A Birthday Repetition Theorem and Complexity of Approximating Dense CSPs[*][†]

Pasin Manurangsi[1] and Prasad Raghavendra[2]

1 **University of California, Berkeley, CA, USA**
`pasin@berkeley.edu`

2 **University of California, Berkeley, CA, USA**
`prasad@berkeley.edu`

Abstract

A $(k \times l)$-*birthday repetition* $\mathcal{G}^{k\times l}$ of a two-prover game $\mathcal{G}$ is a game in which the two provers are sent random sets of questions from $\mathcal{G}$ of sizes k and l respectively. These two sets are sampled independently uniformly among all sets of questions of those particular sizes. We prove the following *birthday repetition theorem*: when $\mathcal{G}$ satisfies some mild conditions, $val(\mathcal{G}^{k\times l})$ decreases exponentially in $\Omega(kl/n)$ where n is the total number of questions. Our result positively resolves an open question posted by Aaronson, Impagliazzo and Moshkovitz [Aaronson et al., CCC, 2014].

As an application of our birthday repetition theorem, we obtain new fine-grained inapproximability results for dense CSPs. Specifically, we establish a tight trade-off between running time and approximation ratio by showing conditional lower bounds, integrality gaps and approximation algorithms; in particular, for any sufficiently large i and for every $k \geq 2$, we show the following:

- We exhibit an $O(q^{1/i})$-approximation algorithm for dense Max k-CSPs with alphabet size q via $O_k(i)$-level of Sherali-Adams relaxation.
- Through our birthday repetition theorem, we obtain an integrality gap of $q^{1/i}$ for $\tilde{\Omega}_k(i)$-level Lasserre relaxation for fully-dense Max k-CSP.
- Assuming that there is a constant $\varepsilon > 0$ such that Max 3SAT cannot be approximated to within $(1-\varepsilon)$ of the optimal in sub-exponential time, our birthday repetition theorem implies that any algorithm that approximates fully-dense Max k-CSP to within a $q^{1/i}$ factor takes $(nq)^{\tilde{\Omega}_k(i)}$ time, almost tightly matching our algorithmic result.

As a corollary of our algorithm for dense Max k-CSP, we give a new approximation algorithm for Densest k-Subhypergraph, a generalization of Densest k-Subgraph to hypergraphs. When the input hypergraph is $O(1)$-uniform and the optimal k-subhypergraph has constant density, our algorithm finds a k-subhypergraph of density $\Omega(n^{-1/i})$ in time $n^{O(i)}$ for any integer $i > 0$.

1998 ACM Subject Classification G.1.6 Linear Programming, G.2.2 Graph Algorithms

Keywords and phrases Birthday Repetition, Constraint Satisfaction Problems, Linear Program

Digital Object Identifier 10.4230/LIPIcs.ICALP.2017.78

1 Introduction

Polynomial-time reductions between computational problems are among the central tools in complexity theory. The rich and vast theory of hardness of approximation emerged out

[*] The full version of this extended abstract can be found at [42], `https://arxiv.org/abs/1607.02986`.

[†] This material is based upon work supported by NSF CCF-1343104 and the Okawa Research Grant.

44th International Colloquium on Automata, Languages, and Programming (ICALP 2017).
Editors: Ioannis Chatzigiannakis, Piotr Indyk, Fabian Kuhn, and Anca Muscholl;
Article No. 78; pp. 78:1–78:15
Leibniz International Proceedings in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

of the celebrated PCP Theorem [6] and the intricate web of polynomial-time reductions developed over the past two decades. During this period, an extensive set of reduction techniques such as parallel repetition and long-codes have been proposed and a variety of mathematical tools including discrete harmonic analysis, information theory and Gaussian isoperimetry have been applied towards analyzing these reductions. These developments have led to an almost complete understanding of the approximability of many fundamental combinatorial optimization problems like Set Cover and Max 3SAT. Yet, there are a few central problems such as computing approximate Nash equlibria, Densest k-Subgraph and Small Set Expansion, that remain out of reach of the web of polynomial-time reductions.

A promising new line of work proposes to understand the complexity of these problems through the lens of *sub-exponential time reductions*. Specifically, the idea is to construct a sub-exponential time reduction from 3SAT to the problem at hand, say, the Approximate Nash Equilibrium problem. Assuming that 3SAT does not admit sub-exponential time algorithms (also known as the Exponential Time Hypothesis (ETH) [35]), this would rule out polynomial time algorithms for the Approximate Nash Equilibrium problem.

At the heart of this line of works, lies the so-called *birthday repetition* of two-prover games. To elaborate on this, we begin by formally defining the notion of two-prover games.

▶ **Definition 1.** A *two-prover game* $\mathcal{G}$ consists of

- A finite set of questions X, Y and corresponding answer sets (aka alphabets) Σ_X, Σ_Y.
- A distribution $\mathcal{Q}$ over pairs of questions $X \times Y$.
- A verification function $P : X \times Y \times \Sigma_X \times \Sigma_Y \to \{0, 1\}$.

The value of $\mathcal{G}$ is the maximum over all strategies $\phi : X \cup Y \to \Sigma_X \cup \Sigma_Y$ of the output of P, i.e., $val(\mathcal{G}) = \max_{\phi: X\cup Y \to \Sigma_X \cup \Sigma_Y} \mathbb{E}_{(x,y)\sim\mathcal{Q}}[P(x, y, \phi(x), \phi(y))]$. We use n and q to denote the number of variables $|X| + |Y|$ and the alphabet size $|\Sigma_X| + |\Sigma_Y|$ respectively.

Two prover games earn their name from the following interpretation of the above definition: The game $\mathcal{G}$ is played between a verifier V and two cooperating provers $Merlin_1$ and $Merlin_2$ who have agreed upon a common strategy, but cannot communicate with each other during the game. The verifier samples two questions $(x, y) \sim \mathcal{Q}$ and sends x to $Merlin_1$ and y to $Merlin_2$. The provers respond with answers $\phi(x)$ and $\phi(y)$, which the verifier accepts or rejects based on the value of the verification function $P(x, y, \phi(x), \phi(y))$.

Two-prover games and, more specifically, a special class of two-prover games known as Label Cover are the starting points for reductions in a large body of hardness of approximation results. The PCP theorem implies that for some constant ε_0, approximating the value of a two prover game to within an additive ε_0 is **NP**-hard. However, this hardness result on its own is inadequate to construct reductions to other combinatorial optimization problems. To this end, this hardness result can be strengthened to imply that it is **NP**-hard to approximate the value of two-prover games to any constant factor, using the *parallel repetition theorem.*

For an integer k, the k-parallel repetition $\mathcal{G}^{\otimes k}$ of $\mathcal{G}$ can be described as follows. The question and answer sets in $\mathcal{G}^{\otimes k}$ consist of k-tuples of questions and answers from $\mathcal{G}$. The distribution over questions in $\mathcal{G}^{\otimes k}$ is given by the product distribution $\mathcal{Q}^k$. The verifier for $\mathcal{G}^{\otimes k}$ accepts the answers if and only if the verifier for $\mathcal{G}$ accepts each of the k individual answers.

Roughly speaking, the parallel repetition theorem asserts that the value of $\mathcal{G}^k$ decays exponentially in k. The theorem forms a key ingredient in obtaining hardness of approximation results, and have aptly received considerable attention in literature [51, 34, 50, 25, 46, 14].

Birthday repetition, introduced by Aaronson et al. [2], is an alternate transformation on two-prover games defined as follows.

▶ **Definition 2.** The $(k \times l)$-*birthday repetition* of a two-prover game $\mathcal{G}$ consists of

- The set of questions in $\mathcal{G}^{k\times l}$ are $\binom{X}{k}$ and $\binom{Y}{l}$ respectively, i.e., each question is a subset $S \subseteq X$ of size k and subset $T \subseteq Y$ of size l.
- The distribution over questions is the uniform product distribution over $\binom{X}{k} \times \binom{Y}{l}$.
- The verifier accepts only if, for every $(x, y) \in S \times T$ such that (x, y) form a valid pair of questions in $\mathcal{G}$, i.e., $(x, y) \in \text{supp}(\mathcal{Q})$, the answers to x and y are accepted $\mathcal{G}$.

The basic idea of birthday repetition can be traced back to the work of Aaronson et al. [1] on quantum multiprover proof systems $\mathbf{QMA}(k)$ for 3SAT. Subsequent work by Aaronson et al. [2] on the classical analogue of $\mathbf{QMA}(k)$ formally defined birthday repetition for two-prover games, and set the stage for applications in hardness of approximation.

Unlike parallel repetition, birthday repetition is only effective for large values of k and l. In particular, if $k, l < o(\sqrt{n})$, then, for most pairs of S and T, there is no $(x, y) \in S \times T$ such that (x, y) belongs to the support of the questions in the original game. However, if we pick $k = l = \omega(\sqrt{n})$, then by the birthday paradox, with high probability the sets S, T contain an edge (x, y) from the original game $\mathcal{G}$. Hence, for this choice of k and l, the game played by the provers is seemingly at least as difficult to succeed, as the original game $\mathcal{G}$. Aaronson et al. [2] confirmed this intuition by proving the following theorem.

▶ **Theorem 3** ([2]). *For any two-prover game $\mathcal{G}$ such that $\mathcal{Q}$ is uniform over its support, if the bipartite graph induced by $(X, Y, \text{supp}(\mathcal{Q}))$ is biregular, then $val(\mathcal{G}^{k\times l}) \leq val(\mathcal{G}) + O(\sqrt{\frac{n}{kl}})$.*

On the one hand, birthday repetition is ineffective in that it has to incur a blowup of $2^{\sqrt{n}}$ in the size, to even simulate the original game $\mathcal{G}$. The distinct advantage of birthday repetition is that the resulting game $\mathcal{G}^{k,l}$ has a distinct structure – in that it is a *free game*.

▶ **Definition 4.** A *free game* is a two-player game such that $\mathcal{Q}$ is uniform over $X \times Y$.

The birthday repetition theorem of [2] immediately implies a hardness of approximation for the value of free games. Specifically, they show that it is ETH-hard to approximate free games to some constant ratio in almost quasi-polynomial time. Interestingly, this lower bound is nearly tight in that free games admit a quasipolynomial time approximation scheme [10, 2].

Following Aaronson et al.'s work, birthday repetition has received numerous applications, which can be broadly classified in to two main themes. On the one hand, there are problems such as computing approximate Nash equilibria [16, 8, 54], approximating free games [2], approximating learning dimensions [43], and approximate symmetric signaling in zero sum games [53], where the underlying problems admit quasipolynomial-time algorithms [26, 38, 28] and birthday repetition can be used to show that such a running time is necessary, assuming ETH. On the other hand, there are computational problems like Densest k-Subgraph [15, 39], injective tensor norms [1, 33, 9], 2-to-4-norms [1, 33, 9] wherein an **NP**-hardness of approximation result seems out of reach of current techniques. But the framework of birthday repetition can be employed to show a quasi-polynomial hardness assuming ETH[1].

Unlike the parallel repetition theorem, the birthday repetition theorem of [2] does not achieve any reduction in the value of the game. It is thus natural to ask whether birthday repetition can also be used to decrease the value of a game. Aaronson et al. conjectured that the value of the birthday repetition game indeed deteriorates exponentially in $\Omega(kl/n)$, which is the expected number of edges between S and T in birthday repetition. Our main contribution is that we resolve the conjecture positively by showing the following theorem.

[1] Although the hardness results for injective tensor norms and 2-to-4-norms build over quantum multiprover proof systems, the basic idea of birthday repetition [1] lies at the heart of these reductions.

▶ **Theorem 5** (Birthday Repetition Theorem (informal)). *Let $\mathcal{G} = (X, Y, \mathcal{Q}, \Sigma_X, \Sigma_Y, P)$ be a two-prover game such that $\mathcal{Q}$ is uniform over its support, $(X, Y, \text{supp}(\mathcal{Q}))$ is biregular and $|\Sigma_X|, |\Sigma_Y|$ are constant. If $val(\mathcal{G}) = 1 - \varepsilon$, then $val(\mathcal{G}^{k \times l}) \leq 2(1 - \varepsilon/2)^{\Omega(\varepsilon^5 kl/n)}$.*

Note that our result is more general than stated above and can handle irregular graphs and non-constant alphabet sizes as well (see Theorem 12 and Theorem 13).

By definition, our theorem immediately implies the following inapproximability of free games.

▶ **Corollary 6.** *Unless ETH is false, no polynomial time algorithm can approximate the value of a free game to within a factor of $2^{\tilde{\Omega}(\log(nq))}$.*

1.1 Dense CSPs

Free games can be viewed as 2-ary constraint satisfaction problems (CSP). From this perspective, free games are *dense*, in that there are constraints on a constant fraction of all pairs of variables. As an application of our birthday repetition theorem, we show almost-tight lower bounds for dense CSPs. To this end, we begin by defining CSPs and its density.

▶ **Definition 7.** A Max k-CSP instance $\mathcal{G}$ consists of

- A finite set of variables V and a finite alphabet set Σ.
- A distribution $\mathcal{Q}$ over k-tuple of variables V^k.
- A predicate $P : V^k \times \Sigma^k \to [0, 1]$.

Similar to two-prover games, $val(\mathcal{G})$ is defined as $\max_{\phi: V \to \Sigma} \mathbb{E}_{S \sim \mathcal{Q}}[P(S, \phi|_S)]$ where $\phi|_S$ is the restriction of the assignment to S, and we use n to denote the number of variables $|V|$ and q to denote the alphabet size $|\Sigma|$ of $\mathcal{G}$. Finally, $\mathcal{G}$ is called Δ-*dense* if $\Delta \cdot \mathcal{Q}(S) \leq 1/|V|^k$ for every $S \in V^k$. The 1-dense instances are also said to be fully-dense.

There has been a long line of works on approximating dense CSPs. Arora et al. were first to devise a polynomial-time approximation scheme for the problem when alphabet size is constant [5]. Since then, numerous algorithms have been invented for approximating dense CSPs using variety of techniques such as combinatorial algorithms with exhaustive sampling [5, 21, 44, 58, 40, 29], subsampling of instances [3, 10], regularity lemmas [30, 20] and LP/SDP hierarchies [22, 11, 31, 60]. Among the known algorithms, the fastest is Yaroslavtsev's [58] that achieves $(1 + \varepsilon)$-approximation in $q^{O_k(\log q)} + (nq)^{O(1)}$ time[2].

Unfortunately, when q is (almost-)polynomial in n, none of the above algorithms run in polynomial time. CSPs in such regime of parameters have long been studied in hardness of approximation (e.g. [12, 52, 7, 24, 47, 45]) and have recently received more attention from the approximation algorithm standpoint, both in the general case [48, 17, 41, 19] and the dense case [40]. The approximabilities of these two cases are vastly different. In the general case, approximating Max 2-CSP to within a factor of $2^{\log^{1-\varepsilon}(nq)}$ is **NP**-hard for any constant $\varepsilon > 0$ [24]. Moreover, the long-standing Sliding Scale Conjecture [12] states that this ratio can be improved to $(nq)^{\varepsilon}$ for some constant $\varepsilon > 0$. On the other hand, aforementioned algorithms for dense CSPs rule out such hardnesses for the dense case.

While the gap between known approximation algorithms and inapproximability results in the general case is tiny ($2^{\log^{\varepsilon}(nq)}$ for any constant $\varepsilon > 0$), the story is different for the dense case, especially when we restrict ourselves to polynomial-time algorithms. Aaronson et al. only ruled out, assuming ETH, polylog(nq)-approximation for such algorithms [2].

[2] [58] states that the algorithm takes $q^{O_k(1)} + (nq)^{O(1)}$ time, which is incorrect [59].

However, for $k > 2$, no non-trivial polynomial-time algorithm for dense Max k-CSP on large alphabet is even known. In this paper, we settle down the complexity of approximating dense Max k-CSP almost completely by answering the following fine-grained question: for each $i \in \mathbb{N}$, what is the best approximation for dense Max k-CSP, achievable in time $(nq)^i$?

Manurangsi and Moshkovitz developed an algorithm for dense Max 2-CSP that, when the instance has value $\Omega(1)$, obtains $O(q^{1/i})$-approximation in $(nq)^{O(i)}$ time [40]. Unfortunately, the algorithm does not work for dense Max k-CSP when $k > 2$. Using a conditioning-based rounding technique developed in [11, 49, 60], we show that the Sherali-Adams (SA) relaxation [56] exhibits a similar approximation even when $k > 2$, as stated below.

▶ **Theorem 8** (Informal). *For every $i > 0$ and any dense* Max k-CSP *instance of value $1-\varepsilon$, an $O_{k,\varepsilon}(i/\Delta)$-level of the SA relaxation yields an $O(q^{1/i})$-approximation for the instance.*

Using our birthday repetition theorem, we prove that the above tradeoff between run-time and approximation ratio cannot be improved even with the stronger Lasserre hierarchy [37]. Specifically, by applying the birthday repetition theorem with $k, l = \Omega(n \log i/i)$ on an $\Omega(n)$-level Lasserre integrality gap for Max 3XOR [55], we show the following.

▶ **Lemma 9** (Informal). *For every sufficiently large $i > 0$, there is a fully-dense* Max k-CSP *instance of value $1/(nq)^{1/i}$ such that the value of $\tilde{\Omega}_k(i)$-level Lasserre relaxation is one.*

Instead, if we assume that there exists a constant $\varepsilon > 0$ so that Max 3SAT cannot be approximated to $1 - \varepsilon$ in sub-exponential time (which we call the Exponential Time Hypothesis for Approximating 3SAT (ETHA)[3]), we can similarly arrive at the following hardness result.

▶ **Lemma 10** (Informal). *Unless ETHA is false, for every sufficiently large $i > 0$, no $(nq)^{\tilde{O}_k(i)}$-time algorithm approximates fully-dense* Max k-CSP *to within a factor of $(nq)^{1/i}$.*

Thus, assuming ETHA, our results resolve complexity of approximating dense CSPs up to a factor of polylog i and a dependency on k in the exponent of the running time.

1.2 Densest k-Subhypergraph

As a by-product of our algorithm for dense Max k-CSP, we give an approximation algorithm for the following Densest k-Subhypergraph problem: given a hypergraph (V, E), find $S \subseteq V$ of k vertices that maximizes the number of edges contained in S.

When the input hypergraph is simply a graph, the problem becomes Densest k-Subgraph, which has been extensively studied dating back to the early '90s [36, 27, 28, 57, 13]. On the other hand, Densest k-Subhypergraph was first studied in 2006, when Hajiaghayi et al. [32] proved that, if 3SAT $\notin \mathbf{DTIME}(2^{n^{3/4+\varepsilon}})$ for some $\varepsilon > 0$, then no polynomial-time algorithm approximates the problem to within a factor of $2^{\log^\delta n}$ for some $\delta > 0$. Later, Applebaum [4] showed, under a cryptographic assumption, that, for sufficiently large d, Densest k-Subhypergraph on d-uniform hypergraph is hard to approximate to a factor of n^ε for some $\varepsilon > 0$. More recently, Chlamtác et al. [18] provided the first non-trivial approximation algorithm for the problem; their algorithm works only on 3-uniform hypergraph and achieves $O(n^{4(4-\sqrt{3})/13+\varepsilon})$-approximation for any constant $\varepsilon > 0$ in polynomial time.

[3] ETHA is also introduced independently as gap-ETH by Dinur [23] who uses it to provide a supporting evidence to the Sliding Scale Conjecture.

Thanks to Charikar et al.'s [17] reduction from Densest k-Subgraph to Max 2-CSP, which can be adapted to reduce Densest k-Subhypergraph on d-uniform hypergraph to Max d-CSP, Theorem 8 implies the following algorithm for Densest k-Subhypergraph.

▶ **Corollary 11** (Informal). *There is a randomized algorithm that, given a d-uniform hypergraph on n vertices whose densest k-subhypergraph is Δ-dense and an integer $i > 0$, runs in $n^{O_d(i/\Delta)}$ time and outputs a k-subhypergraph of density $\Omega_k(\Delta/n^{1/i})$ with high probability.*

Here the density of a d-uniform hypergraph is defined as $d!|E|/|V|^d$. Note that the density condition required is on the optimum not the input. Moreover, when Δ and d are constant, the algorithm provides an $O(n^{1/i})$ approximation in $n^{O(i)}$ time for every $i > 0$. When $d = 2$, this matches the previously known algorithms for Densest k-Subgraph [28, 57, 40].

Organization of the Paper

In Section 2, we provide preliminaries and notations used in the paper. Then, in Section 3, we outline the proofs of our main theorems; the full proofs are deferred to the full version of this work [42]. Next, the algorithm for dense CSPs is described in Section 4. Finally, we conclude by proposing open questions in Section 5. Note that the lower bounds for dense CSPs and the algorithm for Densest k-Subhypergraph are also deferred to the full version.

2 Preliminaries and Notations

For $n \in \mathbb{N}$, we use $[n]$ to denote $\{1, \dots, n\}$. For two sets X and S, define X^S to be the set of tuples $(x_s)_{s \in S}$. We sometimes view $(x_s)_{s \in S}$ as a function from S to X. For a set S and $n \in \mathbb{N}$, $\binom{S}{n}$ denotes the collection of subsets of S of size n. Moreover, let $\binom{S}{0} = \{\emptyset\}$ and $\binom{S}{[n]} = \binom{S}{0} \cup \cdots \cup \binom{S}{n}$. For any bipartite graph (A, B, E) and $S \subseteq A, T \subseteq B$, let $E(S, T)$ denote the set of all edges with one endpoint in S and the other in T.

Let $\mathcal{X}$ be a probability distribution over a finite probability space Θ. We use $x \sim \mathcal{X}$ to denote a random variable x sampled from $\mathcal{X}$. Sometimes we abuse the notation and write Θ in place of the uniform distribution over Θ. For each $\theta \in \Theta$, we denote $\Pr_{x \sim \mathcal{X}}[x = \theta]$ by $\mathcal{X}(\theta)$. The *support* of $\mathcal{X}$ or $\text{supp}(\mathcal{X})$ is the set of all $\theta \in \Theta$ such that $\mathcal{X}(\theta) \neq 0$. For any event E, we use $\mathbb{1}[E]$ to denote the indicator variable for the event.

The *informational divergence* between distributions $\mathcal{X}$ and $\mathcal{Y}$ is defined as $D_{KL}(\mathcal{X} \| \mathcal{Y}) = \sum_{\theta \in \text{supp}(\mathcal{X})} \mathcal{X}(\theta) \log(\mathcal{X}(\theta)/\mathcal{Y}(\theta))$. The *total correlation* between random variables $x_1, \dots, x_n$ is $C(x_1; \dots; x_n) = D_{KL}(\mathcal{X}_{1,\dots,n} \| \mathcal{X}_1 \times \cdots \times \mathcal{X}_n)$ where $\mathcal{X}_{1,\dots,n}$ is the joint distribution of $x_1, \dots, x_n$ and $\mathcal{X}_i$ is the marginal distribution of x_i. Finally, the *conditional total correlation* is defined as $C(x_1; \dots; x_{n-1} | x_n) = \mathbb{E}_{\theta \sim \text{supp}(\mathcal{X}_n)}[C(x_1; \dots; x_{n-1}) | x_n = \theta]$.

For Max k-CSP, we use N to denote the instance size $(nq)^k$. For convenience, we write the predicates as $P_S(\phi|_S)$ instead of $P(S, \phi|_S)$. Moreover, for an assignment ϕ of $\mathcal{G} = (\mathcal{V}, \mathcal{W}, \{P_S\})$, its value is $val_{\mathcal{G}}(\phi) = \mathbb{E}_{S \sim \mathcal{W}}[P_S(\phi|_S)]$. When $\mathcal{G}$ is clear from the context, we simply write $val(\phi)$. Note that $val(\mathcal{G}) = \max_\phi val_{\mathcal{G}}(\phi)$. For any $S, T \subseteq V$, $\phi_S \in \Sigma^S$ and $\phi_T \in \Sigma^T$ are said to be *consistent* if they agree on $S \cap T$ and *inconsistent* otherwise. For consistent ϕ_S, ϕ_T, we define $\phi_S \circ \phi_T \in \Sigma^{S \cup T}$ by $\phi_S \circ \phi_T(x) = \phi_S(x)$ if $x \in S$ and $\phi_S \circ \phi_T(x) = \phi_T(x)$ otherwise. Similar notations are also used for two-prover games. Finally, recall that a game $(X, Y, \mathcal{Q}, \Sigma_X, \Sigma_Y, \{P_{(x,y)}\})$ is a *projection game* if, for each $(x, y) \in \text{supp}(\mathcal{Q})$, there is $f : \Sigma_X \to \Sigma_Y$ such that, for all $\sigma_x \in \Sigma_X, \sigma_y \in \Sigma_Y$, $P_{(x,y)}(\sigma_x, \sigma_y) = \mathbb{1}[f(\sigma_x) = \sigma_y]$.

3 Birthday Repetition Theorem: Proof Overview

In this section, we outline the proofs of our birthday repetition theorems. We first state our main theorems formally, starting with the birthday repetition theorem for general games.

▶ **Theorem 12.** *There exists a constant $\alpha > 0$ such that the following is true. Let $\mathcal{G} = (X, Y, E, \Sigma_X, \Sigma_Y, \{P_{(x,y)}\})$ be any two-prover game of value $1 - \varepsilon$. Let d_{max} be the maximum degree of a vertex in (X, Y, E) and $c = \log |\Sigma_X||\Sigma_Y|$. For all $0 \leq k \leq |X|$ and $0 \leq l \leq |Y|$,*

$$val(\mathcal{G}^{k \times l}) \leq 2(1 - \varepsilon/2)^{\frac{\alpha \varepsilon^5 kl|E|}{d_{max}|X||Y|c^2}}$$

For projection games, we can improve the dependency on ε and avoid the dependency on c:

▶ **Theorem 13.** *There exists a constant $\alpha > 0$ such that the following is true. Let $\mathcal{G} = (X, Y, E, \Sigma_X, \Sigma_Y, \{P_{(x,y)}\})$ be any projection game of value $1 - \varepsilon$. Let d_{max} be the maximum degree of a vertex in (X, Y, E). For all $0 \leq k \leq |X|$ and $0 \leq l \leq |Y|$, we have*

$$val(\mathcal{G}^{k \times l}) \leq 2(1 - \varepsilon/2)^{\frac{\alpha \varepsilon^3 kl|E|}{d_{max}|X||Y|}}$$

In short, we will to show that $\mathcal{G}^{k \times l}$ has small value by "embedding" an $\Omega\left(\frac{kl|E|}{d_{max}|X||Y|}\right)$-parallel repetition game, which has low value by the parallel repetition theorem, into it.

For convenience, let s denote $\frac{kl|E|}{|X||Y|}$, the expected number of edges in $E(S, T)$ when S and T are independently uniformly sampled from $\binom{X}{k}$ and $\binom{Y}{l}$ respectively. Let s_1 and s_2 be $s(1 + \delta)$ and $s(1 - \delta)$ respectively for some $\delta \in [0, 1/2]$ that will be chosen later. We will use $r = \beta s/d_{max}$ rounds of parallel repetition where $\beta \in [0, \delta/40]$ will be specified later. Lastly, let $E^r = \{((x_1, \dots, x_r), (y_1, \dots, y_r)) \mid (x_1, y_1), \dots, (x_r, y_r) \in E\}$.

▶ Remark. δ and β will be chosen based on ε, c and whether $\mathcal{G}$ is a projection game. When ε and c are constant, both δ and β are small constants. This is the most representative case and is good to keep in mind when reading through the proof.

Our overall strategy is to reduce $\mathcal{G}^{\otimes r}$ to $\mathcal{G}^{k \times l}$. Since $val(\mathcal{G}^{\otimes r})$ is exponentially small in $r = \Omega\left(\frac{kl|E|}{d_{max}|X||Y|}\right)$ due to the parallel repetition theorem, such reduction would give a similar upper bound on $val(\mathcal{G}^{k \times l})$. Unfortunately, we do not know how to do this in one step so we will have to go through a sequence of reductions. The sequence of games that we reduce to are $\mathcal{G}^{\otimes r}_{\text{set}}, \mathcal{G}^{k \times l}_{\text{em}}, \mathcal{G}^{k \times l}_{\text{em},[s_1,s_2]}$ and $\mathcal{G}^{k \times l}_{[s_1,s_2]}$ respectively. The game $\mathcal{G}^{\otimes r}_{\text{set}}$ share the same questions, alphabet sets and predicates with $\mathcal{G}^{\otimes r}$ while $\mathcal{G}^{k \times l}_{\text{em}}, \mathcal{G}^{k \times l}_{\text{em},[s_1,s_2]}$ and $\mathcal{G}^{k \times l}_{[s_1,s_2]}$ share those with $\mathcal{G}^{k \times l}$. The distribution of each game is defined as follows.

- The distribution of $\mathcal{G}^{\otimes r}_{\text{set}}$ is uniform over the set E^r_{set} of all $((x_1, \dots, x_r), (y_1, \dots, y_r)) \in E^r$ such that $x_1, \dots, x_r, y_1, \dots, y_r$ are all distinct. Note that this distribution is simply $\mathcal{G}^{\otimes r}$'s distribution conditioned on $x_1, \dots, x_r, y_1, \dots, y_r$ being all distinct.
- We will try to make the distribution $\mathcal{Q}^{k \times l}_{\text{em}}$ of $\mathcal{G}^{k \times l}_{\text{em}}$ reflect an embedding of the game $\mathcal{G}^{\otimes r}_{\text{set}}$. We define $\mathcal{Q}^{k \times l}_{\text{em}}$ based on the following sampling process for $(S, T) \sim \mathcal{Q}^{k \times l}_{\text{em}}$. First, sample $((x_1, \dots, x_r), (y_1, \dots, y_r))$ uniformly at random from E^r_{set}. Then, sample $\tilde{S}$ and $\tilde{T}$ independently uniformly from $\binom{X - \{x_1, \dots, x_r\}}{k - r}$ and $\binom{Y - \{y_1, \dots, y_r\}}{l - r}$ respectively. Finally, set $S = \{x_1, \dots, x_r\} \cup \tilde{S}$ and $T = \{y_1, \dots, y_r\} \cup \tilde{T}$.
- The distribution $\mathcal{Q}^{k \times l}_{\text{em},[s_1,s_2]}$ of $\mathcal{G}^{k \times l}_{\text{em},[s_1,s_2]}$ is the distribution $\mathcal{Q}^{k \times l}_{\text{em}}$ conditioned on the number of edges between the two sets being in the range $[s_1, s_2]$. In other words, $\mathcal{Q}^{k \times l}_{\text{em},[s_1,s_2]}(S, T) = \Pr_{(S',T') \sim \mathcal{Q}^{k \times l}_{\text{em}}}[S = S' \wedge T = T' \mid s_1 \leq |E(S', T')| \leq s_2]$.

- Finally, the distribution of $\mathcal{G}^{k\times l}_{[s_1,s_2]}$ is uniform over the set $E^{k\times l}_{[s_1,s_2]}$ of all (S,T) such that $|E(S,T)| \in [s_1,s_2]$. In other words, we ignore weights in $\mathcal{Q}^{k\times l}_{\text{em},[s_1,s_2]}$ and use the uniform distribution over $\text{supp}(\mathcal{Q}^{k\times l}_{\text{em},[s_1,s_2]})$.

Before we present the overview of the proofs, let us list simple bounds that will be useful in understanding the intuitions. Their proofs can be found in the full version of this work [42].

▶ **Lemma 14.** *Let (X,Y,E) be any bipartite graph with maximum degree d_{max}. For any non-negative integers $k \le |X|$ and $l \le |Y|$, let $s = \frac{kl|E|}{|X||Y|}$. For any $0 \le \gamma < 1/2$, we have*

$$\Pr_{S\sim\binom{X}{k},T\sim\binom{Y}{l}}[|E(S,T)| \notin [(1-\gamma)s,(1+\gamma)s]] \le 4\exp\left(-\frac{\gamma^2 s}{54 d_{max}}\right).$$

▶ **Lemma 15.** *Let $\mathcal{G}$ and $\mathcal{G}'$ be two games on the same questions, alphabets, and predicates but on different distributions $\mathcal{Q}$ and $\mathcal{Q}'$ respectively. If, for some α, $\mathcal{Q}(x,y) \le \alpha \cdot \mathcal{Q}'(x,y)$ for all $x \in X, y \in Y$, then $val(\mathcal{G}) \le \alpha \cdot val(\mathcal{G}')$. In particular, when $\mathcal{Q}$ and $\mathcal{Q}'$ are uniform distributions on some $E \subseteq E'$, $val(\mathcal{G}) \le \frac{|E'|}{|E|} \cdot val(\mathcal{G}')$.*

▶ **Lemma 16.** *Let $\mathcal{G} = (X,Y,\mathcal{Q},\Sigma_X,\Sigma_Y,\{P_{x,y}\}_{(x,y)\in\text{supp}(\mathcal{Q})})$ be any two player game and let A be any event occurring with probability $1-p>0$ (w.r.t. $\mathcal{Q}$). Let $\mathcal{Q}'$ be the conditional probability $\mathcal{Q}$ given A, i.e., $\mathcal{Q}'(\tilde{x},\tilde{y}) = \Pr_{(x,y)\sim\mathcal{Q}}[x=\tilde{x} \wedge y=\tilde{y} \mid A]$. For the game $\mathcal{G}' = (X,Y,\mathcal{Q}',\Sigma_X,\Sigma_Y,\{P_{x,y}\}_{(x,y)\in\text{supp}(\mathcal{Q}')})$, we have $val(\mathcal{G}) - p \le val(\mathcal{G}') \le val(\mathcal{G}) + 2p$.*

We will next give intuitions on why $val(\mathcal{G}^{\otimes r}) \approx val(\mathcal{G}^{\otimes r}_{\text{set}}) \approx val(\mathcal{G}^{k\times l}_{\text{em}}) \approx val(\mathcal{G}^{k\times l}_{\text{em},[s_1,s_2]}) \approx val(\mathcal{G}^{k\times l}_{[s_1,s_2]}) \approx val(\mathcal{G}^{k\times l})$ where each ≈ hides some multiplicative or additive losses in each step. With the right choice of δ and β, we can ensure that each loss is significantly smaller than $val(\mathcal{G}^{\otimes r})$, and, thus, we will be able to bound $val(\mathcal{G}^{k\times l})$. Below, we state these losses more precisely and summarize the overview of each proof.

▶ **Lemma 17.** *$val(\mathcal{G}^{\otimes r}_{set}) \le \left(\frac{1}{1-2\beta}\right)^r \cdot val(\mathcal{G}^{\otimes r})$*

Proof Idea. From Lemma 15, it suffices to lower bound the ratio $|E^r_{\text{set}}|/|E^r|$. This is the probability that r random edges from E do not share any endpoints, which is easy to bound. ◀

▶ **Lemma 18.** *$val(\mathcal{G}^{k\times l}_{em}) \le val(\mathcal{G}^{\otimes r}_{set})$*

Proof Idea. Based on how $\mathcal{Q}^{k\times l}_{\text{em}}$ is defined, it induces a canonical map from each strategy in $\mathcal{G}^{k\times l}_{\text{em}}$ to a "mixed strategy" in $\mathcal{G}^{\otimes r}_{\text{set}}$. We can show that each strategy ϕ in $\mathcal{G}^{k\times l}_{\text{em}}$ has value no more than the value of the mixed strategy in $\mathcal{G}^{\otimes r}_{\text{set}}$ that ϕ maps to, which yields the lemma. ◀

▶ **Lemma 19.** *$val(\mathcal{G}^{k\times l}_{em,[s_1,s_2]}) \le val(\mathcal{G}^{k\times l}_{em}) + 8\exp\left(-\frac{\delta^2 r}{432\beta}\right)$*

Proof Idea. $\mathcal{Q}^{k\times l}_{\text{em},[s_1,s_2]}$ is $\mathcal{Q}^{k\times l}_{\text{em}}$ conditioned on the event $E(S,T) \in [s_1,s_2]$. From Lemma 16, it suffices to bound the probability of such event. From the definition of $\mathcal{Q}^{k\times l}_{\text{em}}$, S and T can be sampled by first sampling $x_1,\dots,x_r,y_1,\dots,y_r$ according to E^r and then sampling the rest of S and T from $X - \{x_1,\dots,x_r\}$ and $Y - \{y_1,\dots,y_r\}$ respectively. When r is small enough, we can show, with the help of Lemma 14, that, for any $x_1,\dots,x_r,y_1,\dots,y_r$, $|E(S,T)|$ concentrates around s. This gives us the desired bound. ◀

▶ **Lemma 20.** *$val(\mathcal{G}^{k\times l}_{[s_1,s_2]}) \le \left(\frac{1+\delta}{1-\delta-2\beta}\right)^{2r} \cdot val(\mathcal{G}^{k\times l}_{em,[s_1,s_2]})$*

Proof Idea. We will show that the two distributions are (multiplicatively) close and evoke Lemma 15 to arrive at the bound. Since the distribution of $\mathcal{G}^{k \times l}_{[s_1, s_2]}$ is uniform, we only need to show that the maximum and the minimum (non-zero) probabilities in $\mathcal{Q}^{k \times l}_{\text{em},[s_1,s_2]}$ are close.

Fortunately, we know that $\mathcal{Q}^{k \times l}_{\text{em},[s_1,s_2]}$ is $\mathcal{Q}^{k \times l}_{\text{em}}$ conditioned on an event. This means that, when $\mathcal{Q}^{k \times l}_{\text{em},[s_1,s_2]}(S, T)$ is not zero, it is proportional to $\mathcal{Q}^{k \times l}_{\text{em}}(S, T)$. The latter, in turn, is proportional to the number of edges $(x_1, y_1), \dots, (x_r, y_r) \in E^r$ such that $x_1, \dots, x_r, y_1, \dots, y_r$ are all distinct and $x_1, \dots, x_r \in S$ and $y_1, \dots, y_r \in T$. In other words, we want to upper bound and lower bound the number of r edges in $E(S, T)$ with distinct endpoints. This is feasible since we know that $|E(S, T)| \in [s_1, s_2]$ and r is so small that with a reasonable probability r edges picked will not share any endpoint with each other. ◀

▶ **Lemma 21.** *$val(\mathcal{G}^{k \times l}) \leq val(\mathcal{G}^{k \times l}_{[s_1,s_2]}) + 4 \exp\left(-\frac{\delta^2 r}{54\beta}\right)$*

Proof Idea. By realising that $\mathcal{G}^{k \times l}_{[s_1,s_2]}$'s distribution is simply $\mathcal{G}^{k \times l}$'s distribution conditioned on $|E(S, T)| \in [s_1, s_2]$, this follows immediately from Lemma 14 and Lemma 16. ◀

We defer proofs of the above lemmas to the full version [42]. Let us now use them to prove the birthday repetition theorems. To avoid repeating arguments for general games and projection games, we prove the following lemma. Its proof, mostly calculations, is deferred to the full version.

▶ **Lemma 22.** *Let $\mathcal{G}$ be any game of value $1 - \varepsilon$ and k, l, β, δ, r be as above. If $val(\mathcal{G}^{\otimes r}) \leq (1-\varepsilon/2)^R$ for some R such that $\frac{200\delta r}{\varepsilon} \leq R \leq \min\{r, \frac{\delta^2 r}{1000\beta\varepsilon}\}$, then $val(\mathcal{G}^{k \times l}) \leq 2(1-\varepsilon/2)^{R/10}$.*

The final ingredient for our main proof is the parallel repetition theorem. For general games, we use Holenstein's version of the theorem [34], which is stated below.

▶ **Theorem 23** ([34]). *There is a constant $C > 0$ such that, for every $k > 0$ and any two-prover game $\mathcal{G} = (X, Y, \mathcal{Q}, \Sigma_X, \Sigma_Y, \{P_{(x,y)}\})$ of value $1-\varepsilon$, $val(\mathcal{G}^{\otimes k}) \leq (1-\varepsilon/2)^{C\varepsilon^2 k/\log(|\Sigma_X||\Sigma_Y|)}$.*

Equipped with Lemma 22 and the parallel repetition theorem, we can now prove our birthday repetition theorems just by selecting the right δ and β.

Proof of Theorem 12. Pick $\delta = \frac{\varepsilon^3 C}{10^3 c}$ and $\beta = \frac{\varepsilon^3 C}{10^{10} c}$ where C is the constant from Theorem 23. From Theorem 23, we have $val(\mathcal{G}^{\otimes r}) \leq (1 - \varepsilon/2)^{C\varepsilon^2 r/c}$. Let $R = C\varepsilon^2 r/c$. We can see that R, δ, β satisfy the conditions in Lemma 22. Hence, we can conclude that $val(\mathcal{G}^{k \times l}) \leq 2(1 - \varepsilon/2)^{R/10} = 2(1 - \varepsilon/2)^{(C^2/10^{11})\left(\frac{\varepsilon^5 kl|E|}{c^2|X||Y|d_{max}}\right)}$ as desired. ◀

In the case of projection games, we can improve dependency on ε and get rid of dependency on c thanks to the stronger bound in Rao's parallel repetition theorem for projection games [50].

▶ **Theorem 24** ([50]). *There exists a constant $C > 0$ such that, for any projection game $\mathcal{G}$ of value $1 - \varepsilon$ and for every $k > 0$, we have $val(\mathcal{G}^{\otimes k}) \leq (1 - \varepsilon/2)^{C\varepsilon k}$.*

Proof of Theorem 13. Pick $\delta = \frac{\varepsilon^2 C}{10^3}$ and $\beta = \frac{\varepsilon^2 C}{10^{10}}$ where C is the constant from Theorem 24. From the theorem, we have $val(\mathcal{G}^{\otimes r}) \leq (1 - \varepsilon/2)^{C\varepsilon r}$. Let $R = C\varepsilon r$. By evoking Lemma 22, we have $val(\mathcal{G}^{k \times l}) \leq 2(1 - \varepsilon/2)^{R/10} = 2(1 - \varepsilon/2)^{(C^2/10^{11})\left(\frac{\varepsilon^3 kl|E|}{|X||Y|d_{max}}\right)}$ as desired. ◀

4 Improved Approximation Algorithm for Dense CSPs

To describe our algorithm, we first explain ingredients central in conditioning-based algorithms: a LP/SDP hierarchy, a conditioning operator, and an independent rounding procedure.

Sherali-Adams (SA) relaxation of Max k-CSP. An *r-level SA solution* μ of $\mathcal{G} = (V, \mathcal{W}, \{P_S\})$ is a collection $\{\mathcal{X}_S\}$ of distributions $\mathcal{X}_S$ on Σ^S for every $S \in \binom{V}{[r]}$ such that, for every $S, T \in \binom{V}{[r]}$, the marginal probability of $\mathcal{X}_S$ and $\mathcal{X}_T$ on $\Sigma^{S \cap T}$ agrees. For $r \geq k$, the value of μ is $val_{SA}(\mu) = \mathbb{E}_{S \sim \mathcal{W}}[\mathbb{E}_{x_S \sim \mu}[P_S(x_S)]]$ where $\mathbb{E}_{x_S \sim \mu}[P_S(x_S)]$ is a shorthand for $\mathbb{E}_{\phi_S \sim \mathcal{X}_{\{i_1,\dots,i_k\}}}[P_S(\phi_S)]$ when $S = (x_{i_1}, \dots, x_{i_k})$. The optimum of the r-level SA relaxation of $\mathcal{G}$, $opt^r_{SA}(\mathcal{G})$, is the maximum value among all the r-level SA solutions. Clearly, finding $opt^r_{SA}(\mathcal{G})$ can be formulated as a LP and can be computed in $(nq)^{O(r)}$ time.

Conditioning SA Solution. Let $\mu = \{\mathcal{X}_S\}$ be any r-level SA solution. For any $T \subseteq V$ and $\phi_T \in \Sigma^T$ such that $\mathcal{X}_T(\phi_T) > 0$, μ conditioned on ϕ_T is $\mu|\phi_T = \{\tilde{\mathcal{X}}_S\}_{|S| \leq r - |T|}$ where

$$\tilde{\mathcal{X}}_S(\phi_S) = \begin{cases} \mathcal{X}_{S \cup T}(\phi_S \circ \phi_T)/\mathcal{X}_T(\phi_T) & \text{if } \phi_S \text{ is consistent with } \phi_T, \\ 0 & \text{otherwise.} \end{cases}$$

It is not hard to see that $\mu|\phi_T$ is an $(r - |T|)$-level SA solution.

Independent Rounding. A natural way to round a SA solution $\{\mathcal{X}_S\}$ is to independently assign each variable x based on $\mathcal{X}_x$. This gives a solution with expected value at least $\mathbb{E}_{S=(x_{i_1},\dots,x_{i_k}) \sim \mathcal{W}} \left[\mathbb{E}_{\phi_S \sim \mathcal{X}_{i_1} \times \cdots \times \mathcal{X}_{i_k}} [P_S(\phi_S)]\right]$ and can be easily derandomized.

Without going into too much detail, conditioning-based algorithms typically proceed as follows. First, solve a LP/SDP relaxation of the problem. As long as the solution has large "total correlation", try conditioning it on an assignment to a random variable. Once the solution has small total correlation, use independent rounding on the solution to get the desired assignment. The intuition here is that, if the solution has large total correlation, conditioning on one variable substantially reduces the total correlation. Hence, after a certain number of rounds of conditioning, the total correlation becomes small. At this point, the solution is quite independent and independent rounding gives a good approximation.

Our algorithm will also follow this framework. In fact, it remains largely unchanged from [60] except that we use a stronger relaxation to avoid arguing about values of conditioned solutions. However, our main contribution lies in the analysis: we will show that independent rounding does well even when the total correlation is large (super-constant). This is in contrast to the previously known conditioning-based algorithms [11, 49, 60], all of which require their measures of correlation to be small constants to get any meaningful result.

The new relaxation, which we call the r-level SA with Conditioning (SAC), is defined below.

$$\begin{aligned} &\text{maximize } \lambda \\ &\text{subject to } \{\mathcal{X}_S\}_{|S| \leq r} \text{ is a valid } r\text{-level SA solution} \\ &\qquad \underset{S \sim \mathcal{W}}{\mathbb{E}}\Big[\underset{\phi_S \sim (\mu|\phi_T)}{\mathbb{E}}[P_S(\phi_S)]\Big] \geq \lambda \qquad \forall T, \phi_T \text{ s.t. } |T| \leq r - k, \mathcal{X}_T(\phi_T) > 0. \end{aligned}$$

If λ is a constant, the program can be easily written as a LP. Thus, the relaxation can be solved to within arbitrarily small error in $(nq)^{O(r)}$ time by binary search on λ.

```
Algorithm 1 Approximation Algorithm for Dense CSPs
Input: a $\Delta$-dense Max $k$-CSP instance $\mathcal{G}$, an integer $i$
Output: An assignment $\phi : V \to \Sigma$
  $r \leftarrow k, \lambda \leftarrow 0$
  while $(r - k)\lambda < k^2 i/\Delta$ and $r < n$ do
      $r \leftarrow r + 1$
      $\mu, \lambda \leftarrow$ solve $r$-level of SAC relaxation for $\mathcal{G}$
  for $T \in \binom{V}{[r-k]}, \phi_T \in \Sigma^T$ do
      $\phi \leftarrow$ independent rounding of $\mu|\phi_T$
  return $\phi$ from the previous step with maximum value
```

Figure 1 Approximation Algorithm for Dense CSPs. The difference between this and the above summary is that we iteratively increase the number of levels r. This is because the number of levels depends on the value of the solution (see Lemma 28). Specifically, we need $r \geq k^2 i/(\Delta\lambda) + k$.

Roughly speaking, our algorithm first solves an $O(\frac{k^2 i}{\Delta} + k)$-level SAC relaxation for the instance. We then try every possible conditioning (i.e., every assignment to $T \subseteq V$ of size $\leq k^2 i/\Delta$). For each conditioned solution, we use independent rounding to arrive at an assignment. Finally, output the best such assignment. The pseudo-code for the full algorithm is shown in Figure 1. This algorithm yields the following approximation for the problem.

▶ **Theorem 25** (Theorem 8, Restated). *On any* Δ*-dense* Max k-CSP *instance of value* $1 - \delta$*, Algorithm 1 outputs an assignment of value at least* $(1 - \delta)\delta^{\frac{\delta}{1-\delta}}/q^{1/i}$ *in time* $N^{O\left(\frac{ki}{(1-\delta)\Delta}\right)}$.

We spend the rest of the section sketching the proof of Theorem 25. First, we define and state a bound on the total correlation of conditioned solutions in Section 4.1. Then, in Subsection 4.2, we state our main contribution of this section, i.e., that even when the total correlation is super-constant, independent rouding still yields non-trivial approximation.

4.1 Total Correlation of Conditioned Sherali-Adams Solution

For a k-level SA solution $\mu = \{\mathcal{X}_S\}$ and a tuple $S = (x_{i_1}, \dots, x_{i_j}) \in V^j$ where $j \leq k$, the total correlation of S is $C_\mu(x_S) = C(\sigma_{i_1}; \dots; \sigma_{i_j})$ where $\sigma_{i_1}, \dots, \sigma_{i_j}$ are jointly sampled from $\mathcal{X}_{\{x_{i_1}, \dots, x_{i_j}\}}$. The total correlation of μ is then defined as $C(\mu) = \mathbb{E}_{S \sim \mathcal{W}}[C_\mu(x_S)]$. μ is said to be κ*-independent* if $C(\mu) \leq \kappa$. Yoshida and Zhou [60] show that, for any $l > 0$ and any $(l + k)$-level SA solution μ, there exists an assignment $\phi_T \in \Sigma^T$ to a subset T of size $\leq l$ such that the total correlation of $(\mu|\phi_T)$ is at most $3^k \log q/(l\Delta)$. Here we can improve this bound as stated below. Since the proof is similar to that of [60], we defer it to the full version of this work [42].

▶ **Lemma 26.** *Let* μ *be a* r*-level SA solution of a* Δ*-dense* Max k-CSP *instance* $(V, \mathcal{W}, \{P_S\})$*. Then, for any* $0 < l \leq r - k$*, there is* $t \leq l$ *such that* $\mathbb{E}_{T \sim V^t, \phi_T \sim \Sigma^T}[C(\mu|\phi_T)] \leq \frac{k^2 \log q}{l\Delta}$.

4.2 New Bound on Rounding κ-independent Solution

For the known conditioning-based algorithms, once the solution is fairly independent, it is easy to show that independent rounding gives a good solution. Specifically, [49] and [60] conclude this step using the Pinsker's inequality, which states that, for any distributions $\mathcal{X}$ and $\mathcal{Y}$, $D_{KL}(\mathcal{X}\|\mathcal{Y}) \geq (2\log 2)\|\mathcal{X} - \mathcal{Y}\|_1^2$. Roughly speaking, $\mathcal{X}$ is the distribution in the LP solution whereas $\mathcal{Y}$ is the distribution from independent rounding. Hence, once $D_{KL}(\mathcal{X}\|\mathcal{Y})$ is at most a small constant ε, it follows that, for any predicate f, $|\mathbb{E}_{x\sim\mathcal{X}}[f(x)] - \mathbb{E}_{y\sim\mathcal{Y}}[f(y)]| \leq \sqrt{\varepsilon/(2\log 2)}$. Thus, if $\mathbb{E}_{x\sim\mathcal{X}}[f(x)]$, the value of the LP solution, is large, then $\mathbb{E}_{y\sim\mathcal{Y}}[f(y)]$, the expected value of a solution from independent rouding, is also large.

While this works for small ε, it completely fails when ε is larger than a certain constant. In this regard, we prove the following lemma, which gives a non-trivial bound even for large ε. For convenience, 0^0 is defined to be 1 and $(\delta^\delta e^{-\kappa})^{\frac{1}{1-\delta}}(1-\delta)$ is defined to be 0 when $\delta = 1$.

▶ **Lemma 27.** *For any two probability distributions $\mathcal{X}, \mathcal{Y}$ over Θ such that $D_{KL}(\mathcal{X}\|\mathcal{Y}) \leq \kappa$ and any $f : \Theta \to [0,1]$, if $\mathbb{E}_{x\sim\mathcal{X}}[f(x)] = 1-\delta$, then $\mathbb{E}_{y\sim\mathcal{Y}}[f(y)] \geq \left(\delta^\delta e^{-\kappa}\right)^{\frac{1}{1-\delta}}(1-\delta)$.*

Lemma 27 can then be used to prove a new lower bound for the value of the output from independent rounding on a κ-independent k-level SA solution as stated below.

▶ **Lemma 28.** *If $\{\mathcal{X}_S\}$ is a κ-independent k-level SA solution of value $1-\delta$ for a* Max *k-CSP instance, then independent rounding gives an assignment of value at least $(\delta^\delta e^{-\kappa})^{\frac{1}{1-\delta}}(1-\delta)$.*

Theorem 25 can now be proved by combining Lemma 26 and 28. Due to space constraint, we omit the proofs of Lemma 27, Lemma 28 and Theorem 25 from this extended abstract.

5 Conclusion and Open Problems

While we settle down the approximability of dense Max k-CSP up to a $k\,\mathrm{polylog}(ki)$ factor in the exponent, our work raises many interesting questions such as the two listed below:

- *Can Lemma 27 be used to prove new approximation guarantees for other problems?* Lemma 27 is a generic bound relating expectations of a function on two distributions based on their informational divergence. Thus, it may help yield new approximation guarantees for other correlation-based algorithms.
- *What is the right dependency on ε and c in the birthday repetition theorem?* It is likely that the dependency of ε and c in our birthday repetition is not tight. In particular, parallel repetition for general games only has $1/c$ factor in the exponent whereas our theorem has $1/c^2$; would it be possible to reduce the dependency to $1/c$ in birthday repetition? Similar question also applies to ε.

Acknowledgements. PM would like to thank Aviad Rubinstein, Dana Moshkovitz, Grigory Yaroslavtsev and Madhur Tulsiani for useful discussions. We also thank Irit Dinur for informing us about her result based on gap-ETH.

References

1 Scott Aaronson, Salman Beigi, Andrew Drucker, Bill Fefferman, and Peter W. Shor. The power of unentanglement. *Theory of Computing*, 5(1):1–42, 2009.

2 Scott Aaronson, Russell Impagliazzo, and Dana Moshkovitz. AM with multiple Merlins. In *IEEE CCC*, pages 44–55, June 2014.

3 Noga Alon, Wenceslas Fernandez de la Vega, Ravi Kannan, and Marek Karpinski. Random sampling and approximation of MAX-CSPs. *J. Comput. Syst. Sci.*, 67(2):212–243, September 2003.

4 Benny Applebaum. Pseudorandom generators with long stretch and low locality from random local one-way functions. *SIAM J. Comput.*, 42(5):2008–2037, 2013.

5 Sanjeev Arora, David Karger, and Marek Karpinski. Polynomial time approximation schemes for dense instances of NP-hard problems. In *ACM STOC*, pages 284–293, 1995.

6 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.

7 Sanjeev Arora and Madhu Sudan. Improved low-degree testing and its applications. *Combinatorica*, 23(3):365–426, 2003.

8 Yakov Babichenko, Christos H. Papadimitriou, and Aviad Rubinstein. Can almost everybody be almost happy? In *ACM ITCS*, pages 1–9, 2016.

9 Boaz Barak, Fernando G. S. L. Brandao, Aram W. Harrow, Jonathan Kelner, David Steurer, and Yuan Zhou. Hypercontractivity, sum-of-squares proofs, and their applications. In *ACM STOC*, pages 307–326, 2012.

10 Boaz Barak, Moritz Hardt, Thomas Holenstein, and David Steurer. Subsampling mathematical relaxations and average-case complexity. In *ACM-SIAM SODA*, pages 512–531, 2011.

11 Boaz Barak, Prasad Raghavendra, and David Steurer. Rounding semidefinite programming hierarchies via global correlation. In *IEEE FOCS*, pages 472–481, 2011.

12 Mihir Bellare, Shafi Goldwasser, Carsten Lund, and Alexander Russell. Efficient probabilistically checkable proofs and applications to approximations. In *ACM STOC*, pages 294–304, 1993.

13 Aditya Bhaskara, Moses Charikar, Eden Chlamtác, Uriel Feige, and Aravindan Vijayaraghavan. Detecting high log-densities: An $O(n^{1/4})$ approximation for densest k-subgraph. In *ACM STOC*, pages 201–210, 2010.

14 Mark Braverman and Ankit Garg. Small value parallel repetition for general games. In *ACM STOC*, pages 335–340, 2015.

15 Mark Braverman, Young Kun Ko, Aviad Rubinstein, and Omri Weinstein. ETH hardness for densest-k-subgraph with perfect completeness. In *ACM-SIAM SODA*, pages 1326–1341, 2017.

16 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 *ACM-SIAM SODA*, pages 970–982, 2015.

17 Moses Charikar, MohammadTaghi Hajiaghayi, and Howard Karloff. Improved approximation algorithms for label cover problems. *Algorithmica*, 61(1):190–206, 2011.

18 Eden Chlamtác, Michael Dinitz, Christian Konrad, Guy Kortsarz, and George Rabanca. The densest k-subhypergraph problem. In *APPROX*, pages 6:1–6:19, 2016.

19 Eden Chlamtác, Pasin Manurangsi, Dana Moshkovitz, and Aravindan Vijayaraghavan. Approximation algorithms for label cover and the log-density threshold. In *ACM-SIAM SODA*, pages 900–919, 2017.

20 Amin Coja-Oghlan, Colin Cooper, and Alan Frieze. An efficient sparse regularity concept. *SIAM J. Discret. Math.*, 23(4):2000–2034, 2010.

21 Wenceslas Fernandez de la Vega, Marek Karpinski, Ravi Kannan, and Santosh Vempala. Tensor decomposition and approximation schemes for constraint satisfaction problems. In *ACM STOC*, pages 747–754, 2005.

22 Wenceslas Fernandez de la Vega and Claire Kenyon-Mathieu. Linear programming relaxations of Maxcut. In *ACM-SIAM SODA*, pages 53–61, 2007.

23 Irit Dinur. Mildly exponential reduction from gap 3SAT to polynomial-gap label-cover. *ECCC*, 23:128, 2016.

24 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.

25 Irit Dinur and David Steurer. Analytical approach to parallel repetition. In *ACM STOC*, pages 624–633, 2014.

26 Shaddin Dughmi. On the hardness of signaling. In *IEEE FOCS*, pages 354–363, 2014.

27 Uriel Feige, David Peleg, and Guy Kortsarz. The dense k-subgraph problem. *Algorithmica*, 29(3), 2001.

28 Uriel Feige and Michael Seltser. On the densest k-subgraph problems, 1997.

29 Dimitris Fotakis, Michael Lampis, and Vangelis Th. Paschos. Sub-exponential approximation schemes for CSPs: From dense to almost sparse. In *STACS*, pages 37:1–37:14, 2016.

30 Alan M. Frieze and Ravi Kannan. The regularity lemma and approximation schemes for dense problems. In *FOCS*, pages 12–20, 1996.

31 Venkatesan Guruswami and Ali Kemal Sinop. Lasserre hierarchy, higher eigenvalues, and approximation schemes for graph partitioning and quadratic integer programming with PSD objectives. In *IEEE FOCS*, pages 482–491, 2011.

32 Mohammad Taghi Hajiaghayi, Kamal Jain, Kishori M. Konwar, Lap Chi Lau, Ion I. Măndoiu, Alexander Russell, Alexander A. Shvartsman, and Vijay V. Vazirani. The minimum k-colored subgraph problem in haplotyping and DNA primer selection. In *IWBRA*, 2006.

33 Aram W. Harrow and Ashley Montanaro. Testing product states, quantum Merlin-Arthur games and tensor optimization. *J. ACM*, 60(1):3:1–3:43, February 2013.

34 Thomas Holenstein. Parallel repetition: Simplification and the no-signaling case. *Theory of Computing*, 5(1):141–172, 2009.

35 Russel Impagliazzo and Ramamohan Paturi. On the complexity of k-sat. *J. Comput. Syst. Sci.*, 62(2):367–375, March 2001.

36 Guy Kortsarz and David Peleg. On choosing a dense subgraph. In *IEEE SFCS*, pages 692–701, 1993.

37 Jean B. Lasserre. Global optimization with polynomials and the problem of moments. *SIAM J. on Optimization*, 11(3):796–817, March 2000.

38 Richard J. Lipton, Evangelos Markakis, and Aranyak Mehta. Playing large games using simple strategies. In *ACM EC*, pages 36–41, 2003.

39 Pasin Manurangsi. Almost-polynomial ratio ETH-hardness of approximating densest k-subgraph. In *ACM STOC*, 2017. To appear.

40 Pasin Manurangsi and Dana Moshkovitz. Approximating dense Max 2-CSPs. In *APPROX*, pages 396–415, 2015.

41 Pasin Manurangsi and Dana Moshkovitz. Improved approximation algorithms for projection games. *Algorithmica*, pages 1–40, 2015.

42 Pasin Manurangsi and Prasad Raghavendra. A birthday repetition theorem and complexity of approximating dense CSPs. *CoRR*, abs/1607.02986, 2016. URL: `https://arxiv.org/abs/1607.02986`.

43 Pasin Manurangsi and Aviad Rubinstein. Inapproximability of VC Dimension and Littlestone's Dimension. Unpublished manuscript, 2017.

44 Claire Mathieu and Warren Schudy. Yet another algorithm for dense max cut: go greedy. In *ACM-SIAM SODA*, pages 176–182, 2008.

45 Dana Moshkovitz. The projection games conjecture and the NP-hardness of $\ln n$-approximating set-cover. In *APPROX 2012*, pages 276–287, 2012.

46 Dana Moshkovitz. Parallel repetition from fortification. In *IEEE FOCS*, pages 414–423, 2014.

47 Dana Moshkovitz and Ran Raz. Two-query PCP with subconstant error. *J. ACM*, 57(5):29:1–29:29, June 2008.

48 David Peleg. Approximation algorithms for the label-cover max and red-blue set cover problems. *Journal of Discrete Algorithms*, 5(1):55–64, 2007.

49 Prasad Raghavendra and Ning Tan. Approximating CSPs with global cardinality constraints using SDP hierarchies. In *ACM-SIAM SODA*, pages 373–387, 2012.

50 Anup Rao. Parallel repetition in projection games and a concentration bound. *SIAM J. Comput.*, 40(6):1871–1891, 2011.

51 Ran Raz. A parallel repetition theorem. *SIAM J. Comput.*, 27(3):763–803, 1998.

52 Ran Raz and Shmuel Safra. A sub-constant error-probability low-degree test, and a sub-constant error-probability PCP characterization of NP. In *ACM STOC*, pages 475–484, 1997.

53 Aviad Rubinstein. ETH-hardness for signaling in symmetric zero-sum games. *CoRR*, abs/1510.04991, 2015.

54 Aviad Rubinstein. Settling the complexity of computing approximate two-player Nash equilibria. In *IEEE FOCS*, pages 258–265, 2016.

55 Grant Schoenebeck. Linear level lasserre lower bounds for certain k-CSPs. In *IEEE FOCS*, pages 593–602, 2008.

56 Hanif D. Sherali and Warren P. Adams. A hierarchy of relaxation between the continuous and convex hull representations. *SIAM J. Discret. Math.*, 3(3):411–430, May 1990.

57 Akiko Suzuki and Takeshi Tokuyama. Dense subgraph problems with output-density conditions. *ACM Trans. Algorithms*, 4(4):43:1–43:18, August 2008.

58 Grigory Yaroslavtsev. Going for speed: Sublinear algorithms for dense r-CSPs. *CoRR*, abs/1407.7887, 2014.

59 Grigory Yaroslavtsev. Personal communication, March 2016.

60 Yuichi Yoshida and Yuan Zhou. Approximation schemes via Sherali-Adams hierarchy for dense constraint satisfaction problems and assignment problems. In *ITCS 2014*, pages 423–438, 2014.