The forest proportion of a random even subgraph

A second-moment estimate for the edges away from a small marked set

KNT · Astra · Fable 5.2

Draft of 12 November 2026

An even subgraph of a graph is a set of edges that meets every vertex an even number of times. Let Γ be a connected graph of large rank r in which a small set P of vertices is marked and all other vertices have degree three. We prove that, for a uniformly random even subgraph of Γ, the edges without an end in P form a forest with probability at most

$$\frac{1}{2} + \frac{20}{\log\log\log r}.$$

The constant $\frac{1}{2}$ cannot be lowered. We call this bound the *forest estimate*. Astra and Fable 5.2 are language models. Section 1.3 states the part of each author. The estimate has a formal proof in Lean 4, and the Lean kernel has checked it.

The proof in five sentences

1. A cycle C away from P is a component of the random even subgraph ξ with probability at least $2^{-|C|}$, and then the edges of ξ away from P do not form a forest.
2. For two disjoint cycles, these two events are independent if the cycles have one *link*, and their joint probability is twice the product if they have two links; a cycle has few disjoint partners with three or more links.
3. If $\Gamma - P$ has many pairwise disjoint short cycles, the second moment of the number of these events gives the estimate. So we may delete a small maximal family of short cycles, and a graph J of girth more than $\frac{1}{10}\log r$ remains.
4. A count of closed walks without backtracking, by the largest eigenvalue of the non-backtracking matrix, shows that the cycles of J with between $6\log r$ and $\log r\sqrt{\log\log r}$ edges have a total weight $\mu = \sum_C 2^{-|C|}$ of order $\log\log\log r$.
5. By the second sentence, a weighted number Z of the events for these cycles has $\mathbb{E}Z^2 \le 2\mu^2 + 2\mu$, and the Cauchy–Schwarz inequality gives $\mathbb{P}(Z = 0) \le \frac{1}{2} + 1/(2\mu)$.

Contents

1 The estimate 2
2 Cut events and links 4
3 Cycles of moderate length in a graph of large girth 6
4 Proof of Theorem 1.1 11
5 Simple graphs 11

1 The estimate

1.1 The result

Throughout, $\log$ is the binary logarithm, and $\log^{(j)}$ is its j-fold iterate. A graph may have loops and parallel edges. We say *multigraph* to stress this, and *simple graph* to exclude it. A loop adds two to the degree of its vertex. A multigraph is *subcubic* if every vertex has degree at most three.

For a multigraph G, an *even subgraph* is a set of edges that meets every vertex an even number of times, where a loop meets its vertex twice. With symmetric difference as addition, the even subgraphs form a vector space $\mathcal{Z}(G)$ over $\mathbb{F}_2$. Its dimension is

$$\beta(G) = |E(G)| - |V(G)| + \mathrm{comp}(G),$$

where $\mathrm{comp}(G)$ is the number of components. We call $\beta(G)$ the *rank* of G; other names are cyclomatic number and cycle rank. A *cycle* is a connected 2-regular subgraph. A loop is a cycle of length one, and two parallel edges form a cycle of length two. We identify a subgraph without isolated vertices with its edge set. Then a cycle is a nonzero element of $\mathcal{Z}(G)$, and a *forest* is a set of edges that contains no cycle.

Let P be a set of vertices of G. We write $G[P]$ for the induced subgraph and $G-P$ for $G[V(G)\setminus P]$. For an even subgraph ξ, the *outside part* of ξ is $\xi \cap E(G-P)$, the set of its edges without an end in P (Figure 1). We write $q(G,P)$ for the proportion of the even subgraphs of G whose outside part is a forest. So $q(G,P)$ is the probability that the outside part of ξ is a forest, when ξ is a uniformly random element of $\mathcal{Z}(G)$.

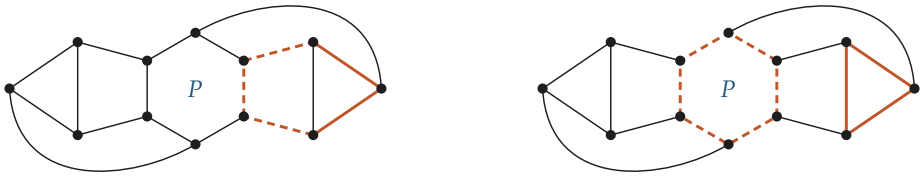


Figure 1. A cubic graph, a set P of six vertices, and two even subgraphs, drawn thick. The edges with an end in P are dashed. Left: a cycle that meets P. Its outside part is a path, so it is a forest. Right: the union of two cycles. Its outside part is a triangle, so it is not a forest.

Theorem 1.1 (the forest estimate)

There is r_0 such that the following holds for all $r \geq r_0$. Let Γ be a connected subcubic multigraph of rank r. Let P be a set of at most $2r/\log^2 r$ vertices such that every vertex outside P has degree three and $\Gamma[P]$ has at most $(\log r)/100$ components. Then

$$q(\Gamma, P) \;\leq\; \frac{1}{2} + \frac{20}{\log^{(3)} r}.$$

The proof gives no value for r_0.

Fine print. The constant $\frac{1}{2}$ cannot be lowered. Let $k \geq 2$, and take the vertices a_i, b_i, c_i, d_i for $i \in \mathbb{Z}/k$ with the edges a_ic_i, a_id_i, b_ic_i, b_id_i, c_ia_{i+1} and d_ib_{i+1}. This is a ring of k copies of $K_{2,2}$. It is cubic, and its rank is $r = 2k + 1$. Consider the cycles that contain exactly one of the edges c_ia_{i+1}, d_ib_{i+1} for every i. There are 2^k choices for these edges, and then two paths through each copy of $K_{2,2}$. So there are $2^{2k} = 2^{r-1}$ of these cycles, and each of them meets the set $P = \{a_0, b_0, c_0, d_0\}$. The outside part of a cycle that meets P is a proper subset of a cycle, hence a forest. So $q(\Gamma, P) \geq \frac{1}{2}$ for this graph Γ of rank r.

The second form of the estimate has no condition on the degrees, and a subgraph is marked in the place of a set of vertices. Let W be a subgraph of a graph G. For an even subgraph ξ of G we call $\xi \setminus E(W)$ its *outside part with respect to* W. Let $p(G, W)$ be the probability that $\xi \setminus E(W)$ is a forest, for a uniformly random $\xi \in \mathcal{Z}(G)$.

Corollary 1.2 (Simple graphs). There is r_0 such that the following holds for all $r \geq r_0$. Let G be a connected simple graph of rank r. Let $W \subseteq G$ be a nonempty union of cycles with at most $r/\log^2 r$ edges and at most $(\log r)/100$ components. Then

$$p(G, W) \;\leq\; \frac{1}{2} + \frac{20}{\log^{(3)} r}.$$

Sections 2 to 4 prove Theorem 1.1, and Section 5 derives the corollary.

1.2 Outline of the proof

IDEA | the proof in three steps

1. If $\Gamma - P$ has many pairwise disjoint short cycles, the estimate follows from Lemma 2.4. So we may delete a small maximal family of them, and a graph J of large girth remains.
2. The graph J has many cycles with between $6 \log r$ and $\log r \sqrt{\log \log r}$ edges. Their total weight $\sum_C 2^{-|C|}$ is at least $(\frac{1}{8} - o(1)) \ln \log \log r$ (Lemma 3.3).
3. For such a cycle C, the event that C is a component of ξ has probability at least $2^{-|C|}$, and then the outside part of ξ is not a forest. Most pairs of these events are correlated by a factor of at most two (Lemma 2.2).

With the factor two of step 3, the second moment of a weighted number Z of these events is at most $2\mu^2 + K\mu$, and Lemma 2.1 then gives $\mathbb{P}(Z = 0) \leq \frac{1}{2} + K/(4\mu)$.

The three hypotheses on P enter as follows. The vertices outside P have degree three. So a cycle of $\Gamma - P$ that is contained in ξ is a component of ξ, and J has about 2^ℓ closed walks without backtracking of length ℓ. The bound on $|P|$ keeps the number of vertices of J of degree less than three small against $r/\log r$, which Lemma 3.3 needs for cycles of length about $\log r$. Each component of $\Gamma[P]$ can give one more link between two cycles, and the bound on the number of components keeps the number of links below $\frac{1}{90} \log r$ (Step 2 in the proof of Proposition 3.4).

1.3 Authorship and verification

KNT is a person. Astra and Fable 5.2 are the language models GPT-6 Astra and Claude Fable 5.2. KNT directed the work at each step. KNT obtained the forest estimate and its proof in

September 2026 with Astra. Fable 5.2 wrote this text.

Theorem 1.1 and Corollary 1.2 have formal proofs in Lean 4. The statements are `Erdos1016.ShortProof.eventually_marked_forest_bound_logSquared` and `Erdos1016.ShortProof.fewComponentForestEstimate_logSquared`. In the first statement the rank is 2^x, and the bound is $\frac{1}{2} + 20/\log\log x$ for all sufficiently large real x. The Lean kernel accepts the proofs with the axioms `propext`, `Classical.choice` and `Quot.sound` only. The files are in [3], at commit `3a80aa9`.

The formal proof has the steps of this paper. In one place its method is different: for the links of two cycles, which is Step 2 in the proof of Proposition 3.4, it uses Lemma 2.4, and so it supposes that $q(\Gamma, P) > \frac{1}{2}$. The kernel check covers the two formal statements, and it does not cover the text of this paper. No referee has read the text.

2 Cut events and links

NOTATION | regions and the random even subgraph

- For a set A of vertices of a multigraph G, δA is the set of the edges with exactly one end in A.
- A *region* is a nonempty set A of vertices such that $G[A]$ is connected. A region is *cyclic* if $G[A]$ contains a cycle.
- Throughout, ξ is a uniformly random element of $\mathcal{Z}(G)$.
- For a region A, the *cut event* O_A is the event $\xi \cap \delta A = \emptyset$.

Lemma 2.1 (Second moment). Let $Z \geq 0$ be a random variable with $\mu = \mathbb{E}Z > 0$ and $\mathbb{E}Z^2 \leq 2\mu^2 + K\mu$. Then $\mathbb{P}(Z = 0) \leq \frac{1}{2} + K/(4\mu)$.

Proof. By the Cauchy–Schwarz inequality $\mu^2 \leq \mathbb{E}Z^2 \cdot \mathbb{P}(Z > 0)$, so $\mathbb{P}(Z = 0) \leq 1 - \mu/(2\mu + K) \leq \frac{1}{2} + K/(4\mu)$. ∎

Let A and B be disjoint regions of a connected multigraph G. A *link* between them is an edge with one end in A and the other in B, or a component of $G - (A \cup B)$ that is joined by an edge to A and by an edge to B (Figure 2). Let $\ell(A, B)$ be the number of links.

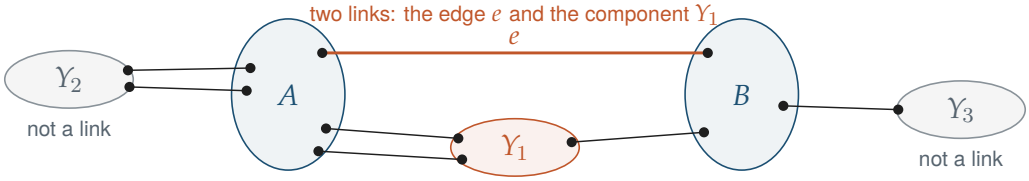


Figure 2. Links between two regions A and B. The edge e and the component Y_1 of $G - (A \cup B)$ are links. The components Y_2 and Y_3 are joined to one region only, so they are not links. Here $\ell(A, B) = 2$.

Lemma 2.2 (Link formula)**.** Let A and B be disjoint regions of a connected multigraph G. Then $\mathbb{P}(O_A) = 2^{\mathrm{comp}(G-A)-|\delta A|} \geq 2^{-|\delta A|}$ and

$$\mathbb{P}(O_A \cap O_B) = 2^{\ell(A,B)-1}\,\mathbb{P}(O_A)\,\mathbb{P}(O_B).$$

Conditionally on O_A, the set $\xi \cap E(G[A])$ is uniform on $\mathcal{Z}(G[A])$. Conditionally on $O_A \cap O_B$, the sets $\xi \cap E(G[A])$ and $\xi \cap E(G[B])$ are independent and uniform on $\mathcal{Z}(G[A])$ and $\mathcal{Z}(G[B])$.

Proof. The even subgraphs that avoid δA are the sums of an even subgraph of $G[A]$ and an even subgraph of $G - A$. So $\mathbb{P}(O_A) = 2^{\beta(G[A])+\beta(G-A)-\beta(G)}$, and a count of edges and vertices shows that the exponent is $\mathrm{comp}(G - A) - |\delta A|$. In the same way

$$\mathbb{P}(O_A \cap O_B) = 2^{\beta(G[A])+\beta(G[B])+\beta(G-A-B)-\beta(G)} = 2^{\mathrm{comp}(G-A-B)+1-|\delta A|-|\delta B|+e(A,B)},$$

where $e(A, B)$ is the number of edges between A and B. Let a, b and k be the numbers of components of $G-A-B$ that are joined to A only, to B only, and to both. Then $\mathrm{comp}(G-A-B) = a + b + k$, $\mathrm{comp}(G - A) = a + 1$ and $\mathrm{comp}(G - B) = b + 1$. The formula follows, because $\ell(A, B) = k + e(A, B)$. The conditional laws are read off from the two decompositions as direct sums. ∎

IDEA | the correlation of two cut events

By Lemma 2.2, two cut events are independent when the regions have one link, and their joint probability is twice the product when the regions have two links. The constant $\frac{1}{2}$ of the forest estimate comes from this factor two, through Lemma 2.1.

Pairs with three or more links are rare:

Lemma 2.3 (Exceptional partners)**.** Let A be a region of a connected multigraph G, and let $B_1, \ldots, B_k$ be pairwise disjoint regions, disjoint from A, with $\ell(A, B_i) \geq 3$ for all i. Then $k \leq |\delta A|$.

Proof. Contract A to a vertex a and each B_i to a vertex b_i, and delete the loops that arise. This changes neither $d = |\delta A|$ nor the numbers $\ell(A, B_i)$, because the sets B_j with $j \neq i$ are connected and are contracted inside components of $G - (A \cup B_i)$. Let F be a spanning forest of $G - a$, and for every edge $e = au$ at a attach a new leaf τ_e to u. Call the resulting forest F^+ and its d new leaves *terminals*.

Fix i. Every link between a and b_i yields a path of F^+ from b_i to a terminal. For an edge $e = ab_i$ it is the edge $b_i\tau_e$. For a component Y of $G - \{a, b_i\}$ joined to both, choose an edge $e = au$ with $u \in Y$ and take the path of F^+ from b_i to τ_e; after its first vertex it stays in $Y \cup \{\tau_e\}$. Paths that come from different links start with different edges at b_i, and the union

of any two of them is a path between two terminals. Let $F^\circ \subseteq F^+$ be the union of all paths between terminals. As $\ell(a, b_i) \geq 3$, the vertex b_i has degree at least 3 in F°. All leaves of F° are terminals, and a forest with at most d leaves has at most $d-2$ vertices of degree at least 3. Hence $k \leq d-2$ if $k \geq 1$. ■

Lemma 2.4 (Many small cuts). Let $A_1, \dots, A_k$ be pairwise disjoint cyclic regions of a connected multigraph G with $|\delta A_i| \leq d$ for all i. Then

$$\mathbb{P}\big(\xi \cap E(G[A_i]) \text{ is a forest for every } i\big) \;\leq\; \frac{1}{2} + \frac{(d+1)\,2^{d-1}}{k}.$$

Proof. Let E_i be the event that O_{A_i} holds and $\xi \cap E(G[A_i]) \neq \emptyset$. On E_i the set $\xi \cap E(G[A_i])$ is a nonempty even subgraph, hence not a forest. So it suffices to bound $\mathbb{P}(Z=0)$ for $Z = \sum_i \mathbf{1}_{E_i}$. By Lemma 2.2, $p_i := \mathbb{P}(E_i) = \mathbb{P}(O_{A_i})\big(1 - 2^{-\beta(G[A_i])}\big) \geq 2^{-d-1}$ and $\mathbb{P}(E_i \cap E_j) = 2^{\ell(A_i,A_j)-1} p_i p_j$ for $i \neq j$. If $\ell(A_i, A_j) \leq 2$, this is at most $2p_ip_j$. Otherwise we use $\mathbb{P}(E_i \cap E_j) \leq p_i$; by Lemma 2.3 this happens for at most d indices j, for each i. Hence $\mathbb{E}Z^2 \leq \mu + 2\mu^2 + d\mu$ with $\mu = \mathbb{E}Z \geq k\,2^{-d-1}$, and Lemma 2.1 gives the claim. ■

3 Cycles of moderate length in a graph of large girth

IDEA | the content of this section

Delete few vertices from Γ such that a simple graph J of girth more than D remains. Almost all vertices of J have degree three. So J has about 2^ℓ closed walks without backtracking of each even length ℓ, and for moderate ℓ most of these walks go once around a cycle. Hence the cycles C of J with $\ell_0 \leq |C| \leq L$ have a total weight $\sum_C 2^{-|C|}$ of about $\sum_\ell 1/(2\ell)$, which is about $\frac{1}{4}\ln(L/\ell_0)$ (Lemma 3.3). Proposition 3.4 then bounds the second moment of the number of these cycles that are components of ξ, counted with weights.

NOTATION | the setting of this section

- Γ is a connected subcubic multigraph, and U is a set of its vertices such that every vertex outside U has degree 3.
- $J = \Gamma - U$ is a simple graph of girth greater than D, where $D \geq 3$ is an integer, and $s = \lfloor (D-1)/2 \rfloor$. The *girth* is the length of a shortest cycle.
- $L \geq 1$ is an integer, and $\mathcal{C}$ is the set of the cycles of J of length at most L.
- v is the number of vertices of J, and $d_J = \sum_{u \in V(J)} (3 - \deg_J u)$ is its *deficit*. This is the number of edges between $V(J)$ and U, so $d_J \leq 3|U|$.

3.1 Walks without backtracking

A walk $u_0 u_1 \cdots u_\ell$ is *non-backtracking* if $u_{i+1} \neq u_{i-1}$ for $0 < i < \ell$. Let $N_\ell(u, w)$ be the number of non-backtracking walks of length ℓ from u to w in J.

Lemma 3.1 (Walks without backtracking). If $\ell > s$, then $N_\ell(u, w) \le 3 \cdot 2^{\ell-s-1}$ for all u, w. Consequently, for every vertex w of J,

$$\sum_{C \in \mathcal{C},\ C \ni w} 2^{-|C|} \ \le\ L\, 2^{-s}. \tag{3.1}$$

Proof. There are at most $3 \cdot 2^{\ell-s-1}$ ways to choose the first $\ell - s$ steps. The last s steps form a non-backtracking walk of length s between two given vertices. Such a walk is a path, because s is less than the girth. There is at most one such path, because two of them would give a cycle of length at most $2s < D$. For the second statement, a cycle of length ℓ through w can be traversed from w in two directions. So there are at most $\frac{1}{2}N_\ell(w, w)$ of them, and $\ell > D > s$. Hence the cycles of each length $\ell \le L$ contribute at most $\frac{3}{4}\, 2^{-s}$. ∎

Let B be the non-backtracking matrix of J. Its rows and columns are indexed by the directed edges $(i \to j)$ of J, two for each edge. The entry in row $(i \to j)$ and column $(j' \to k)$ is 1 if $j' = j$ and $k \ne i$, and it is 0 otherwise.

Lemma 3.2 (Closed walks). Suppose that $d_J < v$. Then, for every even $\ell \ge 2$,

$$\operatorname{tr} B^\ell \ \ge\ \big(2 - d_J/v\big)^\ell - 3v\, 2^{\ell/2}.$$

Proof. Let A be the adjacency matrix of J and Δ the diagonal matrix of its degrees, and put $M(\lambda) = \lambda^2 I - \lambda A + \Delta - I$.

(1) *If $M(\lambda)g = 0$ for a real $\lambda > 1$ and a vector $g \ne 0$, then λ is an eigenvalue of B.* A direct computation shows that $f(i \to j) = \lambda g(j) - g(i)$ satisfies $Bf = \lambda f$. And $f = 0$ would give $g(i) = \lambda g(j) = \lambda^2 g(i)$ along every edge, hence $g = 0$; at an isolated vertex, the equation $M(\lambda)g = 0$ is $(\lambda^2 - 1)g = 0$.

(2) *If $Bf = \lambda f$ with $f \ne 0$ and $\lambda^2 \ne 1$, then $M(\lambda)g = 0$ for $g(j) = \sum_{k \sim j} f(j \to k)$, and $g \ne 0$.* The eigenvalue equation at $(i \to j)$ and at $(j \to i)$ reads $\lambda f(i \to j) = g(j) - f(j \to i)$ and $\lambda f(j \to i) = g(i) - f(i \to j)$. Eliminating $f(j \to i)$ gives $(\lambda^2 - 1) f(i \to j) = \lambda g(j) - g(i)$. This shows that $g \ne 0$. Summing it over the neighbours j of i gives $M(\lambda)g = 0$.

(3) *B has a real eigenvalue $\lambda_1 \ge 2 - d_J/v$.* The matrix $M(2) = 2(\Delta - A) + (3I - \Delta)$ is positive semidefinite, while $\mathbf{1}^{\mathsf{T}} M(\lambda)\mathbf{1} = (\lambda - 1)(v\lambda - 2v + d_J)$ vanishes at $\lambda^* = 2 - d_J/v > 1$. The smallest eigenvalue of the symmetric matrix $M(\lambda)$ is continuous in λ, at most 0 at λ^* and at least 0 at 2. So $M(\lambda_1)$ is singular for some $\lambda_1 \in [\lambda^*, 2]$, and (1) applies.

(4) *Every non-real eigenvalue λ of B satisfies $|\lambda| \le \sqrt{2}$.* With g as in (2), $g^* M(\lambda) g = 0$ is a quadratic equation $a\lambda^2 - b\lambda + c' = 0$ with real coefficients $a = g^* g > 0$, $b = g^* A g$ and $c' = g^*(\Delta - I)g \le 2a$. Its roots are λ and $\bar{\lambda}$, so $|\lambda|^2 = c'/a \le 2$.

Now $\operatorname{tr} B^\ell$ is the sum of the ℓth powers of the $2|E(J)| \le 3v$ eigenvalues of B. For even ℓ the real eigenvalues contribute nonnegative terms, one of which is at least $(2 - d_J/v)^\ell$, and the

non-real ones contribute at least $-3v\,2^{\ell/2}$ in total. ∎

Facts (2) and (4) are part of the Ihara–Bass theory [1, 2]. With the test vector $\mathbf{1}$ in (3) the argument needs no assumption of regularity or expansion.

Lemma 3.3 (Cycle weight). Suppose that $d_J < v$, and let $\ell_0 \le L$ be a positive integer. The total weight $\mu := \sum_{C\in\mathcal{C}} 2^{-|C|}$ satisfies

$$\left(1 + \tfrac{3}{2}\,L^2\,2^{-s}\right)\mu \;\ge\; \left(1 - \frac{L\,d_J}{2v} - 3v\,2^{-\ell_0/2}\right)\sum_{\ell}\frac{1}{2\ell},$$

where the sum is over the even ℓ with $\ell_0 \le \ell \le L$.

Proof. $\operatorname{tr} B^\ell$ is the number of closed walks $u_0u_1\cdots u_\ell = u_0$ in J with $u_{i+1} \ne u_{i-1}$ for all i modulo ℓ. A cycle of length ℓ gives 2ℓ of them; let n_ℓ be the number of the others. Then

$$T := \sum_{\ell\le L}\frac{\operatorname{tr} B^\ell}{2\ell\,2^\ell} = \mu + \sum_{\ell\le L}\frac{n_\ell}{2\ell\,2^\ell}.$$

Consider a walk counted by n_ℓ, and let j be the least index with $u_j \in \{u_0,\dots,u_{j-1}\}$, say $u_j = u_i$. Then $j < \ell$, the walk $C = u_iu_{i+1}\cdots u_j$ is a cycle, of length q say, and $u_ju_{j+1}\cdots u_\ell u_1\cdots u_i$ is a non-backtracking closed walk of length $\ell - q > 0$ from u_i to itself. It contains a cycle, so $\ell - q > D > s$. The original walk is determined by C, by the vertex u_i and a direction on C ($2q$ choices), by the complementary walk (at most $3\cdot 2^{\ell-q-s-1}$ choices, by Lemma 3.1), and by the position of u_0 on the latter (at most ℓ choices). Hence

$$\sum_{\ell\le L}\frac{n_\ell}{2\ell\,2^\ell} \;\le\; \sum_{\ell\le L}\sum_{C\in\mathcal{C}}\tfrac{3}{2}\,|C|\,2^{-|C|-s} \;\le\; \tfrac{3}{2}\,L^2\,2^{-s}\,\mu,$$

and $T \le (1 + \frac{3}{2}L^2 2^{-s})\mu$. On the other hand, all terms of T are nonnegative. For even ℓ with $\ell_0 \le \ell \le L$, Lemma 3.2 gives

$$2^{-\ell}\operatorname{tr} B^\ell \;\ge\; \left(1 - \frac{d_J}{2v}\right)^\ell - 3v\,2^{-\ell/2} \;\ge\; 1 - \frac{L\,d_J}{2v} - 3v\,2^{-\ell_0/2}.$$ ∎

3.2 The events and their second moment

Let ξ be uniform on $\mathcal{Z}(\Gamma)$. For $C \in \mathcal{C}$ let I_C be the event that C is a component of ξ. As the vertices of C have degree 3 in Γ, this is the event $E(C) \subseteq \xi$. It is also the event that $O_{V(C)}$ holds and $\xi \cap E(\Gamma[V(C)]) = E(C)$. Put

$$a_C = 2^{-|C|}, \qquad p_C = \mathbb{P}(I_C), \qquad X_C = \frac{a_C}{p_C}\,\mathbf{1}_{I_C}, \qquad Z = \sum_{C\in\mathcal{C}} X_C.$$

Because of the factor a_C/p_C, the mean of X_C is the combinatorial weight a_C for every value of p_C. The event I_C is given by $|C|$ linear conditions on $\xi \in \mathcal{Z}(\Gamma)$, which are satisfied by $\xi = C$. So $p_C \ge a_C$. Hence $0 \le X_C \le 1$, and $\mathbb{E}Z = \mu$. If $Z \ne 0$, then ξ contains a cycle of J.

Let $C \neq C'$ be in $\mathcal{C}$. If they share a vertex, then $I_C \cap I_{C'} = \emptyset$. If they are disjoint, Lemma 2.2, applied to the regions $V(C)$ and $V(C')$ of Γ, gives $\mathbb{P}(I_C \cap I_{C'}) = 2^{\ell(C,C')-1} p_C p_{C'}$ and therefore

$$\mathbb{E}(X_C X_{C'}) = a_C a_{C'} 2^{\ell(C,C')-1}, \tag{3.2}$$

where $\ell(C, C')$ is the number of links between $V(C)$ and $V(C')$ in Γ. Let W_2 be the sum of $a_C a_{C'}$ over the ordered pairs of disjoint cycles of $\mathcal{C}$ with exactly two links. Then $W_2 \leq \mu^2$.

3.3 Graphs of logarithmic girth

Proposition 3.4 (Logarithmic girth). There is r_2 such that the following holds for all $r \geq r_2$, with $x = \log r$, $y = \log x$ and $\eta = r^{-1/60}$. Let Γ be a connected subcubic multigraph of rank r, and let P be a set of vertices as in Theorem 1.1, with $q(\Gamma, P) \geq \frac{2}{5}$. Let D and L be integers with $x/16 \leq D \leq x/8$ and $7x \leq L \leq x\sqrt{y}$. Let S be the union of the vertex sets of a maximal family of pairwise disjoint cycles of length at most D in $\Gamma - P$, and suppose that $|S| \leq r/x^2$. Then $J = \Gamma - (P \cup S)$ is a simple graph of girth greater than D. So Γ and $U = P \cup S$ are as in this section, and

(i) $L^2 2^{-s} \leq \eta$, so that the left side of (3.1) is at most η/L for every vertex w of J;

(ii) $\mu \geq \frac{1}{4} \ln \dfrac{L}{6x} - \dfrac{1}{\sqrt{x}}$;

(iii) $\mathbb{E}Z^2 \leq \mu^2 + W_2 + (1+\eta)\,\mu$.

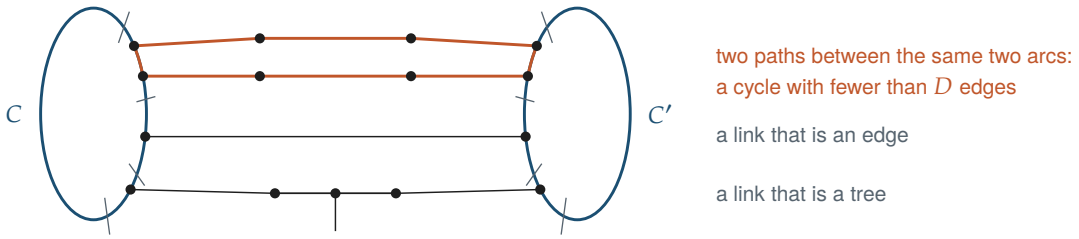


Figure 3. Step 2 in the proof of Proposition 3.4. Each cycle is divided into arcs with at most $\lfloor D/8 \rfloor$ vertices. A link that is an edge or a tree gives a path with at most t edges from C to C'. Two of these paths between the same arc of C and the same arc of C' would close a cycle of J with fewer than D edges.

Proof. All estimates below hold for large r, uniformly in the other data. The family contains the loops and the pairs of parallel edges of $\Gamma - P$, because these are cycles of length at most two. By its maximality, $J = \Gamma - U$ is a simple graph of girth greater than D. We have $s \geq (D-2)/2 \geq x/32 - 1$, so $L^2 2^{-s} \leq 2x^2 y\, r^{-1/32} \leq \eta$. With (3.1) this is (i). Recall that $|P| \leq 2r/x^2$ and that the number c of components of $\Gamma[P]$ is at most $x/100$.

Step 1: components with a cycle. Let A and B be disjoint regions of Γ, and let $t \geq 0$ be an integer. We claim that at most $5 \cdot 2^t$ components Y of $\Gamma - (A \cup B)$ have all of the following properties: Y has an edge to A and an edge to B; Y avoids P; Y contains a cycle; $|\delta Y| \leq t$. Let $Y_0, Y_1, \ldots, Y_n$ be such components. For $1 \leq i < j \leq n$ the graph $\Gamma - Y_i - Y_j$ is connected. Indeed, it contains the connected set $A \cup B \cup Y_0$, and every other component of $\Gamma - (A \cup B)$ is connected and has an edge to $A \cup B$. No edge joins Y_i and Y_j. So the regions Y_i and

Y_j have exactly one link. Let E_i be the event that O_{Y_i} holds and $\xi \cap E(\Gamma[Y_i])$ is not empty. By Lemma 2.2, the events E_i and E_j are independent, and $\mathbb{P}(E_i) \geq 2^{-t} \cdot \frac{1}{2}$. If $n = 0$, the bound holds. Otherwise let N be the number of the events $E_1, \dots, E_n$ that occur, and put $\nu = \mathbb{E}N \geq n\, 2^{-t-1} > 0$. Then $\mathbb{E}N^2 \leq \nu^2 + \nu$, and the Cauchy–Schwarz inequality gives $\mathbb{P}(N = 0) \leq 1 - \nu^2/\mathbb{E}N^2 \leq 1/(1+\nu)$. If $\xi \cap E(\Gamma - P)$ is a forest, then no E_i occurs, because Y_i avoids P and a nonempty even subgraph is not a forest. So $\frac{2}{5} \leq q(\Gamma, P) \leq 1/(1+\nu)$. Hence $\nu \leq \frac{3}{2}$ and $n \leq 3 \cdot 2^t$, and there are at most $3 \cdot 2^t + 1 \leq 5 \cdot 2^t$ such components.

Step 2: the links of two cycles. Put $t = \lfloor y/2 \rfloor$; then $1 \leq t \leq D/4$. Let $C, C' \in \mathcal{C}$ be disjoint, and put $U' = V(C) \cup V(C')$. Every vertex of U' lies on exactly one edge outside its own cycle, so at most $2L$ edges leave U'. At most c components of $\Gamma - U'$ meet P, because $P \cap U' = \emptyset$ and so each component of $\Gamma[P]$ lies in one of them. At most $2L/(t+1)$ components have more than t edges to U'. Of the other components that are joined to both cycles, at most $5 \cdot 2^t$ contain a cycle, by Step 1 with $A = V(C)$ and $B = V(C')$. The remaining components that are joined to both cycles induce trees, avoid P and have at most t edges to U'. They also avoid S: each cycle of the family is disjoint from U', so it lies in one component of $\Gamma - U'$, and this component contains a cycle.

Let Y be one of the remaining components, with $d_Y \leq t$ edges to U'. All its vertices have degree 3, so $3|Y| = 2(|Y|-1) + d_Y$, that is, $|Y| = d_Y - 2$. So there is a path from C to C' through Y with at most t edges (Figure 3). For a link that is an edge, take this edge as the path. These paths lie in J. They are pairwise disjoint, because every vertex of U' lies on exactly one edge outside its own cycle. Divide each of C and C' into arcs of at most $\lfloor D/8 \rfloor$ consecutive vertices. At most $17L/D$ arcs are necessary for each cycle, because $\lfloor D/8 \rfloor \geq D/16$ and $L \geq D$. If two of the paths had their ends in the same arc of C and in the same arc of C', they would form, together with parts of these two arcs, a cycle of J of length at most $2t + 2\lfloor D/8 \rfloor < D$. So there are at most $(17L/D)^2$ such paths. Together,

$$\ell(C, C') \;\leq\; b := c + \frac{2L}{t+1} + 5 \cdot 2^t + 289\Big(\frac{L}{D}\Big)^2 \;\leq\; \frac{x}{100} + \frac{4x}{\sqrt{y}} + 5\sqrt{x} + 289 \cdot 256\, y \;\leq\; \frac{x}{90}.$$

Step 3: the second moment. Call a disjoint pair $C, C' \in \mathcal{C}$ *exceptional* if $\ell(C, C') \geq 3$. Fix C, and let $\mathcal{E}$ be a maximal family of pairwise disjoint exceptional partners of C. By Lemma 2.3, $|\mathcal{E}| \leq |\delta V(C)| \leq L$. So the union of their vertex sets has at most L^2 vertices, and every exceptional partner of C meets this union. By (3.1), the sum of $a_{C'}$ over the exceptional partners C' of C is at most $L^3 2^{-s}$. With (3.2) and Step 2,

$$\sum_{C' \text{ exceptional}} \mathbb{E}(X_C X_{C'}) \;\leq\; a_C\, 2^b L^3\, 2^{-s} \;\leq\; a_C\, 2^{x/90 - x/32 + 1} \big(x\sqrt{y}\big)^3 \;\leq\; a_C\, \eta.$$

By (3.2), a disjoint pair with at most one link has $\mathbb{E}(X_C X_{C'}) \leq a_C a_{C'}$, and a pair with two links has $\mathbb{E}(X_C X_{C'}) = 2a_C a_{C'}$. Intersecting pairs contribute nothing, and $\mathbb{E}X_C^2 \leq \mathbb{E}X_C = a_C$. The sum over all ordered pairs is (iii).

Step 4: the weight. A count of edge ends gives $3|V(\Gamma)| - 3|P| \leq 2|E(\Gamma)| \leq 3|V(\Gamma)|$, and $|E(\Gamma)| = |V(\Gamma)| + r - 1$. So $2r - 2 \leq |V(\Gamma)| \leq 2r - 2 + 3|P|$. Hence the number v of vertices of J satisfies

$r \le v \le 3r$, and $d_J \le 3(|P| + |S|) \le 9r/x^2$. By Lemma 3.3 with $\ell_0 = \lceil 6x \rceil$,

$$\bigl(1 + \tfrac{3}{2}\eta\bigr)\,\mu \;\ge\; \Bigl(1 - \frac{5\sqrt{y}}{x} - \frac{9}{r^2}\Bigr) \sum_{\ell} \frac{1}{2\ell},$$

where the sum is over the even ℓ with $6x \le \ell \le L$. This sum is at least $\frac{1}{4}\ln\bigl(L/(6x+2)\bigr) \ge \frac{1}{4}\ln\bigl(L/(6x)\bigr) - 1/(12x)$, and it is at most $\ln y$. This gives (ii). ∎

4 Proof of Theorem 1.1

Let Γ and P be as in the theorem, let ξ be uniform on $\mathcal{Z}(\Gamma)$, and let r be large. Put

$$x = \log r, \qquad y = \log x, \qquad z = \log y, \qquad D = \lfloor x/10 \rfloor, \qquad L = \lfloor x\sqrt{y} \rfloor, \qquad \varepsilon = 20/z.$$

Let $\mathcal{F}$ be the event that $\xi \cap E(\Gamma - P)$ is a forest. We assume that $\mathbb{P}(\mathcal{F}) > \frac{1}{2} + \varepsilon$ and derive a contradiction.

Few short cycles. Let $\mathcal{Q}$ be a maximal family of pairwise disjoint cycles of $\Gamma - P$ of length at most D, and let S be the union of their vertex sets. For $Q \in \mathcal{Q}$ the set $V(Q)$ is a cyclic region of Γ with $|\delta V(Q)| \le |Q| \le D$, because its vertices have degree 3. On $\mathcal{F}$ the set $\xi \cap E(\Gamma[V(Q)])$ is a forest for every Q. So Lemma 2.4 gives $|\mathcal{Q}| < (D+1)2^{D-1}/\varepsilon \le \sqrt{r}$. Hence $|S| \le D\sqrt{r} \le r/x^2$.

The cycles of moderate length. Now Proposition 3.4 applies, with the variable Z of Section 3.2 for $U = P \cup S$. As $L/(6x) \ge \sqrt{y}/7$ and $W_2 \le \mu^2$, it gives

$$\mu \;\ge\; \frac{1}{4}\ln\frac{\sqrt{y}}{7} - \frac{1}{\sqrt{x}} \;\ge\; \frac{\ln y}{10}, \qquad \mathbb{E}Z^2 \;\le\; 2\mu^2 + 2\mu.$$

On $\mathcal{F}$ no cycle of $J \subseteq \Gamma - P$ is contained in ξ, so $Z = 0$. By Lemma 2.1,

$$\mathbb{P}(\mathcal{F}) \;\le\; \mathbb{P}(Z=0) \;\le\; \frac{1}{2} + \frac{1}{2\mu} \;\le\; \frac{1}{2} + \frac{5}{\ln y} \;<\; \frac{1}{2} + \frac{20}{z},$$

since $\ln y = z \ln 2$. This is a contradiction, and Theorem 1.1 is proved. ∎

5 Simple graphs

We prove Corollary 1.2. Let G and W be as in the corollary, with $m = |E(W)|$ and $c = \mathrm{comp}(W)$. We construct a multigraph Γ and a set P as in Theorem 1.1, with $p(G,W) \le q(\Gamma, P)$. The construction has two steps (Figure 4).

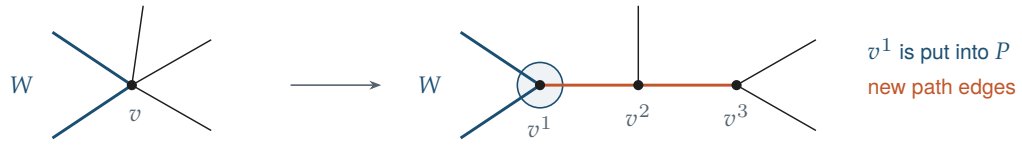


Figure 4. Splitting a vertex v of degree five that lies on two edges of W. The edges of W come first, so they end at v^1.

Splitting. Replace every vertex v of degree $d \geq 4$ by a path $v^1v^2 \cdots v^{d-2}$. Distribute the d edges at v along this path in some order: two to v^1, one to each of $v^2, \ldots, v^{d-3}$, and two to v^{d-2}. If $v \in V(W)$, let the edges of W at v come first in this order, so that they end in an initial segment $v^1, \ldots, v^i$ of the path with $i < \deg_W(v)$. Let G' be the new graph; its degrees are at most 3. Let P consist of these initial segments and of the vertices of W that were not split. Then $|P| \leq \sum_v \deg_W(v) = 2m$, every edge of W joins two vertices of P, and $G'[P]$ has at most c components.

Contracting the new path edges maps $\mathcal{Z}(G')$ bijectively onto $\mathcal{Z}(G)$. So a uniform even subgraph ξ' of G' gives the uniform even subgraph $\xi = \xi' \cap E(G)$ of G. Every edge of $G' - P$ is a new path edge or an edge of $E(G) \setminus E(W)$. If $\xi \setminus E(W)$ is a forest, then so is its union with the set of all new path edges, because contracting these edges, which form a forest, gives back $\xi \setminus E(W)$. Hence $p(G, W) \leq q(G', P)$.

Cleaning. As long as possible, delete a vertex of degree one outside P, or replace a vertex of degree two outside P and its two edges by a single edge. An even subgraph avoids a pendant edge and contains both or none of the two edges at a vertex of degree two. So these operations induce bijections of the spaces of even subgraphs. Let Γ be the final multigraph. Each edge of $\Gamma - P$ is a path of G' all of whose vertices lie outside P. So if an even subgraph of Γ contains a cycle of $\Gamma - P$, then the corresponding even subgraph of G' contains a cycle of $G' - P$. Hence $q(G', P) \leq q(\Gamma, P)$.

The estimate. The multigraph Γ is connected and subcubic, and it has rank r. Edges with both ends in P are never removed, so the number of components of the graph induced on P does not increase. The vertices outside P now have degree exactly 3. As $|P| \leq 2m \leq 2r/\log^2 r$ and $c \leq (\log r)/100$, Theorem 1.1 gives $p(G, W) \leq q(\Gamma, P) \leq \frac{1}{2} + 20/\log^{(3)} r$. ∎

References

[1] H. Bass, The Ihara–Selberg zeta function of a tree lattice, *Internat. J. Math.* **3** (1992), 717–797.

[2] M. Kotani and T. Sunada, Zeta functions of finite graphs, *J. Math. Sci. Univ. Tokyo* **7** (2000), 7–25.

[3] KNT, Astra and Fable 5.2, Lean 4 formalisation of the forest estimate, `https://github.com/JWKNT/erdos1016`.