2013 IEEE 54th Annual Symposium on Foundations of Computer Science

Independent Set, Induced Matching, and Pricing: Connections and Tight (Subexponential Time) Approximation Hardnesses

Parinya Chalermsook
Max-Planck-Institut für Informatik
Germany

Bundit Laekhanukit
McGill University
Canada

Danupon Nanongkai
Nanyang Technological University
Singapore

***Abstract*—We present a series of almost settled inapproximability results for three fundamental problems. The first in our series is the *subexponential-time* inapproximability of the *independent set* problem, a question studied in the area of *parameterized complexity*. The second is the hardness of approximating the *bipartite induced matching* problem on bounded-degree bipartite graphs. The last in our series is the tight hardness of approximating the k-*hypergraph pricing* problem, a fundamental problem arising from the area of *algorithmic game theory*. In particular, assuming the Exponential Time Hypothesis, our two main results are:**

• For any r larger than some constant, any r-approximation algorithm for the independent set problem must run in at least $2^{n^{1-\epsilon}/r^{1+\epsilon}}$ time. This nearly matches the upper bound of $2^{n/r}$ [23]. It also improves some hardness results in the domain of parameterized complexity (e.g., [26], [19]).

• For any k larger than some constant, there is no polynomial time $\min\left\{k^{1-\epsilon}, n^{1/2-\epsilon}\right\}$-approximation algorithm for the k-hypergraph pricing problem , where n is the number of vertices in an input graph. This almost matches the upper bound of $\min\left\{O(k), \tilde{O}(\sqrt{n})\right\}$ (by Balcan and Blum [3] and an algorithm in this paper).

We note an interesting fact that, in contrast to $n^{1/2-\epsilon}$ hardness for polynomial-time algorithms, the k-hypergraph pricing problem admits n^{δ} approximation for any $\delta > 0$ in quasi-polynomial time. This puts this problem in a rare approximability class in which approximability thresholds can be improved significantly by allowing algorithms to run in quasi-polynomial time.

The proofs of our hardness results rely on unexpectedly tight connections between the three problems. First, we establish a connection between the first and second problems by proving a new graph-theoretic property related to an *induced matching number of dispersers*. Then, we show that the $n^{1/2-\epsilon}$ hardness of the last problem follows from nearly tight *subexponential time* inapproximability of the first problem, illustrating a rare application of the second type of inapproximability result to the first one. Finally, to prove the subexponential-time inapproximability of the first problem, we construct a new PCP with several properties; it is sparse and has nearly-linear size, large degree, and small free-bit complexity. Our PCP requires no ground-breaking ideas but rather a very careful assembly of the existing ingredients in the PCP literature.

***Index Terms*—Approximation Algorithms; Subexponential-Time Algorithms; Algorithmic Pricing**

The full version of this paper is available as [16] at http://arxiv.org/abs/1308.2617. P. Chalermsook partially did this work while he was at IDSIA, Lugano, Switzerland, and was supported by the Swiss National Science Foundation project 200020_144491/1. B. Laekhanukit is supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) grant no. 429598 and by Dr&Mrs M.Leong fellowship. D. Nanongkai is supported in part by the following research grants: Nanyang Technological University grant M58110000, Singapore Ministry of Education (MOE) Academic Research Fund (AcRF) Tier 2 grant MOE2010-T2-2-082, and Singapore MOE AcRF Tier 1 grant MOE2012-T1-001-094.

I. Introduction

This paper presents results of two kinds, lying in the intersections between approximation algorithms and other subareas of theoretical computer science. The first kind of our results is a tight hardness of approximating the k-hypergraph pricing problem in polynomial time. This problem arose from the area of algorithmic game theory, and its several variants have recently received attentions from many researchers (see, e.g., [43], [44], [30], [10], [3], [9], [41], [13]). It has, however, resisted previous attempts to improve approximation ratio given by simple algorithms. Indeed, no sophisticated algorithmic techniques have been useful in attacking the problem in its general form. The original motivation of this paper is to show that those simple algorithms are, in fact, the best one can do under a reasonable complexity theoretic assumption. In showing this, we devise a new reduction from another problem studied in discrete mathematics and networking called the *bipartite induced matching* problem. Our reduction, unfortunately, blows up the instance size *exponentially*, and apparently this blowup is unavoidable (this claim will be discussed precisely later). Due to the exponential blowup of our reduction, showing a *tight* polynomial-time hardness of approximating the bipartite induced matching problem is not enough for settling the complexity of the pricing problem. What we need is, roughly speaking, the hardness of approximation result that is tight even for subexponential time approximation algorithms, i.e., proving the lower bound on the approximation ratio that any subexponential time algorithms can achieve.

This motivates us to prove the second type of results: *hardness of subexponential-time approximation*. The subject of subexponential time approximation and the closely related subject of *fixed-parameter tractable (FPT) approximation* have been recently studied in the area of parameterized complexity (e.g., [23], [29], [26], [19]). Our main result of this type is a

0272-5428/13 $26.00 © 2013 IEEE
DOI 10.1109/FOCS.2013.47

sharp trade-off between the running time and approximation ratio for the bipartite induced matching problem, and since our proof crucially relies on the hardness construction for the independent set problem, we obtain a sharp trade-off for approximating the independent set problem as a by-product. The independent set problem is among fundamental problems studied in both approximation algorithms and FPT literature (since it is W[1]-hard), and it is of interest to figure out its subexponential-time approximability. Our trade-off result immediately answers this question, improves previous results in [26], [19] and nearly matches the upper bound in [23].

The main contributions of this paper are the *nearly tight connections* among the aforementioned problems (they are tight in the sense that any further improvements would immediately refute the Exponential Time Hypothesis (ETH)), which essentially imply the nearly tight (subexponential time) hardness of approximation for all of them. Interestingly, our results also illustrate a rare application of the subexponential-time inapproximability to the inapproximability of polynomial-time algorithms. The key ideas of our hardness proofs are simple and algorithmic even though it requires a non-trivial amount of work to actually implement them.

Finally, we found a rather bizarre phenomenon of the k-hypergraph pricing problem (when k is large) in the quasi-polynomial time regime. While both induced matching, independent set and many other natural combinatorial optimization problems do not admit much better approximation ratios in quasi-polynomial time (e.g., $n^{1-\epsilon}$ hardness of approximating the independent set and bipartite induced matching problem still hold against quasi-polynomial time algorithms), the story is completely different for the pricing problem: That is, the pricing problem admits n^{δ} approximation in quasi-polynomial time for any $\delta > 0$, even though it is $n^{1/2-\epsilon}$ hard against polynomial-time approximation algorithms. This contrast puts the pricing problem in a rare approximability class in which polynomial time and quasi-polynomial time algorithms' performances are significantly different.

k-Hypergraph Pricing: In the *unlimited supply k-hypergraph vertex pricing* problem [3], [10], we are given a weighted n-vertex m-edge k-hypergraph (each hyperedge contains at most k vertices) modeling the situation where consumers (represented by hyperedges) with budgets (represented by weights of hyperedges) have their eyes on at most k products (represented by vertices). The goal is to find a price assignment that maximizes the revenue. In particular, there are two variants of this problem with different consumers' buying rules. In the *unit-demand pricing problem* (UDP), we assume that each consumer (represented by a hyperedge e) will buy the cheapest vertex of her interest if she can afford it. In particular, for a given hypergraph H with edge weight $w : E(H) \rightarrow \mathbb{R}_{\geq 0}$ (where $\mathbb{R}_{\geq 0}$ is the set of non-negative reals), our goal is to find a price function $p : V(H) \rightarrow \mathbb{R}_{\geq 0}$ to maximize $\text{profit}_{H,w}(p) = \sum_{e \in E(H)} \text{pay}_e(p)$ where $\text{pay}_e(p) = \min_{v \in e} p(v)$ if $\min_{v \in e} p(v) \leq w(e)$ and 0 otherwise. The other variation is the *single-minded pricing problem* (SMP), where we assume that each consumer will buy *all* vertices if she can afford to; otherwise, she will buy nothing. Thus, the goal is to maximize $\text{profit}_{H,w}(p) = \sum_{e \in E(H)} \text{pay}_e(p)$ where $\text{pay}_e(p) = \sum_{v \in e} p(v)$ if $\sum_{v \in e} p(v) \leq w(e)$ and 0 otherwise.

The pricing problem naturally arose in the area of algorithmic game theory and has important connections to algorithmic mechanism design (e.g., [4], [18]). Its general version (where k could be anything) was introduced by Rusmevichientong et al. [43], [44], and the k-hypergraph version (where k is thought of as some constant) was first considered by Balcan and Blum [3]. (The special case of $k = 2$ has also received a lot of attention [3], [14], [35], [41].) There will be two parameters of interest to us, i.e., n and k. Its current approximation upper bound is $O(k)$ [10], [3] while its lower bound is $\Omega(\min(k^{1/2-\epsilon}, n^{\epsilon}))$ [9], [13], [15].

Bipartite Induced Matching: Informally, an induced matching of an undirected unweighted graph G is a matching M of G such that no two edges in M are joined by an edge in G. To be precise, let $G = (V, E)$ be any undirected unweighted graph. An *induced matching* of G is the set of edges $\mathcal{M} \subseteq E(G)$ such that $\mathcal{M}$ is a matching and for any distinct edges $uu', vv' \in \mathcal{M}$, G has none of the edges in $\{uv, uv', u'v, u'v'\}$. The *induced matching number* of G, denoted by $\mathsf{im}(G)$, is the cardinality of the maximum-cardinality induced matching of G. Our goal is to compute $\mathsf{im}(G)$ of a bipartite graph G.

The notion of induced matching has naturally arisen in discrete mathematics and computer science. It is, for example, studied as the "risk-free" marriage problem in [47] and is a subtask of finding a *strong edge coloring*. This problem and its variations also have connections to various problems such as maximum feasible subsystem [25], [15], maximum expanding sequence [11], storylines extraction [36] and network scheduling, gathering and testing (e.g., [27], [47], [34], [39], [7]). In general graphs, the problem was shown to be NP-complete in [47], [12] and was later shown in [20] to be hard to approximate to within a factor of $n^{1-\epsilon}$ and $d^{1-\epsilon}$ unless $\mathsf{P} = \mathsf{NP}$, where n is the number of vertices and d is the maximum degree of a graph. In bipartite graphs, assuming $\mathsf{P} \neq \mathsf{NP}$, the induced matching problem was shown to be $n^{1/3-\epsilon}$-hard to approximate in [25]. Recently, we [15] showed its tight hardness of $n^{1-\epsilon}$ (assuming $\mathsf{P} \neq \mathsf{NP}$) and a hardness of $d^{1/2-\epsilon}$ on d-degree-bounded bipartite graphs. This hardness leads to tight hardness of several other problems. In this paper, we improve the previous hardness to a tight $d^{1-\epsilon}$ hardness, as well as extending it to a tight approximability/running time trade-off for subexponential time algorithms.

Independent Set: Given a graph $G = (V, E)$, a set of vertices $S \subseteq V$ is *independent* (or *stable*) in G if and only if G has no edge joining any pair of vertices $x, y \in S$. In the independent set problem, we are given an undirected graph $G = (V, E)$, and the goal is to find an independent set S of G with maximum size. Hardness results for the independent set problem heavily rely on developments in the PCP literature. The connection between the independent set problem and the *probabilistic checkable proof system* (PCP) was first discovered by Feige et al. [28] who showed that the independent set problem is hard to approximate to within a factor of $2^{\log^{1-\epsilon} n}$, for any $\epsilon > 0$, unless $\mathsf{NP} \subseteq \mathsf{DTIME}(n^{\mathrm{polylog}(n)})$. The inapproximability result has been improved by Arora and Safra [2] and Arora et al. [1], leading to a polynomial hardness of the problem [1]. Later, Bellare and Sudan [6] introduced the notion of the *free-bit complexity* of a PCP and showed that, given a PCP with logarithmic randomness and free-bit complexity f, the independent set problem is hard to approximate to within a factor of $n^{1/(1+f)-\epsilon}$, for all $\epsilon > 0$, unless $\mathsf{NP} = \mathsf{ZPP}$. There, Bellare and Sudan [6] constructed a PCP with free-bit complexity $f = 3 + \delta$, for all $\delta > 0$, thus proving the hardness of $n^{1/4-\epsilon}$ for the independent set problem. The result has been subsequently improved by Bellare et al. in [5] who gave a construction of a PCP with free-bit complexity $f = 2+\delta$. Finally, Håstad [31] constructed a PCP with arbitrary small free-bit complexity $f > 0$, thus showing the tight hardness (up to the lower order term) of $n^{1-\epsilon}$, for all $\epsilon > 0$, for the independent set problem. A PCP with optimal free-bit complexity was first constructed by Samorodnitsky and Trevisan [45]. The PCP of Samorodnitsky and Trevisan in [45] has imperfect completeness, and this has been improved by Håstad and Khot [32] to a PCP that has both perfect completeness and optimal free-bit complexity. Recently, Moshkovitz and Raz [40] gave a construction of a projective 2-query PCP with nearly-linear size, which can be combined with the result of Samorodnitsky and Trevisan [45] to obtain a PCP with nearly-linear size and optimal free-bit complexity. The soundness of a PCP with optimal free-bit complexity was improved in a very recent breakthrough result of Chan [17]. The complexity assumption of early tight hardness results of the independent set problem (e.g., [31]) is $\mathsf{NP} \neq \mathsf{ZPP}$ because of a random process in the constructions. This process was derandomized in [49] by Zuckerman, thus proving tight hardness under the $\mathsf{P} \neq \mathsf{NP}$ assumption.

Our Results: We present several tight hardness results both for polynomial and subexponential-time algorithms, as summarized in Table I. Most our results rely on a plausible complexity theoretic assumption stronger than $\mathsf{P} \neq \mathsf{NP}$, namely, *Exponential Time Hypothesis* (ETH), which, roughly speaking, states that SAT cannot be decided by any subexponential time algorithm (see Section II for detail).

Our first result, which is our original motivation, is the tight hardness of approximating the k-hypergraph pricing problems in polynomial time. These problems (both UDP and SMP) are known to be $O(k)$-approximable [3], [10] and the hardness of $\Omega(k^{1/2-\epsilon})$ and $\Omega(n^{\delta})$, for some constant $\delta > 0$, are known based on the assumption about hardness of refuting a random 3SAT formula [9]. A series of recent results leads to a disagreement on the right approximability thresholds of the problem. On one hand, the current best approximation algorithm is so simple that one is tempted to believe that a more sophisticated idea would immediately give an improvement on the approximation ratio. On the other hand, no algorithmic approach could go beyond the barrier of $O(k)$ so far, thus leading to a belief that $\Omega(k^{1-\epsilon})$ and $\Omega(n^{1-\epsilon})$ hardness should hold. In this paper, we settle the approximability status of this problem. Somewhat surprisingly, the right hardness threshold of this problem turns out to lie somewhere between the two previously believed numbers: the believed hardness of $\Omega(k^{1-\epsilon})$ was correct but *only for* $k = O(n^{1/2})$.

Theorem I.1. *The k-hypergraph pricing problems (both UDP and SMP) are $\Omega(\min(k^{1-\epsilon}, n^{1/2-\epsilon}))$ hard to approximate in polynomial time unless the ETH is false*[1]. *Moreover, they are $\tilde{O}(\min(k, (n \log n)^{1/2}))$-approximable in polynomial time.*

The main ingredient in proving Theorem I.1 is proving tight hardness thresholds of *subexponential-time* algorithms for the independent set and the induced matching problems in d-degree-bounded bipartite graphs[2]. Besides playing a crucial role in proving Theorem I.1, these results are also of an independent interest. Our first subexponential-time hardness result is for the independent set problem. While the polynomial-time hardness of this problem has been almost settled, the question whether one can do better in subexponential time has only been recently raised in the parameterized complexity community. Cygan et al. [23] and Bourgeois et al. [8] independently observed that better approximation ratios could be achieved if subexponential running time is allowed. In particular, they showed that an r approximation factor can be obtained in $O(2^{n/r}\,\mathrm{poly}(n))$ time. Recently, Chitnis et al. [19] showed that, assuming the ETH, an r-approximation algorithm requires $\Omega(2^{n^{\delta}/r})$ time, for some constant $\delta > 0$[3]. Our hardness of the independent set problem improves upon the lower bound of Chitnis et al. and essentially matches the upper bounds of Cygan et al. and Bourgeois et al.

Theorem I.2. *Any r approximation algorithm for the independent set problem must run in time $2^{n^{1-\epsilon}/r^{1+\epsilon}}$ unless the ETH is false.*

An important immediate step in using Theorem I.2 to prove Theorem I.1 is proving the subexponential-time hardness of the induced matching problem on d-degree-bounded bipartite graphs. The polynomial-time hardness of $n^{1-\epsilon}$ for this problem has only been resolved recently by the authors of this paper in [15], where we also showed a hardness of $d^{1/2-\epsilon}$ when the input graph is bipartite of degree at most d. In this paper, we improve this bound to $d^{1-\epsilon}$ and extend the validity

[1] The $k^{1-\epsilon}$ hardness only requires $\mathsf{NP} \neq \mathsf{ZPP}$ when k is constant.

[2] We note that, indeed, the connection to the pricing problem is via a closely related problem, called the *semi-induced matching* problem, whose hardness follows from the same construction as that of the induced matching problem; see Section II.

[3] We note that their real statement is that for any constant $\epsilon > 0$, there is a constant $F > 0$ depending on ϵ such that CLIQUE (or equivalently the independent set problem) does not have an n^{ϵ}-approximation in $O(2^{\mathsf{OPT}^{F/\epsilon}} \cdot \mathrm{poly}(n))$ time. Their result can be translated into the result we state here.

Problem		Upper	Lower
k-hypergraph pricing (polynomial time)	Previous	$O(k)$ [3], [10]	$\Omega(k^{1/2-\epsilon})$ and $\Omega(n^{\delta})$ [9], [13], [15]
	This paper	$O(\min(k, (n \log n)^{1/2}))$	$\Omega(\min(k^{1-\epsilon}, n^{1/2-\epsilon}))$
k-hypergraph pricing (quasi-polynomial time)	Previous	-	-
	This paper	n^{δ}-approx. algo in $O(2^{(\log m)^{\frac{1-\delta}{\delta}} \log\log m} \operatorname{poly}(n,m))$-time	n^{δ}-approx. algo requires $\Omega(2^{(\log m)^{\frac{1-\delta-\epsilon}{\delta}}})$ time
Independent set (subexpo.-time r-approx. algo)	Previous	$O(2^{n/r} \operatorname{poly}(n))$ time [23]	$\Omega(2^{n^{\delta}/r})$ time [19]
	This paper	-	$\Omega(2^{n^{1-\epsilon}/r^{1+\epsilon}})$ time
Induced matching on d-deg.-bounded bip. graphs (polynomial time)	Previous	$O(d)$ (trivial)	$\Omega(d^{1/2-\epsilon})$ [15]
	This paper	-	$\Omega(d^{1-\epsilon})$
Induced matching on bip. graphs (subexpo.-time r-approx. algo)	Previous	-	-
	This paper	$O(2^{n/r} \operatorname{poly}(n))$ time	$\Omega(2^{n^{1-\epsilon}/r^{1+\epsilon}})$ time

TABLE I
SUMMARY OF RESULTS.

scope of the results to subexponential-time algorithms. The latter result is crucial for proving Theorem I.1.

Theorem I.3. *Let $\epsilon > 0$ be any constant. For any $d \geq c$, for some constant c (depending on ϵ), there is no $d^{1-\epsilon}$ approximation algorithm for the induced matching problem in d-degree-bounded bipartite graphs unless* NP = ZPP. *Moreover, any r approximation algorithm for the bipartite induced matching problem must run in time $2^{n^{1-\epsilon}/r^{1+\epsilon}}$ unless the ETH is false.*

Finally, we note an interesting fact that, while the polynomial-time hardness of the k-hypergraph pricing problem follows from the hardness of the independent set and the bipartite induced matching problems, its subexponential time approximability is quite different from those of the other two problems. In particular, if we want to get an approximation ratio of n^{ϵ} for some constant $\epsilon > 0$. Theorems I.2 and I.3 imply that we still need subexponential time to achieve such an approximation ratio. In contrast, we show that, for the case of the k-hypergraph pricing problem, such an approximation ratio can be achieved in quasi-polynomial time.

Theorem I.4. *For the k-hypergraph pricing problem and any constant $\delta > 0$, there is an algorithm that gives an approximation ratio of $O(n^{\delta})$ and runs in time $O(2^{(\log m)^{\frac{1-\delta}{\delta}} \log\log m} \operatorname{poly}(n,m))$.*

We also prove that the above upper bound is tight.

Techniques: The key ingredients in our proofs are *tight* connections between 3SAT, independent set, induced matching, and pricing problems. Our reductions are fairly tight in the sense that their improvement would violate the ETH. They are outlined in Figure 1 Among several techniques we have to use, the most important new idea is a *simple new property of dispersers*. We show that an operation called *bipartite double cover* will convert a disperser G to another bipartite graph H whose induced matching number is essentially the same as the size of a maximum independent set of G. This property crucially relies on the fact that G is a disperser. We note that the disperser that we study here is not new – it has been around for at least two decades (e.g., [21], [46]) – but the property that we prove is entirely new.

Our hardness proof of the induced matching problem in d-degree-bounded graphs is inspired by the (implicit) reduction of Briest [9] and the previous (explicit) reduction of ours [15]. These previous reductions are not tight in two aspects: (i) they do not result in a tight $d^{1-\epsilon}$ hardness for the bipartite induced matching problem (Briest's reduction in [9] only gives a hardness of d^{δ} for some $\delta > 0$, and our previous reduction in [15] only gives a hardness of $d^{1/2-\epsilon}$), and (ii) they do not give any hardness for subexponential-time approximation algorithms. In this paper, we make use of additional tools to improve these two aspects to get a tight reduction. The first tool is the new property of the disperser as we have discussed. The second tool is a new PCP which results from carefully combining known techniques and ideas from the PCP literature.

Our PCP requires an intricate combination of many known properties: That is, it must be sparse, as well as, having nearly-linear size, large degree, and small free-bit complexity. We explain some of the required properties here. The sparsity and the size of the PCP are required in order to boost hardness of the k-hypergraph pricing problem to $n^{1/2-\epsilon}$ (without these, we would not go beyond n^{δ} for some small δ). The large degree of the PCP is needed to ensure that our randomized construction is successful with high probability. Finally, the small free-bit complexity is needed to get the $d^{1-\epsilon}$ hardness for the bipartite induced matching and the independent set problems in d-degree-bounded graphs; this is the same idea as those used in the literature of proving hardness of the independent set problem.

Our proof of the hardness of the k-hypergraph pricing problem from the hardness of the bipartite induced matching problem is inspired by the previous proofs in [11], [13]. Both previous proofs require a hardness of some special cases of the bipartite induced matching problem (e.g., [15] requires that the input instance is a result of a certain graph product) in order to derive

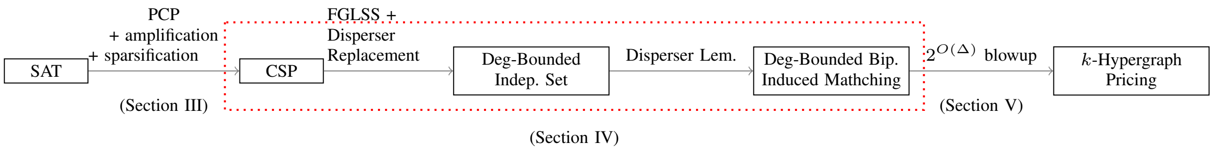

Fig. 1. Hardness proof outline.

the hardness of the k-hypergraph pricing problem. In this paper, we provide insights that lead to a reduction that simply exploits the fact that the input graph has bounded degree, showing a clearer connection between the two problems.

II. Preliminaries

We use standard graph terminologies as in [24]. Denote any graph by $G = (V, E)$. When we consider more than one graph, notations $V(G)$ and $E(G)$ are used for vertices and edges of G, respectively. A set of vertices $S \subseteq V$ is *independent* (or *stable*) in G if and only if G has no edge joining a pair of vertices $u, v \in V$. A set of edges $M \subseteq E$ is a *matching* in G if and only if no two edges of M share an end-vertex, and a matching M is an *induced matching* in G is the subgraph of G induced by M is exactly M.

Semi-Induced Matching: Given a *permutation* (a.k.a, a *total order*) σ of V, a set of edges $M \subseteq E$ is a *σ-semi-induced matching* in G if and only if, for every pair of edges $uv, ab \in M$ such that $\sigma(u) < \sigma(a)$, G has none of the edges ua, ub. Given any graph G and a total order σ, we use the notation $\mathsf{sim}_\sigma(G)$ to denote the size of a maximum σ-semi-induced matching in G, and let $\mathsf{sim}(G) = \max_\sigma \mathsf{sim}_\sigma(G)$. Notice that for any σ, if M is an induced matching in G, then M is also a σ-semi-induced matching in G, so we must have $\mathsf{im}(G) \leq \mathsf{sim}_\sigma(G) \leq \mathsf{sim}(G)$. In the semi-induced matching problem, our goal is to compute $\mathsf{sim}(G)$.

Our hardness proof of the bipartite induced matching problem will, in fact, show a stronger property than just bounding the size of a maximum induced matching. That is, in the completeness case, the reduction guarantees that $\mathsf{im}(G) \geq c$ while in the soundness case, it gives $\mathsf{sim}(G) \leq s$ (where s and c are soundness and completeness parameters, respectively). Notice that $\mathsf{sim}(G) \leq s$ implies $\mathsf{im}(G) \leq s$, so this stronger property implies the (c/s)-hardness of approximating the bipartite induced matching problem as a consequence.

The Pricing Problems: In an equivalent formulation of the k-hypergraph pricing problem, the pricing instance is given by two sets $(\mathcal{C}, \mathcal{I})$ where $\mathcal{C}$ and $\mathcal{I}$ are the sets of consumers and items, respectively. Each consumer $c \in \mathcal{C}$ is associated with a budget B_c and item set $S_c \subseteq \mathcal{I}$, $|S_c| \leq k$. Notice that this formulation is equivalent to the hypergraph formulation, i.e., each vertex corresponds to an item and each edge corresponds to a consumer, and the additional constraint $|S_c| \leq k$ ensures that the size of each hyperedge is at most k. We will be mostly using this formulation.

Constraint Satisfaction Problems: One of the most fundamental problems in theoretical computer science is k-SAT, where we are given a CNF formula φ, and the goal is to decide whether there is an assignment to boolean variables of φ that satisfies all the formula. In the maximization version of k-SAT, the goal is to find an assignment that maximizes the number of clauses of φ satisfied.

The *k-constraints satisfaction problem (k-CSP)* is a generalization of k-SAT, in which each clause is a boolean function Π_j on k variables. The goal of k-CSP is to find an assignment to variables that satisfies as many clauses as possible. That is, the goal is to find an assignment f such that $\Pi_j(f_j) = 1$ for all clause Π_j, where f_j is a *partial assignment* restricted to only variables that appear in Π_j. We use the term *assignment of a clause* Π_j to mean a partial assignment restricted to variables in Π_j.

Hypothesis II.1 (Exponential-Time Hypothesis (ETH))**.** *For any integer $k \geq 3$, there is a constant $0 < q_0(k) < 1$ such that there is no 2^{qN} time algorithm, for all $q < q_0(k)$, that solves k-SAT where N is the size of the instance. In particular, there is no $2^{o(N)}$ time algorithm that solves 3-SAT.*

The ETH was originally stated in terms of the number of variables. Impagliazzo, Paturi and Zane [33] showed that the statement is equivalent for all the parameters, i.e., N in the statement can be the number of variables, the number of clauses or the size of the instance. For more discussion related to the ETH, we refer readers to a comprehensive survey by Lokshtanov et al. [37] and references therein.

III. Nearly-linear size sparse PCP with small free-bit complexity and large degree

We construct the following CSP. Its proof is provided in the full version.

Theorem III.1 (Nearly-linear size sparse PCP with small free-bit complexity and large degree)**.** *Let k, t be parameters and $\epsilon = 1/k$. Also, let $\delta > 0$ be any parameter. There is a randomized polynomial-time algorithm that transforms a 3SAT formula of size N to a (tq)-CSP formula φ, where $q = k^2+2k$, that satisfies the following properties:*

- *(Small Number of Variables and Clauses) The number of variables is at most $qN^{1+\epsilon}$, and the number of clauses is $M = 100q2^{t(k^2+1)}N^{1+\epsilon+\delta}$.*
- *(Big Gap between Completeness and Soundness) The value*

of the YES-INSTANCE *is at least* $c = 1/2^{t+1}$*, and the value of* NO-INSTANCE *is* $s = 2^{-t(k^2+2)}$.

• *(Free-Bit Complexity) For each clause in* φ*, the number of satisfying assignments for such clause is* $w = 2^{2kt}$*. Moreover, for each variable* x_j *that appears in a clause, the number of satisfying assignments for which* $x_j = 0$ *is equal to the number of satisfying assignments for which* $x_j = 1$.

• *(Large Degree) For each variable* x_j*, the total number of clauses in which* x_j *appears is* $M_j \geq N^{\delta} 2^{t(k^2+1)}$.

IV. TIGHT HARDNESS OF SEMI-INDUCED MATCHING

Here we prove the (almost) tight hardness of the induced matching problem on a Δ-degree bounded bipartite graph.

Theorem IV.1 (Hardness of Δ-Degree Bounded Bipartite Semi-induced Matching)**.** *Let* $\epsilon > 0$ *be any constant and* $t > 0$ *be a positive integer. There is a randomized algorithm that transforms a SAT formula* φ *of input size* N *into a* Δ*-degree bounded bipartite graph, where* $\Delta = 2^{t(\frac{1}{\epsilon^2}+O(\frac{1}{\epsilon}))}$ *such that:*

*(*YES-INSTANCE*:)* φ *is satisfiable* $\implies \mathsf{im}(G) \geq \frac{|V(G)|}{\Delta^{\epsilon}}$.
*(*NO-INSTANCE*:)* φ *is not satisfiable* $\implies \mathsf{sim}(G) \leq \frac{|V(G)|}{\Delta^{1-\epsilon}}$.

The construction size is $|V(G)| \leq N^{1+\epsilon}\Delta^{1+\epsilon}$*, and the running time is* $\mathrm{poly}(N, \Delta)$*. Also, as long as* $t \leq 5\epsilon^2 \log N$*, the reduction is guaranteed to be successful with high probability.*

Theorem IV.2 (Hardness of d-Degree-Bounded Independent Set)**.** *Let* $\epsilon > 0$ *be any sufficiently small constant and* $t > 0$ *be a positive integer. There is a randomized algorithm that transforms a SAT formula* φ *on input of size* N *into a* d*-degree-bounded graph* G*, where* $d = 2^{t(\frac{1}{\epsilon^2}+O(\frac{1}{\epsilon}))}$ *such that:*

*(*YES-INSTANCE*:)* φ *is satisfiable* $\implies \alpha(G) \geq \frac{|V(G)|}{d^{\epsilon}}$.
*(*NO-INSTANCE*:)* φ *is not satisfiable* $\implies \alpha(G) \leq \frac{|V(G)|}{d^{1-\epsilon}}$.

The construction size is $|V(G)| = N^{1+\epsilon}d^{1+\epsilon}$*, and the running time is* $\mathrm{poly}(N, d)$*. Also, as long as* $t \leq 5\epsilon^2 \log N$*, the reduction is guaranteed to be successful with high probability.*

The results in Theorems I.2 and I.3 are obtained as a by-product of Theorems IV.1 and IV.2.

The Reduction Our reduction is as follows. Take an instance φ of (qt)-CSP as in Theorem III.1 that has N variables and M clauses.

The FGLSS Graph $\widehat{G}$ *with Disperser Replacement:* First, we construct from φ a graph $\widetilde{G}$ by the FGLSS construction, and then the graph $\widetilde{G}$ will be transformed to graph $\widehat{G}$ by the disperser replacement step. For each clause φ_j of φ, for each possible satisfying assignment C of φ_j, we create in $\widetilde{G}$ a vertex $v(j, C)$ representing the fact that "φ_j is satisfied by assignment C". Then we create an edge $v(j, C)v(j', C') \in E(\widetilde{G})$ if there is a *conflict* between partial assignments C and C', i.e., there is a variable x_i appearing in clauses φ_j and $\varphi_{j'}$ such that C assigns $x_i = 0$ whereas C' assigns $x_i = 1$. So, the total number of vertices is $|V(\widetilde{G})| = w \cdot M$. The independence number of a graph $\widetilde{G}$ corresponds to the number of clauses of φ that can be satisfied. In particular, we can choose at most one vertex from each clause φ_j (otherwise, we would have a conflict between $v(j, C)$ and $v(j, C')$), and we can choose two vertices $v(j, C), v(j', C') \in V(\widetilde{G})$ if and only if the assignment C and C' have no conflict between variables. Thus, the number of satisfiable clauses of φ is the same as the independence number $\alpha(\widetilde{G})$. Hence, in the YES-INSTANCE, we have $\alpha(\widetilde{G}) \geq c \cdot M$, and in NO-INSTANCE, we have $\alpha(\widetilde{G}) \leq s \cdot M$. This gives a hard instance of the independent set problem, but the degree of $\widetilde{G}$ can be very high.

Next, we reduce the degree of $\widetilde{G}$ via the disperser replacement step as in [48]. Consider an additional property of $\widetilde{G}$. For each variable x_i in φ, let O_i and Z_i denote the set of vertices $v(j, C)$ corresponding to the (partial) assignments for which $x_i = 1$ and $x_i = 0$, respectively. It can be deduced from Theorem III.1 that $|O_i| = |Z_i| = M_i/2 \geq 2^{t(k^2+1)}N^{\delta}$, for some constant $\delta > 0$.

There is a conflict between every vertex of O_i and Z_i. This forms a complete bipartite subgraph $\widetilde{G}_i = (O_i, Z_i, \widetilde{E}_i) \subseteq \widetilde{G}$, where $\widetilde{E}_i = \{uw : u \in O_i, w \in Z_i\}$. If we replace each subgraph $\widetilde{G}_i$ of G by a d-degree bounded bipartite graph, the degree of vertices in the resulting graph reduces to qtd. To see this, we may think of each vertex u of $\widetilde{G}$ as a vector with qt coordinates (since it corresponds to an assignment to some clause φ_j which has qt related variables). For each coordinate ℓ of u corresponding to a variable x_i, there are d neighbors of u having a conflict at coordinate ℓ (since the conflict forming in each coordinate are edges in $\widetilde{G}_i$, and we replace $\widetilde{G}_i$ by a d-degree bounded bipartite graph). Thus, each vertex u has at most qtd neighbors. However, as we wish to preserve the independence number of G, i.e., we want $\alpha(\widehat{G}) \approx \alpha(\widetilde{G})$, we require such a d-degree bounded graph to have some additional properties. To be precise, we construct the graph $\widehat{G}$ by replacing each subgraph $\widetilde{G}_i$ of $\widetilde{G}$ by a (d, γ)-*disperser* $H_i = (O_i, Z_i, E_i)$, defined below.

Definition IV.3 (Disperser)**.** *A* (d, γ)*-disperser* $H = (U', W', E')$ *is a* d*-degree bounded bipartite graph on* $n' = |U'| = |W'|$ *vertices such that, for all* $X \subseteq U', Y \subseteq W'$*, if* $|X|, |Y| \geq \gamma n'$*, then there is an edge* $xy \in E'$ *joining a pair of vertices* $x \in X$ *and* $y \in Y$.

Intuitively, the important property of the disperser H_i is that any independent set S in H_i cannot contain a large number of vertices from both O_i and Z_i; otherwise, we would have an edge joining two vertices in S. All these ideas of using disperser to "sparsify" the graphs were used by Trevisan [48] to prove the hardness of the bounded degree independent set problem. The key observation that makes this construction work for our problem is that a similar property that holds for the size of a maximum independent set also holds for the size of a maximum σ-semi-induced matching in $\mathrm{B}[H_i]$, i.e., $\mathrm{B}[H_i]$ cannot contain a large σ-semi-induced matching, for any permutation σ.

Now, we proceed to make the intuition above precise. A (d,γ)-disperser can be constructed by a randomized algorithm. If d is constant, we may construct a (d,γ)-disperser by a deterministic algorithm in [42], which has a running time exponential in terms of d.

Lemma IV.4. *For all $\gamma > 0$ and sufficiently large n, there is a randomized algorithm that with success probability $1 - e^{-n\gamma(\log(1/\gamma)-2)}$, outputs a d-regular bipartite graph $H = (O,Z,E), |O| = |Z| = n$, where $d = (3/\gamma)\log(1/\gamma)$ such that, for all $X \subseteq Z, Y \subseteq O$, if $|X|,|Y| \geq \gamma n$, there is an edge $(x,y) \in E$ joining some vertices $x \in X$ and $y \in Y$.*

The condition that n_i is sufficiently large is satisfied because $|O_i| = |Z_i| \geq M_i \geq N^{\delta}2^{tk^2}$ for all i (since each variable x_i appears in M_i clauses, and for each such clause, there is at least one accepting configuration for which $x_i = 0$ and one for which $x_i = 1$.) Also, since the success probability in constructing each disperser is high (i.e., at least $2^{N^{\delta}}$), we can guarantee that all the dispersers are successfully constructed with high probability. By setting appropriate value of γ (which we will do later) and following analysis in [48], we have the following completeness and soundness parameters with high probability: (YES-INSTANCE:) $\alpha(\widehat{G}) \geq 2^{-t}M$ and (NO-INSTANCE:) $\alpha(\widehat{G}) \leq 2^{-t(k^2+2)}M + \gamma qt(wM)$.

The Final Graph G: We construct the final graph G by transforming $\widehat{G}$ into a bipartite graph as follows: first create two copies V' and V'' of vertices of $\widehat{G}$, i.e., each vertex $u \in V(G)$ has two corresponding copies $u' \in V'$ and $u'' \in V''$; then create an edge joining two vertices $u' \in V'$ and $w'' \in V''$ if and only if there is an edge $uw \in E(\widehat{G})$ or $u = w$. Thus, $G = B[\widehat{G}] = (U \cup W, E_1 \cup E_2)$ where $U = \{(u,1) : u \in V(\widehat{G})\}$, $W = \{(w,2) : w \in V(\widehat{G})\}$, $E_1 = \{(u,1)(u,2) : u \in V(\widehat{G})\}$, and $E_2 = \{(u,1)(w,2) : u,w \in V(\widehat{G}) \wedge (uw \in E(\widehat{G}))\}$.

The graph G is a $(2qtd+1)$-degree bounded bipartite graph on $2|V(\widehat{G})|$ vertices. Edges in G of the form $(u,1)(u,2)$ correspond to a vertex in $\widehat{G}$. Thus, a (semi) induced matching in G whose edges are in this form corresponds to an independent set in $\widehat{G}$. Although this is not the case for every (semi) induced matching $\mathcal{M}$ in G, we will show that we can extract a (semi) induced matching $\mathcal{M}'$ from $\mathcal{M}$ in such a way that $\mathcal{M}'$ maps to an independent set in G, and $|\mathcal{M}'| \geq \Omega(|\mathcal{M}|)$.

Analysis First, we prove the properties of a disperser.

Lemma IV.5 (Disperser Lemma)**.** *Every (d,γ)-disperser $H = (O,Z,E)$ on $2n$ vertices has the following properties: (1) For any independent set S of H, $\min\{|S \cap O|, |S \cap Z|\} \leq \gamma n$. (2) For any permutation (ordering) σ of the vertices of H, the graph $B[H] = (U,W,F)$ has $\mathsf{sim}_\sigma(B[H]) \leq 4\gamma n$.*

Proof: The first property follows from the definition of the (d,γ)-disperser H. That is, letting $X = S \cap O$ and $Y = S \cap Z$, if $|X|,|Y| > \gamma n$, then we must have an edge $xy \in E(H)$ joining some vertex $x \in X$ to some vertex $y \in Y$, contradicting the fact that S is an independent set in H.

Next, we prove the second property. Consider the set of edges $\mathcal{M}$ that form a σ-semi-induced matching in $\mathrm{B}[G]$. We claim that $|\mathcal{M}| \leq 4\gamma n$. By way of contradiction, assume that $|\mathcal{M}| > 4\gamma n$. Observe that, for each edge $(u,1)(v,2) \in \mathcal{M}$, either (1) $u \in O$ and $v \in Z$ or (2) $v \in O$ and $u \in Z$. Since the two cases are symmetric, we analyze only the set of edges of the first case, denoted by $\widehat{\mathcal{M}}$. Also, we assume wlog that at least half of the edges of $\mathcal{M}$ are in $\widehat{\mathcal{M}}$; thus, $|\widehat{\mathcal{M}}| \geq |\mathcal{M}|/2 > 2\gamma n$. Denote by $V(\widehat{\mathcal{M}})$ the set of vertices that are adjacent to some edges in $\widehat{\mathcal{M}}$. Then we get a contradiction from the next claim.

Claim IV.6. *There are two subsets $X \subseteq U \cap V(\widehat{\mathcal{M}}) : |X| = \gamma n$ and $Y = W \cap V(\widehat{\mathcal{M}}) : |Y| \geq \gamma n$ such that $\sigma(x) < \sigma(y)$, for any $x \in X$ and $y \in V(\widehat{\mathcal{M}}) \setminus X$. Moreover, there is no $\widehat{\mathcal{M}}$-edge between vertices in X and Y.*

The second property follows from Claim IV.6: If there were such two sets X,Y, then we can define the "projection" of X and Y onto the graph H by $X' = \{u \in V(H) : (u,1) \in X\}$ and $Y' = \{v \in V(H) : (v,2) \in Y\}$. So, $X' \subseteq O$ and $Y' \subseteq Z$ (due to the definition of $\widehat{\mathcal{M}}$). From the property of disperser, there must be an edge in $E(H)$ joining some $x \in X'$ and $y \in Y'$. This implies there is an edge $(x,1)(y,2) \in E(\mathrm{B}[H])$ where $x \in X$ and $y \in Y$, and there are edges $(x,1)(x',2) \in \widehat{\mathcal{M}}$ and $(y',1)(y,2) \in \widehat{\mathcal{M}}$, contradicting the fact that $\mathcal{M}$ is σ-semi-induced matching. To finish, we prove Claim IV.6 below.

Recall that we have the ordering σ that is defined on the vertices of $\mathrm{B}[H]$, not the vertices of H. We construct X and Y as follows. Order vertices in $U \cap V(\widehat{\mathcal{M}})$ according to the ordering σ and define X to be the first γn vertices according to this ordering. So, we have obtained $X \subseteq U \cap V(\widehat{\mathcal{M}})$ with the property that for any $x \in X$ and $y \in V(\widehat{\mathcal{M}}) \setminus X$, $\sigma(x) < \sigma(y)$. Now, we define $Y \subseteq W \cap V(\widehat{\mathcal{M}})$ as the set of vertices that are not matched by $\widehat{\mathcal{M}}$ with any vertices in X. Since $|X| = \gamma n$, the number of vertices in $W \cap V(\widehat{\mathcal{M}})$ that are matched by $\widehat{\mathcal{M}}$ is only γn, so we can choose arbitrary γn vertices that are not matched as our set Y. ■

As a corollary of Lemma IV.5, we relate the independent number the FGLSS graph $\widetilde{G}$ to the final graph G.

Corollary IV.7. *Let $\widetilde{G}$ and G be the graphs constructed as above. Then, for any permutation (ordering) σ of vertices of G, $\alpha(\widehat{G}) \leq \mathsf{sim}_\sigma(G) \leq \alpha(\widehat{G}) + 4\gamma|V(\widehat{G})|$*

Proof: Recall that $E(G) = E_1 \cup E_2$. To prove the left-hand-side, consider the set of edges E_1. Observe that edges of $E_1 = \{(v,1)(v,2) : u \in V(\widehat{G})\}$ correspond to vertices of $\widehat{G}$ as $\widehat{G}$ and $\widehat{G}$ share the same vertex set. Let S be an independent set in $\widehat{G}$. We claim that the set $E_S = \{(u,1)(u,2) : u \in S\}$ must be an induced matching in G, and this would immediately imply the first inequality: Assume that there was an edge $(u,1)(v,2) \in E(G)$ for some $(u,1)(u,2),(v,1)(v,2) \in E_S$. Then we must have $u,v \in S$ and $uv \in E(\widehat{G}) \subseteq E(\widehat{G})$. This contradicts the fact that S is an independent set.

To prove the right-hand-side, let $\mathcal{M}$ be a σ-semi-induced matching in G. We decompose $\mathcal{M}$ into $\mathcal{M} = \mathcal{M}_1 \cup \mathcal{M}_2$. By an argument similar to the previous one, we clearly have $|\mathcal{M}_1| \leq \alpha(\widehat{G})$. From $\mathcal{M}_1$, we define a set $S \subseteq V(\widetilde{G})$ by $S = \{u \in V(\widetilde{G}) : (u,1)(u,2) \in \mathcal{M}_1\}$, and S must be an independent set in $\widehat{G}$; otherwise, if $uv \in E(G)$ for $u, v \in S$, then we would have $(u,1)(v,2),(v,1)(u,2) \in E(G)$, contradicting to the fact that $\mathcal{M}_1$ is a σ-semi-induced matching.

It suffices to show that $|\mathcal{M}_2| \leq 4\gamma|V(\widehat{G})|$. So, we partition $\mathcal{M}_2$ into $\mathcal{M}_2 = \bigcup_{j=1}^{N} \mathcal{M}_2^j$ where $\mathcal{M}_2^j = \{(u,1)(v,2) \in \mathcal{M}_2, uv \in E(H_j)\}$ (since $\widehat{G}$ is the union of edges of subgraphs H_j). Each set $\mathcal{M}_2^j$ must be a σ_j-induced matching for the ordering σ_j obtained by projecting σ onto the vertices of $B[H_j]$. Hence, Lemma IV.5 implies $|\mathcal{M}_2^j| \leq 4\gamma n_j$ and thus $|\mathcal{M}_2| \leq \sum_{j=1}^{N} |\mathcal{M}_2^j| \leq \sum_{j=1}^{N} 4\gamma n_j \leq 4\gamma qt|V(\widehat{G})|$.

The last inequality follows because of basic counting arguments. Each vertex belongs to exactly qt subgraphs H_j, so summing over all j, we get $\sum_{j=1}^{N} n_j = qt|V(\widehat{G})|$. ■

Completeness and Soundness: The completeness and soundness proofs are now easy. In the YES-INSTANCE, $\alpha(\widehat{G}) \geq c \cdot M$ implies that $\mathsf{sim}_\sigma(G) \geq c \cdot M$, and in the NO-INSTANCE, the fact that $\alpha(\widehat{G}) \leq s \cdot M + \gamma qtwM$ implies that $\mathsf{sim}_\sigma(G) \leq s \cdot M + 5\gamma qtwM$. Now, we choose $\gamma = s/(5qtw)$, which gives $d = O(\frac{1}{\gamma}\log\frac{1}{\gamma}) = O((wqt/s)\log(wqt/s))$. Then we have the final graph G with $n = 2wM$, $\Delta = (2dq+1)$, and hardness gap $g \geq \dfrac{c}{2s}$. Substituting c, s, w, q, M as in Theorem III.1, we get degree $\Delta = O(t^2k^4 2^{t(k^2+2k-1)}) = 2^{t(1/\epsilon^2+\Theta(1/\epsilon))}$, size $|V(G)| = 2^{t(k^2)}N^{1+O(\epsilon)} = \Delta^{1+O(\epsilon)}N^{1+O(\epsilon)}$, and hardness gap $g \geq 2^{t(k^2-1)} \geq \Delta^{1-O(\epsilon)}$.

Success probability of the disperser construction: Notice that the failure probability of the disperser construction given in Lemma IV.4 is large when $N_i\gamma$ is small. In our case, we have $N_i \geq 2^{tk^2}N^\delta$ and $\gamma \geq 2^{-t(k^2+O(k))}$. So, we are guaranteed that $N_i\gamma \geq N^\delta 2^{-O(tk)} = 2^{\delta\log N - O(tk)}$. As long as $t \leq O(\delta\epsilon)\log N$, we would be guaranteed that $N_i\gamma \geq N^{\delta/2}$, so the failure probability in Lemma IV.4 is at most $2^{-N^{\delta/2}}$. Taking a union bound will complete the analysis.

V. HARDNESS OF k-HYPERGRAPH PRICING PROBLEMS

Throughout this section, we use n and m to denote the number of items and consumers respectively. We remark the difference between n (the number of items in the pricing instance) and N (the size of 3SAT formula). We prove the following theorem.

Theorem V.1. *Unless* NP= ZPP*, for any $\epsilon > 0$, there is a constant k_0 (depending on ϵ) such that the k-hypergraph pricing problem for any constant $k > k_0$ is $k^{1-\epsilon}$ hard to approximate. Assuming Hypothesis II.1, for any $\epsilon > 0$, the k-hypergraph pricing problem is hard to approximate to within a factor of* $\min(k^{1-\epsilon}, n^{1/2-\epsilon})$.

Proof Overview and Organization: For any k-hypergraph pricing instance $(\mathcal{C}, \mathcal{I})$, we denote by $\mathsf{OPT}(\mathcal{C}, \mathcal{I})$ the optimal possible revenue that can be collected by any price function. The key in proving Theorem V.1 is the connection between the semi-induced matching and the k-hypergraph pricing problem:

Lemma V.2 (From Semi-induced Matching to Pricing)**.** *There is a randomized reduction that, given a bipartite graph $G = (U, V, E)$ with maximum degree d, outputs an instance $(\mathcal{C}, \mathcal{I})$ of the k-hypergraph pricing problem such that, with high probability,* $(6\ln d/\ln\ln d)\mathsf{sim}(G) \geq \mathsf{OPT}(\mathcal{C}, \mathcal{I}) \geq \mathsf{im}(G)$.

The number of consumers is $|\mathcal{C}| = |U|d^{O(d)}$ and the number of items is $|\mathcal{I}| = |V|$. Moreover, each consumer $c \in \mathcal{C}$ satisfies $|S_c| = d$. The running time of this reduction is $\mathrm{poly}(|\mathcal{C}|, |\mathcal{I}|)$.

Combining the above reduction in Lemma V.2 with the hardness of the induced and semi-induced matching problems in Theorem IV.1 (Section IV) leads to the following intermediate hardness result, which in turn leads to all the hardness results stated in Theorem V.1.

Lemma V.3. *Let $\epsilon > 0$ be any constant. There is a universal constant $d_0 = d_0(\epsilon)$ such that the following holds. For any function $d(\cdot)$ such that $d_0 \leq d(N) \leq N^{1-\epsilon}$, there is a randomized algorithm that transforms an N-variable 3SAT formula φ to a k-hypergraph pricing instance $(\mathcal{C}, \mathcal{I})$ such that: (1) For each consumer c, $|S_c| = d(N)$. (2) The algorithm runs in time $\mathrm{poly}(|\mathcal{C}|, |\mathcal{I}|)$. (3) $|\mathcal{C}| \leq d^{O(d)}N^{1+\epsilon}$ and $|\mathcal{I}| \leq N^{1+\epsilon}d^{1+\epsilon}$. (4) There is a value Z such that (*YES-INSTANCE*) if φ is satisfiable, then $\mathsf{OPT}(\mathcal{C}, \mathcal{I}) \geq Z$, and (*NO-INSTANCE*) if φ is not satisfiable, then $\mathsf{OPT}(\mathcal{C}, \mathcal{I}) \leq Z/d^{1-\epsilon}$.*

The proofs of Theorem V.1 and lemma V.3 can be found in the full version. Here we prove Lemma V.2 by showing a reduction from the semi-induced matching problem on a d-degree bounded bipartite graph to the *k-hypergraph pricing* problem. This reduction is randomized and is successful with a constant probability.

1) The Reduction: Let $G = (U, V, E)$ be a bipartite graph with maximum degree d. We assume wlog that $|U| \leq |V|$. Notice that $\mathsf{sim}(G) \geq \mathsf{im}(G) \geq |U|/d$. For each vertex u of G, let $N_G(u)$ denote the set of neighbors of u in G. If the choice of a graph G is clear from the context, then we will omit the subscript G. Our reduction consists of two phases.

Phase 1: Coloring. We color each vertex $u \in U$ of G by uniformly and independently choosing a random color from $\{1, 2, \ldots, d\}$. We denote by $U_i \subseteq U$, for each $i = 1, 2, \ldots, d$, the set of left vertices that are assigned a color i. We say that a right vertex $v \in V$ is *highly congested* if there is some $i \in [d]$ such that $|N_G(v) \cap U_i| \geq 3\ln d/\ln\ln d$; i.e., v has at least $3\ln d/\ln\ln d$ neighbors of the same color. Let $V_{high} \subseteq V$ be a subset of all vertices that are highly congested and $V' = V \setminus V_{high}$. Thus, V' is the set of vertices in V with highly congested vertices thrown away. Let G' be a subgraph of G induced by (U, V', E). The following property is what we need from this phase in the analysis in Section V-2.

Lemma V.4. *With probability at least* $1/2$, $\mathsf{im}(G') \geq (1 - 2/d)\mathsf{im}(G)$ *and* $\mathsf{sim}(G') \geq (1 - 2/d)\mathsf{sim}(G)$. *In particular, for* $d \geq 4$, $\mathsf{im}(G') \geq \mathsf{im}(G)/2$ *and* $\mathsf{sim}(G') \geq \mathsf{sim}(G)/2$ *with probability at least 1/2.*

The above lemma can be proved by a simple balls-and-bins argument. The proof can be found in the full version.

Phase 2: Finishing. An instance of the k-hypergraph pricing problem is constructed as follows. For each vertex $v \in V'$, we create an item $I(v)$. For each vertex $u \in U_i$, we create d^{3i} consumers; we denote this set of consumers by $\mathcal{C}(u)$. We define the budget of each consumer $c \in \mathcal{C}(u)$ where $u \in U_i$ to be $B_c = d^{-3i}$ and define $S_c = \{I(v) : v \in N_G(u)\}$, so it is immediate that $|S_c| \leq d$. To recap, we have $\mathcal{I} = \{I(v) : v \in V'\}$, $\mathcal{C} = \bigcup_{u \in U} \mathcal{C}(u)$ where $|\mathcal{C}(u)| = d^{3i}$ for $u \in U_i$, $B_u = \dfrac{1}{d^{3i}}$ for each consumer $u \in U_i$, and $S_c = \{I(v) : v \in N_G(u)\}$ for each customer $c \in \mathcal{C}$. In the k-hypergraph formulation, $\mathcal{I}$ is a set of vertices, $\mathcal{C}$ is a set of hyperedges and $k = d$ (since $|S_c| \leq d$ for all $c \in \mathcal{C}$).

2) Analysis: Completeness: We will show that the profit we can collect is at least $\mathsf{im}(G') \geq \mathsf{im}(G)/2$ (by Lemma V.4). Let $\mathcal{M}$ be any induced matching in the graph G'. For each item $I(v)$ with $uv \in \mathcal{M}$ and $u \in U_i$, we set its price to $p(I(v)) = 1/d^{3i}$. For all other items, we set their prices to ∞ for UDP and 0 for SMP. It is easy to check that the total profit is $|\mathcal{M}|$ (more detail in the full version).

Soundness: Now, suppose that an *optimal* price function p yields a profit of r (for either UDP or SMP). We show that $\mathsf{sim}(G) \geq r \log\log d/(12 \log d)$. The proof has two steps. In the first step, we identify a collection of "tight consumers" which roughly correspond to those consumers who pay sufficiently large fraction of their budgets. Then we construct a large semi-induced matching from these tight consumers. We say that a consumer $c \in \mathcal{C}$ is *tight* if she spends at least $1/4d$ fraction of her budget for her desired item. A vertex $u \in U$ is tight if its set of consumers $\mathcal{C}(u)$ contains a tight consumer. Let $\mathcal{C}'$ be the set of tight consumers.

Claim V.5. *The profit made only by tight consumers is* $\geq r/2$.

The proof of Claim V.5 is given in the full version. Now, we construct from the set of tight consumers $\mathcal{C}'$, a σ-semi-induced matching in G for some total order σ. We define σ so that vertices in U is ordered by their colors (increasingly for the case of UDP and decreasingly for the case of SMP).

Let σ be a total order of vertices such that vertices in U_i always precede vertices in U_j if $i < j$ for UDP ($i > j$ for SMP). ($U_i \subseteq U$ is the set of vertices with color i.) Let $U' = \{u \in U : \mathcal{C}(u) \cap \mathcal{C}' \neq \emptyset\}$. By Claim V.5, $|U'| \geq r/2$. For UDP, an edge uv is in $\mathcal{M}$ if $u \in U'$ and a tight consumer in $\mathcal{C}(u)$ buys an item $I(v)$. For SMP, $\mathcal{M}$ contains uv such that $u \in U'$ and $I(v)$ is the most expensive item for a consumer in $\mathcal{C}(u)$. Note that $|\mathcal{M}| \geq |U'| \geq r/2$. This collection $\mathcal{M}$ may not be a σ-semi-induced matching yet, so we will extract from $\mathcal{M}$ a set of edges $\mathcal{M}' \subseteq \mathcal{M}$ that is a σ-semi-induced matching with cardinality $|\mathcal{M}'| \geq r \log\log d/6 \log d$, implying that $\mathsf{sim}(G) \geq r \log\log d/6 \log d$.

We construct $\mathcal{M}'$ by incrementally adding edges from $\mathcal{M}$ as long as $\mathcal{M}'$ remains a σ-semi-induced matching (we do not add any edge that violates a σ-semi-induced matching.) The order of edges we pick from $\mathcal{M}$ depends *reversely* on σ, and we will also do this process separately for different colors of left vertices as follows. We partition $\mathcal{M}$ into $\mathcal{M}_1 \cup \mathcal{M}_2 \cup \ldots \cup \mathcal{M}_d$, where $\mathcal{M}_i = \{uv \in \mathcal{M} : u \in U_i\}$; i.e., $\mathcal{M}_i$ contains edges uv whose end-vertex u is colored i. Then we construct from each set $\mathcal{M}_i$ a set of edges $\mathcal{M}'_i$ as follows. We process each edge $uv \in \mathcal{M}_i$ in the *reverse* order of σ; i.e., an edge uv is processed before an edge $u'v'$ if $\sigma(u) > \sigma(u')$. For each edge $uv \in \mathcal{M}_i$, we remove from $\mathcal{M}_i$ all edges $u'v'$ such that u' is adjacent to v. Then we add uv to the set $\mathcal{M}'_i$ and proceed to the next edge remaining in $\mathcal{M}_i$. Notice that, each time we add an edge uv to $\mathcal{M}'_i$, we remove at most $3 \log d/\log\log d$ edges from $\mathcal{M}_i$ because its end-vertex is not highly congested by the construction of $\mathcal{M}_i$. So, $|\mathcal{M}'_i| \geq |\mathcal{M}_i| \log\log d/3 \log d$. Moreover, it can be seen by the construction that $\mathcal{M}'_i$ is a σ-semi-induced matching. Finally, define $\mathcal{M}' = \bigcup_{i=1}^{d} \mathcal{M}'_i$. Then we have that $|\mathcal{M}'| \geq \frac{|\mathcal{M}| \log\log d}{3 \log d}$. We now claim that $\mathcal{M}'$ is a σ-semi-induced matching. We need two cases for the two models of SMP and UDP. Details appear in the full version.

VI. Open Problems

There are many problems left open. The most fundamental one in algorithmic pricing (from the perspective of approximation algorithms community) is perhaps the *graph pricing problem* which is the k-hypergraph pricing problem where $k = 2$. Currently, only a simple 4-approximation algorithm and a hardness of $2 - \epsilon$ assuming the Unique Game Conjecture, are known [35]. It is interesting to see if the techniques in this paper can be extended to an improved hardness (which will likely to require even tighter connections). Other interesting problems that seem to be unachievable using the current techniques are the *Stackelberg network pricing*, *Stackelberg spanning tree pricing*, and *tollbooth pricing* problems.

Another interesting question is whether our techniques can be used to make a progress in the parameterized complexity domain. In particular, it was conjectured in [38], [19] that the independent set problem parameterized by the size of the solution does not admit an FPT approximation ratio ρ for any function ρ. It might be also interesting to improve our $2^{n^{1-\epsilon}/r^{1+\epsilon}}$ time lower bound for r-approximating the independent set and induced matching problems, e.g., to $2^{(n \cdot \mathrm{polylog}(r))/(r \cdot \mathrm{polylog}(n))}$. Other (perhaps less important) open problems also remain: (1) Is the ETH necessary in proving the lower bound of this problem? For example, can we get a better approximation guarantee for the k-hypergraph pricing problem if there is a subexponential-time algorithm for solving SAT (see, e.g., [22] for similar questions in the exact algorithm domain)? (2) Is it possible to obtain an r-approximation algorithm in $2^{n/r}$ time for the induced matching problem in general graphs?

REFERENCES

[1] S. Arora, C. Lund, R. Motwani, M. Sudan, and M. Szegedy, "Proof verification and the hardness of approximation problems," *J. ACM*, vol. 45, no. 3, pp. 501–555, 1998.

[2] S. Arora and S. Safra, "Probabilistic checking of proofs: A new characterization of NP," *J. ACM*, vol. 45, no. 1, pp. 70–122, 1998.

[3] M.-F. Balcan and A. Blum, "Approximation algorithms and online mechanisms for item pricing," *Theory of Computing*, vol. 3, no. 1, pp. 179–195, 2007.

[4] M.-F. Balcan, A. Blum, J. D. Hartline, and Y. Mansour, "Mechanism design via machine learning," in *FOCS*, 2005, pp. 605–614.

[5] M. Bellare, O. Goldreich, and M. Sudan, "Free bits, PCPs, and nonapproximability-towards tight results," *SIAM J. Comput.*, vol. 27, no. 3, pp. 804–915, 1998.

[6] M. Bellare and M. Sudan, "Improved non-approximability results," in *STOC*, 1994, pp. 184–193.

[7] V. Bonifaci, P. Korteweg, A. Marchetti-Spaccamela, and L. Stougie, "Minimizing flow time in the wireless gathering problem," *ACM Transactions on Algorithms*, vol. 7, no. 3, p. 33, 2011.

[8] N. Bourgeois, B. Escoffier, and V. T. Paschos, "Approximation of max independent set, min vertex cover and related problems by moderately exponential algorithms," *Discrete Applied Mathematics*, vol. 159, no. 17, pp. 1954–1970, 2011.

[9] P. Briest, "Uniform budgets and the envy-free pricing problem," in *ICALP (1)*, ser. Lecture Notes in Computer Science, vol. 5125. Springer, 2008, pp. 808–819.

[10] P. Briest and P. Krysta, "Single-minded unlimited supply pricing on sparse instances," in *SODA*, 2006, pp. 1093–1102.

[11] ——, "Buying cheap is expensive: Approximability of combinatorial pricing problems," *SIAM J. Comput.*, vol. 40, no. 6, pp. 1554–1586, 2011.

[12] K. Cameron, "Induced matchings," *Discrete Appl. Math.*, vol. 24, no. 1-3, pp. 97–102, 1989.

[13] P. Chalermsook, J. Chuzhoy, S. Kannan, and S. Khanna, "Improved hardness results for profit maximization pricing problems with unlimited supply," in *APPROX-RANDOM*, ser. Lecture Notes in Computer Science, vol. 7408. Springer, 2012, pp. 73–84.

[14] P. Chalermsook, S. Kintali, R. J. Lipton, and D. Nanongkai, "Graph pricing problem on bounded treewidth, bounded genus and k-partite graphs," *CoRR*, vol. abs/1203.1940, 2012.

[15] P. Chalermsook, B. Laekhanukit, and D. Nanongkai, "Graph products revisited: Tight approximation hardness of induced matching, poset dimension and more," in *SODA*, 2013, pp. 1557–1576.

[16] ——, "Independent set, induced matching, and pricing: Connections and tight (subexponential time) approximation hardnesses," *CoRR*, vol. abs/1308.2617, 2013. [Online]. Available: http://arxiv.org/abs/1308.2617

[17] S. O. Chan, "Approximation resistance from pairwise independent subgroups," *Electronic Colloquium on Computational Complexity (ECCC)*, vol. 19, p. 110, 2012, to appear in STOC 2013.

[18] S. Chawla, J. D. Hartline, and R. D. Kleinberg, "Algorithmic pricing via virtual valuations," in *ACM Conference on Electronic Commerce*, 2007, pp. 243–251.

[19] R. Chitnis, M. Hajiaghayi, and G. Kortsarz, "Fixed-parameter and approximation algorithms: A new look," *Unpublished Manuscript*, 2013, available at http://www.crab.rutgers.edu/~guyk/pub/1.pdf. [Online]. Available: http://www.crab.rutgers.edu/~guyk/pub/1.pdf

[20] M. Chlebík and J. Chlebíková, "Complexity of approximating bounded variants of optimization problems," *Theor. Comput. Sci.*, vol. 354, no. 3, pp. 320–338, 2006.

[21] A. Cohen and A. Wigderson, "Dispersers, deterministic amplification, and weak random sources (extended abstract)," in *FOCS*, 1989, pp. 14–19.

[22] M. Cygan, H. Dell, D. Lokshtanov, D. Marx, J. Nederlof, Y. Okamoto, R. Paturi, S. Saurabh, and M. Wahlström, "On problems as hard as cnf-sat," in *IEEE Conference on Computational Complexity*, 2012, pp. 74–84.

[23] M. Cygan, L. Kowalik, M. Pilipczuk, and M. Wykurz, "Exponential-time approximation of hard problems," *CoRR*, vol. abs/0810.4934, 2008.

[24] R. Diestel, *Graph Theory (Graduate Texts in Mathematics)*. Springer, August 2005.

[25] K. M. Elbassioni, R. Raman, S. Ray, and R. Sitters, "On the approximability of the maximum feasible subsystem problem with 0/1-coefficients," in *SODA*, 2009, pp. 1210–1219.

[26] B. Escoffier, E. J. Kim, and V. T. Paschos, "Subexponential and FPT-time inapproximability of independent set and related problems," *CoRR*, vol. abs/1211.6656, 2012.

[27] S. Even, O. Goldreich, S. Moran, and P. Tong, "On the NP-completeness of certain network testing problems," *Networks*, vol. 14, no. 1, pp. 1–24, 1984. [Online]. Available: http://dx.doi.org/10.1002/net.3230140102

[28] U. Feige, S. Goldwasser, L. Lovász, S. Safra, and M. Szegedy, "Interactive proofs and the hardness of approximating cliques," *J. ACM*, vol. 43, no. 2, pp. 268–292, 1996.

[29] M. R. Fellows, J. Guo, D. Marx, and S. Saurabh, "Data Reduction and Problem Kernels (Dagstuhl Seminar 12241)," *Dagstuhl Reports*, vol. 2, no. 6, pp. 26–50, 2012. [Online]. Available: http://drops.dagstuhl.de/opus/volltexte/2012/3729

[30] V. Guruswami, J. D. Hartline, A. R. Karlin, D. Kempe, C. Kenyon, and F. McSherry, "On profit-maximizing envy-free pricing," in *SODA*, 2005, pp. 1164–1173.

[31] J. Håstad, "Clique is hard to approximate within $n^{1\text{-epsilon}}$," in *FOCS*, 1996, pp. 627–636.

[32] J. Håstad and S. Khot, "Query efficient PCPs with perfect completeness," *Theory of Computing*, vol. 1, no. 1, pp. 119–148, 2005.

[33] R. Impagliazzo, R. Paturi, and F. Zane, "Which problems have strongly exponential complexity?" *J. Comput. Syst. Sci.*, vol. 63, no. 4, pp. 512–530, 2001.

[34] C. Joo, G. Sharma, N. B. Shroff, and R. R. Mazumdar, "On the complexity of scheduling in wireless networks," *EURASIP J. Wireless Comm. and Networking*, vol. 2010, 2010.

[35] R. Khandekar, T. Kimbrel, K. Makarychev, and M. Sviridenko, "On hardness of pricing items for single-minded bidders," in *APPROX-RANDOM*, 2009, pp. 202–216.

[36] R. Kumar, U. Mahadevan, and D. Sivakumar, "A graph-theoretic approach to extract storylines from search results," in *KDD*, 2004, pp. 216–225.

[37] D. Lokshtanov, D. Marx, and S. Saurabh, "Lower bounds based on the exponential time hypothesis," *Bulletin of the EATCS*, vol. 105, pp. 41–72, 2011.

[38] D. Marx, "Completely inapproximable monotone and antimonotone parameterized problems," *J. Comput. Syst. Sci.*, vol. 79, no. 1, pp. 144–151, 2013.

[39] N. Milosavljevic, "On complexity of wireless gathering problems on unit-disk graphs," in *ADHOC-NOW*, 2011, pp. 308–321.

[40] D. Moshkovitz and R. Raz, "Two-query PCP with subconstant error," *J. ACM*, vol. 57, no. 5, 2010.

[41] P. Popat and Y. Wu, "On the hardness of pricing loss-leaders," in *SODA*, 2012, pp. 735–749.

[42] O. Reingold, S. P. Vadhan, and A. Wigderson, "Entropy waves, the zig-zag graph product, and new constant-degree expanders and extractors," in *FOCS*, 2000, pp. 3–13.

[43] P. Rusmevichientong, S. U. D. of Management Science, and Engineering, *A Non-parametric Approach to Multi-product Pricing Theory and Application*. Stanford University, 2003. [Online]. Available: http://books.google.ca/books?id=KZf3GgAACAAJ

[44] P. Rusmevichientong, B. V. Roy, and P. W. Glynn, "A nonparametric approach to multiproduct pricing," *Operations Research*, vol. 54, no. 1, pp. 82–98, 2006.

[45] A. Samorodnitsky and L. Trevisan, "A PCP characterization of NP with optimal amortized query complexity," in *STOC*, 2000, pp. 191–199.

[46] M. Sipser, "Expanders, randomness, or time versus space," *J. Comput. Syst. Sci.*, vol. 36, no. 3, pp. 379–383, 1988.

[47] L. J. Stockmeyer and V. V. Vazirani, "NP-completeness of some generalizations of the maximum matching problem," *Inf. Process. Lett.*, vol. 15, no. 1, pp. 14–19, 1982.

[48] L. Trevisan, "Non-approximability results for optimization problems on bounded degree instances," in *STOC*, 2001, pp. 453–461.

[49] D. Zuckerman, "Linear degree extractors and the inapproximability of max clique and chromatic number," *Theory of Computing*, vol. 3, no. 1, pp. 103–128, 2007.