Algorithmic Phase Transition for Large Independent Sets in Dense Hypergraphs

Abhishek Dhawan
Department of Mathematics, University of Illinois Urbana-Champaign, IL, USA

Nhi U. Dinh
Department of Mathematics, University of Illinois Urbana-Champaign, IL, USA

Eren C. Kızıldağ
Department of Statistics, University of Illinois Urbana-Champaign, IL, USA

Neeladri Maitra
Department of Mathematics, University of Illinois Urbana-Champaign, IL, USA

Bayram A. Şahin
Department of Statistics, University of Illinois Urbana-Champaign, IL, USA

Abstract

We study the algorithmic tractability of finding large independent sets in dense random hypergraphs. In the sparse regime, much of the natural algorithms can be formulated within either the local or the low-degree polynomial (LDP) framework, and a rich literature has subsequently identified nearly sharp algorithmic thresholds within these classes by exploiting their stability. In the dense setting, however, the algorithmic paradigms are fundamentally different: they are online and thus need not be stable. Perhaps more crucially, even for the classical Erdős–Rényi random graph $G(n, p)$, LDPs are conjectured to fail in the "easy" regime accessible to online algorithms, thereby challenging their viability for dense models.

Our focus is on two models: (i) finding large independent sets in dense r-uniform Erdős–Rényi hypergraphs, where each size-r hyperedge is present independently with probability p, and (ii) the more challenging problem of finding large γ-balanced independent sets in dense r-uniform r-partite hypergraphs, where the vertex set is the disjoint union $V_1 \sqcup \cdots \sqcup V_r$ with $|V_i| = n$ for all i, each hyperedge in $V_1 \times \cdots \times V_r$ is present independently with probability p, and the i-th coordinate of $\gamma \in \mathbb{Q}^r$ specifies the proportion of vertices from V_i in the independent set. For both models, we pinpoint the size of the largest independent set and design online algorithms that achieve a multiplicative approximation factor of $r^{1/(r-1)}$ in the uniform and $(\max_i \gamma_i)^{-1/(r-1)}$ in the r-partite model. Furthermore, we establish matching algorithmic lower bounds, showing that these computational gaps are *sharp*: no online algorithms can breach these gaps.

Our results provide a detailed landscape for dense hypergraphs, thereby completing the picture for dense models in a manner parallel to the sparse counterparts developed recently. Our main technical contribution is twofold: a novel *staged and bucketed greedy* algorithm and a stopping-time argument tailored to the hypergraph and multipartite structure for algorithmic hardness, both of which may be of independent interest. The algorithms and proof techniques in the dense regime differ substantially from those in the sparse, yet the resulting computational gaps are remarkably analogous, pointing to a form of universality. More conceptually, our results corroborate a hypothesis from statistical mechanics linking glassy equilibrium to computational hardness: optimization problems become far more intricate in the presence of global constraints.

2012 ACM Subject Classification Theory of computation → Randomness, geometry and discrete structures; Theory of computation → Online algorithms; Mathematics of computing → Combinatorial optimization; Mathematics of computing → Hypergraphs; Mathematics of computing → Random graphs

Keywords and phrases independent sets, random hypergraphs, online algorithms, overlap gap property, statistical–computational gap, Erdős–Rényi hypergraph

Digital Object Identifier 10.4230/LIPIcs.APPROX/RANDOM.2026.68

Category RANDOM

Related Version *Full Version*: https://arxiv.org/abs/2605.05618 [17]

Funding *Abhishek Dhawan*: Partially supported by the NSF RTG grant DMS-1937241.

1 Introduction

In this paper, we study the statistical and algorithmic landscape of large independent sets in two hypergraph models: (i) the r-uniform Erdős–Rényi hypergraph $H_r(n,p)$, where each size-r subset of $[n] \coloneqq \{1,\dots,n\}$ is present independently with probability p, and (ii) the r-uniform r-partite hypergraph $H(r,n,p)$, whose vertex set is partitioned into r parts $V_1,\dots,V_r$, each of size n, where each edge in $V_1 \times \cdots \times V_r$ is present independently with probability p. Our focus is on the *dense* regime in which p is a constant (as $n \to \infty$). While the landscape for sparse random hypergraphs has been characterized in detail recently [21], the dense regime has remained elusive and requires different techniques. We examine large independent sets in $H_r(n,p)$ and large *balanced* independent sets in $H(r,n,p)$, where the selected vertices are required to be distributed across the r parts according to a prescribed vector $\gamma \in \mathbb{Q}_+^r$. For both models, we pinpoint the largest possible size – referred to as the *statistical threshold*, design polynomial-time algorithms that provably outputs a large independent set, and establish sharp algorithmic lower bounds. Taken together, our results reveal *statistical-computational gaps* analogous to those in the sparse regime, thereby pointing to a form of universality.

Establishing sharp results for dense hypergraphs requires new technical ingredients. To this end, we introduce a novel *staged and bucketed greedy* algorithm tailored to hypergraphs and design a geometrical barrier, equipped with a carefully chosen stopping time, to establish tight algorithmic lower bounds. Informally, the *staged and bucketed* greedy procedure is divided into r stages where in each stage, the algorithm populates the independent set from an "unexhausted" part until a certain target capacity is reached, and then locks that part. As for our algorithmic barrier, it is adapted to the multipartite structure and global balancedness constraint; specifically, it tracks the first time at which the independent set contains sufficiently many vertices from every part of the partition. See Section 1.3 for a more detailed overview. We expect these new tools to be of independent interest for understanding algorithmic tractability in hypergraph models and beyond.

Our algorithmic viewpoint is *online*: vertices are exposed sequentially, and upon the arrival of a vertex v, the algorithm learns the status of hyperedges involving v together with previously exposed vertices, and must irrevocably decide whether to include v in its output. As we explain below, online algorithms are central in the dense setting, where they capture the only known class of polynomial-time algorithms. This stands in stark contrast to the sparse setting, where the predominant algorithmic frameworks are markedly different, see Section 1.1.

We emphasize that, while approximating the largest independent set to within a factor of $n^{1-\epsilon}$ is already NP-hard in the worst-case [38, 47], the average-case picture is substantially richer. For the Erdős–Rényi random graph $G(n,\frac{1}{2})$, the independence number – size of the largest independent set – is approximately $2\log_2 n$, whereas the online greedy algorithm reaches $\log_2 n$, both with high probability (whp) as $n \to \infty$ [11, 37, 54, 55]. In his seminal 1976 work [46], Karp challenged the community to design an efficient algorithm that reaches $(1+\epsilon)\log_2 n$ (whp) for some $\epsilon > 0$, or to prove that this is impossible (modulo a complexity-theoretic assumption such as $P \neq NP$). Nearly five decades later, this question remains open and is widely regarded as a cornerstone of average-case complexity and random graph theory. It has inspired several major developments – such as Jerrum's introduction of the *planted clique* problem [44], and a substantial literature has since uncovered analogous "factor-2 gaps" across a broad range of random graphs models.[1] For an overview of such gaps in random graphs and beyond, see the surveys [29, 30, 34, 7, 63, 27].

1.1 Comparison between Online Algorithms and Low-Degree Polynomials

There is by now a substantial literature on algorithmic phase transitions for the largest independent set problem in sparse random graphs (including the classical Erdős–Rényi model as well as bipartite and multipartite analogues), where nearly sharp lower bounds have been established for local algorithms and low-degree polynomials (LDPs) [35, 60, 62, 59, 21, 31]. In the sparse regime, this framework is particularly natural: local neighborhoods remain small and approximately tree-like, and several natural algorithms can be formulated within either the local or low-degree framework [62, 59, 21, 31].

▶ Remark 1. We note that in the low-degree setting, the local tree-like property is essential to control the variance of certain quantities. We discuss this further in relation to our work below.

The dense regime, however, is fundamentally different. Even for the classical Erdős–Rényi graph $G(n, p)$ with constant p, it is likely that the online greedy algorithm cannot be implemented as an LDP. An informal justification for this is through a certain *invariant*: while LDPs enjoy input stability [62, 31, 59], the online greedy algorithm is shown to be unstable [32], thus unlikely to be implemented as an LDP. The current state of affairs is in fact more subtle. A remarkable conjecture put forth in a recent AIM workshop [1] suggests that LDPs fail even in the "easy" regime:

▶ **Conjecture 2.** *In $G(n, \frac{1}{2})$, no degree-$o(\log^2 n)$ polynomial can find an independent set of size $0.9 \log_2 n$.*

Accordingly, while LDPs provide an important proxy for efficient algorithms in sparse random graphs, their viability in the dense setting, even for the classical Erdős–Rényi random graph $G(n, p)$ with constant p, is questionable.[2] If true, Conjecture 2 would show that LDPs are actually not viable for dense random graphs. A relatively weaker variant of this conjecture was proven in the aforementioned workshop [1]. This motivates the study of online algorithms in the dense regime.

▶ Remark 3. The online greedy algorithm is in fact relevant well beyond the dense setting. Indeed, in the intermediate regime where $p = o(1)$ and $np \to \infty$, the independence number of $G(n, p)$ is $(2 + o(1)) \log(np)/p$, whereas the greedy algorithm finds an independent set of size $(1 + o(1)) \log(np)/p$, both whp [10, 28, 25]. The same phenomenon persists in the sparse regime: in the double limit $n \to \infty$ followed by $d \to \infty$, the independence number of $G(n, d/n)$ is $(2 + o_d(1))n \log d/d$, while greedy reaches $(1 + o_d(1))n \log d/d$ [13]. Thus, throughout dense, intermediate, and sparse regimes, the online greedy yields 1/2-approximation to the optimum (at the level of the leading-order terms).

[1] Importantly, Karp's original conjecture regarding "factor-2 gaps" was formulated for the dense regime, while much of the recent effort has been focused on the sparse setting.
[2] Degree-$o(\log^2 n)$ polynomials are used as a proxy for polynomial-time algorithms; see [40, 51, 63].

Lower bounds for computational intractability of finding large independent sets in $G(n,p)$ within the framework of online algorithms had remained elusive, emerging only very recently [32].[3] At a technical level, they required substantial refinements of the machinery based on the Overlap Gap Property (see below for details). These results were recently extended to the balanced independent set problem in dense bipartite random graphs [19]. As we detail below, the largest balanced independent set problem in bipartite models closely resemble their classical Erdős–Rényi counterparts (even though the unconstrained independent set problem in bipartite random graphs is polynomial-time solvable via a max-flow reduction). As such, a variant of Conjecture 2 plausibly holds for bipartite models – this was in fact put forth in [19, Conjecture 1.6].

Bipartite Random Graphs. For the large balanced independent set problem in sparse random bipartite graphs, Perkins and Wang [59] obtained sharp algorithmic results within both the local and LDP frameworks. Their algorithm arises from bipartiteness, but the sparsity is crucial in placing it into the local or LDP framework – their analysis breaks down in the absence of sparsity. They explicitly note that, since sparse random graphs are locally tree-like, the performance of local algorithms is determined, to first order, by evaluation on a Galton-Watson tree. The same is true also in the bipartite case. More specifically, the analysis of the local algorithm uses the fact that for sparse random graphs, the neighborhoods are asymptotically tree-like and hence can be approximated by a Galton-Watson tree, which they leverage to control certain acceptance probabilities and to bound neighborhood sizes; see [59, Section 2.1]. As for the low-degree analysis, one must control the variance of a certain LDP, for which it is essential to control the size of certain neighborhoods. In [59, Section 2.2], this is handled precisely through a Poisson approximation, which unfortunately holds only in the sparse regime and is no longer valid for the dense setting.

For these reasons, results of [59] for sparse bipartite graphs do not extend to the dense setting. In a dense bipartite graph, even a depth-1 neighborhood already contains $\Theta(n)$ vertices (along with substantial dependencies), so the local limit argument based on Galton-Watson approximation disappears. Consequently, if one tries to adapt the algorithmic framework of [59], the resulting polynomial will unfortunately have a large variance, breaking down the analysis. As for the balanced independent set problem in sparse random multipartite hypergraphs, recent work of Dhawan and Wang [21] devised LDP algorithms, for which the analysis does not transfer to dense regime either.

For these models, the main source of hardness is not multipartiteness per se, but rather the *balancedness* constraint. More broadly, once one imposes a global constraint such as balancedness, the resulting optimization problem becomes substantially more delicate.[4] In particular, the largest balanced independent set problem in dense bipartite or multipartite random graphs is much closer in flavor to the classical largest independent set problem in dense $G(n,p)$ than to the unconstrained independent set problem on bipartite graphs. In light of this, online algorithms offer a natural algorithmic framework in the dense regime, across both the Erdős–Rényi and multipartite settings, whereas the viability of the low-degree framework (or local algorithms) is presently much less clear.[5]

[3] We note, however, that lower bounds for LDPs appeared earlier in [43, 62].

[4] In the language of statistical mechanics, a global constraint such as balancedness is expected to induce "glassy equilibrium" and lead to computational hardness [58, 56, 59, 34]; this has been rigorously verified for certain models, see, e.g., [22, 23].

[5] For related reasons, online algorithms also play an important role in other optimization problems such as coloring over dense random graphs; see [49].

Extension to Hypergraphs and Multipartite Models. The higher uniformity ($r \geq 3$), the multipartite structure, and the balancedness constraint collectively make the rigorous study of random hypergraph models substantially more delicate. To address these challenges, we develop new technical ideas both on the algorithmic side and for establishing sharp algorithmic hardness. These ideas contribute to the technical toolkit for dense random structures, which has only recently begun to emerge; we expect them to be broadly useful for hypergraph models and beyond.

▶ Remark 4. While it is well-established that the independence number of various pseudorandom graph classes attains the computational threshold – a line of inquiry originating with Shearer's foundational work on K_3-free graphs from 1983 [61] and expanding into a robust body of literature [18, 6, 48, 15] – an analogous result for pseudorandom hypergraphs was entirely absent from the literature until the recent work of Dhawan, Methuku, and Vo [20]. There are, however, several earlier results that match the growth rate [4, 26, 16].

These results suggest that finding large independent sets in hypergraphs is inherently harder than the graph case, further motivating the study of hypergraphs.

For a detailed technical overview, see Section 1.3. A summary of our contributions is in order.

1.2 Summary of Main Results

We now provide informal versions of our main results. Recall the descriptions of our models: the first is the dense r-uniform Erdős–Rényi hypergraph $H_r(n,p)$, where each r-element subset of $[n] := \{1,\ldots,n\}$ is present independently with probability p; the second is the dense r-uniform r-partite hypergraph $H(r,n,p)$, where the vertex set is partitioned into r parts $V_1,\ldots,V_r$ of size n and each r-tuple from $V_1 \times \cdots \times V_r$ is included independently with probability p.

For $b := \frac{1}{1-p}$ and $\gamma := (\gamma_1,\ldots,\gamma_r) \in \mathbb{Q}_+^r$ with $\sum_i \gamma_i = 1$, define

$$\alpha_{\mathtt{r}}^{\mathrm{STAT}} := (r!\log_b n)^{\frac{1}{r-1}} \qquad \text{and} \qquad \alpha_{\mathtt{r\text{-}bal}}^{\mathrm{STAT}} := \left(\frac{\log_b n}{\prod_{i=1}^r \gamma_i}\right)^{\frac{1}{r-1}}.$$

Throughout this paper, we focus on constant p, $p = \Theta(1)$ (as $n \to \infty$). We ignore all floor/ceiling operators for simplicity with the understanding that this has no effect on our arguments.

Our first result pins down the size of the largest (γ-balanced) independent set.

▶ **Theorem 5.** *For $H_r(n,p)$, the largest independent set has size $(1 \pm o(1))\,\alpha_r^{\mathrm{STAT}}$ whp. Likewise, for $H(r,n,p)$, the largest γ-balanced independent set has size $(1 \pm o(1))\,\alpha_{r\text{-}bal}^{\mathrm{STAT}}$ whp.*

Theorem 5 identifies $\alpha_{\mathtt{r}}^{\mathrm{STAT}}$ and $\alpha_{\mathtt{r\text{-}bal}}^{\mathrm{STAT}}$ as the *statistical threshold* for independent sets in $H_r(n,p)$ and γ-balanced independent sets in $H(r,n,p)$, respectively. The proof of Theorem 5 is based on the second moment method and is, therefore, non-constructive. This naturally leads to an algorithmic question: can we efficiently find large independent sets in $H_r(n,p)$ or large γ-balanced independent sets in $H(r,n,p)$? To address this question, we let

$$\alpha_{\mathtt{r}}^{\mathrm{COMP}} := ((r-1)!\log_b n)^{\frac{1}{r-1}} \qquad \text{and} \qquad \alpha_{\mathtt{r\text{-}bal}}^{\mathrm{COMP}} := \left(\frac{\log_b n}{\prod_{i=1}^r \gamma_i}\max_{1\le i\le r}\gamma_i\right)^{\frac{1}{r-1}}.$$

Our next result establishes that the thresholds $\alpha_{\mathtt{r\text{-}bal}}^{\mathrm{COMP}}$ and $\alpha_{\mathtt{r}}^{\mathrm{COMP}}$ are achievable in polynomial time via an online algorithm.

▶ **Theorem 6.** *For $H_r(n,p)$, there is an online algorithm which finds, for any $\epsilon > 0$, an independent set of size $(1-\epsilon)\alpha_r^{\rm COMP}$ whp. Likewise, for $H(r,n,p)$, there is an online algorithm which finds, for any $\epsilon > 0$, a γ-balanced independent set of size $(1-\epsilon)\alpha_{r\text{-}bal}^{\rm COMP}$ whp.*

We now compare the statistical thresholds with the algorithmically achievable values. For the dense r-uniform hypergraph $H_r(n,p)$, there is a factor-$r^{1/(r-1)}$ gap between the two, which recovers the well-known "factor-2 gap" for the special case $r=2$. As for $H(r,n,p)$, the gap is of the order $(\max_{1\le i\le r}\gamma_i)^{-1/(r-1)}$. (Note that for $\gamma_1=\cdots=\gamma_r=1/r$, the gaps in both cases are identical.)

Are there fundamental computational barriers in these models? Our final result addresses this question and yields sharp algorithmic lower bounds, suggesting that the barriers above are not spurious and likely inherent.

▶ **Theorem 7.** *For $H_r(n,p)$ and any $\epsilon > 0$, no online algorithm finds an independent set of size $(1+\epsilon)\alpha_r^{\rm COMP}$ with probability $o(1)$. Likewise, for $H(r,n,p)$ and any $\epsilon > 0$, no online algorithm finds a γ-balanced independent set of size $(1+\epsilon)\alpha_{r\text{-}bal}^{\rm COMP}$ with probability $o(1)$.*

Note that aside from onlineness, there is no restriction on the algorithms ruled out by Theorem 7. In particular, so long as the algorithm is online, we impose no restrictions on its runtime. Importantly, Theorem 7 rules out algorithms succeeding even with a vanishing probability $o(1)$. In relevant literature, this is known as *strong hardness*; see [43] as well as [32, 19].

▶ Remark 8. More explicitly, the $o(1)$ guarantee is of the form $\exp(-\Omega(\alpha_{\tt r}^{\rm COMP}\log n)$ for $H_r(n,p)$ and of form $\exp(-\Omega(\alpha_{\tt r\text{-}bal}^{\rm COMP}\log n))$ for $H(r,n,p)$. Both of these guarantees are essentially optimal, since the probability that a randomly chosen subset with size reaching the computational threshold is independent in $H_r(n,p)$ (or a γ-balanced independent set in $H(r,n,p)$) is of the same order.

Taken together, Theorems 6 and 7 sharply characterize the performance of online algorithms in dense r-uniform hypergraphs and dense r-partite hypergraphs, suggesting an inherent computational barrier, precisely at $\alpha_{\tt r}^{\rm COMP}$ and $\alpha_{\tt r\text{-}bal}^{\rm COMP}$. Our result recovers the analogous computational gaps present for $r=2$, generalizing them to $r>2$, thus pointing towards a certain universality.[6]

▶ Remark 9. Note that for $H_r(n,p)$, the computational gap is of the order $r^{1/(r-1)}$, which vanishes as $r\to\infty$. Likewise, the same conclusion also holds for $H(r,n,p)$, using the fact $\max_i\gamma_i\ge 1/r$. Thus, the computational problem gets "easier" when the sizes of the hyperedges become large.

▶ Remark 10. As discussed below, we introduce a novel variant of the greedy algorithm – staged and bucketed greedy – together with a carefully designed stopping time that accounts for both the multipartite structure and the balancedness constraint. Both ingredients are crucial for addressing hypergraphs and multipartite structures, and for moving beyond the classical $G(n,p)$ and bipartite setting. In this sense, our results substantially generalize several existing results in the literature. Indeed, for $H_r(n,p)$ with $r=2$, we recover both Karp's algorithmic result [46] as well as the lower bound of [32]. Likewise, for $H(r,n,p)$, our results with $r=2$ recover the full set of results in [19].

[6] We remark that such a universality arises also in the context of sparse random graphs, see [59, 21, 62].

1.3 Proof Overview

Our proofs for $H_r(n,p)$ and $H(r,n,p)$ follow a consistent three-part strategy: identifying the statistical threshold, demonstrating achievability via online algorithms, and applying OGP to establish matching algorithmic lower bounds. We focus first on $H_r(n,p)$ to clarify the underlying mechanics. We then discuss the technical innovations required for the more complex $H(r,n,p)$ setting, including the *staged and bucketed* greedy procedure and a new stopping time argument, both of which account for the multipartite structure and the global balancedness requirement.

The statistical threshold is established via the second moment method [5]. As for the algorithms, our approach is through an online greedy algorithm. To the best of our knowledge, ours is the first work to analyze the greedy independent set procedure on Erdős–Rényi hypergraphs for $r \geq 3$.

Our impossibility result is positioned in a body of work initiated by Gamarnik and Sudan [35], who investigated local algorithms for independent sets in sparse random graphs using the Overlap Gap Property (OGP); see below for a background. This property yields a dichotomy: the intersection of two "large" independent sets must be either significantly large or very small, creating a forbidden "gap" in overlap sizes. They demonstrated that a successful local algorithm could be leveraged to produce independent sets that violate this OGP. While this approach was refined by Rahman and Virág [60] and extended by Wein [62] to LDPs, these methods fail in dense settings. For dense random graphs, LDPs – and local algorithms, which can be implemented as LDPs – are conjectured to fail; see Conjecture 2. By contrast, the only known algorithmic frameworks that achieve large independent sets in this regime are online. The arguments of [62, 60] crucially rely on algorithmic stability, whereas online algorithms are known to be unstable [32]. Thus, the dense setting calls for substantially different techniques.

In a recent development, Gamarnik, Kızıldağ, and Warnke [32] introduced an OGP-based framework tailored for the online setting. This approach is inherently algorithm-dependent: it utilizes a stopping time to monitor output size, together with interpolation paths that evolve temporally as the algorithm progresses. It yielded sharp lower bounds within online algorithms for the classical Erdős–Rényi random graph $G(n,p)$. Extending these techniques to higher uniformities and multipartite models under a global balancedness constraint turns out to be substantially more challenging, due to additional technical hurdles that are absent in $G(n,p)$ (discussed below). Among other technical contributions, we introduce a new stopping-time argument tailored specifically to hypergraphs, which may be of independent interest. Taken together, our techniques allow us to complete the picture for hypergraph models: while the sparse regime was recently addressed by Dhawan and Wang [21], the dense regime had remained open, which is our focus.

Formally, we construct a family of correlated random hypergraphs that remain identical until a specific random stopping time, after which they evolve independently. Executing the algorithm on these correlated instances yields a collection of candidate independent sets with rigid structural constraints – most notably, they must coincide on all vertices revealed prior to the stopping time. We define these configurations as "forbidden tuples" and prove they are absent from the correlated random hypergraphs with high probability. Since a successful online algorithm would necessarily imply the existence of such tuples, this contradiction establishes our lower bound.

We next extend our analysis to the $H(r,n,p)$ model. Although the statistical threshold is determined via a second-moment argument, the γ-balancedness, a global constraint, adds substantial technical complexity. As for developing an online algorithm, the primary hurdle in this model is the global nature of the balancedness requirement, which must be maintained throughout the construction. To address this, we introduce a *staged and bucketed* greedy algorithm. The procedure is partitioned into r distinct stages; in each stage, the algorithm populates the independent set using vertices from "unexhausted" parts until some such part reaches its target capacity and is "locked." By controlling acceptance probabilities and bounding failure events via careful estimates, we prove that this algorithm yields a γ-balanced independent set of size $(1-\epsilon)\alpha_{\texttt{r-bal}}^{\mathrm{COMP}}$ with high probability.

The derivation of the impossibility result for $H(r,n,p)$ is substantially more involved. The main difficulties stem from two sources: (i) the multipartite nature of $H(r,n,p)$, together with the permutation invariance of the parts in the partition, and (ii) the γ-balancedness constraint, which imposes prescribed cardinality requirements on the intersections of the independent set with the r parts. Among other technical ideas, we introduce a novel stopping time that accounts for both issues simultaneously. Specifically, we stop when the independent set contains sufficiently many vertices from all parts of the partition (by contrast, the stopping time for $H_r(n,p)$ is triggered once the algorithm's output reaches a prescribed size threshold).

Background on OGP. As mentioned earlier, our algorithmic lower bound is based on the OGP framework pioneered by Gamarnik and Sudan [35, 36] (and termed in [33]). The OGP is among the most powerful techniques for identifying computationally hard phases of random optimization problems – i.e., those in which the objective function is random. Its origins lie in earlier works on random CSPs, where the apparent hard phases were observed to exhibit striking geometric phase transitions [3, 2, 57]. The OGP framework systematically converts such geometric features of the solution landscape into formal algorithmic lower bounds; see [29, 32] for surveys on the method.

We now describe the OGP in greater detail for random graphs – the context in which it was originally introduced. For the sparse Erdős–Rényi model $G(n, \frac{d}{n})$, the work of Gamarnik and Sudan [35, 36] established that independent sets of size greater than $(1+1/\sqrt{2})n\frac{\log d}{d}$ must either substantially overlap or be nearly disjoint.[7] As a consequence, they rigorously ruled out local algorithms at this threshold, thereby refuting a well-known conjecture by Hatami, Lovász, and Szegedy [39], widely believed at the time. A subsequent work by Rahman and Virág [60] extended this threshold down to $n\frac{\log d}{d}$ by analyzing the overlap patterns of many independent sets, below which polynomial-time algorithms are known [52]. This approach, now known as the multi OGP (m-OGP), has since proved instrumental in obtaining sharp algorithmic lower bounds in a variety of other models. In the context of random graphs, hardness at the $(1+1/\sqrt{2})n\frac{\log d}{d}$ threshold was later extended to LDPs [31], and sharp lower bounds for LDPs were subsequently obtained in [62] using an asymmetric variant of the m-OGP. Still more delicate refinements of the m-OGP were required in other models, such as random k-SAT [12] and spin glasses [41, 42, 43]. For the latter, this led to the branching OGP, an especially powerful variant organized around an ultrametric tree of solutions. Even more recently, novel variants of the OGP tailored for online algorithms – which need not be stable – have begun to emerge [32, 19].

A summary of our contributions along with potential future directions of inquiry are now in order.

[7] For sparse random graphs, the relevant asymptotic regime is the double limit $n \to \infty$ followed by $d \to \infty$.

1.4 Summary and Open Problems

In this work, we consider the computational hardness of finding large (γ-balanced) independent sets in random hypergraphs. We focus on online algorithms, one of the most natural algorithmic frameworks containing Karp's original algorithm for $G(n,p)$, which motivated his celebrated conjecture regarding statistical-computational gaps for independent sets in $G(n,p)$. Our results imply a computational gap of a multiplicative factor of $r^{1/(r-1)}$ for independent sets in $H_r(n,p)$ and a gap of a multiplicative factor of $(\max_i \gamma_i)^{-1/(r-1)}$ for γ-balanced independent sets in $H(r,n,p)$ with respect to online algorithms; these gaps match those shown for LDPs in the sparse regime [21].

We conclude this introduction with a description of potential future directions of inquiry.

Future Queries. The setting of [32] permits querying a limited set of *future edges* – edges incident to vertices not yet seen – at each step. That is, the decision at time t is based not only on the edges revealed so far, but also on a restricted set of such future edges. In this augmented model, [32] prove both lower bounds and algorithmic guarantees for independent sets in graphs: specifically, they show that algorithms with modest access to future information can in fact exceed the $(1+\epsilon)\log_b n$ threshold, albeit using quasi-polynomial time. Similarly, [19] prove analogous results for γ-balanced independent sets in Erdős–Rényi bipartite graphs.

This naturally raises the following question: can algorithms with limited future information outperform the $(1+\epsilon)\alpha_{\mathtt{r}}^{\mathrm{COMP}}$ threshold for $H_r(n,p)$ and the $(1+\epsilon)\alpha_{\mathtt{r\text{-}bal}}^{\mathrm{COMP}}$ threshold for $H(r,n,p)$? We conjecture that the answer is yes:

▶ **Conjecture 11.** *Let E be the set of all future edges ever revealed to the algorithm and denote by $\binom{\mathcal{A}(G)}{r}$ the set of all r-edges whose vertices are contained in the output $\mathcal{A}(G)$.*

1. *For any $\epsilon > 0$ and $r \in \mathbb{N}$, there exists a constant $c_{\epsilon,r} > 0$ and an online algorithm $\mathcal{A}$ that finds an independent set of size at least $(1+\epsilon)\alpha_{\mathtt{r}}^{\mathrm{COMP}}$ whp in $H_r(n,p)$, provided*

$$\left|E \cap \binom{\mathcal{A}(G)}{r}\right| \le c_{\epsilon,r}\Big(\log_b(np)\Big)^{\frac{r}{r-1}}.$$

2. *For any $\epsilon > 0$ and $r \in \mathbb{N}$, there exists a constant $c_{\epsilon,r} > 0$ and an online algorithm $\mathcal{A}$ that finds a γ-balanced independent set of size at least $(1+\epsilon)\alpha_{r\text{-}bal}^{\mathrm{COMP}}$ whp in $H(r,n,p)$, provided*

$$\left|E \cap \binom{\mathcal{A}(G)}{r}\right| \le c_{\epsilon,r}\Big(\log_b(np)\Big)^{\frac{r}{r-1}}.$$

▶ Remark 12. Due to r-uniformity, the amount of future information required to surpass the computational threshold is $O\left(\log^{\frac{r}{r-1}} n\right)$. For $r = 2$, this is the same scaling arising in [32, 19].

β-Independent Sets. For an integer $1 \le \beta \le r-1$, a β-independent set in a hypergraph $H = (V,E)$ is a set $I \subseteq V$ such that $|e \cap I| \le \beta$ for each $e \in E(H)$. These structures interpolate between the notions of *strong* and *weak* independent sets – the cases $\beta = 1$ and $\beta = r-1$ (note that $\beta = r-1$ corresponds to the usual notion of independent sets). While strong and weak independent sets have been heavily studied in the literature (both in the random and deterministic settings [50, 26, 16, 4, 14]), β-independent sets for $1 < \beta < r-1$ are not as well understood.

Regarding $H_r(n,p)$, the statistical threshold was determined by [50] in the sparse regime.[8] A curious feature of these structures is that the online greedy algorithm is not guaranteed to construct a β-independent set. To see this, consider an instance $H \sim H_r(n,p)$ and suppose, for example, at step t we have constructed an independent set I_t thus far. At this point, we do not know the structure of the subhypergraph $H' \coloneqq H[I_t \cup \{v_{t+1}, \ldots, v_n\}]$. In particular, while I_t is a β-independent set in $H[\{v_1, \ldots, v_t\}]$, it may not be one in H'. Generalizing our results on $H_r(n,p)$ – and, indeed, those of [21] in the sparse regime – to $\beta < r-1$ is an interesting direction.

Distance-k Independent Sets. In a similar flavor, a distance-k independent set in a graph $G = (V,E)$ is a set $I \subseteq V$ such that the vertices in I are pairwise of distance at least $k+1$; equivalently, it is an independent set in G^k (note that one recovers the usual notion of independent sets for $k = 1$). As above, the greedy algorithm is not guaranteed to construct a distance-k independent set in a graph G.

Very recently, the growth rate of the statistical threshold for the distance-k independence number was determined for $G(n,p)$ [24]. They do not pin down the leading constant factor and raise it as an open problem (see the discussion after [24, Conjecture 8]). In the deterministic setting, there are a number of results for special graph classes, most notably for line graphs [45, 8, 53]. We remark that the problem is trivial for random graphs when $p \geq \sqrt{\frac{2\log n}{n}}$ and $k \geq 2$ as such graphs have diameter 2 [9]. It would be interesting to investigate the problem in the sparse regime.

▶ **Question 13.** *Are there statistical–computational gaps for the maximum distance-k independent set problem when $p = o\left(\sqrt{\frac{2\log n}{n}}\right)$ and $k \geq 2$?*

References

1. Problem list of the AIM Workshop Low-degree polynomial methods in average-case complexity (organizers Sam Hopkins, Tselil Schramm, and Alex Wein), 2024. URL: `https://aimath.org/pastworkshops/lowdegreecomplexityrep.pdf`.
2. Dimitris Achlioptas and Amin Coja-Oghlan. Algorithmic barriers from phase transitions. In *2008 49th Annual IEEE Symposium on Foundations of Computer Science*, pages 793–802. IEEE, 2008. `doi:10.1109/FOCS.2008.11`.
3. Dimitris Achlioptas and Federico Ricci-Tersenghi. On the solution-space geometry of random constraint satisfaction problems. In *Proceedings of the thirty-eighth annual ACM symposium on Theory of computing*, pages 130–139, 2006. `doi:10.1145/1132516.1132537`.
4. Miklós Ajtai, János Komlós, Janos Pintz, Joel Spencer, and Endre Szemerédi. Extremal uncrowded hypergraphs. *Journal of Combinatorial Theory, Series A*, 32(3):321–335, 1982. `doi:10.1016/0097-3165(82)90049-8`.
5. Noga Alon and Joel H Spencer. *The probabilistic method.* John Wiley & Sons, 2016.
6. James Anderson, Anton Bernshteyn, and Abhishek Dhawan. Coloring graphs with forbidden bipartite subgraphs. *Combinatorics, Probability and Computing*, 32(1):45–67, 2023. `doi:10.1017/S0963548322000104`.
7. Afonso S Bandeira, Amelia Perry, and Alexander S Wein. Notes on computational-to-statistical gaps: predictions using statistical physics. *arXiv preprint*, 2018. `arXiv:1803.11132`.
8. Richard Bi, Peter Bradshaw, Abhishek Dhawan, and Jingwei Xu. The strong chromatic index of $k\_\{t,t\}$-free graphs. *arXiv preprint*, 2026. `arXiv:2603.15207`.

[8] They actually prove a stronger result in terms of the coloring variant of the parameter.

9 Béla Bollobás. The diameter of random graphs. *Transactions of the American Mathematical Society*, 267(1):41–52, 1981.

10 Béla Bollobás. *Random Graphs*. Cambridge Studies in Advanced Mathematics. Cambridge University Press, 2 edition, 2001.

11 Béla Bollobás and Paul Erdös. Cliques in random graphs. In *Mathematical Proceedings of the Cambridge Philosophical Society*, volume 80(3), pages 419–427. Cambridge University Press, 1976.

12 Guy Bresler and Brice Huang. The algorithmic phase transition of random k-sat for low degree polynomials. *arXiv preprint*, 2021. `arXiv:2106.02129`.

13 Amin Coja-Oghlan and Charilaos Efthymiou. On independent sets in random graphs. *Random Structures & Algorithms*, 47(3):436–486, 2015. `doi:10.1002/RSA.20550`.

14 Jonathan Cutler and AJ Radcliffe. Hypergraph independent sets. *Combinatorics, Probability and Computing*, 22(1):9–20, 2013. `doi:10.1017/S0963548312000454`.

15 Ewan Davies, Matthew Jenssen, Will Perkins, and Barnaby Roberts. On the average size of independent sets in triangle-free graphs. *Proceedings of the American Mathematical Society*, 146(1):111–124, 2018.

16 Abhishek Dhawan. Fractional coloring via entropy. `https://arxiv.org/abs/2603.17730` (preprint), 2026.

17 Abhishek Dhawan, Nhi U Dinh, Eren C Kızıldağ, Neeladri Maitra, and Bayram A Şahin. Algorithmic phase transition for large independent sets in dense hypergraphs. *arXiv preprint*, 2026. `arXiv:2605.05618`.

18 Abhishek Dhawan, Oliver Janzer, and Abhishek Methuku. Independent sets and colorings of $k_{t,t,t}$-free graphs. *arXiv preprint*, 2025. `arXiv:2511.17191`.

19 Abhishek Dhawan, Eren C Kızıldağ, and Neeladri Maitra. Sharp online hardness for large balanced independent sets. *arXiv preprint*, 2025. `arXiv:2508.20785`.

20 Abhishek Dhawan, Abhishek Methuku, and Minh-Quan Vo. The independence number of uncrowded hypergraphs: bounds matching the shattering threshold. *arXiv preprint*, 2026. `arXiv:2606.18048`.

21 Abhishek Dhawan and Yuzhou Wang. The low-degree hardness of finding large independent sets in sparse random hypergraphs. *SIAM Journal on Discrete Mathematics*, 40(1):374–419, 2026. `doi:10.1137/24M1701575`.

22 Uriel Feige. Relations between average case complexity and approximation complexity. In *Proceedings of the thiry-fourth annual ACM symposium on Theory of computing*, pages 534–543, 2002. `doi:10.1145/509907.509985`.

23 Uriel Feige and Shimon Kogan. Hardness of approximation of the balanced complete bipartite subgraph problem. *Dept. Comput. Sci. Appl. Math., Weizmann Inst. Sci., Rehovot, Israel, Tech. Rep. MCS04-04*, 2004.

24 Alan Frieze, Ross Kang, Aditya Raut, Michelle Sweering, and Hilde Verbeek. Coloring powers of random graphs. *arXiv preprint*, 2026. `arXiv:2604.14006`.

25 Alan Frieze and Colin McDiarmid. Algorithmic theory of random graphs. *Random Structures & Algorithms*, 10(1-2):5–42, 1997. `doi:10.1002/(SICI)1098-2418(199701/03)10:1/2\%3C5::AID-RSA2\%3E3.0.CO;2-Z`.

26 Alan Frieze and Dhruv Mubayi. Coloring simple hypergraphs. *Journal of Combinatorial Theory, Series B*, 103(6):767–794, 2013. `doi:10.1016/J.JCTB.2013.09.003`.

27 Alan Frieze and Nicholas C Wormald. Random k-sat: A tight threshold for moderately growing k. *Combinatorica*, 25(3):297–306, 2005. `doi:10.1007/S00493-005-0017-3`.

28 Alan M Frieze. On the independence number of random graphs. *Discrete Mathematics*, 81(2):171–175, 1990. `doi:10.1016/0012-365X(90)90149-C`.

29 David Gamarnik. The overlap gap property: A topological barrier to optimizing over random structures. *Proceedings of the National Academy of Sciences*, 118(41), 2021.

30 David Gamarnik. Turing in the shadows of nobel and abel: an algorithmic story behind two recent prizes. *arXiv preprint*, 2025. `doi:10.48550/arXiv.2501.15312`.

31 David Gamarnik, Aukosh Jagannath, and Alexander S Wein. Low-degree hardness of random optimization problems. In *2020 IEEE 61st Annual Symposium on Foundations of Computer Science (FOCS)*, pages 131–140. IEEE, 2020. doi:10.1109/FOCS46700.2020.00021.

32 David Gamarnik, Eren C Kızıldağ, and Lutz Warnke. Optimal hardness of online algorithms for large independent sets. *arXiv preprint*, 2025. arXiv:2504.11450.

33 David Gamarnik and Quan Li. Finding a large submatrix of a gaussian random matrix. *The Annals of Statistics*, 46(6A):2511–2561, 2018.

34 David Gamarnik, Cristopher Moore, and Lenka Zdeborová. Disordered systems insights on computational hardness. *Journal of Statistical Mechanics: Theory and Experiment*, 2022(11):114015, 2022.

35 David Gamarnik and Madhu Sudan. Limits of local algorithms over sparse random graphs. In *Proceedings of the 5th conference on Innovations in theoretical computer science*, pages 369–376, 2014. doi:10.1145/2554797.2554831.

36 David Gamarnik and Madhu Sudan. Limits of local algorithms over sparse random graphs. *Ann. Probab.*, 45(4):2353–2376, July 2017. doi:10.1214/16-AOP1114.

37 G. R. Grimmett and C. J. H. McDiarmid. On colouring random graphs. *Mathematical Proceedings of the Cambridge Philosophical Society*, 77(2):313–324, 1975. doi:10.1017/S0305004100051124.

38 Johan Håstad. Clique is hard to approximate within $n^{1-\epsilon}$. *Acta Mathematica*, 182(1):105–142, 1999.

39 Hamed Hatami, László Lovász, and Balázs Szegedy. Limits of locally–globally convergent graph sequences. *Geometric and Functional Analysis*, 24(1):269–296, 2014.

40 Samuel Brink Klevit Hopkins. Statistical inference and the sum of squares method, 2018.

41 Brice Huang and Mark Sellke. Tight lipschitz hardness for optimizing mean field spin glasses. *arXiv preprint*, 2021. arXiv:2110.07847.

42 Brice Huang and Mark Sellke. Algorithmic threshold for multi-species spherical spin glasses. *arXiv preprint*, 2023. doi:10.48550/arXiv.2303.12172.

43 Brice Huang and Mark Sellke. Strong low degree hardness for stable local optima in spin glasses. *arXiv preprint*, 2025. doi:10.48550/arXiv.2501.06427.

44 Mark Jerrum. Large cliques elude the metropolis process. *Random Structures & Algorithms*, 3(4):347–359, 1992. doi:10.1002/RSA.3240030402.

45 Tomáš Kaiser and Ross J Kang. The distance-t chromatic index of graphs. *Combinatorics, Probability and Computing*, 23(1):90–101, 2014. doi:10.1017/S0963548313000473.

46 Richard M. Karp. The probabilistic analysis of some combinatorial search algorithms, 1976.

47 Subhash Khot. Improved inapproximability results for maxclique, chromatic number and approximate graph coloring. In *Proceedings 42nd IEEE Symposium on Foundations of Computer Science*, pages 600–609. IEEE, 2001. doi:10.1109/SFCS.2001.959936.

48 Jeong Han Kim. On Brooks' theorem for sparse graphs. *Combinatorics, Probability and Computing*, 4(2):97–132, 1995. doi:10.1017/S0963548300001528.

49 Michael Krivelevich. Coloring random graphs—an algorithmic perspective. In *Mathematics and Computer Science II: Algorithms, Trees, Combinatorics and Probabilities*, pages 175–195. Springer, 2002.

50 Michael Krivelevich and Benny Sudakov. The chromatic numbers of random hypergraphs. *Random Structures & Algorithms*, 12(4):381–403, 1998. doi:10.1002/(SICI)1098-2418(199807)12:4\%3C381::AID-RSA5\%3E3.0.CO;2-P.

51 Dmitriy Kunisky, Alexander S Wein, and Afonso S Bandeira. Notes on computational hardness of hypothesis testing: Predictions using the low-degree likelihood ratio. In *ISAAC Congress (International Society for Analysis, its Applications and Computation)*, pages 1–50. Springer, 2019.

52 Joseph Lauer and Nicholas Wormald. Large independent sets in regular graphs of large girth. *Journal of Combinatorial Theory, Series B*, 97(6):999–1009, 2007. doi:10.1016/J.JCTB.2007.02.006.

53 Mohammad Mahdian. The strong chromatic index of c4-free graphs. *Random Structures & Algorithms*, 17(3-4):357–375, 2000. `doi:10.1002/1098-2418(200010/12)17:3/4\%3C357::AID-RSA9\%3E3.0.CO;2-Y`.

54 David W Matula. On the complete subgraphs of a random graphproc. In *Second Chapel Hill Conf. on Combinatorial Mathematics and its Applications (Univ. North Carolina, Chapel Hill, NC, 1970)*, pages 356–369, 1970.

55 David W Matula. *The largest clique size in a random graph*. Department of Computer Science, Southern Methodist University Dallas, TX, 1976.

56 Marc Mezard. Glassy phases: a possible origin of computational hardness. In *APS March Meeting Abstracts*, volume 2010, pages A6–004, 2010.

57 Marc Mézard, Thierry Mora, and Riccardo Zecchina. Clustering of solutions in the random satisfiability problem. *Physical Review Letters*, 94(19):197205, 2005.

58 Marc Mézard and Giorgio Parisi. Mean-field theory of randomly frustrated systems with finite connectivity. *Europhysics Letters*, 3(10):1067, 1987.

59 Will Perkins and Yuzhou Wang. On the hardness of finding balanced independent sets in random bipartite graphs. In *Proceedings of the 2024 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 2376–2397. SIAM, 2024. `doi:10.1137/1.9781611977912.85`.

60 Mustazee Rahman and Balint Virag. Local algorithms for independent sets are half-optimal. *The Annals of Probability*, 45(3):1543–1577, 2017.

61 James B Shearer. A note on the independence number of triangle-free graphs. *Discrete Mathematics*, 46(1):83–87, 1983. `doi:10.1016/0012-365X(83)90273-X`.

62 Alexander S Wein. Optimal low-degree hardness of maximum independent set. *Mathematical Statistics and Learning*, 4(3):221–251, 2022.

63 Alexander S Wein. Computational complexity of statistics: New insights from low-degree polynomials. *arXiv preprint*, 2025. `doi:10.48550/arXiv.2506.10748`.