hexagon:2611.00677v2

Regular induced subgraphs in bounded independent-edge arrays

Arthur Freitas Ramos* David Barros Hulak† Ruy J. G. B. de Queiroz‡

19 November 2026

Abstract

Let G be a random graph whose edges are independent and whose edge probabilities lie in $[\xi, 1-\xi]$ for a fixed $0 < \xi < 1/2$. With probability at least $1 - n^{-1}$, G contains a regular induced subgraph on at least $c_\xi (\log n)^2$ vertices. If all expected degrees are equal, each prescribed order k up to a small multiple of $n^{2/3}$ occurs with failure probability at most $\exp(-cn/k^{3/2})$, for k sufficiently large. The same probability estimate holds up to order $n^{2/(d+3)}$ for a Lipschitz kernel in dimension d, and up to order $n^{4/(r+7)}$ for an array admitting a bounded symmetric completion of rank at most r. A linear product array has largest regular induced order $\Theta(\sqrt{n})$ with high probability, attaining the scale for dimension one and rank one. The proof weights each candidate set by the reciprocal of its success probability and controls the dependence due to shared edges.

Keywords. Random graphs, induced subgraphs, regular graphs, inhomogeneous edge probabilities.
2020 Mathematics Subject Classification. 05C80, 05C07, 60F05.

1 Introduction

How large a regular induced subgraph must a random graph contain if the edge probabilities can be chosen separately? For $G(n, 1/2)$, Krivelevich, Sudakov and Wormald give a high-probability upper bound of $2n^{2/3}$ and find regular induced sets at subcritical orders. In particular $R = n^{2/3+o(1)}$ with high probability [6, Theorem 1.1]. For unequal edge probabilities there is a second obstruction to regularity. Even before the graph is sampled, the expected degrees on a candidate vertex set may differ by much more than a typical degree fluctuation. We give lower bounds that depend on how much of this variation can be removed by choosing the vertex set.

Write $G(\mathbf{p})$ for the simple graph on $[n]$ in which distinct unordered pairs are present independently, with probabilities $p_{ij} = p_{ji}$. Throughout, $0 < \xi < 1/2$ is fixed and $\xi \le p_{ij} \le 1-\xi$ for $i \neq j$. Write A_k for the event that some set of exactly k vertices induces a regular graph, and put $R = \max\{k : A_k \text{ occurs}\}$. Thus $\{R \ge k\} = \bigcup_{\ell \ge k} A_\ell$. When a domain U is specified, A_k and R refer to the induced array on U. The common degree may depend on the set. All logarithms are natural.

*Microsoft, USA. Corresponding author: arfreita@microsoft.com.
†Independent researcher.
‡Centro de Informática, Universidade Federal de Pernambuco, Brazil.

Theorem 1.1. *The following bounds hold uniformly in the indicated arrays for all sufficiently large n.*

1. *There are $0 < c'_\xi \le c_\xi$ such that*

$$\Pr\{R < c_\xi(\log n)^2\} \le n^{-1}, \qquad \Pr\{R < c'_\xi(\log n)^2\} \le \exp(-\sqrt{n}\log n/8).$$

2. *In each of the three cases below there are constants $a, c > 0$ and k_0 such that, for every integer $k_0 \le k \le an^\beta$,*

$$\Pr(A_k^c) \le \exp(-cn/k^{3/2}). \tag{1.1}$$

Hypothesis on the array	β
All expected degrees $\sum_{j\neq i} p_{ij}$ are equal	$2/3$
Symmetric L-Lipschitz kernel on $[0,1]^d$	$2/(d+3)$
Bounded symmetric matrix completion of rank at most r	$4/(r+7)$

In the kernel case $p_{ij} = W(x_i, x_j)$ for arbitrary deterministic types, W takes values in $[\xi, 1-\xi]$, and $|W(x,y) - W(x',y)| \le L\|x - x'\|_\infty$. A completion is a matrix $P \in [0,1]^{n\times n}$ with the prescribed off-diagonal entries. The constants and the onset depend only on ξ and, where applicable, the fixed parameters (d, L) or r, with $d, r \ge 1$.

For equal expected degrees, (1.1) gives high probability for $k_0 \le k = o(n^{2/3})$. At $k = \lfloor an^{2/3} \rfloor$ it gives a success probability bounded away from zero uniformly in n. For the other two classes, $n/k^{3/2}$ grows as a positive power of n even at the indicated endpoint. In particular every bounded one-dimensional Lipschitz model has a regular induced subgraph of order at least $c\sqrt{n}$ with high probability. Section 6 gives an example attaining this scale.

Relation to earlier work. Erdős, Fajtlowicz and Staton asked whether the largest regular induced subgraph of every n-vertex graph has order $\omega(\log n)$ [1, 6]. A complementary deterministic approach studies modular witnesses and finite absorption certificates [8]. Our first bound concerns a random graph with a fixed probability floor. For instance it applies after independently toggling the adjacency of each pair in a prescribed graph. The resulting induced subgraph is regular in the perturbed graph. In the other direction, Dyson and McKay use inhomogeneous random graphs to give a square-root upper bound for the extremal problem [4].

Levy studies a one-dimensional Lipschitz kernel with independent uniform types and obtains a conditional $n^{1/2-\varepsilon}$ lower bound under local degree and overlap hypotheses [7]. Our one-dimensional bound holds uniformly for every assignment of the types. Conditioning on the types therefore gives a $c\sqrt{n}$ lower bound with high probability for Levy's model, with c depending on the probability floor and the kernel's Lipschitz constant. Levy also conjectures the sharp family-wide constant $\sqrt{2e}$. The constant in our order bound depends on the model parameters.

Where the scales come from. In a domain of N vertices with normalized expected degrees within $O(k^{-1/2})$ of one another, many k-sets can be tilted to the same integer target degree. Two independent uniform k-sets share k^2/N vertices in expectation. If they share j vertices, averaging the tilt over the shared vertices gives a factor of order $\exp(O(j^2/k))$ in the small-overlap range. Thus $k^3/N^2 = O(1)$ is the natural condition for a bounded second moment. Nearly coincident sets require a separate estimate, which is what permits the endpoint $N \asymp k^{3/2}$. Equal row sums supply many such domains by sampling. Kernel and rank assumptions supply them by grouping vertices. With no structural assumption we instead increase the edge density until the mean-square degree variation is small enough for a logarithmic choice of k.

Formalization. The Palomar Registry entry PALOMAR-2026-11-19-000016 (version 1) records the two logarithmic bounds together and each of the three exact-order bounds separately [9]. Its Lean statement file defines the exact-order and at-least-order events for the finite Bernoulli product law. The proof library also contains counterparts of the Fourier estimates and the two-sided product example. The repository's coverage map refers to an earlier 17-page version of this manuscript and uses that version's numbering.

2 The local cost of prescribing degrees

For independent Bernoulli edges on $[k]$ with probabilities q_{ij}, let D be the degree vector, $w_{ij} = q_{ij}(1 - q_{ij})$, and $u_{ij} = e_i + e_j$. A present edge adds u_{ij} to D, so its covariance matrix is

$$Q = \sum_{i<j} w_{ij} u_{ij} u_{ij}^T.$$

The following two estimates are the analytic inputs. The determinant normalization is essential. Replacing it in both estimates by $k^{-k/2}\exp(O(k))$ would cost an exponential factor even for sets sharing only two vertices.

Lemma 2.1 (Full and residual degree atoms)**.** *Fix a floor $\delta \in (0, 1/2)$ for the q_{ij}. If $b = \mathbb{E}D$ is integral with even total, then, uniformly for sufficiently large k,*

$$\Pr(D = b) \geq c_\delta (2\pi)^{-k/2} (\det Q)^{-1/2} \geq \exp(-\tfrac{1}{2} k \log k - C_\delta k).$$

For the upper bound, assume only the same floor on the edge probabilities. Fix $0 < \rho < 1$ and delete all edges inside a set J of $j \leq \rho k$ vertices and let D' and Q' denote the remaining degree vector and covariance. For every sufficiently large k (depending on δ, ρ) and every integral target b', one has

$$\Pr(D' = b') \leq C_{\delta,\rho} (2\pi)^{-k/2} (\det Q')^{-1/2}.$$

The Fourier argument for these estimates is given in Appendix A. In particular the residual target may be far from its mean. Given a realization of the edges internal to J, obtaining $D = b$ requires a single residual target. We record the resulting comparison.

Corollary 2.2. *Under the centering and parity hypotheses above, for every configuration H on the pairs in J,*

$$\Pr(D = b \mid H) / \Pr(D = b) \leq C_{\delta,\rho} \exp(C_{\delta,\rho} j^2 / k).$$

Proof. With $Q_J = Q - Q' \succeq 0$, the trace of Q_J is $2\sum_{a<b\in J} w_{ab} = O(j^2)$. The residual covariance satisfies $Q' \succeq c_{\delta,\rho} k I$ by (A.1) below. Consequently

$$\log(\det Q / \det Q') \leq \operatorname{tr}((Q')^{-1} Q_J) \leq C_{\delta,\rho} j^2 / k.$$

Apply Lemma 2.1 and take the square root of the determinant ratio. □

2.1 Tilting to a common degree

A tilt of the edge logits by $\theta_i + \theta_j$ has a useful exact property. Its likelihood relative to the original law is constant on each degree vector. The following estimates quantify the price of matching the means to that vector.

Lemma 2.3 (Inverse in the maximum norm)**.** *Let $k \geq 4$, let $0 < \lambda \leq 1/4$, and suppose $\lambda \leq w_{ij} = w_{ji} \leq 1/4$ for $i \neq j$. Let W have entries w_{ij} off the diagonal and zero diagonal, and put*

$$D = \operatorname{diag}\Big(\sum_{j\neq i} w_{ij}\Big)_{i=1}^{k}, \qquad H = D + W.$$

Then

$$\lambda(k-2)I \preceq H \preceq \tfrac{1}{2}(k-1)I, \qquad \|H^{-1}\|_{\infty\to\infty} \le C_\lambda/k.$$

Proof. For every $x \in \mathbb{R}^k$,

$$x^T H x = \sum_{i<j} w_{ij}(x_i + x_j)^2, \qquad \sum_{i<j}(x_i+x_j)^2 = (k-2)\|x\|_2^2 + \left(\sum_i x_i\right)^2.$$

These identities prove the two spectral bounds and, in particular, invertibility.

For the maximum norm, set $P = D^{-1}W$. The matrix P is stochastic. Two distinct rows have common mass at least

$$\sum_a \min(P_{ia}, P_{ja}) \ge \frac{4\lambda(k-2)}{k-1} \ge 2\lambda =: \beta.$$

For a real vector x, write $\operatorname{osc}(x) = \max_i x_i - \min_i x_i$. Subtracting the common mass from two probability rows leaves two nonnegative measures of equal mass at most $1-\beta$. Their averages of x differ by at most $(1-\beta)\operatorname{osc}(x)$. Hence

$$\operatorname{osc}(Px) \le (1-\beta)\operatorname{osc}(x).$$

If $(I+P)x = y$, the inequality $\operatorname{osc}(x) \le \operatorname{osc}(y) + \operatorname{osc}(Px)$ implies $\beta \operatorname{osc}(x) \le 2\|y\|_\infty$. Let $m = \min_i x_i$ and $M = \max_i x_i$. At a coordinate attaining M the equation gives $y_i \ge M + m$; at a coordinate attaining m it gives $y_i \le m + M$. Thus $|M+m| \le \|y\|_\infty$, and

$$\|x\|_\infty = \frac{|M+m| + M - m}{2} \le \left(\frac{1}{2} + \frac{1}{\beta}\right)\|y\|_\infty.$$

For $Hx = z$, take $y = D^{-1}z$ and use $D_{ii} \ge \lambda(k-1)$. This proves the asserted inverse bound. □

Lemma 2.4 (A deterministic vertex tilt)**.** *Fix $\xi \in (0, 1/2)$. There are constants $\varepsilon_\xi > 0$, $a_\xi \in (0,1/2)$, $C_\xi < \infty$, and $k_0(\xi)$ such that the following holds. Let $k \ge k_0(\xi)$, let $q_{ij} \in [\xi, 1-\xi]$, and let $\mathbf{d} = (d_1, \dots, d_k) \in \mathbb{Z}^k$ have even total sum. Put*

$$r_i = \sum_{j\ne i} q_{ij} - d_i, \qquad \|r\|_\infty \le \varepsilon_\xi k.$$

There is a vector $\theta \in \mathbb{R}^k$ for which

$$q_{ij}(\theta) = \frac{q_{ij}e^{\theta_i+\theta_j}}{1 - q_{ij} + q_{ij}e^{\theta_i+\theta_j}}$$

satisfies

$$\sum_{j\ne i} q_{ij}(\theta) = d_i \quad (i \in [k]), \qquad q_{ij}(\theta) \in [a_\xi, 1-a_\xi],$$
$$\|\theta\|_\infty \le C_\xi \|r\|_\infty/k, \qquad \|\theta\|_2 \le C_\xi \|r\|_2/k.$$

If $\mathbb{P}_0$ and $\mathbb{P}_\theta$ are the original and tilted product measures, respectively, their relative entropy obeys

$$K := D(\mathbb{P}_\theta \| \mathbb{P}_0) \le C_\xi \|r\|_2^2/k.$$

On every graph configuration with degree vector $\mathbf{d}$, the likelihood ratio $\mathbb{P}_\theta/\mathbb{P}_0$ is the constant e^K. Consequently

$$\mathbb{P}_0(D = \mathbf{d}) \ge \exp\left\{-\frac{1}{2}k\log k - C_\xi k - C_\xi \frac{\|r\|_2^2}{k}\right\}. \tag{2.1}$$

Proof. Define

$$F_i(\theta) = \sum_{j \neq i} q_{ij}(\theta) - d_i, \qquad \Psi(\theta) = \sum_{i<j} \log(1 - q_{ij} + q_{ij} e^{\theta_i + \theta_j}).$$

The gradient of Ψ is the vector of tilted expected degrees, and $H(\theta) := DF(\theta) = \nabla^2 \Psi(\theta)$ has the form of Lemma 2.3, with weights $q_{ij}(\theta)(1 - q_{ij}(\theta))$. For $\|\theta\|_\infty \leq 1$, one may take

$$a_\xi = \frac{\xi e^{-2}}{1 - \xi + \xi e^{-2}} > 0$$

as a lower bound both for $q_{ij}(\theta)$ and for $1 - q_{ij}(\theta)$. Set $\lambda_\xi = a_\xi(1 - a_\xi)$. Throughout that cube,

$$\lambda_\xi (k-2) I \preceq H(\theta) \preceq \tfrac{1}{2}(k-1) I, \qquad \|H(\theta)^{-1}\|_{\infty \to \infty} \leq A_\xi / k. \tag{2.2}$$

We first construct the tilt. Since $F(0) = r$, Taylor expansion gives

$$F(\theta) = r + H(0)\theta + R(\theta).$$

The first two derivatives of the logistic function are bounded absolutely. On the cube of radius $b \leq 1$, each row has $k - 1$ summands, each depending on a sum of two coordinates. The one-variable mean value formula therefore gives

$$\|R(\theta)\|_\infty \leq C k b^2, \qquad \|R(\theta) - R(\phi)\|_\infty \leq C k b \|\theta - \phi\|_\infty.$$

If $r = 0$ take $\theta = 0$. Otherwise put $b = 2A_\xi \|r\|_\infty / k$ and $T(\theta) = -H(0)^{-1}(r + R(\theta))$. Choose ε_ξ so that $b \leq 1$ and $C A_\xi b \leq 1/2$. On the closed cube of radius b, the preceding estimates give $\|T(\theta)\|_\infty \leq b/2 + C A_\xi b^2 \leq b$ and $\mathrm{Lip}(T) \leq 1/2$. The contraction theorem gives a fixed point in that cube. It solves $F(\theta) = 0$, satisfies the stated maximum-norm bound, and preserves the floor a_ξ.

For the Euclidean bound, integrate the Hessian along the segment from zero to the fixed point. Since $\nabla\Psi(\theta) - \nabla\Psi(0) = -r$,

$$-r \cdot \theta = \int_0^1 \theta^T H(t\theta) \theta \, dt \geq \lambda_\xi (k-2) \|\theta\|_2^2.$$

Cauchy–Schwarz yields $\|\theta\|_2 \leq \|r\|_2 / [\lambda_\xi (k-2)]$.

For any graph configuration A, let $D(A)$ be its degree vector. The exact product-measure likelihood ratio is

$$\frac{\mathbb{P}_\theta(A)}{\mathbb{P}_0(A)} = \exp\{\theta \cdot D(A) - \Psi(\theta)\}.$$

Taking its logarithmic expectation under $\mathbb{P}_\theta$, and using $\nabla\Psi(\theta) = \mathbf{d}$ and $\Psi(0) = 0$, gives

$$\begin{aligned} K &= \theta \cdot \nabla\Psi(\theta) - \Psi(\theta) \\ &= \int_0^1 t \, \theta^T H(t\theta) \theta \, dt \leq \frac{k-1}{4} \|\theta\|_2^2 \leq C_\xi \|r\|_2^2 / k. \end{aligned}$$

On $D(A) = \mathbf{d}$, the logarithmic likelihood ratio equals this same K. Hence $\mathbb{P}_0(D = \mathbf{d}) = e^{-K} \mathbb{P}_\theta(D = \mathbf{d})$. Under $\mathbb{P}_\theta$, the target degrees are exactly the means, they are integral with even total, and every edge probability lies in $[a_\xi, 1 - a_\xi]$. Applying the centered degree estimate, Lemma 2.1, with floor a_ξ, and enlarging $k_0(\xi)$ and C_ξ if needed, proves (2.1). □

2.2 Sampling distinct vertices

Lemma 2.5 (Concentration for sampling without replacement)**.** *Let $a_1, \dots, a_M \in [0,1]$, let $\bar{a} = M^{-1}\sum_i a_i$, and let Y be the sum of a uniformly chosen m-element subset of these numbers, where $1 \le m \le M$. Then*

$$\mathbb{E}e^{t(Y-m\bar{a})} \le e^{mt^2/8} \qquad (t \in \mathbb{R}),$$

and for every $s \ge 0$,

$$\mathbb{P}(|Y - m\bar{a}| \ge s) \le 2e^{-2s^2/m}.$$

Also $\operatorname{Var} Y \le m/4$.

Proof. For nonnegative $z_1, \dots, z_M$, their elementary symmetric polynomial satisfies

$$\frac{e_m(z_1, \dots, z_M)}{\binom{M}{m}} \le \left(\frac{1}{M}\sum_i z_i\right)^m. \tag{2.3}$$

For $m \ge 2$, averaging two coordinates at fixed sum increases their product and leaves all terms involving just one coordinate unchanged in aggregate. Thus this operation cannot decrease e_m. Repeatedly averaging the largest and smallest coordinates decreases the sum of squared deviations by half the squared range. The range tends to zero, giving (2.3) by continuity. For $m = 1$ there is equality.

Take $z_i = e^{ta_i}$. The left side of (2.3) is $\mathbb{E}e^{tY}$. For a single uniform draw X, the log moment generating function of $X - \bar{a}$ has value and derivative zero at zero. Its second derivative is the variance under an exponential tilt, at most $1/4$ since $0 \le X \le 1$. Hence

$$\mathbb{E}e^{t(Y-m\bar{a})} \le (\mathbb{E}e^{t(X-\bar{a})})^m \le e^{mt^2/8}.$$

Exponential Markov at $t = \pm 4s/m$ gives the two tails. Comparing the second derivatives at zero gives $\operatorname{Var} Y \le m/4$, since both sides of the centered moment bound have value one and derivative zero there. □

2.3 A common target degree

Lemma 2.6 (Candidate family from mean-squared drift)**.** *Fix $\xi \in (0, 1/2)$. Let $(q_{xy})_{x \ne y}$ be a symmetric array on $[N]$ with entries in $[\xi, 1-\xi]$, and let $p \in [\xi, 1-\xi]$. Write*

$$a_x = \frac{1}{N-1}\sum_{y \ne x} q_{xy} - p.$$

Suppose that

$$\max_x |a_x| \le \alpha, \qquad \frac{1}{N}\sum_x a_x^2 \le \eta^2.$$

Uniformly along sequences with $k \to \infty$, $\alpha \to 0$, and $k \le N$, there is a legal degree d, with kd even and $|d - p(k-1)| \le 2$, and a family $\mathcal{F} \subseteq \binom{[N]}{k}$ of size at least $\frac{1}{2}\binom{N}{k}$, such that every $S \in \mathcal{F}$ has a vertex tilt θ^S satisfying

$$\mathbb{E}_{\theta^S} \deg_{G[S]}(v) = d \quad (v \in S), \qquad q_{uv}(\theta^S) \in [a_\xi, 1 - a_\xi],$$

$$\|\theta^S\|_\infty \le C_\xi\left(\alpha + \sqrt{\frac{\log k}{k}}\right), \qquad \|\theta^S\|_2^2 \le C_\xi(1 + \eta^2 k),$$

$$p_S := \mathbb{P}(G[S] \text{ is } d\text{-regular}) \ge p_* := \exp\{-\tfrac{1}{2}k\log k - C_\xi(k + \eta^2 k^2)\}.$$

Here $a_\xi > 0$ and C_ξ depend only on ξ. There is no restriction on k^2/N in this lemma.

Proof. Choose d before selecting the vertices. If k is even, choose a nearest integer to $p(k-1)$; if k is odd, choose a nearest even integer. For large k, uniformly for $p \in [\xi, 1-\xi]$, this is a legal degree $0 \le d \le k-1$ with the stated parity and error. Set $b = p(k-1) - d$, so $|b| \le 2$.

Choose a uniform ordered tuple $(X_1, \dots, X_k)$ of distinct labels in $[N]$, and put

$$r_i = \sum_{j \neq i} q_{X_i X_j} - d.$$

Conditional on $X_i = x$, the remaining labels form a uniform $(k-1)$-subset of $[N] \setminus \{x\}$. Consequently

$$\mathbb{E}(r_i \mid X_i = x) = (k-1)a_x + b, \qquad \mathrm{Var}(r_i \mid X_i = x) \le (k-1)/4,$$

where the variance assertion follows from Lemma 2.5. The same lemma, with $t = \sqrt{2k \log k}$, and a union bound give

$$\|r\|_\infty \le \alpha(k-1) + 2 + \sqrt{2k \log k} \tag{2.4}$$

with probability at least $1 - 2k^{-3}$. There is no independence assertion between different row sums. Their squared norm is instead controlled directly by conditional second moments:

$$\begin{aligned} \mathbb{E}\|r\|_2^2 &\le k \left\{ \frac{k-1}{4} + 2(k-1)^2 \frac{1}{N} \sum_x a_x^2 + 8 \right\} \\ &\le C(k^2 + \eta^2 k^3). \end{aligned}$$

Markov's inequality therefore gives

$$\|r\|_2^2 \le 4C(k^2 + \eta^2 k^3) \tag{2.5}$$

with probability at least $3/4$. The two displayed drift bounds hold together with probability at least $1/2$ for all sufficiently large k.

On their intersection $\|r\|_\infty / k = o(1)$ uniformly. The tilt in Lemma 2.4 therefore exists, with the fixed floor a_ξ. Its maximum norm is at most $C_\xi \|r\|_\infty / k$, its squared energy at most $C_\xi \|r\|_2^2 / k^2$, and its negative log probability cost beyond the centered atom at most $C_\xi \|r\|_2^2 / k$. Substitution of (2.4) and (2.5) gives all three asserted estimates. Both drift tests are invariant under reorderings, and every k-set has $k!$ orderings, so at least half of the unordered sets qualify. □

3 Shared edges and a weighted count

Let E_S be the event that a candidate S induces a d-regular graph, where the degree d and the family $\mathcal{F}$ come from Lemma 2.6. Weight a success on S by $1/p_S$, where $p_S = \Pr(E_S)$. Then

$$Z = \sum_{S \in \mathcal{F}} \frac{\mathbf{1}_{E_S}}{p_S}, \qquad \mathbb{E}Z = |\mathcal{F}|.$$

This weighting allows different candidates to have very different probabilities. Write $H_{ST} = \Pr(E_S \cap E_T)/(p_S p_T)$ for the normalized joint probability.

Lemma 3.1 (Shared-edge likelihood)**.** *Let $J = S \cap T$, $j = |J|$, and $X_S(J) = \sum_{v \in J} (\theta_v^S)^2$. For $j \le 3k/4$ the normalized joint probability satisfies*

$$\frac{\Pr(E_S \cap E_T)}{p_S p_T} \le K \exp\{D j^2/k + (j/4)[X_S(J) + X_T(J)]\}, \tag{3.1}$$

where K, D depend only on ξ. For $j = 0, 1$ the ratio is one.

Proof. Write H for the shared-edge configuration and $L_S(H)$ for its tilted likelihood relative to its original law. Constancy of the full likelihood on E_S gives

$$\frac{\Pr(E_S \mid H)}{p_S} = L_S(H)\frac{\Pr_{\theta^S}(E_S \mid H)}{\Pr_{\theta^S}(E_S)}.$$

The last ratio is bounded by Corollary 2.2. Conditional on H, the two events depend on disjoint sets of remaining edges. Their normalized joint probability is therefore at most $Ke^{Dj^2/k}\mathbb{E}[L_S(H)L_T(H)]$. For one shared edge with logit increments s, t, put $\psi(u) = \log(1 - q + qe^u)$. Its contribution to the logarithm of the expectation is

$$\psi(s+t) - \psi(s) - \psi(t) = \int_0^s \int_0^t \psi''(u+v)\, dv\, du \le (s^2+t^2)/8$$

since $0 \le \psi'' \le 1/4$. Sum over edges and use $\sum_{a<b\in J}(\theta_a+\theta_b)^2 \le 2j\sum_{a\in J}\theta_a^2$. For $j \le 1$ the two sets share no edge. □

For fixed S and a uniform k-set T in $[N]$, write $P_j = \Pr(|S \cap T| = j)$. A union bound over the j-subsets of S yields

$$P_j \le \binom{k}{j}\frac{(k)_j}{(N)_j} \le \min\{(k^2/N)^j,\, (ek^2/(jN))^j\}, \qquad j \ge 1. \tag{3.2}$$

Here $(k)_j = k(k-1)\cdots(k-j+1)$. The next two results use different parts of this estimate. The first tolerates appreciable mean-square drift when typical overlaps are small. The second reaches the $k^{3/2}$ block scale when every row is nearly balanced.

Theorem 3.2 (Mean-square drift)**.** *Let q be an independent-edge array on N vertices with $q_{ij} \in [\xi, 1-\xi]$. Write $a_i = (N-1)^{-1}\sum_{j\ne i} q_{ij} - p$ for some $p \in [\xi, 1-\xi]$. Suppose $\max_i |a_i| \le \alpha$ and $N^{-1}\sum_i a_i^2 \le \eta^2$. There exists $c_\xi^* > 0$ such that, uniformly when $k \to \infty$, $\alpha \to 0$, $k^2/N \to 0$ and $\eta^2 k \le c_\xi^* \log(N/k^2)$,*

$$\Pr(A_k^c) \le C_\xi(k^2/N)^{3/2} + C_\xi e^{-k\log k/8}.$$

When $\eta = 0$ the first term can be replaced by $C_\xi k^4/N^2$.

Proof. Write $H_{ST} = \Pr(E_S \cap E_T)/(p_S p_T)$. For $j = 0, 1$, $H_{ST} - 1 = 0$. If U_j bounds H_{ST} for all candidate pairs of overlap $j \ge 2$, then

$$\frac{\operatorname{Var} Z}{(\mathbb{E}Z)^2} = \frac{1}{|\mathcal{F}|^2}\sum_{S,T\in\mathcal{F}}(H_{ST}-1) \le \frac{\binom{N}{k}^2}{|\mathcal{F}|^2}\sum_{j=2}^{k} P_j U_j \le 4\sum_{j=2}^{k} P_j U_j.$$

Indeed, exactly $\binom{N}{k}^2 P_j$ ordered pairs of unrestricted k-sets have intersection size j. Write $H_0 = \log(N/k^2)$. For $2 \le j \le 3k/4$, (3.1) and the total tilt energy give a ratio at most $K\exp(D_0 j + D_1\eta^2 kj)$. By (3.2) their contribution to $\operatorname{Var} Z/(\mathbb{E}Z)^2$ is at most $4K\sum_{j\ge2} u^j$, where $u = \exp(-H_0 + D_0 + D_1\eta^2 k)$. Choose $D_1 c_\xi^* \le 1/4$. This sum is $O(e^{-3H_0/2})$, or $O(e^{-2H_0})$ when $\eta = 0$.

For $j > 3k/4$, use $\Pr(E_S \cap E_T)/(p_S p_T) \le 1/\max(p_S, p_T) \le e^B$ with $B = \frac{1}{2}k\log k + C_0(k + \eta^2 k^2)$. Now (3.2) gives $P_j \le g^j$ with $g = 4ek/(3N)$. Since $\log(N/k) = \log k + H_0$,

$$B + (3k/4)\log g \le k\{-\tfrac{1}{4}\log k - \tfrac{3}{4}H_0 + C_0\eta^2 k + O_\xi(1)\} \le -\tfrac{1}{8}k\log k$$

after taking $C_0 c_\xi^* \le 1/4$. Sum the geometric tail, including $j = k$, and apply $\Pr(Z = 0) \le \operatorname{Var} Z/(\mathbb{E}Z)^2$. □

Theorem 3.3 (A balanced block)**.** *Let the independent edges on $[N]$ have probabilities in $[\xi, 1-\xi]$ and define $a_i = (N-1)^{-1}\sum_{j\neq i} q_{ij} - p$ for $p \in [\xi, 1-\xi]$. Fix $M \geq 0$. There are $L_0, p_0 > 0$ depending only on M, ξ such that, for all sufficiently large k, if $N \geq L_0 k^{3/2}$ and $\max_i |a_i| \leq \sqrt{M/k}$, then $\Pr(A_k) \geq p_0$. The successful candidates may all be tested at the same degree d.*

Proof. Apply Lemma 2.6 with $\alpha = \eta = \sqrt{M/k}$. It gives one d and a family occupying at least half of all k-sets. For constants $A \geq 1$ and C depending only on M, ξ, every candidate satisfies

$$\|\theta^S\|_2^2 \leq A, \qquad \|\theta^S\|_\infty^2 \leq A\log k/k, \qquad p_S^{-1} \leq \exp(\tfrac{1}{2}k\log k + Ck).$$

We bound $\mathbb{E}Z^2/(\mathbb{E}Z)^2$ in three overlap ranges.

Small overlaps. For $j \leq t = \lfloor k/(A\log k)\rfloor$, symmetry in (S,T) and the arithmetic–geometric mean inequality bound the sum of $\exp((j/4)(X_S + X_T))$ over candidate pairs of overlap j by the corresponding sum of $\exp(jX_S/2)$. More explicitly, $2e^{j(X_S+X_T)/4} \leq e^{jX_S/2} + e^{jX_T/2}$ and the two sums on the right agree after swapping S, T. Keep $S \in \mathcal{F}$ fixed and extend the other sum to all T. Each j-subset $J \subseteq S$ now occurs $\binom{N-k}{k-j}$ times. Thus, for any nonnegative g_S and any j with $P_j > 0$,

$$\frac{1}{|\mathcal{F}|^2}\sum_{S\in\mathcal{F}}\sum_{\substack{|T|=k\\|S\cap T|=j}} g_S(S\cap T) = \frac{\binom{N}{k}}{|\mathcal{F}|}P_j\,\frac{1}{|\mathcal{F}|}\sum_{S\in\mathcal{F}}\mathbb{E}_J g_S(J),$$

where J on the right is uniform among the j-subsets of S. The factor in front of P_j is at most two. Assign a zero contribution when $P_j = 0$. The elementary symmetric mean bound (2.3) and $e^u \leq 1 + 2u$ on $[0,1]$ give

$$\mathbb{E}_J e^{jX_S(J)/2} \leq \left(k^{-1}\sum_{v\in S} e^{j(\theta_v^S)^2/2}\right)^j \leq (1 + Aj/k)^j \leq e^{Aj^2/k}.$$

The normalized contribution is at most $2K\sum_{j\leq t} P_j e^{(D+A)j^2/k}$. Put $\mu = k^2/N$ and $s = \lceil\max(e^2\mu, \log k)\rceil$. The terms through $\min(s,t)$ sum to at most $\exp((D+A)s^2/k)\sum_j P_j = O(1)$, since $N \geq L_0 k^{3/2}$ gives $s = O(\sqrt{k} + \log k)$. For $s < j \leq t$, (3.2) gives $P_j \leq e^{-j}$ and $(D+A)j^2/k \leq j/2$ eventually.

Intermediate overlaps. For $t < j \leq 3k/4$, use $X_S, X_T \leq A$ in (3.1). The resulting sum is at most $4K\sum_{j>t}(C_1\mu/j)^j = o(1)$ since $\mu/t = O((\log k)/\sqrt{k})$.

Large overlaps. For $j > 3k/4$ put $m = k - j$. Condition on all edges of S. Each vertex of J needs a specified number of edges to $T \setminus S$ to reach d in T. Those j sums involve disjoint sets of fresh edges, so they are conditionally independent. For a sum Y of m independent Bernoulli variables of floor δ, one-dimensional inversion gives

$$\sup_{\ell\in\mathbb{Z}}\Pr(Y = \ell) \leq \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{-2m\delta(1-\delta)\sin^2(t/2)}\,dt \leq \frac{C_\delta}{\sqrt{m+1}}. \tag{3.3}$$

Use $\sin(|t|/2) \geq |t|/\pi$ and take $C_\delta \geq 1$ for $m = 0$. If $\mathcal{G}_S$ denotes the exposed edges, E_S is $\mathcal{G}_S$-measurable and $\Pr(E_T \mid \mathcal{G}_S) \leq (C_\xi/\sqrt{m+1})^j$. Multiplying by $\mathbf{1}_{E_S}$ and taking expectations gives $H_{ST} \leq p_T^{-1}(C_\xi/\sqrt{m+1})^j$, including $S = T$. Multiply by $P_j \leq (4ek/(3N))^j$. The logarithm of the product is at most

$$\tfrac{1}{2}\{m\log k - j\log(m+1)\} + C_2 k - j\log L_0.$$

For $0 \leq m \leq k/4$, concavity gives $\log(m+1) \geq (4m/k)\log(1 + k/4)$. As $4j/k \geq 1$, the braces are at most $m\log[k/(1+k/4)] \leq k\log 4$. The constant C_2 depends only on M, ξ. Choose

$\frac{3}{4}\log L_0 \ge C_2 + 1 + \frac{1}{2}\log 4$. After accounting for the candidate fraction, the entire range contributes at most $4(k+1)e^{-k}$. This last estimate retains the information in the m fresh vertices even when the two candidates nearly coincide.

The normalized second moment is bounded by a constant independent of k, N. Cauchy–Schwarz gives $\Pr(Z > 0) \ge (\mathbb{E}Z)^2/\mathbb{E}Z^2$. □

4 Extraction with small mean-square degree variation

For a symmetric array $q_{ij} \in [0,1]$, $i \ne j \in V$, of order $M \ge 2$, write

$$e_q(S) = \sum_{\{i,j\}\subseteq S} q_{ij}, \qquad d_i = \sum_{j\ne i} q_{ij}, \qquad p = \frac{e_q(V)}{\binom{M}{2}}, \qquad \bar{d} = p(M-1).$$

Its normalized degree variance is

$$v = \frac{1}{M^3}\sum_{i\in V}(d_i - \bar{d})^2.$$

The next argument uses a nonuniform random vertex sample only to establish the existence of a deterministic induced subarray.

Lemma 4.1 (A variance-driven density increment)**.** *Suppose $M \ge 32$. If $v > 16e^{-M/8}$, there is a set $S \subseteq V$ of order at least $M/4$ whose induced weighted density is greater than $p + v/2$.*

Proof. Put $a_i = (d_i - \bar{d})/M$. Then

$$\sum_i a_i = 0, \qquad |a_i| \le 1, \qquad \sum_i a_i^2 = Mv, \qquad \sum_i a_i d_i = M^2 v.$$

Select vertices independently with probabilities $w_i = 1/2 + a_i/4 \in [1/4, 3/4]$, and call the resulting set S. Its expected order is $M/2$. Expansion of the edge expectation gives

$$\mathbb{E}e_q(S) = \frac{pM(M-1)}{8} + \frac{1}{8}\sum_i a_i d_i + \frac{1}{16}\sum_{i<j} q_{ij}a_i a_j.$$

Since

$$\left|\sum_{i<j} q_{ij}a_i a_j\right| \le \frac{1}{2}\left(\sum_i |a_i|\right)^2 \le \frac{M}{2}\sum_i a_i^2 = \frac{1}{2}M^2 v,$$

we obtain

$$\mathbb{E}e_q(S) \ge \frac{pM(M-1)}{8} + \frac{3M^2 v}{32}. \tag{4.1}$$

Also $\sum_i w_i^2 = M/4 + Mv/16$, so independence gives

$$\mathbb{E}\binom{|S|}{2} = \frac{M(M-1)}{8} - \frac{Mv}{32}. \tag{4.2}$$

Subtracting $(p + v/2)$ times (4.2) from (4.1) yields

$$\mathbb{E}\left[e_q(S) - (p + v/2)\binom{|S|}{2}\right] \ge \frac{M^2 v}{32}. \tag{4.3}$$

Here all discarded terms are nonnegative, since $p, v \ge 0$.

The bounded-summand exponential estimate, applied to the independent vertex indicators of total mean $M/2$, gives $\Pr(|S| < M/4) \le e^{-M/8}$. If every set of order at least $M/4$ had weighted density at most $p + v/2$, the random surplus in (4.3) would be nonpositive off that exceptional event. On the event it is at most $e_q(S) \le M^2/2$. Its expectation would therefore be at most $M^2 e^{-M/8}/2$, contradicting $v > 16e^{-M/8}$. The resulting set has at least two vertices since $M \ge 32$. □

Lemma 4.2 (Mean-square extraction)**.** *Let $0 < \eta \le 1/16$, and put*

$$J = \left\lceil \frac{2}{\eta^2} \right\rceil, \qquad M_0 = \max\left\{32,\ 8\log\frac{16}{\eta^2}\right\}.$$

If $n \ge 4^J M_0$, every symmetric array in $[0,1]$ on n vertices has an induced subarray of order

$$N \ge \tfrac{1}{2}\, 4^{-J} n$$

whose degree sums d'_i and average $\bar{d}'$ satisfy

$$\max_i |d'_i - \bar{d}'| \le 3\sqrt{\eta}\,(N-1), \qquad \frac{1}{N}\sum_i (d'_i - \bar{d}')^2 \le 6\eta^2 (N-1)^2. \tag{4.4}$$

In particular the logarithm of the vertex loss is $O(\eta^{-2})$. If the initial entries lie in $[\xi, 1-\xi]$, then so do the extracted entries and their common density $\bar{d}'/(N-1)$.

Proof. Repeatedly restrict the current array to a set supplied by Lemma 4.1 whenever its normalized degree variance exceeds η^2. Through the first J restrictions, whenever defined, the current order after $t \le J$ steps is at least $4^{-t} n \ge M_0$, so $16e^{-M/8} \le \eta^2$ and the increment lemma applies. Each restriction retains at least one quarter of the vertices and increases the density by more than $\eta^2/2$. There are fewer than J restrictions, since J such increases would raise a nonnegative initial density above one. The terminal subarray therefore has order $M \ge 4^{-J} n$ and

$$\sum_i (d_i - \bar{d})^2 \le \eta^2 M^3.$$

Remove the set B of vertices satisfying $|d_i - \bar{d}| > \sqrt{\eta}\, M$. Its size is at most ηM. Let $U = V \setminus B$, $N = |U|$, and let d'_i denote the degree sums of the remaining subarray. Then $N \ge (1-\eta)M \ge 15M/16$. For $i \in U$, write

$$d'_i - \bar{d} = (d_i - \bar{d}) - s_i, \qquad s_i = \sum_{b \in B} q_{ib} \in [0, \eta M].$$

Consequently

$$\max_{i \in U} |d'_i - \bar{d}| \le (\sqrt{\eta} + \eta) M, \qquad \sum_{i \in U} (d'_i - \bar{d})^2 \le 2\eta^2 M^3 + 2N\eta^2 M^2 \le 4\eta^2 M^3.$$

Centering at the new average $\bar{d}'$ cannot increase the sum of squares, and it increases the preceding maximum bound by at most a factor of two. Since $M \ge 32$,

$$N - 1 \ge \frac{15M}{16} - 1 \ge \frac{7M}{8}.$$

Using $\eta \le \sqrt{\eta}/4$, we find

$$\max_i |d'_i - \bar{d}'| \le \frac{5}{2}\sqrt{\eta} M \le \frac{20}{7}\sqrt{\eta}(N-1) < 3\sqrt{\eta}(N-1).$$

Similarly,

$$\frac{\sum_i (d'_i - \bar{d}')^2}{N(N-1)^2} \le 4\eta^2 \frac{M}{N} \left(\frac{M}{N-1}\right)^2 \le 4\eta^2 \frac{16}{15} \left(\frac{8}{7}\right)^2 < 6\eta^2.$$

These are (4.4). The size bound follows from $N \ge 15M/16 \ge M/2$. The entry and density assertions follow by restriction and averaging. □

Theorem 4.3 (A universal squared-logarithmic lower bound)**.** *For every fixed $\xi \in (0, 1/2)$ there are constants $c_\xi > 0$ and $n_0(\xi)$ such that, for all $n \ge n_0(\xi)$,*

$$\sup_{\mathbf{p} \in [\xi, 1-\xi]^{\binom{n}{2}}} \Pr\{R(G(\mathbf{p})) < c_\xi (\log n)^2\} \le n^{-1}.$$

Proof. Put $L = \log n$ and choose $\eta^2 = 32/L$. For sufficiently large n, $\eta \le 1/16$ and

$$J \log 4 \le \left(\frac{L}{16} + 1\right) \log 4 \le L/4, \qquad \log M_0 = O(\log \log L).$$

Thus the onset of Lemma 4.2 holds. Applied directly to the given probability array, it supplies a deterministic induced domain of order $N \ge n^{3/4}/2$. On this domain the normalized maximum degree-sum drift is at most $3\sqrt{\eta} = o(1)$, and the normalized root-mean-square drift is at most $\zeta = \sqrt{6}\eta$, so $\zeta^2 = 192/L$.

Let c^*_ξ be the tolerance constant in Theorem 3.2, choose $a > 0$ with $192a < c^*_\xi/2$, and take

$$k = 2 \left\lfloor \frac{a}{2} L^2 \right\rfloor.$$

Then $k \to \infty$, $k^2/N \to 0$, and uniformly for $N \in [n^{3/4}/2, n]$,

$$\log(N/k^2) \ge L/2, \qquad \zeta^2 k \le 192aL < c^*_\xi \log(N/k^2).$$

Theorem 3.2 therefore gives a regular induced k-set with failure probability at most

$$C_\xi (k^2/N)^{3/2} + C_\xi e^{-k \log k/8} = O_\xi(L^6 n^{-9/8}) + e^{-\Omega_\xi(L^2 \log L)} = o(n^{-1}).$$

After enlarging the onset, this is at most n^{-1}. Since $k \ge (a/2)L^2$ eventually, take $c_\xi = a/2$. All selections and estimates used only the deterministic probability array, and every constant is independent of its individual entries. □

Corollary 4.4 (Amplified probability and almost-sure form)**.** *Let c_ξ be a constant for which Theorem 4.3 holds. For every fixed $0 < c'_\xi < c_\xi/4$ and all sufficiently large n,*

$$\sup_{\mathbf{p} \in [\xi, 1-\xi]^{\binom{n}{2}}} \Pr\{R(G(\mathbf{p})) < c'_\xi (\log n)^2\} \le \exp\{-\tfrac{1}{8}\sqrt{n} \log n\}.$$

Consequently, if (G_n) is any sequence of random graphs on $[n]$ with independent edges within each graph and all its probabilities in $[\xi, 1-\xi]$, then almost surely

$$R(G_n) \ge c'_\xi (\log n)^2 \qquad \textit{for all sufficiently large } n.$$

No independence between different members of the sequence is needed.

Proof. Set $m = \lfloor \sqrt{n} \rfloor$ and $r = \lfloor n/m \rfloor$. Choose r disjoint blocks of m vertices. Each block fails to contain the universal witness with probability at most m^{-1}. The induced graphs on the blocks use disjoint sets of independent edge variables, so the probability that all blocks fail is at most

$$m^{-r} = e^{-r \log m} \le e^{-\sqrt{n} \log n/8}$$

for large n: use $r \geq \sqrt{n}/2$ and $\log m \geq (\log n)/4$. Since $(\log m)^2/(\log n)^2 \to 1/4$, any successful block supplies the asserted order for every fixed $c'_\xi < c_\xi/4$.

The error probabilities are summable in n. The probability of any failure at an index $n \geq N$ is at most the tail of their convergent sum, which tends to zero as $N \to \infty$. Thus almost surely only finitely many failures occur, without a condition on the dependence across indices. □

5 Amplification and structured arrays

The block size in Theorem 3.3 is $m = \lceil L_0 k^{3/2} \rceil$. To obtain (1.1) we need a positive success probability in order n/m disjoint blocks of this size. The blocks must be chosen before revealing any edges. Equal expected degrees allow a random partition argument, while the kernel and rank assumptions allow a deterministic choice.

5.1 Equal expected degrees

Let p be the common normalized expected degree. Use Theorem 3.3 with $M = 1$, and choose a small enough that $n \geq 2m$ when $k \leq an^{2/3}$. Take a uniform permutation and group its first $m\lfloor n/m \rfloor$ positions into blocks of size m. Conditional on a vertex occupying a specified position in a given block, the other $m-1$ vertices there are sampled without replacement from the other $n-1$ vertices. The probability that its block row mean differs from p by more than $1/\sqrt{k}$ is at most $2\exp[-2(m-1)/k]$. The probability that a specified block contains any such row is at most $2m\exp[-2(m-1)/k] < 1/2$ for large k. Taking expectations, some partition has at least half its blocks good. This argument only requires a bound on the fraction of bad blocks, so its onset is independent of n.

Fix that partition. Every good block meets the hypotheses of Theorem 3.3. Since their induced graphs use disjoint edge coordinates, the probability that they all fail is at most $(1-p_0)^{\lfloor n/m \rfloor/2} \leq \exp(-cn/k^{3/2})$. This proves the equal-degree assertion of Theorem 1.1.

5.2 Packing blocks from large cells

Lemma 5.1. *Suppose there are fixed $s \geq 0$, $b > 0$ and $B \geq 0$ with the following property. For each induced domain of order u and each $0 < h \leq 1$ with $buh^s \geq 2$, the domain contains a set of at least buh^s vertices on which the off-diagonal probabilities lie within Bh of a common $p \in [\xi, 1-\xi]$. Then (1.1) holds up to $k \leq an^{2/(s+3)}$, uniformly with constants depending on s, b, B, ξ.*

Proof. Take $h = k^{-1/2}$ and $m = \lceil L_0 k^{3/2} \rceil$, with L_0 from Theorem 3.3 for $M = B^2$. For a small enough, $b(n/2)k^{-s/2} \geq m$ throughout the stated range. While at least $n/2$ vertices remain, find such a cell, take exactly m of its vertices as a block, and remove just that block. Every restriction of the cell still has all pair probabilities within $B/\sqrt{k}$ of p, so all block row means satisfy the same bound. This produces at least $n/(2m)$ disjoint blocks, independently of the graph realization. Apply Theorem 3.3 in each and multiply their failure probabilities. □

Lipschitz types. Partition the type cube into at most $(2/h)^d$ cubes of maximum-norm diameter h. In every domain of u vertices some cube contains at least $u(h/2)^d$ types. Symmetry of W gives the Lipschitz estimate in either argument, so the range of the edge probabilities there is at most $2Lh$. Taking any off-diagonal entry as p verifies the hypothesis of Lemma 5.1 with $s = d$. No assumption on the empirical distribution of the types is required.

Rank and diagonal subdivision. Let P be a bounded symmetric completion of rank at most r. A direct covering of its row coordinates at resolution h would give $s = r$. Symmetry permits the coarser resolution $\sqrt{h}$ in those coordinates, provided one also controls the diagonal. This improves the cell exponent to $(r + 1)/2$.

Lemma 5.2. *Every principal submatrix of order u contains a set of at least $c_r u h^{(r+1)/2}$ vertices on which the range of all entries is at most $(1 + r^2)h$, for $0 < h \le 1$.*

Proof. Write $t \le r$ for the rank of the submatrix, still called P. The case $t = 0$ is immediate. Choose row indices I and column indices J for a t-minor of maximum absolute determinant. Each row is a linear combination of the rows indexed by I. Cramer's rule on the columns J shows that its coefficients $x_i \in \mathbb{R}^t$ lie in $[-1, 1]^t$. By symmetry, with $B = P_{I,I}$,

$$P_{ij} = x_i^T B x_j, \qquad |B_{\ell m}| \le 1.$$

Set $w = \sqrt{h}$ and retain a largest coordinate cube of side w. It has at least $u(w/3)^r$ vertices. On it the diagonal values $x_i^T B x_i$ have range at most $2r^2 w$, by expanding their pairwise differences. Subdivision into intervals of length h now costs at most $(2r^2 + 1)/w$ further classes. Keep a largest class. For any two of its vertices,

$$2P_{ij} = P_{ii} + P_{jj} - (x_i - x_j)^T B (x_i - x_j).$$

The last term has absolute value at most $r^2 w^2 = r^2 h$. The first two diagonal terms each range over an interval of length h, so the pair entries have range at most $(1 + r^2)h$. The retained fraction is at least $w^{r+1}/[3^r(2r^2 + 1)]$. □

On a cell with at least two vertices, choose an off-diagonal entry as the common p. Lemma 5.1 with $s = (r + 1)/2$ gives $\beta = 2/(s + 3) = 4/(r + 7)$, completing the proof of Theorem 1.1.

Conditional environments. The upper bounds on failure are uniform in the array. They therefore remain valid after averaging over a random environment, whenever the hypotheses hold almost surely with the same parameters and the conditional edge law is the indicated product law. For independent toggling of an arbitrary base graph with probability ξ, the two possible edge probabilities are ξ and $1 - \xi$. The unrestricted logarithmic bounds always apply. If the base is regular, the perturbed expected degrees $\xi(n - 1) + (1 - 2\xi)d_{\mathrm{base}}(i)$ are equal, so the balanced bound applies as well.

6 A sharp square-root example

For $a, b > 0$ with $a + b < 1$, take

$$z_i = a + b\, i/n, \qquad p_{ij} = z_i z_j, \quad 1 \le i < j \le n.$$

Alon, Krivelevich and Sudakov used the choice $a = 1/4$, $b = 1/2$ to obtain an upper bound of order $\sqrt{n}(\log n)^{3/4}$ [1, Section 3]. We determine the order for the same product model over the full fixed parameter range. The array has rank-one completion zz^T and is generated by the Lipschitz kernel $W(x, y) = (a + bx)(a + by)$ on the unit interval. Given ξ, choosing $a = \sqrt{\xi}$ and $a + b = \sqrt{1 - \xi}$ realizes that floor. The lower bounds for rank one and dimension one coincide. An upper bound for the very same array shows that neither class admits a larger uniform order of growth.

Theorem 6.1. *There are constants $c, C > 0$, depending only on a, b, such that*

$$\Pr\{c\sqrt{n} \le R \le C\sqrt{n}\} \longrightarrow 1.$$

Take $\xi < \min\{a^2, 1-(a+b)^2, 1/2\}$ and $k = \lfloor \kappa\sqrt{n} \rfloor$ for a sufficiently small fixed $\kappa > 0$. Then (1.1) bounds failure by $\exp(-c_1 n^{1/4})$, and $k \geq (\kappa/2)\sqrt{n}$ eventually. For the upper bound, two effects must be retained at once. A set with little variation in its vertex weights still pays the local probability cost of making all its degrees equal. A set with widely dispersed weights pays an additional cost for the variation in its expected degrees.

Proposition 6.2. *For a fixed k-set S, $k \geq 3$, write $\mu_i = \sum_{j \in S\setminus\{i\}} p_{ij}$, $\bar{\mu} = k^{-1}\sum_i \mu_i$, and $V(S) = \sum_i (\mu_i - \bar{\mu})^2$. Under the fixed floor,*

$$\Pr(G[S] \text{ is regular}) \leq (C_\xi/\sqrt{k})^k \exp[-2V(S)/(k-2)]. \tag{6.1}$$

With arbitrary probabilities in $[0,1]$ the bound $\exp[-2V(S)/(k-2)]$ alone still holds.

Proof. Put $v_i = \mu_i - \bar{\mu}$ and tilt by $\theta_i = -4v_i/(k-2)$. On every regular outcome $\theta \cdot D = 0$. Writing $\Psi(\theta) = \sum_{i<j} \log(1 - p_{ij} + p_{ij}e^{\theta_i+\theta_j})$, the likelihood identity therefore gives $\Pr_0(\text{regular}) = e^{\Psi(\theta)} \Pr_\theta(\text{regular})$. The curvature of every one-edge log partition function is at most $1/4$. As $\sum_i v_i = 0$ and $\sum_{i<j}(v_i+v_j)^2 = (k-2)V(S)$,

$$\Psi(\theta) \leq \sum_i \theta_i\mu_i + \tfrac{1}{8}\sum_{i<j}(\theta_i+\theta_j)^2 = -2V(S)/(k-2).$$

This proves the floor-free estimate, including deterministic edge coordinates. Under the fixed floor the tilt preserves a positive floor depending only on ξ, since $|\theta_i| \leq 4(k-1)/(k-2) \leq 8$. Under the tilted law, fix a degree d and expose the edges in rows $i = 1, \ldots, k$, revealing ij with $j > i$ at step i. Given the previous rows, the required number of successes in the $k-i$ fresh trials is fixed. By (3.3) its conditional probability is at most $C/\sqrt{k-i+1}$, uniformly in the history. Earlier degree constraints are measurable at this point. Iterating conditional expectations gives at most $C^k/\sqrt{k!}$ for the prescribed degree vector, as in the exposure argument of [1, Section 3]. Summing over the k possible degrees, using $k! \geq (k/e)^k$ and $k \leq 2^k$, gives $(C_\xi/\sqrt{k})^k$ for tilted regularity. Multiplying by the exponential penalty proves (6.1). □

Upper bound in Theorem 6.1. For the product array $\mu_i = z_i(\sum_{\ell \in S} z_\ell - z_i)$. Therefore

$$|\mu_i - \mu_j| = |z_i - z_j| \sum_{\ell \in S\setminus\{i,j\}} z_\ell \geq (ab/n)(k-2)|i-j|.$$

Use $V(S) = (2k)^{-1}\sum_{i,j\in S}(\mu_i - \mu_j)^2$ and choose $v \in S$ minimizing $\sum_{i\in S}(i-v)^2$. The minimum is at most its average over v. It follows that

$$\frac{2V(S)}{k-2} \geq \frac{(ab)^2(k-2)}{n^2}\sum_{i\in S}(i-v)^2 \geq \frac{\lambda k}{n^2}\sum_{i\in S}(i-v)^2, \qquad \lambda = (ab)^2/3.$$

For a fixed anchor v, set $w_i = \exp[-\lambda k(i-v)^2/n^2]$. Comparing the decreasing Gaussian on either side of v with its integral gives $\sum_i w_i \leq C_{a,b} n/\sqrt{k}$. The sum of $\prod_{i\in S} w_i$ over the k-sets containing v is at most $(\sum_i w_i)^{k-1}/(k-1)!$, since $w_v = 1$. Sum (6.1) over the n choices of anchor, using the bound just obtained for the sets assigned to each. With $k! \geq (k/e)^k$ this yields

$$\Pr(A_k) \leq k^{3/2}(D_{a,b}n/k^2)^k.$$

Choose C so that $D_{a,b}/C^2 < 1/2$. Summing this bound over *all* $k \geq C\sqrt{n}$ gives $o(1)$ and hence the asserted upper tail for R. □

The unrestricted problem. Does every fixed floor ξ admit a $\beta_\xi > 0$ for which $\inf_{\mathbf{p}} \Pr(R \geq n^{\beta_\xi}) \to 1$? The product example forces $\beta_\xi \leq 1/2$ at the exponent level. The density increment in Section 4 yields a logarithmic power for general arrays. Proposition 6.2 indicates what obstructs a larger witness: for instance $\Pr(A_k) \geq \gamma$ implies that some k-set satisfies $V(S) \leq \frac{k-2}{2} \log(\binom{n}{k}/\gamma)$ by the floor-free bound. To obtain the corresponding test for $R \geq k$, one must sum over every order $\ell \geq k$, as in the proof above.

A Fourier estimates for the degree vector

We give the proof of Lemma 2.1 in a form that treats the complete and deleted arrays together. The saddle-point calculation follows the degree-enumeration method of Barvinok and Hartigan [2]; see also Isaev and McKay [5]. Here the edge logits are arbitrary and the residual upper bound holds at any integral target.

Let $\mathcal{E}$ be the available pairs after deleting the pairs in J (take $J = \varnothing$ for the full array), let $Q_{\mathcal{E}} = \sum_{e \in \mathcal{E}} w_e u_e u_e^T$, and set

$$\Xi_{\mathcal{E}} = (2\pi)^{k/2} (\det Q_{\mathcal{E}})^{-1/2}, \qquad M(t) = \prod_{ab \in \mathcal{E}} |1 - q_{ab} + q_{ab} e^{i(t_a + t_b)}|.$$

Integration is with respect to Lebesgue measure on $\mathbb{T}^k = [-\pi, \pi)^k$. Fourier inversion contributes the separate factor $(2\pi)^{-k}$.

The covariance. Write $m = k - |J| \geq (1-\rho)k$. In the unweighted sum $\sum_{e \in \mathcal{E}} u_e u_e^T$, contrasts within J and within its complement have eigenvalues m and $k-2$ respectively. On the two normalized constant vectors of those sets the matrix is

$$\begin{pmatrix} m & \sqrt{|J|m} \\ \sqrt{|J|m} & k + m - 2 \end{pmatrix}.$$

Its determinant is $2m(m-1)$ and its trace is at most $3k$. Taking the variance floor into account, or using the complete-pair identity of Lemma 2.3 for $J = \varnothing$, gives

$$c_{\delta,\rho} k I \preceq Q_{\mathcal{E}} \preceq (k-1) I/2. \tag{A.1}$$

The two peaks. Put $r_k = (\log k)/\sqrt{k}$. Only the cubes of radius r_k around 0 and $\pi\mathbf{1}$ contribute on the scale $\Xi_{\mathcal{E}}$. Here is the tail estimate, including the case of deletions. Fix $0 < \varepsilon < \pi/1000$ and call an available pair bad when the circular distance of $t_i + t_j$ from zero exceeds ε. Since

$$|1 - q + qe^{iu}|^2 = 1 - 4q(1-q)\sin^2(u/2),$$

the region with at least $k(\log k)^2$ bad pairs has integral at most $(2\pi)^k e^{-ck(\log k)^2} = o(\Xi_{\mathcal{E}})$ by (A.1). Otherwise there is a vertex outside J incident with $O_\rho((\log k)^2)$ bad pairs. Among its good neighbors outside J there are $\Omega_\rho(k^2)$ available pairs, so at least one of them is good. The three pair sums in this triangle put twice the angle of the chosen vertex within 3ε of zero. All but $R_0 = O_\rho((\log k)^2)$ angles are consequently within 3ε of the same $b \in \{0, \pi\}$.

Outside the cube of radius 10ε about b, fix an exceptional angle t_i. It has at least $(1-\rho)k - R_0 - 1$ available neighbors in this core. Replacing t_i by s within ε of b moves their pair sums from distances above 7ε to distances below 4ε (allowing equality at the latter endpoint). Each corresponding factor gains a fixed ratio $e^\gamma > 1$. For the other available neighbors, exclude from the replacement interval a k^{-2}-neighborhood of each $\pi - t_j$. This leaves length at least ε and bounds those factors below by k^{-2}/π, since $|1 - q + qe^{iu}| \geq |\cos(u/2)|$. In particular we can take $R_0 = \lceil 4(1-\rho)^{-1} \log^2 k \rceil + 2$ and fix a core of exactly $k - R_0$ coordinates.

For fixed core, exceptional index i, and all angles except t_i, let I^* be the surviving replacement interval with the exclusions above. For all $s \in I^*$ and all bad values t_i at distance greater than 10ε from b,

$$M(t) \le e^{-\gamma((1-\rho)k-R_0-1)}(\pi k^2)^{R_0} M(t_{-i}, s) \le e^{-ck} M(t_{-i}, s).$$

This is a product inequality and remains valid if a nonincident factor vanishes. Since $|I^*| \ge \varepsilon$, integrating in the two scalar variables bounds the bad-fiber integral by $(2\pi/\varepsilon)e^{-ck}$ times the full fiber integral. Fubini and a union bound over at most $2k\binom{k}{R_0}$ choices of phase, core complement and index show that the mass outside the two fixed cubes is at most $o(\Xi_{\mathcal{E}}) + e^{-c'k}\int M$. The combinatorial cost has logarithm $O_\rho(\log^3 k)$.

Inside the fixed cube about zero, $-\log M$ is even and has Hessian at least $c_{\delta,\rho} kI$: each scalar negative log modulus has second derivative at least $\delta(1-\delta)/2$ near zero, and the available-pair quadratic form satisfies the lower bound above. For the probability measure proportional to M on this cube, each coordinate has tail at r at most $2e^{-c_{\delta,\rho} k r^2/2}$. Indeed, factoring $e^{-c_{\delta,\rho} k\|t\|^2/2}$ from the density leaves a log-concave function on the cube. Prékopa's marginal theorem then gives a coordinate density proportional to $e^{-c_{\delta,\rho} k x^2/2} h_0(x)$, with h_0 even and log-concave (extend the cube density by zero). Thus h_0 decreases in $|x|$. Its covariance with $\mathbf{1}_{\{|x|\ge r\}}$ under the centered Gaussian law is nonpositive, which proves the tail comparison after normalization. A union bound at r_k gives $o(1)$ of the fixed-cube mass outside the small cube. Translation handles the cube at $\pi\mathbf{1}$. Once the small-cube integral of M is $O(\Xi_{\mathcal{E}})$, absorption of $e^{-c'k}\int M$ proves the asserted tail estimate.

The local expansion. Write $U = [-r_k, r_k]^k$ and $s_{ij} = t_i + t_j$. Multiply the characteristic function by the phase that centers at the mean. On U its logarithm is

$$-\tfrac{1}{2} t^T Q_{\mathcal{E}} t + i f(t) + h(t) + O_\delta(k^{-1/2}\log^5 k),$$

where the sums below run over available pairs and

$$f = \tfrac{1}{6}\sum_{ij} w_{ij}(2q_{ij}-1)s_{ij}^3, \qquad h = \tfrac{1}{24}\sum_{ij} w_{ij}(1-6w_{ij})s_{ij}^4.$$

The error comes from summing $O_\delta(|s_{ij}|^5)$ per edge. Let $T \sim N(0, Q_{\mathcal{E}}^{-1})$ and $\nu = \mathbb{E}h(T)$. Its edge-coordinate variances are $O(k^{-1})$, giving $\|h(T)\|_{L^2} = O(1)$ and $\Pr(T \notin U) \le 2ke^{-c\log^2 k}$. On U, $|h| \le C\log^4 k$ and $\|\nabla h\|_2 = O(\log^3 k)$. In standard Gaussian coordinates h therefore has Lipschitz constant $K = O(k^{-1/2}\log^3 k)$. Extend it with the same Lipschitz constant and clip to $[-C\log^4 k, C\log^4 k]$. The extension H has $\mathbb{E}H = \nu + o(1)$ by the Gaussian tail bound. Gaussian concentration gives $\mathbb{E}e^{s(H-\mathbb{E}H)} \le e^{s^2K^2/2}$ for $s = 1, 2$. Together with Jensen's lower bound this gives $\mathbb{E}e^{sH} = e^{s\nu} + o(1)$ for $s = 1, 2$, since $|\nu| = O(1)$. Expanding the square gives $\mathbb{E}(e^H - e^\nu)^2 = o(1)$. The extension agrees with h on U in Gaussian coordinates, and $\Pr(T \notin U) = o(1)$, so $e^{h(T)}\mathbf{1}_U(T) = e^\nu + o_{L^1}(1)$. Integrating the modulus on U gives $\Xi_{\mathcal{E}}(e^\nu + o(1))$, which proves the residual atom bound by inversion and the preceding tail estimate.

The centered atom. For the full array, write $T = Q^{-1/2}Z$ with Z standard Gaussian. For $e = \{i, j\}$, write $q_e = q_{ij}$, $v_e = Q^{-1/2}u_e$, $Y_e = \langle v_e, Z\rangle$ and $a_e = w_e(2q_e - 1)/6$. Put $F = f(T) = \sum_e a_e Y_e^3$ and $\mu = \mathbb{E}F^2$. We show directly that $\mathbb{E}e^{iF} = e^{-\mu/2} + O_\delta(k^{-1/2})$. The proof uses the Gaussian Poincaré inequality twice, in the spirit of [3]. All derivatives below are with respect to Z.

The identity $\sum_e w_e v_e v_e^T = I$ and the variance floor give $\sum_e v_e v_e^T \preceq CI$. In particular $\|\sum_e t_e v_e\|^2 \le C\sum_e t_e^2$ for all scalar coefficients t_e, by duality. Also $\|v_e\|^2 \le C/k$ by (A.1).

Since $\nabla F = 3\sum_e a_e Y_e^2 v_e$, the Gaussian fourth moment gives $\mathbb{E}\|\nabla F\|^2 \le C\sum_e \mathbb{E}Y_e^4 = O(1)$. Thus $\mu = O(1)$ by Gaussian Poincaré and oddness. Let $\ell = \mathbb{E}\nabla F$, so $\|\ell\| = O(1)$.

For the Gaussian generator $\mathcal{L} = \Delta - z\cdot\nabla$ put $u(z) = (F(z) + 2\langle \ell, z\rangle)/3$. Since F is homogeneous cubic and $\Delta F(z) = 2\langle \ell, z\rangle$, one has $-\mathcal{L}u = F$. Gaussian integration by parts now gives

$$\mathbb{E}[Fg(F)] = \mathbb{E}[g'(F)\Gamma], \qquad \Gamma = \tfrac{1}{3}(\|\nabla F\|^2 + 2\langle \ell, \nabla F\rangle), \qquad \mathbb{E}\Gamma = \mu.$$

The last equality follows by taking $g(x) = x$. To bound the variance, put $b_e = \langle v_e, \nabla F + \ell\rangle$. Differentiating Γ and using $\nabla^2 F = 6\sum_e a_e Y_e v_e v_e^T$ gives

$$\|\nabla\Gamma\|^2 = 16\Big\|\sum_e a_e Y_e b_e v_e\Big\|^2 \le C\sum_e Y_e^2 b_e^2.$$

Each b_e is a polynomial of degree at most two. For any such polynomial q and any Gaussian coordinate $Y = \langle v, Z\rangle$, $\mathbb{E}[Y^2 q^2] \le 6\|v\|^2\mathbb{E}q^2$, regardless of their dependence. To see this for $v \neq 0$, rotate $Y/\|v\|$ to a standard coordinate x and condition on the others. Write $q = A(x^2-1) + Bx + D$. Then $\mathbb{E}_x q^2 = 2A^2 + B^2 + D^2$, whereas $\mathbb{E}_x x^2 q^2 = 10A^2 + 3B^2 + D^2 + 4AD \le 6\mathbb{E}_x q^2$. Another application of Poincaré, followed by $\sum_e v_e v_e^T \preceq CI$, gives

$$\operatorname{Var}\Gamma \le \mathbb{E}\|\nabla\Gamma\|^2 \le (C/k)\sum_e \mathbb{E}b_e^2 \le (C'/k)\mathbb{E}\|\nabla F + \ell\|^2 = O(k^{-1}).$$

Finally $\phi(s) = \mathbb{E}e^{isF}$ satisfies $\phi'(s) + \mu s\phi(s) = -s\,\mathbb{E}[(\Gamma - \mu)e^{isF}]$. Integration on $[0,1]$ gives $|\phi(1) - e^{-\mu/2}| \le \frac{1}{2}\sqrt{\operatorname{Var}\Gamma}$. This estimate is uniform even when μ approaches zero.

Combining the cubic and quartic estimates gives the centered integral on U as $\Xi_{\mathcal{E}}(e^{-\mu/2+\nu} + o(1))$. For an integral mean b of even total, the phase factor at $\pi\mathbf{1}$ is $e^{-i\pi\sum b_i} = 1$. The two peaks therefore add. The full centered probability is

$$\frac{2}{(2\pi)^{k/2}\sqrt{\det Q}}(e^{-\mu/2+\nu} + o(1)).$$

The leading correction is bounded away from zero uniformly since $\mu + |\nu| = O_\delta(1)$. The upper spectral bound in (A.1) now gives the elementary exponential lower bound in Lemma 2.1.

AI assistance. OpenAI's Codex assisted with revising the manuscript and reviewing its proofs, including the Fourier and overlap arguments. Codex also assisted in developing the accompanying Lean formalization. A human contributor has checked and understood most of the results stated in the abstract.

References

[1] N. Alon, M. Krivelevich and B. Sudakov, *Large nearly regular induced subgraphs*, SIAM Journal on Discrete Mathematics **22** (2008), 1325–1337. doi:10.1137/070704927.

[2] A. Barvinok and J. A. Hartigan, *The number of graphs and a random graph with a given degree sequence*, Random Structures & Algorithms **42** (2013), 301–348. doi:10.1002/rsa.20409. Preprint: arXiv:1003.0356v3.

[3] S. Chatterjee, *Fluctuations of eigenvalues and second order Poincaré inequalities*, Probability Theory and Related Fields **143** (2009), 1–40. doi:10.1007/s00440-007-0118-6.

[4] P. W. Dyson and B. D. McKay, *Ramsey numbers for regular induced subgraphs*, preprint (2026), arXiv:2604.08215v3.

[5] M. Isaev and B. D. McKay, *Complex martingales and asymptotic enumeration*, Random Structures & Algorithms **52** (2018), 617–661. doi:10.1002/rsa.20754.

[6] M. Krivelevich, B. Sudakov and N. Wormald, *Regular induced subgraphs of a random graph*, Random Structures & Algorithms **38** (2011), 235–250. doi:10.1002/rsa.20324.

[7] A. E. Levy, *The lower-bound problem for regular induced subgraphs of type-based random graphs*, preprint (2026), arXiv:2608.08169v2.

[8] A. F. Ramos, D. B. Hulak and R. J. G. B. de Queiroz, *Pair-trace absorption certificates for regular induced subgraphs*, preprint (2026), arXiv:2604.23882v1.

[9] A. F. Ramos, D. B. Hulak and R. J. G. B. de Queiroz, *Regular induced subgraphs in bounded independent-edge arrays*, Palomar Registry, 19 November 2026, version 1. Registry entry; Lean source at commit `251dd555`; coverage map.