

ETH-Hardness of Approximating 2-CSPs and Directed Steiner Network

Irit Dinur*
Weizmann Institute of Science

Pasin Manurangsi†
UC Berkeley

May 10, 2018

Abstract

We study the 2-ary constraint satisfaction problems (2-CSPs), which can be stated as follows: given a constraint graph $G = (V, E)$, an alphabet set Σ and, for each edge $\{u, v\} \in E$, a constraint $C_{uv} \subseteq \Sigma \times \Sigma$, the goal is to find an assignment $\sigma : V \rightarrow \Sigma$ that satisfies as many constraints as possible, where a constraint C_{uv} is said to be satisfied by σ if $(\sigma(u), \sigma(v)) \in C_{uv}$.

While the approximability of 2-CSPs is quite well understood when the alphabet size $|\Sigma|$ is constant (see e.g. [Rag08]), many problems are still open when $|\Sigma|$ becomes super constant. One open problem that has received significant attention in the literature is whether it is hard to approximate 2-CSPs to within a polynomial factor of both $|\Sigma|$ and $|V|$ (i.e. $(|\Sigma||V|)^{\Omega(1)}$ factor). As a special case of what is now referred to as the Sliding Scale Conjecture, Bellare et al. [BGLR93] suggested that the answer to this question might be positive. Alas, despite many efforts by researchers to resolve this conjecture (e.g. [RS97, AS97, DFK⁺11, DHK15, Mos17]), it still remains open to this day.

In this work, we separate between $|V|$ and $|\Sigma|$ and ask a closely related but weaker question: is it hard to approximate 2-CSPs to within a polynomial factor of $|V|$ (but while $|\Sigma|$ may be super-polynomial in $|V|$)? Assuming the exponential time hypothesis (ETH), we answer this question positively. Specifically, we show that, unless ETH fails, no polynomial time algorithm can approximate 2-CSPs to within a factor of $|V|^{1-1/(\log|V|)^{1/2-\rho}}$ for every $\rho > 0$. Note that our ratio is not only polynomial but also almost linear. This is almost optimal since a trivial algorithm yields an $O(|V|)$ -approximation for 2-CSPs.

Thanks to a known reduction [DK99, CFM17] from 2-CSPs to the Directed Steiner Network (DSN) problem (aka Directed Steiner Forest), our result implies an inapproximability result for the latter with polynomial ratio in terms of the number of demand pairs. Specifically, assuming ETH, no polynomial time algorithm can approximate DSN to within a factor of $k^{1/4-o(1)}$ where k is the number of demand pairs. The ratio is roughly the square root of the best known approximation ratio achieved by polynomial time algorithms due to Chekuri et al. [CEGS11] and Feldman et al. [FKN12], which yield $O(k^{1/2+\epsilon})$ -approximation for every constant $\epsilon > 0$.

Additionally, if we assume a stronger assumption that there exists $\epsilon > 0$ such that no subexponential time algorithm can distinguish a satisfiable 3-CNF formula from one which is not even $(1 - \epsilon)$ -satisfiable (aka Gap-ETH), then, for 2-CSPs, our reduction not only rules out polynomial time (i.e. $(|V||\Sigma|)^{O(1)}$ time) algorithms, but also fixed parameter tractable (FPT) algorithms parameterized by the number of variables $|V|$. These are algorithms with running time $t(|V|) \cdot (|V||\Sigma|)^{O(1)}$ where t can be any function. Similar improvements also apply for DSN parameterized by the number of demand pairs k .

*Email: irit.dinur@weizmann.ac.il. This work was supported by BSF grant 2014371.

†Email: pasin@berkeley.edu. This work was done while the author was visiting Weizmann Institute of Science.

1 Introduction

We study the 2-ary constraint satisfaction problems (2-CSPs), which can be stated as follows: given a constraint graph $G = (V, E)$, an alphabet set Σ and, for each edge $\{u, v\} \in E$, a constraint $C_{uv} \subseteq \Sigma \times \Sigma$, the goal is to find an assignment $\sigma : V \rightarrow \Sigma$ that satisfies as many constraints as possible, where a constraint C_{uv} is said to be satisfied by σ if $(\sigma(u), \sigma(v)) \in C_{uv}$. Throughout the paper, we use k to denote the number of variables $|V|$, n to denote the the alphabet size $|\Sigma|$, and N to denote the instance size nk .

Constraint satisfaction problems and their inapproximability have been studied extensively since the proof of the PCP theorem in the early 90's [AS98, ALM⁺98]. Most of the effort has been directed towards understanding the approximability of CSPs with constant arity and constant alphabet size, leading to a reasonable if yet incomplete understanding of the landscape [Hås01, Kho02, KKMO07, Rag08, AM09, Cha16]. When the alphabet size grows, the sliding scale conjecture of [BGLR93] predicts that the hardness of approximation ratio will grow as well, and be at least polynomial¹ in the alphabet size n . This has been confirmed for values of n up to $2^{(\log N)^{1-\delta}}$, see [RS97, AS97, DFK⁺11]. Proving the same for n that is polynomial in N is the so-called polynomial sliding scale conjecture and is still quite open. Before we proceed, let us note that the aforementioned results of [RS97, AS97, DFK⁺11] work only for arity strictly larger than two and, hence, do not imply inapproximability for 2-CSPs. We will discuss the special case of 2-CSPs in details below.

The polynomial sliding scale conjecture has been approached from different angles. In [DHK15] the authors try to find the smallest arity and alphabet size such that the hardness factor is polynomial in n , and in [Din16] the conjecture is shown to follow (in some weaker sense) from the Gap-ETH hypothesis, which we discuss in more details later. In this work we focus on yet another angle, which is to separate n and k and ask whether it is hard to approximate constant arity CSPs to within a factor that is polynomial in k (but possibly not polynomial in n). Observe here that obtaining NP-hardness of poly(k) factor is likely to be as hard as obtaining one with poly(N); this is because CSPs can be solved exactly in time $n^{O(k)}$, which means that, unless NP is contained in subexponential time (i.e. $\text{NP} \not\subseteq \bigcap_{\epsilon>0} \text{DTIME}(2^{n^\epsilon})$), NP-hard instances of CSPs must have $k = \text{poly}(N)$.

This motivates us to look for hardness of approximation from assumptions stronger than $\text{P} \neq \text{NP}$. Specifically, our result will be based on the Exponential Time Hypothesis (ETH), which states that no subexponential time algorithm can solve 3-SAT (see Conjecture 5). We show that, unless ETH fails, no polynomial time algorithm can approximate 2-CSPs to within an almost linear ratio in k , as stated below. This is almost optimal since there is a straightforward $(k/2)$ -approximation for any 2-CSP, by simply satisfying all constraints that touch the variable with highest degree.

Theorem 1 (Main Theorem) *Assuming ETH, for any constant $\rho > 0$, no algorithm can, given a 2-CSP instance Γ with alphabet size n and k variables such that the constraint graph is the complete graph on the k variables, distinguish between the following two cases in polynomial time:*

- (Completeness) $\text{val}(\Gamma) = 1$, and
- (Soundness) $\text{val}(\Gamma) < 2^{(\log k)^{1/2+\rho}} / k$.

Here $\text{val}(\Gamma)$ denotes the maximum fraction of edges satisfied by any assignment.

To paint a full picture of how our result stands in comparison to previous results, let us state what is known about the approximability of 2-CSPs; due to the vast literature regarding 2-CSPs, we will focus only the regime of large alphabets which is most relevant to our setting. In terms

¹Throughout the paper, we use *polynomial in x* (or $\text{poly}(x)$) to refer to x^c for some real number $c > 0$.

of NP-hardness, the best known inapproximability ratio is $(\log N)^c$ for every constant $c > 0$; this follows from Moshkovitz-Raz PCP [MR10] and the Parallel Repetition Theorem for the low soundness regime [DS14]. Assuming a slightly weaker assumption that NP is not contained in quasipolynomial time (i.e. $\text{NP} \not\subseteq \bigcup_{c>0} \text{DTIME}(n^{(\log n)^c})$), 2-CSP is hard to approximate to within a factor of $2^{(\log N)^{1-\delta}}$ for every constant $\delta > 0$; this can be proved by applying Raz’s original Parallel Repetition Theorem [Raz98] to the PCP Theorem. In [Din16], the author observed that running time for parallel repetition can be reduced by looking at unordered sets instead of ordered tuples. This observation implies that², assuming ETH, no polynomial time $N^{1/(\log \log \log N)^c}$ -approximation algorithm exists for 2-CSPs for some constant $c > 0$. Moreover, under Gap-ETH (which will be stated shortly), it was shown that, for every sufficiently small $\varepsilon > 0$, an N^ε -approximation algorithm must run in time $N^{\Omega(\exp(1/\varepsilon))}$. Note that, while this latest result comes close to the polynomial sliding scale conjecture, it does not quite resolve the conjecture yet. In particular, even the weak form of the conjecture postulates that there exists $\delta > 0$ for which no polynomial time algorithm can approximate 2-CSPs to within N^δ factor of the optimum. This statement does not follow from the result of [Din16]. Nevertheless, the Gap-ETH-hardness of [Din16] does imply that, for any $f = o(1)$, no polynomial time algorithm can approximate 2-CSPs to within a factor of $N^{f(N)}$.

In all hardness results mentioned above, the constructions give 2-CSP instances in which the alphabet size n is smaller than the number of variables k . In other words, even if we aim for an inapproximability ratio in terms of k instead of N , we still get the same ratios as stated above. Thus, our result is the first hardness of approximation for 2-CSPs with $\text{poly}(k)$ factor. Note again that our result rules out any polynomial time algorithm and not just $N^{O(\exp(1/\varepsilon))}$ -time algorithm ruled out by [Din16]. Moreover, our ratio is almost linear in k whereas the result of [Din16] only holds for ε that is sufficiently small depending on the parameters of the Gap-ETH Hypothesis.

An interesting feature of our reduction is that it produces 2-CSP instances with the alphabet size n that is much larger than k . This is reminiscence of the setting of 2-CSPs parameterized by the number of variables k . In this setting, the algorithm’s running time is allowed to depend not only polynomially on N but also on any function of k (i.e. $g(k) \cdot \text{poly}(N)$ running time for some function g); such algorithm is called a *fixed parameter tractable (FPT)* algorithm parameterized by k . The question here is whether this added running time can help us approximate the problem beyond the $O(k)$ factor achieved by the straightforward algorithm. We show that, even in this parameterized setting, the trivial algorithm is still essentially optimal (up to lower order terms). This result holds under the Gap Exponential Time Hypothesis (Gap-ETH), a strengthening of ETH which states that, for some $\varepsilon > 0$, even distinguishing between a satisfiable 3-CNF formula and one which is not even $(1 - \varepsilon)$ -satisfiable cannot be done in subexponential time (see Conjecture 7), as stated below.

Theorem 2 *Assuming Gap-ETH, for any constant $\rho > 0$ and any function g , no algorithm can, given a 2-CSP instance Γ with alphabet size n and k variables such that the constraint graph is the complete graph on the k variables, distinguish between the following two cases in $g(k) \cdot (nk)^{O(1)}$ time:*

- (Completeness) $\text{val}(\Gamma) = 1$, and,
- (Soundness) $\text{val}(\Gamma) < 2^{(\log k)^{1/2+\rho}} / k$.

To the best of our knowledge, the only previous inapproximability result for parameterized 2-CSPs is from [CFM17]. There the authors showed that, assuming Gap-ETH, no $k^{o(1)}$ -approximation $g(k) \cdot (nk)^{O(1)}$ -time algorithm exists; this is shown via a simple reduction from parameterized inapproximability of Densest- k Subgraph from [CCK⁺17] (which is in turn based on a construction

²In [Din16], only the Gap-ETH-hardness result is stated. However, the ETH-hardness result follows rather easily.

from [Man17]). Our result is a direct improvement over this result.

We end our discussion on 2-CSPs by noting that several approximation algorithms have also been devised for 2-CSPs with large alphabets [Pel07, CHK11, KKT16, MM17, CMMV17]. In particular, while our results suggest that the trivial algorithm achieves an essentially optimal ratio in terms of k , non-trivial approximation is possible when we measure the ratio in terms of N instead of k : specifically, a polynomial time $O(N^{1/3})$ -approximation algorithm is known [CHK11].

Direct Steiner Network. As a corollary of our hardness of approximation results for 2-CSPs, we obtain an inapproximability result for Directed Steiner Network with polynomial ratio in terms of the number of demand pairs. In the Directed Steiner Network (DSN) problem (sometimes referred to as the Directed Steiner Forest problem [FKN12, CDKL17]), we are given an edge-weighted directed graph G and a set \mathcal{D} of k demand pairs $(s_1, t_1), \dots, (s_k, t_k) \in V \times V$ and the goal is to find a subgraph H of G with minimum weight such that there is a path in H from s_i to t_i for every $i \in [k]$. DSN was first studied in the approximation algorithms context by Charikar et al. [CCC⁺99] who gave a polynomial time $\tilde{O}(k^{2/3})$ -approximation algorithm for the problem. This ratio was later improved to $O(k^{1/2+\varepsilon})$ for every $\varepsilon > 0$ by Chekuri et al. [CEGS11]. Later, a different approximation algorithm with similar approximation ratio was proposed by Feldman et al. [FKN12].

Algorithms with approximation ratios in terms of the number of vertices n have also been devised [FKN12, BBM⁺13, CDKL17, AB17]. In this case, the best known algorithm is that of Berman et al. [BBM⁺13], which yields an $O(n^{2/3+\varepsilon})$ -approximation for every constant $\varepsilon > 0$ in polynomial time. Moreover, when the graph is unweighted (i.e. each edge costs the same), Abboud and Bodwin recently gave an improved $O(n^{0.5778})$ -approximation algorithm for the problem [AB17].

On the hardness side, there exists a known reduction from 2-CSP to DSN that preserves approximation ratio to within polynomial factor³ [DK99]. Hence, known hardness of approximation of 2-CSPs translate immediately to that of DSN: it is NP-hard to approximate to within any polylogarithmic ratio [MR10, DS14], it is hard to approximate to within $2^{\log^{1-\varepsilon} n}$ factor for every $\varepsilon > 0$ unless $\text{NP} \subseteq \text{QP}$ [Raz98], and it is Gap-ETH-hard to approximate to within $n^{o(1)}$ factor [Din16]. Note that, since k is always bounded above by n^2 , all these hardness results also hold when n is replaced by k in the ratios. Recently, this reduction was also used by Chitnis et al. [CFM17] to rule out $k^{o(1)}$ -FPT-approximation algorithm for DSN parameterized by k assuming Gap-ETH. Alas, none of these hardness results achieve ratios that are polynomial in either n or k and it remains open whether DSN is hard to approximate to within a factor that is polynomial in n or in k .

By plugging our hardness result for 2-CSPs into the reduction, we immediately get ETH-hardness and Gap-ETH-hardness of approximating DSN to within a factor of $k^{1/4-o(1)}$ as stated below.

Corollary 3 *Assuming ETH, for any constant $\rho' > 0$, there is no polynomial time $\frac{k^{1/4}}{2^{(\log k)^{1/2+\rho'}}$ -approximation algorithm for DSN.*

Corollary 4 *Assuming Gap-ETH, for any constant $\rho' > 0$ and any function g , there is no $g(k) \cdot (nk)^{O(1)}$ -time $\frac{k^{1/4}}{2^{(\log k)^{1/2+\rho'}}$ -approximation algorithm for DSN.*

In other words, if one wants a polynomial time approximation algorithm with ratio depending only on k and not on n , then the algorithms of Chekuri et al. [CEGS11] and Feldman et al. [FKN12]

³That is, for any non-decreasing function ρ , if DSN admits $\rho(nk)$ -approximation in polynomial time, then 2-CSP also admits $\rho(nk)^c$ -approximation polynomial time for some absolute constant c .

are roughly within a square of the optimal algorithm. To the best of our knowledge, these are the first inapproximability results of DSN whose ratios are polynomial in terms of k . Again, Corollary 4 is a direct improvement over the FPT inapproximability result from [CFM17] which, under the same assumption, rules out only $k^{o(1)}$ -factor FPT-approximation algorithm.

Agreement tests

Our main result is proved through an agreement testing argument. In agreement testing there is a universe \mathcal{U} , a collection of subsets $S_1, \dots, S_k \subseteq \mathcal{U}$, and for each subset S_i we are given a local function $f_{S_i} : S_i \rightarrow \{0, 1\}$. A pair of subsets are said to *agree* if their local functions agree on every element in the intersection. The goal is, given a non-negligible fraction of agreeing pairs, to deduce the existence of a global function $g : \mathcal{U} \rightarrow \{0, 1\}$ that (approximately) coincides with many of the local functions. For a more complete description see [DK17].

Agreement tests capture a natural local to global statement and are present in essentially all PCPs, for example they appear explicitly in the line vs. line and plane vs. plane low degree tests [RS96, AS97, RS97]. Our reduction is based on a combinatorial agreement test, where the universe is $[n]$ and the subsets S_1, \dots, S_k have $\Omega(n)$ elements each and are “in general position”, namely they behave like subsets chosen independently at random. A convenient feature about this setting is that every pair of subsets intersect.

Since we are aiming for a large gap, the agreement test must work (i.e., yield a global function) with a very small fraction of agreeing pairs, which in our case is close to $1/k$.

In this small agreement regime the idea, as pioneered in the work of Raz-Safra [RS97], is to zero in on a sub-collection of subsets that is (almost) perfectly consistent. From this sub-collection it is easy to recover a global function and show that it coincides almost perfectly with the local functions in the sub-collection. A major difference between our combinatorial setting and the algebraic setting of Raz-Safra is the lack of “distance” in our case: we can not assume that two distinct local functions differ on many points (in contrast, this is a key feature of low degree polynomials). We overcome this by considering different “strengths” of agreement, depending on the fraction of points on which the two subsets agree. This notion too is present in several previous works on combinatorial agreement tests [IKW12, DN17].

Hardness of Approximation through Subexponential Time Reductions. Our result is one of the many results in recent years that show hardness of approximation via subexponential time reductions [AIM14, BKW15, Rub16b, DFS16, Din16, BKRW17, MR17, Man17, Rub16a, Rub16b, Rub17, ARW17, CCK⁺17, KLM18, Rub18, BGKM18]. These results are often based on the Exponential Time Hypothesis (ETH) and its variants. Proposed by Impagliazzo and Paturi [IP01], ETH can be formally stated as follows:

Conjecture 5 (Exponential Time Hypothesis (ETH) [IP01]) *There exists a constant $\delta > 0$ such that no algorithm can decide whether any given 3-CNF formula is satisfiable in time $O(2^{\delta m})$ where m denotes the number of clauses⁴.*

A crucial ingredient in most, but not all⁵, reductions in this line of work is a nearly-linear size PCP Theorem. For the purpose of our work, the PCP Theorem can be viewed as a polynomial

⁴The original conjecture states the lower bound as exponential in terms of the number of variables not clauses. However, thanks to the sparsification lemma [IPZ01], it is by now known that the two versions are equivalent.

⁵The exceptions are [Rub16b, ARW17, Rub18, KLM18, Che18] in which gaps are not created via the PCP Theorem.

time transformation of a 3-SAT instance $\tilde{\Phi}$ to another 3-SAT instance Φ that creates a gap between the YES and NO cases. Specifically, if $\tilde{\Phi}$ is satisfiable, Φ remains satisfiable. On the other hand, if $\tilde{\Phi}$ is unsatisfiable, then Φ is not only unsatisfiable but it is also not even $(1 - \varepsilon)$ -satisfiable for some constant $\varepsilon > 0$ (i.e. no assignment satisfies $(1 - \varepsilon)$ fraction of clauses). The “nearly-linear size” part refers to the size of the new instance Φ compared to that of $\tilde{\Phi}$. Currently, the best known dependency in this form of the PCP Theorem between the two sizes is quasi-linear (i.e. with a polylogarithmic blow-up), as stated below.

Theorem 6 (Quasi-Linear Size PCP [BS08, Din07]) *For some constants $\varepsilon, \Delta, c > 0$, there is a polynomial time algorithm that, given any 3-CNF formula $\tilde{\Phi}$ with m clauses, produces another 3-CNF formula Φ with $O(m \log^c m)$ clauses such that*

- (Completeness) if $\text{val}(\tilde{\Phi}) = 1$, then $\text{val}(\Phi) = 1$, and,
- (Soundness) if $\text{val}(\tilde{\Phi}) < 1$, then $\text{val}(\Phi) < 1 - \varepsilon$, and,
- (Bounded Degree) each variable in Φ appears in at most Δ clauses.

The aforementioned ETH-hardness of approximation proofs typically proceed in two steps. First, the PCP Theorem is invoked to reduce a 3-SAT instance $\tilde{\Phi}$ of size m to an instance of the gap version of 3-SAT Φ of size $m' = O(m \log^c m)$. Second, the gap version of 3-SAT is reduced in subexponential time to the problem at hand. As long as the reduction takes time $2^{o(m'/\log^c m')} = 2^{o(m)}$, we can obtain hardness of approximation result for the latter problem. This is in contrast to proving NP-hardness of approximation for which a polynomial time reduction is required.

Another related but stronger version of ETH that we will also employ is the Gap Exponential Time Hypothesis (Gap-ETH), which states that even the gap version of 3-SAT cannot be solved in subexponential time:

Conjecture 7 (Gap Exponential Time Hypothesis (Gap-ETH) [Din16, MR16]) *There exist constants $\delta, \varepsilon, \Delta > 0$ such that no algorithm can, given any 3-CNF formula Φ such that each of its variable appears in at most Δ clauses⁶, distinguish between the following two cases⁷ in time $O(2^{\delta m})$ time where m denotes the number of clauses:*

- (Completeness) $\text{val}(\Phi) = 1$.
- (Soundness) $\text{val}(\Phi) < 1 - \varepsilon$.

By starting with Gap-ETH instead of ETH, there is no need to apply the PCP Theorem and hence a polylogarithmic loss in the size of the 3-SAT instance does not occur. As demonstrated in previous works, this allows one to improve the ratio in hardness of approximation results [Din16, MR16, Man17] and, more importantly, it can be used to prove inapproximability results for some parameterized problems [BEKP15, CCK⁺17, CFM17]⁸, which are not known to be hard to approximate under ETH. Specifically, for many parameterized problems, the reduction from the gap version of 3-SAT to the problem has size $2^{m'/f(k)}$ for some function f that grows to infinity with k (i.e. $f = \omega(1)$), where m' is the number of clauses in the 3-CNF formula and k is the parameter of the problem. For simplicity, let us focus on the case where $f(k) = k$. If one wishes to derive a meaningful result starting from ETH, $2^{m'/k}$ must be subexponential in terms of m , the number of clauses in the original (no-gap) 3-CNF formula. This means that the term k must dominate the $\log^c m$ factor blow-up from the PCP Theorem. However, since FPT algorithms are allowed to have running time of the form $g(k)$ for any function g , we can pick g to be 2^{2^k} . In this case, the algorithm runs in

⁶This bounded degree assumption can be assumed without loss of generality; see [MR16] for more details.

⁷Note that when Φ satisfies neither case (i.e. $1 - \varepsilon \leq \text{val}(\Phi) < 1$), the algorithm is allowed to output anything.

⁸While [BEKP15] states that the assumption is the existence of a linear-size PCP, Gap-ETH clearly suffices there.

superexponential time in terms of m and we cannot deduce anything regarding the algorithm. On the other hand, if we start from Gap-ETH, we can pick k to be a large constant independent of m , which indeed yields hardness of the form claimed in Theorem 2 and Corollary 4.

Finally, we remark that Gap-ETH would follow from ETH if a linear-size (constant-query) PCP exists. While constructing short PCPs has long been an active area of research [BGH⁺06, BS08, Din07, MR10, BKK⁺16], no linear-size PCP is yet known. On the other hand, there are some supporting evidences for the hypothesis. For instance, it is known that the natural relaxation of 3-SAT in the Sum-of-Squares hierarchy cannot refute Gap-ETH [Gri01, Sch08]. Moreover, Applebaum recently showed that the hypothesis follows from certain cryptographic assumptions [App17]. For a more in-depth discussion on Gap-ETH, please refer to [Din16].

Organization of the Paper. In the next section, we describe our reduction and give an overview of the proof. Then, in Section 3, we define additional notions and state some preliminaries. We proceed to provide the full proof of our main agreement theorem in Section 4. Using this agreement theorem, we deduce the soundness of our reduction in Section 5. We then plug in the parameters and prove the inapproximability results for 2-CSPs in Section 6. In Section 7, we show how the hardness of approximation result for 2-CSPs imply inapproximability for DSN as well. Finally, we conclude our work with some discussions and open questions in Section 8.

2 Proof Overview

Like other (Gap-)ETH-hardness of approximation results, our proof is based on a subexponential time reduction from the gap version of 3-SAT to our problem of interest, 2-CSPs. Before we describe our reduction, let us define more notations for 2-CSPs and 3-SAT, to facilitate our explanation.

2-CSPs. For notational convenience, we will modify the definition of 2-CSPs slightly so that each variable is allowed to have different alphabets; this definition is clearly equivalent to the more common definition used above. Specifically, an instance Γ of 2-CSP now consists of (1) a constraint graph $G = (V, E)$, (2) for each vertex (or variable) $v \in V$, an alphabet set Σ_v , and, (3) for each edge $\{u, v\} \in E$, a constraint $C_{uv} \subseteq \Sigma_u \times \Sigma_v$. Additionally, to avoid confusion with 3-SAT, we refrain from using the word *assignment* for 2-CSPs and instead use *labeling*, i.e., a labeling of Γ is a tuple $\sigma = (\sigma_v)_{v \in V}$ such that $\sigma_v \in \Sigma_v$ for all $v \in V$. An edge $\{u, v\} \in E$ is said to be *satisfied* by a labeling σ if $(\sigma_u, \sigma_v) \in C_{uv}$. The value of a labeling σ , denoted by $\text{val}(\sigma)$, is defined as the fraction of edges that it satisfies, i.e., $|\{\{u, v\} \in E \mid (\sigma_u, \sigma_v) \in C_{uv}\}| / |E|$. The goal of 2-CSPs is to find σ with maximum value; we denote the such optimal value by $\text{val}(\Gamma)$, i.e., $\text{val}(\Gamma) = \max_{\sigma} \text{val}(\sigma)$.

3-SAT. An instance Φ of 3-SAT consists of a variable set X and a clause set \mathcal{C} where each clause is a disjunction of at most three literals. For any assignment $\psi : X \rightarrow \{0, 1\}$, $\text{val}(\psi)$ denotes the fraction of clauses satisfied by ψ . The goal is to find an assignment ψ that satisfies as many clauses as possible; let $\text{val}(\Phi) = \max_{\psi} \text{val}(\psi)$ denote the fraction of clauses satisfied by such assignment. For each $C \in \mathcal{C}$, we use $\text{var}(C)$ to denote the set of variables whose literals appear in C . We extend this notation naturally to sets of clauses, i.e., for every $T \subseteq \mathcal{C}$, $\text{var}(T) = \bigcup_{C \in T} \text{var}(C)$.

Our Construction

Before we state our reduction, let us again reiterate the objective of our reduction. Roughly speaking, given a 3-SAT stance $\Phi = (X, \mathcal{C})$, we would like to produce a 2-CSP instance Γ_{Φ} such that

- (Completeness) If $\text{val}(\Phi) = 1$, then $\text{val}(\Gamma_{\Phi}) = 1$,

- (Soundness) If $\text{val}(\Phi) < 1 - \varepsilon$, then $\text{val}(\Gamma_\Phi) < k^{o(1)}/k$ where k is number of variables of Γ_Φ ,
- (Reduction Time) The time it takes to produce Γ_Φ should be $2^{o(m)}$ where $m = |\mathcal{C}|$,

where $\varepsilon > 0$ is some absolute constant.

Observe that, when plugging a reduction with these properties to Gap-ETH, we directly arrive at the claimed $k^{1-o(1)}$ inapproximability for 2-CSPs. However, for ETH, since we start with a decision version of 3-SAT without any gap, we have to first invoke the PCP theorem to produce an instance of the gap version of 3-SAT before we can apply our reduction. Since the shortest known PCP has a polylogarithmic blow-up in the size (see Theorem 6), the running time lower bound for gap 3-SAT will not be exponential anymore, rather it will be of the form $2^{\Omega(m/\text{polylog}m)}$ instead. Hence, our reduction will need to produce Γ_Φ in $2^{o(m/\text{polylog}m)}$ time. As we shall see later in Section 6, this will also be possible with appropriate settings of parameters.

With the desired properties in place, we now move on to state our reduction. In addition to a 3-CNF formula Φ , the reduction also takes in a collection \mathcal{T} of subsets of clauses of Φ . For now, the readers should think of the subsets in \mathcal{T} as random subsets of \mathcal{C} where each element is included in each subset independently at random with probability α , which will be specified later. As we will see below, we only need two simple properties that the subsets in \mathcal{T} are “well-behaved” enough and we will later give a deterministic construction of such well-behaved subsets. With this in mind, our reduction can be formally described as follows.

Definition 8 (The Reduction) *Given a 3-CNF formula $\Phi = (X, \mathcal{C})$ and a collection \mathcal{T} of subsets of \mathcal{C} , we define a 2-CSP instance $\Gamma_{\Phi, \mathcal{T}} = (G = (V, E), \Sigma, \{C_{uv}\}_{\{u,v\} \in E})$ as follows:*

- The graph G is the complete graph where the vertex set is \mathcal{T} , i.e., $V = \mathcal{T}$ and $E = \binom{\mathcal{T}}{2}$.
- For each $T \in \mathcal{T}$, the alphabet set Σ_T is the set of all partial assignments to $\text{var}(T)$ that satisfies every clause in T , i.e., $\Sigma_T = \{\psi_T : \text{var}(T) \rightarrow \{0, 1\} \mid \forall C \in T, \psi_T \text{ satisfies } C\}$.
- For every $T_1 \neq T_2 \in \mathcal{T}$, (ψ_{T_1}, ψ_{T_2}) is included in $C_{T_1 T_2}$ if and only if they are consistent, i.e., $C_{T_1 T_2} = \{(\psi_{T_1}, \psi_{T_2}) \in \Sigma_{T_1} \times \Sigma_{T_2} \mid \forall x \in \text{var}(T_1) \cap \text{var}(T_2), \psi_{T_1}(x) = \psi_{T_2}(x)\}$.

Let us now examine the properties of the reduction. The number of vertices in $\Gamma_{\Phi, \mathcal{T}}$ is $k = |\mathcal{T}|$. For the purpose of the proof overview, α should be thought of as $1/\text{polylog}m$ whereas k should be thought of as much larger than $1/\alpha$ (e.g. $k = \exp(1/\alpha)$). For such value of k , all random sets in \mathcal{T} will have size $O(\alpha m)$ w.h.p., meaning that the reduction time is $2^{m/\text{polylog}m}$ as desired.

Moreover, when Φ is satisfiable, it is not hard to see that $\text{val}(\Gamma_{\Phi, \mathcal{T}}) = 1$; more specifically, if $\psi : X \rightarrow \{0, 1\}$ is the assignment that satisfies every clause of Φ , then we can label each vertex $T \in \mathcal{T}$ of $\Gamma_{\Phi, \mathcal{T}}$ by $\psi|_{\text{var}(T)}$, the restriction of ψ on $\text{var}(T)$. Since ψ satisfies all the clauses, $\psi|_{\text{var}(T)}$ satisfies all clauses in T , meaning that this is a valid labeling. Moreover, since these are restrictions of the same global assignment ψ , they are all consistent and every edge is satisfied.

Hence, we are only left to show that, if $\text{val}(\Phi) < 1 - \varepsilon$, then $\text{val}(\Gamma_{\Phi, \mathcal{T}}) < k^{o(1)}/k$; this is indeed our main technical contribution. We will show this by contrapositive: assuming that $\text{val}(\Gamma_{\Phi, \mathcal{T}}) \geq k^{o(1)}/k$, we will “decode” back an assignment to Φ that satisfies $1 - \varepsilon$ fraction of clauses.

2.1 Soundness Analysis as an Agreement Theorem

Our task at hand can be viewed as agreement testing. Informally, in agreement testing, the input is a collection $\{f_S\}_{S \in \mathcal{S}}$ of local functions $f_S : S \rightarrow \{0, 1\}$ where \mathcal{S} is a collection of subsets of some universe \mathcal{U} such that, for many pairs S_1 and S_2 , f_{S_1} and f_{S_2} agree, i.e., $f_{S_1}(x) = f_{S_2}(x)$ for all $x \in S_1 \cap S_2$. An agreement theorem says that there must be a global function $g : \mathcal{U} \rightarrow \{0, 1\}$

that coincides (exactly or approximately) with many of the local functions, and thus explains the pairwise “local” agreements. In our case, a labeling $\sigma = \{\sigma_T\}_{T \in \mathcal{T}}$ with high value is exactly a collection of functions $\sigma_T : \text{var}(T) \rightarrow \{0, 1\}$ such that, for many pairs of T_1 and T_2 , σ_{T_1} and σ_{T_2} agrees. The heart of our soundness proof is an agreement theorem that recovers a global function $\psi : X \rightarrow \{0, 1\}$ that approximately coincides with many of the local functions σ_T 's and thus satisfies $1 - \varepsilon$ fraction of clauses of Φ . To discuss the agreement theorem in more details, let us define several additional notations, starting with those for (approximate) agreements of a pair of functions:

Definition 9 For any universe \mathcal{U} , let $f_{S_1} : S_1 \rightarrow \{0, 1\}$ and $f_{S_2} : S_2 \rightarrow \{0, 1\}$ be any two functions whose domains S_1, S_2 are subsets of \mathcal{U} . We use the following notations for (dis)agreements of these two functions:

- Let $\text{disagr}(f_{S_1}, f_{S_2})$ denote the number of $x \in S_1 \cap S_2$ that f_{S_1} and f_{S_2} disagree on, i.e., $\text{disagr}(f_{S_1}, f_{S_2}) = |\{x \in S_1 \cap S_2 \mid f_{S_1}(x) \neq f_{S_2}(x)\}|$.
- For any $\zeta \geq 0$, we say that f_{S_1} and f_{S_2} are ζ -consistent if $\text{disagr}(f_{S_1}, f_{S_2}) \leq \zeta|\mathcal{U}|$, and we say that the two functions are ζ -inconsistent otherwise. For $\zeta = 0$, we sometimes drop 0 and refer to these simply as consistent and inconsistent (instead of 0-consistent and 0-inconsistent).
- We use $f_{S_1} \stackrel{\zeta}{\approx} f_{S_2}$ and $f_{S_1} \not\stackrel{\zeta}{\approx} f_{S_2}$ as shorthands for ζ -consistency and ζ -inconsistency respectively. Again, for $\zeta = 0$, we may drop 0 from the notations and simply use $f_{S_1} \approx f_{S_2}$ and $f_{S_1} \not\approx f_{S_2}$.

Next, we define the notion of agreement probability for any collection of functions:

Definition 10 For any $\zeta \geq 0$ and any collection $\mathcal{F} = \{f_S\}_{S \in \mathcal{S}}$ of functions, the ζ -agreement probability, denoted by $\text{agree}_\zeta(\mathcal{F})$ is the probability that f_S is ζ -consistent with $f_{S'}$ where S and S' are chosen independently uniformly at random from \mathcal{S} , i.e., $\text{agree}_\zeta(\mathcal{F}) = \Pr_{S, S' \in \mathcal{S}}[f_S \stackrel{\zeta}{\approx} f_{S'}]$. When $\zeta = 0$, we will drop 0 from the notation and simply use $\text{agree}(\mathcal{F})$.

Our main agreement theorem, which works when each $S \in \mathcal{S}$ is a large “random” subset, says that, if $\text{agree}(\mathcal{F})$ is noticeably large, then there exists a global function that is approximately consistent with many of the local functions in \mathcal{F} . This is stated more precisely (but still informally) below.

Theorem 11 (Informal; See Theorem 21) Let \mathcal{S} be a collection of k independent random αn -element subsets of $[n]$. The following holds with high probability: for any $\beta > 0$ and any collection of functions $\mathcal{F} = \{f_S\}_{S \in \mathcal{S}}$ such that $\delta \triangleq \text{agree}(\mathcal{F}) \geq k^{\alpha\beta\alpha(1)}/k$, there exist a function $g : [n] \rightarrow \{0, 1\}$ and a subcollection \mathcal{S}' of size $\delta k^{1-\alpha\beta\alpha(1)}$ such that $g \stackrel{\beta}{\approx} f_{S'}$ for all $S' \in \mathcal{S}'$.

To see that Theorem 11 implies our soundness, let us view a labeling $\sigma = \{\sigma_T\}_{T \in \mathcal{T}}$ as a collection $\mathcal{F} = \{f_S\}_{S \in \mathcal{S}}$ where $\mathcal{S} = \{\text{var}(T) \mid T \in \mathcal{T}\}$ and $f_{\text{var}(T)}$ is simply σ_T . Now, when $\text{val}(\sigma)$ is large, $\text{agree}(\mathcal{F})$ is large as well. Moreover, while the sets $S \in \mathcal{S}$ are not random subsets of variables but rather variable sets of random subsets of clauses, it turns out that these sets are “well-behaved” enough for us to apply Theorem 11. This yields a global function $\psi : X \rightarrow \{0, 1\}$ that are β -consistent with many σ_T 's. Note that, if instead of β -consistency we had exact consistency, then we would have been done because ψ must satisfy all clauses that appear in any T such that ψ is consistent with σ_T ; since there are many such T 's and these are random sets, ψ indeed satisfies almost all clauses. A simple counting argument shows that this remains true even with approximate consistency, provided that most clauses appear in at least a certain fraction of such T 's (an assumption which holds for random subsets). Hence, the soundness of our reduction follows from Theorem 11, and we devote the rest of this section to outline an overview of its proof.

Optimality of the parameters of Theorem 11. Before we proceed to the overview, we would like

to note that the size of the subcollection \mathcal{S}' in Theorem 11 is nearly optimal. This is because, we can partition \mathcal{S} into $1/\delta$ subcollections $\mathcal{S}_1, \dots, \mathcal{S}_{1/\delta}$ each of size δk and, for each $i \in [1/\delta]$, randomly select a global function $g_i : [n] \rightarrow \{0, 1\}$ and let each f_S be the restriction of g_i to S for each $S \in \mathcal{S}_i$. In this way, we have $\text{agree}(\mathcal{F}) \geq \delta k$ and any global function can be (approximately) consistent with at most δk local functions. This means that \mathcal{S}' can be of size at most δk in this case and, up to a $k^{o_{\beta, \alpha}(1)}$ multiplicative factor, Theorem 11 yields almost a largest possible \mathcal{S}' .

2.2 A Simplified Proof: $\delta \geq k^{o(1)}/k^{1/2}$ Regime

We now sketch the proof of Theorem 11. Before we describe how we can find g when $\delta \geq k^{o_{\beta, \alpha}(1)}/k$, let us sketch the proof assuming a stronger assumption that $\delta \geq \Theta_{\alpha, \beta}(1)/k^{1/2}$. Note that this simplified proof already implies a $k^{1/2-o(1)}$ factor ETH-hardness of approximating 2-CSPs. In the next subsection, we will then proceed to refine the arguments to handle smaller values of δ .

Let us consider the *consistency graph* of \mathcal{F} . This is the graph $G^{\mathcal{F}}$ whose vertex set is \mathcal{S} and there is an edge between S_1 and S_2 if and only if f_{S_1} and f_{S_2} are consistent. Note that the number of edges in $G^{\mathcal{F}}$ is equal to $\frac{k^2\delta-k}{2}$, where the subtraction of k comes from the fact that $\delta = \text{agree}(\mathcal{F})$ includes the agreement of each set and itself (whereas $G^{\mathcal{F}}$ does not).

Previous works on agreement testers exploit particular structures of the consistency graph to decode a global function. One such property that is relevant to our proof is the notion of *almost transitivity* defined by Raz and Safra in the analysis of their test [RS97]. More specifically, a graph $G = (V, E)$ is said to be q -transitive for some $q > 0$ if, for every non-edge $\{u, v\}$ (i.e. $\{u, v\} \in \binom{V}{2} \setminus E$), u and v can share at most q common neighbors⁹. Raz and Safra showed that their consistency graph is $(k^{1-\Omega(1)})$ -transitive where k denotes the number of vertices of the graph. They then proved a generic theorem regarding $(k^{1-\Omega(1)})$ -transitive graphs that, for any such graph, its vertex set can be partitioned so that the subgraph induced by each partition is a clique and that the number of edges between different partitions is small. Since a sufficiently large clique corresponds to a global function in their setting, they can then immediately deduce that their result.

Observe that, in our setting, a large clique also corresponds to a global function that is consistent with many local functions. In particular, suppose that there exists $\mathcal{S}' \subseteq \mathcal{S}$ of size sufficiently large such that \mathcal{S}' induces a clique in $G^{\mathcal{F}}$. Since $f_{S'}$'s are perfectly consistent with each other for all $S' \in \mathcal{S}'$, there is a global function $g : [n] \rightarrow \{0, 1\}$ that is consistent with all such $f_{S'}$'s. Hence, if we could show that our consistency graph $G^{\mathcal{F}}$ is $(k^{1-\Omega(1)})$ -transitive, then we could use the same argument as Raz and Safra's to deduce our desired result. Alas, our graph $G^{\mathcal{F}}$ does not necessarily satisfy this transitivity property; for instance, consider any two sets $S_1, S_2 \in \mathcal{S}$ and let f_{S_1}, f_{S_2} be such that they disagree on only one variable, i.e., there is a unique $x \in S_1 \cap S_2$ such that $f_{S_1}(x) \neq f_{S_2}(x)$. It is possible that, for every $S \in \mathcal{S}$ that does not contain x , f_S agrees with both f_{S_1} and f_{S_2} ; in other words, every such S can be a common neighbor of S_1 and S_2 . Since each variable x appears roughly in only $\Theta(\alpha)$ fraction of the sets, there can be as many as $(1 - \Theta(\alpha))k = (1 - o(1))k$ common neighbors of S_1 and S_2 even when there is no edge between S_1 and S_2 !

Fortunately for us, a weaker statement holds: if f_{S_1} and f_{S_2} disagree on more than ζn variables (instead of just one variable as above), then S_1 and S_2 have at most $O(\ln(1/\zeta)/\alpha)$ common neighbors in $G^{\mathcal{F}}$. Here ζ should be thought of as β^2 times a small constant which will be specified later. To see why this statement holds, observe that, since every $S \in \mathcal{S}$ is a random subset that includes each

⁹In [RS97], the transitivity parameter q is used to denote the *fraction* of vertices that are neighbors of both u and v rather than the *number* of such vertices as defined here. However, the latter notion will be more convenient for us.

clause $x \in [n]$ with probability α , Chernoff bound implies that, for every subcollection $\tilde{\mathcal{S}} \subseteq \mathcal{S}$ of size $\Omega(\ln(1/\zeta)/\alpha)$, $\bigcup_{S \in \tilde{\mathcal{S}}} S$ contains all but $O(\zeta)$ fraction of variables. Let $\tilde{\mathcal{S}}_{S_1, S_2} \subseteq \mathcal{S}$ denote the set of common neighbors of S_1 and S_2 . It is easy to see that S_1 and S_2 can only disagree on variables that do not appear in $\bigcup_{S \in \tilde{\mathcal{S}}_{S_1, S_2}} S$. If $\tilde{\mathcal{S}}_{S_1, S_2}$ is of size $\Omega(\ln(1/\zeta)/\alpha)$, then $\bigcup_{S \in \tilde{\mathcal{S}}_{S_1, S_2}} S$ contains all but $O(\zeta)$ fraction of variables, which means that S_1 and S_2 disagrees only on $O(\zeta)$ fraction of variables. By selecting the constant appropriately inside $O(\cdot)$, we arrive at the claim statement.

In other words, while the transitive property does not hold for every edge, it holds for the edges $\{S_1, S_2\}$ where f_{S_1} and f_{S_2} are ζ -inconsistent. This motivates us to define a two-level consistency graph, where the edges with ζ -inconsistent are referred to as the *red* edges whereas the original edges in $G^{\mathcal{F}}$ is now referred to as the *blue* edges. We define this formally below.

Definition 12 (Red/blue Graph) A red-blue graph is an undirected graph $G = (V, E = E_r \cup E_b)$ where its edge set E is partitioned into two sets E_r , the set of red edges, and E_b , the set of blue edges. We use the prefixes “blue-” and “red-” to refer to the quantities of the graph (V, E_b) and (V, E_r) respectively; for instance, u is said to be a blue-neighbor of v if $\{u, v\} \in E_b$.

Definition 13 (Two-Level Consistency Graph) Given a collection of functions $\mathcal{F} = \{f_S\}_{S \in \mathcal{S}}$ and a real number $0 \leq \zeta \leq 1$, the two-level consistency graph $G^{\mathcal{F}, \zeta} = (V^{\mathcal{F}, \zeta}, E_r^{\mathcal{F}, \zeta} \cup E_b^{\mathcal{F}, \zeta})$ is a red-blue graph defined as follows.

- The vertex set $V^{\mathcal{F}, \zeta}$ is simply \mathcal{S} .
- The blue edges are the consistent pairs $\{S_1, S_2\}$, i.e., $E_b = \{\{S_1, S_2\} \in \binom{\mathcal{S}}{2} \mid f_{S_1} \approx f_{S_2}\}$.
- The red edges are the ζ -inconsistent pairs $\{S_1, S_2\}$, i.e., $E_r = \{\{S_1, S_2\} \in \binom{\mathcal{S}}{2} \mid f_{S_1} \not\approx_{\zeta} f_{S_2}\}$.

Note that S_1, S_2 constitute neither a blue nor a red edge when $0 < \text{disagr}(f_{S_1}, f_{S_2}) \leq \zeta n$.

Now, the transitivity property we argue above can be stated as follows: for every red-edge $\{S_1, S_2\}$ of $G^{\mathcal{F}, \zeta}$, there are at most $O(\ln(1/\zeta)/\alpha)$ different S 's such that both $\{S, S_1\}$ and $\{S, S_2\}$ are blue edges. For brevity, let us call any red-blue graph $G = (V, E_r \cup E_b)$ *q-red/blue-transitive* if, for every red edge $\{u, v\} \in E_r$, u and v have at most q common blue-neighbors. We will now argue that in any *q-red/blue-transitive* of average blue-degree d , there exists a subset $U \subseteq V$ of size $\Omega(d)$ such that only $O(qk/d^2)$ fraction of pairs of vertices in U are red edges.

Before we prove this, let us state why this is useful for decoding the desired global function g . Observe that such a subset U of vertices in the two-level consistency graph translates to a subcollection $\mathcal{S}' \subseteq \mathcal{S}$ such that, for all but $O(qk/d^2)$ fraction of pairs of sets $S_1, S_2 \subseteq \mathcal{S}'$, $\{S_1, S_2\}$ does not form a red edge. Recall from definition of red edges that, for such S_1, S_2 , f_{S_1} and f_{S_2} disagrees on at most ζn variables. In other words, \mathcal{S}' is similar to a clique in the (not two-level) consistency graph, except that (1) $O(qk/d^2)$ fraction of pairs $\{S_1, S_2\}$ are allowed to disagree on as many variables as they like, and (2) even for the rest of pairs, the guarantee now is that they agree on all but at most ζn variables, instead of total agreement as in the previous case of clique. Fortunately, this still suffices to find g that is $O(\sqrt{qk/d^2} + \zeta)$ -consistent with $\Omega(d)$ functions. One way construct such a global function is to simply assign each $g(x)$ according to the majority of $f_S(x)$ for all $S \in \mathcal{S}'$ such that $x \in S$. (This is formalized in Section 4.3.) Note that in our case $q = O(\ln(1/\zeta)/\alpha)$ and $d = \Omega(\delta k)$. Hence, if we pick $\zeta \ll \beta^2$ and $\delta \gg (q^{1/2}/\beta)/k^{1/2} = O_{\beta, \alpha}(1)/k^{1/2}$, we indeed get a global function g that is β -consistent with $\Omega(\delta k)$ local functions.

We now move on to sketch how one can find such an “almost non-red subgraph”. For simplicity, let us assume that every vertex has the same blue-degree (i.e. (V, E_b) is d -regular). Let us count the

number of *red-blue-blue triangle* (or *rbb triangle*), which is a 3-tuple (u, v, w) of vertices in V such that $\{u, v\}, \{v, w\}$ are blue edges whereas $\{u, w\}$ is a red edge. An illustration of a rbb triangle can be found in Figure 1a. The red/blue transitivity can be used to bound the number of rbb triangles as follows. For each $(u^*, w^*) \in V^2$, since the graph is q -red/blue-transitive there are at most q rbb triangle with $u = u^*$ and $w = w^*$. Hence, in total, there can be at most qk^2 rbb triangles. As a result, there exists $v^* \in V$ such that the number of rbb triangles (u, v, w) such that $v = v^*$ is at most qk . Let us now consider the set $U = N_b(v^*)$ that consists of all blue-neighbors of v^* . There can be at most qk red edges with both endpoints in $N_b(v^*)$ because each such edge corresponds to a rbb triangle with $v = v^*$. From our assumption that every vertex has blue degree d , we indeed have that $|U| = d$ and that the fraction of pairs of vertices in U that are linked by red edges is $O(qk/d^2)$ as desired. This completes our overview for the case $\delta \geq \Theta_{\beta, \alpha}(1)/k^{1/2}$.

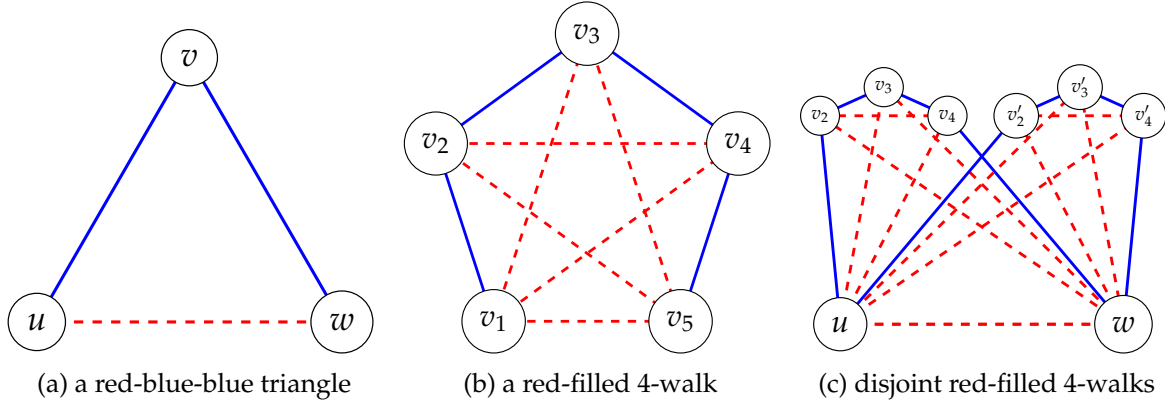


Figure 1: Illustrations of red-filled walks. The red edges are represented by red dashed lines whereas the blue edges are represented by blue solid lines. Figure 1a and Figure 1b demonstrate a red-filled 2 walk (aka rbb triangle) and a red-filled 4-walk respectively. Figure 1c shows two disjoint red-filled 4-walks.

2.3 Towards $\delta = k^{o(1)}/k$ Regime

To handle smaller δ , we need to first understand why the approach above fails to work for $\delta \leq 1/k^{1/2}$. To do so, note that the above proof sketch can be summarized into three main steps:

- (1) Show that the two-level consistency graph $G^{\mathcal{F}}$ is q -red/blue-transitive for some $q = k^{o(1)}$.
- (2) Use red/blue transitivity to find a large subgraph of $G^{\mathcal{F}}$ with few induced red edges.
- (3) Decode a global function from such an “almost non-red subgraph”.

The reason that we need $\delta \gg 1/k^{1/2}$, or equivalently $d \gg k^{1/2}$, lies in Step (2). Although not stated as such earlier, our argument in this step can be described as follows. We consider all length-2 blue-walks, i.e., all $(u, v, w) \in V^3$ such that $\{u, v\}$ and $\{v, w\}$ are both blue edges, and, using the red/blue transitivity of the graph, we argue that, for almost of all these walks, $\{u, w\}$ is not a red edge (i.e. (u, v, w) is not a rbb triangle), which then allows us to find an almost non-red subgraph. For this argument to work, we need the number of length-2 blue-walks to far exceed the number of rbb triangles. The former is kd^2 whereas the latter is bounded above by k^2q in q -red/blue-transitive graphs. This means that we need $kd^2 \gg k^2q$, which implies that $d \gg k^{1/2}$.

To overcome this limitation, we instead consider all length- ℓ blue-walks for $\ell > 2$ and we will define a “rbb-triangle-like” structure on these walks. Our goal is again to show that this structure appears rarely in random length- ℓ blue-walks and we will then use this to find a subgraph that

allows us to decode a good assignment for Φ . Observe that the number of length- ℓ blue walks is kd^ℓ . We also hope that the number of “rbb-triangle-like” structures is still small; in particular, we will still get a similar bound $k^{2+o(1)}$ for such generalized structure, similar to our previous bound for the red-blue-blue triangles. When this is the case, we need $kd^\ell \geq k^{2+o(1)}$, meaning that when $\ell = \omega(1)$ it suffices to select $d = k^{o(1)}$, which yields $k^{1-o(1)}$ factor inapproximability as desired. To facilitate our discussion, let us define notations for ℓ -walks here.

Definition 14 (ℓ -Walks) For any red/blue graph $G = (V, E_r \cup E_b)$ and any integer $\ell \geq 2$, an ℓ -blue-walk in G is an $(\ell + 1)$ -tuple of vertices $(v_1, v_2, \dots, v_{\ell+1}) \in V^{\ell+1}$ such that every pair of consecutive vertices are joined by a blue edge, i.e., $\{v_i, v_{i+1}\} \in E_b$ for every $i \in [\ell]$. For brevity, we sometimes refer to ℓ -blue walks simply as ℓ -walks. We use \mathcal{W}_ℓ^G to denote the set of all ℓ -walks in G .

Note here that a vertex can appear multiple times in a single ℓ -walk.

One detail we have yet to specify in the proof is the structure that generalizes the rbb triangle for ℓ -walks where $\ell > 2$. Like before, this structure will enforce the two end points of the walk to be joined by a red edge, i.e., $\{v_1, v_{\ell+1}\} \in E_r$. Additionally, we require every pair of non-consecutive vertices to be joined by a red edge. We call such a walk a *red-filled ℓ -walk* (see Figure 1b):

Definition 15 (Red-Filled ℓ -Walks) For any red/blue graph $G = (V, E_r \cup E_b)$, a red-filled ℓ -walk is an ℓ -walk $(v_1, v_2, \dots, v_{\ell+1})$ such that every pair of non-consecutive vertices is joined by a red edge, i.e., $\{v_i, v_j\} \in E_r$ for every $i, j \in [\ell + 1]$ such that $j > i + 1$. Let $\widehat{\mathcal{W}}_\ell^G$ denote the set of all red-filled ℓ -walks in G . Moreover, for every $u, v \in V$, let $\widehat{\mathcal{W}}_\ell^G(u, v)$ denote the set of all red-filled ℓ -walks from u to v , i.e., $\mathcal{W}_\ell^G(u, v) = \{(v_1, \dots, v_{\ell+1}) \in \widehat{\mathcal{W}}_\ell^G \mid v_1 = u \wedge v_{\ell+1} = v\}$.

As mentioned earlier, we will need a generalized transitivity property that works not only for rbb triangles but also for our new structure, i.e. the red-filled ℓ -walks. This can be defined analogously to q -red/blue transitivity as follows.

Definition 16 ((q, ℓ) -Red/Blue Transitivity) For any positive integers $q, \ell \in \mathbb{N}$, a red/blue graph $G = (V, E_r \cup E_b)$ is said to be (q, ℓ) -red/blue-transitive if, for every pair of vertices $u, v \in V$ that are joined by a red edge, there exists at most q red-filled ℓ -walks starting at u and ending at v , i.e., $|\widehat{\mathcal{W}}_\ell^G(u, v)| \leq q$.

Using a similar argument to before, we can show that, when \mathcal{S} consists of random subsets where each element is included in a subset with probability α , the two-level agreement graph is (q, ℓ) -red/blue transitive for some parameter q that is a function of only α and ℓ . When $1/\alpha$ and ℓ are small enough in terms of k , q can be made to be $k^{o(1)}$. (The full proof can be found in Section 4.1.1.)

Once this is proved, it is not hard (using a similar argument as before) to show that, when $d \gg (kq)^{1/\ell}$, most ℓ -walks are not red-filled, i.e., $|\mathcal{W}_\ell^G| \gg |\widehat{\mathcal{W}}_\ell^G|$. Even with this, it is still unclear how we can get back a “clique-like” subgraph; in the case of $\ell = 2$ above, this implies that a blue-neighborhood induces few red edges, but the argument does not seem to generalize to larger ℓ . Fortunately, it is still quite easy to find a large subgraph that a non-trivial fraction of pairs of vertices do *not* form red edges; specifically, we will find two subsets $U_1, U_2 \subseteq V$ each of size d such that for at least $1/\ell^2$ fraction of $(u_1, u_2) \in U_1 \times U_2$, $\{u_1, u_2\}$ is not a red edge. To find such sets, observe that, if $|\mathcal{W}_\ell^G| \geq 2|\widehat{\mathcal{W}}_\ell^G|$, then for a random $(v_1, \dots, v_{\ell+1}) \in \mathcal{W}_\ell^G$ the probability that there exists non-consecutive vertex v_i, v_j in the walk that are joined by a red edge is at least $1/2$. Since there are less than $\ell^2/2$ such i, j , union bound implies that there must be non-consecutive i^*, j^* such that the probability that v_{i^*}, v_{j^*} are not joined by a red edge is at least $1/\ell^2$. Let us assume without loss of generality that $i^* < j^*$; since they are not consecutive, we have $i^* + 1 < j^*$.

Let us consider v_{i^*+1}, v_{j^*-1} . By a simple averaging argument, there must be u^* and w^* such that, conditioning on $v_{i^*+1} = u^*$ and $v_{j^*-1} = w^*$, the probability that $\{v_{i^*}, v_{j^*}\} \notin E_r$ is at least $1/\ell^2$. However, this conditional probability is exactly equal to fraction of $(u_1, u_2) \in N_b(u^*) \times N_b(w^*)$ that u_1 and u_2 are not joined by a red edge. Recall again that $N_b(v)$ is used to denote the set of all blue-neighbors of v . Thus, $U_1 = N_b(u^*)$ and $U_2 = N_b(w^*)$ are the sets with desired property.

We are still not done yet since we have to use these sets to decode back the global function g . This is still not obvious: the guarantee we have for our sets U_1, U_2 is rather weak since we only know that at least $1/\ell^2$ of the pairs of vertices from the two sets do not form red edges. This is in contrast to the $\ell = 2$ case where we have a subgraph such that almost all induced edges are *not* red.

To see how to overcome this barrier, recall that a pair S_1, S_2 that does not form a red edge corresponds to $f_{S_1} \stackrel{\zeta}{\approx} f_{S_2}$. As a thought experiment, let us think of the following scenario: if instead of just ζ -consistency, these pairs satisfy (exact) consistency, then we can consider the collection $\tilde{\mathcal{F}} = \{f_S\}_{S \in \tilde{U}}$ where $\tilde{U} = U_1 \cup U_2$. This is a collection of $\Theta(d)$ local functions such that $\text{agree}(\tilde{\mathcal{F}}) \geq \Omega(1/\ell^2)$. Thus, when $d \gg \ell^4$, we are in the regime where $\text{agree}(\tilde{\mathcal{F}}) \gg 1/d^{1/2}$, meaning that we can apply our earlier argument (for the $\delta \geq k^{o(1)}/k^{1/2}$ regime) to recover g !

The approach in the previous paragraph of course does not work directly because we only know that $\Omega(1/\ell^2)$ fraction of the pairs $\{S_1, S_2\} \subseteq \tilde{U}$ are ζ -consistent, not exactly consistent. However, we can still try to mimic the proof in the regime $\delta \geq k^{o(1)}/k^{1/2}$ and define a red/blue graph in such a way that such ζ -consistent pairs are now blue edges. Naturally, the red edges will now be the ζ' -inconsistent pairs for some $\zeta' > \zeta$. In other words, we consider the *generalized two-level consistency graph* defined as follows.

Definition 17 (Generalized Two-Level Consistency Graph) *Given a collection of functions $\mathcal{F} = \{f_S\}_{S \in \mathcal{S}}$ and two real numbers $0 \leq \zeta \leq \zeta' \leq 1$, the generalized two-level consistency graph $G^{\mathcal{F}, \zeta, \zeta'} = (V^{\mathcal{F}, \zeta, \zeta'}, E_r^{\mathcal{F}, \zeta, \zeta'} \cup E_b^{\mathcal{F}, \zeta, \zeta'})$ is a red/blue graph defined as follows.*

- The vertex set $V^{\mathcal{F}, \zeta, \zeta'}$ is simply \mathcal{S} .
- The blue edges are the ζ -consistent pairs $\{S_1, S_2\}$, i.e., $E_b^{\mathcal{F}, \zeta, \zeta'} = \{\{S_1, S_2\} \in \binom{\mathcal{S}}{2} \mid f_{S_1} \stackrel{\zeta}{\approx} f_{S_2}\}$.
- The red edges are the ζ' -inconsistent pairs $\{S_1, S_2\}$, i.e., $E_r^{\mathcal{F}, \zeta, \zeta'} = \{\{S_1, S_2\} \in \binom{\mathcal{S}}{2} \mid f_{S_1} \not\stackrel{\zeta'}{\approx} f_{S_2}\}$.

As its name suggests, the generalized two-level consistency graph is a generalization of the two-level consistency graph from Definition 13; namely $G^{\mathcal{F}, 0, \zeta}$ in the more general definition coincides with $G^{\mathcal{F}, \zeta}$ in the original definition.

Now, it is not hard to show that when $\zeta' \gg \zeta/\alpha$, the graph $G^{\mathcal{F}, \zeta, \zeta'}$ is again q -red/blue transitive for some q that depends only on α and ζ . This means that we can apply our argument from the $\delta \geq 1/k^{1/2-o(1)}$ regime on the graph $G^{\mathcal{F}, \zeta, \zeta'}$, which yields a subset $U \subseteq \tilde{U}$ such that almost all pairs $\{S_1, S_2\} \subseteq U$ are ζ' -consistent. By selecting the parameters appropriately, such an almost non-red subgraph once again gives us the desired global function. This wraps up our proof overview.

Changes from Previous Version. The proof presented in this manuscript differs slightly from that in the conference version of this work [DM18]. Specifically, while both versions follow the same approach to find U_1, U_2 , they diverge afterwards. In [DM18], instead of reapplying the argument of the regime $\delta \gg 1/k^{1/2}$ on U_1, U_2 as we do in this version, we resorted to the Kóvári-Sós-Turán (KST) Theorem [KST54], which roughly states that every dense bipartite graph has a biclique (complete bipartite subgraph) of logarithmic size. By applying this theorem to the bipartite graph between U_1 and U_2 where there is an edge between $u_1 \in U_1$ and $u_2 \in U_2$ if and only if $\{u_1, u_2\}$ is not a red edge,

we arrived at subsets $V_1 \subseteq U_1, V_2 \subseteq U_2$ of reasonable large size such that for all $(u_1, u_2) \in V_1 \times V_2$, u_1 and u_2 are not joined by a red edge. Finally, we observed that such a “non-red biclique” can also be used to decode a global function, by taking the majority of either V_1 or V_2 side.

A disadvantage of the proof in [DM18] is that the agreement theorem there finds a global function that approximately agrees with only $\Omega_{\beta, \alpha}(\log(\delta k))$ of the local functions whereas Theorem 11 finds one that approximately agrees with $\delta k^{1-o_{\beta, \alpha}(1)}$ local functions, which, as discussed earlier, is nearly optimal. This loss in [DM18] also affects the low order term in the inapproximability ratio: while the ETH-hardness in the current manuscript has inapproximability ratio $k/2^{(\log k)^{1/2+\rho}}$ for any $\rho > 0$, the ratio in [DM18] is only $k/2^{(\log k)^{1-\gamma}}$ for some (small) $\gamma > 0$.

3 Preliminaries

3.1 Well-Behaved Subsets

We next define two properties of collections of subsets, which will be needed in our soundness analysis. First, recall that, in our proof overview for the weaker $k^{1/2-o(1)}$ factor hardness, we need the following to show the red/blue transitivity of the consistency graph: for any r subsets from the collection, their union must contain almost all clauses. Here r is a positive integer that effects the red/blue transitivity parameter. Collections with this property are sometimes called *dispersers*. For walks with larger length, we need a stronger property that any union of r intersections of ℓ subsets are large. We call such collections *intersection dispersers*:

Definition 18 (Intersection Disperser) For any universe \mathcal{U} , a collection \mathcal{S} of subsets of \mathcal{U} is an (r, ℓ, η) -intersection disperser if, for any r disjoint subcollections $\mathcal{S}^1, \dots, \mathcal{S}^r \subseteq \mathcal{S}$ each of size at most ℓ , we have

$$\left| \bigcup_{i=1}^r \left(\bigcap_{S \in \mathcal{S}^i} S \right) \right| \geq (1 - \eta)|\mathcal{U}|.$$

Note that in the definition we require $\mathcal{S}^1, \dots, \mathcal{S}^r$ to be disjoint. This is necessary because otherwise we can include a common set $S \in \mathcal{S}$ into all the subcollections. In this case, the union will be contained in S and hence will not cover almost all the universe.

Another property we need is that any sufficiently large subcollection $\tilde{\mathcal{S}}$ of \mathcal{S} is “sufficiently uniform”. This is used when we decode a good assignment from an almost non-red subgraph. More specifically, the uniformity condition requires that almost all clauses appear in not too small number of subsets in $\tilde{\mathcal{S}}$, as formalized below.

Definition 19 (Uniformity) For any universe \mathcal{U} , a collection $\tilde{\mathcal{S}}$ of subsets of \mathcal{U} is (γ, μ) -uniform if, for at least $(1 - \mu)$ fraction of elements $u \in \mathcal{U}$, u appears in at least γ fraction of the subsets in $\tilde{\mathcal{S}}$. In other words, $\tilde{\mathcal{S}}$ is (γ, μ) -uniform if and only if $|\{u \in \mathcal{U} \mid \Pr_{S \in \tilde{\mathcal{S}}}[u \in S] \geq \gamma\}| \geq (1 - \mu)|\mathcal{U}|$.

Using standard concentration bounds, it is not hard to show that, when m is sufficiently large, a collection of random subsets where each element is included in each subset independently with probability α is an $(1/O(\alpha^\ell), \ell, O(1))$ -disperser and every subcollection of size $\Omega(1/\alpha)$ is $(\alpha, O(1))$ uniform. The exact parameter dependencies are shown in the lemma below.

Lemma 20 (Deterministic Construction of Well-Behaved Subsets) For any $0 < \alpha, \mu, \eta < 1$ and any $k, \ell \in \mathbb{N}$, let m_0 be $1000(\log k \log(1/\mu)/(\alpha\mu^2) + \ell \log(1/\eta) \log k/(\alpha^\ell \eta) + 1/\alpha + 1)$. For any integer $m \geq m_0$ and m -element set \mathcal{U} , there is a collection \mathcal{T} of subsets of \mathcal{U} with the following properties.

- (Size) Every subset in \mathcal{T} has size at most $2\alpha m$.
- (Intersection Disperser) \mathcal{T} is a $(\lceil \ln(2/\eta)/\alpha^\ell \rceil, \ell, \eta)$ -disperser.
- (Uniformity) Any subcollection $\tilde{\mathcal{T}} \subseteq \mathcal{T}$ of size $\lceil 8 \ln(2/\mu)/\alpha \rceil$ is $(\alpha/2, \mu)$ -uniform.

Moreover, such a collection \mathcal{T} can be deterministically constructed in time $\text{poly}(m)2^{O(m_0k^2)}$.

Since all techniques involved in the proof of Lemma 20 are standard, we defer it to Appendix A. Let us turn our focus back to our main technical contribution: the agreement testing theorem.

4 The Main Agreement Theorem

The main goal of this section is to prove the following agreement theorem, which is the formal version of Theorem 11 and is also the main technical contribution of this work.

Theorem 21 *For any $0 < \eta, \zeta, \gamma, \mu < 1$ and $r, \ell, k, h, n, d \in \mathbb{N}$ such that $\ell \geq 2$, let \mathcal{S} be any collection of k subsets of $[n]$ such that \mathcal{S} is (r, ℓ, ζ) -intersection disperser and every subcollection $\tilde{\mathcal{S}} \subseteq \mathcal{S}$ of size h is (γ, μ) -uniform, and let $\mathcal{F} = \{f_S\}_{S \in \mathcal{S}}$ be any collection of functions. If $\delta \triangleq \text{agree}(\mathcal{F}) \geq \frac{10+64(r\ell)^2k^{1/\ell}}{k}$, then there exists a subcollection $\mathcal{S}' \subseteq \mathcal{S}$ of size at least $\frac{\delta k}{256\ell^2}$ and a function $g : [n] \rightarrow \{0, 1\}$ such that $g \stackrel{\beta}{\approx} f_S$ for all $S \in \mathcal{S}'$ where*

$$\beta = 2\sqrt{\frac{65536h\ell^6}{\delta k} + \mu + 2\zeta/\gamma}.$$

While the parameters of the theorem can be confusing, when each subset in \mathcal{S} is a random αn -size subset of $[n]$, the parameters we are interested in are as follows: μ and η both go to 0 as n goes to infinity, h and γ depend only on α , and, r is $O(1/\alpha^\ell)$. Since we want the requirement on soundness as weak as possible, we want to minimize $(r\ell)^2k^{1/\ell} = 2^{O_\alpha(\ell + (\log k)/\ell)}$. Hence, our best choice is to let $\ell = \sqrt{\log k}$, which indeed yields the $k/2^{(\log k)^{1/2+\rho}}$ ratio inapproximability for 2-CSPs.

To prove this theorem, we follow the general outline as stated in the proof overview section. In particular, the proof contains five main steps, as elaborated below.

- (1) First, we will show that when \mathcal{S} is an intersection disperser with appropriate parameters, then the two-level consistency graph $G^{\mathcal{F}, \zeta}$ satisfies (q, ℓ) -red/blue transitivity for certain q, ℓ .
- (2) Second, we argue that, for any red/blue transitive graphs that contains sufficiently many blue edges, we can find a large subset \tilde{U} of vertices such that a reasonably large fraction of pairs $\{S_1, S_2\} \subseteq \tilde{U}$ are non-red. This is done by counting red-filled ℓ -walks for an appropriate ℓ .
- (3) We then focus on $\tilde{\mathcal{F}} = \{f_S\}_{S \in \tilde{U}}$ and show, using a uniformity condition of \mathcal{S} , that the generalized two-level consistency graph $G^{\tilde{\mathcal{F}}, \zeta, \zeta'}$ is red/blue transitive with certain parameters.
- (4) Next, counting rbb triangles reveals a large “almost non-red subgraph” in the graph $G^{\tilde{\mathcal{F}}, \zeta, \zeta'}$.
- (5) Finally, we decode a global function from this almost non-red subgraph.

This section is organized as follows. In Subsection 4.1, we show transitivity properties of the two-level and generalized two-level consistency graphs, i.e., Steps (1) and (3). Subsection 4.2 contains a structural lemma regarding an existence of a large subgraph with certain non-red density in red/blue transitive graphs; this lemma is at the heart of Steps (2) and (4). Next, in Subsection 4.3, we prove Step (5). Finally, in Subsection 4.4, we put these parts together and prove Theorem 21.

4.1 Red/Blue-Transitivity of (Generalized) Two-Level Consistency Graph

4.1.1 Red/Blue-Transitivity from Intersection Disperser

The first step in our proof is to show that the two-level consistency graph $G^{\mathcal{F}, \zeta}$ is red/blue-transitive, assuming that \mathcal{S} is an intersection disperser. Specifically, our main lemma is the following:

Lemma 22 *If \mathcal{S} is an (r, ℓ, ζ) -intersection disperser, then, for any $\mathcal{F} = \{f_S\}_{S \in \mathcal{S}}$, $G^{\mathcal{F}, \zeta}$ is $((r\ell)^{2(\ell-1)}, \ell)$ -red/blue-transitive.*

We note here that both in Lemma 22 and Claim 23 below, the transitivity property holds not only for ℓ -walks as specified in the statements, but also for $(\ell + 1)$ -walks. However, since the latter does not yield any improvement to our main results, we work with only ℓ -walks, which makes the calculations cleaner.

In other words, we would like to show that, for every $S_1, S_2 \in \mathcal{S}$ that are joined by a red edge in $G^{\mathcal{F}, \zeta}$, there are at most $(r\ell)^{2(\ell-1)}$ red-filled ℓ -walks from S_1 to S_2 . The intersection disperser does not immediately imply such a bound, due to the requirement in the definition that the subcollections are disjoint. Rather, it only directly implies a bound on number of *disjoint* ℓ -walks from S_1 to S_2 , where two ℓ walks from S_1 to S_2 , $(T_1 = S_1, \dots, T_{\ell+1} = S_2), (T'_1 = S_1, \dots, T'_{\ell+1} = S_2) \in \mathcal{W}_\ell^{G^{\mathcal{F}, \zeta}}(S_1, S_2)$, are said to be *disjoint* if they do not share any vertex except the starting and ending vertices, i.e., $\{T_2, \dots, T_\ell\} \cap \{T'_2, \dots, T'_\ell\} = \emptyset$. Note that multiple walks sharing starting and ending vertices are said to be disjoint if they are mutually disjoint. The following claim follows almost immediately from definition of intersection dispersers:

Claim 23 *If \mathcal{S} is an (r, ℓ, ζ) -intersection disperser, then, for any $\mathcal{F} = \{f_S\}_{S \in \mathcal{S}}$, any integer $2 \leq p \leq \ell$ and any $\{S_1, S_2\} \in E_r^{\mathcal{F}, \zeta}$, there are less than r disjoint p -walks from S_1 to S_2 in $G^{\mathcal{F}, \zeta}$.*

Proof. Suppose for the sake of contradiction that \mathcal{S} is an (r, ℓ, ζ) -intersection disperser but there exist $\mathcal{F} = \{f_S\}_{S \in \mathcal{S}}, 2 \leq p \leq \ell$ and $\{S_1, S_2\} \in E_r^{\mathcal{F}, \zeta}$ such that there are at least r disjoint p -walks from S_1 to S_2 . Let these walks be $(T_{1,1} = S_1, T_{1,2}, \dots, T_{1,p}, T_{1,p+1} = S_2), \dots, (T_{r,1} = S_1, T_{r,2}, \dots, T_{r,p}, T_{r,p+1} = S_2) \in \mathcal{W}_p^{G^{\mathcal{F}, \zeta}}(S_1, S_2)$.

For each $i \in [r]$, consider any $x \in \bigcap_{j=1}^{p+1} T_{i,j}$. A priori this intersection may be empty but since \mathcal{S} is an intersection disperser this usually does not occur. Since $\{T_{i,j}, T_{i,j+1}\} \in E_b^{\mathcal{F}, \zeta}$ for every $j \in [p]$, we have

$$f_{S_1}(x) = f_{T_{i,1}}(x) = f_{T_{i,2}}(x) = \dots = f_{T_{i,p}}(x) = f_{T_{i,p+1}}(x) = f_{S_2}(x).$$

Hence, for every $x \in \bigcup_{i=1}^r \left(\bigcap_{j=1}^{p+1} T_{i,j} \right)$, $f_{S_1}(x) = f_{S_2}(x)$. Let T^* denote $\bigcup_{i=1}^r \left(\bigcap_{j=2}^p T_{i,j} \right)$. Since $T_{i,1} = S_1$ and $T_{i,p+1} = S_2$ for all $i \in [r]$, we have

$$\bigcup_{i=1}^r \left(\bigcap_{j=1}^{p+1} T_{i,j} \right) = (S_1 \cap S_2) \cup T^*.$$

In other words, f_{S_1} and f_{S_2} can only disagree on variables outside of T^* . However, since \mathcal{S} is an (r, ℓ, ζ) -intersection disperser, we have $|T^*| \geq (1 - \zeta)n$. Hence, $\text{disagr}(S_1, S_2) \leq \zeta n$, which contradicts with $\{S_1, S_2\} \in E_r^{\mathcal{F}, \zeta}$. \square

Since all 2-walks from S_1 to S_2 are disjoint, the above claim immediately gives a bound on the number of red-filled 2-walks from S_1 to S_2 . To bound the number of red-filled walks of larger lengths, we will use induction on the length of the walks. Suppose that we have bounded the number of red-filled i -walks sharing starting and ending vertices for $i \leq z - 1$. The key idea in the proof is that we can use this inductive hypothesis to show that, for any $S_1, S_2, S \in \mathcal{S}$, few z -walks from S_1 to S_2 contain a given S . Here we say that a z -walk $(T_1 = S_1, \dots, T_z = S_2)$ from S_1 to S_2 contains S if $S \in \{T_2, \dots, T_z\}$. This implies that for a given z -walk from S_1 to S_2 there are only few walks that are not disjoint from it. This allows us to show that, if there are too many z -walks, then there must also be many disjoint z -walks as well, which would violate Claim 23. A formal proof of Lemma 22 based on this intuition is given below.

Proof of Lemma 22. For every integer i such that $2 \leq i \leq \ell$, let $P(i)$ denote the following statement: for every $S_1, S_2 \in \mathcal{S}$, $|\widehat{\mathcal{W}}_i^{G^{\mathcal{F}, \mathcal{L}}}(S_1, S_2)| \leq (ri)^{2(i-1)}$. For convenience, let $B_i = (ri)^{2(i-1)}$ for every $2 \leq i \leq \ell$.

Base Case. Since every different 2-walks from S_1 to S_2 are disjoint, Claim 23 immediately implies that the number of 2-walks from S_1 to S_2 is at most $r \leq B_2$.

Inductive Step.

Suppose that, for some integer z such that $3 \leq z \leq \ell$, $P(3), \dots, P(z-1)$ are true. We will show that $P(z)$ is true. To do so, let us first prove that, for any fixed starting and ending vertices, any vertex cannot appear in too many red-filled z -walks, as stated in the following claim.

Claim 24 For all $S_1, S_2, S \in \mathcal{S}$, the number of red-filled z -walks from S_1 to S_2 containing S in $G^{\mathcal{F}, \mathcal{L}}$ is at most $B_z/(zr)$.

Proof. First, observe that the number of red-filled z -walks from S_1 to S_2 containing S is at most the sum over all positions $2 \leq j \leq z$ of the number of z -walks from S_1 to S_2 such that the j -th vertex in the walk is S . More formally, the number of red-filled z -walks from S_1 to S_2 containing S is

$$|\{(T_1, \dots, T_{z+1}) \in \widehat{\mathcal{W}}_z^{G^{\mathcal{F}, \mathcal{L}}}(S_1, S_2) \mid \exists 2 \leq j \leq z, T_j = S_j\}| \leq \sum_{j=2}^z |\{(T_1, \dots, T_{z+1}) \in \widehat{\mathcal{W}}_z^{G^{\mathcal{F}, \mathcal{L}}}(S_1, S_2) \mid T_j = S\}|.$$

Now, for each $2 \leq j \leq z$, to bound the number of red-filled z -walks from S_1 to S_2 whose j -th vertex is S , let us consider the following three cases based on the value of j :

1. $3 \leq j \leq z - 1$. Observe that, for any such walk $(T_1 = S_1, T_2, \dots, T_j = S, \dots, T_z, T_{z+1} = S_2)$, the subwalk $(T_1 = S_1, \dots, T_j = S)$ and $(T_j = S, \dots, T_{z+1} = S_2)$ must be red-filled walks as well. Since the numbers of red-filled $(j-1)$ -walks from S_1 to S and red-filled $(z+1-j)$ -walks from S to S_2 are bounded by B_{j-1} and B_{z+1-j} respectively (from the inductive hypothesis), there are at most B_{j-1} choices of $(T_1 = S_1, \dots, T_j = S)$ and B_{z+1-j} choices of $(T_j = S, \dots, T_{z-1}, T_z = S_2)$. Hence, there are at most $B_{j-1}B_{z+1-j}$ red-filled z -walks from S_1 to S_2 whose j -th vertex is S .
2. $j = 2$. In this case, the subwalk $(T_j = S, \dots, T_{z+1} = S_2)$ must be a red-filled $(z-1)$ -walk from S to S_2 . Hence, the number of red-filled z -walks from S_1 to S_2 where $T_j = S$ is bounded above by B_{z-1} .
3. $j = z$. Similar to the previous case, we also have the bound of B_{z-1} .

For convenience, let $B_1 = 1$. The above argument gives us the following bound for every $2 \leq j \leq z$:

$$|\{(T_1, \dots, T_{z+1}) \in \widehat{\mathcal{W}}_z^{G^{\mathcal{F}, \zeta}}(S_1, S_2) \mid T_j = S\}| \leq B_{j-1}B_{z+1-j}.$$

Summing this over j , we have the following upper bound on the number of red-filled z -walks from S_1 to S_2 containing S :

$$\sum_{j=2}^z B_{j-1}B_{z+1-j} = \sum_{j=2}^z (r(j-1))^{2(j-2)}(r(z+1-j))^{2(z-j)} \leq \sum_{j=2}^z (rz)^{2(z-2)} \leq B_z/(zr),$$

which concludes the proof of the claim. \square

Having proved the above claim, it is now easy to show that $P(z)$ is true. Suppose for the sake of contradiction that there exists $S_1, S_2 \in \mathcal{S}$ such that $|\widehat{\mathcal{W}}_z^{G^{\mathcal{F}, \zeta}}(S_1, S_2)| > B_z$. Consider the following procedure of selecting disjoint walks from $\widehat{\mathcal{W}}_z^{G^{\mathcal{F}, \zeta}}(S_1, S_2)$. First, initialize $U = \widehat{\mathcal{W}}_z^{G^{\mathcal{F}, \zeta}}(S_1, S_2)$ and repeat the following process as long as $U \neq \emptyset$: select any $(T_1, \dots, T_{z+1}) \in U$ and remove every (T'_1, \dots, T'_{z+1}) that is not disjoint with (T_1, \dots, T_{z+1}) from U . Observe that, each time a walk (T_1, \dots, T_{z+1}) is selected, the number of walks removed from U is at most B_z/r ; this is because each removed walk must contain at least one of T_2, \dots, T_z , but, from the above claim, each of these vertices are contained in at most $B_z/(zr)$ walks. Since we start with more than B_z walks, at least r walks are picked. These walks are disjoint z -walks starting from S_1 and S_2 , which, due to Claim 23, is a contradiction. Thus, $P(z)$ is true as desired.

Hence, $P(\ell)$ is true, which, by definition, implies that $G_{\mathcal{F}}$ is $((r\ell)^{2(\ell-1)}, \ell)$ -red/blue-transitive. \square

4.1.2 Red/Blue-Transitivity from Uniformity

In Step (3) of our proof, we need to show red/blue-transitivity of the generalized two-level consistency graph $G^{\mathcal{F}, \zeta, \zeta'}$. This is encapsulated in the following lemma.

Lemma 25 *If every subcollection $\tilde{\mathcal{S}} \subseteq \mathcal{S}$ of size r is (γ, μ) -uniform, then, for any $\zeta \geq 0, \zeta' \geq \mu + 2\zeta/\gamma$ and $\mathcal{F} = \{f_S\}_{S \in \mathcal{S}}$, the generalized two-level consistency graph $G^{\mathcal{F}, \zeta, \zeta'}$ is r -red/blue transitive.*

The proof of the lemma is quite simple. The key observation is that, if S_1 and S_2 are joined by a red edge and T is a common blue-neighbor in the graph $G^{\mathcal{F}, \zeta, \zeta'}$, then it means that T only hits a small number (i.e. $2\zeta n$) of the variables on which f_{S_1} and f_{S_2} disagree. In other words, such variables appear less frequently in common blue-neighbors of S_1 and S_2 . If the common-blue neighbor set is of size r , this contradicts the fact that the set is (γ, μ) -uniform. This intuition is formalized below.

Proof of Lemma 25. Suppose for the sake of contradiction that $G^{\mathcal{F}, \zeta, \zeta'}$ is not r -red/blue transitive. That is, there exist $S_1, S_2 \in \mathcal{S}$ that are joined by a red edge such that there are r red-filled 2-walks (i.e. rbb triangle) from S_1 to S_2 . Suppose that these walks are $(S_1, T_1, S_2), (S_1, T_2, S_2), \dots, (S_1, T_r, S_2)$.

For every $i \in [r]$, since (S_1, T_i, S_2) is a 2-walk, $\{S_1, T_i\}$ and $\{S_2, T_i\}$ are blue edges. This implies that

$$\text{disagr}(f_{S_1}, f_{T_i}), \text{disagr}(f_{S_2}, f_{T_i}) \leq \zeta n. \quad (1)$$

On the other hand, we can lower bound $\mathbb{E}_{i \in [r]} [\text{disagr}(f_{S_1}, f_{T_i}) + \text{disagr}(f_{S_2}, f_{T_i})]$ as follows. First, let let X^{disagr} denote the set of all $x \in S_1 \cap S_2$ such that $f_{S_1}(x) \neq f_{S_2}(x)$; since $\{S_1, S_2\}$ is a red edge,

we have $|\mathcal{X}^{\text{disagr}}| > \zeta'n$. We can rearrange $\mathbb{E}_{i \in [r]}[\text{disagr}(f_{S_1}, f_{T_i}) + \text{disagr}(f_{S_2}, f_{T_i})]$ as

$$\begin{aligned}
& \mathbb{E}_{i \in [r]} [\text{disagr}(f_{S_1}, f_{T_i}) + \text{disagr}(f_{S_2}, f_{T_i})] \\
&= \sum_{x \in [n]} \left(\Pr_{i \in [r]} [x \in (S_1 \cap T_i) \wedge f_{S_1}(x) \neq f_{T_i}(x)] + \Pr_{i \in [r]} [x \in (S_2 \cap T_i) \wedge f_{S_2}(x) \neq f_{T_i}(x)] \right) \\
&\geq \sum_{x \in \mathcal{X}^{\text{disagr}}} \left(\Pr_{i \in [r]} [x \in T_i \wedge f_{S_1}(x) \neq f_{T_i}(x)] + \Pr_{i \in [r]} [x \in T_i \wedge f_{S_2}(x) \neq f_{T_i}(x)] \right) \\
&\geq \sum_{x \in \mathcal{X}^{\text{disagr}}} \left(\Pr_{i \in [r]} [x \in T_i \wedge (f_{S_1}(x) \neq f_{T_i}(x) \vee f_{S_2}(x) \neq f_{T_i}(x))] \right) \\
&= \sum_{x \in \mathcal{X}^{\text{disagr}}} \Pr_{i \in [r]} [x \in T_i] \tag{2}
\end{aligned}$$

We remark here that the second inequality comes from union bound, whereas the last equality follows from the fact that $(f_{S_1}(x) \neq f_{T_i}(x)) \vee (f_{S_2}(x) \neq f_{T_i}(x))$ is always true when $f_{S_1}(x) \neq f_{S_2}(x)$.

Recall that $\{T_1, \dots, T_r\} \subseteq \mathcal{S}$ is a subcollection of size r and is thus (γ, μ) -uniform. Let $\mathcal{X}_{\geq \gamma}$ be the set of all $x \in [n]$ that appears in at least γ fraction of T_i 's. The (γ, μ) -uniformity of $\{T_1, \dots, T_r\}$ implies that $|\mathcal{X}_{\geq \gamma}| \geq (1 - \mu)n$. From this and from $|\mathcal{X}^{\text{disagr}}| > \zeta'n$, we can lower bound the right hand side of (2) further as follows:

$$\sum_{x \in \mathcal{X}^{\text{disagr}}} \Pr_{T_i \in \mathcal{T}} [x \in T_i] \geq \sum_{x \in \mathcal{X}^{\text{disagr}} \cap \mathcal{X}_{\geq \gamma}} \Pr_{T_i \in \mathcal{T}} [x \in T_i] \geq \gamma |\mathcal{X}^{\text{disagr}} \cap \mathcal{X}_{\geq \gamma}| > \gamma(\zeta' - \mu)n \geq 2\zeta n \tag{3}$$

where the last inequality comes from our assumption that $\zeta' \geq \mu + 2\zeta/\gamma$.

Finally, combining (1), (2) and (3) yields the desired contradiction. \square

4.2 Finding Almost Non-Red Subgraph in Red/Blue-Transitive Graph

Recall that in two steps of our proofs, we need to utilize the red/blue transitivity of the (generalized) two-level consistency graph to find a large subgraph with certain number of non-red pairs:

- Specifically, in Step (2), we would like to show that, for appropriate values of q and ℓ , any (q, ℓ) -red/blue transitive graph with sufficiently many blue edges must contain a sufficiently large subgraph whose significant (i.e. $1/\ell^2$) fraction of pairs of vertices are non-red.
- Additionally, in Step (4), we need to show that any $o(d^2/k)$ -red/blue transitive graph with sufficiently many blue edges must contain a sufficiently large subgraph such that almost all pairs of its vertices are non-red.

It turns out that a single lemma stated below suffices for both steps. In particular, the lemma below returns a subgraph such that roughly $1/\binom{\ell_0}{2}$ fraction of pairs of its vertices are non-red. Plugging in $\ell_0 = \ell$ recovers our former objective whereas setting $\ell_0 = 2$ satisfies the latter.

Lemma 26 *For every $k_0, q_0, \ell_0, d_0 \in \mathbb{N}$ such that $\ell_0 \geq 2$ and every k_0 -vertex (q_0, ℓ_0) -red/blue-transitive graph $G = (V, E_r \cup E_b)$ such that $|E_b| \geq 2k_0d_0$, there exist subsets of vertices $U_1, U_2 \subseteq V$ each of size at least d_0 such that $|\{(u, v) \in U_1 \times U_2 \mid \{u, v\} \notin E_r\}| \geq |U_1||U_2|(1 - \frac{q_0k_0}{d_0^{\ell_0}})/\binom{\ell_0}{2}$. Moreover, when $\ell_0 = 2$, the previous statement remains true even with an additional requirement that $U_1 = U_2$.*

The proof of Lemma 26 below is exactly as sketched earlier in Subsection 2.

Proof of Lemma 26. We start by preprocessing the graph so that every vertex has blue-degree at least d_0 . In particular, as long as there exists a vertex v whose blue-degree is at most d_0 , we remove v from G . Let $G' = (V', E'_r \cup E'_b)$ be the graph at the end of this process. Note that we remove less than $k_0 d_0$ blue edges in total. Since at the beginning $|E_b| \geq 2k_0 d_0$, we have $|E'_b| \geq k_0 d_0$. Observe also that G' remains (q_0, ℓ_0) -red/blue-transitive.

Since V' is (q_0, ℓ_0) -red/blue-transitive, we can bound the number of red-filled ℓ_0 -walk as follows.

$$|\widehat{\mathcal{W}}_{\ell_0}^{G'}| = \sum_{\substack{u, v \in V' \\ \{u, v\} \in E'_r}} |\widehat{\mathcal{W}}_{\ell_0}^{G'}(u, v)| \leq \sum_{\substack{u, v \in V' \\ \{u, v\} \in E'_r}} q_0 \leq q_0 k_0^2.$$

Moreover, notice that $|\mathcal{W}_{\ell_0}^{G'}| \geq (k_0 d_0) \cdot d_0^{\ell_0 - 1} = k_0 d_0^{\ell_0}$; this is simply because there are at least $k_0 d_0$ choices for (v_1, v_2) (i.e. all blue edges) and, for any (v_1, \dots, v_{i-1}) , there are at least d_0 choices for v_i .

Hence, we have $|\widehat{\mathcal{W}}_{\ell_0}^{G'}| / |\mathcal{W}_{\ell_0}^{G'}| \leq q_0 k_0 / d_0^{\ell_0}$. This implies that $1 - q_0 k_0 / d_0^{\ell_0} \leq \Pr_{(v_1, \dots, v_{\ell_0+1}) \in \mathcal{W}_{\ell_0}^{G'}}[(v_1, \dots, v_{\ell_0+1}) \notin \widehat{\mathcal{W}}_{\ell_0}^{G'}]$. This probability can be further rearranged as follows.

$$\begin{aligned} \Pr_{(v_1, \dots, v_{\ell_0+1}) \in \mathcal{W}_{\ell_0}^{G'}}[(v_1, \dots, v_{\ell_0+1}) \notin \widehat{\mathcal{W}}_{\ell_0}^{G'}] &= \Pr_{(v_1, \dots, v_{\ell_0+1}) \in \mathcal{W}_{\ell_0}^{G'}}[\exists i, j \in [\ell_0 + 1] \text{ such that } j > i + 1, \{v_i, v_j\} \notin E'_r] \\ &\text{(Union Bound)} \leq \sum_{\substack{i, j \in [\ell_0+1] \\ j > i+1}} \Pr_{(v_1, \dots, v_{\ell_0+1}) \in \mathcal{W}_{\ell_0}^{G'}}[\{v_i, v_j\} \notin E'_r]. \end{aligned}$$

Now, note that the number of pairs of $i, j \in [\ell_0 + 1]$ such that $j > i + 1$ is $\binom{\ell_0+1}{2} - \ell_0 = \binom{\ell_0}{2}$. This implies that there exists one such i, j such that $\Pr_{(v_1, \dots, v_{\ell_0+1}) \in \mathcal{W}_{\ell_0}^{G'}}[\{v_i, v_j\} \notin E'_r] \geq (1 - \frac{q_0 k_0}{d_0^{\ell_0}}) / \binom{\ell_0}{2}$. The probability $\Pr_{(v_1, \dots, v_{\ell_0+1}) \in \mathcal{W}_{\ell_0}^{G'}}[\{v_i, v_j\} \notin E'_r]$ can now be bounded as follows.

$$\begin{aligned} &\Pr_{(v_1, \dots, v_{\ell_0+1}) \in \mathcal{W}_{\ell_0}^{G'}}[\{v_i, v_j\} \notin E'_r] \\ &= \sum_{u, v} \Pr_{(v_1, \dots, v_{\ell_0+1}) \in \mathcal{W}_{\ell_0}^{G'}}[\{v_i, v_j\} \notin E'_r \mid v_{i+1} = u \wedge v_{j-1} = v] \Pr_{(v_1, \dots, v_{\ell_0+1}) \in \mathcal{W}_{\ell_0}^{G'}}[v_{i+1} = u \wedge v_{j-1} = v] \\ &\leq \left(\max_{u, v} \Pr_{(v_1, \dots, v_{\ell_0+1}) \in \mathcal{W}_{\ell_0}^{G'}}[\{v_i, v_j\} \notin E'_r \mid v_{i+1} = u \wedge v_{j-1} = v] \right) \left(\sum_{u, v} \Pr_{(v_1, \dots, v_{\ell_0+1}) \in \mathcal{W}_{\ell_0}^{G'}}[v_{i+1} = u \wedge v_{j-1} = v] \right) \\ &= \max_{u, v} \Pr_{(v_1, \dots, v_{\ell_0+1}) \in \mathcal{W}_{\ell_0}^{G'}}[\{v_i, v_j\} \notin E'_r \mid v_{i+1} = u \wedge v_{j-1} = v] \end{aligned}$$

where the summation and maximization is taken over all $u, v \in V'$ such that $\Pr_{(v_1, \dots, v_{\ell_0+1}) \in \mathcal{W}_{\ell_0}^{G'}}[v_{i+1} = u \wedge v_{j-1} = v]$ is non-zero. Hence, we can conclude that there exists $u^*, v^* \in V'$ such that

$$\Pr_{(v_1, \dots, v_{\ell_0+1}) \in \mathcal{W}_{\ell_0}^{G'}}[\{v_i, v_j\} \notin E'_r \mid v_{i+1} = u^* \wedge v_{j-1} = v^*] \geq (1 - \frac{q_0 k_0}{d_0^{\ell_0}}) / \binom{\ell_0}{2}.$$

The expression on the left is exactly $|\{(u, v) \in N_b(u^*) \times N_b(v^*) \mid \{u, v\} \notin E'_r\}| / (|N_b(u^*)| \cdot |N_b(v^*)|)$. From this and from every vertex in G' has blue-degree at least d_0 , $U_1 = N_b(u^*), U_2 = N_b(v^*)$ are the desired sets. Finally, observe that, when $\ell = 2$, we must have $i = 1$ and $j = 3$, resulting in $v_{i+1} = v_{j-1}$; this implies that $u^* = v^*$ and we have $U_1 = U_2$. \square

4.3 Majority Decoding of an Almost Non-Red Subgraph

In the last step of our proof, we will decode a global function g from a sufficiently large almost non-red subgraph in the two-level consistency graph $G^{\mathcal{F}, \zeta, \zeta'}$. Recall that an almost non-red subgraph in $G^{\mathcal{F}, \zeta, \zeta'}$ simply corresponds to a subcollection \mathcal{S}' such that, for almost all pairs $(S_1, S_2) \in \mathcal{S}' \times \mathcal{S}'$, f_{S_1} is ζ' -consistent with f_{S_2} . The main result of this subsection is that, given such \mathcal{S}' , we can find a global function g that approximately agrees with most of the local functions in the subcollection. This is stated more precisely below.

Lemma 27 *Let $\mathcal{F} = \{f_S\}_{S \in \mathcal{S}'}$ be a collection of functions such that $\text{agree}_{\zeta'}(\mathcal{F}) \geq 1 - \kappa$. Then, the function $g : [n] \rightarrow \{0, 1\}$ defined by $g(x) = \text{Majority}_{\substack{S \in \mathcal{S}' \\ x \in S}}(f_S(x))$ satisfies*

$$\mathbb{E}_{S \in \mathcal{S}'} [\text{disagr}(g, f_S)] \leq n \sqrt{\kappa + \zeta'}.$$

Proof. Recall that $\text{agree}_{\zeta'}(\mathcal{F}) \geq 1 - \kappa$ is equivalent to $\Pr_{S_1, S_2 \in \mathcal{S}'} [f_{S_1}(x) \stackrel{\zeta'}{\approx} f_{S_2}(x)] \geq 1 - \kappa$. Hence,

$$\mathbb{E}_{S_1, S_2 \in \mathcal{S}'} [\text{disagr}(f_{S_1}, f_{S_2})] \leq \Pr_{S_1, S_2 \in \mathcal{S}'} [f_{S_1} \not\stackrel{\zeta'}{\approx} f_{S_2}] \cdot n + \Pr_{S_1, S_2 \in \mathcal{S}'} [f_{S_1} \stackrel{\zeta'}{\approx} f_{S_2}] \cdot (\zeta' n) \leq (\kappa + \zeta') n. \quad (4)$$

We can then lower bound the expression on the left hand side as follows.

$$\begin{aligned} \mathbb{E}_{S_1, S_2 \in \mathcal{S}'} [\text{disagr}(f_{S_1}, f_{S_2})] &= \sum_{x \in [n]} \Pr_{S_1, S_2 \in \mathcal{S}'} [x \in S_1 \wedge x \in S_2 \wedge f_{S_1}(x) \neq f_{S_2}(x)] \\ &\geq \sum_{x \in [n]} \Pr_{S_1, S_2 \in \mathcal{S}'} [x \in S_1 \wedge x \in S_2 \wedge f_{S_1}(x) \neq g(x) \wedge f_{S_2}(x) = g(x)] \\ &= \sum_{x \in [n]} \Pr_{S_1 \in \mathcal{S}'} [x \in S_1 \wedge f_{S_1}(x) \neq g(x)] \Pr_{S_2 \in \mathcal{S}'} [x \in S_2 \wedge f_{S_2}(x) = g(x)] \\ (\text{Since } g(x) &= \text{Majority}_{\substack{S \in \mathcal{S}' \\ x \in S}}(f_S(x))) &\geq \sum_{x \in [n]} \Pr_{S_1 \in \mathcal{S}'} [x \in S_1 \wedge f_{S_1}(x) \neq g(x)] \Pr_{S_2 \in \mathcal{S}'} [x \in S_2 \wedge f_{S_2}(x) \neq g(x)] \\ &= \sum_{x \in [n]} \left(\Pr_{S \in \mathcal{S}'} [x \in S \wedge f_S(x) \neq g(x)] \right)^2 \\ (\text{Power Mean Inequality}) &\geq \frac{1}{n} \left(\sum_{x \in [n]} \Pr_{S \in \mathcal{S}'} [x \in S \wedge f_S(x) \neq g(x)] \right)^2 \\ &= \frac{1}{n} \left(\mathbb{E}_{S \in \mathcal{S}'} [\text{disagr}(g, f_S)] \right)^2. \end{aligned} \quad (5)$$

Combining (4) and (5) gives the desired bound. \square

4.4 Putting Things Together: Proof of Theorem 21

Finally, we put all five steps together as outlined at the beginning of this section. This is formalized below. Note that Theorem 21 follows from the theorem below simply by Markov inequality.

Theorem 28 *For any $0 < \eta, \zeta, \gamma, \mu < 1$ and $r, \ell, k, h, n, d \in \mathbb{N}$ such that $\ell \geq 2$, let \mathcal{S} be any collection of k subsets of $[n]$ such that \mathcal{S} is (r, ℓ, ζ) -intersection disperser and every subcollection $\tilde{\mathcal{S}} \subseteq \mathcal{S}$ of size h is*

(γ, μ) -uniform, and let $\mathcal{F} = \{f_S\}_{S \in \mathcal{S}}$ be any collection of functions. If $\delta \triangleq \text{agree}(\mathcal{F}) \geq \frac{10+64(r\ell)^2 k^{1/\ell}}{k}$, then there exists a subcollection $\mathcal{S}' \subseteq \mathcal{S}$ of size at least $\frac{\delta k}{128\ell^2}$ and a function $g : [n] \rightarrow \{0, 1\}$ such that

$$\mathbb{E}_{S \in \mathcal{S}'} [\text{disagr}(g, f_S)] \leq n \sqrt{\frac{65536h\ell^6}{\delta k}} + \mu + 2\zeta/\gamma.$$

Proof. Observe that $\text{agree}(\mathcal{F})$ directly corresponds to the number of blue edges $|E_b^{\mathcal{F}, \zeta}|$ in the two-level consistency graph $G^{\mathcal{F}, \zeta}$. In particular, $\text{agree}(\mathcal{F}) = \delta$ means that the number of blue edges is $(\delta k^2 - k)/2$. Since \mathcal{S} is a (r, ℓ, ζ) -intersection disperser, Lemma 22 implies that $G^{\mathcal{F}, \zeta}$ is $((r\ell)^{2(\ell-1)}, \ell)$ -red/blue-transitive. Let $d = \lfloor \frac{|E_b^{\mathcal{F}, \zeta}|}{2k} \rfloor = \lfloor \frac{\delta k - 1}{4} \rfloor$; since $\delta \geq \frac{10+64(r\ell)^2 k^{1/\ell}}{k}$, we have $d \geq (r\ell)^2 k^{1/\ell}$.

Applying Lemma 26 with $G = G^{\mathcal{F}, \zeta}$, $k_0 = k$, $\ell_0 = \ell$, $q_0 = (r\ell)^{2(\ell-1)}$ and $d_0 = d$, we can conclude that there exist subsets $U_1, U_2 \subseteq V^{\mathcal{F}, \zeta}$ each of size at least d such that

$$\frac{|\{(u, v) \in U_1 \times U_2 \mid \{u, v\} \notin E_r^{\mathcal{F}, \zeta}\}|}{|U_1||U_2|} \geq \frac{1 - (r\ell)^{2(\ell-1)}k/d^\ell}{\binom{\ell}{2}} \geq \frac{1}{\ell^2}$$

where the last inequality follows from our aforementioned lower bound on d and from $\ell \geq 2$.

Next, observe that, if we let U'_1 and U'_2 be random subsets of U_1, U_2 of size d , then we have

$$\mathbb{E}_{U'_1, U'_2} \left[\frac{|\{(u', v') \in U'_1 \times U'_2 \mid \{u', v'\} \notin E_r^{\mathcal{F}, \zeta}\}|}{d^2} \right] = \frac{|\{(u, v) \in U_1 \times U_2 \mid \{u, v\} \notin E_r^{\mathcal{F}, \zeta}\}|}{|U_1||U_2|}.$$

As a result, there exists \tilde{U}_1, \tilde{U}_2 each of size exactly d such that

$$\frac{|\{(\tilde{u}, \tilde{v}) \in \tilde{U}_1 \times \tilde{U}_2 \mid \{\tilde{u}, \tilde{v}\} \notin E_r^{\mathcal{F}, \zeta}\}|}{d^2} \geq \frac{1}{\ell^2}. \quad (6)$$

Now, let $\tilde{U} = \tilde{U}_1 \cup \tilde{U}_2$. (6) implies that the number of $\{\tilde{u}, \tilde{v}\} \subseteq \tilde{U}$ such that $\{\tilde{u}, \tilde{v}\} \notin E_r^{\mathcal{F}, \zeta}$ is at least $d^2/(2\ell^2) - d$ where the factor of 2 comes from the fact that each pair $\{\tilde{u}, \tilde{v}\}$ is double counted in the left hand side of (6) and the subtraction of d comes from the fact that the left hand side of (6) also count the case where $\tilde{u} = \tilde{v}$.

Now, let $\tilde{\mathcal{F}} = \{f_S\}_{S \in \tilde{U}}$, $\zeta' = \mu + 2\zeta/\gamma$ and consider the two-level consistency graph $G^{\tilde{\mathcal{F}}, \zeta'}$. Observe that $\{\tilde{u}, \tilde{v}\}$ is a blue edge in this new graph $G^{\tilde{\mathcal{F}}, \zeta'}$ if and only if it is not a red edge in the original graph $G^{\mathcal{F}, \zeta}$. Hence, the bound derived in the previous paragraph implies that

$$|E_b^{\tilde{\mathcal{F}}, \zeta'}| \geq \frac{d^2}{2\ell^2} - d.$$

Let $d' = |E_b^{\tilde{\mathcal{F}}, \zeta'}|/(2|\tilde{U}|) \geq |E_b^{\tilde{\mathcal{F}}, \zeta'}|/(4d) = d/(8\ell^2) - 1/4$. Recall that $d = \lfloor (\delta k - 1)/4 \rfloor$; from $\delta \geq (10 + 64(r\ell)^2)/k$, we have $d \geq 8\ell^2$ and $d \geq \delta k/8$. Hence, we have $d' \geq d/(16\ell^2) \geq \delta k/(128\ell^2)$.

Furthermore, by Lemma 25 and from our assumption that every subcollection of \mathcal{S} of size h is (γ, μ) -uniform, the graph $G^{\tilde{\mathcal{F}}, \zeta'}$ is h -red/blue transitive. Applying Lemma 26 with $G = G^{\tilde{\mathcal{F}}, \zeta'}$, $k_0 = |\tilde{U}| \leq 2d$, $\ell_0 = 2$, $q_0 = h$ and $d_0 = d'$, there must be a set $U' \subseteq \tilde{U}$ of size at least d' such that

$$\frac{|\{(u', v') \in U' \times U' \mid \{u', v'\} \notin E_r^{\tilde{\mathcal{F}}, \zeta'}\}|}{|U'|^2} \geq 1 - \frac{2hd}{(d')^2} \geq 1 - \frac{512h\ell^4}{d} \geq 1 - \frac{65536h\ell^6}{\delta k} \quad (7)$$

where the last two inequalities follow from $d' \geq d/(16\ell^2)$ and $d' \geq \delta k/(128\ell^2)$ respectively.

Let $\mathcal{F}' = \{f_S\}_{S \in U'}$. Observe that the expression on the left hand side of (7) is simply $\text{agree}_{\zeta'}(\mathcal{F}')$. Hence, by Lemma 27, there exists a function $g : [n] \rightarrow \{0, 1\}$ such that

$$\mathbb{E}_{S \in U'}[\text{disagr}(g, f_S)] \leq n \sqrt{\frac{65536h\ell^6}{\delta k} + \zeta'} = n \sqrt{\frac{65536h\ell^6}{\delta k} + \mu + 2\zeta/\gamma}.$$

In other words, U' is the desired subcollection, which completes our proof. \square

5 Soundness Analysis of the Reduction

We will next use our agreement theorem to analyze the soundness of our reduction. The soundness of our reduction can be stated more precisely as follows:

Theorem 29 *For any $\Delta \in \mathbb{N}$, let Φ be any 3-CNF formula with variable set X and clause set \mathcal{C} such that each variable appears in at most Δ clauses. Moreover, for any $0 < \eta, \zeta, \gamma, \mu < 1$ and $r, \ell, k, h, d \in \mathbb{N}$ such that $\ell \geq 2$, let \mathcal{T} be any collection of k subsets of \mathcal{C} such that \mathcal{T} is (r, ℓ, ζ) -intersection disperser and every subcollection $\tilde{\mathcal{T}} \subseteq \mathcal{T}$ of size h is (γ, μ) -uniform. If $\text{val}(\Phi) < 1 - \mu - (3\Delta/\gamma)\sqrt{4\Delta\mu + 6\Delta\zeta/\gamma}$, then*

$$\text{val}(\Gamma_{\Phi, \mathcal{T}}) < \frac{10 + 64(r\ell)^2 k^{1/\ell} + 65536h\ell^2/\mu}{k}.$$

Again, we will prove the contrapositive that if $\text{val}(\Gamma_{\Phi, \mathcal{T}})$ is large, then $\text{val}(\Phi)$ is also large. Recall that $\text{val}(\Gamma_{\Phi, \mathcal{T}})$ being large implies that there exists a labeling $\sigma = \{\sigma_T\}_{T \in \mathcal{T}}$ with high agreement probability. We would like to apply our agreement testing theorem. Note however that Theorem 28 only applies when the subsets of *variables* are “well-behaved” (i.e. satisfies uniformity and is an intersection disperser). However, in our construction, the subset of variables are not random, rather they are variable set of random subsets of clauses. Hence, we will first need to translate the “well-behavedness” from subsets of clauses to their corresponding variable sets; this is shown in Section 5.1. Once this is in place, we can apply Theorem 28, which gives us a global assignment that approximately agrees with many σ_T 's. We show in Section 5.2 that such assignment satisfies most of the constraint, which implies that $\text{val}(\Phi)$ must be large as desired. The full proof of Theorem 29 can then be found in Section 5.3.

5.1 Well-Behave Subsets of Clauses vs Well-Behave Subsets of Variables

For convenient, let us define an additional notation:

Definition 30 *Let Φ be any 3-CNF formula and \mathcal{T} be any subset of clauses of Φ . We use $\mathcal{S}_{\Phi, \mathcal{T}}$ to denote the collection $\{\text{var}(T)\}_{T \in \mathcal{T}}$ of subsets of variables.*

Note that the subsets in $\mathcal{S}_{\Phi, \mathcal{T}}$ are indeed the variable sets of our labeling $\sigma = \{\sigma_T\}_{T \in \mathcal{T}}$. Moreover, it is rather straightforward to see that both uniformity and intersection disperser conditions translate from \mathcal{T} to $\mathcal{S}_{\Phi, \mathcal{T}}$ with little loss in parameters, provided that each variable appears in bounded number of clauses. These observations are formalized and proved below.

Lemma 31 *Suppose that every variable in Φ appears in at least one and at most Δ clauses. If \mathcal{T} is (γ, μ) -uniform, then $\mathcal{S}_{\Phi, \mathcal{T}}$ is $(\gamma, 3\Delta\mu)$ -uniform.*

Proof. First, observe that, since each variable appears in at most Δ clauses, we have $n \geq m/\Delta$. Now, let $\mathcal{C}_{\geq \gamma} = \{C \in \mathcal{C} \mid \Pr_{T \in \mathcal{T}}[C \in T] \geq \gamma\}$ and $X_{\geq \gamma} = \{x \in X \mid \Pr_{S \in \mathcal{S}_{\Phi, \mathcal{T}}}[x \in S] \geq \gamma\}$. Recall that

(γ, μ) -uniformity of \mathcal{T} implies that $|\mathcal{C}_{\geq \gamma}| \geq (1 - \mu)m$. Observe that any $x \in \text{var}(\mathcal{C}_{\geq \gamma})$ must also be contained in $X_{\geq \gamma}$. Since every variable appears in at least one clause, we have that every variable $x \notin X_{\geq \gamma}$ must be in $\text{var}(\mathcal{C} \setminus \mathcal{C}_{\geq \gamma})$. As a result, $|X \setminus X_{\geq \gamma}| \leq 3\mu m$. From this and from $n \geq m/\Delta$, we arrive at the desired conclusion. \square

Lemma 32 *Suppose that every variable in Φ appears in at least one and at most Δ clauses. If \mathcal{T} is an (r, ℓ, η) -intersection disperser, then $\mathcal{S}_{\Phi, \mathcal{T}}$ is $(r, \ell, 3\Delta\eta)$ -uniform.*

Proof. Consider any r disjoint subcollections $\mathcal{S}^1 = \{S_{1,1}, \dots, S_{1,p_1}\}, \dots, \mathcal{S}^r = \{S_{r,1}, \dots, S_{r,p_r}\} \subseteq \mathcal{S}_{\Phi, \mathcal{T}}$ each of size at most ℓ . From our definition of $\mathcal{S}_{\Phi, \mathcal{T}}$, there is an r disjoint subcollections $\mathcal{T}^1 = \{T_{1,1}, \dots, T_{1,p_1}\}, \dots, \mathcal{T}^r = \{T_{r,1}, \dots, T_{r,p_r}\} \subseteq \mathcal{T}$ such that $S_{i,j} = \text{var}(T_{i,j})$ for all $i \in [r]$ and $j \in [p_i]$. Observe that

$$\bigcup_{i=1}^r \left(\bigcap_{S \in \mathcal{S}^i} S \right) \supseteq \text{var} \left(\bigcup_{i=1}^r \left(\bigcap_{T \in \mathcal{T}^i} T \right) \right).$$

Moreover, since \mathcal{T} is an (r, ℓ, η) -intersection disperser, we have $|\bigcup_{i=1}^r (\bigcap_{T \in \mathcal{T}^i} T)| \geq (1 - \eta)m$. As a result, since each variable appears in at least one clause, we indeed have $|\bigcup_{i=1}^r (\bigcap_{S \in \mathcal{S}^i} S)| \geq n - 3\eta m \geq (1 - 3\Delta\eta)n$ as desired. \square

5.2 Global Function with Many Agreements is a Good Assignment

In this subsection, we show that any global assignment that are approximately consistent with a collection of labels $\{\sigma_T\}_{T \in \mathcal{T}^*}$ must satisfy most of the constraints, assuming that \mathcal{T}^* is sufficiently uniform, which is stated more precisely below.

Lemma 33 *Let \mathcal{T}^* be any (γ, μ) -uniform collection of subsets of clauses and σ be any labeling of \mathcal{T}^* . If there exists $\psi : X \rightarrow \{0, 1\}$ such that $\mathbb{E}_{T \in \mathcal{T}^*} [\text{disagr}(\psi, \sigma_T)] \leq \nu n$, then $\text{val}(\psi) \geq 1 - \mu - 3\nu\Delta/\gamma$.*

The key to proving that ψ violates few clauses is that, if a clause C is violated, then, for each $T \in \mathcal{T}^*$ that contains C , σ_T and ψ must disagree on at least one of $\text{var}(C)$ because σ_T satisfies C but ψ violates it. Hence, if C appears often in \mathcal{T} , then it contributes to many disagreements between σ_T and ψ ; the uniformity condition helps us ensure that most C indeed appear often in \mathcal{T} . Comparing this lower bound against the assumed upper bound on the expected disagreements gives us the desired result. This intuition is formalized below.

Proof. Let $\mathcal{C}_{\geq \gamma}$ denote the set of all clauses that appear in at least γ fraction of $T \in \mathcal{T}^*$, i.e., $\mathcal{C}_{\geq \gamma} = \{C \in \mathcal{C} \mid \Pr_{T \in \mathcal{T}^*}[C \in T] \geq \gamma\}$. Since \mathcal{T}^* is (γ, μ) -uniform, we have $|\mathcal{C}_{\geq \gamma}| \geq (1 - \mu)m$.

Since each variable x appears in at most Δ clauses, we can obtain the following bound:

$$\begin{aligned} \mathbb{E}_{T \in \mathcal{T}^*} [\text{disagr}(\psi, \sigma_T)] &= \sum_{x \in X} \Pr_{T \in \mathcal{T}^*} [x \in \text{var}(T) \wedge \sigma_T(x) \neq \psi(x)] \\ &\geq \frac{1}{\Delta} \sum_{C \in \mathcal{C}_{\geq \gamma}} \sum_{x \in \text{var}(C)} \Pr_{T \in \mathcal{T}^*} [x \in \text{var}(T) \wedge \sigma_T(x) \neq \psi(x)] \\ &\geq \frac{1}{\Delta} \sum_{C \in \mathcal{C}_{\geq \gamma}} \sum_{T \in \mathcal{T}^*} \Pr [C \in T \wedge \sigma_T(x) \neq \psi(x)] \\ \text{(Union Bound)} &\geq \frac{1}{\Delta} \sum_{C \in \mathcal{C}_{\geq \gamma}} \Pr_{T \in \mathcal{T}^*} \left[C \in T \wedge \left(\bigvee_{x \in \text{var}(C)} \sigma_T(x) \neq \psi(x) \right) \right] \end{aligned}$$

$$\geq \frac{\gamma}{\Delta} \sum_{C \in \mathcal{C}_{\geq \gamma}} \Pr_{T \in \mathcal{T}^*} \left[\bigvee_{x \in \text{var}(C)} \sigma_T(x) \neq \psi(x) \mid C \in T \right] \quad (8)$$

Note here that we use the fact that each variable appears in at most Δ clauses in the first inequality and that the last inequality follows from the fact that each $C \in \mathcal{C}_{\geq \gamma}$ appears in at least γ fraction of $T \in \mathcal{T}^*$. The rest of the inequalities are trivial.

Let $\mathcal{C}_{\text{UNSAT}}$ denote the set of clauses violated by ψ . Observe that, for any $C \in \mathcal{C}_{\text{UNSAT}}$ and any $T \in \mathcal{T}^*$ such that $C \in T$, σ_T must disagree with ψ on at least one of $x \in \text{var}(C)$; this is simply because C is satisfied by σ_T but violated by ψ . In other words, for every $C \in \mathcal{C}_{\text{UNSAT}}$, we have

$$\Pr_{T \in \mathcal{T}^*} \left[\bigvee_{x \in \text{var}(C)} \sigma_T(x) \neq \psi(x) \mid C \in T \right] = 1. \quad (9)$$

(8), (9) and the assumption that $\mathbb{E}_{T \in \mathcal{T}^*} [\text{disagr}(\psi, \sigma_T)] \leq \nu n$ imply that

$$\nu \Delta n / \gamma \geq |\mathcal{C}_{\geq \gamma} \cap \mathcal{C}_{\text{UNSAT}}|.$$

Since $\mathcal{C}_{\geq \gamma} \geq m(1 - \mu)$ and $n \leq 3m$ (from every variable appears in at least one clause), we can conclude that $|\mathcal{C}_{\text{UNSAT}}| \leq \mu m + 3\nu \Delta m / \gamma$. As a result, $\text{val}(\psi) \geq 1 - \mu - 3\nu \Delta / \gamma$ as desired. \square

5.3 Putting Things Together: Proof of Theorem 29

Proof of Theorem 29. We may assume w.l.o.g. that each variable appears in at least one clause.

We will prove the theorem by contrapositive. Suppose that $\text{val}(\Gamma_{\Phi, \mathcal{T}}) \geq \frac{10+64(r\ell)^2 k^{1/\ell} + 65536h\ell^2/\mu}{k}$. This means that there exists a labeling $\sigma = \{\sigma_T\}_{T \in \mathcal{T}}$ such that $\text{val}(\sigma) \geq \frac{10+64(r\ell)^2 k^{1/\ell} + 65536h\ell^2/\mu}{k}$; this also means that, if we view σ as a collection of functions $\mathcal{F} = \{f_S\}_{S \in \mathcal{S}_{\Phi, \mathcal{T}}}$ where $f_{\text{var}(T)} = \sigma_T$, then $\text{agree}(\mathcal{F}) \geq \frac{10+64(r\ell)^2 k^{1/\ell} + 65536h\ell^2/\mu}{k}$. Let $\delta = \text{agree}(\mathcal{F})$.

Furthermore, Lemmas 32 and 31 imply that $\mathcal{S}_{\Phi, \mathcal{T}}$ is an $(r, \ell, 3\Delta\zeta)$ -intersection disperser and every subcollection of $\mathcal{S}_{\Phi, \mathcal{T}}$ of size h is $(\gamma, 3\Delta\mu)$ -uniform respectively. This enables us to apply Theorem 28 on \mathcal{F} , which yields a subcollection $\mathcal{S}' \subseteq \mathcal{S}_{\Phi, \mathcal{T}}$ of size at least $\frac{\delta k}{128\ell^2} \geq h$ and $g : X \rightarrow \{0, 1\}$ such that

$$\mathbb{E}_{S \in \mathcal{S}'} [\text{disagr}(g, f_S)] \leq n \sqrt{\frac{65536h\ell^6}{\delta k} + 3\Delta\mu + 6\Delta\zeta/\gamma} \leq n \sqrt{4\Delta\mu + 6\Delta\zeta/\gamma}$$

where the second inequality follows from $\delta k \geq 65536h\ell^2/\mu$.

Let \mathcal{S}^* be the subcollection of \mathcal{S}' of size h that minimizes $\mathbb{E}_{S \in \mathcal{S}^*} [\text{disagr}(g, f_S)]$. Observe that $\mathbb{E}_{S \in \mathcal{S}^*} [\text{disagr}(g, f_S)] \leq \mathbb{E}_{S \in \mathcal{S}'} [\text{disagr}(g, f_S)] \leq n \sqrt{4\Delta\mu + 6\Delta\zeta/\gamma}$. This is equivalent to saying that there exists a subcollection $\mathcal{T}^* \subseteq \mathcal{T}$ of size h such that $\mathbb{E}_{T \in \mathcal{T}^*} [\text{disagr}(g, \sigma_T)] \leq n \sqrt{4\Delta\mu + 6\Delta\zeta/\gamma}$.

Since every subcollection of \mathcal{T} of size h is (γ, μ) -uniform, we can apply Lemma 33 to infer that $\text{val}(\Phi) \geq 1 - \mu - (3\Delta/\gamma) \sqrt{4\Delta\mu + 6\Delta\zeta/\gamma}$ as desired. \square

6 Proof of Inapproximability Results of 2-CSPs

The inapproximability results for 2-CSPs can be shown simply by plugging in appropriate parameters to Theorem 29. More specifically, for ETH-hardness, since there is a polylog m loss in the PCP Theorem (Theorem 6), we need to select our $\alpha = 1/\text{polylog}m$ so that the size (and running time) of the reduction is $2^{o(m)}$. Now, observe that the parameter r in Theorem 29 for the intersection disperser property grows with $(1/\alpha)^\ell$ (see Lemma 20). Since the soundness guarantee in Theorem 29 is of the form $k^{O(1/\ell)}(r\ell)^{O(1)}/k = k^{O(1/\ell)}(1/\alpha)^{O(\ell)}/k$, it is minimized when ℓ is roughly $\sqrt{\log k}$, which yields the soundness of $2^{(\log k)^{1/2+o(1)}}$. Other parameters are chosen accordingly.

Proof of Theorem 1. Let c, ε, Δ be constants from Theorem 6.

For any 3-CNF formula $\tilde{\Phi}$ with m clauses, let us first apply the nearly-linear size PCP from Theorem 6 to produce a 3-CNF formula Φ with $m' = O(m \log^c m)$ clauses. Let us also define the following parameters:

- $\alpha = \frac{1}{(\log m)^{c+1}},$
- $\gamma = \frac{\alpha}{2} = \frac{1}{2(\log m)^{c+1}},$
- $\mu = \frac{\varepsilon^2 \gamma^2}{288 \Delta^3} = \Theta_{\varepsilon, \Delta} \left(\frac{1}{(\log m)^{2c+2}} \right),$
- $\zeta = \frac{\varepsilon^2 \gamma^3}{432 \Delta^3} = \Theta_{\varepsilon, \Delta} \left(\frac{1}{(\log m)^{3c+3}} \right),$
- $\ell = (\log m)^{1/4},$
- $r = \lceil \frac{\ln(2/\zeta)}{\alpha^\ell} \rceil = 2^{\Theta_{\varepsilon, \Delta, c}((\log m)^{1/4} \log \log m)},$
- $h = \lceil 8 \ln(2/\mu)/\alpha \rceil = \Theta_{\varepsilon, \Delta, c}((\log m)^{c+1}),$
- $k = 2^{\ell^2} = 2^{\sqrt{\log m}}.$

We then use Lemma 20 with the above parameters $\alpha, \mu, \zeta, k, \ell$ to construct a collection \mathcal{S} of subsets of clauses of Φ such that the following conditions hold.

- Every subset in \mathcal{S} has size at most $2\alpha m' = o(m)$.
- \mathcal{S} is a (r, ℓ, ζ) -disperser.
- Any subcollection $\tilde{\mathcal{S}} \subseteq \mathcal{S}$ of size h is $(\alpha/2, \mu)$ -uniform.

Note that, for our choices of parameter, the parameter m_0 of Lemma 20 is $2^{O_{\varepsilon, \Delta, c}((\log m)^{1/4} \log \log m)} = 2^{o(\log m)}$. This means that, for sufficiently large m , we indeed have $m' \geq m \geq m_0$ and the running time needed to produce \mathcal{S} is $\text{poly}(m)2^{O(m_0 k^2)} = 2^{o(m)}$. Note that we assume without loss of generality here that $m' \geq m$; if this is not the case, we can simply copy the formula $\Phi \lceil m/m' \rceil$ times using new variables each time, which does not change the value of the formula.

We now consider the 2-CSP instance $\Gamma_{\Phi, \mathcal{S}}$. Observe that the running time used to create $\Gamma_{\Phi, \mathcal{S}}$ (and hence also the size of $\Gamma_{\Phi, \mathcal{S}}$) is no more than $\text{poly}(k) \cdot 2^{o(\alpha m')} = 2^{o(m)}$. Moreover, if $\text{val}(\tilde{\Phi}) = 1$, then $\text{val}(\Phi) = 1$ and it is easy to see that $\text{val}(\Gamma_{\Phi, \mathcal{S}}) = 1$ as well.

On the other hand, if $\text{val}(\tilde{\Phi}) < 1$, then $\text{val}(\Phi) < 1 - \varepsilon$. Due to our choice of parameters, we can apply Theorem 29, which implies that

$$\text{val}(\Gamma_{\Phi, \mathcal{S}}) < \frac{O((r\ell)^2 k^{1/\ell} + h\ell^2/\mu)}{k} = 2^{O_{\varepsilon, \Delta, c}((\log m)^{1/4} \log \log m)}/k = 2^{O_{\varepsilon, \Delta, c}(\sqrt{\log k} \log \log k)}/k.$$

For sufficiently large m (depending only on $c, \varepsilon, \Delta, \rho$), this term is at most $2^{(\log k)^{1/2+\rho}}/k$. Hence, if

there exists a polynomial time that can distinguish the two cases in the theorem statement, we can run this algorithm on $\Gamma_{\Phi, \mathcal{S}}$ to decide whether $\tilde{\Phi}$ is satisfiable in $2^{o(m)}$ time, contradicting ETH. \square

For Gap-ETH-hardness, we do not incur a loss of $\text{polylog } m$ from the PCP Theorem anymore. Thus, it suffices to chose α to be any function that converges to zero as k goes to ∞ (e.g. $\alpha = 1 / \log \log k$), and k can now be independent of m . The rest of the analysis remains unchanged.

Proof of Theorem 2. Let $\delta, \varepsilon, \Delta$ be the constants from Theorem 2. For any positive integer k , define the parameters as follows:

- $\alpha = \frac{1}{\log \log k}$,
- $\gamma = \frac{\alpha}{2} = \frac{1}{2(\log \log k)}$,
- $\mu = \frac{\varepsilon^2 \gamma^2}{288 \Delta^3} = \Theta_{\varepsilon, \Delta} \left(\frac{1}{(\log \log k)^2} \right)$,
- $\zeta = \frac{\varepsilon^2 \gamma^3}{432 \Delta^3} = \Theta_{\varepsilon, \Delta} \left(\frac{1}{(\log \log k)^3} \right)$,
- $\ell = \sqrt{\log k}$,
- $r = \lceil \frac{\ln(2/\zeta)}{\alpha^\ell} \rceil = 2^{\Theta_{\varepsilon, \Delta, c}(\sqrt{\log k} \log \log \log k)}$,
- $h = \lceil 8 \ln(2/\mu) / \alpha \rceil = \Theta_{\varepsilon, \Delta, c}(\log \log k)$.

Consider any 3-CNF formula Φ with m clauses such that each variable appears in at most Δ clauses. We then use Lemma 20 with the above parameters $\alpha, \mu, \eta, k, \ell$ to construct a collection \mathcal{S} of subsets of clauses of Φ such that the following conditions hold.

- Every subset in \mathcal{S} has size at most $2\alpha m$.
- \mathcal{S} is a (r, ℓ, η) -dispenser.
- Any subcollection $\tilde{\mathcal{S}} \subseteq \mathcal{S}$ of size h is $(\alpha/2, \mu)$ -uniform.

Note that, for our choices of parameter, the parameter m_0 of Lemma 20 is a function of k . This means that, for sufficiently large m (which depends on k), we indeed have $m \geq m_0$ and the running time needed to produce \mathcal{S} is $\text{poly}(m)2^{O(m_0 k^2)} = O_k(\text{poly}(m))$.

We now consider the 2-CSP instance $\Gamma_{\Phi, \mathcal{S}}$. Observe that the running time used to create $\Gamma_{\Phi, \mathcal{S}}$ (and hence also the size of $\Gamma_{\Phi, \mathcal{S}}$) is no more than $\text{poly}(k) \cdot 2^{O(\alpha m)} = 2^{O(m/\log \log k)}$. Moreover, if $\text{val}(\Phi) = 1$, it is easy to see that $\text{val}(\Gamma_{\Phi, \mathcal{S}}) = 1$ as well.

Suppose that $\text{val}(\Phi) < 1 - \varepsilon$. Due to our choice of parameters, we can apply Theorem 29, which implies that

$$\text{val}(\Gamma_{\Phi, \mathcal{S}}) < \frac{O(k^{1/\ell}(r\ell)^2) + h\ell^2/\mu}{k} = 2^{O_{\varepsilon, \Delta}(\log \log k / \sqrt{\log k})} / k.$$

For sufficiently large k (depending on $\varepsilon, \Delta, \rho$), this term is at most $2^{(\log k)^{1/2+\rho}} / k$.

If there exists a $g(k) \cdot (nk)^D$ -time algorithm that can distinguish the two cases in the theorem statement for some constant D , then pick sufficiently large k such that the time needed to produce $\Gamma_{\Phi, \mathcal{S}}$ is $O(2^{\delta m})$ and its size is at most $2^{\delta m/D}$, and that $\text{val}(\Gamma_{\Phi, \mathcal{S}}) < 2^{1/(\log k)^{1/2+\rho}} / k$ whenever $\text{val}(\Phi) < 1 - \varepsilon$. When we run this algorithm on $\Gamma_{\Phi, \mathcal{S}}$ for such k , the algorithm can distinguish between $\text{val}(\Phi) = 1$ and $\text{val}(\Phi) < 1 - \varepsilon$ in $O(2^{\delta m})$ time, which contradicts Gap-ETH. \square

7 Inapproximability of Directed Steiner Network

We now move on to prove hardness of approximation of DSN by simply plugging in the our main theorems to known reductions from 2-CSPs to DSN. The properties of the reduction are stated in the lemma below. Note that, while the reduction is attributed to Dodis and Khanna [DK99], the lemma below is extracted from [CFM17] since, in [DK99], the full description and its properties are left out due to space constraint.

Lemma 34 ([CFM17, Lemma 27]) *There exists a polynomial time reduction that, given a 2-CSP instance¹⁰ Γ with the constraint graph being a complete graph on k variables, produces an edge-weighted directed graph $G = (V, E)$ and a set of demand pairs $\mathcal{D} = \{(s_1, t_1), \dots, (s_{k'}, t_{k'})\}$ such that*

- (Completeness) *If $\text{val}(\Gamma) = 1$, then there exists a subgraph H of cost 1 that satisfies all demands.*
- (Soundness) *If $\text{val}(\Gamma) < \gamma$, every subgraph satisfying all demand pairs has cost more than $\sqrt{2/\gamma}$.*
- (Parameter Dependency) $k' = k^2 - k$.

Note that the exponent $1/4$ in the hardness of approximating DSN comes from two places: we lose a square factor in the parameter (i.e. $k' = \Theta(k^2)$) and another square factor in the gap.

Proof of Corollary 3. Suppose for the sake of contradiction that, for some constant $\rho' > 0$, there exists a polynomial time $2^{(\log k')^{1/2+\rho'}} / (k')^{1/4}$ -approximation algorithm where k' is the number of demand pairs; let us call this algorithm \mathbb{A} . Moreover, let ρ be any constant smaller than ρ' .

Given a 2-CSP instance Γ with complete constraint graph on k variables, we invoke Lemma 34 to produce a DSN instance (G, D) where $|D| = k' = k^2 - k$. From the completeness of the construction, we have that, if $\text{val}(\Gamma) = 1$, then the optimum of (G, D) is also 1. On the other hand, if $\text{val}(\Gamma) < 2^{(\log k)^{1/2+\rho}} / k$, then the optimum of (G, D) must be more than $\sqrt{2k} / 2^{(\log k)^{1/2+\rho}}$, which is at least $(k')^{1/4} / 2^{(\log k')^{1/2+\rho'}}$ when k is sufficiently large. Hence, by running algorithm \mathbb{A} , we can distinguish these two cases of Γ in polynomial time. From Theorem 1, this contradicts ETH. \square

Proof of Corollary 4. Suppose for the sake of contradiction that, for some constant $\rho' > 0$ and for some function g , there exists a $g(k') \cdot (nk')^{O(1)}$ -time $2^{(\log k')^{1/2+\rho'}} / (k')^{1/4}$ -approximation algorithm where k' is the number of demand pairs; let us call this algorithm \mathbb{A} . Moreover, let ρ be any constant smaller than ρ' .

Given a 2-CSP instance Γ with complete constraint graph on k variables, we invoke Lemma 34 to produce a DSN instance (G, D) where $|D| = k' = k^2 - k$. From the completeness of the construction, if $\text{val}(\Gamma) = 1$, then the optimum of (G, D) is also 1. On the other hand, if $\text{val}(\Gamma) < 2^{(\log k)^{1/2+\rho}} / k$, then the optimum of (G, D) must be more than $\sqrt{2k} / 2^{(\log k)^{1/2+\rho}}$, which is at least $(k')^{1/4} / 2^{(\log k')^{1/2+\rho'}}$ when k is sufficiently large. Hence, by running algorithm \mathbb{A} , we can distinguish these two cases of Γ in time $t(k) \cdot |\Gamma|^{O(1)}$ where $t(k) = g(k^2 - k)$. From Theorem 2, this contradicts Gap-ETH. \square

8 Conclusion and Discussions

In this work, we show that 2-CSP is ETH-hard to approximate to within a factor of $k^{1-o(1)}$ where k denotes the number of variables. This ratio is nearly optimal since a trivial algorithm yields an $O(k)$ -approximation for the problem. Under Gap-ETH, we strengthen our result by improving the

¹⁰Lemma 27 of [CFM17] states this reduction in terms of Maximum Colored Subgraph Isomorphism. However, it is easy to see that the reduction also works with 2-CSPs as well.

lower order term in the inapproximability factor and ruling out not only polynomial time algorithm but FPT algorithms parameterized by k . Due to a known reduction, our results also imply $k^{1/4-o(1)}$ hardness of approximating DSN where k denotes the number of demand pairs.

Of course the polynomial sliding scale conjecture still remains open after our work and, as touched upon in the introduction, resolving the conjecture will help advance our understanding of approximability of many problems. Even without fully resolving the conjecture, it may still be good to further study the interaction between the number of variables k and the alphabet size n . For instance, while we show the inapproximability result with ratio almost linear in k , the dependency between n and k is quite bad; in particular, in our ETH-hardness reduction, n is $2^{2^{(\log k)^d}}$ for some constant $d > 0$. Would it be possible to improve this dependency (say, to $n = k^{\text{polylog} k}$)?

On the other hand, as explained earlier, k must be independent of n in the parameterized setting and hence our question above does not apply to this regime. However, one intriguing question in this area is whether a parameterized hardness of approximation for 2-CSPs can be proved under any assumption weaker than Gap-ETH (e.g. ETH). This is not known even for a constant inapproximability factor. In fact, it is not hard to see that inapproximability of parameterized 2-CSPs implies inapproximability of parameterized clique; the latter is a well-studied problem that has so far resisted attempts at proving inapproximability from any assumption except Gap-ETH [CHK13, HKK13, KS16]. Note that this is in contrast to some other parameterized problems, such as dominating set, for which W[1]-hardness of approximation is known [CL16, KLM18]¹¹.

Another interesting research direction is to try to prove similar hardness results for other problems. For example, Densest k -Subgraph (DkS) is one such candidate problem; similar to 2-CSPs with k variables, the problem can be approximated trivially to within $O(k)$ -factor and no polynomial time (or even FPT time parameterized by k) $k^{1-\varepsilon}$ -approximation algorithm is known for the problem for any $\varepsilon > 0$. Hence, it may be possible to prove ETH-hardness of factor $k^{1-o(1)}$ for DkS as well.

Acknowledgment

We would like to thank Prahladh Harsha for many useful discussions. Pasin would also like to thank Rajesh Chitnis and Andreas Emil Feldmann for insightful discussions regarding directed steiner network problems.

References

- [AB17] Amir Abboud and Greg Bodwin. Reachability preservers: New extremal bounds and approximation algorithms. *CoRR*, abs/1710.11250, 2017.
- [AIM14] Scott Aaronson, Russell Impagliazzo, and Dana Moshkovitz. AM with multiple merlins. In *CCC*, pages 44–55, 2014.
- [ALM⁺98] Sanjeev Arora, Carsten Lund, Rajeev Motwani, Madhu Sudan, and Mario Szegedy. Proof verification and the hardness of approximation problems. *J. ACM*, 45(3):501–555, May 1998.
- [AM09] Per Austrin and Elchanan Mossel. Approximation resistant predicates from pairwise independence. *Computational Complexity*, 18(2):249–271, 2009.
- [App17] Benny Applebaum. Exponentially-hard gap-CSP and local PRG via local hardcore functions. In *FOCS*, pages 836–847, 2017.

¹¹See also [Lin15] which [CL16] relies heavily on.

- [ARW17] Amir Abboud, Aviad Rubinfeld, and Ryan Williams. Distributed PCP theorems for hardness of approximation in P. In *FOCS*, pages 25–36, 2017.
- [AS97] Sanjeev Arora and Madhu Sudan. Improved low degree testing and its applications. In *STOC*, pages 485–495, May 1997.
- [AS98] Sanjeev Arora and Shmuel Safra. Probabilistic checking of proofs: A new characterization of NP. *J. ACM*, 45(1):70–122, January 1998.
- [BBM⁺13] Piotr Berman, Arnab Bhattacharyya, Konstantin Makarychev, Sofya Raskhodnikova, and Grigory Yaroslavtsev. Approximation algorithms for spanner problems and directed steiner forest. *Inf. Comput.*, 222:93–107, 2013.
- [BEKP15] Edouard Bonnet, Bruno Escoffier, Eun Jung Kim, and Vangelis Th. Paschos. On subexponential and FPT-time inapproximability. *Algorithmica*, 71(3):541–565, 2015.
- [BGH⁺06] Eli Ben-Sasson, Oded Goldreich, Prahladh Harsha, Madhu Sudan, and Salil P. Vadhan. Robust PCPs of proximity, shorter PCPs, and applications to coding. *SIAM J. Comput.*, 36(4):889–974, 2006.
- [BGKM18] Arnab Bhattacharyya, Suprovat Ghoshal, Karthik C. S., and Pasin Manurangsi. Parameterized intractability of even set and shortest vector problem from Gap-ETH. In *ICALP*, 2018. To appear.
- [BGLR93] Mihir Bellare, Shafi Goldwasser, Carsten Lund, and Alexander Russell. Efficient probabilistically checkable proofs and applications to approximations. In *STOC*, pages 294–304, 1993.
- [BKK⁺16] Eli Ben-Sasson, Yohay Kaplan, Swastik Kopparty, Or Meir, and Henning Stichtenoth. Constant rate PCPs for Circuit-SAT with sublinear query complexity. *J. ACM*, 63(4):32:1–32:57, 2016.
- [BKRW17] Mark Braverman, Young Kun-Ko, Aviad Rubinfeld, and Omri Weinstein. ETH hardness for densest- k -subgraph with perfect completeness. In *SODA*, pages 1326–1341, 2017.
- [BKW15] Mark Braverman, Young Kun-Ko, and Omri Weinstein. Approximating the best nash equilibrium in $n^{o(\log n)}$ -time breaks the exponential time hypothesis. In *SODA*, pages 970–982, 2015.
- [BS08] Eli Ben-Sasson and Madhu Sudan. Short PCPs with polylog query complexity. *SIAM J. Comput.*, 38(2):551–607, 2008.
- [CCC⁺99] Moses Charikar, Chandra Chekuri, To-Yat Cheung, Zuo Dai, Ashish Goel, Sudipto Guha, and Ming Li. Approximation algorithms for directed steiner problems. *J. Algorithms*, 33(1):73–91, 1999.
- [CCK⁺17] Parinya Chalermsook, Marek Cygan, Guy Kortsarz, Bundit Laekhanukit, Pasin Manurangsi, Danupon Nanongkai, and Luca Trevisan. From Gap-ETH to FPT-inapproximability: Clique, dominating set, and more. In *FOCS*, pages 743–754, 2017.
- [CDKL17] Eden Chlamtác, Michael Dinitz, Guy Kortsarz, and Bundit Laekhanukit. Approximating spanners and directed steiner forest: Upper and lower bounds. In *SODA*, pages 534–553, 2017.

- [CEGS11] Chandra Chekuri, Guy Even, Anupam Gupta, and Danny Segev. Set connectivity problems in undirected graphs and the directed steiner network problem. *ACM Trans. Algorithms*, 7(2):18:1–18:17, 2011.
- [CFM17] Rajesh Chitnis, Andreas Emil Feldmann, and Pasin Manurangsi. Parameterized approximation algorithms for directed steiner network problems. *CoRR*, abs/1707.06499, 2017.
- [Cha16] Siu On Chan. Approximation resistance from pairwise-independent subgroups. *J. ACM*, 63(3):27:1–27:32, 2016.
- [Che18] Lijie Chen. On the hardness of approximate and exact (bichromatic) maximum inner product. In *CCC*, 2018. To appear.
- [CHK11] Moses Charikar, MohammadTaghi Hajiaghayi, and Howard J. Karloff. Improved approximation algorithms for label cover problems. *Algorithmica*, 61(1):190–206, 2011.
- [CHK13] Rajesh Hemant Chitnis, MohammadTaghi Hajiaghayi, and Guy Kortsarz. Fixed-parameter and approximation algorithms: A new look. In *IPEC*, pages 110–122, 2013.
- [CL16] Yijia Chen and Bingkai Lin. The constant inapproximability of the parameterized dominating set problem. In *FOCS*, pages 505–514, 2016.
- [CMMV17] Eden Chlamtác, Pasin Manurangsi, Dana Moshkovitz, and Aravindan Vijayaraghavan. Approximation algorithms for label cover and the log-density threshold. In *SODA*, pages 900–919, 2017.
- [DFK⁺11] Irit Dinur, Eldar Fischer, Guy Kindler, Ran Raz, and Shmuel Safra. PCP characterizations of NP: Toward a polynomially-small error-probability. *Computational Complexity*, 20(3):413–504, 2011.
- [DFS16] Argyrios Deligkas, John Fearnley, and Rahul Savani. Inapproximability results for approximate nash equilibria. In *WINE*, pages 29–43, 2016.
- [DHK15] Irit Dinur, Prahladh Harsha, and Guy Kindler. Polynomially low error PCPs with polyloglog n queries via modular composition. In *STOC*, pages 267–276, 2015.
- [Din07] Irit Dinur. The PCP theorem by gap amplification. *J. ACM*, 54(3):12, 2007.
- [Din16] Irit Dinur. Mildly exponential reduction from gap 3SAT to polynomial-gap label-cover. *ECCC*, 23:128, 2016.
- [DK99] Yevgeniy Dodis and Sanjeev Khanna. Design networks with bounded pairwise distance. In *STOC*, pages 750–759, 1999.
- [DK17] Irit Dinur and Tali Kaufman. High dimensional expanders imply agreement expanders. In *FOCS*, pages 974–985, 2017.
- [DM18] Irit Dinur and Pasin Manurangsi. ETH-hardness of approximating 2-CSPs and directed steiner network. In *ITCS*, pages 36:1–36:20, 2018.
- [DN17] Irit Dinur and Inbal Livni Navon. Exponentially small soundness for the direct product z-test. In *CCC*, pages 29:1–29:50, 2017.
- [DS14] Irit Dinur and David Steurer. Analytical approach to parallel repetition. In *STOC*, pages 624–633, 2014.

- [FKN12] Moran Feldman, Guy Kortsarz, and Zeev Nutov. Improved approximation algorithms for directed steiner forest. *J. Comput. Syst. Sci.*, 78(1):279–292, 2012.
- [Gri01] Dima Grigoriev. Linear lower bound on degrees of positivstellensatz calculus proofs for the parity. *Theor. Comput. Sci.*, 259(1-2):613–622, 2001.
- [Hås01] Johan Håstad. Some optimal inapproximability results. *J. ACM*, 48(4):798–859, 2001.
- [HKK13] Mohammad Taghi Hajiaghayi, Rohit Khandekar, and Guy Kortsarz. The foundations of fixed parameter inapproximability. *CoRR*, abs/1310.2711, 2013.
- [IKW12] Russell Impagliazzo, Valentine Kabanets, and Avi Wigderson. New direct-product testers and 2-query PCPs. *SIAM J. Comput.*, 41(6):1722–1768, 2012.
- [IP01] Russell Impagliazzo and Ramamohan Paturi. On the complexity of k-SAT. *J. Comput. Syst. Sci.*, 62(2):367–375, 2001.
- [IPZ01] Russell Impagliazzo, Ramamohan Paturi, and Francis Zane. Which problems have strongly exponential complexity? *J. Comput. Syst. Sci.*, 63(4):512–530, 2001.
- [Kho02] Subhash Khot. On the power of unique 2-prover 1-round games. In *STOC*, pages 767–775, 2002.
- [KKMO07] Subhash Khot, Guy Kindler, Elchanan Mossel, and Ryan O’Donnell. Optimal inapproximability results for MAX-CUT and other 2-variable CSPs? *SIAM J. Comput.*, 37(1):319–357, 2007.
- [KKT16] Guy Kindler, Alexandra Kolla, and Luca Trevisan. Approximation of non-boolean 2CSP. In *SODA*, pages 1705–1714, 2016.
- [KLM18] Karthik C. S., Bundit Laekhanukit, and Pasin Manurangsi. On the parameterized complexity of approximating dominating set. In *STOC*, 2018. To appear.
- [KS16] Subhash Khot and Igor Shinkar. On hardness of approximating the parameterized clique problem. In *ITCS*, pages 37–45, 2016.
- [KST54] Tamás Kővári, Vera T. Sós, and Pál Turán. On a problem of K. Zarankiewicz. *Colloquium Mathematicae*, 3(1):50–57, 1954.
- [Lin15] Bingkai Lin. The parameterized complexity of k -Biclique. In *SODA*, pages 605–615, 2015.
- [Man17] Pasin Manurangsi. Almost-polynomial ratio ETH-hardness of approximating densest k -subgraph. In *STOC*, pages 954–961, 2017.
- [MM17] Pasin Manurangsi and Dana Moshkovitz. Improved approximation algorithms for projection games. *Algorithmica*, 77(2):555–594, 2017.
- [Mos17] Dana Moshkovitz. Low-degree test with polynomially small error. *Computational Complexity*, 26(3):531–582, 2017.
- [MR10] Dana Moshkovitz and Ran Raz. Sub-constant error probabilistically checkable proof of almost-linear size. *Computational Complexity*, 19(3):367–422, 2010.
- [MR16] Pasin Manurangsi and Prasad Raghavendra. A birthday repetition theorem and complexity of approximating dense CSPs. *CoRR*, abs/1607.02986, 2016.

- [MR17] Pasin Manurangsi and Aviad Rubinfeld. Inapproximability of VC dimension and littlestone’s dimension. In *COLT*, pages 1432–1460, 2017.
- [Pel07] David Peleg. Approximation algorithms for the label-cover_{max} and red-blue set cover problems. *J. Discrete Algorithms*, 5(1):55–64, 2007.
- [Rag08] Prasad Raghavendra. Optimal algorithms and inapproximability results for every CSP? In *STOC*, pages 245–254, 2008.
- [Raz98] Ran Raz. A parallel repetition theorem. *SIAM J. Comput.*, 27(3):763–803, 1998.
- [RS96] Ronitt Rubinfeld and Madhu Sudan. Robust characterizations of polynomials with applications to program testing. *SIAM J. Comput.*, 25(2):252–271, 1996.
- [RS97] Ran Raz and Shmuel Safra. A sub-constant error-probability low-degree test, and a sub-constant error-probability PCP characterization of NP. In *STOC*, pages 475–484, 1997.
- [Rub16a] Aviad Rubinfeld. Detecting communities is hard, and counting them is even harder. *CoRR*, abs/1611.08326, 2016.
- [Rub16b] Aviad Rubinfeld. Settling the complexity of computing approximate two-player nash equilibria. In *FOCS*, pages 258–265, 2016.
- [Rub17] Aviad Rubinfeld. Honest signaling in zero-sum games is hard, and lying is even harder. In *ICALP*, pages 77:1–77:13, 2017.
- [Rub18] Aviad Rubinfeld. Hardness of approximate nearest neighbor search. In *STOC*, 2018. To appear.
- [Sch08] Grant Schoenebeck. Linear level lasserre lower bounds for certain k-csps. In *FOCS*, pages 593–602, 2008.

A Constructing Well-Behaved Sets

The main goal of this section is to prove Lemma 20. We divide this section into two parts. In the first part (A.1), we use probabilistic argument to show that a collection of random subsets satisfied properties as required in Lemma 20. Then, in the section part (A.2), we show how to construct these subsets deterministically in subexponential time, which is required if a deterministic reduction from 3-SAT to 2-CSP is sought.

A.1 Random Sets Behave Well

For any universe \mathcal{U} and any $0 < \alpha < 1$, let $\mathcal{D}_{\mathcal{U},\alpha}$ denote the distribution on subsets of \mathcal{U} where $S \sim \mathcal{D}_{\mathcal{U},\alpha}$ is generated by including each element $u \in \mathcal{U}$ independently with probability α .

The main lemma of this section is the following, which implies the existential part of Lemma 20.

Lemma 35 *For any $0 < \alpha, \mu, \eta < 1$ and any $m, k, \ell \in \mathbb{N}$, let S_1, \dots, S_k be subsets of an m -element universe \mathcal{U} drawn independently at random from $\mathcal{D}_{\mathcal{U},\alpha}$. Then, with probability at least $1 - 2^{\log k(\lceil 8 \ln(2/\mu)/\alpha \rceil) - \mu^2 m/16} - 2^{\ell \lceil \ln(2/\eta)/\alpha^\ell \rceil \log(2k) - \eta m/6} - 2^{-\alpha m/3}$, the following properties hold:*

- (Size) Every subset has size at most $2\alpha m$.
- (Intersection Disperser) the collection $\mathcal{S} = \{S_1, \dots, S_k\}$ is a $(\lceil \ln(2/\eta)/\alpha^\ell \rceil, \ell, \eta)$ -disperser.

- (Uniformity) Any subcollection $\tilde{\mathcal{S}} \subseteq \mathcal{S}$ of size $\lceil 8 \ln(2/\mu)/\alpha \rceil$ is $(\alpha/2, \mu)$ -uniform.

By union bound, the above lemma is an immediate implication of Lemmas 36, 37 and 38, which are stated and proved below. All proofs consist only of straightforward probabilistic arguments.

Lemma 36 For any $0 < \alpha, \mu < 1$ and any $m, k \in \mathbb{N}$, let S_1, \dots, S_k be subsets of an m -element universe \mathcal{U} drawn independently at random from $\mathcal{D}_{\mathcal{U}, \alpha}$. Then, for every subcollection $\tilde{\mathcal{S}} \subseteq \mathcal{S}$ of size $\lceil 8 \ln(2/\mu)/\alpha \rceil$, $\tilde{\mathcal{S}}$ is $(\alpha/2, \mu)$ -uniform with probability at least $1 - 2^{\log k (\lceil 8 \ln(2/\mu)/\alpha \rceil) - \mu^2 m / 16}$.

Proof. Consider any subcollection $\tilde{\mathcal{S}} = \{S_{j_1}, \dots, S_{j_h}\}$ where $h = \lceil 8 \ln(2/\mu)/\alpha \rceil$. We will calculate the probability that $\tilde{\mathcal{S}}$ is $(\alpha/2, \mu)$ -uniform and use union bound over all such $\tilde{\mathcal{S}}$'s to derived the desired result.

Recall that, for each $u \in \mathcal{U}$, u is included independently in each set S_{j_i} with probability α . Hence, we can apply Chernoff bound to lower bound the probability that u is included in less than $\alpha/2$ fraction of the subsets in $\tilde{\mathcal{S}}$; in particular, this yields the following inequality.

$$\Pr \left[|\{S \in \tilde{\mathcal{S}} \mid u \in S\}| \geq (\alpha/2)|\tilde{\mathcal{S}}| \right] \geq 1 - e^{-\alpha|\tilde{\mathcal{S}}|/8} \geq 1 - \mu/2 \quad (10)$$

where the second inequality comes from $|\tilde{\mathcal{S}}| = h \geq 8 \ln(2/\mu)/\alpha$.

Now, note again that the event $|\{S \in \tilde{\mathcal{S}} \mid u \in S\}| \geq (\alpha/2)|\tilde{\mathcal{S}}|$ is independent for each $u \in \mathcal{U}$. Hence, we can again apply Chernoff bound to lower bound the probability that this event occurs for at least $(1 - \mu)$ fraction of $u \in \mathcal{U}$, which gives the following bound:

$$\Pr \left[|\{u \in \mathcal{U} \mid |\{S \in \tilde{\mathcal{S}} \mid u \in S\}| \geq (\alpha/2)|\tilde{\mathcal{S}}|\}| \geq (1 - \mu)m \right] \geq 1 - e^{-(\mu^2(1-\mu/2)m)/8} \geq 1 - e^{-\mu^2 m / 16}.$$

In other words, for each subcollection \mathcal{S} of size at least $8 \ln(2/\mu)/\alpha$, \mathcal{S} is not $(\alpha/2, \mu)$ -uniform with probability at most $e^{-\mu^2 m / 16}$. Since there are no more than k^h such subcollection, union bound implies that the probability that every subcollection of size h is $(\alpha/2, \mu)$ -uniform is at least $1 - k^h e^{-\mu^2 m / 16} \geq 1 - 2^{h \log k - \mu^2 m / 16}$. \square

Lemma 37 For any $0 < \alpha, \eta < 1$ and any $m, k, \ell \in \mathbb{N}$, let $r = \lceil \ln(2/\eta)/\alpha^\ell \rceil$ and let S_1, \dots, S_k be subsets of an m -element universe \mathcal{U} drawn independently at random from $\mathcal{D}_{\mathcal{U}, \alpha}$. Then, the collection $\mathcal{S} = \{S_1, \dots, S_k\}$ is (r, ℓ, η) -intersection disperser with probability at least $1 - 2^{\ell r \log(2k) - \eta m / 6}$.

Proof. Consider any disjoint subcollections $\mathcal{S}^1 = \{S_{j_{1,1}}, \dots, S_{j_{1,h_1}}\}, \dots, \mathcal{S}^r = \{S_{j_{r,1}}, \dots, S_{j_{r,h_r}}\}$ where $h_1, \dots, h_r \leq \ell$. We will compute the probability that $|\bigcup_{i=1}^r (\bigcap_{S \in \mathcal{S}^i} S)| \geq (1 - \eta)m$ and then use union bound over all choices of $\mathcal{S}^1, \dots, \mathcal{S}^r$.

Let us consider an element $u \in \mathcal{U}$. For each subcollection \mathcal{S}^i , since u is included in each of $S_{j_{i,1}}, \dots, S_{j_{i,h_i}}$ independently with probability α , $\Pr[u \in \bigcap_{S \in \mathcal{S}^i} S] = \alpha^{h_i} \geq \alpha^\ell$. Since the subcollections are disjoint, the event $u \notin \bigcap_{S \in \mathcal{S}^i} S$ is independent for different \mathcal{S}^i . Hence, we have

$$\Pr \left[u \notin \bigcup_{i=1}^r \left(\bigcap_{S \in \mathcal{S}^i} S \right) \right] = \prod_{i=1}^r \Pr \left[u \notin \bigcap_{S \in \mathcal{S}^i} S \right] \leq (1 - \alpha^\ell)^r \leq e^{-\alpha^\ell r} \leq \eta/2 \quad (11)$$

where the last inequality comes from $r \geq \ln(2/\eta)/\alpha^\ell$.

For different $u \in \mathcal{U}$, the event $u \notin \bigcup_{i=1}^r (\bigcap_{S \in \mathcal{S}^i} S)$ is independent. Applying the Chernoff bound, we have

$$\Pr \left[\left| \bigcup_{i=1}^r \left(\bigcap_{S \in \mathcal{S}^i} S \right) \right| < (1 - \eta)m \right] \leq e^{-\eta m/6}.$$

Finally, note that the number of different subcollections $\mathcal{S}_1, \dots, \mathcal{S}_r$ (i.e. $\{j_{1,1}, \dots, j_{1,h_1}\}, \dots, \{j_{r,1}, \dots, j_{r,h_r}\}$) is at most $(2k)^{\ell r}$; this is because there are at most $\binom{k}{0} + \binom{k}{1} + \dots + \binom{k}{\ell} \leq (\ell + 1)k^\ell \leq (2k)^\ell$ choices of \mathcal{S}_i for each $i \in [r]$. As a result, by union bound, the probability that $|\bigcup_{i=1}^r (\bigcap_{S \in \mathcal{S}^i} S)| < (1 - \eta)m$ is at most $(2k)^{\ell r} \cdot e^{-\eta m/6} \geq 2^{\ell r \log(2k) - \eta m/6}$ as desired. \square

Lemma 38 For any $0 < \alpha < 1$ and any $m \in \mathbb{N}$, let S_1, \dots, S_k be subsets of an m -element universe \mathcal{U} drawn independently at random from $\mathcal{D}_{\mathcal{U}, \alpha}$. Then, $|S_1|, \dots, |S_k| \leq 2\alpha m$ with probability at least $1 - 2^{\log k - \alpha m/3}$.

Proof. For each $i \in [k]$, since each $u \in \mathcal{U}$ is included in S_i independently w.p. α , Chernoff bound implies that $\Pr[|S_i| > 2\alpha m] \leq 2^{-\alpha m/3}$. By union bound over all $i \in [k]$, we get the desired bound. \square

A.2 A Deterministic Construction

Proof of Lemma 20. The existence follows immediately from Lemma 35. Now, one way to construct \mathcal{S} is to just enumerate over all choices of S_1, \dots, S_k . However, this is rather slow as there can be as many as 2^m choices of S_i , i.e., the running time can be as high as $\text{poly}(m)2^{km}$. (Saving can be made by consider only S_i 's of size at most $2\alpha m$, but the running time here is still (at least) $\text{poly}(m)2^{\alpha km}$.) We would like to get rid of the dependency of m from the exponent. We do so quite easily by dividing \mathcal{U} into parts of size roughly m_0 , finding well-behaved subsets for each part and put them together. More precisely, let us consider the following algorithm.

- (a) Partition \mathcal{U} into $\mathcal{U}_1 \cup \dots \cup \mathcal{U}_{\lceil m/m_0 \rceil}$ where each \mathcal{U}_i has size between m_0 and $2m_0$ (inclusive).
 - (b) From Lemma 35, for each $i \in [\lceil m/m_0 \rceil]$ there exists a collection \mathcal{S}_i of k subsets of \mathcal{U}_i such that
 - Every subset has size at most $2\alpha |\mathcal{U}_i|$.
 - \mathcal{S}_i is a $(\lceil \ln(2/\eta) / \alpha^\ell \rceil, \ell, \eta)$ -disperser (with respect to \mathcal{U}_i).
 - Any subcollection $\tilde{\mathcal{S}} \subseteq \mathcal{S}_i$ of size $\lceil 8 \ln(2/\mu) / \alpha \rceil$ is $(\alpha/2, \mu)$ -uniform (with respect to \mathcal{U}_i).
- Since \mathcal{U}_i is of size $O(m_0)$, we can enumerate all possible collections of k subsets of \mathcal{U}_i and check whether it satisfies the three properties above to find such a collection \mathcal{S}_i . There are $2^{O(km_0)}$ collections to enumerate over. For each collection it takes at most $\text{poly}(m_0)2^{O(k^2)}$ time to determine whether it is a desired intersection disperser because there are no more than $(k+1)^{2k} = 2^{O(k^2)}$ different disjoint subcollections of k sets. Moreover, we can check the uniformity in $\text{poly}(m_0)2^{O(k)}$ time. In total, this process takes time at most $\text{poly}(m_0)2^{O(m_0 k^2)}$.
- (c) Suppose that $\mathcal{S}_i = \{S_{i,1}, \dots, S_{i,k}\}$ for every $i \in [\lceil m/m_0 \rceil]$. We finally take $\mathcal{S} = \{\bigcup_{i=1}^{\lceil m/m_0 \rceil} S_{i,1}, \dots, \bigcup_{i=1}^{\lceil m/m_0 \rceil} S_{i,k}\}$

It is not hard to see that \mathcal{S} satisfies the desired properties and that the algorithm runs in $\text{poly}(m)2^{O(m_0 k^2)}$ time as desired. \square