MINIMAX SEMANTIC ENTROPY FOR FOUR MEANINGS FROM WEAK PAIRWISE JUDGMENTS

Anonymous authors
Paper under double-blind review

ABSTRACT

Semantic entropy scores a language model's uncertainty by the entropy of meanings among its sampled answers, seen only through a pairwise judge. When the judge is weak, the answers cannot be sorted by meaning, yet the verdicts may still determine the entropy. We model the judge as a known channel on the complete graph: given the meanings, it declares each pair of answers equivalent independently, with probability $q + \Delta$ within a meaning and q across meanings. With at most three meanings, the edge and wedge averages of the verdict graph estimate the pair and triple collision probabilities, which fix the entropy; with four they do not, and the three-star average supplies the missing moment. For n answers, at most four meanings and q bounded away from 0 and 1, the minimax squared risk has order $\min\{1, (n^4\Delta^6)^{-1/3}\}$ on every fixed polynomial window $n\Delta^2 \le n^{-\kappa}$, $\kappa > 0$, and the entropy is uniformly consistently estimable exactly when $n^4\Delta^6 \to \infty$. With $\Delta = n^{-\gamma}$ the critical exponent is $2/3$, against $3/4$ for three meanings in concurrent work, so a judge with $2/3 \le \gamma < 3/4$ suffices for three meanings and not for four. The rate is attained by an estimator that projects the three averages onto the valid moment set and solves a quartic. For the lower bound, a third-order inclusion–exclusion sweep of the likelihood score, whose coefficients obey an exact motif identity, bounds the Fisher information. If only pairs fixed in advance are judged, the risk bound depends on the number of three-stars they form, which hub designs maximize up to a constant factor.

1 INTRODUCTION

Semantic entropy is only as reliable as the judge that sorts a language model's answers by meaning. When the judge separates meanings weakly, sorting fails, and the question becomes how much of the entropy its pairwise verdicts still determine. With at most three meanings the answer runs through the collision probabilities $P_2 = \sum_c \theta_c^2$ and $P_3 = \sum_c \theta_c^3$ of the meaning distribution θ; the edges and wedges of the verdict graph estimate both. With four meanings these two moments no longer determine the entropy: $(\frac{1}{2}, \frac{1}{4}, \frac{1}{4}, 0)$ and $(0.45, 0.4, 0.1, 0.05)$ share $P_2 = 0.375$ and $P_3 = 0.15625$, and their entropies differ by 0.066. We show that the three-stars of the verdict graph supply the missing moment at the sharp rate, and that the critical separation exponent moves from $3/4$ to $2/3$.

Semantic entropy scores a language model's uncertainty about a prompt by the entropy of the distribution of *meanings* among its sampled answers (Kuhn et al., 2023; Farquhar et al., 2024). The meanings are never observed: the discrete estimator (Farquhar et al., 2024) samples n answers, asks an equivalence judge about pairs of them, groups the answers by the verdicts and plugs in the group proportions. We model the judge as a known channel: given the meanings, it declares two answers equivalent with probability $q + \Delta$ if they share a meaning and q otherwise, independently across pairs. Concurrent work (Anonymous, 2026) shows that for three meanings the grouping cannot be recovered once $n\Delta^2 \to 0$: no estimator of the meanings beats assigning every answer to the most likely one. In simulation, a spectral-clustering plug-in forced to four clusters returns estimates near $\log 4$ at all six distributions tested (Table 8 in Appendix I). The entropy stays estimable: with at most three meanings its minimax squared risk has order $\min\{1, (n^3\Delta^4)^{-1/2}\}$ throughout the weakly separated window $n\Delta^2 \le c_1$, and with $\Delta = n^{-\gamma}$ it is uniformly consistently estimable exactly when $\gamma < 3/4$ (Anonymous, 2026). Whether the threshold $3/4$ is the first term of a sequence was left open.

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

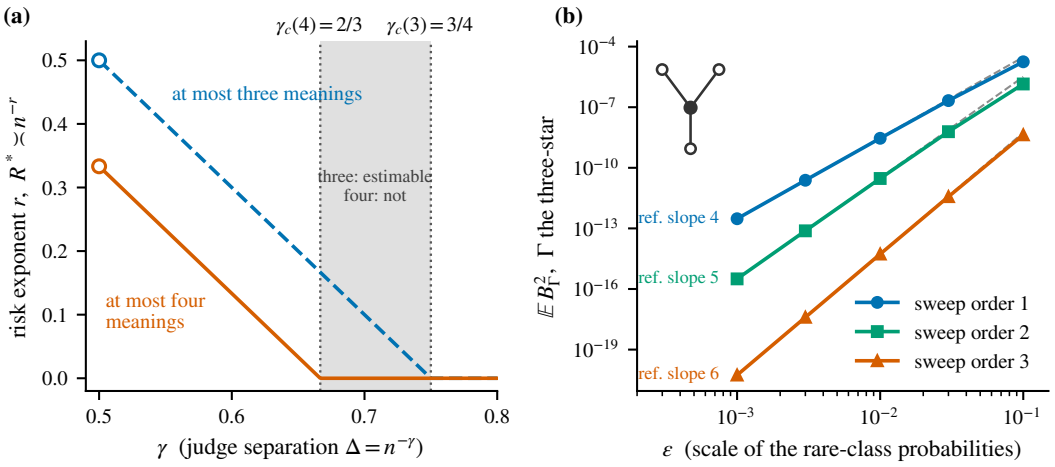


Figure 1: **(a)** Exponent r in $R^* \asymp n^{-r}$ for the semantic entropy with separation $\Delta = n^{-\gamma}$ and $q \in [q_0, 1 - q_0]$ fixed, for at most three meanings (concurrent work) and at most four (this paper, Corollary 3). Uniform consistency fails from $\gamma_c = 3/4$ and $\gamma_c = 2/3$ on, respectively; in the shaded band the entropy is uniformly estimable with at most three meanings and not with four. Open circles mark $\gamma = 1/2$, where $n\Delta^2 = 1$ and the matched range ends. **(b)** Second moment of the coefficient B_Γ of the three-star Γ in the swept score, over labels drawn from the least-favourable distribution $\theta(1/2)$ with rare-class scale ε, for sweeps of order one, two and three. Grey dashed reference lines of slopes 4, 5 and 6 pass through the points at $\varepsilon = 10^{-3}$; the measured slopes between $\varepsilon = 3 \cdot 10^{-3}$ and 10^{-3} are 3.992, 4.986 and 5.996. The third-order sweep reaches the order ε^6 that the matching lower bound needs.

Results. We prove that, with at most four meanings and q in a fixed interval $[q_0, 1-q_0]$, the minimax squared risk has order $\min\{1, (n^4\Delta^6)^{-1/3}\}$ on every fixed polynomial window $n\Delta^2 \le n^{-\kappa}$, and that the entropy is uniformly consistently estimable exactly when $n^4\Delta^6 \to \infty$, with no window condition. The threshold for $\Delta = n^{-\gamma}$ is therefore $\gamma_c(4) = 2/3$ (Figure 1(a)): for $2/3 \le \gamma < 3/4$, entropy with at most three meanings is uniformly estimable and entropy with four is not. The rate is attained by a tuning-free estimator. The average over the three-stars of the centred verdict graph is an unbiased estimate of P_4; together with the edge and wedge averages it is projected onto the set of valid moment vectors, and a quartic returns the distribution. Near a point mass the entropy is Hölder-$\frac{1}{3}$ in these moments, which turns a squared moment error of order $(n^4\Delta^6)^{-1}$ into the entropy rate.

The lower bound. Two distributions with the same P_2 and P_3 give verdict graphs with the same edge and wedge statistics in mean; they differ through P_4, by ε^3 for an entropy gap of order ε. We join such a pair inside the fibre of fixed (P_2, P_3) and bound the Fisher information of the verdict graph along it. Averaging labels out of the likelihood score, a Rao–Blackwell sweep, removes most of its fluctuation, but sweeps of order one and two do not reach the order ε^6 the bound needs (Figure 1(b)). A third-order inclusion–exclusion sweep does: its Taylor coefficients obey an exact motif identity and gain a factor ε for every majority-labelled vertex, and pairing overlapping motifs, together with a Taylor-remainder bound, controls the full likelihood score, not only its low-degree part. The three-meaning lower bound instead combines a one-sweep representation of the score with nonperturbative influence bounds on the whole window $n\Delta^2 \le c_1$; the argument here is finite-order and gives an information-theoretic lower bound on every fixed polynomial window $n\Delta^2 \le n^{-\kappa}$.

Contributions.

- The minimax rate of semantic entropy for at most four meanings, $R_4^* \asymp \min\{1, (n^4\Delta^6)^{-1/3}\}$ on every fixed polynomial window, and the consistency threshold $\gamma_c(4) = 2/3$ (Theorems 1 and 2 and Corollary 3). The upper bound holds for every channel and every $n \ge 4$ with explicit constants;

its statistics cost $O(n^2)$ arithmetic operations, followed by a fixed-dimensional projection and a quartic.

- A finite-order route to an information-theoretic lower bound: an exact motif identity for the third-order sweep, a factor ε for every majority-labelled vertex, and pairing with remainder control, which bound the Fisher information of the full likelihood (Proposition 5 and Lemma 6).
- An upper bound for query design: on any query graph fixed in advance, the risk is bounded in terms of the number S_3 of three-stars, and for $n \geq 6$ and $n-1 \leq Q \leq \binom{n}{2}$ queries, hub designs reach $S_3 \geq n^2 Q/48$, within a constant factor of the largest possible value (Theorem 7 and Corollary 8).

Section 2 states the model and the main results, Section 3 explains the lower bound, Section 4 treats query design, and Section 5 discusses related work; all proofs are in the appendix.

2 Setup and main results

Model. A language model is sampled n times on one prompt. Answer i has a meaning $Z_i \in [4] := \{0,1,2,3\}$, and $Z_1, \dots, Z_n$ are i.i.d. from $\theta \in \Theta_4 := \{\theta \in [0,1]^4 : \sum_c \theta_c = 1\}$, so at most four meanings occur and zero coordinates are allowed. We also call Z_i the label of answer i and the elements of $[4]$ classes. The target is the entropy $H(\theta) := -\sum_c \theta_c \log \theta_c$. A judge compares every pair of answers and reports $A_{ij} \in \{0,1\}$; given the meanings the verdicts are conditionally independent, with

$$A_{ij} \mid Z \;\sim\; \mathrm{Bern}\big(q + \Delta\, \mathbf{1}\{Z_i = Z_j\}\big), \qquad 1 \leq i < j \leq n. \tag{1}$$

The within-meaning and between-meaning agreement probabilities $p := q + \Delta$ and q are the same for all pairs and known, with $0 \leq q < p \leq 1$. We write $\delta := n\Delta^2$ and $\rho := \max\{p(1-p), q(1-q)\}/\Delta^2$. The upper bound holds for every channel; the lower bounds assume $q \in [q_0, 1-q_0]$ for a fixed $q_0 \in (0, 1/2]$, and constants written c, C depend only on q_0 unless a subscript indicates more. We study weak separation, $n\Delta^2 \to 0$. The minimax risk is

$$R_4^*(n; q, \Delta) := \inf_{\widehat{H}} \sup_{\theta \in \Theta_4} \mathbb{E}_\theta \big(\widehat{H} - H(\theta)\big)^2,$$

the infimum running over all estimators $\widehat{H} = \widehat{H}(A, V)$, where V is auxiliary randomness independent of (Z, A).

Moments. Let $P_k(\theta) := \sum_c \theta_c^k$, the probability that k independent answers share a meaning, and let $X_{ij} := (A_{ij} - q)/\Delta$ be the centred verdicts. Since $\mathbb{E}[X_{ij} \mid Z] = \mathbf{1}\{Z_i = Z_j\}$ and distinct pairs are conditionally independent, the product of the X_e over the edges of any tree on k vertices has mean P_k. With at most four meanings, Newton's identities give the elementary symmetric polynomials of θ from its moments,

$$e_2 = \tfrac{1}{2}(1 - P_2), \qquad e_3 = \tfrac{1}{6}(1 - 3P_2 + 2P_3), \qquad e_4 = \tfrac{1}{24}\big(1 - 6P_2 + 3P_2^2 + 8P_3 - 6P_4\big), \tag{2}$$

and $e_1 = 1$, so (P_2, P_3, P_4) determines θ up to permutation as the four roots of $x^4 - x^3 + e_2 x^2 - e_3 x + e_4$, and with it $H(\theta)$. The pair (P_2, P_3) does not (Section 1).

Algorithm 1 Three-star moment estimator

Require: verdicts $(A_{ij})_{i<j}$, channel (q, Δ)

1: $X_{ij} \leftarrow (A_{ij} - q)/\Delta$ for $i \neq j$

2: $\widehat{P}_2 \leftarrow \binom{n}{2}^{-1} \sum_{i<j} X_{ij}, \quad \widehat{P}_3 \leftarrow \frac{1}{n(n-1)(n-2)} \sum_i \sum_{j \neq k,\ j,k \neq i} X_{ij} X_{ik}$

3: $\widehat{P}_4 \leftarrow \big(n\binom{n-1}{3}\big)^{-1} \sum_i \sum_{j<k<l,\ j,k,l \neq i} X_{ij} X_{ik} X_{il}$

4: $\widetilde{P} \leftarrow$ a nearest point to $(\widehat{P}_2, \widehat{P}_3, \widehat{P}_4)$ in $\mathcal{M}_4 := \{(P_2, P_3, P_4)(\theta) : \theta \in \Theta_4\}$

5: $\widetilde{\theta} \leftarrow$ the four roots of $x^4 - x^3 + \tilde{e}_2 x^2 - \tilde{e}_3 x + \tilde{e}_4$, with $\tilde{e}_j$ computed from $\widetilde{P}$ by (2)

6: **return** $\widehat{H} := H(\widetilde{\theta})$

Algorithm 1 averages the centred verdicts over edges, ordered wedges and three-stars. With the row power sums $s_m(i) := \sum_{j \neq i} X_{ij}^m$, the inner sums of lines 2 and 3 are $s_1^2 - s_2$ and $(s_1^3 - 3s_1 s_2 + 2s_3)/6$,

so the three statistics cost $O(n^2)$ arithmetic operations. The projection minimizes a polynomial of degree 8 over the three-dimensional simplex and the quartic has degree four; computed on a grid of $O(n^{3/2})$ points, the projection keeps the bound of Theorem 1 with a larger constant (Remark 32). In simulation (Appendix I), along $n\Delta^2 = 1$ the mean squared error of this estimator falls with n at each of the six distributions tested, to at most 5.6×10^{-3} at $n = 1600$. A spectral-clustering plug-in forced to four clusters instead returns estimates near $\log 4$ at all six, and its error approaches that of the constant estimate $\log 4$ as n grows. Each verdict is a call to the judge: $\binom{n}{2}$ calls on the complete graph and fewer on the query graphs of Section 4. The grid depends only on n, so one build, about 2.5 seconds at $n = 1600$, serves every prompt with n answers, and an estimate then takes about one millisecond on one CPU core.

Theorem 1 (upper bound). *Let $n \geq 4$ and $0 \leq q < q + \Delta \leq 1$. For every $\theta \in \Theta_4$ the estimator of Algorithm 1 satisfies*

$$\mathbb{E}_\theta\big(\widehat{H} - H(\theta)\big)^2 \;\leq\; \min\big\{(\log 4)^2,\; 155\, V_4^{1/3}\big\}, \qquad V_4 := \frac{1}{n} + \frac{\rho}{n^2} + \frac{\rho^2}{n^3} + \frac{\rho^3}{n^4}.$$

If moreover $n\Delta^2 \leq 1$, then $\mathbb{E}_\theta(\widehat{H} - H(\theta))^2 \leq 171 \min\{1, (n^4\Delta^6)^{-1/3}\}$.

The exponent $1/3$ comes from two facts proved in Appendix G. The moment estimate is unbiased, with $\mathbb{E}\|\widehat{P} - P\|^2 \leq 78 V_4$, and $V_4 \leq 2(n^4\Delta^6)^{-1}$ when $n\Delta^2 \leq 1$, because $\rho \leq 1/(4\Delta^2)$. And with $D_e(\theta, \theta') := \sum_{j=2}^4 |e_j(\theta) - e_j(\theta')|$,

$$|H(\theta) - H(\theta')| \;\leq\; D_e \log 2 + \frac{2\pi}{\sqrt{3}}\, D_e^{1/3}, \tag{3}$$

a consequence of the integral representation $H(\theta) - H(\theta') = \int_0^\infty \log \frac{\prod_c (u+\theta_c)}{\prod_c (u+\theta'_c)}\, \mathrm{d}u$ of Jozsa & Mitchison (2015). The projection at most doubles the moment error, so the squared entropy error is of order $(\mathbb{E}\|\widehat{P} - P\|^2)^{1/3}$. The exponent is intrinsic: along the path of Section 3 the moments move by $O(\varepsilon^3)$ while the entropy moves by order ε.

Theorem 2 (lower bound). *For every $\kappa \in (0, 1/3]$ there are $c_\kappa > 0$ and n_κ, depending only on (q_0, κ), such that for all $n \geq n_\kappa$, $q \in [q_0, 1 - q_0]$ and $\Delta > 0$ with $n\Delta^2 \leq n^{-\kappa}$,*

$$R_4^*(n; q, \Delta) \;\geq\; c_\kappa \min\big\{1,\; (n^4\Delta^6)^{-1/3}\big\}.$$

Corollary 3 (rate and threshold). *(i) Fix $\gamma \in (1/2, 2/3]$ and $q \in [q_0, 1 - q_0]$. With $\Delta = n^{-\gamma}$, $R_4^*(n; q, \Delta) \asymp n^{-(4-6\gamma)/3}$ for $n \geq n_\gamma$, with n_γ and the constants depending only on (q_0, γ).*

(ii) Let $q_n \in [q_0, 1 - q_0]$ and $\Delta_n > 0$ with $q_n + \Delta_n \leq 1$. Estimators $\widehat{H}_n$ with $\sup_{\theta \in \Theta_4} \mathbb{E}_\theta(\widehat{H}_n - H(\theta))^2 \to 0$ exist if and only if $n^4\Delta_n^6 \to \infty$. In particular, for $\Delta_n = n^{-\gamma}$ the entropy is uniformly consistently estimable exactly when $\gamma < 2/3$.

Part (ii) needs no window condition: if $n^4\Delta_n^6$ stays bounded along a subsequence, then $n\Delta_n^2 \leq n^{-1/4}$ eventually and Theorem 2 applies with $\kappa = 1/4$, while sufficiency follows from the first bound of Theorem 1. With at most three meanings the threshold is $\gamma_c(3) = 3/4$ (Anonymous, 2026); for $2/3 \leq \gamma < 3/4$ the entropy is therefore uniformly estimable with at most three meanings and not with four. Table 1 summarizes the two cases, and Section 5 and Appendix A separate what the two analyses share from what is new here.

Calibration. The estimator uses the channel (q, Δ). Run with (q', Δ') instead, it computes the centred verdicts $sX_{ij} - \beta$, where $s := \Delta/\Delta'$ and $\beta := (q' - q)/\Delta'$, so its three statistics are an explicit affine image of the correct ones, and its risk is at most $\min\{(\log 4)^2, 904\,(V_4 + (|s-1| + |\beta|)^2)^{1/3}\}$ for $|s - 1| \leq 1/2$ and $|\beta| \leq 1$ (Corollary 31). The rate $\min\{1, (n^4\Delta^6)^{-1/3}\}$ is therefore kept when $|q' - q| + |\Delta' - \Delta|$ is of order $\Delta(n^4\Delta^6)^{-1/2}$, and when $n^4\Delta^6 \to \infty$ the estimator stays uniformly consistent as long as $|q' - q| = o(\Delta)$ and $\Delta'/\Delta \to 1$. The channel can be estimated from labelled pairs. Let q' and $q' + \Delta'$ be the agreement frequencies of the judge on m pairs of answers known to differ and on m pairs known to share a meaning, independent of the data, and run the estimator with (q', Δ') when $\Delta' > 0$ and return $\log 2$ otherwise (Corollary 31(ii)). For $n\Delta^2 \leq 1$ the rate is then kept once $m \geq n^4\Delta^4 = n^2\delta^2$, and when $n^4\Delta^6 \to \infty$ consistency is kept once $m\Delta^2 \to \infty$. With $m = \lceil n^2\delta^2 \rceil$ and $\delta \to 0$, these $2m$ calls are a vanishing fraction of the $\binom{n}{2}$ calls on the data, but they need pairs whose meanings are known.

Table 1: At most three meanings (concurrent work) and at most four meanings (this paper), for n answers observed through a known pairwise judge on the complete graph with $q \in [q_0, 1-q_0]$; $P_k := \sum_c \theta_c^k$ and, for the threshold row, $\Delta = n^{-\gamma}$.

	at most three meanings (Anonymous, 2026)	at most four meanings (this paper)
moments that determine θ up to permutation	P_2, P_3	P_2, P_3, P_4
statistics of the verdicts	edges, wedges	edges, wedges, three-stars
entropy modulus near a vertex	Hölder-$\frac{1}{2}$	Hölder-$\frac{1}{3}$
minimax squared-risk rate, up to constants, on the windows below	$\min\{1, (n^3\Delta^4)^{-1/2}\}$	$\min\{1, (n^4\Delta^6)^{-1/3}\}$
consistency threshold γ_c	$3/4$	$2/3$
lower-bound mechanism	one sweep and nonperturbative influence bounds	third-order sweep, motif identity, pairing and remainder control
matching lower bound proved on	the whole window $n\Delta^2 \le c_1$	every window $n\Delta^2 \le n^{-\kappa}$, $\kappa > 0$ fixed

3 Why the lower bound holds

This section explains the proof of Theorem 2; the details are in Appendices B to F. Throughout, $q \in [q_0, 1-q_0]$, κ is fixed, and $\varepsilon > 0$ is a small amplitude chosen at the end.

3.1 A path inside a fibre

Lemma 4 (least-favourable path). *Let $g(z) := z(z-\varepsilon/2)^2(z-1+\varepsilon)$, let r_ε be its critical point in $(0, \varepsilon/2)$, and for $u \in [1/4, 3/4]$ let $\theta(u)$ be the four roots of $g(z) = u\, g(r_\varepsilon)$, with θ_0 the largest. There is an absolute $\varepsilon_P > 0$ such that for $\varepsilon \le \varepsilon_P$: all four roots are positive and simple; P_2 and P_3 are constant along the path while $P_4(\theta(\frac{3}{4})) - P_4(\theta(\frac{1}{4})) = -2\varepsilon^3 m_\varepsilon$ with $m_\varepsilon = 1/54 + O(\varepsilon)$; the rare coordinates are $\theta_c = \varepsilon x_c(u)(1 + O(\varepsilon))$, where x_1, x_2, x_3 are the roots of $x(x-\frac{1}{2})^2 = u/54$; $|\dot\theta_c/\theta_c| \le C$ for $c \ge 1$, $|\dot\theta_0| \le C\varepsilon^3$; and $H(\theta(\frac{3}{4})) - H(\theta(\frac{1}{4})) = h_*\varepsilon + O(\varepsilon^2)$ with $h_* = 0.078447894650829\ldots > 0$ given by an explicit integral.*

Only the constant term of $g(z) - u\, g(r_\varepsilon)$ depends on u, so e_1, e_2, e_3 stay fixed and, by Newton's identities, so do P_2 and P_3; P_4 moves by order ε^3 while the entropy moves by order ε. By a Hellinger bound along the path and Le Cam's two-point argument (Lemmas 12 and 13), $\int_{1/4}^{3/4} I(u)\,\mathrm{d}u \le 1$ for the Fisher information $I(u)$ of the verdict graph implies $R_4^* \ge h_*^2\varepsilon^2/64$. The rate follows once the information is bounded at the amplitude $\varepsilon \asymp (n^4\Delta^6)^{-1/6}$, that is, once $I(u) = O(\varepsilon^6 n^4 \Delta^6)$.

3.2 Rewriting the score and sweeping

Write $J_{ij} := \log\frac{1-p}{1-q} + A_{ij}\log\frac{p(1-q)}{q(1-p)}$ for the log-likelihood-ratio weight of a verdict, $F_{ic} := \sum_{j\neq i} J_{ij}\mathbf{1}\{Z_j = c\}$ for the field of class c at answer i, and $\psi_c := \dot\theta_c/\theta_c$. The observed score $S_{\mathrm{obs}} := \partial_u \log P_u(A)$ equals $\mathbb{E}_u[\widetilde{S}_1 \mid A]$ with

$$\widetilde{S}_1 := \sum_{i=1}^n \Phi(F_i), \qquad \Phi(F) := \frac{\sum_c \dot\theta_c e^{F_c}}{\sum_c \theta_c e^{F_c}},$$

by the Glauber identity (Lemma 11). The proxy $\widetilde{S}_1$ has the right conditional mean but reacts to every individual label. For $U \subseteq [n]$ let $\mathcal{P}_U f := \mathbb{E}_u[f \mid A, Z_{-U}]$, which averages the labels in U over their posterior, and for $S \subseteq [n]$ let $\Lambda_S := \sum_{U\subseteq S}(-1)^{|S|-|U|}\mathcal{P}_U$. The sweep of order r is

$$\widetilde{S}_{\mathrm{RB}r} := \sum_{S\subseteq[n],\ |S|\le r} \Lambda_S \widetilde{S}_1.$$

Every Λ_S with $S \neq \varnothing$ has conditional mean zero given A, so $\mathbb{E}_u[\widetilde{S}_{\mathrm{RB}r} \mid A] = S_{\mathrm{obs}}$ and $I(u) \leq \mathbb{E}_u \widetilde{S}^2_{\mathrm{RB}r}$ (Lemma 16). The sweeps interpolate between $\widetilde{S}_1$ and S_{obs}: summing Λ_S over all $S \subseteq [n]$ gives $\mathcal{P}_{[n]}\widetilde{S}_1 = S_{\mathrm{obs}}$.

3.3 Why the order of the sweep matters

For fixed labels, $\widetilde{S}_{\mathrm{RB}r}$ is an explicit analytic function of the edge weights near $J = 0$ (Appendix C). Write $B^{(r)}_\Gamma(z)$ for its Taylor coefficient of the monomial $J^\Gamma := \prod_e J_e^{m_e}$ of an edge multiset Γ (a motif). Since each J_e is of order Δ, a motif with v vertices and d edges, none repeated, contributes about $n^v \Delta^{2d}\, \mathbb{E} B^{(r)}_\Gamma(Z)^2$ to the second moment: $n^3\Delta^4$ for wedges and $n^4\Delta^6$ for three-stars and three-edge paths; a repeated edge adds further factors of Δ. As $\varepsilon \to 0$ a posterior average resets a label to the majority class, and a sweep of order r cancels, at leading order, every term that depends on at most r rare labels. Within this construction the first-order sweep still leaves wedge coefficients with $\mathbb{E}B^2 \asymp \varepsilon^4$; by the count above their contribution is of order $\varepsilon^4 n^3 \Delta^4$, which at the critical amplitude is $n^{1/3}$ times the required order $\varepsilon^6 n^4 \Delta^6$. The second-order sweep leaves three-star coefficients with $\mathbb{E}B^2 \asymp \varepsilon^5$, a factor ε^{-1} above the order ε^6 that the lower bound needs. Both are computed explicitly in Remark 23, and Figure 1(b) shows the three orders on the three-star. We use the third-order sweep and write $B_\Gamma := B^{(3)}_\Gamma$.

3.4 The motif identity

For a motif Γ with multiplicities m_e put $\Gamma! := \prod_e m_e!$, let $V(\Gamma)$ be its vertex set and $v(\Gamma) := |V(\Gamma)|$, and let κ_Γ be the joint cumulant, under $Z \sim \theta^{\otimes n}$, of the co-membership indicators $W_e := \mathbf{1}\{Z_k = Z_l\}$, $e = \{k, l\} \in \Gamma$, each repeated m_e times.

Proposition 5 (motif identity, third-order sweep)**.** *For every motif Γ there is $\varepsilon_\Gamma > 0$ such that for $\varepsilon \leq \varepsilon_\Gamma$, $u \in [1/4, 3/4]$ and every labelling z:*

(i) $B_\Gamma(z)$ *depends only on* $z_{V(\Gamma)}$*, and* $B_\Gamma \equiv 0$ *if* Γ *is disconnected;*

(ii) $B_\Gamma = \partial_u \kappa_\Gamma/\Gamma! - \sum_{S \subseteq V(\Gamma),\, |S| \geq 4} [J^\Gamma] \Lambda_S \widetilde{S}_1$*, where* $[J^\Gamma]$ *extracts the coefficient of* J^Γ *and* Λ_S *acts on the system restricted to* $V(\Gamma)$*;*

(iii) $B_\Gamma \equiv 0$ *if* $v(\Gamma) \leq 3$*, and* $|\partial_u \kappa_\Gamma| \leq C_\Gamma \varepsilon^3$*;*

(iv) $|[J^\Gamma]\Lambda_S \widetilde{S}_1(z)| \leq C_\Gamma \varepsilon^{1+|M|}$ *for* $S \subseteq V(\Gamma)$*, where* M *is the set of vertices of* S *labelled* 0*;*

(v) $\mathbb{E}_\theta B_\Gamma(Z)^2 \leq C_\Gamma \varepsilon^6$ *if* Γ *is connected with* $v(\Gamma) \geq 4$*.*

The identity (ii) is the heart of the argument. On the system restricted to $V(\Gamma)$, the full sweep $\sum_{S \subseteq V(\Gamma)} \Lambda_S$ equals $\mathcal{P}_{V(\Gamma)}$, and by the Glauber identity $\mathcal{P}_{V(\Gamma)}\widetilde{S}_1 = \partial_u \log \mathbb{E}_\theta \exp(\sum_e J_e W_e)$, whose Taylor coefficients are the path derivatives of joint cumulants. A joint cumulant of co-membership indicators is a polynomial in $P_2, \ldots, P_{v(\Gamma)}$, which gives (iii): it is constant along the path when $v(\Gamma) \leq 3$, and its derivative is $O(\varepsilon^3)$ otherwise because $\dot{P}_k = O(\varepsilon^3)$ for $k \geq 4$. The terms with $|S| \geq 4$ are small by (iv): a sweep over a set of mostly majority-labelled vertices changes little, and each majority vertex in S costs a factor ε. We prove (iv) by continuing the posterior weights to a complex polydisc in the edge weights and applying Cauchy estimates. A labelling with m rare labels on $V(\Gamma)$ has probability of order ε^m, and every S with $|S| \geq 4$ then contains at least $4 - m$ majority vertices, so the second moment of the difference in (ii) is at most $\sum_m \varepsilon^m \cdot \varepsilon^{2+2(4-m)_+} \lesssim \varepsilon^6$, which gives (v). The same count for a sweep of order r gives $\mathbb{E}(B^{(r)}_\Gamma - \partial_u \kappa_\Gamma/\Gamma!)^2 \leq C_\Gamma \varepsilon^{r+3}$ (Remark 22), consistent with the slopes in Figure 1(b).

3.5 From motifs to the full score

Lemma 6 (information on a fixed window)**.** *For every $\kappa \in (0, 1/3]$ there are $a_0, \varepsilon_1 > 0$ and n_κ, depending only on (q_0, κ), such that for $n \geq n_\kappa$, $q \in [q_0, 1 - q_0]$, $n\Delta^2 \leq n^{-\kappa}$ and $\varepsilon :=$*

$\min\{\varepsilon_1,\ a_0(n^4\Delta^6)^{-1/6}\}$,

$$\sup_{u\in[1/4,3/4]} \mathbb{E}_u \widetilde{S}^2_{\mathrm{RB3}} \ \le\ 1.$$

Fix the degree $D := \max\{15, \lceil 14/(3\kappa)\rceil + 1\}$ and split $\widetilde{S}_{\mathrm{RB3}}$ into its Taylor polynomial of degree D in all $\binom{n}{2}$ edge weights and a remainder. The polynomial is $\sum_\Gamma J^\Gamma B_\Gamma(Z)$. Terms of vertex-disjoint motifs are independent; two motifs that share a vertex form a connected multigraph K with $v(K) \ge 4$, and $|\mathbb{E}[J_e^a \mid Z]| \le (2\Delta/q_0)^{\max\{2,a\}}$ together with Proposition 5(v) bounds their covariance by $C_D\varepsilon^6\Delta^{2d(K)}$, $d(K)$ the number of distinct edges. There are at most $n^{v(K)} \le n^{d(K)+1}$ placements, and $n^{d(K)+1}\Delta^{2d(K)} = n\delta^{d(K)} \le n\delta^3 = n^4\Delta^6$, so the polynomial has variance at most $C_D\varepsilon^6 n^4\Delta^6$. The remainder is bounded without any cancellation. The swept score splits into local pieces $\Lambda_S\Phi(F_i)$ that depend on the edge weights inside $\{i\}\cup S$ and on the fields from outside; on the event that every field is at most $\sqrt{\delta}\log n$, a multivariate Taylor bound gives $|\mathrm{Rem}_D| \le C_D\varepsilon n(n\Delta)^3(\sqrt{\delta}\log n)^{D-2}$, and the event fails with probability $e^{-c\log^2 n}$, where a polynomial bound on the remainder suffices. Since $\mathbb{E}_u\widetilde{S}_{\mathrm{RB3}} = 0$, $\mathbb{E}_u\widetilde{S}^2_{\mathrm{RB3}} \le 2\mathrm{Var}(\mathrm{Poly}_D) + 4\mathbb{E}\,\mathrm{Rem}_D^2$. At the critical amplitude the square of the remainder bound is of order $n^{14/3}\delta^D(\log n)^{2D-4}$, which tends to zero for fixed D exactly because $\delta \le n^{-\kappa}$ and $\kappa D > 14/3$. This is the only place where the window enters: when $n\Delta^2$ does not decay polynomially, no fixed D makes this remainder bound vanish. With Lemma 6, $\int_{1/4}^{3/4} I \le 1/2$, and Lemma 4 gives Theorem 2 in both regimes, $n^4\Delta^6$ large (vanishing risk) and bounded (constant risk).

Checks of the coefficient structure. We evaluated the exact Taylor coefficients numerically by enumerating every labelling of each motif (Appendix I). The full sweep agrees with $\partial_u\kappa_\Gamma/\Gamma!$ on eight motifs at $\varepsilon \in \{0.05, 0.01\}$ with maximum absolute residual 3.83×10^{-18}. Under the third-order sweep we evaluated sixteen motifs on up to seven vertices; for the fifteen whose coefficients are resolved in floating point, the local slope of $\log \mathbb{E}B_\Gamma^2$ against $\log\varepsilon$ between $\varepsilon = 3\cdot10^{-3}$ and 10^{-3} lies between 5.91 and 8.05 (Table 2).

4 Which pairs to query

Judge calls cost money, so a practitioner may query only some pairs. Let G be a simple graph on $[n]$ with Q edges and degrees d_i, fixed in advance independently of the labels and verdicts; only A_e, $e \in E(G)$, are observed. Algorithm 1 runs unchanged with the averages taken over the edges, ordered wedges and three-stars of G; write $\widehat{H}_G$ for its output and $S_3(G) := \sum_i \binom{d_i}{3}$ for the number of three-stars.

Theorem 7 (query design). *Let $q \in [q_0, 1-q_0]$, $n\Delta^2 \le q_0^2/4$ and $S_3(G) > 0$. For every $\theta \in \Theta_4$,*

$$\mathbb{E}_\theta\big(\widehat{H}_G - H(\theta)\big)^2 \ \le\ \min\big\{(\log 4)^2,\ 14\,(\Delta^6 S_3(G))^{-1/3}\big\}.$$

Corollary 8 (hub designs). *(i) Every graph with Q edges has $S_3(G) \le n^2Q/3$.*

(ii) If $n \ge 6$ and $n-1 \le Q \le \binom{n}{2}$, the hub design, which joins $\lfloor Q/(n-1)\rfloor$ vertices to all others and places the remaining edges arbitrarily, has $S_3 \ge n^2Q/48$; its uncapped bound $14(\Delta^6 S_3)^{-1/3}$ is within a factor $16^{1/3}$ of the smallest value that expression can take with Q queries.

(iii) A d-regular design with $d = 2Q/n \ge 3$ has $S_3 = n\binom{d}{3}$, and its uncapped bound exceeds that of the hub design by a factor of order $(n^2/Q)^{2/3}$.

By (i) and (ii), the bound of Theorem 7 can be made to tend to zero with Q queries if and only if $n^2Q\Delta^6 \to \infty$, that is, $Q\delta^3/n \to \infty$, and hub designs then attain it. For $\Delta = n^{-\gamma}$ with $\gamma \in (1/2, 2/3)$ the condition reads $Q/n^{6\gamma-2} \to \infty$: more than linearly many queries, which can still be a vanishing fraction of the $\binom{n}{2}$ pairs. These comparisons are between upper bounds; Theorem 2 covers only the complete graph. The hub design spends its budget on few answers of high degree, which makes the number of three-stars, the statistic that carries P_4, of the largest possible order.

5 Related work

Semantic entropy. Kuhn et al. (2023) introduced semantic entropy with bidirectional-entailment clustering, and Farquhar et al. (2024) used its discrete form, the entropy of the cluster proportions, to detect hallucinations. Later work estimates the same cluster entropy from fewer samples, by Bayesian estimation (Ciosek et al., 2025) or by Bayesian estimation with adaptive stopping and a guided exploration of the space of meanings (Sun et al., 2026), adjusts it for meanings not yet seen in the sample (McCabe et al., 2026; McCabe & Huang, 2026), or changes the uncertainty functional itself. Kernel language entropy replaces hard clusters by a semantic kernel (Nikitin et al., 2024), Nguyen et al. (2025) work with pairwise similarities directly, and Lin et al. (2024) score the degree of a similarity graph. Chen et al. (2026) argue that these methods amount to unsupervised clustering. None of them models the noise of the judge. Liu (2026) analyses the same observation model with a spectral-clustering plug-in, whose guarantee applies where clustering errors are small; the concurrent three-meaning analysis (Anonymous, 2026) gives the minimax rate below that regime.

Three meanings. That analysis is the closest to this paper (Table 1). For the upper bound it already has the edge and wedge averages with their variance bounds; the integral representation behind the entropy transfer, which gives the exponent $1/2$ there and $1/3$ here; the calibration from labelled pairs; the error transfer for an approximate projection; and query design on d-regular graphs. New here are the three-star average and its variance, the grid on the three-dimensional simplex and query design through the number of three-stars. The two lower bounds share their general tools (Lemmas 9 to 13) and differ in mechanism. Here the two hypotheses share P_2 and P_3 and differ through P_4 at order ε^3, so the Fisher information must be $O(\varepsilon^6 n^4 \Delta^6)$, which sweeps of order one and two do not reach (Section 3). The path inside the fibre, the higher-order sweeps with the motif identity and the majority-vertex factor, and the pairing with remainder control are new. The three-meaning lower bound combines a single sweep with nonperturbative influence bounds and holds on the whole window $n\Delta^2 \le c_1$; the finite-order argument here holds on every fixed polynomial window $n\Delta^2 \le n^{-\kappa}$.

Clustering from noisy pairwise verdicts. When the separation is large, the verdict graph is a stochastic block model whose partition can be recovered, and recovery guarantees are known in detail for block models and for clustering with noisy same-cluster queries (Lei & Rinaldo, 2015; Zhang & Zhou, 2016; Mazumdar & Saha, 2017; Del Pia et al., 2022; Xia & Huang, 2022), including repeated adaptive queries to the same channel (Teo et al., 2026). These results target the partition. For three meanings, Anonymous (2026) show that once $n\Delta^2 \to 0$ no estimator of the meanings beats assigning every answer to the most likely one, while the entropy stays consistently estimable as long as $n^3\Delta^4 \to \infty$; the present paper works in the same regime.

Moments of block models and symmetric functionals. Finite collections of small-subgraph statistics identify block-model parameters, generically up to label permutations (Allman et al., 2011) or, for a known number of blocks, under nondegeneracy conditions (Bickel et al., 2011); edge and triad moments identify the connectivity parameters of binary affiliation models with known group proportions (Ambroise & Matias, 2012); and the graph pencil method maps star and double-star densities to block parameters under degree separation (Gunderson et al., 2023). With observed samples, symmetric functionals such as the entropy can be estimated by local moment matching (Han et al., 2018), collision counts estimate integer-order Rényi entropies (Acharya et al., 2017), and Wu & Yang (2020) project estimated moments onto the set of valid moments before inverting. We use these ingredients for a functional observed through noisy pairwise verdicts and evaluated at the boundary of the simplex, where the rate is set by the Hölder-$\frac{1}{3}$ modulus, and pair them with a matching lower bound.

Entropy in symmetric coordinates. The integral representation of the entropy through $\prod_c (u+\theta_c)$ is due to Jozsa & Mitchison (2015), and our transfer (3) follows from it. Roots of a polynomial of degree d are Hölder continuous in its coefficients with exponent $1/d$ (Rainer, 2009); this is background for the exponent, which we prove directly.

Lower bounds. Low-degree polynomial methods give computational lower bounds for estimation (Schramm & Wein, 2022), including for distinguishing two planted models, as in community counting

(Rush et al., 2026). Our bound is information-theoretic: it controls the full likelihood score of this model. Detecting block structure against an Erdős–Rényi null is well studied (Banks et al., 2016; Gao & Lafferty, 2017; Jin et al., 2021), as is reconstruction in the planted partition model (Mossel et al., 2015); here both hypotheses are planted, with equal edge and wedge densities. Castillo & Orbanz (2022) show that uniform estimation of the connectivity parameters of block models is slow; our target is a different functional, the entropy of the class proportions. With observed labels the entropy on a fixed alphabet has minimax rate $1/n$ (Wu & Yang, 2016; Jiao et al., 2015).

Query design. Non-adaptive and adaptive designs for clustering with noisy queries aim at recovering the partition (Mazumdar & Saha, 2017; Xia & Huang, 2022); motif counting under subgraph sampling has been studied without verdict noise (Klusowski & Wu, 2018); and extremal star counts are known asymptotically (Reiher & Wagner, 2018).

6 Conclusion and Limitations

With at most four meanings and $q \in [q_0, 1 - q_0]$, the semantic entropy observed through weakly separated pairwise judgments has minimax squared risk of order $\min\{1, (n^4\Delta^6)^{-1/3}\}$ on every fixed polynomial window $n\Delta^2 \le n^{-\kappa}$, attained by edge, wedge and three-star averages followed by a projection and a quartic, and its consistency threshold is $\gamma_c(4) = 2/3$. The matching lower bound rests on a third-order sweep whose Taylor coefficients obey an exact motif identity: every connected motif pays ε^6, and pairing with a remainder bound turns this into a bound on the Fisher information of the full likelihood.

Limitations. The matching lower bound holds on every fixed polynomial window $n\Delta^2 \le n^{-\kappa}$, with constants that depend on κ; when $n\Delta^2$ decays more slowly than any power of n, or stays constant, the remainder bound fails and the rate is not established; covering the whole window $n\Delta^2 \le c_1$ needs nonperturbative control of the remainder. The results cover at most four meanings. The lower bound is proved for the complete graph, and the design comparisons are between upper bounds for designs fixed in advance. The model assumes common within-meaning and between-meaning agreement probabilities and conditionally independent verdicts. Corollary 31 covers errors in these two probabilities and their estimation from labelled pairs, but neither agreement probabilities that depend on the meanings or on the pair nor dependence between verdicts on pairs that share an answer. The lower bound treats the channel as known, so the labelled pairs are not part of its budget. The model has not been validated on real judges. The explicit upper constants, 155 and 14, are loose, and no numerical value of the lower-bound constant is given. For more than four meanings, neither the least-favourable path nor the entropy transfer is proved.

Outlook. We conjecture that K meanings need a sweep of order $2K - 5$ and that the threshold is $\gamma_c(K) = K/(2K - 2)$, which recovers $3/4$ and $2/3$ and tends to $1/2$, as the counting behind Remark 22 suggests.

Reproducibility Statement

The model is (1), the estimator is Algorithm 1, and Remark 32 gives the implementation used in the simulations. Every result is proved in the appendix with its hypotheses restated: the general tools in Appendix A, the lower bound and the threshold in Appendices B to F, the upper bound in Appendix G and the query designs in Appendix H. Appendix I describes the computations: the symbolic certificate in the proof of Remark 23(b), the interval enclosure of h_*, evaluations of exact coefficient formulas in floating point, a Monte Carlo variance check, a simulation of the estimator's risk and a measurement of its cost. The supplementary material contains the code of each computation and of the arithmetic behind the explicit constants, with the output each produced for this paper, and the code that draws Figure 1 and typesets Tables 2 to 8. The code runs on a CPU with pinned package versions and fixed random seeds, and its documentation maps each script to the result it supports.

AI use statement

In this work, we used a large language model to polish the text in the paper, making it more consistent with native-language expression and improving the readability of the paper.

REFERENCES

Jayadev Acharya, Alon Orlitsky, Ananda Theertha Suresh, and Himanshu Tyagi. Estimating Rényi entropy of discrete distributions. *IEEE Transactions on Information Theory*, 63(1):38–56, 2017. doi: 10.1109/TIT.2016.2620435.

Elizabeth S. Allman, Catherine Matias, and John A. Rhodes. Parameter identifiability in a class of random graph mixture models. *Journal of Statistical Planning and Inference*, 141(5):1719–1736, 2011. doi: 10.1016/j.jspi.2010.11.022.

Christophe Ambroise and Catherine Matias. New consistent and asymptotically normal parameter estimates for random-graph mixture models. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 74(1):3–35, 2012. doi: 10.1111/j.1467-9868.2011.01009.x.

Anonymous. Concurrent submission, 2026.

Jess Banks, Cristopher Moore, Joe Neeman, and Praneeth Netrapalli. Information-theoretic thresholds for community detection in sparse networks. In *29th Annual Conference on Learning Theory*, volume 49 of *Proceedings of Machine Learning Research*, pp. 383–416. PMLR, 2016.

Peter J. Bickel, Aiyou Chen, and Elizaveta Levina. The method of moments and degree distributions for network models. *The Annals of Statistics*, 39(5):2280–2301, 2011. doi: 10.1214/11-AOS904.

Ismaël Castillo and Peter Orbanz. Uniform estimation in stochastic block models is slow. *Electronic Journal of Statistics*, 16(1):2947–3000, 2022. doi: 10.1214/22-EJS2014.

Tiejin Chen, Longchao Da, Xiaoou Liu, and Hua Wei. Position: Uncertainty quantification in LLMs is just unsupervised clustering. In *Proceedings of the 43rd International Conference on Machine Learning*, 2026. Position Paper Track; arXiv:2605.19220.

Kamil Ciosek, Nicolò Felicioni, and Sina Ghiassian. Hallucination detection on a budget: Efficient Bayesian estimation of semantic entropy. *Transactions on Machine Learning Research*, 2025. arXiv:2504.03579.

Alberto Del Pia, Mingchen Ma, and Christos Tzamos. Clustering with queries under semi-random noise. In *Proceedings of Thirty Fifth Conference on Learning Theory*, volume 178 of *Proceedings of Machine Learning Research*, pp. 5278–5313. PMLR, 2022.

Sebastian Farquhar, Jannik Kossen, Lorenz Kuhn, and Yarin Gal. Detecting hallucinations in large language models using semantic entropy. *Nature*, 630(8017):625–630, 2024. doi: 10.1038/s41586-024-07421-0.

Chao Gao and John Lafferty. Testing for global network structure using small subgraph statistics. *arXiv preprint arXiv:1710.00862*, 2017.

Lee M. Gunderson, Gecia Bravo-Hermsdorff, and Peter Orbanz. The graph pencil method: Mapping subgraph densities to stochastic block models. In *Advances in Neural Information Processing Systems 36*, pp. 79583–79593, 2023. doi: 10.52202/075280-3484.

Yanjun Han, Jiantao Jiao, and Tsachy Weissman. Local moment matching: A unified methodology for symmetric functional estimation and distribution estimation under Wasserstein distance. In *Proceedings of the 31st Conference On Learning Theory*, volume 75 of *Proceedings of Machine Learning Research*, pp. 3189–3221. PMLR, 2018.

Jiantao Jiao, Kartik Venkat, Yanjun Han, and Tsachy Weissman. Minimax estimation of functionals of discrete distributions. *IEEE Transactions on Information Theory*, 61(5):2835–2885, 2015. doi: 10.1109/TIT.2015.2412945.

Jiashun Jin, Zheng Tracy Ke, and Shengming Luo. Optimal adaptivity of signed-polygon statistics for network testing. *The Annals of Statistics*, 49(6):3408–3433, 2021. doi: 10.1214/21-AOS2089.

Richard Jozsa and Graeme Mitchison. Symmetric polynomials in information theory: Entropy and subentropy. *Journal of Mathematical Physics*, 56(6):062201, 2015. doi: 10.1063/1.4922317.

Jason M. Klusowski and Yihong Wu. Counting motifs with graph sampling. In *Proceedings of the 31st Conference On Learning Theory*, volume 75 of *Proceedings of Machine Learning Research*, pp. 1966–2011. PMLR, 2018.

Lorenz Kuhn, Yarin Gal, and Sebastian Farquhar. Semantic uncertainty: Linguistic invariances for uncertainty estimation in natural language generation. In *The Eleventh International Conference on Learning Representations*, 2023.

Jing Lei and Alessandro Rinaldo. Consistency of spectral clustering in stochastic block models. *The Annals of Statistics*, 43(1):215–237, 2015. doi: 10.1214/14-AOS1274.

Zhen Lin, Shubhendu Trivedi, and Jimeng Sun. Generating with confidence: Uncertainty quantification for black-box large language models. *Transactions on Machine Learning Research*, 2024. arXiv:2305.19187.

Yi Liu. A statistically consistent measure of semantic uncertainty using language models. In *2026 8th International Conference on Natural Language Processing (ICNLP)*, pp. 305–321. IEEE, 2026. doi: 10.1109/ICNLP69856.2026.11527557. arXiv:2502.00507.

Arya Mazumdar and Barna Saha. Clustering with noisy queries. In *Advances in Neural Information Processing Systems 30*, 2017.

Lucas H. McCabe and H. Howie Huang. SENECA: Small-sample discrete entropy estimation via self-consistent missing mass. *arXiv preprint arXiv:2605.00668*, 2026.

Lucas H. McCabe, Rimon Melamed, Thomas Hartvigsen, and H. Howie Huang. Estimating semantic alphabet size for LLM uncertainty quantification. In *The Fourteenth International Conference on Learning Representations*, 2026. arXiv:2509.14478.

Elchanan Mossel, Joe Neeman, and Allan Sly. Reconstruction and estimation in the planted partition model. *Probability Theory and Related Fields*, 162(3–4):431–461, 2015. doi: 10.1007/s00440-014-0576-6.

Dang Nguyen, Ali Payani, and Baharan Mirzasoleiman. Beyond semantic entropy: Boosting LLM uncertainty quantification with pairwise semantic similarity. In *Findings of the Association for Computational Linguistics: ACL 2025*, pp. 4530–4540. Association for Computational Linguistics, 2025. doi: 10.18653/v1/2025.findings-acl.234.

Alexander Nikitin, Jannik Kossen, Yarin Gal, and Pekka Marttinen. Kernel language entropy: Fine-grained uncertainty quantification for LLMs from semantic similarities. In *Advances in Neural Information Processing Systems 37*, pp. 8901–8929, 2024.

Armin Rainer. Perturbation of complex polynomials and normal operators. *Mathematische Nachrichten*, 282(12):1623–1636, 2009. doi: 10.1002/mana.200910837.

Christian Reiher and Stephan Wagner. Maximum star densities. *Studia Scientiarum Mathematicarum Hungarica*, 55(2):238–259, 2018. doi: 10.1556/012.2018.55.2.1395.

Cynthia Rush, Fiona Skerman, Alexander S. Wein, and Dana Yang. Is it easier to count communities than find them? *Random Structures & Algorithms*, 68(4):e70071, 2026. doi: 10.1002/rsa.70071.

Tselil Schramm and Alexander S. Wein. Computational barriers to estimation from low-degree polynomials. *The Annals of Statistics*, 50(3):1833–1858, 2022. doi: 10.1214/22-aos2179.

Qiyao Sun, Xingming Li, Xixiang He, Ao Cheng, Xuanyu Ji, Hailun Lu, Runke Huang, and Qingyong Hu. Efficient hallucination detection: Adaptive Bayesian estimation of semantic entropy with guided semantic exploration. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pp. 33117–33125, 2026. doi: 10.1609/aaai.v40i39.40595. arXiv:2603.22812.

Rachel S. Y. Teo, P. N. Karthik, Ramya Korlakai Vinayak, and Vincent Y. F. Tan. Almost asymptotically optimal active clustering through pairwise observations. In *2026 IEEE International Symposium on Information Theory (ISIT)*, pp. 1–6. IEEE, 2026. doi: 10.1109/ISIT62367.2026.11653853.

Yihong Wu and Pengkun Yang. Minimax rates of entropy estimation on large alphabets via best polynomial approximation. *IEEE Transactions on Information Theory*, 62(6):3702–3720, 2016. doi: 10.1109/TIT.2016.2548468.

Yihong Wu and Pengkun Yang. Optimal estimation of Gaussian mixtures via denoised method of moments. *The Annals of Statistics*, 48(4):1981–2007, 2020. doi: 10.1214/19-AOS1873.

Jinghui Xia and Zengfeng Huang. Optimal clustering with noisy queries via multi-armed bandit. In *Proceedings of the 39th International Conference on Machine Learning*, volume 162 of *Proceedings of Machine Learning Research*, pp. 24315–24331. PMLR, 2022.

Anderson Y. Zhang and Harrison H. Zhou. Minimax rates of community detection in stochastic block models. *The Annals of Statistics*, 44(5):2252–2280, 2016. doi: 10.1214/15-AOS1428.

A NOTATION AND TOOLS

This appendix fixes the notation used in all proofs and states, with proofs, the general tools that the lower bound uses, so that the paper is self-contained.

Relation to the three-meaning analysis. The concurrent three-meaning work (Anonymous, 2026) uses three-class forms of the edge weights, the Glauber identity, the score representation, the Hellinger path bound and the two-point bound (Lemmas 9 to 13); here Lemma 9(iii) extends its bounds on the edge weights to moments of every order, and Lemma 10 extends its identity for one answer to a block of answers. It also has the same variance bounds for the edge and wedge averages (Lemma 28(ii),(iii)); the integral representation behind the entropy transfer, which gives the exponent $1/2$ there (Lemma 29); the affine form of miscalibrated statistics and the calibration of the channel from labelled pairs, for the edge and wedge averages (Corollary 31); the error transfer for an approximate projection, which it computes on a grid along the boundary curves of its two-dimensional moment set (Remark 32); and query design on d-regular graphs, for the edge and wedge averages (Theorem 7). The following have no counterpart there: the path along which P_2 and P_3 stay fixed (Lemma 15); the sweeps of every order and the motif identity with the majority-vertex bound (Proposition 5 and Appendices C and D); the pairing of overlapping motifs and the global degree-D Taylor remainder that control the full likelihood score (Lemmas 24 to 27); the variance of the three-star average (Lemma 28(iv)); the grid on the three-dimensional simplex with its tolerance bound (Remark 32); and query design on arbitrary fixed graphs through the number of three-stars, with the hub designs (Theorem 7 and Corollary 8). The three-meaning lower bound rests instead on a single sweep and influence bounds for the rewritten score. That work also treats two questions that are not treated here: it shows that no estimator is uniformly consistent in its weakly separated regime when the background rate q is unknown, and it gives the exact bias of its wedge average when verdicts that share an answer are dependent.

Model and notation. The classes are $[4] := \{0,1,2,3\}$ and $\Theta_4 := \{\theta \in [0,1]^4 : \sum_c \theta_c = 1\}$. The labels $Z_1, \dots, Z_n$ are i.i.d. from θ, $W_{ij} := \mathbf{1}\{Z_i = Z_j\}$, and given Z the verdicts A_{ij}, $i < j$, are independent with $A_{ij} \mid Z \sim \mathrm{Bern}(q + \Delta W_{ij})$; $p := q + \Delta$, $\delta := n\Delta^2$, $\rho := \max\{p(1-p), q(1-q)\}/\Delta^2$ and $P_k := \sum_c \theta_c^k$. We write $\mathbf{e}_c \in \mathbb{R}^4$ for the c-th basis vector and $\mathbf{1}_c$ for the indicator of class c. In Appendices A to F we assume $q \in [q_0, 1-q_0]$ with $q_0 \in (0, 1/2]$ and $0 < \Delta \le q_0/2$; then $p \le 1 - q_0/2$. Letters C, c denote positive constants that depend only on q_0, and a subscript such as C_D or C_Γ records additional dependence; their values may change from line to line.

Edge weights and fields. Put $\lambda := \log(p/q) > 0$, $\nu := \log\frac{1-p}{1-q} < 0$ and $J_{ij} := \nu + (\lambda - \nu)A_{ij}$ for $i \ne j$. For a vertex i and a class c the field is $F_{ic} := \sum_{j \ne i} J_{ij}\mathbf{1}\{Z_j = c\}$, so $F_i = \sum_{j \ne i} J_{ij}\mathbf{e}_{Z_j} \in \mathbb{R}^4$, and

$$\pi_{ic} := \frac{\theta_c e^{F_{ic}}}{\sum_b \theta_b e^{F_{ib}}}.$$

Since $\mathrm{Bern}(a; q+\Delta)/\mathrm{Bern}(a; q) = \exp(\nu + (\lambda - \nu)a)$ for $a \in \{0,1\}$,

$$\mathbb{P}(A \mid Z) = \prod_{i<j} \mathrm{Bern}(A_{ij}; q) \cdot \exp\Big(\sum_{i<j} J_{ij} W_{ij}\Big). \tag{4}$$

Lemma 9 (edge weights). *(i)* $0 < \lambda \le \Delta/q_0$, $0 < -\nu \le 2\Delta/q_0$ *and* $|J_{ij}| \le 2\Delta/q_0$.

(ii) Given Z, the weights J_{ij}, $i < j$, are conditionally independent, $|\mathbb{E}[J_{ij} \mid Z]| \le 2\Delta^2/q_0^2$ *and* $\mathrm{Var}(J_{ij} \mid Z) \le 3\Delta^2/q_0^2$.

(iii) For every integer $a \ge 1$,

$$\big|\mathbb{E}[J_{ij}^a \mid Z]\big| \le \Big(\frac{2\Delta}{q_0}\Big)^{c(a)}, \qquad c(1) := 2, \quad c(a) := a \ \ (a \ge 2). \tag{5}$$

Proof. (i) $\lambda = \log(1 + \Delta/q) \le \Delta/q \le \Delta/q_0$. With $x := \Delta/(1-q) \in (0,1)$ and $\log(1-x) \ge -x/(1-x)$ we get $\nu \ge -\Delta/(1-p) \ge -2\Delta/q_0$. Hence $|J_{ij}| \le \max\{\lambda, -\nu\} \le 2\Delta/q_0$.

(ii) Conditional independence holds because each J_{ij} is a function of A_{ij} alone. Given Z, $\mathbb{E}[J_{ij} \mid Z] = \nu + (\lambda - \nu)r$ with $r \in \{q, p\}$; for $r = p$ this is the Kullback–Leibler divergence $D(p \,\|\, q)$ and for $r = q$ it is $-D(q \,\|\, p)$. Using $D(x \,\|\, y) \le (x-y)^2/(y(1-y))$, $q(1-q) \ge q_0/2$, $p \ge q_0$ and $1-p \ge q_0/2$, both are at most $2\Delta^2/q_0^2$ in absolute value. Finally $\mathrm{Var}(J_{ij} \mid Z) = (\lambda-\nu)^2 r(1-r) \le \frac{1}{4}(3\Delta/q_0)^2 \le 3\Delta^2/q_0^2$.

(iii) For $a \ge 2$ use (i); for $a = 1$ use (ii) and $2\Delta^2/q_0^2 \le (2\Delta/q_0)^2$. □

Lemma 10 (Glauber identity). *Let $\theta_c > 0$ for all c, and let $U \subseteq [n]$. Given (A, Z_{-U}), the conditional law of Z_U is proportional to*

$$w(z'_U; Z_{-U}, J) := \prod_{k \in U} \theta_{z'_k} \cdot \exp\Big(\sum_{e=\{k,l\}:\, e \cap U \neq \varnothing} J_e \mathbf{1}\{z_k = z_l\} \Big), \qquad z := (z'_U, Z_{-U}). \tag{6}$$

In particular $\mathbb{P}(Z_i = c \mid A, Z_{-i}) = \pi_{ic}$.

Proof. By (4), $\mathbb{P}(Z_U = z'_U, A \mid Z_{-U})$ is $\prod_{k \in U} \theta_{z'_k}$ times $\prod_{i<j} \mathrm{Bern}(A_{ij}; q) \exp(\sum_e J_e W_e(z))$; the factors of the edges that do not meet U and the Bernoulli factors do not depend on z'_U. Normalizing gives the first claim; for $U = \{i\}$ the weight is $\theta_c e^{F_{ic}}$. □

Lemma 11 (score representation). *Let $u \mapsto \theta(u) \in \Theta_4$ be continuously differentiable on an interval, with all $\theta_c(u) > 0$ and $\sum_c \dot\theta_c = 0$, and write P_u for the law of A under $\theta(u)$, $S_{\mathrm{obs}}(A) := \partial_u \log P_u(A)$, $\psi_c := \dot\theta_c/\theta_c$ and*

$$\Phi(F) := \frac{\sum_c \dot\theta_c e^{F_c}}{\sum_c \theta_c e^{F_c}} = \sum_c \pi_c(F)\, \psi_c, \qquad \widetilde{S}_1 := \sum_{i=1}^n \Phi(F_i).$$

Then $S_{\mathrm{obs}} = \mathbb{E}_u[\sum_i \psi_{Z_i} \mid A] = \mathbb{E}_u[\widetilde{S}_1 \mid A]$, $\mathbb{E}_u S_{\mathrm{obs}} = 0$, and $\Phi(0) = 0$.

Proof. $P_u(A) = \sum_z \prod_i \theta_{z_i}(u)\, \mathbb{P}(A \mid Z = z)$ is a polynomial in $\theta(u)$, and differentiating term by term gives $\partial_u P_u(A) = \sum_z \big(\sum_i \psi_{z_i}\big) \prod_i \theta_{z_i} \mathbb{P}(A \mid z)$, that is, $S_{\mathrm{obs}} = \mathbb{E}_u[\sum_i \psi_{Z_i} \mid A]$. By Lemma 10 and the tower property, $\mathbb{E}_u[\psi_{Z_i} \mid A] = \mathbb{E}_u[\sum_c \pi_{ic}\psi_c \mid A] = \mathbb{E}_u[\Phi(F_i) \mid A]$. Next $\mathbb{E}_u S_{\mathrm{obs}} = \sum_A \partial_u P_u(A) = 0$, and $\Phi(0) = \sum_c \dot\theta_c = 0$. □

Lemma 12 (Hellinger path bound). *Let $\{P_u\}_{u \in [u_0, u_1]}$, $u_1 - u_0 \le 1$, be distributions on a finite set with $P_u(A) > 0$ and $u \mapsto P_u(A)$ continuously differentiable, and let $I(u) := \sum_A P_u(A)(\partial_u \log P_u(A))^2$. Then*

$$\mathrm{TV}(P_{u_0}, P_{u_1}) \;\le\; \frac{1}{2}\Big(\int_{u_0}^{u_1} I(u)\, \mathrm{d}u \Big)^{1/2}.$$

Proof. $\sqrt{P_{u_1}(A)} - \sqrt{P_{u_0}(A)} = \int_{u_0}^{u_1} \frac{1}{2}\sqrt{P_u(A)}\, \partial_u \log P_u(A)\, \mathrm{d}u$, whose square is at most $\frac{1}{4}\int_{u_0}^{u_1} P_u(A)(\partial_u \log P_u(A))^2 \mathrm{d}u$ by Cauchy–Schwarz on an interval of length at most one. Summing over A, $\sum_A (\sqrt{P_{u_1}} - \sqrt{P_{u_0}})^2 \le \frac{1}{4}\int I$. Finally $\sum_A |P - Q| \le \big(\sum_A (\sqrt{P} - \sqrt{Q})^2\big)^{1/2} \big(\sum_A (\sqrt{P} + \sqrt{Q})^2\big)^{1/2} \le 2\big(\sum_A (\sqrt{P} - \sqrt{Q})^2\big)^{1/2}$, so $\mathrm{TV} = \frac{1}{2}\sum_A |P - Q| \le \frac{1}{2}(\int I)^{1/2}$. □

Lemma 13 (two points). *Let $\omega > 0$ and $\theta^{(0)}, \theta^{(1)} \in \Theta_4$ with $|H(\theta^{(0)}) - H(\theta^{(1)})| \ge \omega$ and $\mathrm{TV}(P_{\theta^{(0)}}, P_{\theta^{(1)}}) \le 1/2$. Then every estimator $\widehat{H} = \widehat{H}(A, V)$, where V is an auxiliary random variable independent of (Z, A), satisfies $\max_{k \in \{0,1\}} \mathbb{E}_{\theta^{(k)}}(\widehat{H} - H(\theta^{(k)}))^2 \ge \omega^2/16$.*

Proof. The pair (A, V) has laws $P_{\theta^{(k)}} \otimes \mu$, whose total variation distance equals $\mathrm{TV}(P_{\theta^{(0)}}, P_{\theta^{(1)}})$. The events $E_k := \{|\widehat{H} - H(\theta^{(k)})| < \omega/2\}$ are disjoint, so $\mathbb{P}_0(E_0^c) + \mathbb{P}_1(E_1^c) \ge \mathbb{P}_0(E_0^c) + \mathbb{P}_1(E_0) \ge 1 - \mathrm{TV} \ge 1/2$. By Markov's inequality, $\sum_k \mathbb{E}_k(\widehat{H} - H(\theta^{(k)}))^2 \ge (\omega/2)^2 \cdot \frac{1}{2}$, and the maximum is at least half the sum. □

Lemma 14 (derivatives of Φ)**.** *Let $\pi(F)$ be the tilted law $\pi_c(F) = \theta_c e^{F_c}/\sum_b \theta_b e^{F_b}$. For $F \in \mathbb{C}^4$ with $\sum_b \theta_b e^{F_b} \neq 0$ the same formula defines complex weights, and $\Phi(F) = \sum_c \pi_c(F)\psi_c$. For every $m \geq 1$ and classes $c_1, \dots, c_m$,*

$$\partial_{c_1} \cdots \partial_{c_m} \Phi(F) = \kappa_{\pi(F)}\big(\psi; \mathbf{1}_{c_1}, \dots, \mathbf{1}_{c_m}\big),$$

the joint cumulant under $\pi(F)$. If $F \in \mathbb{R}^4$ and $g := \max_c |F_c|$, then $|\Phi(F)| \leq \|\dot\theta\|_1 e^{2g}$ and $|\partial_{c_1} \cdots \partial_{c_m} \Phi(F)| \leq C_m \|\dot\theta\|_1 e^{2g}$ with $C_m := \mathrm{Bell}(m+1)\, m!$. Moreover $\partial_c \partial_d \Phi(0) = \mathbf{1}\{c = d\}\dot\theta_c - \dot\theta_c \theta_d - \theta_c \dot\theta_d$.

Proof. $\mathcal{K}(F) := \log \sum_b \theta_b e^{F_b}$ is the cumulant generating function of the four-point family $\pi(F)$, so $\partial_{c_0} \cdots \partial_{c_m} \mathcal{K} = \kappa_\pi(\mathbf{1}_{c_0}, \dots, \mathbf{1}_{c_m})$; differentiating $\Phi = \sum_{c_0} \psi_{c_0} \partial_{c_0} \mathcal{K}$ and using multilinearity gives the cumulant formula. In the moment expansion of the cumulant over partitions of $\{0, 1, \dots, m\}$, ψ enters only through the block containing 0, whose moment is either $\Phi(F)$ or $\pi_c \psi_c \mathbf{1}\{\cdot\} = \dot\theta_c e^{F_c}/\sum_b \theta_b e^{F_b} \cdot \mathbf{1}\{\cdot\}$; for real F both are at most $\|\dot\theta\|_1 e^{2g}$ in absolute value, since $\sum_b \theta_b e^{F_b} \geq e^{-g}$. The other blocks are probabilities, at most one. There are $\mathrm{Bell}(m+1)$ partitions and the coefficients are at most $m!$ in absolute value. The formula at $F = 0$ follows from $\partial_c \Phi = \pi_c(\psi_c - \Phi)$, $\partial_d \pi_c = \pi_c(\mathbf{1}\{c = d\} - \pi_d)$, $\pi(0) = \theta$ and $\Phi(0) = 0$. □

B THE LEAST-FAVOURABLE PATH

Let $\varepsilon \in (0, \varepsilon_P]$ with ε_P a small absolute constant, put $a := \varepsilon/2$, $b := 1 - \varepsilon$ and

$$g(z) := z(z-a)^2(z-b) = z^4 - z^3 + (a^2 + 2ab)z^2 - a^2 b\, z,$$

where $2a + b = 1$ was used. Then $g'(z) = (z-a)\big(4z^2 - (2a+3b)z + ab\big)$, and $r_\varepsilon := \big[(2a+3b) - \sqrt{(2a+3b)^2 - 16ab}\big]/8$ is the unique critical point of g in $(0, a)$, with $r_\varepsilon = \varepsilon/6 + O(\varepsilon^2)$. Since $g < 0$ on $(0, a)$, $g(r_\varepsilon) < 0$; put $m_\varepsilon := -g(r_\varepsilon)/\varepsilon^3$, so that $m_\varepsilon = 1/54 + O(\varepsilon)$ and m_ε is analytic in ε near 0. For $u \in [1/4, 3/4]$ let $\theta(u)$ be the four roots of $g(z) = u\, g(r_\varepsilon)$, with θ_0 the largest, and write $\psi_c := \dot\theta_c/\theta_c$ with the dot denoting ∂_u.

Lemma 15 (restatement of Lemma 4)**.** *There are absolute constants $\varepsilon_P > 0$ and C_P such that for $\varepsilon \leq \varepsilon_P$ and $u \in [1/4, 3/4]$:*

(i) the four roots are positive and simple and $\sum_c \theta_c = 1$; for $c = 1, 2, 3$, $\theta_c = \varepsilon x_c(u)(1 + O(\varepsilon))$, where $x_1(u) > x_2(u) > x_3(u) > 0$ are the roots of $x(x - \frac{1}{2})^2 = u/54$, and $\theta_0 = 1 - \sum_{c \geq 1} \theta_c$;

(ii) e_1, e_2, e_3 are constant along the path and $e_4 = \varepsilon^3 m_\varepsilon u$; hence P_2 and P_3 are constant,

$$\sum_c \dot\theta_c = \sum_c \theta_c \dot\theta_c = \sum_c \theta_c^2 \dot\theta_c = 0, \qquad P_4(\theta(\tfrac{3}{4})) - P_4(\theta(\tfrac{1}{4})) = -2\varepsilon^3 m_\varepsilon; \tag{7}$$

(iii) $\dot\theta_c = g(r_\varepsilon)/g'(\theta_c)$; $|\psi_c| \leq C_P$ for $c = 1, 2, 3$, $|\dot\theta_0| \leq C_P \varepsilon^3$ and $\|\dot\theta\|_1 \leq C_P \varepsilon$;

(iv) $H(\theta(\frac{3}{4})) - H(\theta(\frac{1}{4})) = h_* \varepsilon + O(\varepsilon^2)$, where $h_* := h(x(\frac{3}{4})) - h(x(\frac{1}{4}))$ and $h(x) := -\sum_c x_c \log x_c$; moreover

$$h_* = \int_0^\infty \log \frac{t^3 + t^2 + t/4 + 1/72}{t^3 + t^2 + t/4 + 1/216}\, \mathrm{d}t \;\geq\; \log \frac{489}{487} \;>\; 0,$$

and interval arithmetic gives

$$h_* \in [0.078447894650829996 46908,\ 0.078447894650829996 46909].$$

Proof. (i) Put $z = \varepsilon x$. Since $g(\varepsilon x) = -\varepsilon^3 x(x - \frac{1}{2})^2\big(1 - \varepsilon(1 + x)\big)$, the small roots solve

$$G(x, \varepsilon) := x(x - \tfrac{1}{2})^2\big(1 - \varepsilon(1 + x)\big) - u\, m_\varepsilon = 0.$$

At $\varepsilon = 0$ this is $\phi(x) := x(x - \frac{1}{2})^2 = u/54$. The cubic ϕ increases on $(0, 1/6)$, decreases on $(1/6, 1/2)$ and increases on $(1/2, \infty)$, with $\phi(1/6) = 1/54$ and $\phi(1/2) = 0$; so for $u \in (0, 1)$ it has

three simple roots $x_3 < 1/6 < x_2 < 1/2 < x_1$, and on the compact set $u \in [1/4, 3/4]$ their pairwise gaps and the values $|\phi'(x_c)|$ are bounded below by a positive absolute constant. By the analytic implicit function theorem, the three roots $x_c(u, \varepsilon)$ of G are analytic on a complex neighbourhood of $[1/4, 3/4] \times \{0\}$, and $\varepsilon x_c(u, \varepsilon)$ are three roots of $g = u\, g(r_\varepsilon)$. The four roots sum to 1 because the coefficient of z^3 is -1, so the fourth root is $1 - \varepsilon \sum_c x_c = 1 - O(\varepsilon)$, the largest one.

(ii) $\prod_c (z - \theta_c) = g(z) - u\, g(r_\varepsilon)$, so e_1, e_2, e_3 are coefficients of g and $e_4 = -u\, g(r_\varepsilon) = \varepsilon^3 m_\varepsilon u$. Newton's identities give $P_2 = e_1^2 - 2e_2$, $P_3 = e_1^3 - 3e_1 e_2 + 3e_3$ and $P_4 = e_1 P_3 - e_2 P_2 + e_3 P_1 - 4e_4$; the first two are constant, and P_4 changes by $-4\varepsilon^3 m_\varepsilon(\frac{3}{4} - \frac{1}{4})$. The three identities in (7) are $\dot P_1 = \dot P_2 = \dot P_3 = 0$.

(iii) Differentiating $g(\theta_c(u)) = u\, g(r_\varepsilon)$ gives $\dot\theta_c$. For $c \geq 1$, $\dot\theta_c = \varepsilon\, \partial_u x_c$ and $\psi_c = \partial_u x_c / x_c$, bounded by the uniform bounds of (i). From $\sum_c \theta_c^2 \dot\theta_c = 0$, $\theta_0^2 \dot\theta_0 = -\sum_{c\geq 1} \theta_c^2 \dot\theta_c = O(\varepsilon^3)$, and $\theta_0 \geq 1/2$. Finally $\|\dot\theta\|_1 \leq |\dot\theta_0| + \sum_{c\geq 1} \theta_c |\psi_c| \leq C_P \varepsilon$.

(iv) $H = -\theta_0 \log \theta_0 - \sum_{c\geq 1} \varepsilon x_c (\log \varepsilon + \log x_c)$. By (iii) the endpoint values of θ_0 differ by $O(\varepsilon^3)$, so $-\theta_0 \log \theta_0$ and $-\varepsilon \log \varepsilon \sum_{c\geq 1} x_c = -(1 - \theta_0) \log \varepsilon$ change by $O(\varepsilon^3 \log(1/\varepsilon))$; the remaining term is $\varepsilon h(x(u, \varepsilon)) = \varepsilon h(x(u)) + O(\varepsilon^2)$. By (i) the coordinates of $x(u)$ are positive, and since $\phi(x) - u/54 = x^3 - x^2 + x/4 - u/54$ they sum to 1 and satisfy $\prod_c (t + x_c(u)) = t^3 + t^2 + t/4 + u/54$. The integral representation in the proof of Lemma 29, applied to $(x(\frac{3}{4}), 0)$ and $(x(\frac{1}{4}), 0)$ in Θ_4, whose zero coordinates contribute the same factor t to both polynomials, gives the integral for h_*. Its integrand equals $\log\big(1 + \frac{1}{108}(t^3 + t^2 + t/4 + 1/216)^{-1}\big)$, which is positive and decreasing in t and equals $\log \frac{489}{487}$ at $t = 1$, so h_* is at least its integral over $[0, 1]$, which is at least $\log \frac{489}{487}$. For the enclosure of h_*, each root $x_c(1/4)$ and $x_c(3/4)$ is bracketed by an interval of width $2 \cdot 10^{-40}$ with rational endpoints, at which the cubic $\phi(x) - u/54$ takes opposite signs in exact rational arithmetic, and h is evaluated in interval arithmetic on the brackets. □

Table 4 in Appendix I reports the path numerically: along a grid of 101 values of u, $\max_{c\geq 1} |\psi_c|$ approaches 4.3658 and $\max_u |\psi_0|/\varepsilon^3$ approaches 0.0185 as ε decreases to 10^{-4}; at $\varepsilon = 10^{-4}$ the normalized entropy gap is 0.078445 against $h_* = 0.0784478946\ldots$, and the normalized P_4 gap is -0.037033 against the limit $-1/27$.

C Rewriting and sweeps

Fix the path of Appendix B and $u \in [1/4, 3/4]$. For $U \subseteq [n]$ let $\mathcal{P}_U f := \mathbb{E}_u[f \mid A, Z_{-U}]$, with $\mathcal{P}_\varnothing := \mathrm{Id}$, and for $S \subseteq [n]$ let $\Lambda_S := \sum_{U\subseteq S} (-1)^{|S|-|U|} \mathcal{P}_U$. The sweep of order r is $\widetilde S_{\mathrm{RB}r} := \sum_{|S|\leq r} \Lambda_S \widetilde S_1$, with $\widetilde S_1 = \sum_i \Phi(F_i)$ as in Lemma 11.

Lemma 16 (sweeps). *(i) For every integrable $f(A, Z)$ and every non-empty S, $\mathbb{E}_u[\Lambda_S f \mid A] = 0$. Consequently $\mathbb{E}_u[\widetilde S_{\mathrm{RB}r} \mid A] = S_{\mathrm{obs}}$, $\mathbb{E}_u \widetilde S_{\mathrm{RB}r} = 0$ and $I(u) \leq \mathbb{E}_u \widetilde S^2_{\mathrm{RB}r}$.*

(ii) $\sum_{S\subseteq[n]} \Lambda_S = \mathcal{P}_{[n]}$; in particular $\sum_{S\subseteq[n]} \Lambda_S \widetilde S_1 = S_{\mathrm{obs}}$.

Proof. (i) By the tower property $\mathbb{E}_u[\mathcal{P}_U f \mid A] = \mathbb{E}_u[f \mid A]$ for every U, so $\mathbb{E}_u[\Lambda_S f \mid A] = \mathbb{E}_u[f \mid A] \sum_{U\subseteq S} (-1)^{|S|-|U|} = 0$ for $S \neq \varnothing$. With Lemma 11, $\mathbb{E}_u[\widetilde S_{\mathrm{RB}r} \mid A] = \mathbb{E}_u[\widetilde S_1 \mid A] = S_{\mathrm{obs}}$, and conditional Jensen gives $S^2_{\mathrm{obs}} \leq \mathbb{E}_u[\widetilde S^2_{\mathrm{RB}r} \mid A]$; also $\mathbb{E}_u S_{\mathrm{obs}} = 0$. (ii) $\sum_S \sum_{U\subseteq S} (-1)^{|S\setminus U|} \mathcal{P}_U = \sum_U \mathcal{P}_U \sum_{S\supseteq U} (-1)^{|S\setminus U|} = \mathcal{P}_{[n]}$, since the inner sum vanishes unless $U = [n]$; and $\mathcal{P}_{[n]} \widetilde S_1 = \mathbb{E}_u[\widetilde S_1 \mid A] = S_{\mathrm{obs}}$. □

The sweep as an analytic function of the edge weights. Fix a labelling $z \in [4]^n$. The operators $\mathcal{P}_U$ depend on A only through J. For a real vector $J = (J_{kl})_{k<l}$ let

$$w(z) := \prod_v \theta_{z_v} \prod_{k<l} e^{J_{kl} W_{kl}(z)}, \qquad \mathcal{P}_U f(z) := \frac{\sum_{z'_U} w(z'_U, z_{-U})\, f(z'_U, z_{-U})}{\sum_{z'_U} w(z'_U, z_{-U})},$$

which at $J = J(A)$ is the conditional expectation by Lemma 10. With $\widetilde{S}_1$ written as the function $\sum_i \Phi(\sum_{j\neq i} J_{ij}\mathbf{e}_{z_j})$ of (J, z), the sweep becomes $\widetilde{S}_{\mathrm{RB}r} = \mathcal{S}_r(J; z)$ for an explicit function $\mathcal{S}_r$ that is a finite combination of ratios of exponential polynomials in J. All denominators are positive at $J = 0$, so $\mathcal{S}_r$ is analytic on a complex neighbourhood of $J = 0$. For an edge multiset Γ with multiplicities m_e we write $J^\Gamma := \prod_e J_e^{m_e}$, $\Gamma! := \prod_e m_e!$, $|\Gamma| := \sum_e m_e$, $V(\Gamma)$ for the set of vertices touched by Γ and $v(\Gamma) := |V(\Gamma)|$, and

$$B_\Gamma^{(r)}(z) := [J^\Gamma]\mathcal{S}_r := \frac{1}{\Gamma!}\,\partial^\Gamma \mathcal{S}_r(J; z)\Big|_{J=0}, \qquad B_\Gamma := B_\Gamma^{(3)}.$$

The same notation $[J^\Gamma]$ is used for the coefficient of J^Γ in any function of J that is analytic near 0.

Motif systems. For a finite vertex set V, the motif system on V is the same construction with only the vertices of V: the weight $w_V(z) := \prod_{v\in V}\theta_{z_v}\prod_{\{k,l\}\subseteq V} e^{J_{kl}W_{kl}}$, the fields, Φ, the operators $\mathcal{P}_U^{(V)}$ for $U \subseteq V$, their inclusion–exclusion combinations $\Lambda_S^{(V)}$, and $\widetilde{S}_1^{(V)} := \sum_{i\in V}\Phi(F_i^{(V)})$, all use only the edges inside V.

D THE MOTIF LEMMA

This appendix proves Proposition 5. We work with the sweep of a general order r where this costs nothing, and write $B_\Gamma^{(r)}$ for the coefficient of J^Γ in $\widetilde{S}_{\mathrm{RB}r}$ (Appendix C).

D.1 LOCALITY

Lemma 17 (locality)**.** *Let Γ be an edge multiset, $V := V(\Gamma)$, $z \in [4]^n$ and $S \subseteq [n]$.*

(i) If $S \not\subseteq V$, then $[J^\Gamma]\Lambda_S\widetilde{S}_1 = 0$.

(ii) If $S \subseteq V$, then $[J^\Gamma]\Lambda_S\widetilde{S}_1(z) = [J^\Gamma]\Lambda_S^{(V)}\widetilde{S}_1^{(V)}(z_V)$.

(iii) Hence $B_\Gamma^{(r)}(z)$ depends only on z_V, and $B_\Gamma^{(r)} = [J^\Gamma]\sum_{S\subseteq V,\,|S|\le r}\Lambda_S^{(V)}\widetilde{S}_1^{(V)}$.

(iv) If Γ, as a multigraph, is disconnected, then $B_\Gamma^{(r)} = 0$.

Proof. The coefficient $[J^\Gamma]$ depends only on the restriction of a function to the subspace $\{J_e = 0,\ e \notin \Gamma\}$, and we compute there. (i) Take $w \in S \setminus V$. On the subspace every edge weight at w vanishes, so $\widetilde{S}_1$ does not depend on z_w (the fields of the other vertices give z_w the weight $J_{iw} = 0$, and $\Phi(F_w) = \Phi(0) = 0$), and in the posterior weight (6) of $Z_{U\cup\{w\}}$ the coordinate z_w is not coupled to the others. Hence $\mathcal{P}_{U\cup\{w\}}\widetilde{S}_1 = \mathcal{P}_U\widetilde{S}_1$ for every $U \subseteq S\setminus\{w\}$, and $\Lambda_S = \sum_{U\subseteq S\setminus\{w\}}(-1)^{|S\setminus\{w\}|-|U|}(\mathcal{P}_{U\cup\{w\}} - \mathcal{P}_U)$ annihilates $\widetilde{S}_1$. (ii) On the subspace the vertices outside V are isolated: their terms Φ vanish, and the fields and posteriors of the vertices of V involve only edges inside V. (iii) follows from (i) and (ii). (iv) Let Γ split into non-empty vertex-disjoint parts Γ_1, Γ_2. On the subspace there is no edge between $V(\Gamma_1)$ and $V(\Gamma_2)$. For $i \in V(\Gamma_1)$, $\Phi(F_i)$ contains no weight of Γ_2, and the posterior of Z_U factorizes between the two parts, so $\mathcal{P}_U\Phi(F_i) = \mathcal{P}_{U\cap V(\Gamma_1)}\Phi(F_i)$ contains no weight of Γ_2; the case $i \in V(\Gamma_2)$ is symmetric. So no term contains the monomial J^Γ. □

D.2 THE EXACT IDENTITY

Let κ_Γ be the joint cumulant of $\{W_e\}_{e\in\Gamma}$, each W_e repeated m_e times, under $Z \sim \theta^{\otimes n}$; equivalently

$$\log \mathbb{E}_\theta \exp\Big(\sum_e J_e W_e\Big) = \sum_\Gamma \frac{J^\Gamma}{\Gamma!}\,\kappa_\Gamma.$$

Lemma 18 (exact identity)**.** *For every edge multiset Γ, $V := V(\Gamma)$ and z,*

$$B_\Gamma^{(r)}(z) = \frac{\partial_u \kappa_\Gamma}{\Gamma!} - \sum_{S\subseteq V,\ |S|>r}[J^\Gamma]\big(\Lambda_S^{(V)}\widetilde{S}_1^{(V)}\big)(z). \tag{8}$$

Proof. On the motif system, Lemma 16(ii) gives $\sum_{S\subseteq V}\Lambda_S^{(V)}=\mathcal{P}_V^{(V)}$, so by Lemma 17(iii) $B_\Gamma^{(r)}=[J^\Gamma]\mathcal{P}_V^{(V)}\widetilde{S}_1^{(V)}-\sum_{|S|>r}[J^\Gamma]\Lambda_S^{(V)}\widetilde{S}_1^{(V)}$. The operator $\mathcal{P}_V^{(V)}$ averages all labels of V with the weight w_V. The Glauber identity is algebraic: as a function of z_i, $w_V(z)=\theta_{z_i}e^{F_{i,z_i}}$ times a factor that does not depend on z_i, so $\sum_{z_i}w_V\,\psi_{z_i}=\Phi(F_i)\sum_{z_i}w_V$. Therefore

$$\mathcal{P}_V^{(V)}\widetilde{S}_1^{(V)}=\frac{\sum_z w_V(z)\sum_{i\in V}\psi_{z_i}}{\sum_z w_V(z)}=\partial_u\log\sum_z w_V(z)=\partial_u\log\mathbb{E}_\theta\exp\Big(\sum_{\{k,l\}\subseteq V}J_{kl}W_{kl}\Big),$$

using $\partial_u w_V=w_V\sum_{i\in V}\psi_{z_i}$. The function $\log\sum_z w_V(z)$ is jointly analytic in (u,J) for J near 0, because the path is analytic in u (proof of Lemma 15(i)) and $\sum_z w_V=1$ at $J=0$; so $[J^\Gamma]$ commutes with ∂_u, and taking $[J^\Gamma]$ gives $\partial_u\kappa_\Gamma/\Gamma!$. □

Corollary 19. *(i) If $v(\Gamma)\le 3$ and $r\ge 3$, then $B_\Gamma^{(r)}\equiv 0$; in particular $B_\Gamma\equiv 0$.*

(ii) $|\partial_u\kappa_\Gamma|\le C_\Gamma\varepsilon^3$ for every Γ.

Proof. A joint cumulant is a polynomial in the joint moments of sub-multisets, and $\mathbb{E}\prod_{e\in\Gamma'}W_e=\prod_Q P_{v(Q)}$, the product running over the connected components Q of Γ', since all vertices of a component must share a class. So κ_Γ is a polynomial in $P_2,\dots,P_{v(\Gamma)}$ with coefficients depending only on Γ. (i) For $v(\Gamma)\le 3$ only P_2,P_3 occur, whose derivatives vanish by (7), and (8) has no term with $|S|>r\ge 3$. (ii) $\partial_u\kappa_\Gamma=\sum_{k\ge 4}(\partial\kappa_\Gamma/\partial P_k)\dot{P}_k$ and, for $k\ge 4$, $|\dot{P}_k|=k|\theta_0^{k-1}\dot{\theta}_0+\sum_{c\ge 1}\theta_c^{k-1}\dot{\theta}_c|\le C_k\varepsilon^3$ by Lemma 15(iii). □

D.3 Majority vertices

Lemma 20 (one factor ε per majority vertex)**.** *Fix a vertex set V, a subset $S\subseteq V$ and $z\in[4]^V$, and let $M:=\{k\in S: z_k=0\}$. There are $r_0>0$, C and ε_0, depending only on $|V|$, such that for $\varepsilon\le\varepsilon_0$ and every complex J in the polydisc $\max_e|J_e|\le r_0$,*

$$\big|\Lambda_S^{(V)}\widetilde{S}_1^{(V)}(z;J)\big|\;\le\;C\,\varepsilon^{1+|M|}.$$

By the Cauchy estimates, $|[J^\Gamma]\Lambda_S^{(V)}\widetilde{S}_1^{(V)}(z)|\le C\,r_0^{-|\Gamma|}\varepsilon^{1+|M|}$ for every edge multiset Γ on V.

Proof. All expressions below are rational functions of $t_e:=e^{J_e}$, holomorphic in J where their denominators do not vanish; identities that hold for small real J, where the weights form probability measures, extend to the polydisc by analytic continuation. Choose r_0 so small that on the polydisc every normalized partition function $\sum_{z'_U}\prod_{k\in U}\theta_{z'_k}e^{(\cdots)}$, which equals 1 at $J=0$, has modulus at least $1/2$. Then: (a) every posterior "probability" has modulus at most C; (b) the posterior probability that all vertices of a set $R\subseteq U$ carry rare classes has modulus at most $(C\varepsilon)^{|R|}$, since its numerator carries $\prod_{k\in R}\theta_{z'_k}$ and $\sum_{c\ge 1}\theta_c\le C\varepsilon$; (c) $\widetilde{S}_1^{(V)}$ is linear in $\dot{\theta}$ and has modulus at most $C\varepsilon$, because $|\Phi(F)|\le\|\dot{\theta}\|_1\max_c|e^{F_c}|/|\sum_b\theta_b e^{F_b}|$ and $\|\dot{\theta}\|_1\le C\varepsilon$.

Let $T:=S\setminus M$ and write $U\subseteq S$ as $U_T\cup U_M$ with $U_T\subseteq T$, $U_M\subseteq M$:

$$\Lambda_S\widetilde{S}_1(z)=\sum_{U_T\subseteq T}(-1)^{|T\setminus U_T|}\sum_{U_M\subseteq M}(-1)^{|M\setminus U_M|}\mathcal{P}_{U_T\cup U_M}\widetilde{S}_1(z).$$

Fix U_T, let $g:=\mathcal{P}_{U_T}\widetilde{S}_1$, a function of the labels outside U_T with $|g|\le C\varepsilon$, and let μ be the conditional law of $Y:=Z_M$ when the labels outside $U_T\cup M$ are fixed at z and Z_{U_T} is integrated out. By the tower property and $z_M=0$, $\mathcal{P}_{U_T\cup U_M}\widetilde{S}_1(z)=\mathbb{E}_\mu[g(Y)\mid Y_{M\setminus U_M}=0]$. Decompose g by resetting coordinates of M to 0:

$$g(y)=\sum_{R\subseteq M}g_R(y_R),\qquad g_R(y_R):=\sum_{R'\subseteq R}(-1)^{|R\setminus R'|}g(y_{R'},0_{M\setminus R'});$$

g_R vanishes as soon as $y_k = 0$ for some $k \in R$, and $|g_R| \le 2^{|R|}C\varepsilon$. If $R \not\subseteq U_M$, the condition $Y_{M\setminus U_M} = 0$ forces $g_R = 0$. Writing $C_0 := M \setminus U_M$ and $Q := M \setminus R$, the inner sum over U_M becomes

$$\sum_{R\subseteq M}\ \sum_{C_0\subseteq Q} (-1)^{|C_0|}\varphi_R(C_0), \qquad \varphi_R(C_0) := \mathbb{E}_\mu\big[g_R(Y_R) \mid Y_{C_0} = 0\big] = \frac{N(C_0)}{D(C_0)},$$

with $N(C_0) := \mathbb{E}_\mu[g_R \prod_{k\in C_0}(1-\eta_k)]$, $D(C_0) := \mathbb{E}_\mu[\prod_{k\in C_0}(1-\eta_k)]$ and $\eta_k := \mathbf{1}\{Y_k \neq 0\}$. Expanding the products, $N(C_0) = \sum_{C'\subseteq C_0}(-1)^{|C'|}a_{C'}$ and $D(C_0) = 1 - \sum_{\varnothing\neq C'\subseteq C_0}(-1)^{|C'|+1}b_{C'}$, where $a_{C'} := \mathbb{E}_\mu[g_R\prod_{k\in C'}\eta_k]$ and $b_{C'} := \mathbb{E}_\mu[\prod_{k\in C'}\eta_k]$. By (b), $|a_{C'}| \le C\varepsilon\,(C\varepsilon)^{|R|+|C'|}$, since $g_R \neq 0$ requires Y_R to be rare, and $|b_{C'}| \le (C\varepsilon)^{|C'|}$. For small ε, $|1 - D(C_0)| \le 1/2$, so $\varphi_R(C_0) = N(C_0)\sum_{j\ge 0}(1-D(C_0))^j$. Expanded, this is a sum of products $\pm a_{C'_0}b_{C'_1}\cdots b_{C'_j}$, $j \ge 0$, over $C'_0 \subseteq C_0$ and non-empty $C'_1, \dots, C'_j \subseteq C_0$, with signs that do not depend on C_0. There are at most $2^{|Q|(j+1)}$ such tuples of subsets of Q for each j, and each product has modulus at most $C\varepsilon(C\varepsilon)^{|R|}(C\varepsilon)^{\sum_l |C'_l|}$ with $\sum_l |C'_l| \ge j$; so once $2^{|Q|}C\varepsilon \le 1/2$ the expansion converges absolutely, uniformly on the polydisc, and the alternating sum over $C_0 \subseteq Q$ may be taken product by product. A product survives it only if $C'_0 \cup \cdots \cup C'_j = Q$, because $\sum_{A\subseteq C_0\subseteq Q}(-1)^{|C_0|} = 0$ for $A \subsetneq Q$; then $\sum_l |C'_l| \ge \max\{|Q|, j\}$, so a surviving product has modulus at most $C\varepsilon(C\varepsilon)^{|R|+|Q|}(C\varepsilon)^{(j-|Q|)_+}$, and summing over j and the tuples gives at most $C\varepsilon(C\varepsilon)^{|R|+|Q|}$. As $|R| + |Q| = |M|$, the inner sum has modulus at most $C\varepsilon^{1+|M|}$, and the sum over U_T costs only a constant factor. □

D.4 Weighted second moments

Lemma 21 (weighted coefficient bound)**.** *Let Γ be connected with $v(\Gamma) \ge 4$. For $u \in [1/4, 3/4]$ and $\varepsilon \le \varepsilon_\Gamma$, $\mathbb{E}_\theta B_\Gamma(Z)^2 \le C_\Gamma\varepsilon^6$. More precisely, if z carries exactly m rare labels on $V(\Gamma)$, then $|B_\Gamma(z) - \partial_u\kappa_\Gamma/\Gamma!| \le C_\Gamma\varepsilon^{1+(4-m)_+}$.*

Proof. By (8) with $r = 3$, the difference is a sum over $S \subseteq V(\Gamma)$ with $|S| \ge 4$; each such S contains at least $|S| - m \ge 4 - m$ majority vertices, and Lemma 20 gives the second claim. Since a labelling with m rare labels has probability at most $C_\Gamma\varepsilon^m$,

$$\mathbb{E}\big(B_\Gamma - \partial_u\kappa_\Gamma/\Gamma!\big)^2 \le C_\Gamma \sum_{m=0}^{v(\Gamma)} \varepsilon^{m+2+2(4-m)_+} \le C_\Gamma\varepsilon^6,$$

the exponent being $10 - m \ge 6$ for $m \le 4$ and $m + 2 \ge 6$ for $m \ge 4$. Add Corollary 19(ii). □

Proof of Proposition 5. (i) is Lemma 17(iii)–(iv) with $r = 3$; (ii) is Lemma 18 with $r = 3$; (iii) is Corollary 19; (iv) is Lemma 20 applied to the motif system on $V(\Gamma)$, whose constants depend only on Γ; and (v) is Lemma 21. □

Remark 22 (sweep order). Replacing $|S| \le 3$ by $|S| \le r$, the same count gives $\mathbb{E}\big(B_\Gamma^{(r)} - \partial_u\kappa_\Gamma/\Gamma!\big)^2 \le C_\Gamma\varepsilon^{r+3}$ for every connected Γ: each S with $|S| > r$ contains at least $r + 1 - m$ majority vertices, and $m + 2 + 2(r + 1 - m)_+ \ge r + 3$ for every m. With Corollary 19(ii), $\mathbb{E}(B_\Gamma^{(r)})^2 \le C_\Gamma(\varepsilon^6 + \varepsilon^{r+3})$, that is, at most of order ε^4, ε^5 and ε^6 for $r = 1, 2, 3$.

D.5 Lower-order sweeps

For the swept score of this construction, the next remark shows that the orders in Remark 22 are attained for $r = 1, 2$ and that the three-star keeps order ε^6 under the third-order sweep.

Remark 23 (explicit coefficients). Write $h_{cd} := \partial_c\partial_d\Phi(0) = \mathbf{1}\{c = d\}\dot\theta_c - \dot\theta_c\theta_d - \theta_c\dot\theta_d$ (Lemma 14), and $x_c, \dot x_c := \partial_u x_c$ for the leading-order path of Lemma 15(i).

(a) *Wedge, first order.* For the wedge $\Gamma = \{ij, ik\}$ and every labelling,

$$B_\Gamma^{(1)}(z) = -h_{z_jz_k} + 2\theta_{z_j}\dot\theta_{z_j} + 2\theta_{z_k}\dot\theta_{z_k} - P_2\big(\dot\theta_{z_j} + \dot\theta_{z_k}\big). \tag{9}$$

Consequently $\mathbb{E}B_\Gamma^{(1)}(Z)^2 = L_1(u)\varepsilon^4 + O(\varepsilon^5)$ with $L_1(u) := \sum_{a,b\geq 1} x_a x_b\big(\mathbf{1}\{a=b\}\dot{x}_a + \dot{x}_a + \dot{x}_b\big)^2 > 0$.

(b) *Three-star, first and second order.* For the three-star with centre i and leaves k, l, m, and a rare class a,

$$B^{(1)}(0,a,a,0) = 3\dot{\theta}_a + R_1, \qquad B^{(2)}(0,a,a,a) = 7\dot{\theta}_a + R_2,$$

the arguments listing the labels of (i,k,l,m), where R_1, R_2 are polynomials in $(\theta_b, \dot{\theta}_b)_{b\geq 1}$ and $\dot{\theta}_0$ (with $\theta_0 = 1 - \sum_{b\geq 1}\theta_b$) that contain no monomial of total degree below two. Hence $R_1, R_2 = O(\varepsilon^2)$ on the path, and since these labellings have probabilities $\theta_0^2\theta_a^2 \asymp \varepsilon^2$ and $\theta_0\theta_a^3 \asymp \varepsilon^3$, $\mathbb{E}(B^{(1)})^2 \geq c\varepsilon^4$ and $\mathbb{E}(B^{(2)})^2 \geq c\varepsilon^5$.

(c) *Three-star, third order.* $\mathbb{E}B_\Gamma = \dot{P}_4 + O(\varepsilon^5)$ for the three-star, so $\mathbb{E}B_\Gamma^2 \geq (\mathbb{E}B_\Gamma)^2 = \dot{P}_4^2(1 + O(\varepsilon^2)) \asymp \varepsilon^6$.

The constants c and those implicit in $O(\cdot)$ and $\asymp$ do not depend on $u \in [1/4, 3/4]$.

Proof. (a) By Lemma 17 the coefficient is computed on the motif system on $\{i,j,k\}$ with $J_{jk} = 0$, and the first-order sweep there is $\widetilde{S}_1 + \sum_{v\in\{i,j,k\}}(\mathcal{P}_v - \mathrm{Id})\widetilde{S}_1$. Only $\Phi(F_i)$ contains $J_{ij}J_{ik}$, with coefficient $h_{z_j z_k}$. In $\mathcal{P}_j\widetilde{S}_1$ the posterior weight of z_j' depends only on J_{ij}; the coefficient of $J_{ij}J_{ik}$ collects the prior average $\sum_c \theta_c h_{cz_k} = \theta_{z_k}\dot{\theta}_{z_k} - P_2\dot{\theta}_{z_k}$ (using $\sum_c \theta_c\dot{\theta}_c = 0$) and the product of the first-order terms of the weight and of $[J_{ik}]\widetilde{S}_1 = \dot{\theta}_{z_k} + \dot{\theta}_{z_i}$; the latter vanishes because $[J_{ik}]\widetilde{S}_1$ does not depend on z_j' while the first-order terms of the normalized weights sum to zero. The case of $\mathcal{P}_k$ is symmetric. In $\mathcal{P}_i\widetilde{S}_1$ the weight of z_i' depends on J_{ij} and J_{ik}: the prior average of $h_{z_j z_k}$, which does not depend on z_i', returns $h_{z_j z_k}$, and the two cross terms give $\sum_c \theta_c(\mathbf{1}\{c = z_j\} - \theta_{z_j})\dot{\theta}_c = \theta_{z_j}\dot{\theta}_{z_j}$ and the same with z_k, again by $\sum_c \theta_c\dot{\theta}_c = 0$. Adding the four contributions gives (9). On the path, (9) equals $-3\dot{\theta}_a + O(\varepsilon^2)$ if $z_j = z_k = a \geq 1$, $-(\dot{\theta}_a + \dot{\theta}_b) + O(\varepsilon^2)$ if $z_j = a \neq b = z_k$ are rare, $O(\varepsilon^2)$ if exactly one of z_j, z_k is 0 and $O(\varepsilon^3)$ if both are. With $\theta_a = \varepsilon x_a(1 + O(\varepsilon))$ and $\dot{\theta}_a = \varepsilon\dot{x}_a(1 + O(\varepsilon))$, the labellings with two rare leaves give $L_1(u)\varepsilon^4$ and the others $O(\varepsilon^5)$. $L_1(u) > 0$ because its diagonal terms $9x_a^2\dot{x}_a^2$ are positive.

(b) The coefficients are exact polynomials in $(\theta, \dot{\theta})$: each $\mathcal{P}_U$ is a ratio whose denominator is 1 at $J = 0$, so every Taylor coefficient is a polynomial. We computed $B^{(1)}(0,1,1,0)$ and $B^{(2)}(0,1,1,1)$ symbolically by differentiating the exact motif-system expressions (Appendix I); their constant parts vanish and their parts of degree one are $3\dot{\theta}_1$ and $7\dot{\theta}_1$, and relabelling the rare classes gives the statement for every a. The leading factors have a transparent origin: at leading order a posterior average resets a label to 0, the coefficient of $J_{ik}J_{il}J_{im}$ in $\Phi(F_i)$ at a majority centre with j leaves of class a and the others majority equals $\dot{\theta}_a$, $-\dot{\theta}_a$, $\dot{\theta}_a$ for $j = 3, 2, 1$ and vanishes at leading order for $j = 0$, and inclusion–exclusion over the resets gives $2\dot{\theta}_a - (-\dot{\theta}_a) = 3\dot{\theta}_a$ for $(0,a,a,0)$ under one sweep and $3\dot{\theta}_a - 2(-2\dot{\theta}_a) + 0 = 7\dot{\theta}_a$ for $(0,a,a,a)$ under two.

(c) For the three-star, $\kappa_\Gamma = P_4 - 3P_2P_3 + 2P_2^3$, so $\partial_u\kappa_\Gamma = \dot{P}_4$. By (8) with $r = 3$, $\mathbb{E}B_\Gamma = \dot{P}_4 - \mathbb{E}[J^\Gamma]\Lambda_V\widetilde{S}_1$, and by Lemma 20 a labelling with m rare labels contributes at most $C\varepsilon^m \cdot \varepsilon^{1+(4-m)} = C\varepsilon^5$. Finally $\dot{P}_4 = -4\varepsilon^3 m_\varepsilon$ by Lemma 15(ii). The constants are uniform in $u \in [1/4, 3/4]$ because x_a and $\dot{x}_a$ are continuous and bounded away from zero there, and the path bounds of Lemma 15 are uniform in u. □

Table 3 confirms (a) and (b) numerically: the closed form (9) agrees with the exact coefficient on all 64 labellings to within 9×10^{-18}, $\mathbb{E}B^2/\varepsilon^4$ approaches $L_1(u)$, and $B^{(2)}(0,a,a,a)/\dot{\theta}_a$ approaches 7. The computed mean of the third-order three-star coefficient equals $\dot{P}_4$ to all printed digits.

E THE INFORMATION BOUND

Lemma 24 (information bound, full form)**.** *For every integer $D \geq 15$ there are C_D, $\varepsilon_D > 0$ and n_D, depending on q_0 and D, such that if $n \geq n_D$, $q \in [q_0, 1-q_0]$, $\delta \leq 1$, $\sqrt{\delta}\log n \leq 1/C$ and $\varepsilon \leq \varepsilon_D$,*

then for every $u \in [1/4, 3/4]$

$$\mathbb{E}_u \widetilde{S}^2_{\mathrm{RB3}} \;\le\; C_D\Big[\varepsilon^6 n\delta^3 + \varepsilon^2 n^5 \delta^3 \big(\delta \log^2 n\big)^{D-2} + n^{4D+2} e^{-q_0^2 \log^2 n/50}\Big]. \tag{10}$$

Here $n\delta^3 = n^4\Delta^6$, and $\Delta = \sqrt{\delta/n} \le q_0/2$ once $n \ge 4/q_0^2$, so the standing assumptions of Appendix A hold. Fix z and write $\widetilde{S}_{\mathrm{RB3}} = \mathcal{S}_3(J; z)$ as its Taylor polynomial of degree D plus a remainder,

$$\widetilde{S}_{\mathrm{RB3}} = \mathrm{Poly}_D + \mathrm{Rem}_D, \qquad \mathrm{Poly}_D := \sum_{|\Gamma| \le D} J^\Gamma B_\Gamma(Z),$$

the sum running over the edge multisets Γ of the complete graph with $|\Gamma| \le D$.

E.1 The polynomial part

Lemma 25 (pairing)**.** $\mathrm{Var}(\mathrm{Poly}_D) \le C_D\, \varepsilon^6 n\delta^3$.

Proof. Write $X_\Gamma := J^\Gamma B_\Gamma(Z)$. By Lemma 17(iv) and Corollary 19(i), only connected Γ with $v(\Gamma) \ge 4$ contribute. X_Γ is a function of $(Z_{V(\Gamma)}, A_{E(\Gamma)})$, and each A_e depends only on the labels of its endpoints and on independent edge noise, so vertex-disjoint Γ, Γ' give independent $X_\Gamma, X_{\Gamma'}$ and contribute no covariance. For Γ, Γ' sharing a vertex, $K := \Gamma \uplus \Gamma'$, with multiplicities $a_e := m_e + m'_e$, is connected. Given Z the edges are independent, so by (5) and Cauchy–Schwarz,

$$\big|\mathbb{E}[X_\Gamma X_{\Gamma'}]\big| = \Big|\mathbb{E}_Z\Big[B_\Gamma B_{\Gamma'} \prod_e \mathbb{E}[J_e^{a_e} \mid Z]\Big]\Big| \le \Big(\frac{2\Delta}{q_0}\Big)^{\sum_e c(a_e)} \|B_\Gamma\|_2 \|B_{\Gamma'}\|_2,$$

and likewise $|\mathbb{E}X_\Gamma\, \mathbb{E}X_{\Gamma'}| \le (2\Delta/q_0)^{\sum_e c(m_e)+\sum_e c(m'_e)} \|B_\Gamma\|_2 \|B_{\Gamma'}\|_2$, where $c(m) + c(m') \ge c(m+m')$. By Lemma 21, $\|B_\Gamma\|_2 \|B_{\Gamma'}\|_2 \le C_\Gamma C_{\Gamma'} \varepsilon^6$. Since $c(a) \ge 2$ for all $a \ge 1$ and $2\Delta/q_0 \le 1$, both covariance terms are at most $C_D \varepsilon^6 \Delta^{2d(K)}$, where $d(K)$ is the number of distinct edges of K. For given shapes of Γ, Γ' and a given overlap pattern, there are at most $n^{v(K)}$ placements in the complete graph; K is connected, so $v(K) \le d(K) + 1$ and

$$n^{v(K)} \Delta^{2d(K)} = n^{v(K)-d(K)} \delta^{d(K)} \le n\, \delta^{d(K)} \le n\delta^3,$$

because $v(K) \ge 4$ forces $d(K) \ge 3$ and $\delta \le 1$. There are finitely many shapes and overlap patterns, their number depending on D. □

E.2 Local decomposition

For $i \in [n]$ and $S \subseteq [n]$ with $|S| \le 3$, let $L := \{i\} \cup S$, let E_L be the set of pairs inside L, and define the external fields $F_v^{\mathrm{ext}} := \sum_{m \notin L} J_{vm} \mathbf{e}_{z_m} \in \mathbb{R}^4$ for $v \in L$. By linearity,

$$\widetilde{S}_{\mathrm{RB3}} = \sum_{i=1}^{n} \sum_{|S| \le 3} \Psi_{i,S}, \qquad \Psi_{i,S} := \Lambda_S \Phi(F_i) = \psi_{i,S}(x, f; z_L), \tag{11}$$

with $x := J_{E_L}$ and $f := (F_v^{\mathrm{ext}})_{v \in L}$, where $\psi_{i,S}$ is explicit: for $U \subseteq S$, the conditional weight of Z_U is proportional to

$$\Big(\prod_{k \in U} \theta_{z'_k}\Big) \exp\Big[\sum_{k \in U} F^{\mathrm{ext}}_{k,z'_k} + \sum_{k \in U,\ l \in L \setminus U} J_{kl} \mathbf{1}\{z'_k = z_l\} + \sum_{k<l,\ k,l \in U} J_{kl} \mathbf{1}\{z'_k = z'_l\}\Big],$$

and $F_i = F_i^{\mathrm{ext}} + \sum_{k \in L \setminus \{i\}} J_{ik} \mathbf{e}_{z_k}$. The variables x and f are separate arguments of $\psi_{i,S}$, and f is linear in J.

Lemma 26 (connected decomposition)**.** *If the graph* $(L, \mathrm{supp}\, x)$ *is disconnected, then* $\psi_{i,S}(x, f) = 0$. *Consequently*

$$\psi_{i,S}(x, f) = \sum_{B \subseteq E_L:\ (L,B)\ \textit{connected}} x^B\, h_B(x_B, f),$$

$$h_B(x_B, f) := \int_{[0,1]^B} \partial_{x_B} \psi_{i,S}\big(t \odot x_B, 0_{E_L \setminus B}, f\big)\, \mathrm{d}t.$$

Proof. Let L split into non-empty parts $L_1 \ni i$ and L_2 with no edge of $\operatorname{supp} x$ between them. Then $\Phi(F_i)$ does not depend on z_{L_2}, and the conditional weight of Z_U factorizes between L_1 and L_2: the local weights between the parts vanish, and the external fields are fixed vectors that do not couple labels. Hence $\mathcal{P}_U\Phi(F_i) = \mathcal{P}_{U\cap L_1}\Phi(F_i)$ and

$$\Lambda_S\Phi(F_i) = \sum_{U_1\subseteq S\cap L_1} (-1)^{|S\cap L_1|-|U_1|}\mathcal{P}_{U_1}\Phi(F_i) \sum_{U_2\subseteq S\cap L_2} (-1)^{|S\cap L_2|-|U_2|} = 0,$$

since $S\cap L_2 = L_2 \neq \varnothing$. For the decomposition, Möbius inversion in the coordinates of x gives $\psi = \sum_{B\subseteq E_L} g_B(x_B)$ with $g_B := \sum_{B'\subseteq B}(-1)^{|B\setminus B'|}\psi(x_{B'}, 0, f)$. If (L, B) is disconnected, so is (L, B') for every $B'\subseteq B$, and $g_B = 0$. If it is connected, g_B vanishes whenever $x_e = 0$ for some $e \in B$, so applying the fundamental theorem of calculus in each coordinate of B gives $g_B = x^B h_B$, because $\partial_{x_B} g_B = \partial_{x_B}\psi(\cdot, 0, f)$. □

Derivative bounds. $\psi_{i,S}$ is a signed sum of at most 8 functions $\mathcal{P}_U\Phi(F_i)$, $U\subseteq S$. Each is a sum over z'_U of normalized exponential weights times Φ evaluated at $F_i^{\mathrm{ext}} + \sum_{k\in L\setminus\{i\}} x_{ik}\mathbf{e}_{z''_k}$, where z''_k is z'_k or z_k. The normalized weights are softmax functions of logits that are affine in (x, f) with bounded coefficients, so all their derivatives are bounded by constants depending only on the order. By Lemma 14 the derivatives of Φ of order m are at most $C_m\|\dot\theta\|_1 e^{2g}$ with $g \le \|F_i^{\mathrm{ext}}\|_\infty + |x|_1$, and $\|\dot\theta\|_1 \le C\varepsilon$. By the Leibniz and chain rules, for every multi-index α,

$$\left|\partial^\alpha\psi_{i,S}(x, f)\right| \;\le\; C_{|\alpha|}\,\varepsilon\,\exp\Big(2\big(\|F_i^{\mathrm{ext}}\|_\infty + |x|_1\big)\Big), \tag{12}$$

and h_B obeys the same bound with $C_{|\alpha|+|B|}$ in place of $C_{|\alpha|}$, by differentiating under the integral, whose integrand is smooth on the compact cube $[0, 1]^B$. Separately, $|\Phi| \le \max_c|\psi_c| \le C_P$ everywhere, since Φ is an average of ψ under a probability vector, so $|\Psi_{i,S}| \le 8C_P$.

E.3 The remainder

Because f is linear in J, the Taylor polynomial of degree D of $\Psi_{i,S}$ in J is the Taylor polynomial of degree D of $\psi_{i,S}$ in $y := (x, f)$, evaluated at $y = y(J)$; and the Taylor polynomial of $\mathcal{S}_3$ is the sum of those of the $\Psi_{i,S}$. By Lemma 26 and $T_D(x^B h) = x^B T_{D-|B|}h$, where T_k denotes the Taylor polynomial of degree k,

$$\mathrm{Rem}_D = \sum_i \sum_{|S|\le 3} \sum_{B\text{ connected}} x^B\big(h_B - T_{D-|B|}h_B\big)(y), \tag{13}$$

and Taylor's formula with integral remainder together with (12) gives $|h_B - T_k h_B|(y) \le C_D\,\varepsilon\,|y|_1^{k+1}\sup_{t\in[0,1]}\exp\big(2(\|f\|_\infty + |x|_1)\big)|_{ty}$.

The small-field event. Let $\mathcal{G} := \{\max_{v,c}|F_{vc}| \le \sqrt{\delta}\log n\}$. Given Z, F_{vc} is a sum of at most n conditionally independent terms; after centring, each is bounded by $4\Delta/q_0$, the conditional variances sum to at most $3\delta/q_0^2$, and the conditional mean has modulus at most $2\delta/q_0^2$ (Lemma 9). For large n the mean is at most $\frac{1}{2}\sqrt{\delta}\log n$, and Bernstein's inequality, with $\Delta/\sqrt{\delta} = n^{-1/2}$, gives $\mathbb{P}(|F_{vc}| > \sqrt{\delta}\log n) \le 2\exp(-q_0^2\log^2 n/50)$. A union bound over v and c gives $\mathbb{P}(\mathcal{G}^c) \le 8n\,e^{-q_0^2\log^2 n/50}$.

Lemma 27 (remainder)**.** *On $\mathcal{G}$, $|\mathrm{Rem}_D| \le C_D\,\varepsilon\,n\,(n\Delta)^3(\sqrt{\delta}\log n)^{D-2}$; everywhere, $|\mathrm{Rem}_D| \le C_D n^{2D}$.*

Proof. On $\mathcal{G}$, $\|F_v^{\mathrm{ext}}\|_\infty \le \sqrt{\delta}\log n + 6\Delta/q_0$ because at most three edge weights, each at most $2\Delta/q_0$, are removed from F_v. Hence $|y|_1 \le 40\sqrt{\delta}\log n$ for large n, and the exponential factor is at most $e^{C\sqrt{\delta}\log n} \le 2$. Let s' be the number of vertices of L other than i. A connected B has at least s' edges; for a fixed i there are at most $2n^{s'}$ sets S with a given s', since S may contain i or not, and each pair (i, S) has at most 2^6 sets B. Since $|x^B| \le (2\Delta/q_0)^{|B|}$ and $\Delta \le \sqrt{\delta}\log n$,

$$|x^B|\,|y|_1^{D+1-|B|} \;\le\; C_D\,\Delta^{s'}\big(\sqrt{\delta}\log n\big)^{D+1-s'}.$$

Summing over i and S gives at most $C_D \varepsilon n \sum_{s' \le 3} (n\Delta)^{s'} (\sqrt{\delta} \log n)^{D+1-s'}$, and since $n\Delta = \sqrt{n\delta} \ge \sqrt{\delta} \log n$ the term $s' = 3$ dominates. Everywhere, $|\widetilde{S}_{\mathrm{RB3}}| \le n \sum_{s \le 3} \binom{n}{s} \cdot 8C_P \le Cn^4$, while $|\mathrm{Poly}_D| \le \#\{\Gamma : |\Gamma| \le D\} \cdot \max_\Gamma |B_\Gamma| \cdot (2\Delta/q_0)^{|\Gamma|} \le C_D n^{2D}$ by Lemma 20; hence $|\mathrm{Rem}_D| \le C_D n^{2D}$. □

Proof of Lemma 24. By Lemma 16, $\mathbb{E}_u \widetilde{S}_{\mathrm{RB3}} = 0$, so $\mathbb{E}\mathrm{Poly}_D = -\mathbb{E}\mathrm{Rem}_D$ and

$$\mathbb{E}_u \widetilde{S}^2_{\mathrm{RB3}} \le 2\mathbb{E}\mathrm{Poly}_D^2 + 2\mathbb{E}\mathrm{Rem}_D^2 \le 2\mathrm{Var}(\mathrm{Poly}_D) + 4\mathbb{E}\mathrm{Rem}_D^2.$$

By Lemma 27, $\mathbb{E}\mathrm{Rem}_D^2 \le C_D \varepsilon^2 n^2 (n\Delta)^6 (\delta \log^2 n)^{D-2} + C_D n^{4D} \cdot 8n\, e^{-q_0^2 \log^2 n/50}$, and $n^2(n\Delta)^6 = n^5\delta^3$. Together with Lemma 25 this is (10). □

F Proofs of the lower bound and the threshold

Proof of Lemma 6. Let $D := \max\{15, \lceil 14/(3\kappa) \rceil + 1\}$, so that $\kappa D > 14/3$, and let C_D, ε_D, n_D be as in Lemma 24. Let $\varepsilon_H > 0$ be such that the entropy gap of Lemma 15(iv) is at least $h_*\varepsilon/2$ for $\varepsilon \le \varepsilon_H$, which exists because $h_* > 0$. Put $a_0 := (4C_D)^{-1/6}$ and $\varepsilon_1 := \min\{\varepsilon_D, \varepsilon_P, \varepsilon_H\}$, and let $\varepsilon := \min\{\varepsilon_1, a_0(n^4\Delta^6)^{-1/6}\}$. The window gives $\delta \le n^{-\kappa}$, so $\sqrt{\delta} \log n \le 1/C$ for n large and Lemma 24 applies; recall $n^4\Delta^6 = n\delta^3$.

If $a_0(n\delta^3)^{-1/6} \le \varepsilon_1$, then $\varepsilon = a_0(n\delta^3)^{-1/6}$ and $\varepsilon^6 n\delta^3 = a_0^6 = 1/(4C_D)$, while

$$\varepsilon^2 n^5 \delta^3 (\delta \log^2 n)^{D-2} = a_0^2\, n^{14/3} \delta^2 (\delta \log^2 n)^{D-2} \le a_0^2\, n^{14/3-\kappa D} (\log n)^{2D} \longrightarrow 0.$$

Otherwise $\varepsilon = \varepsilon_1$ and $n\delta^3 < (a_0/\varepsilon_1)^6$, so $C_D \varepsilon_1^6 n\delta^3 \le C_D a_0^6 = 1/4$; moreover $\delta < (a_0/\varepsilon_1)^2 n^{-1/3}$, so $\varepsilon_1^2 n^5 \delta^3 (\delta \log^2 n)^{D-2} \le C_\kappa n^{4-(D-2)/3} (\log n)^{2D} \to 0$ because $D \ge 15$, with C_κ depending only on (q_0, κ). In both cases the last term of (10) tends to zero, so for $n \ge n_\kappa$, $\mathbb{E}_u \widetilde{S}^2_{\mathrm{RB3}} \le \frac{1}{4} + \frac{3}{4} \le 1$ uniformly in $u \in [1/4, 3/4]$. □

Proof of Theorem 2. Let $\varepsilon, \varepsilon_1$ and ε_H be as in the proof of Lemma 6; since $\varepsilon \le \varepsilon_1 \le \varepsilon_H$, the entropy gap along the path is at least $h_*\varepsilon/2$. The family $u \mapsto P_{\theta(u)}$ lives on the finite set of graphs, is positive because $0 < q < p < 1$, and is continuously differentiable because $P_u(A)$ is a polynomial in the smooth path $\theta(u)$. By Lemma 16 and Lemma 6, $I(u) \le \mathbb{E}_u \widetilde{S}^2_{\mathrm{RB3}} \le 1$ on $[1/4, 3/4]$, so Lemma 12 gives $\mathrm{TV}(P_{\theta(1/4)}, P_{\theta(3/4)}) \le \frac{1}{2}(1/2)^{1/2} < \frac{1}{2}$. Lemma 13 with $\omega = h_*\varepsilon/2$ gives

$$R_4^* \;\ge\; \frac{h_*^2 \varepsilon^2}{64} = \frac{h_*^2}{64} \min\left\{\varepsilon_1^2,\; a_0^2 (n^4\Delta^6)^{-1/3}\right\} \;\ge\; c_\kappa \min\left\{1, (n^4\Delta^6)^{-1/3}\right\},$$

with $c_\kappa := \frac{h_*^2}{64} \min\{\varepsilon_1^2, a_0^2\}$, which depends only on (q_0, κ). The case $n^4\Delta^6 < (a_0/\varepsilon_1)^6$, in which $\varepsilon = \varepsilon_1$ is fixed, gives the constant-risk part of the bound. □

Proof of Corollary 3. (i) $n\Delta^2 = n^{-(2\gamma-1)}$, so Theorem 2 applies with $\kappa := \min\{2\gamma - 1, 1/3\}$ for $n \ge n_\gamma := n_\kappa$, and Theorem 1 applies since $n\Delta^2 \le 1$ and $q + \Delta \le 1 - q_0/2$ (Appendix E). As $n^4\Delta^6 = n^{4-6\gamma} \ge 1$ for $\gamma \le 2/3$, both bounds are of order $n^{-(4-6\gamma)/3}$.

(ii) Necessity. Suppose $n^4\Delta_n^6 \not\to \infty$. Along a subsequence $n^4\Delta_n^6 \le K$ for some $K < \infty$, so $n\Delta_n^2 \le K^{1/3} n^{-1/3} \le n^{-1/4}$ for large n, and Theorem 2 with $\kappa = 1/4$ gives $R_4^* \ge c_{1/4} \min\{1, K^{-1/3}\} > 0$ along the subsequence; no estimator sequence is uniformly consistent. Sufficiency is Corollary 30, which needs no window condition. □

G Proof of the upper bound

In this appendix $n \ge 4$, $0 \le q < q + \Delta \le 1$ and $\theta \in \Theta_4$ are arbitrary; no window condition is used. Write $W_{ij} := \mathbf{1}\{Z_i = Z_j\}$, $\xi_{ij} := X_{ij} - W_{ij}$ and $v_{ij} := (q + \Delta W_{ij})(1 - q - \Delta W_{ij})$. Given Z, the ξ_e are independent with mean zero and variance $v_e/\Delta^2 \le \rho$; we also use $\bar{\rho} := \mathbb{E}v_{12}/\Delta^2 \le \rho$.

G.1 THE THREE MOMENT STATISTICS

Lemma 28 (moments on the complete graph). *(i)* $\mathbb{E}\widehat{P}_k = P_k$ *for* $k = 2, 3, 4$.

(ii) $\mathrm{Var}\widehat{P}_2 \le \frac{2n-3}{2n(n-1)} + \frac{2\rho}{n(n-1)} \le 3V_4$.

(iii) $\mathrm{Var}\widehat{P}_3 \le \frac{9}{4n} + \frac{8\rho}{n(n-1)} + \frac{2\rho^2}{n(n-1)(n-2)} \le 11V_4$.

(iv) $\mathrm{Var}\widehat{P}_4 \le \frac{4}{n} + \frac{18\rho}{n(n-1)} + \frac{18\rho^2}{n(n-1)(n-2)} + \frac{6\rho^3}{n(n-1)(n-2)(n-3)} \le 64V_4$.

Hence $\mathbb{E}\|\widehat{P} - P\|^2 \le 78V_4$.

Proof. (i) $\mathbb{E}[X_{ij} \mid Z] = W_{ij}$, and distinct pairs are conditionally independent, so the products over the edges of an edge, an ordered wedge $(i; j, k)$ and a three-star $(i; \{j, k, l\})$ have conditional means W_{ij}, $W_{ij}W_{ik}$ and $W_{ij}W_{ik}W_{il}$, the indicators that 2, 3 and 4 answers share a meaning, with expectations P_2, P_3 and P_4.

(ii) By the law of total variance, $\mathrm{Var}\widehat{P}_2 = \mathrm{Var}(\binom{n}{2}^{-1}\sum_e W_e) + \binom{n}{2}^{-1}\bar{\rho}$. In the first term, a pair of equal edges contributes $P_2(1-P_2)$, each of the $\binom{n}{2}\cdot 2(n-2)$ ordered pairs of edges sharing one endpoint contributes $P_3 - P_2^2$, and disjoint edges are independent. With $P_2(1-P_2) \le 1/4$, $P_3 - P_2^2 = \mathrm{Var}(\theta_{Z_1}) \le 1/4$ and $\bar{\rho} \le \rho$, $\mathrm{Var}\widehat{P}_2 \le \frac{1/2+2\rho+(n-2)}{n(n-1)}$, which is the first bound. For $n \ge 4$, $\frac{2n-3}{2n(n-1)} \le \frac{1}{n}$ and $\frac{2\rho}{n(n-1)} \le \frac{3\rho}{n^2}$.

(iii) and (iv) For an s-leaf star average ($s = 2$ gives $\widehat{P}_3$ over ordered wedges, $s = 3$ gives $\widehat{P}_4$), expand $\prod_e(W_e + \xi_e)$ over the set T of edges that carry ξ and collect the terms with $|T| = t$ into M_t, with $M_0 =: L$ a function of Z. Given Z, each M_t with $t \ge 1$ has mean zero, and for $t \ne t'$, $\mathbb{E}[M_t M_{t'} \mid Z] = 0$, since a product of ξ's has nonzero conditional mean only if every edge occurs at least twice, which forces the two edge sets to coincide. Hence $\mathrm{Var} = \mathrm{Var}L + \sum_{t\ge 1}\mathbb{E}M_t^2$.

For $\widehat{P}_3$: $L = \binom{n}{3}^{-1}\sum_{|T_3|=3}\mathbf{1}\{T_3 \text{ monochromatic}\}$; changing Z_a changes only the $\binom{n-1}{2}$ indicators containing a, so L moves in an interval of length at most $3/n$, and Popoviciu's inequality with the Efron–Stein inequality gives $\mathrm{Var}L \le n(3/n)^2/4 = 9/(4n)$. The single-noise term is $M_1 = \frac{2}{N_3}\sum_{e=\{i,k\}} c_e\xi_e$ with $N_3 := n(n-1)(n-2)$ and $c_e = \sum_{j\notin\{i,k\}}(W_{ij} + W_{kj}) \le 2(n-2)$, so $\mathbb{E}M_1^2 \le \frac{4}{N_3^2}\binom{n}{2}\,4(n-2)^2\rho = \frac{8\rho}{n(n-1)}$. The double-noise term sums over the $N_3/2$ unordered wedges, which give distinct pairs of edges and hence orthogonal terms: $\mathbb{E}M_2^2 \le \frac{4}{N_3^2}\frac{N_3}{2}\rho^2 = \frac{2\rho^2}{n(n-1)(n-2)}$.

For $\widehat{P}_4$, with $N_s := n\binom{n-1}{3} = 4\binom{n}{4}$: every four-set corresponds to exactly four stars, so $L = \binom{n}{4}^{-1}\sum_{|T_4|=4}\mathbf{1}\{T_4 \text{ monochromatic}\}$, which moves by at most $\binom{n-1}{3}/\binom{n}{4} = 4/n$ when one label changes, and $\mathrm{Var}L \le n(4/n)^2/4 = 4/n$. The single-noise term is $M_1 = N_s^{-1}\sum_e c_e\xi_e$, where c_e, a sum over the stars containing e of the product of the other two indicators, satisfies $0 \le c_e \le 2\binom{n-2}{2}$; hence $\mathbb{E}[M_1^2 \mid Z] \le N_s^{-2}\binom{n}{2}(n-2)^2(n-3)^2\rho = \frac{18\rho}{n(n-1)}$. The double-noise term groups into wedges $(i; \{j, k\})$ with coefficient $\sum_l W_{il} \le n-3$, distinct wedges giving orthogonal terms, so $\mathbb{E}M_2^2 \le N_s^{-2}\,n\binom{n-1}{2}(n-3)^2\rho^2 = \frac{18\rho^2}{n(n-1)(n-2)}$. Distinct stars give distinct triples of edges, so $\mathbb{E}M_3^2 \le \rho^3/N_s = \frac{6\rho^3}{n(n-1)(n-2)(n-3)}$.

The bounds by multiples of V_4 use $n-1 \ge 3n/4$, $n-2 \ge n/2$ and $n-3 \ge n/4$ for $n \ge 4$. □

G.2 THE HÖLDER-$\frac{1}{3}$ TRANSFER

Lemma 29 (transfer). *Let* $\theta, \theta' \in \Theta_4$*, zero coordinates allowed, and* $D_e := \sum_{j=2}^4 |e_j(\theta) - e_j(\theta')|$. *Then* $D_e \le D_e^{\max} := \frac{3}{8} + \frac{1}{16} + \frac{1}{256}$ *and*

$$|H(\theta) - H(\theta')| \;\le\; D_e \log 2 + \frac{2\pi}{\sqrt{3}}\, D_e^{1/3}.$$

Proof. Let $p_\theta(u) := \prod_c (u+\theta_c) = u^4+u^3+e_2u^2+e_3u+e_4$, where $\sum_c \theta_c = 1$ was used; since $e_j \geq 0$, $p_\theta(u) \geq u^4+u^3$. For $t \in [0,1]$ and $M > 0$, $\int_0^M \log(u+t)\,\mathrm{d}u = (M+t)\log(M+t) - M - t\log t$ (with $0\log 0 = 0$), so summing over c and subtracting the same identity for θ',

$$\int_0^M \log \frac{p_\theta(u)}{p_{\theta'}(u)}\,\mathrm{d}u = H(\theta) - H(\theta') + \sum_c \big[\phi_M(\theta_c) - \phi_M(\theta'_c)\big], \qquad \phi_M(t) := (M+t)\log(M+t).$$

By Taylor's theorem $\phi_M(t) = M\log M + t(\log M + 1) + t^2/(2M) + O(t^3/M^2)$ for $t \in [0,1]$, so the bracket is $(P_2(\theta) - P_2(\theta'))/(2M) + O(M^{-2}) \to 0$; this is the integral representation of Jozsa & Mitchison (2015), $H(\theta) - H(\theta') = \int_0^\infty \log(p_\theta/p_{\theta'})\,\mathrm{d}u$, the integrand being absolutely integrable by the bounds below. For $x, y > 0$, $|\log(x/y)| \leq \log(1 + |x-y|/\min(x,y))$, and $p_\theta - p_{\theta'} = \delta_2 u^2 + \delta_3 u + \delta_4$ with $\delta_j := e_j(\theta) - e_j(\theta')$. For $u \geq 1$ the difference is at most $D_e u^2$, and the integrand is at most $D_e u^2/(u^4+u^3) = D_e/(u(u+1))$, whose integral over $[1,\infty)$ is $D_e \log 2$. For $0 < u \leq 1$ the difference is at most D_e and $u^4+u^3 \geq u^3$, so the contribution is at most $\int_0^1 \log(1 + D_e/u^3)\,\mathrm{d}u \leq D_e^{1/3}\int_0^\infty \log(1+t^{-3})\,\mathrm{d}t = D_e^{1/3}\int_0^\infty \frac{3\,\mathrm{d}t}{1+t^3} = \frac{2\pi}{\sqrt{3}} D_e^{1/3}$, by the substitution $u = D_e^{1/3}t$ and an integration by parts. Finally $P_2 \geq 1/4$ gives $e_2 \leq 3/8$, and the AM–GM inequality gives $e_3 \leq 4\cdot 4^{-3} = 1/16$ and $e_4 \leq 4^{-4} = 1/256$. □

G.3 Proof of Theorem 1

Proof. The estimator is well defined. $\mathcal{M}_4$ is the image of the compact set Θ_4 under a continuous map, so every point has a nearest point in $\mathcal{M}_4$, selected by a fixed rule. If $\widetilde{P} = P(\widetilde{\theta})$ with $\widetilde{\theta} \in \Theta_4$, then by (2) the quartic of Algorithm 1 equals $\prod_c (x - \widetilde{\theta}_c)$, so its roots are the coordinates of $\widetilde{\theta}$; any θ'' with the same moments has the same elementary symmetric polynomials and hence the same multiset of coordinates. So $\widehat{H} = H(\widetilde{\theta})$ is well defined and lies in $[0, \log 4]$.

Moment error. As $P \in \mathcal{M}_4$, $\|\widetilde{P} - P\| \leq \|\widetilde{P} - \widehat{P}\| + \|\widehat{P} - P\| \leq 2\|\widehat{P} - P\|$. Write $\delta P_k := \widetilde{P}_k - P_k$. By (2) and $|\widetilde{P}_2^2 - P_2^2| \leq 2|\delta P_2|$,

$$D_e(\widetilde{\theta}, \theta) \leq \tfrac{3}{2}|\delta P_2| + \tfrac{2}{3}|\delta P_3| + \tfrac{1}{4}|\delta P_4| \leq c_e\,\|\widehat{P} - P\|, \qquad c_e := 2\big(\tfrac{9}{4} + \tfrac{4}{9} + \tfrac{1}{16}\big)^{1/2} \leq 3.321.$$

Assembly. By Lemma 29 and $D_e \leq D_e^{\max}$, $(\widehat{H} - H)^2 \leq K_H D_e^{2/3}$ with $K_H := ((D_e^{\max})^{2/3}\log 2 + 2\pi/\sqrt{3})^2 \leq 16.24$. By Jensen's inequality and Lemma 28, $\mathbb{E}D_e^{2/3} \leq (\mathbb{E}D_e^2)^{1/3} \leq (78 c_e^2 V_4)^{1/3}$, so the risk is at most $K_H(78c_e^2)^{1/3}V_4^{1/3} \leq 155 V_4^{1/3}$ (numerically 154.41). The cap $(\log 4)^2$ holds because $\widehat{H}, H \in [0, \log 4]$.

The window form. If $n\Delta^2 \leq 1$, then $\rho \leq 1/(4\Delta^2)$ gives $n^4\Delta^6 V_4 \leq n^3\Delta^6 + \frac{n^2\Delta^4}{4} + \frac{n\Delta^2}{16} + \frac{1}{64} \leq \frac{85}{64}$, so the risk is at most $155(85/64)^{1/3}(n^4\Delta^6)^{-1/3} \leq 171(n^4\Delta^6)^{-1/3}$, and with the cap $(\log 4)^2 \leq 171$ it is at most $171\min\{1, (n^4\Delta^6)^{-1/3}\}$. □

Corollary 30 (uniform consistency)**.** *If $\Delta_n \in (0,1]$ and $n^4\Delta_n^6 \to \infty$, then $\sup_q \sup_{\theta\in\Theta_4} \mathbb{E}_\theta(\widehat{H} - H(\theta))^2 \to 0$, the supremum over q running over $0 \leq q < q + \Delta_n \leq 1$.*

Proof. $n \to \infty$ because $\Delta_n \leq 1$. From $\rho \leq 1/(4\Delta_n^2)$: $\rho/n^2 \leq \frac{1}{4}(n^4\Delta_n^6)^{-1/3}n^{-2/3}$, $\rho^2/n^3 \leq \frac{1}{16}(n^4\Delta_n^6)^{-2/3}n^{-1/3}$ and $\rho^3/n^4 \leq (64 n^4\Delta_n^6)^{-1}$. These bounds do not depend on (q,θ), so $V_4 \to 0$ uniformly, and Theorem 1 applies. □

Corollary 31 (miscalibrated channel)**.** *Let $n \geq 4$ and $0 \leq q < q + \Delta \leq 1$.*

(i) Run Algorithm 1 with (q', Δ'), $\Delta' > 0$, in place of (q, Δ). Put $s := \Delta/\Delta'$ and $\beta := (q' - q)/\Delta'$, and suppose $|s - 1| \leq 1/2$ and $|\beta| \leq 1$. Then its output $\widehat{H}'$ satisfies, for every $\theta \in \Theta_4$,

$$\mathbb{E}_\theta\big(\widehat{H}' - H(\theta)\big)^2 \;\leq\; \min\big\{(\log 4)^2,\; 904\,\big(V_4 + (|s-1| + |\beta|)^2\big)^{1/3}\big\}.$$

(ii) Let $\hat{p}$ and $\hat{q}$ be the fractions of agreements of the judge on m pairs of answers known to share a meaning and on m pairs known to differ, whose $2m$ verdicts are independent with laws $\mathrm{Bern}(p)$ and $\mathrm{Bern}(q)$ and independent of (Z, A). Run Algorithm 1 with $(\hat{q}, \hat{p}-\hat{q})$ if $\hat{p} > \hat{q}$, and return $\log 2$ otherwise. Its output $\widehat{H}^{\mathrm{cal}}$ satisfies, for every $\theta \in \Theta_4$,

$$\mathbb{E}_\theta\big(\widehat{H}^{\mathrm{cal}} - H(\theta)\big)^2 \;\le\; \min\Big\{(\log 4)^2,\; 904\,\Big(V_4 + \frac{4}{m\Delta^2}\Big)^{1/3} + 4(\log 4)^2 e^{-m\Delta^2/32}\Big\},$$

and if $n\Delta^2 \le 1$ and $m \ge n^4\Delta^4$, the right-hand side is at most $1600\min\{1, (n^4\Delta^6)^{-1/3}\}$.

Proof. (i) The centred verdicts computed with (q', Δ') are $X'_{ij} = (A_{ij} - q')/\Delta' = sX_{ij} - \beta$. On the complete graph every edge lies in $2(n-2)$ unordered wedges and in $2\binom{n-2}{2}$ three-stars, and every unordered wedge in $n-3$ three-stars; the average over ordered wedges equals the average over unordered ones because $X_{ij}X_{ik}$ is symmetric in j, k. Averaging the products of $sX_e - \beta$ therefore gives

$$\widehat{P}'_2 = s\widehat{P}_2 - \beta, \qquad \widehat{P}'_3 = s^2\widehat{P}_3 - 2s\beta\widehat{P}_2 + \beta^2, \qquad \widehat{P}'_4 = s^3\widehat{P}_4 - 3s^2\beta\widehat{P}_3 + 3s\beta^2\widehat{P}_2 - \beta^3.$$

Thus $\widehat{P}' = M\widehat{P} + c$ with a lower-triangular M and a constant vector c, and $\widehat{P}' - P = M(\widehat{P} - P) + b$ with $b := MP + c - P$. For $s \in [1/2, 3/2]$ and $|\beta| \le 1$ the entries of M give $\|M\| \le \|M\|_F \le 10$, and since $P_k \in [0,1]$, $|b_2| \le |s-1| + |\beta|$, $|b_3| \le 2.5|s-1| + 4|\beta|$ and $|b_4| \le 4.75|s-1| + 12.25|\beta|$, so $\|b\| \le 13(|s-1| + |\beta|)$. With Lemma 28,

$$\mathbb{E}\|\widehat{P}' - P\|^2 \le 2\|M\|^2\, 78V_4 + 2\|b\|^2 \le 15600\big(V_4 + (|s-1| + |\beta|)^2\big).$$

The projection and transfer steps of the proof of Theorem 1 do not use the form of $\widehat{P}$ and give $\mathbb{E}(\widehat{H}' - H)^2 \le K_H c_e^{2/3}(\mathbb{E}\|\widehat{P}' - P\|^2)^{1/3}$, and $K_H c_e^{2/3} 15600^{1/3} \le 904$. The cap holds because $\widehat{H}', H \in [0, \log 4]$.

(ii) Condition on the $2m$ labelled verdicts; this fixes $\hat{q}$ and $\hat{\Delta} := \hat{p} - \hat{q}$ and leaves the law of (Z, A) unchanged. On the event $\mathcal{G} := \{|\hat{p}-p| \le \Delta/8,\ |\hat{q}-q| \le \Delta/8\}$ we have $|\hat{\Delta} - \Delta| \le |\hat{p}-p| + |\hat{q}-q| \le \Delta/4$, so $\hat{\Delta} \ge 3\Delta/4 > 0$, $|s-1| = |\Delta - \hat{\Delta}|/\hat{\Delta} \le 1/3$ and $|\beta| = |\hat{q} - q|/\hat{\Delta} \le 1/6$; thus (i) applies, with

$$|s-1| + |\beta| \;\le\; \frac{|\hat{p}-p| + 2|\hat{q}-q|}{\hat{\Delta}} \;\le\; \frac{4}{3\Delta}\big(|\hat{p}-p| + 2|\hat{q}-q|\big).$$

Since $(a+2b)^2 \le 3a^2 + 6b^2$ and $\mathbb{E}(\hat{p}-p)^2, \mathbb{E}(\hat{q}-q)^2 \le 1/(4m)$, this gives $\mathbb{E}\big[(|s-1|+|\beta|)^2 \mathbf{1}_{\mathcal{G}}\big] \le \frac{16}{9\Delta^2} \cdot \frac{9}{4m} = \frac{4}{m\Delta^2}$, and since $x \mapsto 904(V_4 + x)^{1/3}$ is nonnegative, increasing and concave, the contribution of $\mathcal{G}$ to the risk is at most $904(V_4 + 4/(m\Delta^2))^{1/3}$. Off $\mathcal{G}$ the squared error is at most $(\log 4)^2$, and Hoeffding's inequality for the two averages gives $\mathbb{P}(\mathcal{G}^c) \le 4e^{-2m(\Delta/8)^2} = 4e^{-m\Delta^2/32}$. The cap holds because $\widehat{H}^{\mathrm{cal}}, H \in [0, \log 4]$. For the last claim let $x := n^4\Delta^6$. When $n\Delta^2 \le 1$, $xV_4 \le 85/64$ as in the proof of Theorem 1; $m \ge n^4\Delta^4$ gives $4/(m\Delta^2) \le 4/x$ and $e^{-m\Delta^2/32} \le e^{-x/32} \le 2x^{-1/3}$, because $x^{1/3}e^{-x/32} \le (32/(3e))^{1/3} < 2$. So the bound is at most $\big(904(85/64 + 4)^{1/3} + 8(\log 4)^2\big)x^{-1/3} \le 1600\, x^{-1/3}$, and $(\log 4)^2 \le 1600$. □

Remark 32 (computing the projection). The nearest point $\widetilde{P}$ minimizes the polynomial $\|P(\theta) - \widehat{P}\|^2$ of degree 8 over the three-dimensional simplex Θ_4. If an implementation returns $\widetilde{P} \in \mathcal{M}_4$ with $\|\widetilde{P} - \widehat{P}\| \le \mathrm{dist}(\widehat{P}, \mathcal{M}_4) + \eta$, then $\|\widetilde{P} - P\| \le 2\|\widehat{P} - P\| + \eta$, and since $t \mapsto t^{2/3}$ is subadditive the bound of Theorem 1 acquires the additional term $K_H(c_e\eta/2)^{2/3} \le 23\,\eta^{2/3}$. A grid attains $\eta \le V_4^{1/2}$. Let $\hat{\theta}$ minimize $\|P(\theta) - \widehat{P}\|$ over $\Theta_4^N := \{\theta \in \Theta_4 : N\theta \in \mathbb{Z}^4\}$. Rounding the coordinates of $N\theta$ to integers with the same sum moves each of them by less than one, so every $\theta \in \Theta_4$ lies within ℓ_1-distance $4/N$ of Θ_4^N; and $|P_k(\theta) - P_k(\theta')| \le k\|\theta - \theta'\|_1$, so $\|P(\theta) - P(\theta')\| \le \sqrt{29}\,\|\theta - \theta'\|_1$. Hence $\widetilde{P} := P(\hat{\theta})$ satisfies $\|\widetilde{P} - \widehat{P}\| \le \mathrm{dist}(\widehat{P}, \mathcal{M}_4) + \eta$ with $\eta \le 4\sqrt{29}/N \le 22/N$. For $N := \lceil 22\sqrt{n} \rceil$, $V_4 \ge 1/n$ gives $\eta \le V_4^{1/2}$, so $\widehat{H} := H(\hat{\theta})$ satisfies Theorem 1 with 178 in place of 155 and 196 in place of 171; the grid has $\binom{N+3}{3} = O(n^{3/2})$ points, and since $\widetilde{\theta} = \hat{\theta}$ up to permutation no quartic needs to be solved.

H PROOFS FOR QUERY DESIGN

Let G be a simple graph on $[n]$, fixed independently of (Z, A), with $m = Q$ edges, degrees d_i, $d_{\max} := \max_i d_i$, ordered wedges $(i; j, k)$ with $\{i, j\}, \{i, k\} \in E(G)$ and $j \neq k$, whose number is $N_\Lambda := \sum_i d_i(d_i - 1)$, and $S_3(G) = \sum_i \binom{d_i}{3}$ three-stars. The estimator $\widehat{H}_G$ uses $\widehat{P}_2^G := m^{-1} \sum_{e \in E(G)} X_e$, $\widehat{P}_3^G := N_\Lambda^{-1} \sum_{(i;j,k)} X_{ij} X_{ik}$ and $\widehat{P}_4^G := S_3(G)^{-1} \sum_{(i;\{j,k,l\})} X_{ij} X_{ik} X_{il}$, followed by lines 4 to 6 of Algorithm 1. Let $\kappa_a := d_a(d_a - 1) + 2 \sum_{i \in N(a)} (d_i - 1)$, the number of ordered wedges containing a, and $K(G) := \sum_a \kappa_a^2 / N_\Lambda^2$.

Lemma 33 (moments on a query graph)**.** *Let $0 \leq q < q + \Delta \leq 1$, $\theta \in \Theta_4$ and $S_3(G) > 0$. Then some $d_i \geq 3$, so $m \geq 3$ and $N_\Lambda \geq 6$, and the three statistics are defined.*

(i) The three statistics are unbiased.

(ii) $\mathrm{Var} \widehat{P}_2^G \leq \frac{1}{4m} + \frac{N_\Lambda}{4m^2} + \frac{\rho}{m}$.

(iii) $\mathrm{Var} \widehat{P}_3^G \leq \frac{K(G)}{4} + \frac{8(d_{\max} - 1)\rho}{N_\Lambda} + \frac{2\rho^2}{N_\Lambda}$.

(iv) $\mathrm{Var} \widehat{P}_4^G \leq \left(\frac{2}{3} n^3 + 3n^2 \rho + 3n\rho^2 + \rho^3\right) / S_3(G) \leq (n + \rho)^3 / S_3(G)$.

Proof. (i) is as in Lemma 28(i). (ii) In the law of total variance, the conditional variance contributes $\bar{\rho}/m \leq \rho/m$; in $\mathrm{Var}(m^{-1} \sum_e W_e)$ the diagonal gives $m P_2(1 - P_2)$ and the N_Λ ordered pairs of distinct edges sharing a vertex give $P_3 - P_2^2$ each, both at most $1/4$.

(iii) The decomposition $\widehat{P}_3^G = L + M_1 + M_2$ and its orthogonality are as in the proof of Lemma 28, since they use only the conditional independence of distinct edges. Changing Z_a changes at most κ_a terms of L, so L moves in an interval of length κ_a / N_Λ, and Efron–Stein with Popoviciu gives $\mathrm{Var} L \leq K(G)/4$. The single-noise term is $M_1 = \frac{2}{N_\Lambda} \sum_{e=\{i,k\}} c_e \xi_e$ with $c_e \leq (d_i - 1) + (d_k - 1)$, and $\sum_{\{i,k\} \in E(G)} (d_i + d_k - 2)^2 \leq 2 \sum_i d_i (d_i - 1)^2 \leq 2(d_{\max} - 1) N_\Lambda$, so $\mathbb{E} M_1^2 \leq 8(d_{\max} - 1)\rho / N_\Lambda$. The $N_\Lambda/2$ unordered wedges give distinct pairs of edges, so $\mathbb{E} M_2^2 \leq \frac{4}{N_\Lambda^2} \frac{N_\Lambda}{2} \rho^2$.

(iv) The same decomposition over three-stars gives $\widehat{P}_4^G = L + M_1 + M_2 + M_3$. Changing Z_a changes at most $\kappa_a^{(3)} := \binom{d_a}{3} + \sum_{i \in N(a)} \binom{d_i - 1}{2}$ indicators, the stars centred at a and those with a as a leaf; $\kappa_a^{(3)} \leq \binom{n-1}{3} + (n-1)\binom{n-2}{2} = 4\binom{n-1}{3} \leq \frac{2}{3} n^3$ and $\sum_a \kappa_a^{(3)} = 4 S_3$. Hence $\mathrm{Var} L \leq \sum_a (\kappa_a^{(3)} / S_3)^2 / 4 \leq \max_a \kappa_a^{(3)} / S_3 \leq 2n^3 / (3 S_3)$. For the single-noise term $M_1 = S_3^{-1} \sum_e c_e \xi_e$, each c_e is at most the number of stars containing e, so $c_e \leq d_{\max}^2$ and $\sum_e c_e \leq 3 S_3$, whence $\mathbb{E} M_1^2 \leq 3 d_{\max}^2 \rho / S_3$. The double-noise term groups into wedges $(i; \{j, k\})$ of G with coefficient at most $d_i - 2$, distinct wedges being orthogonal; since $\binom{d}{2}(d-2)^2 = 3\binom{d}{3}(d-2)$, $\mathbb{E} M_2^2 \leq S_3^{-2} \sum_i \binom{d_i}{2} (d_i - 2)^2 \rho^2 \leq 3 d_{\max} \rho^2 / S_3$. Finally $\mathbb{E} M_3^2 \leq \rho^3 / S_3$. Summing, with $d_{\max} \leq n$, gives the claim. □

Proof of Theorem 7. The proof of Theorem 1 applies verbatim with $78 V_4$ replaced by $V_G := \mathrm{Var} \widehat{P}_2^G + \mathrm{Var} \widehat{P}_3^G + \mathrm{Var} \widehat{P}_4^G$, and gives $\mathbb{E}(\widehat{H}_G - H)^2 \leq \min\{(\log 4)^2, K_H (c_e^2 V_G)^{1/3}\}$, all three statistics being defined by Lemma 33. Under $q \in [q_0, 1 - q_0]$ and $n\Delta^2 \leq q_0^2/4$, $\rho \geq q_0(1 - q_0)/\Delta^2 \geq 4n(1 - q_0)/q_0 \geq 4n$, so $n \leq \rho/4$ and $\rho \geq 1$. We use $S_3 \leq \sum_i d_i^3 / 6 \leq d_{\max}^2 m / 3 \leq n^2 m / 3$; $S_3 \leq d_{\max} N_\Lambda / 6$, since $\binom{d}{3} = \binom{d}{2} \frac{d-2}{3}$; $N_\Lambda \leq 2 d_{\max} m \leq 2nm$; and $K(G) \leq \max_a \kappa_a \sum_a \kappa_a / N_\Lambda^2 \leq 3n^2 \cdot 3 N_\Lambda / N_\Lambda^2 = 9n^2 / N_\Lambda$. Term by term,

$$\frac{1}{4m} \leq \frac{n^2}{12 S_3}, \quad \frac{N_\Lambda}{4m^2} \leq \frac{n^3}{6 S_3}, \quad \frac{\rho}{m} \leq \frac{n^2 \rho}{3 S_3},$$
$$\frac{K(G)}{4} \leq \frac{3n^3}{8 S_3}, \quad \frac{8(d_{\max} - 1)\rho}{N_\Lambda} \leq \frac{4n^2 \rho}{3 S_3}, \quad \frac{2\rho^2}{N_\Lambda} \leq \frac{n\rho^2}{3 S_3},$$

and $(n + \rho)^3 \leq (5\rho/4)^3$. With $n \leq \rho/4$ and $\rho \geq 1$ the total is at most $\left(\frac{1}{192} + \frac{1}{384} + \frac{1}{48} + \frac{3}{512} + \frac{1}{12} + \frac{1}{12} + \frac{125}{64}\right) \frac{\rho^3}{S_3} = \frac{1103}{512} \frac{\rho^3}{S_3} \leq \frac{3\rho^3}{S_3}$. Hence the risk is at most $K_H (3c_e^2)^{1/3} \rho\, S_3^{-1/3} \leq 52.2\, \rho\, S_3^{-1/3}$, and $\rho \leq 1/(4\Delta^2)$ gives the constant $13.05 \leq 14$. □

Proof of Corollary 8. (i) was shown above. (ii) Let $s := \lfloor m/(n-1) \rfloor$; then $1 \le s \le n/2$ and $s \ge m/(2(n-1))$. Joining s hubs to all other vertices uses $s(n-1) - \binom{s}{2} \le m$ edges; adding the remaining edges anywhere does not decrease S_3. Each hub has degree $n-1$, so $S_3 \ge s\binom{n-1}{3} \ge \frac{m}{2(n-1)} \cdot \frac{(n-1)(n-2)(n-3)}{6} = \frac{m(n-2)(n-3)}{12} \ge \frac{n^2 m}{48}$ for $n \ge 6$. By (i), $14(\Delta^6 S_3)^{-1/3} \ge 14(3/(n^2 m \Delta^6))^{1/3}$ for every graph, while the hub design gives at most $14(48/(n^2 m \Delta^6))^{1/3}$; the ratio is $16^{1/3}$. (iii) $S_3 = n\binom{d}{3}$, and $nd^3/27 \le n\binom{d}{3} \le nd^3/6$ for $d \ge 3$, with $d = 2m/n$; so the uncapped expressions are of order $(m^3\Delta^6/n^2)^{-1/3}$ for the regular design and $(n^2 m \Delta^6)^{-1/3}$ for the hub design, and their ratio is of order $(n^2/m)^{2/3}$. □

I CERTIFICATES AND NUMERICAL CHECKS

The symbolic computation of Remark 23(b) is a certificate used in its proof. The interval enclosure of h_* supplies its digits; $h_* > 0$ is proved in Lemma 15(iv). The other computations check the coefficient structure and the variance of the three-star average, simulate the risk of the estimator and measure its cost. All run on a CPU. The supplementary material contains their code and full outputs, among them the rational brackets of the six roots and both polynomials of Remark 23(b).

Exact Taylor coefficients. For a motif Γ on the vertex set V, the swept score of the motif system is a finite combination of ratios of exponential polynomials in the edge weights of Γ. We compute its Taylor expansion exactly in the ring of polynomials in these weights truncated at the multiplicities of Γ, enumerating all $4^{v(\Gamma)}$ labellings, and read off $B_\Gamma^{(r)}(z)$. On the motif system the order-r sweep collapses to $\sum_{|U| \le r} c_r(|U|) \mathcal{P}_U^{(V)} \widetilde{S}_1^{(V)}$ with $c_r(k) := \sum_{j=0}^{r-k} (-1)^j \binom{|V|-k}{j}$. The path point $\theta(u)$ is computed with at least 60 significant digits and the ring arithmetic is carried out in double precision; second moments are $\mathbb{E}B^2 = \sum_z \prod_{v \in V} \theta_{z_v} B(z)^2$. Three checks use this computation, and it cross-checks the certificate for Remark 23(b).

- *The identity of Lemma 18.* For eight motifs (three-star, three-edge path, triangle, four-star, the spider with legs $(2,1,1)$, four-edge path, four-cycle and the paw) at $\varepsilon \in \{0.05, 0.01\}$ and $u = 1/2$, the full sweep ($r = v(\Gamma)$) agrees with the Taylor coefficient of the motif-system score $\partial_u \log \mathbb{E}_\theta \exp(\sum_e J_e W_e)$, computed directly, with maximum absolute residual 3.83×10^{-18} over all labellings. Under the third-order sweep, the triangle coefficient is zero to floating-point accuracy and the mean coefficient of the three-star and of the three-edge path equals $\dot{P}_4$ to all printed digits.
- *The third-order scan.* We evaluate $\mathbb{E}B_\Gamma^2/\varepsilon^6$ for sixteen connected motifs on four to seven vertices at $\varepsilon \in \{0.1, 0.03, 0.01, 0.003, 0.001\}$; Table 2 shows the values at 10^{-1}, 10^{-2} and 10^{-3} and the local slope between $3 \cdot 10^{-3}$ and 10^{-3}. Motifs on four or five vertices are evaluated at $u \in \{1/4, 1/2, 3/4\}$ and those on six or seven vertices at $u = 1/2$. At the finest scale the local slopes lie between 5.91 and 8.05; at larger ε the normalized values still move in both directions. The stars with four to six leaves carry the largest values, dominated by labellings with four rare leaves.
- *Lower-order sweeps.* The closed form (9) agrees with the exact first-order wedge coefficient on all 64 labellings at $\varepsilon \in \{0.1, 0.01\}$ and $u \in \{1/4, 1/2, 3/4\}$, with maximum difference 8.7×10^{-18}; Table 3 compares $\mathbb{E}B^2/\varepsilon^4$ with $L_1(u)$ and reports $B^{(2)}(0,a,a,a)/\dot{\theta}_a$. For the three-star, the sweeps of order one, two and three give the slopes 3.992, 4.986 and 5.996 between $\varepsilon = 3 \cdot 10^{-3}$ and 10^{-3} (Figure 1(b)).
- *Certificate for Remark 23(b).* Using a computer-algebra system, we compute $B^{(1)}(0,1,1,0)$ and $B^{(2)}(0,1,1,1)$ for the three-star as exact polynomials in $\theta_1, \theta_2, \theta_3, \dot{\theta}_0, \dot{\theta}_1, \dot{\theta}_2, \dot{\theta}_3$, with $\theta_0 = 1 - \theta_1 - \theta_2 - \theta_3$, by differentiating the exact motif-system expressions in the three edge weights at zero. Both have zero constant part; their parts of degree one are $3\dot{\theta}_1$ and $7\dot{\theta}_1$; their total degrees are 5 and 6. At two path points ($\varepsilon = 0.05$, $u = 1/2$ and $\varepsilon = 0.01$, $u = 0.3$) the polynomials agree with the ring computation to twelve significant digits.

The path. Table 4 evaluates Lemma 15 on a grid of 101 equally spaced values of $u \in [1/4, 3/4]$, with the path computed to 60 significant digits. The enclosure of h_* brackets each root of $x(x - \frac{1}{2})^2 = u/54$, $u \in \{1/4, 3/4\}$, in an interval of width 2×10^{-40} with rational endpoints, at which the sign

Table 2: Third-order sweep. For sixteen connected motifs Γ with $v(\Gamma)$ vertices, $\mathbb{E}B_\Gamma^2/\varepsilon^6$, where B_Γ is the coefficient of J^Γ in the swept score and the expectation is over independent vertex labels drawn from the path point $\theta(u)$ of rare-class scale ε; the exact coefficient expressions, with all labellings of the motif enumerated, are evaluated numerically. Also shown: the local slope of $\log \mathbb{E}B_\Gamma^2$ against $\log \varepsilon$ between $\varepsilon = 10^{-3}$ and $3 \cdot 10^{-3}$, and the number m of rare-class labels in the stratum that dominates $\mathbb{E}B_\Gamma^2$ at $\varepsilon = 10^{-3}$. Values are at finite ε and are not limiting constants. †The entries for the six-edge path are unresolved floating-point residuals at the rounding floor; no slope or stratum is read from them.

motif Γ	$v(\Gamma)$, u	$\mathbb{E}B_\Gamma^2/\varepsilon^6$ at $\varepsilon =$			slope	m
		10^{-1}	10^{-2}	10^{-3}		
star $K_{1,3}$	4, 1/4	4.28×10^{-3}	5.36×10^{-3}	5.47×10^{-3}	5.996	0
star $K_{1,3}$	4, 1/2	4.28×10^{-3}	5.36×10^{-3}	5.47×10^{-3}	5.996	0
star $K_{1,3}$	4, 3/4	4.28×10^{-3}	5.36×10^{-3}	5.47×10^{-3}	5.996	0
path, 3 edges	4, 1/4	4.28×10^{-3}	5.36×10^{-3}	5.47×10^{-3}	5.996	0
path, 3 edges	4, 1/2	4.28×10^{-3}	5.36×10^{-3}	5.47×10^{-3}	5.996	0
path, 3 edges	4, 3/4	4.28×10^{-3}	5.36×10^{-3}	5.47×10^{-3}	5.996	0
star $K_{1,4}$	5, 1/4	0.270	1.53	1.78	5.970	4
star $K_{1,4}$	5, 1/2	0.149	0.825	0.960	5.970	4
star $K_{1,4}$	5, 3/4	0.116	0.640	0.748	5.969	4
spider $(2,1,1)$	5, 1/4	0.100	0.0373	0.0190	6.186	0
spider $(2,1,1)$	5, 1/2	0.0564	0.0272	0.0179	6.106	0
spider $(2,1,1)$	5, 3/4	0.0446	0.0248	0.0177	6.084	0
path, 4 edges	5, 1/4	0.0113	2.89×10^{-3}	3.05×10^{-3}	5.981	0
path, 4 edges	5, 1/2	6.68×10^{-3}	2.81×10^{-3}	3.05×10^{-3}	5.979	0
path, 4 edges	5, 3/4	5.56×10^{-3}	2.79×10^{-3}	3.05×10^{-3}	5.978	0
cycle C_4	4, 1/4	0.0219	0.0457	0.0490	5.986	0
cycle C_4	4, 1/2	0.0219	0.0457	0.0490	5.986	0
cycle C_4	4, 3/4	0.0219	0.0457	0.0490	5.986	0
paw	4, 1/4	0.0201	0.0204	0.0218	5.985	0
paw	4, 1/2	0.0153	0.0203	0.0218	5.985	0
paw	4, 3/4	0.0144	0.0203	0.0218	5.985	0
star $K_{1,5}$	6, 1/2	0.340	3.74	4.71	5.956	4
path, 5 edges	6, 1/2	7.28×10^{-4}	4.31×10^{-6}	3.52×10^{-8}	8.046	0
double star $S_{2,2}$	6, 1/2	0.196	0.0772	0.0389	6.190	0
spider $(2,1,1,1)$	6, 1/2	0.120	0.0746	0.0388	6.184	0
spider $(2,2,1)$	6, 1/2	0.0154	2.76×10^{-3}	3.02×10^{-3}	5.965	0
spider $(3,1,1)$	6, 1/2	0.0141	2.72×10^{-3}	3.02×10^{-3}	5.965	0
star $K_{1,6}$	7, 1/2	14.6	8.35	13.7	5.909	4
spider $(2,2,2)$	7, 1/2	7.69×10^{-4}	6.95×10^{-6}	5.93×10^{-8}	8.038	0
path, 6 edges†	7, 1/2	1.26×10^{-26}	7.52×10^{-23}	2.92×10^{-19}	n/a	n/a

Table 3: Lower-order sweeps along the least-favourable path, with $\dot\theta_a := \partial_u \theta_a$. Top: first-order sweep on the wedge, $\mathbb{E}B^2/\varepsilon^4$ against its limit $L_1(u)$. Bottom: second-order sweep on the three-star whose centre carries the majority label 0 and whose three leaves carry the rare label a, the ratio $B(0,a,a,a)/\dot\theta_a$.

u	wedge, first order: $\mathbb{E}B^2/\varepsilon^4$ at $\varepsilon =$			$L_1(u)$
	10^{-2}	10^{-3}	10^{-4}	
1/4	0.173697	0.177859	0.178280	0.178326
1/2	0.098095	0.100548	0.100795	0.100823
3/4	0.087890	0.090268	0.090508	0.090535
u	three-star, second order: $B(0,a,a,a)/\dot\theta_a$ at $\varepsilon = 10^{-4}$			limit
	$a=1$	$a=2$	$a=3$	
1/4	6.995782	6.996492	6.997828	7
1/2	6.995662	6.996701	6.997740	7
3/4	6.995573	6.996909	6.997620	7

Table 4: The least-favourable path, with class 0 the majority class and $\psi_c := \partial_u \log \theta_c$. Over the grid of 101 equally spaced $u \in [1/4, 3/4]$, $\bar{\psi} := \max_{u,\, c\geq 1} |\psi_c|$ and $\bar{\psi}_0 := \max_u |\psi_0|$; the endpoint differences are $\delta H := H(\theta(\frac{3}{4})) - H(\theta(\frac{1}{4}))$ and $\delta P_4 := P_4(\theta(\frac{3}{4})) - P_4(\theta(\frac{1}{4}))$. The last row gives the leading-order coefficients.

ε	$\bar{\psi}$	$\bar{\psi}_0/\varepsilon^3$	$\delta H/\varepsilon$	$\delta P_4/\varepsilon^3$
0.1	4.36888	0.02796	0.076036	−0.032720
0.05	4.36727	0.02258	0.077255	−0.034877
0.02	4.36635	0.02002	0.077970	−0.036173
0.01	4.36605	0.01925	0.078208	−0.036605
0.001	4.36578	0.01859	0.078424	−0.036994
0.0001	4.36576	0.01853	0.078445	−0.037033
leading order			$h_* = 0.0784478947$	$-1/27 = -0.037037$

of the cubic is computed exactly, and evaluates h in interval arithmetic on the brackets; the integral of Lemma 15(iv), evaluated by quadrature, lies in the enclosure.

Variance of the three-star average on query graphs. Table 5 checks the leading term $\rho^3/S_3(G)$ of Lemma 33(iv) by Monte Carlo on five graphs with $n = 60$: the complete graph, a circulant 20-regular graph, the hub design, a uniform random graph with $Q = 600$ edges and a union of random three-stars with $Q = 600$ edges. One graph per design is held fixed across replicates and separations; $q = 0.3$; the labels are i.i.d. from the path point $\theta(1/2)$ with $\varepsilon = 0.1$; each cell uses 20,000 replicates, and the standard error of the ratio is computed from the empirical fourth central moment. The three-noise term has the exact normalized value $R_3(\Delta) = \sum_c \theta_c \big((v_0 + (v_1 - v_0)\theta_c)/\max(v_0, v_1)\big)^3$ with $v_0 := q(1-q)$ and $v_1 := (q+\Delta)(1-q-\Delta)$, which equals $1 - \frac{3(1-P_2)(1-2q)}{q(1-q)}\Delta + O(\Delta^2)$, here $1 - \frac{37}{35}\Delta + O(\Delta^2)$. The estimated ratios decrease across the six separations and approach 1; several cross below 1 within Monte Carlo error, as expected from $R_3(\Delta) < 1$. At the larger separations the ratios exceed 1 because the other terms of Lemma 33(iv), of relative orders $d_{\max}/\rho$, $d_{\max}^2/\rho^2$ and n^3/ρ^3, are not yet small. Lemma 33 holds for every channel, so all six separations test it; only the last two also meet the condition $n\Delta^2 \leq q_0^2/4$ of Theorem 7, which here, with $q_0 \leq 0.3$, requires $\Delta \leq 0.0194$.

Risk of the estimator. Table 6 reports the mean squared error of Algorithm 1 in simulation, with the grid projection of Remark 32: $N = \lceil 22\sqrt{n} \rceil$, the points of Θ_4^N with sorted coordinates (the moments are symmetric), and the nearest point to $\widehat{P}$ found with a k-d tree over their moment vectors. The verdicts follow (1) on the complete graph with $q = 0.3$; n runs from 100 to 1600 along $n\Delta^2 = 1$, and Δ from 0.025 to 0.2 at $n = 400$; each cell uses 500 replicates. The six distributions are an interior point, a point with a zero coordinate, a point with the same P_2 and P_3 and a different entropy (the pair of Section 1), a vertex, and the path points $\theta(\frac{1}{4})$, $\theta(\frac{3}{4})$ of Lemma 15 with rare-class scale $\varepsilon_n = \min\{0.3, \frac{1}{2}(n^4\Delta^6)^{-1/6}\}$, of the order of the least-favourable amplitude. Along $n\Delta^2 = 1$ the error decreases at every distribution. The largest of the six errors falls from 3.43×10^{-2} at $n = 100$ to 5.62×10^{-3} at $n = 1600$, a slope of -0.65 in $\log n$, and the slopes for the individual distributions lie between -0.61 and -0.80. This decay is faster than $n^{-1/3}$, the order of the bound of Theorem 1 along $n\Delta^2 = 1$, and in every cell the error is below $0.26\,(n^4\Delta^6)^{-1/3}$. The rate is determined only up to constants and no numerical value of the lower-bound constant is given, so these ratios do not show whether a tested distribution is close to the worst case, and the simulation does not test the rate. The term $178V_4^{1/3}$ of the bound of Remark 32 is at least 16.5 in every cell, above the value $(\log 4)^2$ that caps the bound. The projection is active: an exact test, which checks by root isolation over the rationals that every root of the quartic of Algorithm 1 built from $\widehat{P}$ is real and nonnegative, finds $\widehat{P}$ outside $\mathcal{M}_4$ in 61% to 100% of the replicates of a cell. The estimate is biased at the boundary distributions tested (Table 7): the squared bias is up to 41% of the mean squared error at the vertex and 36% at the point with a zero coordinate. The certified tolerance $4\sqrt{29}/N$ lies between 0.025 and 0.098; refining the grid point with a local solver lowers the distance to $\widehat{P}$ by at most 6.2×10^{-3} and changes the error of a cell by at most 2.6%.

Table 5: Variance of the three-star average $\hat{P}_4^G$ on five query graphs G with Q edges and $S_3(G) = \sum_i \binom{d_i}{3}$ three-stars, normalized by its predicted leading term $\rho^3/S_3(G)$, $\rho = \max\{p(1-p), q(1-q)\}/\Delta^2$: $n = 60$, $q = 0.3$, labels i.i.d. from the path point $\theta(1/2)$ with $\varepsilon = 0.1$, one graph per design held fixed across replicates and separations, 20,000 Monte Carlo replicates per cell; standard errors in parentheses, computed from the empirical fourth central moment. The row $R_3(\Delta)$ is the exact graph-independent factor of the three-noise term.

design	Q	$S_3(G)$	$S_3(G)\,\mathrm{Var}(\hat{P}_4^G)/\rho^3$ at $\Delta =$		
			0.2	0.1	0.05
complete graph	1770	1950540	265.4 (2.8)	17.2 (0.2)	3.003 (0.043)
20-regular (circulant)	600	68400	25.4 (0.3)	3.639 (0.052)	1.464 (0.020)
hub design	600	366529	132.8 (1.6)	12.5 (0.2)	2.741 (0.064)
Erdős–Rényi $G(n, Q)$	600	78898	31.2 (0.4)	4.048 (0.058)	1.539 (0.024)
union of random three-stars	600	81729	32.6 (0.4)	4.259 (0.065)	1.575 (0.026)
three-noise factor $R_3(\Delta)$			0.9185	0.9351	0.9590
design	Q	$S_3(G)$	$S_3(G)\,\mathrm{Var}(\hat{P}_4^G)/\rho^3$ at $\Delta =$		
			0.025	0.0125	0.00625
complete graph	1770	1950540	1.419 (0.021)	1.120 (0.017)	1.004 (0.013)
20-regular (circulant)	600	68400	1.124 (0.016)	1.000 (0.013)	0.992 (0.014)
hub design	600	366529	1.299 (0.029)	1.070 (0.025)	1.038 (0.022)
Erdős–Rényi $G(n, Q)$	600	78898	1.045 (0.016)	0.998 (0.014)	0.982 (0.014)
union of random three-stars	600	81729	1.125 (0.019)	1.036 (0.015)	0.989 (0.015)
three-noise factor $R_3(\Delta)$			0.9768	0.9876	0.9936

Table 6: Mean squared error ($\times 10^3$) of the entropy estimate of Algorithm 1 with the grid projection of Remark 32, standard errors in parentheses, 500 replicates per cell, $q = 0.3$. Top: $n\Delta^2 = 1$. Bottom: $n = 400$, so that $n\Delta^2 = 0.25, 1, 4, 16$. The distributions are $(0.4, 0.3, 0.2, 0.1)$, $(\frac{1}{2}, \frac{1}{4}, \frac{1}{4}, 0)$, $(0.45, 0.4, 0.1, 0.05)$, $(1, 0, 0, 0)$ and the path points of Lemma 15 with $\varepsilon_n = \min\{0.3, \frac{1}{2}(n^4\Delta^6)^{-1/6}\}$. The last two rows of each block give the reference order $(n^4\Delta^6)^{-1/3}$ and, over the six distributions, the smallest and largest error of the plug-in entropy of the true labels. The cell $n = 400$, $\Delta = 0.05$ appears in both blocks as two independent runs.

n, with $n\Delta^2 = 1$	100	200	400	800	1600
interior point	19.3 (1.29)	11.6 (0.88)	7.79 (0.54)	4.54 (0.33)	2.56 (0.21)
zero coordinate	30.4 (1.73)	17.3 (1.03)	12.5 (0.74)	6.95 (0.43)	5.62 (0.32)
same P_2, P_3	26.2 (1.66)	14.6 (0.87)	9.84 (0.51)	5.91 (0.28)	4.67 (0.22)
vertex	18.4 (1.41)	10.9 (0.89)	5.77 (0.46)	3.62 (0.32)	2.27 (0.19)
path point $\theta(\frac{1}{4})$	26.4 (1.50)	14.5 (0.83)	9.06 (0.47)	6.00 (0.30)	3.79 (0.20)
path point $\theta(\frac{3}{4})$	34.3 (2.18)	16.6 (0.82)	9.91 (0.52)	6.81 (0.40)	3.76 (0.23)
$(n^4\Delta^6)^{-1/3}$	215.4	171.0	135.7	107.7	85.5
true labels, range	0.00–8.35	0.00–3.97	0.00–2.10	0.00–1.24	0.00–0.62
Δ, with $n = 400$	0.025	0.05	0.1	0.2	
interior point	17.1 (1.28)	7.46 (0.54)	2.73 (0.19)	0.83 (0.05)	
zero coordinate	27.1 (1.57)	10.9 (0.70)	4.27 (0.28)	1.60 (0.11)	
same P_2, P_3	21.4 (1.37)	8.66 (0.50)	4.61 (0.27)	1.98 (0.12)	
vertex	24.8 (1.88)	5.61 (0.45)	2.02 (0.17)	0.62 (0.05)	
path point $\theta(\frac{1}{4})$	22.3 (1.06)	8.36 (0.44)	3.70 (0.21)	2.02 (0.14)	
path point $\theta(\frac{3}{4})$	24.9 (1.39)	9.81 (0.54)	4.17 (0.21)	2.16 (0.13)	
$(n^4\Delta^6)^{-1/3}$	542.9	135.7	33.9	8.48	
true labels, range	0.00–1.74	0.00–2.43	0.00–2.29	0.00–1.91	

Table 7: Signed bias ($\times 10^3$) of the entropy estimate of Algorithm 1 in the cells of Table 6, standard errors in parentheses.

n, with $n\Delta^2 = 1$	100	200	400	800	1600
interior point	$-29.5\,(6.1)$	$-17.4\,(4.8)$	$-15.5\,(3.9)$	$-9.3\,(3.0)$	$-7.7\,(2.2)$
zero coordinate	$+52.8\,(7.4)$	$+52.9\,(5.4)$	$+52.1\,(4.4)$	$+40.5\,(3.3)$	$+45.2\,(2.7)$
same P_2, P_3	$-7.6\,(7.2)$	$-18.3\,(5.4)$	$-22.6\,(4.3)$	$-22.4\,(3.3)$	$-33.7\,(2.7)$
vertex	$+82.1\,(4.8)$	$+64.0\,(3.7)$	$+45.9\,(2.7)$	$+34.8\,(2.2)$	$+29.7\,(1.7)$
path point $\theta(\frac{1}{4})$	$-33.3\,(7.1)$	$-25.2\,(5.3)$	$-27.8\,(4.1)$	$-15.8\,(3.4)$	$-13.5\,(2.7)$
path point $\theta(\frac{3}{4})$	$-64.2\,(7.8)$	$-51.7\,(5.3)$	$-41.1\,(4.1)$	$-35.9\,(3.3)$	$-25.8\,(2.5)$
Δ, with $n = 400$	0.025	0.05	0.1	0.2	
interior point	$-30.4\,(5.7)$	$-13.3\,(3.8)$	$-4.2\,(2.3)$	$+0.2\,(1.3)$	
zero coordinate	$+74.2\,(6.6)$	$+45.8\,(4.2)$	$+31.4\,(2.6)$	$+17.4\,(1.6)$	
same P_2, P_3	$-1.2\,(6.5)$	$-19.9\,(4.1)$	$-18.0\,(2.9)$	$-6.8\,(2.0)$	
vertex	$+100.4\,(5.4)$	$+45.6\,(2.7)$	$+27.5\,(1.6)$	$+14.4\,(0.9)$	
path point $\theta(\frac{1}{4})$	$-42.6\,(6.4)$	$-21.8\,(4.0)$	$-19.8\,(2.6)$	$-11.5\,(1.9)$	
path point $\theta(\frac{3}{4})$	$-58.8\,(6.6)$	$-34.3\,(4.2)$	$-21.8\,(2.7)$	$-15.0\,(2.0)$	

Table 8: Mean squared error ($\times 10^3$) of the spectral-clustering plug-in forced to four clusters, on the verdicts of Table 6, standard errors in parentheses; below each distribution, the squared error $(\log 4 - H(\theta))^2$ ($\times 10^3$) of the constant estimate $\log 4$. The column k gives the number of nonzero coordinates of θ; the path points change with the cell through ε_n.

n, with $n\Delta^2 = 1$	k	100	200	400	800	1600
interior point	4	7.54 (0.12)	9.14 (0.08)	9.98 (0.05)	10.6 (0.02)	11.0 (0.01)
constant $\log 4$		11.3	11.3	11.3	11.3	11.3
zero coordinate	3	105.8 (0.50)	111.9 (0.37)	115.7 (0.17)	117.7 (0.08)	118.8 (0.05)
constant $\log 4$		120.1	120.1	120.1	120.1	120.1
same P_2, P_3	4	66.3 (0.44)	71.9 (0.25)	74.9 (0.14)	76.8 (0.06)	77.7 (0.03)
constant $\log 4$		78.6	78.6	78.6	78.6	78.6
vertex	1	1868.9 (1.98)	1889.7 (1.72)	1905.2 (0.65)	1913.0 (0.35)	1917.2 (0.18)
constant $\log 4$		1921.8	1921.8	1921.8	1921.8	1921.8
path point $\theta(\frac{1}{4})$	4	422.4 (1.02)	503.2 (0.66)	583.0 (0.33)	660.6 (0.18)	736.7 (0.09)
constant $\log 4$		448.5	519.6	592.4	665.9	739.3
path point $\theta(\frac{3}{4})$	4	399.8 (1.11)	481.5 (0.64)	562.8 (0.33)	641.1 (0.18)	717.9 (0.10)
constant $\log 4$		426.2	498.0	571.6	646.1	720.6
Δ, with $n = 400$	k	0.025	0.05	0.1	0.2	
interior point	4	10.0 (0.05)	9.93 (0.05)	9.97 (0.05)	6.77 (0.12)	
constant $\log 4$		11.3	11.3	11.3	11.3	
zero coordinate	3	115.6 (0.17)	115.7 (0.15)	115.8 (0.16)	103.0 (1.09)	
constant $\log 4$		120.1	120.1	120.1	120.1	
same P_2, P_3	4	75.1 (0.12)	74.9 (0.14)	74.5 (0.14)	69.4 (0.29)	
constant $\log 4$		78.6	78.6	78.6	78.6	
vertex	1	1903.9 (0.66)	1905.1 (0.62)	1906.1 (0.64)	1906.4 (0.57)	
constant $\log 4$		1921.8	1921.8	1921.8	1921.8	
path point $\theta(\frac{1}{4})$	4	295.0 (0.25)	583.5 (0.35)	1005.3 (1.06)	1344.1 (0.45)	
constant $\log 4$		302.4	592.4	1018.7	1356.6	
path point $\theta(\frac{3}{4})$	4	273.1 (0.23)	562.6 (0.35)	992.3 (0.46)	1334.0 (0.54)	
constant $\log 4$		279.6	571.6	1004.5	1348.3	

On the same verdicts we also run a spectral-clustering plug-in forced to four clusters: k-means with $K = 4$ on the four leading eigenvectors of $A - q(\mathbf{1}\mathbf{1}^\top - I)$, followed by the entropy of the cluster proportions (the kind of plug-in that Liu (2026) analyses where clustering errors are small). Its mean estimates lie between 1.358 and 1.385 in every cell, near $\log 4 \approx 1.386$ at all six distributions. With K fixed at four it splits the answers into four groups even at the vertex, so the comparison covers this one choice of K on the cells tested. Table 8 gives its error at each distribution, with the number k of nonzero coordinates of θ, next to the squared error $(\log 4 - H(\theta))^2$ of the constant estimate $\log 4$, which ignores the verdicts. In every cell its error lies below the constant's, by at most 5.3×10^{-2}, and along $n\Delta^2 = 1$ the difference shrinks with n at every distribution, to at most 3.1% of the constant's error at $n = 1600$. Compared with Algorithm 1, its error is smaller only at the interior point, whose entropy 1.28 is near $\log 4$, for $n \le 200$ along $n\Delta^2 = 1$ and at $n\Delta^2 = 1/4$; at $n = 1600$ it is 4.3 to 845 times larger. The plug-in entropy of the true labels, which uses information the verdicts do not carry, has error at most 8.4×10^{-3}.

Cost. Each verdict is one call to the judge. The complete graph takes $\binom{n}{2}$ calls, 1.28×10^6 at $n = 1600$; a query graph takes one call per edge (Section 4); and a channel estimated from labelled pairs adds $2m$ calls on pairs whose meanings are known (Corollary 31(ii)). The grid of Remark 32 and its k-d tree depend only on n, so one build serves every prompt with n answers, and each estimate then needs the three statistics and one nearest-point query. For n from 100 to 1600 the grid has 7.9×10^4 to 4.8×10^6 points with sorted coordinates. On one core of an AMD Ryzen Threadripper PRO 5965WX, one measurement at a time, building the grid with its tree takes 0.07 to 2.5 seconds and raises the peak memory of the process by 11 to 640 MiB, about 140 bytes per point, and the median time per estimate is 0.08 to 0.95 milliseconds, of which the statistics take at most 0.6. At $n = 1600$ one build takes as long as about 2700 estimates.