# SEMANTIC ENTROPY IS ESTIMABLE BELOW THE RECOVERY THRESHOLD: MINIMAX RATES UNDER A NOISY JUDGE

**Anonymous authors**
Paper under double-blind review

## ABSTRACT

Semantic entropy scores a language model's uncertainty by the entropy of the distribution from which its $n$ sampled answers draw their meanings, and those meanings are never observed: an entailment judge is called on pairs and the answers are grouped by its verdicts. That grouping can fail when the judge is too weak for the available sample size. We model the judge as a channel that, given the meanings, declares each pair equivalent independently with probability $q+\Delta$ when the two answers share a meaning and $q$ otherwise. For at most three meanings, a known channel with $q \in [q_0, 1-q_0]$ and a weakly separated judge, $n\Delta^2 \le c_1$, we determine the minimax squared risk of the entropy: $R_n^* \asymp \min\{1, (n^3\Delta^4)^{-1/2}\}$, with both constants depending only on $q_0$. The rate is attained by an estimator that never groups the answers. It reads the collision probability off the agreement density and the triple-collision probability off a centred statistic in the squared degrees, projects the pair onto the set of valid moments, and solves a cubic. The matching lower bound is proved along a family that holds the agreement density fixed while the entropy changes, so every linear statistic of the verdicts has the same mean at its endpoints. With $\Delta = n^{-\gamma}$, the entropy is uniformly consistently estimable exactly for $\gamma < 3/4$, and on $\gamma \in (1/2, 3/4)$ the minimax risk is of order $n^{(4\gamma-3)/2}$. For every $\gamma > 1/2$, below the recovery threshold $n\Delta^2 \asymp 1$, no estimator of the meanings improves asymptotically on assigning each answer to the most likely one, even when given all other answers' meanings. On $\gamma \in (1/2, 3/4)$ the entropy is therefore estimable while the grouping is unrecoverable.

## 1 INTRODUCTION

Semantic entropy scores a language model's uncertainty about a question by the entropy of the distribution over the *meanings* of its sampled answers, rather than over their surface forms (Kuhn et al., 2023). The meanings are never observed. In practice a pipeline samples $n$ answers, calls an entailment judge on pairs of them, groups the answers by the judge's verdicts, and plugs the group proportions into the entropy. Everything therefore rests on the grouping, and the grouping rests on how sharply the judge separates two answers to the same question.

Model the judge as a channel: two answers sharing a meaning are declared equivalent with probability $p$, two answers with different meanings, with probability $q < p$, and the verdicts are conditionally independent given the meanings. When the separation $\Delta = p - q$ is large, the verdict graph is a stochastic block model with recoverable blocks, and plugging in their proportions is sound (Liu, 2026). To ask what happens when it is not, let $\Delta = \Delta_n$ shrink. Two limits appear: once $n\Delta_n^2 \to 0$, below the recovery threshold $n\Delta^2 \asymp 1$, no estimator of the meanings beats the majority rule (every answer to the most likely meaning), while the entropy stays estimable as long as $n^3\Delta_n^4 \to \infty$.

The reason is that the entropy needs only the proportions of the meanings, and their moments can be estimated where the grouping cannot be recovered. The density of agreements estimates the collision probability $P_2 = \sum_c \theta_c^2$, a centred statistic in the squared degrees estimates the triple-collision probability $P_3 = \sum_c \theta_c^3$, and with at most three meanings those two numbers determine the entropy. Both average over all $\binom{n}{2}$ verdicts, so they tolerate a per-verdict signal far below what deciding one answer's meaning requires. This paper determines exactly how far. The rate is set



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.

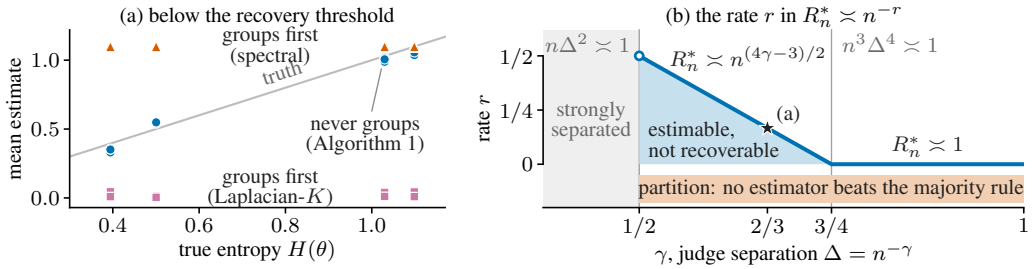



Figure 1: Where the semantic entropy is estimable and where the meaning partition is not. **(a)** Mean outputs in the six settings of Table 1 ($\Delta = n^{-2/3}$, $n\Delta^2 \le 0.13$), four distributions, each at its true entropy: Algorithm 1 tracks it, while spectral and Laplacian-$K$, which group first, stay near $\log 3$ and 0 on all four. **(b)** For at most three meanings and a known channel, $q \in [q_0, 1-q_0]$, $\Delta = n^{-\gamma}$, the minimax squared risk of the entropy is $R_n^* \asymp n^{-r}$ with $r = (3-4\gamma)/2$ for $\gamma \in (1/2, 3/4)$ and $r = 0$ for $\gamma \ge 3/4$ (Corollary 3 and Proposition 6). At a fixed $\theta$ with $\max_c \theta_c < 1$ and every $\gamma > 1/2$, no estimator of the meanings beats the majority rule, per answer and for the whole grouping under the permutation-invariant Hamming loss (Propositions 7 and 8). The star marks the setting of (a).

near a point mass, where the entropy is only Hölder-$\frac{1}{2}$ in $(P_2, P_3)$: the lower bound compares two distributions with equal $P_2$, which give every linear statistic of the verdicts the same mean (Section 4.2).

**Contributions.** Throughout, the channel $(q, \Delta)$ is known, $q \in [q_0, 1-q_0]$ for a fixed $q_0 \in (0, 1/2]$, there are at most three meanings, and the weakly separated regime is $n\Delta^2 \le c_1$ with $c_1$ depending only on $q_0$. The channel can also be calibrated from labelled pairs at no cost in rate.

- We determine the minimax squared risk of the semantic entropy on the whole weakly separated region, $R_n^* \asymp \min\{1, (n^3\Delta^4)^{-1/2}\}$ with constants depending only on $q_0$; it vanishes exactly for $\gamma < 3/4$ when $\Delta_n = n^{-\gamma}$.
- The rate is attained by a projected two-moment estimator that reads the graph once and solves a cubic, with no tuning parameter and no spectral step; its upper bound holds with explicit constants at every $n$ and every channel.
- The matching lower bound holds over every function of the verdicts. Its core is a path at constant agreement density, along which a Glauber identity and a blockwise Efron–Stein bound control the Fisher information of the graph despite the hidden meanings.
- On $\gamma \in (1/2, 3/4)$ the entropy is estimable at rate $n^{(4\gamma-3)/2}$, while at every fixed $\theta$ with $\max_c \theta_c < 1$ no estimator of the meanings improves asymptotically on the majority rule, per answer even with an oracle that reveals every other meaning, and for the whole grouping under the permutation-invariant Hamming loss.

The method covers at most three meanings. With two meanings the agreement density already determines the entropy; with four, two explicit distributions share both moments and differ in entropy by 0.066, so no function of $(P_2, P_3)$ can be the entropy (Section 5.2). Three is the first case in which the density alone is not enough and the last in which two moments are.

## 2 Related work

**Semantic entropy for language-model uncertainty.** Kuhn et al. (2023) introduced the object studied here and the bidirectional-entailment clustering that produces it; Farquhar et al. (2024) scaled it to hallucination detection and gave the discrete form estimated here, the entropy of the cluster proportions. Later work varies the object or the cost, not the observation model: Nikitin et al. (2024) replace hard clusters by a similarity kernel and a von Neumann entropy, Nguyen et al. (2025) use pairwise similarities in place of a partition, Lin et al. (2024) score uncertainty by the mean degree of the similarity graph, affine in $\widehat{P}_2$ for binary verdicts, Ciosek et al. (2025) estimate the same entropy from fewer samples with a Bayesian posterior, McCabe et al. (2026) and McCabe & Huang (2026)

target the alphabet size and the small-sample regime, and Chen et al. (2026) argue that the enterprise is unsupervised clustering. None of them models the judge's noise and asks what remains estimable.

**The closest prior result.** Liu (2026) is, to our knowledge, the only prior work with this observation model and this target. Its estimator builds the Laplacian of the verdict matrix, takes its first $K$ eigenvectors, assigns answers to clusters by an approximate $k$-means step, and plugs the cluster proportions into the Shannon entropy. Its main guarantee (Theorem 3.4 of arXiv version 3) bounds the absolute error of the population entropy with probability $1 - 3/n$, at leading order $\sqrt{\log n/n}$ for a fixed number of classes with positive masses; exact recovery is not needed, since the misclustering count enters that bound linearly and a spectral-clustering error rate is substituted for it. The two analyses cover different regimes. That analysis applies where a spectral-clustering error is small, needs $K$ but not the channel, and proves no lower bound. We work where no grouping beats the majority rule (Propositions 7 and 8), need the channel but only the bound of three meanings, and pair the upper bound with a matching lower bound over every estimator.

**Moment methods for graph mixtures.** The two identities we use appear in Ambroise & Matias (2012), who credit the moment-equation approach to Frank & Harary (1982); they and Bickel et al. (2011) estimate block-model parameters from edge and subgraph counts, with asymptotic normality at nondegenerate fixed parameters and $\sqrt{n}$-consistency in sparse graphs, respectively; Ambroise & Matias also note that for at most three groups, $(P_2, P_3)$ fixes the unordered proportions, which they estimate without recovering labels. Neither gives a risk bound for a functional or any lower bound, and neither treats a contrast vanishing at a constant edge density, where the Bernoulli noise dominates the label signal and the entropy loses smoothness at the boundary of the simplex. Allman et al. (2011) study identifiability for this family, and Castillo & Orbanz (2022) show that another block-model functional is also estimated slowly, by a different mechanism. Closest in construction is Wu & Yang (2020), who project estimated moments onto the set of valid moment vectors before inverting, as Bickel et al. do for graph moments; we use the same device, where the Hölder-$\frac{1}{2}$ modulus of the entropy in the moment pair sets the rate.

**Where recovery works, and the minimax toolbox.** Recovery guarantees and thresholds for block models and for clustering from noisy pairwise queries are known in detail (Lei & Rinaldo, 2015; Zhang & Zhou, 2016; Mazumdar & Saha, 2017; Del Pia et al., 2022; Xia & Huang, 2022), including repeated, adaptive queries of the same pair (Teo et al., 2026); Section 5 shows non-recovery with one verdict per pair. Below the threshold of the sparse two-block model, even its connection parameters are not consistently estimable (Mossel et al., 2015); with $q$ unknown, Proposition F.5 is a related obstruction for the entropy in the dense weak-contrast regime. With observed labels the entropy on a fixed alphabet has minimax rate $1/n$ (Wu & Yang, 2016; Jiao et al., 2015); seeing the labels only through the judge costs a factor $n^{2\gamma-1/2}$ for $1/2 < \gamma < 3/4$, due to the Hölder-$\frac{1}{2}$ modulus rather than to polynomial approximation. Detecting block structure against an Erdős–Rényi null is well studied (Banks et al., 2016; Gao & Lafferty, 2017; Jin et al., 2021); a common null reaches only the saturated branch (Theorem 5), so we compare two planted models with equal edge density. Appendix A adds notation and related work.

## 3 Setup and the estimator

**Model.** Let $Z_1, \dots, Z_n$ be independent with $\mathbb{P}(Z_i = c) = \theta_c$, where $\theta \in \Delta_3 := \{\theta \in [0,1]^3 : \sum_c \theta_c = 1\}$; zero coordinates are allowed, so a distribution on one or two meanings is a legitimate parameter. The judge is called once on each pair, and given $Z$ its verdicts are conditionally independent,

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

with $A_{ji} := A_{ij}$ and $A_{ii} := 0$. Write $p := q + \Delta$, so $\Delta = p - q$ is the judge's separation. The estimand is the population entropy $H(\theta) := -\sum_c \theta_c \log \theta_c$, not the entropy of the class frequencies realized by these $n$ answers, and the loss is squared error:

$$R_n^*(q, \Delta) \;:=\; \inf_{\widehat{H}} \sup_{\theta \in \Delta_3} \mathbb{E}_\theta \big(\widehat{H} - H(\theta)\big)^2, \tag{2}$$

the infimum over all functions of $A$, randomized independently of $A$ if desired. Throughout, $P_k := \sum_c \theta_c^k$; $e_2 := (1 - P_2)/2$ and $e_3 := (1 - 3P_2 + 2P_3)/6$ are the elementary symmetric polynomials

**Algorithm 1** Projected two-moment entropy estimator. $M_3 := \{(P_2(\theta), P_3(\theta)) : \theta \in \Delta_3\}$.

1: from $A$ compute $|E|$ and $d_1, \dots, d_n$, and form $\widehat{P} = (\widehat{P}_2, \widehat{P}_3)$ by (4)
2: if $\widehat{P} \in M_3$ set $\tilde{P} := \widehat{P}$; else let $\tilde{P}$ be a nearest point of $M_3$ to $\widehat{P}$ in Euclidean distance
3: solve $x^3 - x^2 + \tilde{e}_2 x - \tilde{e}_3 = 0$, giving $\widetilde{\theta} \in \Delta_3$
4: return $\widehat{H} := H(\widetilde{\theta})$

of $\theta$; $d_i := \sum_j A_{ij}$, $|E| := \sum_{i<j} A_{ij}$, $N_2 := \binom{n}{2}$, $N_3 := n(n-1)(n-2)$; and

$$\rho := \frac{\max\{p(1-p),\, q(1-q)\}}{\Delta^2}, \qquad V_n := \frac{1}{n} + \frac{\rho}{n^2} + \frac{\rho^2}{n^3}. \tag{3}$$

**Assumption 1.** The labels are i.i.d. from $\theta \in \Delta_3$; given the labels the verdicts are conditionally independent and follow (1); the channel $(q, \Delta)$ and the bound of three meanings are known to the estimator, and $q \in [q_0, 1 - q_0]$ for a fixed $q_0 \in (0, 1/2]$ wherever a lower bound is stated.

The estimator knows the bound of three meanings but not the support size of $\theta$.

**Two moments determine the entropy, and both are visible.** For at most three classes the pair $(P_2, P_3)$ determines $(e_2, e_3)$, hence the monic cubic $x^3 - x^2 + e_2 x - e_3$, hence the multiset $\{\theta_1, \theta_2, \theta_3\}$ and therefore $H(\theta)$. The agreement density estimates the collision probability $P_2$, and a centred statistic in the squared degrees estimates the triple-collision probability $P_3$:

$$\widehat{P}_2 := \frac{1}{\Delta}\Big(\frac{|E|}{N_2} - q\Big), \qquad \widehat{P}_3 := \frac{1}{N_3 \Delta^2} \sum_{i=1}^{n} \Big[\big(d_i - q(n-1)\big)^2 - d_i(1-q)^2 - (n-1-d_i)q^2\Big]. \tag{4}$$

Both are unbiased at every $\theta$, with variances $O(V_n)$ (Lemma B.1); the subtracted terms in $\widehat{P}_3$ remove the diagonal contribution of the Bernoulli noise. $\widehat{P}_3$ uses the *squared* degrees: the degree sum alone equals $2|E|$ and adds nothing to $\widehat{P}_2$. Neither statistic decides what any individual answer means.

Every step is explicit. $M_3$ is the semialgebraic set $\{e_2 \geq 0,\ e_3 \geq 0,\ \mathrm{disc} \geq 0\}$, where disc is the discriminant of the cubic, and its boundary is the union of the two curves traced by $(1-t, t, 0)$ and $(a, a, 1-2a)$, so step 2 minimizes over two curves in the plane; the roots in step 3 are automatically non-negative and sum to one (Appendix B). Reading the graph dominates, so Algorithm 1 costs $O(n^2)$ and has no tuning parameter, no spectral decomposition and no approximation order.

**A non-asymptotic upper bound.** The entropy is uniformly Hölder-$\frac{1}{2}$ in the moment pair.

**Lemma 1** (proved in Appendix B). *For $\theta, \theta' \in \Delta_3$ put $D := |e_2(\theta) - e_2(\theta')| + |e_3(\theta) - e_3(\theta')|$. Then $D \leq 10/27$, and $H(\theta) = H(\theta')$ when $D = 0$, while for $D > 0$, $|H(\theta) - H(\theta')| \leq D \log 2 + \log(1 + D) + 2\sqrt{D} \arctan(1/\sqrt{D}) \leq \pi\sqrt{D}$.*

The proof compares $\prod_c (u + \theta_c)$ with $\prod_c (u + \theta'_c)$ inside $H(\theta) - H(\theta') = \int_0^\infty \log \frac{\prod_c (u+\theta_c)}{\prod_c (u+\theta'_c)} \mathrm{d}u$, splitting at $u = 1$. The exponent cannot be improved: $M_3$ has a cusp at $(1, 1)$, the moment pair of a point mass and the origin of Figure 2, and $\theta_\varepsilon = (1 - \varepsilon, \varepsilon, 0)$ and $\theta'_\varepsilon = (1 - \varepsilon', \varepsilon'/2, \varepsilon'/2)$ with the same $P_2$ lie on its two sides, with $\sqrt{D} \sim \varepsilon/2$ while their entropies differ by $\varepsilon \log 2$ to first order (Remark B.3). That singular modulus at the boundary of the simplex, not the variance of the moments, is where the exponent of the minimax rate comes from.

**Theorem 2** (proved in Appendix B). *For every $n \geq 3$, every $0 \leq q < q + \Delta \leq 1$ and every $\theta \in \Delta_3$,*

$$\mathbb{E}_\theta\big(\widehat{H} - H(\theta)\big)^2 \ \leq\ \min\Big\{(\log 3)^2,\ 81\, V_n^{1/2}\Big\}.$$

The bound holds at every $n$ and every channel, with an explicit numerical constant; the cap $(\log 3)^2$ is the squared diameter of the range of $H$ on $\Delta_3$. When $\rho \gg n$ the last term of $V_n$ dominates and $V_n^{1/2} \sim \max\{p(1-p), q(1-q)\}\,(n^3\Delta^4)^{-1/2}$, the rate the next section shows is optimal.

# 4 THE MINIMAX RATE

Fix $q_0 \in (0, 1/2]$ and set $v_0 := \frac{q_0}{2}(1 - \frac{q_0}{2})$ and $c_1 := \min\{q_0^2/4,\ v_0/20\}$. The weakly separated regime is $n\Delta^2 \le c_1$: the degree shift a meaning produces, of order $n\Delta$, is smaller than the standard deviation $\sqrt{n\,q(1-q)}$ of the degree noise by a constant factor. A channel held fixed as $n$ grows leaves this regime, in which case Theorem 2 alone already gives risk of order $n^{-1/2}$.

## 4.1 THE RATE

**Corollary 3** (proved in Appendix D). *Let $q \in [q_0, 1 - q_0]$, $n \ge 3$ and $n\Delta^2 \le c_1$. Then*

$$c_\star \min\left\{1,\ (n^3\Delta^4)^{-1/2}\right\} \ \le\ R_n^*(q, \Delta) \ \le\ C_\star \min\left\{1,\ (n^3\Delta^4)^{-1/2}\right\},$$

*with $C_\star = \max\{81(c_1^2 + c_1/4 + 1/16)^{1/2}, (\log 3)^2\}$ and $c_\star > 0$ depending only on $q_0$. The upper bound is attained by Algorithm 1.*

So the squared risk is determined, up to constants depending only on $q_0$, by the single quantity $n^3\Delta^4$, on the whole weakly separated region. With $\Delta = n^{-\gamma}$ this is $n^{3-4\gamma}$, so the risk vanishes at rate $n^{(4\gamma-3)/2}$ for $\gamma < 3/4$ and is bounded below by a constant for $\gamma \ge 3/4$ (Figure 1). The upper half is Theorem 2 specialized to $\Delta \le \sqrt{c_1}n^{-1/2}$. The lower half has two pieces: Theorem 4 uses the constant-$P_2$ path below; Theorem 5 uses a common Erdős–Rényi reference.

**Theorem 4** (proved in Appendix D). *There are finite constants $C_0$ and $K_0$ depending only on $q_0$ such that for every $q \in [q_0, 1 - q_0]$ and every $(n, \Delta)$ with*

$$C_0^{1/4}\, n^{-3/4} \ \le\ \Delta \ \le\ \sqrt{c_1}\, n^{-1/2}, \tag{5}$$

*the minimax risk satisfies $R_n^*(q, \Delta) \ \ge\ \frac{(\log 2)^2}{64} K_0^{-1/2} (n^3\Delta^4)^{-1/2}$.*

**Theorem 5** (proved in Appendix C). *For every $q \in [q_0, 1 - q_0]$, $n \ge 2$ and $n\Delta^2 \le c_1$, $R_n^*(q, \Delta) \ \ge\ \frac{v_0^2(\log 2)^2}{576} \min\{1, (n^3\Delta^4)^{-1}\}$.*

In the exponent parametrization the window (5) covers the entire range $\gamma \in [1/2, 3/4]$: every fixed $\gamma$ strictly inside it satisfies (5) once $n$ is large, and the endpoints are reached with the prefactors $\sqrt{c_1}$ and $C_0^{1/4}$. Outside the window, where $n^3\Delta^4 < C_0$, the risk is of constant order and Theorem 5 already matches the cap of Theorem 2 up to a constant. Combining the three gives Corollary 3 with $c_\star = \min\{\frac{(\log 2)^2}{64} K_0^{-1/2},\ \frac{v_0^2(\log 2)^2}{576} \min\{1, C_0^{-1/2}\}\}$ and leaves no segment of the weakly separated region unmatched. The constant of Theorem 5 is an explicit number; the constant $K_0$ of Theorem 4, and with it $c_\star$ and the threshold $C_0$, is finite and depends on nothing but $q_0$.

**Proposition 6** (proved in Appendix C). *Let $q_n \in [q_0, 1 - q_0]$ and $0 < \Delta_n \le 1 - q_n$. Estimators with $\sup_{\theta \in \Delta_3} \mathbb{E}_\theta(\widehat{H}_n - H(\theta))^2 \to 0$ exist if and only if $n^3\Delta_n^4 \to \infty$. With $\Delta_n = n^{-\gamma}$ the critical exponent is $\gamma_c(3) = 3/4$, and no consistent estimator exists for $\gamma \ge 3/4$.*

Sufficiency is Theorem 2 and necessity is Theorem 5. The threshold is inclusive on the right: at $\gamma = 3/4$ the quantity $n^3\Delta_n^4$ equals 1 and does not diverge.

## 4.2 WHERE THE EXPONENT COMES FROM

The lower bound is a two-point argument, and the work goes into making the two points as far apart in entropy as possible while keeping the two graph distributions indistinguishable. The arithmetic is short:

$$\left|H(\theta_\varepsilon) - H(\theta'_\varepsilon)\right| \asymp \varepsilon, \qquad I(t) \ \le\ K_0\, \varepsilon^4\, n^3\Delta^4, \qquad \varepsilon \asymp (n^3\Delta^4)^{-1/4} \implies R_n^* \gtrsim \varepsilon^2, \tag{6}$$

where $I(t)$ is the Fisher information of the observed graph along a path joining the two points. Taking $\varepsilon$ as large as the middle bound allows makes the total variation at most $\frac{1}{2}$, and Le Cam's two-point inequality turns an entropy gap of order $\varepsilon$ into a squared risk of order $\varepsilon^2 = (n^3\Delta^4)^{-1/2}$, against every estimator that sees the whole graph. Four ingredients produce the middle bound.

**A path at constant agreement density.** The endpoints are $\theta_\varepsilon = (1-\varepsilon, \varepsilon, 0)$ and $\theta'_\varepsilon = (1-\varepsilon', \varepsilon'/2, \varepsilon'/2)$ with $\varepsilon'$ chosen so that $P_2$ agrees, joined by $\theta(t) = (1-e,\ e-\eta,\ \eta)$ with $\eta(t) = \frac{\varepsilon'}{2}t^2$, along which $P_2$ is constant (Figure 2). The expected edge count is therefore constant along the family, removing the one mean shift of order $n^2\Delta$.

**From total variation to Fisher information.** Along a smooth path the Hellinger distance, and hence the total variation, is controlled by $\int_0^1 \sqrt{I(t)}\,\mathrm{d}t$. This replaces a likelihood ratio on $2^{\binom{n}{2}}$ outcomes by a second moment at each $t$.

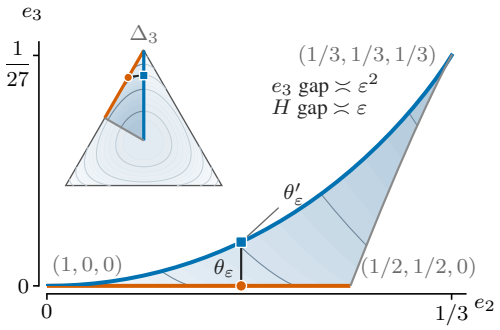



Figure 2: $M_3$ in $(e_2, e_3)$, shaded by $H$, and (inset) the part $\theta_1 \geq \theta_2 \geq \theta_3$ of $\Delta_3$ mapped onto it one-to-one. Its sides $\theta_3 = 0$ and $\theta_2 = \theta_3$ meet at an angle at $(1,0,0)$ and are tangent at its image; the pair $\theta_\varepsilon, \theta'_\varepsilon$ ($\varepsilon = 0.2$) spans the cusp.

**Rewriting the score.** The observed score is a conditional expectation over the hidden labels with no closed form. A Glauber identity rewrites it as an explicit function $\tilde{S}_1 = \sum_i \Phi(F_i)$ of the complete data, where $F_i$ is a local field at answer $i$; one Rao–Blackwell sweep then yields a statistic $\tilde{S}_{\mathrm{RB}}$ with $I(t) \leq \mathbb{E}_t \tilde{S}^2_{\mathrm{RB}}$, using orthogonality against the edge count and conditional Jensen. What remains is a second-moment estimate for an explicit function of $(Z, A)$.

**Efron–Stein, blockwise.** That second moment splits into the influences of one edge and of one label. Each is estimated with the labels held fixed, under the conditional law, and only then averaged over the label counts, which keeps the combinatorics of three classes out of the concentration step. The result is $\mathbb{E}_t \tilde{S}^2_{\mathrm{RB}} \leq K_0 \varepsilon^4 n \delta^2$ with $\delta = n\Delta^2$, the middle inequality of (6). This step needs $n\varepsilon \geq 1$, automatic at the amplitude of Theorem 4; Remark D.1 explains why the argument needs it.

## 5 Estimation without recovery

The rate of Corollary 3 is attained by an estimator that never decides what any single answer means. Once $n\Delta^2 \to 0$, no procedure recovers those decisions: under the losses of Propositions 7 and 8, a grouping step would report a partition no better than the majority rule.

### 5.1 The meaning partition is not recoverable below the recovery threshold

Fix $\theta \in \Delta_3$ with $\max_c \theta_c < 1$, so that assigning every answer to the most likely meaning errs with probability $1 - \max_c \theta_c > 0$. That is the baseline a recovery procedure has to beat.

**Proposition 7** (proved in Appendix E)**.** *Under Assumption 1, let $n \geq n_0$ and $n\Delta^2 \leq c_1$. For every randomized $\widehat{Z}_1 = \widehat{Z}_1(A, Z_{-1})$,*

$$\mathbb{P}\big(\widehat{Z}_1 \neq Z_1\big) \ \geq\ 1 - \max_c \theta_c - C\sqrt{n\Delta^2},$$

*with $C$ and $n_0$ depending only on $(q_0, c_1)$. The same bound holds for any estimator that sees only $A$, since $\sigma(A) \subseteq \sigma(A, Z_{-1})$.*

Here $Z_{-1}$ is the oracle: it reveals every other answer's meaning. The proof computes the posterior of $Z_1$ given $(A, Z_{-1})$ in closed form through the Glauber identity of Section 4.2 and shows that the field driving it has size $O(\sqrt{n\Delta^2})$, so the posterior sits within that distance of the prior. The grouping as a whole is not recoverable either, even up to relabelling.

**Proposition 8** (proved in Appendix E)**.** *Under Assumption 1 with $n \geq 2$ and no further condition on $(n, \Delta)$, let $L(\widehat{Z}, Z) := \min_{\pi \in S_3} \frac{1}{n} \sum_i \mathbf{1}\{\pi(\widehat{Z}_i) \neq Z_i\}$ be the permutation-invariant Hamming loss. Every randomized $\widehat{Z} = \widehat{Z}(A)$ satisfies*

$$\mathbb{E}\, L(\widehat{Z}, Z) \ \geq\ 1 - \max_c \theta_c - \sqrt{\frac{(n-1)\Delta^2}{4q_0(1-q_0)} + \frac{\log 6}{2n}}.$$

The proof bounds $\sum_i I(Z_i; A) \le I(Z; A) \le \binom{n}{2} \frac{\Delta^2}{q_0(1-q_0)}$ by the variational form of mutual information, converts it by Pinsker, and pays $\log 6$ for the freedom to relabel. As $n\Delta^2 \to 0$ the bound tends to the majority baseline.

**Corollary 9** (proved in Appendix E). *Let $q_n \in [q_0, 1-q_0]$, $\Delta_n = n^{-\gamma}$ with $\gamma \in (1/2, 3/4)$, and $\theta$ fixed with $\max_c \theta_c < 1$. Then $n\Delta_n^2 \to 0$ and $n^3\Delta_n^4 \to \infty$, so Algorithm 1 is uniformly consistent at the rate $R_n^* \asymp n^{(4\gamma-3)/2} \to 0$, while the optimal expected loss of any label estimator tends to $1 - \max_c \theta_c$, per answer even against the oracle of Proposition 7 and for the whole grouping under $L$.*

Every fixed $\gamma \in (1/2, 3/4)$ therefore lies where the rate is matched and the grouping cannot be recovered (Figure 1); at $\Delta \asymp n^{-1/2}$ the oracle can beat the baseline by a constant (Remark E.4). Recovery stays at the baseline above $3/4$ too, but there the entropy is no longer estimable.

These statements bound recovery of the partition, not the entropy risk of every procedure that outputs one: a procedure may estimate the proportions accurately and then assign answers arbitrarily, recovering nothing while reporting a good entropy. And the least-favourable family of Theorem 4 moves toward a vertex of the simplex, where the majority rule itself has vanishing loss, so that family cannot carry a recovery statement; hence Propositions 7 and 8 fix $\theta$ with $\max_c \theta_c < 1$.

### 5.2 Why three meanings

With $K = 2$ meanings, $\theta = (1-t, t)$ gives $P_2 = 1 - 2t + 2t^2$, whose solutions $t = \frac{1}{2}(1 \pm \sqrt{2P_2 - 1})$ are mirror images of each other and carry the same entropy, so the edge density already determines $H$ and $P_3$ is unnecessary. With $K = 3$, the pair $(P_2, P_3)$ determines the multiset of coordinates, and the resulting map is only Hölder-$\frac{1}{2}$ at the boundary (Lemma 1), which sets the exponent $-1/2$ of Corollary 3. With $K = 4$, two moments no longer determine the entropy:

$$\theta = (0.5,\ 0.25,\ 0.25,\ 0) \qquad \text{and} \qquad \theta' = (0.45,\ 0.4,\ 0.1,\ 0.05)$$

have the same $P_2 = 0.375$ and the same $P_3 = 0.15625$, while $H(\theta) = 1.039721$ and $H(\theta') = 1.105890$ differ by $0.066169$.

This is the boundary of the two-moment method, not of identifiability: with four meanings $H$ is a function of $(P_2, P_3, P_4)$, and $P_4$ is in principle estimable from four-cycle counts (Remark B.9).

## 6 Numerical study

The numerics below illustrate the shape of the theorems above. Everything runs on CPU with synthetic graphs from (1); Appendix G gives every configuration, distribution and repetition count, with standard errors for the tabulated risks.

**The predicted shape.** Figure 3 measures the risk of Algorithm 1 on a grid laid out directly in $n^3\Delta^4$, which spans a factor of 200 across the 22 cells while $n\Delta^2$ stays inside one decade. The ratio of measured risk to $V_n^{1/2}$ lies in $[0.139, 0.348]$ over all cells and in $[0.139, 0.181]$, with median $0.156$, over the ten with $n^3\Delta^4 \ge 45$; at fixed $n^3\Delta^4 \ge 90$ its point estimates agree across $n$ within $5\%$, inside their standard errors of 6 to $8\%$. Against $(n^3\Delta^4)^{-1/2}$ the ratio varies by a factor of $1.9$ on those ten cells, so the data follow the finite-sample form $V_n$ rather than its leading term; below $n^3\Delta^4 = 45$ the ratio to $V_n^{1/2}$ climbs as $n^3\Delta^4$ falls. Every cell lies in the reach of Theorem 2; the lower bounds are proved for $n\Delta^2 \le c_1$, which at $q_0 = 0.3$ is $0.006375$, below every cell.

**Against clustering baselines.** Table 1 compares Algorithm 1 with the route that groups first, in two instantiations: the centred adjacency matrix, and the unnormalized Laplacian of the verdict matrix with its first $K$ eigenvectors, the construction of Liu (2026). Algorithm 1's worst risk is $16.9$ to $75.9$, $38.9$ to $182.4$ and $10.4$ to $46.4$ times smaller than those of the first, the second and the data-free constant, and it improves with $n$ while none of them does. The gap is consistent with Propositions 7 and 8: over the 24 combinations the first route's mean output lies between $1.0918$ and $1.0977$ against $\log 3 = 1.0986$, the second's between $0.0045$ and $0.0447$. The first splits into three nearly balanced blocks and saturates; the second collapses onto one (Figure 5 in Appendix G). The first has lower risk than Algorithm 1 on the uniform distribution in all six settings and on $(0.5, 0.3, 0.2)$ in five.

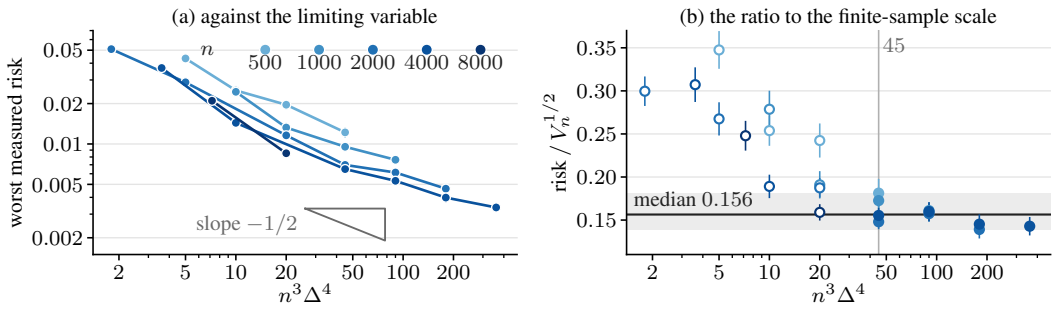



Figure 3: Maximum measured risk on 22 cells, $n \in \{500, \dots, 8000\}$, $n^3\Delta^4$ from 1.8 to 360, $n\Delta^2$ from 0.030 to 0.300, $q = 0.3$, 400 repetitions per cell and distribution. Risk is the largest mean squared error over five distributions, three fixed and the two endpoints of the least-favourable pair at that cell's amplitude; each error bar is one standard error of that particular mean squared error, not of the maximum. **(a)** against the limiting variable, one series per $n$. **(b)** the risk divided by $V_n^{1/2}$ from (3), the scale of Theorem 2's bound; the line is the median 0.156 over the ten cells with $n^3\Delta^4 \geq 45$ (filled markers) and the band their range, against the constant 81 of that bound.

Table 1: Largest mean squared error over $\theta = (0.9, 0.05, 0.05)$, $(0.8, 0.2, 0)$, $(0.5, 0.3, 0.2)$ and $(1/3, 1/3, 1/3)$ at $\Delta = n^{-2/3}$, 400 repetitions each. *moment* is Algorithm 1; *constant* always outputs $\frac{1}{2}\log 3$. The clustering baselines take ten k-means restarts and plug in the block proportions: *spectral* splits the centred adjacency matrix into three blocks, *spectral-$K$* into the true number of meanings (both maxima fall on the near-vertex distribution, where $K = 3$ and the two coincide), and *Laplacian-$K$* uses the first $K$ eigenvectors of the unnormalized Laplacian of the verdict matrix. *degree-CL* never groups: it maximizes, on a grid, the binomial-mixture composite likelihood of the degrees. *oracle* plugs in the true label frequencies, a comparator, not a lower bound.

| $n$ | $q$ | $n\Delta^2$ | moment | degree-CL | spectral | spectral-$K$ | Laplacian-$K$ | constant | oracle |
|---|---|---|---|---|---|---|---|---|---|
| 500 | 0.10 | 0.126 | 1.226e−2 | 1.134e−2 | 4.864e−1 | 4.864e−1 | 1.130e+0 | 3.017e−1 | 1.294e−3 |
| 500 | 0.30 | 0.126 | 2.896e−2 | 2.232e−2 | 4.908e−1 | 4.908e−1 | 1.127e+0 | 3.017e−1 | 1.587e−3 |
| 500 | 0.45 | 0.126 | 2.728e−2 | 2.236e−2 | 4.917e−1 | 4.917e−1 | 1.119e+0 | 3.017e−1 | 1.490e−3 |
| 2000 | 0.10 | 0.079 | 6.508e−3 | 7.453e−3 | 4.942e−1 | 4.942e−1 | 1.187e+0 | 3.017e−1 | 3.669e−4 |
| 2000 | 0.30 | 0.079 | 1.165e−2 | 9.494e−3 | 4.945e−1 | 4.945e−1 | 1.187e+0 | 3.017e−1 | 3.831e−4 |
| 2000 | 0.45 | 0.079 | 1.342e−2 | 1.077e−2 | 4.946e−1 | 4.946e−1 | 1.187e+0 | 3.017e−1 | 3.996e−4 |

Degree-CL and the triplet composite likelihood of Ambroise & Matias (2012) never group either, and their worst risks are 0.77 to 1.17 times that of Algorithm 1 (Appendix G); neither carries a guarantee in this regime.

A strong-signal control at $\Delta = 0.8$ checks the implementations: the clustering routes, given the true $K$, come within a factor of 3.6 of the oracle's risk and degree-CL within 1.06 (Appendix G).

**The least-favourable pair.** At $\varepsilon = 0.6\,(n^3\Delta^4)^{-1/4}$, the scaling of Theorem 4, the two endpoints differ in entropy by 0.1590 at $n = 500$ and 0.1496 at $n = 2000$, while the estimator's mean outputs differ by less than one standard error at both. A battery of up to 17 tests, trained and evaluated on separate samples, has largest observed power 0.223 over all amplitudes here against 1.000 for a positive control outside the window (Appendix G).

## 7 Calibration and the query budget

Algorithm 1 takes $(q, \Delta)$ as input, measured in practice on held-out pairs with known meanings, so the estimator runs with $(\hat{q}, \hat{\Delta})$. Knowing $q$ is indispensable: with $q$ unknown, no function of the graph is uniformly consistent over $(q, \theta)$ when $n\Delta^2 \leq c_1$ (Proposition F.5).

**Corollary 10** (proved in Appendix F). *Let $(\hat{q}, \hat{\Delta})$ be independent of $A$ and condition on their values. If $n^3\Delta^4 \geq 4$, $n\Delta^2 \leq c_1$ and $|\hat{q} - q| + |\hat{\Delta} - \Delta| \leq 1/(n^{3/2}\Delta)$, then running Algorithm 1 with $(\hat{q}, \hat{\Delta})$ gives $\mathbb{E}_\theta(\widehat{H}^{\mathrm{cal}} - H(\theta))^2 \leq C'(n^3\Delta^4)^{-1/2}$ with $C' \leq 2.2 \times 10^3$, the same order as Theorem 2.*

The underlying lemma is exact rather than perturbative: a misspecified channel distorts the moment pair affinely, with coefficients in Appendix F, contributing a bias of at most $19\eta$ with $\eta = (|\hat{q} - q| + |\hat{\Delta} - \Delta|)/\Delta$, and inflating the moment variances of Lemma B.1 at most 132-fold. Corollary 10 requires $(\hat{q}, \hat{\Delta})$ to come from a calibration set independent of $A$. Order $n^3\Delta^2$ labelled pairs of each kind calibrate the channel at no cost in rate (Corollary F.4).

Judging all $\binom{n}{2}$ pairs is usually the dominant cost, and the method extends to a restricted query set. Querying only the edges of a fixed simple $d$-regular graph chosen without looking at the data, with $2 \leq d \leq n-1$ and $m = nd/2$ queries, gives risk at most $\min\{(\log 3)^2,\ 47(1/n + \rho/m + n\rho^2/m^2)^{1/2}\}$, needing neither connectivity nor expansion (Appendix F.6).

## 8 Conclusion

Semantic entropy is computed in practice by grouping sampled answers with a judge and plugging in the group proportions. We have shown that when the judge separates meanings only weakly the grouping step is unnecessary, that below the recovery threshold the partition it aims at is not recoverable, and that the entropy itself still has a determined minimax rate. Grouping is unnecessary because, with at most three meanings, the entropy depends only on the collision and triple-collision probabilities, and averages over all pairs and triples of answers estimate both. The exponent comes from geometry rather than from variance: near a point mass the entropy is only Hölder-$\frac{1}{2}$ in these probabilities, so the minimax squared risk is of order $\min\{1, (n^3\Delta^4)^{-1/2}\}$. The estimator reads the graph once and solves a cubic, keeps its rate with a channel calibrated from labelled pairs, and extends to a $d$-regular design of $nd/2$ judge calls.

**Limitations.** Conditional dependence can bias $\widehat{P}_3$ (Proposition F.6), and a robust estimator at the same rate is open. Class-dependent rates, $p_c$ within meaning $c$ and $q_{cd}$ between meanings $c \neq d$, move the agreement density off $q + \Delta P_2$, which Corollary 10, correcting two scalars, does not reach. We measure neither on a real judge: the numerics use synthetic judges, no language model appears here, and the lower bounds are over functions of the verdict graph, not of the answers. The lower bound's constant is not evaluated.

Two directions follow: a $(K-1)$-moment estimator with a matched least-favourable family, which would give the rate for general $K$ and say whether $3/4$ begins a sequence $\gamma_c(K)$, and a minimax treatment of the calibration set as part of the experiment.

### Reproducibility statement

The model is (1) with Assumption 1 and the estimator is Algorithm 1; every statement in the main body is proved in the appendix with its hypotheses restated. All numerics use synthetic data on CPU, and Appendix G describes the numerical procedures, the seeding of the simulated graphs, every configuration, distribution and repetition count, the fitting and held-out sample sizes of the detection battery, standard errors for the tabulated risks, the strong-signal control and the moving pair, and confidence intervals for the selected detection tests. Code for every experiment is in the supplementary material.

### 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.

Debapratim Banerjee and Zongming Ma. Optimal hypothesis testing for stochastic block models with growing degrees. *arXiv preprint arXiv:1705.05305*, 2017.

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*, volume 306 of *Proceedings of Machine Learning Research*. PMLR, 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.

Ove Frank and Frank Harary. Cluster inference by using transitivity indices in empirical graphs. *Journal of the American Statistical Association*, 77(380):835–840, 1982. doi: 10.1080/01621459.1982.10477895.

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.

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*, pp. 5788–5799, 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. doi: 10.52202/079017-0283.

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 FURTHER RELATED WORK

| symbol | meaning |
|---|---|
| $n$ | number of sampled answers |
| $K$ | number of meanings, at most three here |
| $Z_i \in \{1,2,3\}$ | the meaning of answer $i$, never observed |
| $\theta \in \Delta_3$ | law of a single answer's meaning; zero coordinates allowed |
| $q,\ p = q + \Delta$ | judge agreement probability between and within meanings |
| $\Delta = p - q$ | separation of the judge; known |
| $A_{ij}$ | the judge's verdict on the pair $\{i, j\}$ |
| $d_i,\ \lvert E\rvert$ | degree of $i$ and total number of agreements |
| $N_2 = \binom{n}{2},\ N_3 = n(n-1)(n-2)$ | unordered pairs, ordered triples |
| $P_k = \sum_c \theta_c^k$ | power sums; $P_2$ collision, $P_3$ triple collision |
| $e_2,\ e_3$ | elementary symmetric polynomials of $\theta$ |
| $H(\theta) = -\sum_c \theta_c \log \theta_c$ | the target |
| $\widehat{P}_2, \widehat{P}_3$ | the two moment estimators of (4) |
| $M_3$ | the set of moment pairs attainable on $\Delta_3$ |
| $\widetilde{\theta},\ \widehat{H} = H(\widetilde{\theta})$ | recovered distribution and final estimate |
| $\rho = \max\{p(1-p), q(1-q)\}/\Delta^2$ | noise-to-signal scale |
| $V_n = 1/n + \rho/n^2 + \rho^2/n^3$ | the variance proxy of (3) |
| $\delta = n\Delta^2$ | weak-separation parameter |
| $R_n^*(q, \Delta)$ | minimax squared risk, (2) |
| $q_0, v_0 = \frac{q_0}{2}(1 - \frac{q_0}{2}), c_1 = \min\{q_0^2/4, v_0/20\}$ | constants of the regime |
| $C_0,\ K_0$ | constants of Theorem 4, finite, depending only on $q_0$ |
| $n_0$ | threshold on $n$ in Proposition 7, fixed in Appendix D |
| $L(\widehat{Z}, Z)$ | permutation-invariant Hamming loss on the partition |
| $G(n, r)$ | Erdős–Rényi graph with edge probability $r$ |

The appendix is organized as follows. Appendix B proves the moment identities, the Hölder modulus and the upper bound of Theorem 2. Appendix C proves the fallback lower bound of Theorem 5 and the consistency threshold of Proposition 6. Appendix D proves the matching lower bound of Theorem 4 and assembles Corollary 3. Appendix E proves the recovery bounds of Propositions 7 and 8 and the separation of Corollary 9. Appendix F proves the calibration guarantee of Corollary 10, counts the labelled pairs that reach it, shows that the channel cannot be dropped, and bounds the effect of verdicts that share an answer; Appendix F.6 records the query-budget design. Appendix G gives the numerical details.

**Further related work.** Without a judge, the power sum $P_\alpha$ is estimated from the $\alpha$-wise coincidences among the samples, and for integer $\alpha > 1$ this attains the optimal sample complexity of the Rényi entropy of order $\alpha$ (Acharya et al., 2017); $\widehat{P}_2$ and $\widehat{P}_3$ estimate the same collision probabilities for $\alpha = 2, 3$ when every coincidence is seen only through the channel. With the labels observed, Han et al. (2018) give a unified construction, local moment matching, that attains the optimal sample complexity for the Shannon entropy, for the power sums of order $\alpha \in (0, 1)$, and for the support size under a known lower bound on the non-zero probabilities. For a given number $K$ of blocks, Gunderson et al. (2023) map the densities of stars with up to $2K - 1$ edges and of bistars exactly to the parameters of a block model whose blocks have distinct expected degrees; here two blocks share a degree whenever their proportions coincide, and with the channel known, stars with at most two edges suffice for $K \leq 3$. Banerjee & Ma (2017) test an Erdős–Rényi graph against a symmetric two-block model as the degrees grow, approximating the signed-cycle likelihood-ratio test by spectral statistics that attain the optimal asymptotic power in the contiguous regime once the average degree grows faster than $n^{1/3}$; our lower bounds compare two planted models with the same edge density instead (Section 4).

## B PROOFS FOR THE UPPER BOUND

Throughout this appendix the graph $A$ follows the model (1). We write $W_{ij} := \mathbf{1}\{Z_i = Z_j\}$ for the co-membership indicator, so that $A_{ij} \mid Z \sim \mathrm{Bern}(q+\Delta W_{ij})$; we write $v_{ij} := (q+\Delta W_{ij})(1-q-\Delta W_{ij})$

for the conditional variance of $A_{ij}$ given $Z$, and we abbreviate

$$\bar{\rho} \;:=\; \frac{P_2\,p(1-p)+(1-P_2)\,q(1-q)}{\Delta^2} \;\le\; \rho. \tag{7}$$

### B.1 Moment identities and variances

**Lemma B.1.** *Let $n \ge 3$ and $0 \le q < q+\Delta \le 1$, and let $A$ follow (1) for a $\theta \in \Delta_3$: the labels $Z_1,\dots,Z_n$ are i.i.d. from $\theta$, given the labels the verdicts are conditionally independent, and the pair $(q,\Delta)$ is known. Then:*

1. $\mathbb{E}\widehat{P}_2 = P_2$ *and* $\mathbb{E}\widehat{P}_3 = P_3$.
2. $\mathrm{Var}(\widehat{P}_2) = \dfrac{2\,(P_2(1-P_2)+\bar{\rho})}{n(n-1)} + \dfrac{4\,(P_3-P_2^2)(n-2)}{n(n-1)}$.
3. $\mathrm{Var}(\widehat{P}_3) \;\le\; \dfrac{9}{4n} + \dfrac{8\rho}{n(n-1)} + \dfrac{2\rho^2}{n(n-1)(n-2)} \;\le\; 12\,V_n$.
4. $\mathrm{Var}(\widehat{P}_2) \;\le\; \dfrac{2n-3}{2n(n-1)} + \dfrac{2\rho}{n(n-1)} \;\le\; 3V_n$, *and consequently* $\mathrm{Var}(\widehat{P}_2)+\mathrm{Var}(\widehat{P}_3) \le v_n \le 15\,V_n$, *where*

$$v_n \;:=\; \frac{2n-3}{2n(n-1)} + \frac{9}{4n} + \frac{10\rho}{n(n-1)} + \frac{2\rho^2}{n(n-1)(n-2)}. \tag{8}$$

Part 4 records the explicit form of the variance sum that Theorem B.4 uses.

*Proof.* Put $X_{ij} := (A_{ij}-q)/\Delta$ and $\xi_{ij} := X_{ij} - W_{ij}$ for $i \ne j$, both symmetric in $i$ and $j$. Given $Z$, the variables $\{\xi_e\}$ are independent with $\mathbb{E}[\xi_e \mid Z] = 0$ and $\mathrm{Var}(\xi_e \mid Z) = v_e/\Delta^2 \le \rho$.

*Step 1 (representation of $\widehat{P}_3$ over triples).* Fix $i$. From $d_i - q(n-1) = \sum_{j\ne i}(A_{ij}-q)$,

$$\big(d_i - q(n-1)\big)^2 = \sum_{j\ne i}(A_{ij}-q)^2 + \sum_{\substack{j\ne k\\ j,k\ne i}} (A_{ij}-q)(A_{ik}-q).$$

Since $A_{ij} \in \{0,1\}$ we have $(A_{ij}-q)^2 = A_{ij}(1-q)^2 + (1-A_{ij})q^2$, and summing over $j \ne i$ gives $d_i(1-q)^2 + (n-1-d_i)q^2$. Substituting back into the definition (4),

$$\widehat{P}_3 = \frac{1}{N_3}\sum_{(i;j,k)} X_{ij}X_{ik}, \tag{9}$$

the sum running over all $N_3$ ordered triples. In the same way $\widehat{P}_2 = N_2^{-1}\sum_{i<j} X_{ij}$.

*Step 2 (unbiasedness).* We have $\mathbb{E}[X_{ij} \mid Z] = W_{ij}$, so $\mathbb{E}\widehat{P}_2 = \mathbb{E}W_{12} = \mathbb{P}(Z_1 = Z_2) = P_2$. For $j \ne k$ the edges $\{i,j\}$ and $\{i,k\}$ are distinct and therefore conditionally independent given $Z$, so $\mathbb{E}[X_{ij}X_{ik} \mid Z] = W_{ij}W_{ik} = \mathbf{1}\{Z_i = Z_j = Z_k\}$, whose expectation is $\sum_c \theta_c^3 = P_3$.

*Step 3 (the exact variance of $\widehat{P}_2$).* By the law of total variance, $\mathrm{Var}\widehat{P}_2 = \mathrm{Var}(\mathbb{E}[\widehat{P}_2 \mid Z]) + \mathbb{E}\,\mathrm{Var}(\widehat{P}_2 \mid Z)$.

- Conditional variance. Given $Z$ the $X_e$ are independent, so $\mathrm{Var}(\widehat{P}_2 \mid Z) = N_2^{-2}\sum_e v_e/\Delta^2$. Moreover $\mathbb{E}v_{12} = P_2\,p(1-p) + (1-P_2)\,q(1-q) = \Delta^2\bar{\rho}$ by (7), so $\mathbb{E}\,\mathrm{Var}(\widehat{P}_2 \mid Z) = \bar{\rho}/N_2 = 2\bar{\rho}/(n(n-1))$.
- Conditional mean. Here $\mathbb{E}[\widehat{P}_2 \mid Z] = N_2^{-1}\sum_e W_e$. The ordered pairs of edges $(e,f)$ fall into three cases. There are $N_2$ pairs with $e = f$, of covariance $\mathrm{Var}(W_{12}) = P_2(1-P_2)$. There are $N_2 \cdot 2(n-2)$ pairs sharing exactly one endpoint, of covariance $\mathbb{E}[W_{12}W_{13}] - P_2^2 =$

$P_3 - P_2^2$. Disjoint pairs of edges depend on disjoint sets of labels, hence are independent and have covariance zero. Therefore

$$\mathrm{Var}\big(\mathbb{E}[\widehat{P}_2 \mid Z]\big) = \frac{P_2(1-P_2) + 2(n-2)(P_3 - P_2^2)}{N_2}.$$

Adding the two parts gives part 2.

*Step 4 (decomposition of $\widehat{P}_3$ and orthogonality).* Substituting $X_{ij}X_{ik} = W_{ij}W_{ik} + W_{ij}\xi_{ik} + \xi_{ij}W_{ik} + \xi_{ij}\xi_{ik}$ into (9) gives $\widehat{P}_3 = L + M + Q$, where

$$L := \frac{1}{N_3}\sum_{(i;j,k)} W_{ij}W_{ik}, \qquad M := \frac{1}{N_3}\sum_{(i;j,k)} \big(W_{ij}\xi_{ik} + \xi_{ij}W_{ik}\big),$$

$$Q := \frac{1}{N_3}\sum_{(i;j,k)} \xi_{ij}\xi_{ik}.$$

The term $L$ is a function of $Z$. Next, $\mathbb{E}[M \mid Z] = 0$, and every term of $Q$ is a product of conditionally independent mean-zero variables carried by two distinct edges, so $\mathbb{E}[Q \mid Z] = 0$. Hence $\mathrm{Cov}(L, M+Q) = 0$. Consider now $\mathbb{E}[MQ \mid Z]$: it is a linear combination of terms $\mathbb{E}[\xi_e\xi_f\xi_g \mid Z]$ with $f \neq g$. When $e \notin \{f, g\}$ this is the expectation of a product of three independent mean-zero variables and vanishes; when $e = f$ it equals $\mathbb{E}[\xi_f^2 \mid Z]\,\mathbb{E}[\xi_g \mid Z] = 0$, and the case $e = g$ is the same. So $\mathbb{E}[MQ \mid Z] = 0$ and

$$\mathrm{Var}\widehat{P}_3 = \mathrm{Var}L + \mathbb{E}M^2 + \mathbb{E}Q^2. \tag{10}$$

*Step 5 (the label term).* Each unordered triple $\{a, b, c\}$ occurs exactly 6 times in the sum over ordered triples (three choices of the centre, two orders of the two ends), and $W_{ij}W_{ik}$ is the indicator that the three points share a class, so

$$L = \binom{n}{3}^{-1} \sum_{\{a,b,c\}} \mathbf{1}\{Z_a = Z_b = Z_c\}.$$

If the labels other than $Z_i$ are held fixed and only $Z_i$ is changed, the only indicators that can change are the $\binom{n-1}{2}$ ones involving $i$, each valued in $\{0, 1\}$. So, as a function of $Z_i$, $L$ takes values in an interval of length at most $\binom{n-1}{2}/\binom{n}{3} = 3/n$, and Popoviciu's inequality gives $\mathrm{Var}_i L \le (3/n)^2/4$, where $\mathrm{Var}_i$ denotes the conditional variance in $Z_i$ with the remaining labels held fixed. The Efron–Stein inequality, applicable since $Z_1, \ldots, Z_n$ are independent and $L$ is bounded, gives

$$\mathrm{Var}L \;\le\; \sum_{i=1}^{n} \mathbb{E}\,\mathrm{Var}_i L \;\le\; n \cdot \frac{9}{4n^2} \;=\; \frac{9}{4n}. \tag{11}$$

*Step 6 (the single-noise term).* The set of ordered triples is invariant under the exchange $j \leftrightarrow k$, so $\sum_{(i;j,k)} \xi_{ij}W_{ik} = \sum_{(i;j,k)} W_{ij}\xi_{ik}$ and therefore $M = (2/N_3)\sum_{(i;j,k)} W_{ij}\xi_{ik}$. Collecting the terms by the unordered edge $e = \{i, k\}$, which occurs both with $i$ as the centre and $k$ as an end, and with $k$ as the centre and $i$ as an end,

$$M = \frac{2}{N_3}\sum_{e=\{i,k\}} c_e\,\xi_e, \qquad c_e := \sum_{j \notin \{i,k\}} \big(W_{ij} + W_{kj}\big) \in [0,\ 2(n-2)].$$

Given $Z$ the coefficient $c_e$ is a constant, and the $\xi_e$ are independent with mean zero and variance at most $\rho$, so

$$\mathbb{E}[M^2 \mid Z] = \frac{4}{N_3^2}\sum_e c_e^2\,\frac{v_e}{\Delta^2} \;\le\; \frac{4}{N_3^2}\cdot\frac{n(n-1)}{2}\cdot 4(n-2)^2\rho \;=\; \frac{8\rho}{n(n-1)}. \tag{12}$$

*Step 7 (the double-noise term).* The ordered pairs $(j, k)$ and $(k, j)$ give the same product, so $Q = (2/N_3)\sum_w \xi_{ij}\xi_{ik}$, the sum running over the wedges $w = (i; \{j, k\})$, of which there are $n\binom{n-1}{2} = N_3/2$. Two distinct edges share at most one endpoint, so the pair of edges $\{\{i, j\}, \{i, k\}\}$

determines the centre $i$ and the two ends $\{j,k\}$, and distinct wedges give distinct pairs of edges. Take the pairs of edges $\{e,f\} \neq \{g,h\}$ (with $e \neq f$ and $g \neq h$) of two distinct wedges: the two two-element sets are unequal, so $\{e,f\} \not\subseteq \{g,h\}$, and we may assume $e \notin \{g,h\}$; then $e$ occurs exactly once in the product $\xi_e\xi_f\xi_g\xi_h$ and the conditional expectation vanishes. Therefore

$$\mathbb{E}[Q^2 \mid Z] = \frac{4}{N_3^2}\sum_w \frac{v_{ij}\,v_{ik}}{\Delta^4} \;\le\; \frac{4}{N_3^2}\cdot\frac{N_3}{2}\cdot\rho^2 \;=\; \frac{2\rho^2}{n(n-1)(n-2)}. \tag{13}$$

*Step 8 (assembly and simplification).* Steps 4 to 7, that is (10) together with (11), (12) and (13), give the first inequality of part 3. For $n \geq 3$ we have $n-1 \geq 2n/3$ and $n-2 \geq n/3$, hence $n(n-1) \geq 2n^2/3$ and $n(n-1)(n-2) \geq 2n^3/9$, so that $8\rho/(n(n-1)) \leq 12\rho/n^2$ and $2\rho^2/(n(n-1)(n-2)) \leq 9\rho^2/n^3$; together with $9/(4n) \leq 12/n$ this yields $\mathrm{Var}\widehat{P}_3 \leq 12V_n$.

For part 4, first $P_2(1-P_2) \leq 1/4$. Next, $\mathbb{E}\,\theta_{Z_1} = P_2$ and $\mathbb{E}\,\theta_{Z_1}^2 = P_3$ give $P_3 - P_2^2 = \mathrm{Var}(\theta_{Z_1})$, and since $\theta_{Z_1} \in [0,1]$ Popoviciu's inequality gives $P_3 - P_2^2 \leq 1/4$; finally $\bar{\rho} \leq \rho$. Substituting these into part 2,

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

Since $2n-3 \leq 2(n-1)$, the first term is at most $1/n$, and $2\rho/(n(n-1)) \leq 3\rho/n^2$, so $\mathrm{Var}\widehat{P}_2 \leq 3V_n$. Adding the first inequality of part 3 gives $v_n$. Finally $(2n-3)/(2n(n-1)) + 9/(4n) \leq 13/(4n)$, $10\rho/(n(n-1)) \leq 15\rho/n^2$ and $2\rho^2/(n(n-1)(n-2)) \leq 9\rho^2/n^3$, so $v_n \leq 15V_n$. □

### B.2 Proof of Lemma 1

**Lemma B.2** (restatement of Lemma 1)**.** *Let $\theta, \theta' \in \Delta_3$, zero coordinates being allowed, and put $D := |e_2(\theta) - e_2(\theta')| + |e_3(\theta) - e_3(\theta')|$. Then $D \leq 10/27$, and if $D > 0$,*

$$\big|H(\theta) - H(\theta')\big| \;\le\; D\log 2 + \log(1+D) + 2\sqrt{D}\,\arctan\frac{1}{\sqrt{D}} \;\le\; \pi\sqrt{D}; \tag{14}$$

*if $D = 0$ then $H(\theta) = H(\theta')$. The exponent $1/2$ cannot be improved, by Remark B.3.*

*Proof.* For $u \geq 0$ write $p_\theta(u) := \prod_{c=1}^3 (u+\theta_c) = u^3 + u^2 + e_2(\theta)\,u + e_3(\theta)$, where $\sum_c \theta_c = 1$ has been used, and put $\delta_2 := e_2(\theta) - e_2(\theta')$, $\delta_3 := e_3(\theta) - e_3(\theta')$, so that $D = |\delta_2| + |\delta_3|$.

*Step 0 (the range of $D$).* By Cauchy–Schwarz, $P_2 \geq 1/3$, so $e_2 = (1-P_2)/2 \in [0, 1/3]$; by AM–GM, $e_3 = \theta_1\theta_2\theta_3 \in [0, 1/27]$. Hence $|\delta_2| \leq 1/3$, $|\delta_3| \leq 1/27$ and $D \leq 10/27$.

*Step 1 (an integral identity).* For $t \in [0,1]$ and $M > 0$,

$$\int_0^M \log(u+t)\,\mathrm{d}u = (M+t)\log(M+t) - M - t\log t,$$

where for $t = 0$ the left-hand side is a convergent improper integral and the right-hand side is read with $0\log 0 = 0$. Summing over $c = 1,2,3$ and using $\sum_c \theta_c = 1$,

$$\int_0^M \log p_\theta(u)\,\mathrm{d}u = \sum_c \phi_M(\theta_c) - 3M + H(\theta), \qquad \phi_M(t) := (M+t)\log(M+t).$$

Writing the same identity for $\theta'$ and subtracting,

$$\int_0^M \log\frac{p_\theta(u)}{p_{\theta'}(u)}\,\mathrm{d}u = H(\theta) - H(\theta') + \sum_c \big[\phi_M(\theta_c) - \phi_M(\theta_c')\big].$$

Here $\phi_M'(t) = \log(M+t) + 1$, $\phi_M''(t) = 1/(M+t)$ and $\phi_M'''(t) = -1/(M+t)^2$, so for $t \in [0,1]$ Taylor's theorem with Lagrange remainder gives $\phi_M(t) = M\log M + t(\log M + 1) + t^2/(2M) + R_M(t)$ with $|R_M(t)| \leq t^3/(6M^2)$. Since $\sum_c \theta_c = \sum_c \theta_c' = 1$, the bracketed sum equals $(P_2(\theta) - P_2(\theta'))/(2M) + O(M^{-2})$, which tends to zero as $M \to \infty$. Steps 3 and 4 show that the integrand is absolutely integrable on $(0,\infty)$, so the limit of the left-hand side is the integral over $(0,\infty)$, and

$$H(\theta) - H(\theta') = \int_0^\infty \log\frac{p_\theta(u)}{p_{\theta'}(u)}\,\mathrm{d}u. \tag{15}$$

*Step 2 (a pointwise bound).* For $x, y > 0$ one has $|\log(x/y)| \le \log\big(1 + |x-y|/\min(x,y)\big)$: if $x \ge y$ the left-hand side equals $\log(1 + (x-y)/y)$, and the case $x < y$ is symmetric. For $u > 0$ we have $p_\theta(u), p_{\theta'}(u) \ge u^3 + u^2 > 0$, because $e_2, e_3 \ge 0$, and $p_\theta(u) - p_{\theta'}(u) = \delta_2 u + \delta_3$, so

$$\Big|\log \frac{p_\theta(u)}{p_{\theta'}(u)}\Big| \;\le\; \log\Big(1 + \frac{|\delta_2|\,u + |\delta_3|}{u^3 + u^2}\Big). \tag{16}$$

*Step 3 (the range $u \ge 1$).* There $|\delta_2|u + |\delta_3| \le Du$. Using $\log(1+x) \le x$, the right-hand side of (16) is at most $Du/(u^3+u^2) = D/(u(u+1))$, and $\int_1^\infty \mathrm{d}u/(u(u+1)) = \log 2$. The contribution of this range is therefore at most $D \log 2$.

*Step 4 (the range $0 < u \le 1$).* There $|\delta_2|u + |\delta_3| \le D$ and $u^3 + u^2 \ge u^2$, so the right-hand side of (16) is at most $\log(1 + D/u^2)$. Let $D > 0$. From

$$\frac{\mathrm{d}}{\mathrm{d}u}\Big[u \log\Big(1 + \frac{D}{u^2}\Big) + 2\sqrt{D}\,\arctan\frac{u}{\sqrt{D}}\Big] = \log\Big(1 + \frac{D}{u^2}\Big) - \frac{2D}{u^2 + D} + \frac{2D}{u^2 + D} = \log\Big(1 + \frac{D}{u^2}\Big)$$

and $\lim_{u \to 0^+} u \log(1 + D/u^2) = 0$ we obtain

$$\int_0^1 \log\Big(1 + \frac{D}{u^2}\Big)\,\mathrm{d}u = \log(1+D) + 2\sqrt{D}\,\arctan\frac{1}{\sqrt{D}}. \tag{17}$$

Steps 3 and 4 also supply the absolute integrability used in Step 1.

*Step 5 (assembly).* If $D = 0$ then $p_\theta = p_{\theta'}$, the two multisets of coordinates coincide, and $H(\theta) = H(\theta')$. Let $D > 0$. Steps 1 to 4 give

$$\big|H(\theta) - H(\theta')\big| \;\le\; D\log 2 + \log(1+D) + 2\sqrt{D}\,\arctan\frac{1}{\sqrt{D}}.$$

Since $\arctan(1/x) = \pi/2 - \arctan x$ for $x > 0$, the last term equals $\pi\sqrt{D} - 2\sqrt{D}\arctan\sqrt{D}$. Using $\log(1+D) \le D$ and $\arctan x \ge x - x^3/3$ for $x \ge 0$, which holds because the two sides agree at 0 and the derivative of their difference is $x^4/(1+x^2) \ge 0$,

$$D\log 2 + \log(1+D) - 2\sqrt{D}\arctan\sqrt{D} \;\le\; D\log 2 + D - 2D + \tfrac{2}{3}D^2 \;=\; -D\big(1 - \log 2 - \tfrac{2}{3}D\big) \;\le\; 0,$$

the last step by Step 0: $\frac{2}{3}D \le 20/81 < 0.247 < 1 - \log 2$. Hence $|H(\theta) - H(\theta')| \le \pi\sqrt{D}$, which is (14). □

*Remark* B.3 (the exponent $1/2$ cannot be improved). For $\varepsilon \in (0, 1/2]$ let $\theta_\varepsilon := (1-\varepsilon, \varepsilon, 0)$ and $\theta'_\varepsilon := (1-\varepsilon', \varepsilon'/2, \varepsilon'/2)$, where $\varepsilon'$ is the smaller root of $f(x) := \frac{3}{2}x^2 - 2x + 2\varepsilon(1-\varepsilon)$. Since $f(x) = P_2(\theta'_x) - P_2(\theta_\varepsilon)$, the two distributions have the same $P_2$.

- *The location of $\varepsilon'$.* The function $f$ is strictly decreasing on $(-\infty, 2/3]$, with $f(\varepsilon) = -\varepsilon^2/2 < 0$ and $f(\varepsilon - \varepsilon^2) = \frac{3}{2}\varepsilon^2(1-\varepsilon)^2 > 0$, so $\varepsilon - \varepsilon^2 < \varepsilon' < \varepsilon$; in particular $\varepsilon' \ge \varepsilon/2$ and $\varepsilon - \varepsilon' \le \varepsilon^2$.
- *The order of $D$.* The two values of $e_2$ are equal and $e_3(\theta_\varepsilon) = 0$, so $D = e_3(\theta'_\varepsilon) = (1-\varepsilon')\varepsilon'^2/4$ and $\sqrt{D} = \frac{\varepsilon}{2}(1 + O(\varepsilon))$.
- *The order of the entropy gap.* Write $h(x) := -x\log x - (1-x)\log(1-x)$, so that $H(\theta'_\varepsilon) - H(\theta_\varepsilon) = \varepsilon'\log 2 + h(\varepsilon') - h(\varepsilon)$. For $x \in (0, 1/2]$ we have $|h'(x)| = \log((1-x)/x) \le \log(1/x)$, so by the mean value theorem $|h(\varepsilon) - h(\varepsilon')| \le (\varepsilon - \varepsilon')\log(1/\varepsilon') \le \varepsilon^2 \log(2/\varepsilon)$. Hence $H(\theta'_\varepsilon) - H(\theta_\varepsilon) = \varepsilon\log 2 + O(\varepsilon^2\log(1/\varepsilon))$.

Consequently $|H(\theta'_\varepsilon) - H(\theta_\varepsilon)|/\sqrt{D} \to 2\log 2 > 0$ as $\varepsilon \to 0$. If there were a $\gamma > 1/2$ and a constant $C$ with $|H(\theta) - H(\theta')| \le CD^\gamma$ for all $\theta, \theta'$, this ratio would be at most $CD^{\gamma - 1/2} \to 0$, a contradiction.

### B.3 Proof of Theorem 2

**Theorem B.4** (restatement of Theorem 2)**.** *Let $n \geq 3$ and $0 \leq q < q + \Delta \leq 1$, under the hypotheses of Lemma B.1. For every $\theta \in \Delta_3$,*

$$\mathbb{E}_\theta\big(\widehat{H} - H(\theta)\big)^2 \;\leq\; \min\Big\{(\log 3)^2,\; 2\sqrt{10/9}\,\pi^2\, v_n^{1/2}\Big\} \;\leq\; \min\Big\{(\log 3)^2,\; 81\, V_n^{1/2}\Big\}, \tag{18}$$

*with $v_n$ as in part 4 of Lemma B.1.*

The bound needs neither $q$ bounded away from 0 and 1 nor a window condition.

*Proof.* Write $P := (P_2(\theta), P_3(\theta))$, let $\|\cdot\|$ be the Euclidean norm on $\mathbb{R}^2$, and put $\tilde{D} := |e_2(\widetilde{\theta}) - e_2(\theta)| + |e_3(\widetilde{\theta}) - e_3(\theta)|$.

*Step 1 (the estimator is well defined).* The set $M_3$ is the image of the compact set $\Delta_3$ under the continuous map $\theta \mapsto (P_2(\theta), P_3(\theta))$, hence compact and non-empty, so every $\widehat{P} \in \mathbb{R}^2$ has a nearest point in $M_3$. The graph $A$ takes finitely many values, and $\tilde{P}$, selected by a fixed rule, is a function of $A$, so no question of measurability arises. If $\theta, \theta'' \in \Delta_3$ have the same $(P_2, P_3)$, then they have the same $e_1 = 1$, $e_2$ and $e_3$, the monic polynomials $\prod_c (x - \theta_c)$ and $\prod_c (x - \theta''_c)$ coincide, the multisets of coordinates coincide, and therefore $H(\theta) = H(\theta'')$. Hence $\widehat{H} = H(\widetilde{\theta})$ does not depend on the choice of $\widetilde{\theta}$, and $\widehat{H} \in [0, \log 3]$.

*Step 2 (the projection at most doubles the distance).* Since $P \in M_3$, $\|\tilde{P} - \widehat{P}\| = \operatorname{dist}(\widehat{P}, M_3) \leq \|P - \widehat{P}\|$, and therefore

$$\|\tilde{P} - P\| \;\leq\; \|\tilde{P} - \widehat{P}\| + \|\widehat{P} - P\| \;\leq\; 2\,\|\widehat{P} - P\|. \tag{19}$$

*Step 3 (the difference of the coefficients).* Write $\delta P_k := \tilde{P}_k - P_k$. From $e_2 = (1 - P_2)/2$ and $e_3 = (1 - 3P_2 + 2P_3)/6$,

$$\begin{aligned} \tilde{D} = \frac{|\delta P_2|}{2} + \frac{|3\,\delta P_2 - 2\,\delta P_3|}{6} \;&\leq\; |\delta P_2| + \frac{|\delta P_3|}{3} \\ &\leq\; \sqrt{1 + \tfrac{1}{9}}\;\|\tilde{P} - P\| \;\leq\; 2\sqrt{10/9}\;\|\widehat{P} - P\|, \end{aligned} \tag{20}$$

where the penultimate step is Cauchy–Schwarz and the last one is (19).

*Step 4 (assembly).* Since $\widetilde{\theta}, \theta \in \Delta_3$, Lemma B.2 and (20) give $(\widehat{H} - H(\theta))^2 \leq \pi^2 \tilde{D} \leq 2\sqrt{10/9}\,\pi^2\,\|\widehat{P} - P\|$. Taking expectations, and using Jensen's inequality together with Lemma B.1, by which $\widehat{P}$ is unbiased and hence $\mathbb{E}\|\widehat{P} - P\|^2 = \operatorname{Var}\widehat{P}_2 + \operatorname{Var}\widehat{P}_3 \leq v_n$,

$$\mathbb{E}\big(\widehat{H} - H(\theta)\big)^2 \;\leq\; 2\sqrt{10/9}\,\pi^2\,\big(\mathbb{E}\|\widehat{P} - P\|^2\big)^{1/2} \;\leq\; 2\sqrt{10/9}\,\pi^2\, v_n^{1/2}.$$

Moreover $\widehat{H}, H(\theta) \in [0, \log 3]$, so $(\widehat{H} - H(\theta))^2 \leq (\log 3)^2$ always, which gives the first inequality of (18). From $v_n \leq 15 V_n$ and $2\sqrt{10/9} \cdot \sqrt{15} \cdot \pi^2 = 80.585\ldots \leq 81$ the second inequality follows. □

**Corollary B.5** (the weakly separated scale)**.** *Let $b > 0$ and $0 < \Delta \leq b n^{-1/2}$. Then, for every $q$ and every $\theta \in \Delta_3$,*

$$\mathbb{E}_\theta\big(\widehat{H} - H(\theta)\big)^2 \;\leq\; 81\Big(b^4 + \frac{b^2}{4} + \frac{1}{16}\Big)^{1/2}\,\big(n^3\Delta^4\big)^{-1/2}.$$

*Proof.* From $x(1 - x) \leq 1/4$ we get $\rho \leq 1/(4\Delta^2)$, so

$$n^3\Delta^4\, V_n \;\leq\; n^2\Delta^4 + \frac{n\Delta^2}{4} + \frac{1}{16} \;\leq\; b^4 + \frac{b^2}{4} + \frac{1}{16},$$

and substituting this into the second inequality of (18) gives the claim. □

**Corollary B.6** (sufficiency in Proposition 6)**.** *Let $\Delta_n \in (0, 1]$ with $n^3\Delta_n^4 \to \infty$. Then*

$$\sup_q \sup_{\theta \in \Delta_3} \mathbb{E}_\theta\big(\widehat{H} - H(\theta)\big)^2 \;\longrightarrow\; 0,$$

*where $q$ ranges over the values with $0 \leq q < q + \Delta_n \leq 1$.*

*Proof.* Since $\Delta_n \le 1$ we have $n^3 \ge n^3\Delta_n^4 \to \infty$, so $1/n \to 0$. From $\rho \le 1/(4\Delta_n^2)$ we get $\rho/n^2 \le (4n^2\Delta_n^2)^{-1} = \frac{1}{4}\,(n \cdot n^3\Delta_n^4)^{-1/2} \to 0$ and $\rho^2/n^3 \le (16\,n^3\Delta_n^4)^{-1} \to 0$. These bounds do not depend on $q$ or on $\theta$, so $V_n \to 0$ uniformly in $(q, \theta)$, and the conclusion follows from Theorem B.4. □

*Remark* B.7 (computing the projection). If $\widehat{P} \in M_3$ then $\tilde{P} = \widehat{P}$. If $\widehat{P} \notin M_3$, the nearest point $m^*$ lies on $\partial M_3$: otherwise $m^*$ would be an interior point, and moving it a little in the direction $\widehat{P} - m^*$ would keep it inside $M_3$ and bring it closer to $\widehat{P}$. The set $\{e_2 > 0,\ e_3 > 0,\ \mathrm{disc} > 0\}$ is defined by strict polynomial inequalities, hence open, and by Remark B.8 it is contained in $M_3$, so all of its points are interior points of $M_3$. The only points of $M_3$ with $e_2 = 0$ are the images of the vertices, where $e_3 = 0$ as well. Hence $\partial M_3 \subseteq \{e_3 = 0\} \cup \{\mathrm{disc} = 0\}$, the first set being traced by the points $(1-t, t, 0)$ and the second by the points $(a, a, 1-2a)$ with $t, a \in [0, 1/2]$, up to permutation. It is therefore enough to search for the nearest point on the images of these two curves. If an implementation only guarantees $\tilde{P} \in M_3$ with $\|\tilde{P} - \widehat{P}\| \le \mathrm{dist}(\widehat{P}, M_3) + \eta$, then Step 2 of the proof above becomes $\|\tilde{P} - P\| \le 2\|\widehat{P} - P\| + \eta$, and the bound of Theorem B.4 acquires the additional term $\sqrt{10/9}\,\pi^2\,\eta$.

*Remark* B.8 (the semialgebraic description of $M_3$ and the computation of $\widetilde{\theta}$). We have $M_3 = \{(P_2, P_3) \in \mathbb{R}^2 : e_2 \ge 0,\ e_3 \ge 0,\ \mathrm{disc} \ge 0\}$, where $e_2 = (1 - P_2)/2$, $e_3 = (1 - 3P_2 + 2P_3)/6$ and $\mathrm{disc} = e_2^2 - 4e_2^3 - 4e_3 + 18e_2e_3 - 27e_3^2$ is the discriminant of $f(x) := x^3 - x^2 + e_2x - e_3$.

- If $(P_2, P_3)$ comes from a $\theta \in \Delta_3$, then $f(x) = \prod_c (x - \theta_c)$, so $e_2 = \sum_{c<d} \theta_c\theta_d \ge 0$, $e_3 = \prod_c \theta_c \ge 0$ and $\mathrm{disc} = \prod_{c<d}(\theta_c - \theta_d)^2 \ge 0$.
- Conversely, if a real cubic $f$ has a real root $r$ and a pair of non-real conjugate roots $\alpha \pm i\beta$ with $\beta \ne 0$, then $\mathrm{disc} = \left[(r-\alpha)^2 + \beta^2\right]^2 (2i\beta)^2 < 0$; so $\mathrm{disc} \ge 0$ forces all three roots to be real. For $x < 0$ we have $f(x) = x^3 - x^2 + e_2x - e_3 < 0$, so there is no negative root. The three roots are non-negative and sum to 1 by Vieta's formulas, hence they form some $\theta \in \Delta_3$ whose $e_2$ and $e_3$ are the given values; by Newton's identities $P_2 = 1 - 2e_2$ and $P_3 = 1 - 3e_2 + 3e_3$, so its moments are the given $(P_2, P_3)$.

The coordinates of $\widetilde{\theta}$ are therefore the three roots of $x^3 - x^2 + \tilde{e}_2x - \tilde{e}_3$.

### B.4 Why at most three meanings

*Remark* B.9 (why $K \le 3$). The bound $K \le 3$ is the exact boundary of the route on which two moments suffice.

- $K = 2$*: one moment suffices.* For $\theta = (1-t, t)$ one has $P_2 = 1 - 2t + 2t^2$, whose solutions $t = \frac{1}{2}(1 \pm \sqrt{2P_2 - 1})$ are mirror images of each other and carry the same entropy. The edge density alone therefore determines the entropy, and $P_3$ is not used.
- $K = 3$*: two moments suffice exactly, and the modulus of the entropy is singular.* By Step 1 of the proof of Theorem B.4, two distributions $\theta$ with the same $(P_2, P_3)$ have the same $e_1 = 1$, $e_2$ and $e_3$, hence the same monic polynomial $\prod_c (x - \theta_c)$, hence the same multiset of coordinates and the same entropy. So $H$ is a function of $(P_2, P_3)$, which is what makes the estimator of Section 3 meaningful. At the same time, Lemma 1 and Remark B.3 show that this function has only Hölder-$\frac{1}{2}$ regularity at the boundary and that the exponent cannot be improved, which is the source of the $(n^3\Delta^4)^{-1/2}$ rather than $(n^3\Delta^4)^{-1}$ in the rate.
- $K = 4$*: two moments no longer suffice.* Take

  $$\theta = (0.5,\ 0.25,\ 0.25,\ 0), \qquad \theta' = (0.45,\ 0.4,\ 0.1,\ 0.05).$$

  The two have exactly the same $P_2 = 0.375$ and the same $P_3 = 0.15625$, while $H(\theta) = 1.039721$ and $H(\theta') = 1.105890$ differ by $0.066169$. On four classes there is therefore no function mapping $(P_2, P_3)$ to the entropy, and the estimator of Section 3 has no verbatim four-class version.

This remark describes the boundary of the *route*, not the boundary of identifiability: with four classes $H$ is still determined by $(P_2, P_3, P_4)$, and $P_4$ is in principle estimable from a statistic counting

four-cycles. Carrying the upper and lower bounds of this paper to general $K$ requires an estimator that handles $K-1$ power sums at once together with a matching least-favourable family, not an adjustment of the constants in the present proofs. The statements therefore cover $K \le 3$, the first case in which the edge density alone no longer determines the entropy.

# C THE FALLBACK LOWER BOUND AND THE CONSISTENCY THRESHOLD

This appendix proves the fallback lower bound of Theorem 5 and the consistency threshold of Proposition 6. Both rest on two-point arguments in which the two candidate graph distributions are compared through a common Erdős–Rényi reference, and each comparison is carried out by a replica computation of the $\chi^2$ divergence.

Throughout, for distributions $P$ and $Q$ on a finite sample space we write $\mathrm{TV}(P,Q) := \frac{1}{2}\sum_A |P(A)-Q(A)|$, and $\chi^2(P\,\|\,Q) := \sum_A P(A)^2/Q(A) - 1$ whenever $Q$ is everywhere positive. As in the main text, $P_\theta$ is the law of the graph $A$ under (1) with parameter $\theta \in \Delta_3$, and $G(n,r)$ is the Erdős–Rényi graph whose $\binom{n}{2}$ edges are independent $\mathrm{Bern}(r)$ variables.

One auxiliary family is needed that the main text does not use.

**Definition C.1.** For $\tau, \varepsilon \in [0,1]$, the single planted-subset model $R_\tau(\varepsilon)$ is the following distribution on graphs. Let $Y_1, \dots, Y_n$ be i.i.d. $\mathrm{Bern}(\varepsilon)$ and put $S := \{i : Y_i = 1\}$. Given $Y$ the edges are conditionally independent, and

$$\mathbb{P}(A_{ij} = 1 \mid Y) \;=\; \begin{cases} q + \Delta\tau, & i,j \in S, \\ q, & \text{exactly one of } i,j \text{ lies in } S, \\ p, & i,j \notin S. \end{cases}$$

Taking $Y_i := \mathbf{1}\{Z_i = 2\}$ shows that $P_\theta = R_1(\varepsilon)$ when $\theta = (1-\varepsilon, \varepsilon, 0)$, so one endpoint of the hard pair of Remark B.3 is already a planted-subset model; the other endpoint is brought into the family by Lemma C.4.

## C.1 A REPLICA $\chi^2$ BOUND FOR THE LABEL MODEL

**Lemma C.2.** *Let $n \ge 2$, $0 \le q < q + \Delta \le 1$ and $\theta \in \Delta_3$, and write $s := P_2(\theta)$, $r := q + \Delta s$ and $m := \max\{s, 1-s\}$.*

1. *$r > 0$; and $r = 1$ if and only if $p = 1$ and $\theta$ is a vertex of $\Delta_3$, in which case $P_\theta = G(n,1)$, the point mass on the complete graph.*
2. *If $r < 1$, then*

$$\chi^2\big(P_\theta \,\big\|\, G(n,r)\big) \;\le\; \exp\Big\{\frac{n(n-1)^2 m^2 \Delta^4}{8\, r^2(1-r)^2}\Big\} - 1. \tag{21}$$

The statement needs nothing beyond the observation model (1) and the stated parameter range; in particular $q$ is not required to stay away from 0 and 1, since the factor $r^2(1-r)^2$ already carries the dependence on $q$.

*Proof.* Write $\lambda := \Delta/(r(1-r))$ and $\mu := \Delta^2/(r(1-r))$, let $Z'$ be an independent replica of $Z$, and put $W'_{ij} := \mathbf{1}\{Z'_i = Z'_j\}$, $X_{ij} := (W_{ij} - s)(W'_{ij} - s)$ and $S_{\mathrm{rep}} := \sum_{i<j} X_{ij}$. We write $\mathrm{Bern}(a; x)$ for the probability that a $\mathrm{Bern}(x)$ variable takes the value $a$.

*Step 1 (the range of $r$).* Since $P_2 \ge 1/3$ and $\Delta > 0$, we have $r \ge q + \Delta/3 > 0$. From $s \le 1$ and $q + \Delta \le 1$ we get $r \le 1$, and $r = 1$ forces $s = 1$ and $q + \Delta = 1$. Now $s = P_2 = 1$ means that $\theta$ is a vertex, so $W_{ij} \equiv 1$ and every edge is present with probability $p = 1$; then $P_\theta$ is the point mass on the complete graph, which is $G(n,1)$. From here on $0 < r < 1$.

*Step 2 (edgewise likelihood ratio).* For $w, a \in \{0,1\}$, the identity $q + \Delta w = r + \Delta(w - s)$ gives

$$\frac{\mathrm{Bern}(a;\, q + \Delta w)}{\mathrm{Bern}(a;\, r)} \;=\; 1 + \lambda\,(w - s)(a - r). \tag{22}$$

Indeed, for $a = 1$ the left-hand side is $1 + \Delta(w - s)/r$ and the right-hand side is $1 + \lambda(w - s)(1 - r)$, which agree; for $a = 0$ the left-hand side is $1 - \Delta(w - s)/(1 - r)$ and the right-hand side is $1 - \lambda(w - s)r$, which again agree. Given $Z$ the edges are conditionally independent, so

$$\frac{P_\theta(A)}{G(n,r)(A)} \;=\; \mathbb{E}_Z \prod_{i<j} \big(1 + \lambda(W_{ij} - s)(A_{ij} - r)\big).$$

*Step 3 (replica representation).* The sample space is finite, so sums and expectations may be interchanged:

$$\begin{aligned} 1 + \chi^2 \;&=\; \sum_A G(n,r)(A) \Big(\frac{P_\theta(A)}{G(n,r)(A)}\Big)^2 \\ &=\; \mathbb{E}_{Z,Z'} \sum_A G(n,r)(A) \prod_{i<j} \big(1 + \lambda(W_{ij} - s)(A_{ij} - r)\big)\big(1 + \lambda(W'_{ij} - s)(A_{ij} - r)\big). \end{aligned}$$

Under $G(n, r)$ the edges are independent, so the sum over $A$ factorizes over edges; on a single edge $\mathbb{E}[A_{ij} - r] = 0$ and $\mathbb{E}[(A_{ij} - r)^2] = r(1 - r)$, whence

$$1 + \chi^2 \;=\; \mathbb{E}_{Z,Z'} \prod_{i<j} \big(1 + \mu\, X_{ij}\big). \tag{23}$$

*Step 4 (exponential bound).* Each factor $1 + \mu X_{ij}$ is the expectation under $\mathrm{Bern}(r)$ of a product of two non-negative likelihood ratios, hence is non-negative; and $1 + x \le e^x$ for every real $x$. A product of non-negative numbers is at most the product of their upper bounds, so $1 + \chi^2 \le \mathbb{E} \exp(\mu S_{\mathrm{rep}})$.

*Step 5 (bounded differences).* The quantity $S_{\mathrm{rep}}$ is a function of the i.i.d. variables $\xi_i := (Z_i, Z'_i)$, $i = 1, \dots, n$. Since $Z$ and $Z'$ are independent and $\mathbb{E}W_{ij} = P_2 = s$, we have $\mathbb{E}X_{ij} = 0$ and $\mathbb{E}S_{\mathrm{rep}} = 0$. Changing $\xi_i$ alone can change only the $n - 1$ terms $X_{ij}$ with $j \neq i$. Each such term takes values in $\{s^2,\ -s(1 - s),\ (1 - s)^2\}$, whose largest element minus its smallest is $\max\{s^2 + s(1 - s),\ (1 - s)^2 + s(1 - s)\} = \max\{s, 1 - s\} = m$, so $S_{\mathrm{rep}}$ changes by at most $c_i := (n - 1)m$. The bounded-difference bound for the moment generating function of a function $f$ of independent variables, $\mathbb{E}e^{t(f - \mathbb{E}f)} \le \exp(t^2 \sum_i c_i^2/8)$, therefore gives

$$\mathbb{E} \exp(\mu S_{\mathrm{rep}}) \;\le\; \exp\Big(\frac{\mu^2}{8} \cdot n(n - 1)^2 m^2\Big) \;=\; \exp\Big\{\frac{n(n - 1)^2 m^2 \Delta^4}{8 r^2 (1 - r)^2}\Big\}.$$

Combining this with Step 4 gives (21). □

A fixed-pair lower bound already follows from Lemma C.2: two parameters with the same second moment but different entropies are both close to the same Erdős–Rényi reference as soon as $n^3\Delta^4$ is bounded by a constant. It supplies the necessity half of Proposition 6 at a fixed pair, and Proposition C.10 below extends it to the uniform statement.

**Proposition C.3.** *Let $n \ge 2$ and $0 \le q < q + \Delta \le 1$, and put $r := q + \Delta/2$, $c_0 := 32 \log(5/4) \approx 7.1406$, $\theta_a := (1/2, 1/2, 0)$, $\theta_b := (2/3, 1/6, 1/6)$ and $\omega := \log 3 - \frac{4}{3} \log 2 \approx 0.17442$. If $n(n - 1)^2\Delta^4 \le c_0\, r^2(1 - r)^2$, and in particular if $n^3\Delta^4 \le c_0\, r^2(1 - r)^2$, then every estimator $\widehat{H}$ satisfies*

$$\max_{\theta \in \{\theta_a, \theta_b\}} \mathbb{E}_\theta \big(\widehat{H} - H(\theta)\big)^2 \;\ge\; \frac{\omega^2}{16},$$

*and consequently $R_n^*(q, \Delta) \ge \omega^2/16 \approx 0.0019$.*

*Proof. Step 1 (moments and entropies of the two points).* We have $P_2(\theta_a) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}$ and $P_2(\theta_b) = \frac{4}{9} + \frac{2}{36} = \frac{1}{2}$, so both points have $s = 1/2$ and $m = 1/2$, and both give the same $r = q + \Delta/2$. From $\Delta > 0$ and $q + \Delta \le 1$ we get $0 < r < 1$. Moreover $H(\theta_a) = \log 2$ and $H(\theta_b) = \frac{2}{3} \log \frac{3}{2} + \frac{1}{3} \log 6 = \log 3 - \frac{1}{3} \log 2$, so $H(\theta_b) - H(\theta_a) = \omega > 0$.

*Step 2 (distance of each point to the reference).* By part 2 of Lemma C.2 and the hypothesis, for $\theta \in \{\theta_a, \theta_b\}$,

$$\chi^2\big(P_\theta \,\big\|\, G(n,r)\big) \;\le\; \exp\Big\{\frac{n(n - 1)^2\Delta^4}{32\, r^2(1 - r)^2}\Big\} - 1 \;\le\; e^{c_0/32} - 1 \;=\; \frac{1}{4}.$$

Since $n(n-1)^2 \le n^3$, the condition $n^3\Delta^4 \le c_0 r^2(1-r)^2$ implies the hypothesis.

*Step 3 (total variation between the two points).* By Cauchy–Schwarz, $\sum_A |P(A) - Q(A)| = \sum_A Q(A)\,|P(A)/Q(A) - 1| \le \sqrt{\chi^2(P \,\|\, Q)}$, that is,

$$\mathrm{TV}(P, Q) \;\le\; \tfrac{1}{2}\sqrt{\chi^2(P \,\|\, Q)}. \tag{24}$$

Hence $\mathrm{TV}(P_{\theta_a}, G(n,r)) \le 1/4$ and $\mathrm{TV}(P_{\theta_b}, G(n,r)) \le 1/4$, and the triangle inequality gives $\mathrm{TV}(P_{\theta_a}, P_{\theta_b}) \le 1/2$.

*Step 4 (external randomization).* If $\widehat{H} = \widehat{H}(A, U)$ with $U$ independent of $A$ and with a law that does not depend on $\theta$, then the total variation between the joint laws of $(A, U)$ at the two points equals $\mathrm{TV}(P_{\theta_a}, P_{\theta_b})$. The argument below is carried out for the joint laws and therefore covers randomized estimators as well.

*Step 5 (Le Cam's two-point method).* Let $E := \{|\widehat{H} - H(\theta_a)| < \omega/2\}$. On $E$ we have $|\widehat{H} - H(\theta_b)| \ge \omega - |\widehat{H} - H(\theta_a)| > \omega/2$. Therefore

$$\begin{aligned}\mathbb{E}_{\theta_a}\big(\widehat{H} - H(\theta_a)\big)^2 + \mathbb{E}_{\theta_b}\big(\widehat{H} - H(\theta_b)\big)^2 \;&\ge\; \frac{\omega^2}{4}\big(\mathbb{P}_{\theta_a}(E^c) + \mathbb{P}_{\theta_b}(E)\big) \\ &\ge\; \frac{\omega^2}{4}\big(1 - \mathrm{TV}(P_{\theta_a}, P_{\theta_b})\big) \;\ge\; \frac{\omega^2}{8},\end{aligned}$$

where the second inequality uses $\mathbb{P}_{\theta_a}(E^c) + \mathbb{P}_{\theta_b}(E) = 1 - (\mathbb{P}_{\theta_a}(E) - \mathbb{P}_{\theta_b}(E))$. The two terms sum to at least $\omega^2/8$, so the larger of them is at least $\omega^2/16$. The supremum defining $R_n^*$ runs over all of $\Delta_3$ and is therefore at least the maximum over the two points. □

### C.2 The split is invisible

The second endpoint of the hard pair splits the minority class in two. The next lemma says that at weak separation this split cannot be seen: the law of the graph under $\theta' = (1-\varepsilon', \varepsilon'/2, \varepsilon'/2)$ is close in total variation to the planted-subset model $R_{1/2}(\varepsilon')$, which leaves the minority block homogeneous.

**Lemma C.4.** *Let $n \ge 2$, $0 \le q < q + \Delta \le 1$ and $\varepsilon' \in [0,1]$, and put $\theta' := (1-\varepsilon', \varepsilon'/2, \varepsilon'/2)$, $r_S := q + \Delta/2$, $c_S := \Delta^2/(4r_S(1-r_S))$ and $\kappa := \sqrt{e^{1/8} - 1} \approx 0.3649$. If $n\,c_S \le 1/2$, then*

$$\mathrm{TV}\big(P_{\theta'},\; R_{1/2}(\varepsilon')\big) \;\le\; \kappa\, n\, c_S\, \varepsilon'. \tag{25}$$

*Proof.* Under $P_{\theta'}$ let $S := \{i : Z_i \ne 1\}$, and for $i \in S$ let $\sigma_i := +1$ if $Z_i = 2$ and $\sigma_i := -1$ if $Z_i = 3$. Write $\mathrm{SBM}_k$ for the distribution on graphs with $k$ vertices in which the signs $\sigma_1, \dots, \sigma_k$ are independent and uniform on $\{\pm 1\}$ and, given the signs, the edges are independent with $\mathbb{P}(A_{ij} = 1 \mid \sigma) = r_S + \frac{\Delta}{2}\sigma_i\sigma_j$, and write $\mathrm{ER}_k := G(k, r_S)$. Let $\sigma'$ be an independent replica of $\sigma$, and put $\tau_i := \sigma_i\sigma_i'$, $T := \sum_{i=1}^k \tau_i$ and $x := c_S k$.

*Step 1 (a common latent representation).* Under $P_{\theta'}$ the indicators $Y_i := \mathbf{1}\{i \in S\}$ are i.i.d. $\mathrm{Bern}(\varepsilon')$, and given $Y$ the signs $\{\sigma_i\}_{i \in S}$ are mutually independent, uniform and independent of $Y$. The edge probabilities are $p$ when neither endpoint lies in $S$, $q$ when exactly one endpoint lies in $S$, and $q + \Delta\mathbf{1}\{\sigma_i = \sigma_j\} = r_S + \frac{\Delta}{2}\sigma_i\sigma_j$ when both endpoints lie in $S$. Under $R_{1/2}(\varepsilon')$ the law of $Y$ is the same and the edges outside $S$ and across $S$ have the same probabilities, while the edges inside $S$ have probability $r_S$.

*Step 2 (reduction to a fixed $S$).* Given $S$, under either model the conditional law of $A$ is a product of the law of the subgraph inside $S$ and the law of the remaining edges; the latter factor is the same under the two models, while the subgraph inside $S$ follows $\mathrm{SBM}_{|S|}$ under $P_{\theta'}$ and $\mathrm{ER}_{|S|}$ under $R_{1/2}(\varepsilon')$. When two product measures share one factor, their total variation equals the total variation of the other factor, so the conditional total variation given $S$ is $\mathrm{TV}(\mathrm{SBM}_{|S|}, \mathrm{ER}_{|S|})$. The two marginals are mixtures with the same weights $\mathbb{P}(S)$, so by the triangle inequality

$$\mathrm{TV}\big(P_{\theta'}, R_{1/2}(\varepsilon')\big) \;\le\; \sum_S \mathbb{P}(S)\; \mathrm{TV}\big(\mathrm{SBM}_{|S|}, \mathrm{ER}_{|S|}\big). \tag{26}$$

When $|S| \le 1$ there are no edges inside $S$ and the corresponding term vanishes.

*Step 3 (replica representation for a fixed $k = |S| \geq 2$).* Here $0 < r_S < 1$, because $\Delta > 0$ and $q + \Delta \leq 1$. As in (22), for $u \in \{\pm 1\}$,

$$\frac{\mathrm{Bern}(a;\, r_S + \frac{\Delta}{2}u)}{\mathrm{Bern}(a;\, r_S)} \;=\; 1 + \frac{\Delta}{2}\, u\, \frac{a - r_S}{r_S(1 - r_S)},$$

and exactly as in (23),

$$1 + \chi^2\big(\mathrm{SBM}_k \,\big\|\, \mathrm{ER}_k\big) \;=\; \mathbb{E}_{\sigma,\sigma'} \prod_{i<j} \big(1 + c_S\, \tau_i \tau_j\big),$$

where $\tau_1, \ldots, \tau_k$ are independent and uniform on $\{\pm 1\}$, since the product of two independent uniform signs is again a uniform sign, independently over $i$.

*Step 4 (exponential bound).* Here $c_S \leq 1/(2n) < 1$, so every factor is positive, and $1 + y \leq e^y$ together with $\tau_i^2 = 1$ gives

$$\prod_{i<j} (1 + c_S \tau_i \tau_j) \;\leq\; \exp\Big(c_S \sum_{i<j} \tau_i \tau_j\Big) \;=\; \exp\Big(\frac{c_S}{2}\big(T^2 - k\big)\Big).$$

*Step 5 (a Gaussian integral).* Let $g \sim N(0,1)$ be independent of $\tau$. The Gaussian moment generating function gives $\exp(\frac{c_S}{2} T^2) = \mathbb{E}_g \exp(\sqrt{c_S}\, g\, T)$, and the integrand is non-negative, so the expectations may be interchanged:

$$\mathbb{E}_\tau \exp\Big(\frac{c_S}{2} T^2\Big) \;=\; \mathbb{E}_g\big[\cosh(\sqrt{c_S}\, g)^k\big] \;\leq\; \mathbb{E}_g \exp\Big(\frac{x\, g^2}{2}\Big) \;=\; (1 - x)^{-1/2},$$

using $\cosh y \leq e^{y^2/2}$ and $x = c_S k \leq n c_S \leq 1/2 < 1$. Together with Steps 3 and 4,

$$1 + \chi^2\big(\mathrm{SBM}_k \,\big\|\, \mathrm{ER}_k\big) \;\leq\; e^{-x/2}(1 - x)^{-1/2}.$$

*Step 6 (passing to total variation).* For $x \in [0, 1/2]$ the function $\phi(x) := -x - \log(1 - x) - x^2$ satisfies $\phi(0) = 0$ and $\phi'(x) = x(2x - 1)/(1 - x) \leq 0$, so $-\frac{x}{2} - \frac{1}{2}\log(1 - x) \leq \frac{x^2}{2}$, that is, $\log(1 + \chi^2) \leq x^2/2$. The map $y \mapsto e^y - 1$ is convex on $[0, 1/8]$ and therefore lies below the chord joining $(0, 0)$ to $(1/8, e^{1/8} - 1)$, while $x^2/2 \leq 1/8$, so

$$\chi^2 \;\leq\; e^{x^2/2} - 1 \;\leq\; 8\big(e^{1/8} - 1\big) \cdot \frac{x^2}{2} \;=\; 4\kappa^2 x^2, \qquad \mathrm{TV}\big(\mathrm{SBM}_k, \mathrm{ER}_k\big) \;\leq\; \tfrac{1}{2}\sqrt{\chi^2} \;\leq\; \kappa\, c_S\, k.$$

*Step 7 (averaging over $S$).* Every $S$ satisfies $c_S|S| \leq n c_S \leq 1/2$, so Step 6 applies to each term of (26), and

$$\mathrm{TV}\big(P_{\theta'}, R_{1/2}(\varepsilon')\big) \;\leq\; \kappa\, c_S\, \mathbb{E}|S| \;=\; \kappa\, n\, c_S\, \varepsilon'. \qquad \square$$

**Corollary C.5.** *Let $q_0 \in (0, 1/2]$, $q \in [q_0, 1 - q_0]$ and $\Delta \leq q_0$, and recall $v_0 = \frac{q_0}{2}(1 - \frac{q_0}{2})$. If $n\Delta^2 \leq 2v_0$, then $nc_S \leq 1/2$ and*

$$\mathrm{TV}\big(P_{\theta'},\; R_{1/2}(\varepsilon')\big) \;\leq\; \frac{\kappa}{4v_0}\, n\Delta^2\, \varepsilon';$$

*in particular the right-hand side is at most $1/8$ as soon as $n\Delta^2\varepsilon' \leq v_0/(2\kappa)$.*

*Proof.* From $\Delta \leq q_0$ and $q \in [q_0, 1 - q_0]$ we get $r_S \in [q_0,\, 1 - q_0/2]$. The map $r \mapsto r(1 - r)$ is concave, so its minimum over that interval is attained at an endpoint, and the two endpoint values are $q_0(1 - q_0)$ and $v_0$; since $q_0 \leq 1/2$ gives $v_0 \leq q_0(1 - q_0)$, we obtain $r_S(1 - r_S) \geq v_0$ and hence $nc_S \leq n\Delta^2/(4v_0) \leq 1/2$. Substituting into (25) gives the claim. $\square$

*Remark* C.6. At the scale of the hard pair, $\varepsilon^2 = a^2(n^3\Delta^4)^{-1/2}$ with a constant $a$, one has $\varepsilon' \leq \varepsilon$ by Remark B.3, so $n\Delta^2\varepsilon' \leq n\Delta^2\varepsilon = a\, n^{1/4}\Delta$. If $n\Delta^2 \leq c_1$, the right-hand side is at most $a\sqrt{c_1}\, n^{-1/4}$, so the contribution of this step tends to zero as $n$ grows; to reach the constant $1/8$ it is enough that $n\Delta^2\varepsilon' \leq v_0/(2\kappa)$. A single small constant can secure both hypotheses of Corollary C.5 at once: for $n \geq 2$, take $c_1 \leq \min\{2v_0,\, 2q_0^2\}$, so that $n\Delta^2 \leq c_1$ implies $n\Delta^2 \leq 2v_0$ and $\Delta^2 \leq c_1/2 \leq q_0^2$; taking in addition $c_1\varepsilon_0 \leq v_0/(2\kappa)$, where $\varepsilon_0$ is any cap on the amplitude, so that $\varepsilon \leq \varepsilon_0$, gives the constant $1/8$.

### C.3 A REPLICA $\chi^2$ BOUND FOR THE PLANTED-SUBSET MODEL

The bound of Lemma C.2 is not sharp enough for the fallback lower bound, because it does not profit from a small planted fraction. The next lemma repeats the replica computation for $R_\tau(\varepsilon)$ and keeps the centring, which makes the leading term carry a factor $v^2 \asymp \varepsilon^2$.

**Lemma C.7.** *Let $n \geq 2$, $0 \leq q < q + \Delta \leq 1$, $\tau \in [0,1]$ and $\varepsilon \in (0, 1/2]$, and write*

$$k := 1+\tau, \quad b := 1-2\varepsilon, \quad v := \varepsilon(1-\varepsilon), \quad \beta := 1-k\varepsilon, \quad d := \frac{1-\tau}{2}, \quad \ell_\tau := n\beta^2 - 2d^2,$$

$$s_\tau := 1-2\varepsilon+k\varepsilon^2, \quad r := q+\Delta s_\tau, \quad \sigma^2 := r(1-r), \quad \mu := \frac{\Delta^2}{\sigma^2}, \quad \Theta := \frac{n\Delta^2}{\sigma^2}, \quad \Gamma := \frac{n^{3/2}\Delta^2 v}{\sigma^2}.$$

*Then $0 < r < 1$. Let $\alpha > 1$ and let $\alpha' := \alpha/(\alpha-1)$ be its conjugate exponent, and assume $\frac{5\alpha\Theta}{12} < 1$ and $\frac{\alpha'\Theta}{4} < 1$. Then*

$$1 + \chi^2\big(R_\tau(\varepsilon) \,\big\|\, G(n,r)\big) \;\leq\; \exp\Big\{\frac{33\,\alpha\,\Gamma^2}{64\big(1-\frac{5\alpha\Theta}{12}\big)}\Big\} \cdot \Big(1 - \frac{\alpha'\Theta}{4}\Big)^{-\frac{1}{2\alpha'}}. \tag{27}$$

*The right-hand side depends on $\tau$ only through $\Theta$ and $\Gamma$, and is therefore uniform over any range of $\tau$, in particular over $[1/2, 1]$.*

*Proof.* Let $Y_1, \dots, Y_n$ be i.i.d. $\mathrm{Bern}(\varepsilon)$ as in Definition C.1, let $Y'$ be an independent replica, and put $u_i := Y_i - \varepsilon$ and $u'_i := Y'_i - \varepsilon$. Write $T := \sum_i u_i u'_i$, $A := \sum_i u_i$ and $A' := \sum_i u'_i$ for the centred label sums; the graph itself enters only in Steps 1 and 2, through its entries $A_{ij}$. Let $w_{ij}$ be the block weight of the edge $ij$ under $R_\tau(\varepsilon)$, and put $X_{ij} := w_{ij} - s_\tau$ and $S_{\mathrm{rep}} := \sum_{i<j} X_{ij} X'_{ij}$, together with

$$L := \ell_\tau T + kvd\,(A + A'), \qquad U := T - \frac{\beta}{k}(A + A'), \qquad \lambda := \frac{\Delta}{\sigma^2} \quad (\text{so that } \mu = \lambda^2\sigma^2).$$

*Step 1 (block weights, $s_\tau$, and the range of $r$).* Under $R_\tau(\varepsilon)$ the edges are conditionally independent given $Y$ with $\mathbb{P}(A_{ij} = 1 \mid Y) = q + \Delta w_{ij}$, where

$$w_{ij} \;=\; \tau Y_i Y_j + (1-Y_i)(1-Y_j) \;=\; 1 - Y_i - Y_j + k\,Y_i Y_j.$$

The three cases are checked one by one: when $Y_i = Y_j = 1$ both expressions equal $\tau$; when exactly one of them equals 1 both vanish; when $Y_i = Y_j = 0$ both equal 1. Hence $w_{ij} \in [0,1]$ and $q + \Delta w_{ij} \in [0,1]$, and

$$\mathbb{E}\, w_{ij} \;=\; 1 - 2\varepsilon + k\varepsilon^2 \;=\; s_\tau.$$

From $\varepsilon \leq 1/2$ and $k \leq 2$ we get $k\varepsilon \leq 1$, so $s_\tau = 1 - \varepsilon(2 - k\varepsilon) < 1$, and also $s_\tau \geq 1 - 2\varepsilon + \varepsilon^2 = (1-\varepsilon)^2 > 0$. Therefore

$$0 \;<\; q + \Delta(1-\varepsilon)^2 \;\leq\; r = q + \Delta s_\tau \;<\; q + \Delta \;\leq\; 1.$$

*Step 2 (edgewise likelihood ratio and replica representation).* For any $w$ with $q + \Delta w \in [0,1]$ and any $a \in \{0,1\}$, the identity $q + \Delta w = r + \Delta(w - s_\tau)$ gives

$$\frac{\mathrm{Bern}(a;\, q + \Delta w)}{\mathrm{Bern}(a;\, r)} \;=\; 1 + \lambda\,(w - s_\tau)(a - r),$$

the cases $a = 1$ and $a = 0$ being one line of algebra each, exactly as in (22). Given $Y$ the edges are conditionally independent, so the likelihood ratio of $R_\tau(\varepsilon)$ with respect to $G(n,r)$ at a graph with entries $A_{ij}$ equals $\mathbb{E}_Y \prod_{i<j}(1 + \lambda X_{ij}(A_{ij} - r))$. The sample space is finite, so sums and expectations may be interchanged, and under $G(n,r)$ the sum factorizes over edges, a single edge contributing $\mathbb{E}(A_{ij} - r) = 0$ and $\mathbb{E}(A_{ij} - r)^2 = \sigma^2$, with $\lambda^2\sigma^2 = \mu$. As in (23),

$$1 + \chi^2\big(R_\tau(\varepsilon) \,\big\|\, G(n,r)\big) \;=\; \mathbb{E}_{Y,Y'} \prod_{i<j} \big(1 + \mu\, X_{ij} X'_{ij}\big).$$

*Step 3 (the exact replica sum).* Substituting $Y_i = u_i + \varepsilon$,

$$w_{ij} \;=\; 1 - 2\varepsilon - u_i - u_j + k\big(u_i u_j + \varepsilon(u_i + u_j) + \varepsilon^2\big) \;=\; s_\tau - \beta(u_i + u_j) + k\,u_i u_j,$$

that is, $X_{ij} = -\beta(u_i + u_j) + k\,u_i u_j$. Since $Y_i \in \{0,1\}$ we have $u_i^2 = (1-2\varepsilon)Y_i + \varepsilon^2 = b\,u_i + v$, and likewise $u_i'^2 = b\,u_i' + v$. Expanding $X_{ij}X'_{ij}$ and summing term by term,

$$\sum_{i<j}(u_i+u_j)(u_i'+u_j') \;=\; (n-1)T + \sum_{i\neq j} u_i u_j' \;=\; (n-2)T + AA',$$

$$\sum_{i<j}(u_i+u_j)u_i'u_j' \;=\; \sum_{i\neq j} u_i u_i' u_j' \;=\; TA' - \sum_i u_i u_i'^2 \;=\; TA' - bT - vA,$$

$$\sum_{i<j} u_i u_j (u_i'+u_j') \;=\; TA - bT - vA',$$

$$\sum_{i<j} u_i u_j u_i' u_j' \;=\; \tfrac{1}{2}\Big(T^2 - \sum_i u_i^2 u_i'^2\Big) \;=\; \tfrac{1}{2}\big(T^2 - b^2 T - bv(A+A') - nv^2\big),$$

the last line using $u_i^2 u_i'^2 = (bu_i+v)(bu_i'+v) = b^2 u_i u_i' + bv(u_i+u_i') + v^2$. Combining,

$$S_{\mathrm{rep}} \;=\; \beta^2\big[(n-2)T + AA'\big] - \beta k\big[T(A+A') - 2bT - v(A+A')\big] + \frac{k^2}{2}\big[T^2 - b^2T - bv(A+A') - nv^2\big].$$

The coefficient of $T$ is $\beta^2(n-2) + 2\beta k b - \frac{k^2 b^2}{2} = n\beta^2 - 2\big(\beta - \frac{kb}{2}\big)^2$, and

$$\beta - \frac{kb}{2} \;=\; 1 - k\varepsilon - \frac{k}{2}(1-2\varepsilon) \;=\; 1 - \frac{k}{2} \;=\; \frac{1-\tau}{2} \;=\; d,$$

so that coefficient equals $\ell_\tau$; the coefficient of $(A+A')$ is $\beta k v - \frac{k^2 b v}{2} = kv\big(\beta - \frac{kb}{2}\big) = kvd$. Hence

$$S_{\mathrm{rep}} \;=\; \ell_\tau\, T + \beta^2 AA' - \beta k\, T(A+A') + kvd\,(A+A') + \frac{k^2}{2}\big(T^2 - nv^2\big). \tag{28}$$

For $\tau = 1$ one has $d = 0$, $\beta = b$ and $\ell_1 = nb^2$, and (28) degenerates to $S_{\mathrm{rep}} = b^2(nT + AA') - 2bT(A+A') + 2(T^2 - nv^2)$.

*Step 4 (exponential bound).* Each factor $1 + \mu X_{ij}X'_{ij}$ is the expectation under $\mathrm{Bern}(r)$ of a product of two non-negative likelihood ratios and is therefore non-negative; by $1 + x \le e^x$ and the monotonicity of products of non-negative numbers,

$$1 + \chi^2 \;\le\; \mathbb{E}\exp(\mu S_{\mathrm{rep}}).$$

*Step 5 (completing the square).* Since $(A+A')^2 - 2AA' = A^2 + A'^2$,

$$\frac{k^2}{2}T^2 - \beta k\,T(A+A') + \beta^2 AA' \;=\; \frac{k^2}{2}\Big(T - \frac{\beta}{k}(A+A')\Big)^2 - \frac{\beta^2}{2}\big(A^2+A'^2\big) \;=\; \frac{k^2}{2}U^2 - \frac{\beta^2}{2}\big(A^2+A'^2\big).$$

Here $\mu > 0$, while the subtracted quantities $\frac{\beta^2}{2}(A^2+A'^2)$ and $\frac{k^2 n v^2}{2}$ are both non-negative, so (28) gives

$$\mu S_{\mathrm{rep}} \;\le\; \mu L + \frac{\mu k^2}{2}U^2.$$

*Step 6 (one Hölder step and two exponential moments).* By Hölder's inequality, for $\alpha > 1$ and $\alpha' = \alpha/(\alpha-1)$,

$$\mathbb{E}e^{\mu S_{\mathrm{rep}}} \;\le\; \big(\mathbb{E}e^{\alpha\mu L}\big)^{1/\alpha}\Big(\mathbb{E}\exp\big(\tfrac{\alpha'\mu k^2}{2}U^2\big)\Big)^{1/\alpha'}. \tag{29}$$

*(6a) The linear term.* Here $L = \sum_i \xi_i$ with $\xi_i := \ell_\tau u_i u_i' + kvd\,(u_i + u_i')$ i.i.d. in $i$. From $\mathbb{E}u_i = \mathbb{E}u_i' = 0$, $\mathbb{E}u_i^2 = v$ and the independence of $u$ and $u'$ we get $\mathbb{E}\xi_i = 0$ and $\mathbb{E}[u_i u_i'(u_i + u_i')] = 0$, so

$$\mathrm{Var}(\xi_i) \;=\; \ell_\tau^2 v^2 + 2k^2 v^2 d^2 \cdot v \;=\; v^2\big(\ell_\tau^2 + 2k^2 v d^2\big).$$

From $\varepsilon \le 1/2$ and $k \le 2$ we get $\beta \in [0,1]$, and $d \in [0, 1/2]$, so $-\frac{1}{2} \le \ell_\tau \le n$, that is, $|\ell_\tau| \le n$; and $k^2 d^2 = \frac{(1-\tau^2)^2}{4} \le \frac{1}{4}$ together with $v \le \frac{1}{4}$ gives $2k^2 v d^2 \le \frac{v}{2} \le \frac{1}{8}$, so that, using $n \ge 2$,

$$\mathrm{Var}(\xi_i) \;\le\; v^2\Big(n^2 + \frac{1}{8}\Big) \;\le\; \frac{33}{32}\,n^2 v^2.$$

Moreover $|u_i| \le 1-\varepsilon \le 1$ and $kvd \le \frac{1}{4}$ give $|\xi_i| \le |\ell_\tau| + \frac{1}{2} \le n + \frac{1}{2} \le \frac{5}{4}n$. For a mean-zero random variable with $|\xi| \le C$ and for $0 \le \lambda < 3/C$, the bounds $m! \ge 2 \cdot 3^{m-2}$ and $\mathbb{E}|\xi|^m \le \mathrm{Var}(\xi)\, C^{m-2}$ for $m \ge 2$ give

$$\mathbb{E}e^{\lambda\xi} \;\le\; 1 + \frac{\lambda^2 \mathrm{Var}(\xi)}{2} \sum_{j\ge 0} \Big(\frac{\lambda C}{3}\Big)^j \;=\; 1 + \frac{\lambda^2 \mathrm{Var}(\xi)}{2(1-\lambda C/3)} \;\le\; \exp\Big\{\frac{\lambda^2 \mathrm{Var}(\xi)}{2(1-\lambda C/3)}\Big\}.$$

Taking $\lambda = \alpha\mu$ and $C = \frac{5}{4}n$, so that $\lambda C/3 = \frac{5\alpha\Theta}{12} < 1$,

$$\frac{1}{\alpha}\log \mathbb{E}e^{\alpha\mu L} \;\le\; \frac{\alpha\,\mu^2\, n \cdot \frac{33}{32} n^2 v^2}{2\big(1 - \frac{5\alpha\Theta}{12}\big)} \;=\; \frac{33\,\alpha\,\Gamma^2}{64\big(1 - \frac{5\alpha\Theta}{12}\big)}.$$

*(6b) The square term.* Here $U = \sum_i \zeta_i$ with $\zeta_i := u_i u_i' - \frac{\beta}{k}(u_i + u_i')$ i.i.d. in $i$ and of mean zero. It takes only three values:

$$\zeta_i \;=\; \varepsilon^2 + \frac{2\beta\varepsilon}{k}\;(Y_i = Y_i' = 0), \qquad -v - \frac{\beta b}{k}\;(Y_i \ne Y_i'), \qquad (1-\varepsilon)^2 - \frac{2\beta(1-\varepsilon)}{k}\;(Y_i = Y_i' = 1).$$

Since $\frac{\beta}{k} = \frac{1}{k} - \varepsilon$, the first value minus the second equals $\varepsilon + \frac{\beta}{k} = \frac{1}{k}$, and the third minus the second equals $(1-\varepsilon) - \frac{\beta}{k} = 1 - \frac{1}{k}$; for $k \in [1,2]$ we have $\frac{1}{k} \ge \frac{1}{2} \ge 1 - \frac{1}{k} \ge 0$, so the range of $\zeta_i$ has length exactly $\frac{1}{k}$. Hoeffding's lemma gives $\mathbb{E}e^{t\zeta_i} \le \exp\{t^2/(8k^2)\}$ and hence $\mathbb{E}e^{tU} \le \exp\{nt^2/(8k^2)\}$. For $c > 0$ let $g \sim N(0,1)$ be independent of $U$; the Gaussian moment generating function gives $e^{cU^2/2} = \mathbb{E}_g e^{\sqrt{c}\,gU}$, and since the integrand is non-negative, Tonelli's theorem allows the expectations to be interchanged:

$$\mathbb{E}e^{cU^2/2} \;\le\; \mathbb{E}_g \exp\Big\{\frac{nc\,g^2}{8k^2}\Big\} \;=\; \Big(1 - \frac{nc}{4k^2}\Big)^{-1/2}, \qquad \frac{nc}{4k^2} < 1.$$

Taking $c = \alpha'\mu k^2$, so that $\frac{nc}{4k^2} = \frac{\alpha'\Theta}{4} < 1$, gives $\big(\mathbb{E}e^{\alpha'\mu k^2 U^2/2}\big)^{1/\alpha'} \le \big(1 - \frac{\alpha'\Theta}{4}\big)^{-1/(2\alpha')}$.

*Step 7 (assembly).* Substituting (6a) and (6b) into (29) and combining with Step 4 gives (27). The right-hand side depends on $\tau$ only through $\Theta$ and $\Gamma$, so it is uniform over any range of $\tau$. □

**Corollary C.8.** *If $\Theta \le 1/20$ and $\Gamma \le 1/3$, then, taking $\alpha = \alpha' = 2$ in Lemma C.7,*

$$\chi^2\big(R_\tau(\varepsilon) \,\big\|\, G(n,r)\big) \;\le\; 0.13417, \qquad \mathrm{TV}\big(R_\tau(\varepsilon),\, G(n,r)\big) \;\le\; 0.18315.$$

*Proof.* With $\alpha = \alpha' = 2$ we have $\frac{5\alpha\Theta}{12} = \frac{5\Theta}{6} \le \frac{1}{24}$ and $\frac{\alpha'\Theta}{4} = \frac{\Theta}{2} \le \frac{1}{40}$, so (27) gives

$$1 + \chi^2 \;\le\; \exp\Big\{\frac{66\,\Gamma^2}{64 \cdot \frac{23}{24}}\Big\}\Big(1 - \frac{1}{40}\Big)^{-1/4} \;\le\; e^{1.07610/9} \cdot 1.00635 \;<\; 1.13417,$$

that is, $\chi^2 \le 0.13417$; then (24) gives $\mathrm{TV} \le 0.18315$. □

### C.4 The fallback lower bound

**Theorem C.9** (restatement of Theorem 5)**.** *Let $q_0 \in (0, 1/2]$, $q \in [q_0, 1-q_0]$, $n \ge 2$ and $0 < \Delta \le 1 - q$, and recall*

$$v_0 = \frac{q_0}{2}\Big(1 - \frac{q_0}{2}\Big), \qquad c_1 = \min\Big\{\frac{q_0^2}{4},\, \frac{v_0}{20}\Big\}.$$

*If $n\Delta^2 \le c_1$, then*

$$R_n^*(q,\Delta) \;\ge\; \frac{v_0^2 (\log 2)^2}{576}\, \min\big\{1,\; (n^3\Delta^4)^{-1}\big\}.$$

*Proof. Step 1 (the parameter range).* From $n \ge 2$ and $n\Delta^2 \le c_1 \le q_0^2/4$ we get $\Delta^2 \le c_1/2 \le q_0^2/8$, so $\Delta \le q_0/(2\sqrt{2}) < q_0/2$. Every quantity of the form $r = q + \Delta s$ with $s \in [0,1]$ therefore lies in $[q_0,\; 1 - q_0 + q_0/2] = [q_0,\; 1 - q_0/2]$. The map $x \mapsto x(1-x)$ is concave, so its minimum over

that interval is attained at an endpoint, and the two endpoint values are $q_0(1-q_0)$ and $v_0$; since $q_0 \le 1/2 \le 2/3$ gives $v_0 \le q_0(1-q_0)$,

$$\sigma^2 = r(1-r) \;\ge\; v_0, \qquad \Theta = \frac{n\Delta^2}{\sigma^2} \;\le\; \frac{c_1}{v_0} \;\le\; \frac{1}{20}.$$

*Step 2 (the two points and a common reference).* Take

$$\varepsilon \;:=\; \min\Big\{\frac{1}{12},\; \frac{v_0}{3}\,(n^3\Delta^4)^{-1/2}\Big\} \;\in\; (0, 1/12], \tag{30}$$

let $\varepsilon'$ be as in Remark B.3, so that $\varepsilon - \varepsilon^2 < \varepsilon' < \varepsilon$, and take the two points $\theta_\varepsilon = (1-\varepsilon, \varepsilon, 0)$ and $\theta'_\varepsilon = (1-\varepsilon', \varepsilon'/2, \varepsilon'/2)$. A direct computation gives

$$P_2(\theta_\varepsilon) = 1 - 2\varepsilon + 2\varepsilon^2 = s_1(\varepsilon), \qquad P_2(\theta'_\varepsilon) = 1 - 2\varepsilon' + \tfrac{3}{2}\varepsilon'^2 = s_{1/2}(\varepsilon'),$$

and the defining property of $\varepsilon'$ is precisely $P_2(\theta'_\varepsilon) = P_2(\theta_\varepsilon)$. Write $s$ for this common value and put $r := q + \Delta s$. Then $R_1(\varepsilon)$ and $R_{1/2}(\varepsilon')$ have the same reference distribution $G(n, r)$ in Lemma C.7. Moreover $P_{\theta_\varepsilon} = R_1(\varepsilon)$ by Definition C.1.

*Step 3 (three pieces of total variation).* By Step 1, $\Theta \le 1/20$.

(a) At $\tau = 1$ with planting probability $\varepsilon$, the quantity $\Gamma$ satisfies $\Gamma \le \frac{n^{3/2}\Delta^2\varepsilon}{v_0} \le \frac{1}{3}$, using $v \le \varepsilon$, $\sigma^2 \ge v_0$ and the choice (30). Corollary C.8 then gives $\mathrm{TV}(R_1(\varepsilon), G(n, r)) \le 0.18315$.

(b) Since $\varepsilon' < \varepsilon \le 1/2$ implies $\varepsilon'(1-\varepsilon') < v$, the quantity $\Gamma$ at $\tau = 1/2$ with planting probability $\varepsilon'$ is smaller, and again $\mathrm{TV}(R_{1/2}(\varepsilon'), G(n, r)) \le 0.18315$.

(c) From $\Delta \le q_0$ and $n\Delta^2 \le c_1 \le v_0/20 \le 2v_0$, Corollary C.5 applies and gives

$$\mathrm{TV}\big(P_{\theta'_\varepsilon},\, R_{1/2}(\varepsilon')\big) \;\le\; \frac{\kappa}{4v_0}\, n\Delta^2\, \varepsilon' \;\le\; \frac{\kappa}{4v_0}\cdot\frac{v_0}{20}\cdot\frac{1}{12} \;=\; \frac{\kappa}{960} \;<\; 3.81\times 10^{-4}.$$

The triangle inequality now gives

$$\mathrm{TV}\big(P_{\theta_\varepsilon},\, P_{\theta'_\varepsilon}\big) \;\le\; 0.18315 + 0.18315 + 0.00039 \;=\; 0.36669 \;<\; \frac{1}{2}.$$

*Step 4 (the entropy gap).* By Remark B.3, $H(\theta_\varepsilon) = h(\varepsilon)$ and $H(\theta'_\varepsilon) = h(\varepsilon') + \varepsilon'\log 2$, where $h(x) := -x\log x - (1-x)\log(1-x)$, and $|h(\varepsilon) - h(\varepsilon')| \le (\varepsilon - \varepsilon')\log(1/\varepsilon') \le \varepsilon^2 \log(2/\varepsilon)$, so

$$\omega_\varepsilon \;:=\; H(\theta'_\varepsilon) - H(\theta_\varepsilon) \;\ge\; (\varepsilon - \varepsilon^2)\log 2 - \varepsilon^2 \log\frac{2}{\varepsilon}.$$

The right-hand side is at least $\frac{\varepsilon\log 2}{2}$ if and only if $\log 2\,(\frac{1}{2} - \varepsilon) \ge \varepsilon \log\frac{2}{\varepsilon}$; the left-hand side decreases in $\varepsilon$ and the right-hand side increases on $(0, 2/e)$, while at $\varepsilon = 1/12$ the left side equals $0.28881$ and the right-hand side equals $0.26484$, so the inequality holds on $(0, 1/12]$. Hence $\omega_\varepsilon \ge \frac{\varepsilon\log 2}{2} > 0$.

*Step 5 (Le Cam's two-point method).* Step 3 gives $\mathrm{TV} \le 1/2$, and Steps 4 and 5 of the proof of Proposition C.3, including the treatment of external randomization, apply verbatim to the two points $\theta_\varepsilon, \theta'_\varepsilon$ with gap $\omega_\varepsilon$. They give

$$R_n^*(q, \Delta) \;\ge\; \frac{\omega_\varepsilon^2}{16} \;\ge\; \frac{\varepsilon^2(\log 2)^2}{64}.$$

*Step 6 (substituting $\varepsilon$).* For $A, B > 0$ and $x > 0$ one has $\min\{A, Bx\} \ge \min\{A, B\}\min\{1, x\}$. Taking $A = 1/144$, $B = v_0^2/9$ and $x = (n^3\Delta^4)^{-1}$, and using $v_0 \le 1/4$, so that $\frac{v_0^2}{9} \le \frac{1}{144}$, (30) gives

$$\varepsilon^2 \;=\; \min\Big\{\frac{1}{144},\; \frac{v_0^2}{9}(n^3\Delta^4)^{-1}\Big\} \;\ge\; \frac{v_0^2}{9}\min\big\{1, (n^3\Delta^4)^{-1}\big\}.$$

Combining with Step 5 completes the proof. □

### C.5 THE CONSISTENCY THRESHOLD

**Proposition C.10.** *Let $q_0 \in (0, 1/2]$ and $C < \infty$. If $q \in [q_0, 1 - q_0]$, $n^3\Delta^4 \le C$ and $n \ge C/c_1^2$, then*

$$R_n^*(q, \Delta) \;\ge\; \frac{v_0^2(\log 2)^2}{576}\,\min\{1,\, C^{-1}\} \;>\; 0.$$

*Proof.* From $n^3\Delta^4 \le C$ we get $(n\Delta^2)^2 = n^3\Delta^4/n \le C/n$, so $n \ge C/c_1^2$ gives $n\Delta^2 \le \sqrt{C/n} \le c_1$ and Theorem C.9 applies; moreover $\min\{1, (n^3\Delta^4)^{-1}\} \ge \min\{1, C^{-1}\}$. □

**Proposition C.11** (restatement of Proposition 6)**.** *Let $q_0 \in (0, 1/2]$ and consider sequences $(q_n, \Delta_n)$ with $q_n \in [q_0, 1 - q_0]$ and $0 < \Delta_n \le 1 - q_n$. There is a sequence of estimators $\widehat{H}_n$ with*

$$\sup_{\theta \in \Delta_3} \mathbb{E}_\theta\big(\widehat{H}_n - H(\theta)\big)^2 \;\longrightarrow\; 0$$

*if and only if $n^3\Delta_n^4 \to \infty$; sufficiency is furnished by the projected moment estimator of Algorithm 1. In particular, for $\Delta_n = n^{-\gamma}$ the critical exponent is $\gamma_c(3) = 3/4$.*

*Proof.* Sufficiency: when $n^3\Delta_n^4 \to \infty$, Theorem 2 makes the worst-case risk of the projected moment estimator tend to zero, and its bound is uniform in $q$, so it applies in particular to $q_n \in [q_0, 1 - q_0]$.

Necessity: suppose $n^3\Delta_n^4 \not\to \infty$. Then there are $C < \infty$ and a subsequence $n_1 < n_2 < \cdots$ with $n_j^3\Delta_{n_j}^4 \le C$. Only finitely many terms of the subsequence are smaller than $C/c_1^2$, and for all the others Proposition C.10 gives $R_{n_j}^* \ge \frac{v_0^2(\log 2)^2}{576}\min\{1, C^{-1}\} > 0$, so no sequence of estimators has worst-case risk tending to zero.

Finally, for $\Delta_n = n^{-\gamma}$ we have $n^3\Delta_n^4 = n^{3-4\gamma} \to \infty$ if and only if $\gamma < 3/4$, so $\gamma_c(3) = 3/4$. □

## D PROOF OF THE MATCHING LOWER BOUND

This appendix proves the second-moment estimate that carries the middle inequality of (6), the indistinguishability of the least-favourable pair that follows from it, and the two corollaries that turn that pair into a rate: first on the window (5), then on the whole weakly separated region. Figure 4 shows how its parts fit together.

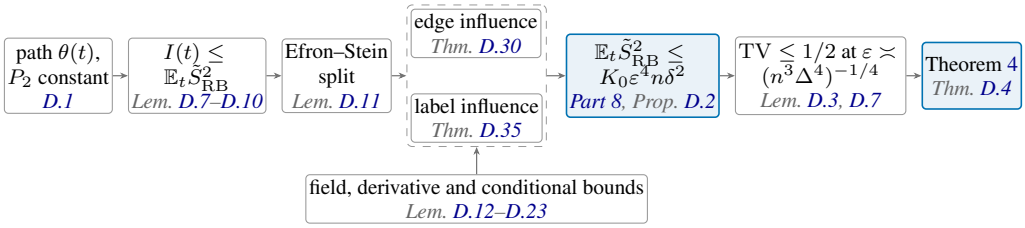



Figure 4: The proof of Theorem 4. Each box is a step of the argument; below it, in italics, is where this appendix proves it.

### D.1 SETTING AND NOTATION

**Standing hypotheses.** Throughout this appendix the observations follow (1), and

$$q \in [q_0, 1 - q_0], \qquad n \ge n_0, \qquad \delta := n\Delta^2 \le c_1, \qquad \varepsilon \le \tfrac{1}{12},$$

with

$$s_0 := 96, \qquad n_0 := \max\{3,\; 16 s_0^2 c_1/q_0^2\}.$$

Here $q_0 \in (0, 1/2]$, $v_0 = \frac{q_0}{2}(1 - \frac{q_0}{2})$ and $c_1$ are the constants of Section 4; everything below uses only that $c_1 \in (0, 1]$ and $c_1 \le \min\{q_0^2/4,\; v_0/20\}$, so the value $c_1 = \min\{q_0^2/4, v_0/20\}$ fixed in the main body is admissible. Proposition D.2 needs one further hypothesis, $n\varepsilon \ge 1$ (Remark D.1).

The bound $c_1 \le \min\{q_0^2/4, v_0/20\}$ is what lets the imported results used here run under a single window constant, and shrinking $c_1$ does not affect them. Taking $c_1 := \min\{q_0^2/(3s_0^2),\ v_0/20\}$ gives the cleanest constants: the quantity $\Lambda_*$ of Lemma D.18 is then at most $6e^2$ and $n_0 \le 16/3$, so that $n \ge 6$ suffices.

The hypothesis $\varepsilon \le 1/12$ is not implied by the others, none of which bounds $\varepsilon$ from above: taking $\delta = a_0^2/(E^2\sqrt{n})$ produces $\varepsilon = E$ with all remaining conditions satisfied once $n$ is large. It is used for the existence of the path $\theta(t)$, for the path scales of Part 3 and for the entropy-gap estimate of Part 8. Inside the parameter range of Theorem D.4 it is guaranteed by $n^3\Delta^4 \ge C_0 \ge (12a_0)^4$.

**The hard pair and the path.** For $\varepsilon \in (0, 1/2]$ write $\theta_\varepsilon = (1-\varepsilon, \varepsilon, 0)$ and $\theta'_\varepsilon = (1-\varepsilon', \varepsilon'/2, \varepsilon'/2)$, where $\varepsilon'$ is determined by $P_2(\theta'_\varepsilon) = P_2(\theta_\varepsilon) =: s$ and satisfies $\varepsilon - \varepsilon^2 < \varepsilon' < \varepsilon$. For $\eta \in [0, \varepsilon'/2]$ let $e(\eta)$ be the smaller root of $(1-e)^2 + (e-\eta)^2 + \eta^2 = s$, that is

$$e(\eta) = \tfrac{1}{2}\big[(1+\eta) - \sqrt{(1-2\varepsilon)^2 + 2\eta - 3\eta^2}\big], \qquad e(0) = \varepsilon, \quad e(\varepsilon'/2) = \varepsilon'. \tag{31}$$

The path is $\eta(t) := \frac{\varepsilon'}{2}t^2$ and $\theta(t) := (1-e,\ e-\eta,\ \eta)$ for $t \in [0,1]$, so that $\theta(0) = \theta_\varepsilon$ and $\theta(1) = \theta'_\varepsilon$. Its derivatives are $\dot\eta = \varepsilon' t$, $\dot e = \dot\eta(e-2\eta)/(2e-\eta-1)$ and $\dot\theta = (-\dot e,\ \dot e - \dot\eta,\ \dot\eta)$, and along the whole path

$$\sum_c \dot\theta_c = 0, \qquad \sum_c \theta_c \dot\theta_c = 0 \quad \text{(that is, } P_2 \text{ is constant)}. \tag{32}$$

We write $P_t := P_{\theta(t)}$ and let $\mathbb{E}_t$ be the joint expectation of $(Z, A)$ under $\theta(t)$. The letter $s$ is reused locally three more times below, always in a fixed context: for the order of an exponential moment in Parts 3 and 4, for the interpolation parameter of a fractional label in Part 7, and for the abbreviation $s = n^3\Delta^4$ in the proof of Corollary D.6.

**Local fields and posteriors.** Put $\lambda := \log(p/q) > 0$ and $\nu := \log\frac{1-p}{1-q} < 0$, and $J_{ij} := \nu + (\lambda - \nu)A_{ij} \in \{\nu, \lambda\}$ for $i \ne j$, $J_{ii} := 0$. With $\mathbf{e}_c$ the $c$-th standard basis vector and $E_k := \mathbf{e}_{Z_k}$, set

$$F_i := \sum_{j \ne i} J_{ij} E_j, \qquad \bar F_i := \sum_b \theta_b F_{ib}, \qquad G_{ic} := F_{ic} - \bar F_i, \qquad \gamma_i := \max_c |F_{ic}|.$$

The quantity $\gamma_i$ is the largest absolute value of the field at vertex $i$ and is unrelated to the exponent $\gamma$ of the parametrization $\Delta = n^{-\gamma}$. In expressions such as $\sum_b \theta_b e^{F_b}$ the index $b$ is a dummy class index; $a$ and $b$ appear as vertices only in the context of one fixed edge $\{a, b\}$. Further,

$$\pi_{ic} := \frac{\theta_c e^{F_{ic}}}{\sum_b \theta_b e^{F_{ib}}}, \qquad \rho_{ic} := \frac{\pi_{ic}}{\theta_c}, \qquad V_k := \operatorname{diag}(\pi_k) - \pi_k \pi_k^\top, \qquad \psi_c := \frac{\dot\theta_c}{\theta_c}, \qquad w_c := \theta_c \dot\theta_c.$$

**The score and its rewriting.** Let $S_{\mathrm{obs}}(A) := \partial_t \log P_t(A)$ and let $I(t) := \mathbb{E}_t S_{\mathrm{obs}}^2$ be the Fisher information of the observed graph along the path. Set

$$\Phi(F) := \frac{\sum_b \dot\theta_b e^{F_b}}{\sum_b \theta_b e^{F_b}}, \qquad \tilde S_1 := \sum_i \Phi(F_i), \qquad \mathcal{P}_k f := \mathbb{E}_t[f \mid A, Z_{-k}],$$

$$\tilde S_{\mathrm{RB}} := \sum_k \mathcal{P}_k \tilde S_1 - (n-1)\tilde S_1,$$

so that $\tilde S_{\mathrm{RB}}$ is one Rao–Blackwell sweep applied to $\tilde S_1$. The centred edge count is

$$\mathcal{E}(A) := \sum_{i<j} A_{ij} - \binom{n}{2}(q + \Delta s).$$

**$L^{\mathsf{p}}$ notation and label counts.** For label information $\mathcal{L}$ write $\|X\|_{\mathsf{p},Z} := (\mathbb{E}[|X|^{\mathsf{p}} \mid \mathcal{L}])^{1/\mathsf{p}}$, where $\mathsf{p}$ is the order of the norm and is unrelated to the edge probability $p = q + \Delta$. From Part 4 on, $\mathcal{L}$ is $\sigma(Z)$ or its join with a resampled label and with the event of one edge, as the context dictates. Let

$$N_c := \#\{k : Z_k = c\}, \qquad N_{\min} := N_2 + N_3, \qquad \nu_c := \frac{N_c}{n},$$

$$\nu_c^+ := \nu_c + \frac{2}{n}, \qquad \nu_{\min}^+ := \nu_2^+ + \nu_3^+.$$

In this appendix $N_2$ and $N_3$ have only this meaning: the counts $\binom{n}{2}$ and $n(n-1)(n-2)$ of Section 3 are never denoted by these two symbols here. The class frequencies $\nu_c$ always carry a class subscript and are not the edge weight $\nu$.

**Scales.** Write $\delta := n\Delta^2$ and $a^4 := \varepsilon^4 n\delta^2 = \varepsilon^4\, n^3\Delta^4$. In this appendix $C$ and $C_m$ denote finite positive constants that depend only on $(q_0, c_1)$ and may take different values at different occurrences. Every $L^{\mathrm{p}}$ order used below lies in the fixed range $[2, 32]$, and the order of an exponential weight never exceeds $s_0 = 96$.

*Remark* D.1. The blockwise estimate needs $n\varepsilon \geq 1$, in three places: Lemma D.23, where it controls the fluctuation of the number of minority-class vertices; the exceptional-edge term of Proposition D.29, where $\Delta^2\varepsilon \leq \sqrt{\delta}\,\varepsilon^2$ uses $1/n \leq \varepsilon$; and the absorption of $\varepsilon^3/n$ in Theorem D.30. Part 6 exhibits a block whose own second moment is of order $1/(n\varepsilon)$ rather than $O(\delta)$, so the requirement is tight for the blockwise route rather than an artefact of it. At the fixed amplitude $\varepsilon = a_0(n^3\Delta^4)^{-1/4}$ at which Lemma D.3 and Theorem D.4 call Proposition D.2, the requirement is automatic: $n\varepsilon = a_0 n^{3/4}\delta^{-1/2} \geq a_0 n^{3/4} \geq 1$ as soon as $n \geq a_0^{-4/3}$, which the window (5) enforces.

## D.2 Statements

**Proposition D.2** (second moment of the rewritten score)**.** *There is a $K_0 < \infty$ depending only on $(q_0, c_1)$ such that, as soon as*

$$n \geq n_0, \qquad q \in [q_0, 1-q_0], \qquad \delta := n\Delta^2 \leq c_1, \qquad \varepsilon \leq \tfrac{1}{12}, \qquad n\varepsilon \geq 1,$$

*one has, for every $t \in (0, 1]$,*

$$\mathbb{E}_t\big(\tilde{S}_{\mathrm{RB}}\big)^2 \;\leq\; K_0\,\varepsilon^4 n\delta^2. \tag{33}$$

**Lemma D.3** (indistinguishability at the fixed amplitude)**.** *Put $a_0 := K_0^{-1/4}$ and*

$$C_0 := \max\big\{(12a_0)^4,\; c_1^2 \max\{n_0,\; a_0^{-4/3}\}\big\}. \tag{34}$$

*Let $q \in [q_0, 1-q_0]$, let $(n, \Delta)$ satisfy $n^3\Delta^4 \geq C_0$ and $n\Delta^2 \leq c_1$, and take $\varepsilon := a_0(n^3\Delta^4)^{-1/4}$. Then*

$$\mathrm{TV}\big(P_{\theta_\varepsilon},\; P_{\theta'_\varepsilon}\big) \;\leq\; \frac{1}{2}.$$

**Theorem D.4** (restatement of Theorem 4)**.** *On the same parameter range, that is, for*

$$q \in [q_0, 1-q_0], \qquad C_0^{1/4}\, n^{-3/4} \;\leq\; \Delta \;\leq\; \sqrt{c_1}\, n^{-1/2},$$

*one has*

$$R_n^*(q, \Delta) \;\geq\; \frac{(\log 2)^2}{64}\, K_0^{-1/2}\, \big(n^3\Delta^4\big)^{-1/2}.$$

*Once $c_1$ is taken to be the function of $q_0$ fixed in Appendix D.1, the two constants $c_2 := C_0^{1/4}$ and $c := \frac{(\log 2)^2}{64} K_0^{-1/2}$ depend only on $q_0$.*

**Corollary D.5** (the rate inside the window)**.** *On the parameter range of Theorem D.4, together with the specialization of Theorem 2 to $\Delta \leq \sqrt{c_1} n^{-1/2}$,*

$$\frac{(\log 2)^2}{64} K_0^{-1/2} \;\leq\; \frac{R_n^*(q, \Delta)}{(n^3\Delta^4)^{-1/2}} \;\leq\; 81\Big(c_1^2 + \frac{c_1}{4} + \frac{1}{16}\Big)^{1/2},$$

*that is, $R_n^*(q, \Delta) \asymp (n^3\Delta^4)^{-1/2}$, the upper bound being attained by the projected moment estimator of Algorithm 1.*

**Corollary D.6** (restatement of Corollary 3)**.** *Let $q_0$, $c_1$, $C_0$ and $K_0$ be as above, and let $q \in [q_0, 1-q_0]$, $n \geq 3$ and $n\Delta^2 \leq c_1$; the requirement $n^3\Delta^4 \geq C_0$ is dropped. Then*

$$c_\star \min\big\{1,\; (n^3\Delta^4)^{-1/2}\big\} \;\leq\; R_n^*(q, \Delta) \;\leq\; C_\star \min\big\{1,\; (n^3\Delta^4)^{-1/2}\big\},$$

$$c_\star := \min\Big\{\frac{(\log 2)^2}{64} K_0^{-1/2},\; \frac{v_0^2 (\log 2)^2}{576} \min\{1, C_0^{-1/2}\}\Big\},$$

$$C_\star := \max\Big\{81\Big(c_1^2 + \frac{c_1}{4} + \frac{1}{16}\Big)^{1/2},\; (\log 3)^2\Big\},$$

*and both constants depend only on $q_0$. Thus $R_n^*(q, \Delta) \asymp \min\{1, (n^3\Delta^4)^{-1/2}\}$ holds on the whole of $n\Delta^2 \leq c_1$, with no unmatched segment left.*

### D.3 PART 1: THE PATH, THE SCORE AND THE ONE-SWEEP REWRITING

**Lemma D.7** (Hellinger path bound). *Let $\{P_t\}_{t\in[0,1]}$ be a family of distributions on a finite set with $P_t(A) > 0$ and with $t \mapsto P_t(A)$ continuously differentiable on $[0,1]$. Write $I(t) := \sum_A P_t(A)(\partial_t \log P_t(A))^2$ and $H^2(P,Q) := \frac{1}{2}\sum_A(\sqrt{P(A)} - \sqrt{Q(A)})^2$ for the squared Hellinger distance, a two-argument symbol that is not the entropy $H(\cdot)$. Then $H^2(P_0,P_1) \le \frac{1}{8}\int_0^1 I(t)\,\mathrm{d}t$ and $\mathrm{TV}(P_0,P_1) \le \sqrt{2}\,H(P_0,P_1)$, so that*

$$\mathrm{TV}(P_0,P_1) \;\le\; \frac{1}{2}\Big(\int_0^1 I(t)\,\mathrm{d}t\Big)^{1/2}. \tag{35}$$

*Proof.* We have $\sqrt{P_1(A)} - \sqrt{P_0(A)} = \int_0^1 \frac{1}{2}\sqrt{P_t(A)}\,\partial_t \log P_t(A)\,\mathrm{d}t$, and by Cauchy–Schwarz on the unit interval its square is at most $\frac{1}{4}\int_0^1 P_t(A)(\partial_t \log P_t(A))^2\mathrm{d}t$; summing over $A$ gives $2H^2 \le \frac{1}{4}\int_0^1 I$. On the other hand $\sum_A |P-Q| = \sum_A |\sqrt{P}-\sqrt{Q}|(\sqrt{P}+\sqrt{Q}) \le (2H^2)^{1/2}(\sum_A(\sqrt{P}+\sqrt{Q})^2)^{1/2} \le (2H^2)^{1/2}\cdot 2$, whence $\mathrm{TV} \le \sqrt{2}H$. □

**Applicability.** When $0 < q < p < 1$, every graph has positive probability under every label configuration, and $P_{\theta(t)}(A) = \sum_Z \prod_i \theta_{Z_i}(t) \prod_{i<j} \mathrm{Bern}(A_{ij}; q + \Delta\mathbf{1}\{Z_i = Z_j\})$ is a polynomial in $\theta(t)$, while $\theta(t)$ is continuously differentiable on $[0,1]$: the discriminant under the root in (31) satisfies $(1-2\varepsilon)^2 + 2\eta - 3\eta^2 > 0$ because $\eta \le \varepsilon'/2 \le 1/24$.

**Lemma D.8** (Glauber identity). *For $t \in (0,1]$, a vertex $i$ and a class $c$ one has $\mathbb{P}_t(Z_i = c \mid A, Z_{-i}) = \pi_{ic}$, and for every function $\chi(A, Z_{-i})$, $\mathbb{E}_t[\mathbf{1}\{Z_i = c\}\chi \mid A] = \mathbb{E}_t[\pi_{ic}\chi \mid A]$.*

*Proof.* Given $Z_{-i}$ we have $\mathbb{P}_t(Z_i = c, A \mid Z_{-i}) = \theta_c \prod_{j\neq i} \mathrm{Bern}(A_{ij}; q + \Delta\mathbf{1}\{Z_j = c\}) \cdot$ (the factors of the edges not meeting $i$). After division by $\prod_{j\neq i}\mathrm{Bern}(A_{ij};q)$, which does not depend on $c$, the relative weight in $c$ is $\theta_c \exp(\sum_{j\neq i}\mathbf{1}\{Z_j = c\}J_{ij}) = \theta_c e^{F_{ic}}$, and normalizing gives $\pi_{ic}$. For the second identity, condition first on $(A, Z_{-i})$, where $\chi$ comes out of the expectation, then apply the first identity, then take the conditional expectation given $A$. □

**Lemma D.9** (edge-count orthogonality and a second-moment bound). *Let $t \in (0,1]$. If $\tilde{S}(A,Z)$ satisfies $\mathbb{E}_t[\tilde{S} \mid A] = S_{\mathrm{obs}}(A)$, then for all constants $\beta, \beta_0$,*

$$I(t) \;\le\; \mathbb{E}_t\big[(\tilde{S} - \beta\mathcal{E} - \beta_0)^2\big].$$

*In particular, taking $\beta = \beta_0 = 0$, $I(t) \le \mathbb{E}_t\tilde{S}^2$.*

*Proof.* First, $\mathbb{E}_t S_{\mathrm{obs}} = \sum_A \partial_t P_t(A) = \partial_t 1 = 0$. Next, $\mathbb{E}_t[A_{ij}] = q + \Delta P_2(\theta(t)) = q + \Delta s$ does not depend on $t$ by (32), so $\mathbb{E}_t[S_{\mathrm{obs}}\mathcal{E}] = \partial_t \mathbb{E}_t[\mathcal{E}] = 0$. Hence $\mathbb{E}_t[(S_{\mathrm{obs}} - \beta\mathcal{E} - \beta_0)^2] = I(t) + \mathbb{E}_t[(\beta\mathcal{E} + \beta_0)^2] \ge I(t)$. Finally $S_{\mathrm{obs}} - \beta\mathcal{E} - \beta_0 = \mathbb{E}_t[\tilde{S} - \beta\mathcal{E} - \beta_0 \mid A]$, and by conditional Jensen its second moment is at most $\mathbb{E}_t[(\tilde{S} - \beta\mathcal{E} - \beta_0)^2]$. □

**Lemma D.10** (explicit rewriting). *For $t \in (0,1]$:*

*(i) all $\theta_c(t) > 0$, and $S_{\mathrm{obs}}(A) = \mathbb{E}_t\big[\sum_i \dot{\theta}_{Z_i}/\theta_{Z_i} \,\big|\, A\big]$;*

*(ii) $\tilde{S}_1 = \sum_i \Phi(F_i) = \sum_i \sum_c \dot{\theta}_c(\rho_{ic} - 1)$ satisfies $\mathbb{E}_t[\tilde{S}_1 \mid A] = S_{\mathrm{obs}}(A)$;*

*(iii) $\tilde{S}_{\mathrm{RB}}$ satisfies $\mathbb{E}_t[\tilde{S}_{\mathrm{RB}} \mid A] = S_{\mathrm{obs}}(A)$, hence $\mathbb{E}_t\tilde{S}_{\mathrm{RB}} = 0$ and $\mathbb{E}_t\tilde{S}_{\mathrm{RB}}^2 = \mathrm{Var}_t(\tilde{S}_{\mathrm{RB}})$.*

*Proof.* (i) Differentiate the polynomial expression for $P_{\theta(t)}(A)$ term by term.

(ii) Write $\psi_{Z_i} = \sum_c \mathbf{1}\{Z_i = c\}\psi_c$. Lemma D.8 with $\chi \equiv 1$ and the identity $\pi_{ic}\psi_c = \dot{\theta}_c\rho_{ic}$ give $\mathbb{E}_t[\psi_{Z_i} \mid A] = \mathbb{E}_t[\sum_c \dot{\theta}_c\rho_{ic} \mid A]$; and $\rho_{ic} = e^{F_{ic}}/\sum_b \theta_b e^{F_{ib}}$ together with $\sum_c \dot{\theta}_c = 0$ give $\sum_c \dot{\theta}_c\rho_{ic} = \sum_c \dot{\theta}_c(\rho_{ic} - 1) = \Phi(F_i)$. Combine with (i).

(iii) Write $\tilde{S}_{\mathrm{RB}} = \tilde{S}_1 + \sum_k(\mathcal{P}_k\tilde{S}_1 - \tilde{S}_1)$. Each term $\mathcal{P}_k\tilde{S}_1 - \tilde{S}_1$ has conditional mean zero given $A$ by the tower property, so the conditional expectation is unchanged; then use $\mathbb{E}_t S_{\mathrm{obs}} = 0$. □

This rewriting replaces the dependence of $\tilde{S}_1$ on a single label by the local conditional probability exactly, at all orders, with no Taylor truncation. It is the vehicle for an upper bound on $I(t)$, and the last identity of (iii) is the entry point to Part 2: the second moment is a variance, so it can be estimated coordinate by coordinate without centring the components first.

### D.4 Part 2: reduction by Efron–Stein

**Lemma D.11** (Efron–Stein reduction)**.** *Under $P_t$,*

$$\mathbb{E}_t \tilde{S}_{\mathrm{RB}}^2 \;\le\; \frac{n(n-1)}{4}\,\mathbb{E}_t\big[(\tilde{S}_{\mathrm{RB}} - \tilde{S}_{\mathrm{RB}}^{(12)})^2\big] \;+\; \frac{n}{2}\,\mathbb{E}_t\big[(\tilde{S}_{\mathrm{RB}} - \tilde{S}_{\mathrm{RB}}^{(1)})^2\big],$$

*where $\tilde{S}_{\mathrm{RB}}^{(ij)}$ resamples the uniform variable of the edge $\{i,j\}$ once and $\tilde{S}_{\mathrm{RB}}^{(v)}$ resamples the label $Z_v$ once, all other inputs being left unchanged.*

*Proof.* Use the independent-input representation: $Z_1,\dots,Z_n$ are i.i.d. from $\theta(t)$, $\{U_e\}_e$ are i.i.d. uniform on $[0,1]$ and independent of $Z$, and $A_{ij} := \mathbf{1}\{U_{ij} < q + \Delta \mathbf{1}\{Z_i = Z_j\}\}$; this gives the correct joint law. Then $\tilde{S}_{\mathrm{RB}}$ is a function $f$ of these $n + \binom{n}{2}$ independent inputs. The Efron–Stein inequality gives $\mathrm{Var}_t(f) \le \frac{1}{2}\sum_i \mathbb{E}_t[(f - f^{(i)})^2]$, the sum running over all input coordinates and $f^{(i)}$ replacing the $i$-th coordinate by an i.i.d. copy, while Lemma D.10(iii) gives $\mathrm{Var}_t(f) = \mathbb{E}_t \tilde{S}_{\mathrm{RB}}^2$. The joint law of $(Z,U)$ and the function $f$ are both invariant under permutations of the vertices, since $\tilde{S}_1 = \sum_i \Phi(F_i)$ and $\sum_k \mathcal{P}_k - (n-1)$ are symmetric in the vertices, so the $\binom{n}{2}$ edge terms are equal to one another and the $n$ label terms are equal to one another. Substituting gives the claim. □

**The two targets.** To obtain $\mathbb{E}_t \tilde{S}_{\mathrm{RB}}^2 \le C\varepsilon^4 n\delta^2$ it suffices to prove

$$\text{(ES-E)} \qquad \mathbb{E}_t\big[(\tilde{S}_{\mathrm{RB}} - \tilde{S}_{\mathrm{RB}}^{(ab)})^2\big] \;\le\; C\Delta^2\varepsilon^4\delta, \tag{36}$$

$$\text{(ES-L)} \qquad \mathbb{E}_t\big[(\tilde{S}_{\mathrm{RB}} - \tilde{S}_{\mathrm{RB}}^{(v)})^2\big] \;\le\; C\varepsilon^4\delta^2, \tag{37}$$

since then the edge term is at most $\frac{n^2}{4} C\Delta^2\varepsilon^4\delta = \frac{C}{4}\varepsilon^4\delta \cdot n^2\Delta^2 = \frac{C}{4}\varepsilon^4 n\delta^2$ and the label term is at most $\frac{C}{2}\varepsilon^4 n\delta^2$. Both right-hand sides are exactly a constant multiple of $a^4$, so the split leaves no slack: if either term exceeded its target by a diverging factor, the whole route would fail. Equation (36) is proved in Part 6 (Theorem D.30) and (37) in Part 7 (Theorem D.35).

### D.5 Part 3: field bounds and derivative bounds for $\Phi$

Throughout this part $q \in [q_0, 1-q_0]$ and $0 < \Delta \le q_0/2$, the latter being guaranteed by $\delta \le c_1 \le q_0^2/4$ together with $n \ge 3$; hence $p \le 1 - q_0/2$.

**Lemma D.12** (deterministic bounds)**.**

$$0 < \lambda \le \frac{\Delta}{q_0}, \qquad 0 < -\nu \le \frac{2\Delta}{q_0}, \qquad |J_{ij}| \le \frac{2\Delta}{q_0}, \qquad 0 < \lambda - \nu \le \frac{3\Delta}{q_0}.$$

*Proof.* $\lambda = \log(1 + \Delta/q) \le \Delta/q \le \Delta/q_0$. For $\nu = \log(1 - \frac{\Delta}{1-q})$ use $\log(1-x) \ge -x/(1-x)$ for $x \in [0,1)$ with $x = \Delta/(1-q)$, which gives $\nu \ge -\Delta/(1-q-\Delta) = -\Delta/(1-p) \ge -2\Delta/q_0$. The remaining two bounds follow from the first two. □

**Lemma D.13** (first and second order given the labels)**.** *Given $Z$, the weights $J_{ij}$ with $i < j$ are conditionally independent, and*

$$\mathbb{E}[J_{ij} \mid Z] = \begin{cases} D(p \,\|\, q), & Z_i = Z_j, \\ -D(q \,\|\, p), & Z_i \ne Z_j, \end{cases} \qquad \big|\mathbb{E}[J_{ij} \mid Z]\big| \le \frac{2\Delta^2}{q_0^2}, \qquad \mathrm{Var}(J_{ij} \mid Z) \le \frac{3\Delta^2}{q_0^2},$$

*where $D(\cdot \,\|\, \cdot)$ is the binary Kullback–Leibler divergence.*

*Proof.* We have $\mathbb{E}[J_{ij} \mid Z] = \nu + (\lambda - \nu) r_{ij}$ with $r_{ij} \in \{q, p\}$. For $r_{ij} = p$ this is $p \log \frac{p}{q} + (1 - p) \log \frac{1-p}{1-q} = D(p\|q)$, and for $r_{ij} = q$ it is $-D(q\|p)$. Now use $D(x\|y) \le \frac{(x-y)^2}{y(1-y)}$:

$$D(p\|q) \le \frac{\Delta^2}{q(1-q)} \le \frac{\Delta^2}{q_0(1-q_0)} \le \frac{2\Delta^2}{q_0}, \qquad D(q\|p) \le \frac{\Delta^2}{p(1-p)} \le \frac{\Delta^2}{q_0 \cdot q_0/2} = \frac{2\Delta^2}{q_0^2},$$

using $1 - q_0 \ge 1/2$, $p \ge q \ge q_0$ and $1 - p \ge q_0/2$; and $2/q_0 \le 2/q_0^2$ because $q_0 \le 1/2$. For the variance, $\mathrm{Var}(J_{ij} \mid Z) = (\lambda - \nu)^2 r_{ij}(1 - r_{ij}) \le \frac{1}{4}(3\Delta/q_0)^2 \le 3\Delta^2/q_0^2$. □

**Lemma D.14** (exponential moments of the field)**.** *Let $\delta \le c_1$ and $n \ge 16 s_0^2 c_1 / q_0^2$. Then for every $i$, every $c$ and every $|s| \le s_0$,*

$$\mathbb{E}_t\big[e^{s F_{ic}}\big] \le \exp\Big\{\frac{2|s|\delta}{q_0^2} + \frac{9 s^2 \delta}{4 q_0^2}\Big\}, \qquad \mathbb{E}_t\big[e^{s|F_{ic}|}\big] \le 2 \exp\Big\{\frac{2|s|\delta}{q_0^2} + \frac{9 s^2 \delta}{4 q_0^2}\Big\} =: \Lambda(s).$$

*Here $\Lambda$ depends only on $(q_0, s_0, c_1)$ and tends to 2 as $c_1 \to 0$. The same bounds hold, with the same constants, for $G_{ic} = F_{ic} - \bar{F}_i$.*

*Proof.* Fix $Z$. Then $F_{ic} = \sum_{j \ne i} \mathbf{1}\{Z_j = c\} J_{ij}$ is a sum of at most $n$ conditionally independent terms. By Lemma D.13, $|\mathbb{E}[F_{ic} \mid Z]| \le 2n\Delta^2/q_0^2 = 2\delta/q_0^2$ and $v := \mathrm{Var}(F_{ic} \mid Z) \le 3\delta/q_0^2$, and each centred term is bounded by $b := 4\Delta/q_0$. Bernstein's bound on the moment generating function gives, for the centred sum $W$, $\mathbb{E}[e^{sW} \mid Z] \le \exp\{\frac{v s^2}{2(1 - b|s|/3)}\}$ when $|s| < 3/b$. From $\Delta = \sqrt{\delta/n} \le \sqrt{c_1/n}$ and $n \ge 16 s_0^2 c_1/q_0^2$ we get $b|s| \le \frac{4 s_0}{q_0}\sqrt{c_1/n} \le 1$, hence $1 - b|s|/3 \ge 2/3$ and $\mathbb{E}[e^{sW} \mid Z] \le \exp\{3 v s^2/4\} \le \exp\{9 s^2 \delta/(4 q_0^2)\}$. Multiplying by $e^{s\mathbb{E}[F_{ic}|Z]}$ and taking the expectation over $Z$ gives the first bound; the second follows from $e^{s|x|} \le e^{sx} + e^{-sx}$. Finally $G_{ic} = \sum_{j \ne i} (\mathbf{1}\{Z_j = c\} - \theta_{Z_j}) J_{ij}$ is again a sum of conditionally independent terms with coefficients of absolute value at most 1, so the same proof applies verbatim. □

Recall $\tilde{S}_1 = \sum_i \Phi(F_i)$ with $\Phi(F) = \sum_b \dot{\theta}_b e^{F_b} / \sum_b \theta_b e^{F_b}$.

**Lemma D.15** (exponential-family representation)**.** *Write $\mathcal{K}(F) := \log \sum_b \theta_b e^{F_b}$. Then $\pi_c(F) = \partial_c \mathcal{K}(F)$,*

$$\Phi(F) = \mathbb{E}_{\pi(F)}[\psi] = \sum_c \psi_c \, \partial_c \mathcal{K}(F),$$

*and for every $m \ge 1$ and all class indices $c^{(1)}, \dots, c^{(m)} \in \{1, 2, 3\}$, $\partial_{c^{(1)}} \cdots \partial_{c^{(m)}} \Phi(F) = \kappa_{\pi(F)}(\psi;\ \mathbf{1}_{c^{(1)}}, \dots, \mathbf{1}_{c^{(m)}})$, the joint cumulant of $\psi$ and the $m$ indicators under $\pi(F)$.*

*Proof.* The first identity is a rewriting of the definition: $\sum_c \psi_c \pi_c = \sum_c \frac{\dot{\theta}_c}{\theta_c} \cdot \frac{\theta_c e^{F_c}}{\sum_b \theta_b e^{F_b}} = \Phi(F)$. Since $\mathcal{K}$ is the cumulant generating function of this three-point exponential family, $\partial_{c^{(0)}} \partial_{c^{(1)}} \cdots \partial_{c^{(m)}} \mathcal{K} = \kappa_\pi(\mathbf{1}_{c^{(0)}}, \dots, \mathbf{1}_{c^{(m)}})$; differentiating $\Phi = \sum_{c^{(0)}} \psi_{c^{(0)}} \partial_{c^{(0)}} \mathcal{K}$ and using the linearity of a cumulant in its first argument gives the claim. □

**Lemma D.16** (derivative bounds)**.** *The function $\Phi$ depends on $F$ only through $G := F - \sum_b \theta_b F_b$. Write $g := \max_c |G_c|$. Then*

$$|\Phi(F)| \le 2\|\dot{\theta}\|_1 \, g \, e^{2g}, \qquad \big|\partial_{c^{(1)}} \cdots \partial_{c^{(m)}} \Phi(F)\big| \le C_m \|\dot{\theta}\|_1 (1 + g) e^{2g}, \quad C_m := 2 \, \mathrm{Bell}(m+1) \, m!.$$

*Proof.* Shift invariance: adding the same constant to all three components of $F$ does not change $\pi(F)$, and $G$ is exactly $F$ minus the constant $\sum_b \theta_b F_b$, so $\pi(F) = \pi(G)$ and $\Phi(F) = \Phi(G)$. From $\sum_b \theta_b = 1$ and $|G_b| \le g$ we get $e^{-g} \le \sum_b \theta_b e^{G_b} \le e^g$, hence $e^{-2g} \le \rho_c \le e^{2g}$.

(i) Since $\sum_c \dot{\theta}_c = 0$ we have $\Phi = \sum_c (\rho_c - 1)\dot{\theta}_c$; and $\rho_c - 1 = \sum_b \theta_b (e^{G_c} - e^{G_b}) / \sum_b \theta_b e^{G_b}$ with $|e^x - e^y| \le |x - y| e^{\max(x,y)} \le 2g e^g$ and denominator at least $e^{-g}$, so $|\rho_c - 1| \le 2g e^{2g}$.

(ii) By Lemma D.15 and the moment-to-cumulant partition formula,

$$\kappa_\pi(\psi; \mathbf{1}_{c^{(1)}}, \dots, \mathbf{1}_{c^{(m)}}) = \sum_{\sigma \vdash \{0,1,\dots,m\}} (-1)^{|\sigma|-1} (|\sigma| - 1)! \prod_{B \in \sigma} \mathbb{E}_\pi\Big[\prod_{l \in B} Y_l\Big],$$
$$Y_0 := \psi, \ Y_l := \mathbf{1}_{c^{(l)}}.$$

Each block moment is of one of three kinds: a block not containing 0 gives $\pi_c \mathbf{1}\{\text{all indices of the block equal } c\} \le 1$; a block containing 0 and different from $\{0\}$ gives $\pi_c\psi_c \mathbf{1}\{\cdots\} = \rho_c\dot\theta_c \mathbf{1}\{\cdots\}$, of absolute value at most $e^{2g}\|\dot\theta\|_\infty$; and $B = \{0\}$ gives $\Phi$, of absolute value at most $2\|\dot\theta\|_1 g e^{2g}$. The function $\psi$ is itself unbounded, since $\psi_3 = 2/t \to \infty$ as $t \to 0$, but in every partition it appears only in the form $\pi_c\psi_c = \rho_c\dot\theta_c$ or $\mathbb{E}_\pi\psi = \Phi$, both of which are controlled by $\|\dot\theta\|_1$ times an exponential weight, so the singularity at $t \to 0$ disappears altogether and the bound is uniform over $t \in (0, 1]$. Each product contains exactly one factor carrying $\psi$ and at most $m$ factors bounded by 1; the number of partitions is $\mathrm{Bell}(m+1)$ and $(|\sigma|-1)! \le m!$. □

**Lemma D.17** (explicit first and second derivatives)**.**

$$\partial_c\Phi(F) = \pi_c(\psi_c - \Phi) = \rho_c\dot\theta_c - \pi_c\Phi, \qquad \partial_d\partial_c\Phi(F) = \pi_c(\mathbf{1}\{c=d\} - \pi_d)(\psi_c - \Phi) - \pi_c\pi_d(\psi_d - \Phi);$$

*moreover* $|\partial_c\Phi(F)| \le e^{2g}|\dot\theta_c| + 2\theta_c\|\dot\theta\|_1 g e^{4g}$*, and* $\Phi(0) = 0$*,* $\nabla\Phi(0) = \dot\theta$*, while* $\nabla^2\Phi(0) =: H_\Phi$ *satisfies* $(H_\Phi)_{cd} = \mathbf{1}\{c=d\}\dot\theta_c - \dot\theta_c\theta_d - \theta_c\dot\theta_d$*.*

*Proof.* For $m = 1$, Lemma D.15 gives $\partial_c\Phi = \mathrm{Cov}_\pi(\psi, \mathbf{1}_c) = \pi_c\psi_c - \pi_c\Phi$, and $\pi_c\psi_c = \rho_c\dot\theta_c$. Differentiating once more in $d$ and using $\partial_d\pi_c = \pi_c(\mathbf{1}\{c=d\} - \pi_d)$ and $\partial_d\Phi = \pi_d(\psi_d - \Phi)$ gives the second identity. At $F = 0$ one has $\pi = \theta$, $\Phi = 0$ and $\theta_c\psi_c = \dot\theta_c$; substituting into the two identities gives $\nabla\Phi(0)$ and $H_\Phi$. □

**Scales along the path.** For $\varepsilon \le 1/12$ and $t \in (0, 1]$,

$$\|\dot\theta\|_1 \le 4\varepsilon, \qquad \|\dot\theta\|_\infty \le 2\varepsilon, \qquad |\dot\theta_1| \le 4\varepsilon^2, \qquad \theta_2 + \theta_3 = e \le \varepsilon, \tag{38}$$

and, writing $\dot\theta = (r, -r-h, h)$ with $r, h \ge 0$, the identity $\sum_c \theta_c\dot\theta_c = 0$ gives $\theta_1 r + \theta_3 h = \theta_2(r+h)$, hence

$$\sum_c |w_c| = \sum_c |\theta_c\dot\theta_c| = 2\theta_2(r+h) = \theta_2\|\dot\theta\|_1 \le \varepsilon\|\dot\theta\|_1 \le 4\varepsilon^2. \tag{39}$$

So every $\|\dot\theta\|_1$ in Lemma D.16 and Lemma D.17 may be replaced by $4\varepsilon$. The edge influence and the label influence are both linear in $\dot\theta$ and carry one such factor only; the $\varepsilon^4 = (\varepsilon^2)^2$ of the target needs one further $\varepsilon$, which comes from the weight of the minority classes and from the cancellation of Part 5.

**Numerical check.** Lemma D.12 and Lemma D.13 hold cell by cell on $q_0 \in \{0.1, 0.3, 0.5\}$, $q \in \{q_0, 1/2, 1-q_0\}$, $n \in \{200, 800, 3200\}$ and $\delta \in \{0.02, 0.06, 0.15\}$, the worst ratio of actual value to bound being 0.9975 for Lemma D.12, nearly an equality at $\lambda \le \Delta/q_0$, and 0.2504 and 0.3340 for the mean and the variance of Lemma D.13. Lemma D.14 was checked on the 81 cells given by $q = q_0 = 0.3$, the same $(n, \delta)$ grid, $t \in \{0.25, 0.5, 1\}$ and $s \in \{0.5, 1, 2\}$ with 4000 samples per cell: the largest ratio of the empirical $\mathbb{E}e^{s|F_{1c}|}$ to the bound is 0.4006. The three path scales of (38), checked at $\varepsilon \in \{1/12, 0.05, 0.02, 0.005, 0.001\}$ and 400 values of $t$, have worst ratios 0.4999, 0.4999 and 0.1098.

### D.6 Part 4: the conditional toolkit

Every estimate of Parts 4 to 7 holds the labels fixed first: given $Z$ all the randomness comes from mutually independent edges and the bounds of Part 3 are uniform in $Z$, so each quantity is first estimated under $\mathbb{P}(\cdot \mid \mathcal{L})$, with the bound written as an explicit function of the label counts, and only then averaged over $Z$.

**Assumption 2** (admissible weights)**.** Given $\mathcal{L}$, the weights $\{J_{ij}\}_{i<j}$ are mutually independent, take values in $[\nu, \lambda]$, and satisfy $|\mathbb{E}[J_{ij} \mid \mathcal{L}]| \le 2\Delta^2/q_0^2$. One exceptional edge $\{a, b\}$ is allowed, whose weight is a fixed number $u \in [\nu, \lambda]$; and one vertex $v$ is allowed to carry a fractional label $E_v \in \Delta_3$, the remaining ones having $E_k = \mathbf{e}_{Z_k}$, the field still being $F_i = \sum_{j \ne i} J_{ij}E_j$.

The lemmas of this part are stated under Assumption 2, which is slightly weaker than the model, so that the same estimates can be used again in Part 7. The model itself satisfies Assumption 2 by Lemma D.12 and Lemma D.13, and so does the system obtained in Lemma D.24 by replacing $J_{ab}$ with $u$. Under Assumption 2 one has $|J_{ij}| \le 2\Delta/q_0$ and $\mathrm{Var}(J_{ij} \mid \mathcal{L}) \le (\lambda - \nu)^2/4 \le 9\Delta^2/(4q_0^2)$.

**Lemma D.18** (exponential moments, uniformly in the labels)**.** *Under Assumption 2, for every $i$ and every $0 \le s \le s_0$,*

$$\mathbb{E}\big[e^{s\gamma_i} \mid \mathcal{L}\big] \;\le\; \Lambda_* := 6\exp\Big\{\big(2s_0 + \tfrac{9}{4}s_0^2\big)\frac{c_1}{q_0^2} + 1\Big\},$$

*and the same holds after deleting any number of terms from $F_i$. In particular $\|\gamma_i\|_{\mathsf{p},Z} \le C$; and by Lemma D.19 with $\xi \equiv 1$, which does not use the present lemma, also $\|F_{ic}\|_{\mathsf{p},Z} \le C\sqrt{\delta}$.*

*Proof.* The proof of Lemma D.14 was carried out with the labels held fixed and used only the conditional independence, the bound $4\Delta/q_0$ on a single term, and the bounds on the conditional mean and the conditional variance. Here the coefficients are $E_{jc} \in [0,1]$, so that proof applies verbatim to the non-exceptional edges under Assumption 2 and gives $\mathbb{E}[e^{s|F_{ic}|} \mid \mathcal{L}] \le \Lambda(s)$. The exceptional edge moves $F_{ic}$ by at most $2\Delta/q_0$, and $2s_0\Delta/q_0 \le 1/2$ when $n \ge n_0$. Conclude with $e^{s\gamma_i} \le \sum_c e^{s|F_{ic}|}$. □

**Lemma D.19** (row sums)**.** *Under Assumption 2 fix a vertex $a$ and let $\mathcal{G}_{-a}$ be the $\sigma$-field generated by $\mathcal{L}$ and $\{J_{ij} : i,j \ne a\}$. Let $S \subset [n] \setminus \{a\}$ be determined by the labels and let $\xi_k$, $k \in S$, be $\mathcal{G}_{-a}$-measurable, measurability not being required at the other endpoint $k = b$ of the exceptional edge. Then for $\mathsf{p} \in [2, 32]$,*

$$\Big\| \sum_{k\in S} J_{ak}\,\xi_k \Big\|_{\mathsf{p},Z} \;\le\; C\,\Delta\sqrt{|S|}\,\max_{k\in S}\|\xi_k\|_{\mathsf{p},Z}.$$

*For vector-valued $\xi_k$ the bound holds componentwise.*

*Proof.* Write $m_k := \mathbb{E}[J_{ak} \mid \mathcal{L}]$ and $\bar{J}_{ak} := J_{ak} - m_k$. Given $\mathcal{L}$ the weights of row $a$ are independent of the others, so given $\mathcal{G}_{-a}$ the variables $\{\bar{J}_{ak}\xi_k\}_{k\in S\setminus\{b\}}$ are independent and centred, with the $\xi_k$ acting as constants. Rosenthal's inequality gives

$$\mathbb{E}\Big[\Big|\sum_k \bar{J}_{ak}\xi_k\Big|^{\mathsf{p}} \,\Big|\, \mathcal{G}_{-a}\Big] \le C\Big[\Big(\sum_k \xi_k^2 \mathrm{Var} J_{ak}\Big)^{\mathsf{p}/2} + \sum_k |\xi_k|^{\mathsf{p}}\,\mathbb{E}|\bar{J}_{ak}|^{\mathsf{p}}\Big] \le C\Delta^{\mathsf{p}}\Big(\sum_k \xi_k^2\Big)^{\mathsf{p}/2},$$

the last step using $|\bar{J}_{ak}| \le 4\Delta/q_0$ and the embedding $\|\cdot\|_{\ell^{\mathsf{p}}} \le \|\cdot\|_{\ell^2}$. Minkowski's inequality in $L^{\mathsf{p}/2}(\cdot \mid \mathcal{L})$ then gives $\|\sum_k \xi_k^2\|_{\mathsf{p}/2,Z} \le \sum_k \|\xi_k\|_{\mathsf{p},Z}^2 \le |S|\max_k \|\xi_k\|_{\mathsf{p},Z}^2$. For the mean part, $|\sum_k m_k\xi_k| \le (2\Delta^2/q_0^2)\sum_k |\xi_k|$, whose norm is at most $(2\Delta^2/q_0^2)|S|\max_k\|\xi_k\|_{\mathsf{p},Z}$, and $\Delta^2|S| = \Delta\sqrt{|S|}\cdot\Delta\sqrt{|S|} \le \Delta\sqrt{|S|}\sqrt{\delta}$. The exceptional term $|u\,\xi_b| \le (2\Delta/q_0)|\xi_b|$ is absorbed directly. □

**Lemma D.20** (posterior bounds)**.** *For an arbitrary field $F$ with $\gamma := \max_c |F_c|$, $\pi = \pi(F)$, $\rho = \pi/\theta$ and $V = \mathrm{diag}(\pi) - \pi\pi^\top$, and for arbitrary $z \in \Delta_3$ and $x \in \{1,2,3\}$:*

*(i)* $e^{-2\gamma} \le \rho_c \le e^{2\gamma}$, $|\rho_c - 1| \le 2\gamma e^{2\gamma}$, $\pi_c \le \theta_c e^{2\gamma}$, $1 - \pi_1 \le \varepsilon e^{2\gamma}$;

*(ii)* $|Vz|_1 \le 2(1-\pi_1) \le 2\varepsilon e^{2\gamma}$ *and* $|\partial_x(Vz)|_1 \le 3(1-\pi_1) \le 3\varepsilon e^{2\gamma}$;

*(iii)* $|\partial_x \rho_c| \le e^{2\gamma}$ *and* $|\partial_x \pi_c| \le \min\{\theta_c, \varepsilon\}e^{2\gamma}$.

*Proof.* (i) From $e^{-\gamma} \le \sum_b \theta_b e^{F_b} \le e^{\gamma}$ we get the two bounds on $\rho_c$; $\rho_c - 1 = \sum_b \theta_b(e^{F_c} - e^{F_b})/\sum_b \theta_b e^{F_b}$ with $|e^{F_c} - e^{F_b}| \le 2\gamma e^{\gamma}$; and $1 - \pi_1 = \pi_2 + \pi_3 \le (\theta_2 + \theta_3)e^{2\gamma}$, where $\theta_2 + \theta_3 = e \le \varepsilon$ along the path.

(ii) We have $(Vz)_c = \pi_c(z_c - \langle \pi, z\rangle)$; the two terms with $c \in \{2,3\}$ add up to at most $\pi_2 + \pi_3$, and for $c = 1$, $|z_1 - \langle \pi, z\rangle| = |\sum_b \pi_b(z_1 - z_b)| \le 1 - \pi_1$. Differentiating, $\partial_x(Vz)_c = \pi_c(\mathbf{1}\{c = x\} - \pi_x)(z_c - \langle \pi, z\rangle) - \pi_c\pi_x(z_x - \langle \pi, z\rangle)$; summing the first part over $c$ again gives at most $2(1-\pi_1)$, and the second part equals $\pi_x|z_x - \langle \pi, z\rangle| \le 1 - \pi_1$, using $\pi_x \le 1 - \pi_1$ when $x \ne 1$ and the estimate above when $x = 1$.

(iii) Here $\partial_x \rho_c = \rho_c(\mathbf{1}\{c = x\} - \pi_x)$ and $\partial_x \pi_c = \pi_c(\mathbf{1}\{c = x\} - \pi_x)$; for $c = 1$ use $\pi_1|\mathbf{1}\{x = 1\} - \pi_x| \le 1 - \pi_1$. □

**Lemma D.21** (the pairing bound for $\Phi$)**.** *One has $|\Phi(F)| \le e^{2\gamma}\sum_c |\dot{\theta}_c|\,|F_c|$, and for $1 \le m \le 4$ and arbitrary class indices, $|\partial_{c^{(1)}} \cdots \partial_{c^{(m)}}\Phi(F)| \le C\varepsilon e^{6\gamma}$.*

*Proof.* From $\sum_c \dot\theta_c = 0$ we get $\Phi(F) = \sum_c \dot\theta_c (e^{F_c} - 1)/\sum_b \theta_b e^{F_b}$, where the denominator is at least $e^{-\gamma}$ and $|e^{F_c} - 1| \le |F_c| e^{\gamma}$. The second bound is Lemma D.16 combined with $\|\dot\theta\|_1 \le 4\varepsilon$, $g \le 2\gamma$ and $(1+g)e^{2g} \le e^{3g}$. □

In the first bound, each $\dot\theta_c$ is multiplied by the field $F_c$ of its own class only, and the fields of the minority classes are small, namely $\|F_{ic}\|_{\mathrm{p},Z} \le C\Delta\sqrt{N_c}$ by Lemma D.19.

**Lemma D.22** (leave-one-out)**.** *Under Assumption 2, let $\varphi$ be a smooth function of the field, let $k \ne a$ and put $F_k^{(-a)} := F_k - J_{ka}E_a$, which is $\mathcal{G}_{-a}$-measurable. Then*

$$\varphi(F_k) = \varphi\big(F_k^{(-a)}\big) + J_{ka}\int_0^1 \big\langle \nabla\varphi\big(F_k^{(-a)} + rJ_{ka}E_a\big),\ E_a\big\rangle\,\mathrm{d}r,$$

*so that for the $\mathcal{G}_{-a}$-measurable $\xi_k^\circ := \varphi(F_k^{(-a)})$,*

$$\Big|\sum_{k\in S} J_{ak}\varphi(F_k) - \sum_{k\in S} J_{ak}\xi_k^\circ\Big| \le \frac{4\Delta^2}{q_0^2}\sum_{k\in S}\sup_{r\in[0,1]}\big|\nabla\varphi\big(F_k^{(-a)} + rJ_{ka}E_a\big)\big|_\infty,$$

*where each field inside a supremum differs from $F_k$ by at most $2\Delta/q_0$ in every component, so that any weight of the form $e^{s\gamma}$ may be replaced by $e^{1/2}e^{s\gamma_k}$.*

*Proof.* The first identity is the fundamental theorem of calculus along the segment, with $|E_a|_1 = 1$; the second uses $J_{ak}^2 \le 4\Delta^2/q_0^2$. □

**Lemma D.23** (label statistics)**.** *Let $Z_1,\dots,Z_n$ be i.i.d. from $\theta(t)$, let $\varepsilon \le 1/12$ and $n\varepsilon \ge 1$. Put $\mathcal{D} := \sum_c |\dot\theta_c|\sqrt{\nu_c^+}$, $\mathcal{D}_2 := \sum_c |\dot\theta_c|\nu_c^+$ and $\mathcal{M} := \sqrt{\nu_{\min}^+} + \varepsilon$. Then for every fixed vertex $b$ and every minority class $y \in \{2,3\}$,*

$$\mathbb{E}\mathcal{D}^2 \le C\varepsilon^3,\quad \mathbb{E}\mathcal{D}_2^2 \le C\varepsilon^4,\quad \mathbb{E}\mathcal{M}^2 \le C\varepsilon,\quad \mathbb{E}[\mathcal{D}^2\mathcal{M}^2] \le C\varepsilon^4,\quad \mathbb{E}(\nu_{\min}^+)^2 \le C\varepsilon^2,$$
$$\mathbb{E}\big[\mathbf{1}\{Z_b = y\}\mathcal{D}^2\big] \le C\varepsilon^4,\qquad \dot\theta_y^2\,\mathbb{E}\big[\mathbf{1}\{Z_b = y\}\mathcal{M}^2\big] \le C\varepsilon^4,\qquad \dot\theta_y^2\theta_y \le C\varepsilon^3.$$

*Proof.* Here $N_c \sim \mathrm{Bin}(n,\theta_c)$, so $\mathbb{E}\nu_c = \theta_c$, $\mathbb{E}\nu_c^2 \le \theta_c^2 + \theta_c/n$, $\mathbb{E}[\nu_c\nu_{c'}] \le \theta_c\theta_{c'}$ for $c \ne c'$, and $\mathbb{E}[\mathbf{1}\{Z_b = y\}\nu_c] \le \theta_y(\theta_c + 1/n)$; and $n\varepsilon \ge 1$ gives $2/n \le 2\varepsilon$, so replacing $\nu_c$ by $\nu_c^+$ changes only the constants. The path scales give $|\dot\theta_1| \le 4\varepsilon^2$, $|\dot\theta_c| \le 2\varepsilon$ and $\theta_2 + \theta_3 \le \varepsilon$, whence $\sum_c \dot\theta_c^2\theta_c \le C\varepsilon^3$ and $\sum_c \dot\theta_c^2(\theta_c^2 + \theta_c/n) \le C\varepsilon^4 + C\varepsilon^3/n \le C\varepsilon^4$, the last step using $n\varepsilon \ge 1$. Substituting into $\mathcal{D}^2 \le 3\sum_c \dot\theta_c^2\nu_c^+$, $\mathcal{D}_2^2 \le 3\sum_c \dot\theta_c^2(\nu_c^+)^2$ and $\mathcal{M}^2 \le 2\nu_{\min}^+ + 2\varepsilon^2$ term by term gives each of the bounds; for instance, for $c \in \{2,3\}$, $\dot\theta_c^2\,\mathbb{E}[\nu_c^+\nu_{\min}^+] \le 4\varepsilon^2 \cdot C(\theta_c\varepsilon + \varepsilon/n + 1/n^2) \le C\varepsilon^4$, while the term with $c = 1$ follows directly from $\dot\theta_1^2 \le 16\varepsilon^4$ and $\mathbb{E}\mathcal{M}^2 \le C\varepsilon$. □

### D.7 Part 5: Exact decomposition of the edge influence

Treat $J_{ab}$ as a free variable $u$, holding the other $J_{ij}$ and all labels fixed, and write $\tilde S_{\mathrm{RB}}(u)$ for the resulting value; the true value is $u = J_{ab} \in \{\nu,\lambda\}$.

**Lemma D.24** (reduction to a derivative)**.**

$$\mathbb{E}_t\big[(\tilde S_{\mathrm{RB}} - \tilde S_{\mathrm{RB}}^{(ab)})^2\big] \ \le\ \frac{9\Delta^2}{q_0^2}\sup_{u\in[\nu,\lambda]}\mathbb{E}_t\big[(\partial_u\tilde S_{\mathrm{RB}}(u))^2\big].$$

*Proof.* Resampling $U_{ab}$ changes only $A_{ab}$ and therefore only moves $u$ between $\nu$ and $\lambda$, all other inputs being fixed. The map $u \mapsto \tilde S_{\mathrm{RB}}(u)$ is continuously differentiable on $[\nu,\lambda]$, being a rational function of finitely many $e^u$ with an everywhere positive denominator, so $|\tilde S_{\mathrm{RB}} - \tilde S_{\mathrm{RB}}^{(ab)}| \le \int_\nu^\lambda |\partial_u\tilde S_{\mathrm{RB}}(u)|\mathrm{d}u$. By Cauchy–Schwarz on $[\nu,\lambda]$ the square of the right-hand side is at most $(\lambda-\nu)\int_\nu^\lambda(\partial_u\tilde S_{\mathrm{RB}}(u))^2\mathrm{d}u$. Take expectations, the integrand being non-negative so that Tonelli applies, and use $\lambda - \nu \le 3\Delta/q_0$ from Lemma D.12. □

The expectation is taken under the true law of $(Z,A)$, while $\partial_u$ acts on the function obtained after replacing $J_{ab}$ by $u$; the law itself is not modified. Since $|u| \le \max(|\nu|,\lambda) \le 2\Delta/q_0$, all the field bounds of Parts 3 and 4 remain valid for the modified $F_a$ and $F_b$.

**Notation.** Put $F_i^{k\to c} := F_i + J_{ik}(\mathbf{e}_c - E_k)$, $\Psi_{ik} := \sum_c \pi_{kc}\Phi(F_i^{k\to c}) = \mathbb{E}_t[\Phi(F_i) \mid A, Z_{-k}]$ and $\Phi_i^{k\to\cdot} := (\Phi(F_i^{k\to c}))_{c=1,2,3}$. Put further $f_a := \sum_{j\neq a,b} J_{aj}E_j$ and $f_b := \sum_{j\neq a,b} J_{bj}E_j$, so that $F_a = f_a + u\mathbf{e}_{Z_b}$ and $F_b = f_b + u\mathbf{e}_{Z_a}$, with $f_a$ and $f_b$ free of $u$.

**Lemma D.25** (blocks and derivative)**.** *The field $F_k$ does not depend on $Z_k$, so*

$$\tilde{S}_{\mathrm{RB}} = \sum_i R_i, \qquad R_i := \Phi(F_i) + \sum_{k\neq i}\big(\Psi_{ik} - \Phi(F_i)\big) = \sum_{k\neq i}\Psi_{ik} - (n-2)\Phi(F_i), \tag{40}$$

*and $\partial_u\pi_a = V_a\mathbf{e}_{Z_b}$, $\partial_u\pi_b = V_b\mathbf{e}_{Z_a}$, while the other $\pi_k$ do not depend on $u$. Grouping by the place where $u$ occurs,*

$$\partial_u\tilde{S}_{\mathrm{RB}} = (\mathrm{I}) + (\mathrm{II})_a + (\mathrm{II})_b + (\mathrm{III}),$$

$$(\mathrm{I}) := \sum_{i\notin\{a,b\}}\Big[\big\langle V_a\mathbf{e}_{Z_b},\ \Phi_i^{a\to\cdot}\big\rangle + \big\langle V_b\mathbf{e}_{Z_a},\ \Phi_i^{b\to\cdot}\big\rangle\Big],$$

$$(\mathrm{II})_a := \sum_{k\notin\{a,b\}}\Big\langle \sum_c \pi_{kc}\nabla\Phi\big(F_a^{k\to c}\big) - \nabla\Phi(F_a),\ \mathbf{e}_{Z_b}\Big\rangle, \qquad (\mathrm{II})_b := (a\leftrightarrow b),$$

$$(\mathrm{III}) := \underbrace{\big\langle V_b\mathbf{e}_{Z_a}, \Phi^a(u)\big\rangle + \sum_c \pi_{bc}\partial_c\Phi(f_a + u\mathbf{e}_c)}_{=:\ (\mathrm{III})_a\ =\ \partial_u\Psi_{ab}} + \underbrace{\big\langle V_a\mathbf{e}_{Z_b}, \Phi^b(u)\big\rangle + \sum_c \pi_{ac}\partial_c\Phi(f_b + u\mathbf{e}_c)}_{=:\ (\mathrm{III})_b\ =\ \partial_u\Psi_{ba}},$$

*where $\Phi^a(u) := (\Phi(f_a + u\mathbf{e}_c))_c$ and $\Phi^b(u) := (\Phi(f_b + u\mathbf{e}_c))_c$. Regrouped by rows,*

$$\partial_u\tilde{S}_{\mathrm{RB}} = (\mathrm{I}) + B_a + B_b, \quad (\mathrm{I}) = \sum_{i\notin\{a,b\}}\partial_u R_i,$$

$$B_a := \partial_u R_a = (\mathrm{II})_a + (\mathrm{III})_a, \quad B_b := \partial_u R_b = (\mathrm{II})_b + (\mathrm{III})_b.$$

*Proof.* For the first identity, $\tilde{S}_{\mathrm{RB}} = \sum_k \mathcal{P}_k\tilde{S}_1 - (n-1)\tilde{S}_1$; the operator $\mathcal{P}_k$ takes the conditional expectation over $Z_k$ only, and $F_k$ does not involve $Z_k$, so $\mathcal{P}_k\Phi(F_k) = \Phi(F_k)$ and $\mathcal{P}_k\Phi(F_i) = \Psi_{ik}$ for $i \neq k$. Collecting the coefficients at a fixed $i$: the term $\Phi(F_i)$ occurs once in $\sum_k \mathcal{P}_k\tilde{S}_1$, at $k = i$, and $-(n-1)$ times in $-(n-1)\tilde{S}_1$, so that its coefficient is $-(n-2)$.

For $\partial_u\pi_a$: $\pi_{ac} = \theta_c e^{F_{ac}}/\sum_b \theta_b e^{F_{ab}}$ and $\partial_u F_{ac} = (\mathbf{e}_{Z_b})_c$, so the first derivative of the exponential family gives $\partial_u\pi_{ac} = \pi_{ac}((\mathbf{e}_{Z_b})_c - \pi_{a,Z_b})$, that is $\partial_u\pi_a = V_a\mathbf{e}_{Z_b}$.

For the grouping: $u$ occurs only in $F_a$ and $F_b$, through $\mathbf{e}_{Z_b}$ and $\mathbf{e}_{Z_a}$, in $\pi_a$ and $\pi_b$, and in $J_{ab}$ itself. Checking $\Psi_{ik}$ and $\Phi(F_i)$ case by case:

- $i, k \notin \{a, b\}$: $\Psi_{ik}$ is free of $u$; and for $i \notin \{a, b\}$ so is $\Phi(F_i)$.
- $i \notin \{a, b\}$, $k \in \{a, b\}$: only $\pi_k$ carries $u$, neither $F_i^{k\to c}$ nor $J_{ik}$ does, which gives (I).
- $i = a$, $k \notin \{a, b\}$: only $F_a^{k\to c} = F_a + J_{ak}(\mathbf{e}_c - \mathbf{e}_{Z_k})$ carries $u$, and $\partial_u$ acts as the directional derivative $\langle\nabla\Phi, \mathbf{e}_{Z_b}\rangle$, since $b \neq a$ and $b \neq k$ so that $u\mathbf{e}_{Z_b}$ stays inside $F_a^{k\to c}$. These $n-2$ terms combine with $-(n-2)\partial_u\Phi(F_a) = -(n-2)\langle\nabla\Phi(F_a), \mathbf{e}_{Z_b}\rangle$ into $(\mathrm{II})_a$.
- $i = b$, $k \notin \{a, b\}$: symmetric, giving $(\mathrm{II})_b$.
- $i = a$, $k = b$: $\Psi_{ab} = \sum_c \pi_{bc}\Phi(f_a + u\mathbf{e}_c)$, where $u$ occurs both in $\pi_b$ and in the explicit $u\mathbf{e}_c$, and differentiating gives $(\mathrm{III})_a$; the case $i = b$, $k = a$ is symmetric and gives $(\mathrm{III})_b$.

Regrouping by rows merely collects these terms according to the $R_i$ they belong to. □

**The estimates have to be made after grouping by rows.** The block $(\mathrm{II})_a$ pairs the field of vertex $a$ with the bare label $\mathbf{e}_{Z_b}$ of vertex $b$, while $(\mathrm{III})_a$ pairs the same field with the conditional law $\pi_b$ of $b$. Each of the two contains a term of order $\varepsilon^{3/2}\sqrt{\delta}$, namely the coefficient $\dot{\theta}_2$ times the minority-class field $F_{a2}$, whose conditional variance given the labels is $\#\{j \neq a : Z_j = 2\}(\lambda - \nu)^2 r(1-r)$ with $r \in \{q, p\}$, of order $\delta\theta_2/(q(1-q))$ after averaging over the labels. The two cancel inside $B_a = (\mathrm{II})_a + (\mathrm{III})_a$. Taking the norm of each of the two blocks separately and then adding leaves a factor $\varepsilon^{-1}$; Remark D.36 gives the numbers. Lemma D.26 therefore expands $R_a$ as a whole and never passes through that split.

**Lemma D.26** (second-order integral remainder and its derivative)**.** *For $i \neq k$ put $d_{kc} := \mathbf{e}_c - E_k$ and*

$$q_{ik} := \sum_c \pi_{kc} \int_0^1 (1-r)\, \nabla^2\Phi\big(F_i + rJ_{ik}d_{kc}\big)[d_{kc}, d_{kc}]\, \mathrm{d}r,$$
$$M_i := \sum_{k\neq i} J_{ik}(E_k - \pi_k), \quad Q_i := \sum_{k\neq i} J_{ik}^2 q_{ik}.$$

*Then $\sum_c M_{ic} = 0$, and pointwise*

$$R_i \;=\; \underbrace{\Phi(F_i) - \big\langle \nabla\Phi(F_i),\, M_i\big\rangle}_{=:\,\Xi_i} + Q_i. \tag{41}$$

*Taking $i = a$, $u = J_{ab}$, $x := Z_a$, $y := Z_b$ and writing $\Phi_a := \Phi(F_a)$, one has $\partial_u \Xi_a = T_1 - T_2 + T_3$ with*

$$T_1 := \big\langle \nabla\Phi(F_a), \pi_b\big\rangle = \sum_c w_c(\rho_{ac}\rho_{bc} - 1) - \Phi_a\langle \pi_a, \pi_b\rangle,$$

$$T_2 := \big\langle \nabla^2\Phi(F_a)\mathbf{e}_y,\; M_a\big\rangle = \pi_{ay}(\psi_y - \Phi_a)\big(M_{ay} - \langle \pi_a, M_a\rangle\big) - \pi_{ay}\big\langle \pi_a \odot (\psi - \Phi_a),\; M_a\big\rangle,$$

$$T_3 := u\,\big\langle \nabla\Phi(F_a),\; V_b\mathbf{e}_x\big\rangle.$$

*Proof.* Taylor's formula with integral remainder gives $\Phi(F_i + J_{ik}d) = \Phi(F_i) + J_{ik}\langle \nabla\Phi(F_i), d\rangle + J_{ik}^2 \int_0^1 (1-r)\nabla^2\Phi(F_i + rJ_{ik}d)[d,d]\mathrm{d}r$. Summing over $c$ with the weights $\pi_{kc}$ and using $\sum_c \pi_{kc}d_{kc} = \pi_k - E_k$ gives $\Psi_{ik} - \Phi(F_i) = -J_{ik}\langle \nabla\Phi(F_i), E_k - \pi_k\rangle + J_{ik}^2 q_{ik}$; summing over $k \neq i$ and substituting into $R_i$ from Lemma D.25 gives (41). Also $\sum_c M_{ic} = 0$, because $E_k$ and $\pi_k$ are both probability vectors.

For the derivative: $\partial_u F_a = \mathbf{e}_y$; for $k \notin \{a, b\}$ the vector $\pi_k$ is free of $u$, while $\partial_u \pi_b = V_b\mathbf{e}_x$, so $\partial_u M_a = \partial_u[u(\mathbf{e}_y - \pi_b)] = (\mathbf{e}_y - \pi_b) - uV_b\mathbf{e}_x$. Hence

$$\partial_u \Xi_a = \langle \nabla\Phi(F_a), \mathbf{e}_y\rangle - \langle \nabla^2\Phi(F_a)\mathbf{e}_y, M_a\rangle - \langle \nabla\Phi(F_a), \mathbf{e}_y - \pi_b\rangle + u\langle \nabla\Phi(F_a), V_b\mathbf{e}_x\rangle = T_1 - T_2 + T_3.$$

The second form of $T_1$ follows from $\partial_c\Phi = \pi_c(\psi_c - \Phi)$ (Lemma D.17), from $\pi_{bc}\pi_{ac}\psi_c = w_c\rho_{ac}\rho_{bc}$ and from $\sum_c w_c = 0$. The second form of $T_2$ is the second-order formula of Lemma D.17 with $d = y$, paired with $M_a$. □

**Lemma D.27** (exact cancellation at the majority endpoint)**.** *Write $\chi(s) := e^s(s-1) + 1 = \int_0^s re^r \mathrm{d}r$ and $\mathcal{N}(F) := \sum_b \theta_b e^{F_b}$. For every field $F$,*

$$\sum_c \rho_c(F)\,\dot\theta_c\, F_c \;-\; \Phi(F) \;=\; \frac{1}{\mathcal{N}(F)} \sum_c \dot\theta_c\, \chi(F_c), \qquad 0 \le \chi(s) \le \tfrac{1}{2}s^2 e^{|s|}. \tag{42}$$

*Consequently, for $y = 1$ one has $T_1 - T_2 = A_1 + A_2 + A_3 + A_4 + A_5 + A_6$ with*

$$A_1 := \sum_c w_c(\rho_{ac}\rho_{bc} - 1), \qquad A_2 := \frac{\pi_{a1}}{\mathcal{N}(F_a)} \sum_c \dot\theta_c \chi(F_{ac}), \qquad A_3 := \Phi_a\big[\pi_{a1} - \langle \pi_a, \pi_b\rangle\big],$$

$$A_4 := -\pi_{a1} \sum_c w_c\rho_{ac} \sum_{k\neq a} J_{ak}\rho_{kc}, \qquad A_5 := -\rho_{a1}\dot\theta_1\big(M_{a1} - \langle \pi_a, M_a\rangle\big),$$
$$A_6 := \pi_{a1}\Phi_a\big[M_{a1} - 2\langle \pi_a, M_a\rangle\big];$$

*and for $y \in \{2, 3\}$ one has $T_1 - T_2 = A_1 + B_2 + B_3 + B_4 + B_5$ with*

$$B_2 := -\Phi_a\langle \pi_a, \pi_b\rangle, \quad B_3 := -\rho_{ay}\dot\theta_y\big(M_{ay} - \langle \pi_a, M_a\rangle\big), \quad B_4 := \pi_{ay}\Phi_a\big(M_{ay} - \langle \pi_a, M_a\rangle\big),$$

$$B_5 := \pi_{ay}\big\langle \rho_a \odot \dot\theta - \Phi_a\pi_a,\; M_a\big\rangle.$$

*Proof.* Since $\rho_c = e^{F_c}/\mathcal{N}$ and $\Phi = \sum_c \dot\theta_c \rho_c$, we get $\sum_c \dot\theta_c(\rho_c F_c - \rho_c) = \mathcal{N}^{-1}\sum_c \dot\theta_c e^{F_c}(F_c - 1) = \mathcal{N}^{-1}\sum_c \dot\theta_c(\chi(F_c) - 1)$, and the trailing $-\mathcal{N}^{-1}\sum_c \dot\theta_c$ vanishes because $\sum_c \dot\theta_c = 0$. The bound on $\chi$ is $|\int_0^s re^r \mathrm{d}r| \le e^{|s|}s^2/2$.

For $y = 1$: since $\pi_{a1}\psi_1 = \rho_{a1}\dot\theta_1$,

$$-T_2 = A_5 + \pi_{a1}\Phi_a\big(M_{a1} - \langle \pi_a, M_a\rangle\big) + \pi_{a1}\sum_c \rho_{ac}\dot\theta_c M_{ac} - \pi_{a1}\Phi_a\langle \pi_a, M_a\rangle.$$

Substituting $M_{ac} = F_{ac} - \sum_k J_{ak}\pi_{kc}$ and $\dot\theta_c\pi_{kc} = w_c\rho_{kc}$ into the third term turns it into $\pi_{a1}\sum_c \rho_{ac}\dot\theta_c F_{ac} + A_4$. Combining with the term $-\Phi_a\langle \pi_a, \pi_b\rangle$ of $T_1$,

$$\pi_{a1}\sum_c \rho_{ac}\dot\theta_c F_{ac} - \Phi_a\langle \pi_a, \pi_b\rangle = \pi_{a1}\Big[\sum_c \rho_{ac}\dot\theta_c F_{ac} - \Phi_a\Big] + \Phi_a\big[\pi_{a1} - \langle \pi_a, \pi_b\rangle\big] = A_2 + A_3,$$

by (42). For $y \in \{2, 3\}$ the displayed formulas only rename the terms of $T_1$ and $T_2$, using $\pi_{ay}\psi_y = \rho_{ay}\dot\theta_y$. □

**This step is the crux.** On the layer $y = 1$, where the weight is close to 1, the quantity $T_1$ contains $-\Phi_a$ and $T_2$ contains $\sum_c \rho_{ac}\dot\theta_c F_{ac}$, each of order $\varepsilon^{3/2}\sqrt{\delta}$, the coefficient $\dot\theta_2$ times a minority-class field, which exceeds the target by a factor $\varepsilon^{-1/2}$. Their difference is exactly $\mathcal{N}^{-1}\sum_c \dot\theta_c\chi(F_{ac})$, which is diagonal in the class index: each term is $\dot\theta_c$ times the square of the field of the same class, and no cross term such as $\dot\theta_2 F_{a2}F_{a1}$ appears. Hence there is no contribution of order $\delta/\varepsilon$.

**Numerical check.** (a) The block decomposition of Lemma D.25: computing $\tilde S_{\mathrm{RB}}(u)$ directly from its definition for $n \in \{8, 11, 14\}$, taking the conditional expectation over the three classes exactly for each $k$, and comparing with the sum of the blocks over the 72 cells given by $\varepsilon \in \{0.2, 0.05\}$, $t \in \{0.3, 0.7, 1\}$, $\delta \in \{0.05, 0.2\}$ and two seeds, the worst relative difference against a five-point central difference is $1.05 \times 10^{-7}$, the cells with the smallest derivative having reached the noise floor of double-precision differencing; redoing four of those cells at 40 digits of precision gives a worst relative difference of $5.5 \times 10^{-15}$. (b) Lemma D.26 and Lemma D.27: over the 144 configurations given by $n \in \{12, 40\}$, $\varepsilon \in \{0.08, 0.01\}$, $t \in \{0.3, 1\}$, $\delta \in \{0.05, 0.3\}$ and each of the nine label layers forced once, the worst relative difference between $T_1 - T_2 + T_3$ and a five-point central difference of $\Xi_a(u)$ is $4.0 \times 10^{-9}$, and between $T_1 - T_2$ and the sum of the $A$ and $B$ terms it is $1.1 \times 10^{-13}$.

### D.8 Part 6: termwise bounds for the edge influence, and ES-E

This part works under Assumption 2 with integer labels throughout, the exceptional edge being $\{a, b\}$ with weight $u \in [\nu, \lambda]$, and assumes $n\varepsilon \ge 1$. The quantities $\mathcal{D}$, $\mathcal{D}_2$ and $\mathcal{M}$ are the label statistics of Lemma D.23; the summand $2/n$ inside $\nu_c^+$ absorbs the one term added by the exceptional edge and the one term added by the fractional label of Part 7.

**Lemma D.28** (basic field bounds)**.** *For every vertex $i$, every class $c$ and every $\mathsf{p} \in [2, 16]$,*

$$\|F_{ic}\|_{\mathsf{p},Z} \le C\sqrt{\delta}\sqrt{\nu_c^+}, \qquad \|\Phi(F_i)\|_{\mathsf{p},Z} \le C\sqrt{\delta}\,\mathcal{D}, \qquad \|M_{ic}\|_{\mathsf{p},Z} \le C\sqrt{\delta}\,\mathcal{M},$$
$$\Big\|\sum_{k\ne i} J_{ik}\rho_{kc}\Big\|_{\mathsf{p},Z} \le C\sqrt{\delta}.$$

*The same holds when the field carries one fractional label, and when several indices $k$ are deleted from the sum.*

*Proof.* First bound: Lemma D.19 with $S = \{k \ne i : Z_k = c\}$, $\xi \equiv 1$ and $|S| \le N_c$; the exceptional edge and the fractional label each add at most one term $2\Delta/q_0$, which the $2/n$ inside $\nu_c^+$ absorbs. Second bound: the pairing bound of Lemma D.21, Hölder's inequality and Lemma D.18 give $\|\Phi(F_i)\|_{\mathsf{p},Z} \le \|e^{2\gamma_i}\|_{2\mathsf{p},Z}\sum_c |\dot\theta_c|\|F_{ic}\|_{2\mathsf{p},Z}$, where $2\mathsf{p} \le 32$ and the order of the exponential weight is $4\mathsf{p} \le 64 \le s_0$. Fourth bound: Lemma D.22 with $\varphi = \rho_c$ and $\xi_k^\circ = \rho_c(F_k^{(-i)})$, where

$\|\xi_k^\circ\|_{\mathsf{p},Z} \le C$, and Lemma D.19 give a main term $C\Delta\sqrt{n}$, while the leave-one-out remainder is at most $(4\Delta^2/q_0^2)\sum_k e^{2\gamma_k+1/2}$ by Lemma D.20(iii), of norm at most $C\delta$. Third bound: for $c \in \{2,3\}$,

$$M_{ic} = \sum_{k\neq i:\ Z_k=c} J_{ik}(1-\pi_{kc}) \ - \sum_{k\neq i:\ Z_k\neq c} J_{ik}\pi_{kc}.$$

In the first sum, $\varphi = 1-\pi_c \in [0,1]$, giving a main term $C\Delta\sqrt{N_c}$ and a leave-one-out remainder $C\Delta^2 N_c$. In the second, $\varphi = \pi_c \le \theta_c e^{2\gamma}$, giving a main term $C\Delta\sqrt{n}\,\theta_c$ and, by Lemma D.20(iii), a leave-one-out remainder at most $C\Delta^2 n\theta_c$. Together these are $C\sqrt{\delta}(\sqrt{\nu_c^+}+\theta_c)$. The case $c=1$ follows from $M_{i1} = -M_{i2} - M_{i3}$. □

All three components of $M_i$ are of minority-class order: a majority-class neighbour enters only with the weight $1-\pi_{k1} \le \varepsilon e^{2\gamma_k}$, and $\|1-\pi_{k1}\|_{\mathsf{p},Z} \le C\varepsilon$ by Lemma D.18 and Hölder, while a pointwise $O(\varepsilon)$ would be false because the exponential weight cannot be removed; and there are only $N_{\min}$ minority-class neighbours.

**Proposition D.29** (conditional bound on the edge derivative)**.** *In the setting above, for* $\mathsf{p} \in [2,4]$,

$$\left\|\partial_u \tilde{S}_{\mathrm{RB}}\right\|_{\mathsf{p},Z} \ \le\ C\sqrt{\delta}\,\big[\mathcal{B}_{a|b}(Z) + \mathcal{B}_{b|a}(Z)\big],$$

$$\mathcal{B}_{a|b}(Z) := \varepsilon^2 + \varepsilon\mathcal{D} + \sqrt{\delta}\big(\mathcal{D}_2 + \mathcal{D}\mathcal{M} + \varepsilon\nu_{\min}^+\big) + \mathbf{1}\{Z_b \neq 1\}\Big(\mathcal{D} + |\dot\theta_{Z_b}|\mathcal{M} + \frac{\varepsilon}{\sqrt{n}}\Big),$$

*where* $\mathcal{B}_{b|a}$ *is the same quantity with* $a$ *and* $b$ *interchanged.*

*Proof.* By Lemma D.25 and Lemma D.26, $\partial_u \tilde{S}_{\mathrm{RB}} = (\mathrm{I})_a + (\mathrm{I})_b + B_a + B_b$, where $(\mathrm{I})_a := \sum_{i\notin\{a,b\}}\langle V_a \mathbf{e}_y, \Phi_i^{a\to\cdot}\rangle$ and $B_a = T_1 - T_2 + T_3 + \partial_u Q_a$. Only the half carrying $a$ is written out. Every application of Hölder's inequality below splits a product into at most four factors, and the order of an exponential weight never exceeds $s_0$.

*The block* $(\mathrm{I})_a$. Here $\langle V_a\mathbf{e}_y, \Phi_i^{a\to\cdot}\rangle = \pi_{ay}\sum_{c\neq y}\pi_{ac}[\Phi(F_i^{a\to y}) - \Phi(F_i^{a\to c})]$ with $F_i^{a\to c} = F_i^{(-a)} + J_{ia}\mathbf{e}_c$. Taylor's formula gives

$$\Phi\big(F_i^{a\to y}\big) - \Phi\big(F_i^{a\to c}\big) = J_{ia}\xi_i + J_{ia}^2 r_i, \qquad \xi_i := (\partial_y - \partial_c)\Phi\big(F_i^{(-a)}\big), \quad |r_i| \le C\varepsilon e^{6\gamma_i+1},$$

where $\xi_i$ is $\mathcal{G}_{-a}$-measurable with $|\xi_i| \le C\varepsilon e^{6\gamma_i+1}$ by Lemma D.21. Lemma D.19 gives $\|\sum_i J_{ia}\xi_i\|_{2\mathsf{p},Z} \le C\sqrt{\delta}\varepsilon$, and the sum of the remainders has norm at most $C\delta\varepsilon$. As for the prefactor, when $c \neq y$ at least one of the two indices is a minority class, so $\pi_{ay}\pi_{ac} \le \theta_y\theta_c e^{4\gamma_a} \le \varepsilon e^{4\gamma_a}$. Hence $\|(\mathrm{I})_a\|_{\mathsf{p},Z} \le C\varepsilon^2\sqrt{\delta}$.

*The terms* $T_3$ *and* $\partial_u Q_a$. We have $|T_3| \le (2\Delta/q_0)|\nabla\Phi(F_a)|_\infty |V_b\mathbf{e}_x|_1 \le C\Delta\varepsilon^2 e^{6\gamma_a+2\gamma_b}$, of norm at most $C\varepsilon^2\sqrt{\delta}$. For $k \notin \{a,b\}$ the vector $\pi_k$ is free of $u$, and

$$\partial_u q_{ak} = \sum_c \pi_{kc}\int_0^1 (1-r)\nabla^3\Phi\big(F_a + rJ_{ak}d_{kc}\big)[d_{kc}, d_{kc}, \mathbf{e}_y]\mathrm{d}r, \qquad |\partial_u q_{ak}| \le C\varepsilon e^{6\gamma_a+1}(1-\pi_{kZ_k}),$$

because $d_{kc} = 0$ when $c = Z_k$; and $1-\pi_{kZ_k} \le \mathbf{1}\{Z_k \neq 1\} + \varepsilon e^{2\gamma_k}$. Hence $\|\sum_{k\notin\{a,b\}} J_{ak}^2\partial_u q_{ak}\|_{\mathsf{p},Z} \le C\Delta^2\varepsilon(N_{\min} + n\varepsilon) \le C\sqrt{\delta}\cdot\sqrt{\delta}\varepsilon(\nu_{\min}^+ + \varepsilon)$. The term $k=b$ is $\partial_u(u^2 q_{ab}) = 2uq_{ab} + u^2\partial_u q_{ab}$, where $|q_{ab}| \le C\varepsilon e^{6\gamma_a+1}(1-\pi_{by}) \le C\varepsilon e^{6\gamma_a+2\gamma_b+1}(\mathbf{1}\{y\neq 1\}+\varepsilon)$ and $|\partial_u q_{ab}| \le C\varepsilon e^{6\gamma_a+2\gamma_b+1}$, since $u$ enters in three places, through $\pi_b$, through $F_a$ and through the displacement $ru\,d_{bc}$, the first handled by Lemma D.20(ii) and the other two by the third-order bound of Lemma D.21. Using $\Delta = \sqrt{\delta/n}$ and $1/n \le \varepsilon$, the norm of this term is at most $C\sqrt{\delta}[\varepsilon^2 + \mathbf{1}\{y\neq 1\}\varepsilon/\sqrt{n}]$.

*The term* $T_1 - T_2$ *for* $y = 1$. Use the six terms of Lemma D.27 together with Lemma D.28 and Lemma D.20:

- $|A_1| \le \sum_c |w_c|\cdot 2(\gamma_a+\gamma_b)e^{2\gamma_a+2\gamma_b}$ and $\sum_c |w_c| \le 4\varepsilon^2$ by (39), so $\|A_1\|_{\mathsf{p},Z} \le C\varepsilon^2\sqrt{\delta}$;
- $|A_2| \le \frac{1}{2}e^{2\gamma_a}\sum_c |\dot\theta_c| F_{ac}^2$, so $\|A_2\|_{\mathsf{p},Z} \le C\sum_c |\dot\theta_c|\|F_{ac}\|_{4\mathsf{p},Z}^2 \le C\delta\mathcal{D}_2$;

- $|\pi_{a1}-\langle\pi_a,\pi_b\rangle| = |\pi_{a1}(1-\pi_{b1})-\pi_{a2}\pi_{b2}-\pi_{a3}\pi_{b3}| \le (1-\pi_{a1})+(1-\pi_{b1}) \le \varepsilon(e^{2\gamma_a}+e^{2\gamma_b})$, so $\|A_3\|_{\mathsf{p},Z} \le C\varepsilon\sqrt{\delta}\mathcal{D}$;
- $\|A_4\|_{\mathsf{p},Z} \le \sum_c |w_c| \|e^{2\gamma_a}\|_{2\mathsf{p},Z} \| \sum_k J_{ak}\rho_{kc}\|_{2\mathsf{p},Z} \le C\varepsilon^2\sqrt{\delta}$;
- $|A_5| \le e^{2\gamma_a}\cdot 4\varepsilon^2 \cdot 2|M_a|_\infty$, so $\|A_5\|_{\mathsf{p},Z} \le C\varepsilon^2\sqrt{\delta}$;
- $|A_6| \le 3|\Phi_a||M_a|_\infty$, so $\|A_6\|_{\mathsf{p},Z} \le C\delta\mathcal{D}\mathcal{M}$.

*The term $T_1 - T_2$ for $y \in \{2,3\}$.* The bound $\|A_1\|_{\mathsf{p},Z} \le C\varepsilon^2\sqrt{\delta}$ is as above; $|B_2| \le |\Phi_a|$ gives $C\sqrt{\delta}\mathcal{D}$; $|B_3| \le e^{2\gamma_a}|\dot\theta_y|\cdot 2|M_a|_\infty$ gives $C|\dot\theta_y|\sqrt{\delta}\mathcal{M}$; $|B_4| \le 2|\Phi_a||M_a|_\infty$ gives $C\delta\mathcal{D}\mathcal{M}$; and $\pi_{ay} \le \varepsilon e^{2\gamma_a}$ together with $|\rho_a \odot \dot\theta - \Phi_a\pi_a|_\infty \le C\varepsilon e^{6\gamma_a}$ gives $\|B_5\|_{\mathsf{p},Z} \le C\varepsilon^2\sqrt{\delta}$.

Adding the terms gives the claim. □

**Theorem D.30** (ES-E). *Let $q \in [q_0, 1-q_0]$, $\delta \le c_1 \le 1$, $\varepsilon \le 1/12$, $n \ge n_0$ and $n\varepsilon \ge 1$. Then for all $t \in (0,1]$,*

$$\sup_{u\in[\nu,\lambda]} \mathbb{E}_t\big[(\partial_u \tilde S_{\mathrm{RB}}(u))^2\big] \le C\varepsilon^4\delta, \qquad \mathbb{E}_t\big[(\tilde S_{\mathrm{RB}} - \tilde S_{\mathrm{RB}}^{(ab)})^2\big] \le C\Delta^2\varepsilon^4\delta,$$

*with $C$ depending only on $(q_0, c_1)$.*

*Proof.* Fix $u \in [\nu,\lambda]$. After replacing $J_{ab}$ by $u$, Assumption 2 holds and the law of the remaining inputs is unchanged. Taking $\mathsf{p} = 2$ in Proposition D.29 and averaging over $Z$,

$$\mathbb{E}_t\big[(\partial_u \tilde S_{\mathrm{RB}}(u))^2\big] \le C\delta\,\mathbb{E}\big[(\mathcal{B}_{a|b} + \mathcal{B}_{b|a})^2\big].$$

By Lemma D.23,

$$\begin{aligned}\mathbb{E}\mathcal{B}_{a|b}^2 &\le C\Big[\varepsilon^4 + \varepsilon^2\mathbb{E}\mathcal{D}^2 + \delta\big(\mathbb{E}\mathcal{D}_2^2 + \mathbb{E}[\mathcal{D}^2\mathcal{M}^2] + \varepsilon^2\mathbb{E}(\nu_{\min}^+)^2\big) \\ &\qquad + \mathbb{E}\big[\mathbf{1}\{Z_b \ne 1\}\mathcal{D}^2\big] + \sum_{y\in\{2,3\}} \dot\theta_y^2\mathbb{E}\big[\mathbf{1}\{Z_b = y\}\mathcal{M}^2\big] + \frac{\varepsilon^3}{n}\Big] \\ &\le C\varepsilon^4.\end{aligned}$$

The bound is uniform in $u$, so no continuity estimate in $u$ is needed. The second inequality follows from Lemma D.24. □

**Where $n\varepsilon \ge 1$ is used.** In three places: in Lemma D.23, where it controls the fluctuation of the number of minority-class vertices, since for small $n\varepsilon$ the quantity $\mathbb{E}\nu_c^2$ is dominated by $\theta_c/n$ rather than by $\theta_c^2$; in the exceptional-edge term of Proposition D.29, where $\Delta^2\varepsilon \le \sqrt{\delta}\varepsilon^2$ requires $1/n \le \varepsilon$; and in the absorption of $\varepsilon^3/n$ in Theorem D.30. Taken in isolation, the exceptional-edge term of the block $\partial_u Q_a$ is not small: its leading term is $u\,h(Z_b)$ with $h(z) := \sum_c \theta_c\langle H_\Phi(\mathbf{e}_c - \mathbf{e}_z), \mathbf{e}_c - \mathbf{e}_z\rangle$, and for fixed $n, \varepsilon, t$ as $\delta \to 0$ one has $\mathbb{E}[(\partial_u Q_a|_{u=J_{ab}})^2]/(\varepsilon^4\delta) \to \sum_y \theta_y h(y)^2/(n\varepsilon^4 q(1-q))$, whose minority-class part is of order $1/(n\varepsilon)$ rather than $O(\delta)$. Here the derivative is evaluated at the true value $u = J_{ab}$, using $\mathbb{E}J_{ab}^2 = \Delta^2/(q(1-q)) + O(\Delta^3)$ with an $O(\Delta^3)$ constant depending only on $q_0$; for a deterministic $u = u(\Delta)$ with $u(\Delta)/\Delta \to \kappa$ the limit is instead $\kappa^2\sum_y \theta_y h(y)^2/(n\varepsilon^4)$. This is the second moment of that block by itself, not a summand inside the second moment of the full derivative: the cross terms with the other blocks cancel it, since in the first-order expansion of the full derivative in $J$ the coefficient of the single edge $u$ is $c_x + c_y$ with $c_z = 2(w_z - \mu_3) = O(\varepsilon^2)$ and $\mu_3 := \sum_c \theta_c^2\dot\theta_c$, as follows from the second-order expansion of $\Phi$ at $F = 0$ together with $\sum_c \theta_c (H_\Phi)_{cc} = -2\mu_3$. A blockwise proof cannot see that cancellation, which is why $n\varepsilon \ge 1$ is used.

### D.9 Part 7: the label influence, and ES-L

**Coupling and split.** The input of Lemma D.11 is $(Z, U)$. Resampling $Z_v$ produces $Z' := Z^{v \to Z'_v}$ with $Z'_v \sim \theta(t)$ independent of all other inputs; $U$ is left unchanged, so only the edges of row $v$ can change, $A'_{vj} = \mathbf{1}\{U_{vj} < q + \Delta \mathbf{1}\{Z'_v = Z_j\}\}$. Writing $J$ and $J'$ for the corresponding weights and $f := \tilde{S}_{\mathrm{RB}}$,

$$f(Z, J) - f(Z', J') \;=\; \underbrace{\big[f(Z, J) - f(Z', J)\big]}_{D_{\mathrm{lab}}} \;+\; \underbrace{\big[f(Z', J) - f(Z', J')\big]}_{D_{\mathrm{flip}}}.$$

Both terms vanish when $Z'_v = Z_v$, and $\mathbb{P}(Z'_v \neq Z_v) = 1 - P_2 \leq 2\varepsilon$. Given $(Z, Z')$ the weights $J$ are still mutually independent, take values in $\{\nu, \lambda\}$ and have conditional means bounded by $2\Delta^2/q_0^2$, so $(Z', J)$ satisfies Assumption 2 even though it is not a sample from the model.

**Lemma D.31** (the flipped-edge part)**.** $\mathbb{E}\, D_{\mathrm{flip}}^2 \leq C\varepsilon^4\delta^2(\Delta + \delta)$.

*Proof.* Write $\eta_j := (J'_{vj} - J_{vj})/(\lambda - \nu) \in \{0, \pm 1\}$. One has $\eta_j \neq 0$ exactly when $U_{vj} \in [q, p)$ and $\mathbf{1}\{Z_v = Z_j\} \neq \mathbf{1}\{Z'_v = Z_j\}$; given $(Z, Z')$ the $\eta_j$ are independent with $\mathbb{P}(\eta_j \neq 0 \mid Z, Z') \leq \Delta$. Put $J(s) := J + s(J' - J)$ and $N_f := \sum_j \eta_j^2$. The function $f$ is continuously differentiable in each $J_{vj}$, so

$$D_{\mathrm{flip}} = -(\lambda - \nu) \int_0^1 \sum_j \eta_j\, \partial_{vj} f\big(Z', J(s)\big)\mathrm{d}s,$$

$$D_{\mathrm{flip}}^2 \leq (\lambda - \nu)^2 N_f \sum_j \eta_j^2 \int_0^1 \big(\partial_{vj} f(Z', J(s))\big)^2 \mathrm{d}s,$$

the second inequality being Cauchy–Schwarz in $j$ followed by Jensen in $s$. Fix $j$ and $s$ and condition on $(Z, Z')$ and on the event $\{\eta_j \neq 0\}$: then $J_{vj}(s)$ is a fixed value in $[\nu, \lambda]$, equal to $\nu + s(\lambda - \nu)$ or to $\lambda - s(\lambda - \nu)$ according to the direction of the flip, the remaining weights are still independent with values in $[\nu, \lambda]$, and the conditional means in row $v$ are convex combinations of $\mathbb{E}J_{vl}$ and $\mathbb{E}J'_{vl}$, of absolute value at most $2\Delta^2/q_0^2$. This is exactly Assumption 2, with exceptional edge $\{v, j\}$ and the integer labels $Z'$. Proposition D.29 with $\mathsf{p} = 4$ gives

$$\big\|\partial_{vj} f(Z', J(s))\big\|_4 \leq C\sqrt{\delta}\big[\mathcal{B}_{v|j}(Z') + \mathcal{B}_{j|v}(Z')\big],$$

the norm being taken under that conditional law. Under the same conditional law $N_f = 1 + \sum_{l \neq j} \eta_l^2$, the sum being a sum of independent indicators with success probability at most $\Delta$, so $\|N_f\|_2 \leq 1 + n\Delta + \sqrt{n\Delta} \leq 2(1 + n\Delta)$. By Cauchy–Schwarz,

$$\mathbb{E}\big[D_{\mathrm{flip}}^2 \mid Z, Z'\big] \leq \frac{9\Delta^2}{q_0^2} \sum_j \Delta \cdot 2(1 + n\Delta) \cdot C\delta\big[\mathcal{B}_{v|j}(Z') + \mathcal{B}_{j|v}(Z')\big]^2.$$

The components of $Z'$ are i.i.d. from $\theta(t)$, so the proof of Theorem D.30 gives $\mathbb{E}[\mathcal{B}_{v|j}(Z') + \mathcal{B}_{j|v}(Z')]^2 \leq C\varepsilon^4$. Hence $\mathbb{E}D_{\mathrm{flip}}^2 \leq C\Delta^3(1 + n\Delta)\, n\, \delta\varepsilon^4 = C(\Delta\delta + \delta^2)\delta\varepsilon^4$. □

The flipped-edge part is smaller than the target by a factor $\Delta + \delta$: about $n\Delta$ edges flip at once, each with an influence of the order given by Theorem D.30 times $\lambda - \nu$, so even if all of them added with the same sign the total would be only $n\Delta \cdot \Delta \cdot \varepsilon^2\sqrt{\delta} = \varepsilon^2\delta^{3/2}$.

**An exact formula for the relabelling part.** Write $c := Z_v$, $c' := Z'_v$ and $d := \mathbf{e}_{c'} - \mathbf{e}_c$. Give the vertex $v$ the fractional label $E_v(s) := \mathbf{e}_c + sd \in \Delta_3$, all other labels and all of $J$ being held fixed; the corresponding fields are $F_i(s) = F_i^\circ + J_{iv}E_v(s)$, where $F_i^\circ := \sum_{j \neq i,v} J_{ij}\mathbf{e}_{Z_j}$ is the field after deleting the vertex $v$ from the graph. All the quantities carrying a $\circ$, such as $\pi_k^\circ := \pi(F_k^\circ)$, $V_k^\circ$ and $M_i^\circ := \sum_{k \neq i,v} J_{ik}(\mathbf{e}_{Z_k} - \pi_k^\circ)$, belong to the system of the $n - 1$ vertices obtained by deleting $v$: that system also satisfies Assumption 2, is independent of row $v$, and does not depend on $s$.

**Lemma D.32** (differentiating along the fractional label). $D_{\mathrm{lab}} = -\int_0^1 (L_1 + L_2 + L_3 + L_4)(s)\,\mathrm{d}s$, *where*

$$L_1 := \sum_{k\neq v} J_{kv}\big\langle V_k d,\ \Phi_v^{k\to\cdot}\big\rangle, \qquad L_2 := -\sum_{i\neq v} J_{iv}\big\langle \nabla^2\Phi(F_i)d,\ M_i^{(-v)}\big\rangle,$$

$$L_3 := \sum_{i\neq v}\Big\langle \nabla\Phi(F_i),\ \sum_{k\neq i,v} J_{ik}J_{kv}V_k d\Big\rangle, \qquad L_4 := \sum_{i\neq v}\sum_{k\neq i,v} J_{ik}^2\,\partial_s q_{ik},$$

*with* $M_i^{(-v)} := \sum_{k\neq i,v} J_{ik}(\mathbf{e}_{Z_k} - \pi_k)$*, all quantities on the right being evaluated at the fractional label* $E_v(s)$.

*Proof.* The field $F_v$ does not involve $Z_v$, so $R_v = \Phi(F_v) + \sum_{k\neq v}(\Psi_{vk} - \Phi(F_v))$ depends on $s$ only through the $\pi_k$ inside $\Psi_{vk} = \langle \pi_k, \Phi_v^{k\to\cdot}\rangle$. For $i \neq v$ the term $\Psi_{iv} = \sum_c \pi_{vc}\Phi(F_i^\circ + J_{iv}\mathbf{e}_c)$ does not depend on $s$, so $R_i = \Psi_{iv} + \Theta_i$ with $\Theta_i := \sum_{k\neq i,v}(\Psi_{ik} - \Phi(F_i))$, and the Taylor identity of Lemma D.26, summed over $k \neq i, v$, gives $\Theta_i = -\langle \nabla\Phi(F_i), M_i^{(-v)}\rangle + \sum_{k\neq i,v} J_{ik}^2 q_{ik}$. These identities remain valid at a fractional label, since they use nothing but the smoothness of $\Phi$. Differentiating with $\partial_s F_i = J_{iv}d$ and $\partial_s \pi_k = J_{kv}V_k d$: the first block gives $L_1$, and the derivative of $-\langle \nabla\Phi(F_i), M_i^{(-v)}\rangle$ is $-J_{iv}\langle \nabla^2\Phi(F_i)d, M_i^{(-v)}\rangle + \langle \nabla\Phi(F_i), \sum_k J_{ik}J_{kv}V_k d\rangle$. Finally $f(Z', J) - f(Z, J) = \int_0^1 \partial_s f\,\mathrm{d}s$. □

**Lemma D.33** (bounds on the four blocks). *The norms below are taken under the conditional law given* $(Z, Z')$ *and are uniform in* $s \in [0,1]$*. The statistics carrying a subscript* $-v$ *are computed from the labels other than* $Z_v$*, that is,* $\nu^+_{c,-v} := \#\{k \neq v : Z_k = c\}/n + 2/n$, $\nu^+_{\min,-v} := \nu^+_{2,-v} + \nu^+_{3,-v}$ *and* $\mathcal{M}_{-v} := \sqrt{\nu^+_{\min,-v}} + \varepsilon$*; they are independent of* $(Z_v, Z'_v)$ *and Lemma D.23 applies to them as well. Then, for* $\mathsf{p} = 2$*,*

$$\|L_1\| \le C\delta\varepsilon^2, \qquad \|L_2\| \le C\delta\varepsilon\,\mathcal{M}_{-v}, \qquad \|L_3\| \le C\delta\varepsilon^2, \qquad \|L_4\| \le C\delta^{3/2}\varepsilon\big(\nu^+_{\min,-v} + \varepsilon\big).$$

*Proof.* We first record two row and column sums that are used repeatedly. For $i \neq v$ and $z \in \Delta_3$ put

$$m_i[z] := \sum_{k\neq i,v} J_{ik}J_{kv}\,\tilde v_k, \qquad \tilde v_k \in \Big\{V_k^\circ z,\ \int_0^1 V\big(F_k^\circ + rJ_{kv}E_v\big)z\,\mathrm{d}r,\ V_k z\Big\},$$

a row sum of row $i$ in which the factor $J_{kv}$ of the coefficient $J_{kv}\tilde v_k$ is $\mathcal{G}_{-i}$-measurable and $\tilde v_k$ depends on row $i$ only through $J_{ki}$. Lemma D.22, leaving out $i$ and bounding the gradient by the $3\varepsilon e^{2\gamma}$ of Lemma D.20(ii), together with Lemma D.19, gives

$$\|m_i[z]\|_{\mathsf{p}} \le C\Delta\sqrt{n}\cdot\Delta\varepsilon + C\Delta^2 n\cdot\Delta\varepsilon \le C\sqrt{\delta}\,\Delta\,\varepsilon, \qquad \mathsf{p}\in[2,8]. \tag{43}$$

In the same way, for $k \neq v$ the column sums $Y_k := \sum_{i\neq k,v} J_{ik}\nabla\Phi(F_i)$ and $Y_k^\circ := \sum_{i\neq k,v} J_{ik}\nabla\Phi(F_i^\circ)$ satisfy

$$\|Y_k\|_{\mathsf{p}} + \|Y_k^\circ\|_{\mathsf{p}} \le C\sqrt{\delta}\,\varepsilon, \qquad \mathsf{p}\in[2,8], \tag{44}$$

leaving out $k$ and using $|\nabla\Phi| \le C\varepsilon e^{6\gamma}$ and $|\nabla^2\Phi| \le C\varepsilon e^{6\gamma}$.

*The block* $L_1$. Since $\sum_c (V_k d)_c = 0$ we have $\langle V_k d, \Phi_v^{k\to\cdot}\rangle = \sum_c (V_k d)_c[\Phi(F_v^{(-k)} + J_{vk}\mathbf{e}_c) - \Phi(F_v^{(-k)})]$, of absolute value at most $|V_k d|_1 |J_{vk}| C\varepsilon e^{6\gamma_v+1} \le C\Delta\varepsilon^2 e^{2\gamma_k+6\gamma_v+1}$. Multiplying by $|J_{kv}|$ and summing over $k$, every index carries two weight factors, so taking absolute values directly suffices: $\|L_1\| \le Cn\Delta^2\varepsilon^2 = C\delta\varepsilon^2$.

*The block* $L_2$. Split $L_2 = L_2^\circ + \mathcal{R}_1 + \mathcal{R}_2$, where

$$L_2^\circ := -\sum_{i\neq v} J_{iv}\xi_i^\circ, \quad \xi_i^\circ := \big\langle \nabla^2\Phi(F_i^\circ)d,\ M_i^\circ\big\rangle;$$

$$\mathcal{R}_1 := -\sum_{i\neq v} J_{iv}\big\langle [\nabla^2\Phi(F_i) - \nabla^2\Phi(F_i^\circ)]d,\ M_i^{(-v)}\big\rangle;$$

$$\mathcal{R}_2 := -\sum_{i\neq v} J_{iv}\big\langle \nabla^2\Phi(F_i^\circ)d,\ M_i^{(-v)} - M_i^\circ\big\rangle.$$

The variable $\xi_i^\circ$ is $\mathcal{G}_{-v}$-measurable with $|\xi_i^\circ| \le C\varepsilon e^{6\gamma_i^\circ}|M_i^\circ|_\infty$, so Lemma D.28, applied to the system with $v$ deleted, gives $\|\xi_i^\circ\|_2 \le C\varepsilon\sqrt{\delta}\mathcal{M}_{-v}$, and Lemma D.19 on row $v$ gives $\|L_2^\circ\| \le C\Delta\sqrt{n}\cdot\varepsilon\sqrt{\delta}\mathcal{M}_{-v} = C\delta\varepsilon\mathcal{M}_{-v}$. For $\mathcal{R}_1$: $|\nabla^2\Phi(F_i) - \nabla^2\Phi(F_i^\circ)| \le |J_{iv}|C\varepsilon e^{6\gamma_i+1}$, so each $i$ carries $J_{iv}^2$, while $\|M_i^{(-v)}\|_4 \le C\sqrt{\delta}\mathcal{M}_{-v}$ by Lemma D.28 with a fractional label in the field; hence $\|\mathcal{R}_1\| \le Cn\Delta^2\cdot\varepsilon\cdot\sqrt{\delta}\mathcal{M}_{-v} = C\delta^{3/2}\varepsilon\mathcal{M}_{-v}$. For $\mathcal{R}_2$: $\pi_k - \pi_k^\circ = J_{kv}\int_0^1 V(F_k^\circ + rJ_{kv}E_v)E_v\,\mathrm{d}r$, so $M_i^{(-v)} - M_i^\circ = -m_i[E_v]$ with the second choice of $\tilde v_k$, and (43) gives

$$\|\mathcal{R}_2\| \le \sum_{i\neq v}\frac{2\Delta}{q_0}\cdot C\varepsilon\cdot\big\||m_i|_1 e^{6\gamma_i^\circ}\big\|_2 \le Cn\Delta\cdot\varepsilon\cdot\sqrt{\delta}\Delta\varepsilon = C\delta^{3/2}\varepsilon^2.$$

*The block $L_3$.* Exchanging the order of summation, $L_3 = \sum_{k\neq v} J_{kv}\langle Y_k, V_k d\rangle$. Split $L_3 = L_3^\circ + \mathcal{R}_3 + \mathcal{R}_4$ with

$$L_3^\circ := \sum_{k\neq v} J_{kv}\langle Y_k^\circ, V_k^\circ d\rangle, \qquad \mathcal{R}_3 := \sum_{k\neq v} J_{kv}\langle Y_k, (V_k - V_k^\circ)d\rangle, \qquad \mathcal{R}_4 := \sum_{k\neq v} J_{kv}\langle Y_k - Y_k^\circ, V_k^\circ d\rangle.$$

The variable $\langle Y_k^\circ, V_k^\circ d\rangle$ is $\mathcal{G}_{-v}$-measurable, and by (44) and $|V_k^\circ d|_1 \le 4\varepsilon e^{2\gamma_k^\circ}$ its norm is at most $C\sqrt{\delta}\varepsilon^2$, so Lemma D.19 gives $\|L_3^\circ\| \le C\delta\varepsilon^2$. For $\mathcal{R}_3$: $|(V_k - V_k^\circ)d|_1 \le |J_{kv}|\cdot 6\varepsilon e^{2\gamma_k+1}$, so each $k$ carries $J_{kv}^2$ and $\|\mathcal{R}_3\| \le C\delta\cdot\varepsilon\cdot\sqrt{\delta}\varepsilon$. For $\mathcal{R}_4$: $\nabla\Phi(F_i) - \nabla\Phi(F_i^\circ) = J_{iv}h_i$ with $h_i := \int_0^1 \nabla^2\Phi(F_i^\circ + rJ_{iv}E_v)E_v\,\mathrm{d}r$ and $|h_i|_\infty \le C\varepsilon e^{6\gamma_i+1}$; exchanging the order of summation,

$$\mathcal{R}_4 = \sum_{i\neq v} J_{iv}\big\langle h_i,\ m_i[d]\big\rangle \quad \text{(with the first choice of } \tilde v_k\text{)}, \qquad \|\mathcal{R}_4\| \le Cn\Delta\cdot\varepsilon\cdot\sqrt{\delta}\Delta\varepsilon = C\delta^{3/2}\varepsilon^2.$$

The terms $\mathcal{R}_2$ and $\mathcal{R}_4$ have the same structure, a triangular sum over two summation indices with three weight factors, for which taking absolute values directly would produce a spurious $\sqrt{n}$; the remedy is to pull out the common factor, the $\nabla^2\Phi$ depending only on $F_i$, to apply Lemma D.19 to the inner vector-valued row sum $m_i$, and only then to take absolute values in the outer sum.

*The block $L_4$.* Here $\partial_s q_{ik} = J_{kv}a_{ik} + J_{iv}b_{ik}$ with

$$a_{ik} := \sum_c (V_k d)_c \int_0^1 (1-r)\nabla^2\Phi(F_i + rJ_{ik}d_{kc})[d_{kc}, d_{kc}]\mathrm{d}r,$$

$$b_{ik} := \sum_c \pi_{kc}\int_0^1 (1-r)\nabla^3\Phi(F_i + rJ_{ik}d_{kc})[d_{kc}, d_{kc}, d]\mathrm{d}r,$$

and $|a_{ik}| \le C\varepsilon^2 e^{2\gamma_k + 6\gamma_i + 1}$, $|b_{ik}| \le C\varepsilon e^{6\gamma_i+1}(\mathbf{1}\{Z_k \neq 1\} + \varepsilon e^{2\gamma_k})$. Hence $L_4 = \sum_{k\neq v} J_{kv}\beta_k + \sum_{i\neq v} J_{iv}\beta_i'$ with $\beta_k := \sum_{i\neq k,v} J_{ik}^2 a_{ik}$ and $\beta_i' := \sum_{k\neq i,v} J_{ik}^2 b_{ik}$, where $\|\beta_k\| \le C\delta\varepsilon^2$ and $\|\beta_i'\| \le C\delta\varepsilon(\nu^+_{\min,-v} + \varepsilon)$. Replacing $V_k$, $\pi_k$ and $F_i$ inside $\beta_k$ and $\beta_i'$ by their $\circ$ versions makes them $\mathcal{G}_{-v}$-measurable, and Lemma D.19 gives the main terms $C\sqrt{\delta}\cdot\delta\varepsilon^2$ and $C\sqrt{\delta}\cdot\delta\varepsilon(\nu^+_{\min,-v} + \varepsilon)$; each difference produced by the replacement carries one more factor $J_{kv}$ or $J_{iv}$, by Lemma D.20(ii), (iii) and the third- and fourth-order bounds of Lemma D.21, and is of the form $\sum_{i,k}|J_{kv}|J_{ik}^2|J_{iv}|C\varepsilon^2$ or $\sum_{i,k} J_{iv}^2 J_{ik}^2 C\varepsilon(\mathbf{1}\{Z_k \neq 1\} + \varepsilon)$, where every index carries two weight factors, of norm at most $C\delta^2\varepsilon(\nu^+_{\min,-v} + \varepsilon)$. □

**Lemma D.34** (the relabelling part)**.** $\mathbb{E}\,D_{\mathrm{lab}}^2 \le C\varepsilon^4\delta^2$.

*Proof.* By Lemma D.32, Lemma D.33 and Minkowski's inequality, $\|D_{\mathrm{lab}}\|_{2,(Z,Z')} \le C\delta\varepsilon(\mathcal{M}_{-v} + \varepsilon + \sqrt{\delta}\nu^+_{\min,-v}) \le C\delta\varepsilon\mathcal{M}_{-v}$, using $\varepsilon \le \mathcal{M}_{-v}$ and $\nu^+_{\min,-v} \le \mathcal{M}_{-v}^2 \le 3\mathcal{M}_{-v}$. Now $D_{\mathrm{lab}}$ vanishes off $\{Z_v' \neq Z_v\}$, while $\mathcal{M}_{-v}$ is independent of $(Z_v, Z_v')$, so

$$\mathbb{E}\,D_{\mathrm{lab}}^2 \le C\delta^2\varepsilon^2\cdot\mathbb{P}(Z_v' \neq Z_v)\cdot\mathbb{E}\mathcal{M}_{-v}^2 \le C\delta^2\varepsilon^2\cdot 2\varepsilon\cdot C\varepsilon.$$ □

The dominant term is $L_2$: an $\varepsilon$ from the coefficient of $\nabla^2\Phi$, times $\sqrt{\varepsilon\delta}$ from the minority-class order of $M_i^\circ$, times $\sqrt{\delta}$ from the row sum of row $v$, times the square root of the relabelling probability $2\varepsilon$. This is the surviving degenerate pairwise interaction, of order exactly $\varepsilon^4\delta^2$, with no slack.

**Theorem D.35** (ES-L). *Under the hypotheses of Theorem D.30,*

$$\mathbb{E}_t\big[(\tilde{S}_{\mathrm{RB}} - \tilde{S}_{\mathrm{RB}}^{(v)})^2\big] \;\le\; 2\,\mathbb{E}D_{\mathrm{lab}}^2 + 2\,\mathbb{E}D_{\mathrm{flip}}^2 \;\le\; C\varepsilon^4\delta^2,$$

*with $C$ depending only on $(q_0, c_1)$.*

*Proof.* The first inequality is $(x+y)^2 \le 2x^2 + 2y^2$ applied to $\tilde{S}_{\mathrm{RB}} - \tilde{S}_{\mathrm{RB}}^{(v)} = D_{\mathrm{lab}} + D_{\mathrm{flip}}$, and the second combines Lemma D.34 with Lemma D.31, in which $\Delta + \delta \le C$. □

**Numerical check.** Over the 288 configurations given by $n \in \{12, 40\}$, $\varepsilon \in \{0.08, 0.01\}$, $t \in \{0.3, 1\}$, $\delta \in \{0.05, 0.3\}$, the six pairs $(c, c')$ and $s \in \{0, 0.37, 1\}$, the worst relative difference between $L_1 + L_2 + L_3$ and a five-point central difference in $s$ of the part not containing $Q$ is $5.9 \times 10^{-8}$, which is the differencing noise.

### D.10 Part 8: assembly

*Proof of Proposition D.2.* By Lemma D.11, Theorem D.30 and Theorem D.35,

$$\mathbb{E}_t\tilde{S}_{\mathrm{RB}}^2 \;\le\; \frac{n^2}{4}\cdot C\Delta^2\varepsilon^4\delta \;+\; \frac{n}{2}\cdot C\varepsilon^4\delta^2 \;=\; \Big(\frac{C}{4} + \frac{C}{2}\Big)\varepsilon^4 n\delta^2.$$

The hypothesis $n \ge n_0$ guarantees $\Delta \le \sqrt{c_1/n} \le q_0/2$, so the bounds of Part 3 are available. Take $K_0$ to be the coefficient on the right; it depends only on $(q_0, c_1)$. □

*Proof of Lemma D.3.* The identity $n^3\Delta^4 = n\delta^2 \le nc_1^2$ together with $n^3\Delta^4 \ge C_0$ forces $n \ge C_0/c_1^2 \ge \max\{n_0, a_0^{-4/3}\}$ by (34). Next, $\varepsilon = a_0(n^3\Delta^4)^{-1/4} \le a_0 C_0^{-1/4} \le 1/12$ because $C_0 \ge (12a_0)^4$, and $\delta = n\Delta^2 \le c_1$, and

$$n\varepsilon = a_0\, n\,(n\delta^2)^{-1/4} = a_0\, n^{3/4}\delta^{-1/2} \;\ge\; a_0 n^{3/4} \;\ge\; 1$$

using $\delta \le c_1 \le 1$ and $n \ge a_0^{-4/3}$. So all the hypotheses of Proposition D.2 hold, and $\varepsilon^4 n\delta^2 = a_0^4(n^3\Delta^4)^{-1}\cdot n\delta^2 = a_0^4$, whence

$$\mathbb{E}_t\tilde{S}_{\mathrm{RB}}^2 \;\le\; K_0 a_0^4 \;=\; 1 \qquad (t \in (0,1]).$$

By Lemma D.10(iii), which gives $\mathbb{E}_t[\tilde{S}_{\mathrm{RB}} \mid A] = S_{\mathrm{obs}}$, and by Lemma D.9, we get $I(t) \le \mathbb{E}_t\tilde{S}_{\mathrm{RB}}^2 \le 1$ for every $t \in (0,1]$; the single point $t = 0$ does not affect the integral, since the applicability paragraph of Part 1 gives $P_t(A) > 0$ and $t \mapsto P_t(A)$ continuously differentiable on $[0,1]$, so that $I$ is continuous on $[0,1]$. By Lemma D.7,

$$\mathrm{TV}\big(P_{\theta_\varepsilon}, P_{\theta'_\varepsilon}\big) = \mathrm{TV}\big(P_{\theta(0)}, P_{\theta(1)}\big) \le \frac{1}{2}\Big(\int_0^1 I(t)\,\mathrm{d}t\Big)^{1/2} \le \frac{1}{2}.$$ □

*Proof of Theorem D.4.* The parameter range $C_0^{1/4}n^{-3/4} \le \Delta \le \sqrt{c_1}n^{-1/2}$ is exactly $n^3\Delta^4 \ge C_0$ and $n\Delta^2 \le c_1$, so Lemma D.3 applies. Take the two points $\theta_\varepsilon$ and $\theta'_\varepsilon$ with $\varepsilon = a_0(n^3\Delta^4)^{-1/4} \le 1/12$. By the entropy-gap inequality established in the proof of Theorem 5, which uses only $\varepsilon \le 1/12$ and the defining property of $\varepsilon'$, the gap

$$\omega_\varepsilon := H(\theta'_\varepsilon) - H(\theta_\varepsilon) \;\ge\; \frac{\varepsilon\log 2}{2} \;>\; 0.$$

Since $\mathrm{TV} \le 1/2$, Le Cam's two-point method in the form established there, external randomization included, applies verbatim to these two points with separation $\omega_\varepsilon$ and gives $R_n^*(q,\Delta) \ge \omega_\varepsilon^2/16 \ge \varepsilon^2(\log 2)^2/64$. Substituting $\varepsilon^2 = a_0^2(n^3\Delta^4)^{-1/2}$ and $a_0^2 = K_0^{-1/2}$ gives the claim. □

*Proof of Corollary D.5.* The lower bound is Theorem D.4. For the upper bound, $\Delta \le \sqrt{c_1} n^{-1/2}$, and the specialization of Theorem 2 with the constant $\sqrt{c_1}$ gives $\mathbb{E}_\theta(\widehat{H} - H(\theta))^2 \le 81(c_1^2 + c_1/4 + 1/16)^{1/2}(n^3\Delta^4)^{-1/2}$ for every $\theta \in \Delta_3$, so $R_n^*$ is at most the same quantity. □

*Proof of Corollary D.6.* Write $s := n^3\Delta^4 = n\delta^2$.

*Upper bound.* For $s \ge 1$ one has $\min\{1, s^{-1/2}\} = s^{-1/2}$, and the specialization of Theorem 2 with the constant $\sqrt{c_1}$ gives $R_n^* \le 81(c_1^2 + c_1/4 + 1/16)^{1/2} s^{-1/2}$. For $s < 1$ one has $\min\{1, s^{-1/2}\} = 1$, and Theorem 2 gives $R_n^* \le (\log 3)^2$. In both cases the bound is at most $C_\star \min\{1, s^{-1/2}\}$.

*Lower bound for $s \ge C_0$.* Theorem D.4 gives $R_n^* \ge \frac{(\log 2)^2}{64} K_0^{-1/2} s^{-1/2} \ge \frac{(\log 2)^2}{64} K_0^{-1/2} \min\{1, s^{-1/2}\}$.

*Lower bound for $s < C_0$.* The hypothesis $n\Delta^2 \le c_1 \le \min\{q_0^2/4, v_0/20\}$ is exactly the hypothesis of Theorem 5, and the requirement placed on $c_1$ in Appendix D.1 is there for this step, so $R_n^* \ge \frac{v_0^2(\log 2)^2}{576} \min\{1, s^{-1}\}$. For $s \le 1$ one has $\min\{1, s^{-1}\} = 1 = \min\{1, s^{-1/2}\}$, and for $1 < s < C_0$ one has $\min\{1, s^{-1}\} / \min\{1, s^{-1/2}\} = s^{-1/2} > C_0^{-1/2}$. In both cases $\min\{1, s^{-1}\} \ge \min\{1, C_0^{-1/2}\} \min\{1, s^{-1/2}\}$.

Taking the smaller of the two lower bounds gives $c_\star$. □

On the segment where $n^3\Delta^4$ is bounded, both bounds are of constant order and their ratio is controlled by the single constant $\min\{1, C_0^{-1/2}\}$, which is why the rate matches on the whole weakly separated region with no further lower-bound machinery.

*Remark* D.36 (why the estimates are made after grouping by rows). Take $n = 800$, $\delta = \varepsilon$, $t = 0.5$ and $q = 0.3$; the table lists the second moments of the blocks in units of $\varepsilon^4\delta$, estimated with a stratified estimator. The blocks $(\mathrm{II})_a$, $(\mathrm{III}) = (\mathrm{III})_a + (\mathrm{III})_b$ and $B_a$ are those of Lemma D.25.

| $\varepsilon$ | 0.16 | 0.08 | 0.04 | 0.02 | 0.01 |
|---|---|---|---|---|---|
| $(\mathrm{II})_a$ | 7.14 | 19.06 | 37.75 | 66.92 | 129.51 |
| $(\mathrm{III})$ | 8.31 | 22.84 | 54.82 | 113.04 | 232.40 |
| $B_a$ | 5.89 | 9.78 | 11.24 | 10.86 | 9.73 |

Over these four halvings of $\varepsilon$ the two ungrouped blocks grow by factors between 1.77 and 2.75 per step, consistent with growth of order $1/\varepsilon$, while the grouped $B_a$ varies by at most a factor of 1.91 across the whole range. The table illustrates the contrast rather than establishing a rate, and that contrast is the reason we take the norm after grouping: an argument that bounds the two blocks separately and then adds them would leave a factor $\varepsilon^{-1}$ behind.

# E Proofs for the separation from recovery

The three statements proved here are made at a fixed $\theta \in \Delta_3$ and do not use the least-favourable path of Appendix D. They do reuse its local fields and posteriors: $J_{ij}$, $F_i$, $\pi_i$, $\rho_i$ and $\gamma_i$ are the quantities introduced there, and $\delta := n\Delta^2$.

The standing hypotheses are those of (1) together with $q \in [q_0, 1 - q_0]$, $n \ge n_0$ and $\delta \le c_1$, the constants $q_0 \in (0, 1/2]$, $c_1$ and $n_0$ being the ones fixed in Appendix D. The distribution $\theta \in \Delta_3$ is arbitrary and need not lie on the path used there. Proposition E.2 needs neither $n \ge n_0$ nor $\delta \le c_1$. Here $\widehat{Z}_1$ is an estimator of $Z_1$ with values in $\{1, 2, 3\}$ that may depend on $(A, Z_{-1})$ and on a random seed independent of $(A, Z)$, and $\mathcal{L}(Z_i \mid Y)$ denotes the conditional law of $Z_i$ given $Y$.

## E.1 The per-answer bound

We restate Proposition 7.

**Proposition E.1** (per-answer recovery lower bound)**.** *Let $q_0 \in (0, 1/2]$, $q \in [q_0, 1 - q_0]$, $n \geq n_0$, $\delta \leq c_1$ and $\theta \in \Delta_3$, and let $\widehat{Z}_1$ be an arbitrary, possibly randomized, function of $(A, Z_{-1})$. Then*

$$\mathbb{P}\big(\widehat{Z}_1 \neq Z_1\big) \;\geq\; 1 - \max_c \theta_c - C\sqrt{\delta}, \tag{45}$$

*where $C$ depends only on $(q_0, c_1)$. The same bound holds for an estimator that sees only $A$, since $\sigma(A) \subset \sigma(A, Z_{-1})$.*

*Proof. Step 1 (the posterior is $\pi_1$).* The proof of Lemma D.8 uses only Bayes' rule and the conditional independence of the edges given $Z$, and does not use the path, so for every $\theta \in \Delta_3$

$$\mathbb{P}(Z_1 = c \mid A, Z_{-1}) = \pi_{1c}$$

(for a class with $\theta_c = 0$ both sides vanish).

*Step 2 (the Bayes error).* For an arbitrary randomized $\widehat{Z}_1 = g(A, Z_{-1}, W)$, with $W$ independent of $(A, Z)$, conditioning on $W$ first and then using Step 1 gives

$$\mathbb{P}(\widehat{Z}_1 \neq Z_1) = \mathbb{E}\big[1 - \pi_{1,\widehat{Z}_1}\big] \;\geq\; \mathbb{E}\Big[1 - \max_c \pi_{1c}\Big].$$

*Step 3 (the deviation of the posterior from the prior).* If $\theta_c = 0$ then $\pi_{1c} = 0$ by Step 1 and $|\pi_{1c} - \theta_c| = 0$, so assume $\theta_c > 0$, where $\rho_{1c} = \pi_{1c}/\theta_c$ is defined. Since $\pi_{1c} = \theta_c \rho_{1c}$,

$$|\pi_{1c} - \theta_c| = \theta_c |\rho_{1c} - 1| \leq |\rho_{1c} - 1| \leq 2\gamma_1 e^{2\gamma_1},$$

the last inequality being the second inequality of part (i) of Lemma D.20, which holds for an arbitrary field and does not use the path. Hence

$$\max_c \pi_{1c} \;\leq\; \max_c \theta_c + 2\gamma_1 e^{2\gamma_1}.$$

*Step 4 (the size of the field).* By the field bounds of Appendix D.5, the model satisfies the admissible-weight hypothesis under which the tools of Appendix D.6 are stated. Taking $a = 1$, $S = \{k \neq 1 : Z_k = c\}$ and $\xi \equiv 1$ in Lemma D.19 gives $\|F_{1c}\|_{2,Z} \leq C\Delta\sqrt{|S|} \leq C\Delta\sqrt{n} = C\sqrt{\delta}$, and taking the expectation over $Z$ gives $\|F_{1c}\|_2 \leq C\sqrt{\delta}$, so that $\|\gamma_1\|_2 \leq \sum_c \|F_{1c}\|_2 \leq C\sqrt{\delta}$. The exponential moment bound of Appendix D.6, applied with $s = 4 \leq s_0$, gives $\mathbb{E}e^{4\gamma_1} \leq \Lambda_*$. By Cauchy–Schwarz,

$$\mathbb{E}\big[2\gamma_1 e^{2\gamma_1}\big] \;\leq\; 2\,\|\gamma_1\|_2 \left(\mathbb{E}e^{4\gamma_1}\right)^{1/2} \;\leq\; 2C\sqrt{\Lambda_*}\,\sqrt{\delta}.$$

Combining Steps 2 to 4 gives (45). □

### E.2 The bound for the whole partition

We restate Proposition 8.

**Proposition E.2** (recovery lower bound for the whole partition)**.** *Let $q \in [q_0, 1 - q_0]$, $\theta \in \Delta_3$ and $n \geq 2$; neither $\delta \leq c_1$ nor $n \geq n_0$ is needed. For an arbitrary, possibly randomized, $\widehat{Z} = \widehat{Z}(A, W) \in \{1, 2, 3\}^n$, write*

$$L(\widehat{Z}, Z) := \min_{\pi \in S_3} \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}\{\pi(\widehat{Z}_i) \neq Z_i\}$$

*for the permutation-invariant Hamming loss. Then*

$$\mathbb{E}\, L(\widehat{Z}, Z) \;\geq\; 1 - \max_c \theta_c - \sqrt{\frac{(n-1)\Delta^2}{4q_0(1-q_0)} + \frac{\log 6}{2n}}. \tag{46}$$

*In particular, when $n\Delta^2 \to 0$ the right-hand side tends to the prior error $1 - \max_c \theta_c$: even the step of aligning the label names to the most favourable permutation does not rescue the estimator.*

*Proof. Step 1 (give the optimal permutation to the estimator).* Let $\pi^*$ attain the minimum defining $L$; it is a function of $(\widehat{Z}, Z, W)$ with values in $S_3$, and $|S_3| = 6$. Then $L = \frac{1}{n}\sum_i \mathbf{1}\{\pi^*(\widehat{Z}_i) \neq Z_i\}$, and $\pi^*(\widehat{Z}_i)$ is a function of $Y := (A, W, \pi^*)$. Handing $\pi^*$ to the estimator for free can only decrease the error, so it suffices to prove the lower bound for per-answer estimators measurable with respect to $Y$.

*Step 2 (the per-answer Bayes error).* For every $i$ and every $Y$-measurable $\tilde{Z}_i$,

$$\mathbb{P}(\tilde{Z}_i \neq Z_i) \;\geq\; \mathbb{E}\Big[1 - \max_c \mathbb{P}(Z_i = c \mid Y)\Big] \;\geq\; 1 - \max_c \theta_c - \mathbb{E}\big\|\mathcal{L}(Z_i \mid Y) - \theta\big\|_{\mathrm{TV}},$$

the second inequality by $\max_c p_c \leq \max_c \theta_c + \|p - \theta\|_{\mathrm{TV}}$, which compares $\ell^\infty$ with $\ell^1$.

*Step 3 (Pinsker and Jensen).* $\mathbb{E}\|\mathcal{L}(Z_i \mid Y) - \theta\|_{\mathrm{TV}} \leq \mathbb{E}\sqrt{\frac{1}{2}\mathrm{KL}(\mathcal{L}(Z_i \mid Y) \,\|\, \theta)} \leq \sqrt{\frac{1}{2}I(Z_i; Y)}$, the last step by Jensen's inequality and $\mathbb{E}_Y \mathrm{KL}(\mathcal{L}(Z_i \mid Y) \,\|\, \theta) = I(Z_i; Y)$.

*Step 4 (average over $i$).* By Steps 2 and 3 and Jensen's inequality, the square root being concave,

$$\mathbb{E}\, L \;\geq\; 1 - \max_c \theta_c - \frac{1}{n}\sum_i \sqrt{\tfrac{1}{2}I(Z_i; Y)} \;\geq\; 1 - \max_c \theta_c - \sqrt{\frac{1}{2n}\sum_i I(Z_i; Y)}. \tag{47}$$

*Step 5 (the information budget).* The labels $Z_1, \ldots, Z_n$ are independent, so $\sum_i I(Z_i; Y) \leq I(Z; Y)$. Since $W$ is independent of $(Z, A)$ we have $I(Z; A, W) = I(Z; A)$, and therefore $I(Z; Y) = I(Z; A, W) + I(Z; \pi^* \mid A, W) \leq I(Z; A) + H(\pi^*) \leq I(Z; A) + \log 6$. Finally, the variational representation of mutual information, namely $I(Z; A) \leq \mathbb{E}_Z \mathrm{KL}(P_{A|Z} \,\|\, Q)$ for any reference distribution $Q$, applied with $Q = G(n, q)$, together with the conditional independence of the edges given $Z$ and the Bernoulli bound $\mathrm{KL}(x \,\|\, y) \leq \frac{(x-y)^2}{y(1-y)}$, gives

$$I(Z; A) \;\leq\; \binom{n}{2}\frac{\Delta^2}{q(1-q)} \;\leq\; \binom{n}{2}\frac{\Delta^2}{q_0(1-q_0)},$$

the last step because $x \mapsto x(1-x)$ attains its minimum over $[q_0, 1-q_0]$ at the endpoints. Substituting into (47),

$$\frac{1}{2n}\sum_i I(Z_i; Y) \;\leq\; \frac{1}{2n}\Big[\binom{n}{2}\frac{\Delta^2}{q_0(1-q_0)} + \log 6\Big] = \frac{(n-1)\Delta^2}{4q_0(1-q_0)} + \frac{\log 6}{2n},$$

which is (46). □

### E.3 The separation

We restate Corollary 9.

**Corollary E.3** (separation of estimation from recovery)**.** *Let $q_n \in [q_0, 1-q_0]$, let $\Delta_n = n^{-\gamma}$ with $\gamma \in (1/2, 3/4)$, and let $\theta \in \Delta_3$ be fixed with $\max_c \theta_c < 1$. Then*

- *(i) $n\Delta_n^2 = n^{1-2\gamma} \to 0$, so every sequence of estimators $\widehat{Z}_1^{(n)}$ satisfies $\liminf_n \mathbb{P}(\widehat{Z}_1^{(n)} \neq Z_1) \geq 1 - \max_c \theta_c$: the per-answer class assignment is no better than the prior, and not even the oracle that reveals all the remaining labels rescues it; the same limit holds for the whole partition, since $\liminf_n \mathbb{E}L(\widehat{Z}^{(n)}, Z) \geq 1 - \max_c \theta_c$ by Proposition E.2, and the majority assignment attains it;*
- *(ii) $n^3\Delta_n^4 = n^{3-4\gamma} \to \infty$, so the projected moment estimator of Section 3 is uniformly consistent by Corollary B.6, and $R_n^* \asymp (n^3\Delta_n^4)^{-1/2} \to 0$ by Corollary 3.*

*That is, throughout $\gamma \in (1/2, 3/4)$ the entropy is consistently estimable while no estimator of the labels improves asymptotically on the prior baseline $1 - \max_c \theta_c$, which is positive because $\max_c \theta_c < 1$. The open right endpoint matters: the window of Theorem 4 reaches $\Delta_n = C_0^{1/4} n^{-3/4}$, where $n^3\Delta_n^4 = C_0$ and the matched risk is of constant order rather than vanishing. The baseline is not beaten above $3/4$ either, but there the entropy is no longer estimable, so the interval is where the two behaviours separate rather than the whole region in which recovery fails.*

*Proof.* (i) From $\gamma > 1/2$ we get $\delta_n = n^{1-2\gamma} \to 0$, so $\delta_n \le c_1$ and $n \ge n_0$ once $n$ is large enough; Proposition E.1 then applies and $C\sqrt{\delta_n} \to 0$. (ii) From $\gamma < 3/4$ we get $n^3\Delta_n^4 \to \infty$, so Corollary B.6 gives uniform consistency; and when $\delta_n \le c_1$, Corollary 3 gives $R_n^* \asymp \min\{1, (n^3\Delta_n^4)^{-1/2}\}$, whose right-hand side equals $(n^3\Delta_n^4)^{-1/2}$ for $n$ large. □

*Remark* E.4 (the left end of the window). Part (i) of Corollary E.3 needs $\delta_n \to 0$, and the left end of the window shows why. The window of Theorem 4 contains $\Delta_n = b\, n^{-1/2}$ with $0 < b \le \sqrt{c_1}$, where the rate is still matched, but there $\delta_n = b^2$, the bound (45) stays $Cb$ below the baseline, and the oracle does beat the baseline. Take $\theta = (1/2, 1/2, 0)$ and a fixed $q \in [q_0, 1-q_0]$, write $S_c := \{j \ne 1 : Z_j = c\}$, and let $\widehat{Z}_1 := 1$ if $U > 0$ and $\widehat{Z}_1 := 2$ otherwise, where

$$U := \sum_{j \in S_1} (A_{1j} - q) - \sum_{j \in S_2} (A_{1j} - q)$$

is a function of $(A, Z_{-1})$. Given $Z_{-1}$ and $Z_1 = 1$, the $n-1$ summands of $U$ are independent and bounded by 1, with total mean $\Delta_n|S_1|$ and total variance $(n-1)q(1-q)(1+O(\Delta_n))$; given $Z_1 = 2$ the mean is $-\Delta_n|S_2|$ and the variance is the same up to the same factor. Draw $(Z_j)_{j \ge 2}$ once as an i.i.d. sequence, independent of $Z_1$, and at sample size $n$ read $S_c$ as $\{2 \le j \le n : Z_j = c\}$; then $|S_c|/n \to 1/2$ almost surely. Conditionally on almost every such sequence and on either value of $Z_1$, the centred summands are independent and bounded by 1 in absolute value while their total variance is asymptotic to $nq(1-q)$, so the Lindeberg condition holds. Since $\Delta_n\sqrt{n} = b$, the conditional law of $U/\sqrt{nq(1-q)}$ converges to $N(\mu, 1)$ for $Z_1 = 1$ and to $N(-\mu, 1)$ for $Z_1 = 2$, where $\mu := b/(2\sqrt{q(1-q)})$. By dominated convergence over the labels,

$$\mathbb{P}\big(\widehat{Z}_1 \ne Z_1\big) \;\to\; \mathbb{P}(g > \mu) \;<\; \tfrac{1}{2} \;=\; 1 - \max_c \theta_c, \qquad g \sim N(0,1).$$

The gain $\frac{1}{2} - \mathbb{P}(g > \mu) = \mathbb{P}(0 < g \le \mu)$ is of order $b = \sqrt{\delta_n}$ for small $b$, so the order $\sqrt{\delta}$ of the correction in (45) is attained. The separation is therefore a statement about $\delta_n \to 0$, which holds for every fixed $\gamma \in (1/2, 3/4)$ and fails at the constant-prefactor end $\gamma = 1/2$ of the window.

# F Calibration and the query budget

Suppose the operator runs Algorithm 1 with a measured pair $(\hat{q}, \hat{\Delta})$ in place of $(q, \Delta)$. Write $\widehat{P}^{\mathrm{cal}} := (\widehat{P}_2^{\mathrm{cal}}, \widehat{P}_3^{\mathrm{cal}})$ for the statistics (4) with $(q, \Delta)$ replaced by $(\hat{q}, \hat{\Delta})$, and $\widehat{H}^{\mathrm{cal}}$ for the entropy obtained from them through the projection and the cubic solve of Section 3. Let $\|\cdot\|$ be the Euclidean norm on $\mathbb{R}^2$ and $P := (P_2(\theta), P_3(\theta))$.

The standing hypotheses are those of (1) with $n \ge 3$; the statements that use the weakly separated regime $\delta = n\Delta^2 \le c_1$ say so. Throughout, $(\hat{q}, \hat{\Delta})$ comes from a calibration set independent of $A$; Lemma F.1 assumes in addition $|\hat{\Delta} - \Delta| \le \Delta/2$ and $\eta \le 1$.

## F.1 Sensitivity of the plug-in estimator

**Lemma F.1** (sensitivity to a misspecified channel, given the calibration values)**.** *Every expectation in this lemma is taken given $(\hat{q}, \hat{\Delta})$, that is, with $(\hat{q}, \hat{\Delta})$ treated as a pair of fixed numbers; when $(\hat{q}, \hat{\Delta})$ is random and independent of $A$, take the expectation of the bounds below over it (Corollary F.2). Put*

$$v := \frac{\Delta}{\hat{\Delta}}, \qquad u := \frac{\hat{q} - q}{\hat{\Delta}}, \qquad \eta := \frac{|\hat{q} - q| + |\hat{\Delta} - \Delta|}{\Delta}.$$

*Then the conditional means of $\widehat{P}_2^{\mathrm{cal}}$ and $\widehat{P}_3^{\mathrm{cal}}$ are*

$$\mathbb{E}_\theta\big[\widehat{P}_2^{\mathrm{cal}} \,\big|\, \hat{q}, \hat{\Delta}\big] = v\, P_2 - u, \qquad \mathbb{E}_\theta\big[\widehat{P}_3^{\mathrm{cal}} \,\big|\, \hat{q}, \hat{\Delta}\big] = v^2 P_3 - 2vu\, P_2 + u^2, \tag{48}$$

*and if $|\hat{\Delta} - \Delta| \le \Delta/2$ and $\eta \le 1$, then*

$$\big\|\mathbb{E}_\theta\big[\widehat{P}^{\mathrm{cal}} \,\big|\, \hat{q}, \hat{\Delta}\big] - P\big\| \;\le\; 19\,\eta, \qquad \mathbb{E}_\theta\big[(\widehat{H}^{\mathrm{cal}} - H(\theta))^2 \,\big|\, \hat{q}, \hat{\Delta}\big] \;\le\; \min\Big\{(\log 3)^2,\; 10^3\big(V_n^{1/2} + \eta\big)\Big\}, \tag{49}$$

*with $V_n$ as in (3). The constants are explicit: a misspecified channel inflates the variance of Lemma B.1 by an absolute constant factor and adds a bias of $19\eta$.*

*Proof. Step 1 (T is quadratic in q).* Write

$$T(x) := \sum_i \Big[ \big(d_i - x(n-1)\big)^2 - d_i(1-x)^2 - (n-1-d_i)x^2 \Big],$$

so that $\widehat{P}_3^{\mathrm{cal}} = T(\hat{q})/(N_3\hat{\Delta}^2)$ with $N_3 = n(n-1)(n-2)$. Expanding term by term,

$$\begin{aligned} \big(d_i - x(n-1)\big)^2 &= d_i^2 - 2x(n-1)d_i + x^2(n-1)^2, \\ -d_i(1-x)^2 &= -d_i + 2xd_i - x^2 d_i, \\ -(n-1-d_i)x^2 &= -x^2(n-1) + x^2 d_i, \end{aligned}$$

and adding the three lines, in which the terms $x^2 d_i$ cancel, leaves $d_i^2 - d_i - 2x(n-2)d_i + x^2(n-1)(n-2)$, so that

$$T(x) = \sum_i (d_i^2 - d_i) \;-\; 2x(n-2)\sum_i d_i \;+\; n(n-1)(n-2)\,x^2. \tag{50}$$

*Step 2 (the means of the two moment estimators).* Write $\bar{A} := |E|/N_2$ with $N_2 = \binom{n}{2}$, and $b := \hat{q} - q$. Part 1 of Lemma B.1 gives $\mathbb{E}_\theta \bar{A} = q + \Delta P_2$ and $\mathbb{E}_\theta T(q)/(N_3\Delta^2) = P_3$. For the first identity,

$$\mathbb{E}_\theta \widehat{P}_2^{\mathrm{cal}} = \frac{\mathbb{E}_\theta \bar{A} - \hat{q}}{\hat{\Delta}} = \frac{\Delta P_2 - b}{\hat{\Delta}} = v\,P_2 - u.$$

For the second, (50) and $\sum_i d_i = 2|E| = n(n-1)\bar{A}$ give

$$T(\hat{q}) - T(q) = -2(n-2)b\,n(n-1)\bar{A} + n(n-1)(n-2)\big(2qb + b^2\big) = -2N_3\,b\,(\bar{A} - q) + N_3\,b^2,$$

so that

$$\widehat{P}_3^{\mathrm{cal}} = \frac{T(q)}{N_3\hat{\Delta}^2} - \frac{2b(\bar{A}-q)}{\hat{\Delta}^2} + \frac{b^2}{\hat{\Delta}^2} = v^2 \cdot \frac{T(q)}{N_3\Delta^2} - \frac{2b}{\hat{\Delta}} \cdot \frac{\bar{A}-q}{\hat{\Delta}} + u^2. \tag{51}$$

Taking expectations and using $\mathbb{E}_\theta \frac{2b(\bar{A}-q)}{\hat{\Delta}^2} = \frac{2b\,\Delta P_2}{\hat{\Delta}^2} = 2uv\,P_2$ gives $\mathbb{E}_\theta \widehat{P}_3^{\mathrm{cal}} = v^2 P_3 - 2uv\,P_2 + u^2$, which is (48).

*Step 3 (the size of the bias).* From $|\hat{\Delta} - \Delta| \le \Delta/2$ we get $\hat{\Delta} \ge \Delta/2$, hence $v = \Delta/\hat{\Delta} \le 2$, $|v-1| = |\Delta - \hat{\Delta}|/\hat{\Delta} \le 2|\hat{\Delta} - \Delta|/\Delta \le 2\eta$ and $|u| \le 2|b|/\Delta \le 2\eta$. Using $P_2, P_3 \in [0,1]$ and $\eta \le 1$,

$$\begin{aligned} \big|\mathbb{E}\widehat{P}_2^{\mathrm{cal}} - P_2\big| &\le |v-1|P_2 + |u| \le 4\eta, \\ \big|\mathbb{E}\widehat{P}_3^{\mathrm{cal}} - P_3\big| &\le |v^2-1|P_3 + 2v|u|P_2 + u^2 \le 6\eta + 8\eta + 4\eta^2 \le 18\eta, \end{aligned}$$

where $|v^2 - 1| = |v-1|(v+1) \le 6\eta$. Hence $\|\mathbb{E}\widehat{P}^{\mathrm{cal}} - P\| \le \sqrt{4^2 + 18^2}\,\eta \le 19\eta$.

*Step 4 (the variance).* First, $\mathrm{Var}\widehat{P}_2^{\mathrm{cal}} = \mathrm{Var}(\bar{A})/\hat{\Delta}^2 \le 4\mathrm{Var}\widehat{P}_2$. By the decomposition (51) and $(x+y)^2 \le 2x^2 + 2y^2$,

$$\mathrm{Var}\widehat{P}_3^{\mathrm{cal}} \;\le\; 2v^4\mathrm{Var}\widehat{P}_3 + 8u^2v^2\,\frac{\mathrm{Var}(\bar{A})}{\Delta^2} \;\le\; 32\mathrm{Var}\widehat{P}_3 + 128\eta^2\mathrm{Var}\widehat{P}_2 \;\le\; 32\mathrm{Var}\widehat{P}_3 + 128\mathrm{Var}\widehat{P}_2,$$

so that $\mathrm{Var}\widehat{P}_2^{\mathrm{cal}} + \mathrm{Var}\widehat{P}_3^{\mathrm{cal}} \le 132\,(\mathrm{Var}\widehat{P}_2 + \mathrm{Var}\widehat{P}_3) \le 132\,v_n$ by part 4 of Lemma B.1.

*Step 5 (transfer to the entropy).* Steps 1 to 4 of the proof of Theorem B.4 use the single quantity $\mathbb{E}\|\widehat{P} - P\|^2$ and nowhere use that $\widehat{P}$ is unbiased; unbiasedness enters only at the last inequality of its Step 4, where that quantity is replaced by $v_n$. Therefore

$$\begin{aligned} \mathbb{E}_\theta\big(\widehat{H}^{\mathrm{cal}} - H(\theta)\big)^2 &\le 2\sqrt{10/9}\,\pi^2\Big(\mathbb{E}_\theta\big\|\widehat{P}^{\mathrm{cal}} - P\big\|^2\Big)^{1/2} \\ &\le 20.808\big(132\,v_n + (19\eta)^2\big)^{1/2} \\ &\le 20.808\big(\sqrt{132}\,v_n^{1/2} + 19\eta\big), \end{aligned}$$

and $v_n \le 15V_n$, again by part 4 of Lemma B.1, gives $\le 930\,V_n^{1/2} + 396\,\eta \le 10^3(V_n^{1/2} + \eta)$. Finally $\widehat{H}^{\mathrm{cal}}, H(\theta) \in [0, \log 3]$ gives the cap $(\log 3)^2$, since Step 1 of the proof of Theorem B.4 applies to any $\widehat{P} \in \mathbb{R}^2$ and does not depend on the channel being the correct one. This is (49). □

### F.2 Random calibration values

**Corollary F.2** (random calibration values)**.** *Let $(\hat{q}, \hat{\Delta})$ be independent of $A$. If $|\hat{\Delta} - \Delta| \le \Delta/2$ and $\eta \le 1$ hold almost surely, then taking the expectation over $\eta$ gives*

$$\mathbb{E}_\theta\big(\widehat{H}^{\mathrm{cal}} - H(\theta)\big)^2 \;\le\; \min\Big\{(\log 3)^2,\; 10^3\big(V_n^{1/2} + \mathbb{E}\eta\big)\Big\}.$$

*If the two conditions hold only with probability $1 - p_{\mathrm{fail}}$, the right-hand side gains a further $(\log 3)^2 p_{\mathrm{fail}}$, the failure event being capped by $\widehat{H}^{\mathrm{cal}}, H \in [0, \log 3]$. For that cap to apply, the estimator has to be defined on the whole failure event, and $\hat{\Delta} = 0$ is not excluded there, so we set $\widehat{H}^{\mathrm{cal}} := 0$ when $\hat{\Delta} = 0$; any fixed value in $[0, \log 3]$ serves equally well.*

*Proof.* Since $(\hat{q}, \hat{\Delta})$ is independent of $A$, the conditional bound (49) holds for almost every calibration value. Averaging its affine part in $\eta$ and keeping the deterministic cap $(\log 3)^2$ gives the claim. On the failure event use $(\widehat{H}^{\mathrm{cal}} - H)^2 \le (\log 3)^2$ as a cap, the estimator being defined there by the convention above. □

Both identities (48) and the variance bound of Lemma F.1 are conditional. When $u$ and $v$ are random and integrable, which the almost-sure conditions of Corollary F.2 guarantee, $\mathbb{E}\widehat{P}_2^{\mathrm{cal}} = P_2\mathbb{E}v - \mathbb{E}u$, and the variance picks up the fluctuation of the calibration values themselves. Without an integrability assumption the unconditional moments need not exist: with $\hat{q} = q$ and $\hat{\Delta}$ uniform on $(0, 2\Delta)$ one has $\mathbb{E}(1/\hat{\Delta}) = \infty$ and $\mathbb{E}\widehat{P}_2^{\mathrm{cal}}$ is undefined, although the entropy bound of Corollary F.2 is unaffected because it never uses these raw moments. In the integrable case, for instance, if $\hat{\Delta} = \Delta$ and $\hat{q} = q \pm b$ with equal probabilities and independently of $A$, then

$$\mathbb{E}\widehat{P}_2^{\mathrm{cal}} = P_2 \qquad \text{while} \qquad \mathrm{Var}\widehat{P}_2^{\mathrm{cal}} = \mathrm{Var}\widehat{P}_2 + (b/\Delta)^2.$$

### F.3 Calibration accuracy that preserves the order

We restate Corollary 10.

**Corollary F.3** (accuracy preserving the order of the known-channel upper bound)**.** *Let $n^3\Delta^4 \ge 4$ and $n\Delta^2 \le c_1$, let $(\hat{q}, \hat{\Delta})$ be given and independent of $A$, and let*

$$|\hat{q} - q| + |\hat{\Delta} - \Delta| \;\le\; \frac{1}{n^{3/2}\Delta}.$$

*Then $\mathbb{E}_\theta(\widehat{H}^{\mathrm{cal}} - H(\theta))^2 \le C'(n^3\Delta^4)^{-1/2}$ with*

$$C' := 10^3\big[(c_1^2 + c_1/4 + 1/16)^{1/2} + 1\big] \;\le\; 2.2 \times 10^3,$$

*the same order as Theorem 2.*

*Proof.* The hypothesis $n^3\Delta^4 \ge 4$ gives

$$|\hat{\Delta} - \Delta| \;\le\; \frac{1}{n^{3/2}\Delta} \;\le\; \frac{\Delta}{2}, \qquad \eta = \frac{|\hat{q} - q| + |\hat{\Delta} - \Delta|}{\Delta} \;\le\; \frac{1}{n^{3/2}\Delta^2} = \big(n^3\Delta^4\big)^{-1/2} \;\le\; 1,$$

the second inequality of the first chain being equivalent to $n^3\Delta^4 \ge 4$. So Lemma F.1 applies. Taking $b = \sqrt{c_1}$ in Corollary B.5 gives $V_n^{1/2} \le (c_1^2 + c_1/4 + 1/16)^{1/2}(n^3\Delta^4)^{-1/2}$, and substituting this into (49) gives the claim. □

So the order of the known-channel upper bound is preserved at this calibration accuracy. The condition is sufficient, and nothing above shows that the order degrades once the calibration error exceeds this scale. The next corollary counts the labelled pairs that reach it.

**Corollary F.4** (calibration from labelled pairs)**.** *Let $n\Delta^2 \le c_1$. Suppose the calibration set holds $m$ pairs of answers known to share a meaning and $m$ pairs known to differ, each judged once through the channel $(q, \Delta)$, independently of one another and of $A$, and let $\hat{p}$ and $\hat{q}$ be the fractions of agreements*

*in the two groups and* $\hat{\Delta} := \hat{p} - \hat{q}$. *If* $m \geq \frac{9}{4}n^3\Delta^2$, *then running Algorithm 1 with* $(\hat{q}, \hat{\Delta})$, *with the convention of Corollary F.2 at* $\hat{\Delta} = 0$, *gives*

$$\mathbb{E}_\theta\big(\widehat{H}^{\mathrm{cal}} - H(\theta)\big)^2 \;\leq\; (C' + 6)\,(n^3\Delta^4)^{-1/2},$$

*with* $C'$ *as in Corollary F.3. So* $\frac{9}{2}n^3\Delta^2$ *labelled pairs, half of each kind, preserve the order of Theorem 2.*

*Proof.* Let $\mathcal{G} := \{|\hat{q}-q| \leq \Delta/6,\ |\hat{p}-p| \leq \Delta/6\}$. On $\mathcal{G}$ we have $|\hat{\Delta}-\Delta| \leq |\hat{p}-p|+|\hat{q}-q| \leq \Delta/3$ and $\eta \leq (\Delta/6 + \Delta/3)/\Delta = 1/2$, so both conditions of Corollary F.2 hold there. Hoeffding's inequality for the two averages of $m$ independent Bernoulli variables, and $m\Delta^2/18 \geq n^3\Delta^4/8$, give

$$\mathbb{P}(\mathcal{G}^c) \;\leq\; 4\exp\big(-2m(\Delta/6)^2\big) \;\leq\; 4\exp\big(-n^3\Delta^4/8\big) \;\leq\; 8e^{-1/2}\,(n^3\Delta^4)^{-1/2},$$

the last step because $x^{1/2}e^{-x/8} \leq 2e^{-1/2}$ for every $x > 0$. Next, $\mathbb{E}|\hat{q} - q| \leq (q(1-q)/m)^{1/2} \leq 1/(2\sqrt{m})$, likewise for $\hat{p}$, and $|\hat{\Delta} - \Delta| \leq |\hat{p} - p| + |\hat{q} - q|$, so

$$\mathbb{E}\eta \;\leq\; \frac{2\mathbb{E}|\hat{q} - q| + \mathbb{E}|\hat{p} - p|}{\Delta} \;\leq\; \frac{3}{2\sqrt{m}\,\Delta} \;\leq\; (n^3\Delta^4)^{-1/2},$$

the last step by $m \geq \frac{9}{4}n^3\Delta^2$. Corollary F.2 with $p_{\mathrm{fail}} = \mathbb{P}(\mathcal{G}^c)$ now gives $\mathbb{E}_\theta(\widehat{H}^{\mathrm{cal}} - H(\theta))^2 \leq 10^3(V_n^{1/2}+\mathbb{E}\eta)+(\log 3)^2\mathbb{P}(\mathcal{G}^c)$. Finally $V_n^{1/2} \leq (c_1^2+c_1/4+1/16)^{1/2}(n^3\Delta^4)^{-1/2}$ by Corollary B.5 with $b = \sqrt{c_1}$, and $(\log 3)^2 \cdot 8e^{-1/2} < 5.86$. □

The count is per kind. Pairs drawn at random from the answer distribution differ in meaning with probability $1 - P_2$, so collecting $m$ differing pairs that way takes about $m/(1 - P_2)$ draws, which grows without bound as $\theta$ approaches a vertex. Pairs taken from other questions serve only if the channel is the same there.

### F.4 The channel cannot be dropped

Some information on the channel is necessary, and not only for this estimator. If the background rate $q$ is unknown and nothing else informs it, no function of the graph estimates the entropy uniformly consistently over $(q, \theta)$ in the weakly separated regime, even where $n^3\Delta^4 \to \infty$ and a known channel gives a vanishing risk.

**Proposition F.5** (an unknown background rate)**.** *Let* $q_0 \in (0, 1/2)$, $n \geq 2$ *and* $0 < \Delta \leq \min\{1/2 - q_0,\ (2n)^{-1/2}\}$. *Suppose the estimator is given* $\Delta$ *and the bound of three meanings but not* $q$, *which may take any value in* $[q_0, 1 - q_0]$ *with* $q + \Delta \leq 1$. *Then every estimator* $\widehat{H} = \widehat{H}(A)$, *randomized independently of* $A$ *if desired, satisfies*

$$\sup_{q \in [q_0,\, \min\{1-q_0,\, 1-\Delta\}]}\ \sup_{\theta \in \Delta_3}\ \mathbb{E}_{\theta,q}\big(\widehat{H} - H(\theta)\big)^2 \;\geq\; \frac{(\log 2)^2}{16}.$$

*In particular no estimator is uniformly consistent along a sequence with* $n\Delta_n^2 \leq c_1$, *including sequences with* $n^3\Delta_n^4 \to \infty$, *on which Corollary 3 gives a vanishing risk once* $q$ *is known.*

*Proof.* Put $q_A := \frac{1}{2} - \Delta$ and $q_B := \frac{1}{2} - \frac{\Delta}{2}$; both lie in $[q_0, 1 - q_0]$ because $\Delta \leq 1/2 - q_0$, and both are at most $\frac{1}{2} \leq 1 - \Delta$. Under $\theta_A := (1, 0, 0)$ and $q = q_A$ every pair shares its meaning, so the verdicts are independent $\mathrm{Bern}(q_A + \Delta) = \mathrm{Bern}(1/2)$ and the law of the graph is exactly $G(n, 1/2)$. Under $\theta_B := (0, \frac{1}{2}, \frac{1}{2})$ and $q = q_B$, apply Lemma C.4 with $\varepsilon' = 1$: then $r_S = q_B + \Delta/2 = 1/2$, $c_S = \Delta^2/(4r_S(1 - r_S)) = \Delta^2$ and $nc_S = n\Delta^2 \leq 1/2$, and $R_{1/2}(1)$ puts every answer in the planted subset, so by Definition C.1 it is $G(n, 1/2)$. Hence

$$\mathrm{TV}\big(P_{\theta_A,q_A},\, P_{\theta_B,q_B}\big) \;=\; \mathrm{TV}\big(G(n,1/2),\, P_{\theta_B,q_B}\big) \;\leq\; \kappa\, n\Delta^2 \;\leq\; \kappa/2 \;<\; 1/2.$$

The entropies are $H(\theta_A) = 0$ and $H(\theta_B) = \log 2$. For any estimator the events $\{|\widehat{H}| \geq \frac{1}{2}\log 2\}$ and $\{|\widehat{H} - \log 2| \geq \frac{1}{2}\log 2\}$ cover every outcome, so their probabilities under the first and the second law add up to at least $1 - \mathrm{TV} \geq 1/2$; one of them is therefore at least $1/4$, and the corresponding squared risk is at least $(\frac{1}{2}\log 2)^2/4$. A randomized estimator is handled by conditioning on its randomization, which is independent of $A$ and leaves the total variation unchanged. The final claim follows because $c_1 \leq 1/2$, and $\Delta_n \leq 1/2 - q_0$ once $n$ is large. □

The two laws share the edge density $1/2$, and at $n\Delta^2 \le 1/2$ the second hides its partition, so without $q$ a single meaning judged at agreement rate $1/2$ cannot be told apart from two equally likely meanings judged at rates $\frac{1}{2} \pm \frac{\Delta}{2}$. The minimax problem that treats the calibration set as part of the experiment, with its size as a budget, is not solved here.

### F.5 Verdicts that share an answer

The transfer in Step 5 of the proof of Lemma F.1 also bounds the risk when the channel is right on average but the verdicts are not conditionally independent. The price is a bias in $\widehat{P}_3$, which the weak judge amplifies.

**Proposition F.6** (dependent verdicts)**.** *Let $n \ge 3$ and $0 \le q < q + \Delta \le 1$, let the labels be i.i.d. from $\theta \in \Delta_3$, and let the verdicts, not necessarily conditionally independent, satisfy $\mathbb{E}[A_{ij} \mid Z] = q + \Delta W_{ij}$ for $i < j$. Put*

$$\zeta \;:=\; \frac{1}{N_3} \sum_{(i;j,k)} \mathbb{E}\,\mathrm{Cov}\big(A_{ij}, A_{ik} \,\big|\, Z\big),$$

*the average over the ordered triples of (9), and let $T := \mathrm{Var}\widehat{P}_2 + \mathrm{Var}\widehat{P}_3$ under this law. Then $\mathbb{E}\widehat{P}_2 = P_2$, $\mathbb{E}\widehat{P}_3 = P_3 + \zeta/\Delta^2$, and*

$$\mathbb{E}_\theta\big(\widehat{H} - H(\theta)\big)^2 \;\le\; \min\Big\{(\log 3)^2,\; 2\sqrt{10/9}\,\pi^2\big(T + \zeta^2/\Delta^4\big)^{1/2}\Big\}.$$

*Proof.* The representation (9) and $\widehat{P}_2 = N_2^{-1}\sum_{i<j} X_{ij}$ use only $A_{ij} \in \{0,1\}$, so they hold under any law of the verdicts. Here $\mathbb{E}[X_{ij} \mid Z] = W_{ij}$, and for $j \ne k$

$$\mathbb{E}[X_{ij}X_{ik} \mid Z] \;=\; W_{ij}W_{ik} + \frac{\mathrm{Cov}(A_{ij}, A_{ik} \mid Z)}{\Delta^2}.$$

Averaging over $Z$, with $\mathbb{E}[W_{ij}W_{ik}] = P_3$, and over the ordered triples gives the two means. For the risk, the proof of Theorem B.4 up to the last inequality of its Step 4 gives $\mathbb{E}_\theta(\widehat{H} - H(\theta))^2 \le 2\sqrt{10/9}\,\pi^2(\mathbb{E}\|\widehat{P} - P\|^2)^{1/2}$ and the cap $(\log 3)^2$ without using the law of the verdicts, as in Step 5 of the proof of Lemma F.1, and $\mathbb{E}\|\widehat{P} - P\|^2 = T + (\mathbb{E}\widehat{P}_3 - P_3)^2 = T + \zeta^2/\Delta^4$. □

Lemma B.1 no longer bounds $T$, because pairs of verdicts that share an answer add covariance to both statistics as well; the bias term, however, is exact. Three departures show its size. In each, the answers carry i.i.d. effects independent of $Z$, bounded so that every rate below lies in $[0,1]$; given the labels and the effects, verdicts on distinct pairs are independent, and every verdict keeps the mean $q + \Delta W_{ij}$ given $Z$ alone, so $\widehat{P}_2$ stays unbiased.

- *Additive tendencies*, $A_{ij} \mid (Z, \beta) \sim \mathrm{Bern}(q + \Delta W_{ij} + \beta_i + \beta_j)$ with $\mathbb{E}\beta_i = 0$ and $\mathrm{Var}\beta_i = s^2$: then $\zeta = s^2$, and the bias of $\widehat{P}_3$ is $s^2/\Delta^2$.
- *Effects on the separation*, $A_{ij} \mid (Z, a) \sim \mathrm{Bern}(q + \Delta a_i a_j W_{ij})$ with $\mathbb{E}a_i = 1$ and $\mathrm{Var}a_i = \tau^2$: then $\zeta = \tau^2\Delta^2 P_3$, and the bias $\tau^2 P_3$ is not amplified.
- *Effects on the whole rate*, $A_{ij} \mid (Z, a) \sim \mathrm{Bern}(a_i a_j (q + \Delta W_{ij}))$ with the same $a$: then $\zeta/\Delta^2 = \tau^2(q^2/\Delta^2 + 2qP_2/\Delta + P_3)$, amplified again.

In the additive case the means of the two statistics do not separate the tendencies from the label signal. There $\mathbb{E}[d_i \mid Z_i, \beta_i] = (n-1)(q + \Delta\theta_{Z_i} + \beta_i)$, so the normalized expected degrees $\mathbb{E}[d_i \mid Z_i, \beta_i]/(n-1)$ vary across answers through the label, with variance $\Delta^2(P_3 - P_2^2)$, and through the tendency, with variance $s^2$. Accordingly the two statistics of (4) have the same means under $(\theta, s)$ as under conditional independence at any $\theta'$ with $P_2(\theta') = P_2(\theta)$ and $P_3(\theta') = P_3(\theta) + s^2/\Delta^2$, whenever such a $\theta'$ exists. Whether some statistic of the graph removes such effects and still attains the rate of Corollary 3 is open. Appendix G.4 measures the three cases at the settings of Table 1.

### F.6 A $d$-REGULAR QUERY DESIGN

Querying every pair costs $\binom{n}{2}$ judgments. When that number is a budget rather than a given, the natural restriction is to call the judge only on the edges of a fixed graph. Throughout this subsection $G$ is a simple $d$-regular graph on the $n$ answers, with edge set $E(G)$ and $m := nd/2$ edges, and only $\{A_{ij}\}_{\{i,j\}\in E(G)}$ is observed. The graph is fixed before the data are seen: it is independent of $A$ and of $Z$, and it is not chosen adaptively. As in Section 3, the labels are i.i.d. from $\theta \in \Delta_3$, the verdicts are conditionally independent given the labels and follow (1), the channel $(q, \Delta)$ is known and satisfies $0 \le q < q + \Delta \le 1$, and $n \ge 3$. Nothing below requires $q$ to stay away from 0 and 1, and nothing below asks for a weak-separation condition on $\Delta$.

Write $X_{ij} := (A_{ij} - q)/\Delta$ and $W_{ij} := \mathbf{1}\{Z_i = Z_j\}$, so that $\mathbb{E}[X_{ij} \mid Z] = W_{ij}$, and let $\xi_{ij} := X_{ij} - W_{ij}$ be the centred noise. Given $Z$ the variables $\{\xi_e\}_{e\in E(G)}$ are independent with mean zero and $\mathrm{Var}(\xi_e \mid Z) = v_e/\Delta^2 \le \rho$, where $v_e$ is the conditional variance of $A_e$ given $Z$ and $\mathbb{E}v_e = \Delta^2\bar\rho$ (Lemma B.1). Let $N_G(i)$ be the set of neighbours of $i$ in $G$, so that $|N_G(i)| = d$, let

$$\Lambda(G) \;:=\; \big\{(i;j,k) \;:\; \{i,j\},\{i,k\} \in E(G),\; j \ne k\big\}$$

be the ordered wedge set of $G$, with $N_\Lambda := |\Lambda(G)| = n\,d(d-1)$, and let $\Lambda_u(G)$ be the unordered wedge set, with $|\Lambda_u(G)| = n\binom{d}{2} = N_\Lambda/2$.

**Corollary F.7.** *Let $n \ge 3$, let $0 \le q < q + \Delta \le 1$, and let $G$ be a simple $d$-regular query graph as above with $2 \le d \le n-1$. Put*

$$\widehat P_2^G := \frac{1}{m}\sum_{\{i,j\}\in E(G)} X_{ij}, \qquad \widehat P_3^G := \frac{1}{N_\Lambda}\sum_{(i;j,k)\in\Lambda(G)} X_{ij}X_{ik},$$

*and let $\widehat H_G$ be the estimator obtained from the pair $(\widehat P_2^G, \widehat P_3^G)$ by the remaining steps of Section 3: project onto $M_3$, invert the moments to $\widetilde\theta$, and take the entropy. Then*

1. *$\mathbb{E}\widehat P_2^G = P_2$ and $\mathbb{E}\widehat P_3^G = P_3$;*
2. *the variance of the second-moment estimator is*

$$\mathrm{Var}(\widehat P_2^G) \;=\; \frac{P_2(1-P_2) + 2(d-1)(P_3 - P_2^2) + \bar\rho}{m} \;\le\; \frac{2d-1}{2nd} + \frac{\rho}{m} \;\le\; \frac{1}{n} + \frac{\rho}{m}; \tag{52}$$

3. *the variance of the third-moment estimator satisfies*

$$\mathrm{Var}(\widehat P_3^G) \;\le\; \frac{9}{4n} + \frac{4\rho}{m} + \frac{2\rho^2}{n\,d(d-1)}; \tag{53}$$

4. *for every $\theta \in \Delta_3$,*

$$\mathbb{E}_\theta\big(\widehat H_G - H(\theta)\big)^2 \le \min\Big\{(\log 3)^2,\; 2\sqrt{10/9}\,\pi^2\Big(\frac{2d-1}{2nd} + \frac{9}{4n} + \frac{5\rho}{m} + \frac{2\rho^2}{n\,d(d-1)}\Big)^{1/2}\Big\} \tag{54}$$

$$\le \min\Big\{(\log 3)^2,\; 47\Big(\frac{1}{n} + \frac{\rho}{m} + \frac{n\rho^2}{m^2}\Big)^{1/2}\Big\}. \tag{55}$$

The design needs neither connectivity nor an expansion property. At $d = n-1$, the complete graph, one has $m = N_2$, and (52) and (53) become the corresponding statements of Lemma B.1; the bracket in (54) is then exactly the quantity $v_n$ of Lemma B.1, so (54) reduces to the first inequality of Theorem B.4.

*Proof.* **Step 1 (unbiasedness).** From $\mathbb{E}[X_{ij} \mid Z] = W_{ij}$ we get $\mathbb{E}\widehat P_2^G = \mathbb{E}W_{12} = P_2$. For $(i;j,k) \in \Lambda(G)$ the pairs $\{i,j\} \ne \{i,k\}$ are two distinct edges of $G$ and are conditionally independent given $Z$, so $\mathbb{E}[X_{ij}X_{ik} \mid Z] = W_{ij}W_{ik} = \mathbf{1}\{Z_i = Z_j = Z_k\}$, whose expectation is $P_3$.

**Step 2 (the variance of $\widehat P_2^G$).** Use the law of total variance. For the conditional-variance part, given $Z$ the variables $\{X_e\}_{e\in E(G)}$ are independent, so $\mathbb{E}\mathrm{Var}(\widehat P_2^G \mid Z) = m^{-2}\sum_{e\in E(G)}\mathbb{E}v_e/\Delta^2 = \bar\rho/m$.

For the conditional-mean part, $\mathbb{E}[\widehat{P}_2^G \mid Z] = m^{-1}\sum_e W_e$, and the ordered pairs $(e, f)$ of edges split into three cases. There are $m$ pairs with $e = f$, of covariance $P_2(1-P_2)$; there are $2m(d-1)$ pairs sharing exactly one endpoint, since each endpoint of each edge carries $d-1$ further edges, of covariance $P_3 - P_2^2$; and a disjoint pair of edges depends on disjoint sets of labels and has covariance zero. Adding the two parts gives the equality in (52). Now use $P_2(1-P_2) \le 1/4$, $P_3 - P_2^2 = \mathrm{Var}(\theta_{Z_1}) \le 1/4$ (Lemma B.1) and $\bar{\rho} \le \rho$:

$$\mathrm{Var}(\widehat{P}_2^G) \;\le\; \frac{1/4 + (d-1)/2}{m} + \frac{\rho}{m} \;=\; \frac{2d-1}{4m} + \frac{\rho}{m} \;=\; \frac{2d-1}{2nd} + \frac{\rho}{m} \;\le\; \frac{1}{n} + \frac{\rho}{m}.$$

**Step 3 (decomposition and orthogonality).** Substituting $X_{ij}X_{ik} = W_{ij}W_{ik} + W_{ij}\xi_{ik} + \xi_{ij}W_{ik} + \xi_{ij}\xi_{ik}$ gives $\widehat{P}_3^G = L_G + M_G + Q_G$, where

$$L_G := \frac{1}{N_\Lambda}\sum_{(i;j,k)\in\Lambda(G)} W_{ij}W_{ik}, \qquad M_G := \frac{1}{N_\Lambda}\sum_{(i;j,k)\in\Lambda(G)} \big(W_{ij}\xi_{ik} + \xi_{ij}W_{ik}\big),$$
$$Q_G := \frac{1}{N_\Lambda}\sum_{(i;j,k)\in\Lambda(G)} \xi_{ij}\xi_{ik}$$

are the label term, the single-noise term and the double-noise term of $\widehat{P}_3^G$, defined as in Lemma B.1 with the sum taken over $\Lambda(G)$. Here $\mathbb{E}[M_G \mid Z] = 0$; every summand of $Q_G$ is a product of two conditionally independent mean-zero variables carried by two distinct edges of $G$, so $\mathbb{E}[Q_G \mid Z] = 0$; and $\mathbb{E}[M_G Q_G \mid Z]$ is a linear combination of terms $\mathbb{E}[\xi_e\xi_f\xi_g \mid Z]$ with $f \ne g$, each of which vanishes by the case analysis of Lemma B.1. Therefore

$$\mathrm{Var}(\widehat{P}_3^G) \;=\; \mathrm{Var}(L_G) + \mathbb{E}M_G^2 + \mathbb{E}Q_G^2. \tag{56}$$

**Step 4 (the label term).** The ordered pairs $(j,k)$ and $(k,j)$ give the same indicator, so $L_G = \frac{2}{N_\Lambda}\sum_{(i;\{j,k\})\in\Lambda_u(G)} \mathbf{1}\{Z_i = Z_j = Z_k\}$. Holding all labels but $Z_a$ fixed, only the unordered wedges containing $a$ can change: there are $\binom{d}{2}$ of them with $a$ at the centre and $d(d-1)$ with $a$ as a leaf, the centre being one of the $d$ neighbours of $a$ and the other leaf one of the remaining $d-1$ neighbours of that centre, hence $\frac{3}{2}d(d-1)$ in all. Each indicator takes values in $\{0,1\}$, so as a function of $Z_a$ the statistic $L_G$ takes values in an interval of length at most

$$\frac{2}{N_\Lambda}\cdot\frac{3}{2}d(d-1) \;=\; \frac{3d(d-1)}{n\,d(d-1)} \;=\; \frac{3}{n}.$$

Popoviciu's inequality gives $\mathrm{Var}_a(L_G) \le 9/(4n^2)$, where $\mathrm{Var}_a$ is the variance in $Z_a$ with the other labels held fixed, and the Efron–Stein inequality gives $\mathrm{Var}(L_G) \le 9/(4n)$. This bound does not depend on $d$.

**Step 5 (the single-noise term).** The set $\Lambda(G)$ is invariant under $j \leftrightarrow k$, so $M_G = \frac{2}{N_\Lambda}\sum_{(i;j,k)\in\Lambda(G)} W_{ij}\xi_{ik}$. Collect the terms by the edge $e = \{u,v\} \in E(G)$ that carries the noise: there are $d-1$ wedges with centre $u$ and leaf $v$, the third vertex running over $N_G(u)\setminus\{v\}$, and by symmetry $d-1$ wedges with centre $v$, so

$$M_G = \frac{2}{N_\Lambda}\sum_{e=\{u,v\}\in E(G)} c_e\,\xi_e, \qquad c_e := \sum_{j\in N_G(u)\setminus\{v\}} W_{uj} + \sum_{j\in N_G(v)\setminus\{u\}} W_{vj} \;\in\; [0,\; 2(d-1)].$$

Given $Z$ the $c_e$ are constants and the $\xi_e$ are independent with mean zero and variance at most $\rho$, so

$$\mathbb{E}[M_G^2 \mid Z] = \frac{4}{N_\Lambda^2}\sum_e c_e^2\,\frac{v_e}{\Delta^2} \;\le\; \frac{4}{N_\Lambda^2}\cdot m\cdot 4(d-1)^2\rho \;=\; \frac{16\,m\,\rho}{n^2d^2} \;=\; \frac{8\rho}{nd} \;=\; \frac{4\rho}{m},$$

the last two steps using $N_\Lambda = n\,d(d-1)$ and $m = nd/2$.

**Step 6 (the double-noise term).** Likewise $Q_G = \frac{2}{N_\Lambda}\sum_{(i;\{j,k\})\in\Lambda_u(G)} \xi_{ij}\xi_{ik}$. Two distinct edges share at most one endpoint, so the pair of edges $\{\{i,j\},\{i,k\}\}$ determines the centre $i$ and the leaves $\{j,k\}$, and distinct unordered wedges correspond to distinct pairs of edges; taking two distinct

wedges, some edge occurs exactly once in the product and the conditional expectation vanishes. Hence

$$\mathbb{E}[Q_G^2 \mid Z] = \frac{4}{N_\Lambda^2} \sum_{(i;\{j,k\})\in\Lambda_u(G)} \frac{v_{ij}\, v_{ik}}{\Delta^4} \;\le\; \frac{4}{N_\Lambda^2}\cdot\frac{N_\Lambda}{2}\cdot\rho^2 \;=\; \frac{2\rho^2}{N_\Lambda} \;=\; \frac{2\rho^2}{n\,d(d-1)}.$$

Combining Steps 3 to 6 through (56) gives (53).

**Step 7 (assembly).** The passage from the moment pair to the entropy in the proof of Theorem 2, namely the projection bound, the affine map to $(e_2, e_3)$ and the modulus of Lemma 1, uses only that $\widehat{P} = (\widehat{P}_2, \widehat{P}_3)$ is unbiased, that $P \in M_3$ and the geometry of $M_3$. None of this involves the observation design, so it applies verbatim to $(\widehat{P}_2^G, \widehat{P}_3^G)$:

$$\mathbb{E}_\theta\big(\widehat{H}_G - H(\theta)\big)^2 \;\le\; 2\sqrt{10/9}\,\pi^2\big(\mathrm{Var}(\widehat{P}_2^G) + \mathrm{Var}(\widehat{P}_3^G)\big)^{1/2},$$

and $\widehat{H}_G, H(\theta) \in [0, \log 3]$ gives the cap $(\log 3)^2$. Adding (52) and (53) yields $\frac{2d-1}{2nd} + \frac{9}{4n} + \frac{5\rho}{m} + \frac{2\rho^2}{nd(d-1)}$, which at $d = n-1$ is exactly the quantity $v_n$ of Lemma B.1; this is (54). Using $\frac{2d-1}{2nd} \le \frac{1}{n}$ bounds it by $\frac{13}{4n} + \frac{5\rho}{m} + \frac{2\rho^2}{nd(d-1)}$. Finally $d \ge 2$ gives $d(d-1) \ge d^2/2$, and $d = 2m/n$, so

$$\frac{2\rho^2}{n\,d(d-1)} \;\le\; \frac{4\rho^2}{n\,d^2} \;=\; \frac{n\rho^2}{m^2},$$

whence the bracket is at most $5\big(\frac{1}{n} + \frac{\rho}{m} + \frac{n\rho^2}{m^2}\big)$, and $2\sqrt{10/9}\,\pi^2\sqrt{5} = 46.5258\ldots \le 47$. □

*Remark* F.8 (a perfect matching identifies only $P_2$). Let $n$ be even and let $G$ be a perfect matching, so that $d = 1$. The edges of a matching have pairwise disjoint endpoints, so the pairs $\{(Z_i, Z_j)\}_{\{i,j\}\in E(G)}$ are mutually independent and the $W_{ij}$ are i.i.d. $\mathrm{Bern}(P_2)$; since the verdicts are conditionally independent given the labels, $\{A_{ij}\}_{\{i,j\}\in E(G)}$ is i.i.d. $\mathrm{Bern}(q+\Delta P_2)$. The observation distribution depends on $\theta$ only through $P_2$. Take $\theta_a = (1/2, 1/2, 0)$ and $\theta_b = (2/3, 1/6, 1/6)$: they share $P_2 = 1/2$ and their entropies differ by $\omega = \log 3 - \frac{4}{3}\log 2$, so under any matching design they induce the same observation distribution. For any estimator $\widehat{H}$, external randomization included, the two risks are then computed under one and the same distribution, and $x^2 + y^2 \ge (x-y)^2/2$ gives

$$\mathbb{E}\big(\widehat{H} - H(\theta_a)\big)^2 + \mathbb{E}\big(\widehat{H} - H(\theta_b)\big)^2 \;\ge\; \frac{\omega^2}{2},$$

so the worst-case risk is at least $\omega^2/4$. The condition $d \ge 2$ is the weakest one under which wedges exist in Corollary F.7, and hence under which $P_3$ is estimable.

*Remark* F.9 (saturation and the weak-separation window). The three terms of the last bound in (55) are equal at $m = n\rho$: there $\rho/m = 1/n$ and $n\rho^2/m^2 = 1/n$, so raising the budget above $n\rho$ no longer improves this bound beyond a constant factor. The budget is capped by $m \le N_2$, so saturation is reachable only if $\rho \le (n-1)/2$.

Turning this condition into a condition on $\Delta$ needs an extra assumption. Since $\rho = \max\{p(1-p),\, q(1-q)\}/\Delta^2$, one has $\rho \asymp \Delta^{-2}$ only when $q$ is fixed in the interior, that is, $q \in [q_0, 1-q_0]$ with $\Delta$ small, and then $\rho \le (n-1)/2$ is of the same order as $n\Delta^2 \gtrsim 1$, which falls outside the window $n\Delta^2 \le c_1$. Without that assumption saturation can occur inside the window: take $q = 0$, $n = k^3$ and $\Delta = k^{-2}$ with $k \ge 2$, so that $p(1-p)/\Delta^2 = k^2 - 1 = \rho$; the choice $d = 2(k^2-1) \le n-1$ gives $m = nd/2 = n\rho$ while $n\Delta^2 = 1/k \to 0$.

Inside the window, whether the bound is useful depends on the design and not on the window alone. At $q = 0.3$ and $\Delta = n^{-2/3}$, where $n\Delta^2 = n^{-1/3} \to 0$ and the window holds, the complete-graph design has $\rho \asymp n^{4/3}$ and $m \asymp n^2$, the bracket is of order $n^{-1/3}$ and the bound tends to zero like $n^{-1/6}$, in agreement with Theorem 2 since $n^3\Delta^4 = n^{1/3} \to \infty$. A fixed $d$ behaves in the opposite way: at the same $(q, \Delta)$ one has $n\rho^2/m^2 \asymp n^{5/3}/d^2 \to \infty$, and (55) reduces to the cap $(\log 3)^2$ once $n$ is large.

## G Numerical details

All experiments run on CPU with synthetic graphs drawn from (1), or, in Appendix G.4, from the departures of Appendix F.5. They use no external data or language model and illustrate the theorems.

Tables 2 and 3 give the standard errors of the mean squared errors behind the reported maxima; standard errors for the strong-signal control and the moving pair, and confidence intervals for the selected detection tests, are given below.

### G.1 Implementation

**The estimator.** The statistics (4) are computed from the edge count and the degree sequence, and membership in $M_3$ is tested with the three inequalities of Remark B.8. Inside $M_3$ the cubic of step 3 of Algorithm 1 is solved through the eigenvalues of its companion matrix. Its roots are real in exact arithmetic; the implementation takes their real parts, clips them to $[0,1]$ and rescales them to sum to one. Over $60{,}232$ points of $M_3$ with known $\theta$, including points within $10^{-7}$ of a double root or of a face of the simplex, the returned entropy is within $1.2\times10^{-8}$ of $H(\theta)$. Outside $M_3$, step 2 searches the two boundary curves of Remark B.7, each on a grid of $200{,}001$ equally spaced parameter values in $[0,1/2]$, locates the nearest grid point exactly with a $k$-d tree, and returns the entropy of the distribution that maps to it. The two curves are $m_A(t)=(1-2t+2t^2,\,1-3t+3t^2)$ and $m_B(a)=(1-4a+6a^2,\,1-6a+12a^2-6a^3)$; on $[0,1/2]$ their speeds satisfy $\|m_A'\|_2\le\sqrt{13}$ and $\|m_B'\|_2\le 2\sqrt{13}$, and the parameter spacing is $h=2.5\times10^{-6}$, so every point of either curve lies within $\sqrt{13}\,h<9.1\times10^{-6}$ of a grid point. The nearest-grid query therefore meets the condition of Remark B.7 with $\eta<9.1\times10^{-6}$, and the bound of Theorem B.4 grows by less than $\sqrt{10/9}\,\pi^2\cdot9.1\times10^{-6}<10^{-4}$. This bounds the discretization error of the grid in exact arithmetic, not the floating-point error of the cubic branch, which the check above measures. The fixed spacing adds at most $10^{-4}$, which is negligible at the sizes simulated here but does not vanish with $n$; an implementation used along a sequence $(n,\Delta_n)$ needs a spacing $h_n\le V_n^{1/2}/\sqrt{13}$, which gives $\eta_n\le V_n^{1/2}$ and keeps the extra term of Remark B.7 within a constant factor of the bound of Theorem 2.

**Random numbers.** Repetition $i\in\{0,\dots,399\}$ of the $c$-th combination of cell and distribution in the rate sweep and in the baseline comparison draws its graph from NumPy's default generator seeded with $1{,}000{,}003\,c+i$, the combinations being numbered from 1 in the order they run; the second family of the rate sweep numbers its combinations from $10{,}001$. In the detection battery, graph $k$ under hypothesis $h$, with $h=0$ for the null and $h=1$ for the alternative, uses NumPy's seed sequence $[20260916,\,r,\,n,\,\mathrm{round}(1000a),\,h,\,k]$, where $r=0$ for $\Delta=n^{-2/3}$ and $r=1$ for $\Delta=n^{-1/3}$; the graph is drawn from that sequence and the random start of the mean-field iteration from its first spawned child. Graphs $k=0,\dots,299$ form the fitting set, $k=300,\dots,599$ the held-out set, and, under the null, $k=600,\dots,899$ the calibration set.

### G.2 The rate sweep of Figure 3

The grid is laid out directly in $n^3\Delta^4$, because that is the variable the rate is stated in, and the sample sizes are chosen so that neighbouring values of $n$ share targets and the collapse across $n$ can be measured rather than assumed. Two families of cells:

- $n\in\{500,1000,2000,4000\}$, each taking four values of $n^3\Delta^4$ from the shared ladder $\{5,10,20,45,90,180,360\}$, shifted by one step as $n$ doubles so that consecutive sample sizes share three of their four targets. This keeps $n\Delta^2$ between $0.100$ and $0.300$ at every cell;
- $n\in\{2000,4000,8000\}$ with $n\Delta^2$ at $0.050$ and $0.030$, which pushes the ratio $\rho/n$ from $0.711$ up to $7.051$, and so raises the share of $V_n$ carried by the term $\rho^2/n^3$ from $0.228$ to $0.861$. Here $\rho/n$ is $\max\{p(1-p),q(1-q)\}/(n\Delta^2)$ and so slightly exceeds $q(1-q)/(n\Delta^2)$.

Throughout, $q=0.3$, and each of the 22 cells is run at five distributions with 400 repetitions each: the interior point $(0.5,0.3,0.2)$, the near-vertex point $(0.9,0.05,0.05)$, the boundary point $(0.8,0.2,0)$, and the two endpoints $\theta_\varepsilon$, $\theta'_\varepsilon$ of the least-favourable pair at that cell's amplitude $\varepsilon=a_0(n^3\Delta^4)^{-1/4}$ with $a_0=0.6$. The risk plotted for a cell is the largest of the five mean squared errors, and the error bar is the standard error of that particular mean squared error, not a standard error of the maximum, which would require a separate estimate of the selection effect.

Table 2: Every cell of the rate sweep. The risk is the largest mean squared error over the five distributions described above, and the last column names which of them attained it.

| $n$ | $n^3\Delta^4$ | $n\Delta^2$ | risk | s.e. | worst $\theta$ |
|---|---|---|---|---|---|
| 500 | 5 | 0.100 | 4.340e−2 | 2.7e−3 | $\theta_\varepsilon$ |
| 500 | 10 | 0.141 | 2.506e−2 | 1.7e−3 | $\theta_\varepsilon$ |
| 500 | 20 | 0.200 | 1.961e−2 | 1.6e−3 | near-vertex |
| 500 | 45 | 0.300 | 1.220e−2 | 1.1e−3 | near-vertex |
| 1000 | 10 | 0.100 | 2.447e−2 | 1.9e−3 | $\theta_\varepsilon$ |
| 1000 | 20 | 0.141 | 1.326e−2 | 1.1e−3 | near-vertex |
| 1000 | 45 | 0.212 | 9.532e−3 | 6.0e−4 | $\theta'_\varepsilon$ |
| 1000 | 90 | 0.300 | 7.624e−3 | 4.4e−4 | $\theta'_\varepsilon$ |
| 2000 | 1.8 | 0.030 | 5.092e−2 | 2.9e−3 | $\theta_\varepsilon$ |
| 2000 | 5 | 0.050 | 2.882e−2 | 2.1e−3 | $\theta_\varepsilon$ |
| 2000 | 20 | 0.100 | 1.159e−2 | 7.0e−4 | $\theta'_\varepsilon$ |
| 2000 | 45 | 0.150 | 6.981e−3 | 4.1e−4 | $\theta'_\varepsilon$ |
| 2000 | 90 | 0.212 | 6.127e−3 | 3.7e−4 | $\theta'_\varepsilon$ |
| 2000 | 180 | 0.300 | 4.643e−3 | 3.5e−4 | boundary |
| 4000 | 3.6 | 0.030 | 3.685e−2 | 2.4e−3 | $\theta_\varepsilon$ |
| 4000 | 10 | 0.050 | 1.437e−2 | 1.0e−3 | $\theta_\varepsilon$ |
| 4000 | 45 | 0.106 | 6.497e−3 | 3.7e−4 | $\theta'_\varepsilon$ |
| 4000 | 90 | 0.150 | 5.315e−3 | 3.7e−4 | boundary |
| 4000 | 180 | 0.212 | 3.984e−3 | 3.1e−4 | boundary |
| 4000 | 360 | 0.300 | 3.364e−3 | 2.5e−4 | boundary |
| 8000 | 7.2 | 0.030 | 2.100e−2 | 1.5e−3 | $\theta_\varepsilon$ |
| 8000 | 20 | 0.050 | 8.527e−3 | 5.1e−4 | $\theta'_\varepsilon$ |

Table 2 lists every cell. Over all 22, the ratio of measured risk to $V_n^{1/2}$ runs from $0.1391$ to $0.3475$; over the ten cells with $n^3\Delta^4 \geq 45$ it runs from $0.1391$ to $0.1813$ with median $0.1564$, which is the line drawn in Figure 3(b). At the targets 90 and 180 the point estimates of that ratio have relative ranges, maximum over minimum minus one, of about $2.2\%$ and $4.4\%$; the risks behind them carry relative standard errors of about $5.8$ to $7.8\%$, so these comparisons do not resolve differences at the $5\%$ scale. The corresponding ratio to the limiting form $(n^3\Delta^4)^{-1/2}$ varies by a factor of $1.88$ on the same ten cells, against $1.30$ for the ratio to $V_n^{1/2}$. All 22 cells have $n\Delta^2$ between $0.030$ and $0.300$, above the $c_1 = 0.006375$ with which the lower bounds are proved at $q_0 = 0.3$; since $n^3\Delta^4 = n(n\Delta^2)^2$, the condition $n\Delta^2 \leq c_1$ caps $n^3\Delta^4$ at $0.3251$ for $n \leq 8000$, inside the saturated branch.

### G.3 The baseline comparison of Table 1

$\Delta = n^{-2/3}$, $n \in \{500, 2000\}$, $q \in \{0.1, 0.3, 0.45\}$, four distributions (interior, uniform, near-vertex, boundary), 400 repetitions per combination. The spectral baseline centres the symmetric adjacency matrix by subtracting its mean entry, takes the eigenvectors of the three largest algebraic eigenvalues of the result, ordered by decreasing eigenvalue, scales each by the square root of the absolute value of its eigenvalue, runs k-means with ten restarts and keeps the one with the smallest within-cluster sum of squares, and plugs the cluster proportions into the entropy. The restarts are there so that the comparison measures the signal rather than one k-means initialization. The spectral-$K$ variant is the same procedure given the true number of meanings, which is 2 on the boundary distribution and 3 elsewhere; at $K = 2$ it clusters the leading coordinate alone. The Laplacian-$K$ baseline is the construction of Liu (2026): it forms the unnormalized Laplacian $D - A$ of the verdict matrix, takes the eigenvectors of its $K$ smallest eigenvalues, runs the same k-means with ten restarts on their rows, and plugs in the cluster proportions. It is given the same true $K$ as spectral-$K$. Like the moment estimator, the degree-CL baseline never groups. It maximizes over $\Delta_3$, approximately and by the grid search described below, the composite log-likelihood $\sum_i \log \sum_c \theta_c\, b(d_i; n-1, q+\Delta\theta_c)$, which treats the degrees as independent draws from the law each degree has on its own, with $(q, \Delta)$ known as for the moment estimator, and plugs the maximizer into the entropy; here $b(\cdot\,; N, s)$ is the binomial probability function. The degrees are not independent, so this is a composite likelihood rather than the likelihood of the graph. The criterion is symmetric under relabelling, so the search runs over

sorted $\theta$: a grid of spacing $0.01$, then two refinements of spacing $5 \times 10^{-4}$ and $2.5 \times 10^{-5}$ around each of the three best grid points. It is evaluated on the same graphs as the other estimators.

The triplet composite likelihood of Ambroise & Matias (2012), triplet-CL below, never groups either. For the binary affiliation model they maximize the sum, over the triplets of answers, of the log-probability of the triplet's three verdicts, a mixture over the labels of its answers. With $(q, \Delta)$ known, that law depends on $\theta$ only through $(P_2, P_3)$: the three labels agree with probability $P_3$, a given two agree and the third differs with probability $P_2 - P_3$ for each of the three choices of the two, and all three differ with probability $1 - 3P_2 + 2P_3$. The three verdicts of a triplet are then exchangeable, so the criterion reads the graph only through the numbers of triplets carrying zero to three agreements, that is, through the edge count, the number of paths of length two, which the squared degrees determine, and the triangle count. We maximize it over $\Delta_3$ by the grid search of degree-CL, on the same graphs, and plug the maximizer into the entropy. The consistency theorem of Ambroise & Matias (2012) holds at fixed parameters and assumes identifiability and a compact parameter space, with every proportion bounded away from zero and both connection probabilities bounded away from zero and one; it does not cover a vanishing separation.

Table 3 gives the standard error of every entry of Table 1, each that of the mean squared error attaining the maximum over the four distributions, as in Table 2. At the precision of the reported values, the largest standard error relative to its entry is about $9.7\%$ for the moment estimator, $8.2\%$ for degree-CL and $7.6\%$ for the oracle, and below $0.6\%$ for the three clustering baselines; the constant has no sampling error because its output is fixed. Every ratio between the moment estimator and a clustering baseline or the constant quoted in Section 6, the smallest being $10.4$, is therefore far outside sampling error. The comparison with degree-CL needs the paired standard error instead: both estimators run on the same graphs and their errors are positively correlated, so combining the two marginal standard errors would overstate the uncertainty of their difference. On the paired scale the two trade off across distributions. Degree-CL has the lower risk on all six near-vertex cells, by $1.2$ to $4.2$ paired standard errors, and the higher risk on $14$ of the other $18$ cells, $11$ of them by more than two. The worst risk of the moment estimator falls on the near-vertex distribution in five of the six settings, which is why degree-CL has the lower worst risk in those five. Where both worst risks fall on the same distribution, in the first four rows of Table 1, their difference is $1.60$, $4.22$, $3.11$ and $4.67$ paired standard errors, the first three in favour of degree-CL; these are standard errors of a difference at a fixed distribution, not of the difference between two maxima.

Triplet-CL behaves like degree-CL (Table 4). Its worst risk is $0.79$ to $1.17$ times that of the moment estimator and lower in five of the six settings; on the paired scale it has the lower risk on all six near-vertex cells, by $2.5$ to $4.8$ paired standard errors, and the higher risk on $13$ of the other $18$ cells, $8$ of them by more than two. The two composite likelihoods read different statistics, the whole degree sequence against the triplet counts, which besides the triangles are the two statistics of Algorithm 1. What they share is a likelihood criterion maximized over $\Delta_3$ in place of solving for two moments. The optimality proved here concerns the rate, not the constant.

The data-free constant outputs $\frac{1}{2}\log 3 = 0.5493$, close to the entropy $0.5004$ of the boundary distribution $(0.8, 0.2, 0)$. There its mean squared error, $2.392 \times 10^{-3}$, is below those of the three estimators that never group in all six settings of Table 1, which run from $6.508 \times 10^{-3}$ to $2.216 \times 10^{-2}$; on the uniform distribution it is $3.017 \times 10^{-1}$, the maximum reported in Table 1.

Table 3: Standard errors of the entries of Table 1. Each is the standard error of the mean squared error that attains the maximum over the four distributions, from $400$ repetitions, not a standard error of the maximum.

| $n$ | $q$ | $n\Delta^2$ | moment | degree-CL | spectral | spectral-$K$ | Laplacian-$K$ | constant | oracle |
|---|---|---|---|---|---|---|---|---|---|
| 500 | 0.10 | 0.126 | 9.8e−4 | 9.3e−4 | 4.6e−4 | 4.6e−4 | 1.2e−3 | 0 | 9.3e−5 |
| 500 | 0.30 | 0.126 | 2.1e−3 | 1.5e−3 | 2.5e−4 | 2.5e−4 | 3.2e−3 | 0 | 1.1e−4 |
| 500 | 0.45 | 0.126 | 2.1e−3 | 1.7e−3 | 2.0e−4 | 2.0e−4 | 5.9e−3 | 0 | 1.1e−4 |
| 2000 | 0.10 | 0.079 | 4.6e−4 | 5.1e−4 | 8.4e−5 | 8.4e−5 | 1.9e−4 | 0 | 2.8e−5 |
| 2000 | 0.30 | 0.079 | 1.1e−3 | 7.5e−4 | 7.0e−5 | 7.0e−5 | 2.1e−4 | 0 | 2.5e−5 |
| 2000 | 0.45 | 0.079 | 1.3e−3 | 7.6e−4 | 6.3e−5 | 6.3e−5 | 1.9e−4 | 0 | 2.7e−5 |

Table 4: The three estimators that never group, at the settings of Table 1: the largest mean squared error over the four distributions, with the standard error of that entry in parentheses. The first two columns repeat Table 1; triplet-CL is the triplet composite likelihood of Ambroise & Matias (2012), given the channel.

| $n$ | $q$ | moment | degree-CL | triplet-CL |
|---|---|---|---|---|
| 500 | 0.10 | 1.226e−2 (9.8e−4) | 1.134e−2 (9.3e−4) | 1.118e−2 (9.2e−4) |
| 500 | 0.30 | 2.896e−2 (2.1e−3) | 2.232e−2 (1.5e−3) | 2.282e−2 (1.6e−3) |
| 500 | 0.45 | 2.728e−2 (2.1e−3) | 2.236e−2 (1.7e−3) | 2.186e−2 (1.7e−3) |
| 2000 | 0.10 | 6.508e−3 (4.6e−4) | 7.453e−3 (5.1e−4) | 7.621e−3 (5.0e−4) |
| 2000 | 0.30 | 1.165e−2 (1.1e−3) | 9.494e−3 (7.5e−4) | 9.907e−3 (7.7e−4) |
| 2000 | 0.45 | 1.342e−2 (1.3e−3) | 1.077e−2 (7.6e−4) | 1.112e−2 (7.4e−4) |

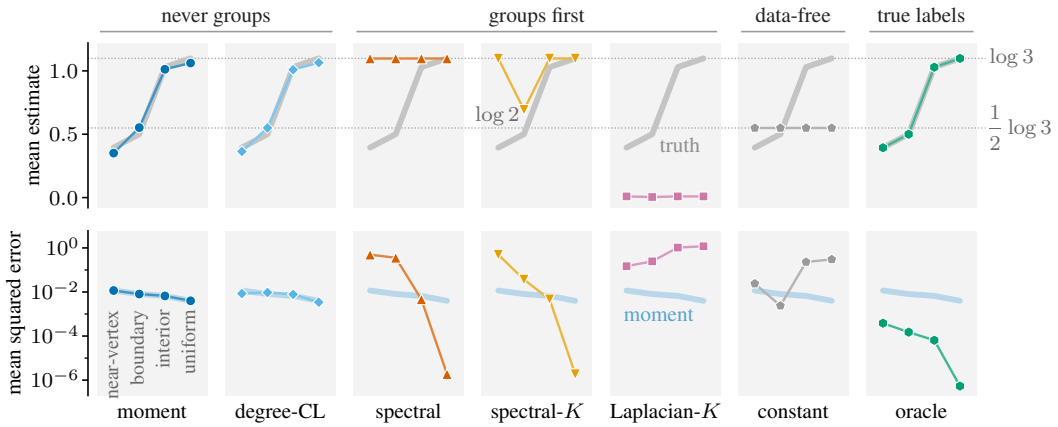



Figure 5: The six estimators and the oracle at $n = 2000$, $q = 0.3$, $\Delta = n^{-2/3}$, one column each. Within a column the four distributions appear in order of true entropy: near-vertex $(0.9, 0.05, 0.05)$, boundary $(0.8, 0.2, 0)$, interior $(0.5, 0.3, 0.2)$ and uniform $(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$. **Top row:** mean output, with the true entropy in grey. The spectral baseline sits near $\log 3$ on every distribution, which is what the absence of signal looks like after a forced three-way split, and spectral-$K$, given the two meanings of the boundary distribution, sits near $\log 2$ there; the Laplacian baseline collapses onto a single block and reports almost zero. **Bottom row:** mean squared error on a logarithmic scale, with that of the moment estimator in pale blue in every column. A clustering baseline or the constant beats the moment estimator only where its nearly fixed output happens to sit near the truth: $\log 3$ for the spectral baselines on the uniform and interior distributions, and $\frac{1}{2}\log 3$ for the constant on the boundary one. Degree-CL, which never groups, stays close to the moment estimator on all four, below it on the uniform and near-vertex distributions and above it on the other two.

**Strong-signal control.** $n = 500$, $q = 0.1$ and $\Delta = 0.8$, far outside the weakly separated regime, with 400 repetitions per distribution; each mean squared error below is followed by its standard error. Spectral-$K$, Laplacian-$K$ and the oracle agree to the printed precision, in mean output as well, on the interior distribution, $2.846 \times 10^{-4}$ ($2.1 \times 10^{-5}$), the uniform one, $9.268 \times 10^{-6}$ ($1.0 \times 10^{-6}$), and the boundary one, $6.066 \times 10^{-4}$ ($4.4 \times 10^{-5}$); on the boundary distribution, which has two meanings, spectral-$K$ clusters two groups in one dimension. The spectral baseline, which forces three blocks, agrees on the interior and uniform distributions and reaches $3.049 \times 10^{-1}$ ($6.8 \times 10^{-4}$) on the boundary one. On the near-vertex distribution spectral and spectral-$K$ give $2.063 \times 10^{-3}$ ($6.4 \times 10^{-4}$), Laplacian-$K$ gives $5.145 \times 10^{-3}$ ($8.7 \times 10^{-4}$) and the oracle $1.447 \times 10^{-3}$ ($1.0 \times 10^{-4}$), risk ratios of about 1.4 and 3.6. The two excesses over the oracle's mean squared error are about 0.96 and 4.25 times the standard error of the corresponding baseline's mean squared error; these use the marginal standard errors, not those of paired differences. Given a strong signal and the true number of meanings, the implementations therefore reach the oracle's risk on three distributions and stay within a factor of 3.6 of it on the fourth, whereas in Table 1 every clustering entry is about 300 to 3200 times the oracle's. Degree-CL gives $2.868 \times 10^{-4}$ ($2.1 \times 10^{-5}$), $5.172 \times 10^{-6}$ ($8.4 \times 10^{-7}$), $1.477 \times 10^{-3}$ ($1.1 \times 10^{-4}$) and $6.404 \times 10^{-4}$ ($4.7 \times 10^{-5}$) on the interior, uniform, near-vertex and boundary

distributions, at most 1.06 times the oracle's risk, triplet-CL gives $2.891 \times 10^{-4}$ ($2.1 \times 10^{-5}$), $4.488 \times 10^{-6}$ ($7.0 \times 10^{-7}$), $1.460 \times 10^{-3}$ ($1.0 \times 10^{-4}$) and $6.232 \times 10^{-4}$ ($4.5 \times 10^{-5}$), at most 1.03 times, and the moment estimator gives $2.906 \times 10^{-4}$ ($2.1 \times 10^{-5}$), $4.500 \times 10^{-6}$ ($7.1 \times 10^{-7}$), $1.466 \times 10^{-3}$ ($1.0 \times 10^{-4}$) and $6.221 \times 10^{-4}$ ($4.6 \times 10^{-5}$).

**The moving least-favourable pair.** At $q = 0.3$ and amplitude $a_0 = 0.6$, with 400 repetitions per endpoint: at $n = 500$ the pair is $\varepsilon = 0.3575$ against $\varepsilon' = 0.2949$, true entropy gap 0.1590, estimator mean outputs 0.7559 and 0.7570 with standard errors 0.0074 and 0.0077, difference 0.0011 with standard error 0.011; at $n = 2000$ it is $\varepsilon = 0.3185$ against $\varepsilon' = 0.2729$, gap 0.1496, means 0.7079 and 0.7083 with standard errors 0.0047 and 0.0043, difference 0.0004 with standard error 0.0064. In both cases the difference is below one standard error, while the entropy gap is more than 14 of them. Degree-CL separates the endpoints slightly more: its mean outputs differ by 0.0163 at $n = 500$ and 0.0149 at $n = 2000$, which are 1.6 and 2.1 standard errors and about a tenth of the entropy gap, and triplet-CL by 0.0212 and 0.0125, which are 2.0 and 1.8 standard errors. Within its parameter range, the proof of Theorem 4 bounds the total variation between the endpoints by $1/2$, which allows unequal mean outputs; these amplitudes exceed its limit $\varepsilon \leq 1/12$.

### G.4 Verdicts that share an answer

Proposition F.6 gives the bias of $\widehat{P}_3$ exactly; this subsection measures what it does to the risk, at the settings of Table 1. The three departures of Appendix F.5 are driven by signs $\sigma_i = \pm 1$, equally likely and independent of the labels: $\beta_i = s\sigma_i$ for additive tendencies and $a_i = 1 + \tau\sigma_i$ for the two multiplicative effects, so that $\mathrm{Var}\beta_i = s^2$ and $\mathrm{Var}a_i = \tau^2$. The levels are $s \in \{0.001, 0.003, 0.01\}$, $\tau \in \{0.1, 0.3\}$ on the separation and $\tau \in \{0.001, 0.003, 0.01\}$ on the whole rate. Repetition $i$ of a combination reuses the labels and the uniform variables of the graph with the same seed in the baseline comparison and draws the signs from NumPy's default generator seeded with the pair formed by that seed and 1; every level thresholds the same uniforms, so the conditionally independent level reproduces the moment column of Table 1 and every change in risk is measured on paired draws.

Over the 216 combinations of level, setting and distribution, the mean of $\widehat{P}_3 - P_3$ lies within 2.3 standard errors of $\zeta/\Delta^2$ and the mean of $\widehat{P}_2 - P_2$ within 2.3 of zero, and on the same draws the shift of $\widehat{P}_3$ away from conditional independence tracks $\zeta/\Delta^2$ more closely still. Table 5 shows both, and the consequence for the risk, at $q = 0.3$. Additive tendencies matter most, and more so as $n$ grows, because $\Delta = n^{-2/3}$ shrinks: $s = 0.003$ adds 0.036 to $\widehat{P}_3$ at $n = 500$ and 0.227 at $n = 2000$, where it raises the worst risk from $1.165 \times 10^{-2}$ to $5.982 \times 10^{-2}$. At $s = 0.01$ and $n = 2000$ the bias of 2.52 puts the moment pair where its projection onto $M_3$ is the vertex $(1, 1)$ in every repetition, so the estimate is 0 and the worst risk is $(\log 3)^2$, attained on the uniform distribution. Effects of size $\tau = 0.3$ on the separation, whose bias $\tau^2 P_3$ is not amplified, raise the worst risk only to $1.972 \times 10^{-2}$, while effects on the whole rate act like additive tendencies with $s \approx \tau q$, since $\zeta/\Delta^2 \approx \tau^2 q^2/\Delta^2$. At $q = 0.1$ and 0.45 the pattern is the same, with the effects on the whole rate weaker at the first and stronger at the second, as $s \approx \tau q$ predicts.

Table 5: Dependent verdicts at $q = 0.3$ and the settings of Table 1: the exact bias $\zeta/\Delta^2$ of $\widehat{P}_3$, the measured shift of $\widehat{P}_3$ from conditional independence on the same draws at the interior distribution $(0.5, 0.3, 0.2)$, and the largest mean squared error of Algorithm 1 over the four distributions, 400 repetitions each, with the standard error of that entry in parentheses. The first row is conditional independence and repeats the moment column of Table 1.

| | | $n = 500$ | | | $n = 2000$ | | |
|---|---|---|---|---|---|---|---|
| departure | level | $\zeta/\Delta^2$ | shift of $\widehat{P}_3$ | worst risk | $\zeta/\Delta^2$ | shift of $\widehat{P}_3$ | worst risk |
| independent | | 0.0000 | 0.0000 | 2.896e−2 (2.1e−3) | 0.0000 | 0.0000 | 1.165e−2 (1.1e−3) |
| additive, $s$ | 0.001 | 0.0040 | 0.0046 | 2.966e−2 (2.1e−3) | 0.0252 | 0.0254 | 1.426e−2 (1.3e−3) |
| | 0.003 | 0.0357 | 0.0378 | 3.453e−2 (2.3e−3) | 0.2268 | 0.2281 | 5.982e−2 (2.8e−3) |
| | 0.01 | 0.3969 | 0.4027 | 1.068e−1 (3.0e−3) | 2.5198 | 2.5218 | 1.207e+0 (0) |
| separation, $\tau$ | 0.1 | 0.0016 | 0.0020 | 2.965e−2 (2.1e−3) | 0.0016 | 0.0017 | 1.227e−2 (1.1e−3) |
| | 0.3 | 0.0144 | 0.0152 | 3.972e−2 (2.5e−3) | 0.0144 | 0.0143 | 1.972e−2 (1.6e−3) |
| whole rate, $\tau$ | 0.001 | 0.0004 | 0.0004 | 2.911e−2 (2.1e−3) | 0.0023 | 0.0022 | 1.180e−2 (1.0e−3) |
| | 0.003 | 0.0033 | 0.0040 | 2.959e−2 (2.1e−3) | 0.0207 | 0.0210 | 1.378e−2 (1.2e−3) |
| | 0.01 | 0.0372 | 0.0390 | 3.503e−2 (2.3e−3) | 0.2304 | 0.2319 | 6.237e−2 (2.9e−3) |

### G.5 The detection battery

The two endpoints of the least-favourable family at $q = 0.3$, $\Delta = n^{-2/3}$ and $\varepsilon = a\, n^{-1/12}$ are compared by 17 tests built from 23 features: the first four moments of the standardized degree sequence and its empirical variance, skewness and excess kurtosis, the two moment estimates $\widehat{P}_2$ and $\widehat{P}_3$, a mixture log-likelihood ratio between the two hypotheses, signed triangle and signed four-cycle counts, the five largest and the five smallest eigenvalues of the adjacency matrix centred at its expected edge probability and divided by $\sqrt{n\, r(1-r)}$ with $r = q + \Delta P_2$, and a mean-field evidence-bound difference. Fifteen of these features are tested one at a time; the eight interior eigenvalues enter only through the two multivariate tests. The grid is $n \in \{500, 1000, 2000, 4000, 8000\}$ and $a \in \{0.1, 0.2, 0.4, 0.8\}$, and each cell simulates three independent sets of graphs: 300 per hypothesis for fitting, 300 per hypothesis held out, and a further 300 under the null for calibration. A scalar statistic takes its rejection direction from the fitting set and its level-0.05 critical value from the calibration set; the two multivariate tests, a logistic regression and a gradient-boosted tree ensemble, are fitted together with their standardizer on the fitting set, frozen, and then thresholded on the same calibration set. Power and balanced accuracy are reported on the held-out set, which no fitting or thresholding step has seen. The battery shrinks with $n$ for cost reasons: 17 tests up to $n = 2000$, 15 at $n = 4000$ after dropping the signed triangle and four-cycle counts, and 12 at $n = 8000$ after also dropping the eigenvalue and mean-field features. The positive control repeats the grid at $\Delta = n^{-1/3}$, at all four amplitudes for $n \le 1000$ and at $a = 0.1$ above, which is what makes 31 cells in all. Across those cells and the 492 test instances they contain, the mean realized level is 0.0496, with a minimum of 0.007 and a maximum of 0.113.

**Features and tests.** The mixture log-likelihood ratio treats the degrees as independent draws from the law $\sum_c \theta_c \,\mathrm{Bin}(n-1, q + \Delta\theta_c)$ that each degree has on its own, and takes the difference of that log-likelihood between the two hypotheses. For each hypothesis the evidence-bound feature evaluates the naive mean-field lower bound

$$\mathcal{F}_\theta(R) := \sum_{i,\, c:\theta_c>0} R_{ic} \log \frac{\theta_c}{R_{ic}} + \sum_{i<j} J_{ij} \sum_{c:\theta_c>0} R_{ic} R_{jc}$$

on the log-likelihood of the graph, up to a term the two hypotheses share, where each row of $R$ is a probability vector on the support of $\theta$ and $J_{ij}$ are the pair weights of Appendix D. Starting from $R_{ic} = \theta_c$ it iterates

$$R_{ic}^{(t+1)} = \tfrac{1}{2} R_{ic}^{(t)} + \tfrac{1}{2} \frac{\theta_c \exp\big(\sum_{j\neq i} J_{ij} R_{jc}^{(t)}\big)}{\sum_{d:\theta_d>0} \theta_d \exp\big(\sum_{j\neq i} J_{ij} R_{jd}^{(t)}\big)}$$

until $\max_{i,c} |R_{ic}^{(t+1)} - R_{ic}^{(t)}| < 10^{-6}$ or for 500 updates; for the three-class hypothesis a second run starts from rows proportional to $\theta_c \exp(0.5\,\xi_{ic})$ with independent $\xi_{ic} \sim N(0,1)$. The feature

is the larger terminal value of the two alternative runs minus the terminal value of the null run; these are values of the bound at feasible points, not certified maxima. Both classifiers standardize each feature by its mean and its population standard deviation over the 600 fitting graphs, adding $10^{-12}$ to the latter. The logistic regression has an $\ell_2$ penalty of inverse strength 1 and at most 5000 iterations; the tree ensemble is scikit-learn's histogram-based gradient boosting with 200 iterations, at most 15 leaves per tree, learning rate 0.05, no early stopping and random seed 0, its other settings at the library defaults. A test rejects when its score strictly exceeds the 286th smallest of the 300 null calibration scores; for fixed fitted scores, exchangeability bounds the null rejection probability, averaged over the calibration draw, by $15/301 < 0.05$. A scalar statistic is multiplied by the sign of its alternative-minus-null mean difference on the fitting set, taken as $+1$ at zero. For balanced accuracy a classifier declares the alternative at probability at least $1/2$. For a scalar statistic, let $v_1 < \cdots < v_m$ be the distinct fitting scores after the sign change; among the candidates $v_1 - 1$, $(v_1 + v_2)/2, \ldots, (v_{m-1} + v_m)/2$ and $v_m + 1$, taken in increasing order, the threshold is the first that maximizes balanced accuracy on the fitting set, a score strictly above it declaring the alternative.

On the weak-separation grid the largest observed power is 0.097 for $a \le 0.2$ and 0.223 over all amplitudes. For $n = 500, 1000, 2000, 4000, 8000$ the largest power in a cell is 0.067, 0.083, 0.070, 0.070, 0.067 at $a = 0.1$, and 0.183, 0.143, 0.170, 0.163, 0.223 at $a = 0.8$. The positive control puts the same pair at $\Delta = n^{-1/3}$, outside the window, where the per-answer signal grows with $n$; there the largest powers at $a = 0.1$ are 0.810, 0.990, 1.000, 1.000 and 0.983, the last from the reduced feature set.

**Intervals.** Once a test is fitted and thresholded, its number of held-out rejections is binomial with 300 trials, and the intervals below are two-sided 95% Clopper–Pearson intervals unless stated otherwise. On the weak-separation grid no test at $a \le 0.4$ has its power interval entirely above its own level interval; at $a = 0.8$ the numbers of tests that do are 6, 7, 7, 7, 7 in increasing order of $n$. These marginal comparisons establish neither equal power and level in the other cells nor a power that is constant in $n$. The largest power for $a \le 0.2$, 0.097 with interval $[0.066, 0.136]$, belongs to the gradient-boosted ensemble at $a = 0.2$, $n = 500$, whose realized level is 0.073 with interval $[0.047, 0.109]$. The largest power on the grid, $67/300 \approx 0.223$ with interval $[0.177, 0.275]$, belongs to the second moment of the standardized degrees at $a = 0.8$, $n = 8000$, whose realized level is $18/300 = 0.060$ with interval $[0.036, 0.093]$. In the positive control at $a = 0.1$, $n = 2000$, the evidence-bound feature has power 1.000 with interval $[0.988, 1.000]$ at realized level 0.050 with interval $[0.028, 0.081]$. For the grid maximum, intervals at error probability $0.05/12$ over the 12 tests of its cell are $[0.159, 0.299]$ and $[0.028, 0.110]$; to cover selection over all 492 test instances, error probability $0.05/984$ for each of their 984 power and level intervals gives $[0.136, 0.332]$ and $[0.019, 0.134]$, still disjoint. That test therefore has power above its level even under this simultaneous adjustment, whereas the counts of six and seven are not simultaneous discoveries, and every maximum quoted in this subsection is a selected point estimate that may be biased upward. The weak-separation cells have $n\Delta^2 = n^{-1/3}$ between 0.050 and 0.126, above the $c_1 = 0.006375$ with which the lower bounds are proved at $q_0 = 0.3$, and for $a \ge 0.2$ they have $\varepsilon > 1/12$, beyond the amplitudes that the proof of Theorem 4 allows. They follow the exponent of the least-favourable family but not the constants of the theorem.

### G.6 The explicit constants at finite $n$

Theorem 2 is non-asymptotic, so its bound can be compared at every $n$ with the trivial benchmark $(\log 3)^2/4 = 0.3017$, the squared error of the midpoint of the range of $H$. At a fixed channel the crossover is at $n = 15{,}454$ for the sharper uncapped expression $2\sqrt{10/9}\,\pi^2 v_n^{1/2}$ of Theorem B.4 at $\rho = 0$, and at $n = 72{,}063$ for the weakened form $81V_n^{1/2}$; those two numbers move by less than 0.5% across the channels $(q, \Delta) \in \{(0.3, 0.6), (0.3, 0.3), (0.3, 0.1), (0.1, 0.3)\}$.

Along a shrinking separation, which is the case this paper treats, the crossover is much later: at $\Delta_n = n^{-1/2}$ and $q = 0.3$ it is $n = 25{,}986$ for the uncapped expression and $n = 90{,}510$ for $81V_n^{1/2}$, and at $\Delta_n = n^{-2/3}$, which is inside the exponent window where the rate is matched, it is $n = 86{,}394{,}886$ at $q = 0.3$ and $n = 1{,}536{,}966$ at $q = 0.1$, with the weakened form later still. Below those sizes, and so at every size simulated here, the explicit bound exceeds the benchmark, though not necessarily the cap: at half the largest crossover, $n = 43{,}197{,}443$, the two expressions evaluate

to $0.341$ and $0.914$, both under the cap. At all $22$ cells of Table 2, $81 V_n^{1/2}$ exceeds the cap $(\log 3)^2$, which is $24$ to $359$ times the measured risk. By contrast, Figure 3 measures the coefficient at $0.156$ against $81$, a factor of about $520$: the constant of Theorem 2 is conservative, and its rate is tight.