-coloured complete graphs and long rainbow cycles*, Israel J. Math., 222 (2017), pp. 317–331.
- [2] N. ALON AND J. H. SPENCER, *The Probabilistic Method*, 4th ed., Wiley Ser. Discrete Math. Optim., John Wiley & Sons, Hoboken, NJ, 2016.
- [3] I. BALLA, A. POKROVSKIY, AND B. SUDAKOV, *A remark on Hamilton cycles with few colors*, Mosc. J. Comb. Number Theory, 7 (2017), pp. 73–77.
- [4] F. BENZING, A. POKROVSKIY, AND B. SUDAKOV, *Long directed rainbow cycles and rainbow spanning trees*, European J. Combin., 88 (2020), 103102.
- [5] B. BUKH, *Random algebraic construction of extremal graphs*, Bull. Lond. Math. Soc., 47 (2015), pp. 939–945.
- [6] B. BUKH AND D. CONLON, *Rational exponents in extremal graph theory*, J. Eur. Math. Soc. (JEMS), 20 (2018), pp. 1747–1757.
- [7] B. BUKH AND M. TAIT, *Turán numbers of theta graphs*, Combin. Probab. Comput., 29 (2020), pp. 495–507.
- [8] D. CONLON, *Graphs with few paths of prescribed length between any two vertices*, Bull. Lond. Math. Soc., 51 (2019), pp. 1015–1021.
- [9] D. CONLON, O. JANZER, AND J. LEE, *More on the extremal number of subdivisions*, Combinatorica, to appear.
- [10] D. CONLON AND J. LEE, *On the extremal number of subdivisions*, Int. Math. Res. Not. IMRN, 2021 (2021), pp. 9122–9145.
- [11] M. COULSON AND G. PERARNAU, *A rainbow Dirac’s theorem*, SIAM J. Discrete Math., 34 (2020), pp. 1670–1692, <https://doi.org/10.1137/18M1218881>.
- [12] S. EHARD, S. GLOCK, AND F. JOOS, *A rainbow blow-up lemma for almost optimally bounded edge-colourings*, Forum Math. Sigma, 8 (2020), e37.
- [13] R. J. FAUDREE AND M. SIMONOVITS, *On a class of degenerate extremal graph problems*, Combinatorica, 3 (1983), pp. 83–93.
- [14] P. GAO, R. RAMADURAI, I. WANLESS, AND N. WORMALD, *Full Rainbow Matchings in Graphs and Hypergraphs*, preprint, <https://arxiv.org/abs/1709.02665>, 2018.
- [15] S. GLOCK AND F. JOOS, *A rainbow blow-up lemma*, Random Structures Algorithms, 56 (2020), pp. 1031–1069.
- [16] S. GLOCK, D. KÜHN, R. MONTGOMERY, AND D. OSTHUS, *Decompositions into isomorphic rainbow spanning trees*, J. Combin. Theory Ser. B, 146 (2021), pp. 439–484.
- [17] A. GRZESIK, O. JANZER, AND Z. L. NAGY, *The Turán number of blow-ups of trees*, J. Combin. Theory Ser. B, to appear.
- [18] T. GUSTAVSSON, *Decompositions of Large Graphs and Digraphs with High Minimum Degree*, Doctoral Dissertation, Department of Mathematics, University of Stockholm, 1991.
- [19] O. JANZER, *Improved bounds for the extremal number of subdivisions*, Electron. J. Combin., 26 (2019), P3.3.
- [20] P. KEEVASH AND K. STADEN, *Ringel’s Tree Packing Conjecture in Quasirandom Graphs*, preprint, <https://arxiv.org/abs/2004.09947>, 2020.
- [21] P. KEEVASH AND L. YEPREMYAN, *Rainbow matchings in properly-colored multigraphs*, SIAM J. Discrete Math., 32 (2018), pp. 1577–1584, <https://doi.org/10.1137/17M1151742>.
- [22] J. KIM, D. KÜHN, A. KUPAVSKII, AND D. OSTHUS, *Rainbow structures in locally bounded colourings of graphs*, Random Structures Algorithms, 56 (2020), pp. 1171–1204.
- [23] R. KRUEGER, *Generalized Ramsey numbers: Forbidding paths with few colors*, Electron. J. Combin., 27 (2020), 1.44.
- [24] S. LANG AND A. WEIL, *Number of points of varieties in finite fields*, Amer. J. Math., 76 (1954), pp. 819–827.

- [25] R. MONTGOMERY, A. POKROVSKIY, AND B. SUDAKOV, *Decompositions into spanning rainbow structures*, Proc. Lond. Math. Soc., 119 (2019), pp. 899–959.
- [26] R. MONTGOMERY, A. POKROVSKIY, AND B. SUDAKOV, *Embedding rainbow trees with applications to graph labelling and decompositions*, J. Eur. Math. Soc. (JEMS), 22 (2020), pp. 3101–3132.
- [27] R. MONTGOMERY, A. POKROVSKIY, AND B. SUDAKOV, *A Proof of Ringel’s Conjecture*, preprint, <https://arxiv.org/abs/2001.02665>, 2020.
- [28] A. POKROVSKIY, *An approximate version of a conjecture of Aharoni and Berger*, Adv. Math., 333 (2018), pp. 1197–1241.
- [29] A. POKROVSKIY AND B. SUDAKOV, *Linearly many rainbow trees in properly edge-coloured complete graphs*, J. Combin. Theory Ser. B, 132 (2018), pp. 134–156.
- [30] A. POKROVSKIY AND B. SUDAKOV, *A counterexample to Stein’s equi- n -square conjecture*, Proc. Amer. Math. Soc., 147 (2019), pp. 2281–2287.
- [31] R. M. WILSON, *An existence theory for pairwise balanced designs. III. Proof of the existence conjectures*, J. Combin. Theory Ser. A, 18 (1975), pp. 71–79.