Electronic Colloquium on Computational Complexity, Report No. 150 (2006)

# Towards Hardness of Envy-Free Pricing

Patrick Briest*

Dept. of Computer Science
Dortmund University
Otto-Hahn-Str. 14
Dortmund, Germany
patrick.briest@cs.uni-dortmund.de

**Abstract**

We consider the envy-free pricing problem, in which we want to compute revenue maximizing prices for a set of products $\mathcal{P}$ assuming that each consumer from a set of consumer samples $\mathcal{C}$ will buy the product maximizing her personal utility, i.e., the difference between her respective budget and the product's price. We show that assuming specific hardness of the balanced bipartite independent set problem in constant degree graphs or hardness of refuting random 3CNF formulas, the envy-free pricing problem cannot be approximated in polynomial time within $\mathcal{O}(\log^{\varepsilon}|\mathcal{C}|)$ for some $\varepsilon > 0$. This is the first result giving evidence that envy-free pricing might be hard to approximate within essentially better ratios than the logarithmic ratio obtained so far. Additionally, it gives another example of how average case complexity is connected to the worst case approximation complexity of notorious optimization problems.

## 1 Introduction

Inspired by the possibility of gathering large amounts of data about the preferences and budgets of a company's potential customers by web sites designed for this purpose, Rusmevichientong [15] and Glynn et al. [11] introduced a class of so called *multi-product pricing* problems that aim at computing optimal pricing schemes for a company's product range. In the original version of the problem each consumer is represented by a budget and a set of products she is interested in. Given fixed prices for the products, she decides to buy one of the products she is interested in with a price not exceeding her budget. The decision is made corresponding to either the *min-buying*, *max-buying*, or *rank-buying* model, where the consumer buys the product with lowest price not exceeding the budget, highest price not exceeding the budget, or highest rank according to some consumer specific ranking, respectively. All these problems are usually referred to as *unit-demand pricing*, since consumers will buy exactly one product if they can afford to do so.

Aggarwal et al. [1] extend the problem definition and allow consumers with different budgets for the different products they are interested in. Assuming that a *price ladder constraint*, i.e., a predefined order on the prices of all products, is given, they derive a polynomial time approximation scheme for the max-buying and rank-buying (under another reasonable assumption) models. They also show how to obtain logarithmic approximation ratios for all three models if no price ladder is given. Briest and Krysta [7] show that both of these algorithms are essentially best possible.

*The author is supported by DFG grant Kr 2332/1-2 within Emmy Noether program.



ISSN 1433-8092

Guruswami et al. [12] consider a different selection rule, which has already been proposed in [1]. In the *max-gain* model, a consumer buys the product maximizing her personal utility, i.e., the difference between the product's price and her respective budget. In the case of limited product supply, the definition in [12] requires that each consumer must obtain the product she desires most whenever she can afford any product at all. Thus, the resulting pricing scheme must be *envy-free* and we obtain the *envy-free pricing* problem.

Another problem introduced in [12] is so called *combinatorial* or *single-minded pricing*, which is inspired by single-minded combinatorial auction design. In this scenario each consumer has a single budget value and buys the whole set of products she is interested in if the sum of prices does not exceed her budget. Among other results, Guruswami et al. show that techniques similar to those of [1] yield a logarithmic approximation for this problem, which is proven to be close to best possible by Demaine et al. [8]. Balcan and Blum [3] and Briest and Krysta [6] give improved approximation results for a number of restricted versions of the problem.

## 2 New Results

The main open problem in the field is settling the approximation complexity of the envy-free and max-gain pricing problems, which have resisted all attempts so far. In the envy-free pricing problem a consumer $c$ is characterized by her budgets $b(c, e)$ for different products $e$. Given fixed prices $p(e)$, every consumer $c$ must be allocated the product maximizing $b(c, e) - p(e)$, i.e., consumer $c$'s utility from receiving the product at its specified price. Note, that the last condition (referred to as *eny-freeness*) is only an issue if we assume that product supply is limited. In this paper we consider the most restricted version of these problems, in which every consumer has only a single budget for all the products she desires and product supply is unlimited. Clearly, this problem can just as well be seen as a restriction of the unit-demand min-buying problem, since with uniform budgets it is always the product with lowest absolute price that maximizes a consumer's utility.

**Definition 1** *In the* Unit-Demand Pricing Problem *with* Min-Buying Model *(*UDP-MIN*), unlimited product supply and uniform budgets, we are given products $\mathcal{P}$ and consumer samples $\mathcal{C}$ consisting of budgets $b_c \in \mathbb{R}_0^+$ and product sets $S_c$ for all $c \in \mathcal{C}$. For a price assignment $p : \mathcal{P} \to \mathbb{R}_0^+$ we let $\mathcal{A}(p) = \{c \in \mathcal{C} \mid \exists e \in S_c : p(e) \leq b_c\}$ refer to the set of consumers that can afford to buy any product under $p$. We want to find prices $p$ that maximize*

$$\sum_{c \in \mathcal{A}(p)} \min\{p(e) \mid e \in S_c \wedge p(e) \leq b_c\}.$$

For the rest of this paper UDP-MIN refers to the problem defined above, i.e., we implicitly assume unlimited product supply and uniform budgets. As the main result of this paper, we show a reduction from the *Balanced Bipartite Independent Set Problem* (BBIS) in constant degree bipartite graphs to UDP-MIN. This shows that, assuming there are no randomized polynomial time algorithms of a certain kind approximating constant degree BBIS within arbitrarily small constant factors, there are no polynomial time algorithms approximating UDP-MIN within $\mathcal{O}(\log^{\varepsilon} |\mathcal{C}|)$ for some $\varepsilon > 0$.

Up to now, no explicit hardness results have been proven for BBIS in constant degree graphs, although the problem has been receiving a lot of attention. The first result for general BBIS using a quite moderate complexity theoretic assumption was obtained by Khot [13]. Previous results by Feige [9] and Feige and Kogan [10] are deriving hardness of BBIS under more specific assumptions. In [9], Feige shows an interesting connection between the average case complexity of refuting 3CNF-formulas and the worst case approximation complexity of several notorious optimization problems including BBIS. To put our result into a somewhat wider context, we formulate a slightly stronger version of the hypothesis in [9] and show that this is enough for our purposes.

Before stating the hypothesis we need to describe the random sampling procedure used to obtain random 3CNF formulas in [9]. Given $n$ variables we create formulas consisting of $m = \Delta n$ clauses for some large constant $\Delta \in \mathbb{N}$. Each literal of every clause is picked uniformly at random from the set of $2n$ literals. Thus, every clause consists of 3 (not necessarily different) literals that are picked independently at random. When $\Delta$ is large enough, every truth assignment satisfies roughly $(7/8)m$ clauses of a random 3CNF formula. Thus, a *typical* random 3CNF formula does not have more than $(7/8)m$ simultaneously satisfiable clauses. On the other hand, a formula with $(1-\varepsilon)m$ simultaneously satisfiable clauses can be considered *exceptional*. Hypothesis 1 states that it is hard to detect exceptional formulas on average.

**Hypothesis 1** *For every fixed $\varepsilon > 0$ and sufficiently large constant $\Delta \in \mathbb{N}$, there is no polynomial time (randomized) algorithm that, given a random 3CNF formula with $n$ variables and $m = \Delta n$ clauses, outputs* typical *with probability at least $1/2$, but outputs* exceptional *on every formula with $(1-\varepsilon)m$ simultaneously satisfiable clauses with probability at least $1 - 1/2^{poly(n)}$.*

The difference between Hypothesis 1 and the hypothesis in [9] is that we allow randomized algorithms that have exponentially small error probability when it comes to detecting exceptional formulas. We need this stronger version as a result of our reduction from BBIS to UDP-MIN, which is partially based on a random construction that introduces an exponentially small one-sided error probability for detecting large independent sets. In analogy to [9] we define a notion of hardness based on Hypothesis 1. We use slightly different notation compared to [9] to reflect the difference in the underlying hypotheses.

**Definition 2** *A problem is said to be R3SAT$^\star$-hard, if having a (randomized) polynomial time algorithm (with exponentially small failure probability) for it refutes Hypothesis 1.*

**Definition 3** *In the* Balanced Bipartite Independent Set Problem *(BBIS) we are given a bipartite graph $G = (V, W, E)$. We want to find maximum cardinality subsets of vertices $V' \subset V$, $W' \subset W$ with $|V'| = |W'|$, such that $\{v, w\} \notin E$ for all $v \in V'$, $w \in W'$.*

The first step in our proof is mostly identical to proofs given in [9], where hardness of general BBIS is derived. We do a slightly more careful analysis and obtain R3SAT$^\star$-hardness of BBIS in constant degree graphs. We point out that this part of the proof can be replaced by the following weaker hypothesis, which states that the gap variant of BBIS in constant degree graphs does not have randomized polynomial time algorithms with one-sided error. More formally, let $\mathcal{G}(a, d)$, $\mathcal{G}(b, d)$ be two families of bipartite graphs on $2n$ vertices with constant degree $d \in \mathbb{N}$ and maximum BBIS of size at most $an$ or at least $bn$, respectively. Given $0 < a < b < 1$ and $d \in \mathbb{N}$ the problem BBIS($a, b, d$) requires deciding whether $G \in \mathcal{G}(a, d)$ or $G \in \mathcal{G}(b, d)$ for a given graph $G$. For our purposes Hypothesis 2 is fully sufficient.

**Hypothesis 2** *There exist constants $0 < a < b < 1$ and $d \in \mathbb{N}$, such that BBIS($a, b, d$) $\notin$ RP.*

Having hardness of constant degree BBIS we apply the method of derandomized graph products [2] to obtain semi-logarithmic hardness of approximation for BBIS in graphs with roughly logarithmic maximum degree. The main part of the proof consists of the reduction to UDP-MIN. As an intermediate step in the reduction we modify the BBIS instance by adding a number of random edges and interpret vertices on one side of the bipartition as sets. The connection to UDP-MIN is made by considering sequences of these sets that have a certain expansion property. This is formalized in the following definition.

**Definition 4** *In the* Maximum Expanding Sequence Problem *(*MES*) we are given an ordered collection* $S_1, \ldots, S_m$ *of sets. An expanding sequence* $\phi = (\phi(1) < \cdots < \phi(\ell))$ *of length* $|\phi| = \ell$ *is a selection of sets* $S_{\phi(1)}, \ldots, S_{\phi(\ell)}$*, such that*

$$S_{\phi(j)} \nsubseteq \bigcup_{i=1}^{j-1} S_{\phi(i)}$$

*for* $2 \leq j \leq \ell$*.* MES *asks for finding such a sequence of maximum length.*

We are not aware that MES has been considered explicitly before. We briefly point out that a reduction similar to the one given in Section 3.3 yields hardness of approximation under a standard assumption.

**Theorem 1** MES *is hard to approximate within* $\mathcal{O}(m^{\varepsilon})$ *for some* $\varepsilon > 0$*, unless* $NP \subseteq \bigcap_{\varepsilon>0}$ *BPTIME*$(2^{n^{\varepsilon}})$*.*

It is relatively straightforward to encode MES in terms of UDP-MIN, since MES models precisely the dependence between different price levels in the pricing problem. The only difficulty lies in ensuring that the resulting UDP-MIN instances are of polynomial size. In fact, if we consider an extended version of the problem, we obtain hardness even under standard assumptions.

**Theorem 2** *Let* UDP-MIN$^{\star}$ *be the version of* UDP-MIN *in which each consumer* $c$ *gives an additional number* $m_c$ *defining the number of copies she will buy of the cheapest product she can afford.* UDP-MIN$^{\star}$ *is hard to approximate within* $\mathcal{O}(|\mathcal{C}|^{\varepsilon})$ *for some* $\varepsilon > 0$*, unless* $NP \subseteq \bigcap_{\varepsilon>0} BPTIME(2^{n^{\varepsilon}})$*.*

The formal proofs of Theorems 1 and 2 are ommitted from this manuscript. They are, however, completely analogous to the reductions presented in Sections 3.3 and 3.4. These are used in combination with the known hardness results for general BBIS from [13].

In order to reduce MES to UDP-MIN we have to restrict our attention to severely restricted problem instances. BBIS instances with semi-logarithmic maximum degree yield MES instances in which both the size of every set and the frequency of every element is at most semi-logarithmic. Ordering sets in an appropriate way we obtain an instance that consists of a logarithmic number of blocks, each of which contains only pairwise disjoint sets. This allows encoding the problem with a logarithmic number of different prices and, thus, a polynomial number of consumers.

**Theorem 3** *There exists* $\varepsilon > 0$*, such that it is R3SAT*$^{\star}$*-hard to approximate* UDP-MIN *with uniform budgets within* $\mathcal{O}(\log^{\varepsilon} |\mathcal{C}|)$*. Hardness of approximation holds even under the weaker assumption of Hypothesis 2.*

As UDP-MIN is just a special case of the envy-free pricing problem, we obtain identical hardness results for the general case.

**Corollary 1** *There exists* $\varepsilon > 0$*, such that it is R3SAT*$^{\star}$*-hard to approximate the envy-free pricing problem within* $\mathcal{O}(\log^{\varepsilon} |\mathcal{C}|)$*. Hardness of approximation holds even under the weaker assumption of Hypothesis 2.*

Section 3 contains the proof of Theorem 3. Section 3.1 shows the relation between refutation of random 3CNF-formulas and hardness of BBIS in constant degree graphs. Section 3.2 states how to amplify the hardness gap to the desired level using derandomized graph products. The reduction to UDP-MIN using the concept of expanding sequences is found in Sections 3.3 and 3.4.

# 3 Proof of Theorem 3

## 3.1 R3SAT$^\star$-hardness of Constant Degree BBIS

We show a reduction from MAX-3AND. Given a collection of clauses, each of which contains 3 (not necessarily distinct) literals and is satisfied if all 3 literals are assigned the boolean value true, we want to determine the maximum number of simultaneously satisfiable clauses. The remainder of this part of the proof is roughly identical to the one in [9], except for the fact that a small change in the reduction yields graphs of constant degree. Lemma 1 is explicitly stated in [9] for the case of their underlying hypothesis and extends easily to our notion of R3SAT$^\star$-hardness. We note that if we talk about random MAX-3AND instances, we assume the sampling procedure as described in Section 1.

**Lemma 1** *For every fixed $\varepsilon > 0$ and sufficiently large constant $\Delta \in \mathbb{N}$, the following problem is R3SAT$^\star$-hard. Given a random 3AND formula with $n$ variables and $m = \Delta n$ clauses, output* typical *with probability at least $1/2$, but output* exceptional *on every formula with $(1/4 - \varepsilon)m$ simultaneously satisfiable clauses.*

We want to show that if we have some good approximation algorithm for BBIS in constant degree graphs, then we can use it to design a refutation algorithm for MAX-3AND, which contradicts Hypothesis 1. Let us have a closer look at the random formulas we are given as an input. Clearly, in expectation each literal will appear $(3/2)\Delta$ times in the formula. Now let $V_i$ be a random variable counting the number of occurrences of literal $\ell_i$. Applying the Chernoff bound [14] we have that

$$\Pr\left[(1-\delta)\frac{3}{2}\Delta \le V_i \le (1+\delta)\frac{3}{2}\Delta\right] \ge 1 - 2\mathrm{e}^{-(3/4)\delta^2\Delta}$$

for any $0 < \delta < 1$. For every literal we define an additional random variable $X_i \in \{0,1\}$ that indicates whether the above condition is satisfied and let $X = X_1 + \cdots + X_{2n}$. By linearity of expectation it obviously holds that

$$\mathrm{E}\left[X\right] \ge \left(1 - 2\mathrm{e}^{-(3/4)\delta^2\Delta}\right)2n.$$

This implies that

$$\Pr\left[X < (1-\sqrt{2}\mathrm{e}^{-(3/8)\delta^2\Delta})2n\right] \le \sqrt{2}\mathrm{e}^{-(3/8)\delta^2\Delta},$$

since any higher probability would imply that

$$E[X] < \sqrt{2}\mathrm{e}^{-(3/8)\delta^2\Delta}(1-\sqrt{2}\mathrm{e}^{-(3/8)\delta^2\Delta})2n + (1-\sqrt{2}\mathrm{e}^{-(3/8)\delta^2\Delta})2n = (1-2\mathrm{e}^{-(3/4)\delta^2\Delta})2n,$$

a contradiction. Now fix any $\gamma > 0$ and observe that by choosing $\Delta$ sufficiently large we can ensure that $\sqrt{2}\mathrm{e}^{-(3/8)\delta^2\Delta} \le \gamma$.

**Fact 1** *With probability $1-\gamma$ a $(1-\gamma)$-fraction of the literals appear between $(1-\delta)\frac{3}{2}\Delta$ and $(1+\delta)\frac{3}{2}\Delta$ times in a random MAX-3AND formula.*

The first step of our refutation algorithm for MAX-3AND consists of checking the above condition. If too many literals deviate from their expected number of occurrences, the algorithm outputs *exceptional*. If this is not the case, we continue by removing the few problematic literals from the formula. More precisely, we remove every clause that contains a literal appearing more than $(1+\delta)(3/2)\Delta$ times.

Let $\mu = 3(\delta + \gamma)$. We know that $(1 - \gamma)2n$ *good* literals appear at least $(1 - \delta)(3/2)\Delta$ times within the formula. Thus, a total number of at least

$$(1 - \gamma)2n(1 - \delta)(3/2)\Delta \geq (1 - \delta - \gamma)3\Delta n$$

literal occurrences belong to *good* literals. This leaves at most $(\delta + \gamma)3\Delta n = \mu m$ literal occurrences belonging to *bad* literals and, consequently, gives an upper bound on the number of clauses that are removed from the formula. For the rest of the reduction to BBIS we need two more facts. (Fact 3 is explicitly proven in [9].)

**Fact 2** *If the original MAX-3AND formula had $(1/4-\varepsilon)m$ satisfiable clauses, then the number of satisfiable clauses in our modified formula is bounded below by $(1/4 - \varepsilon - \mu)m$*

**Fact 3** *For every $\varepsilon > 0$, sufficiently large $\Delta \in \mathbb{N}$ and $n$ large enough, the following holds. With high probability every set of $(1/8+\varepsilon)m$ clauses in a random MAX-3AND formula with $m = \Delta n$ clauses contains at least $n + 1$ different literals.*

We transform the modified formula into an instance of BBIS as follows. On both sides of the bipartition we have a vertex for every clause of the formula. Vertices on opposite sides are connected by an edge, if the corresponding clauses contain conflicting literals, i.e., if some variable appears in positive form in one clause and in negative form in the other. Thus, two vertices are connected if and only if the corresponding clauses cannot be satisfied simultaneously.

It is straightforward to argue that $(1/4-\varepsilon-\mu)m$ satisfiable clauses result in a maximum BBIS of at least the same size, while for random formulas the size of the maximum BBIS is bounded above by $(1/8 + \varepsilon)m$ with high probability. Additionally we know that, since every clause contains 3 literals and every literal appears at most $(1 + \delta)(3/2)\Delta$ times, the resulting bipartite graph has a maximum degree of at most $(1 + \delta)(9/2)\Delta$.

Assume now we had some polynomial time algorithm that can distinguish the two cases with an error probability exponentially close to 0. By applying this algorithm to the above BBIS instance we immediately obtain a polynomial time refutation algorithm for MAX-3AND with exponentially small failure probability for detecting exceptional formulas. If the BBIS algorithm returns a BBIS of size at least $(1/8 + \varepsilon)m$, we output *exceptional*. Otherwise, we output *typical*. The failure probability for detecting typical formulas is dominated by the probability that the formula has too many literals deviating from their expected number of occurrences and, thus, can be made an arbitrarily small constant.

**Lemma 2** *Let $\mathcal{G}(a, d)$, $\mathcal{G}(b, d)$ be the families of bipartite graphs on $2n$ vertices with maximum degree bounded by $d \in \mathbb{N}$ and a maximum balanced bipartite independent set of size at most $an$ or at least $bn$, respectively. There exist $0 < a < b < 1$ and $d \in \mathbb{N}$, such that deciding whether a given graph $G \in \mathcal{G}(a, d) \cup \mathcal{G}(b, d)$ belongs to $\mathcal{G}(a, d)$ or $\mathcal{G}(b, d)$ is R3SAT$^\star$-hard.*

### 3.2 Gap-Amplification for Bounded Degree BBIS

For a bipartite graph $G = (V, W, E)$, $|V| = |W| = n$, let $\alpha(G)$ refer to the size of a maximum balanced bipartite independent set in $G$. Let $\mathcal{G}(a, d)$ and $\mathcal{G}(b, d)$ be two families of bipartite graphs with maximum degree bounded by $d$ and $\alpha(G) \leq an$ for $G \in \mathcal{G}(a, d)$, $\alpha(G) \geq bn$ for $G \in \mathcal{G}(b, d)$. From the previous section we know that we can choose constants $a$, $b$ and $d$, such that deciding whether a given graph is from $\mathcal{G}(a, d)$ or $\mathcal{G}(b, d)$ is hard assuming Hypothesis 1 holds. The following definition is in analogy to [2, 4].

**Definition 5** *Let $G = (V, W, E)$, $|V| = |W| = n$, be a bipartite graph and $k \in \mathbb{N}$. The $k$*-fold graph product *$G^k = (V^k, W^k, E_k)$ is defined by cartesian products $V^k$, $W^k$ and $\{(v_1, \dots, v_k), (w_1, \dots, w_k)\} \in E_k$ if and only if $\{v_1, \dots, v_k, w_1, \dots, w_k\}$ is not a bipartite independent set in $G$.*

Berman and Schnitger [4] and Blum [5] consider so-called *randomized graph products*, which are obtained as the subgraph induced by a random sample of the vertices of $G^k$, to amplify hardness of the independent set problem in non-bipartite graphs. Alon et al. [2] show how this construction can be derandomized by replacing the sampling procedure of [4]. We briefly describe the application of their approach to bipartite graphs. Given $G = (V, W, E)$, $|V| = |W| = n$, we construct a non-bipartite $\delta$-regular Ramanujan graph $H$ on $n$ vertices and constant degree $\delta$ (depending only on constants $a$ and $b$). Vertices $V^k$ and $W^k$ of the *derandomized graph product $DG^k$* are obtained by choosing a vertex of $H$ uniformly at random and taking a random walk of length $k-1$ starting at this vertex. For $k = O(\log n)$ the number $n\delta^{k-1}$ of such random walks is polynomial and, thus, $DG^k$ can be constructed deterministically in polynomial time. The edges of $DG^k$ are defined as before. Now let $\delta A$ be the (symmetric) adjacency matrix of $H$, where $\lambda_0 \geq \lambda_1 \geq \cdots \geq \lambda_{n-1}$ are eigenvalues of matrix $A$, and let $\lambda = \max\{\lambda_1, |\lambda_{n-1}|\}$. The following is a slightly simplified version of Theorem 1 of [2], which extends easily to the BBIS problem.

**Theorem 4 ([2])** *For every balanced bipartite graph $G$ and any $k$ it holds that*

$$\alpha(G)\delta^{k-1}\left(\frac{\alpha(G)}{n} - \lambda\right)^{k-1} \leq \alpha(DG^k) \leq \alpha(G)\delta^{k-1}\left(\frac{\alpha(G)}{n} + \lambda\right)^{k-1}.$$

We next state a slightly extended version of Theorem 3 of [2]. We include the proof to demonstrate applicability in the context of BBIS and the possibility to obtain maximum degrees that are parameterized in the number of vertices of $DG^k$.

**Theorem 5** *Let $d(n) = \mathcal{O}(\sqrt{\log n})$, $\mathcal{G}(a(n), d(n))$ and $\mathcal{G}(b(n), d(n))$ the families of balanced bipartite graphs on $2n$ vertices, maximum degree bounded by $d(n)$ and maximum BBIS of size at most $a(n) \cdot n$ or at least $b(n) \cdot n$, respectively. There exist $0 < a(n) < b(n) < 1$ with $b(n)/a(n) = \Omega(\log^{\varepsilon} n)$ for some $\varepsilon > 0$, such that given $G \in \mathcal{G}(a(n), d(n)) \cup \mathcal{G}(b(n), d(n))$ it is R3SAT$^\star$-hard to decide whether $G \in \mathcal{G}(a(n), d(n))$ or $G \in \mathcal{G}(b(n), d(n))$.*

*Proof:* Let $\mathcal{G}(a, d)$ and $\mathcal{G}(b, d)$ be defined as above and let $G \in \mathcal{G}(a, d) \cup \mathcal{G}(b, d)$, $G = (V, W, E)$, $|V| = |W| = n$. Choosing $0 < a < b < 1$ appropriately it is R3SAT$^\star$-hard to decide whether $G \in \mathcal{G}(a, d)$ or $G \in \mathcal{G}(b, d)$ by Lemma 2. We now consider the $k$-fold derandomized graph product $DG^k = (DV, DW, DE)$.

In $DG^k$ an edge $\{(v_1, \dots, v_k), (w_1, \dots, w_k)\}$ exists only if there are indices $i$ and $j$, such that $\{v_i, w_j\} \in E$. We fix $(v_1, \dots, v_k)$ and count the maximum number of adjacent vertices. There are $k^2$ possibilities to select $i$ and $j$. Fixing indices fixes $v_i$ as well and, by the fact that $G$ has maximum degree $d$, there are at most $d$ possible choices for $w_j$. Finally, there remain $\delta^{k-1}$ possibilities to choose the random walk generating $(w_1, \dots, w_k)$. Thus, $DG^k$ has maximum degree $\Delta \leq dk^2\delta^{k-1}$.

For $\delta$-regular Ramanujan graphs it is known that $\lambda \approx 2\sqrt{\delta - 1}/\delta$. By choosing the constant degree $\delta \geq 2$ of $H$ sufficiently large we have that

$$\lambda < \frac{2}{\sqrt{\delta}} \leq \frac{1}{3}(b - a).$$

By Theorem 4 the gap between the cases that $G \in \mathcal{G}(a, d)$ and $G \in \mathcal{G}(b, d)$ is then amplified to

$$\frac{bn\delta^{k-1}(b-\lambda)^{k-1}}{an\delta^{k-1}(a+\lambda)^{k-1}} \geq \left(\frac{b-\lambda}{a+\lambda}\right)^k > (1+\lambda)^k.$$

Using the fact that $\delta \approx 4/\lambda^2$ and choosing a constant $\gamma$, such that $(4/\lambda^2)^\gamma \approx (1+\lambda)$, we obtain that $(1+\lambda)^k \geq \delta^{\gamma k}$.

Given $G \in \mathcal{G}(a, d) \cup \mathcal{G}(b, d)$, $G = (V, W, E)$ and $|V| = |W| = n$, we choose (for the remainder of the proof $\log$ is to the base of $\delta$) $k = (1/2 - \varepsilon) \log\log n$. Thus, the number of vertices $N$ on one side of the bipartition of $DG^k$ is lower bounded by

$$N \geq n \cdot \delta^{(1/2-\varepsilon)\log\log n} = \Omega(n).$$

The maximum degree $\Delta$ of $DG^k$ is upper bounded by

$$d((1/2)\log\log n)^2 \delta^{(1/2-\varepsilon)\log\log n} = \mathcal{O}(\sqrt{\log n}).$$

Thus, we have that $\Delta = O(\sqrt{\log N})$ as desired. The gap between the cases $G \in \mathcal{G}_a$ and $G \in \mathcal{G}_b$ is amplified to

$$\delta^{\gamma k} = \log^{(1/2-\varepsilon)\gamma} n = \Omega\left(\log^{(1/2-\varepsilon)\gamma} N\right),$$

where we use that $\log N = \mathcal{O}(\log n)$. Clearly, it is not difficult to calculate the precise values of $a(n)$ and $b(n)$ as needed to prove our claim. □

We want to remark that by construction the constant degree graphs obtained by the reduction in Section 3.1 are symmetric in the sense that we can rename vertices $V = \{v_1, \ldots, v_n\}$ and $W = \{w_1, \ldots, w_n\}$, such that $\{v_i, w_j\} \in E$ if and only if $\{v_j, w_i\} \in E$. This property is not lost during gap amplification, since we can use the same expander graph to obtain the vertices on both sides of the graph product.

### 3.3 Maximum Expanding Sequences

Let $G \in \mathcal{G}(a(n), d(n)) \cup \mathcal{G}(b(n), d(n))$, $G = (V, W, E)$,$|V| = |W| = n$, with $a(n)$, $b(n)$ and $d(n)$ as in Theorem 5 be given. We will reduce the problem of deciding whether $G \in \mathcal{G}(a(n), d(n))$ or $G \in \mathcal{G}(b(n), d(n))$ to solving a restricted instance of MES. We start by adding a couple of random edges to the graph. More precisely, every possible edge is added to $G$ with probability $(b(n)n)^{-1}$. We do not allow multiple edges and, thus, edges that have already been present in $G$ will not be duplicated.

Afterwards we remove vertices with too high degree from the graph. In expectation the random experiment tries to add $b(n)^{-1}$ new edges to every vertex $v \in V \cup W$. We remove a vertex $v$ if more than $c \cdot b(n)^{-1}$ edges are added to it, where $c$ is some sufficiently large constant to be determined later. Let $N_v$ be the random variable counting the number edges added to $v$ and $R_v \in \{0, 1\}$ a random variable indicating whether $v$ is removed after the random experiment. Applying the Chernoff bound we obtain

$$\Pr[R_v = 1] = \Pr[N_v \geq c \cdot b(n)^{-1}] \leq \mathrm{e}^{-(c-1)/(3b(n))}$$

for any constant $c \geq 2$. We denote the modified graph by $G' = (V', W', E')$. For every vertex $v_i \in V'$ we define a corresponding set $S_i$ by

$$S_i = \left\{w_j \in W' \,|\, \{v_i, w_j\} \in E'\right\},$$

i.e., vertices $V'$ correspond to sets over universe $W'$ in our MES instance. In order to obtain a feasible MES instance we need to define an order on sets $S_i$, which we do next. Observe that vertices in $G'$ have degree at most

$$d'(n) \leq d(n) + c \cdot b(n)^{-1} = \mathcal{O}(\sqrt{\log n}),$$

where we use the fact that bipartite graphs with bounded degree $d(n)$ have a BBIS of size at least $n/(d(n)+1)$ and, thus, it must be the case that $b(n)^{-1} = \mathcal{O}(d(n))$. Furthermore, if the maximum degree of $G'$ is $d'(n)$, then the sets $S_i$ can be partitioned into $d'(n)^2$ many classes, such that sets in each class do not intersect. To see this, note, that every set contains at most $d'(n)$ elements, each of which is contained in at most $d'(n) - 1$ further sets. Thus, starting with $d'(n)^2$ empty classes and adding sets one by one, the number of classes to which a specific set cannot be added is always bounded above by $d'(n)(d'(n) - 1)$.

Let $\mathcal{C}_1, \ldots, \mathcal{C}_{q(n)}$ denote the classes of sets obtained in this way and observe that $q(n) = \mathcal{O}(\log n)$. We reorder sets according to the classes and finally obtain an MES instance $S_1, \ldots, S_m$ for which it holds that $\mathcal{C}_j = \{S_{k(j)}, S_{k(j)+1} \ldots S_{\ell(j)}\}$, i.e., sets belonging to a single class form a non-interrupted block in the ordering. This property is not required for the remainder of this section, but will be of immense importance for the reduction to UDP-MIN in Section 3.4.

**Definition 6** *We say that an* MES *instance $S_1, \ldots, S_m$ is* log-separable *if it can be separated into $q(m) = \mathcal{O}(\log m)$ subsets $\mathcal{C}_1, \ldots, \mathcal{C}_{q(m)}$, such that $\mathcal{C}_j = \{S_{k(j)}, S_{k(j)+1} \ldots S_{\ell(j)}\}$, where $k(1) = 1$, $\ell(q(m)) = m$, $k(j+1) = \ell(j) + 1$ for $1 \leq j \leq q - 1$ and each $\mathcal{C}_j$ contains only non-intersecting sets.*

**Soundness:** Let $G \in \mathcal{G}(b(n), d(n))$. Assume for the moment that no vertices are removed from $G$ and w.l.o.g. let $\mathcal{S}^* = \{S_1, \ldots, S_{b(n)n}\}$ be the sets in the MES instance corresponding to vertices from $V$ that belong to a maximum BBIS. Let $W^* \subset W$ denote the vertices from $W$ belonging to the BBIS. Finally, let $\mathcal{S}^*_r$ and $W^*_r$ denote the subsets of $\mathcal{S}^*$ and $W^*$ that are not removed from the instance due to the degree constraint.

We construct an expanding sequence by considering sets $S_1, \ldots, S_{b(n)n}$ one by one. Set $S_j$ is selected for the sequence if it has not been removed, i.e., if $S_j \in \mathcal{S}^*_r$, and contains exactly one element from $W^*$, which has not been covered by previously selected sets and is not removed from the instance. For $1 \leq j \leq b(n)n/2$ we estimate the probability that $S_j$ is not selected using the union bound. We have already seen that the probability of $S_j$ or any specific $v \in W^*$ being removed is bounded above by $\mathrm{e}^{-(c-1)/(3b(n))}$. Every selected set contains exactly 1 element from $W^*$. Thus, for $j \leq b(n)n/2$ there are at least $b(n)n/2$ uncovered elements remaining when we try to select $S_j$. The probability that $S_j$ contains exactly one previously uncovered element from $W^*$ is therefore at least

$$\sum_{w \text{ uncovered}} \frac{1}{b(n)n} \left(1 - \frac{1}{b(n)n}\right)^{b(n)n} \geq \frac{b(n)n}{2} \frac{1}{\mathrm{e}b(n)n} = \frac{1}{2\mathrm{e}}$$

and, thus, the failure probability is bounded above by $1 - 1/2\mathrm{e}$. Applying the union bound we obtain that

$$\Pr\left[S_j \text{ is selected}\right] \geq \frac{1}{2\mathrm{e}} - 2\mathrm{e}^{-(c-1)/(3b(n))} \approx \frac{1}{2\mathrm{e}}$$

for sufficiently large constant $c$. Let $Y$ denote the number of selected sets. It clearly holds that $\mathrm{E}[Y] \geq (1/4\mathrm{e})b(n)n$ and applying the same argumentation as in Section 3.1 we get that $\Pr[Y \leq (1/(8\mathrm{e}))b(n)n] \leq 1 - (1/(8\mathrm{e}))$, since otherwise it had to be true that

$$\mathrm{E}[Y] < \left(1 - \frac{1}{8\mathrm{e}}\right) \frac{1}{8\mathrm{e}} b(n)n + \frac{1}{8\mathrm{e}} b(n)n \leq \frac{1}{4\mathrm{e}} b(n)n,$$

a contradiction. This implies that with probability $\Omega(1)$ there exists an expanding sequence of length $\Omega(b(n)n)$.

**Completeness:** Let $G \in \mathcal{G}(a(n), d(n))$ and consider any expanding sequence $\phi$ in $S_1, \ldots, S_m$. Since the maximum BBIS in $G$ is of size $a(n)n$, every selection of $a(n)n + 1$ vertices from $V$ must be adjacent to all but $a(n)n$ vertices from $W$. Thus, the first $a(n)n + 1$ sets from $\phi$ leave at most $a(n)n$ elements uncovered and it follows that $|\phi| \leq 2a(n)n + 1$.

We have shown a randomized reduction with constant one-sided error probability. By repeating the algorithm a polynomial number of times, we obtain error probabilities that are exponentially close to $0$. This yields the following result.

**Lemma 3** *The restricted version of* MES *with* $\log$*-separable instances is R3SAT*$^\star$*-hard to approximate within* $\mathcal{O}(\log^\varepsilon m)$ *for some constant* $\varepsilon > 0$.

### 3.4 Reduction to UDP-MIN

The final step in the proof of Theorem 3 consists of reducing $\log$-separable MES to UDP-MIN with uniform budgets. Let MES instance $S_1, \ldots, S_m$ be separable into $\mathcal{C}_1, \ldots, \mathcal{C}_{q(m)}$ with $q(m) = \mathcal{O}(\log m)$.

For each element $e$ in the universe of the MES instance we have a corresponding product, to which we will also refer as $e$. For every set $S_i$ in class $\mathcal{C}_k$ we define a collection of $2^{k-1}$ identical consumers $C_i = \{c_i^1, c_i^2, \ldots\}$. Each of these consumers has budget $b_i = 2^{1-k}$ and is interested in products from set $S_i$.

**Soundness:** Let $\phi = (\phi(1) < \cdots < \phi(\ell))$ be an expanding sequence of length $\ell$. For every $1 \leq i \leq \ell$ let $N_i$ denote the elements that are newly covered by $S_i$. Now we repeat the following for $i = 1, \ldots, \ell$. Determine $N_i$, then set the prices of all elements $e \in N_i$ to $b_i$. As a result, all consumers belonging to a set $S_i$ in the expanding sequence will buy at their budget values. Thus, the overall profit from the price assignment is at least $\ell$.

**Completeness:** Assume that we are given a price assignment resulting in overall revenue $r$. First observe that w.l.o.g. all prices are from the set of distinct budget values, i.e., all prices are powers of 2. Then note that w.l.o.g. revenue at least $r/2$ is due to consumers buying at their budget values, since otherwise we could increase overall revenue by multiplying all prices by 2. Finally, it's not difficult to see that consumers buying at their budget values form an expanding sequence. It follows that we obtain an expanding sequence $\phi$ of length at least $r/2$.

This finishes the proof of Theorem 3.

## 4 Conclusions and Open Problems

We have shown that assuming specific hardness of constant degree BBIS or hardness on average of refuting random 3CNF-formulas, the envy-free pricing problem does not allow approximation guarantees essentially beyond the known results. This leaves open the question whether one can prove inapproximability under some standard complexity theoretic assumption. Our current results suggest a number of ways to approach this task.

1. Prove hardness of approximation for constant degree BBIS. Applying our reduction this immediately yields hardness of envy-free pricing.

2. We believe that the MES problem essentially characterizes one of the main difficulties of envy-free pricing, namely the interaction between different price levels and the way that combinations of consumers block each other. Sticking to this approach, maybe it is possible to prove hardness of approximation of (very restricted) MES without using BBIS.

3. Although MES captures one of the difficulties of pricing, there are others, as well. Our reduction shows hardness of the most restricted version of envy-free pricing. Considering also limited product supply, the problem becomes even more involved because of the envy-freeness condition. Maybe exploiting this feature of the problem one might find another proof of hardness of approximation.

On the other hand, one may try to refute Hypotheses 1 or 2. Especially settling the approximation complexity of BBIS in constant degree graphs either way is an important open problem.

## References


[1] G. Aggarwal, T. Feder, R. Motwani, and A. Zhu. Algorithms for Multi-Product Pricing. In *Proc. of 31st ICALP*, 2004.

[2] N. Alon, U. Feige, A. Wigderson, and D. Zuckerman. Derandomized Graph Products. *Computational Complexity*, **5**(1): 60–75, 1995.

[3] N. Balcan and A. Blum. Approximation Algorithms and Online Mechanisms for Item Pricing. In *Proc. of 7th EC*, 2006.

[4] P. Berman and G. Schnitger. On the Complexity of Approximating the Independent Set Problem. *Information and Computation*, **96**(1): 77–94, 1992.

[5] A. Blum. *Algorithms for Approximate Graph Coloring.* Ph.D. thesis, MIT Laboratory for Computer Science MIT/LCS/TR-506, May 1991.

[6] P. Briest and P. Krysta. Single-Minded Unlimited-Supply Pricing on Sparse Instances. In *Proc. of 17th SODA*, 2006.

[7] P. Briest and P. Krysta. Buying Cheap is Expensive: Hardness of Non-Parametric Multi-Product Pricing. To appear in *Proc. of 18th SODA*, 2007.

[8] E.D. Demaine, U. Feige, M.T. Hajiaghayi, and M.R. Salavatipour. Combination Can Be Hard: Approximability of the Unique Coverage Problem. In *Proc. of 17th SODA*, 2006.

[9] U. Feige. Relations between Average Case Complexity and Approximation Complexity. In *Proc. of 34th STOC*, 2002.

[10] U. Feige and S. Kogan. Hardness of Approximation of the Balanced Complete Bipartite Subgraph Problem. *Technical Report MCS04-04*, Dept. of Computer Science and Applied Math., The Weizmann Institute of Science, 2004.

[11] P. Glynn, P. Rusmevichientong, and B. Van Roy. A Non-Parametric Approach to Multi-Product Pricing. *Operations Research*, **54**(1):82–98, 2006.

[12] V. Guruswami, J.D. Hartline, A.R. Karlin, D. Kempe, C. Kenyon, and F. McSherry. On Profit-Maximizing Envy-Free Pricing. In *Proc. of 16th SODA*, 2005.

[13] S. Khot. Ruling out PTAS for Graph Min-Bisection, Densest Subgraph and Bipartite Clique. In *Proc. of 45th FOCS*, 2004.

[14] R. Motwani and P. Raghavan. *Randomized Algorithms.* Cambridge University Press, 1995.

[15] P. Rusmevichientong. *A Non-Parametric Approach to Multi-Product Pricing: Theory and Application.* Ph.D. thesis, 2003.



ECCC ISSN 1433-8092

http://eccc.hpi-web.de/