Some easy optimization problems have the overlap-gap property

Shuangping Li FIFALSP@STANFORD.EDU and **Tselil Schramm** TSELIL@STANFORD.EDU
390 Jane Stanford Way, Stanford University, Stanford, CA 94305

Editors: Nika Haghtalab and Ankur Moitra

Abstract

We show that the shortest s-t path problem has the overlap-gap property in (i) sparse $\mathbb{G}(n,p)$ graphs and (ii) complete graphs with i.i.d. Exponential edge weights. Furthermore, we demonstrate that in sparse $\mathbb{G}(n,p)$ graphs, shortest path is solved by $O(\log n)$-degree polynomial estimators, and a uniform approximate shortest path can be sampled in polynomial time. This constitutes the first example in which the overlap-gap property is not predictive of algorithmic intractability for a (non-algebraic) average-case optimization problem.

Keywords: Overlap-gap property, low-degree polynomial lower bounds, average-case complexity.

1. Introduction

An instance $\mathcal{I}$ of an optimization problem is said to have the *overlap-gap property* (OGP) if its near-optimal solutions form multiple well-separated clusters. That is, every pair of near-optimal solutions S_1, S_2 are either close, with $\operatorname{dist}(S_1, S_2) \leqslant \varepsilon$ (in some distance metric over solutions), or separated, with $\operatorname{dist}(S_1, S_2) \geqslant 1-\delta$, and furthermore $\mathcal{I}$ has at least one pair of far apart solutions.

The definition was inspired by complexity-theoretic heuristics in statistical physics, developed in the context of random CSPs such as k-SAT (Mézard et al. (2005); Achlioptas and Ricci-Tersenghi (2006)). Suppose you have an interacting particle system whose low-energy states are the near-optimal solutions to the optimization problem $\mathcal{I}$. If $\mathcal{I}$ has low-energy states S_1, S_2 that are far apart, then in equilibrium the system should occasionally transition between S_1, S_2; if the solutions to $\mathcal{I}$ are clustered according to the OGP, then traversing from S_1 to S_2 requires passing through high-energy states, which should be hard for "natural" algorithms.

This heuristic picture can be translated into *unconditional* lower bounds against specific algorithms, particularly in the context of average-case optimization where one can precisely characterize the optimization landscape. Gamarnik and Sudan (2017) were the first to show that the OGP implies that local algorithms cannot find large independent sets in sparse random graphs. This simple and powerful idea has since been extended, and variations of the OGP are known to imply unconditional lower bounds against sufficiently "smooth" or Lipschitz algorithms such as Langevin Dynamics and Approximate Message Passing (Gamarnik et al. (2024); Huang and Sellke (2022)), low-degree polynomial algorithms (Gamarnik et al. (2024); Wein (2022)), and generally any algorithm which is stable to perturbations in the input. A detailed explanation and further references are available in the survey of Gamarnik (2021). A strengthening of the property known as *disorder chaos* has recently been used to give unconditional lower bounds for sampling from Ising models (El Alaoui et al. (2022)).

Remarkably, in many average-case optimization problems, the overlap-gap property kicks in at the known computational threshold. For example, consider the maximum independent set problem in a random graph of average degree d. The size of the maximum independent set concentrates very well around $2\log d/d$. Despite decades of effort, the best known algorithms can reliably find a set half this size, but no larger. Strikingly, a sequence of works (Gamarnik and Sudan (2017); Rahman and Virág (2017); Gamarnik et al. (2024); Wein (2022)) has shown that a type of overlap-gap property holds for independent sets of measure $(1+\varepsilon)\log d/d$, from which it is possible to conclude that no local/smooth/low-degree algorithm can reliably find independent sets of measure $> \log d/d$ (which is precisely where present day polynomial-time algorithms are stuck). The OGP is similarly consistent with computational limits for random k-SAT (Bresler and Huang (2022)), random Ising models and other spin glasses (Gamarnik et al. (2024); Huang and Sellke (2022, 2023); El Alaoui et al. (2022)), random knapsack (Gamarnik and Kızıldağ (2023)), and more. The only example where OGP is known to give an incorrect prediction of hardness is the k-XOR problem, which has the OGP even while it is easy to solve via Gaussian elimination Ibrahimi et al. (2012). The algebraic structure of k-XOR is known to throw off heuristics for predicting average-case complexity, for example for the planted analogue of k-XOR or for the related problem of "Learning Parity with Noise," so researchers have mostly shrugged this example off. In the last five years or so, it is the authors' sense that researchers have come to accept the overlap-gap property as evidence for computational intractability, perhaps not just for local/smooth/low-degree algorithms, but for algorithmic methods more generally.

The purpose of this paper is to caution against complacency regarding OGP lower bounds. Our main result is that the algorithmically easy shortest path problem has the overlap gap property in random graphs. Indeed, we show that the shortest path problem can even be solved by low-degree polynomial algorithms in sparse random graphs, and further, that approximate short paths are easy to sample.

1.1. Main results

We study the problem of the shortest s-t path in the Erdös-Rényi graph $\mathbb{G}(n,q)$, for $q=\Theta(\frac{\log n}{n})$ above the connectivity threshold. Polynomial-time algorithms for s-t path are a staple of undergraduate algorithms courses. Yet a simple application of the first moment method demonstrates that this problem has the OGP.

Theorem 1 (Informal version of Theorem 6) *The shortest (s,t)-path problem in $\mathbb{G}(n,C\frac{\log n}{n})$ has the overlap-gap property with high probability when $C>1$. Furthermore, if $\boldsymbol{G},\boldsymbol{G}'\sim\mathbb{G}(n,C\frac{\log n}{n})$ independently, then with high probability all near-shortest (s,t)-paths in $\boldsymbol{G}$ and $\boldsymbol{G}'$ are almost disjoint.*

We say that an algorithm is "stable" or "smooth" if it is Lipschitz in its inputs; that is, if $\boldsymbol{G},\boldsymbol{G}'$ are two graphs which differ in an α fraction of edges, then algorithm A is L-smooth if the output paths $A(\boldsymbol{G}),A(\boldsymbol{G}')$ differ in at most an $L\cdot\alpha$ fraction of their edges. This notion of smoothness can be generalized in a straightforward way to apply to randomized algorithms, etc. Via established techniques, we can show that the OGP implies that no smooth algorithm can reliably solve the (s,t)-path problem in $\mathbb{G}(n,q)$.

Corollary 2 (Informal version of Theorem 7) *Suppose A is a "stable" algorithm in the sense that when it runs and succeeds on graphs G,G' which agree on a ρ-fraction of edges, it produces paths $A(G),A(G')$ overlapping in a $\Omega(\varepsilon)$-fraction of their edges. Then A cannot have success probability better than $\frac{1-\rho}{6}$ in computing $(1+\varepsilon)$-approximate (s,t)-shortest paths in $\mathbb{G}(n,C\frac{\log n}{n})$.*

One concrete class of algorithms which behave stably on random inputs is *low-degree polynomial algorithms*, that is, algorithms A where $A(G)$ is a vector-valued polynomial, each entry of which is a bounded-degree polynomial in the adjacency matrix of graph G. Intuitively, when applied to correlated random graphs $\boldsymbol{G}, \boldsymbol{G}'$, a low-degree polynomial ought to be somewhat stable since low-degree functions are relatively resilient to noise.

This is an interesting class of algorithms because $O(\log n)$-degree polynomials appear to work as well as any other polynomial-time algorithm for a large class of average-case *planted problems*—a good overview is given in the thesis of Hopkins (2018) and in the survey by Kunisky et al. (2019), though there have been many developments since these were written. Mild stability of bounded-degree polynomials has been established in prior work on OGP; usually the resulting bounds imply that degree-D polynomials must fail with probability $\geqslant n^{-O(D)}$ (Gamarnik et al. (2024); Wein (2022)). Unfortunately the methods appearing in the literature were too quantitatively weak to apply in our context,[1] but by appealing to symmetry and an invariance principle of Caravenna et al. (2023), we are able to show that Theorem 1 implies the following lower bound for low-degree polynomials.

Proposition 3 (Informal version of Theorem 11) *For any fixed $D \in \mathbb{N}$, there exists $\delta = 2^{-O(D)}$ such that any degree-D polynomial approximation of the shortest (s,t)-path in $\mathbb{G}(n,q)$ fails with probability $\geqslant \delta$.*

We believe that our invariance principle-based techniques may be useful for improving low-degree polynomial lower bounds in other sparse random models (though we are cognizant of the fact that in light of our other results, this may be a moot point).

The lower bound from Theorem 3 is typical of low-degree lower bounds derived from OGP, in that one only rules out degree-D algorithms with exponentially small failure probability in D. But in most contexts, one would be perfectly content with a polynomial-time computable degree-$O(\log n)$ algorithm which succeeds with probability $\Omega(1)$. The shortest (s,t)-path problem witnesses this to be more than just a weird technicality:

Lemma 4 (Informal version of Theorem 12) *There is a degree-$O(\frac{\log n}{\log\log n})$ efficiently computable polynomial which exactly computes the indicator that edge (i,j) participates in a near-shortest (s,t)-path in $\mathbb{G}(n, C\frac{\log n}{n})$ with success probability $1-o(1)$.*

Our result highlights a potential brittleness of OGP lower bounds. The OGP implies unconditional lower bounds, but the subtle issue is that it only rules out algorithms with ultra-high success probability. Previously it was plausible that this was simply an artifact of the proof technique. Our results demonstrate that sometimes there are indeed algorithms which, despite being smooth, succeed with decent probability, even where lower bounds rule out ultra-high success probability.

Two short months after our manuscript was posted on arXiv, the work of Huang and Sellke (2025) has partially addressed this concern, showing that in some cases, if the OGP holds with probability $1-f(n)$, then using a new argument one can rule out algorithms of success probability $\Omega(1)$ and degree $o(\log\frac{1}{f(n)})$.

[1] We are working with sparse random graphs, and the OGP does not hold when there is an edge between s,t, which happens with probability $q = \Omega(\log n/n)$. The works of Gamarnik et al. (2024); Wein (2022) need OGP to hold with probability $1 - 1/n^{\Omega(D)}$ to derive a nontrivial conclusion for polynomials of degree D.

Dense models. Sparse average-case models are known to sometimes exhibit anomalous behavior. In the appendix we establish that the same overlap-gap phenomenon exists for shortest (s,t)-path in the dense average-case model wherein the complete graph K_n has i.i.d. Exponential edge weights, also known as *first passage percolation*. We show that this implies lower bounds against stable algorithms.

Sampling. OGP lower bounds have played a role in inspiring the related notion of "disorder chaos," which roughly asks that if $\boldsymbol{G}, \boldsymbol{G}'$ are correlated samples from $\mathbb{G}(n,q)$, then the uniform distributions π, π' over their $(1+\varepsilon)$-approximate shortest paths are far apart in Wasserstein distance. This property is known to imply lower bounds for sampling from π with smooth algorithms, such as Langevin Dynamics or Approximate Message Passing. In the appendix we show that shortest (s,t)-path in sparse random graphs does indeed have disorder chaos, and therefore cannot be sampled by smooth algorithms. On the other hand, as we further observe in the appendix, the set of all approximate (s,t)-shortest paths can be enumerated in polynomial time, and therefore one can also sample in polynomial time. To our knowledge this is the first known example of a problem which exhibits disorder chaos, but for which sampling is easy.

1.2. Discussion

The OGP gives unconditional lower bounds against smooth and stable algorithms, with hardness results that qualitatively match known algorithmic thresholds for a number of average case problems. But in the wake of our result, a number of questions need to be reconsidered.

1. *Are stable algorithms even appropriate for non-planted average case optimization problems?*

   The OGP is only appropriate for optimization problems where there are multiple near-optimal solutions, and furthermore the near-optimal solutions themselves are brittle to the addition of noise. Given this, it seems clear that smooth and stable algorithms are actually ill-suited to these tasks from the start. This is in sharp contrast to planted models, where we expect the planted optimal solutions to be noise stable. Perhaps we should be trying to rule out non-stable algorithms, such as linear programs and semidefinite programs, instead.

   At least in planted average-case models, logarithmic-degree polynomial bounds (so far) appear to give a good indication of the computational complexity of optimization (see e.g. Hopkins (2018); Kunisky et al. (2019)). Under the best of circumstances, polynomials of degree-D are only $\exp(\Theta(D))$-smooth, so it is plausible that degree-$O(\log n)$ polynomials are a reasonable model to focus on. In our original arxiv preprint, we asked whether one can rule out degree-$O(\log n)$ algorithms which succeed with constant probability. Two months later, the work of Huang and Sellke (2025) gives an affirmative answer for many problems which exhibit the OGP, so long as the OGP holds with probability $1-o(1/\log n)$ and the "ensemble OGP" argument has Markovian structure. Their results apply to e.g. maximum independent set in $\mathbb{G}(n, \frac{d}{n})$, and to max clique in $\mathbb{G}(n, \frac{1}{2})$.

2. *Is the class of degree-$O(\log n)$ polynomials expressive in the context of non-planted optimization? Concretely, can an algorithm for finding a $.9\log n$-sized clique in $\mathbb{G}(n,\frac{1}{2})$ be implemented as a degree-$O(\log n)$ polynomial?*

   It is still not clear how powerful degree-$O(\log n)$ polynomials are in the context of non-planted average case optimization problems. The best algorithms known for spin glasses,

k-SAT, and max-independent set in $\mathbb{G}(n, \frac{d}{n})$ can be implemented by low-degree polynomials (Bayati et al. (2015); Montanari (2021); Ivkov and Schramm (2024); Bresler and Huang (2022); Wein (2022)). However, we do not know how to implement the state-of-the art algorithm for max clique in $\mathbb{G}(n, \frac{1}{2})$ in the $O(\log n)$-degree model. To check if OGP lower bounds are, in general, meaningful for non-planted optimization problems, we should make sure that the algorithms ruled out by OGP in the "hard regime" are actually useful in the easy regime.

3. *Are there additional conditions under which we expect that OGP is a good heuristic for hardness?*

   As was mentioned above, problems with algebraic structure, like k-XOR, are widely considered to be "black sheep" in average-case complexity, and so heuristics like OGP are not considered to be indicative of the underlying computational complexity.

   Though (s,t)-path in unweighted graph can be solved by Gaussian Elimination, *shortest* path cannot Berlekamp et al. (1978). Still, the average-case shortest (s,t)-path problem has some structural features that set is apart from e.g. Ising models or max independent set. Salient differences include: (i) the size of the set of approximate solutions is $\mathrm{poly}(n)$ rather than exponential or subexponential; (ii) the probability of the overlap-gap property is only $1 - O(1/\operatorname{polylog} n)$ rather than $1 - O(1/\mathrm{poly}(n))$; (iii) an (s,t) path is subject to the "hard constraint" that it actually be an unbroken path, whereas in e.g. Ising models there are no such constraints. Perhaps one can attribute the failure of OGP to accurately predict hardness to one of these structural features, or to some other property of (s,t)-paths? The new work of Huang and Sellke (2025) suggests that (ii) might play a role, at least in some models.

   It seems that in order to understand if the OGP is a good heuristic prediction of hardness (beyond the concrete lower bounds one is able to prove from it), one must understand which features of shortest path distinguish it from other problems. A possible way to go about this is to revisit the "easy" problems from Algorithms 101 one at a time, and check if each of them has the OGP.

1.3. Technical overview

The proofs of most of our results are quite short. For the convenience of the reader, we give a high-level sense of the arguments here.

Overlap-gap property. We establish the overlap-gap property for (s,t)-paths in $\mathbb{G}(n,q)$ with $q = \frac{C \log n}{n}$ using a simple first moment argument.

We want to show that it is unlikely that two near-shortest paths from s to t overlap in a constant fraction of their edges.

The reason we expect this to be true is that graphs from $\mathbb{G}(n,q)$ have only a few short cycles, and if there were two s,t paths with near-minimum length which overlap in only, say, 50% of their edges, then this means there must be a short cycle in the neighborhood of s or t, which is not very likely. Formalizing this amounts to some relatively simple counting arguments. The existence of near-disjoint near-optimal paths comes from a second moment method lower bound on the number of near-optimal s,t paths involving vertices only in

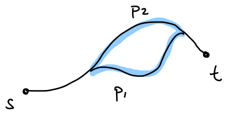

Figure 1: Paths overlapping on a constant fraction of edges create a short cycle.

$[n/2]$, and those involving vertices only in $\{n/2+1,\dots,n\}$. The arguments in First Passage Percolation are similar.

Stability of low-degree polynomials. In previous works, it was shown that low-degree polynomials in sparse random graphs are stable with probability $\Omega(1/\text{poly}(n))$ (Wein (2022); Gamarnik et al. (2024)). This is too weak for us, as our OGP can never hold with probability larger than $1-O(\log n/n)$; if there happens to be an edge between s and t, all bets are off.

To show that degree-$O(1)$ polynomials in the entries of the adjacency matrix of $\boldsymbol{G} \sim \mathbb{G}(n,q)$ are stable, we prove an invariance principle for arbitrary *symmetric* polynomials in correlated sparse Bernoulli random variables. Prior works such as Wein (2022); Gamarnik et al. (2024) did not exploit symmetry; however the symmetry of the model $\mathbb{G}(n,q)$ under vertex relabeling implies that any asymmetric polynomial can be symmetrized without hurting its performance. Furthermore, symmetric polynomials are more stable. To show stability, we are able to appeal to the invariance principle/interpolation argument of Caravenna et al. (2023), which applies to functions of sufficiently low influence, or rather, functions whose gradient has sufficiently low ℓ_4 norm. Bounding the gradient of an arbitrary symmetric function of $\boldsymbol{G}$ requires some nontrivial combinatorial arguments.

Low-degree algorithms and sampling. The length of the shortest path in $\boldsymbol{G} \sim \mathbb{G}(n,q)$ is $L = o(\log n)$ with high probability. The indicator for the existence of a specific path $v_1,\dots,v_L$ of length L in $\boldsymbol{G}$ can be written as a degree $L-1$ polynomial: $\prod_{i=1}^{L-1} \mathbf{1}[\boldsymbol{G}_{v_i,v_{i+1}} = 1]$. A good estimator for the indicator that edge (u,v) participates in the shortest path from s to t is the sum over the indicators of all paths of length L from s to t which include (u,v). Formalizing this requires some easy concentration arguments.

Sampling from the set of near-shortest paths can be done in polynomial time because the number of paths of length $(1+\varepsilon)L$ starting at vertex s is polynomial in n with high probability. One simply enumerates all of these paths, deletes the ones that do not end in t, and then samples one uniformly at random.

1.4. Notation

Throughout, we will use boldface font, such as $\boldsymbol{X}$, to denote that $\boldsymbol{X}$ is a random variable. We will use $n^{\underline{k}}$ to denote the falling factorial, $n^{\underline{k}} = n(n-1)\cdots(n-k+1)$. We use $\mathbb{G}(n,q)$ to denote the Erdős-Rényi distribution; almost everywhere, we will be in the setting where $q = \frac{C\log n}{n}$ for $C > 1$ a constant independent of n.

2. Shortest path in a random graph

Here we will provide some of the details for our main results regarding the OGP for shortest paths in sparse random graphs, with some proof details omitted. For the readers' convenience, the full manuscript has been reproduced in the appendix; there one may also find our results for disorder chaos and sampling, and for first passage percolation on randomly weighted complete graphs.

In this section, let $\boldsymbol{G} \sim \mathbb{G}(n,q)$, with $q = \frac{C\log n}{n}$ and $C > 1$. We will study the shortest (s,t)-path problem in $\boldsymbol{G}$. By symmetry of $\mathbb{G}(n,q)$, we can assume $s=1$ and $t=2$ without loss of generality.

With high probability, the shortest path between vertices 1 and 2 has length $\text{OPT} = (1+o(1))\frac{\log n}{\log nq}$ (we'll show as much, using the second moment method). Furthermore, we will show in Appendix A.2 that if we let $\mathcal{P}_\varepsilon(\boldsymbol{G})$ be the set of all paths of length $(1+\varepsilon)\mathrm{OPT}$, then with high probability $\mathcal{P}_\varepsilon(\boldsymbol{G})$ has the overlap-gap property: each pair of paths $p_1, p_2 \in \mathcal{P}_\varepsilon(\boldsymbol{G})$ overlaps on either an $O(\varepsilon)$ fraction of edges or on all of the edges; that is, there exists a constant C such that for all ε sufficiently small,

$$\frac{|p_1 \cap p_2|}{\sqrt{|p_1| \cdot |p_2|}} \in [0, C\varepsilon) \cup \{1\},$$

and further there exist paths p_1, p_2 which are almost disjoint. By-now standard arguments then imply that any sufficiently "smooth" algorithm cannot reliably find shortest paths in $\mathbb{G}(n,q)$. This demonstrates that efficient optimization methods can be successful for mean-field optimization problems, even in the presence of an overlap-gap structure.

Furthermore, we will show that the same is true for low-degree polynomials. For average-case optimization problems, especially planted problems, the best degree-$O(\log n)$ polynomial estimators (such as spectral algorithms), anecdotally, achieve the same computational thresholds as any polynomial-time algorithm. For this reason, lower bounds against degree $\omega(\log n)$ polynomials have become a heuristic for predicting information-computation gaps.

The OGP also gives lower bounds against degree-$O(1)$ polynomial estimators, because they behave smoothly on average-case inputs. We will show this is the case in Appendix A.3; unfortunately, our OGP only holds with probability $1 - 1/\operatorname{polylog} n$, so we will not be able to apply black-box arguments and we will need to do some work in order to apply an invariance principle (proven in the appendix). Though better than nothing, this is weaker evidence than a lower bound against a polynomial of degree-$\omega(\log n)$. In Appendix A.3 we will also show that the shortest path problem constitutes a cautionary example, by demonstrating that polynomials of degree $\Theta(\frac{\log n}{\log\log n})$ can indeed approximate the shortest path in $\mathbb{G}(n,q)$, despite the presence of an overlap gap.

2.1. Characterizing approximate shortest paths in Erdös-Rényi graphs

We begin by characterizing the length, OPT, of the shortest path in $\mathbb{G}(n,q)$, and the number of $(1+\varepsilon)\mathrm{OPT}$-length paths in $\mathbb{G}(n,q)$.

Lemma 5 (Shortest path in $\mathbb{G}(n,q)$) *If $\boldsymbol{G} \sim \mathbb{G}(n, \frac{C\log n}{n})$, then with probability $\geqslant 1 - O(\frac{1}{\log^d n})$ the length of the shortest path between $1, 2$ in $\boldsymbol{G}$ is $\frac{\log n}{\log\log n + \log C} \pm d$, and with probability $1 - O(\frac{\log\log n}{\log n})$, $|\mathcal{P}_\varepsilon(\boldsymbol{G})| = (1 \pm o(1))n^\varepsilon$.*

The proof follows from an application of the first and second moment method, and we defer it to the appendix.

2.2. The overlap-gap property in random graphs

Now we will use the second moment method to show that shortest paths in $\mathbb{G}(n,q)$ have the overlap-gap property. We will also derive a lower bound against any sufficiently stable algorithm.

Theorem 6 *Let $\boldsymbol{G}, \boldsymbol{G}' \sim \mathbb{G}(n,q)$ for $q = \frac{B\log n}{n}$ and $B > 1$, where $\boldsymbol{G}'$ is sampled from $\boldsymbol{G}$ by re-sampling each edge or non-edge with probability $1-\rho$. Then there exists a constant C such that* *for all $\varepsilon > 0$ sufficiently small, with probability $1 - O(\frac{\log\log n}{\log n})$ all pairs $(p, p') \in \mathcal{P}_\varepsilon(\boldsymbol{G}) \times \mathcal{P}_\varepsilon(\boldsymbol{G}')$ have*

$$\frac{|p \cap p'|}{\sqrt{|p| \cdot |p'|}} \in \begin{cases} [0, C\varepsilon) & \rho < (\frac{1}{\log n})^{2\varepsilon} \\ [0, C\varepsilon) \cup \{1\} & \textit{otherwise}. \end{cases}$$

Furthermore, with probability $\geqslant 1 - O(\frac{\log\log n}{\log n})$ there exist p, p' of overlap $< C\varepsilon$.

Proof Condition first on the outcome of Theorem 5. Now, let $N_{m,m',k}(\boldsymbol{G}, \boldsymbol{G}')$ be the number of pairs (p, p') where p is a 1-2 path of length m in $\boldsymbol{G}$, p' is a 1-2 path of length m' in $\boldsymbol{G}'$, $p \cap p'$ contains k edges, and suppose first $k < m, m'$ (ruling out overlap 1). We make use of the following claim, which we prove in the appendix:

Claim 2.1 *Let M_{k,d_1,d_2} be the number of pairs of paths p_1, p_2 with $|p_1| = d_1 + k$, $|p_2| = d_2 + k$, and $|p_1 \cap p_2| = k$ in K_n, with $d_1, d_2 > 0$. Then if $k, d_1, d_2 \ll n^{1/3}$, for all n sufficiently large,*

$$M_{k,d_1,d_2} \leqslant \left(\frac{k+1}{n^2} + \left(\frac{100k^3(d_1+k)(d_2+k)}{n} \right)^3 \right) n^{d_1+d_2+k}.$$

Applying Claim 2.1, we have

$$\begin{aligned} \mathbf{E}[N_{m,m',k}(\boldsymbol{G}, \boldsymbol{G}')] &= M_{k,m-k,m'-k} \cdot q^{m+m'-k} \cdot (\rho + (1-\rho)q)^k \\ &\leqslant \left(\frac{k+1}{n^2} + \frac{\log^{15} n}{n^3} \right) (nq)^{m+m'-k} (\rho + (1-\rho)q)^k \\ &\leqslant 2\frac{k+1}{n^2} (nq)^{m+m'-k} (\rho + (1-\rho)q)^k \leqslant 2(k+1)n^{2\varepsilon} \left(\frac{\rho + (1-\rho)q}{nq} \right)^k, \end{aligned}$$

where we have used that $m, m' \leqslant (1+\varepsilon)\frac{\log n}{\log nq}$ by Theorem 5. This is $O(1/\log^5 n)$ when $k > \frac{2\varepsilon\log n + 5\log\log n}{\log nq}$. Taking a union bound over all $O(\log^3 n)$ values of $\frac{\log n - 100}{\log nq} \leqslant m, m' \leqslant (1+\varepsilon)\frac{\log n}{\log nq}$ and $m > k > 2\varepsilon\frac{\log n}{\log nq}$ gives the first part of the result.

Now, suppose $k = m = m'$ (notice that if $k = m$, it must be the case that $m = m'$, since the endpoints of the paths are equal). In this case,

$$\mathbf{E}[N_{m,m,m}] \leqslant n^{m-1} q^m (\rho + (1-\rho)q)^m = \frac{(nq(\rho + (1-\rho)q))^m}{n}$$

which is $O(\frac{1}{\log n})$ when $\log(nq(\rho + (1-\rho)q)) < -1.5\varepsilon\log(nq)$, which occurs for $\rho \leqslant \frac{1}{(\log n)^{2\varepsilon}}$.

The final remark, regarding the existence of low-overlap pairs, follows from the lower bound on $|\mathcal{P}_\varepsilon(\boldsymbol{G})|$ in Theorem 5, along with the following observation: if one partitions $[n] \setminus \{1, 2\}$ into two disjoint and equally-sized sets of vertices, A and B, then Theorem 5 assures us that with high probability there will be $(n/2)^{\Omega(\varepsilon)}$ near-optimal paths with all vertices in A, and similarly in B. These paths will be disjoint. ■

The overlap-gap property allows us to rule out stable algorithms for shortest path:

Corollary 7 *Let $\rho \in [0, 1)$ be bounded away from 1, $\varepsilon > 0$ be sufficiently small, and n be sufficiently large. Then there can be no algorithm for ε-approximate shortest path which simultaneously* *(i) has failure probability $\leqslant \frac{1-\rho}{6}$, and (ii) is stable, in the sense that if $\boldsymbol{G}, \boldsymbol{G}' \sim \mathbb{G}(n, \frac{C \log n}{n})$ and the edges of $\boldsymbol{G}'$ are at least ρ-correlated with the edges of $\boldsymbol{G}$, then conditioned on $\mathcal{A}$ succeeding on its inputs, $\frac{|\mathcal{A}(\boldsymbol{G}) \cap \mathcal{A}(\boldsymbol{G}')|}{\sqrt{|\mathcal{A}(\boldsymbol{G})| \cdot |\mathcal{A}(\boldsymbol{G}')|}} > C\varepsilon$.*

Proof Choose $T = \lceil \frac{1}{1-\rho} \rceil = O(1)$, and note $\rho \geqslant 1 - \frac{1}{T}$. We sample a sequence of graphs $\boldsymbol{G}_0, \boldsymbol{G}_1, \ldots, \boldsymbol{G}_T$, with each $\boldsymbol{G}_t \sim \mathbb{G}(n, q)$, $\boldsymbol{G}_0, \boldsymbol{G}_T$ are independent, and each pair $\boldsymbol{G}_t, \boldsymbol{G}_{t+1}$ is marginally $(1-\frac{1}{T})$-correlated (that is, $\boldsymbol{G}_t, \boldsymbol{G}_{t+1}$ can be coupled with a pair $\boldsymbol{G}, \boldsymbol{G}'$ with $\boldsymbol{G} \sim \mathbb{G}(n, q)$ and $\boldsymbol{G}'$ obtained by resampling each edge with probability $\frac{1}{T}$). First, sample $\boldsymbol{G}_0, \boldsymbol{G}_T$ independently. Additionally, for each $(i, j) \in \binom{[n]}{2}$, sample an independent random variable $\boldsymbol{U}_{ij} \sim \mathrm{Unif}([0, 1])$. Now, for each t, let $\boldsymbol{G}_t(i, j) = \boldsymbol{G}_0(i, j) \cdot \mathbf{1}[\boldsymbol{U}_{ij} > \frac{t}{T}] + \boldsymbol{G}_T(i, j) \cdot \mathbf{1}[\boldsymbol{U}_{ij} \leqslant \frac{t}{T}]$. Clearly, $\boldsymbol{G}_t \sim \mathbb{G}(n, q)$. Also, edge (i, j) is resampled going from $\boldsymbol{G}_t$ to $\boldsymbol{G}_{t+1}$ if and only if $\boldsymbol{U}_{ij} \in (\frac{t}{T}, \frac{t+1}{T}]$, which is true for each edge independently with probability $\frac{1}{T}$. Thus, $\boldsymbol{G}_t, \boldsymbol{G}_{t+1}$ are marginally a $1 - \frac{1}{T}$-correlated pair.

Let $\boldsymbol{p}_t = \mathcal{A}(\boldsymbol{G}_t)$, and suppose $\mathcal{A}$ fails with probability δ. From Theorem 6, we have that with probability at least $1 - O(\frac{T \log \log n}{\log n}) \geqslant \frac{1}{2}$, each $\boldsymbol{p}_t$ must have overlap with $\boldsymbol{p}_0$ which is either equal to 1, or at most $C\varepsilon$, and further the overlap of $\boldsymbol{p}_0$ and $\boldsymbol{p}_T$ must be at most $C\varepsilon$ (as they are independent). Hence, with high probability there must exist some $t \in [T]$ with $\frac{|\boldsymbol{p}_0 \cap \boldsymbol{p}_t|}{\sqrt{|\boldsymbol{p}_0||\boldsymbol{p}_t|}} = 1$ but $\frac{|\boldsymbol{p}_0 \cap \boldsymbol{p}_{t+1}|}{\sqrt{|\boldsymbol{p}_0||\boldsymbol{p}_{t+1}|}} \leqslant C\varepsilon$, implying that $\frac{|\boldsymbol{p}_t \cap \boldsymbol{p}_{t+1}|}{\sqrt{|\boldsymbol{p}_t||\boldsymbol{p}_{t+1}|}} \leqslant C\varepsilon$. This is a contradiction, unless $\mathcal{A}$ did not succeed on some input, so the success probability on all inputs cannot exceed $\frac{1}{2}$. By a union bound $\mathcal{A}$ must have been successful on all $T + 1$ inputs with probability at least $1 - \delta(T + 1)$, so it must be the case that $\delta > \frac{1}{2(T+1)} \geqslant \frac{1-\rho}{6}$. ∎

2.3. Low-degree polynomial estimators

As discussed above, the shortest s, t-path problem can be solved, even exactly, in polynomial time. The overlap-gap property means that it cannot be solved by any sufficiently smooth algorithm. In particular, in Gamarnik et al. (2024); Wein (2022) it is shown that polynomials of degree $O(1)$ are smooth with probability $\Omega(1/\mathrm{poly}(n))$ when the input is an adjacency matrix of a graph sampled from $\mathbb{G}(n, q)$.[2] Here, we will apply an invariance principle to conclude that symmetric degree-$O(1)$ polynomials are smooth with probability $\Omega(1)$, which will allow us to prove a lower bound against degree-$O(1)$ polynomials.

But while degree $O(1)$ polynomials are smooth, degree $\omega(1)$ polynomials are not necessarily smooth. In Appendix A.3.1 we will show that the OGP indeed implies that degree-$O(1)$ polynomials cannot estimate s, t-paths, and in Appendix A.3.2 we'll show that polynomials of degree $\Omega(\log n / \log \log n)$ can estimate the shortest s, t-path problem very well.

We identify each path p from 1 to 2 by its indicator vector in $\{0, 1\}^{\binom{[n]}{2}}$.

[2] They also observe that this is the case for e.g. polynomials in Normal-valued variables, where the fact is just a consequence of hypercontractivity. In sparse random graphs, matters are a bit more complex.

Definition 8 *We say that $f : \{0,1\}^{\binom{[n]}{2}} \to \mathbb{R}^{\binom{n}{2}}$ is an (α,β)-approximation of the $(1+\varepsilon)$-approximate shortest path between $1,2$ in $\boldsymbol{G} \sim \mathbb{G}(n,q)$ if $\mathbf{E}[\|f\|^2] = \mathrm{OPT}$,*[3] *and*

$$\mathbf{Pr}\left[\exists p \in \mathcal{P}_\varepsilon(\boldsymbol{G}) \text{ s.t. } \frac{\langle f, p\rangle}{\sqrt{\mathbf{E}[\|f\|^2]}\cdot\|p\|} \geqslant \alpha, \quad \text{and} \quad \alpha \leqslant \frac{\|f\|^2}{\mathbf{E}[\|f\|^2]} \leqslant \frac{1}{\alpha}\right] \geqslant \beta.$$

2.3.1. LOWER BOUND FOR CONSTANT DEGREE POLYNOMIALS

Degree-$O(1)$ polynomial estimators are smooth algorithms. This has already been established in the literature, but extant results such as Gamarnik et al. (2024); Wein (2022) establish that stability holds with probability at least $1/\mathrm{poly}(n)$; here, because our overlap-gap property holds only with probability $1 - \frac{1}{\mathrm{polylog}\, n}$, we require a new statement which takes advantage of the fact that an estimator f is, without loss of generality, symmetric under vertex-relabelings.

Lemma 9 *Suppose $f : \{0,1\}^{\binom{[n]}{2}} \to \mathbb{R}^{\binom{[n]}{2}}$ is an approximation to the shortest $1,2$-path. For each $\pi \in S_n$, define $f_\pi(G) = \pi^{-1} f(\pi(G))$, and let $f^{\mathrm{sym}}(G) = \mathbf{E}_{\pi\sim \mathrm{Unif}(S_{[n]\setminus\{1,2\}})}\, \pi^{-1} f(\pi(G))$. Then for any coupling of graphs $\boldsymbol{G} \sim \mathbb{G}(n,q)$ and $s-t$ paths $\boldsymbol{p}$ in $\boldsymbol{G}$,*

$$\frac{\mathbf{E}\langle f(\boldsymbol{G}), \boldsymbol{p}\rangle}{\sqrt{\mathbf{E}\,\|f(\boldsymbol{G})\|^2}} \leqslant \frac{\mathbf{E}\langle f^{\mathrm{sym}}(\boldsymbol{G}), \boldsymbol{p}\rangle}{\sqrt{\mathbf{E}\,\|f^{\mathrm{sym}}(\boldsymbol{G})\|^2}}.$$

Proof The numerators are the same, because $\boldsymbol{G}$ and its $1-2$ paths have a distribution which is symmetric under permutations of $[n]$ which fix $1,2$, so

$$\mathop{\mathbf{E}}_{\boldsymbol{G},\boldsymbol{p}} \langle f(\boldsymbol{G}), \boldsymbol{p}\rangle = \mathop{\mathbf{E}}_{\boldsymbol{G},\boldsymbol{p}} \mathop{\mathbf{E}}_{\pi} \langle f(\pi(\boldsymbol{G})), \pi(\boldsymbol{p})\rangle = \mathop{\mathbf{E}}_{\boldsymbol{G},\boldsymbol{p}} \langle f^{\mathrm{sym}}(\boldsymbol{G}), \boldsymbol{p}\rangle.$$

The denominator on the right-hand side is only smaller by Jensen's inequality. ■

If f is a $(1-\eta, 1-\delta)$-approximation of the $(1+\varepsilon)$ shortest s,t-path for small enough $\eta, \delta, \varepsilon$, then by an averaging argument $\langle f(\boldsymbol{G}), \boldsymbol{p}\rangle$ will come close to saturating the Cauchy-Schwarz inequality with good probability. In this case, Theorem 9 implies that $f^{\mathrm{sym}}(G) = \mathbf{E}_\pi\, \pi^{-1}(f(\pi(G)))$ is at least as close to saturating Cauchy-Schwarz in expectation, from which another averaging argument implies that f^{sym} is also a $(1-\eta', 1-\delta')$-approximation to the s,t-shortest path, for η', δ' going to zero with η, δ. Since we will only be concerned with the small η, δ regime, from now on without loss of generality we consider only symmetric f, as ruling out f^{sym} suffices to rule out any f.

Symmetric functions are more stable than asymmetric functions, which allows us to give better quantitative guarantees than Gamarnik et al. (2024); Wein (2022). We will prove the following smoothness result in the appendix.

Theorem 10 *For $\rho \geqslant 1 - 1/T$, $\gamma \geqslant 3D(6e)^D/T$, and f a degree-D polynomial which is fixed by the action of $S_{[n]\setminus\{1,2\}}$, for ρ-correlated graphs $\boldsymbol{G}, \boldsymbol{G}' \sim \mathbb{G}(n,q)$, we have that*

$$\mathbf{Pr}\left[\|f(\boldsymbol{G}) - f(\boldsymbol{G}')\|_2^2 \geqslant \gamma\, \mathbf{E}[\|f(\boldsymbol{G})\|^2]\right] \leqslant \exp\left(-\frac{D}{3e}\left(\frac{\gamma T}{3D}\right)^{1/D}\right) + o_n(1).$$

[3] This is just a normalization convention; the parameters are easily adjusted for rescaling by $(1+\varepsilon)$.

Lower bound from overlap-gap property

Theorem 11 *For any integer $D \geqslant 0$ and $0 < \eta < 0.1$, there is a small enough constant $\delta > 0$ such that there is no degree-D polynomial which is a $(1-\eta, 1-\delta)$-approximation of the shortest path between $1, 2$ in $\mathbb{G}(n, q)$.*

Proof We give a proof by contradiction. Suppose there exists a degree-D polynomial which is a $(1-\eta, 1-\delta)$-approximation of the shortest path between $1, 2$ in $\mathbb{G}(n, q)$. Choose $T = \lceil \frac{1}{1-\rho} \rceil = O(1)$, and note $\rho \geqslant 1 - \frac{1}{T}$. We sample a sequence of graphs $\boldsymbol{G}_0, \boldsymbol{G}_1, \ldots, \boldsymbol{G}_T$ as in the proof of Theorem 7, where each $\boldsymbol{G}_t \sim \mathbb{G}(n, q)$, $\boldsymbol{G}_0, \boldsymbol{G}_T$ are independent, and each pair $\boldsymbol{G}_t, \boldsymbol{G}_{t+1}$ is marginally $(1 - \frac{1}{T})$-correlated. For shorthand, call $L = \frac{\log n}{\log \log n}$ to be the typical approximate length of the shortest path in $\boldsymbol{G}$. We define four events:

1. Let $\mathcal{E}_{\text{stable}}$ be the event that $\|f(\boldsymbol{G}_t) - f(\boldsymbol{G}_{t+1})\| \leqslant s\sqrt{\mathbf{E}[\|f(\boldsymbol{G})\|^2]}$ for all $0 \leqslant t \leqslant T-1$.
2. Let $\mathcal{E}_{\text{alg}}$ be the event that f is a $(1-\eta, 1-\delta)$-approximation of a path in $\mathcal{P}_\varepsilon(\boldsymbol{G}_t)$ for all $0 \leqslant t \leqslant T$.
3. Let $\mathcal{E}_{\text{gap}}$ be the event that for all $t \leqslant T-1$ and $(p_t, p_{t+1}) \in \mathcal{P}_\varepsilon(\boldsymbol{G}_t) \times \mathcal{P}_\varepsilon(\boldsymbol{G}_{t+1})$, $\frac{1}{L}\|p_t - p_{t+1}\|^2 \notin (0, 2-2C\varepsilon)$, and that for all $(p_0, p_T) \in \mathcal{P}_\varepsilon(\boldsymbol{G}_0) \times \mathcal{P}_\varepsilon(\boldsymbol{G}_T)$, $\frac{1}{L}\|p_0 - p_T\|^2 \geqslant 2 - 2C\varepsilon$.
4. Let $\mathcal{E}_{\text{paths}}$ be the event that for all $t \leqslant T$, $\mathcal{P}_\varepsilon(\boldsymbol{G}_t) \neq \emptyset$ and contains only paths of size in $[1, 1+\varepsilon]L$.

We argue that these events cannot happen simultaneously for small enough constant s and ε. Indeed, assume $\mathcal{E}_{\text{paths}}$ occurs, so that each set of paths is non-empty and the paths have the expected size; for shorthand, assume as well that $\mathcal{E}_{\text{alg}}$ occurs, so that with each output $f(\boldsymbol{G}_t)$ we may associate a path p_t with $\|f(\boldsymbol{G}_t) - p_t\|^2 \leqslant \|f(\boldsymbol{G}_t)\|^2 + \|p_t\|^2 - 2(1-\eta)\sqrt{\mathbf{E}\,\|f(\boldsymbol{G}_t)\|^2}\|p_t\| \leqslant 4\eta L$. Assume as well that $\mathcal{E}_{\text{gap}}$ occurs, so that $\frac{1}{L}\|p_0 - p_T\|^2 \geqslant 2 - 2C\varepsilon$, and for all t, $\frac{1}{L}\|p_t - p_{t+1}\|^2 \notin (0, 2-2C\varepsilon)$. This implies there must be a first time $t^* \in [T]$ where $\frac{1}{L}\|p_0 - p_{t^*}\|^2 \geqslant 2 - 2C\varepsilon$. But now if $\mathcal{E}_{\text{stable}}$ occurs, by triangle inequality

$$\begin{aligned}\|p_0 - p_{t^*}\| &\leqslant \|p_0 - p_{t^*-1}\| + \|p_{t^*-1} - p_{t^*}\| \leqslant \|p_{t^*-1} - p_{t^*}\| \\ &\leqslant \|p_{t^*-1} - f(\boldsymbol{G}_{t^*-1})\| + \|f(\boldsymbol{G}_{t^*-1}) - f(\boldsymbol{G}_{t^*})\| + \|f(\boldsymbol{G}_{t^*}) - p_{t^*}\| \\ &\leqslant 4\sqrt{\eta}\sqrt{L} + O(s\sqrt{L}) < \sqrt{(2-2C\varepsilon)L},\end{aligned}$$

for small constant s and ε, which is a contradiction.

Now, we lower-bound the probability of the events one-by-one. By construction, each $\boldsymbol{G}_t \sim \mathbb{G}(n, q)$ and each pair $\boldsymbol{G}_t, \boldsymbol{G}_{t+1}$ is marginally $(1-\frac{1}{T})$-correlated. By Theorem 10 and a union bound over pairs $\boldsymbol{G}_t, \boldsymbol{G}_{t+1}$, we have $\mathbf{Pr}[\overline{\mathcal{E}_{\text{stable}}}] \leqslant T\exp(-\frac{D}{3e}(\frac{s^2 T}{3D})^{1/D}) + o_n(1)$. The failure probability of the algorithm is at most δ by assumption, so by a union bound $\mathbf{Pr}[\overline{\mathcal{E}_{\text{alg}}}] \leqslant (T+1)\delta$. Finally, since $\boldsymbol{G}_0$ and $\boldsymbol{G}_T$ are independent, from Theorem 6 and Theorem 5 we have that $\mathbf{Pr}[\overline{\mathcal{E}_{\text{gap}}}] + \mathbf{Pr}[\overline{\mathcal{E}_{\text{paths}}}] \leqslant (T+1) \cdot O\left(\frac{\log\log n}{\log n}\right)$. Thus, it must be the case that

$$\begin{aligned}0 &\geqslant \mathbf{Pr}[\mathcal{E}_{\text{stable}} \cap \mathcal{E}_{\text{alg}} \cap \mathcal{E}_{\text{gap}} \cap \mathcal{E}_{\text{paths}}] \geqslant 1 - \mathbf{Pr}[\overline{\mathcal{E}_{\text{stable}}}] - \mathbf{Pr}[\overline{\mathcal{E}_{\text{alg}}}] - \mathbf{Pr}[\overline{\mathcal{E}_{\text{gap}}}] - \mathbf{Pr}[\overline{\mathcal{E}_{\text{paths}}}] \\ &\geqslant 1 - T\exp\left(-\frac{D}{3e}\left(\frac{s^2 T}{3D}\right)^{1/D}\right) + o_n(1) - (T+1)\cdot\delta - T \cdot O\left(\frac{\log\log n}{\log n}\right),\end{aligned}$$

the positivity of the right-hand-side for $\delta = \frac{1}{100(T+1)}$ and large enough T yields a contradiction. ■

2.3.2. A POLYNOMIAL ESTIMATOR OF LOGARITHMIC DEGREE

Here we will argue that the shortest s, t-path may be computed with reasonable success probability by degree $o(\log n)$ polynomials. This estimator does not, strictly speaking, fulfill Theorem 8, because of an issue of breaking symmetry. However,it gives almost full information about whether an edge participates in a short s, t-path.[4]

Lemma 12 *There exists a degree-$O(\frac{\log n}{\log \log n})$ polynomial in the entries of the adjacency matrix of $\boldsymbol{G}$ which achieves correlation $1 - o(1)$ with the indicator that (i, j) participates a path of length $m = \lceil \frac{\log n}{\log nq} \rceil + 1$ in $\boldsymbol{G}$. Furthermore, with high probability the polynomial is computable in polynomial time.*

This is also proven in the appendix. The idea is to define the polynomial

$$f_{ij}(G) = \sum_{\substack{p \in \mathcal{P}_{(m)}(K_n) \\ p \ni (i,j)}} \prod_{(a,b) \in p} \mathbf{1}[(a, b) \in E(G)],$$

where $\mathcal{P}_{(m)}(K_n)$ is the set of all m-paths from s to t in the complete graph. This is a degree-m polynomial in the adjancency matrix of G, which precisely counts how many m-paths between s, t in G contain the edge (i, j). Using simple concentration arguments, we can show that in $\boldsymbol{G} \sim \mathbb{G}(n, q)$ this polynomial is very well-correlated with the indicator that (i, j) participates in a path of length m from s to t, which gives the proof.

Acknowledgments

We thank David Gamarnik, Andrea Montanari, Guy Bresler, and Sam Hopkins for helpful conversations.

References

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.

Noga Alon and Joel H Spencer. *The probabilistic method*. John Wiley & Sons, 2016.

Noga Alon, Raphael Yuster, and Uri Zwick. Color-coding. *J. ACM*, 42(4):844–856, jul 1995. ISSN 0004-5411. doi: 10.1145/210332.210337. URL https://doi.org/10.1145/210332.210337.

Vikraman Arvind and Venkatesh Raman. Approximation algorithms for some parameterized counting problems. In *Algorithms and Computation: 13th International Symposium, ISAAC 2002 Vancouver, BC, Canada, November 21–23, 2002 Proceedings 13*, pages 453–464. Springer, 2002.

Dominique Bakry, Ivan Gentil, and Michel Ledoux. *Analysis and geometry of Markov diffusion operators*, volume 103. Springer, 2014.

[4] This raises another question about the usual OGP-based low-degree polynomial lower bounds: they may be brittle to the specific definition of a low-degree polynomial estimator.

Mohsen Bayati, Marc Lelarge, and Andrea Montanari. Universality in polytope phase transitions and message passing algorithms. *Annals of applied probability: an official journal of the Institute of Mathematical Statistics*, 25(2):753–822, 2015.

ER Berlekamp, RJ Mceliece, and HCA Van Tilborg. On the inherent intractability of certain coding problems. *IEEE Transactions on Information Theory*, 1978.

Shankar Bhamidi and Remco van der Hofstad. Weak disorder asymptotics in the stochastic mean-field model of distance. *The Annals of Applied Probability*, pages 29–69, 2012.

Béla Bollobás*, David Gamarnik, Oliver Riordan, and Benny Sudakov. On the value of a random minimum weight steiner tree. *Combinatorica*, 24(2):187–207, 2004.

Matthew S Brennan, Guy Bresler, Sam Hopkins, Jerry Li, and Tselil Schramm. Statistical query algorithms and low degree tests are almost equivalent. In *Conference on Learning Theory*, pages 774–774. PMLR, 2021.

Guy Bresler and Brice Huang. The algorithmic phase transition of random k-sat for low degree polynomials. In *2021 IEEE 62nd Annual Symposium on Foundations of Computer Science (FOCS)*, pages 298–309. IEEE, 2022.

Francesco Caravenna, Rongfeng Sun, and Nikos Zygouras. The critical 2d stochastic heat flow. *Inventiones mathematicae*, 233(1):325–460, 2023.

Ahmed El Alaoui, Andrea Montanari, and Mark Sellke. Sampling from the Sherrington-Kirkpatrick Gibbs measure via algorithmic stochastic localization. In *2022 IEEE 63rd Annual Symposium on Foundations of Computer Science (FOCS)*, pages 323–334. IEEE, 2022.

David Gamarnik. The overlap gap property: A topological barrier to optimizing over random structures. *Proceedings of the National Academy of Sciences*, 118(41):e2108492118, 2021.

David Gamarnik and Eren C Kızıldağ. Algorithmic obstructions in the random number partitioning problem. *The Annals of Applied Probability*, 33(6B):5497–5563, 2023.

David Gamarnik and Madhu Sudan. Limits of local algorithms over sparse random graphs. *Annals of probability: An official journal of the Institute of Mathematical Statistics*, 45(4):2353–2376, 2017.

David Gamarnik, Aukosh Jagannath, and Alexander S Wein. Hardness of random optimization problems for Boolean circuits, low-degree polynomials, and Langevin dynamics. *SIAM Journal on Computing*, 53(1):1–46, 2024.

Samuel Hopkins. *Statistical inference and the sum of squares method.* PhD thesis, Cornell University, 2018.

Brice Huang and Mark Sellke. Tight Lipschitz hardness for optimizing mean field spin glasses. In *2022 IEEE 63rd Annual Symposium on Foundations of Computer Science (FOCS)*, pages 312–322. IEEE, 2022.

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

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

Morteza Ibrahimi, Yashodhan Kanoria, Matt Kraning, and Andrea Montanari. The set of solutions of random XORSAT formulae. In *Proceedings of the twenty-third annual ACM-SIAM symposium on Discrete Algorithms*, pages 760–779. SIAM, 2012.

Misha Ivkov and Tselil Schramm. Semidefinite programs simulate approximate message passing robustly. In *Proceedings of the 56th Annual ACM Symposium on Theory of Computing*, pages 348–357, 2024.

Dmitriy Kunisky. Low coordinate degree algorithms i: Universality of computational thresholds for hypothesis testing. *arXiv preprint arXiv:2403.07862*, 2024.

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.

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

Andrea Montanari. Optimization of the Sherrington–Kirkpatrick Hamiltonian. *SIAM Journal on Computing*, (0):FOCS19–1, 2021.

Mustazee Rahman and Bálint Virág. Local algorithms for independent sets are half-optimal. *Annals of Probability*, 45(3), 2017.

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

In this appendix, Appendix A contains the full content of section 2 of our paper: we give full details of the proof that the shortest $s-t$ path problem in sparse random graphs has the overlap-gap property and deriving consequences for smooth algorithms, low-degree algorithms, and sampling. We provide similar results for first-passage percolation in the complete graph in Appendix B. In Appendix C we prove an invariance principle for symmetric polynomials of fixed degree, which is needed for the results in Appendix A.3.

Appendix A. Shortest path in a random graph

In this section, let $\boldsymbol{G} \sim \mathbb{G}(n,q)$, with $q = \frac{C \log n}{n}$ and $C > 1$. We will study the shortest (s,t)-path problem in $\boldsymbol{G}$. By symmetry of $\mathbb{G}(n,q)$, we can assume $s = 1$ and $t = 2$ without loss of generality.

With high probability, the shortest path between vertices 1 and 2 has length $\mathrm{OPT} = (1 + o(1))\frac{\log n}{\log nq}$ (we'll show as much, using the second moment method). Furthermore, we will show in Appendix A.2 that if we let $\mathcal{P}_\varepsilon(\boldsymbol{G})$ be the set of all paths of length $(1+\varepsilon)\mathrm{OPT}$, then with high probability $\mathcal{P}_\varepsilon(\boldsymbol{G})$ has the overlap-gap property: each pair of paths $p_1, p_2 \in \mathcal{P}_\varepsilon(\boldsymbol{G})$ overlaps on either an $O(\varepsilon)$ fraction of edges or on all of the edges; that is, there exists a constant C such that for all ε sufficiently small,

$$\frac{|p_1 \cap p_2|}{\sqrt{|p_1| \cdot |p_2|}} \in [0, C\varepsilon) \cup \{1\},$$

and further there exist paths p_1, p_2 which are almost disjoint. By-now standard arguments then imply that any sufficiently "smooth" algorithm cannot reliably find shortest paths in $\mathbb{G}(n,q)$. This demonstrates that efficient optimization methods can be successful for mean-field optimization problems, even in the presence of an overlap-gap structure.

Furthermore, we will show that the same is true for low-degree polynomials. For average-case optimization problems, especially planted problems, the best degree-$O(\log n)$ polynomial estimators (such as spectral algorithms), anecdotally, achieve the same computational thresholds as any polynomial-time algorithm. For this reason, lower bounds against degree $\omega(\log n)$ polynomials have become a heuristic for predicting information-computation gaps.

The OGP also gives lower bounds against degree-$O(1)$ polynomial estimators, because they behave smoothly on average-case inputs. We will show this is the case in Appendix A.3; unfortunately, our OGP only holds with probability $1 - 1/\operatorname{polylog} n$, so we will not be able to apply black-box arguments and we will need to do some work in order to apply an invariance principle (we do this in Appendix C). Though better than nothing, this is weaker evidence than a lower bound against a polynomial of degree-$\omega(\log n)$. In Appendix A.3 we will also show that the shortest path problem constitutes a cautionary example, by demonstrating that polynomials of degree $\Theta(\frac{\log n}{\log\log n})$ can indeed approximate the shortest path in $\mathbb{G}(n,q)$, despite the presence of an overlap gap.

Finally, we will argue in Appendix A.4 that the uniform distribution over $\mathcal{P}_\varepsilon(\boldsymbol{G})$ exhibits disorder chaos, but that sampling from this distribution is easy because one can enumerate $\mathcal{P}_\varepsilon(\boldsymbol{G})$ in polynomial time.

A.1. Characterizing approximate shortest paths in Erdös-Rényi graphs

We begin by characterizing the length, OPT, of the shortest path in $\mathbb{G}(n,q)$, and the number of $(1+\varepsilon)\mathrm{OPT}$-length paths in $\mathbb{G}(n,q)$.

Lemma 13 (Shortest path in $\mathbb{G}(n,q)$) *If $\boldsymbol{G} \sim \mathbb{G}(n, \frac{C \log n}{n})$, then with probability $\geqslant 1 - O(\frac{1}{\log^d n})$ the length of the shortest path between $1, 2$ in $\boldsymbol{G}$ is $\frac{\log n}{\log\log n + \log C} \pm d$, and with probability $1 - O(\frac{\log \log n}{\log n})$, $|\mathcal{P}_\varepsilon(\boldsymbol{G})| = (1 \pm o(1))n^{\varepsilon}$.*

Proof Let $N_m(\boldsymbol{G})$ be the number of paths of length m between $1, 2$ in $\boldsymbol{G}$. In expectation,

$$\mathbf{E}[N_m(\boldsymbol{G})] = n^{\underline{m-1}} q^m \leqslant \frac{1}{n}(nq)^m.$$

Hence by Markov's inequality, if $m < \frac{\log n - \log \frac{1}{\delta}}{\log nq}$, $N_m(\boldsymbol{G}) = 0$ with probability $\geqslant 1 - \delta$. This gives the probable upper bound on the length of the shortest path. Henceforth, define $\ell^* = \frac{\log n}{\log nq}$.

Now, we argue that for $m = (1+\varepsilon)\ell^*$, N_m has expectation $\sim n^{\varepsilon}$, and concentrates around its expectation. Since $n^{\underline{m}} \geqslant n^m(1 - \frac{m}{n})^m \geqslant n^m(1 - \frac{m^2}{n})$, we have that for $m \leqslant \log n$,

$$\mathbf{E}[N_m(\boldsymbol{G})] \geqslant \frac{(nq)^m}{n}\left(1 - \tfrac{\log^2 n}{n}\right) \geqslant n^{m/\ell^* - 1}\left(1 - \tfrac{\log^2 n}{n}\right) \geqslant n^{\varepsilon}\left(1 - \tfrac{\log^2 n}{n}\right).$$

We now bound the variance of N_m. Let $\mathcal{P}_{(m)}$ be the set of all paths of length m between $1, 2$ in K_n. Then

$$\mathbf{E}[N_m^2] = \sum_{p_1, p_2 \in \mathcal{P}_{(m)}} \mathbf{Pr}[p_1, p_2 \in \boldsymbol{G}] = \sum_{p_1, p_2 \in \mathcal{P}_{(m)}} q^{2m - |p_1 \cap p_2|}, \tag{1}$$

where the notation $|p_1 \cap p_2|$ refers to the number of common edges. We will parameterize the above sum according to the number of edges in $p_1 \cap p_2$, and according to the number of connected components in $p_1 \cap p_2$. We give a more general characterization than we need here, as it will be useful in later proofs.

Claim A.1 *Let M_{k,d_1,d_2} be the number of pairs of paths p_1, p_2 with $|p_1| = d_1 + k$, $|p_2| = d_2 + k$, and $|p_1 \cap p_2| = k$ in K_n, with $d_1, d_2 > 0$. Then if $k, d_1, d_2 \ll n^{1/3}$, for all n sufficiently large,*

$$M_{k,d_1,d_2} \leqslant \left(\frac{k+1}{n^2} + \left(\frac{100k^3(d_1+k)(d_2+k)}{n}\right)^3\right) n^{d_1+d_2+k}.$$

Proof Suppose that the intersection graph $p_1 \cap p_2$ has $c \geqslant 2$ components (at least two because $d_1, d_2 > 0$, and the endpoints $1, 2$ are present in both paths), and k edges.

First, if $c = 2$, then one can only choose how many edges to include in the component containing the source vertex, 1. There are at most $k+1$ choices for this. Then, there are $d_1 + d_2 + 2k - (k + c) = d_1 + d_2 + k - c$ choices for the vertex labels, giving a total of at most $(k+1)n^{d_1+d_2+k-2}$ such graphs.

Now, suppose that $c > 2$, and suppose that δ of the components contain no edges. By a "stars and bars" argument, the number of choices for the number of edges in each component is $\binom{k-1}{c-\delta-1}$. Since p_1, p_2 must intersect on sub-paths, the edges participating in each common component can be chosen by (i) choosing if the component appears in lexicographic or reverse-lexicographic order in p_2 (with respect to p_1), (ii) choosing a first edge in p_1 and either a first or last edge in p_2, depending on the lexicographic choice. The number of such choices is $2^{c-\delta} \cdot \binom{d_1+k}{c-\delta}(d_2+k)^{\underline{c-\delta}}$. Now, label the vertices: since there are c components and k edges shared in $p_1 \cap p_2$, and the intersection forms a tree, there are $k+c$ vertices in $p_1 \cap p_2$. So the total number of vertices to label is $d_1+d_2+2k-(k+c) = d_1+d+2+k-c$, to avoid double-counting overlapping vertices (and accounting for the fact that the endpoints $1,2$ are already labeled).

This counts all valid overlapping paths, and may overcount since sometimes the components end up overlapping.

The total number is thus at most

$$\binom{k-1}{c-\delta-1} \cdot 2^{c-\delta}(c-\delta)! \binom{d_1+k}{c-\delta}\binom{d_2+k}{c-\delta} \cdot n^{d_1+d_2+k-c} \leqslant \left(\frac{2e^3k(d_1+k)(d_2+k)}{n}\right)^c n^{d_1+d_2+k},$$

where we have used Stirling's approximation.

Summing over all $\leqslant k$ choices of c and all $\leqslant k$ choices of δ, for n sufficiently large,

$$M_{k,d_1,d_2} \leqslant \left(\frac{k+1}{n^2} + \left(\frac{100k^3(d_1+k)(d_2+k)}{n}\right)^3\right) n^{d_1+d_2+k}.$$

■

Returning to Equation (1), we have

$$\begin{aligned}
\mathbf{E}[N_m^2] &\leqslant n^{2m-2}q^{2m} + n^{m-1}q^m + \sum_{k=1}^{m-1} M_{k,m-k,m-k} \cdot q^{2m-k} \\
&\leqslant \left(1+O(\tfrac{1}{n^\varepsilon})\right) n^{2m-2}q^{2m} + \sum_{k=1}^{m-1}\left(\frac{k+1}{n^2} + \frac{\log^{15} n}{n^3}\right)(nq)^{2m-k} \\
&\leqslant \left(1+O(\tfrac{1}{n^\varepsilon})\right) n^{2m-2}q^{2m} + n^{2m-2}q^{2m}\sum_{k=1}^{m-1}\left(\frac{k+1}{(nq)^k} + \frac{\log^{15} n}{n(nq)^k}\right) \\
&\leqslant \left(1+O(\tfrac{1}{n^\varepsilon})\right) n^{2m-2}q^{2m} + O\left(\tfrac{1}{\log n}\right) \cdot n^{2m-2}q^{2m} = \left(1+O(\tfrac{1}{\log n})\right)\mathbf{E}[N_m]^2.
\end{aligned}$$

Hence applying Chebyshev's inequality we have that $N_{(1+\varepsilon)\ell^*} \geqslant \frac{1}{2}n^\varepsilon$ with probability $\geqslant 1 - O(\frac{1}{\log n})$.

From the concentration of the N_m quantity and $m = (1+O(\frac{1}{\log n}))\ell^*$ and $(1+\varepsilon)\ell^*$, we conclude that $\mathrm{OPT} \leqslant \ell^*+1 = (1+O(\frac{1}{\log\log n}))\ell^*$ with probability $\geqslant 1 - O(\frac{\log\log n}{\log n})$, and that $|\mathcal{P}_\varepsilon(\boldsymbol{G})| = (1 \pm o(1))n^\varepsilon$ with probability $\geqslant 1 - O(\frac{\log\log n}{\log n})$. ■

A.2. The overlap-gap property in random graphs

Now we will use the second moment method to show that shortest paths in $\mathbb{G}(n,q)$ have the overlap-gap property. We will also derive a lower bound against any sufficiently stable algorithm.

Theorem 14 *Let $\boldsymbol{G}, \boldsymbol{G}' \sim \mathbb{G}(n,q)$ for $q = \frac{B\log n}{n}$ and $B > 1$, where $\boldsymbol{G}'$ is sampled from $\boldsymbol{G}$ by re-sampling each edge or non-edge with probability $1-\rho$. Then there exists a constant C such that* *for all $\varepsilon > 0$ sufficiently small, with probability $1 - O(\frac{\log\log n}{\log n})$ all pairs $(p, p') \in \mathcal{P}_\varepsilon(\boldsymbol{G}) \times \mathcal{P}_\varepsilon(\boldsymbol{G}')$ have*

$$\frac{|p \cap p'|}{\sqrt{|p| \cdot |p'|}} \in \begin{cases} [0, C\varepsilon) & \rho < (\frac{1}{\log n})^{2\varepsilon} \\ [0, C\varepsilon) \cup \{1\} & \textit{otherwise}. \end{cases}$$

Furthermore, with probability $\geqslant 1 - O(\frac{\log\log n}{\log n})$ there exist p, p' of overlap $< C\varepsilon$.

Proof Condition first on the outcome of Theorem 13. Now, let $N_{m,m',k}(\boldsymbol{G}, \boldsymbol{G}')$ be the number of pairs (p, p') where p is a 1-2 path of length m in $\boldsymbol{G}$, p' is a 1-2 path of length m' in $\boldsymbol{G}'$, $p \cap p'$ contains k edges, and suppose first $k < m, m'$ (ruling out overlap 1). Applying Claim A.1, we have

$$\begin{aligned} \mathbf{E}[N_{m,m',k}(\boldsymbol{G}, \boldsymbol{G}')] &= M_{k,m-k,m'-k} \cdot q^{m+m'-k} \cdot (\rho + (1-\rho)q)^k \\ &\leqslant \left(\frac{k+1}{n^2} + \frac{\log^{15} n}{n^3}\right) (nq)^{m+m'-k} (\rho + (1-\rho)q)^k \\ &\leqslant 2\frac{k+1}{n^2} (nq)^{m+m'-k} (\rho + (1-\rho)q)^k \\ &\leqslant 2(k+1)n^{2\varepsilon} \left(\frac{\rho + (1-\rho)q}{nq}\right)^k, \end{aligned}$$

where we have used that $m, m' \leqslant (1+\varepsilon)\frac{\log n}{\log nq}$ by Theorem 13. This is $O(1/\log^5 n)$ when $k > \frac{2\varepsilon \log n + 5 \log\log n}{\log nq}$. Taking a union bound over all $O(\log^3 n)$ values of $\frac{\log n - 100}{\log nq} \leqslant m, m' \leqslant (1+\varepsilon)\frac{\log n}{\log nq}$ and $m > k > 2\varepsilon\frac{\log n}{\log nq}$ gives the first part of the result.

Now, suppose $k = m = m'$ (notice that if $k = m$, it must be the case that $m = m'$, since the endpoints of the paths are equal). In this case,

$$\mathbf{E}[N_{m,m,m}] \leqslant n^{m-1} q^m (\rho + (1-\rho)q)^m = \frac{(nq(\rho + (1-\rho)q))^m}{n}$$

which is $O(\frac{1}{\log n})$ when $\log(nq(\rho + (1-\rho)q)) < -1.5\varepsilon \log(nq)$, which occurs for $\rho \leqslant \frac{1}{(\log n)^{2\varepsilon}}$.

The final remark, regarding the existence of low-overlap pairs, follows from the lower bound on $|\mathcal{P}_\varepsilon(\boldsymbol{G})|$ in Theorem 13, along with the following observation: if one partitions $[n] \setminus \{1, 2\}$ into two disjoint and equally-sized sets of vertices, A and B, then Theorem 13 assures us that with high probability there will be $(n/2)^{\Omega(\varepsilon)}$ near-optimal paths with all vertices in A, and similarly in B. These paths will be disjoint. ■

The overlap-gap property allows us to rule out stable algorithms for shortest path:

Corollary 15 *Let $\rho \in [0, 1)$ be bounded away from 1, $\varepsilon > 0$ be sufficiently small, and n be sufficiently large. Then there can be no algorithm for ε-approximate shortest path which simultaneously (i) has failure probability $\leqslant \frac{1-\rho}{6}$, and (ii) is stable, in the sense that if $\boldsymbol{G}, \boldsymbol{G}' \sim \mathbb{G}(n, \frac{C\log n}{n})$ and the edges of $\boldsymbol{G}'$ are at least ρ-correlated with the edges of $\boldsymbol{G}$, then conditioned on $\mathcal{A}$ succeeding on its inputs, $\frac{|\mathcal{A}(\boldsymbol{G}) \cap \mathcal{A}(\boldsymbol{G}')|}{\sqrt{|\mathcal{A}(\boldsymbol{G})| \cdot |\mathcal{A}(\boldsymbol{G}')|}} > C\varepsilon$.*

Proof Choose $T = \lceil \frac{1}{1-\rho} \rceil = O(1)$, and note $\rho \geqslant 1 - \frac{1}{T}$. We sample a sequence of graphs $\boldsymbol{G}_0, \boldsymbol{G}_1, \ldots, \boldsymbol{G}_T$, with each $\boldsymbol{G}_t \sim \mathbb{G}(n, q)$, $\boldsymbol{G}_0, \boldsymbol{G}_T$ are independent, and each pair $\boldsymbol{G}_t, \boldsymbol{G}_{t+1}$ is marginally $(1-\frac{1}{T})$-correlated (that is, $\boldsymbol{G}_t, \boldsymbol{G}_{t+1}$ can be coupled with a pair $\boldsymbol{G}, \boldsymbol{G}'$ with $\boldsymbol{G} \sim \mathbb{G}(n,q)$ and $\boldsymbol{G}'$ obtained by resampling each edge with probability $\frac{1}{T}$). First, sample $\boldsymbol{G}_0, \boldsymbol{G}_T$ independently. Additionally, for each $(i,j) \in \binom{[n]}{2}$, sample an independent random variable $\boldsymbol{U}_{ij} \sim \mathrm{Unif}([0,1])$. Now, for each t, let $\boldsymbol{G}_t(i,j) = \boldsymbol{G}_0(i,j) \cdot \mathbf{1}[\boldsymbol{U}_{ij} > \frac{t}{T}] + \boldsymbol{G}_T(i,j) \cdot \mathbf{1}[\boldsymbol{U}_{ij} \leqslant \frac{t}{T}]$. Clearly, $\boldsymbol{G}_t \sim \mathbb{G}(n,q)$. Also, edge (i,j) is resampled going from $\boldsymbol{G}_t$ to $\boldsymbol{G}_{t+1}$ if and only if $\boldsymbol{U}_{ij} \in (\frac{t}{T}, \frac{t+1}{T}]$, which is true for each edge independently with probability $\frac{1}{T}$. Thus, $\boldsymbol{G}_t, \boldsymbol{G}_{t+1}$ are marginally a $1-\frac{1}{T}$-correlated pair.

Let $\boldsymbol{p}_t = \mathcal{A}(\boldsymbol{G}_t)$, and suppose $\mathcal{A}$ fails with probability δ. From Theorem 14, we have that with probability at least $1 - O(\frac{T \log\log n}{\log n}) \geqslant \frac{1}{2}$, each $\boldsymbol{p}_t$ must have overlap with $\boldsymbol{p}_0$ which is either equal to 1, or at most $C\varepsilon$, and further the overlap of $\boldsymbol{p}_0$ and $\boldsymbol{p}_T$ must be at most $C\varepsilon$ (as they are independent). Hence, with high probability there must exist some $t \in [T]$ with $\frac{|\boldsymbol{p}_0 \cap \boldsymbol{p}_t|}{\sqrt{|\boldsymbol{p}_0||\boldsymbol{p}_t|}} = 1$ but $\frac{|\boldsymbol{p}_0 \cap \boldsymbol{p}_{t+1}|}{\sqrt{|\boldsymbol{p}_0||\boldsymbol{p}_{t+1}|}} \leqslant C\varepsilon$, implying that $\frac{|\boldsymbol{p}_t \cap \boldsymbol{p}_{t+1}|}{\sqrt{|\boldsymbol{p}_t||\boldsymbol{p}_{t+1}|}} \leqslant C\varepsilon$. This is a contradiction, unless $\mathcal{A}$ did not succeed on some input, so the success probability on all inputs cannot exceed $\frac{1}{2}$. By a union bound $\mathcal{A}$ must have been successful on all $T+1$ inputs with probability at least $1-\delta(T+1)$, so it must be the case that $\delta > \frac{1}{2(T+1)} \geqslant \frac{1-\rho}{6}$. ■

A.3. Low-degree polynomial estimators

As discussed above, the shortest s,t-path problem can be solved, even exactly, in polynomial time. The overlap-gap property means that it cannot be solved by any sufficiently smooth algorithm. In particular, in Gamarnik et al. (2024); Wein (2022) it is shown that polynomials of degree $O(1)$ are smooth with probability $\Omega(1/\mathrm{poly}(n))$ when the input is an adjacency matrix of a graph sampled from $\mathbb{G}(n,q)$.[5] Here, we will apply an invariance principle to conclude that symmetric degree-$O(1)$ polynomials are smooth with probability $\Omega(1)$, which will allow us to prove a lower bound against degree-$O(1)$ polynomials.

But while degree $O(1)$ polynomials are smooth, degree $\omega(1)$ polynomials are not necessarily smooth. In Appendix A.3.1 we will show that the OGP indeed implies that degree-$O(1)$ polynomials cannot estimate s,t-paths, and in Appendix A.3.2 we'll show that polynomials of degree $\Omega(\log n/\log\log n)$ can estimate the shortest s,t-path problem very well.

We identify each path p from 1 to 2 by its indicator vector in $\{0,1\}^{\binom{[n]}{2}}$.

Definition 16 *We say that $f : \{0,1\}^{\binom{[n]}{2}} \to \mathbb{R}^{\binom{n}{2}}$ is an (α,β)-approximation of the $(1+\varepsilon)$-approximate shortest path between $1,2$ in $\boldsymbol{G} \sim \mathbb{G}(n,q)$ if $\mathbf{E}[\|f\|^2] = \mathrm{OPT}$,[6] and*

$$\mathbf{Pr}\left[\exists p \in \mathcal{P}_\varepsilon(\boldsymbol{G}) \text{ s.t. } \frac{\langle f,p\rangle}{\sqrt{\mathbf{E}[\|f\|^2]\cdot\|p\|}} \geqslant \alpha, \quad \textit{and} \quad \alpha \leqslant \frac{\|f\|^2}{\mathbf{E}[\|f\|^2]} \leqslant \frac{1}{\alpha}\right] \geqslant \beta.$$

A.3.1. Lower bound for constant degree polynomials

Degree-$O(1)$ polynomial estimators are smooth algorithms. This has already been established in the literature, but extant results such as Gamarnik et al. (2024); Wein (2022) establish that stability

[5] They also observe that this is the case for e.g. polynomials in Normal-valued variables, where the fact is just a consequence of hypercontractivity. In sparse random graphs, matters are a bit more complex.

[6] This is just a normalization convention; the parameters are easily adjusted for rescaling by $(1+\varepsilon)$.

holds with probability at least $1/\mathrm{poly}(n)$; here, because our overlap-gap property holds only with probability $1 - \frac{1}{\mathrm{polylog}\, n}$, we require a new statement which takes advantage of the fact that an estimator f is, without loss of generality, symmetric under vertex-relabelings.

Lemma 17 *Suppose $f : \{0,1\}^{\binom{[n]}{2}} \to \mathbb{R}^{\binom{[n]}{2}}$ is an approximation to the shortest $1,2$-path. For each $\pi \in S_n$, define $f_\pi(G) = \pi^{-1} f(\pi(G))$, and let $f^{\mathrm{sym}}(G) = \mathbf{E}_{\pi \sim \mathrm{Unif}(S_{[n]\setminus\{1,2\}})} \pi^{-1} f(\pi(G))$. Then for any coupling of graphs $\boldsymbol{G} \sim \mathbb{G}(n,q)$ and $s-t$ paths $\boldsymbol{p}$ in $\boldsymbol{G}$,*

$$\frac{\mathbf{E}\langle f(\boldsymbol{G}), \boldsymbol{p}\rangle}{\sqrt{\mathbf{E}\,\|f(\boldsymbol{G})\|^2}} \leqslant \frac{\mathbf{E}\langle f^{\mathrm{sym}}(\boldsymbol{G}), \boldsymbol{p}\rangle}{\sqrt{\mathbf{E}\,\|f^{\mathrm{sym}}(\boldsymbol{G})\|^2}}.$$

Proof The numerators are the same, because $\boldsymbol{G}$ and its $1-2$ paths have a distribution which is symmetric under permutations of $[n]$ which fix $1, 2$, so

$$\mathop{\mathbf{E}}_{\boldsymbol{G},\boldsymbol{p}} \langle f(\boldsymbol{G}), \boldsymbol{p}\rangle = \mathop{\mathbf{E}}_{\boldsymbol{G},\boldsymbol{p}} \mathop{\mathbf{E}}_{\pi} \langle f(\pi(\boldsymbol{G})), \pi(\boldsymbol{p})\rangle = \mathop{\mathbf{E}}_{\boldsymbol{G},\boldsymbol{p}} \langle f^{\mathrm{sym}}(\boldsymbol{G}), \boldsymbol{p}\rangle.$$

The denominator on the right-hand side is only smaller by Jensen's inequality. ■

If f is a $(1-\eta, 1-\delta)$-approximation of the $(1+\varepsilon)$ shortest s,t-path for $\eta, \delta, \varepsilon$ sufficiently small, then by an averaging argument $\langle f(\boldsymbol{G}), \boldsymbol{p}\rangle$ will come close to saturating the Cauchy-Schwarz inequality with good probability. In this case, Theorem 17 implies that $f^{\mathrm{sym}}(G) = \mathbf{E}_\pi\, \pi^{-1}(f(\pi(G)))$ is at least as close to saturating Cauchy-Schwarz in expectation, from which another averaging argument implies that f^{sym} is also a $(1-\eta', 1-\delta')$-approximation to the s,t-shortest path, for η', δ' going to zero with η, δ. Since we will only be concerned with the small η, δ regime, from now on without loss of generality we consider only symmetric f, as ruling out f^{sym} suffices to rule out any f.

Symmetric functions are more stable than asymmetric functions, which allows us to give better quantitative guarantees than Gamarnik et al. (2024); Wein (2022). We will prove the following smoothness result in Appendix C.

Theorem 18 *For $\rho \geqslant 1 - 1/T$, $\gamma \geqslant 3D(6e)^D/T$, and f a degree-D polynomial which is fixed by the action of $S_{[n]\setminus\{1,2\}}$, for ρ-correlated graphs $\boldsymbol{G}, \boldsymbol{G}' \sim \mathbb{G}(n,q)$, we have that*

$$\mathbf{Pr}\left[\|f(\boldsymbol{G}) - f(\boldsymbol{G}')\|_2^2 \geqslant \gamma\, \mathbf{E}[\|f(\boldsymbol{G})\|^2]\right] \leqslant \exp\left(-\frac{D}{3e}\left(\frac{\gamma T}{3D}\right)^{1/D}\right) + o_n(1).$$

Lower bound from overlap-gap property

Theorem 19 *For any integer $D \geqslant 0$ and $0 < \eta < 0.1$, there is a small enough constant $\delta > 0$ such that there is no degree-D polynomial which is a $(1-\eta, 1-\delta)$-approximation of the shortest path between $1, 2$ in $\mathbb{G}(n,q)$.*

Proof We give a proof by contradiction. Suppose there exists a degree-D polynomial which is a $(1-\eta, 1-\delta)$-approximation of the shortest path between $1, 2$ in $\mathbb{G}(n,q)$. Choose $T = \lceil \frac{1}{1-\rho} \rceil = O(1)$, and note $\rho \geqslant 1 - \frac{1}{T}$. We sample a sequence of graphs $\boldsymbol{G}_0, \boldsymbol{G}_1, \ldots, \boldsymbol{G}_T$ as in the proof of Theorem 15, where each $\boldsymbol{G}_t \sim \mathbb{G}(n,q)$, $\boldsymbol{G}_0, \boldsymbol{G}_T$ are independent, and each pair $\boldsymbol{G}_t, \boldsymbol{G}_{t+1}$ is marginally $(1-\frac{1}{T})$-correlated. For shorthand, call $L = \frac{\log n}{\log\log n}$ to be the typical approximate length of the shortest path in $\boldsymbol{G}$. We define four events:

1. Let $\mathcal{E}_{\text{stable}}$ be the event that $\|f(\boldsymbol{G}_t) - f(\boldsymbol{G}_{t+1})\| \leqslant s\sqrt{\mathbf{E}[\|f(\boldsymbol{G})\|^2]}$ for all $0 \leqslant t \leqslant T-1$.
2. Let $\mathcal{E}_{\text{alg}}$ be the event that f is a $(1-\eta, 1-\delta)$-approximation of a path in $\mathcal{P}_\varepsilon(\boldsymbol{G}_t)$ for all $0 \leqslant t \leqslant T$.
3. Let $\mathcal{E}_{\text{gap}}$ be the event that for all $t \leqslant T-1$ and $(p_t, p_{t+1}) \in \mathcal{P}_\varepsilon(\boldsymbol{G}_t) \times \mathcal{P}_\varepsilon(\boldsymbol{G}_{t+1})$, $\frac{1}{L}\|p_t - p_{t+1}\|^2 \notin (0, 2-2C\varepsilon)$, and that for all $(p_0, p_T) \in \mathcal{P}_\varepsilon(\boldsymbol{G}_0) \times \mathcal{P}_\varepsilon(\boldsymbol{G}_T)$, $\frac{1}{L}\|p_0 - p_T\|^2 \geqslant 2-2C\varepsilon$.
4. Let $\mathcal{E}_{\text{paths}}$ be the event that for all $t \leqslant T$, $\mathcal{P}_\varepsilon(\boldsymbol{G}_t) \neq \emptyset$ and contains only paths of size in $[1, 1+\varepsilon]L$.

We argue that these events cannot happen simultaneously for small enough constant s and ε. Indeed, assume $\mathcal{E}_{\text{paths}}$ occurs, so that each set of paths is non-empty and the paths have the expected size; for shorthand, assume as well that $\mathcal{E}_{\text{alg}}$ occurs, so that with each output $f(\boldsymbol{G}_t)$ we may associate a path p_t with $\|f(\boldsymbol{G}_t) - p_t\|^2 \leqslant \|f(\boldsymbol{G}_t)\|^2 + \|p_t\|^2 - 2(1-\eta)\sqrt{\mathbf{E}\,\|f(\boldsymbol{G}_t)\|^2}\|p_t\| \leqslant 4\eta L$. Assume as well that $\mathcal{E}_{\text{gap}}$ occurs, so that $\frac{1}{L}\|p_0 - p_T\|^2 \geqslant 2-2C\varepsilon$, and for all t, $\frac{1}{L}\|p_t - p_{t+1}\|^2 \notin (0, 2-2C\varepsilon)$. This implies there must be a first time $t^* \in [T]$ where $\frac{1}{L}\|p_0 - p_{t^*}\|^2 \geqslant 2-2C\varepsilon$. But now if $\mathcal{E}_{\text{stable}}$ occurs, by triangle inequality

$$\begin{aligned}
\|p_0 - p_{t^*}\| &\leqslant \|p_0 - p_{t^*-1}\| + \|p_{t^*-1} - p_{t^*}\| \\
&= \|p_{t^*-1} - p_{t^*}\| \\
&\leqslant \|p_{t^*-1} - f(\boldsymbol{G}_{t^*-1})\| + \|f(\boldsymbol{G}_{t^*-1}) - f(\boldsymbol{G}_{t^*})\| + \|f(\boldsymbol{G}_{t^*}) - p_{t^*}\| \\
&\leqslant 4\sqrt{\eta}\sqrt{L} + O(s\sqrt{L}) < \sqrt{(2-2C\varepsilon)L},
\end{aligned}$$

for small constant s and ε, which is a contradiction.

Now, we lower-bound the probability of the events one-by-one. By construction, each $\boldsymbol{G}_t \sim \mathbb{G}(n,q)$ and each pair $\boldsymbol{G}_t, \boldsymbol{G}_{t+1}$ is marginally $(1-\frac{1}{T})$-correlated. By Theorem 18 and a union bound over pairs $\boldsymbol{G}_t, \boldsymbol{G}_{t+1}$, we have $\mathbf{Pr}[\overline{\mathcal{E}_{\text{stable}}}] \leqslant T\exp(-\frac{D}{3e}(\frac{s^2T}{3D})^{1/D}) + o_n(1)$. The failure probability of the algorithm is at most δ by assumption, so by a union bound $\mathbf{Pr}[\overline{\mathcal{E}_{\text{alg}}}] \leqslant (T+1)\delta$. Finally, since $\boldsymbol{G}_0$ and $\boldsymbol{G}_T$ are independent, from Theorem 14 and Theorem 13 we have that $\mathbf{Pr}[\overline{\mathcal{E}_{\text{gap}}}] + \mathbf{Pr}[\overline{\mathcal{E}_{\text{paths}}}] \leqslant (T+1)\cdot O\left(\frac{\log\log n}{\log n}\right)$. Thus, it must be the case that

$$\begin{aligned}
0 &\geqslant \mathbf{Pr}[\mathcal{E}_{\text{stable}} \cap \mathcal{E}_{\text{alg}} \cap \mathcal{E}_{\text{gap}} \cap \mathcal{E}_{\text{paths}}] \geqslant 1 - \mathbf{Pr}[\overline{\mathcal{E}_{\text{stable}}}] - \mathbf{Pr}[\overline{\mathcal{E}_{\text{alg}}}] - \mathbf{Pr}[\overline{\mathcal{E}_{\text{gap}}}] - \mathbf{Pr}[\overline{\mathcal{E}_{\text{paths}}}] \\
&\geqslant 1 - T\exp\left(-\frac{D}{3e}\left(\frac{s^2T}{3D}\right)^{1/D}\right) + o_n(1) - (T+1)\cdot\delta - T\cdot O\left(\frac{\log\log n}{\log n}\right),
\end{aligned}$$

the positivity of the right-hand-side for $\delta = \frac{1}{100(T+1)}$ and large enough T yields our contradiction. ■

A.3.2. A POLYNOMIAL ESTIMATOR OF LOGARITHMIC DEGREE

Here we will argue that the shortest s,t-path may be computed with reasonable success probability by degree $o(\log n)$ polynomials. This estimator does not, strictly speaking, fulfill Theorem 16, because of an issue of breaking symmetry. However,it gives almost full information about whether an edge participates in a short s,t-path.[7]

[7] This raises another question about the usual OGP-based low-degree polynomial lower bounds: they may be brittle to the specific definition of a low-degree polynomial estimator.

Lemma 20 *There exists a degree-$O(\frac{\log n}{\log\log n})$ polynomial in the entries of the adjacency matrix of $\boldsymbol{G}$ which achieves correlation $1 - o(1)$ with the indicator that (i,j) participates a path of length $m = \lceil \frac{\log n}{\log nq} \rceil + 1$ in $\boldsymbol{G}$. Furthermore, with high probability the polynomial is computable in polynomial time.*

Proof Define $m = \lceil \frac{\log n}{\log nq} \rceil + 1 = \frac{\log n}{\log nq} + 1 + \xi$. By Theorem 13, the length of the shortest path in $\boldsymbol{G}$ will be close to m, and with $1 - o(1)$ probability there will be a 1-2 path of length m. Define $\boldsymbol{p}_{ij}$ to be the indicator that (i,j) participates in an 1-2 path of length m. Define the polynomial

$$f_{ij}(\boldsymbol{G}) = \sum_{\substack{p \in \mathcal{P}_{(m)} \\ (i,j) \in p}} \prod_{(a,b) \in p} \mathbf{1}[(a,b) \in E(\boldsymbol{G})],$$

where we recall that $\mathcal{P}_{(m)}$ is the set of all m-edge 1-2 paths in K_n.

We now verify that $f_{ij}(\boldsymbol{G})$ correlates well with $\boldsymbol{p}_{ij}$, in the sense that

$$\frac{\mathbf{E}[f_{ij}(\boldsymbol{G})\boldsymbol{p}_{ij}]}{\sqrt{\mathbf{E}[f_{ij}(\boldsymbol{G})^2]\,\mathbf{E}[\boldsymbol{p}_{ij}^2]}} = 1 - o(1). \tag{2}$$

We must consider separately the case when $\{i,j\} \cap \{1,2\} = \emptyset$ from the case when $|\{i,j\} \cap \{1,2\}| = 1$; for brevity's sake we consider only the latter case, as the former case is similar (and actually simpler).

To compute the correlation, we use that $f_{ij}(\boldsymbol{G})$ takes on non-negative integer values, and is zero only if (i,j) participates in no 1-2 paths of length m. Therefore

$$\mathbf{E}[f_{ij}(\boldsymbol{G})\boldsymbol{p}_{ij}] = \mathbf{E}[f_{ij}] = 2(m-2)n^{\underline{m-3}} \cdot q^m,$$

as there are $2(m-2)n^{\underline{m-3}}$ paths from 1 to 2 that involve the edge i,j (choose the location and orientation of (i,j), then the remaining vertex labels). By our choice of m, this quantity is $O(\frac{\log^{2+\xi} n}{n^2})$.

To compute the $\mathbf{E}[\boldsymbol{p}_{ij}^2] = \mathbf{E}[\boldsymbol{p}_{ij}] = \mathbf{Pr}[(i,j) \text{ in an } m-\text{path}]$, we apply the Janson inequality (see e.g. Chapter 8 of Alon and Spencer (2016)):

Fact 21 (Corollary of Janson Inequalities) *If $\{H_\alpha\}_\alpha$ is a set of edge-induced subgraphs of K_n, the probability that none of the subgraphs appear in $\mathbb{G}(n,q)$ is at least $\prod_\alpha (1 - q^{|E(H_\alpha)|})$.*

This implies that

$$\mathbf{E}[\boldsymbol{p}_{ij}] = 1 - \mathbf{Pr}[(i,j) \text{ not in } m-\text{path}] \leqslant 1 - (1-q^m)^{2(m-2)n^{\underline{m-3}}} \leqslant (1+o(1)) \cdot 2(m-2)n^{\underline{m-3}} q^m,$$

where we have used that $q^{2m} \ll 1$.

Finally, we compute the second moment of f_{ij}. For this, we will need to count the number of pairs of paths from $1,2$ that intersect on the edge (i,j).

Claim A.2 *Let $W_{k,m}$ be the number of pairs of paths from 1 to 2 that overlap on a total of k edges, one of which is the edge (i,j). Then*

$$W_{k,m} \leqslant \begin{cases} 4(m-2)^2 n^{2m-6} & k = 1, \\ k(k+1)n^{2m-k-4} + O\left(k^4 m^6\right) n^{2m-k-5} & k > 1. \end{cases}$$

Proof Let c be the number of components in $p \cap p'$. As $k < m$, there must be at least two components.

In the case $c = 2$, only $k > 1$ is possible, since we have assumed $\{i, j\}$ does not intersect with $\{1, 2\}$. So one chooses $k+1$ possibilities for the size of the component containing 1, k possibilities for the edge (i, j), and $2m - k - 4$ vertex labels, for a total of $k(k+1)n^{2m-k-4}$ choices.

In the case $c = 3$, $k = 1$, there are $4(m-2)^2 n^{2m-6}$ pairs: choose the location of i, j in each path, the orientation of i, j in each path, and then all the remaining vertex labels.

Finally, if $c \geqslant 3$ we argue just as in Claim A.1, except that we choose one of the k edges to be (i, j), choose its 2 orientations, and then lose $n(n-1)$ choices of vertex labels since i, j are fixed. Hence summing over all $c \geqslant 3$, in this case there are at most $O(k^4 m^6 n^{2m-k-5})$ paths. ■

Thus by Claim A.1,

$$
\begin{aligned}
&\mathbf{E}[f_{ij}(\boldsymbol{G})^2] = \sum_{(i,j)\in p, p' \in \mathcal{P}_{(m)}} q^{2m-|p\cap p'|} \\
&\leqslant 2(m-2)n^{m-3}q^m + 4(m-2)^2 n^{2m-6} q^{2m-1} + \sum_{k=2}^{m-1} W_{k,m} q^{2m-k} \\
&\leqslant 2(m-2)n^{m-3}q^m + 4(m-2)^2 n^{2m-6} q^{2m-1} + \sum_{k=2}^{m-1} \left(k(k+1)n^{2m-k-4} + O(k^4 m^6) n^{2m-k-5} \right) q^{2m-k} \\
&\leqslant 2(m-2)n^{m-3}q^m + 4(m-2)^2 n^{2m-6} q^{2m-1} + n^{m-3}q^m \sum_{k=2}^{m-1} \frac{(nq)^m}{n \cdot (nq)^k} \left(k(k+1) + O\left(\frac{k^4 m^6}{n} \right) \right) \\
&\leqslant 2(m-2)n^{m-3}q^m + 4(m-2)^2 n^{2m-6} q^{2m-1} + n^{m-3}q^m \sum_{k=2}^{m-1} \frac{(C\log n)^\xi}{n \cdot (C \log n)^k} \left(k(k+1) + O\left(\frac{k^4 m^6}{n} \right) \right) \\
&\leqslant (2(m-2) + o(1)) \cdot n^{m-3} q^m.
\end{aligned}
$$

Putting these together, the correlation Equation (2) is $\geqslant 1 - o(1)$, as desired.

Clearly, f_{ij} is a polynomial of degree $m = o(\log n)$ in the entries of the adjacency matrix of $\boldsymbol{G}$. The polynomial can be computed approximately in polynomial time, either approximately computed using color-coding Alon et al. (1995); Arvind and Raman (2002), or, alternatively, by breadth-first search in G, on the event that G is sparse (more discussion in the proof of Theorem 23). ■

A.4. Sampling despite disorder chaos

Disorder chaos is a stronger version of the overlap-gap property: it is a more quantitative overlap-gap statement about samples from the distribution over good solutions. In our context, we would say that the shortest-paths problem exhibits disorder chaos if with high probability over ρ-correlated $\boldsymbol{G}, \boldsymbol{G}' \sim \boldsymbol{G}(n, q)$,

$$
\lim_{\rho \to 1} \mathop{\mathbf{E}}_{\boldsymbol{p}, \boldsymbol{p}' \sim \mathrm{Unif}(\mathcal{P}_\varepsilon(\boldsymbol{G})) \otimes \mathrm{Unif}(\mathcal{P}_\varepsilon(\boldsymbol{G}'))} \left[\left(\frac{\langle \boldsymbol{p}, \boldsymbol{p}' \rangle}{\ell^*} \right)^2 \right] = o_n(1),
$$

Where $\ell^* = \frac{\log n}{\log nq}$ is the length of a typical path. One could instead consider a different measure over $\mathcal{P}_\varepsilon(\boldsymbol{G})$, but we take the uniform measure here for simplicity.

Disorder chaos sometimes implies lower bounds against stable algorithms for sampling from $\mathrm{Unif}(\mathcal{P}_\varepsilon(\boldsymbol{G}))$ (see El Alaoui et al. (2022)). The argument goes by showing that, under certain conditions, disorder chaos implies a lower bound on the Wasserstein distance between $\mathrm{Unif}(\mathcal{P}_\varepsilon(\boldsymbol{G}))$ and $\mathrm{Unif}(\mathcal{P}_\varepsilon(\boldsymbol{G}'))$. Let $\Pi(\boldsymbol{G}, \boldsymbol{G}')$ be the space of couplings of $\mathrm{Unif}(\mathcal{P}_\varepsilon(\boldsymbol{G}))$ and $\mathrm{Unif}(\mathcal{P}_\varepsilon(\boldsymbol{G}'))$. If one can show that with high probability,

$$\lim_{\rho \to 1} \inf_{\pi \in \Pi(\boldsymbol{G}, \boldsymbol{G}')} \mathop{\mathbf{E}}_{(\boldsymbol{p}, \boldsymbol{p}') \sim \pi} \left[\frac{1}{\ell^*} \|\boldsymbol{p} - \boldsymbol{p}'\|^2 \right] > c - o_n(1) \tag{3}$$

for some universal constant $c > 0$, then a concise argument[8] of El Alaoui et al. (2022) yields lower bounds against smooth algorithms for sampling from $\mathrm{Unif}(\mathcal{P}_\varepsilon(\boldsymbol{G}))$ in the Wasserstein distance.

Here, we will establish that shortest paths in $\mathbb{G}(n, q)$ have the Wasserstein distance bound as in Equation (3), implying that stable algorithms cannot sample. Furthermore, we will show that there is an efficient, non-stable algorithm that, with high probability over $\boldsymbol{G} \sim \mathbb{G}(n, q)$ succeeds in producing exact samples from $\mathrm{Unif}(\mathcal{P}_\varepsilon(\boldsymbol{G}))$.

A lower bound on the Wasserstein distance. Here we establish that Equation (3) holds for shortest path in $\mathbb{G}(n, q)$, implying that stable algorithms cannot sample from $\mathrm{Unif}(\mathcal{P}_\varepsilon(\boldsymbol{G}))$.

Lemma 22 *Suppose $\rho < 1$, and $\boldsymbol{G}, \boldsymbol{G}' \sim \mathbb{G}(n, q)$ are ρ-correlated, with $q = \frac{C \log n}{n}$ and $C \geqslant 1$. Let $\Pi(\boldsymbol{G}, \boldsymbol{G}')$ be the set of all couplings of $\mathrm{Unif}(\mathcal{P}_\varepsilon(\boldsymbol{G})), \mathrm{Unif}(\mathcal{P}_\varepsilon(\boldsymbol{G}'))$. Then for all $\delta > 0$, there exists n sufficiently large so that with probability $1 - o_n(1)$,*

$$\inf_{\pi \in \Pi(\boldsymbol{G}, \boldsymbol{G}')} \mathop{\mathbf{E}}_{(\boldsymbol{p}, \boldsymbol{p}') \sim \pi} \left[\frac{1}{\ell^*} \|\boldsymbol{p} - \boldsymbol{p}'\|^2 \right] \geqslant 1 - \delta.$$

Proof Suppose p is a 1-2 path of length $L \in [1, 1+\varepsilon]\ell^*$, where $\ell^* = \frac{\log n}{\log nq}$. We will argue that conditioned on $p \in E(\boldsymbol{G})$, with high probability, there is no $\boldsymbol{p}' \in \mathcal{P}_\varepsilon(\boldsymbol{G}')$ for which $\langle p, \boldsymbol{p}' \rangle \geqslant \frac{1}{2}\ell^*$. This is enough to establish Equation (3) with $c = 1$, as for all but a vanishing measure of $\boldsymbol{p} \sim \mathrm{Unif}(\mathcal{P}_\varepsilon(\boldsymbol{G}))$, there is no $\boldsymbol{p}' \in \mathcal{P}_\varepsilon(\boldsymbol{G}')$ to which they could be transported with cost < 1.

The argument is a first moment argument similar to the one used to establish the OGP. Conditioned on $p \in E(\boldsymbol{G})$, for any given 1-2 path p' of length $L' \in [1, 1+\varepsilon]\ell^*$ which overlaps with p in k edges,

$$\mathbf{Pr}[p' \in \boldsymbol{G}' \mid p \in \boldsymbol{G}] = (\rho + (1-\rho)q)^k q^{L'-k}.$$

[8] The triangle inequality.

By Claim A.1, the number of such paths p' is at most $\frac{1}{n^{L-1}} M_{k,L-k,L'-k} \leqslant (1+o_n(1))n^{L'-k}\left(\frac{k+1}{n^2}+\frac{\log^{15} n}{n^3}\right)$ when $k < \min(L, L')$, and at most 1 when $k = L = L'$. So, summing from $k = \frac{1}{2}\ell^*, \dots, L'$,

$$\begin{aligned}
&\mathbf{E}[\#p' \in \mathcal{P}_\varepsilon(\boldsymbol{G}),\, \langle p', p\rangle \geqslant \frac{1}{2}\ell^*] \\
&\leqslant (\rho+(1-\rho)q)^L + \sum_{k=\frac{1}{2}\ell^*}^{L'-1} (\rho+(1-\rho)q)^k q^{L'-k}(1+o_n(1))n^{L'-k}\left(\frac{k+1}{n^2}+\frac{\log^{15} n}{n^3}\right) \\
&\leqslant o_n(1) + O\left(\frac{(nq)^{L'}\log n}{n^2}\right)\sum_{k=\frac{1}{2}\ell^*}^{L'}\left(\frac{(1+o_n(1))\rho}{nq}\right)^k \\
&= o_n(1) + O\left(\frac{\log n}{n^{1-\varepsilon}}\right).
\end{aligned}$$

Thus, by Markov's inequality, with high probability p will have no $\boldsymbol{p}' \in \mathcal{P}_\varepsilon(\boldsymbol{G}')$ which overlaps by $\frac{1}{2}\ell^*$ or more, as desired. ■

Polynomial-time algorithm for sampling.

Lemma 23 *There is a polynomial-time algorithm which with high probability succeeds in sampling from* $\mathrm{Unif}(\mathcal{P}_\varepsilon(\boldsymbol{G}))$ *when* $\boldsymbol{G} \sim \mathbb{G}(n,q)$ *for* $q \geqslant \frac{\log n}{n}$.

Remark 24 *We remark that the proof of Theorem 23 can easily be extended to efficiently sample from any distribution on* $\mathcal{P}_\varepsilon(\boldsymbol{G})$ *where the weight of each path is polynomial-time computable from its description.*

Proof [Proof of Theorem 23] From Theorem 13, the length of the shortest path concentrates around $\ell^* = \frac{\log n}{\log nq}$ with high probability. So the algorithm is as follows: one simply enumerates all walks of length $L = (1+\varepsilon)\ell^*$ starting at vertex 1, and then deletes from the list any walk that either is not simple, or did not encounter vertex 2. Finally, one of these paths is chosen uniformly at random (or, if one wishes, one can add weights which depend on the path length, as discussed in Theorem 24).

This can be implemented with a stack or a queue, in a manner similar to depth-first search or breadth-first search, except that one must keep track of the length of the path explored so far at every vertex, and the stopping criterion is based on the length of the path rather than on whether vertices have already been explored. If $\boldsymbol{d}_{\max}$ is the maximum degree of a vertex in $\boldsymbol{G}$, then this algorithm comprises of at most $\boldsymbol{d}_{\max}^L$ operations, each requiring the recording of $O(\log n)$ bits of information. And in the end, there are at most $\boldsymbol{d}_{\max}^L$ paths to work with. So provided that $\boldsymbol{d}_{\max}^L = \mathrm{poly}(n)$, the algorithm is polynomial time.

We now argue that this holds with high probability. Indeed, the degree $\boldsymbol{d}_i$ of vertex i is distributed as $\mathrm{Bin}(n-1,q)$, and so applying Markov's inequality with the moment generating function,

$$\mathbf{Pr}[\boldsymbol{d}_i > 3q(n-1)] \leqslant \frac{\mathbf{E}\, e^{\boldsymbol{d}_i}}{e^{3q(n-1)}} = \frac{(1-q+qe)^{n-1}}{e^{3q(n-1)}} \leqslant \frac{e^{(e-1)q(n-1)}}{e^{3q(n-1)}} \leqslant e^{(e-4)q(n-1)} \ll \frac{1}{n^{1.1}},$$

as $q \geqslant \frac{\log n}{n}$. Hence by a union bound, with high probability the maximum degree is no larger than $3nq$. Then with high probability,

$$\boldsymbol{d}_{\max}^L \leqslant (3nq)^L = (3nq)^{(1+\varepsilon)\frac{\log n}{\log nq}} \leqslant n^{1+\varepsilon+o(1)},$$

and we have our conclusion. ■

Appendix B. First passage percolation on the complete graph

In this section, we establish an overlap-gap property for shortest paths in the complete graph with random non-negative edge weights. This problem, too, is solved in polynomial time by the shortest path algorithm.

Our setting will be as follows: Sample a vector $\boldsymbol{\ell} \in \mathbb{R}_+^{\binom{[n]}{2}}$ with i.i.d. $\mathrm{Exp}(1)$ entries. We will use $\boldsymbol{\ell}$ as an assignment of edge lengths to the edges of K_n, and denote the resulting randomly weighted graph by $K_n^{\boldsymbol{\ell}}$. For each path p in K_n, we identify p with the vector in $\{0,1\}^{\binom{[n]}{2}}$ whose entries indicate membership in p. We define the length of p in $K_n^{\boldsymbol{\ell}}$ to be $\mathrm{len}_{\boldsymbol{\ell}}(p) = \langle p, \boldsymbol{\ell} \rangle$, and we define the *hopcount* of p to be $\mathrm{hop}(p) = \|p\|^2$.

In this case our object of study will be minimum-length paths between vertices 1 and 2 in $K_n^{\boldsymbol{\ell}}$ (note that in $\mathbb{G}(n,q)$ the hopcount and length coincide). We will begin by listing (and deriving) some properties of the approximate min-length paths, describe a natural distribution over correlated instances, derive the overlap-gap property and lower bounds against smooth algorithms, and finally comment on the existence of efficient algorithms for sampling from $\mathcal{P}_\varepsilon(\boldsymbol{\ell})$ and the possibility of a low-degree polynomial estimator.

B.1. Characterizing approximate shortest paths in weighted complete graphs

The properties of *the* minimum-length path are studied under the name "first passage percolation" on the complete graph. The behavior of the length and hopcount of the shortest path is well-understood Bollobás* et al. (2004), as summarized in the following theorem (Theorem 1.1 and 1.2 in Bhamidi and van der Hofstad (2012)):

Theorem *If $\boldsymbol{p}$ is the shortest path between $1,2$ in $K_n^{\boldsymbol{\ell}}$, then*

$$\left(n \cdot \mathrm{len}_{\boldsymbol{\ell}}(\boldsymbol{p}) - \log n, \frac{\mathrm{hop}(\boldsymbol{p}) - \log n}{\sqrt{\log n}}\right) \xrightarrow{d} (\boldsymbol{X}, \boldsymbol{Z}),$$

for independent random variables $\boldsymbol{Z}, \boldsymbol{X}$.[9]

Here, we are interested in the set of ε-approximate shortest paths,

$$\mathcal{P}_\varepsilon(\boldsymbol{\ell}) := \left\{ p \mid \mathrm{len}_{\boldsymbol{\ell}}(p) \leqslant (1+\varepsilon)\tfrac{\log n}{n} \right\}.$$

For paths p, p', we define their overlap in terms of the hopcount overlap $\frac{\langle p, p' \rangle}{\|p\| \cdot \|p'\|}$. To demonstrate the overlap-gap property for $\mathcal{P}_\varepsilon(\boldsymbol{\ell})$, it will be helpful to have a bound on the hopcount of an approximate shortest path.

Lemma 25 *Let $\varepsilon \in (0,1)$ be a sufficiently small constant. If $p \in \mathcal{P}_\varepsilon(\boldsymbol{\ell})$, then with probability $1 - o_n(1)$, $\left|\frac{\mathrm{hop}(p)}{\log n} - 1\right| \leqslant 2\sqrt{\varepsilon}$.*

[9] Further, $\boldsymbol{Z}$ is a standard Normal random variable. The distribution of $\boldsymbol{X}$ is more involved.

Proof Let $\boldsymbol{X}^{(k)}$ be the number of paths of hopcount k and length at most $L = (1+\varepsilon)\frac{\log n}{n}$ between 1 and 2. The number of paths p of hopcount k between $1, 2$ is precisely $(n-2)^{\underline{k-1}}$, and the length $\boldsymbol{X}_p$ of each such path is distributed as the sum of k independent $\mathrm{Exp}(1)$ random variables, that is, $\boldsymbol{X}_p$ is $\mathrm{Gamma}(k,1)$-distributed. Hence, integrating the density function directly,

$$\mathbf{Pr}[\boldsymbol{X}_p \leqslant L] = \int_0^L \frac{e^{-a}a^{k-1}}{(k-1)!}da \leqslant \frac{1}{k!}L^k.$$

Thus, combining the above and applying Stirling's formula,

$$\mathbf{E}[\boldsymbol{X}^{(k)}] \leqslant (n-2)^{\underline{k-1}} \cdot \mathbf{Pr}[\boldsymbol{X}_p \leqslant L] \leqslant n^{\underline{k-1}} \frac{1}{k!}\left(\frac{(1+\varepsilon)\log n}{n}\right)^k \leqslant \left(\frac{e(1+\varepsilon)\log n}{k}\right)^{k-1} \cdot \frac{\log n}{n}.$$

In the case that $k > e^2(1+\varepsilon)\log n$, $\mathbf{E}[X^{(k)}] \leqslant \frac{1}{n^2}$, so union bounding over the at most n choices of such k, we have by Markov's inequality that the probability that any k-hop path with $k > e^2(1+\varepsilon)\log n$ is $(1+\varepsilon)$-approximately minimum: length is at most $1/n$.

Henceforth, we may focus on k of order $\log n$. Writing $k = \beta \log n$, we have

$$\log(\mathbf{E}[\boldsymbol{X}^{(k)}]) \leqslant (\beta + \log(1+\varepsilon)\beta - \beta\log\beta - 1)\cdot \log n + \log\log n.$$

This function is concave in β, maximized at $\beta = (1+\varepsilon)$. Hence it suffices to verify that the value is negative at $\beta = (1 - 2\sqrt{\varepsilon})$ and $\beta = (1+2\sqrt{\varepsilon})$, when ε is sufficiently small. Indeed, if $\beta = 1+\delta$ with $|\delta| < 1$, using the series expansion for $\log(1+\delta)$,

$$\begin{aligned} 1+\delta+(1+\delta)\log(1+\varepsilon)-(1+\delta)\log(1+\delta)-1 &\leqslant \delta+\varepsilon+\varepsilon\delta-(1+\delta)(\delta-\frac{\delta^2}{2}+\frac{\delta^3}{3}-\cdots) \\ &= \varepsilon+\varepsilon\delta-\frac{1}{2}\delta^2+O(\delta^3). \end{aligned}$$

For $|\delta| = 2\sqrt{\varepsilon}$ and ε sufficiently small, this is negative. Thus, for $k \leqslant (1-2\sqrt{\varepsilon})\log n$ and $(1+2\sqrt{\varepsilon})\log n \leqslant k \leqslant e^2(1+\varepsilon)\log n$, $\mathbf{E}[\boldsymbol{X}^{(k)}] \leqslant n^{-c}$ for $c > 0$ a constant depending only on ε. Applying Markov's inequality and taking a union bound over all $O(\log n)$ such k completes the proof. ■

B.2. Correlated instances

In order to state our overlap-gap property for correlated instances, we introduce a canonical joint distribution over sets of correlated edge weights $\boldsymbol{\ell}$.

Given an exponential random variable $\boldsymbol{X}_0 \sim \mathrm{Exp}(1)$ and a time $t \in \mathbb{R}_+$, running the following Langevin diffusion process for time t produces an exponential random variable $\boldsymbol{X}_t \sim \mathrm{Exp}(1)$, by taking:

$$d\boldsymbol{X}_s = -\mathbf{1}[\boldsymbol{X}_s \geqslant 0]ds + \sqrt{2}d\boldsymbol{B}_s$$

for $(\boldsymbol{B}_s)_{s\geqslant 0}$ a standard Brownian motion. For a function $f : \mathbb{R}_+ \to \mathbb{R}$, we use the notation

$$P_t f(x) = \mathbf{E}[f(\boldsymbol{X}_t) \mid \boldsymbol{X}_0 = x].$$

We consider correlated weights $\boldsymbol{\ell}^0$, $\boldsymbol{\ell}^t$ given by sampling $\boldsymbol{\ell}^0 \sim (\mathrm{Exp}(1))^{\otimes\binom{[n]}{2}}$, and then applying the time-t Langevin process to each coordinate $e \in \binom{[n]}{2}$ independently to produce $\boldsymbol{\ell}^t$. One can compute that

$$\mathrm{Corr}(\boldsymbol{\ell}_e^0, \boldsymbol{\ell}_e^t) = \frac{\mathbf{E}[(\boldsymbol{\ell}_e^0 - 1)(\boldsymbol{\ell}_e^t - 1)]}{\mathbf{Var}\,\boldsymbol{\ell}_e} = e^{-t}.$$

Lemma 26 *Let $\mathcal{F}, \mathcal{G} \subset \mathbb{R}^{\binom{[n]}{2}}$ be two events. Let $(\boldsymbol{\ell}^s)_{s \geqslant 0}$ be vectors sampled as described above. Then for any $t \geqslant 0$,*

$$\mathbf{Pr}[\mathcal{F}(\boldsymbol{\ell}^0), \mathcal{G}(\boldsymbol{\ell}^t)] \leqslant (1 - e^t)\,\mathbf{Pr}[\mathcal{F}(\boldsymbol{\ell}^0)]\,\mathbf{Pr}[\mathcal{G}(\boldsymbol{\ell}^0)] + e^{-t}\,\mathbf{Pr}[\mathcal{F}(\boldsymbol{\ell}^0), \mathcal{G}(\boldsymbol{\ell}^0)].$$

Proof Let $f(\ell) = \mathbf{1}[\mathcal{F}(\ell)]$ and let $g(\ell) = \mathbf{1}[\mathcal{G}(\ell)]$. As is explained in e.g. Bakry et al. (2014) (see section 2.7), there is a basis of polynomials, the *Laguerre polynomials*, $\{L_k\}_{k\in\mathbb{N}}$, which form an orthonormal basis for functions in L^2 with respect to the exponential measure, and such that for each $e \in \binom{[n]}{2}$,

$$\mathbf{E}[L_k(\boldsymbol{\ell}_e^0) P_t L_m(\boldsymbol{\ell}_e^0)] = \mathbf{E}[L_k(\boldsymbol{\ell}_e^0) L_m(\boldsymbol{\ell}_e^t)] = \mathbf{1}[k = m] e^{-tk}.$$

Moreover the variables corresponding to each edge are independent. Therefore we may expand $f(\ell) = \sum_{\alpha \in \mathbb{N}^{\binom{[n]}{2}}} \hat{f}_\alpha \cdot \prod_{e \in \binom{[n]}{2}} L_{\alpha(e)}(\ell_e)$, noting that $\hat{f}_0 = \mathbf{E}[f(\boldsymbol{\ell})] = \mathbf{Pr}[\mathcal{F}(\boldsymbol{\ell}^0)]$ and $\mathbf{Var}(f) = \sum_{|\alpha| \geqslant 1} \hat{f}_\alpha^2$, and similarly for $g(\ell)$. Then we may compute

$$\begin{aligned}
\mathbf{Pr}[\mathcal{F}(\boldsymbol{\ell}^0), \mathcal{G}(\boldsymbol{\ell}^t)] &= \mathbf{E}[f(\boldsymbol{\ell}^0) g(\boldsymbol{\ell}^t)] \\
&= \sum_{\alpha, \beta \in \mathbb{N}^{\binom{[n]}{[2]}}} \hat{f}_\alpha \hat{g}_\beta \prod_{e \in \binom{[n]}{2}} \mathbf{E}[L_{\alpha(e)}(\boldsymbol{\ell}_e^0) L_{\beta(e)}(\boldsymbol{\ell}_e^t)] \\
&= \sum_\alpha \hat{f}_\alpha \hat{g}_\alpha e^{-t|\alpha|} \leqslant \mathbf{E}[f]\,\mathbf{E}[g] + e^{-t}\left(\mathbf{E}[f(\boldsymbol{\ell}^0) g(\boldsymbol{\ell}^0)] - \mathbf{E}[f]\,\mathbf{E}[g]\right).
\end{aligned}$$

The conclusion now follows because f and g are indicators. ■

We note that the Langevin dynamics evolves the weights smoothly:

Claim B.1 *For any fixed $t > 0$, if $\boldsymbol{\ell}^0$ and $\boldsymbol{\ell}^t$ are sampled as above, then for all n large enough,*

$$\|\boldsymbol{\ell}^0 - \boldsymbol{\ell}^t\|_2 \leqslant \sqrt{2(1 - e^{-t}) + o_n(1)}\|\boldsymbol{\ell}^0 - \mathbf{E}[\boldsymbol{\ell}^0]\|_2 \leqslant \sqrt{1 - e^{-t} + o_n(1)}\|\boldsymbol{\ell}^0\|_2.$$

with probability $\geqslant 1 - \frac{1}{\mathrm{poly} n}$.

Proof As mentioned above, the correlation between $\boldsymbol{X}_0$ and $\boldsymbol{X}_t$ can be computed as e^{-t}. The claim now follows from the classic equality $\mathbf{E}[(\boldsymbol{X} - \boldsymbol{X}')^2] = 2\,\mathbf{Var}[\boldsymbol{X}] - 2\,\mathbf{Cov}[\boldsymbol{X}, \boldsymbol{X}']$ for $\boldsymbol{X}, \boldsymbol{X}'$ with the same marginal distribution, combined with concentration of sums of independent random variables, and finally the fact that $\mathbf{Var}(\boldsymbol{X}) = 1$ when $\boldsymbol{X} \sim \mathrm{Exp}(1)$. ■

B.3. The overlap-gap property for first passage percolation

We now show that any two paths $p_1 \neq p_2$ in $\mathcal{P}_\varepsilon(\boldsymbol{\ell}^0) \times \mathcal{P}_\varepsilon(\boldsymbol{\ell}^t)$ have the overlap-gap property, and that when t is large enough large overlap is improbable.

Lemma 27 *For any $t \geqslant 0$, let $\boldsymbol{\ell}^0 \sim \mathrm{Exp}(1)^{\otimes\binom{[n]}{2}}$ and let $\boldsymbol{\ell}^t$ be the outcome of a time-t Langevin diffusion starting from $\boldsymbol{\ell}^0$. There exist universal constants $C_1, C_2 > 0$ and $\varepsilon_0 > 0$ such that for all $\varepsilon < \varepsilon_0$, it holds with probability $1 - 1/\mathrm{poly}(n)$ that all pairs $(p, p') \in \mathcal{P}_\varepsilon(\boldsymbol{\ell}^0) \times \mathcal{P}_\varepsilon(\boldsymbol{\ell}^t)$ have*

$$\frac{|p \cap p'|}{|p| \cdot |p'|} \in \begin{cases} [0, C_1\sqrt{\varepsilon}) & t > (1 + C_2\sqrt{\varepsilon}) \log n \\ [0, C_1\sqrt{\varepsilon}) \cup \{1\} & \textit{otherwise.} \end{cases}$$

Furthermore, with probability $1 - O(\frac{1}{\log n})$ there exist $p, p' \in \mathcal{P}_\varepsilon(\boldsymbol{\ell}^0) \times \mathcal{P}_\varepsilon(\boldsymbol{\ell}^0)$ with overlap $\leqslant C_1\sqrt{\varepsilon}$.

Proof To establish the overlap-gap property, we bound the expected number of overlapping paths. We begin by giving a bound for non-full overlaps in the case when $t = 0$; for simplicity's sake we call $\boldsymbol{\ell} = \boldsymbol{\ell}^0$. By Theorem 26, the bound will transfer to any $t \geqslant 0$. We will then get a sharper bound on the full-overlap case when $t > 0$.

Suppose paths p_1, p_2 (both with endpoints 1,2) overlap on k edges, and that $\mathrm{hop}(p_1) = k + m_1$ and $\mathrm{hop}(p_2) = k + m_2$, with $m_1, m_2 \neq 0$. We will bound the expected number of pairs of such paths with $\mathrm{len}_{\boldsymbol{\ell}}(p_1), \mathrm{len}_{\boldsymbol{\ell}}(p_2) \leqslant (1 + \varepsilon)\frac{\log n}{n}$.

First, we compute the probability that $\mathrm{len}_{\boldsymbol{\ell}}(p_1), \mathrm{len}_{\boldsymbol{\ell}}(p_2) \leqslant (1 + \varepsilon)\frac{\log n}{n}$. The length of path p_1 is the sum of the lengths of the k shared edges, plus the sum of the lengths of the m_1 edges unique to p_1 (and the same for p_2). So letting $L = (1 + \varepsilon)\frac{\log n}{n}$ and defining the independent random variables $\boldsymbol{A} \sim \mathrm{Gamma}(k, 1)$, $\boldsymbol{B} \sim \mathrm{Gamma}(m_1, 1)$, and $\boldsymbol{C} \sim \mathrm{Gamma}(m_2, 1)$, we have by direct integration of the density functions that

$$\begin{aligned}
&\mathbf{Pr}[\mathrm{len}_{\boldsymbol{\ell}}(p_1), \mathrm{len}_{\boldsymbol{\ell}}(p_2) \leqslant L] &(4)\\
&= \mathbf{Pr}[\boldsymbol{A} + \boldsymbol{B} \leqslant L,\ \boldsymbol{A} + \boldsymbol{C} \leqslant L] \\
&= \int_0^L \frac{a^{k-1}e^{-a}}{(k-1)!} \left(\int_0^{L-a} \frac{b^{m_1-1}e^{-b}}{(m_1-1)!} db\right) \left(\int_0^{L-a} \frac{c^{m_2-1}e^{-c}}{(m_2-1)!} dc\right) da \\
&\leqslant \frac{1}{(k-1)!(m_1-1)!(m_2-1)!} \int_0^L a^{k-1} \int_0^{L-a} b^{m_1-1} db \int_0^{L-a} c^{m_2-1} dc\, da \\
&= \frac{1}{(k-1)!m_1!m_2!} \int_0^L a^{k-1}(L-a)^{m_1+m_2} da \\
&= \frac{1}{(k-1)!m_1!m_2!} \frac{(k-1)!(m_1+m_2)!}{(m_1+m_2+k)!} L^{k+m_1+m_2} &(5)\\
&\leqslant \binom{m_1+m_2}{m_1} \cdot \left(\frac{eL}{k+m_1+m_2}\right)^{k+m_1+m_2},
\end{aligned}$$

where in the last line we have applied Stirling's approximation.

Applying Claim 2.1, so long as $m_1, m_2 \neq 0$, the expected number of pairs of approximate shortest paths, in this case, is at most

$$\begin{aligned}
&\mathbf{E}[\#\{p_1, p_2\}] = M_{k, m_1-k, m_2-k} \cdot \mathbf{Pr}[\mathrm{len}_{\boldsymbol{\ell}}(p_1), \mathrm{len}_{\boldsymbol{\ell}}(p_2) \leqslant (1+\varepsilon)\tfrac{\log n}{n}] \\
&\leqslant \left(\frac{k+1}{n^2} + \frac{O(\log^{15} n)}{n^3}\right) \cdot n^{k+m_1+m_2} \cdot \binom{m_1+m_2}{m_1} \cdot \left(\frac{e(1+\varepsilon)\log n}{(k+m_1+m_2)n}\right)^{k+m_1+m_2}.
\end{aligned}$$

Without loss of generality, from Theorem 25, we may assume that the hoplength of p_1 and p_2 is $(1 \pm 2\sqrt{\varepsilon}) \log n$, and we will do so from now on to simplify the computation. Suppose $k = (1-\lambda)\log n$, $m_1 = (\lambda + \delta_1)\log n$, and $m_2 = (\lambda+\delta_2)\log n$ for $|\delta_1|, |\delta_2| \leqslant 2\sqrt{\varepsilon}$. Bounding the display above, for ε sufficiently small,

$$\mathbf{E}[\#\{p_1, p_2\}] \leqslant \left(\frac{O(\log n)}{n^2}\right) \left(\frac{e\,(1+\varepsilon)}{1+\lambda-4\sqrt{\varepsilon}}\right)^{(1+\lambda+4\sqrt{\varepsilon})\log n} \cdot 2^{(2\lambda+4\sqrt{\varepsilon})\log n}.$$

Taking logarithms,

$$\begin{aligned}\frac{\log \mathbf{E}[\#\{p_1,p_2\}]}{\log n} &\leqslant -2 + (1+\lambda+4\sqrt{\varepsilon})\log\left(\frac{e(1+\varepsilon)}{1+\lambda-4\sqrt{\varepsilon}}\right) + (2\lambda+4\sqrt{\varepsilon})\log 2 + o_n(1) \\ &\leqslant -2 + C\sqrt{\varepsilon} + (1+\lambda)\log\frac{e}{(1+\lambda)} + 2\lambda\log 2 + o_n(1),\end{aligned}$$

for C a universal constant, whenever ε is sufficiently small. Taking a first derivative in λ, one can see that the above expression is increasing in λ when $\lambda \in (-1, 3)$. Further, evaluating the expression at $\lambda = 1-\beta$ for $\beta < \frac{1}{2}$, we obtain an upper bound:

$$\begin{aligned}&-2 + C\sqrt{\varepsilon} + (2-\beta)\,(1-\log(2-\beta)) + 2(1-\beta)\log 2 \\ &= C\sqrt{\varepsilon} - \beta + 2\log\frac{2}{2-\beta} + \beta\log(2-\beta) - 2\beta\log 2\end{aligned}$$

Applying the identity $\log\frac{1}{1-x} \leqslant x + x^2$ valid for all $0 \leqslant x \leqslant \frac{1}{2}$,

$$\begin{aligned}&\leqslant C\sqrt{\varepsilon} - \beta + (\beta + \frac{1}{2}\beta^2) + (\beta\log 2 + \beta\log(1-\beta/2)) - 2\beta\log 2 \\ &\leqslant C\sqrt{\varepsilon} + \frac{1}{2}\beta^2 - \beta\log 2 + \beta\log(1-\beta/2)\end{aligned}$$

And finally applying the identity $\log(1-x) \leqslant -x$,

$$\leqslant C\sqrt{\varepsilon} - \beta\log 2.$$

Hence the quantity is negative when $\beta \geqslant C\sqrt{\varepsilon}/2\log 2$. Combined with the discussion above, when $-1 < \lambda < 1 - C\frac{\sqrt{\varepsilon}}{\log 2}$ and $(1-\lambda)\log n < |p_1|, |p_2|$, $\mathbf{E}[\#\{p_1,p_2\}] \leqslant n^{-\gamma}$ for $\gamma > 0$ a universal constant, for all n sufficiently large. Applying Markov's inequality and taking a union bound over the $O(\log^4 n)$ possible values of c, m_1, m_2, k, we have that with high probability, there do not exist $p_1, p_2 \in \mathcal{P}_\varepsilon(\boldsymbol{\ell})$ of overlap $\alpha = \frac{1-\lambda}{\sqrt{(1+\delta_1)(1+\delta_2)}} \in (C_1\sqrt{\varepsilon}, 1)$. The same is true if $t > 0$, only the probability term changes, and by Theorem 26 it can only decrease.

Now, in the case when $m_1 = m_2 = 0$, we consider $t > 0$. Using again our bound on the hoplength from Theorem 25, $k = (1 \pm 2\sqrt{\varepsilon})\log n$, and we have that

$$\begin{aligned}\mathbf{E}[\#p_1, p_2] &\leqslant n^k \left(e^{-t}\left(\frac{e(1+\varepsilon)\log n}{nk}\right)^{2k} + \left(\frac{e(1+\varepsilon)\log n}{nk}\right)^{2k}\right) \\ &\leqslant e^{-t}\left(\frac{e(1+\varepsilon)}{1-2\sqrt{\varepsilon}}\right)^{(1+2\sqrt{\varepsilon})\log n} + \frac{1}{\text{poly}(n)},\end{aligned}$$

which is $o_n(1)$ for all ε sufficiently small when $t \geqslant (1 + C_2\sqrt{\varepsilon}) \log n$ for some constant C_2. The conclusion follows by Markov's inequality.

To obtain the guarantee for the existence of near-disjoint paths, we argue that $|\mathcal{P}_\varepsilon(\boldsymbol{\ell}^0)| \gg 1$ with high probability via a second moment argument. Given that paths must overlap either fully or by at most $C_1\sqrt{\varepsilon}$, this implies there must exist two paths of low overlap, by the pigeonhole principle.

Let $\boldsymbol{N}_{h,L}$ be the number of paths of hoplength $h = \log n$ and cost at most $L = (1+\varepsilon)\frac{\log n}{n}$. Notice that $|\mathcal{P}_\varepsilon(\boldsymbol{\ell})| \geqslant \boldsymbol{N}_{h,L}$. We will show that $\boldsymbol{N}_{h,L}$ concentrates around its expectation, which is at least

$$\mathbf{E}[\boldsymbol{N}_{h,L}] = n^{\underline{h-1}} \int_0^L \frac{1}{(h-1)!} a^{h-1} e^{-a} da \geqslant n^{h-1} e^{-L} \frac{L^h}{h!}(1-o(1)) \geqslant \frac{1}{n}(e(1+\varepsilon))^{\log n}(1-o(1)) \gg 1.$$

The line Equation (5) above, implies that

$$\begin{aligned}
\mathbf{E}[\boldsymbol{N}_{h,L}^2] - \mathbf{E}[\boldsymbol{N}_{h,L}]^2 &\leqslant \mathbf{E}[\boldsymbol{N}_{h,L}] + \sum_{k=1}^{h-1} O\left(\frac{k}{n^2}\right) n^{2h-k} \binom{2h-2k}{h-k} \left(\frac{(1+\varepsilon)\log n}{n}\right)^{2h-k} \frac{1}{(2h-k)!} \\
&\leqslant \mathbf{E}[\boldsymbol{N}_{h,L}] + \sum_{k=1}^{h-1} O\left(\frac{k}{n^2}\right) \binom{2h-2k}{h-k} \frac{1}{(2h-k)!} \left((1+\varepsilon)\log n\right)^{2h-k}.
\end{aligned}$$

If we define $A_k = k\binom{2h-2k}{h-k}/(2h-k)!$, then we have that $A_{k+1}/A_k = \frac{(k+1)(h-k)}{2k(2h-2k-1)(2h-k)} \leqslant 3/4$ for any $1 \leqslant k \leqslant h-1$, since $(k+1)/(2k) \leqslant 3/4$ for any $k \geqslant 2$ and the inequality holds trivially when $k = 1$. Therefore, $\sum_{k=1}^{h-1} A_k \leqslant 4A_1$ and thus

$$\begin{aligned}
\mathbf{E}[\boldsymbol{N}_{h,L}^2] - \mathbf{E}[\boldsymbol{N}_{h,L}]^2 &\leqslant \mathbf{E}[\boldsymbol{N}_{h,L}] + O\left(\frac{1}{n^2}\right) \binom{2h-2}{h-1} \frac{1}{(2h-1)!} \left((1+\varepsilon)\log n\right)^{2h-1} \\
&\leqslant \mathbf{E}[\boldsymbol{N}_{k,L}] + O\left(\frac{1}{hn^2}\right) \frac{1}{(h-1)!(h-1)!} \left((1+\varepsilon)\log n\right)^{2h-1} \\
&\leqslant \mathbf{E}[\boldsymbol{N}_{h,L}] + O\left(\frac{h}{n^2}\right) \frac{1}{h!h!} \frac{1}{\log n} \left((1+\varepsilon)\log n\right)^{2h} \\
&\leqslant \mathbf{E}[\boldsymbol{N}_{h,L}] + O\left(\frac{h}{n^2}\right) \frac{1}{\log n} \frac{1}{\log n} \left(e(1+\varepsilon)\right)^{2h} \\
&\leqslant \mathbf{E}[\boldsymbol{N}_{h,L}] + O\left(\frac{1}{n^2 \log n}\right) \left(e(1+\varepsilon)\right)^{2\log n} = O\left(\frac{1}{\log n}\right) \mathbf{E}[\boldsymbol{N}_{h,L}]^2.
\end{aligned}$$

Hence, $\mathbf{Var}[N_{h,L}] = O(\frac{1}{\log n})\, \mathbf{E}[N_{h,L}]^2$, so $\boldsymbol{N}_{h,L} \geqslant \frac{1}{2}\, \mathbf{E}[\boldsymbol{N}_{h,L}] \gg 1$ with probability $1 - O(\frac{1}{\log n})$, as desired. ■

We can leverage the above OGP to prove a lower bound for stable algorithms.

Lemma 28 *Let $t > 0$, $\varepsilon > 0$ be sufficiently small. Then for all n sufficiently large, there can be no algorithm for ε-approximate shortest path in $K_n(\boldsymbol{\ell})$ which simultaneously (i) has success probability $\geqslant 1 - o(\frac{1}{\log n})$, and (ii) is stable, in the sense that if $\boldsymbol{\ell}^0, \boldsymbol{\ell}^t$ are sampled as described above (and hence $\|\boldsymbol{\ell}^0 - \boldsymbol{\ell}^t\| \leqslant \sqrt{(1-e^{-t}) + o_n(1)}\|\boldsymbol{\ell}^0\|$ with high probability), then when $\mathcal{A}$ succeeds on its inputs, $\frac{|\mathcal{A}(\boldsymbol{\ell}^0) \cap \mathcal{A}(\boldsymbol{\ell}^t)|}{\sqrt{|\mathcal{A}(\boldsymbol{\ell}^0)| \cdot |\mathcal{A}(\boldsymbol{\ell}^t)|}} > C\sqrt{\varepsilon}$.*

Proof Sample a sequence $(\boldsymbol{\ell}_s)_{s\geqslant 0}$ by sampling $\boldsymbol{\ell}_0 \sim \mathrm{Exp}(1)^{\otimes\binom{[n]}{2}}$ and then running the Langevin dynamics described above. Let $\boldsymbol{p}_s = \mathcal{A}(\boldsymbol{\ell}_s)$. Choose $K = \lceil \frac{2\log n}{t} \rceil$. From Theorem 6, we have that with probability at least $1 - O(\frac{K}{\mathrm{poly} n})$, for each $k \in [K]$, the path $\boldsymbol{p}_{tk}$ must have overlap with $\boldsymbol{p}_0$ which is either equal to 1, or at most $C\sqrt{\varepsilon}$. Furthermore, since $tK \geqslant 2\log n$, the overlap of $\boldsymbol{p}_0$ and $\boldsymbol{p}_T$ must be at most $C\sqrt{\varepsilon}$. Hence, with high probability there must exist some $k \in [K]$ with $\frac{|\boldsymbol{p}_0 \cap \boldsymbol{p}_{tk}|}{\sqrt{|\boldsymbol{p}_0||\boldsymbol{p}_{tk}|}} = 1$ but $\frac{|\boldsymbol{p}_0 \cap \boldsymbol{p}_{t(k+1)}|}{\sqrt{|\boldsymbol{p}_0||\boldsymbol{p}_{t(k+1)}|}} \leqslant C\sqrt{\varepsilon}$, implying that $\frac{|\boldsymbol{p}_{tk} \cap \boldsymbol{p}_{t(k+1)}|}{\sqrt{|\boldsymbol{p}_{tk}||\boldsymbol{p}_{t(k+1)}|}} \leqslant C\sqrt{\varepsilon}$. This is a contradiction, unless $\mathcal{A}$ did not succeed on some input. By a union bound $\mathcal{A}$ must have been successful on all $K+1$ inputs with probability at least $1 - \delta(K+1)$, so it must be the case that $\delta > \frac{1}{K+1} = \Theta(\frac{1}{\log n})$. ■

B.4. Low-degree polynomial estimators and sampling

Bounded-degree polynomial estimators. It is not clear whether a degree-$O(\log n)$ polynomial estimator of the first passage percolation path exists. A first attempt may be as follows: to estimate whether (i,j) participates in the path, sum over all 1-2 paths of hoplength $(1+2\sqrt{\varepsilon})\log n$ containing (i,j), a low-degree approximation for the indicator that the total length of the same path is at most $\frac{(1+\varepsilon)\log n}{n}$. However, the degree of the polynomial approximator would have to be $\Omega(n/\log n)$ to have the decision boundary on the order of $\log n/n$.

It is perhaps possible to have an estimator of low *coordinate degree* Hopkins (2018); Brennan et al. (2021); Kunisky (2024), meaning that the estimator is a linear combination of functions each depending on $O(\log n)$ coordinates of $\boldsymbol{\ell}$. For example, if we know the law W of the edge weights that participate in the first passage percolation path, one can choose some parameter k, obtain k even-measure quantile values $Q_1, \dots, Q_k$ of W, and attempt the estimate

$$f_{ij}(\boldsymbol{\ell}) = \sum_{\|p\|^2 \leqslant (1+2\sqrt{\varepsilon})\log n} \sum_{t \vdash_k E(p)} \prod_{(i,j)\in p} \mathbf{1}[\boldsymbol{\ell}_{ij} \leqslant Q_{t(i,j)}],$$

where the notation $t \vdash_k E(p)$ means that we partition $E(p)$ into k parts of equal size.

The estimator is of coordinate degree $O(\log n)$, and so long as $k = O(1)$, the number of terms is $2^{O(\log^2 n)}$. We find it plausible that this estimator may work, but we will not undertake its analysis here.

Sampling. We remark that the sampling algorithm described in Theorem 23 will also work for sampling short paths in $K_n^{\boldsymbol{\ell}}$, given the following observation: all edges participating in paths of length at most $(1+\varepsilon)\frac{\log n}{n}$ must have weight at most $(1+\varepsilon)\frac{\log n}{n}$. Therefore, we may sparsify K_n by first deleting all edges of weight $> (1+\varepsilon)\frac{\log n}{n}$. The support of the resulting graph is the same as the support of $\mathbb{G}(n,q)$ for $q = \Theta(\frac{\log n}{n})$, therefore the algorithm described in Theorem 23 can be used to produce a list of all potential length $\leqslant (1+\varepsilon)\frac{\log n}{n}$ paths from 1 to 2, which will have polynomial size with high probability. The list can further be filtered to remove paths that are too long.

We believe that a Wasserstein lower bound in the style of Equation (3) will rule out smooth sampling algorithms in this case too; however we did not undertake the (at this point, somewhat tedious) verification of this fact.

Appendix C. Concentration of symmetric low-degree polynomials of sparse random variables

In this section, we show the concentration of symmetric low-degree polynomials of sparse random variables, which is needed for the proof of Theorem 10. We will appeal to a Gaussian invariance principle for correlated random variables proven in Caravenna et al. (2023) via an interpolation argument, along with the result for Gaussian polynomials in Gamarnik et al. (2024) (which is a straightforward argument based on hypercontractivity). Our main goal will be to establish bounds on the expected fourth moment of the gradient of a function associated with our polynomial.

Suppose that $f : \{0,1\}^{\binom{[n]}{2}} \to \mathbb{R}^{\binom{[n]}{2}}$ is a degree-D polynomial approximation to the shortest path, symmetric under the action of $S_{[n]\setminus\{1,2\}}$. Suppose $\boldsymbol{G}, \boldsymbol{G}'$ are ρ-correlated samples from $\mathbb{G}(n,q)$ with $\rho \geqslant 0$. With probability $\geqslant (1-q)^2$, the edge $(1,2)$ is in neither of $\boldsymbol{G}, \boldsymbol{G}'$. Henceforth, we condition on the event that $(1,2)$ is in neither $\boldsymbol{G}, \boldsymbol{G}'$.

Let $\mathcal{H}$ be the set of all labeled graphs $H = (V(H), E(H))$, with up to 2 special vertices labeled $s(H), t(H)$, and $|E(H)| \leqslant D$. Let $\hookrightarrow$ denote an injection, and for $H \in \mathcal{H}$ let $\mathcal{L}(H)$ denote the set of all labelings $\ell : V(H) \hookrightarrow [n]$ such that $\ell(s(H)) = 1$ and $\ell(t(H)) = 2$.

We say a graph H is of uniform edge density at most α if all subgraphs $J \subseteq H$ have edge density at most α. Note that if H is of uniform edge density at most 1, each of its connected components contains at most one cycle.

Definition 29 *Let the active graph shapes be $\mathcal{H}$, the set of all graphs H satisfying the following conditions: (1) $|E(H)| \leqslant D$, (2) the vertices of H are unlabeled save for at most four "special" vertices labeled s, t, u, v, (3) H is of uniform edge density at most 1, and (4) any component of H containing the special vertices s or t can contain at most one of them, and is a tree.*

Lemma 30 *If $f : \{0,1\}^{\binom{[n]}{2}} \to \mathbb{R}^N$ is a vector-valued polynomial of degree $D = O(1)$, then there exists a constant $\delta > 0$ depending only on D such that so long as $q_n \leqslant \frac{1}{n^{1-\delta}}$, with probability $1 - o_n(1)$ over $\boldsymbol{G} \sim \mathbb{G}(n, q_n)$,*

$$f(\boldsymbol{G}) = f^{\mathcal{H}}(\boldsymbol{G}),$$

where $f^{\mathcal{H}}$ is the restriction of f to monomials of the form $\prod_{e \in E(H)} \mathbf{1}[\boldsymbol{G}_e = 1]$ for subgraphs H which are isomorphic to some active graph shape $H^* \in \mathcal{H}$, when the special label s is identified with 1 and the special label t is identified with 2.

Proof For any H, $\prod_{e \in E(H)} \mathbf{1}[\boldsymbol{G}_e = 1]$ is nonzero only when, for all $J \subseteq H$, $\prod_{e \in E(J)} \mathbf{1}[\boldsymbol{G}_e = 1]$ is nonzero. So the statement holds if we can show that with probability $1 - o_n(1)$, $\boldsymbol{G}$ contains no subgraphs J with at most D edges which either (a) have density > 1, (b) is connected and contains both vertex 1 and 2, or (c) contains one of 1 or 2, is connected, and is denser than a tree. The argument will be via the first moment method in all three cases.

Case (a): if J has k vertices and $m > k$ edges for $m \leqslant D$, then the expected number of appearances of J is $\leqslant n^k q_n^m \leqslant n^k q_n^{k+1} \leqslant \frac{1}{n^{1-\delta(k+1)}} = o_n(1)$ if δ is small enough as a function of D.

Case (b): if J has k vertices and $m \geqslant k-1$ edges (the minimum necessary for it to be connected) for $m \leqslant D$, and two of the vertices are required to map to 1 and 2, then the expected number of appearances of J is $\leqslant n^{k-2} q_n^m \leqslant n^{k-2} q_n^{k-1} \leqslant \frac{1}{n^{1-\delta(k-1)}} = o_n(1)$ if δ is small enough as a function of D.

Case (c): if J has k vertices and $m \geqslant k$ edges (the minimum necessary for it to be connected and have a cycle), and at least one of the vertices must map to 1 or 2, then the expected number of appearances of J is $\leqslant n^{k-1} \cdot 2 \cdot q_n^k \leqslant 2\frac{1}{n^{1-\delta(k)}} = o_n(1)$ (again when δ is small enough as a function of D.

Taking a union bound over the at most $\exp(D^2)$ graphs on D vertices completes the proof. ■

Henceforth, we shall assume f is supported on active graphs, and we write $f = f^{\mathcal{H}}$ (we will drop the superscript $\mathcal{H}$ to keep the notation clean).

We let $s_H = |V(H) \cap \{s,t\}|$, $u_H = |V(H) \cap \{u,v\}|$, $v_H = |V(H) \setminus \{s,t,u,v\}|$, and $e_H = |E(H)|$. For each $H \in \mathcal{H}$ and $i \neq j \in [n]$, let $\mathcal{L}_{H,i,j}$ be the set of one-to-one maps $\ell : V(H) \hookrightarrow [n]$ which satisfy that $\ell(s) = 1, \ell(t) = 2, \ell(u) = i, \ell(v) = j$; each map appears with multiplicity one, meaning that if $\ell(H)$ and $\ell'(H)$ are isomorphic then $\ell = \ell'$.

When f is symmetric under the action of $S_{[n]\setminus\{1,2\}}$, we can always write

$$f_{ij}(\boldsymbol{G}) = \sum_{H \in \mathcal{H}} \hat{f}_H \sum_{\ell \in \mathcal{L}_{H,i,j}} \chi_{\ell(H)}(\boldsymbol{G}),$$

where the functions χ_α are the orthonormal basis of Walsh-Hadamard characters, with $\chi_\alpha(G) = \prod_{(i,j)\in\alpha} \frac{G_{ij}-q}{\sqrt{q(1-q)}}$.

We assume f is normalized so that $\mathbf{E}[\|f(\boldsymbol{G})\|^2] = 1$, which means that

$$\begin{aligned}
1 &= \sum_{ij} \mathbf{E}\left[\left(\sum_{H \in \mathcal{H}} \hat{f}_H \sum_{\ell \in \mathcal{L}_{H,i,j}} \chi_{\ell(H)}(\boldsymbol{G})\right)^2\right] \\
&= \sum_{ij} \sum_{H \in \mathcal{H}} \hat{f}_H^2 \cdot \frac{n^{v_H}}{\mathrm{aut}(H)} \\
&= \sum_{H \in \mathcal{H}} \hat{f}_H^2 \cdot \frac{n^{v_H + u_H}}{\mathrm{aut}(H)}.
\end{aligned}$$

where $\mathrm{aut}(H)$ is the number of automorphisms of H. Thus $|\hat{f}_H| \lesssim \left(\frac{1}{n}\right)^{\frac{1}{2}(v_H+u_H)}$ when $D = O(1)$.

We now compare the polynomial of $\boldsymbol{G}$ with that of correlated Gaussians. Recall that $\boldsymbol{G}, \boldsymbol{G}'$ are ρ-correlated samples from $\mathbb{G}(n,q)$. The inner product $c(\boldsymbol{G},\boldsymbol{G}') = \langle f(\boldsymbol{G}), f(\boldsymbol{G}')\rangle$ can be written as

$$c(\boldsymbol{G},\boldsymbol{G}') = \sum_{i,j} \sum_{H_1,H_2 \in \mathcal{H}} \hat{f}_{H_1}\hat{f}_{H_2} \sum_{\ell_1 \in \mathcal{L}_{H_1,i,j}, \ell_2 \in \mathcal{L}_{H_2,i,j}} \chi_{\ell_1(H_1)}(\boldsymbol{G})\chi_{\ell_2(H_2)}(\boldsymbol{G}').$$

Note that for each edge a, if we denote $\boldsymbol{\chi}_a^{(1)} = \chi_a(\boldsymbol{G})$ and $\boldsymbol{\chi}_a^{(2)} = \chi_a(\boldsymbol{G}')$, then $c(\boldsymbol{G},\boldsymbol{G}') = \Phi(\boldsymbol{\chi}^{(1)}, \boldsymbol{\chi}^{(2)})$, where $\Phi(\cdot)$ is a multi-linear polynomial on $2\binom{n}{2}$ variables $\boldsymbol{\chi}^{(1)}$ and $\boldsymbol{\chi}^{(2)}$. Further,

$$\|f(\boldsymbol{G}) - f(\boldsymbol{G}')\|_2^2 = c(\boldsymbol{G},\boldsymbol{G}) + c(\boldsymbol{G}',\boldsymbol{G}') - 2c(\boldsymbol{G},\boldsymbol{G}').$$

Therefore similarly, if we denote $\boldsymbol{\chi}_a^{(1)} = \boldsymbol{\chi}_a^{(3)} = \chi_a(\boldsymbol{G})$ and $\boldsymbol{\chi}_a^{(2)} = \boldsymbol{\chi}_a^{(4)} = \chi_a(\boldsymbol{G}')$, then define Ψ as

$$\begin{aligned}
\|f(\boldsymbol{G}) - f(\boldsymbol{G}')\|_2^2 &=: \Psi(\boldsymbol{\chi}^{(1)}, \boldsymbol{\chi}^{(2)}, \boldsymbol{\chi}^{(3)}, \boldsymbol{\chi}^{(4)}) && (6)\\
&= \Phi(\boldsymbol{\chi}^{(1)}, \boldsymbol{\chi}^{(3)}) + \Phi(\boldsymbol{\chi}^{(2)}, \boldsymbol{\chi}^{(4)}) - \Phi(\boldsymbol{\chi}^{(1)}, \boldsymbol{\chi}^{(4)}) - \Phi(\boldsymbol{\chi}^{(2)}, \boldsymbol{\chi}^{(3)}), && (7)
\end{aligned}$$

where $\Psi(\cdot)$ is a multi-linear polynomial on $4\binom{n}{2}$ variables $\boldsymbol{\chi}^{(1)}$, $\boldsymbol{\chi}^{(2)}$, $\boldsymbol{\chi}^{(3)}$, and $\boldsymbol{\chi}^{(4)}$. Let $\boldsymbol{Z}^{(1)}$, $\boldsymbol{Z}^{(2)}$, $\boldsymbol{Z}^{(3)}$, and $\boldsymbol{Z}^{(4)}$ be length $\binom{n}{2}$ mean-0 jointly Gaussian vectors such that they are independent of the $\boldsymbol{\chi}$'s and has the same covariance structure. For simplicity, write $\boldsymbol{\chi} = (\boldsymbol{\chi}^{(1)}, \boldsymbol{\chi}^{(2)}, \boldsymbol{\chi}^{(3)}, \boldsymbol{\chi}^{(4)})$ and $\boldsymbol{Z} = (\boldsymbol{Z}^{(1)}, \boldsymbol{Z}^{(2)}, \boldsymbol{Z}^{(3)}, \boldsymbol{Z}^{(4)})$.

Lemma 31 *Let $h : \mathbb{R} \to \mathbb{R}$ be a bounded function with bounded first three derivatives. Then*

$$|\,\mathbf{E}\,h(\Psi(\boldsymbol{\chi})) - \mathbf{E}\,h(\Psi(\boldsymbol{Z}))| \leqslant O\left(\frac{\|h'''(x)\|_\infty}{n\sqrt{q}}\right).$$

Proof [Proof of Theorem 31] For $\mathsf{s}, \mathsf{t} \in [0,1]$, $k \in \{1,2,3,4\}$, and any edge a, we define $\boldsymbol{W}_{\mathsf{s},\mathsf{t}}^{(k),a}$ to be a length $\binom{n}{2}$ vector such that for any edge e,

$$\left(\boldsymbol{W}_{\mathsf{s},\mathsf{t}}^{(k),a}\right)_e = \mathsf{s}\sqrt{\mathsf{t}}\boldsymbol{\chi}_e^{(k)}\mathbf{1}(e=a) + \sqrt{\mathsf{t}}\boldsymbol{\chi}_e^{(k)}\mathbf{1}(e \neq a) + \sqrt{1-\mathsf{t}}\boldsymbol{Z}_e.$$

Write $\boldsymbol{W}_{\mathsf{s},\mathsf{t}}^{a} = (\boldsymbol{W}_{\mathsf{s},\mathsf{t}}^{(1),a}, \boldsymbol{W}_{\mathsf{s},\mathsf{t}}^{(2),a}, \boldsymbol{W}_{\mathsf{s},\mathsf{t}}^{(3),a}, \boldsymbol{W}_{\mathsf{s},\mathsf{t}}^{(4),a})$. By a result of Caravenna et al. (2023), it suffices to show that the expected fourth moment of gradients of functions in $\boldsymbol{W}$ are bounded.

Lemma 32 (Lemma A.4 in Caravenna et al. (2023)) *Let $h : \mathbb{R} \to \mathbb{R}$ be a bounded function with bounded first three derivatives. Then there exists an absolute constant C such that*

$$|\,\mathbf{E}\,h(\Psi(\boldsymbol{\chi})) - \mathbf{E}\,h(\Psi(\boldsymbol{Z})))| \leqslant C\|h'''(x)\|_\infty \sup_a \mathbf{E}[|\chi_a(\boldsymbol{G})|^3] \sum_a \sum_{k=1}^{4} \sup_{\mathsf{s},\mathsf{t}\in[0,1]} \mathbf{E}\left[|\partial_{a,k}\Psi(\boldsymbol{W}_{\mathsf{s},\mathsf{t}}^{a})|^3\right],$$

here if we write $\Psi(\cdot)$ as a function on $x^{(1)}, x^{(2)}, x^{(3)}, x^{(4)}$ variables, then $\partial_{a,k}$ denotes the partial derivative with respect to $x_a^{(k)}$.

The term $\mathbf{E}[|\chi_a(\boldsymbol{G})|^3] = \frac{1-2q}{\sqrt{q(1-q)}}$ by direct computation. It then remains to figure out the last expectation term. By Hölder's inequality and Equation (6), we know that for some absolute constant C,

$$\sup_{\mathsf{s},\mathsf{t}\in[0,1]} \mathbf{E}\left[|\partial_{a,k}\Psi(\boldsymbol{W}_{\mathsf{s},\mathsf{t}}^{a})|^3\right] \leqslant \sup_{\mathsf{s},\mathsf{t}\in[0,1]} \mathbf{E}\left[|\partial_{a,k}\Psi(\boldsymbol{W}_{\mathsf{s},\mathsf{t}}^{a})|^4\right]^{3/4} \tag{8}$$

$$\leqslant C \sum_{i\in\{1,2\}} \sum_{j\in\{3,4\}} \sup_{\mathsf{s},\mathsf{t}\in[0,1]} \mathbf{E}\left[|\partial_{a,k}\Phi(\boldsymbol{W}_{\mathsf{s},\mathsf{t}}^{(i),a}, \boldsymbol{W}_{\mathsf{s},\mathsf{t}}^{(j),a})|^4\right]^{3/4}. \tag{9}$$

Now we bound the fourth moment. Note that the partial derivative $\partial_{a,k}\Phi(\boldsymbol{W}_{\mathsf{s},\mathsf{t}}^{(i),a}, \boldsymbol{W}_{\mathsf{s},\mathsf{t}}^{(j),a})$ is only non-zero when $k = i$ or j and if so, $\partial_{a,i}\Phi(\boldsymbol{W}_{\mathsf{s},\mathsf{t}}^{(i),a}, \boldsymbol{W}_{\mathsf{s},\mathsf{t}}^{(j),a}) = \partial_{a,j}\Phi(\boldsymbol{W}_{\mathsf{s},\mathsf{t}}^{(i),a}, \boldsymbol{W}_{\mathsf{s},\mathsf{t}}^{(j),a})$, so we will drop the k in the subscript and only use ∂_a. To simplify notation, we use $\boldsymbol{\chi}$ and $\tilde{\boldsymbol{\chi}}$ to denote $\boldsymbol{\chi}^{(i)}$ and $\boldsymbol{\chi}^{(j)}$, and similarly for $\boldsymbol{Z}$ and $\boldsymbol{W}$. So when $(i,j) = (1,3)$ or $(2,4)$, the correlation between $\boldsymbol{\chi}$ and $\tilde{\boldsymbol{\chi}}$ is 1, and when $(i,j) = (1,4)$ or $(2,3)$, the correlation may be less than one (but non-negative). We will give a bound that applies to any correlation value. Recall that by definition,

$$\Phi(\boldsymbol{W}_{\mathsf{s},\mathsf{t}}^{a}, \tilde{\boldsymbol{W}}_{\mathsf{s},\mathsf{t}}^{a}) = \sum_{i,j} \sum_{H_1,H_2\in\mathcal{H}} \hat{f}_{H_1}\hat{f}_{H_2} \sum_{\ell_1\in\mathcal{L}_{H_1,i,j},\ell_2\in\mathcal{L}_{H_2,i,j}} (\boldsymbol{W}_{\mathsf{s},\mathsf{t}}^{a})_{\ell_1(H_1)}(\tilde{\boldsymbol{W}}_{\mathsf{s},\mathsf{t}}^{a})_{\ell_2(H_2)}.$$

Thus

$$\partial_a \Phi(\boldsymbol{W}^a_{\mathsf{s},\mathsf{t}}, \tilde{\boldsymbol{W}}^a_{\mathsf{s},\mathsf{t}}) = \sum_{i,j} \sum_{H_1,H_2 \in \mathcal{H}} \hat{f}_{H_1} \hat{f}_{H_2} \sum_{\substack{\ell_1 \in \mathcal{L}_{H_1,i,j}, \ell_2 \in \mathcal{L}_{H_2,i,j} \\ a \in \ell_1(H_1)}} (\boldsymbol{W}^a_{\mathsf{s},\mathsf{t}})_{\ell_1(H_1) \setminus a} (\tilde{\boldsymbol{W}}^a_{\mathsf{s},\mathsf{t}})_{\ell_2(H_2)},$$

where each term $(\boldsymbol{W}^a_{\mathsf{s},\mathsf{t}})_{\ell_1(H_1) \setminus a} (\tilde{\boldsymbol{W}}^a_{\mathsf{s},\mathsf{t}})_{\ell_2(H_2)}$ equals

$$\prod_{b_1 \in \ell_1(H_1) \setminus a} (\sqrt{\mathsf{t}} \boldsymbol{\chi}_{b_1} + \sqrt{1-\mathsf{t}} \boldsymbol{Z}_{b_1}) \prod_{b_2 \in \ell_2(H_2)} (\mathsf{s}\sqrt{\mathsf{t}} \tilde{\boldsymbol{\chi}}_{b_2} \mathbf{1}(b_2 = a) + \sqrt{\mathsf{t}} \tilde{\boldsymbol{\chi}}_{b_2} \mathbf{1}(b_2 \neq a) + \sqrt{1-\mathsf{t}} \tilde{\boldsymbol{Z}}_{b_2}). \tag{10}$$

We compute

$$\mathbf{E}\, \partial_a \Phi(\boldsymbol{W}^a_{\mathsf{s},\mathsf{t}}, \tilde{\boldsymbol{W}}^a_{\mathsf{s},\mathsf{t}})^4 = \mathbf{E} \left(\sum_{i,j} \sum_{H_1,H_2 \in \mathcal{H}} \hat{f}_{H_1} \hat{f}_{H_2} \sum_{\substack{\ell_1 \in \mathcal{L}_{H_1,i,j}, \ell_2 \in \mathcal{L}_{H_2,i,j} \\ a \in \ell_1(H_1)}} (\boldsymbol{W}^a_{\mathsf{s},\mathsf{t}})_{\ell_1(H_1) \setminus a} (\tilde{\boldsymbol{W}}^a_{\mathsf{s},\mathsf{t}})_{\ell_2(H_2)} \right)^4$$

$$= \sum_{H_1,\cdots,H_8 \in \mathcal{H}} \prod_{k=1}^{8} \hat{f}_{H_k} \sum_{\substack{i_1,i_2,i_3,i_4 \\ j_1,j_2,j_3,j_4}} \sum_{\ell} \mathbf{E} \left[\prod_{k=1}^{4} (\boldsymbol{W}^a_{\mathsf{s},\mathsf{t}})_{\ell_{2k-1}(H_{2k-1}) \setminus a} (\tilde{\boldsymbol{W}}^a_{\mathsf{s},\mathsf{t}})_{\ell_{2k}(H_{2k})} \right],$$

where the summation $\sum_\ell$ is over all $\ell_1 \in \mathcal{L}_{H_1,i_1,j_1}$, $\ell_2 \in \mathcal{L}_{H_2,i_1,j_1}$, $\ell_3 \in \mathcal{L}_{H_3,i_2,j_2}$, $\ell_4 \in \mathcal{L}_{H_4,i_2,j_2}$, $\ell_5 \in \mathcal{L}_{H_5,i_3,j_3}$, $\ell_6 \in \mathcal{L}_{H_6,i_3,j_3}$, $\ell_7 \in \mathcal{L}_{H_7,i_4,j_4}$, $\ell_8 \in \mathcal{L}_{H_8,i_4,j_4}$, $a \in \ell_1(H_1)$, $a \in \ell_3(H_3)$, $a \in \ell_5(H_5)$, $a \in \ell_7(H_7)$. To further simplify, we use equation (10) to expand the product. Note that since we have assumed the correlation is non-negative, all the terms have non-negative expectation, therefore we can bound $\mathsf{s}\sqrt{\mathsf{t}}$ by $\sqrt{\mathsf{t}}$ and have that

$$\begin{aligned} &\mathbf{E} \left[\prod_{k=1}^{4} (\boldsymbol{W}^a_{\mathsf{s},\mathsf{t}})_{\ell_{2k-1}(H_{2k-1}) \setminus a} (\tilde{\boldsymbol{W}}^a_{\mathsf{s},\mathsf{t}})_{\ell_{2k}(H_{2k})} \right] \\ &\leqslant \mathbf{E} \left[\prod_{k=1}^{4} (\sqrt{\mathsf{t}} \boldsymbol{\chi} + \sqrt{1-\mathsf{t}} \boldsymbol{Z})_{\ell_{2k-1}(H_{2k-1}) \setminus a} (\sqrt{\mathsf{t}} \tilde{\boldsymbol{\chi}} + \sqrt{1-\mathsf{t}} \tilde{\boldsymbol{Z}})_{\ell_{2k}(H_{2k})} \right]. \end{aligned}$$

Next we control the right hand side.

Lemma 33 *For any $k_1, k_2 \geqslant 0$ such that $2 \leqslant k = k_1 + k_2 \leqslant 8$, and any edge b,*

$$\mathbf{E} \left[(\sqrt{\mathsf{t}} \boldsymbol{\chi}_b + \sqrt{1-\mathsf{t}} Z_b)^{k_1} (\sqrt{\mathsf{t}} \tilde{\boldsymbol{\chi}}_b + \sqrt{1-\mathsf{t}} \tilde{\boldsymbol{Z}}_b)^{k_2} \right] \leqslant 2(q/2)^{-(k-2)/2}.$$

Proof [Proof of Theorem 33] By direct computation, $\mathbf{E}[\boldsymbol{\chi}_b] = 0$ and for any $2 \leqslant k \leqslant 8$,

$$\mathbf{E}\left[\boldsymbol{\chi}_b^k\right] = \left(\frac{1-q}{\sqrt{q(1-q)}} \right)^k q + \left(\frac{-q}{\sqrt{q(1-q)}} \right)^k (1-q) \leqslant \frac{(1-q)^{k-1} + q^{k-1}}{(q(1-q))^{(k-2)/2}} \leqslant q^{-(k-2)/2},$$

where the last inequality is equality for $k = 2$ and when $k \geqslant 3$, since $q = o(1)$, $(1-q)^{k-1} + q^{k-1} = 1 - q(k-1) + O(q^2) < 1 - q(k-2)/2 \leqslant (1-q)^{(k-2)/2}$. For correlated terms, we let $\boldsymbol{\chi}$

and $\tilde{\boldsymbol{\chi}}$ be ρ correlated. Then we write $p_1 = \rho q(1-q) + q^2$, $p_2 = \rho q(1-q) + (1-q^2)$, and $p_3 = (1-\rho)q(1-q)$, and denote $q_1 = (1-q)/\sqrt{q(1-q)}$ and $q_2 = -q/\sqrt{q(1-q)}$. For $k_1, k_2 \geqslant 0$ such that $2 \leqslant k = k_1 + k_2 \leqslant 8$,

$$\begin{aligned}\mathbf{E}\left[\boldsymbol{\chi}_b^{k_1}\tilde{\boldsymbol{\chi}}_b^{k_2}\right] &= q_1^k p_1 + q_2^k p_2 + q_1^{k_1} q_2^{k_2} p_3 + q_2^{k_1} q_1^{k_2} p_3 \leqslant q_1^k p_1 + |q_2|^k p_2 + (q_1^k + |q_2|^k) p_3 \\ &= \frac{(1-q)^{k-1} + q^{k-1}}{(q(1-q))^{(k-2)/2}} \leqslant q^{-(k-2)/2}.\end{aligned}$$

For the Gaussian term $\boldsymbol{Z}_b$, $\mathbf{E}[\boldsymbol{Z}_b] = 0$, $\mathbf{E}[\boldsymbol{Z}_b^2] = 1$ and for any $3 \leqslant k \leqslant 8$, $\mathbf{E}[\boldsymbol{Z}_b^k] \leqslant 105 \leqslant q^{-(k-2)/2}$. And by Wick's theorem, this is also true for correlated terms that $\mathbf{E}[\boldsymbol{Z}_b^{k_1}\tilde{\boldsymbol{Z}}_b^{k_2}] \leqslant \mathbf{E}[\boldsymbol{Z}_b^{k_1+k_2}] \leqslant q^{-(k-2)/2}$ for $2 \leqslant k = k_1 + k_2 \leqslant 8$. Now we claim that for any $0 \leqslant c_1 \leqslant k_1$ and $0 \leqslant c_2 \leqslant k_2$, with $2 \leqslant k = k_1 + k_2 \leqslant 8$, we have

$$\mathbf{E}\left[\boldsymbol{\chi}_b^{c_1}\tilde{\boldsymbol{\chi}}_b^{c_2}\right]\mathbf{E}\left[\boldsymbol{Z}_b^{k_1-c_1}\tilde{\boldsymbol{Z}}_b^{k_2-c_2}\right] \leqslant q^{-(k-2)/2}.$$

If $c_1 + c_2 \leqslant 2$ or $k_1 + k_2 - c_1 - c_2 \leqslant 2$, then the above holds by considering different cases and combining the previous estimates for $\mathbf{E}\left[\boldsymbol{\chi}_b^{c_1}\tilde{\boldsymbol{\chi}}_b^{c_2}\right]$ and $\mathbf{E}\left[\boldsymbol{Z}_b^{k_1-c_1}\tilde{\boldsymbol{Z}}_b^{k_2-c_2}\right]$. Else,

$$\mathbf{E}\left[\boldsymbol{\chi}_b^{c_1}\tilde{\boldsymbol{\chi}}_b^{c_2}\right]\mathbf{E}\left[\boldsymbol{Z}_b^{k_1-c_1}\tilde{\boldsymbol{Z}}_b^{k_2-c_2}\right] \leqslant q^{-(c_1+c_2-2)/2} q^{-(k_1+k_2-c_1-c_2-2)/2} \leqslant q^{-(k-4)/2} \leqslant q^{-(k-2)/2}.$$

Therefore,

$$\begin{aligned}&\mathbf{E}\left[(\sqrt{\mathsf{t}}\boldsymbol{\chi}_b + \sqrt{1-\mathsf{t}}\boldsymbol{Z}_b)^{k_1}(\sqrt{\mathsf{t}}\tilde{\boldsymbol{\chi}}_b + \sqrt{1-\mathsf{t}}\tilde{\boldsymbol{Z}}_b)^{k_2}\right] \\ &= \sum_{c_1=0}^{k_1}\binom{k_1}{c_1}\sqrt{\mathsf{t}}^{c_1}\sqrt{1-\mathsf{t}}^{k_1-c_1}\sum_{c_2=0}^{k_2}\binom{k_2}{c_2}\sqrt{\mathsf{t}}^{c_2}\sqrt{1-\mathsf{t}}^{k_2-c_2}\,\mathbf{E}\left[\boldsymbol{\chi}_b^{c_1}\tilde{\boldsymbol{\chi}}_b^{c_2}\right]\mathbf{E}\left[\boldsymbol{Z}_b^{k_1-c_1}\tilde{\boldsymbol{Z}}_b^{k_2-c_2}\right] \\ &\leqslant \sum_{c_1=0}^{k_1}\binom{k_1}{c_1}\sqrt{\mathsf{t}}^{c_1}\sqrt{1-\mathsf{t}}^{k_1-c_1}\sum_{c_2=0}^{k_2}\binom{k_2}{c_2}\sqrt{\mathsf{t}}^{c_2}\sqrt{1-\mathsf{t}}^{k_2-c_2} q^{-(k-2)/2} \\ &\leqslant (\sqrt{\mathsf{t}} + \sqrt{1-\mathsf{t}})^k q^{-(k-2)/2} \leqslant 2(q/2)^{-(k-2)/2},\end{aligned}$$

which completes the proof of Theorem 33. ■

Finally, we will require the following combinatorial lemma for bounding the number of summands for each collection of subgraphs.

Lemma 34 *Suppose $H_1, \ldots, H_m \in \mathcal{H}$. Let $H_{\text{sep}} = \bigcup_{i=1}^m H_i$ have ν_{sep} vertices and μ_{sep} edges (the special vertices s, t, if present in several of the H_i, are counted only with multiplicity one). For each $i \in [m]$, let $\ell_i : V(H_i) \hookrightarrow \mathbb{N}$ be a one-to-one labeling of the vertices, satisfying $\ell_i(s) = 1$ and $\ell_i(t) = 2$. Call H_{combo} the labeled simple graph $\bigcup_{i=1}^m \ell_i(H_i)$, with ν_{combo} vertices, and μ_{combo} edges. Then if every edge in H_{combo} is covered at least twice,*

$$\nu_{\text{combo}} \leqslant \frac{1}{2}\left(\nu_{\text{sep}} + \mathbf{1}[s \in V(H_{\text{sep}})] + \mathbf{1}[t \in V(H_{\text{sep}})] - (\mu_{\text{sep}} - 2\mu_{\text{combo}})\right).$$

Proof If a graph has ν vertices, μ edges, κ components, then

$$\nu = \kappa + \mu - \gamma,$$

where γ is the number of edges one must remove to obtain the spanning forest of the graph. We will apply this identity to H_{combo}, then use the fact that each H_i is of uniform density at most 1 to relate $\kappa_{\text{combo}}, \mu_{\text{combo}}, \gamma_{\text{combo}}$ to the related quantities of the H_i.

First, some observations about edges. Say that H_{combo} has $\tilde{\mu}_\ell$ edges covered by ℓ edges in H_{sep} for each $\ell \in \{2, \dots, m\}$. Since each edge in H_{sep} maps to an edge in H_{combo}, and each edge in H_{combo} is covered twice, $\tilde{\mu}_1 = 0$, so

$$\mu_{\text{sep}} = \sum_{\ell=2}^{m} \ell \tilde{\mu}_\ell \quad \text{and} \quad \mu_{\text{combo}} = \sum_{\ell=2}^{m} \tilde{\mu}_\ell.$$

Now, some observations about components. In H_{sep}, every component comes from some distinct H_i, except for up to two components C_s, C_t which contain the special vertices s and t; these may be composed of several H_i. Since the $H_i \in \mathcal{H}$, and since all graphs in $\mathcal{H}$ have the property that the components including s or t must be disjoint and trees, the components C_s, C_t of H_{sep} are also trees (or possibly are just the empty graph). We will say that in H_{combo}, for each $\ell \in \mathbb{N}$, there are $\tilde{\kappa}_\ell$ components which result from combining ℓ components from the H_{sep}. Since by assumption each edge in H_{combo} is covered at least twice, and each H_i has all-distinct edges, $\tilde{\kappa}_1 \leqslant \mathbf{1}[C_s \neq \emptyset] + \mathbf{1}[C_t \neq \emptyset]$, as any other component in H_{combo} has to combine at least two components to be double-covered. Thus,

$$\kappa_{\text{combo}} = \sum_{\ell=1}^{\infty} \tilde{\kappa}_\ell, \quad \tilde{\kappa}_1 \leqslant \mathbf{1}[C_s \neq \emptyset] + \mathbf{1}[C_t \neq \emptyset], \quad \text{and} \quad \kappa_{\text{sep}} = \sum_{\ell=2}^{\infty} \ell \tilde{\kappa}_\ell.$$

Finally, we make some observations about cycles. Let ξ_ℓ be the number of components in H_{combo} which contain the image of exactly ℓ cycles from H_{sep}. Then because each connected component in H_{sep} contains at most once cycle, $\gamma_{\text{sep}} = \sum_{\ell=1}^{m} \ell \xi_\ell$. We also have that

$$\gamma_{\text{combo}} \geqslant \sum_{\ell=1}^{m} \xi_\ell,$$

as one must remove at least one edge from each cycle in H_{sep} to get a forest, and if there are only ℓ cycles which map together to a component of H_{combo}, then at least one edge must be removed.

Now, we have that

$$\begin{aligned}
&\nu_{\text{combo}} \\
&= \kappa_{\text{combo}} + \mu_{\text{combo}} - \gamma_{\text{combo}} \\
&\leqslant \sum_{\ell=1}^{\infty} \tilde{\kappa}_\ell + \sum_{\ell=2}^{m} \tilde{\mu}_\ell - \sum_{\ell=1}^{m} \xi_\ell \\
&= \sum_{\ell=1}^{\infty} \tilde{\kappa}_\ell + \sum_{\ell=2}^{m} \tilde{\mu}_\ell - \sum_{\ell=1}^{m} \xi_\ell + \frac{1}{2}\left(\kappa_{\text{sep}} - \sum_{\ell=2}^{\infty} \ell\tilde{\kappa}_\ell\right) + \frac{1}{2}\left(\mu_{\text{sep}} - \sum_{\ell=2}^{m} \ell\tilde{\mu}_\ell\right) + \frac{1}{2}\left(\sum_{\ell=1}^{m} \ell\xi_\ell - \gamma_{\text{sep}}\right) \\
&= \frac{1}{2}(\kappa_{\text{sep}} + \mu_{\text{sep}} - \gamma_{\text{sep}}) + \frac{1}{2}\tilde{\kappa}_1 - \sum_{\ell=3}^{\infty} \frac{\ell-2}{2}\tilde{\kappa}_\ell - \sum_{\ell=3}^{m} \frac{\ell-2}{2}\tilde{\mu}_\ell + \sum_{\ell=1}^{m} \frac{\ell-2}{2}\xi_\ell \\
&= \frac{1}{2}\nu_{\text{sep}} + \frac{1}{2}\tilde{\kappa}_1 - \sum_{\ell=3}^{m} \frac{\ell-2}{2}\tilde{\mu}_\ell + \sum_{\ell=3}^{\infty} \frac{\ell-2}{2}\left(\xi_\ell - \tilde{\kappa}_\ell\right) \\
&\leqslant \frac{1}{2}\nu_{\text{sep}} + \frac{1}{2}(\mathbf{1}[C_s \neq \emptyset] + \mathbf{1}[C_t \neq \emptyset]) - \frac{1}{2}(\mu_{\text{sep}} - 2\mu_{\text{combo}}) + \sum_{\ell=3}^{\infty} \frac{\ell-2}{2}\left(\xi_\ell - \tilde{\kappa}_\ell\right),
\end{aligned}$$

We'll argue that the rightmost sum is at most zero, which will conclude the proof. Indeed, note that in H_{sep}, each connected component contained at most one cycle. Hence, any connected component of H_{combo} with ℓ cycles must be in the image of ℓ or more components from H_{sep}. Hence the positive contribution of each component to ξ_ℓ is negated by its negative contribution to $\tilde{\kappa}_{\ell'}$ for some $\ell' \geqslant \ell$, completing the proof. ■

Now back to the computation of the fourth moment. Call $U = \cup_{i=1}^{8} H_i$ to be the union graph of the unlabeled shapes, in which only the s, t vertices are identified if present. Let $\mathcal{C}(H_1, \cdots, H_8)$ be the set of all possible combination graphs of the H_i, such that the edge a must be covered in all H_1, H_3, H_5, H_7 and all edges are covered at least twice in $U_a = \left(\bigcup_{i\in\{1,3,5,7\}} H_i \backslash a\right) \cup \left(\bigcup_{i\in\{2,4,6,8\}} H_i\right)$. Let C be any graph in $\mathcal{C}(H_1, \ldots, H_8)$, and write C_a as the corresponding combination graph of U_a. The special vertices $\{i_{2k-1}, i_{2k}\}$ of H_{2k-1}, H_{2k} are identified if present, and the special vertices $\{s, t\}$ of all eight graphs are identified if present. Let v_C be the number of vertices in C excluding the special vertices $\{s, t\}$, and s_C be the number of $\{s, t\}$ vertices in C, and let e_C be the number of edges in C, and define v, s, e similarly for C_a and the union graphs U and U_a. We will show that

$$v_C \leqslant \frac{1}{2}v_U - \frac{1}{2}e_U + e_{C_a}. \tag{11}$$

Indeed, for $C \in \mathcal{C}(H_1, \cdots, H_8)$, we consider eight graphs $H_1\backslash a, H_2, H_3\backslash a, H_4, H_5, H_6, H_7, H_8$, and define U' as their union and C' as their combination. Note that for $H_1\backslash a$ and $H_3\backslash a$, we only delete the edge a but not vertices. Then C' and U' satisfy the conditions of Theorem 34, and we have that

$$v_{C'} + s_{C'} \leqslant \frac{1}{2}(v_{U'} + s_{U'}) + \frac{1}{2}s_{U'} - \frac{1}{2}(e_{U'} - 2e_{C'}) \implies v_{C'} \leqslant \frac{1}{2}v_{U'} - \frac{1}{2}e_{U'} + e_{C'}.$$

Therefore equation (11) follows by noticing that $v_C = v_{C'}$, $v_U = v_{U'}$, $e_U = e_{U'} + 2$, $e_{C_a} \geqslant e_{C'} - 1$. Let a_1 and a_2 be the two vertices of the edge a. Then by Theorem 33 and equation (11), if

$\{a_1, a_2\} \cap \{s, t\} = \emptyset$, then

$$
\begin{aligned}
&\mathbf{E}\, \partial_a \Phi(\boldsymbol{W}^a_{s,t}, \tilde{\boldsymbol{W}}^a_{s,t})^4 \\
&= \sum_{H_1,\cdots,H_8 \in \mathcal{H}} \prod_{k=1}^{8} \hat{f}_{H_k} \sum_{\ell} \mathbf{E} \left[\prod_{k=1}^{4} (\boldsymbol{W}^a_{s,t})_{\ell_{2k-1}(H_{2k-1}) \backslash a} (\tilde{\boldsymbol{W}}^a_{s,t})_{\ell_{2k}(H_{2k})} \right] \\
&\leqslant \sum_{H_1,\cdots,H_8 \in \mathcal{H}} \prod_{k=1}^{8} \hat{f}_{H_k} \sum_{C \in \mathcal{C}(H_1,\cdots,H_8)} n^{v_C - 2} \cdot 2(q/2)^{-\frac{1}{2}(e_{U_a} - 2e_{C_a})} \\
&= \sum_{H_1,\cdots,H_8 \in \mathcal{H}} \prod_{k=1}^{8} O\left(n^{-\frac{1}{2}(v_{H_k} + u_{H_k})}\right) \sum_{C \in \mathcal{C}(H_1,\cdots,H_8)} n^{\frac{1}{2}(e_{U_a} - 2e_{C_a}) + v_C - 2} (nq/2)^{-\frac{1}{2}(e_{U_a} - 2e_{C_a})} \\
&\leqslant O\left(n^{-\frac{1}{2}\sum_{k=1}^{8}(v_{H_k} + u_{H_k}) + \frac{1}{2}(e_{U_a} - 2e_{C_a}) + v_C - 2}\right) \leqslant O(n^{-4}),
\end{aligned}
$$

where the summation $\sum_\ell$ is over all $\ell_1 \in \mathcal{L}_{H_1,i_1,j_1}, \ell_2 \in \mathcal{L}_{H_2,i_1,j_1}, \ell_3 \in \mathcal{L}_{H_3,i_2,j_2}, \ell_4 \in \mathcal{L}_{H_4,i_2,j_2}, \ell_5 \in \mathcal{L}_{H_5,i_3,j_3}, \ell_6 \in \mathcal{L}_{H_6,i_3,j_3}, \ell_7 \in \mathcal{L}_{H_7,i_4,j_4}, \ell_8 \in \mathcal{L}_{H_8,i_4,j_4}, a \in \ell_1(H_1), a \in \ell_3(H_3), a \in \ell_5(H_5), a \in \ell_7(H_7)$. In line 4 we have used our bound on the $\hat{f}_H$, and the final line uses that $-\frac{1}{2}\sum_{k=1}^{8}(v_{H_k} + u_{H_k}) + \frac{1}{2}(e_{U_a} - 2e_{C_a}) + v_C = -\frac{1}{2}v_U + \frac{1}{2}(e_U - 4) - e_{C_a} + v_C \leqslant -2$.

If $|\{a_1, a_2\} \cap \{s, t\}| = 1$, then we can replace in the second line n^{v_C-2} by n^{v_C-1}, and get $O(n^{-3})$. Similarly, if $\{a_1, a_2\} = \{s, t\}$, then we can replace in the second line n^{v_C-2} by n^{v_C}, and get $O(n^{-2})$. Combining with Theorem 32 and Equation (9), we have that

$$
\begin{aligned}
|\mathbf{E}\, h(\Psi(\boldsymbol{\chi})) - \mathbf{E}\, h(\Psi(\boldsymbol{Z}))| &\leqslant \frac{\|h'''(x)\|_\infty}{\sqrt{q}} \left(n^2 O(n^{-4\cdot\frac{3}{4}}) + nO(n^{-3\cdot\frac{3}{4}}) + 1O(n^{-2\cdot\frac{3}{4}})\right) \\
&= O\left(\frac{\|h'''(x)\|_\infty}{n\sqrt{q}}\right),
\end{aligned}
$$

which completes the proof of Theorem 31. ■

We next prove Theorem 10.

Proof [Proof of Theorem 10] Recall that

$$
f_{ij}(\boldsymbol{G}) = \sum_{H \in \mathcal{H}} \hat{f}_H \sum_{\ell \in \mathcal{L}_{H,i,j}} \boldsymbol{\chi}_{\ell(H)}(\boldsymbol{G}).
$$

We do a change of variables and define $g_{ij}(\boldsymbol{\chi}) = f_{ij}(\boldsymbol{G})$. Then g is also a vector-valued polynomial of degree D. And $\Psi(\boldsymbol{\chi}) = \|g(\boldsymbol{\chi}) - g(\boldsymbol{\chi}')\|_2^2$ and $\Psi(\boldsymbol{Z}) = \|g(\boldsymbol{Z}) - g(\boldsymbol{Z}')\|_2^2$, where $\boldsymbol{\chi}$ and $\boldsymbol{\chi}'$ are ρ-correlated and the same for $\boldsymbol{Z}$ and $\boldsymbol{Z}'$. By Theorem 31, we know that

$$
|\mathbf{E}\, h(\|g(\boldsymbol{\chi}) - g(\boldsymbol{\chi}')\|_2^2) - \mathbf{E}\, h(\|g(\boldsymbol{Z}) - g(\boldsymbol{Z}')\|_2^2)| \leqslant O\left(\frac{\|h'''(x)\|_\infty}{n\sqrt{q}}\right), \tag{12}
$$

for bounded h with bounded first three derivatives. We define a smooth function ϕ

$$
\phi(x) = \begin{cases} \exp(-\frac{1}{x}), & \text{if } x > 0, \\ 0, & \text{if } x \leqslant 0. \end{cases}
$$

For any $\delta > 0$, we define a smooth transition function

$$h_\delta(x) = \frac{\phi(x/\delta)}{\phi(1 - x/\delta) + \phi(x/\delta)}.$$

Note that h_δ is smooth, increasing, $h_\delta(x) = 0$ when $x \leqslant 0$, and $h_\delta(x) = 1$ when $x \geqslant \delta$. Moreover, there is a universal constant K such that $|h_\delta'''| \leqslant K/\delta^3$. By Equation (12), this implies that

$$\begin{aligned}
&\mathbf{Pr}\left[\|f(\boldsymbol{G}) - f(\boldsymbol{G}')\|_2^2 \geqslant \gamma\right] = \mathbf{Pr}\left[\|g(\boldsymbol{\chi}) - g(\boldsymbol{\chi}')\|_2^2 \geqslant \gamma\right] \leqslant \mathbf{E}\left[h_\delta\left(\|g(\boldsymbol{\chi}) - g(\boldsymbol{\chi}')\|_2^2 - \gamma + \delta\right)\right]\\
&\leqslant O\left(\frac{1}{n\sqrt{q}\delta^3}\right) + \mathbf{E}\left[h_\delta\left(\|g(\boldsymbol{Z}) - g(\boldsymbol{Z}')\|_2^2 - \gamma + \delta\right)\right]\\
&\leqslant O\left(\frac{1}{n\sqrt{q}\delta^3}\right) + \mathbf{Pr}\left[\|g(\boldsymbol{Z}) - g(\boldsymbol{Z}')\|_2^2 \geqslant \gamma - \delta\right].
\end{aligned}$$

To control the last probability term, we cite the following result

Lemma 35 (Theorem 3.1 in Gamarnik et al. (2024)) *Let $0 \leqslant \rho \leqslant 1$. Let $\boldsymbol{Z}, \boldsymbol{Z}'$ be a pair of standard Gaussian random vectors on $\mathbb{R}^d$ that are ρ-correlated. Let $g : \mathbb{R}^d \to \mathbb{R}^k$ be a (deterministic) polynomial of degree at most D with $\mathbf{E}\,\|g(\boldsymbol{Z})\|_2^2 = 1$. For any $t \geqslant (6e)^D$,*

$$\mathbf{Pr}\left(\|g(\boldsymbol{Z}) - g(\boldsymbol{Z}')\|_2^2 \geqslant 2t\left(1 - \rho^D\right)\right) \leqslant \exp\left(-\frac{D}{3e}t^{1/D}\right).$$

For $\rho \geqslant 1 - \frac{1}{T}$, we take $t = \frac{\gamma}{3(1-\rho^D)}$ and $\delta = \gamma/3$. Then

$$\begin{aligned}
\mathbf{Pr}\left[\|f(\boldsymbol{G}) - f(\boldsymbol{G}')\|_2^2 \geqslant \gamma\right] &\leqslant O\left(\frac{1}{n\sqrt{q}\gamma^3}\right) + \exp\left(-\frac{D}{3e}\left(\frac{\gamma}{3(1-\rho^D)}\right)^{1/D}\right)\\
&\leqslant O\left(\frac{1}{n\sqrt{q}\gamma^3}\right) + \exp\left(-\frac{D}{3e}\left(\frac{\gamma T}{3D}\right)^{1/D}\right).
\end{aligned}$$

This, in combination with our normalization of f, completes the proof of Theorem 10. ■