THE CONSTANT STEINER KERNEL IS NOT OPTIMAL AT LARGE GIRTH

MATTHEW KWAN AND HUY TUAN PHAM

Abstract. In the paper "nonlinear lower-tail large deviations at criticality", we bound the number of order-n Steiner triple systems with no Pasch configuration. In doing so, we show that a certain variational problem is minimised by a constant kernel. Using ChatGPT 6 Pro, we also discovered that the analogous variational problems for Steiner triple systems of girth at least g, for large g, are *not* minimised by the constant kernel, and we included a brief remark to this effect. In this note (not intended for publication), we present the details of this proof.

Note: This proof was discovered with ChatGPT 6 Pro, and we are not attempting to claim any kind of credit for it; this note only exists to share information that may be of interest to the community. Much of the text in this note was produced by ChatGPT 6 Pro and Claude Opus 5.5 (at maximum effort).

hexagon:2611.00166v1

1. Introduction

A *Steiner triple system* (STS) of order n is a collection of triples from $[n]$ such that every pair of points lies in exactly one triple; it has $\binom{n}{2}/3$ triples. For $k \geq 4$, an *Erdős k-configuration* is a set of k triples spanning only $k+2$ points, such that every subset of s of its triples with $2 \leq s < k$ spans at least $s+3$ points. (For $s = 2$ this says that two triples share at most one point, as they must in an STS.) The smallest one is the Pasch configuration, four triples on six points. Erdős configurations play a role for triple systems similar to that of cycles for graphs, and an STS has *girth at least g* if it contains no Erdős k-configuration with $4 \leq k < g$. We write $\mathcal{E}_k$ for the set of isomorphism classes of Erdős k-configurations.

Glock, Kühn, Lo and Osthus [2] conjectured an asymptotic formula for the number of STS of order n with girth at least g, and Kwan, Sah, Sawhney and Simkin [4] proved the corresponding lower bound. In [3, Theorem 1.17], we proved the matching upper bound for $g = 5$. Our proof reduces the count to a variational problem over "profiles" of triples, and then shows [3, Lemma 8.10] that the uniform profile is optimal. In [3, Remark 1.19] we report that for large g the uniform profile is not optimal for the corresponding variational problems, so that our method, in this form, gives upper bounds weaker than conjectured. This note proves a precise version of that statement. As remarked in [3], we suspect that this provides evidence that the conjecture is false.

1.1. Theorem statement.

A *Steiner kernel* on a probability space (Ω, μ) is a symmetric measurable function $W\colon \Omega^3 \to [0,\infty)$ such that, for almost every $(x,y) \in \Omega^2$, we have $\int W(x,y,z)\,\mathrm{d}\mu(z) = 1$; we call this the *pair constraint*.

For a Steiner kernel W, we define the entropy cost

$$D(W) = \int_{\Omega^3} W \log W \,\mathrm{d}\mu^3.$$

This quantity is nonnegative because $\int W = 1$, and it vanishes only for the constant kernel $W \equiv 1$. Also, for a 3-graph F, define the *homomorphism density*

$$t_F(W) = \int_{\Omega^{V(F)}} \prod_{e \in E(F)} W(x_e)\,\mathrm{d}\mu^{V(F)}(x).$$

and let $\mathrm{Aut}(F)$ be the automorphism group of F. For $F \in \mathcal{E}_k$, let

$$a_F = \frac{6}{|\mathrm{Aut}(F)|}.$$

and define the girth-g functional

$$I_g(W) = D(W) + \sum_{4 \leq k < g} \sum_{F \in \mathcal{E}_k} a_F\, t_F(W). \tag{1.1}$$

Matthew Kwan was supported by ERC Starting Grant "RANDSTRUCT" No. 101076777. Huy Tuan Pham was supported by a Clay Research Fellowship, a Sloan Research Fellowship and NSF Career grant DMS-2543870.

At the constant kernel, this functional equals

$$C_g := I_g(1) = \sum_{4 \le k < g} \sum_{F \in \mathcal{E}_k} a_F,$$

the predicted number of forbidden configurations per triple in a uniformly random STS.

For $g = 5$ the only forbidden configuration is the Pasch configuration, so $C_5 = 1/4$. In [3, Lemma 8.10] we prove the inequality $I_5(W) \ge 1/4$ for kernels on finite uniform spaces, even without the pair constraint (only $\int W = 1$ is needed). For general g the analogous statement would be

$$I_g(W) \ge C_g \qquad \text{for every Steiner kernel } W, \tag{1.2}$$

that is, the constant kernel would minimise I_g. It does not.

Theorem 1.1. *There is a symmetric measurable function U with $|U| \le 1$ on a probability space, with zero pair marginals, $\int U(x,y,z)\,\mathrm{d}\mu(z) = 0$ for all x, y, such that for all sufficiently large g the Steiner kernel*

$$W_g = 1 + g^{-g/5}\,U$$

satisfies $I_g(W_g) < C_g$.

To comment on the connection to the Glock–Kühn–Lo–Osthus conjecture: we would conjecture that for a random order-n Steiner triple system S,

$$\log \mathbb{P}[S \text{ has girth at least } g] = -\frac{\binom{n}{2}}{3} \cdot \min_W I_g(W) + o(n^2).$$

If this conjecture were true, then Theorem 1.1 would contradict the conjectured formula in [2] for the number of STS with girth at least g. It would be a very significant undertaking to actually prove this conjecture (in addition to the ideas in [3], presumably one would need some version of the techniques in [4] or [1]), but we nonetheless view Theorem 1.1 as strong evidence that the Glock–Kühn–Lo–Osthus conjecture is false.

To prove Theorem 1.1, the approach is to consider a slight perturbation $1 + \varepsilon U$. This perturbation costs about ε^2 in entropy, whereas the configuration terms only respond at order ε^4. What changes with g is the number of configuration terms, which grows like g^g. We look for a sub-configuration that has negative density in some admissible direction U, and that is contained in so many Erdős configurations that the resulting negative terms add up to more than the entropy cost, even though ε is tiny.

Section 2 shows that the pair constraint severely limits which sub-configurations can play this role: the first ones whose contributions get multiplied by the number of configurations have six edges, and there are only two of them, the 3×3 grid and one other. Section 3 builds U from a theorem of Sidorenko, which provides a signed kernel with negative grid density, together with a three-colour lift that enforces the pair constraint and switches the other six-edge configuration off. Section 4 proves Theorem 1.1 assuming that grids are abundant inside Erdős configurations, and Section 5 proves this by completing a grid at random.

2. What a perturbation of the constant kernel sees

Let $U\colon \Omega^3 \to [-1,1]$ be symmetric with zero pair marginals. Then $W = 1 + \varepsilon U$ is a Steiner kernel for $0 \le \varepsilon \le 1$. To compare $I_g(1+\varepsilon U)$ with C_g, expand each density by multiplying out $\prod_e (1 + \varepsilon U(x_e))$:

$$t_F(1 + \varepsilon U) = \sum_{A \subseteq E(F)} \varepsilon^{|A|}\, t_A(U), \tag{2.1}$$

where $t_A(U)$ is the density of the 3-graph formed by the edges in A (vertices of F covered by no edge of A simply integrate to 1).

The pair constraint now does most of the work. If some vertex v lies in exactly one edge $e \in A$, then integrating over x_v first replaces $U(x_e)$ by its pair marginal, which is zero, so $t_A(U) = 0$. Call A *leafless* if every vertex covered by A lies in at least two edges of A. Only leafless sets contribute to (2.1): the empty set contributes 1, and single edges contribute nothing.

Which leafless edge sets does an Erdős configuration contain? The whole configuration is one. Indeed, if a vertex of $F \in \mathcal{E}_k$ lay in only one edge e, then the other $k-1$ edges would span at most $k+1$ vertices, contrary to the definition. So $A = E(F)$ contributes $\varepsilon^k\, t_F(U)$. Proper subsets are far more restricted, because the definition of $\mathcal{E}_k$ forces them to be sparse while leaflessness forces them to be dense.

Lemma 2.1. *Let $F \in \mathcal{E}_k$, and let $A \subsetneq E(F)$ be nonempty and leafless. Then $|A| \geq 6$. If $|A| = 6$, then A covers exactly nine vertices, each lying in exactly two edges of A, and A is isomorphic either to the 3×3 grid G or to the configuration P dual to the triangular prism.*

Here the *grid* G has the nine cells of a 3×3 array as its vertices, and its three rows and three columns as its edges. The configuration P is described in the proof.

Proof. Let $j = |A|$, and let v be the number of vertices covered by A. A single edge has leaves, so $j \geq 2$, and then the definition of $\mathcal{E}_k$ gives $v \geq j+3$. Since each covered vertex lies in at least two of the j edges, $3j \geq 2v$. Together these give $3j \geq 2j+6$, so $j \geq 6$; and if $j = 6$ both inequalities are equalities, so $v = 9$ and every vertex lies in exactly two edges of A.

Such an A is determined by a graph. Let A^* be the graph whose vertices are the six edges of A, with one edge for each vertex of A, joining the two edges of A that contain it. Each edge of A has three vertices, so A^* is 3-regular; it has no loops, and no parallel edges because two edges of F share at most one vertex. Conversely, A is recovered from A^*: its vertices are the edges of A^*, and its edges are the stars of the vertices of A^*. There are exactly two simple 3-regular graphs on six vertices, $K_{3,3}$ and the triangular prism. If $A^* = K_{3,3}$ with parts $\{r_1, r_2, r_3\}$ and $\{c_1, c_2, c_3\}$, then the vertices of A are the nine edges $r_i c_j$, and its edges are the stars of the r_i and of the c_j. These are the rows and columns of a 3×3 array, so $A \cong G$. The prism gives the other configuration, which we call P. □

Summing (2.1) over configurations, $I_g(1+\varepsilon U) - C_g$ consists of the entropy term $D(1+\varepsilon U)$ together with one term $a_F\, \varepsilon^{|A|}\, t_A(U)$ for each $F \in \mathcal{E}_k$ $(4 \leq k < g)$ and each nonempty leafless $A \subseteq E(F)$. By Theorem 2.1 these terms come in three kinds:

- whole configurations with at most six edges, a fixed finite family contributing $O(\varepsilon^4)$ independently of g;
- proper six-edge subsets, which are copies of G or of P and contribute $\varepsilon^6\, t_G(U)$ or $\varepsilon^6\, t_P(U)$ each;
- sets with at least seven edges.

The entropy term is of order ε^2, with a constant independent of g. The six-edge terms are therefore the first ones whose total coefficient grows with g, since the numbers of copies of G and P grow with the number of configurations. This tells us what to ask of U:

(U1) U is symmetric, $|U| \leq 1$, and U has zero pair marginals;
(U2) $t_G(U) < 0$;
(U3) $t_P(U) = 0$.

Condition (U3) spares us from comparing the number of copies of P with the number of copies of G. Under (U1)–(U3), the six-edge terms add up to $-|t_G(U)|\, b_g\, \varepsilon^6$, where

$$b_g = \sum_{4 \leq k < g} \sum_{F \in \mathcal{E}_k} a_F\, N_G(F),$$

and $N_G(F)$ denotes the number of copies of G in F. We will see that b_g, like C_g, is $g^{(1+o(1))g}$. The grid terms then beat the entropy cost as soon as $b_g \varepsilon^4 \to \infty$, i.e. for ε somewhat larger than $g^{-g/4}$. They also beat the terms with at least seven edges, whose total weight is at most $2^g C_g$ and which are each $O(\varepsilon^7)$, provided $\varepsilon \leq e^{-Cg}$ for a suitable constant C. The choice $\varepsilon = g^{-g/5}$ lies comfortably inside this window.

3. An admissible direction with negative grid density

Condition (U2) asks for a direction in which the grid has negative density, and this is provided by a theorem of Sidorenko. In his terminology it says that the grid is not *positive*; this is the failure of positivity mentioned in [3, Remark 1.19].

Theorem 3.1 (Sidorenko)**.** *There are a probability space (Ω_0, μ_0) and a bounded symmetric measurable function $V\colon \Omega_0^3 \to \mathbb{R}$ with $t_G(V) < 0$.*

For a proof, see [5, Proposition 1.5]. Dividing V by $\|V\|_\infty$, we may assume $|V| \leq 1$. Nothing else about V will matter: it may be nonzero when two of its arguments coincide, and its pair marginals may be arbitrary.

Such a V is not yet an admissible direction. Nothing forces its pair marginals to vanish, and without that $1+\varepsilon V$ is not a Steiner kernel, and the expansion of Section 2 would no longer discard edge sets with leaves. We repair this by tensoring with a kernel whose pair marginals do vanish. Given a symmetric function K on Λ^3 for a probability space (Λ, ν), set

$$U\big((x,\alpha),(y,\beta),(z,\gamma)\big) = V(x,y,z)\, K(\alpha,\beta,\gamma) \qquad \text{on } (\Omega_0 \times \Lambda)^3.$$

Integrating over the third argument factorises, so U has zero pair marginals as soon as K does, whatever V is; and $t_A(U) = t_A(V)\,t_A(K)$ for every 3-graph A. Thus (U2) and (U3) become conditions on K: we need $t_G(K) > 0$, to keep the sign of $t_G(V)$, and $t_P(K) = 0$.

Both G and P are 2-regular: each vertex lies in exactly two edges. This makes the most obvious choice useless. On $\Lambda = \mathbb{Z}/2\mathbb{Z}$, the kernel $K(\alpha,\beta,\gamma) = (-1)^{\alpha+\beta+\gamma}$ has zero pair marginals, but in $\prod_{e\in A} K(x_e)$ each colour of a 2-regular A appears twice, so $t_A(K) = 1$: this K cannot tell G from P. And P does have to be switched off: we know nothing about $t_P(V)$, and if it were positive, copies of P would contribute with the wrong sign.

It turns out that three colours suffice. On $\Lambda = \mathbb{Z}/3\mathbb{Z}$ with the uniform measure, let $\omega = e^{2\pi i/3}$ and

$$K(\alpha,\beta,\gamma) = \operatorname{Re}\omega^{\alpha+\beta+\gamma} = \begin{cases} 1 & \text{if } \alpha+\beta+\gamma \equiv 0 \pmod 3, \\ -\frac{1}{2} & \text{otherwise.} \end{cases}$$

Its pair marginals vanish because $\sum_\gamma \omega^{\alpha+\beta+\gamma} = 0$. The point of a cube root of unity is that the two edges at a vertex can only cancel each other's characters if they carry conjugate ones, and this turns $t_A(K)$ into a count of 2-colourings.

Lemma 3.2. *Let A be a connected linear 3-graph in which every vertex lies in exactly two edges, and let A^* be the graph defined in the proof of Theorem 2.1. Then $t_A(K) = 2^{1-|A|}$ if A^* is bipartite, and $t_A(K) = 0$ otherwise.*

Proof. Write $K = \frac{1}{2}(\chi + \bar\chi)$ with $\chi(\alpha,\beta,\gamma) = \omega^{\alpha+\beta+\gamma}$, and expand the product over the edges of A as a sum over choices $\sigma \in \{\pm 1\}^A$ of χ or $\bar\chi$ for each edge:

$$\prod_{e\in A} K(x_e) = 2^{-|A|} \sum_\sigma \prod_{e\in A} \omega^{\sigma_e \sum_{v\in e} x_v} = 2^{-|A|} \sum_\sigma \prod_v \omega^{(\sigma_e+\sigma_{e'})x_v},$$

where e and e' denote the two edges containing v. Averaging over $x_v \in \mathbb{Z}/3\mathbb{Z}$ gives 1 if $\sigma_e + \sigma_{e'} \equiv 0 \pmod 3$ and 0 otherwise; as $\sigma_e + \sigma_{e'} \in \{-2, 0, 2\}$, the condition is $\sigma_e = -\sigma_{e'}$. So $t_A(K)$ is $2^{-|A|}$ times the number of σ taking opposite values on any two edges that share a vertex, which is the number of proper 2-colourings of A^*. A connected graph has two of them if it is bipartite, and none otherwise. □

The grid's graph G^* is $K_{3,3}$, which is bipartite, so $t_G(K) = 2^{-5}$; the prism contains triangles, so $t_P(K) = 0$. Collecting what we have, U satisfies (U1)–(U3):

Proposition 3.3. *Let V be as in Theorem 3.1 with $|V| \le 1$, let $\Omega = \Omega_0 \times \mathbb{Z}/3\mathbb{Z}$ with the product of μ_0 and the uniform measure, and let $U = V \otimes K$ be as above. Then U is symmetric, $|U| \le 1$, U has zero pair marginals, and*

$$t_G(U) = \frac{t_G(V)}{32} < 0, \qquad t_P(U) = 0.$$

Proof. U is symmetric because V and K are, and $|U| \le 1$ because $|V| \le 1$ and $|K| \le 1$. The rest follows from $t_A(U) = t_A(V)\,t_A(K)$ and Theorem 3.2. □

4. Proof of Theorem 1.1, given that grids are abundant

Let U be as in Theorem 3.3, and write $a = -t_G(U) > 0$. The bookkeeping of Section 2 gives, for $g \ge 7$ and $0 \le \varepsilon \le 1$, the exact identity

$$I_g(1+\varepsilon U) - C_g = D(1+\varepsilon U) + \sum_{k=4}^{6} \sum_{F\in\mathcal{E}_k} a_F\,\varepsilon^k\,t_F(U) - a\,b_g\,\varepsilon^6 + R_g(\varepsilon), \tag{4.1}$$

where $R_g(\varepsilon)$ is the sum of $a_F\,\varepsilon^{|A|}\,t_A(U)$ over all $F \in \mathcal{E}_k$ with $4 \le k < g$ and all leafless $A \subseteq E(F)$ with $|A| \ge 7$. Indeed, $t_F(1+\varepsilon U) - 1$ is the sum of $\varepsilon^{|A|}\,t_A(U)$ over nonempty leafless $A \subseteq E(F)$. The sets $A = E(F)$ with $k \le 6$ give the second term. By Theorem 2.1, proper leafless sets have at least six edges, and those with exactly six are copies of G, each contributing $\varepsilon^6\,t_G(U)$, or of P, contributing nothing. Everything else has at least seven edges.

Every term except the grid term is small:

- Since $\log(1+u) \le u$, we have $(1+u)\log(1+u) - u \le u^2$ for $u \ge -1$. As $\int \varepsilon U = 0$, this gives

$$D(1+\varepsilon U) = \int \big[(1+\varepsilon U)\log(1+\varepsilon U) - \varepsilon U\big] \le \varepsilon^2 \int U^2 \le \varepsilon^2.$$

- As $|U| \le 1$, every $|t_A(U)|$ is at most 1. The second term in (4.1) is therefore at most $C_7\,\varepsilon^4$ in absolute value, where $C_7 = \sum_{k=4}^{6} \sum_{F \in \mathcal{E}_k} a_F$ is an absolute constant. We never need to know its sign.
- A configuration with $k < g$ edges has fewer than 2^g edge subsets, so $|R_g(\varepsilon)| \le 2^g C_g\,\varepsilon^7$.

Hence

$$I_g(1+\varepsilon U) - C_g \le (1+C_7)\,\varepsilon^2 - a\,b_g\,\varepsilon^6 + 2^g C_g\,\varepsilon^7. \tag{4.2}$$

Everything now depends on how large b_g is compared with C_g.

Lemma 4.1. *As $g \to \infty$,*

$$C_g \le \exp\bigl(g\log g + O(g)\bigr) \qquad \textit{and} \qquad b_g \ge \exp\bigl(g\log g - O(g)\bigr).$$

The upper bound is a crude count of all possible configurations. The lower bound says that the weighted number of pairs (Erdős configuration, grid inside it) is as large as the weighted number of configurations, up to a factor $e^{O(g)}$. This does not require grids to be typical: heuristically, only a polynomially small proportion of Erdős configurations contain a grid. But polynomial, and even exponential, losses are invisible on the scale $g\log g$.

Proof of Theorem 1.1, assuming Theorem 4.1. Take $\varepsilon = g^{-g/5} = \exp(-\frac{1}{5}g\log g)$ in (4.2). Then

$$\begin{aligned}(1+C_7)\,\varepsilon^2 &= \exp\bigl(-\tfrac{2}{5}g\log g + O(1)\bigr),\\ a\,b_g\,\varepsilon^6 &\ge \exp\bigl(-\tfrac{1}{5}g\log g - O(g)\bigr),\\ 2^g C_g\,\varepsilon^7 &\le \exp\bigl(-\tfrac{2}{5}g\log g + O(g)\bigr).\end{aligned}$$

The middle term exceeds the other two by a factor $\exp(\frac{1}{5}g\log g - O(g)) \to \infty$, so $I_g(1+\varepsilon U) < C_g$ for all sufficiently large g. □

As anticipated in Section 2, the exponent $\frac{1}{5}$ has no special role: the same computation works for $\varepsilon = g^{-\theta g}$ with any $\theta \in (0, \frac{1}{4})$.

5. Grids are abundant: proof of Theorem 4.1

Both bounds are statements about labelled configurations. By the orbit–stabiliser theorem, an Erdős k-configuration F has $(k+2)!/|\operatorname{Aut}(F)|$ labelled copies on the vertex set $[k+2]$. Hence

$$\sum_{F \in \mathcal{E}_k} a_F = \frac{6\,L_k}{(k+2)!} \qquad \text{and} \qquad \sum_{F \in \mathcal{E}_k} a_F\,N_G(F) = \frac{6}{(k+2)!}\sum_{F'} N_G(F'), \tag{5.1}$$

where L_k is the number of Erdős k-configurations on $[k+2]$, and F' ranges over them.

Upper bound. An Erdős k-configuration on $[k+2]$ is a set of k triples from $[k+2]$, so

$$L_k \le \binom{\binom{k+2}{3}}{k} \le \frac{(k+2)^{3k}}{k!} = \exp\bigl(2k\log k + O(k)\bigr).$$

Since $(k+2)! = \exp(k\log k + O(k))$, (5.1) gives $\sum_{F \in \mathcal{E}_k} a_F \le \exp(k\log k + O(k))$, and summing over $k < g$ gives $C_g \le \exp(g\log g + O(g))$.

Lower bound: completing a grid at random. Fix $k = g-1$, and a copy of G on nine of the vertices in $[k+2]$. By (5.1), $b_g \ge 6L_k^G/(k+2)!$, where L_k^G is the number of Erdős k-configurations on $[k+2]$ that contain the fixed grid. So we need $L_k^G \ge \exp(2k\log k - O(k))$: up to a factor $e^{O(k)}$, as many configurations through a fixed grid as the upper bound allows for all configurations. We obtain them by completing the grid at random.

What should a completion look like? In an Erdős k-configuration every vertex lies in at least two edges (as we saw in Section 2), and the degrees sum to $3k = 2(k+2)+(k-4)$. So if all degrees are 2 or 3, exactly six vertices have degree 2. We build configurations of this kind using the so-called *configuration model* (see e.g. [6]). Besides the nine grid vertices, we let six vertices have degree 2, and the remaining $k-13$ have degree 3; each grid vertex also gets degree 3, i.e. one incidence besides its two grid lines. Represent the missing incidences by *stubs*: one at each grid vertex, two at each degree-2 vertex and three at each other vertex, so

$$M = 9 + 12 + 3(k-13) = 3k - 18$$

in total. Add $k-6$ new edges with three stubs each, again M stubs in total, and match vertex stubs to edge stubs by a uniformly random bijection. Call the resulting multi-hypergraph F.

Every Erdős configuration on $[k+2]$ which contains the fixed grid and has the prescribed degrees arises from exactly

$$(k-6)! \cdot (3!)^{k-6} \cdot (2!)^6 \cdot (3!)^{k-13}$$

bijections: the new edges can be labelled in $(k-6)!$ ways, and the stubs at each new edge and at each vertex can be ordered arbitrarily. Therefore

$$L_k^G \geq \mathbb{P}[F \in \mathcal{E}_k] \cdot \frac{(3k-18)!}{(k-6)!\, 2^6\, 6^{2k-19}},$$

and by Stirling's formula the fraction is $\exp(3k\log k - k\log k + O(k)) = \exp(2k\log k + O(k))$. So the lower bound in Theorem 4.1 follows from the next proposition. (Any bound of the form $\mathbb{P}[F \in \mathcal{E}_k] \geq e^{-O(k)}$ would do, but the natural proof gives a constant.)

Proposition 5.1. *$\mathbb{P}[F \in \mathcal{E}_k] \to e^{-6}$ as $k \to \infty$.*

The multi-hypergraph F is an Erdős k-configuration if and only if

(i) F is linear, and
(ii) no set S of edges with $3 \leq |S| \leq k-1$ spans at most $|S|+2$ vertices.

(Linearity rules out repeated vertices in an edge and repeated edges, and it is the case $|S| = 2$ of the definition; F spans all $k+2$ vertices because each has degree at least 2.) Condition (i) is local, and it fails with probability bounded away from zero: it is the source of the constant e^{-6}. Condition (ii) is global, but it fails only with probability $O(1/k)$.

Linearity. Think of F through its incidence graph: the bipartite multigraph with the vertices of F on one side, the edges of F on the other, and one link for each incidence. F is linear exactly when this graph has no cycle of length 2 (a vertex matched twice to the same edge) and no cycle of length 4 (two edges sharing two vertices).

Short cycles meeting the grid are rare. The grid itself has none. A grid vertex has a single stub, so it lies on no 2-cycle, and a 4-cycle through it must consist of a grid line and a new edge containing two vertices of that line. There are 18 pairs of vertices on a common grid line, and the stubs of a given pair go to the same new edge with probability $O(1/k)$.

The remaining short cycles are formed by stubs alone, and all but fifteen vertices have three stubs. Let X and Y be the numbers of 2-cycles and 4-cycles formed by stub pairs. A 2-cycle is specified by a vertex, a new edge, and two stub pairs between them; a 4-cycle by two vertices, two new edges, and four stub pairs forming the cycle. A given set of t stub pairs is present with probability $1/(M)_t = (1+O(1/k))(3k)^{-t}$ for bounded t, where $(M)_t = M(M-1)\cdots(M-t+1)$. Since almost every vertex has three stubs,

$$\mathbb{E}X = (1+O(1/k))\, k \cdot k \cdot \frac{3\cdot 3\cdot 2}{(3k)^2} \to 2, \qquad \mathbb{E}Y = (1+O(1/k))\binom{k}{2}^2 \frac{(3\cdot 2)^4}{(3k)^4} \to 4.$$

The method of moments upgrades this to a Poisson limit. For fixed i and j, the factorial moment $\mathbb{E}[(X)_i (Y)_j]$ counts ordered tuples of i distinct 2-cycles and j distinct 4-cycles that are all present. Tuples of pairwise disjoint cycles contribute $2^i 4^j + o(1)$, by the computation above. In every other tuple, the union of the cycles has a component containing two distinct cycles, and so has more links than nodes. Since each node can be chosen in $O(k)$ ways and each link costs a factor $O(1/k)$ in probability, these tuples contribute $O(1/k)$. Hence (X, Y) converges in distribution to a pair of independent Poisson random variables with means 2 and 4, and

$$\mathbb{P}[F \text{ is linear}] = \mathbb{P}[X = Y = 0] + O(1/k) \to e^{-2}e^{-4} = e^{-6}.$$

No dense edge sets. Call a set S of edges of F *bad* if $3 \leq |S| \leq k-1$ and S spans at most $|S|+2$ vertices. We show that the expected number of bad sets is $O(1/k)$, handling sets with at most $k/2$ edges directly and larger ones through their complements. In both cases a bad set forces all stubs of some r new edges into a prescribed set of roughly $3r$ stubs, at a cost of order k^{-3r}, while the sets involved can be chosen in about k^{2r} ways. The remaining bounded powers of k are controlled by the condition $|S| \geq 3$ and by the fact that the grid is itself sparse.

Small sets. Suppose S is bad and $s = |S| \leq k/2$. Let A be the set of grid lines in S, say p rows and q columns, $h = p + q$ in total, and let the other $r = s - h$ edges of S be new. Choose a set Y of exactly $s+2$ vertices containing every vertex of S. Then Y contains the $3h - pq$ vertices covered by A, and all $3r$ stubs of the new edges in S are matched to stubs at vertices of Y.

The grid is sparse in the same sense as an Erdős configuration: one grid line covers 3 vertices, and $h \geq 2$ grid lines cover $3h - pq \geq h + 3$ (one checks the few cases). Consequently $r \geq 1$; more precisely

$r \geq 3$ if $h = 0$ (as $s \geq 3$), $r \geq 2$ if $h = 1$, and $r \geq 1$ if $h \geq 2$. Moreover, the number of vertices of Y not forced by A is $s + 2 - (3h - pq) = r + c$, where $c = 2$ if $h = 0$, $c = 0$ if $h = 1$, and $c \leq -1$ if $h \geq 2$.

Now fix A and r. The new edges of S can be chosen in at most $\binom{k}{r}$ ways and the rest of Y in at most $\binom{k+2}{r+c}$ ways, and the $3r$ stubs land among the at most $3(s+2)$ stubs at Y with probability at most $(3s+6)_{3r}/(M)_{3r}$. For bounded r this gives $O(k^r \cdot k^{r+c} \cdot k^{-3r}) = O(k^{c-r})$, and $c - r \leq -1$ in all three cases.

Large sets. Suppose S is bad and $|S| > k/2$, and let T be its complement, with $m = |T| < k/2$. Since S spans at most $|S| + 2 = k + 2 - m$ vertices, at least m vertices lie in no edge of S, so all their edges are in T. Choose a set Z of exactly m such vertices. Let T contain $h = p + q$ grid lines (p rows and q columns) and $r = m - h$ new edges. A grid vertex can belong to Z only if both grid lines through it are in T, so Z contains at most $pq \leq 2h$ grid vertices. If Z contains j grid vertices and d vertices of degree 2, then the vertices of Z carry $3m - 2j - d$ stubs, all of which must be matched to the $3r$ stubs of the new edges in T. In particular $r \geq 1$.

Fix the grid lines in T, and j, d and r. The new edges of T can be chosen in at most $\binom{k}{r}$ ways and Z in $O(\binom{k}{m-j-d})$ ways, and the stub condition has probability $(3r)_{3m-2j-d}/(M)_{3m-2j-d}$. For bounded r this gives $O(k^{r+(m-j-d)-(3m-2j-d)}) = O(k^{j-2h-r})$, which is $O(k^{-r}) = O(k^{-1})$ because $j \leq 2h$.

Summing over r. The estimates above are for fixed r; it remains to show that large r contribute little. In both cases the bound for a given $r \leq k/2$ is at most $k^{O(1)}\binom{k}{r}^2/\binom{3k}{3r}$, with an absolute implied constant. To see this, write $(y)_x/(M)_x = \binom{y}{x}/\binom{M}{x}$. In both cases $x \leq y$ (otherwise the probability is zero) and x, y differ from $3r$ by at most 24, so $\binom{y}{x} \leq (3k)^{24}$ and $\binom{M}{x} \geq M^{-24}\binom{M}{3r}$; moreover $\binom{M}{3r} \geq 3^{-18}\binom{3k}{3r}$ for $r \leq k/2$ and k large, and the second binomial coefficient in each count is at most $k^{O(1)}\binom{k}{r}$, because its lower index differs from r by a bounded amount and $r \leq k/2$.

Heuristically, choosing r edges and about r vertices is worth $\binom{k}{r}^2$, while forcing $3r$ stubs into a set of about $3r$ stubs costs $1/\binom{3k}{3r}$. With $\eta(x) = -x\log x - (1-x)\log(1-x)$, the standard estimates $\binom{k}{r} \leq e^{k\eta(r/k)}$ and $\binom{3k}{3r} \geq e^{3k\eta(r/k)}/(3k+1)$ make this precise:

$$\binom{k}{r}^2 \Big/ \binom{3k}{3r} \leq (3k+1)\, e^{-k\eta(r/k)} \leq (3k+1)\Big(\frac{r}{k}\Big)^r,$$

using $\eta(x) \geq x\log(1/x)$. So the bound for each $r \leq k/2$ is at most $k^{c_0}(r/k)^r$ for an absolute constant c_0. For $r_0 \leq r \leq \sqrt{k}$ this is at most $k^{c_0 - r/2}$, and for $\sqrt{k} \leq r \leq k/2$ it is at most $k^{c_0}2^{-\sqrt{k}}$. Choosing $r_0 \geq 2c_0 + 4$, the terms with $r \geq r_0$ add up to $O(k^{-2})$, while the finitely many $r < r_0$ contribute $O(1/k)$ by the fixed-r estimates. Since there are boundedly many choices of the grid lines involved and of j and d, the expected number of bad sets is $O(1/k)$.

Proof of Theorem 5.1. By the two estimates, $\mathbb{P}[F \in \mathcal{E}_k] = \mathbb{P}[F \text{ is linear}] - O(1/k) \to e^{-6}$. □

This completes the proof of Theorem 4.1, and with it the proof of Theorem 1.1.

References

[1] M. Delcourt and L. Postle, *Proof of the high girth existence conjecture via refined absorption*, arXiv preprint arXiv:2402.17856 (2024).

[2] S. Glock, D. Kühn, A. Lo, and D. Osthus, *On a conjecture of Erdős on locally sparse Steiner triple systems*, Combinatorica **40** (2020), no. 3, 363–403.

[3] M. Kwan and H. T. Pham, *Nonlinear lower-tail large deviations at criticality*, arXiv preprint arXiv:2610.10233 (2026).

[4] M. Kwan, A. Sah, M. Sawhney, and M. Simkin, *High-girth Steiner triple systems*, Ann. of Math. (2) **200** (2024), no. 3, 1059–1156.

[5] A. Sidorenko, *On positive hypergraphs*, European J. Combin. **106** (2022), Paper No. 103574, 5.

[6] N. C. Wormald, *Models of random regular graphs*, Surveys in combinatorics, 1999 (Canterbury), London Math. Soc. Lecture Note Ser., vol. 267, Cambridge Univ. Press, Cambridge, 1999, pp. 239–298.

Institute of Science and Technology Austria (ISTA).
Email address: matthew.kwan@ist.ac.at

Department of Mathematics, University of Chicago.
Email address: htpham@uchicago.edu