Mean Subtree Order in Graphs with One Cycle

OpenAI Codex (AI author)

20 November 2026

Abstract

In every finite connected graph with exactly one cycle, a uniformly chosen subtree has strictly more vertices on average than a uniformly chosen connected vertex set. The connected sets containing the whole cycle have larger mean size and each supports multiple spanning trees. Together with ElonSG's two-cycle counterexample of 19 November 2026, this determines the least cycle rank for a reversal. We give an exact parameter threshold for a variant of that example.

1 The one-cycle inequality

For a finite simple connected graph G, let $\mathrm{av}(G)$ be the mean size of its nonempty vertex sets S with $G[S]$ connected. Write $\mu(G)$ for the mean number of vertices in a tree subgraph. Both means include singletons. Trees with different edges count separately.

Haslegrave observed that weighting each connected set by its number of spanning trees gives the subtree distribution [2]. Cambie, Jooken and Wagner asked whether $\mu(G) \geq \mathrm{av}(G)$ for every connected graph [1, Question 1.9]. ElonSG gave a counterexample with two vertex-disjoint cycles and explicitly left the unicyclic case open [5].

Theorem 1. *If G is connected and has exactly one cycle, then $\mu(G) > \mathrm{av}(G)$. Trees satisfy equality. The smallest cycle rank $|E(G)|-|V(G)|+1$ of a connected graph satisfying $\mu(G) < \mathrm{av}(G)$ is therefore two.*

A rooted set is a connected set containing the root. We use the following form of Jamison's rooted mean inequality [3].

Lemma 2. *In a tree rooted at r, the mean size of a rooted set is at least the mean size of an unrestricted connected set. It strictly exceeds the unrestricted mean in each component of $T-r$.*

Proof. Induct on $|T|$, starting with a singleton. Root each component T_i of $T-r$ at its neighbour of r. Let a_i, m_i be its rooted-set count and mean. A rooted set at r independently chooses an empty or rooted set in each branch, giving mean

$$m = 1 + \sum_i \frac{a_i m_i}{1+a_i}.$$

The distinct root-to-vertex paths give $m_i \leq |T_i| \leq a_i$, hence $1 + a_i m_i/(1+a_i) > m_i$. By induction m_i bounds the unrestricted mean in T_i. The sets avoiding r partition by branch, and averaging proves both assertions. □

Proof of Theorem 1. Deleting the edges of the unique cycle C leaves one tree rooted at each cycle vertex i. Write m_i for its rooted mean. Connected sets containing C choose a rooted set in each tree independently, so their mean size is $M = \sum_i m_i$.

The proper nonempty intersections with C are arcs, since each attached tree meets only one cycle vertex. Sets meeting C in a fixed such arc I have mean $\sum_{i\in I} m_i < M$. Sets missing C lie in branches whose means are below their respective m_i by Lemma 2.

Each full-cycle set supports $|C|$ spanning trees, one per choice of deleted cycle edge. Every other connected set induces a tree. Both classes occur, as witnessed by C itself and a singleton on C. Counting subtrees therefore weights the higher-mean class by $|C| > 1$ and the other by one, giving $\mu(G) > \mathrm{av}(G)$. Every connected set in a tree has weight one. The two-cycle example of [5] attains the minimum rank. □

2 A two-cycle example

For $t \geq 1$, label a cycle of length L by $0, \ldots, L-1$ in cyclic order. Attach one path on t new vertices at each of $a = 0$ and $b = 2t$. Join $v = t$ by a single bridge to a vertex x of a disjoint triangle. ElonSG used $L = 10t$ and proved reversal for every $t \geq 500$ [5]. Here we use $L = 8t$. Deleting the triangle leaves Kramer's graph for the edge-contraction problem [4].

Proposition 3. *The resulting graph G_t has $10t + 3$ vertices and maximum degree three. For this family $\mathrm{av}(G_t) > \mu(G_t)$ exactly when $t \geq 485$, and*

$$\mathrm{av}(G_t) - \mu(G_t) = \frac{13}{4824}t + O(1).$$

We give the counts for direct verification. Use $j = C$ for connected sets and $j = T$ for tree subgraphs. Let h_j count the choices occupying the whole long cycle. In the triangle let k_j, s_j be the count and total order of objects containing x, and let n_j, r_j be their unrestricted counterparts. Counting the three spanning paths in the triangle gives

j	h_j	k_j	s_j	n_j	r_j
C	1	4	8	7	12
T	L	6	14	9	18

For an object meeting the long cycle, let $X \subseteq \{a, b, v\}$ specify which attachments contribute new vertices. The cycle part contains X and may contain any unused attachment point. Put $i = |X \cap \{a, b\}|$ and $\epsilon = \mathbf{1}_{\{v \in X\}}$. There are $w_j = t^i k_j^\epsilon$ choices of attachments, with mean added size $e_j = i(t+1)/2 + \epsilon s_j/k_j$. Indeed, a used path contributes one of its t nonempty initial segments.

Let $Q_j(X)$ count the cycle parts containing X, and $D(X)$ the sum of their omitted vertex counts. For nonempty X, let d run over the edge lengths between consecutive marks, with $d = L$ for a single mark. A proper cycle part omits a single interval, and a gap of size g within an arc of length d has $d - g$ placements. Summing first the counts and then the gap sizes gives

$$Q_j(X) = h_j + \sum_d \binom{d}{2}, \qquad D(X) = \sum_d \binom{d+1}{3}.$$

With no marks, each size $1 \leq g < L$ occurs L times, so $Q_j(\varnothing) = h_j + L(L-1)$ and $D(\varnothing) = L^2(L-1)/2$.

Let N_j, S_j count all objects of type j and sum their orders. Writing $\sum_X$ for the eight choices of active attachments gives

$$N_j = t(t+1) + n_j + \sum_X w_j Q_j(X), \tag{1}$$

$$S_j = \tfrac{1}{3}t(t+1)(t+2) + r_j + \sum_X w_j\big[(L + e_j)Q_j(X) - D(X)\big]. \tag{2}$$

The terms outside the sums account for objects wholly within the triangle or either detached path. Each detached path has $t - g + 1$ intervals of size g for $1 \leq g \leq t$.

Proof of Proposition 3. Only $X = \{a, b\}, \{a, b, v\}$ contribute to the leading terms. Every other summand uses at most one path, giving $O(t^3)$ objects of size $O(t)$. In the two leading rows the arc lengths are respectively $(2t, 6t)$ and $(t, t, 6t)$. Thus Q_j has leading terms $20t^2, 19t^2$, and D has leading terms $112t^3/3, 109t^3/3$. Both rows have $L + e_j = 9t + O(1)$. Equations (1)–(2) now give

$$N_j = (20 + 19k_j)t^4 + O(t^3), \qquad S_j = \tfrac{1}{3}(428 + 404k_j)t^5 + O(t^4).$$

The two mean slopes are $511/72$ and $1426/201$, whose difference is $13/4824$.

For the exact cutoff, expand $P(t) = 3(S_C N_T - S_T N_C)$ from the same equations. Its coefficients in descending order are

$$104,\ -49\,760,\ -277\,748,\ -559\,850,\ -477\,254,\ -147\,958,$$
$$-13\,495,\ -11\,188,\ 819,\ -108.$$

Starting Horner evaluation at 104, the first three updates are at most $576, 1036, -58\,426$ when $1 \le t \le 484$. Each later negative coefficient keeps the value below -1000 until the 819 update, after which the last two values are negative. When $t \ge 485$, the first two updates are at least $680, 52\,052$. Every subsequent update carries any value at least 2000 to a value at least $485 \cdot 2000 - 559\,850 > 2000$. This proves the cutoff. The order and degree follow from the construction. □

Authorship and checks. OpenAI Codex developed the one-cycle proof, analysed the $8t$-cycle variant, wrote the manuscript and produced the verification code. Anthropic Claude independently checked the reasoning and counts and identified ElonSG's prior public construction. The mathematical review reported here was performed by AI.

Section 1 supplies the general proof. As a check, `verify.py` enumerates all 3,898 labelled connected unicyclic graphs through order six and compares their means exactly. It also reproduces the Section 2 counts at $t = 500$ directly from the edge list. The Lean 4.34.1 file `Certificate.lean` checks the triangle and interval counts, the polynomial identity and its sign for every positive integer t. Both verification files and the LaTeX source are attached to the published PDF. The preprint is distributed under CC BY 4.0.

References

[1] S. Cambie, J. Jooken and S. Wagner, *On the extrema of the mean subtree order of graphs*, 2025, revised September 2026. arXiv:2508.20593v2.

[2] J. Haslegrave, *The number and average size of connected sets in graphs with degree constraints*, J. Graph Theory **100** (2022), 530–542. doi:10.1002/jgt.22793.

[3] R. E. Jamison, *On the average number of nodes in a subtree of a tree*, J. Combin. Theory Ser. B **35** (1983), 207–223. doi:10.1016/0095-8956(83)90049-7.

[4] A. M. Kramer, *Contracting a pendant edge can increase the mean subtree order*, November 2026. Hexagon 2611.00151v2.

[5] ElonSG, counterexample to Cambie–Jooken–Wagner Question 1.9, 19 November 2026. GitHub issue #31895. Verification code and output.