

HOMOMORPHIC-CORE PHASE TRANSITION THRESHOLD IN ERDŐS-RÉNYI RANDOM GRAPHS

JIAHENG WANG

ABSTRACT. It is shown in this manuscript that a random graph G drawn from the Erdős-Rényi model $\mathcal{G}(n, p)$ with

$$p = p(n) \leq 1/2, \quad \lim_{n \rightarrow +\infty} (np - \log n - \log \log n) = +\infty,$$

is a homomorphic core, i.e., every homomorphism from G to itself is an automorphism. This implies tight ETH-based lower bounds of the subgraph isomorphism problem for almost all k -vertex patterns with polynomial average degree.

DISCLOSURE OF AI TOOL USAGE. The core¹ idea of the proof was found by ChatGPT 5.6 Sol, while the manuscript is written by the *homo sapiens* author in a traditional manner. The author shall *forgo* claiming intellectual property of the key technical contribution in this manuscript. The author, however, takes full responsibility of endorsing the correctness and proper citations.

The total number of hours that the model worked on this task was approximately 7 hours, summing up the cost all interactions. The model ran at xHigh and Max levels. This information is disclosed on the basis of potential concern with fair and eco-friendly usage of AI tools.

The disclosure is intentionally put before the main text of this manuscript for visibility.

1. INTRODUCTION

1.1. Motivation: the complexity of homomorphisms. Graph homomorphisms [Lov12, HN04] play a fundamental role in theoretical computer science because they provide a uniform language for expressing structure-preserving assignments. Many seemingly unrelated problems, including graph colouring, database-query evaluation, and constraint satisfaction problems, can be formulated as asking whether a graph homomorphism exists and how many such homomorphisms there are [CM77, Sch78, FV98, KV00, FGRZ25]. This motivates the study of the complexity of deciding whether graph homomorphisms exist and of counting them. Recall that a graph homomorphism from G to H is a map from $V(G)$ to $V(H)$ that preserves adjacency. We denote by $\# \text{hom}(G \rightarrow H)$ the number of homomorphisms from G to H .

We focus on homomorphism problems *from the left*, in which the host graph H is the input. The task $\text{HOM}(G)$ asks whether such a homomorphism exists, and $\# \text{HOM}(G)$ asks how many exist. These problems are naturally *parameterised* by G . By contrast, when H is fixed and G is the input, complexity dichotomies are typically achievable: depending on whether the host graph H satisfies certain criteria, the problems are either solvable in polynomial time or hard with respect to their complexity classes (**NP**-hard, $\# \mathbf{P}$ -hard, $\mathbf{modP}_p$ -hard); see for example [HN90, DG00, CC17, GGJ17, GLS21, BK22, Bul17, Zhu20].

On n -vertex input graphs, $\# \text{HOM}(G)$ can always be solved in time $n^{\Theta(|V(G)|)}$ by brute-force enumeration. For certain graphs, this running time is optimal. If G is a k -clique, then $\text{HOM}(G)$ (respectively, its counting variant) is equivalent to the *subgraph isomorphism problem* $\text{SUB}(G)$, which asks whether the input graph contains a subgraph isomorphic to G (not necessarily an induced

(Jiaheng Wang) UNIVERSITY OF HELSINKI, FINLAND. Financially supported by HELSINKI INSTITUTE OF INFORMATION TECHNOLOGY (HIIT).

¹This is a pun. The author is aware that explaining a joke is the best way to ruin the joke, however.

one). The Exponential-Time Hypothesis (ETH) [IP01, IPZ01] implies an $n^{\Omega(k)}$ -time lower bound for SUB(G) when G is a k -clique [CCF⁺05, CHKX06].

Now consider the case in which G is the k -by- k biclique $K_{k,k}$ (the complete bipartite graph). Although the counting problem $\# \text{HOM}(K_{k,k})$ is intractable under standard complexity assumptions [FG04, DJ04], its decision counterpart behaves very differently. In fact, the decision problem $\text{HOM}(K_{k,k})$ is trivial. Indeed, if H contains at least one edge uv , then there is a homomorphism from $K_{k,k}$ to H : map every vertex in its left part to u and every vertex in its right part to v . Otherwise, H has no edges, and no homomorphism from $K_{k,k}$ to H exists.

This argument extends to broader graph families. Suppose that a pattern graph G has a *proper* subgraph G' and that there exists a homomorphism f from G to G' . If the host graph H contains a homomorphic copy of G , then restricting this copy to G' yields a homomorphic copy of G' . Conversely, if H contains a homomorphic copy of G' through a mapping g , then the composition $g \circ f$ is a homomorphism from G to H , giving a homomorphic copy of G . Therefore, $\text{HOM}(G)$ and $\text{HOM}(G')$ are essentially equivalent. In other words, whereas the complexity of $\# \text{HOM}(G)$ is governed by the *treewidth* of G [FG04, DJ04], the complexity of $\text{HOM}(G)$ depends on the treewidth of the *homomorphic core* of G rather than on that of G itself (see the references [CR00, Fre90, DKV02] for algorithms and the work by Grohe [Gro07] for the lower bound).

Recall that a homomorphic core² is a graph that admits no homomorphism to any proper subgraph. The homomorphic core of a graph G is the smallest subgraph of G that is itself a core and to which G admits a homomorphism. Every graph has a unique homomorphic core up to isomorphism [HN92], providing an alternative characterisation of the useful, intensively studied notion of *homomorphic equivalence*.

1.2. Motivation: the complexity of subgraph isomorphism. Thus far, we have seen how homomorphic cores help us understand the complexity of homomorphism problems. They also turn out to be useful in the context of subgraph isomorphism and counting. See [CDM17] for a framework that connects subgraph counts and homomorphism counts.

The computational task SUB(G) asks whether an input graph contains G as a subgraph. A well-studied variant is the *colourful* problem COLSUB(G), defined as follows. Let G have k vertices labelled by $1, \dots, k$. The input consists of a graph H together with a partition of its vertex set, $V(H) = \sqcup_{i \in [k]} V_i$. The problem asks whether there exist vertices $v_1, \dots, v_k$ with $v_i \in V_i$ such that $v_i v_j \in E(H)$ for every $ij \in E(G)$. Intuitively, in the colourful setting, the vertices of the pattern graph G receive distinct colours, the input graph H is coloured using the same colours, and the embedding must preserve these colours. Equivalently, COLSUB(G) can be viewed as a 2-CSP (a constraint satisfaction problem of arity 2), with variables x_i for $i \in [k]$, relations R_e for $e \in E(G)$, and domains $\mathcal{D}(x_i) = V_i$.

The complexity of COLSUB(G) is much better understood than that of its uncoloured counterpart SUB(G). For COLSUB(G), the treewidth of G fully governs the complexity in a nearly fine-grained sense: if $\text{tw}(G) = t$, dynamic programming over a tree decomposition solves COLSUB(G) in time $O(n^{\Theta(t)})$; conversely, assuming ETH, there exists a universal constant $C > 0$ such that, for every fixed graph G of treewidth t , COLSUB(G) admits no $O(n^{C \cdot (t/\log t)})$ -time algorithm [Mar10, KMPS24, CDNW26].

The above lower bound for COLSUB(G) can be transferred to the uncoloured problem SUB(G) whenever G is a homomorphic core. When G is not a homomorphic core, however, the picture becomes far less clear [MP14]. This difficulty arises, in particular, for non-trivial bipartite graphs. Even for highly specific pattern families, determining the complexity has been a major challenge, let alone obtaining a complete complexity dichotomy for SUB(G). For instance, the parameterised

²We refrain from abbreviating this term as merely *core*, because a core may also refer to a k -core in the sense of graph degeneracy. Graph degeneracy will be used later in the proof.

complexity of the k -biclique problem was resolved only relatively recently [Lin18], even though, as a standard textbook on parameterised complexity puts it, “almost everyone considers that this problem should *obviously* be $\mathbf{W}[1]$ -hard” [DF13]. The same was true for grid graphs [CGL17]. It is worth noting that both papers received best-paper awards at their respective conferences.

Thus, $\text{SUB}(G)$ is a challenging problem to study for *every* G . Nevertheless, we may study its complexity for a *typical* pattern drawn from the Erdős–Rényi random graph model $\mathcal{G}(k, p)$.

For $\text{COLSUB}(G)$, this setting is well understood when the edge probability lies slightly above the phase-transition threshold for connectivity; that is, when $p \geq (1 + \varepsilon) \log k/k$ for any fixed $\varepsilon > 0$. In this regime, let $\rho = \log(k)/\log(kp)$ denote the *typical diameter*. Then, with high probability over a random graph $G \sim \mathcal{G}(k, p)$, one obtains an $\Omega(n^{C \cdot k/\rho})$ lower bound for $\text{COLSUB}(G)$ [CDNW26]. To transfer this lower bound to the uncoloured setting, it would therefore suffice to show that a random pattern is a homomorphic core with high probability in the desired regime.

1.3. Main results. The first result in this direction was established by Hell and Nešetřil [HN04], who showed that a random graph drawn from $\mathcal{G}(n, 1/2)$ is a homomorphic core with high probability. For simplicity, we assume $p \leq 1/2$ throughout the remainder of the discussion. Bonato and Prałat [BP09] extended this result to the regime $p > n^{-1/3} \log^2 n$. (See also the work of Parsa and Kayll [PK21] on the directed version.)

On the other hand, observe that any n -vertex graph ($n \geq 2$) with at least one isolated vertex is not a homomorphic core: the isolated vertex can be mapped to any other vertex of the graph. Furthermore, any n -vertex graph ($n \geq 3$) with a degree-1 vertex v is not a homomorphic core. Indeed, if the vertex adjacent to v has degree at least 2, then v can be mapped to another neighbour of that vertex. Consequently, in any regime in which the typical *minimum degree* is 0 or 1, a random graph is not a core with high probability.

The phase-transition threshold for minimum degree at least 2 is $p = (\log n + \log \log n \pm \omega(1))/n$ [ER61], so the phase-transition threshold for being a homomorphic core cannot be lower. It was conjectured that the phase-transition threshold for being a homomorphic core coincides with that for minimum degree at least 2 [BP09]. Our main result confirms this conjecture.

Theorem 1. *Let $p = p(n) \leq 1/2$. If*

$$\lim_{n \rightarrow +\infty} (np - \log n - \log \log n) = +\infty,$$

then with high probability $G \sim \mathcal{G}(n, p)$ is a homomorphic core. Conversely, if

$$\lim_{n \rightarrow +\infty} (np - \log n - \log \log n) = -\infty,$$

then with high probability $G \sim \mathcal{G}(n, p)$ is not a homomorphic core.

By the discussion in the previous subsection, our result yields a lower bound for the subgraph isomorphism problem $\text{SUB}(G)$ when the pattern $G \sim \mathcal{G}(k, p)$ is random and $p \geq (1 + \varepsilon) \log k/k$.

Corollary 2. *Assuming ETH, there exists an absolute constant $c > 0$ such that, for every constant $\varepsilon > 0$ and every $p \geq (1 + \varepsilon) \log k/k$, the following holds: with high probability over an Erdős–Rényi random graph $H \sim \mathcal{G}(k, p)$, the problem $\text{SUB}(H)$ admits no $O(n^{c \cdot k/\rho})$ -time algorithm on n -vertex input graphs. Here, $\rho = \log(k)/\log(kp)$ denotes the typical distance in $\mathcal{G}(k, p)$ [CL01].*

In particular, if $p = k^{-\alpha}$ is an inverse polynomial in k with a fixed $\alpha > 0$ and lies within the regime of the corollary, then the typical distance ρ is bounded by a constant. In this case, the corollary implies that, under ETH and with high probability, no $O(n^{c' \cdot k})$ -time algorithm exists, for some constant $c' > 0$ depending only on α and c . In other words, almost every pattern with polynomial average degree is *maximally* hard: unless ETH fails, no algorithm can fundamentally beat the $n^{O(k)}$ -time brute-force algorithm.

1.4. Proof overview. An equivalent definition of homomorphic core is through the *retraction*. A homomorphism r from a graph G to its induced subgraph $G[S_r]$ with image S_r is a retraction if every element in S_r is a fixed point, i.e., $r(V(G)) = S_r$ and $r(s) = s$ for every $s \in S_r$. It is *proper* if $S_r \neq V(G)$. It can be shown that a graph is a homomorphic core if and only if it has no proper retraction. We therefore prove the upper-threshold direction by showing that, with high probability, no map onto a proper induced subgraph can be a retraction.

The above-critical regime. (Section 3) We first consider the simpler regime in which the expected degree np is much larger than $\log n$. For a fixed candidate retraction r , one can view the drawing procedure of a randomly graph G in two stages: we reveal the graph induced by its image S_r before revealing the remaining edges. Conditional on the revealed subgraph, all other edges remain independent. Any pair that the candidate maps to a non-edge must itself be absent, so the probability that the candidate is a retraction is controlled by the number of edges it forces to be absent.

We condition on typical bounds for the *maximum degree* and *clique number* of the random graph. If the image of the candidate is large, the maximum-degree bound implies that every vertex outside the image forces many edges across the boundary to be absent. This probability penalty dominates the number of possible maps. If the image is small, the clique-number bound, together with the *Motzkin–Straus inequality*, shows that many pairs outside the image must be absent. The resulting penalty is strong enough to take a union bound over all candidate maps, even if we just apply a super trivial upper bound on the number of candidate maps with a fixed image.

The near-critical regime. (Section 4) Near the threshold, the preceding union bound becomes too crude, especially for retractions whose images contain almost every vertex. We therefore condition on a more detailed description of the typical random graph. These typical properties first imply that the boundary of a retraction image must be small. If a vertex outside the image has two neighbours inside it, then those vertices, together with its image under the retraction, form a 4-cycle. Since few vertices lie on such cycles, only a small number of outside vertices can have several neighbours in the image.

The main difficulty is to rule out retractions with very large images; the other cases can be ruled out due to the typical properties. Unlike the proof strategy for the above-critical regime, a trivial bound on the number of candidate maps does not suffice. The property that the minimum degree is at least 2 with high probability, however, allows us to design *witnesses* that narrows down the count. Suppose we have a fixed image S . For every vertex outside the image S , once the random graph is fully revealed, choose two incident edges in a canonical way, whose existence is guaranteed by the minimum-degree condition. The resulting sparse witness substantially restricts the candidate map: after the graph on S has been revealed, the map must send the part of the witness outside S homomorphically into the revealed graph. A spanning-forest argument then gives a much sharper count of the possible maps than the trivial bound. Combining this count with the probabilities that the witness edges are present and that all edges forbidden by the retraction are absent yields a desired bound.

2. PRELIMINARIES

2.1. Basic definitions. $[n] := \{1, 2, \dots, n\}$, and $\log$ is the natural logarithm.

Unless stated otherwise, every graph in this manuscript is finite, simple, loopless and undirected. Additionally, every n -vertex graph is vertex-labelled with $[n]$. Denote by $N_G(v)$ the set of neighbours of v in G . The *degree* of a vertex v in a graph G is $\deg_G(v) = |N_G(v)|$, the number of neighbours of v in G . The subscripts for $N_G(v)$ and $\deg_G(v)$ shall be dropped when it is clear from the context. The (maximum) *degree* $\Delta(G)$ of a graph G is the maximum degree of any vertex in G . The *minimum degree* $\delta(G)$ of a graph G is the minimum degree of any vertex in G . Specially, $\delta(G) = 0$ if G

contains any isolated vertex. The *clique number* $\omega(G)$ of a graph G is the number of vertices of its largest clique.³

Suppose a random variable $X \sim \text{Bin}(n, p)$ follows the binomial distribution with expectation $\mu = np$. The (absolute form of) Chernoff bound states

$$\begin{aligned} \Pr[X \geq t] &\leq \left(\frac{e\mu}{t}\right)^t && \text{for any } t \geq \mu; \\ \Pr[X \leq t] &\leq \exp\{-\mu\} \left(\frac{e\mu}{t}\right)^t && \text{for any } t \leq \mu. \end{aligned} \tag{1}$$

2.2. Graph morphisms. Here we collect basic definitions of graph morphisms.

An *isomorphism* is a bijection map $g: V(G) \rightarrow V(H)$ such that

$$xy \in E(G) \iff g(x)g(y) \in E(H).$$

An *automorphism* of a graph G is an isomorphism from G to itself.

A *homomorphism* from G to H is a map $f: V(G) \rightarrow V(H)$ such that

$$xy \in E(G) \implies f(x)f(y) \in E(H).$$

An *endomorphism* of a graph G is a homomorphism from G to itself.

A *retraction* r from a graph G to its induced subgraph $G[S]$ is an endomorphism such that

$$r(V(G)) = S \quad \text{and} \quad r(s) = s \text{ for every } s \in S.$$

It is *proper* if $S \neq V(G)$.

A standard fact is the following; see [HN04, Proposition 1.31].

Proposition 3 (homomorphic core). *The two statements are equivalent for any graph G .*

- (1) *Every endomorphism of G is an automorphism.*
- (2) *G has no proper retraction.*

A graph G is called a *homomorphic core* if it satisfies (any of) the above two statements.

Proof. The direction (1) $\implies$ (2) is trivial: any retraction by definition is an endomorphism.

To show (2) $\implies$ (1), assume G has an endomorphism to a proper induced subgraph. Let H be the induced subgraph with the minimum number of vertices that G is homomorphic to. Thus by the choice of H , any endomorphism of H must be an automorphism. Now let $f: V(G) \rightarrow V(H)$ be a homomorphism from G to H , and $g: V(H) \rightarrow V(H)$ be the restriction of f on H , i.e., $f(s) = g(s)$ for any $s \in V(H)$. Trivially, g is an endomorphism, and hence an automorphism, and thus it has an inverse automorphism g^{-1} . Then $g^{-1} \circ f$ is a retraction of G to H . $\square$

2.3. Random graphs. The Erdős–Rényi random graph model $\mathcal{G}(n, p)$ [ER60] is a distribution over n -vertex graphs with each pair of distinct vertices joined by an edge with probability p independently. Erdős and Rényi [ER61] located the phase transition threshold for minimum degree 2:

Theorem 4 (minimum degree phase transition). *Let $p = p(n) \leq 1/2$. If*

$$\lim_{n \rightarrow +\infty} (np - \log n - \log \log n) = +\infty,$$

then w.h.p. $\delta(G) \geq 2$. Conversely, if

$$\lim_{n \rightarrow +\infty} (np - \log n - \log \log n) = -\infty,$$

then w.h.p. $\delta(G) \leq 1$.

³Be aware of the notation overload of ω . However, it is clear from the context whether it means the clique number or the asymptotic notation.

Proof. Set $d = np$, and let X be the number of vertices of degree at most 1. For every vertex v ,

$$\Pr[\deg(v) \leq 1] = (1-p)^{n-1} + (n-1)p(1-p)^{n-2} \leq (1+d) \exp\{-d+1\}.$$

By linearity of expectation, $\mathbb{E}[X] \leq n(1+d) \exp\{-d+1\}$. Suppose first that

$$w_n := d - \log n - \log \log n \longrightarrow +\infty.$$

If $d \leq 2 \log n$, then

$$\mathbb{E}[X] \leq n(1+d) \exp\{-d+1\} = \exp\{1-w_n\} \frac{1+d}{\log n}.$$

If $d > 2 \log n$, then, since $(1+x) \exp\{-x\}$ is decreasing for $x > 0$,

$$\mathbb{E}[X] \leq n(1+2 \log n) \exp\{-2 \log n + 1\} = O\left(\frac{\log n}{n}\right).$$

Since both right-hand-sides of the above two inequalities are $o(1)$, we have $\mathbb{E}[X] = o(1)$. Thus, by Markov's inequality, $X = 0$ with high probability, and hence $\delta(G) \geq 2$.

Conversely, we apply a second-moment analysis. Suppose that

$$w_n := \log n + \log \log n - d \longrightarrow +\infty.$$

We may assume $d \geq (\log n)/2$; this stochastically dominates the case when d is smaller than this. Because $d = O(\log n)$, we have $p = d/n = o(1)$ and $np^2 = d^2/n = o(1)$. Therefore,

$$\mathbb{E}[X] = n(1-p)^{n-2}((1-p) + (n-1)p) = (1+o(1))n(1+d) \exp\{-d\} = (1+o(1)) \frac{1+d}{\log n} \exp\{w_n\} \rightarrow +\infty.$$

For the second-moment bound, fix two distinct vertices u and v . Conditioning on whether uv is an edge gives

$$\begin{aligned} \Pr[\deg(u) \leq 1, \deg(v) \leq 1] &= (1-p) \left((1-p)^{n-2} + (n-2)p(1-p)^{n-3} \right)^2 + p(1-p)^{2n-4} \\ &\leq \left((1-p)^{n-2} + (n-2)p(1-p)^{n-3} \right)^2. \end{aligned}$$

Moreover,

$$\begin{aligned} \Pr[\deg(v) \leq 1] &= (1-p) \left((1-p)^{n-2} + (n-2)p(1-p)^{n-3} \right) + p(1-p)^{n-2} \\ &\geq (1-p) \left((1-p)^{n-2} + (n-2)p(1-p)^{n-3} \right). \end{aligned}$$

Consequently,

$$\Pr[\deg(u) \leq 1, \deg(v) \leq 1] \leq \frac{\Pr[\deg(v) \leq 1]^2}{(1-p)^2} = (1+o(1)) \Pr[\deg(v) \leq 1]^2,$$

where the last equality follows from $p = o(1)$.

It follows that

$$\begin{aligned} \mathbb{E}[X^2] &= \mathbb{E}[X] + \sum_{u \neq v} \Pr[\deg(u) \leq 1, \deg(v) \leq 1] \\ &\leq \mathbb{E}[X] + (1+o(1))n(n-1) \left(\frac{\mathbb{E}[X]}{n} \right)^2 \\ &\leq \mathbb{E}[X] + (1+o(1))\mathbb{E}[X]^2. \end{aligned}$$

Since $\mathbb{E}[X] \rightarrow \infty$, the second-moment inequality yields

$$\Pr[X > 0] \geq \frac{\mathbb{E}[X]^2}{\mathbb{E}[X^2]} = 1 - o(1).$$

Therefore $\delta(G) \leq 1$ with high probability. □

3. THE ABOVE-CRITICAL REGIME

With a multiplicative gap away from the phase transition threshold in Theorem 1, a (comparably) simpler proof can be obtained.

Theorem 5. *Let $p = p(n) \leq 1/2$ with*

$$\lim_{n \rightarrow +\infty} \frac{np}{\log n} = +\infty.$$

Then $G \sim \mathcal{G}(n, p)$ is a homomorphic core w.h.p.

3.1. Typical structures. Recall the following crude estimations on the degree and the clique number of a random graph in this regime.

Lemma 6. *In the same regime of $p = p(n)$ as in Theorem 5, the following holds simultaneously w.h.p. for a randomly drawn $\mathcal{G}(n, p)$:*

- (a) (maximum degree) $\Delta(G) \leq 7n/12$;
- (b) (clique number) $\omega(G) \leq \lceil 4 \log n / (\log 1/p) \rceil$.

Proof. It suffices to show that each event fails with probability $o(1)$, and then apply union bound.

For (a), note that every vertex degree is dominated by the binomial distribution $\text{bin}(n-1, 1/2)$, and then apply Hoeffding's inequality.

To show (b), for any $W \in \mathbb{N}_{\geq 1}$, by union bound,

$$\Pr[\omega(G) \geq W] \leq \binom{n}{W} p^{\binom{W}{2}} \leq \exp \left\{ W \log \frac{en}{W} - \frac{W(W-1) \log 1/p}{2} \right\}. \quad (2)$$

Now let $W = \lceil 4 \log n / (\log 1/p) \rceil + 1$, so we have $W-1 \geq 4 \log n / (\log 1/p)$. The above is further upper bounded by

$$(2) \leq \exp\{W(1 - \log n - \log W)\} \leq \exp\{W(1 - \log n)\} = o(1). \quad \square$$

3.2. Conditioned on a typical structure. Over a randomly drawn $G \sim \mathcal{G}(n, p)$, define a “global” random event

$$\mathcal{A} \quad := \quad \Delta(G) \leq \frac{7n}{12} \quad \text{and} \quad \omega(G) \leq \left\lceil \frac{4 \log n}{\log 1/p} \right\rceil.$$

By Lemma 6, $\Pr[\neg \mathcal{A}] = o(1)$.

For any map (candidate *proper* retraction) $r: [n] \rightarrow S_r$ that is identity on $S_r \neq [n]$, define the random event

$$\mathcal{R}_r \quad := \quad r \text{ is a retraction.}$$

By the definition of homomorphic cores (Proposition 3), we need to show $\Pr[\bigvee_r \mathcal{R}_r] = o(1)$. But since $\Pr[\neg \mathcal{A}] = o(1)$ and

$$\Pr \left[\bigvee_r \mathcal{R}_r \right] \leq \Pr \left[\left(\bigvee_r \mathcal{R}_r \right) \wedge \mathcal{A} \right] + \Pr[\neg \mathcal{A}] = \Pr \left[\bigvee_r (\mathcal{R}_r \wedge \mathcal{A}) \right] + \Pr[\neg \mathcal{A}] \leq \sum_r \Pr[\mathcal{R}_r \wedge \mathcal{A}] + \Pr[\neg \mathcal{A}],$$

it suffices to prove that, summing over all proper r ,

$$\sum_r \Pr[\mathcal{R}_r \wedge \mathcal{A}] = o(1). \quad (3)$$

We first bound each term $\Pr[\mathcal{R}_r \wedge \mathcal{A}]$. Fix the map r and let $G[S_r]$ be the random induced subgraph of G on S_r . By the law of total probability,

$$\Pr[\mathcal{R}_r \wedge \mathcal{A}] = \sum_H \Pr[\mathcal{R}_r \wedge \mathcal{A} \mid G[S_r] = H] \cdot \Pr[G[S_r] = H], \quad (4)$$

where the summation goes over all graphs H on the vertex set S_r . Then observe that, assuming the global event $\mathcal{A}$ happens, the induced subgraph $G[S_r]$ must also satisfy

$$\Delta(G[S_r]) \leq \frac{7n}{12} \quad \text{and} \quad \omega(G[S_r]) \leq \left\lceil \frac{4 \log n}{\log 1/p} \right\rceil,$$

and therefore, only terms with graphs H satisfying the above condition survives in the summation of (4). Let $\mathcal{H}$ be the collection of all such H 's. (4) can be rewritten as

$$\Pr[\mathcal{R}_r \wedge \mathcal{A}] = \sum_{H \in \mathcal{H}} \Pr[\mathcal{R}_r \wedge \mathcal{A} \mid G[S_r] = H] \cdot \Pr[G[S_r] = H]. \quad (5)$$

Dropping $\mathcal{A}$ yields an upper bound:

$$\Pr[\mathcal{R}_r \wedge \mathcal{A}] \leq \sum_{H \in \mathcal{H}} \Pr[\mathcal{R}_r \mid G[S_r] = H] \cdot \Pr[G[S_r] = H] \leq \max_{H \in \mathcal{H}} \Pr[\mathcal{R}_r \mid G[S_r] = H]. \quad (6)$$

3.3. Probability of a candidate retraction. It is left for us to characterise the quantity

$$\Pr[\mathcal{R}_r \mid G[S_r] = H].$$

Fix the map r . A random graph $G \sim \mathcal{G}(n, p)$ can be viewed as drawn in two independent stages: reveal all the edges among S_r , and then reveal the remaining edges. The candidate r is invalidated if and only if any remaining edge is mapped to a non-edge in H under r . Let

$$B(H, r) := \{uv : r(u)r(v) \notin E(H)\} \subseteq \binom{[n]}{2}$$

be the set of edges forced to be absent. Then $\Pr[\mathcal{R}_r \mid G[S_r] = H] = (1 - p)^{B(H, r)}$ and hence

$$\Pr[\mathcal{R}_r \wedge \mathcal{A}] \leq \max_{H \in \mathcal{H}} (1 - p)^{B(H, r)}. \quad (7)$$

The quantity $B(H, r)$ has an explicit formula.

Lemma 7. *For every $s \in S_r$, define*

$$b_s := |\{x \in [n] \setminus S_r : r(x) = s\}|$$

the number of vertices mapped by r to a vertex s except itself. Then $B(H, r) = B_1(H, r) + B_2(H, r)$ where

$$B_1(H, r) := \sum_{s \in S_r} b_s (|S_r| - \deg_H(s)); \quad (8)$$

$$B_2(H, r) := \binom{n - |S_r|}{2} - \sum_{st \in E(H)} b_s b_t. \quad (9)$$

Intuitively speaking, B_1 and B_2 together count all edges forced to be absent to make r a retraction. B_1 includes edges with exactly one endpoint in the image S_r , and B_2 includes edges outside S_r .

Proof of Lemma 7. We first count $B_1(H, r)$. Consider $x \in [n] \setminus S_r$ with $r(x) = s$, and another vertex $t \in S_r$. Under r , the edge xt would map to st , and therefore, if $st \notin E(H)$, then the edge xt must be absent, or otherwise r is no longer a homomorphism. For a fixed vertex $s \in S$, there are $|S_r| - \deg_H(s)$ vertices t in S_r for which st is not an edge, including $t = s$. There are b_s vertices outside S_r that get mapped to s , so B_1 counts all edges between S_r and $[n] \setminus S_r$ forced to be absent. All the other edges between S_r and $[n] \setminus S_r$ are free.

For $B_2(H, r)$, consider a pair of vertices $x \neq y \in [n] \setminus S_r$ with $r(x) = s$ and $r(y) = t$. The edge xy is permitted if and only if $st \in E(H)$. There are $\binom{n - |S_r|}{2}$ pairs outside $|S_r|$. For each edge $st \in E(H)$, exactly $b_s b_t$ pairs have one endpoint mapped to s and the other endpoint mapped to t under r .

These are the only pairs outside S_r allowed to be edges, and hence B_2 counts all edges outside S_r forced to be absent. $\square$

In the next two subsections, we bound $B(H, r)$ in two cases based on the image size $|S_r|$ of r . This partitions the summation (3) into two parts.

Lemma 8 (large image). *We have*

$$\sum_{\substack{r: \\ 2n/3 \leq |S_r| \leq n-1}} \max_{H \in \mathcal{H}} (1-p)^{B(H, r)} = o(1).$$

Lemma 9 (small image). *We have*

$$\sum_{\substack{r: \\ 1 \leq |S_r| < 2n/3}} \max_{H \in \mathcal{H}} (1-p)^{B(H, r)} = o(1).$$

Theorem 5 follows once the above two lemmata are proved.

3.4. Large image. This subsection proves Lemma 8. For any $H \in \mathcal{H}$, plugging the condition $2n/3 \leq |S_r| \leq n-1$ and the upper bound of $\Delta(H)$ into (8) in Lemma 7, we have

$$B_1(H, r) = \sum_{s \in S_r} b_s(|S_r| - \deg_H(s)) \geq \sum_{s \in S_r} b_s \left(\frac{2n}{3} - \frac{7n}{12} \right) = \frac{n}{12} \sum_{s \in S_r} b_s = \frac{(n - |S_r|)n}{12}$$

where the last equation is by the definition of b_s . Trivially, we have $B(H, r) \geq B_1(H, r)$ and hence

$$(1-p)^{B(H, r)} \leq (1-p)^{(n-|S_r|)n/12} \leq \exp \left\{ -\frac{(n-|S_r|)np}{12} \right\}. \quad (10)$$

Consider the summation in Lemma 8. For a fixed size $2n/3 \leq \ell \leq n-1$, the number of choices of a set $S \subseteq [n]$ with $|S| = \ell$ is $\binom{n}{\ell}$. Once S is fixed, the number of maps that fixes S and maps $[n] \setminus S$ into S is $\ell^{n-\ell}$. Therefore, the number of retractions is at most $\binom{n}{\ell} \ell^{n-\ell}$. Summing (10) over all ℓ in this range gives

$$\begin{aligned} & \sum_{\substack{r: \\ 2n/3 \leq |S_r| \leq n-1}} \max_{H \in \mathcal{H}} (1-p)^{B(H, r)} = \sum_{2n/3 \leq \ell \leq n-1} \exp \left\{ -\frac{(n-\ell)np}{12} \right\} \binom{n}{\ell} \ell^{n-\ell} \\ & \stackrel{(a)}{=} \sum_{1 \leq k \leq n/3} \exp \left\{ -\frac{knp}{12} \right\} \binom{n}{k} (n-k)^k \stackrel{(b)}{\leq} \sum_{1 \leq k \leq n/3} \exp \left\{ -\frac{knp}{12} \right\} \left(\frac{en^2}{k} \right)^k \\ & = \sum_{1 \leq k \leq n/3} \exp \left\{ k \left(2 \log n + 1 - \log k - \frac{np}{12} \right) \right\} \\ & \leq \sum_{k \geq 1} \exp \left\{ k \left(2 \log n + 1 - \frac{np}{12} \right) \right\} = \frac{\exp \{ 2 \log n + 1 - np/12 \}}{1 - \exp \{ 2 \log n + 1 - np/12 \}} \\ & \stackrel{(c)}{=} n^{-\omega(1)} = o(1). \end{aligned}$$

Here, (a) is the substitution $k = n - \ell$, (b) is due to the inequality $\binom{n}{k} (n-k)^k \leq \left(\frac{en^2}{k} \right)^k$, and (c) is because of the condition $\frac{np}{\log n} = \omega(1)$. This concludes the proof of Lemma 8.

3.5. Small image. This subsection proves Lemma 9. The following Motzkin–Straus inequality [MS65] from combinatorial optimisation is later used. A proof is provided here for completeness.

Lemma 10 (Motzkin–Straus inequality). *Let H be a graph with clique number $\omega(H)$, together with non-negative weights $\{b_v\}_{v \in V(H)}$ on vertices satisfying $\sum_{v \in V(H)} b_v = k > 0$. Then*

$$\sum_{uv \in E(H)} b_u b_v \leq \frac{k^2}{2} \left(1 - \frac{1}{\omega(H)}\right).$$

Proof. It suffices to prove for $k = 1$ by normalising all weights.

Repeat the following step, until there does not exist any pair of vertices x, y of positive weights b_x, b_y that are non-adjacent. We may write

$$\sum_{uv \in E(H)} b_u b_v = \left(\sum_{u \in N(x)} b_u \right) b_x + \left(\sum_{v \in N(y)} b_v \right) b_y + C$$

where C does not depend on b_x or b_y . If $\sum_{u \in N(x)} b_u > \sum_{v \in N(y)} b_v$, we update $b_x \leftarrow b_x + b_y$ and $b_y \leftarrow 0$; otherwise, update $b_x \leftarrow 0$ and $b_y \leftarrow b_x + b_y$. This procedure strictly increase the total count $\sum_{uv \in E(H)} b_u b_v$.

The above procedure eventually halts because it strictly decreases the size of the positive support. And when it halts, the support is a clique K_r of some size $r \leq \omega(H)$. Let $\{b'_u\}$ be the final weights. On this clique, the Cauchy–Schwarz inequality gives

$$\sum_{uv \in E(H)} b_u b_v \leq \sum_{xy \in E(K_r)} b'_x b'_y = \frac{1}{2} \left(1 - \sum_{x \in V(K_r)} b'^2_x\right) \stackrel{\text{C-S}}{\leq} \frac{1}{2} \left(1 - \frac{1}{r}\right) \leq \frac{1}{2} \left(1 - \frac{1}{\omega(H)}\right). \quad \square$$

Write $W = \left\lceil \frac{4 \log n}{\log 1/p} \right\rceil$ so $\omega(H) \leq W$ for any $H \in \mathcal{H}$. We have

$$\begin{aligned} B_2(H, r) &= \binom{n - |S_r|}{2} - \sum_{st \in E(H)} b_s b_t \stackrel{(d)}{\geq} \binom{n - |S_r|}{2} - \frac{(n - |S_r|)^2}{2} \left(1 - \frac{1}{W}\right) \\ &= \frac{n - |S_r|}{2W} (n - |S_r| - W) \geq \frac{n^2}{36W} \end{aligned}$$

for all sufficiently large n since $W = O(\log n)$. Here, (d) is due to Lemma 10. Again, we trivially have $B(H, r) \geq B_2(H, r)$ and hence

$$(1 - p)^{B(H, r)} \leq (1 - p)^{n^2/(36W)} \leq \exp \left\{ -\frac{n^2 p}{36W} \right\}. \quad (11)$$

Consider the summation in Lemma 9. Trivially, the number of maps from $[n]$ to $[n]$ is n^n . Using (11), we have

$$\sum_{\substack{r: \\ 1 \leq |S_r| < 2n/3}} \max_{H \in \mathcal{H}} (1 - p)^{B(H, r)} \leq n^n \exp \left\{ -\frac{n^2 p}{36W} \right\} = \exp \left\{ -n \log n \left(\frac{np}{36W \log n} - 1 \right) \right\}.$$

We claim that $np/(W \log n) = \omega(1)$, which immediately implies that the above quantity is $o(1)$ and therefore finishes the proof of Lemma 9. Indeed, for all n such that $p(n) < \sqrt{n}/n$, we have $W \leq 8$, and so $np/(W \log n) \geq np/(8 \log n)$. For all n such that $p(n) \geq \sqrt{n}/n$, since $p(n) \leq 1/2$, we have $W \leq \lceil (4/\log 2) \log n \rceil \leq 8 \log n$, and so $np/(W \log n) \geq \sqrt{n}/(8 \log^2 n)$. This gives

$$\frac{np}{W \log n} \geq \min \left\{ \frac{np}{8 \log n}, \frac{\sqrt{n}}{8 \log^2 n} \right\}.$$

Both terms are $\omega(1)$ in the regime of p as Theorem 5.

4. THE NEAR-CRITICAL REGIME

For the rest of this manuscript, assume

$$\begin{aligned} d := np = \log n + \log \log n + w \quad \text{with} \quad \lim_{n \rightarrow +\infty} w(n) = +\infty, \quad \text{and} \\ d \leq \log^2 n, \end{aligned} \tag{12}$$

as otherwise Theorem 5 applies.

4.1. Typical structure near the critical threshold. We need some more standard facts about certain graph parameters in the above regime.

Recall the following basic definitions. The *independence number* $\alpha(G)$ of a graph G is the number of vertices of its largest independent set, or alternatively, the clique number of its complement graph. The *chromatic number* $\chi(G)$ of a graph G is the minimum number q that G has a proper q -vertex-colouring.

Lemma 11. *Under the condition (12), the following holds simultaneously w.h.p. for a randomly drawn $G \sim \mathcal{G}(n, p)$:*

- (a) (minimum degree) $\delta(G) \geq 2$;
- (b) (maximum degree) $\Delta(G) \leq 4d$;
- (c) (4-cycle count) at most $4d^5$ vertices lie on a 4-cycle;
- (d) (chromatic number) $\chi(G) \geq d/(5 \log d)$;
- (e) (subgraph edge density) with $r_0 := 30 \log d / \log \log d$, every $U \subseteq [n]$ with $|U| \leq 8/p$ satisfies $|E(G[U])| < r_0 |U|/2$.

Proof. By a union bound, it suffices to prove that the failure probability of each item is $o(1)$.

- (a) Exactly the same as Theorem 4.
- (b) This is derived from the union bound and Chernoff bound (1):

$$\Pr[\Delta(G) \geq 4d] \leq n \cdot \Pr[\deg_G(v) \geq 4d] \leq n \cdot 0.08^d \leq n \cdot 0.08^{\log n} = o(n^{-1.52}) = o(1).$$

- (c) By the linearity of expectation, the expected number Z of 4-cycles is $p^4 \cdot n(n-1)(n-2)(n-3)/8 \leq d^4/8$. By Markov's inequality, the probability $\Pr[Z > d^5] \leq 1/(8d) = o(1)$, where at most $4d^5$ vertices lie on a 4-cycle.
- (d) The expected number of independent sets of size k is

$$\binom{n}{k} (1-p)^{\binom{k}{2}} \leq \exp \left\{ k \log \frac{en}{k} - \frac{pk(k-1)}{2} \right\}.$$

For any $k \geq 5 \log d/p$, the above quantity tends to 0, meaning w.h.p. the independence number $\alpha(G) < 5 \log d/p$, and therefore the chromatic number $\chi(G) \geq d/(5 \log d)$.

- (e) Let $u = |U| \leq 8/p$. The number of edges $|E(G[U])|$ is subject to the binomial distribution $\text{Bin}(u(u-1)/2, p)$. By a union bound over all U of size u , the probability that (e) is violated by a fixed u is at most

$$\binom{n}{u} \Pr \left[\text{Bin} \left(\frac{u(u-1)}{2}, p \right) \geq \frac{r_0 u}{2} \right].$$

By (1), this quantity is bounded above by

$$\binom{n}{u} \exp \left\{ \frac{r_0 u}{2} \log \frac{epu}{r_0} \right\} \leq \exp \left\{ u \left(\log \frac{ed}{pu} + \frac{r_0}{2} \log \frac{epu}{r_0} \right) \right\}. \tag{13}$$

Here, the last inequality due to $\binom{n}{u} \leq (en/u)^u$ and $d = np$. The function $\log(ed/x) + r_0/2 \cdot \log(ex/r_0)$ is increasing since $r_0 > 2$. At $x = 8$, its value is bounded by

$$\log \frac{ed}{8} - \frac{r_0}{2} \log \frac{r_0}{8e} \leq \log \frac{ed}{8} - \frac{r_0}{4} \log \log d \leq -6 \log d$$

for sufficiently large d . As $pn \leq 8$, we upper bound (13) by

$$\leq \exp\{-6u \log d\} = d^{-6u}.$$

Summing all possible u yields a bound $\sum_{u=1}^{8/p} d^{-6u} \leq d^{-6} = o(1)$ on the probability that (e) is violated. $\square$

An interesting consequence of the above typical graph parameters, especially Item (a)(b)(c), is the boundary size of the image S_r of a retraction r .

Lemma 12. *Let r be a retraction to S_r . If events (a)(b)(c) in Lemma 11 happen, then the number of edges between S_r and $[n] \setminus S_r$ is upper bounded by $16d^6 + n - |S_r|$.*

Proof. By Lemma 11(a), every vertex $x \in [n] \setminus S_r$ has at least two distinct neighbours. Suppose x has two distinct neighbours $u, v \in S$. By the definition of retraction, the edges $r(x)r(u) = r(x)u$ and $r(x)r(v) = r(x)v$ exist, and $r(x) \neq x, u, v$. Therefore, the 4-cycle $x \rightarrow u \rightarrow r(x) \rightarrow v \rightarrow x$ exist, so x lies on a 4-cycle. Consequently, any vertex in $[n] \setminus S_r$ that does not lie on any 4-cycle must have at most one neighbours in S . Trivially, by Lemma 11(b), any vertex in $[n] \setminus S_r$ that lies on a 4-cycle has at most $4d$ neighbours in S . The lemma then follows by Lemma 11(c). $\square$

4.2. Conditioned on a typical structure. In a similar manner as the proof of Theorem 5, with $G \sim \mathcal{G}(n, p)$ define the “global” event

$$\mathcal{B} \quad := \quad \text{All items in Lemma 11 happen simultaneously.}$$

By Lemma 11, we have $\Pr[\mathcal{B}] = 1 - o(1)$.

Like in Section 3.3, we subdivide the event that a proper retraction exists into three cases depending on the size of the retraction’s image. For any map (candidate proper retraction) $r: [n] \rightarrow S_r$ that is identity on $S_r \neq [n]$, we call the map is of Type I, II or III, based on the following criteria.

$$\begin{aligned} 1 &\leq |[n] \setminus S_r| < 64d^6; & \text{(Type I)} \\ 64d^6 &\leq |[n] \setminus S_r| \leq n - 8/p; & \text{(Type II)} \\ n - 8/p &< |[n] \setminus S_r| \leq n - 1. & \text{(Type III)} \end{aligned}$$

We assume n large enough so the above inequalities are well-defined. Define $\mathcal{R}_r$ to be the random event that r is a retraction. For $\star \in \{I, II, III\}$, define the events

$$\mathcal{E}_\star \quad := \quad \text{There is an } r \text{ of type } \star \text{ that } \mathcal{R}_r \text{ happens.}$$

Similar to (3), we only need to bound

$$\left(\sum_{r: \text{Case I}} \Pr[\mathcal{R}_r \wedge \mathcal{B}] \right) + \Pr[\mathcal{E}_{II} \wedge \mathcal{B}] + \Pr[\mathcal{E}_{III} \wedge \mathcal{B}] = o(1). \quad (14)$$

Lemma 13. *We have*

$$\sum_{r: \text{Case I}} \Pr[\mathcal{R}_r \wedge \mathcal{B}] = o(1). \quad (15)$$

Lemma 14. *We have $\Pr[\mathcal{E}_{II} \wedge \mathcal{B}] = o(1)$.*

Lemma 15. *We have $\Pr[\mathcal{E}_{III} \wedge \mathcal{B}] = o(1)$.*

The first half of Theorem 1 follows after the above three lemmata. The second half of Theorem 1 is implied by Theorem 4.

4.3. **Case I.** This subsection proves Lemma 13 where the image of the candidate retraction is very large.

We begin with changing the order of summation in (15). For any $S \subsetneq [n]$, if $\mathcal{B}$ happens, then $\Delta(G[S]) \leq 4d$ for any induced subgraph $G[S]$ of G . Let $\mathcal{H}_S$ be the collection of all graphs H on the vertex set S such that $\Delta(H) \leq 4d$. Analogously to (5), we have

$$\begin{aligned} \sum_{r: \text{Case I}} \Pr[\mathcal{R}_r \wedge \mathcal{B}] &\leq \sum_{1 \leq \ell < 64d^6} \sum_{\substack{S \subseteq [n] \\ |S|=n-\ell}} \sum_{H \in \mathcal{H}_S} \Pr[G[S] = H] \cdot Z(H) \\ &\leq \sum_{1 \leq \ell < 64d^6} \sum_{\substack{S \subseteq [n] \\ |S|=n-\ell}} \max_{H \in \mathcal{H}_S} Z(H) \end{aligned} \tag{16}$$

with the definition

$$Z(H) := \sum_{\substack{r: \\ S_r=S}} \Pr[\mathcal{R}_r \wedge \mathcal{B} \mid G[S] = H] \tag{17}$$

for $H \in \mathcal{H}_S$.

Fix S and $H \in \mathcal{H}_S$. From this point onward, H is fixed rather than random. Conditional on $G[S] = H$, all edges that are not fully in S remain mutually independent.

Suppose that $\mathcal{R}_r \wedge \mathcal{B}$ occurs for some candidate map r with $S_r = S$. Since $\delta(G) \geq 2$ under $\mathcal{B}$, choose two distinct edges incident with each vertex $x \in [n] \setminus S$, and let F be the simple graph formed by the union of all these chosen edges. For definiteness, one may choose the two edges whose other endpoints have the smallest labels. Formally, the edge set of F is

$$\bigcup_{x \in [n] \setminus S} \{xy_1, xy_2 : y_1 \neq y_2 \text{ are the smallest indices that } xy_1, xy_2 \in E(G)\}.$$

Thus F is determined by G , although in the estimate below we simply take a union bound over all possible F . The graph F is served as a *witness*.

As in the notation above, partition

$$F_{\text{out}} := F[[n] \setminus S], \quad F_{\text{boundary}} := F - F_{\text{out}}.$$

Set

$$e_2 := |E(F_{\text{out}})|, \quad e_1 := |E(F_{\text{boundary}})|.$$

So F_{out} is an *induced* subgraph of F that collects edges in $[n] \setminus S$, and F_{boundary} is a bipartite subgraph collecting edges between S and $[n] \setminus S$. Let κ be the number of connected components of F_{out} , including isolated vertices.

Lemma 16. *Any witness F constructed above satisfies*

$$e_1 + e_2 \leq 2(n - |S|) \quad \text{and} \quad e_1 + e_2 - \kappa \geq \frac{2(n - |S|)}{3}.$$

Proof. The first inequality follows because two edges were selected at each vertex of $[n] \setminus S$, and repeated selections produce only one edge in the simple graph F .

Consider a component J of F_{out} . Suppose that J has a vertices and b edges, and that c witness edges join $V(J)$ to S . Every vertex of J is incident with at least two edges of F , so

$$2b + c \geq 2a.$$

If J is a tree, then $b = a - 1$, and hence $c \geq 2$. It follows that $b + c - 1 \geq a$. If J is not a tree, then $b \geq a$; moreover $a \geq 3$, because the graph is simple. Therefore $b + c - 1 \geq a - 1 \geq 2a/3$. Summing these inequalities over all κ components gives the desired bound. $\square$

If the subgraph $G[S]$ has been revealed to be H , then a retraction r restricted on $[n] \setminus S$ is a homomorphism from F_{out} to H , since $r(x)r(y) \in E(H)$ for any $xy \in E(F_{\text{out}})$. This significantly reduces the number of candidate maps if the F_{out} part of the witness is fixed.

To rewrite the summation for $Z(H)$ in (17), we enumerate the possible witnesses and candidate maps. It is important that H has already been fixed before this enumeration.

First choose F_{out} . If it has e_2 edges, the number of choices is at most

$$\binom{\binom{n-|S|}{2}}{e_2} \leq (n-|S|)^{2e_2}.$$

Next, for a fixed F_{out} , we claim that the number of maps from $[n] \setminus S$ to S that are homomorphisms from F_{out} to H is at most

$$|S|^\kappa (\Delta(H))^{n-|S|-\kappa} \leq |S|^\kappa (4d)^{n-|S|-\kappa}.$$

To see this, fix one rooted spanning tree in each component of F_{out} . Choosing the image $r(x) \in S$ of one root $x \in F_{\text{out}}$ in each component gives $|S|^\kappa$ so many options. Each of the remaining vertex in F_{out} must map to a neighbour of its parent's image. This gives at most $\Delta(H)$ options, and $\Delta(H) \leq 4d$ because $\mathcal{B}$ happens.

Now fix one of these compatible maps r . We next choose F_{boundary} . A pair xv where $x \in [n] \setminus S$ and $y \in S$ can belong to F_{boundary} only if $r(x)v \in E(H)$. Otherwise the present edge xv would be mapped to a non-edge. The number of such allowed pairs is at most

$$\sum_{x \in [n] \setminus S} \deg_H(r(x)) \leq 4d(n-|S|).$$

Thus, if F_{boundary} has e_1 edges, there are at most

$$(4d(n-|S|))^{e_1}$$

choices for it.

It remains to charge the probabilities of the required present and absent edges. By (8), the map r forces at least

$$B_1(H, r) \geq (n-|S|)(|S|-4d)$$

edges between S and $[n] \setminus S$ to be absent. These forced-absent edges are disjoint from the witness edges in F : indeed, the edges of F_{out} have both endpoints outside S , and every edge xv of F_{boundary} was chosen only when $r(x)v \in E(H)$, meaning the edge is not forced to be absent. Therefore, conditional on $G[S] = H$, the independence of the remaining edges gives

$$p^{e_1+e_2}(1-p)^{(n-|S|)(|S|-4d)}$$

as an upper bound on the probability that all witness edges are present and all these $B_1(H, r)$ edges are absent. We discard any additional conditions required for r to be a retraction, which only enlarges the probability.

Combining the preceding enumeration steps and applying a union bound over all witnesses gives

$$Z(H) \leq (1-p)^{(n-|S|)(|S|-4d)} \sum_{\substack{e_1, e_2 \geq 0, 1 \leq \kappa \leq n-|S| \\ e_1+e_2 \leq 2(n-|S|) \\ e_1+e_2-\kappa \geq 2(n-|S|)/3}} (n-|S|)^{2e_2+e_1} |S|^\kappa (4d)^{n-|S|-\kappa+e_1} p^{e_1+e_2}. \quad (18)$$

Here, the summation over e_1, e_2, κ is truncated by Lemma 16.

We finish by summing (18). In (16), $|S| = n - \ell$ with $1 \leq \ell < 64d^6$. Since $p = d/n$ and $d \leq \log^2 n$,

$$p(\ell + 4d) = O(d^7/n) = o(1),$$

and hence

$$(1-p)^{\ell(n-\ell-4d)} \leq \exp\{-p\ell(n-\ell-4d)\} = \exp\{-d\ell + o(\ell)\}. \quad (19)$$

For every term in (18),

$$|S|^\kappa p^{e_1+e_2} \leq n^\kappa \left(\frac{d}{n}\right)^{e_1+e_2} \leq n^{-2\ell/3} d^{e_1+e_2},$$

where the last inequality uses $e_1 + e_2 - \kappa \geq 2\ell/3$ from Lemma 16. Furthermore,

$$(4d)^{\ell-\kappa+e_1} \leq (4d)^{3\ell}, \quad d^{e_1+e_2} \leq d^{2\ell}, \quad \ell^{2e_2+e_1-\ell} \leq \ell^{3\ell}.$$

There are at most $O(\ell^3)$ triples (e_1, e_2, κ) in the summation. Using $\binom{n}{\ell} \leq (en/\ell)^\ell$, $d = \log n + \log \log n + w$, and (19), the total contribution in (16) from all S with $n - |S| = \ell$ is at most

$$O(\ell^3) \left(C n^{-2/3} \ell^3 d^5 e^{-w+o(1)} \right)^\ell \leq O(\ell^3) \left(C n^{-2/3} d^{23} \right)^\ell$$

for an absolute constant C . Since $d \leq \log^2 n$, this is $o(n^{-\ell/2})$ uniformly for $1 \leq \ell < 64d^6$. Summing over ℓ proves (15).

4.4. Case II. This subsection proves Lemma 14. Recall the condition on the image size of a candidate retraction $64d^6 \leq |[n] \setminus S_r| \leq n - 8/p$. We assume n large enough, so $3/(4p) \leq n/2 \leq n - 8/p$. This is possible in the regime of p in (12).

Suppose that $\mathcal{E}_{\text{II}} \wedge \mathcal{B}$ occurs; that is, $\mathcal{B}$ happens and there is a proper retraction r in Case II. By Lemma 12, the number of edges between S_r and $[n] \setminus S_r$ is at most $\ell + 16d^6 \leq 5\ell/4$ for large enough n . Here, $\ell = n - |S_r|$. Since $\delta(G) \geq 2$, the sum of the degrees of the vertices in $[n] \setminus S_r$ is at least 2ℓ , and so the number of edges of G inside $[n] \setminus S_r$ is at least $3\ell/8$. Therefore, the occurrence of a Case-II retraction implies the existence of a set $U \subseteq [n]$ of size $64d^6 \leq |U| \leq n/2$ such that

$$X_1 := |E(G[U])| \geq \frac{3|U|}{8} \quad \text{and} \quad X_2 := |E_G(U, [n] \setminus U)| \leq \frac{5|U|}{4}, \quad (20)$$

where $E_G(U, [n] \setminus U)$ denotes the set of edges having one endpoint in U and the other endpoint in $[n] \setminus U$. We show that this is unlikely to happen, and hence prove Lemma 14.

First fix a set U of size ℓ . Note that the random edges in $E(G[U])$ and $E_G(U, [n] \setminus U)$ are mutually independent, and so the two counts $X_1 \sim \text{Bin}(\ell(\ell-1)/2, p)$ and $X_2 \sim \text{Bin}(\ell(n-\ell), p)$ are independent binomial distributions. To bound the probabilities of the two events in (20), we apply the Chernoff bound (1).

For X_1 , because $\mathbb{E}[X_1] < p\ell^2/2$, we have

$$\Pr \left[X_1 \geq \frac{3\ell}{8} \right] \leq \left(\frac{ep\ell^2/2}{3\ell/8} \right)^{3\ell/8} = \left(\frac{4ep\ell}{3} \right)^{3\ell/8}$$

by the upper tail bound when $\ell \leq 3/(4p)$. If this condition does not hold, then the right hand side is greater than 1, so the probability is trivially bounded by 1. For X_2 , because $\mathbb{E}[X_2] = d\ell(1 - \ell/n) \geq d\ell/2 > 5\ell/4$ for all sufficiently large n , we can always apply the lower tail bound and have

$$\Pr \left[X_2 \leq \frac{5\ell}{4} \right] \leq \exp \left\{ -d\ell \left(1 - \frac{\ell}{n} \right) \right\} \left(\frac{4ed(1 - \ell/n)}{5} \right)^{5\ell/4}. \quad (21)$$

As X_1 and X_2 are independent, the probability that the event in (20) happens for a fixed U is at most the product of the above two bounds.

By union bound, the probability of (20) happening for any U is at most

$$\sum_{64d^6 \leq \ell \leq 3/(4p)} \left(\frac{4ep\ell}{3} \right)^{3\ell/8} \exp \left\{ -d\ell \left(1 - \frac{\ell}{n} \right) \right\} \left(\frac{4ed(1 - \ell/n)}{5} \right)^{5\ell/4} \quad (22)$$

$$+ \sum_{3/(4p) \leq \ell \leq n/2} \exp \left\{ -d\ell \left(1 - \frac{\ell}{n} \right) \right\} \left(\frac{4ed(1 - \ell/n)}{5} \right)^{5\ell/4} \quad (23)$$

$$+ \sum_{n/2 \leq \ell \leq n-8/p} \exp \left\{ -d\ell \left(1 - \frac{\ell}{n} \right) \right\} \left(\frac{4ed(1 - \ell/n)}{5} \right)^{5\ell/4}. \quad (24)$$

The above summation is well-defined for large enough n ; that is, $3/(4p) \leq n/2 \leq n-8/p$ for n large enough.

To show (22) is $o(1)$, perform a straightforward calculation to see that the summand is $(e^\star)^\ell$ where

$$\star \leq 1.454 + \frac{3}{8} \log p\ell + \frac{5}{4} \log d \left(1 - \frac{\ell}{n} \right) - d \left(1 - \frac{\ell}{n} \right) \leq 1.347 + \frac{5}{4} \log d - \frac{d}{2}.$$

Here we use the condition $\ell \leq 3/(4p) \leq n/2$. This gives that (22) is upper bounded by $e^{-d/3} = o(1)$ for sufficiently large n . The second term (23) can be bounded by following almost the same lines.

For the third term, note that the function $-x + \frac{5}{4} \log x$ is decreasing for $x \geq 5/4$, and the condition on ℓ gives $d(1 - \ell/n) \geq 8 > 5/4$. Therefore, the summand in (24) is $(e^\star)^\ell$ with

$$\star = \frac{5}{4} \log \frac{4e}{5} + \frac{5}{4} \log d \left(1 - \frac{\ell}{n} \right) - d \left(1 - \frac{\ell}{n} \right) \leq \frac{5}{4} \log \frac{4e}{5} + \frac{5}{4} \log 8 - 8 \leq -4.42,$$

and the summation in (24) is therefore $o(1)$ as n grows.

4.5. Case III. This subsection proves Lemma 15.

A graph has *degeneracy* k if every non-empty subgraph has at least one vertex of degree at most k . As a standard fact, a k -degenerate graph has chromatic number at most $k+1$ by an induction on the number of vertices: by removing a vertex of degree at most k and colouring the rest of the graph recursively, the removed vertex will always have at most k neighbors already coloured, leaving at least one colour free out of $k+1$ total colours.

Suppose one such retraction r exists under $\mathcal{B}$. Recall the condition on the size of retraction $1 \leq |S_r| < 8/p$, i.e., it is extremely small. By Lemma 11(d), the chromatic number of the random graph G satisfies is at most $d/(5 \log d)$ for all sufficiently large n under $\mathcal{B}$.

On the other hand, Lemma 11(e) applies to every subset $U \subseteq S_r$. It shows that every non-empty induced subgraph of $G[S_r]$ has average degree strictly less than r_0 , and therefore has a vertex of degree at most $\lfloor r_0 \rfloor$. Consequently, the degeneracy of G is at most $\lfloor r_0 \rfloor$, meaning the graph $G[S_r]$ is $(\lfloor r_0 \rfloor + 1)$ -colourable. Since

$$\lfloor r_0 \rfloor + 1 = \left\lfloor \frac{30 \log d}{\log \log d} \right\rfloor + 1 < \frac{d}{5 \log d}$$

for all sufficiently large n , we obtain

$$\chi(G[S_r]) < \chi(G).$$

This contradicts the existence of the retraction: the retraction $G \rightarrow G[S_r]$ implies $\chi(G) \leq \chi(G[S_r])$, while the fact that $G[S_r]$ is an induced subgraph trivially implies the reverse inequality. Hence a retraction would force $\chi(G) = \chi(G[S_r])$. This proves Lemma 15.

REFERENCES

- [BK22] Andrei A. Bulatov and Amirhossein Kazeminia. Complexity classification of counting graph homomorphisms modulo a prime number. In Stefano Leonardi and Anupam Gupta, editors, *STOC '22: 54th Annual ACM SIGACT Symposium on Theory of Computing, Rome, Italy, June 20 - 24, 2022*, pages 1024–1037. ACM, 2022. doi:10.1145/3519935.3520075. 1
- [BP09] Anthony Bonato and Paweł Prałat. The good, the bad, and the great: homomorphisms and cores of random graphs. *Discrete Math.*, 309(18):5535–5539, 2009. doi:10.1016/j.disc.2008.03.026. 3
- [Bul17] Andrei A. Bulatov. A dichotomy theorem for nonuniform csp. In Chris Umans, editor, *58th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2017, Berkeley, CA, USA, October 15-17, 2017*, pages 319–330. IEEE Computer Society, 2017. doi:10.1109/FOCS.2017.37. 1
- [CC17] Jin-Yi Cai and Xi Chen. Complexity of counting CSP with complex weights. *J. ACM*, 64(3):19:1–19:39, 2017. doi:10.1145/2822891. 1
- [CCF⁺05] Jianer Chen, Benny Chor, Mike Fellows, Xiuzhen Huang, David W. Juedes, Iyad A. Kanj, and Ge Xia. Tight lower bounds for certain parameterized np-hard problems. *Inf. Comput.*, 201(2):216–231, 2005. URL: <https://doi.org/10.1016/j.ic.2005.05.001>, doi:10.1016/J.IC.2005.05.001. 2
- [CDM17] Radu Curticapean, Holger Dell, and Daniel Marx. Homomorphisms are a good basis for counting small subgraphs. In Hamed Hatami, Pierre McKenzie, and Valerie King, editors, *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing, STOC 2017, Montreal, QC, Canada, June 19-23, 2017*, pages 210–223. ACM, 2017. doi:10.1145/3055399.3055502. 2
- [CDNW26] Radu Curticapean, Simon Döring, Daniel Neuen, and Jiaheng Wang. Can you link up with treewidth? *J. Comput. Syst. Sci.*, 162:103835, 2026. URL: <https://doi.org/10.1016/j.jcss.2026.103835>, doi:10.1016/J.JCSS.2026.103835. 2, 3
- [CGL17] Yijia Chen, Martin Grohe, and Bingkai Lin. The hardness of embedding grids and walls. In Hans L. Bodlaender and Gerhard J. Woeginger, editors, *Graph-Theoretic Concepts in Computer Science - 43rd International Workshop, WG 2017, Eindhoven, The Netherlands, June 21-23, 2017, Revised Selected Papers*, volume 10520 of *Lecture Notes in Computer Science*, pages 180–192. Springer, 2017. doi:10.1007/978-3-319-68705-6_14. 3
- [CHKX06] Jianer Chen, Xiuzhen Huang, Iyad A. Kanj, and Ge Xia. Strong computational lower bounds via parameterized complexity. *J. Comput. Syst. Sci.*, 72(8):1346–1367, 2006. URL: <https://doi.org/10.1016/j.jcss.2006.04.007>, doi:10.1016/J.JCSS.2006.04.007. 2
- [CL01] Fan Chung and Linyuan Lu. The diameter of sparse random graphs. *Adv. in Appl. Math.*, 26(4):257–279, 2001. doi:10.1006/aama.2001.0720. 3
- [CM77] Ashok K. Chandra and Philip M. Merlin. Optimal implementation of conjunctive queries in relational data bases. In John E. Hopcroft, Emily P. Friedman, and Michael A. Harrison, editors, *Proceedings of the 9th Annual ACM Symposium on Theory of Computing, May 4-6, 1977, Boulder, Colorado, USA*, pages 77–90. ACM, 1977. doi:10.1145/800105.803397. 1
- [CR00] Chandra Chekuri and Anand Rajaraman. Conjunctive query containment revisited. *Theor. Comput. Sci.*, 239(2):211–229, 2000. doi:10.1016/S0304-3975(99)00220-0. 2
- [DF13] Rodney G. Downey and Michael R. Fellows. *Fundamentals of parameterized complexity*. Texts in Computer Science. Springer, London, 2013. doi:10.1007/978-1-4471-5559-1. 3
- [DG00] Martin E. Dyer and Catherine S. Greenhill. The complexity of counting graph homomorphisms. *Random Struct. Algorithms*, 17(3-4):260–289, 2000. doi:10.1002/1098-2418(200010/12)17:3/4<260::AID-RSA5>3.0.CO;2-W. 1
- [DJ04] Víctor Dalmau and Peter Jonsson. The complexity of counting homomorphisms seen from the other side. *Theor. Comput. Sci.*, 329(1-3):315–323, 2004. URL: <https://doi.org/10.1016/j.tcs.2004.08.008>, doi:10.1016/J.TCS.2004.08.008. 2
- [DKV02] Víctor Dalmau, Phokion G. Kolaitis, and Moshe Y. Vardi. Constraint satisfaction, bounded treewidth, and finite-variable logics. In Pascal Van Hentenryck, editor, *Principles and Practice of Constraint Programming - CP 2002, 8th International Conference, CP 2002, Ithaca, NY, USA, September 9-13, 2002, Proceedings*, volume 2470 of *Lecture Notes in Computer Science*, pages 310–326. Springer, 2002. doi:10.1007/3-540-46135-3_21. 2
- [ER60] P. Erdős and A. Rényi. On the evolution of random graphs. *Magyar Tud. Akad. Mat. Kutató Int. Közl.*, 5:17–61, 1960. 5
- [ER61] P. Erdős and A. Rényi. On the strength of connectedness of a random graph. *Acta Math. Acad. Sci. Hungar.*, 12:261–267, 1961. doi:10.1007/BF02066689. 3, 5
- [FG04] Jörg Flum and Martin Grohe. The parameterized complexity of counting problems. *SIAM J. Comput.*, 33(4):892–922, 2004. doi:10.1137/S0097539703427203. 2

- [FGRZ25] Jacob Focke, Leslie Ann Goldberg, Marc Roth, and Stanislav Zivný. Approximately counting answers to conjunctive queries with disequalities and negations. *ACM Trans. Algorithms*, 21(1):8:1–8:29, 2025. doi:[10.1145/3689634](https://doi.org/10.1145/3689634). 1
- [Fre90] Eugene C. Freuder. Complexity of k-tree structured constraint satisfaction problems. In Howard E. Shrobe, Thomas G. Dietterich, and William R. Swartout, editors, *Proceedings of the 8th National Conference on Artificial Intelligence. Boston, Massachusetts, USA, July 29 - August 3, 1990, 2 Volumes*, pages 4–9. AAAI Press / The MIT Press, 1990. URL: <http://www.aaai.org/Library/AAAI/1990/aaai90-001.php>. 2
- [FV98] Tomás Feder and Moshe Y. Vardi. The computational structure of monotone monadic SNP and constraint satisfaction: A study through datalog and group theory. *SIAM J. Comput.*, 28(1):57–104, 1998. doi:[10.1137/S0097539794266766](https://doi.org/10.1137/S0097539794266766). 1
- [GGJ17] Andreas Galanis, Leslie Ann Goldberg, and Mark Jerrum. A complexity trichotomy for approximately counting list H -colorings. *ACM Trans. Comput. Theory*, 9(2):9:1–9:22, 2017. doi:[10.1145/3037381](https://doi.org/10.1145/3037381). 1
- [GLS21] Andreas Göbel, J. A. Gregor Lagodzinski, and Karen Seidel. Counting homomorphisms to trees modulo a prime. *ACM Trans. Comput. Theory*, 13(3):19:1–19:33, 2021. doi:[10.1145/3460958](https://doi.org/10.1145/3460958). 1
- [Gro07] Martin Grohe. The complexity of homomorphism and constraint satisfaction problems seen from the other side. *J. ACM*, 54(1):1:1–1:24, 2007. doi:[10.1145/1206035.1206036](https://doi.org/10.1145/1206035.1206036). 2
- [HN90] Pavol Hell and Jaroslav Nešetřil. On the complexity of H -coloring. *J. Combin. Theory Ser. B*, 48(1):92–110, 1990. doi:[10.1016/0095-8956\(90\)90132-J](https://doi.org/10.1016/0095-8956(90)90132-J). 1
- [HN92] Pavol Hell and Jaroslav Nešetřil. The core of a graph. *Discrete Mathematics*, 109(1):117–126, 1992. URL: <https://www.sciencedirect.com/science/article/pii/0012365X9290282K>, doi:[10.1016/0012-365X\(92\)90282-K](https://doi.org/10.1016/0012-365X(92)90282-K). 2
- [HN04] Pavol Hell and Jaroslav Nešetřil. *Graphs and homomorphisms*, volume 28 of *Oxford Lecture Series in Mathematics and its Applications*. Oxford University Press, Oxford, 2004. doi:[10.1093/acprof:oso/9780198528173.001.0001](https://doi.org/10.1093/acprof:oso/9780198528173.001.0001). 1, 3, 5
- [IP01] Russell Impagliazzo and Ramamohan Paturi. On the complexity of k-sat. *J. Comput. Syst. Sci.*, 62(2):367–375, 2001. URL: <https://doi.org/10.1006/jcss.2000.1727>, doi:[10.1006/JCSS.2000.1727](https://doi.org/10.1006/JCSS.2000.1727). 2
- [IPZ01] Russell Impagliazzo, Ramamohan Paturi, and Francis Zane. Which problems have strongly exponential complexity? *J. Comput. Syst. Sci.*, 63(4):512–530, 2001. URL: <https://doi.org/10.1006/jcss.2001.1774>, doi:[10.1006/JCSS.2001.1774](https://doi.org/10.1006/JCSS.2001.1774). 2
- [KMPS24] Karthik C. S., Dániel Marx, Marcin Pilipczuk, and Uéverton S. Souza. Conditional lower bounds for sparse parameterized 2-csp: A streamlined proof. In Merav Parter and Seth Pettie, editors, *2024 Symposium on Simplicity in Algorithms, SOSA 2024, Alexandria, VA, USA, January 8-10, 2024*, pages 383–395. SIAM, 2024. doi:[10.1137/1.9781611977936.35](https://doi.org/10.1137/1.9781611977936.35). 2
- [KV00] Phokion G. Kolaitis and Moshe Y. Vardi. Conjunctive-query containment and constraint satisfaction. *J. Comput. Syst. Sci.*, 61(2):302–332, 2000. URL: <https://doi.org/10.1006/jcss.2000.1713>, doi:[10.1006/JCSS.2000.1713](https://doi.org/10.1006/JCSS.2000.1713). 1
- [Lin18] Bingkai Lin. The parameterized complexity of the k -biclique problem. *J. ACM*, 65(5):34:1–34:23, 2018. doi:[10.1145/3212622](https://doi.org/10.1145/3212622). 3
- [Lov12] László Lovász. *Large networks and graph limits*, volume 60 of *American Mathematical Society Colloquium Publications*. American Mathematical Society, Providence, RI, 2012. doi:[10.1090/coll/060](https://doi.org/10.1090/coll/060). 1
- [Mar10] Dániel Marx. Can you beat treewidth? *Theory Comput.*, 6(1):85–112, 2010. URL: <https://doi.org/10.4086/toc.2010.v006a005>, doi:[10.4086/TOC.2010.V006A005](https://doi.org/10.4086/TOC.2010.V006A005). 2
- [MP14] Dániel Marx and Michał Pilipczuk. Everything you always wanted to know about the parameterized complexity of subgraph isomorphism (but were afraid to ask). In Ernst W. Mayr and Natacha Portier, editors, *31st International Symposium on Theoretical Aspects of Computer Science, STACS 2014, Lyon, France, March 5-8, 2014*, volume 25 of *LIPIcs*, pages 542–553. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2014. URL: <https://doi.org/10.4230/LIPIcs.STACS.2014.542>, doi:[10.4230/LIPICS.STACS.2014.542](https://doi.org/10.4230/LIPICS.STACS.2014.542). 2
- [MS65] T. S. Motzkin and E. G. Straus. Maxima for graphs and a new proof of a theorem of Turán. *Canadian J. Math.*, 17:533–540, 1965. doi:[10.4153/CJM-1965-053-6](https://doi.org/10.4153/CJM-1965-053-6). 10
- [PK21] Esmail Parsa and P. Mark Kayll. On random digraphs and cores. *Australas. J. Combin.*, 79:371–379, 2021. 3
- [Sch78] Thomas J. Schaefer. The complexity of satisfiability problems. In Richard J. Lipton, Walter A. Burkhard, Walter J. Savitch, Emily P. Friedman, and Alfred V. Aho, editors, *Proceedings of the 10th Annual ACM Symposium on Theory of Computing, May 1-3, 1978, San Diego, California, USA*, pages 216–226. ACM, 1978. doi:[10.1145/800133.804350](https://doi.org/10.1145/800133.804350). 1
- [Zhu20] Dmitriy Zhuk. A proof of the CSP dichotomy conjecture. *J. ACM*, 67(5):30:1–30:78, 2020. doi:[10.1145/3402029](https://doi.org/10.1145/3402029). 1