The minimum of the graph likelihood

Simone Severini
UCL

Eric W. Weisstein
Wolfram Research

August 12, 2026

Abstract

The likelihood of a finite simple undirected graph G on n vertices is the probability that the uniform sequential attachment process, which at each step joins a new vertex to a uniformly random subset of uniformly random size of the vertices already present, outputs a graph isomorphic to G. Dervovic, Mocherla and Severini conjectured that the likelihood is minimised by the balanced complete bipartite graph. We prove that, among complete bipartite graphs of a given order, the balanced one uniquely minimises the likelihood. Exact computation shows that it also minimises over all graphs for every order from 6 through 14, and that the first counterexample occurs at $n = 15$. The blow-up of the five cycle by independent sets of size three, equivalently the circulant on fifteen vertices with connection set $\{1, 4, 6\}$, has likelihood $0.20128\ldots$ times that of $K_{7,8}$, and it is again triangle-free. We show that the failure is not sporadic by proving that the likelihood of the balanced complete bipartite graph is $2^{-(1/2-1/(8\ln 2)+o(1))n^2}$, whereas the minimum over all graphs of order n is $2^{-(1/2+o(1))n^2}$, so the conjectured minimiser exceeds the minimum by a factor exponential in n^2. We also determine the Shannon entropy of the process to leading order, namely $n^2/(4\ln 2)$ bits, which shows that the conjectured minimiser is in fact more likely than a typical output of the process. The proofs rest on a vertex deletion recurrence which evaluates the likelihood in time $O(n\,2^n)$ and which closes on the blow-ups of any fixed base graph.

1 Introduction

Let $G_1 = K_1$ and construct a random graph G_t from G_{t-1} in three substeps: (1) choose k uniformly from $\{0, 1, \ldots, t-1\}$; (2) choose a k-element subset W of the vertex set $V(G_{t-1})$ uniformly; and (3) add a new vertex adjacent to exactly the vertices of W. For a finite simple undirected graph G on n vertices, its *likelihood* is the probability that G_n is isomorphic to G:

$$\mathcal{L}(G) \;=\; \Pr\big[G_n \cong G\big].$$

This invariant was introduced by Banerji, Mansour and Severini [1] as a graph-theoretic analogue of the infinite monkey theorem, the underlying process being the uniform case of a family of sequential attachment models studied by Janson and Severini [3] in the context of graph limits. The likelihood is a probability distribution on the isomorphism classes of graphs of order n, so $\sum_G \mathcal{L}(G) = 1$, where the sum runs over one representative of each isomorphism class of graphs with n vertices. It quantifies how easily a graph is produced by a mechanism that uses no information about the graph built so far. Writing $\overline{G}$ for the complement of G, explicit values include $\mathcal{L}(K_n) = \mathcal{L}(\overline{K_n}) = 1/n!$ and $\mathcal{L}(K_{1,n-1}) = n(n!)^{-2}\sum_{i=0}^{n-1} i!$, and the invariant satisfies $\mathcal{L}(G) = \mathcal{L}(\overline{G})$ together with the bounds [1, Cor. 18]

$$\frac{1}{|\mathrm{Aut}(G)|\prod_{i=1}^{n}\binom{i-1}{\lfloor (i-1)/2\rfloor}} \;\le\; \mathcal{L}(G) \;\le\; \frac{1}{|\mathrm{Aut}(G)|}. \tag{1}$$

Two extremal problems were left open in [1]. Weisstein [11] computed $\mathcal{L}(G)$ for every graph of order at most 10 in 2013 and observed that for $n \le 10$, with the exceptions $n = 3$ and $n = 5$, the minimum is attained by $K_{\lfloor n/2\rfloor,\lceil n/2\rceil}$ and its complement. The minimum values are recorded by numerator and denominator in A234234 and A234235, and are listed as fractions for $4 \le n \le 14$ in Table 1 below. (The analogous maximum numerator and denominator sequences are A234236 and A234237 [5].) On this numerical evidence Dervovic, Mocherla and Severini [2, §5] proposed the following assertion. Since $n = 3$ and $n = 5$ were already known exceptions, we understand their intended claim to apply to the nonexceptional range $n \ge 6$. It is this claim that we refute below.

Conjecture 1 ([2])**.** *For $n \ge 6$, if $n = 2p$ the minimum of $\mathcal{L}$ over graphs of order n is attained by $K_{p,p}$, and if $n = 2p-1$ it is attained by $K_{p-1,p}$.*

The computations below extend all four sequences with the $n = 11$ and $n = 12$ terms, and the two minimum sequences with the $n = 13$ and $n = 14$ terms. They verify Conjecture 1 through $n = 14$. Our central result, however, is that the conjecture is false: we show that its first failure is at $n = 15$ and prove that counterexamples exist at every sufficiently large order.

The heuristic offered in [2] is that these graphs have a relatively large automorphism group, which by (1) pushes $\mathcal{L}$ down, and that by Mantel's theorem they are the triangle-free graphs with the largest number of edges. Our purpose is to determine how far this reasoning reaches. The results below show both its force and its limitations.

Our first results are positive. Section 3 settles the conjecture inside the family in which it is stated.

Theorem 2. *For every $n \ge 4$ the function $a \mapsto \mathcal{L}(K_{a,n-a})$, $0 \le a \le \lfloor n/2\rfloor$, attains its unique minimum at $a = \lfloor n/2\rfloor$.*

The 2013 enumeration already covered all orders through $n = 10$. Section 4 adds $n = 11$ and $n = 12$, thereby verifying the conjecture for $6 \le n \le 12$. Section 5 adds the remaining cases $n = 13$ and $n = 14$ using a separate large-automorphism search. We regard a self-contained verification in exact arithmetic as worth recording; all our computations use integer or rational arithmetic only.

Theorem 3. *For $4 \le n \le 12$ the minimum of $\mathcal{L}$ over graphs of order n is attained exactly by the two isomorphism classes $K_{\lfloor n/2\rfloor,\lceil n/2\rceil}$ and its complement, except for $n = 5$, where the unique minimiser is C_5. The minimum values are listed in Table 1.*

The two remaining smaller orders can be settled without enumerating all graphs, by an exact enumeration of the graphs whose automorphism groups are large enough to meet the lower bound in (1); details are given in Section 5.

Theorem 4. *For $n = 13$ and $n = 14$, the minimum of $\mathcal{L}$ is attained exactly by $K_{\lfloor n/2\rfloor,\lceil n/2\rceil}$ and its complement. Consequently $n = 15$ is the first order $n \ge 6$ for which Conjecture 1 fails.*

n	$\min_G \mathcal{L}(G)$	minimiser
4	1/36	$K_{2,2}$
5	1/270	C_5
6	23/259200	$K_{3,3}$
7	319/54432000	$K_{3,4}$
8	319/15240960000	$K_{4,4}$
9	76441/115221657600000	$K_{4,5}$
10	76441/145179288576000000	$K_{5,5}$
11	20692621/201218493966336000000	$K_{5,6}$
12	20692621/1115555330549366784000000	$K_{6,6}$
13	368997387881/16080060756670792571289600000000	$K_{6,7}$
14	368997387881/3863073796182591207326613504000000000	$K_{7,7}$

Table 1: Minimum likelihoods among all simple graphs on n vertices. Minimisers are unique up to complementation. Values through $n = 10$ were computed in 2013 [11, 5]; the values for $n = 11$ through $n = 14$ are new here.

Direct exhaustive search is feasible only through $n = 12$, but at $n = 15$ a counterexample does occur. Write $H[\overline{K_m}]$ for the blow-up of H in which every vertex is replaced by an independent set of size m, adjacency being inherited.

Theorem 5. *Conjecture 1 is false for $n = 15$. Indeed*

$$\mathcal{L}\big(C_5[\overline{K_3}]\big) = \frac{63977511069907}{43503039787261205233506826321920000000000} = 1.4706\ldots \times 10^{-27}$$

while

$$\mathcal{L}(K_{7,8}) = \frac{7628328998218493}{1044072954894268925604163831726080000000000} = 7.3063\ldots \times 10^{-27},$$

so

$$\frac{\mathcal{L}(C_5[\overline{K_3}])}{\mathcal{L}(K_{7,8})} = \frac{1535460265677768}{7628328998218493} = 0.20128\ldots < 1.$$

The witness is instructive. It is a triangle-free vertex-transitive graph, so it does not contradict the first half of the heuristic of [2]. However, it does contradict the second half, since $C_5[\overline{K_3}]$ has 45 edges against the 56 of $K_{7,8}$. Blow-ups of C_5 are natural candidates: they remain triangle-free, nearly balanced blow-ups retain substantial symmetry, and the underlying odd cycle keeps them far from bipartite. It is precisely this family that overtakes the complete bipartite graphs.

Section 6 explains why counterexamples must eventually occur and shows that Theorem 5 is the visible edge of an asymptotic phenomenon. Automorphism-group size contributes only $O(n \log n)$ to the logarithm of the likelihood, whereas the difference between the conjectured value and the minimum is of order n^2. Symmetry therefore cannot make the balanced complete bipartite graph a global minimiser for large n.

Theorem 6. *As $n \to \infty$,*

$$\log_2 \frac{1}{\mathcal{L}(K_{\lfloor n/2\rfloor,\lceil n/2\rceil})} = \Big(\frac{1}{2} - \frac{1}{8\ln 2}\Big)n^2 + O(n\log n) = (0.31966\ldots)\, n^2 + O(n\log n),$$

while

$$\log_2 \frac{1}{\min_G \mathcal{L}(G)} = \frac{n^2}{2} + O(n \log n).$$

Consequently $\mathcal{L}(K_{\lfloor n/2\rfloor,\lceil n/2\rceil})/\min_G \mathcal{L}(G) = 2^{(1/(8\ln 2)+o(1))n^2}$*, and Conjecture 1 fails for every sufficiently large* n*.*

Section 7 completes the picture by computing the entropy of the process, which identifies the likelihood of a typical output and shows that the conjectured minimiser sits above it.

Theorem 7. *The Shannon entropy of the distribution* $\mathcal{L}$ *on isomorphism classes of graphs of order* n *equals* $\frac{n^2}{4\ln 2} + O(n\log n) = (0.36067\ldots)n^2 + O(n\log n)$ *bits, against the* $\frac{n^2}{2} + O(n\log n)$ *bits of the uniform distribution on isomorphism classes. Moreover, if* G_n *is the output of the process, then*

$$-\frac{1}{n^2}\log_2 \mathcal{L}(G_n) \;\xrightarrow{\mathrm{p}}\; \frac{1}{4\ln 2}.$$

Since $\frac{1}{2} - \frac{1}{8\ln 2} < \frac{1}{4\ln 2}$, the graph $K_{\lfloor n/2\rfloor,\lceil n/2\rceil}$ is asymptotically *more* likely than a typical output of the process, which is a second way of seeing that it cannot be a minimiser. Section 8 collects remarks on the maximum, where we show that no analogue of Theorem 2 is available, and open problems.

The tool behind all of this is a vertex deletion recurrence for $\mathcal{L}$, stated in §2. It appears in this form in [11] and we claim no novelty for it; what we add is the observation that it closes on the blow-ups of any fixed base graph, which is what makes both Theorem 2 and Theorem 5 accessible.

2 The deletion recurrence and blow-ups

For a graph G of order n, set

$$A(G) \;=\; n!\,|\mathrm{Aut}(G)|\,\mathcal{L}(G).$$

By [1, Thm. 16], unwinding the process along all vertex orderings,

$$A(G) \;=\; \sum_{\sigma\in\mathfrak{S}_n} \prod_{i=1}^{n} \binom{i-1}{d_i(\sigma)}^{-1}, \tag{2}$$

where $d_i(\sigma)$ is the number of neighbours of $\sigma(i)$ among $\sigma(1),\ldots,\sigma(i-1)$. In particular $A(G) = A(\overline{G})$ and $A(K_n) = A(\overline{K_n}) = n!$.

Proposition 8 ([11])**.** $A(K_1) = 1$*, and for every graph* G *of order* $n \geq 2$*,*

$$A(G) \;=\; \sum_{v\in V(G)} \frac{A(G-v)}{\binom{n-1}{\deg_G v}}. \tag{3}$$

Proof. Partition the sum (2) according to $\sigma(n) = v$. For fixed v, the first $n-1$ factors run over all orderings of $V(G)\setminus\{v\}$ and reproduce $A(G-v)$, while the last factor is $\binom{n-1}{\deg_G v}^{-1}$. $\square$

Three consequences deserve to be isolated.

Corollary 9. *Fix a labelling of $V(G)$. Put $F(\{v\}) = 1$ and, for $|S| \geq 2$,*

$$F(S) = \sum_{v \in S} \frac{F(S \setminus \{v\})}{\binom{|S|-1}{\deg_{G[S]} v}}.$$

Then

$$\mathcal{L}(G) = \frac{F(V(G))}{n!\,|\mathrm{Aut}(G)|}.$$

Hence $\mathcal{L}(G)$ is computable exactly in $O(n\,2^n)$ arithmetic operations and $O(2^n)$ space.

This replaces the enumeration of the $n!/|\mathrm{Aut}(G)|$ construction paths used in [1]. Our implementation combines graph processing in the Wolfram Language with a compiled C helper for the main exact computation. The helper evaluates an integer-scaled form of the subset recurrence modulo several 64-bit primes, after which the exact integer is reconstructed by the Chinese remainder theorem. At the time of writing, this implementation has been used to compute exact likelihoods for every graph represented in the development version of `GraphData` [12] through order 29, as well as selected larger graphs whose structure permits effective preprocessing. In particular, it returns the value in Theorem 5 for the circulant $C_{15}(1, 4, 6)$. Some individual graphs of order 30 require roughly four days on current hardware and have not yet all been computed. A separate C plugin integrated with `geng` makes the exhaustive search through order 12 feasible. The complexity of $\mathcal{L}$ remains open; see Problem 26.

Corollary 10. *A is determined by the vertex-deleted deck $\{G - v : v \in V(G)\}$, counted with multiplicity. In particular, hypomorphic graphs have equal A, and $\mathcal{L}$ distinguishes them only through the order of their automorphism groups.*

Proof. The conclusion is immediate for graphs of order at most 2. For graphs of order at least 3, the right-hand side of (3) depends only on the multiset of isomorphism types $G - v$ together with the numbers $\deg_G v = e(G) - e(G - v)$, and $e(G)$ is determined by the deck by Kelly's lemma. □

The third consequence is the one we shall use. Let H be a graph with $V(H) = \{1, \ldots, p\}$ and let $m = (m_1, \ldots, m_p) \in \mathbb{Z}^p_{\geq 0}$. The *blow-up* $H[m]$ is obtained by replacing vertex i by an independent set V_i of size m_i and joining V_i to V_j completely whenever $ij \in E(H)$. Deleting a vertex of $H[m]$ yields $H[m - e_i]$ for some i, so blow-ups of a fixed base form a family closed under vertex deletion and (3) closes on it.

Proposition 11. *With $n = \sum_i m_i$ and $A(H[m]) = 1$ when $n = 1$,*

$$A(H[m]) \;=\; \sum_{i\,:\,m_i > 0} \frac{m_i\, A(H[m - e_i])}{\binom{n-1}{\sum_{j \sim i} m_j}}. \tag{4}$$

Proof. Every vertex of V_i has degree $\sum_{j \sim i} m_j$ and deleting it gives $H[m - e_i]$; apply (3) and group the n terms into the p classes $V_1, \ldots, V_p$. □

Taking $H = K_p$ gives all complete multipartite graphs, indexed by partitions of n; taking $H = K_2$ gives all complete bipartite graphs; taking $H = C_5$ gives the family of Theorem 5. For the automorphism groups of blow-ups we use the theorem of Sabidussi [9, 10] on lexicographic products, in the form given by Grech and Kisielewicz [6, §1]. If no two vertices of H have the same neighbourhood and no two have the same closed neighbourhood, then $\mathrm{Aut}(H[\overline{K_m}]) = \mathrm{Aut}(H) \ltimes \mathfrak{S}_m^p$, of order $|\mathrm{Aut}(H)|\,(m!)^p$.

3 Complete bipartite graphs

Specialising (4) to $H = K_2$ and normalising, put

$$b_a^{(n)} \;=\; \frac{A(K_{a,n-a})}{a!\,(n-a)!}, \qquad 0 \le a \le n,$$

so that, since $|\mathrm{Aut}(K_{a,n-a})| = a!\,(n-a)!\,(1 + [\,2a = n\,])$,

$$\mathcal{L}(K_{a,n-a}) = \frac{b_a^{(n)}}{n!\,\bigl(1 + [\,2a = n\,]\bigr)}. \tag{5}$$

Proposition 11 becomes, using $\binom{n-1}{n-a-1} = \binom{n-1}{a}$,

$$b_a^{(n)} = \frac{b_{a-1}^{(n-1)}}{\binom{n-1}{a-1}} + \frac{b_a^{(n-1)}}{\binom{n-1}{a}} \qquad (1 \le a \le n-1), \tag{6}$$

with $b_0^{(n)} = b_n^{(n)} = 1$ for all $n \ge 1$ and $b_0^{(1)} = b_1^{(1)} = 1$. Note $b_a^{(n)} = b_{n-a}^{(n)}$.

Lemma 12. *For $n \ge 2$ we have $b_1^{(n)} = b_{n-1}^{(n)} = \dfrac{1}{(n-1)!}\displaystyle\sum_{k=0}^{n-1} k!$, and $1 < b_1^{(n)} \le 2$ with equality if and only if $n \in \{2, 3\}$.*

Proof. $K_{1,n-1}$ is the star, and $\mathcal{L}(K_{1,n-1}) = n(n!)^{-2}\sum_{k=0}^{n-1} k!$ by [1, Prop. 20]; substituting into (5) with $|\mathrm{Aut}| = (n-1)!$ gives the formula. For the bounds, isolate the top term and observe that $\sum_{k=0}^{n-2} k! \le (n-2)! + \sum_{k=0}^{n-3} k! \le (n-2)! + (n-2)(n-3)! = 2(n-2)!$, which is smaller than $(n-1)! = (n-1)(n-2)!$ as soon as $n \ge 4$. Hence $b_1^{(n)} = 1 + \sum_{k=0}^{n-2} k!/(n-1)! < 2$ for $n \ge 4$, while $b_1^{(2)} = b_1^{(3)} = 2$; and $b_1^{(n)} > 1$ because the term $k = n-1$ alone contributes 1. $\square$

Lemma 13. *For all $n \ge 1$ and all a we have $b_a^{(n)} \le 2$. Moreover $b_a^{(n)} \le 4/(n-1)$ whenever $n \ge 4$ and $2 \le a \le n-2$.*

Proof. Induction on n. For $n \le 3$ every value is 1 or 2 by Lemma 12. Let $n \ge 4$ and assume the first claim for $n-1$. If $a \in \{0, n\}$ then $b_a^{(n)} = 1$, and if $a \in \{1, n-1\}$ then $b_a^{(n)} \le 2$ by Lemma 12. If $2 \le a \le n-2$ then $1 \le a-1$ and $a \le n-2$, so $\binom{n-1}{a-1} \ge n-1$ and $\binom{n-1}{a} \ge n-1$, and (6) gives $b_a^{(n)} \le 2/(n-1) + 2/(n-1) = 4/(n-1) \le 2$. $\square$

Lemma 14. *For $1 \le a \le n-2$,*

$$b_a^{(n)} - b_{a+1}^{(n)} = \frac{b_{a-1}^{(n-1)}}{\binom{n-1}{a-1}} - \frac{b_{a+1}^{(n-1)}}{\binom{n-1}{a+1}}.$$

Proof. Subtract the instance $a+1$ of (6) from the instance a; the two middle terms $b_a^{(n-1)}/\binom{n-1}{a}$ cancel. $\square$

Proposition 15. *For every $n \ge 4$ the sequence $b_1^{(n)}, b_2^{(n)}, \ldots, b_{\lfloor n/2\rfloor}^{(n)}$ is strictly decreasing.*

Proof. Induction on n. For $n = 4$ we have $b_1 = 5/3 > b_2 = 4/3$, and for $n = 5$, $b_1 = 17/12 > b_2 = 23/36$. Let $n \geq 6$, assume the statement for $n-1$, and fix a with $1 \leq a \leq \lfloor n/2 \rfloor - 1$. By Lemma 14 it suffices to prove

$$\frac{b_{a-1}^{(n-1)}}{\binom{n-1}{a-1}} > \frac{b_{a+1}^{(n-1)}}{\binom{n-1}{a+1}}. \tag{7}$$

Assume first $a \geq 2$, so that $1 \leq a-1 < a+1 \leq \lfloor n/2 \rfloor$. If $a+1 \leq \lfloor (n-1)/2 \rfloor$ then the inductive hypothesis gives $b_{a-1}^{(n-1)} > b_{a+1}^{(n-1)}$, while $a+1 \leq \lfloor (n-1)/2 \rfloor$ forces $\binom{n-1}{a-1} < \binom{n-1}{a+1}$, and (7) follows. Otherwise $a+1 > \lfloor (n-1)/2 \rfloor$ together with $a+1 \leq \lfloor n/2 \rfloor$ forces n even and $a+1 = n/2$, so $n-1 = 2\mu+1$ with $\mu = n/2 - 1 = a$. Then $b_{a+1}^{(n-1)} = b_{n-1-(a+1)}^{(n-1)} = b_a^{(n-1)}$ and $\binom{n-1}{a+1} = \binom{2\mu+1}{\mu+1} = \binom{2\mu+1}{\mu} = \binom{n-1}{a}$, so (7) reduces to $b_{a-1}^{(n-1)}/\binom{n-1}{a-1} > b_a^{(n-1)}/\binom{n-1}{a}$, which holds because $1 \leq a-1 < a \leq \lfloor (n-1)/2 \rfloor$ gives $b_{a-1}^{(n-1)} > b_a^{(n-1)}$ by induction and $\binom{n-1}{a-1} < \binom{n-1}{a}$.

Now let $a = 1$. Then $b_0^{(n-1)} = 1$ and $\binom{n-1}{0} = 1$, so (7) reads $b_2^{(n-1)} < \binom{n-1}{2}$. Since $n \geq 6$, Lemma 13 gives $b_2^{(n-1)} \leq 4/(n-2) \leq 1 < \binom{n-1}{2}$. □

Proof of Theorem 2. By (5) and Proposition 15, $\mathcal{L}(K_{a,n-a})$ is strictly decreasing for $1 \leq a \leq \lfloor n/2 \rfloor - 1$, and the passage from $a = \lfloor n/2 \rfloor - 1$ to $a = \lfloor n/2 \rfloor$ decreases it as well, strictly, since for n even the additional factor $(1 + [2a = n])^{-1} = 1/2$ only helps. It remains to compare with $a = 0$, where $\mathcal{L}(\overline{K_n}) = 1/n!$. For $n \geq 6$ we have $2 \leq \lfloor n/2 \rfloor \leq n-2$, so Lemma 13 gives $b_{\lfloor n/2 \rfloor}^{(n)} \leq 4/(n-1) < 1$; for $n = 4$, $b_2^{(4)}/2 = 2/3 < 1$ and for $n = 5$, $b_2^{(5)} = 23/36 < 1$. □

Remark 16. The function $a \mapsto \mathcal{L}(K_{a,n-a})$ is not monotone on $0 \leq a \leq \lfloor n/2 \rfloor$; the star $a = 1$ is the maximum of the family and lies above the empty graph $a = 0$. For $n = 10$ the values are 2.76×10^{-7}, 3.11×10^{-7}, 3.63×10^{-8}, 1.20×10^{-9}, 2.29×10^{-11}, 5.27×10^{-13}.

4 Exhaustive verification for $n \leq 12$

The structured searches below use the following elementary screening, which combines the deletion recurrence, the lower bound in (1), and the fact that automorphisms preserve degrees. Write $P_n = \prod_{i=1}^{n} \binom{i-1}{\lfloor (i-1)/2 \rfloor}$ and let $m_d(G)$ be the number of vertices of G of degree d.

Lemma 17. *If $n \geq 2$ and $\mathcal{L}(G) \leq \lambda$, then*

$$\prod_{d \geq 0} m_d(G)! \;\geq\; |\mathrm{Aut}(G)| \;\geq\; \frac{1}{n\lambda P_{n-1}} \sum_{v \in V(G)} \binom{n-1}{\deg_G v}^{-1} \;\geq\; \frac{1}{\lambda P_n}.$$

Proof. The left inequality holds because $\mathrm{Aut}(G)$ embeds in the direct product of the symmetric groups on the degree classes. Applying (1) to each graph $G - v$ gives

$$A(G-v) = (n-1)!|\mathrm{Aut}(G-v)|\mathcal{L}(G-v) \geq \frac{(n-1)!}{P_{n-1}}.$$

Substitution in (3), followed by $\mathcal{L}(G) = A(G)/(n!|\mathrm{Aut}(G)|) \leq \lambda$, gives the middle inequality. The last follows from $\binom{n-1}{\deg_G v} \leq \binom{n-1}{\lfloor (n-1)/2 \rfloor}$ and the definition of P_n. □

For the exhaustive proof we use one implementation uniformly through order 12. Setting $D_i = \operatorname{lcm}\{\binom{i-1}{k} : 0 \le k \le i-1\}$ and $\widehat{F}(S) = F(S)\prod_{i \le |S|} D_i$, the recurrence of Corollary 9 becomes

$$\widehat{F}(S) = \sum_{v \in S} \widehat{F}(S \setminus \{v\}) \frac{D_{|S|}}{\dbinom{|S|-1}{\deg_{G[S]} v}},$$

in which every factor is a positive integer. For $n \le 12$, $\widehat{F}(V) \le n! \prod_{i \le n} D_i < 2^{90}$. The integrated `geng` plugin uses checked unsigned 128-bit arithmetic for this recurrence, exact `nauty` automorphism orders, and exact 192-bit cross-products for likelihood comparisons. It rejects larger orders. It enumerates every unlabelled graph through order 11 and, using complement symmetry, one representative from every complement orbit at order 12. As a global check, it reproduces the extrema previously known through order 10.

At order 11, the program streams all 1 018 997 864 unlabelled graphs and obtains

$$\min_G \mathcal{L}(G) = \frac{20692621}{201218493966336000000}.$$

The two minimisers are $K_{5,6}$ and its complement. Graph6 representations of the complement and $K_{5,6}$ are `J~{?GKF@wN_` and `J??F~z{~Fw?`, respectively.

For $n = 12$, complement symmetry reduces the search to one representative from each complement orbit. There are exactly 82 545 586 656 = (165 091 172 592+720)/2 such orbits, since there are 165 091 172 592 unlabelled graphs of order 12, of which 720 are self-complementary. The search obtains

$$\min_G \mathcal{L}(G) = \frac{20692621}{111555533054936678400000000}.$$

The two minimisers are $K_{6,6}$ and its complement, with graph6 representations `KsaCB|}^b{No` and `K~~w?CB?wF_^`. This proves Theorem 3.

Remark 18. The minimum is well separated. For $n = 10$ the second smallest value is 8.39×10^{-12}, larger than the minimum by a factor of about 16.

5 A counterexample

Proof of Theorem 5. Let $H = C_5$ with $V(H) = \mathbb{Z}_5$ and $m = (3,3,3,3,3)$, so $G = C_5[\overline{K_3}]$ is a 6-regular graph of order 15 with 45 edges; it is isomorphic to the circulant $C_{15}(1,4,6)$, the parts being the cosets of the subgroup of order 3 in $\mathbb{Z}_{15}$. No two vertices of C_5 share a neighbourhood or a closed neighbourhood, so $|\mathrm{Aut}(G)| = |\mathrm{Aut}(C_5)|\,(3!)^5 = 10 \times 6^5 = 77760$ by the theorem of Sabidussi. Iterating (4) over the 4^5 vectors $m' \le m$ in exact rational arithmetic gives

$$A(G) = \frac{63977511069907}{427822618422142944000000},$$

and $\mathcal{L}(G) = A(G)/(15! \times 77760)$ is the stated value. Independently, Corollary 9 applied to G with 2^{15} subsets and rational arithmetic returns the same $A(G)$. Theorem 2 or a direct application of (6) gives $\mathcal{L}(K_{7,8})$, and the comparison is a rational inequality. $\square$

Further exact computations locate the example.

Proposition 19. *For $n \in \{13, 14\}$, no independent blow-up $H[m]$ with $|V(H)| \le 12$ has likelihood smaller than $K_{\lfloor n/2 \rfloor, \lceil n/2 \rceil}$. For $n \in \{16, 18\}$ the same holds with $|V(H)| \le 8$. For $n = 15$ the blow-up $C_5[\overline{K_3}]$ does. Also, no vertex-transitive graph of order between 6 and 14 has likelihood smaller than $K_{\lfloor n/2 \rfloor, \lceil n/2 \rceil}$, whereas the circulant $C_{15}(1, 4, 6)$ does.*

For the new cases this is a finite exact computation using (4). Through base order 9, every unlabelled base and every positive part vector is checked. At base order 9 this gives totals of 135 960 660 and 353 497 716 blow-ups for $n = 13$ and $n = 14$, respectively. In each case the minimum is attained by $K_{\lfloor n/2 \rfloor, \lceil n/2 \rceil}$. Blow-ups of smaller bases are included, since a part can be split by duplicating its base vertex as a false twin.

Base order 10 is handled by an exact automorphism screen. A base with false twins can be reduced by merging them and is therefore already covered by a smaller base. If H is false-twin-free, its fibres are precisely the false-twin classes of $H[m]$, and hence

$$|\mathrm{Aut}(H[m])| = \left(\prod_i m_i!\right) |\mathrm{Stab}_{\mathrm{Aut}(H)}(m)| \le \left(\prod_i m_i!\right) |\mathrm{Aut}(H)|.$$

There are 12 005 168 unlabelled bases of order 10, of which 1 990 286 have false twins. A strict competitor at order 13 must have automorphism order at least 183. Since the largest fibre-factorial product among the 220 positive compositions of 13 into 10 parts is $4! = 24$, only false-twin-free bases with automorphism order at least 8 need be kept; there are 50 714. The part-vector-specific bound leaves 2 368 100 targets for an exact automorphism computation, of which 528 765 meet the necessary target threshold and require likelihood evaluation. None beats $K_{6,7}$. At order 14 the corresponding target threshold is 2552 and the maximum fibre-factorial product is $5! = 120$. There are 7 498 false-twin-free bases with automorphism order at least 22; the refined screen computes exact automorphism orders for 526 705 targets and exact likelihoods for the surviving 80 258. None beats $K_{7,7}$.

At base order 11 there are 1 018 997 864 unlabelled bases, of which 106 088 988 have false twins. At order 13 there are only 66 positive part vectors and their largest fibre-factorial product is $3! = 6$. The target automorphism threshold 183 therefore leaves only false-twin-free bases with automorphism order at least 31; exactly 41 737 bases survive. The part-vector-specific screen requires 2 210 912 exact target-automorphism computations and 555 612 likelihood evaluations. None beats $K_{6,7}$. At order 14 there are 286 positive part vectors, the largest fibre-factorial product is $4! = 24$, and the target threshold 2552 requires base automorphism order at least 107. Exactly 5 700 bases survive; 497 915 target automorphism orders and 79 908 likelihoods are evaluated, and none beats $K_{7,7}$. The 41 737 bases retained during the order-13 stream were saved and reused for the order-14 search, so the full order-11 `geng` stream was generated only once.

At base order 12 there are 165 091 172 592 unlabelled bases, of which 10 454 883 132 have false twins. At order 13 there are 12 positive part vectors, each with fibre-factorial product 2. The target threshold 183 therefore requires base automorphism order at least 92, and an exhaustive 4 096-shard stream retains exactly 140 474 bases. The resulting 1 685 688 target representations require exact automorphism computations; 927 658 are discarded by the target threshold and the remaining 758 030 receive exact likelihood evaluations. None beats $K_{6,7}$.

At order 14 there are 78 positive part vectors and the largest fibre-factorial product is $3! = 6$, so a possible competitor must have base automorphism order at least 426. The bases saved during the order-13 stream therefore contain every possibility; the stronger base bound retains 9 224 of them. Of the resulting 719 472 target representations, 487 416 require exact target-automorphism computations and 121 413 survive for exact likelihood evaluation. None beats $K_{7,7}$.

The corresponding searches through base order 7 at $n = 16$ and $n = 18$ evaluate 5 225 220 and 12 920 544 blow-ups, respectively, and again return $K_{\lfloor n/2\rfloor,\lceil n/2\rceil}$. At $n = 16$, extending the search through base order 8 checks every positive eight-part vector for all 12 346 unlabelled bases. After omitting bases with false twins, which reduce to smaller bases, and applying the exact automorphism screen, it evaluates 34 372 082 likelihoods; none beats $K_{8,8}$. The analogous base-order-8 search at $n = 18$ evaluates all 156 498 056 likelihoods; none beats $K_{9,9}$. As a further check at base order 9, the minimum among all 274 668 graphs $H[\overline{K_2}]$ with $|V(H)| = 9$ is attained when H is the 3×3 rook graph. It comes close, but remains above the conjectured minimum by the exact factor

$$\frac{44277431617180521729 75}{44009069025899830241 66} = 1.0060978\ldots.$$

A separate search considers every graph $rH + tK_1$, and its complement, where H is connected of order at most 7, $r \geq 2$, and the total order is 13 or 14. There are 156 such representations at order 13 and 1010 at order 14. At order 13 the smallest likelihood in this family occurs at $2K_6 + K_1$ and is $17.1364\ldots$ times $\mathcal{L}(K_{6,7})$; at order 14 the minimum occurs at $2K_7 = \overline{K_{7,7}}$.

The vertex-transitive assertion is automatic from Theorem 3 through order 12. At orders 13 and 14, respectively, enumeration from the transitive permutation groups gives 14 and 56 isomorphism classes of vertex-transitive graphs. At order 13 the smallest likelihood among them is attained by the Paley graph of order 13 and exceeds $\mathcal{L}(K_{6,7})$ by the exact factor

$$\frac{7061981141760}{368997387881} = 19.1382\ldots.$$

At order 14 the only minima in this class are $K_{7,7}$ and its complement.

The lower bound in (1) also gives a useful restriction on any unrestricted counterexample. For $n = 13, 14, 16, 18$, respectively, a graph G satisfying $\mathcal{L}(G) < \mathcal{L}(K_{\lfloor n/2\rfloor,\lceil n/2\rceil})$ must have

$$|\mathrm{Aut}(G)| \geq 183, \quad 2552, \quad 1556, \quad 462.$$

Indeed, before taking the next integer, the four lower bounds are $182.255\ldots$, $2551.577\ldots$, $1555.153\ldots$ and $461.923\ldots$. This makes enumeration by large automorphism group a possible alternative to enumerating every graph. The exact blow-up searches also rule out an additional tie with $K_{\lfloor n/2\rfloor,\lceil n/2\rceil}$. They show that any remaining graph of order 13 with likelihood at most $\mathcal{L}(K_{6,7})$, as well as its complement, must be false-twin-free; equivalently, it can have neither false twins nor true twins. At order 14, any remaining graph with likelihood at most $\mathcal{L}(K_{7,7})$ can have at most one false-twin pair and at most one true-twin pair. If either occurs, the corresponding quotient of order 13 must have automorphism order at least 1276.

We complete the search by enumerating permutation groups rather than graphs. For a subgroup $\Gamma \leq \mathfrak{S}_n$, every Γ-invariant graph is a union of orbits of Γ on the unordered vertex pairs. Moreover, if two vertices u, v have the property that $\{u, w\}$ and $\{v, w\}$ belong to the same pair orbit for every $w \notin \{u, v\}$, then every such graph makes u, v either false twins or true twins, according as uv is absent or present.

At order 13, the complete table of marks for $\mathfrak{S}_{13}$ in TomLib [8] contains 20 832 conjugacy classes of subgroups, of which 9634 have order at least 183. The forced-twin test leaves 26 classes and 3296 orbital graph representations. Direct twin testing leaves 464, and canonical labelling reduces these to 100 isomorphism classes. Their exact likelihoods have minimum ratio

$$\frac{\mathcal{L}(G)}{\mathcal{L}(K_{6,7})} = \frac{38307963944498}{368997387881} > 1.$$

For a graph of order 14 with a false-twin pair, uniqueness of that pair forces its automorphism group to preserve the pair. After factoring out its transposition, the induced group on the remaining 12 vertices has order at least 1276. The complete table for $\mathfrak{S}_{12}$ has 10 723 subgroup classes, of which 956 meet this bound. Their pair orbits, together with their vertex orbits describing the common neighbourhood of the twin pair, give 7 523 958 representations. Exact twin testing leaves 6 088 090, and canonical labelling leaves 122 914 graphs. Their minimum exact ratio is

$$\frac{\mathcal{L}(G)}{\mathcal{L}(K_{7,7})} = \frac{17952366553322}{368997387881} > 1.$$

Complementation covers the case of a true-twin pair.

It remains to treat graphs of order 14 having neither kind of twin. Starting from $\mathfrak{S}_{14}$, we recursively enumerate conjugacy-class representatives of maximal subgroups whose orders remain at least 2552, identifying subgroups conjugate in $\mathfrak{S}_{14}$. Every subgroup above the threshold occurs in such a chain. GAP returns no failed maximal-subgroup computations and gives 6280 classes. The forced-twin test leaves four classes and 40 orbital representations; direct twin testing and canonical labelling leave four graphs. Their minimum exact ratio is again

$$\frac{38307963944498}{368997387881} > 1.$$

Together with the blow-up search, these computations prove Theorem 4.

The same subgroup method substantially narrows the order-16 problem but does not yet settle it. Here the automorphism threshold is 1556. A checkpointed maximal-subgroup traversal of $\mathfrak{S}_{16}$ finds 116 597 subgroup classes above the threshold, with no failed maximal-subgroup computations. Of these, 116 553 force a false- or true-twin pair in every invariant graph. The remaining 44 classes give 4580 orbital graph representations; direct twin testing leaves 738, and canonical labelling leaves 128 isomorphism classes. None has likelihood below $\mathcal{L}(K_{8,8})$. Thus any counterexample at order 16 must contain false or true twins. The existing blow-up searches cover such graphs only through base order 8, so the forced-twin cases with larger quotients remain to be checked and the unrestricted order-16 problem remains open.

At the degree-sequence level, the final inequality of Lemma 17 leaves 608 239 of the 836 315 graphical sequences at order 13 and 1 041 046 of the 3 166 852 sequences at order 14. The strengthened middle inequality reduces these counts to 152 691 and 175 878. Applying the deletion recurrence once more and minimizing over every possible neighbour-degree assignment, a relaxation that remains a rigorous lower bound, leaves 69 797 and 75 888 sequences, or 34 949 and 37 944 orbits under complementation. Thus degree information alone substantially narrows the search but does not settle either order.

Thus $n = 15$ is the first failure of the conjecture. Second, the advantage of the C_5 blow-ups grows very quickly with n; for $n = 25$,

$$\mathcal{L}\big(C_5[\overline{K_5}]\big) = 1.202\ldots \times 10^{-77} \qquad \text{against} \qquad \mathcal{L}(K_{12,13}) = 6.960\ldots \times 10^{-72},$$

a factor of about 5.8×10^5. Exact enumeration of every positive part vector, modulo the dihedral symmetries of C_5, shows that through $n = 50$ the minimum among positive C_5 part vectors is always attained by a nearly balanced vector, meaning one whose coordinates differ by at most one. It is smaller than $K_{\lfloor n/2\rfloor,\lceil n/2\rceil}$ at $n = 15$, at $n = 17$, and at every order from 19 through 50. Exact evaluation of the best nearly balanced vectors continues this sequence of counterexamples without a gap through $n = 100$; selected early examples are listed in Table 2.

n	m	$\mathcal{L}(C_5[m])/\mathcal{L}(K_{\lfloor n/2\rfloor,\lceil n/2\rceil})$
17	$(3,3,3,4,4)$	$0.32136\ldots$
19	$(3,4,4,4,4)$	$0.072155\ldots$
20	$(4,4,4,4,4)$	$0.088189\ldots$
21	$(4,4,4,4,5)$	$0.018452\ldots$
22	$(4,4,4,5,5)$	$0.26319\ldots$
23	$(4,5,4,5,5)$	$0.0040481\ldots$
24	$(4,5,5,5,5)$	$0.060271\ldots$
25	$(5,5,5,5,5)$	$1.7271\ldots \times 10^{-6}$

Table 2: Selected further counterexamples among blow-ups of C_5. The displayed ratios are rounded from exact rational computations.

For the positive vectors in Table 2, the five parts are the twin classes, so $|\mathrm{Aut}(C_5[m])| = (\prod_{i=1}^{5} m_i!)|\mathrm{Stab}_{\mathrm{Aut}(C_5)}(m)|$; together with (4), this gives every ratio in the table exactly. Within the $C_5[m]$ family, $K_{\lfloor n/2\rfloor,\lceil n/2\rceil}$ is again smaller at $n = 16$ and $n = 18$, where the five parts cannot all have equal size; the failure of the conjecture is therefore not monotone in n at this scale. The unrestricted cases $n = 16$ and $n = 18$ remain open.

6 Asymptotics

We first evaluate $\mathcal{L}(K_{m,m})$. Iterating (6) downwards from (m, m) expresses $b_m^{(2m)}$ as a sum over lattice paths. Formally, call a *deletion path* from (a, b) a sequence of steps each decreasing one coordinate by 1, terminating when a coordinate reaches 0, and give the step that decreases the first coordinate at state (a', b') the cost $\log_2 \binom{a'+b'-1}{b'}$, the other step the cost $\log_2 \binom{a'+b'-1}{a'}$. Unwinding (6) and using $b_0^{(k)} = 1$,

$$b_m^{(2m)} = \sum_{\pi} 2^{-\mathrm{cost}(\pi)}, \tag{8}$$

the sum being over all deletion paths from (m, m). Let $W(m) = \min_\pi \mathrm{cost}(\pi)$.

Lemma 20. $W(m) = \sum_{j=0}^{m-1} \log_2 \binom{m+j}{j}$*, attained by the path that empties one side first.*

Proof. Let $f(a, b)$ denote the minimum cost of a deletion path from (a, b), with $f(a, 0) = f(0, b) = 0$. We show by induction on $a + b$ that for $a \le b$,

$$f(a,b) = S(a,b) := \sum_{j=0}^{a-1} \log_2 \binom{b+j}{j},$$

which is the cost of emptying the first side; by symmetry this suffices. The cases $a=0$ and $a+b\le 2$ are clear. Let $1\le a\le b$. Emptying the first side has cost $\log_2\binom{a+b-1}{b}+S(a-1,b)=S(a,b)$, since $\binom{a+b-1}{b}=\binom{a+b-1}{a-1}$. So $f(a,b)\le S(a,b)$, and it remains to exclude the other first step, that is to show

$$S(a,b)\ \le\ \log_2\binom{a+b-1}{a}+f(a,b-1). \tag{9}$$

Suppose first $a\le b-1$, so that $f(a,b-1)=S(a,b-1)$ by induction. Since $\binom{b+j}{j}/\binom{b-1+j}{j}=(b+j)/b$, we get

$$\begin{aligned} S(a,b)-S(a,b-1) &= \sum_{j=0}^{a-1}\log_2\frac{b+j}{b} \\ &= \log_2\frac{\prod_{j=0}^{a-1}(b+j)}{b^a}, \\ \log_2\binom{a+b-1}{a} &= \log_2\frac{\prod_{j=0}^{a-1}(b+j)}{a!}, \end{aligned}$$

so (9) is equivalent to $a!\le b^a$, which holds because $a\le b$. If instead $a=b$, then $f(a,b-1)=S(a-1,a)$ by induction and $S(a,a)-S(a-1,a)=\log_2\binom{2a-1}{a-1}=\log_2\binom{2a-1}{a}$, so (9) holds with equality, as symmetry demands. □

Lemma 21. *$W(m)=\big(2-\frac{1}{2\ln 2}\big)m^2+O(m\log m)$.*

Proof. By Lemma 20, $W(m)\ln 2=\sum_{j=0}^{m-1}\big(\ln(m+j)!-\ln m!-\ln j!\big)$. Stirling gives $\ln k!=k\ln k-k+O(\log k)$, and comparing the three sums with the integrals $\int x\ln x\,dx$ yields

$$\sum_{j=0}^{m-1}\ln(m+j)!=\tfrac{3}{2}m^2\ln m+(2\ln 2-\tfrac{9}{4})m^2+O(m\log m),$$

$$\begin{aligned} m\ln m! &= m^2\ln m-m^2+O(m\log m), \\ \sum_{j=0}^{m-1}\ln j! &= \tfrac{1}{2}m^2\ln m-\tfrac{3}{4}m^2+O(m\log m). \end{aligned}$$

The $m^2\ln m$ terms cancel and the remainder is $(2\ln 2-\frac{1}{2})m^2+O(m\log m)$. □

Proof of Theorem 6. There are at most $\binom{2m}{m}\le 4^m$ deletion paths from (m,m), so (8) gives

$$\begin{gathered} 2^{-W(m)}\le b_m^{(2m)}\le 4^m 2^{-W(m)}, \\ \log_2(1/b_m^{(2m)})=W(m)+O(m). \end{gathered}$$

By (5),

$$\log_2\frac{1}{\mathcal{L}(K_{m,m})}=\log_2\frac{1}{b_m^{(2m)}}+\log_2(2\times(2m)!).$$

Lemma 21 with $n = 2m$ and $m^2 = n^2/4$ therefore gives

$$\log_2 \frac{1}{\mathcal{L}(K_{m,m})} = \Big(\tfrac{1}{2} - \tfrac{1}{8\ln 2}\Big)n^2 + O(n \log n).$$

For odd n the same argument applied to $(m, m+1)$ changes only the $O(n \log n)$ term.

For the second statement, let $g(n)$ be the number of isomorphism classes of graphs of order n. Since $\mathcal{L}$ is a probability distribution on these classes, $\min_G \mathcal{L}(G) \le 1/g(n)$, and $g(n) \ge 2^{\binom{n}{2}}/n!$, so $\log_2(1/\min_G \mathcal{L}) \ge \binom{n}{2} - \log_2 n! = \frac{n^2}{2} + O(n \log n)$. Conversely (1) with $|\mathrm{Aut}(G)| \le n!$ gives $\min_G \mathcal{L} \ge 1/(n! P_n)$, and $\log_2 P_n = \sum_{i=1}^n \log_2 \binom{i-1}{\lfloor (i-1)/2 \rfloor} = \binom{n}{2} + O(n \log n)$, whence $\log_2(1/\min_G \mathcal{L}) \le \frac{n^2}{2} + O(n \log n)$.

Finally $\frac{1}{2} - \frac{1}{8\ln 2} < \frac{1}{2}$ strictly, so the two exponents differ in the leading term and the ratio is $2^{(1/(8\ln 2)+o(1))n^2}$; in particular $\mathcal{L}(K_{\lfloor n/2 \rfloor, \lceil n/2 \rceil}) > \min_G \mathcal{L}(G)$ for all large n. □

Remark 22. Numerically $\log_2(1/\mathcal{L}(K_{\lfloor n/2 \rfloor, \lceil n/2 \rceil}))/(n^2/2)$ equals 0.8158, 0.7515, 0.7130, 0.6852 and 0.6554 for $n = 10, 40, 80, 160, 640$, descending towards $2 \times (\frac{1}{2} - \frac{1}{8\ln 2}) = 1 - \frac{1}{4\ln 2} = 0.6393\ldots$, and $W(m)/m^2$ equals 1.2615 at $m = 320$ against the limit $2 - \frac{1}{2\ln 2} = 1.27865\ldots$.

7 Entropy

Proof of Theorem 7. The random choices are independent across steps, so the entropy of the whole sequence of choices is

$$\mathcal{H}_n = \sum_{t=1}^{n} \left(\log_2 t + \mathbb{E}_k \log_2 \binom{t-1}{k} \right),$$

with k uniform on $\{0, \ldots, t-1\}$. Since $\log_2 \binom{N}{k} = N H_2(k/N) + O(\log N)$ uniformly, where H_2 is the binary entropy in bits, and $\int_0^1 H_2(x)\,dx = 1/(2\ln 2)$, we get

$$\mathbb{E}_k \log_2 \binom{t-1}{k} = \frac{t}{2\ln 2} + O(\log t), \qquad \mathcal{H}_n = \frac{n^2}{4\ln 2} + O(n \log n).$$

The isomorphism class is a function of the choices, so its entropy is at most $\mathcal{H}_n$; conversely, conditioned on the class, the number of choice sequences producing it is at most $n!$, so the conditional entropy is at most $\log_2 n! = O(n \log n)$. The comparison with $\log_2 g(n) = \frac{n^2}{2} + O(n \log n)$ is as in the proof of Theorem 6.

It remains to justify the assertion about a typical output. Let Y_n denote the complete choice history and put

$$Z_n = -\log_2 \Pr(Y_n) = \sum_{t=1}^{n} \left(\log_2 t + \log_2 \binom{t-1}{K_t} \right),$$

where the K_t are independent and uniform on $\{0, \ldots, t-1\}$. The summands are independent and bounded in absolute value by $O(t)$, so $\mathrm{Var}(Z_n) = O(n^3)$. Chebyshev's inequality and the preceding entropy computation give

$$Z_n = \frac{n^2}{4\ln 2} + o_{\mathrm{p}}(n^2).$$

Now set $I_n = -\log_2 \mathcal{L}(G_n)$. Since G_n is a function of Y_n,

$$D_n := Z_n - I_n = -\log_2 \Pr(Y_n \mid G_n) \geq 0.$$

Moreover $\mathbb{E} D_n = H(Y_n \mid G_n) \leq \log_2 n! = O(n \log n)$, because each isomorphism class has at most $n!$ choice histories. Markov's inequality therefore gives $D_n = o_{\mathrm{p}}(n^2)$, and hence $I_n/n^2 \xrightarrow{\mathrm{p}} 1/(4\ln 2)$. $\square$

Exact values of the entropy for $n \leq 9$, computed from all isomorphism classes, are

$$0, 1, 1.918, 3.198, 4.793, 6.841, 9.444, 12.717, 16.729$$

bits, against

$$\log_2 g(n) = 0, 1, 2, 3.46, 5.09, 7.29, 10.03, 13.59, 18.07.$$

The ratio decreases slowly from 0.939 at $n = 6$ to 0.925 at $n = 9$, consistent with the limiting ratio $1/(2\ln 2) = 0.7213\ldots$ implied by Theorem 7. The concentration statement implies that there is a sequence $\varepsilon_n \downarrow 0$ such that the typical set

$$\mathcal{T}_n = \left\{ G : \left| -\frac{1}{n^2} \log_2 \mathcal{L}(G) - \frac{1}{4\ln 2} \right| \leq \varepsilon_n \right\}$$

satisfies $\Pr(G_n \in \mathcal{T}_n) = 1 - o(1)$. Bounding the likelihood of each member of $\mathcal{T}_n$ above and below and summing shows that $|\mathcal{T}_n| = 2^{n^2/(4\ln 2) + o(n^2)}$. Thus the process concentrates on a $2^{-(1/2 - 1/(4\ln 2))n^2 + o(n^2)}$ fraction of all isomorphism classes while still spreading over $2^{\Theta(n^2)}$ of them, which is the precise sense in which the uniform monkey is a poor but not degenerate sampler.

8 The maximum, and open problems

For the maximum no analogue of Theorem 2 is available, and the situation is genuinely irregular. Up to complementation, the maximisers of $\mathcal{L}$ are threshold graphs for $n \leq 6$; for $n = 7$ and $n = 8$ they are split but not threshold; and for $n = 9$ the maximiser is the disjoint union of two isolated vertices with a caterpillar on seven vertices, which is not split, being obtained from P_6 by attaching a pendant vertex at the third vertex. Thus neither the threshold graphs nor the broader class of split graphs contains the maximisers for all orders. The next three exact maximum values are

$$\max_{|V(G)|=10} \mathcal{L}(G) = \frac{219903017}{4938071040000},$$

$$\max_{|V(G)|=11} \mathcal{L}(G) = \frac{7472883629}{1140694410240000},$$

$$\max_{|V(G)|=12} \mathcal{L}(G) = \frac{18096017946691}{17566693917696000000}.$$

The complementary maximising pairs have graph6 representations `I????GICo` and `Ihuz~~~~w` for $n = 10$; `J?????A@OR?` and `Jhuz~~~~~~_` for $n = 11$; and `K????C?@?PAE` and `Kh~u||~~~~~~` for $n = 12$. What does appear stable is the order of magnitude: $-\log_2 \max_G \mathcal{L}(G)$ equals

$$0, 1, 1.585, 2.469, 3.815, 5.487, 7.439, 9.684, 12.021, 14.455, 17.220, 19.889$$

for $n \leq 12$, consistent with $\max_G \mathcal{L}(G) = 2^{\Theta(n)}/n!$.

Problem 23. Determine $\arg\min_G \mathcal{L}(G)$ for $n \geq 15$. In particular, is $C_5[\overline{K_3}]$ a minimiser at $n = 15$? Are the minimisers always blow-ups, and can their base graphs be chosen from a fixed finite family, or must their orders grow with n? Theorem 6 shows that $|\mathrm{Aut}|$ is of secondary importance on the n^2 logarithmic scale and suggests looking for graphs whose typical vertex orderings have back degrees near the centre of their range.

Problem 24. Settle Conjecture 1 for the remaining small orders $n = 16$ and $n = 18$.

Problem 25. Prove $A(G) \leq C^n|\mathrm{Aut}(G)|$ for an absolute constant C. Equivalently, prove

$$\max_G \mathcal{L}(G) = 2^{O(n)}/n!.$$

Problem 26. Settle the complexity of $\mathcal{L}$, raised in [1] and repeated in [2]. Is exact evaluation of A #P-hard? Corollary 9 is a dynamic program over vertex subsets, which suggests that A is computable in polynomial time on graphs of bounded pathwidth.

Problem 27. Study graph likelihood for a general attachment kernel, replacing the uniform choice of k by a distribution ν_t on $\{0, \dots, t-1\}$ as in [3]. Assuming $\nu_1(0) = 1$, the subset dynamic program extends by putting $F_\nu(\{v\}) = 1$ and

$$F_\nu(S) = |S| \sum_{v \in S} \frac{\nu_{|S|}(\deg_{G[S]} v) F_\nu(S \setminus \{v\})}{\binom{|S|-1}{\deg_{G[S]} v}}, \qquad \mathcal{L}_\nu(G) = \frac{F_\nu(V(G))}{n!\,|\mathrm{Aut}(G)|}.$$

Determine how the extrema and entropy depend on ν. The same recurrence turns $\mathcal{L}_\nu$ into an exactly computable marginal likelihood of an unlabelled graph, which is the quantity that likelihood-based inference for mechanistic network models must otherwise approximate.

Computational verification

All computations use exact integer or rational arithmetic, with `geng` and `nauty` [4] for graph generation and automorphism groups. The primary exhaustive certificate for Theorem 3 is an independently developed C plugin compiled directly into `geng`. It uses the exact automorphism-group order returned by `nauty`, checked unsigned 128-bit arithmetic for the integer-scaled subset recurrence, and exact 192-bit cross-products for rational likelihood comparisons. It rejects orders above 12 rather than risk overflow. The same program reproduced the previously known extrema through order 10 and independently reobtained the order-11 result before the order-12 production run; thus the claims through $n = 12$ do not depend on the preliminary screening programs supplied with the initial manuscript.

At order 12, complement symmetry reduces the computation to one representative from each of the 82 545 586 656 complement orbits of the 165 091 172 592 unlabelled graphs; the latter count includes 720 self-complementary graphs. The computation was divided into 4 800 exact `geng` residue shards and run with 14 workers. All shard outputs completed, with no failed shards or nonempty error files. Their equivalent full graph counts sum to 165 091 172 592 before the merge reports the global extrema. Aggregate helper CPU time was about 817 core-hours, and elapsed wall time was approximately 68 hours.

The principal implementation for individual larger graphs performs graph processing in the Wolfram Language and invokes a C helper that computes the scaled recurrence modulo multiple 64-bit primes; the exact result is reconstructed by the Chinese remainder theorem. It returns

$$A(C_5[\overline{K_3}]) = \frac{63977511069907}{4278226184221429440000000},$$

agreeing with (4) evaluated over the 4^5 vectors $m' \leq (3,3,3,3,3)$ and with the direct rational recurrence over the 2^{15} subsets.

The blow-up searches in Proposition 19 use a separate C program that reads the base graphs streamed by `geng`, enumerates the specified positive part vectors, and evaluates an integer-scaled form of (4). Dynamic-programming quantities use checked unsigned 256-bit arithmetic, automorphism orders are computed exactly by `nauty`, and likelihood comparisons use exact unsigned 320-bit cross-products. The program exits rather than returning a value if any checked bound is exceeded. It reproduces the likelihoods of the known minimising graphs through $n = 12$, the value of $C_5[\overline{K_3}]$ above, and the $n = 18$ rook-graph ratio reported after Proposition 19.

For the extended $C_5[m]$ calculation, a separate arbitrary-precision implementation enumerates positive five-part vectors modulo the dihedral symmetries of C_5 and evaluates (4) with exact rationals. It reproduces the ratios in Table 2, exhaustively checks every positive vector through order 50, and evaluates every nearly balanced vector through order 100. Independent exact evaluations in the Wolfram Language reproduce the resulting ratios at orders 26, 30 and 50.

For each base order p and target order n, the search checks the $\binom{n-1}{p-1}$ positive ordered part vectors for every base. The rows through $p = 9$ are exhaustive, while those for $10 \leq p \leq 12$ apply the exact necessary screens described above before likelihood evaluation. The screened order-11 and order-12 bases from the $n = 13$ searches are reused for $n = 14$. The order-12 stream was divided into 4 096 shards; the merged input counts sum to all 165 091 172 592 unlabelled bases and the retained file contains 140 474 distinct graph6 strings. The order-12, $n = 13$ CPU entry includes 125 939.981 seconds for the shared base screen and 67.805 seconds for the composition search. Parallel runs are merged only after verifying the total numbers of bases and blow-ups. Table 3 records all exact composition searches used here. CPU times are aggregate process times for the blow-up search program; they exclude CPU time used by `geng` for graph generation and are not elapsed wall times.

The remaining order-13 and order-14 searches use the complete tables of marks for $\mathfrak{S}_{12}$ and $\mathfrak{S}_{13}$ in TomLib. For $\mathfrak{S}_{14}$, subgroup classes above the required threshold are obtained by recursively taking maximal-subgroup class representatives and identifying subgroups conjugate in $\mathfrak{S}_{14}$. The traversal has 6280 classes and no failed maximal-subgroup computations. A separate C program expands the group orbits on unordered pairs using disjoint checked unsigned 128-bit masks and applies the exact twin screens described above. Graphs are canonically labelled by `nauty`, deduplicated, and passed to the same checked likelihood program used for the blow-up computations. Independent residue shards sum to the stated orbital representation counts before the canonical merge.

For the partial order-16 calculation, a checkpointed maximal-subgroup traversal above automorphism threshold 1556 produces 116 597 subgroup classes and no failed maximal-subgroup computations. The forced-twin test leaves 44 twin-free-possible classes and 4580 orbital representations. Direct twin testing leaves 738 representations, canonical labelling leaves 128 graphs, and exact likelihood evaluation finds no graph below $K_{8,8}$. This certifies the twin-free part only; the 116 553 subgroup classes that force a twin pair are not expanded into graphs by this calculation.

For the vertex-transitive search, the GAP 4.15.1 transitive group library was used [7]. At orders 13 and 14 there are 9 and 63 transitive permutation groups. Enumerating all unions of their orbits on unordered vertex pairs gives 152 and 1684 graph representations, which

n	p	bases	vectors/base	likelihood evals	CPU seconds
13	7	1 044	924	964 656	13.685
14	7	1 044	1 716	1 791 504	33.791
16	7	1 044	5 005	5 225 220	171.449
18	7	1 044	12 376	12 920 544	625.523
13	8	12 346	792	9 778 032	186.062
14	8	12 346	1 716	21 185 736	557.457
16	8	12 346	6 435	34 372 082	1 440.603
18	8	12 346	19 448	156 498 056	12 447.239
13	9	274 668	495	135 960 660	4 078.112
14	9	274 668	1 287	353 497 716	17 078.575
13	10	12 005 168	220	528 765	31.630
14	10	12 005 168	715	80 258	13.454
13	11	1 018 997 864	66	555 612	598.681
14	11	1 018 997 864	286	79 908	9.805
13	12	165 091 172 592	12	758 030	126 007.786
14	12	165 091 172 592	78	121 413	18.206

Table 3: Exact independent-blow-up searches. Here p is the base order; the order-8 rows for $n = 13, 14$ record intermediate staged checks. Rows through $p = 9$ evaluate every base/part-vector representation. The rows with $10 \leq p \leq 12$ use the necessary automorphism screens described in the text; their bases and vectors/base columns give the full unscreened mathematical input counts. The $p = 11$ and $p = 12$, $n = 14$ searches reused the bases saved during the corresponding $n = 13$ streams, so their CPU times exclude the shared initial scans.

canonical labelling reduces to 14 and 56 isomorphism classes. Likelihoods are then evaluated by the integer-scaled subset recurrence above.

The degree-sequence calculation enumerates all graphical sequences and applies Lemma 17 by integer cross multiplication. For its second deletion, the neighbours of each possible deleted vertex are assigned among the remaining degree classes so as to minimise the resulting lower bound; allowing the minimizing assignment independently for each vertex is a relaxation and therefore cannot discard a genuine competitor.

References

[1] C. R. S. Banerji, T. Mansour and S. Severini, *A notion of graph likelihood and an infinite monkey theorem*, arXiv:1304.3600 (2013); J. Phys. A: Math. Theor. **47** (2014) 035101, https://doi.org/10.1088/1751-8113/47/3/035101.

[2] D. Dervovic, A. Mocherla and S. Severini, *Constructing graphs with limited resources*, arXiv:1802.09844 (2018), https://arxiv.org/abs/1802.09844.

[3] S. Janson and S. Severini, *An example of graph limits of growing sequences of random graphs*, J. Comb. **4** (2013), no. 1, 67–80, https://doi.org/10.4310/JOC.2013.v4.n1.a3.

[4] B. D. McKay and A. Piperno, *Practical graph isomorphism, II*, J. Symbolic Comput. **60** (2014) 94–112, https://doi.org/10.1016/j.jsc.2013.09.003.

[5] OEIS Foundation Inc., *The On-Line Encyclopedia of Integer Sequences*, https://oeis.org; minimum and maximum likelihood by vertex count: A234234/A234235 and A234236/A234237, respectively.

[6] M. Grech and A. Kisielewicz, *Wreath product in automorphism groups of graphs*, J. Graph Theory **101** (2022) 29–51, https://doi.org/10.1002/jgt.22808.

[7] The GAP Group, *GAP—Groups, Algorithms, and Programming*, Version 4.15.1 (2025), https://www.gap-system.org.

[8] T. Merkwitz, L. Naughton and G. Pfeiffer, *TomLib—The GAP Library of Tables of Marks*, Version 1.2.11 (2024), https://gap-packages.github.io/tomlib.

[9] G. Sabidussi, *The composition of graphs*, Duke Math. J. **26** (1959) 693–696, https://doi.org/10.1215/S0012-7094-59-02667-5.

[10] G. Sabidussi, *The lexicographic product of graphs*, Duke Math. J. **28** (1961) 573–578, https://doi.org/10.1215/S0012-7094-61-02857-5.

[11] E. W. Weisstein, *Graph likelihood*, MathWorld—A Wolfram Resource, https://mathworld.wolfram.com/GraphLikelihood.html.

[12] Wolfram Research, `GraphData`, Wolfram Language documentation, https://reference.wolfram.com/language/ref/GraphData.html.