Math. Stat. Learn. 4 (2021), 221–251
DOI 10.4171/MSL/25

Optimal low-degree hardness of maximum independent set

Alexander S. Wein

Abstract. We study the algorithmic task of finding a large independent set in a sparse Erdős–Rényi random graph with n vertices and average degree d. The maximum independent set is known to have size $(2\log d/d)n$ in the double limit $n \to \infty$ followed by $d \to \infty$, but the best known polynomial-time algorithms can only find an independent set of half-optimal size $(\log d/d)n$. We show that the class of *low-degree polynomial algorithms* can find independent sets of half-optimal size but no larger, improving upon a result of Gamarnik, Jagannath, and the author. This generalizes earlier work by Rahman and Virág, which proved the analogous result for the weaker class of *local algorithms*.

1. Introduction

We consider the problem of finding a large independent set (i.e., a set of vertices such that no two are adjacent) in the sparse Erdős–Rényi graph $G(n, d/n)$, where each of the $\binom{n}{2}$ potential edges on vertex set $[n]$ occurs independently with probability d/n. In the double limit $n \to \infty$ followed by $d \to \infty$, the largest independent set $S_{\max}$ is known to have asymptotic size $(2\log d/d)n$. More precisely, as $n \to \infty$ with $d > 0$ fixed we have $\frac{1}{n}|S_{\max}| \to \alpha_d$ with high probability, for some α_d satisfying $\alpha_d = (1 + o_d(1))(2\log d/d)$ as $d \to \infty$ [6, 20]. We will be interested in the associated algorithmic task: give a polynomial-time algorithm that takes as input a graph drawn from $G(n, d/n)$ and outputs (with high probability) a large independent set. We assume d is known to the algorithm, although it can be estimated easily from the total number of edges. The influential work of Karp [31] showed that a simple greedy algorithm can find an independent set of asymptotic size $(\log d/d)n$, which is half of the optimum. Decades later, we still do not know a polynomial-time algorithm to find an independent set of size $(1 + \varepsilon)(\log d/d)n$ for any fixed $\varepsilon > 0$ (independent of both d and n). Moreover, evidence has emerged to suggest that no such algorithm exists. It was shown by Coja-Oghlan and Efthymiou [15] (building on [1]) that the

2020 Mathematics Subject Classification. 68Q87.
Keywords. Low-degree polynomial, independent set, random graph, overlap gap property.

independent sets of size larger than half-optimal are "clustered" in a way that implies slow mixing of the Metropolis process for sampling such sets. Furthermore, it was shown by Rahman and Virág [39] (building on [24, 33]) that the class of *local algorithms* can find independent sets of half-optimal size *and no larger*. Here, a local algorithm (also called *i.i.d. factors*) allows each vertex to decide whether or not to include itself in the set based only on its local neighborhood in the graph (of constant radius) along with i.i.d. random variables attached to the vertices (see Section 3 for a formal definition).

The above results suggest that $(\log d/d)n$ may be the fundamental limit for polynomial-time algorithms. In this work we provide further evidence for this by showing that $(\log d/d)n$ is the fundamental limit for the class of *low-degree polynomial algorithms* (to be defined formally in the next section) where each vertex's membership (or non-membership) in the independent set is determined by thresholding a low-degree multivariate polynomial of the edge-indicator variables that describe the input graph. This class of low-degree algorithms includes the class of local algorithms mentioned above (see Remark 3.2), and also (as discussed in [22, Appendix A]) includes other popular algorithmic paradigms such as approximate message passing[1] (e.g., [7, 18, 30, 35, 37]) and power iteration[2]. Furthermore, starting from the influential line of work [5, 26, 27, 29], it has been established that low-degree algorithms (with degree logarithmic in the dimension) are precisely as powerful as the best known polynomial-time algorithms for a number of problems in high-dimensional statistics including planted clique, sparse PCA, community detection, tensor PCA, and many others [3, 4, 11, 12, 14, 17, 26, 27, 29, 32, 34, 41]. Thus, failure of low-degree algorithms is a form of concrete evidence for computational hardness of statistical problems. For more on low-degree algorithms, we refer the reader to [32] (for a survey on the setting of *hypothesis testing*), [41] (for the setting of *estimation*), or [22] (for the setting of random optimization problems, which is the relevant setting for this work).

Most prior work on low-degree algorithms has focused on problems with a "planted" signal, in which case failure of low-degree algorithms can be shown via a direct linear-algebraic computation. This technique does not apply to "non-planted" problems such as the maximum independent set problem that we consider here, and so a different approach is needed which leverages structural properties of the solution

[1] Notably, approximate message passing (AMP) algorithms include the algorithms of [18, 37] (which build on the earlier works [2, 42]) for optimizing the Sherrington–Kirkpatrick and p-spin models of spin glasses.

[2] Notably, low-degree algorithms capture power iteration on *any* matrix that is itself low-degree in the input. This allows for non-trivial spectral methods such as the *tensor unfolding* method for tensor PCA [28, 40], which outperforms more "standard" algorithms such as message passing and gradient descent [8, 40].

space (see Section 1.2). For non-planted problems, the first results for low-degree algorithms were given by Gamarnik, Jagannath, and the author [22] (building on [21]), who showed that low-degree algorithms cannot find independent sets of size exceeding $(1 + 1/\sqrt{2})(\log d/d)n$ in $G(n, d/n)$. Here we improve this to the optimal threshold $(\log d/d)n$. We also provide the matching positive result, showing that $(\log d/d)n$ is achievable by low-degree algorithms (following a proof sketch given in [22]). This is the first non-planted problem for which matching upper and lower bounds have been obtained on the objective value attainable by low-degree algorithms (apart from trivial cases where the global optimum value can be reached). One conceptual advantage of our results over the existing results for local algorithms is that low-degree algorithms offer a unified framework to explain computational hardness in a wide variety of high-dimensional problems, whereas local algorithms are specific to problems involving sparse graphs. This is exemplified by the fact that our impossibility result can be extended to the case of dense graphs such as $G(n, 1/2)$; see Section 1.3.

1.1. Main results

We now formally define the problem setup, following [22]. We say that a function $f: \mathbb{R}^m \to \mathbb{R}^n$ is a *polynomial of degree (at most) D* if it may be written in the form

$$f(Y) = (f_1(Y), \dots, f_n(Y)), \tag{1}$$

where each $f_i: \mathbb{R}^m \to \mathbb{R}$ is a multivariate polynomial (in the usual sense) of degree at most D with real coefficients. We also define a *random polynomial* $f: \mathbb{R}^m \to \mathbb{R}^n$ in the same way but where the coefficients may be random (but independent from the input Y): formally, for some probability space $(\Omega, \mathbb{P}_\omega)$, f is a map $f: \mathbb{R}^m \times \Omega \to \mathbb{R}^n$ such that $f(\cdot, \omega)$ is a degree-D polynomial for each "seed" $\omega \in \Omega$. (We will see that randomness does not actually help; see Lemma 2.11.)

For our purposes, the input to f will be an n-vertex graph encoded as $Y \in \{0,1\}^m$ with $m = \binom{n}{2}$, where each entry of Y is the indicator variable for the presence of a particular edge. We write $Y \sim G(n, d/n)$ for an Erdős–Rényi graph, i.e., the entries of Y are i.i.d. Bernoulli(d/n).

We need to define what it means for a polynomial $f: \mathbb{R}^m \to \mathbb{R}^n$ to find an independent set in a graph Y. Instead of asking $f(Y)$ to be the indicator vector of an independent set, we relax this somewhat and ask only for a "near-indicator vector" of a "near-independent set". More precisely, the following "rounding" procedure from [22] will be used to extract an independent set from the output of f.

Definition 1.1. Let $f: \mathbb{R}^m \to \mathbb{R}^n$ be a random polynomial with $m = \binom{n}{2}$. For $Y \in \{0,1\}^m$, and $\eta \geq 0$, let $V_f^\eta(Y, \omega)$ be the independent set in the graph Y obtained by

the following procedure. Let

$$A = \{i \in [n] : f_i(Y, \omega) \geq 1\},$$
$$\widetilde{A} = \{i \in A : i \text{ has no neighbors in } A \text{ in the graph } Y\},$$

and

$$B = \{i \in [n] : f_i(Y, \omega) \in (1/2, 1)\}.$$

Then define

$$V_f^{\eta}(Y, \omega) = \begin{cases} \widetilde{A} & \text{if } |A \setminus \widetilde{A}| + |B| \leq \eta n, \\ \emptyset & \text{otherwise.} \end{cases} \tag{2}$$

Informally speaking, f_i should output a value ≥ 1 to indicate that vertex i is in the independent set and should output a value $\leq 1/2$ to indicate that it is not; the set A can be thought of as a candidate independent set. We allow a small number of "errors": there can be up to ηn vertices where either $f_i(Y) \in (1/2, 1)$ (this happens for $i \in B$) or the independent set constraint is violated (this happens for $i \in A \setminus \widetilde{A}$). Vertices that violate the independent set constraint (that is, $i \in A \setminus \widetilde{A}$) are thrown out, and if too many errors are made then the output is the empty set $\emptyset$ (which is thought of as a "failure" event). While the choice of thresholds 1 and $1/2$ is somewhat arbitrary, the interval $(1/2, 1)$ of disallowed outputs is important for our impossibility result (Theorem 1.3), as this ensures that a small change in $f(Y, \omega)$ cannot induce a large change in the resulting independent set $V_f^{\eta}(Y, \omega)$ without encountering the failure event $\emptyset$. On the other hand, our achievability result (Theorem 1.4) will give a low-degree polynomial for which most outputs $f_i(Y)$ lie in $\{0, 1\}$ exactly, i.e., it succeeds even under the more stringent definitions

$$A = \{i \in [n] : f_i(Y, \omega) = 1\} \quad \text{and} \quad B = \{i \in [n] : f_i(Y, \omega) \notin \{0, 1\}\}.$$

Definition 1.2. For parameters $k > 0$, $\delta \in [0, 1]$, $\gamma \geq 1$, and $\eta > 0$, a random polynomial $f : \mathbb{R}^m \to \mathbb{R}^n$ is said to $(k, \delta, \gamma, \eta)$-*optimize* the independent set problem in $G(n, d/n)$ if the following are satisfied when $Y \sim G(n, d/n)$:

- $\mathbb{E}_{Y,\omega}\left[\|f(Y, \omega)\|^2\right] \leq \gamma k$, and
- $\mathbb{P}_{Y,\omega}\left[|V_f^{\eta}(Y, \omega)| \geq k\right] \geq 1 - \delta$.

Here, k is the size of the independent set that is produced, δ is the algorithm's failure probability, γ is a normalization parameter, and η is the error tolerance of the rounding procedure V_f^{η}.

We now state our main results. Theorem 1.3 shows that no low-degree polynomial can find an independent set of size $(1 + \varepsilon)\frac{\log d}{d} n$, while Theorem 1.4 shows that some low-degree polynomial can find an independent set of size $(1 - \varepsilon)\frac{\log d}{d} n$. The proofs

are given in Sections 2 and 3, respectively. The results are interpreted in the remarks below.

Theorem 1.3 (Impossibility). *For any $\varepsilon > 0$, there exists $d^* > 0$ such that for any $d \geq d^*$, there exists $n^* > 0$, $\eta > 0$, $C_1 > 0$, and $C_2 > 0$ (depending on ε, d) such that the following holds. Let $n \geq n^*$, $\gamma \geq 1$, and $1 \leq D \leq \frac{C_1 n}{\gamma \log n}$, and suppose $\delta \geq 0$ satisfies*

$$\delta \leq \exp(-C_2 \gamma D \log n). \tag{3}$$

Then for $k = (1+\varepsilon)\frac{\log d}{d} n$, there is no random degree-D polynomial that $(k, \delta, \gamma, \eta)$-optimizes the independent set problem in $G(n, d/n)$.

We emphasize that the constants η, C_1, C_2 above do not depend on n or D.

Theorem 1.4 (Achievability). *For any $\varepsilon > 0$ there exists $d^* > 0$ such that for any $d \geq d^*$ and any $\eta > 0$ there exists $n^* > 0$, $D > 0$, $\gamma \geq 1$, and $C > 0$ (depending on ε, d, η) such that the following holds for all $n \geq n^*$. For $k = (1-\varepsilon)\frac{\log d}{d} n$ and $\delta = \exp(-C n^{1/3})$, there exists a (deterministic) degree-D polynomial that $(k, \delta, \gamma, \eta)$-optimizes the independent set problem in $G(n, d/n)$.*

Remark 1.5. The results are non-asymptotic but can be thought of as capturing the double limit $n \to \infty$ followed by $d \to \infty$. In other words, d is a large constant depending on ε, and n must then be chosen sufficiently large (where "sufficiently large" depends on d). In the sequel, asymptotic notation such as $O(\cdot)$ pertains to the limit $n \to \infty$ with all other parameters fixed; parameters not depending on n are considered "constants".

Remark 1.6. The "tolerance" parameter η should be thought of as a small constant. The impossibility result shows that *some* $\eta > 0$ (depending on ε, d) is *not* achievable (for very small failure probability δ verifying (3); see Remark 1.7), whereas the achievability result shows that *any* $\eta > 0$ is achievable. The "normalization" parameter γ should be thought of as a large constant. The impossibility result shows that *any* $\gamma \geq 1$ is not achievable (again, for small δ), whereas the achievability result shows that *some* $\gamma \geq 1$ (depending on ε, d, η) is achievable.

Remark 1.7. Typically, when proving impossibility results for low-degree algorithms, the goal is to rule out any degree $D = O(\log n)$ because polynomials of this degree can capture the best known algorithms for a wide array of problems. In our case, a constant degree D (depending on ε, d, η) is sufficient for the achievability result. On the other hand, our impossibility result rules out a much wider range of D values: $D \lesssim n/\log n$.

However, the requirement (3) gives an additional tradeoff between D and the failure probability δ. This is present for technical reasons, and ideally we would

replace (3) by a milder condition such as $\delta = o(1)$. Still, note that the parameters $D = O(1)$ and $\delta = \exp(-\Omega(n^{1/3}))$ in our achievability result lie well within the set of (D, δ) pairs ruled out by our impossibility result. More explicitly, if $D > 0$ and $\gamma \geq 1$ are constants and $\delta = \exp(-Cn^{1/3})$ for a constant $C > 0$ (these are the parameters from the achievability result), then (3) is satisfied for all sufficiently large n (that is, for all $n \geq n^*$ where n^* is a constant depending on D, γ, C and on the constant C_2 from (3)).

Remark 1.8. In the achievability result, the value of δ can likely be improved from $\exp(-Cn^{1/3})$ to $\exp(-Cn)$. This can perhaps be accomplished by using the powerful machinery of [9] in place of Corollary 3.5, but we do not attempt this here.

1.2. Proof techniques

We now give an overview of the proof techniques and discuss their relation to prior work. We first discuss the achievability result (Theorem 1.4). It is known already that local algorithms can find independent sets of size $(1-\varepsilon)\frac{\log d}{d}n$ [39]. Our proof transfers this to a result about low-degree algorithms by showing that *any* local algorithm can be well-approximated by a constant-degree polynomial. A proof sketch of this reduction was given already in Appendix A of [22], but here we give the full details and determine the values of the parameters D, δ, γ, η. The main difficulty lies in establishing that the failure probability δ is very small; for this we appeal to a result of [10] that gives tail bounds for certain "local" functions on sparse random graphs.

We now discuss the impossibility result (Theorem 1.3), which is our main contribution. This result falls into a line of work initiated by Gamarnik and Sudan [24], who showed that local algorithms fail to find independent sets larger than $(1+1/\sqrt{2})\frac{\log d}{d}n$. Their proof harnessed the so-called *overlap gap property (OGP)*: in a typical graph drawn from $G(n, d/n)$, there are no two independent sets that each have size exceeding $(1+1/\sqrt{2})\frac{\log d}{d}n$ and have intersection size ("overlap") roughly $\frac{\log d}{d}n$. They used an interpolation argument to show that if a hypothetical local algorithm were to succeed at finding independent sets larger than $(1+1/\sqrt{2})\frac{\log d}{d}n$, this could be used to construct two independent sets violating the OGP, leading to a contradiction. This proof technique was subsequently extended in two important ways. First, Rahman and Virág [39] improved the threshold for failure of local algorithms down to $\frac{\log d}{d}n$, which is optimal. The proof involves establishing a more intricate "forbidden" structure that involves many independent sets with a particular intersection pattern (in contrast to the OGP, which involves only two sets). Again, a hypothetical local algorithm can be used to construct this forbidden structure, leading to a contradiction. This idea inspired further work in the area of random constraint satisfaction problems [16, 23].

A separate line of work [13, 21, 22] extended the ideas of Gamarnik and Sudan [24] in a different direction: instead of the basic OGP discussed above, they consider an "ensemble" variant of OGP in which a particular overlap between two large independent sets is forbidden even when the independent sets do not come from the same graph but from two correlated random graphs. This variant of OGP can be used not only to rule out local algorithms, but also to rule out any sufficiently "stable" algorithm (which roughly means that a small change to the input only causes a small change to the output); this idea was used by [21] to rule out message-passing algorithms and later by [22] to show that low-degree algorithms – which are stable – cannot find independent sets larger than $(1 + 1/\sqrt{2})\frac{\log d}{d} n$.

To prove our impossibility result, we combine the two main ideas discussed above: we consider a forbidden structure that involves many independent sets and also involves many correlated random graphs. The crux of the proof lies in the specific choice of this forbidden structure (see Proposition 2.3), which is carefully chosen so that (i) with high probability, no instance of this structure occurs, and (ii) a hypothetical stable algorithm can be used to construct an instance of this structure, leading to a contradiction. On a technical level, our forbidden structure is quite different from the one used by Rahman and Virág [39] in that theirs is highly symmetric, e.g., any two of the sets involved have the same intersection size. This is suitable for their purposes because due to special properties of local algorithms, a hypothetical local algorithm can be used to construct such a symmetric collection of sets. In our case, however, it is not clear that a hypothetical low-degree algorithm can be used to construct a symmetric collection of sets; we instead define a new class of forbidden structures that are not necessarily symmetric. Finally, we remark that the only property of low-degree polynomials that we use is their "stability" (in the sense of Proposition 2.6), and so the proof actually rules out all "stable" algorithms.

1.3. Extensions and future directions

In this work we have given the first techniques for obtaining sharp impossibility results for low-degree algorithms in random optimization problems (with no planted signal). Hopefully these techniques can be adapted to other non-planted settings such as random constraint satisfaction problems (e.g., [1, 16, 23]) and spin glass optimization problems [18, 21, 22, 37, 42]. Low-degree algorithms are a promising candidate for a unified framework to explain computational hardness in a wide array of non-planted problems, analogous to the more established low-degree framework for planted problems.

One possible extension of our results is to consider the same independent set problem but in denser graphs. For instance, in $G(n, 1/2)$ the largest independent set has size $2\log_2 n$, but the best known polynomial-time algorithm is a simple greedy algo-

rithm which can find an independent set of half-optimal size $\log_2 n$ [31]. An argument nearly identical to the proof of Theorem 1.3 yields the following result which shows that low-degree algorithms cannot improve upon this.

Theorem 1.9. *For any $\varepsilon > 0$ there exists $n^* > 0$, $\widetilde{\eta} > 0$, $C_1 > 0$, and $C_2 > 0$ (depending on ε) such that the following holds. Let $n \geq n^*$, $\gamma \geq 1$, and $1 \leq D \leq \frac{C_1 \log^2 n}{\gamma}$, and suppose $\delta \geq 0$ satisfies*

$$\delta \leq \exp(-C_2 \gamma D - 2 \log n).$$

Then for $k = (1+\varepsilon)\log_2 n$ and $\eta = \widetilde{\eta} \log_2 n / n$, there is no random degree-D polynomial that $(k, \delta, \gamma, \eta)$-optimizes the independent set problem in $G(n, 1/2)$.

However, the matching achievability result remains open: it is not clear how to write the greedy algorithm as a low-degree polynomial or otherwise give a low-degree algorithm that finds an independent set of size $(1-\varepsilon)\log_2 n$. We expect that it should be possible to obtain such a low-degree algorithm (perhaps of degree $D = O(\log n)$ and failure probability $\delta = \exp(-\Omega(\log^2 n))$) via the approximate message passing framework, which has been successful in other non-planted settings [18, 19, 37, 38].

Notation

Asymptotic notation such as $o(1)$ or $\Omega(n)$ pertains to the limit $n \to \infty$ with all other parameters (such as d) held fixed; in other words, parameters not depending on n are considered "constants" and may be hidden by this notation. On the other hand, $o_d(1)$ denotes a quantity that depends on d but not n, and tends to 0 as $d \to \infty$ (with all other parameters held fixed).

Throughout, we will use the shorthand $m = \binom{n}{2}$ and $\Phi = \frac{\log d}{d} n$. We define $[n] = \{1, 2, \dots, n\}$ and use $\|\cdot\|$ for the ℓ^2-norm of a vector. All logarithms use the natural base unless stated otherwise. All graphs are assumed to have no self-loops nor parallel edges.

2. Proof of impossibility

In this section we prove our main impossibility result (Theorem 1.3) which shows that low-degree algorithms cannot find independent sets of size $(1+\varepsilon)\frac{\log d}{d} n$.

2.1. Interpolation path

Here we define a sequence of correlated random graphs that will be central to the argument. We will represent a graph on vertex set $[n]$ by $Y \in \{0,1\}^m$ where $m = \binom{n}{2}$.

Here $Y_1, \dots, Y_m$ are indicator variables for the edges (where 0 indicates a non-edge and 1 indicates an edge), listed in some fixed but arbitrary order.

Definition 2.1. For $T \in \mathbb{N}$, consider the *length-T interpolation path* $Y^{(0)}, \dots, Y^{(T)}$ sampled as follows. First, $Y^{(0)} \sim G(n, d/n)$. Then for $1 \le t \le T$, $Y^{(t)}$ is obtained from $Y^{(t-1)}$ by resampling coordinate $\sigma(t) \in [m]$ from Bernoulli(d/n). Here $\sigma(t) = t - k_t m$, where k_t is the unique integer for which $1 \le \sigma(t) \le m$.

2.2. Forbidden structures

The proof will hinge on the non-existence of certain structures (primarily the one defined in Proposition 2.3) with high probability over the interpolation path. The following standard bounds will be used repeatedly:

$$\binom{n}{k} \le \left(\frac{en}{k}\right)^k \qquad \text{for all integers } 1 \le k \le n, \tag{4}$$

$$(1-x)^r \le \exp(-rx) \quad \text{for all } x \in \mathbb{R},\ r > 0. \tag{5}$$

We start with a well-known upper bound on the size of the maximum independent set in $G(n, d/n)$.

Lemma 2.2. *Fix $\varepsilon > 0$. If $d > 0$ is a sufficiently large constant (depending on ε), then with probability $1 - \exp(-\Omega(n))$ there is no independent set in $G(n, d/n)$ of size exceeding $(2+\varepsilon)\frac{\log d}{d} n$.*

Lemma 2.2 is a folklore result that can be proved via a straightforward first moment calculation. We include the proof for completeness.

Proof. Let $\Phi = \frac{\log d}{d} n$ and define $a \ge 2 + \varepsilon$ so that $a\Phi = \lceil (2+\varepsilon)\Phi \rceil$. Let N denote the number of independent sets of size exactly $a\Phi$; our goal is to show $N = 0$ with high probability. The proof will use a simple first moment method: we compute $\mathbb{E}[N]$ and show that it is exponentially small. We have

$$\begin{aligned}
\mathbb{E}[N] &= \binom{n}{a\Phi}(1-d/n)^{\binom{a\Phi}{2}} \\
&\le \left(\frac{en}{a\Phi}\right)^{a\Phi} \exp\left(-\frac{d}{n}\binom{a\Phi}{2}\right) && \text{using (4) and (5)} \\
&= \exp\left[a\Phi \log\left(\frac{ed}{a \log d}\right) - \frac{da^2\Phi^2}{2n} + O(1)\right] \\
&= \exp\left[\Phi \log d \left(a - \frac{a^2}{2} + o(1) + o_d(1)\right)\right]
\end{aligned}$$

$$\begin{aligned} &\le \exp\left[\Phi \log d\left(-\varepsilon + o(1) + o_d(1)\right)\right] && \text{using } a \ge 2+\varepsilon \\ &= \exp(-\Omega(n)) \end{aligned}$$

for sufficiently large d. The result follows by Markov's inequality. ∎

The forbidden structure defined in the following result will be the crux of the proof.

Proposition 2.3. *Fix constants $\varepsilon > 0$ and $K \in \mathbb{N}$ with $K \ge 1 + 5/\varepsilon^2$. Consider the interpolation path $Y^{(0)}, \dots, Y^{(T)}$ from Definition 2.1, of any length $T = n^{O(1)}$. If $d > 0$ is a sufficiently large constant (depending on ε, K), then with probability $1 - \exp(-\Omega(n))$ there does not exist a sequence of sets $S_1, \dots, S_K \subseteq [n]$ satisfying the following properties:*

(i) *for each $k \in [K]$ there exists $0 \le t_k \le T$ such that S_k is an independent set in $Y^{(t_k)}$,*

(ii) *$|S_k| \ge (1+\varepsilon)\frac{\log d}{d} n$ for all $k \in [K]$, and*

(iii) *$|S_k \setminus (\cup_{\ell<k} S_\ell)| \in \left[\frac{\varepsilon}{4}\frac{\log d}{d} n, \frac{\varepsilon}{2}\frac{\log d}{d} n\right]$ for all $2 \le k \le K$.*

Proof. Let N denote the number of sequences $(S_1, \dots, S_K)$ satisfying the properties (i)–(iii). The proof will use the first moment method: we compute $\mathbb{E}[N]$ and show that it is exponentially small. Let $\Phi = \frac{\log d}{d} n$. Let a_k and b_k be defined by

$$|S_k| = a_k \Phi \quad \text{and} \quad |S_k \setminus (\cup_{\ell<k} S_\ell)| = b_k \Phi,$$

and note that (ii) and (iii) state that $a_k \ge 1+\varepsilon$ and $b_k \in \left[\frac{\varepsilon}{4}, \frac{\varepsilon}{2}\right]$. Also let c be defined by $|\cup_k S_k| = c\Phi$, and note that (iii) implies $c \le a_1 + (K-1)\frac{\varepsilon}{2}$. By Lemma 2.2, we can assume $a_k \le 2+\varepsilon$. Thus, c is upper-bounded by a constant $C(\varepsilon, K)$ that does not depend on d. We need to count the number of sequences $(S_1, \dots, S_K)$. There are at most n^{2K} choices for the values $\{a_k\}$ and $\{b_k\}$. Once $\{a_k\}$ and $\{b_k\}$ are fixed, the number of ways to choose $\{S_k\}$ is at most

$$\begin{aligned} &\binom{n}{a_1\Phi} \prod_{k=2}^{K} \binom{n}{b_k\Phi} \binom{c\Phi}{(a_k-b_k)\Phi} \\ &\quad \le \left(\frac{en}{a_1\Phi}\right)^{a_1\Phi} \prod_{k=2}^{K} \left(\frac{en}{b_k\Phi}\right)^{b_k\Phi} \left(\frac{ec}{a_k-b_k}\right)^{(a_k-b_k)\Phi} && \text{using (4)} \\ &\quad = \exp\left\{a_1\Phi \log\left(\frac{ed}{a_1 \log d}\right)\right. \\ &\qquad \left. + \sum_{k=2}^{K}\left[b_k\Phi \log\left(\frac{ed}{b_k \log d}\right) + (a_k-b_k)\Phi \log\left(\frac{ec}{a_k-b_k}\right)\right]\right\} \end{aligned}$$

$$= \exp\left\{\Phi \log d\left(a_1 + \sum_{k=2}^{K} b_k + o_d(1)\right)\right\},$$

where we have used $a_k \in [1+\varepsilon, 2+\varepsilon]$, $b_k \in \left[\frac{\varepsilon}{4}, \frac{\varepsilon}{2}\right]$, and $c \le C(\varepsilon, K)$ to conclude that certain terms are $o_d(1)$.

Now for a fixed $\{S_k\}$ satisfying (ii) and (iii), we need to upper-bound the probability that (i) is satisfied. We will take a union bound over the possible choices of $\{t_k\}$ in property (i); there are $(T+1)^K$ such choices. Let E be the number of edges $j \in \binom{[n]}{2}$ of the complete graph such that there exists k such that both endpoints of j lie within S_k. For fixed $\{S_k\}$ and $\{t_k\}$, property (i) occurs iff a certain collection of (at least) E independent non-edges occur in the sampling of $\{Y^{(t)}\}$; this happens with probability at most $(1-d/n)^E \le \exp(-Ed/n)$. Furthermore, we have

$$\begin{aligned} E &\ge \binom{a_1\Phi}{2} + \sum_{k=2}^{K} b_k(a_k - b_k)\Phi^2 \\ &= \frac{a_1^2\Phi^2}{2} - O(n) + \sum_{k=2}^{K} b_k(a_k - b_k)\Phi^2 \\ &= \frac{n}{d} \cdot \Phi \log d\left(\frac{a_1^2}{2} + \sum_{k=2}^{K} b_k(a_k - b_k) - o(1)\right), \end{aligned}$$

where in the first step, the first term counts edges within S_1 and the kth term of the sum counts edges within S_k that have exactly one endpoint in $\cup_{\ell<k} S_\ell$. (Note that no edges are double-counted here.)

Putting it all together, we have

$$\mathbb{E}[N] \le n^{2K}(T+1)^K \sup_{\{a_k\},\{b_k\}} \exp\left\{\Phi \log d\left(a_1 + \sum_{k=2}^{K} b_k + o_d(1)\right)\right\} \exp\left(-\frac{d}{n}E\right)$$

where $\{a_k\}$ and $\{b_k\}$ are subject to the constraints $a_k \ge 1+\varepsilon$ and $b_k \in \left[\frac{\varepsilon}{4}, \frac{\varepsilon}{2}\right]$

$$\begin{aligned} &\le n^{2K}(T+1)^K \sup_{\{a_k\},\{b_k\}} \exp\left\{\Phi \log d\left(a_1 + \sum_{k=2}^{K} b_k - \frac{a_1^2}{2}\right.\right. \\ &\qquad \left.\left. - \sum_{k=2}^{K} b_k(a_k - b_k) + o(1) + o_d(1)\right)\right\} \end{aligned}$$

$$= n^{2K}(T+1)^K \sup_{\{a_k\},\{b_k\}} \exp\left\{\Phi \log d\left(a_1 - \frac{a_1^2}{2} - \sum_{k=2}^{K} b_k(a_k - b_k - 1) + o(1) + o_d(1)\right)\right\},$$

$$\le n^{2K}(T+1)^K \exp\left\{\Phi \log d\left(\frac{1}{2} - \sum_{k=2}^{K} \frac{\varepsilon^2}{8} + o(1) + o_d(1)\right)\right\}$$

where we have used the fact $\sup_{a\in\mathbb{R}}\left(a - \frac{a^2}{2}\right) = \frac{1}{2}$ along with $a_k \ge 1+\varepsilon$ and $b_k \in \left[\frac{\varepsilon}{4}, \frac{\varepsilon}{2}\right]$

$$\le n^{2K}(T+1)^K \exp\left\{\Phi \log d\left(-\frac{1}{8} + o(1) + o_d(1)\right)\right\},$$

where we have used $K \ge 1 + 5/\varepsilon^2$

$$= \exp(-\Omega(n))$$

for sufficiently large d. The result follows by Markov's inequality. ∎

Finally, we will need the following simple result which states that no independent set of $G(n, d/n)$ has large intersection with a fixed set of vertices.

Lemma 2.4. *Fix constants $\varepsilon > 0$ and $a > 0$. Fix $S \subseteq [n]$ with $|S| \le a\frac{\log d}{d}n$. If $d > 0$ is a sufficiently large constant (depending on ε, a), then with probability $1-\exp(-\Omega(n))$ there is no independent set S' in $G(n, d/n)$ satisfying $|S \cap S'| \ge \varepsilon\frac{\log d}{d}n$.*

Proof. The proof is similar to that of Lemma 2.2. As usual, define $\Phi = \frac{\log d}{d}n$. We again use the first moment method. Let N be the number of subsets $U \subseteq S$ such that $|U| = \lceil \varepsilon\Phi \rceil =: b\Phi$ and U is an independent set in $G(n, d/n)$; it is sufficient to show $N = 0$ with high probability. We have

$$\begin{aligned}
\mathbb{E}[N] &= \binom{|S|}{b\Phi}(1-d/n)^{\binom{b\Phi}{2}} \\
&\le \left(\frac{ea}{b}\right)^{b\Phi} \exp\left(-\frac{d}{n}\binom{b\Phi}{2}\right) && \text{using (4) and (5)} \\
&= \exp\left[b\Phi \log\left(\frac{ea}{b}\right) - \frac{db^2\Phi^2}{2n} + O(1)\right] \\
&= \exp\left[\Phi \log d\left(-\frac{b^2}{2} + o(1) + o_d(1)\right)\right] && \text{using } b \in [\varepsilon, a] \\
&= \exp(-\Omega(n))
\end{aligned}$$

for sufficiently large d. The result follows by Markov's inequality. ∎

2.3. Stability of low-degree polynomials

The main result of this section (Proposition 2.6) states that the output of a low-degree polynomial is resilient to changes in the input, in a particular sense. Throughout this section we will use the shorthand $p := d/n$. We think of $Y \sim G(n, p)$ as simply $Y \in \{0,1\}^m$ with i.i.d. Bernoulli(p) coordinates, where $m = \binom{n}{2}$; the graph structure will not be used in this section. We consider the hypercube graph with vertex set $\{0,1\}^m$ and an edge (y, y') whenever y, y' differ on exactly one coordinate.

Definition 2.5. Let $f : \{0,1\}^m \to \mathbb{R}^n$ and let $c > 0$. An edge (y, y') of the hypercube $\{0,1\}^m$ is said to be *c-bad* for f if

$$\|f(y) - f(y')\|^2 \geq c\, \mathbb{E}_{Y \sim G(n,p)} \left[\|f(Y)\|^2\right].$$

Also, for $y \in \{0,1\}^m$, let $B_i^f(y)$ denote the event that the edge traversed by flipping the ith coordinate of y is c-bad for f.

The interpolation path (Definition 2.1) can be thought of as a random walk on the hypercube graph (which is allowed to either remain in place or traverse an edge at each step). The following main result of this section shows that with non-trivial probability, this walk encounters no bad edges. This result is similar to Theorem 4.2 of [22] (which corresponds to the case $L = 1$).

Proposition 2.6. *Let $L \in \mathbb{N}$ and $c > 0$. Consider the interpolation path $Y^{(0)}, \ldots, Y^{(T)}$ from Definition 2.1 of length $T = Lm$, with $p := d/n \leq 1/2$. Let $f : \{0,1\}^m \to \mathbb{R}^n$ be a degree-D polynomial. Then*

$$\mathbb{P}\left[\text{no edge of } Y^{(0)}, \ldots, Y^{(T)} \text{ is } c\text{-bad for } f\right] \geq p^{4LD/c}.$$

The proof will follow from the following two lemmas. The first is essentially an upper bound on the total number (weighted by the measure $G(n, p)$) of bad edges that a low-degree polynomial can have. This was proved in [22] based on standard facts about the *total influence* of low-degree polynomials.

Lemma 2.7 ([22, Lemma 4.3]). *If $p \leq 1/2$ and $f : \{0,1\}^m \to \mathbb{R}^n$ is a degree-D polynomial then*

$$\frac{cp}{2} \sum_{i=1}^{m} \mathbb{P}_{Y \sim G(n,p)} \left[B_i^f(Y)\right] \leq D, \tag{6}$$

where $B_i^f(y)$ is defined in Definition 2.5.

The next lemma gives an inequality that can be interpreted as follows. Roughly speaking, the right-hand side is large if there are many bad edges, and the left-hand side is large if the probability of having no bad edges on the interpolation path is

small. Therefore, the inequality tells us that if the total number of bad edges is small then it is likely for the interpolation path to have no bad edges.

Lemma 2.8. *Consider the interpolation path $Y^{(0)}, \dots, Y^{(T)}$ and the associated function $\sigma\colon [T] \to [m]$ from Definition 2.1. Let $q(y)$ denote the probability that no edge of the interpolation path is bad, conditioned on the starting point $Y^{(0)} = y$. Then*

$$-\mathbb{E}_{Y\sim G(n,p)} \log q(Y) \le H(p) \sum_{t=1}^{T} \mathbb{P}_{Y\sim G(n,p)} \left[B_{\sigma(t)}(Y)\right], \tag{7}$$

where H is the binary entropy function $H(p) = -p \log p - (1-p)\log(1-p)$.

Remark 2.9. The proof of Lemma 2.8 does not make use of the specific notion of c-bad from Definition 2.5. The result still holds if any arbitrary subset of the hypercube edges are designated "bad" (so long as $q(y)$ and $B_i^f(y)$ both use the same notion of "bad").

Remark 2.10. Lemma 2.8 holds not just for the specific choice of σ from Definition 2.1 but for any sequence $\sigma\colon [T] \to [m]$ of coordinates to resample. (In fact, this level of generality will be important for the inductive argument in the proof.)

Proof of Lemma 2.8. Proceed by induction on T. The base case $T = 0$ is immediate. For the case $T \ge 1$, define $\widetilde{q}(y)$ to be the probability that the sub-walk $Y^{(1)}, \dots, Y^{(T)}$ never traverses a bad edge, conditioned on the starting point $Y^{(1)} = y$. Write y_{-i} for the all-but-ith coordinates of y, and write $y_{-i}[b] \in \{0,1\}^m$ to denote the vector obtained from y_{-i} by setting coordinate i to the value $b \in \{0,1\}$. Note that the event $B_i^f(y)$ does not depend on y_i, so we can write

$$B_i^f(y_{-i}) := B_i^f(y).$$

Let $j = \sigma(1)$ be the coordinate resampled in the first step. For any fixed value of y_{-j}, we will consider

$$\varphi(y_{-j}) := -(1-p)\log q\big(y_{-j}[0]\big) - p \log q\big(y_{-j}[1]\big),$$

which can be thought of as the contribution from y_{-j} to the left-hand side of (7). If the event $B_j^f(y_{-j})$ holds then

$$q\big(y_{-j}[0]\big) = (1-p)\widetilde{q}\big(y_{-j}[0]\big) \quad \text{and} \quad q\big(y_{-j}[1]\big) = p\,\widetilde{q}\big(y_{-j}[1]\big),$$

and so

$$\begin{aligned}\varphi(y_{-j}) &= -(1-p)\log\big[(1-p)\widetilde{q}\big(y_{-j}[0]\big)\big] - p\log\big[p\,\widetilde{q}\big(y_{-j}[1]\big)\big] \\ &= H(p) - (1-p)\log\widetilde{q}\big(y_{-j}[0]\big) - p\log\widetilde{q}\big(y_{-j}[1]\big).\end{aligned}$$

On the other hand, if the complement event $\overline{B_j^f(y_{-j})}$ holds then

$$q\big(y_{-j}[0]\big) = q\big(y_{-j}[1]\big) = (1-p)\widetilde{q}\big(y_{-j}[0]\big) + p\,\widetilde{q}\big(y_{-j}[1]\big),$$

and so

$$\begin{aligned}\varphi(y_{-j}) &= -\log\big[(1-p)\widetilde{q}\big(y_{-j}[0]\big) + p\,\widetilde{q}\big(y_{-j}[1]\big)\big] \\ &\le -(1-p)\log\widetilde{q}\big(y_{-j}[0]\big) - p\log\widetilde{q}\big(y_{-j}[1]\big),\end{aligned}$$

where we have used convexity of $x \mapsto -\log x$. Therefore in general we have

$$\varphi(y_{-j}) \le H(p)\,\mathbb{1}_{B_j^f(y_{-j})} - (1-p)\log\widetilde{q}\big(y_{-j}[0]\big) - p\log\widetilde{q}\big(y_{-j}[1]\big).$$

Now, with $Y \sim G(n,p)$, we can write

$$\begin{aligned}-\mathbb{E}\log q(Y) &= \mathbb{E}\,\varphi(Y_{-j}) \\ &\le \mathbb{E}\left[H(p)\,\mathbb{1}_{B_j^f(Y_{-j})} - (1-p)\log\widetilde{q}\big(Y_{-j}[0]\big) - p\log\widetilde{q}\big(Y_{-j}[1]\big)\right] \\ &= H(p)\,\mathbb{P}\left[B_j^f(Y)\right] - \mathbb{E}\log\widetilde{q}(Y).\end{aligned}$$

By the inductive hypothesis,

$$-\mathbb{E}\log\widetilde{q}(Y) \le H(p)\sum_{t=2}^{T}\mathbb{P}\left[B_{\sigma(t)}(Y)\right],$$

so this completes the proof. ∎

Proof of Proposition 2.6. We will combine Lemmas 2.7 and 2.8. First note that since $p \le 1/2$ we have $-p\log p \ge -(1-p)\log(1-p)$, and so

$$H(p) \le -2p\log p. \tag{8}$$

Define $q(y)$ as in Lemma 2.8. The probability that no edge of the interpolation path is c-bad is $\mathbb{E}\,q(Y)$, where $Y \sim G(n,p)$. We have

$$\begin{aligned}-\log\mathbb{E}\,q(Y) &\le -\mathbb{E}\log q(Y) && \text{by Jensen's inequality} \\ &\le H(p)\sum_{t=1}^{T}\mathbb{P}[B_{\sigma(t)}(Y)] && \text{by Lemma 2.8} \\ &= H(p)\cdot L\sum_{i=1}^{m}\mathbb{P}[B_i^f(Y)] && \text{by Definition 2.1}\end{aligned}$$

$$\begin{aligned} &\leq H(p) \cdot L \cdot \frac{2D}{cp} && \text{by Lemma 2.7} \\ &\leq -4c^{-1} L D \log p && \text{by (8)}, \end{aligned}$$

which can be rearranged to yield the result. ∎

2.4. Putting it together

As in [22], we start by observing that a random polynomial can be converted to a deterministic polynomial that works almost as well.

Lemma 2.11. *Suppose f is a random degree-D polynomial that $(k, \delta, \gamma, \eta)$-optimizes the independent set problem in $G(n, d/n)$. Then for any $c > 2$ there exists a deterministic degree-D polynomial that $(k, c\delta, c\gamma, \eta)$-optimizes the independent set problem in $G(n, d/n)$.*

Proof. By definition, we have

$$\mathbb{E}_{Y,\omega}\left[\|f(Y,\omega)\|^2\right] \leq \gamma k \quad \text{and} \quad \mathbb{P}_{Y,\omega}\left[|V_f^\eta(Y,\omega)| < k\right] \leq \delta.$$

By Markov's inequality,

$$\begin{aligned} \mathbb{P}_\omega\left[\mathbb{E}_Y\left[\|f(Y,\omega)\|^2\right] \geq c\gamma k\right] &\leq \frac{1}{c} < \frac{1}{2}, \\ \mathbb{P}_\omega\left[\mathbb{P}_Y\left[|V_f^\eta(Y,\omega)| < k\right] \geq c\delta\right] &\leq \frac{1}{c} < \frac{1}{2}, \end{aligned}$$

and so there exists a seed $\omega^* \in \Omega$ for which the resulting deterministic polynomial $f(\cdot) = f(\cdot, \omega^*)$ satisfies

$$\mathbb{E}_Y\left[\|f(Y)\|^2\right] \leq c\gamma k \quad \text{and} \quad \mathbb{P}_Y\left[|V_f^\eta(Y)| < k\right] \leq c\delta,$$

as desired. ∎

We now prove our main impossibility result.

Proof of Theorem 1.3. For any given $\varepsilon > 0$, set $K = \lceil 1 + 5/\varepsilon^2 \rceil$, $T = (K-1)m$, and $\eta = \frac{\varepsilon \log d}{16d}$. The constant $d^* = d^*(\varepsilon) \geq 1$ will be chosen so that d is sufficiently large to apply Lemma 2.2, Proposition 2.3, and Lemma 2.4 in the sequel. Let $\Phi = \frac{\log d}{d} n$.

Assume on the contrary that the random polynomial that we wish to rule out, exists. By Lemma 2.11, there exists a deterministic degree-D polynomial f that satisfies

$$\mathbb{E}_Y\left[\|f(Y)\|^2\right] \leq 3\gamma(1+\varepsilon)\Phi \quad \text{and} \quad \mathbb{P}_Y\left[|V_f^\eta(Y)| < (1+\varepsilon)\Phi\right] \leq 3\delta. \tag{9}$$

Sample the interpolation path $Y^{(0)}, \dots, Y^{(T)}$ as in Definition 2.1, and let $U_t = V_f^{\eta}(Y^{(t)})$ be the resulting independent sets. Consider the following process to construct a sequence of sets $S_1, \dots, S_K \subseteq [n]$. Let $S_1 = U_0$. Then for $k = 2, 3, \dots, K$, let $S_k = U_{t_k}$, where $t_k \in [T]$ is the minimum t for which $|U_t \setminus (\cup_{\ell<k} S_\ell)| \geq \frac{\varepsilon}{4}\Phi$; if no such t exists then the process fails. We will show that with positive probability (over the interpolation path), the following events all occur simultaneously:

(i) $|U_t| \geq (1+\varepsilon)\Phi$ for all $0 \leq t \leq T$, and the process $S_1, \dots, S_K$ succeeds,

(ii) no edge on the interpolation path is c-bad for f, where $c = \frac{\varepsilon}{96\gamma(1+\varepsilon)}$,

(iii) the conclusion of Proposition 2.3 holds (i.e., no instance of the forbidden structure exists).

We will first show that events (i)–(iii) occur simultaneously with positive probability, and then we will show that this yields a contradiction. By Proposition 2.6, event (ii) occurs with probability at least $(d/n)^{4(K-1)D/c}$. By Proposition 2.3, event (iii) occurs with probability $1 - \exp(-\Omega(n))$. It remains to consider event (i).

For each fixed t we have that $Y^{(t)}$ is distributed as $G(n, d/n)$, so by combining Lemma 2.2 with the second property of f from (9), we have

$$(1+\varepsilon)\Phi \leq |U_t| \leq (2+\varepsilon)\Phi$$

with probability at least $1 - 3\delta - \exp(-\Omega(n))$; we will take a union bound over t. Now suppose that for some $0 \leq T' \leq T - m$, $Y^{(0)}, \dots, Y^{(T')}$ have been sampled so far, and $0 = t_1 < t_2 < \cdots < t_{K'}$ are the indices of the sets $S_k = U_{t_k}$ selected so far ($t_{K'} \leq T'$). Note that $Y^{(T'+m)}$ is independent from $\{Y^{(t)}\}_{t \leq T'}$ and so, provided

$$|S_k| \leq (2+\varepsilon)\Phi$$

for $1 \leq k \leq K'$, Lemma 2.4 (with $S = \cup_{k \leq K'} S_k$ and $a = (2+\varepsilon)K'$) implies

$$|U_{T'+m} \cap (\cup_{k \leq K'} S_k)| \leq \varepsilon\Phi$$

with probability $1 - \exp(-\Omega(n))$. Provided $|U_{T'+m}| \geq (1+\varepsilon)\Phi$, this implies

$$|U_{T'+m} \setminus (\cup_{k \leq K'} S_k)| \geq \Phi \geq \frac{\varepsilon}{4}\Phi,$$

and so $t_{K'+1} \leq T' + m$; thus by induction, $t_k \leq (k-1)m$ for all $k \in [K]$ and so the process $\{S_k\}$ succeeds by timestep $T = (K-1)m$. We therefore conclude that event (i) holds with probability at least $1 - 3\delta(T+1) - \exp(-\Omega(n))$.

Using $3\delta(T+1) = 3\delta[(K-1)m + 1] \leq 3\delta K m$, we now have that events (i)–(iii) occur simultaneously with positive probability, provided

$$(d/n)^{4(K-1)D/c} > 3\delta K m + \exp(-\Omega(n)). \tag{10}$$

For sufficiently large n, the term $\exp(-\Omega(n))$ is at most $\exp(-Cn)$ for some constant $C = C(\varepsilon, d) > 0$. Also recall $m = \binom{n}{2} < \frac{n^2}{2}$. Thus, to satisfy (10), it is sufficient to have

$$(d/n)^{4(K-1)D/c} \geq 3\delta K n^2 \quad \text{and} \quad (d/n)^{4(K-1)D/c} \geq 2\exp(-Cn). \tag{11}$$

For $d \geq 1$, the second condition in (11) is implied by

$$D \leq (Cn - \log 2)\frac{c}{4(K-1)\log n}.$$

For sufficiently large n, and using $c = \frac{\varepsilon}{96\gamma(1+\varepsilon)}$, this is implied by $D \leq \frac{C_1 n}{\gamma \log n}$, where $C_1 = C_1(\varepsilon, d) > 0$ is a constant. For $d \geq 1$, the first condition in (11) is implied by

$$\delta \leq \exp\left[-\frac{4(K-1)D}{c}\log n - 2\log n - \log(3K)\right].$$

Since $\gamma \geq 1$ and $D \geq 1$, for sufficiently large n this is implied by $\delta \leq \exp(-C_2 \gamma D \log n)$ for another constant $C_2 = C_2(\varepsilon, d) > 0$.

To complete the proof, it remains to show that if events (i)–(iii) occur simultaneously, this results in a contradiction. The idea is to use the stability property from (ii) to show that the sets $S_1, \dots, S_K$ from (i) are an instance of the forbidden structure that is disallowed by (iii).

We will first show $|U_t \triangle U_{t-1}| \leq \frac{\varepsilon}{4}\Phi$ for all $1 \leq t \leq T$, where $\triangle$ denotes the symmetric difference operation on sets. From (i) we know that the failure event in V_f^η (the second case of (2)) does not occur on any of the inputs $Y^{(t)}$ (because otherwise the output U_t of V_f^η would be $\emptyset$, violating (i)). By definition of symmetric difference, there are exactly $|U_t \triangle U_{t-1}|$ coordinates $i \in [n]$ such that the indicator $\mathbb{1}_{i \in U_t}$ differs from the indicator $\mathbb{1}_{i \in U_{t-1}}$; call these coordinates $J \subseteq [n]$. From the definition of V_f^η (Definition 1.1), each $i \in J$ falls into one of two cases (or both):

(a) among the values $f_i(Y^{(t)})$ and $f_i(Y^{(t-1)})$, one is ≥ 1 and the other is $\leq 1/2$; or

(b) i lies in the set $(A \setminus \widetilde{A}) \cup B$ (see Definition 1.1) for either $V_f^\eta(Y^{(t)})$ or $V_f^\eta(Y^{(t-1)})$.

Recall from above that the failure event in V_f^η (the second case of (2)) does not occur on any of the inputs $Y^{(t)}$, and so (using (2)) we have

$$|A \setminus \widetilde{A}| + |B| \leq \eta n$$

for each t. Also note that $|(A \setminus \widetilde{A}) \cup B| = |A \setminus \widetilde{A}| + |B|$ because A and B are disjoint by definition. This means there are at most $2\eta n$ coordinates $i \in J$ for which case (b) occurs. Therefore, the number of coordinates for which case (a) occurs must be at least

$$|J| - 2\eta n = |U_t \triangle U_{t-1}| - 2\eta n,$$

and so there are at least $|U_t \triangle U_{t-1}| - 2\eta n$ coordinates $i \in [n]$ for which

$$|f_i(Y^{(t)}) - f_i(Y^{(t-1)})| \geq 1/2.$$

This gives the bound

$$\|f(Y^{(t)}) - f(Y^{(t-1)})\|^2 \geq \frac{1}{4}\big(|U_t \triangle U_{t-1}| - 2\eta n\big). \tag{12}$$

Using event (ii) along with the definition of c-bad (Definition 2.5) and the first property of f from (9), we also have

$$\|f(Y^{(t)}) - f(Y^{(t-1)})\|^2 \leq c\,\mathbb{E}_{Y\sim G(n,d/n)}\left[\|f(Y)\|^2\right] \leq 3c\gamma(1+\varepsilon)\Phi. \tag{13}$$

Now (12) and (13) can be combined to give

$$|U_t \triangle U_{t-1}| \leq 12c\gamma(1+\varepsilon)\Phi + 2\eta n = \frac{\varepsilon}{4}\Phi$$

as desired, where we have used $c = \frac{\varepsilon}{96\gamma(1+\varepsilon)}$ and $\eta = \frac{\varepsilon \log d}{16d}$.

Recall that for $2 \leq k \leq K$, S_k is defined to be the *first* U_t for which

$$|U_t \setminus (\cup_{\ell<k} S_\ell)| \geq \frac{\varepsilon}{4}\Phi,$$

which means $|U_{t-1} \setminus (\cup_{\ell<k} S_\ell)| < \frac{\varepsilon}{4}\Phi$. Using the fact $|U_t \triangle U_{t-1}| \leq \frac{\varepsilon}{4}\Phi$ from above (in other words, the sets U_t and U_{t-1} only differ on at most $\frac{\varepsilon}{4}\Phi$ entries), this means

$$|S_k \setminus (\cup_{\ell<k} S_\ell)| \leq \frac{\varepsilon}{2}\Phi.$$

Combining this with event (i) and the fact that S_k is an independent set in $Y^{(t_k)}$, we have that $S_1, \ldots, S_K$ satisfies the properties of the forbidden structure from event (iii). This yields a contradiction and completes the proof. ∎

3. Proof of achievability

In this section we prove our main achievability result (Theorem 1.4) which shows that low-degree algorithms can find independent sets of size $(1-\varepsilon)\frac{\log d}{d}n$. We begin by defining some terminology pertaining to local algorithms on graphs. Throughout this section we will consider graphs $G = (V, E)$ with possibly-infinite vertex set V, but which are *locally finite*, i.e., each vertex has a finite number of neighbors. We will consider functions that take as input (G, v), where $G = (V, E)$ is a graph and $v \in V$ is a designated "root" vertex; let Λ denote the set of such (G, v) pairs. We will also

consider functions that take as input (G, v, X), where G and v are as before and $X: V \to [0, 1]$ is a labelling of the vertices; let $\widetilde{\Lambda}$ denote the set of such (G, v, X) pairs.

For a graph $G = (V, E)$ and a vertex $v \in V$, the *r-neighborhood* of v, denoted $N_r(G, v)$, is the rooted graph with root v that contains all vertices reachable from v by a path of length $\leq r$, along with all edges on such paths. We will use $|N_r(G, v)|$ to denote the number of edges in the r-neighborhood. Two rooted graphs are said to be isomorphic if there is a root-preserving graph isomorphism between them. A function g with domain Λ is said to be r-local if $g(G, v)$ depends only on the isomorphism class of $N_r(G, v)$. (Informally, g has access to the "shape" of the r-neighborhood but not the identity of the specific vertices.)

In the presence of vertex labels $X: V \to [0, 1]$, we generalize the above notions as follows. The *labeled r-neighborhood* of v in G, denoted $\widetilde{N}_r(G, v, X)$, is the r-neighborhood along with the vertex labels given by X (restricted to the r-neighborhood). Two rooted labeled graphs are said to be isomorphic if there is a root-preserving and label-preserving graph isomorphism between them. A function h with domain $\widetilde{\Lambda}$ is said to be r-local if $h(G, v, X)$ depends only on the isomorphism class of $\widetilde{N}_r(G, v, X)$.

The *Poisson Galton–Watson tree* with parameter $d > 0$, denoted $\mathrm{PGW}(d)$, is the distribution over rooted (possibly-infinite) trees (T, o) generated as follows:

- Start with a root vertex o at level 0.
- For $\ell = 0, 1, 2, \ldots$, each vertex at level ℓ independently spawns $\mathrm{Pois}(d)$ child vertices at level $\ell + 1$.
- Every vertex (except the root) is connected to its parent by an edge.

It is well known that the distribution of the r-neighborhood of any fixed vertex in $G(n, d/n)$ converges to the r-neighborhood of the root in $\mathrm{PGW}(d)$ as $n \to \infty$ with r held fixed (as discussed in e.g., [39]); see Lemma 3.3 below for one precise sense in which this convergence holds.

An *r-local algorithm* for the maximum independent set problem is an r-local function $h: \widetilde{\Lambda} \to \{0, 1\}$ with the property that $\{v \in V : h(G, v, X) = 1\}$ is an independent set for any graph $G = (V, E)$ with any vertex labels X. A line of prior work [24, 25, 33, 39] has considered the problem of choosing h to maximize the expected size of the independent set when $G \sim G(n, d/n)$ and X is i.i.d. $\mathrm{Unif}([0, 1])$. Due to the convergence of local neighborhoods to $\mathrm{PGW}(d)$, this task is equivalent (up to sub-leading terms in n) to maximizing the probability that $h(T, o, X) = 1$ when $(T, o) \sim \mathrm{PGW}(d)$ and X is again i.i.d. $\mathrm{Unif}([0, 1])$.

The following result of [39] shows that local algorithms can produce large independent sets in $\mathrm{PGW}(d)$. As discussed in Section 4 of [39], this implies that local algorithms can produce independent sets of expected size $(1 - \varepsilon)\frac{\log d}{d} n$ in $G(n, d/n)$.

Theorem 3.1 ([39, Theorem 4.1]). *For any $\varepsilon > 0$ and any sufficiently large d (depending on ε), there exists $r = r(\varepsilon, d)$ and an r-local function $h\colon \widetilde{\Lambda} \to \{0,1\}$ satisfying the following. If $(T, o) \sim \mathrm{PGW}(d)$ and vertex labels $\{X_v\}_{v \in V(T)}$ are drawn i.i.d. from the uniform distribution on $[0,1]$, then*

- *the vertex set $\{v \in V(T) : h(T, o, X) = 1\}$ is an independent set in T with probability 1, and*
- $\mathbb{E}\left[h(T, o, X)\right] \geq (1-\varepsilon)\frac{\log d}{d}$.

Remark 3.2. Our proof of Theorem 1.4 will show how to approximate the local algorithm from Theorem 3.1 by a low-degree algorithm. We will not use any specifics of the local algorithm, and so our proof actually shows how to approximate *any* local algorithm by a low-degree algorithm. More precisely: for any fixed $\varepsilon > 0$, $\eta > 0$, and $d \geq 1$, if we are given an r-local algorithm h for independent sets with

$$\mathbb{E}\left[h(T, o, X)\right] \geq \alpha,$$

then for any $n \geq n^*(\varepsilon, \eta, d, r, \alpha)$ we can produce a deterministic degree-D polynomial that $(k, \delta, \gamma, \eta)$-optimizes the independent set problem in $G(n, d/n)$ with parameters $k = (1-\varepsilon)\alpha n$ and $\delta = \exp(-Cn^{1/3})$, where $D > 0$, $\gamma \geq 1$, $C > 0$ are constants depending on $\varepsilon, \eta, d, r, \alpha$.

The next result, which is a special case of Lemma 12.4 of [10], quantifies the convergence of local neighborhoods of $G(n, d/n)$ to $\mathrm{PGW}(d)$.

Lemma 3.3 (see [10, Lemma 12.4]). *Let $G \sim G(n, d/n)$, and let $(T, o) \sim \mathrm{PGW}(d)$. Let $g\colon \Lambda \to [-1, 1]$ be an r-local function. For all sufficiently large n (depending on d, r) and for any $v \in [n]$,*

$$\left|\mathbb{E}\left[g(G, v)\right] - \mathbb{E}\left[g(T, o)\right]\right| \leq cn^{-1/4}\log n,$$

where $c > 0$ is a universal constant.

The next result is a special case of (the first statement in) Proposition 12.3 of [10].

Proposition 3.4 (see [10, Proposition 12.3]). *Let $G \sim G(n, d/n)$ with $d \geq 1$. Let $g\colon \Lambda \to [-1, 1]$ be an r-local function. For all $p \geq 2$,*

$$\mathbb{E}\left[\left|\sum_{v \in [n]} g(G, v) - \mathbb{E}\sum_{v \in [n]} g(G, v)\right|^p\right] \leq \left(c\sqrt{n}\, p^{3/2}(2d)^r\right)^p,$$

where $c > 0$ is a universal constant.

A simple consequence of the above moment inequality is a tail bound for local functions.

Corollary 3.5. *Let $G \sim G(n, d/n)$ with $d \geq 1$. Let $g\colon \Lambda \to [-1, 1]$ be an r-local function. For a universal constant $c > 0$ and for all $t \geq (2e)^{3/2} c\sqrt{n}(2d)^r$,*

$$\mathbb{P}\left[\Big|\sum_{v\in[n]} g(G, v) - \mathbb{E}\sum_{v\in[n]} g(G, v)\Big| \geq t\right] \leq \exp\left(-\frac{3t^{2/3}}{2ec^{2/3}n^{1/3}(2d)^{2r/3}}\right).$$

Proof. Let c be the constant from Proposition 3.4. Choosing

$$p = e^{-1}\left[c\sqrt{n}(2d)^r/t\right]^{-2/3} \geq 2,$$

we have

$$\begin{aligned}
&\mathbb{P}\left[\Big|\sum_{v\in[n]} g(G, v) - \mathbb{E}\sum_{v\in[n]} g(G, v)\Big| \geq t\right] \\
&\quad = \mathbb{P}\left[\Big|\sum_{v\in[n]} g(G, v) - \mathbb{E}\sum_{v\in[n]} g(G, v)\Big|^p \geq t^p\right] \\
&\quad \leq t^{-p}\,\mathbb{E}\left[\Big|\sum_{v\in[n]} g(G, v) - \mathbb{E}\sum_{v\in[n]} g(G, v)\Big|^p\right] \\
&\quad \leq t^{-p}\left(c\sqrt{n}\,p^{3/2}(2d)^r\right)^p \\
&\quad = \exp\left(-\frac{3}{2}p\right) = \exp\left(-\frac{3t^{2/3}}{2ec^{2/3}n^{1/3}(2d)^{2r/3}}\right),
\end{aligned}$$

as desired. ∎

We will also need the following standard multiplicative version of the Chernoff bound [36].

Proposition 3.6. *Suppose $Z_1, \dots, Z_n$ are independent random variables taking values in $\{0, 1\}$. Let $Z = \sum_i Z_i$ and $\mu = \mathbb{E}[Z]$. For any $0 \leq \delta \leq 1$,*

$$\mathbb{P}\left[Z \leq (1-\delta)\mu\right] \leq \exp\left(-\frac{\delta^2\mu}{2}\right).$$

Also, for any $\delta \geq 0$,

$$\mathbb{P}\left[Z \geq (1+\delta)\mu\right] \leq \exp\left(-\frac{\delta^2\mu}{2+\delta}\right),$$

and so for $\delta \geq 1$,

$$\mathbb{P}\left[Z \geq (1+\delta)\mu\right] \leq \exp\left(-\frac{\delta\mu}{3}\right).$$

Proof of Theorem 1.4. Given $\varepsilon > 0$, apply Theorem 3.1 (with $\varepsilon/5$ in place of ε) to obtain $d^*(\varepsilon) > 1$, $r = r(\varepsilon, d)$ and an r-local function $h: \widetilde{\Lambda} \to \{0, 1\}$ that outputs independent sets with

$$\mathbb{E}\left[h(T, o, X)\right] \geq (1 - \varepsilon/5)\frac{\log d}{d}$$

when $(T, o) \sim \mathrm{PGW}(d)$ and X is i.i.d. Unif($[0, 1]$).

By Lemma 2.11, it is sufficient to prove the result for a *random* polynomial instead of a deterministic one (up to a change in the constants γ, C). We will construct a random polynomial $f: \{0, 1\}^{\binom{n}{2}} \to \mathbb{R}^n$ as follows. The input Y to f encodes a graph on vertex set $[n]$. The internal randomness of f samples vertex labels $\{X_v\}_{v \in [n]}$ i.i.d. from Unif($[0, 1]$). We will construct f with the following property:

$$\begin{aligned} &\text{for any } v \in [n], \text{ if } N_r(Y, v) \text{ is a tree with } |N_r(Y, v)| \leq s, \\ &\text{then } f_v(Y, X) = h(Y, v, X), \end{aligned} \tag{14}$$

where $s = s(\varepsilon, d, \eta)$ is a constant to be chosen later.

Concretely, we construct f as follows. Let $\mathcal{G}_{v,r,s}$ be the collection of graphs G on vertex set $[n]$ for which $|E(G)| \leq s$ and every non-isolated vertex is reachable from v by a path of length $\leq r$. (In other words, $\mathcal{G}_{v,r,s}$ consists of all possible r-neighborhoods for v of size $\leq s$.) Let

$$f_v(Y, X) = \sum_{G \in \mathcal{G}_{v,r,s}} \alpha(G, v, X) \prod_{e \in E(G)} Y_e, \tag{15}$$

where the coefficients $\alpha(G, v, X)$ are chosen in the following manner so that (14) is satisfied. Suppose that $G \in \mathcal{G}_{v,r,s}$ is the true r-neighborhood of v. Then

$$f_v(Y, X) = \sum_{\substack{G' \in \mathcal{G}_{v,r,s} \\ E(G') \subseteq E(G)}} \alpha(G', v, X).$$

Since h is r-local, we have $h(Y, v, X) = h(G, v, X)$. Therefore, once the values $\alpha(G', v, X)$ have been determined for all G' such that $E(G') \subsetneq E(G)$, there is a unique choice for $\alpha(G', v, X)$ that will satisfy (14), namely

$$\alpha(G, v, X) = h(G, v, X) - \sum_{\substack{G' \in \mathcal{G}_{v,r,s} \\ E(G') \subsetneq E(G)}} \alpha(G', v, X). \tag{16}$$

We therefore define $\alpha(G, v, X)$ for all $G \in \mathcal{G}_{v,r,s}$ according to the recursive definition (16), and this ensures that (14) is satisfied.

Note f is a degree-s polynomial in Y, so we will take $D = s$. It remains to check that f $(k, \delta, \gamma, \eta)$-optimizes the independent set problem in $G(n, d/n)$ with the desired parameters k, δ, γ, η, by verifying the two conditions in Definition 1.2. This is carried out in Lemmas 3.10 and 3.9 below. ∎

Next we will need a concentration result for neighborhood counts in a random graph. For a constant r, let $\mathcal{T}^r$ be a set of rooted trees consisting of one representative from each isomorphism class of rooted trees of depth at most $2r$. Let $\mathcal{T}^r_s \subseteq \mathcal{T}^r$ contain only those trees with at most s edges. Let $Y \sim G(n, d/n)$, and for $T \in \mathcal{T}^r$, let n_T denote the number of occurrences of the neighborhood T in Y, i.e.,

$$n_T = |\{v \in [n] : N_{2r}(Y, v) \cong T\}|$$

where $\cong$ denotes isomorphism of rooted graphs. Also, for $T \in \mathcal{T}^r$, let p_T denote the probability that T occurs as the neighborhood of the root in $\mathrm{PGW}(d)$, i.e.,

$$p_T = \mathbb{P}_{(U,o)\sim \mathrm{PGW}(d)}\left[N_{2r}(U, o) \cong T\right].$$

Also, let ϕ_T denote the probability over X that

$$h(Y, v, X) = 1$$

conditioned on $N_{2r}(Y, v) \cong T$. (Note that the event $\{h(Y, v, X) = 1\}$ depends only on X and $N_r(Y, v)$ since h is r-local.)

Lemma 3.7. *For any $\tau > 0$ and any $T \in \mathcal{T}^r$, for sufficiently large n (depending on d, r, τ) we have*

$$\mathbb{P}\left[|n_T - p_T n| \geq \tau n\right] \leq \exp(-C n^{1/3})$$

for some $C = C(d, r, \tau) > 0$.

Proof. By applying Lemma 3.3 to the function $g(G, v) = \mathbb{1}_{N_{2r}(G,v)\cong T}$, we have

$$|\mathbb{E}[n_T] - p_T n| \leq c n^{3/4} \log n \tag{17}$$

for sufficiently large n (depending on d, r). By Corollary 3.5, for any

$$t \geq (2e)^{3/2} c \sqrt{n} (2d)^{2r},$$

we have

$$\mathbb{P}\left[|n_T - \mathbb{E}[n_T]| \geq t\right] \leq \exp\left(-\frac{3t^{2/3}}{2ec^{2/3} n^{1/3} (2d)^{2r/3}}\right). \tag{18}$$

Combining (17) and (18) gives the result. ∎

Next we will show that with high probability, the rounding procedure $V_f^\eta(Y, X)$ does not encounter the failure event (the second case of (2)).

Lemma 3.8. *For any $\eta > 0$, the following holds for sufficiently large s (depending on ε, d, r, η). Let f be defined as in (15) and let $A, \widetilde{A}, B$ be defined as in Definition 1.1. With probability $1 - \exp(-\Omega(n^{1/3}))$, we have*

$$|A \setminus \widetilde{A}| + |B| \le \eta n.$$

Here $\Omega(\cdot)$ hides a constant depending on $\varepsilon, d, r, \eta, s$.

Proof. Suppose some vertex v is such that $N_{2r}(Y, v)$ is a tree with $|N_{2r}(Y, v)| \le s$. Then for all $u \in N_1(Y, v)$ we have that $N_r(Y, u)$ is a tree with $|N_r(Y, u)| \le s$ and so by (14),

$$f_u(Y, X) = h(Y, u, X) \in \{0, 1\}.$$

Since h outputs independent sets, it follows that v is not in the "bad" set $(A \setminus \widetilde{A}) \cup B$ from the definition of V_f^η (Definition 1.1). We have now shown that $(A \setminus \widetilde{A}) \cup B$ is disjoint from the set

$$V_s := \bigcup_{T \in \mathcal{T}_s^r} \{v \in [n] : N_{2r}(Y, v) \cong T\}.$$

For each $T \in \mathcal{T}_s^r$, we have from Lemma 3.7 that $|n_T - p_T n| \le \eta n/(2|\mathcal{T}_s^r|)$ with probability $1 - \exp(-\Omega(n^{1/3}))$. Choose s large enough so that

$$\sum_{T \in \mathcal{T}_s^r} p_T \ge 1 - \eta/2.$$

Noting that $A \cap B = \emptyset$ by definition, we now have

$$\begin{aligned} |A \setminus \widetilde{A}| + |B| = |(A \setminus \widetilde{A}) \cup B| &\le n - |V_s| = n - \sum_{T \in \mathcal{T}_s^r} n_T \\ &\le n - \sum_{T \in \mathcal{T}_s^r} \left(p_T n - \frac{\eta n}{2|\mathcal{T}_s^r|} \right) = \left(1 - \sum_{T \in \mathcal{T}_s^r} p_T \right) n + \frac{\eta n}{2} \le \eta n, \end{aligned}$$

as desired. ∎

Next we will show that the independent set produced by rounding f is large with high probability.

Lemma 3.9. *The following holds for sufficiently large $s > 0$ (depending on ε, d, r, η). Let f be defined as in (15). The independent set $I := V_f^\eta(Y, X)$ has size*

$$|I| \ge (1 - \varepsilon) \frac{\log d}{d} n$$

with probability $1 - \exp(-\Omega(n^{1/3}))$ over both Y and X.

Proof. In light of Lemma 3.8 and the definition of V_f^η (Definition 1.1), it suffices to show $|\widetilde{A}| \geq (1-\varepsilon)\frac{\log d}{d}n$ with probability $1-\exp(-\Omega(n^{1/3}))$. From the guarantees on h,

$$\left(1-\frac{\varepsilon}{5}\right)\frac{\log d}{d} \leq \mathbb{E}_{(T,o)\sim \mathrm{PGW}(d)}\left[h(T,o,X)\right] = \sum_{T\in\mathcal{T}^r} p_T\phi_T.$$

Choose s large enough so that

$$\sum_{T\in\mathcal{T}_s^r} p_T \geq 1-\frac{\varepsilon}{5}\frac{\log d}{d}.$$

Since $\phi_T \in [0,1]$, this implies

$$\sum_{T\in\mathcal{T}_s^r} p_T\phi_T \geq \left(\sum_{T\in\mathcal{T}^r} p_T\phi_T\right) - \frac{\varepsilon}{5}\frac{\log d}{d} \geq \left(1-\frac{2\varepsilon}{5}\right)\frac{\log d}{d}.$$

Again using Lemma 3.7, with probability $1-\exp(-\Omega(n^{1/3}))$ over Y,

$$\begin{aligned}\sum_{T\in\mathcal{T}_s^r} n_T\phi_T &\geq \sum_{T\in\mathcal{T}_s^r}\left(p_T n - \frac{\varepsilon}{5|\mathcal{T}_s^r|}\frac{\log d}{d}n\right)\phi_T \\ &\geq \left(\sum_{T\in\mathcal{T}_s^r} p_T\phi_T\right)n - \frac{\varepsilon}{5}\frac{\log d}{d}n \geq \left(1-\frac{3\varepsilon}{5}\right)\frac{\log d}{d}n. \qquad (19)\end{aligned}$$

Now fix Y satisfying (19) and consider the randomness of X. Recall from the proof of Lemma 3.8 that $(A\setminus\widetilde{A})\cup B$ is disjoint from V_s. Thus, for v satisfying $N_{2r}(Y,v)\cong T$ for some $T\in\mathcal{T}_s^r$, we have $v\in\widetilde{A}$ iff $h(Y,v,X)=1$, which occurs with probability $\phi_v := \phi_T$ (over the randomness of X). We will partition the elements of V_s into "bins" $W_1,\dots,W_{s+1}$ such that for each bin W_i, the vertices in W_i have disjoint r-neighborhoods and so the random variables $\{\mathbb{1}_{v\in\widetilde{A}}\}_{v\in W_i}$ are independent (conditioned on Y). Each vertex $v\in V_s$ has at most $s+1$ vertices in its $2r$-neighborhood, and so there are at most s vertices $u\in V_s$ such that $u\neq v$ and $N_r(Y,v)\cap N_r(Y,u)\neq\emptyset$. Since there are $s+1$ bins, we can greedily assign vertices to bins in order to achieve the desired disjointness property. Now that the bins $\{W_i\}$ have been constructed, we have by the Chernoff bound (Proposition 3.6) that for each i,

$$\mathbb{P}_X\left[\sum_{v\in W_i}\mathbb{1}_{v\in\widetilde{A}} \leq \left(1-\frac{\varepsilon}{5}\right)\mu_i\right] \leq \exp\left(-\frac{1}{2}\left(\frac{\varepsilon}{5}\right)^2\mu_i\right), \qquad (20)$$

where

$$\mu_i = \mathbb{E}_X\sum_{v\in W_i}\mathbb{1}_{v\in\widetilde{A}} = \sum_{v\in W_i}\phi_v. \qquad (21)$$

Call a bin W_i "large" if $\mu_i \geq \frac{\varepsilon}{5(s+1)} \frac{\log d}{d} n$ and "small" otherwise. Using (20) and a union bound over i, we have with probability $1 - \exp(-\Omega(n))$ that every large bin W_i satisfies $\sum_{v \in W_i} \mathbb{1}_{v \in \widetilde{A}} \geq \left(1 - \frac{\varepsilon}{5}\right)\mu_i$. Provided this holds, we now have

$$\begin{aligned}
|\widetilde{A}| &\geq \sum_{v \in V_s} \mathbb{1}_{v \in \widetilde{A}} \geq \sum_{i : W_i \text{ large}} \sum_{v \in W_i} \mathbb{1}_{v \in \widetilde{A}} \geq \sum_{i : W_i \text{ large}} \left(1 - \frac{\varepsilon}{5}\right)\mu_i \\
&= \left(1 - \frac{\varepsilon}{5}\right)\left[\sum_i \mu_i - \sum_{i : W_i \text{ small}} \mu_i\right] \\
&\geq \left(1 - \frac{\varepsilon}{5}\right)\left[\left(\sum_i \mu_i\right) - \frac{\varepsilon}{5}\frac{\log d}{d} n\right] && \text{using the definition of "small"} \\
&= \left(1 - \frac{\varepsilon}{5}\right)\left[\left(\sum_{v \in V_s} \phi_v\right) - \frac{\varepsilon}{5}\frac{\log d}{d} n\right] && \text{using the definition of } \mu_i \text{ (21)} \\
&= \left(1 - \frac{\varepsilon}{5}\right)\left[\left(\sum_{T \in \mathcal{T}_s^r} n_T \phi_T\right) - \frac{\varepsilon}{5}\frac{\log d}{d} n\right] \\
&\geq \left(1 - \frac{\varepsilon}{5}\right)\left[\left(1 - \frac{3\varepsilon}{5}\right)\frac{\log d}{d} n - \frac{\varepsilon}{5}\frac{\log d}{d} n\right] && \text{using (19)} \\
&= \left(1 - \frac{\varepsilon}{5}\right)\left(1 - \frac{4\varepsilon}{5}\right)\frac{\log d}{d} n \geq (1 - \varepsilon)\frac{\log d}{d} n,
\end{aligned}$$

completing the proof. ∎

Finally, we need to check the normalization condition.

Lemma 3.10. *There exists a constant $\gamma = \gamma(\varepsilon, d, \eta) \geq 1$ such that*

$$\mathbb{E}_{Y,X}\left[\|f(Y, X)\|^2\right] \leq \gamma(1 - \varepsilon)\frac{\log d}{d} n.$$

Proof. By linearity of expectation, it is sufficient to show $\mathbb{E}_{Y,X}[f_v(Y, X)^2] = O(1)$ uniformly over v. Fix a vertex $v \in [n]$ and define the random variable $N = |N_r(Y, v)|$. Recall the expansion (15) for f_v. For each $G \in \mathcal{G}_{v,r,s}$, the corresponding term in the sum can be nonzero only if G is a subgraph of $N_r(Y, v)$. Thus, the number of nonzero terms is at most

$$\binom{N}{\leq s} = \sum_{i=0}^{s} \binom{N}{i} \leq \sum_{i=0}^{s} N^i \leq (N + 1)^s.$$

Furthermore, we can see from (16) that the coefficient of each term is bounded by a constant, uniformly over v and X, that is, $|\alpha(G, v, X)| \leq a$ for some $a = a(r, s)$. This means

$$f_v(Y, X)^2 \leq \left[a(N + 1)^s\right]^2 = a^2(N + 1)^{2s}. \tag{22}$$

In order to bound the expectation of this quantity, we will need a tail bound for N. Starting from $m_0 = 1$, let m_i be the number of vertices whose distance in Y from v is exactly i. Conditioned on m_i, we have that m_{i+1} is stochastically dominated by Binomial$(m_i n, d/n)$. Using the Chernoff bound (Proposition 3.6), for fixed $m_i \geq 1$ and any $\delta \geq 1$,

$$\mathbb{P}\left[m_{i+1} \geq (1+\delta) d m_i\right] \leq \exp\left(-\frac{\delta d m_i}{3}\right) \leq \exp\left(-\frac{\delta d}{3}\right).$$

Therefore, with probability at least $1 - r\exp(-\delta d/3)$, we have $m_i < [(1+\delta)d]^i$ for all $0 \leq i \leq r$ and so

$$N < \sum_{i=0}^{r} [(1+\delta)d]^i \leq \left[(1+\delta)d + 1\right]^r.$$

For $\delta \geq 1$ and $d \geq 1$ we have $(1+\delta)d + 1 \leq 2\delta d + 1 \leq 3\delta d$, so we can rewrite the above as

$$\mathbb{P}\left[N \geq (3\delta d)^r\right] \leq r\exp(-\delta d/3).$$

Letting $t = (3\delta d)^r$, we now have a tail bound for N: for all $t \geq (3d)^r$,

$$\mathbb{P}[N \geq t] \leq r\exp(-t^{1/r}/9).$$

Finally, combining this with (22), we have

$$\begin{aligned}\mathbb{E}_{Y,X}\left[f_v(Y,X)^2\right] &\leq \sum_{t=0}^{\infty} a^2(t+1)^{2s}\,\mathbb{P}[N=t] \\ &\leq \sum_{t=0}^{\lceil (3d)^r \rceil} a^2(t+1)^{2s} + \sum_{t=\lceil (3d)^r \rceil}^{\infty} a^2(t+1)^{2s} r\exp(-t^{1/r}/9),\end{aligned}$$

which is finite and independent of n. This completes the proof. ∎

Acknowledgments. The author is grateful to Charles Bordenave for helpful discussions regarding concentration of local functions on graphs (Proposition 3.4 and Corollary 3.5).

Funding. Partially supported by NSF grant DMS-1712730 and by the Simons Collaboration on Algorithms and Geometry. Part of this work was done while the author was visiting the Simons Institute for the Theory of Computing.

References

[1] D. Achlioptas and A. Coja-Oghlan, Algorithmic barriers from phase transitions. In *2008 49th Annual IEEE Symposium on Foundations of Computer Science (Philadelphia, PA, USA, 2008)*, pp. 793–802, IEEE, 2008

[2] L. Addario-Berry and P. Maillard, The algorithmic hardness threshold for continuous random energy models. *Math. Stat. Learn.* **2** (2019), no. 1, 77–101 Zbl 1434.68182 MR 4073148

[3] A. S. Bandeira, J. Banks, D. Kunisky, C. Moore, and A. S. Wein, Spectral planting and the hardness of refuting cuts, colorability, and communities in random graphs. 2020, arXiv:2008.12237

[4] A. S. Bandeira, D. Kunisky, and A. S. Wein. Computational hardness of certifying bounds on constrained PCA problems. In *11th Innovations in Theoretical Computer Science Conference (Seattle, WA, USA, 2020)*, Dagstuhl Publishing, Saarbrücken-Wadern, Germany, 2020

[5] B. Barak, S. Hopkins, J. Kelner, P. K. Kothari, A. Moitra, and A. Potechin, A nearly tight sum-of-squares lower bound for the planted clique problem. *SIAM J. Comput.* **48** (2019), no. 2, 687–735 Zbl 1421.68056 MR 3945259

[6] M. Bayati, D. Gamarnik, and P. Tetali, Combinatorial approach to the interpolation method and scaling limits in sparse random graphs. In *STOC'10—Proceedings of the 2010 ACM International Symposium on Theory of Computing*, pp. 105–114, ACM, New York, 2010 Zbl 1293.05350 MR 2743259

[7] M. Bayati and A. Montanari, The dynamics of message passing on dense graphs, with applications to compressed sensing. *IEEE Trans. Inform. Theory* **57** (2011), no. 2, 764–785 Zbl 1366.94079 MR 2810285

[8] G. Ben Arous, R. Gheissari, and A. Jagannath, Algorithmic thresholds for tensor PCA. *Ann. Probab.* **48** (2020), no. 4, 2052–2087 Zbl 1444.62080 MR 4124533

[9] C. Bordenave and P. Caputo, Large deviations of empirical neighborhood distribution in sparse random graphs. *Probab. Theory Related Fields* **163** (2015), no. 1-2, 149–222 Zbl 1327.60067 MR 3405616

[10] C. Bordenave, S. Coste, and R. R. Nadakuditi, Detection thresholds in very sparse matrix completion. 2020, arXiv:2005.06062

[11] M. Brennan and G. Bresler. Reducibility and statistical-computational gaps from secret leakage. 2020, arXiv:2005.08099

[12] M. Brennan, G. Bresler, S. B. Hopkins, J. Li, and T. Schramm, Statistical query algorithms and low-degree tests are almost equivalent. 2020, arXiv:2009.06107

[13] W.-K. Chen, D. Gamarnik, D. Panchenko, and M. Rahman, Suboptimality of local algorithms for a class of max-cut problems. *Ann. Probab.* **47** (2019), no. 3, 1587–1618 Zbl 1466.60200 MR 3945754

[14] Y. Cherapanamjeri, S. B. Hopkins, T. Kathuria, P. Raghavendra, and N. Tripuraneni, Algorithms for heavy-tailed statistics: regression, covariance estimation, and beyond. In *STOC '20—Proceedings of the 52nd Annual ACM SIGACT Symposium on Theory of Computing*, pp. 601–609, ACM, New York, 2020 Zbl 07298273 MR 4141785

[15] A. Coja-Oghlan and C. Efthymiou, On independent sets in random graphs. *Random Structures Algorithms* **47** (2015), no. 3, 436–486 Zbl 1325.05147 MR 3385742

[16] A. Coja-Oghlan, A. Haqshenas, and S. Hetterich, `Walksat` stalls well below satisfiability. *SIAM J. Discrete Math.* **31** (2017), no. 2, 1160–1173 Zbl 1371.68109 MR 3656499

[17] Y. Ding, D. Kunisky, A. S. Wein, and A. S. Bandeira, Subexponential-time algorithms for sparse PCA. 2019, arXiv:1907.11635

[18] A. El Alaoui, A. Montanari, and M. Sellke, Optimization of mean-field spin glasses. *Ann. Probab.* **49** (2021), no. 6, 2922–2960 MR 4348682

[19] A. El Alaoui and M. Sellke, Algorithmic pure states for the negative spherical perceptron. 2020, arXiv:2010.15811

[20] A. M. Frieze, On the independence number of random graphs. *Discrete Math.* **81** (1990), no. 2, 171–175 Zbl 0712.05052 MR 1054975

[21] D. Gamarnik and A. Jagannath, The overlap gap property and approximate message passing algorithms for p-spin models. *Ann. Probab.* **49** (2021), no. 1, 180–205 Zbl 1470.60277 MR 4203336

[22] D. Gamarnik, A. Jagannath, and A. S. Wein, Low-degree hardness of random optimization problems. In *2020 IEEE 61st Annual Symposium on Foundations of Computer Science*, pp. 131–140, IEEE Computer Soc., Los Alamitos, CA, 2020 MR 4232029

[23] D. Gamarnik and M. Sudan, Performance of the survey propagation-guided decimation algorithm for the random NAE-K-SAT problem. 2014, arXiv:1402.0052

[24] D. Gamarnik and M. Sudan, Limits of local algorithms over sparse random graphs. *Ann. Probab.* **45** (2017), no. 4, 2353–2376 Zbl 1371.05265 MR 3693964

[25] H. Hatami, L. Lovász, and B. Szegedy, Limits of locally-globally convergent graph sequences. *Geom. Funct. Anal.* **24** (2014), no. 1, 269–296 Zbl 1294.05109 MR 3177383

[26] S. Hopkins, Statistical inference and the sum of squares method, PhD thesis, Cornell University, 2018 MR 3864930

[27] S. B. Hopkins, P. K. Kothari, A. Potechin, P. Raghavendra, T. Schramm, and D. Steurer, The power of sum-of-squares for detecting hidden structures. In *58th Annual IEEE Symposium on Foundations of Computer Science—FOCS 2017*, pp. 720–731, IEEE Computer Soc., Los Alamitos, CA, 2017 MR 3734275

[28] S. B. Hopkins, J. Shi, and D. Steurer, Tensor principal component analysis via sum-of-squares proofs. In *Proceedings of The 28th Conference on Learning Theory (Paris, France, 2015)*, pp. 956–1006, Proceedings of Machine Learning Research 40, PMLR, 2015.

[29] S. B. Hopkins and D. Steurer, Efficient Bayesian estimation from few samples: community detection and related problems. In *58th Annual IEEE Symposium on Foundations of Computer Science—FOCS 2017*, pp. 379–390, IEEE Computer Soc., Los Alamitos, CA, 2017 MR 3734245

[30] A. Javanmard and A. Montanari, State evolution for general approximate message passing algorithms, with applications to spatial coupling. *Inf. Inference* **2** (2013), no. 2, 115–144 Zbl 1335.94015 MR 3311445

[31] R. M. Karp, The probabilistic analysis of some combinatorial search algorithms. In *Algorithms and complexity (Pittsburgh, PA, USA, 1976)*, pp. 1–19, Academic Press, New York, 1976 Zbl 0368.68035 MR 0445898

[32] D. Kunisky, A. S. Wein, and A. S. Bandeira, Notes on computational hardness of hypothesis testing: Predictions using the low-degree likelihood ratio. 2019, arXiv:1907.11636

[33] J. Lauer and N. Wormald, Large independent sets in regular graphs of large girth. *J. Combin. Theory Ser. B* **97** (2007), no. 6, 999–1009 Zbl 1183.05058 MR 2354714

[34] Y. Luo and A. R. Zhang, Tensor clustering with planted structures: Statistical optimality and computational limits. 2020, arXiv:2005.10743

[35] A. Maleki, *Approximate Message Passing Algorithms for Compressed Sensing*. ProQuest LLC, Ann Arbor, MI, 2010 MR 4158199 D. L. Donoho, A. Maleki, and A. Montanari, Message-passing algorithms for compressed sensing, *Proc. Natl. Acad. Sci. U.S.A.* **106** (2009), no. 45, 18914–18919

[36] M. Mitzenmacher and E. Upfal, *Probability and computing*. Cambridge Univ. Press, Cambridge, 2005 Zbl 1092.60001 MR 2144605

[37] A. Montanari, Optimization of the Sherrington–Kirkpatrick Hamiltonian. In *2019 IEEE 60th Annual Symposium on Foundations of Computer Science*, pp. 1417–1433, IEEE Comput. Soc. Press, Los Alamitos, CA, 2019 MR 4228234

[38] A. Montanari and E. Richard, Non-negative principal component analysis: message passing algorithms and sharp asymptotics. *IEEE Trans. Inform. Theory* **62** (2016), no. 3, 1458–1484 Zbl 1359.62224 MR 3472260

[39] M. Rahman and B. Virág, Local algorithms for independent sets are half-optimal. *Ann. Probab.* **45** (2017), no. 3, 1543–1577 Zbl 1377.60049 MR 3650409

[40] E. Richard and A. Montanari, A statistical model for tensor PCA. In *Advances in Neural Information Processing Systems (Montreal, Canada, 2014)*, pp. 2897–2905, Curran Associates, Inc., 2014.

[41] T. Schramm and A. S. Wein, Computational barriers to estimation from low-degree polynomials. 2020, arXiv:2008.02269

[42] E. Subag, Following the ground states of full-RSB spherical spin glasses. *Comm. Pure Appl. Math.* **74** (2021), no. 5, 1021–1044 Zbl 1467.82097 MR 4230065

Received 12 November 2020; revised 11 September 2021.

Alexander S. Wein
Department of Mathematics, Courant Institute of Mathematical Sciences,
New York University, New York, NY 10012-1110, USA; awein@cims.nyu.edu