Open Problems Column
Edited by William Gasarch

This Issue's Column! This issue's Open Problem Column is by Erika Melder. It is about lower bounds on approximating set cover. The overarching question is: are there easier proofs?

Request for Columns! I invite any reader who has knowledge of some area to contact me and arrange to write a column about open problems in that area. That area can be (1) broad or narrow or anywhere inbetween, and (2) really important or really unimportant or anywhere inbetween.

Wanted: Easier Proofs of Lower Bounds on Approximating Set Cover
by Erika Melder

1 Introduction

1.1 The Set Cover Problem

The optimization problem SET COVER is defined as follows:

Definition 1.1 (SET COVER)**.** Given a set U (a *universe*) of n elements and sets $S_1, S_2, \ldots, S_k \subseteq U$, with the guarantee that every element of U is in at least one of the S_i, find a minimum-size collection of subsets whose union is U.

This problem is well-known to be NP-hard. As such, the best we can hope for is a good approximation algorithm, which finds a solution to any instance with no more than some function of the smallest possible number of sets. The approximation version of SET COVER is defined as follows:

Definition 1.2 (SET COVER Approximation)**.** For a SET COVER instance $\mathcal{S}$, let $\mathrm{OPT}(\mathcal{S})$ be the number of sets S_i used in a minimum-size solution of $\mathcal{S}$. An algorithm is said to $a(n)$*-approximate* SET COVER, where a is a function of the number of universe elements n in $\mathcal{S}$, if it can generate a solution to any SET COVER instance $\mathcal{S}$ using at most $a(n) \cdot \mathrm{OPT}(\mathcal{S})$ sets.

In particular, we are interested in approximations where $a(n)$ is minimized, and which run in polynomial time.

Chvatal [5] showed that there is a polynomial time approximation algorithm for SET COVER with $a(n) = \ln n$, using a simple greedy algorithm. Since then, no one has developed an improved approximation. Hence, one wonders whether approximating SET COVER better than $\ln n$ is NP-hard. The answer turns out to be yes, and is the culmination of decades of refinements of impossibility results. The proofs of these impossibility results are nontrivial, relying heavily on the machinery of interactive proof systems [3, 9] and the subsequent framework of *probabilistically checkable proofs* (PCP) [1, 2]. This column will trace key developments in proofs of SET COVER inapproximability, beginning with the initial inapproximability result of Lund and Yannakakis [10], and progressing chronologically to explore what techniques were necessary to arrive at the current results. The final result is the following:

Theorem 1.3 (Inapproximability of Set Cover)**.** Unless P = NP, Set Cover cannot be approximated within $c \ln n$ for any $0 < c < 1$.

In this paper we present brief descriptions of seven key papers which progressively improved on the bound for approximating Set Cover, the final paper proving Theorem 1.3. (Full proofs are omitted for brevity, but can be found in Melder [11].) Along the way, we will pose open questions which arise. At the end, we compile these open problems and add some additional ones.

1.2 A Note on the Timeline

The papers are presented in chronological order of their results, but the years of publication listed are those of the respective journals that the papers are ultimately published in. Much of the work on this problem was done collaboratively, with initial results sparking debate and further work before those initial results were printed, and with several papers receiving revisions to reflect newly-obtained results of their successors reliant upon the initial results. For ease of reference, the papers in the main portion of this column will be referred to by the date that they were published, though this will cause some of these papers to seemingly appear out of order. Appendix A presents timelines of these papers arranged by rough order of initial results, in the form of a rough sketch of which results were strictly necessary to produce others.

2 Lund and Yannakakis, 1994 [10]

An initial result was the following theorem:

Theorem 2.1. Unless $\mathrm{NP} \subseteq \mathrm{DTIME}(n^{\mathrm{polylog}(n)})$, Set Cover cannot be approximated within a ratio of $c \log n$ for any positive $c < \frac{1}{4}$.

The proof of this result is nontrivial, relying on a modification of a Feige-Lovász proof system for SAT, originally developed in [8]. This section presents a simplified, descriptive version of the proof. We omit complex proofs of background information and treat them as black boxes to present only the core logic and structures behind the key result. Our sketch of of the Lund-Yannakakis result will be longer than those that follow, because most of the later papers are structurally similar to it. Hence, we will often be referring back to it.

Lund and Yannakakis, building on the work of Feige and Lovász, described a two-prover, one-round interactive proof system for SAT in [10], which was subsequently used as a component of the gap-preserving reduction necessary to obtain Theorem 2.1. The definition of such a proof system is as follows:

Definition 2.2 (Interactive Proof System)**.** Fix an input size n. A *two-prover, one-round interactive proof system* is a system which takes in an input of size n and outputs

Yes or No. It has two kinds of components: a *verifier* and *provers*.

- A *verifier* consists of the following:
  - A finite input alphabet Σ.
  - A finite set R of random seeds of length $\mathcal{O}(\text{polylog}(n))$.
  - A finite set of queries of length $\mathcal{O}(\text{polylog}(n))$ that can be asked to each prover. In this case, there will be two provers, so the verifier will have two such sets, Q_1 and Q_2.
  - A finite set of answers of length $\mathcal{O}(\text{polylog}(n))$ that can be received from each prover. Again, there will be two such sets in this case, A_1 and A_2.
  - A polynomial-time computable function $f : \Sigma^n \times R \to Q_1 \times Q_2$, which generates a pair of queries based on a random seed and an input.
  - A polynomial-time computable function $\Pi : \Sigma^n \times R \times A_1 \times A_2 \to \{\mathsf{Yes}, \mathsf{No}\}$ which determines the output based on the input, seed, and query answers.
- The ith *prover* in the system is a function from Q_i to A_i.

The verifier takes an input $x \in \Sigma^n$, and selects a random seed r uniformly randomly from R. It computes $f(x, r)$ and receives a query pair (q_1, q_2). It then provides these two queries to the provers, who in turn provide the answer pair (a_1, a_2); this process of sending queries and receiving answers is known as *message passing*, and the "one-round" designation of the prover refers to using only one iteration of message passing. After this, the verifier computes $\Pi(x, r, a_1, a_2)$, and accepts or rejects the input based on the result.

Lund and Yannakakis obtain the following theorem:

Theorem 2.3 (Proof System for SAT)**.** There exists a two-prover, one-round interactive proof system for SAT that takes an input formula φ and has the following properties:

- If $\varphi \in$ SAT, then there exists a pair of provers such that the verifier will always output Yes.
- If $\varphi \notin$ SAT, then for all pairs of provers, the verifier accepts with probability at most $\frac{1}{n}$, where n is the input size.
- For a fixed prover P_i, the verifier's queries to that prover are uniformly random across the possible query set Q_i.
- The choice of q_2 is independent of the choice of q_1.
- For any pair (r, a_1) of a seed and an answer from prover 1, there is at most one $a_2 \in A_2$ that would result in a pair of accepting answers (i.e. $\Pi(\varphi, r, a_1, a_2) = \mathsf{Yes}$).
- $|Q_1| = |Q_2|$, meaning that the verifier has the same number of potential queries for each prover. (This can be assumed without loss of generality.)

The existence of such a proof system is a cornerstone of the reduction. Note that the input size is fixed for verifiers of this nature, so we simply select an appropriate verifier to accommodate the size of our input. In this case, the size may be considered to be the number of literals in the formula. Note also that the proof of this theorem is nonconstructive.

Open Problem 2.4. Is there an explicit construction of the proof system guaranteed in Theorem 2.3?

Another critical construction is a special set system (which is a variant of an (n, k)-universal set as defined in [14] and described in Definition 5.1), defined as follows:

Definition 2.5 (Special Set System)**.** A *special set system* $\beta_{m,d} = \{B; C_1, C_2, \ldots, C_m\}$ consists of the following:

- A set B, referred to as the *universe* of the set system.
- m subsets $C_1, C_2, \ldots, C_m \subseteq B$, referred to as *special sets.*

Additionally, it has the following property, known as its *special property*:

- No collection of d or fewer indices $i \in \{1, 2, \ldots, m\}$ satisfies $\bigcup_i D_i = B$, where $D_i = C_i$ or $\overline{C_i}$.

Lund and Yannakakis prove the following lemma:

Lemma 2.6. For any positive integers m and d, there exists a special set system $\beta_{m,d}$ with those parameters, and its universe B has $|B| = \mathcal{O}(2^{2d}m^2)$. Moreover, such a set system can be explicitly constructed for any m, d in DTIME$(n^{\text{polylog}(n)})$.

The construction given by Lund and Yannakakis results in a set system whose elements are bit-vectors. For the purposes of the proof, the actual elements contained in the set system are irrelevant; only the special property is required.

One crucial corollary follows from taking the contrapositive of the special property:

Corollary 2.7. Let $\beta_{m,d} = \{B; C_1, C_2, \ldots, C_m\}$ be a special set system, and let $A \subseteq \{C_1, \ldots, C_m, \overline{C_1}, \ldots, \overline{C_m}\}$ be a collection of C_i's and their complements, with $|A| \leq d$. Then for some $j \in \{1, 2, \ldots, m\}$, both C_j and $\overline{C_j}$ are in A.

Lund and Yannakakis devote the remainder of their paper to the following lemma, proving it using the constructions previously defined.

Lemma 2.8. For a Set Cover instance $\mathcal{S}$, let $\mathsf{OPT}(\mathcal{S})$ denote the smallest number of sets needed to fully cover the universe of $\mathcal{S}$. Then, for any CNF formula φ, we can construct an instance of Set Cover $\mathcal{S}_\varphi$ in DTIME$(n^{\text{polylog}(n)})$ such that:

- If $\varphi \in$ SAT, then $\mathsf{OPT}(\mathcal{S}_\varphi) = |Q_1| + |Q_2|$, where Q_i is the set of queries that may be asked to prover i in an appropriate Feige-Lovász SAT proof system for φ, and
- If $\varphi \notin$ SAT, then $\mathsf{OPT}(\mathcal{S}_\varphi) \geq c \log n(|Q_1| + |Q_2|)$ for any $0 < c < \frac{1}{4}$.

This would complete a gap-inducing reduction from SAT to Set Cover. Because SAT is NP-complete, this would prove that Set Cover is inapproximable within $c\log(n)\cdot$ OPT for any $0 < c < \frac{1}{4}$, unless NP $\subseteq$ DTIME($n^{\text{polylog}(n)}$). The second possibility is because the special set system construction can be done in DTIME($n^{\text{polylog}(n)}$) as shown in [10]. This would subsequently prove Theorem 2.1.

2.1 Using a Randomized Construction

By using a randomized construction for the special set system, rather than a deterministic one, Lund and Yannakakis are able to construct it in ZTIME($n^{\text{polylog}(n)}$), where ZTIME is zero-error probabilistic polynomial time (the class of languages decidable via Las Vegas algorithms). The size of the new special set is $|B| = (d + d\ln m + 2)2^d$, which is approximately the square root of the size of the deterministic algorithm, which is $|B| = \mathcal{O}(2^{2d}m^2)$. Using this, they achieve the following result.

Theorem 2.9. Unless NP $\subseteq$ ZTIME($n^{\text{polylog}(n)}$), Set Cover cannot be approximated within a ratio of $c\log n$ for any positive $c < \frac{1}{2}$.

This is a stronger result than using the deterministic construction, but requires a stronger assumption in turn. As such, it is effectively incomparable to the deterministic result in terms of quality.

3 Bellare, Goldwasser, Lund, and Russell, 1993 [4]

3.1 Inapproximability Result

The main theorem that this paper achieves is as follows:

Theorem 3.1.

1. Unless P = NP, Set Cover cannot be approximated within a ratio of c for any constant c.
2. Unless NP $\subseteq$ DTIME($n^{\mathcal{O}(\log\log n)}$), Set Cover cannot be approximated within a ratio of $c\log n$ for any positive $c < \frac{1}{8}$.

These results come from increasing the number of provers in the system to four, and providing a slightly different construction in place of the special set system. Though the assumptions of these results are weaker, the constants are also worse; therefore, this result is effectively incomparable with the result of Section 2.

3.2 The Sliding Scale Conjecture

The paper also formulates the following conjecture, a weaker version of which was ultimately proven in [6].

Conjecture 3.2 (Sliding Scale Conjecture)**.** There exist two-prover, one-round proof systems for SAT that have logarithmic randomness, logarithmic answer sizes, and $\frac{1}{n}$ error probability.

Contingent on this conjecture, the authors posit that better Set Cover results are achievable. In particular, they observe that a Set Cover result arises from a weaker form of the conjecture.

Conjecture 3.3 (Weak Sliding Scale Conjecture for Set Cover)**.** Suppose that for some constant p and some function $\epsilon(n) = \frac{1}{\log^{\omega(1)} n}$, SAT has p-prover proof systems with logarithmic randomness, logarithmic answer sizes, and error $\epsilon(n)$. Then, unless P = NP, Set Cover cannot be approximated within a ratio of $c \log n$ for any positive $c < \frac{1}{2p}$.

While this conjecture remains unproven, it forms the basis of the Projection Games Conjecture in [12], presented in this paper as Conjecture 7.6.

Open Problem 3.4 (The Sliding Scale Conjecture)**.**

1. Is the Sliding Scale Conjecture (Conjecture 3.2) true? If so, then the Weak Sliding Scale Conjecture follows.
2. Is the Weak Sliding Scale Conjecture (Conjecture 3.3) true?

4 Raz, 1998 [15]

Raz does not focus on Set Cover in this paper; rather, he offers a breakthrough in error of multi-prover interactive proof systems in the form of the following lemma:

Lemma 4.1 (Raz Repetition Lemma)**.** If a two-prover, one-round proof system is repeated ℓ times independently in parallel, then the error is $2^{-c\ell}$, where $c > 0$ is constant and is dependent only on the error in the original proof system and the length of the answers in that proof system.

This implies that parallel repetition of a two-prover one-round proof system reduces the error of the system exponentially fast. The proof of the theorem is extremely complicated, but this result is critical for achieving the lower bounds of future proofs.

Open Problem 4.2.

(a) Does there exist a simpler proof of the Raz Repetition Lemma?

(b) Does there exist a proof which does not rely on parallel repetition to reduce error? Such a proof would eliminate reliance on the Raz Repetition Lemma altogether.

Taken in combination with Theorems 2.1 and 2.9, the hardness assumptions can be relaxed, yielding the following:

Theorem 4.3.

1. Unless NP $\subseteq$ DTIME($n^{\mathcal{O}(\log \log n)}$), Set Cover cannot be approximated within a ratio of $c \log n$ for any positive $c < \frac{1}{4}$.
2. Unless NP $\subseteq$ ZTIME($n^{\mathcal{O}(\log \log n)}$), Set Cover cannot be approximated within a ratio of $c \log n$ for any positive $c < \frac{1}{2}$.

Both parts of this theorem are strict improvements on the results of [10]; the first part is also a strict improvement over the second part of Theorem 3.1 from [4].

5 Naor, Schulman, and Srinivasan, 1995 [14]

5.1 (n, k)-Universal Sets

The key component of this paper was the development of new techniques for constructing (n, k)-universal sets, which are defined as follows:

Definition 5.1 ((n, k)-Universal Set)**.** An (n, k)-universal set $T \subseteq \{0, 1\}^n$ is a minimal set of bitstrings such that, for any set of indices $S \subseteq [n]$ with $|S| = k$, the projection of T on S (defined as the set of subsequences of bitstrings of T taken at the indices in S) is exactly the set $\{0, 1\}^k$.

These sets are the basis for the Special Set System in Lund and Yannakakis (see Definition 2.5). Naor, Schulman, and Srinivasan developed new deterministic procedures for constructing these sets in [14]. Their main result is the following:

Theorem 5.2. There is a deterministic, explicit construction of (n, k)-universal sets of size $2^k k^{\mathcal{O}(\log k)} \log n$.

The theorem is a corollary of their more general result on k-restriction problems. Their procedure makes the deterministic construction much more efficient, yielding performance on par with the randomized construction. This unifies the deterministic and randomized lower bounds, yielding the following refinement of Theorem 4.3:

Theorem 5.3. Unless $\mathrm{NP} \subseteq \mathrm{DTIME}(n^{\mathcal{O}(\log \log n)})$, Set Cover cannot be approximated within a ratio of $c \log n$ for any positive $c < \frac{1}{2}$.

5.2 (n, k, b)-Anti-Universal Sets

In addition to (n, k)-universal sets, another result of Naor *et al.* on restriction games involves a structure known as (n, k, b)-anti-universal sets, which were initially proposed by Feige and are a cornerstone of the reduction in [7]. They are defined as follows:

Definition 5.4 ((n, k, b)-Anti-Universal Set)**.** An (n, k, b)-anti-universal set $\mathcal{T} \subseteq [n] \times [b]$ is a family of functions $t_i : [n] \to [b]$ where, for every pair of vectors $u \in [n]^k$ and $v \in [b]^k$, there is a function $t_j \in \mathcal{T}$ which transforms u into a vector that disagrees with v on each coordinate (meaning that t_j takes every element $u_i \in u$ to something *other* than the corresponding element $v_i \in v$).

The name "anti-universal" is rather misleading. They are simply a generalization of (n, k)-universal sets, which are the case where $b = 2$. Naor *et al.* achieve the following result:

Theorem 5.5. For any fixed b, there is a deterministic, explicit construction of (n, k, b)-anti-universal sets of size $\left(\frac{b}{b-1}\right)^k k^{\mathcal{O}(\log k)} \log n$.

In combination with the reduction of Section 6, this is sufficient to prove Theorem 6.1.

Open Problem 5.6. Theorem 2.1 relies on (n, k)-universal sets, while Theorem 6.1 relies on the slightly more general (n, k, b)-anti-universal sets. However, both of these structures are variants on the more general notion of k-restriction problems. Is it possible to unify the Lund-Yannakakis result with the Feige result in a parametrized framework relating to restriction games? Could such a framework yield better results, or simpler proofs of these results?

6 Feige, 1998 [7]

Feige provides the first crucial innovation by generalizing the frameworks of [10] and by beginning from MAX 3SAT-5 rather than conventional SAT. The ultimate result is as follows:

Theorem 6.1. Unless NP $\subseteq$ DTIME($n^{\mathcal{O}(\log\log n)}$), Set Cover cannot be approximated within a ratio of $c \ln n$ for any positive $c < 1$.

We omit the complete proof for brevity, but as with Section 2, we describe several constructions and concepts necessary for Feige's proof. Many of these are generalizations of constructions in Section 2.

The problem of MAX 3SAT-5 is a particular member of a family of SAT problems with restrictions on the number of clauses each variable appears in. The problem statement is:

Definition 6.2 (MAX 3SAT-5)**.** Consider a CNF boolean formula φ on n variables and $\frac{5}{3}n$ clauses, where the following three conditions hold (such formulas are known as 3CNF-5 formulas):

- Each clause contains 3 literals.
- Each variable appears in exactly 5 clauses.
- No variable appears in the same clause more than once.

The problem of MAX 3SAT-5 is to determine the maximum number of clauses that are simultaneously satisfiable.

The following is a theorem of Feige:

Theorem 6.3. For some $\varepsilon > 0$, it is NP-hard to distinguish between 3SAT-5 formulas that are completely satisfiable, and those with at most a $(1 - \varepsilon)$ fraction of clauses simultaneously satisfiable.

It is this gap in distinguishability that will ultimately lead to the gap reduction. The choice of MAX 3SAT-5 also leads to several open questions.

Open Problem 6.4.

(a) Is it possible to reframe the Lund-Yannakakis proof in terms of a restricted version of SAT to simplify it?

(b) Is it possible to reframe Feige's proof in terms of generic SAT? Would such an argument lead to techniques that could be applied to streamline other proofs?

(c) Could aCNF-b formulas for other values of a and b yield simpler constructions? What about formulas where a or b (or both) is an upper bound, rather than an exact bound as in MAX 3SAT-5? Exploring aCNF-b formulas in general may yield new insights.

Feige also presents an explicit protocol for a two-prover system for MAX 3SAT-5. (See Definition 2.2 for the structure of such a system.) This is different from the methodology of Lund and Yannakakis, who used a theorem asserting the existence of such a protocol without actually describing it. Analysis of Feige's proof system results in the following theorem:

Theorem 6.5. Let φ be a 3CNF-5 formula, and let ε be the fraction of unsatisfied clauses in the assignment that satisfies the *most* possible clauses. Then, under the optimal strategy of the provers, the verifier accepts with probability $1 - \frac{\varepsilon}{3}$.

Note that the error in this proof system is still high, but applying Lemma 4.1 reduces the error to $2^{-c\ell}$ for a universal constant c.

Feige is able to generalize this system to k provers, and establishes two notions of acceptance with which to generate a gap.

Feige also defines a *partition system*, a generalization of the special set system in Definition 2.5 and an invocation of the (n, k, b)-anti-universal set of Definition 5.4.

Definition 6.6 (Partition System)**.** A *partition system* $\beta(m, \mathcal{L}, k, d)$ has the following properties:

- There is a universe set β, with $|\beta| = m$.
- There is a collection of $\mathcal{L}$ distinct partitions of β, $p_1, \ldots, p_{\mathcal{L}}$.
- For $1 \leq i \leq \mathcal{L}$, partition p_i is a collection of k disjoint subsets of β whose union is β.
- Any cover of β by subsets such that no two subsets are from the same partition requires at least d subsets.

Feige obtains the following results:

Lemma 6.7. For every $c \geq 0$ and sufficiently large m, there exists a partition system $\beta(m, \mathcal{L}, k, d)$ such that all of the following hold:

- $\mathcal{L} \simeq (\log m)^c$.
- k may be chosen arbitrarily as long as $k < \ln \frac{m}{3} \ln \ln m$.

- $d = (1 - f(k))k \ln m$, where $\lim_{k \to \infty} f(k) = 0$.

Moreover, such a partition system can be constructed in ZTIME($m^{\mathcal{O}(\log m)}$). By Theorem 5.5, there also exists a deterministic construction taking time linear in m and satisfies $m = \left(\frac{k}{k-1}\right)^d d^{\mathcal{O}(\log d)} \log \mathcal{L}$ for arbitrary constant k.

Compiling these results and unifying them with those of Raz and Naor *et al.* is the key to Feige's proof of Theorem 6.1.

7 Moshkovitz, 2015 [12]

Moshkovitz further generalizes the framework of Feige to achieve the following result:

Theorem 7.1. Unless P = NP and assuming the Projection Games Conjecture holds, Set Cover cannot be approximated within a ratio of $c \ln n$ for any positive $c < 1$.

She does so by shifting the underlying framework to PCPs, which can be analyzed in terms of projection games. These are defined as follows:

Definition 7.2 (Projection Game)**.** A *projection game* takes the following inputs:

- A bipartite graph $G = ((A, B), E)$.
- Finite alphabets Σ_A, Σ_B.
- A set Φ containing a function $\pi_e : \Sigma_A \to \Sigma_B$ for every edge $e \in E$. These functions π_e are the *projections*.

The *size* of the game is defined as $|A| + |B| + |E|$. The following two concepts are also introduced:

- A *labeling* is a function mapping a vertex set to its corresponding alphabet, either $\phi_A : A \to \Sigma_A$ or $\phi_B : B \to \Sigma_B$.
- A pair of labelings ϕ_A and ϕ_B *satisfies* edge $e = (a, b)$ if $\pi_e(\phi_A(a)) = \phi_B(b)$; that is, if the projection π_e takes the label of a to the label of b.

The goal of the game is to find labelings ϕ_A and ϕ_B that satisfy as many edges as possible.

The following theorem is due to Moshkovitz:

Theorem 7.3. Let $\mathcal{P}$ be a projection game of size n with alphabets of size k and soundness error ε, where k and ε may be functions of n. Then, it is NP-hard to distinguish between the case where all edges can be satisfied, and the case where at most a ε fraction of edges can be satisfied.

Unlike the result of Feige, here, the error is explicitly related to the soundness error of the underlying projection game.

The paradigm shift yields the following open problem:

Open Problem 7.4. Proofs up to and including that of Feige rely on interactive proof systems (Definition 2.2). However, these can be converted to probabilistically checkable proofs, analyzed similarly to the projection game in Moshkovitz. Can these older proofs be rewritten in terms of PCP? Could the proofs be simplified or improved using PCP analysis techniques that did not exist at the time of the original publication?

7.0.1 The Projection Games Conjecture

In [13], Moshkovitz and Raz proved the following theorem

Theorem 7.5 (Almost-Linear-Size PCP Theorem)**.** There exists $c > 0$ such that for every $\varepsilon \geq xN^{-c}$, SAT with an input size of n can be reduced to a projection game of input size $N = n^{1+o(1)} \cdot \text{poly}\left(\frac{1}{\varepsilon}\right)$ over an alphabet of size $\exp\left(\frac{1}{\varepsilon}\right)$ and soundness error ε.

In [12], Moshkovitz adapts the Sliding Scale Conjecture of Bellare *et al.* (Conjecture 3.2) to this framework, and conjectures that the alphabet size can be made polynomial in $\frac{1}{\varepsilon}$ rather than exponential. The formal statement of the conjecture is as follows:

Conjecture 7.6 (Projection Games Conjecture)**.** There exists $c > 0$ such that for every $\varepsilon \geq N^{-c}$, SAT with an input size of n can be reduced to a projection game of input size $N = n^{1+o(1)} \cdot \text{poly}\left(\frac{1}{\varepsilon}\right)$ over an alphabet of size $\text{poly}\left(\frac{1}{\varepsilon}\right)$ and soundness error ε.

In both of the above, we may assume without loss of generality that the projection game is *bi-regular*, meaning that all $a \in A$ share a common degree and so do all $b \in B$, though they need not be the same degree for both A and B.

The core lemma of Moshkovitz is given in terms of this conjecture. The full form of the conjecture is not needed to prove that Set Cover is inapproximable; rather, only a certain weaker version of it is necessary. This version in particular was proven by Dinur and Steurer in [6], which leads to the tightest possible lower bound of $\ln(n)$ and concludes development of the bound. From here, Moshkovitz is able to instantiate a projection game and analyze its completeness and soundness to complete her proof.

The Projection Games Conjecture remains open:

Open Problem 7.7 (The Projection Games Conjecture)**.**

1. Is Conjecture 7.6 true? If so, how does the proof differ from that of Theorem 8.1? If not, what part of the conjecture causes the gap in possibility?
2. The Projection Games Conjecture arose from the Sliding Scale Conjecture (presented here as Conjecture 3.2). How does the proof of Theorem 8.1 impact the Sliding Scale Conjecture?

8 Dinur and Steurer, 2013 [6]

The contribution of Dinur and Steurer is a proof of the weaker version of Conjecture 7.6 necessary to achieve the results of Moshkovitz. In particular, they prove the following result about the Label Cover problem:

Theorem 8.1. For every constant $c > 0$, given a LABEL COVER instance of size n with alphabet size at most n, it is NP-hard to decide if its value is 1 or at most $\varepsilon = (\log n)^{-c}$.

This is sufficient to prove Moshkovitz's main claim without reliance on a conjecture, hence proving that SET COVER is inapproximable within $(1 - \varepsilon) \ln n$ for any $0 < \varepsilon < 1$. We therefore arrive at the current result of Theorem 1.3.

9 Open Problems

Many new developments and refinements in this framework are still open.

(1) Is there an explicit construction for the proof system whose existence is guaranteed by Theorem 2.3?

(2) *The Sliding Scale Conjecture*:

   (a) Is the Sliding Scale Conjecture (Conjecture 3.2) true? If so, then the Weak Sliding Scale Conjecture follows.

   (b) Is the Weak Sliding Scale Conjecture (Conjecture 3.3) true?

(3) *Regarding the Raz Repetition Lemma*:

   (a) Does there exist a simpler proof of the Raz Repetition Lemma (presented here as Lemma 4.1?)

   (b) Does there exist a proof which does not rely on parallel repetition to reduce error? Such a proof would eliminate reliance on the Raz Repetition Lemma altogether.

(4) Theorem 2.1 relies on (n, k)-universal sets, while Theorem 6.1 relies on the slightly more general (n, k, b)-anti-universal sets (Definitions 5.1 and Definition 5.4, respectively). However, both of these structures are variants on the more general notion of k-restriction problems. Is it possible to unify the Lund-Yannakakis result with the Feige result in a parametrized framework relating to restriction games? Could such a framework yield better results, or simpler proofs of these results?

(5) Feige's proof uses MAX 3SAT-5 rather than SAT as its initial problem.

   (a) Is it possible to reframe the Lund-Yannakakis proof in terms of a restricted version of SAT to simplify it?

   (b) Is it possible to reframe Feige's proof in terms of generic SAT? Would such an argument lead to techniques that could be applied to streamline other proofs?

   (c) Could aCNF-b formulas for other values of a and b yield simpler constructions? What about formulas where a or b (or both) is an upper bound, rather than an exact bound as in MAX 3SAT-5?

(6) Proofs up to and including that of Feige rely on interactive proof systems (Definition 2.2). However, these can be converted to probabilistically checkable proofs,

analyzed similarly to the projection game in Moshkovitz. Can these older proofs be rewritten in terms of PCP? Could the proofs be simplified or improved using PCP analysis techniques that did not exist at the time of the original publication?

(7) *The Projection Games Conjecture*:

   (a) Is Conjecture 7.6 true? If so, how does the proof differ from that of Theorem 8.1? If not, what part of the conjecture causes the gap in possibility?

   (b) The Projection Games Conjecture arose from the Sliding Scale Conjecture. How does the proof of Theorem 8.1 impact the Sliding Scale Conjecture?

(8) The three core proofs of results (Lund and Yannakakis, Feige, Moshkovitz) each rely on some form of large, bipartite construction whose members correspond in some way to a universe of sets. Is there a simpler way to formulate these proofs? If so, do the simpler constructions enable smaller space complexity of the reductions?

(9) The union of these results provides a complete proof that P $\neq$ NP implies that $a(n) = \ln n$ is an exact bound on polynomial-time approximation for SET COVER. However, the proof is extremely difficult. Would a stronger assumption lead to a simpler proof? We suggest three assumptions:

   - The Unique Games Conjecture
   - The Exponential Time Hypothesis
   - The Strong Exponential Time Hypothesis

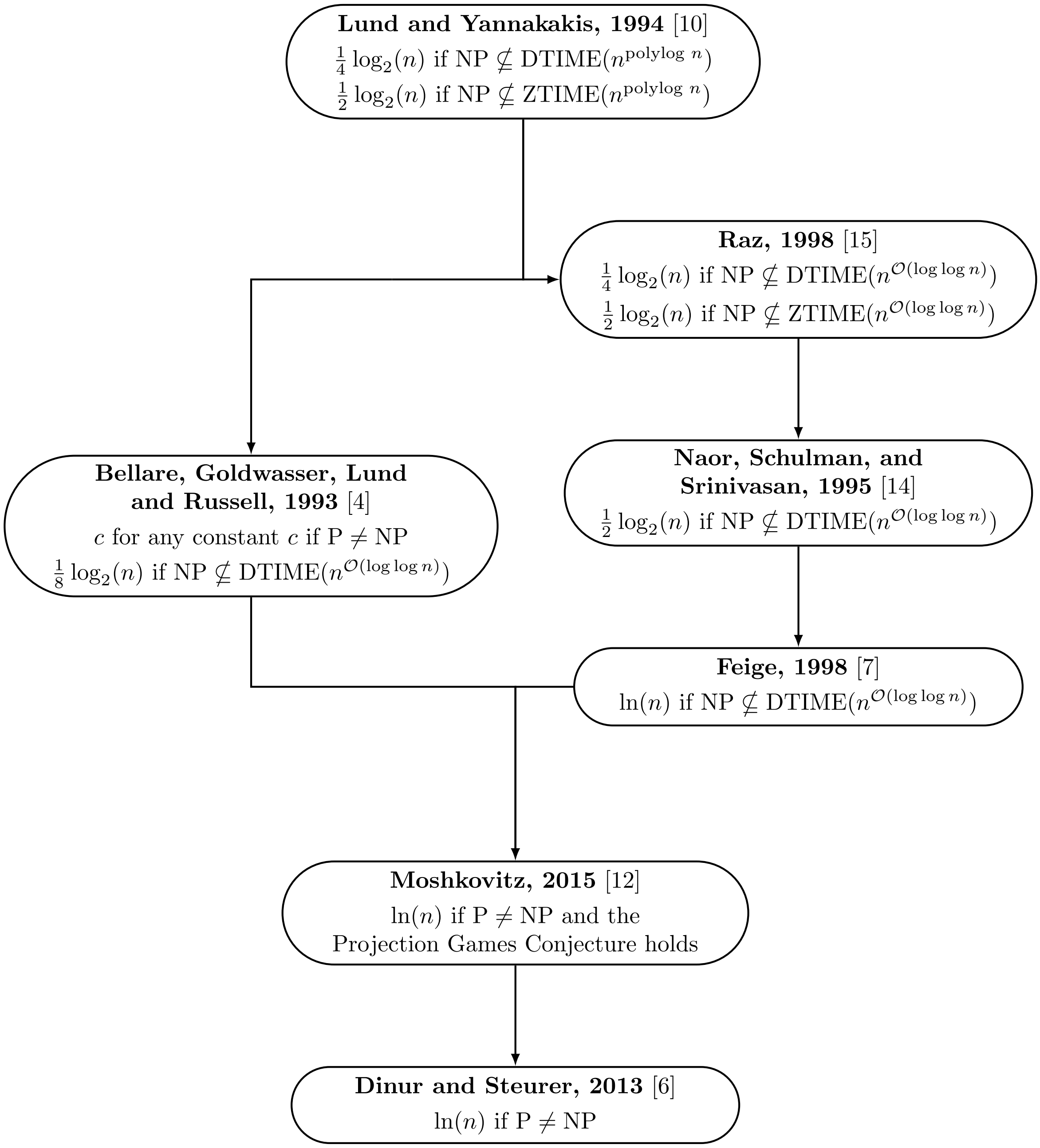
10 Acknowledgements

The author would like to thank William Gasarch for guidance on this project and proof-reading.

Appendix A: Chronological Timelines

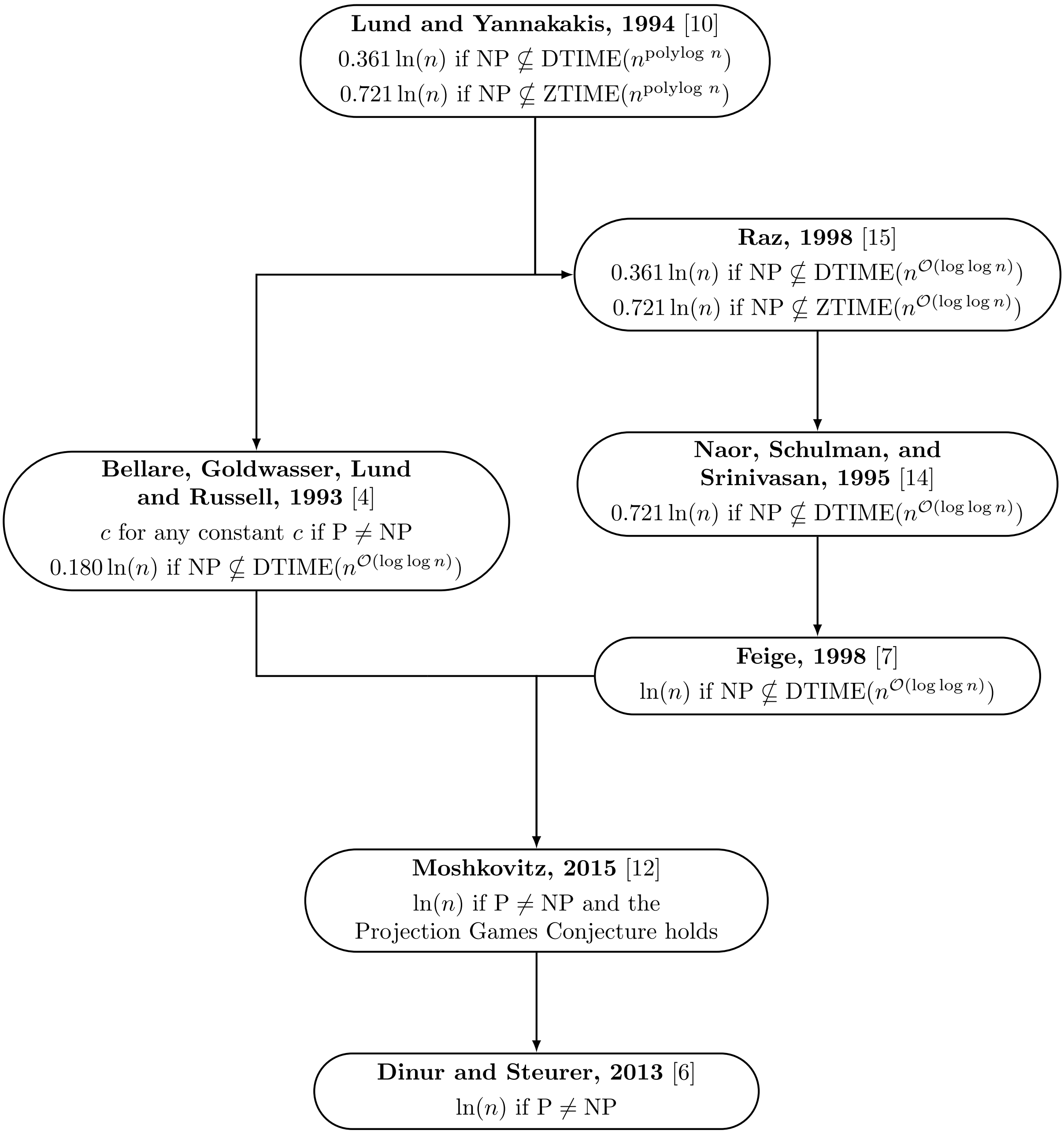
I Chronological Timeline with Results as Stated

The following timeline presents approximate order of *initial* results, with the years representing the final dates of publication of each paper. The values given are the new lower bounds on approximability of Set Cover; note that the main paper presents theorems largely in terms of *upper* bounds on *inapproximability*. See Section 1.2 for more.



II Chronological Timeline in Terms of $\ln(n)$

The following timeline again presents approximate order of *initial* results, with dates given being actual publication dates. Because logarithms differ from one another by only a constant factor, it is possible to present all results of the timeline in terms of $\ln(n)$. This timeline is presented as a way of clearly observing the improvement in results. Decimals are rounded to the nearest thousandth.



References

[1] Sanjeev Arora, Carsten Lund, Rajeev Motwani, Madhu Sudan, and Mario Szegedy. "Proof Verification and the Hardness of Approximation Problems". In: *Journal of the ACM* 45.3 (1998). https://doi.org/10.1145/278298.278306, pp. 501–555.

[2] Sanjeev Arora and Shmuel Safra. "Probabilistic Checking of Proofs: A New Characterization of NP". In: *Journal of the ACM* 45.1 (1998). https://doi.org/10.1145/273865.273901, pp. 70–122.

[3] László Babai and Shlomo Moran. "Arthur-Merlin Games: A Randomized Proof System, and a Hierarchy of Complexity Classes". In: *J. Comput. Syst. Sci.* 36 (1988), pp. 254–276.

[4] M. Bellare, S. Goldwasser, C. Lund, and A. Russell. "Efficient Probabilistically Checkable Proofs and Applications to Approximations". In: *Proceedings of the Twenty-Fifth Annual ACM Symposium on Theory of Computing*. STOC '93. https://doi.org/10.1145/167088.167174. San Diego, California, USA: Association for Computing Machinery, 1993, pp. 294–304. ISBN: 0897915917. DOI: 10.1145/167088.167174.

[5] V. Chvatal. "A greedy heuristic for the set-covering problem". In: *Mathematics of Operations Research* 4 (1979). https://www.jstor.org/stable/3689577, pp. 233–235.

[6] Irit Dinur and David Steurer. "Analytical approach to parallel repetition". In: *Proceedings of the 46th annual ACM Symposium on Theory of Computing*. https://dl.acm.org/doi/pdf/10.1145/2591796.2591884. 2013, pp. 624–633.

[7] Uriel Feige. "A Threshold of $\ln n$ for Approximating Set Cover". In: *J. ACM* 45.4 (July 1998). https://doi.org/10.1145/285055.285059, pp. 634–652. ISSN: 0004-5411. DOI: 10.1145/285055.285059.

[8] Uriel Feige and László Lovász. "Two-Prover One-Round Proof Systems: Their Power and Their Problems (Extended Abstract)". In: *Proceedings of the Twenty-Fourth Annual ACM Symposium on Theory of Computing*. STOC '92. https://doi.org/10.1145/129712.129783. Victoria, British Columbia, Canada: Association for Computing Machinery, 1992, pp. 733–744. ISBN: 0897915119. DOI: 10.1145/129712.129783.

[9] S. Goldwasser, S. Micali, and C. Rackoff. "The Knowledge Complexity of Interactive Proof Systems". In: *SIAM J. Comput.* 18.1 (Feb. 1989). https://doi.org/10.1137/0218012, pp. 186–208. ISSN: 0097-5397. DOI: 10.1137/0218012.

[10] Carsten Lund and Mihalis Yannakakis. "On the Hardness of Approximating Minimization Problems". In: *Journal of the ACM* 41.5 (1994). https://dl.acm.org/doi/pdf/10.1145/185675.306789, pp. 960–981.

[11] Erika Melder. "A Chronology of Set Cover Inapproximability Results". In: *CoRR* abs/2111.08100 (2021). https://arxiv.org/abs/2111.08100. arXiv: 2111.08100.

[12] Dana Moshkovitz. "The Projection Games Conjecture and the NP-Hardness of ln n-Approximating Set-Cover". In: *Theory of Computing* 11.7 (2015). http://www.theoryofcomputing.org/articles/v011a007, pp. 221–235. DOI: 10.4086/toc.2015.v011a007.

[13] Dana Moshkovitz and Ran Raz. "Two-Query PCP with Subconstant Error". In: *J. ACM* 57.5 (June 2008). https://doi.org/10.1145/1754399.1754402. ISSN: 0004-5411. DOI: 10.1145/1754399.1754402.

[14] M. Naor, L.J. Schulman, and A. Srinivasan. "Splitters and near-optimal derandomization". In: *Proceedings of IEEE 36th Annual Foundations of Computer Science.* 1995, pp. 182–191. DOI: 10.1109/SFCS.1995.492475.

[15] Ran Raz. "A Parallel Repetition Theorem". In: *SIAM Journal on Computing* 27.3 (1998). https://doi.org/10.1137/S0097539795280895, pp. 763–803. DOI: 10.1137/S0097539795280895. eprint: https://doi.org/10.1137/S0097539795280895.