hexagon:2611.00763v1

A COUNTEREXAMPLE TO ERDŐS PROBLEM 108 VIA ARC GRAPHS

JENSEN KOHLMEYER AND LIAM KRUER

Abstract. For every positive integer M, we construct a finite triangle-free graph G with $\chi(G) \geq M$ such that every ordinary subgraph of G containing no four-cycle is 6-colorable. Consequently, no chromatic threshold forces a subgraph of girth at least five and chromatic number at least seven. This gives a negative answer to Erdős problem 108.

The proof combines two ingredients. First, a four-cycle-free subgraph of an arc graph can be partitioned into a part controlled by a bounded-degree subgraph of the underlying graph and a forest. Second, a multipartite random construction gives graphs of large chromatic number for which every bounded-degree subgraph is 4-colorable, while retaining an acyclic orientation whose outdegree bound is independent of the prescribed degree cap. We give explicit integer part sizes and finite probability estimates. The accompanying Lean proof verifies the negative answer to the pinned Formal Conjectures statement using the permitted standard axioms. The written proof includes the triangle-free strengthening.

1. Introduction

The high-girth subgraph problem of Erdős and Hajnal asks whether, for every pair of integers $r \geq 4$ and $k \geq 2$, there is an integer $f(k, r)$ such that every graph of chromatic number at least $f(k, r)$ contains a subgraph of girth at least r and chromatic number at least k. The question appears as Problem 108 in Bloom's catalogue [1] and is formalized in the Formal Conjectures library [2].

Throughout this paper, a subgraph may be obtained by deleting vertices and edges. A four-cycle-free graph excludes a copy of C_4 as an ordinary subgraph. A four-cycle with chords is still forbidden.

Theorem 1.1. *For every integer $M \geq 1$, there is a finite nonempty triangle-free simple graph G such that $\chi(G) \geq M$ and every four-cycle-free subgraph H of G satisfies $\chi(H) \leq 6$.*

Corollary 1.2. *There is no integer f such that every graph of chromatic number at least f contains a subgraph of girth at least five and chromatic number at least seven. In particular, the assertion in Erdős problem 108 is false.*

Proof. Given a proposed threshold f, apply Theorem 1.1 with $M = \max\{1, f\}$. Every subgraph of girth at least five is four-cycle-free, so its chromatic number is at most six. □

The construction starts with the multipartite random model used by Janzer, Steiner and Sudakov in their work on chromatic number and regular subgraphs [3]. We adapt its separated-part-size method to obtain a statement with two independent parameters, a chromatic parameter q and a degree cap K. The outdegree bound depends only on q. This permits us to choose K after fixing that bound and then apply an arc-graph argument. All estimates needed for this adaptation are proved here.

Section 2 gives the arc-graph lemmas. Section 3 constructs the base graph, including the size estimates and both probability bounds. Section 4 proves Theorem 1.1. Section 5 describes the formal proof and its verification records.

Date: September 15, 2026.

2020 *Mathematics Subject Classification.* Primary 05C15; Secondary 05C38, 05C80.

Key words and phrases. Erdős–Hajnal conjecture, chromatic number, girth, arc graph, shift graph, probabilistic method, Lean formalization.

Notation. All graphs used in the construction are finite and simple. We write $e_F(X)$ for the number of edges in $F[X]$, and $e_F(X, Y)$ for the number of edges between disjoint vertex sets X and Y. The maximum degree of a graph is denoted by Δ. A graph is 3-degenerate if every nonempty subgraph has a vertex of degree at most three. Such a graph is 4-colorable. For an acyclic digraph D, let $U(D)$ be its underlying undirected graph. Forests have infinite girth in the written proof.

2. Arc graphs and bounded-degree selections

Definition 2.1. Let D be an acyclic orientation of a simple graph. Its arc graph $L(D)$ has the arcs of D as vertices, with two vertices adjacent precisely when the corresponding arcs concatenate. Thus (u, v) and (v, w) are adjacent. We orient this edge from (u, v) to (v, w).

Fix a topological ordering of D. Along every oriented edge of $L(D)$, the tail of the corresponding arc strictly increases in this ordering. The displayed orientation of $L(D)$ is therefore acyclic.

Lemma 2.2. *The graph $L(D)$ is triangle-free.*

Proof. An acyclic orientation of a triangle has vertices a, b, c with edges $a \to b$, $b \to c$ and $a \to c$. If $a = (u, v)$ and $b = (v, w)$, the edge $b \to c$ makes the tail of c equal to w, whereas $a \to c$ makes it equal to v. Hence $v = w$, contradicting the fact that b is an arc of a simple graph. □

Lemma 2.3. *If $L(D)$ is t-colorable, then $U(D)$ is 2^t-colorable. In particular,*

$$\chi(U(D)) \leq 2^{\chi(L(D))}. \tag{1}$$

Proof. Fix a proper coloring of $L(D)$ with colors in $\{1, \dots, t\}$. Assign to each vertex v of D the set of colors occurring on arcs directed out of v. If (u, v) has color c, the set at u contains c. Every arc directed out of v is adjacent to (u, v), so the set at v excludes c. Adjacent vertices of $U(D)$ therefore receive distinct subsets of the t-element color set. □

Lemma 2.4 (Arc selection)**.** *Suppose that D is acyclic, has maximum outdegree at most d, and every subgraph of $U(D)$ with maximum degree at most*

$$K_d = d + \binom{d}{2} \tag{2}$$

is a-colorable. Then every four-cycle-free subgraph of $L(D)$ is $(a + 2)$-colorable.

Proof. Let $H \subseteq L(D)$ be four-cycle-free, and restrict the natural orientation to H. Partition its vertex set into S and T, where S consists of the arc vertices having at least two outgoing neighbors in H, and T consists of the remaining arc vertices.

Fix a vertex v of D. For every incoming arc (u, v) belonging to S, choose an unordered pair of its outgoing neighbors in H. These neighbors are arcs of the form (v, w). Two distinct incoming arcs cannot receive the same pair, since the two incoming arcs and the two chosen outgoing arcs would form a four-cycle. Consequently at most

$$\binom{d_D^+(v)}{2} \leq \binom{d}{2}$$

incoming arcs at v belong to S.

Let D_S retain precisely the arcs of D corresponding to vertices in S. Its outdegrees are at most d and its indegrees are at most $\binom{d}{2}$. Hence $\Delta(U(D_S)) \leq K_d$. By hypothesis there is a proper a-coloring c of $U(D_S)$. Color an arc vertex $(u, v) \in S$ by $c(u)$. If (u, v) and (v, w) are adjacent in $H[S]$, the edge uv belongs to $U(D_S)$, so these two arc vertices receive different colors.

The inherited orientation of $H[T]$ is acyclic and every vertex has outdegree at most one. Such a graph is a forest. In an acyclic orientation any undirected cycle has a source on that cycle and therefore a vertex with two outgoing edges. Color $H[T]$ with two colors disjoint from the first palette. This gives an $(a + 2)$-coloring of H. □

Remark 2.5. The partition and the selected graph D_S depend on the edge-deleted subgraph H. The argument works for ordinary subgraphs. The indegree bound applies to the selected digraph D_S.

3. The multipartite base graph

Theorem 3.1. *For every pair of integers $q \geq 2$ and $K \geq 4$, there is a finite simple graph F with an acyclic orientation D such that*

(1) $\chi(F) > 2q$,
(2) *the maximum outdegree of D is at most $64q^2 - 1$, and*
(3) *every subgraph $J \subseteq F$ with $\Delta(J) \leq K$ is 3-degenerate, and hence 4-colorable.*

In particular, the outdegree bound in (2) is independent of K.

3.1. The random model and explicit sizes.

Put

$$C = 64q^2, \qquad h = 2(K+1), \qquad A = 16CK(K+1), \qquad b = h+2. \tag{3}$$

Choose the positive integer R, then define B and the part sizes by

$$R = (16C)^h 3^{2h+1} A + 3A + 8C + 2, \tag{4}$$

$$B = b^{C+1}, \qquad n_i = R^{B-b^{i+1}} \quad (0 \leq i < C). \tag{5}$$

Take disjoint sets $V_0, \dots, V_{C-1}$ with $|V_i| = n_i$ and no edges inside a part. For every $v \in V_i$ and every $j > i$, independently choose a uniformly distributed vertex $w(v,j) \in V_j$ and insert the edge $vw(v,j)$. Orient each edge from its earlier part to its later part. This orientation is acyclic and has maximum outdegree at most $C-1$.

Write $N = \sum_i n_i$ and $N_0 = R^B$. The following elementary estimates are the reason for the choice of exponents.

Lemma 3.2. *The sizes in* (5) *satisfy*

$$n_i \geq R \geq 8C, \qquad Rn_j \leq n_i \quad (i < j), \tag{6}$$

$$N \leq Cn_0 \leq An_0, \qquad N \leq N_0, \tag{7}$$

$$\sum_{j>i} n_j \leq n_i, \qquad \sum_{j \geq i} n_j \leq 2n_i. \tag{8}$$

For $1 \leq i < C$, put $L = n_{i-1}$. For every integer x with $1 \leq x \leq An_i$, one also has

$$3x \leq L, \qquad \frac{3^{2h+1} N_0^h x}{L^{h+1}} \leq (16C)^{-h}. \tag{9}$$

Proof. Since $b > 1$ and $i < C$, the exponent $B - b^{i+1}$ is a positive integer. For $i < j$ the two exponents differ by at least one, which proves (6). The sizes are decreasing. Since $R \geq C$ and each $n_i \leq R^{B-1}$, we obtain $N \leq CR^{B-1} \leq R^B = N_0$. The other inequalities in (7) use $A \geq C$. Moreover, $Rn_j \leq n_i$ for each $j > i$, so $R\sum_{j>i} n_j \leq Cn_i \leq Rn_i$, proving (8).

Fix $i \geq 1$. Since $R \geq 3A$, the successive-size inequality gives $3x \leq 3An_i \leq Rn_i \leq n_{i-1} = L$. Next, the identity $b = h+2$ gives

$$(h+1)(B - b^i) - hB - (B - b^{i+1}) = b^i.$$

Thus $L^{h+1} = R^{b^i} N_0^h n_i \geq RN_0^h n_i$. It follows that

$$\frac{3^{2h+1} N_0^h x}{L^{h+1}} \leq \frac{3^{2h+1} A}{R} \leq (16C)^{-h},$$

where the last inequality is immediate from (4). □

3.2. **The deterministic cut.** Consider the following local sparsity condition on a realization of F:

$$X \subseteq V_0 \cup \cdots \cup V_{i-1}, \quad 0 < |X| \le An_i \quad \Longrightarrow \quad e_F(X) < (1 + 1/h)|X|. \tag{10}$$

For $i = 0$ the condition is vacuous.

Lemma 3.3. *If* (10) *holds for every i, then every subgraph of F with maximum degree at most K is 3-degenerate.*

Proof. Suppose that $J \subseteq F$ has minimum degree at least four and maximum degree at most K, and let $s = |V(J)| > 0$. By (7), there is an index i with $s \le An_i$. Choose the largest such index. Partition $V(J)$ into X, Y, Z, according as the part index is less than, equal to, or greater than i. If $i = C - 1$, then $Z = \varnothing$. Otherwise, maximality gives $An_{i+1} < s$, and the decreasing sizes give $|Z| \le Cn_{i+1} < Cs/A = s/(16K(K + 1))$. In either case we may use the weaker bound

$$|Z| < \frac{2s}{16K(K+1)}. \tag{11}$$

Each vertex of X has at most one neighbor in V_i, while the degree cap gives at most $K|Y|$ edges incident with Y. Hence $e_J(X, Y) \le |X|$ and $e_J(X, Y) \le K|Y|$. Taking the weighted average with weights $K/(K + 1)$ and $1/(K + 1)$ yields

$$e_J(X, Y) \le \frac{K}{K+1}(|X| + |Y|) \le \frac{K}{K+1}s. \tag{12}$$

The set Y is independent. At most $K|Z|$ edges have an endpoint in Z, and the minimum-degree assumption gives $e(J) \ge 2s$. Therefore

$$\begin{aligned} e_J(X) &\ge 2s - e_J(X, Y) - K|Z| > \left(2 - \frac{K}{K+1} - \frac{1}{8(K+1)}\right) s \\ &= \left(1 + \frac{7}{8(K+1)}\right) s > (1 + 1/h)|X|. \end{aligned} \tag{13}$$

In particular X is nonempty, and $|X| \le s \le An_i$. Since $e_J(X) \le e_F(X)$, this contradicts (10). Thus there is no subgraph with all degrees between four and K. Every subgraph of a graph of maximum degree at most K still has maximum degree at most K, so every such graph is 3-degenerate. □

3.3. **The probability of a dense earlier set.**

Lemma 3.4. *The probability that* (10) *fails for some i is at most* $1/8$.

Proof. Fix $i \ge 1$ and $1 \le x \le An_i$, and put $L = n_{i-1}$, $m = \lceil (1 + 1/h)x \rceil$, and $\delta = 1/(16C)$. For a specified collection of m distinct possible edges among the earlier parts, either two edges demand different choices from the same pair (v, j), in which case the event is impossible, or they impose m independent choices. Every target part has size at least L, so the probability that all the specified edges occur is at most L^{-m}.

There are at most $\binom{N_0}{x}$ possible sets X of size x, and at most $\binom{x^2}{m}$ choices of the required edges. Thus the probability that an earlier set of size x violates (10) is at most

$$\binom{N_0}{x}\binom{x^2}{m}L^{-m} \le \left(\frac{3N_0}{x}\right)^x \left(\frac{3x}{L}\right)^m. \tag{14}$$

Here we used $\binom{n}{k} \le (en/k)^k \le (3n/k)^k$ for $k \ge 1$, together with $m \ge x$. If the prescribed number of edges is impossible, the same estimate holds with probability zero.

Set $a = 3N_0/x$ and $z = 3x/L$. By Lemma 3.2, $0 < z \le 1$. Since $hm \ge (h + 1)x$, raising the right side of (14) to the integer power h gives

$$(a^x z^m)^h \le (a^h z^{h+1})^x = \left(\frac{3^{2h+1}N_0^h x}{L^{h+1}}\right)^x \le \delta^{hx}.$$

It follows that (14) is at most δ^x. Since $\delta \le 1/2$, summing over the possible positive sizes gives a failure probability at most

$$\sum_{x\ge 1} \delta^x = \frac{\delta}{1-\delta} \le 2\delta = \frac{1}{8C}$$

for a fixed i. Summing over at most C indices proves the result. □

3.4. **The probability of a heavy independent set.** Assign weight $1/n_i$ to each vertex of V_i. For a vertex set I, write

$$p_i(I) = \frac{|I \cap V_i|}{n_i}, \qquad w(I) = \sum_{i=0}^{C-1} p_i(I). \tag{15}$$

The total vertex weight is C.

Lemma 3.5. *The probability that F has an independent set I with $w(I) \ge C/(2q)$ is at most $1/8$.*

Proof. Suppose that $w(I) \ge C/(2q)$. Choose the first index i with $p_i(I) \ge 1/(4q)$. Such an index exists. Earlier parts contribute less than $C/(4q)$ and $p_i(I) \le 1$, so

$$\sum_{j>i} p_j(I) > \frac{C}{4q} - 1 = 16q - 1 \ge 15q. \tag{16}$$

The part of I in indices at least i is still independent and preserves both inequalities. We may therefore count only sets in this suffix.

Fix a suffix set I satisfying these inequalities, and put $t = |I \cap V_i| \ge n_i/(4q)$. For I to be independent, all choices from its t vertices in V_i to each later part V_j must avoid $I \cap V_j$. These choices are independent, so

$$\mathbb{P}(I \text{ is independent}) \le \prod_{j>i} (1 - p_j(I))^t \le \exp\left(-t \sum_{j>i} p_j(I)\right) \le e^{-3n_i}. \tag{17}$$

Indeed, (16) and $t \ge n_i/(4q)$ give an exponent of absolute value at least $15n_i/4$.

By (8), the entire suffix has at most $2n_i$ vertices. There are therefore at most $2^{2n_i} \le e^{2n_i}$ possible suffix sets. The union bound and (17) give probability at most e^{-n_i} for a witness with a fixed first heavy index i. Finally, $n_i \ge 8C$ and $e^{n_i} \ge 1 + n_i$ imply $e^{-n_i} \le 1/(8C)$. Summing over all i proves the lemma. □

Proof of Theorem 3.1. Lemmas 3.4 and 3.5 bound the two failure probabilities by $1/8$ each. Their sum is less than one, so there is a realization satisfying (10) in which every independent set has weight less than $C/(2q)$. Lemma 3.3 proves (3), and the orientation of the construction proves (2).

If F were $2q$-colorable, its color classes would be independent sets of total weight strictly less than $2q \cdot C/(2q) = C$. But the color classes partition a vertex set of weight C. This contradiction proves $\chi(F) > 2q$, completing (1). □

4. Proof of the main theorem

Proof of Theorem 1.1. Fix $M \ge 1$ and put

$$q = 2^M + 2, \qquad C = 64q^2, \qquad d = C - 1, \qquad K = d + \binom{d}{2}. \tag{18}$$

These parameters satisfy $q \ge 2$ and $K \ge 4$. Apply Theorem 3.1 with this pair (q, K) to obtain $F = U(D)$. Every subgraph of F with maximum degree at most K is 4-colorable, and the maximum outdegree of D is at most d. The choice of K is legitimate because the outdegree bound in Theorem 3.1 is independent of K.

Set $G = L(D)$. Lemma 2.4, with $a = 4$, shows that every four-cycle-free subgraph of G is 6-colorable. Lemma 2.2 shows that G is triangle-free. If $\chi(G) \leq M$, then Lemma 2.3 gives

$$\chi(F) \leq 2^{\chi(G)} \leq 2^M < 2q,$$

contrary to Theorem 3.1. Thus $\chi(G) > M$. The graph G is finite, and this inequality also ensures that it is nonempty. □

Remark 4.1. List the vertices of D in a topological order. The map sending an arc to the ordered pair of its endpoint indices identifies $L(D)$ with an induced subgraph of the canonical shift graph on ordered pairs (i, j) with $i < j$, where (i, j) is adjacent to (j, k). The proof of Corollary 1.2 uses only the conclusions of Theorem 1.1.

5. Lean formalization and verification

The accompanying source edition [5, 6] formalizes the arc-graph construction and the negative answer. The theorem `Bounty.arc_counterexample_family` gives a nonempty arc-graph witness for each $M \in \mathbb{N}$, with chromatic number at least M and with every four-cycle-free subgraph 6-colorable. The triangle-free conclusion of Theorem 1.1 follows from Lemma 2.2 in the written proof. The checked `Bounty.target` proves the negation of the original `Erdos108.erdos_108` type.

The formalization represents vertices of the multipartite model as pairs (i, x) with $i < C$ and $x < n_i$, in lexicographic order. A sample is a finite dependent product of choices. Unused choices into earlier parts are retained so that the sample space is rectangular. Uniform probabilities are finite-set cardinalities divided by the cardinality of the sample space. The existence argument uses exact finite counting and real-valued estimates in Lean.

The predicate `FourCycleFree` excludes two distinct vertices having two distinct common neighbors. In a simple graph this is exactly the absence of an ordinary four-cycle. The type `SimpleGraph.Subgraph` includes both vertex and edge deletions. A graph relation $J \leq G$ on the same vertex type permits edge deletions and can represent vertex deletions by treating the deleted vertices as isolated.

Table 1 lists the selected correspondences. The Lean proof establishes the 4-colorability conclusion of Theorem 3.1 directly by finite-set induction using the low-degree vertex supplied by the cut argument. The library chromatic number takes values in the extended natural numbers. Its natural-valued girth is zero for an acyclic graph, whereas the written proof assigns infinite girth to forests. The formal proof uses the implication from library girth at least five to four-cycle-freeness. The statement about four-cycle-free subgraphs also covers forests. The final family is transported by `ULift` to the universe quantified over by the target.

Table 1. Selected correspondences with the submitted Lean file. Declaration names lie in namespace `Bounty`.

Mathematical statement	Lean declaration
Lemma 2.3	`base_colorable_of_arc_colorable`
Lemma 2.4	`arc_subgraph_colorable_bound`
Lemma 3.2	`model_sizes_exist`
Lemma 3.3, coloring conclusion	`colorable_four_of_sparse_cuts`
Lemma 3.4	`model_badSparse_probability_le`
Lemma 3.5	`model_badColor_probability_le`
Theorem 3.1, needed conclusions	`model_base_exists`
Theorem 1.1, family and coloring	`arc_counterexample_family`
Corollary 1.2, pinned target	`target`

Pinned environment and scope of verification. The supplied checks use Lean 4.33.1 [4], mathlib revision `0df444a360ea`, and Formal Conjectures task revision `8432eac99811`. The unwrapped proof ends with the declaration below, where `Not` is written as Lean's negation symbol.

```
theorem target : Not (fcTypeOfName% "Erdos108.erdos_108") := by
  exact counterexample_of_family counterexample_family
```

The task's type extractor supplies the statement. The local axiom audit for `Bounty.target` reports exactly `propext`, `Classical.choice`, and `Quot.sound`. The pinned source edition supplies the accepted proof body, the task wrapper, and the local build log and digest [5]. Conjectures.io records production verification and public certification of the negative answer [6]. These records apply to the original target. The original manuscript [7] names two additional Lean declarations for the triangle-free strengthening. The source documentation records that those declarations are absent from the supplied proof [8].

6. Concluding remarks

The obstruction is uniform. The bound six does not depend on the chromatic number of the host graph. Hence the same family excludes a threshold for every girth parameter $r \geq 5$ at chromatic parameter $k = 7$. The argument does not establish whether six is the smallest possible uniform bound.

The sizes in (4)–(5) were chosen to simplify the proof and its formalization. They specify a finite construction, but are far too large to make enumeration of the required realizations practical. The conclusion is an existence theorem proved by the finite probability estimates above.

Acknowledgements and AI-use disclosure. This work was developed with substantial assistance from OpenAI ChatGPT/Codex, including mathematical exploration, the preparation and checking of estimates, construction and debugging of the Lean proofs, and drafting and typesetting of the manuscript. Jensen Kohlmeyer and Liam Kruer are responsible for the mathematical content and its presentation. The formal source and the verification record make the machine-checked claims inspectable.

References

[1] T. F. Bloom, Erdős Problem #108, Erdős Problems.
[2] Google DeepMind and contributors, Formal Conjectures, file `FormalConjectures/ErdosProblems/108.lean`, repository. The task revision is specified in Section 5.
[3] B. Janzer, R. Steiner and B. Sudakov, Chromatic number and regular subgraphs, *Bull. London Math. Soc.* (2026), DOI, preprint arXiv:2410.02437.
[4] Lean community, The Lean 4 theorem prover, version 4.33.1.
[5] Pinned original-source edition and verification report.
[6] Conjectures.io, Erdős problem 108, production verification record.
[7] J. Kohlmeyer and L. Kruer, Original arc-graph manuscript, September 15, 2026, linked from Justin Sun Prize submission JSP-000120.
[8] Source notes on manuscript correspondence, pinned at the manuscript revision.

Purdue University, West Lafayette, Indiana, USA
Email address: `jenwin@purdue.edu`

Purdue University, West Lafayette, Indiana, USA
Email address: `kruerl@purdue.edu`