Tight Results on Minimum Entropy Set Cover

Jean Cardinal* Samuel Fiorini† Gwenaël Joret‡

Abstract

In the minimum entropy set cover problem, one is given a collection of k sets which collectively cover an n-element ground set. A feasible solution of the problem is a partition of the ground set into parts such that each part is included in some of the k given sets. Such a partition defines a probability distribution, obtained by dividing each part size by n. The goal is to find a feasible solution minimizing the (binary) entropy of the corresponding distribution. Halperin and Karp have recently proved that the greedy algorithm always returns a solution whose cost is at most the optimum plus a constant. We improve their result by showing that the greedy algorithm approximates the minimum entropy set cover problem within an additive error of 1 nat $= \log_2 e$ bits $\simeq 1.4427$ bits. Moreover, inspired by recent work by Feige, Lovász and Tetali on the minimum sum set cover problem, we prove that no polynomial-time algorithm can achieve a better constant, unless P = NP. We also discuss some consequences for the related minimum entropy coloring problem.

1 Introduction

Let V be an n-element ground set and $\mathscr{S} = \{S_1, \ldots, S_k\}$ be a collection of subsets of V whose union is V. A *cover* is an assignment $f : V \to \mathscr{S}$ of each point of V to a set of $\mathscr{S}$ such that $v \in f(v)$ for all $v \in V$. For each $i = 1, \ldots, k$, we let $q_i = q_i(f)$ denote the fraction of points assigned by f to the i-th set of $\mathscr{S}$, i.e.,

$$q_i := \frac{|f^{-1}(S_i)|}{n}. \qquad (1)$$

*Université Libre de Bruxelles, Département d'Informatique, c.p. 212, B-1050 Bruxelles, Belgium, jcardin@ulb.ac.be.

†Université Libre de Bruxelles, Département de Mathématique, c.p. 216, B-1050 Bruxelles, Belgium, sfiorini@ulb.ac.be.

‡Université Libre de Bruxelles, Département d'Informatique, c.p. 212, B-1050 Bruxelles, Belgium, gjoret@ulb.ac.be. G. Joret is a Research Fellow of the Fonds National de la Recherche Scientifique (FNRS).

The *minimum entropy set cover problem* (MESC) asks to find a cover f minimizing the entropy of the distribution $(q_1,\dots,q_k)$. Letting $\mathrm{ENT}(f)$ denote this latter quantity, we have

$$\mathrm{ENT}(f) := -\sum_{i=1}^{k} q_i \log q_i. \tag{2}$$

Note that, throughout, all logarithms are to base 2. Note also that, for definiteness, we set $x \log x = 0$ when $x = 0$.

The minimum entropy set cover problem is an NP-hard variant of the classical minimum cardinality set cover problem. Its recent introduction by Halperin and Karp [8] was motivated by various applications in computational biology. The problem is closely related to the minimum entropy coloring problem, which itself originates from the problem of source coding with side information in information theory, see Alon and Orlitsky [1].

The well-known greedy algorithm readily applies to MESC. It iteratively assigns to some set of $\mathscr{S}$ all unassigned points in that set, until all points are assigned. In each iteration, the algorithm chooses a set that contains a maximum number of unassigned points. Halperin and Karp [8] studied the performance of the greedy algorithm for MESC. They proved that the entropy of the cover returned by the algorithm is at most the optimum plus some constant[1]. Approximations within an additive error are considered because the entropy is a logarithmic measure. In the case of MESC, the optimum value always lies between 0 and $\log n$.

In this paper, we revisit the greedy algorithm and give a simple proof that it approximates MESC within 1 *nat*, that is, $\log e \simeq 1.4427$ bits. We then show that the problem is NP-hard to approximate to within $(1-\varepsilon)\log e$ for all positive ε. At the end of the paper, we discuss some consequences for the minimum entropy coloring problem.

At first sight, it might seem surprising that MESC can be approximated so well whereas its father problem, the minimum cardinality set cover problem, is notoriously difficult to approximate, see Feige [3]. We conclude the introduction by offering an intuitive explanation to this phenomenon. A consequential difference between the two problems is the penalty incurred for using too many sets. A minimum entropy cover is allowed to use a lot more sets than a minimum cardinality cover, provided the parts of these extra sets are small.

The same phenomenon also appears when one compares the minimum cardinality set cover problem to the *minimum sum set cover problem* (MSSC), see Feige, Lovász and Tetali [5]. The approximability status of the latter problem is similar

[1]They claim that the greedy algorithm gives a 3 bits approximation (which is correct). However, their proof is flawed (e.g., see their Lemma 6). A straightforward fix gives an approximation guarantee of $3 + 2\log e \simeq 5.8854$ bits.

to that of MESC: the greedy algorithm approximates it within a factor of 4 and achieving a factor of $4-\varepsilon$ is NP-hard, for all positive ε. Furthermore, the techniques used here for proving the corresponding results on MESC are comparable to the ones used in [5] for MSSC, especially for the inapproximability result.

2 Analysis of the Greedy Algorithm

We begin this section by exhibiting a family of instances on which the greedy algorithm perfoms poorly, namely, returns a solution whose cost exceeds the optimum by roughly $\log e$ bits. Below, we use the following bounds on the factorial. These bounds are implied by the more precise bounds given, e.g., in [6].

Lemma 1. *For any positive integer ℓ, we have*

$$\left(\frac{\ell}{e}\right)^{\ell} < \ell! < 2\sqrt{2\pi\ell}\left(\frac{\ell}{e}\right)^{\ell}.$$

Let ℓ be a positive integer. We let the points of V be the cells of a $\ell \times \ell!$ array and $\mathscr{S}$ be the union of two collections $\mathscr{S}_{col}$ and $\mathscr{S}_{line}$ each of which partitions V. The sets in $\mathscr{S}_{col}$ are the $\ell!$ columns of the array. For each $i = 1,\ldots,\ell$, collection $\mathscr{S}_{line}$ contains $\ell!/i$ sets of size i which partition the i-th line of the array. An illustration is given in Figure 1. (While in the figure each set of $\mathscr{S}_{line}$ consists of contiguous cells, we do not require this in general.) Each of the collections $\mathscr{S}_{col}$ and $\mathscr{S}_{line}$ directly yields a feasible solution for MESC, which we denote respectively by f_{col} and f_{line}. Clearly, f_{line} is one of the possible outcomes of the greedy algorithm (sets are produced from bottom to top on Figure 1).

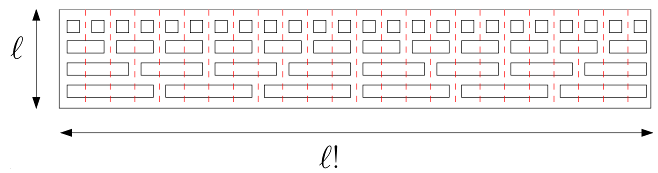

Figure 1: The sets forming $\mathscr{S}_{line}$

The respective costs of f_{col} and f_{line} are as follows:

$$\mathrm{ENT}(f_{col}) = -\sum_{j=1}^{\ell!} \frac{1}{\ell!}\log\frac{1}{\ell!} = \log \ell!,$$

$$\mathrm{ENT}(f_{line}) = -\sum_{i=1}^{\ell} \frac{\ell!}{i}\,\frac{i}{\ell\cdot\ell!}\log\frac{i}{\ell\cdot\ell!} = \log\ell + \log\ell! - \frac{1}{\ell}\log\ell!.$$

By the second inequality of Lemma 1, we then have

$$\mathrm{ENT}(f_{line}) \geq \log \ell + \log \ell! - \frac{1}{\ell} \log \left[2\sqrt{2\pi\ell}\left(\frac{\ell}{e}\right)^{\ell}\right] = \mathrm{ENT}(f_{col}) + \log e - o(1).$$

This implies that the cost of f_{line} is at least the optimum plus $\log e - o(1)$. We now show that the previous instances are essentially the worst for the greedy algorithm. Because the two formulations of MESC given above are equivalent to each other, we can regard a cover f as a partition of the ground set. Accordingly, we refer to the sets $f^{-1}(S_i)$ as the *parts* of f.

Theorem 1. *Let f_{OPT} and f_G be a cover of minimum entropy and a cover returned by the greedy algorithm, respectively. Then we have $ENT(f_G) \leq ENT(f_{OPT}) + \log e$.*

Proof. For $i = 1, \ldots, k$, we let X_i denote the i-th part of f_{OPT} and $x_i = |X_i|$. For $v \in V$, we let a_v be the size of the part of f_G containing v. We claim that the following holds for all v and all i:

$$\prod_{v \in X_i} a_v \geq x_i!. \tag{3}$$

Let us consider the points of X_i in the order in which they were assigned to sets of $\mathscr{S}$ by the greedy algorithm, breaking ties arbitrarily. Consider the j-th element of X_i assigned, say v. In the iteration when v was assigned, the greedy algorithm could have picked set S_i. Because at that time at most $j-1$ points of X_i were assigned, at least $x_i - j + 1$ points of S_i were unassigned, and we have $a_v \geq x_i - j + 1$. This implies the claim.

We now rewrite the entropy of f_G as follows:

$$\mathrm{ENT}(f_G) = -\frac{1}{n}\sum_{v \in V} \log \frac{a_v}{n} = -\frac{1}{n}\sum_{i=1}^{k}\sum_{v \in X_i} \log \frac{a_v}{n} = -\frac{1}{n}\sum_{i=1}^{k} \log \prod_{v \in X_i} \frac{a_v}{n}.$$

By Inequality (3) and the first inequality of Lemma 1, we then have:

$$\mathrm{ENT}(f_G) \leq -\frac{1}{n}\sum_{i=1}^{k} \log \frac{x_i!}{n^{x_i}} \leq -\frac{1}{n}\sum_{i=1}^{k} \log \frac{x_i^{x_i}}{n^{x_i} e^{x_i}} \leq \mathrm{ENT}(f_{OPT}) + \log e.$$

□

Finally, we mention that MESC has a natural weighted version in which each point $v \in V$ has some associated probability p_v. Again, we can associate to each cover f a probability distribution $(q_1, \ldots, q_k)$. This time, we let q_i denote the probability that a random point is assigned to S_i by f, that is,

$$q_i := \sum_{v \in f^{-1}(S_i)} p_v.$$

The goal is then to minimize (2), just as in the unweighted version. The greedy algorithm easily transposes to the weighted case, and so does our analysis. This is easily seen when the probabilities are rational. Indeed, let K be a positive integer such that Kp_v is integral for all points v. Now replicate each point in the ground set $Kp_v - 1$ times. Thus we obtain an unweighted instance which is equivalent to the original weighted instance, in the following sense. The optimum values of the two instances are equal (Lemma 2, given below, forbids replicated versions of a point to be assigned to different sets) and the behavior of the greedy algorithm on the new instance is identical to its behavior on the original instance. The case of real probabilities follows by a continuity argument.

3 Hardness of Approximation

Before turning to the main theorem of this section, we state a lemma which helps deriving good lower bounds on the optimum. Let $q = (q_i)$ and $r = (r_i)$ be two probability distributions over $\mathbb{N}^+$. If $\sum_{i=1}^{\ell} r_i \geq \sum_{i=1}^{\ell} q_i$ holds for all ℓ, we say that q is *dominated* by r. The lemma tells us that in such a case, the entropy of q is at least that of r, provided that q is non-increasing (see, e.g., [9] for a proof).

Lemma 2. *Let $q = (q_i)$ and $r = (r_i)$ be two probability distributions over $\mathbb{N}^+$ with finite support. Assume that q is non-increasing, that is, $q_i \geq q_{i+1}$ for $i \geq 1$. If q is dominated by r, then we have $ENT(q) \geq ENT(r)$.*

We now prove that no polynomial-time algorithm for MESC can achieve a better constant approximation guarantee than the greedy algorithm, unless P = NP. Halperin and Karp [8] gave a polynomial time approximation scheme (PTAS) for the problem. Our result does not contradict theirs since the PTAS they designed is multiplicative, i.e., returns a solution whose cost is most $(1-\varepsilon)$ times the optimum.

Theorem 2. *For every $\varepsilon > 0$, it is NP-hard to approximate the minimum entropy set cover problem within an additive term of $(1-\varepsilon)\log e$. This remains true on instances such that every point is in the same number of sets and every set has the same size.*

Proof. A *3SAT-6 formula* is a CNF formula in which every clause contains exactly three literals, every literal appears in exactly three clauses, and a variable appears at most once in each clause. Such a formula is said to be δ-*satisfiable* if at most a δ-fraction of its clauses are satisfiable. It is known that distinguishing between a satisfiable 3SAT-6 formula and one which is δ-satisfiable is NP-hard for some δ with $0 < \delta < 1$, see Feige et al. [5]. In the latter reference, the authors slightly modified a reduction due to Feige [3] to design a polynomial-time reduction associating to any 3SAT-6 formula φ a corresponding set system $\mathbf{S}(\varphi) = (V, \mathscr{S})$. They used the new reduction to prove that the minimum sum set cover problem is NP-hard to approximate to within $2 - \varepsilon$ on uniform regular hypergraphs (see Theorem 12 in that paper). For any given constants $c > 0$ and $\lambda > 0$, it is possible to set the values of the parameters of the reduction in such a way that:

- the sets of $\mathscr{S}$ have all the same size n/t, where n denotes the number of points in V, and every point of V is contained in the same number of sets;
- if φ is satisfiable, then V can be covered by t disjoint sets of $\mathscr{S}$;
- if φ is δ-satisfiable, then every i sets chosen from $\mathscr{S}$ cover at most a $1 - (1 - 1/t)^i + \lambda$ fraction of the points of V, for $1 \leq i \leq ct$.

Suppose from now on that φ is a 3SAT-6 formula which is either satifiable or δ-satisfiable, and denote by f_{OPT} an optimal solution of MESC with input $\mathbf{S}(\varphi)$. For $1 \leq i \leq k$, let $q_i = q_i(f_{OPT})$ be defined as in (1). For $i > k$, we let $q_i = 0$. Letting q denote the sequence (q_i), we assume without loss of generality that q is non-increasing.

If φ is satisfiable, then it follows from Lemma 2 that the optimal solution consists in covering V with t disjoint sets. Hence, $\mathrm{ENT}(f_{OPT}) = \mathrm{ENT}(q) = \log t$ in this case. Assume now that φ is δ-satisfiable. Let $\alpha = \varepsilon/2$, $\lambda = \alpha^2/2 - \alpha^3/6$ and $c = -\ln \lambda$.

Claim 1. *The following lower bound on the optimum holds:*

$$ENT(q) \geq \log t + (1 - \varepsilon/2) \log e + o(1),$$

where $o(1)$ tends to zero when t tends to infinity.

Claim 1 implies that any algorithm approximating MESC within an additive term of $(1 - \varepsilon) \log e$ can be used to decide whether φ is satisfiable or δ-satisfiable. Indeed, as noted in [5], t may be assumed to be larger than any fixed constant. The theorem then follows.

In order to prove the claim, we define a sequence $r = (r_i)$ as follows (see Figure 2 for an illustration):

$$r_i = \begin{cases} 1/t & \text{for } 1 \le i \le \lceil \alpha t \rceil, \\ (1-1/t)^{i-1}/t & \text{for } \lceil \alpha t \rceil + 1 \le i \le \lfloor \tilde{c} t \rfloor, \\ 1 - \sum_{i=1}^{\lfloor \tilde{c} t \rfloor} r_i & \text{for } i = \lfloor \tilde{c} t \rfloor + 1, \\ 0 & \text{otherwise,} \end{cases}$$

where $\tilde{c}$ is a real such that

$$\frac{\lceil \alpha t \rceil}{t} + (1-1/t)^{\lceil \alpha t \rceil} - (1-1/t)^{\tilde{c} t} = 1. \tag{4}$$

By our choice of parameters, we can assume $\lceil \alpha t \rceil + 1 \le \lfloor \tilde{c} t \rfloor$ by lowering ε if necessary. From the definition of $\tilde{c}$ we have

$$\sum_{i=1}^{\lfloor \tilde{c} t \rfloor} r_i = \frac{\lceil \alpha t \rceil}{t} + (1-1/t)^{\lceil \alpha t \rceil} - (1-1/t)^{\lfloor \tilde{c} t \rfloor} \le 1.$$

Therefore, the sequence r is a probability distribution over $\mathbb{N}^+$.

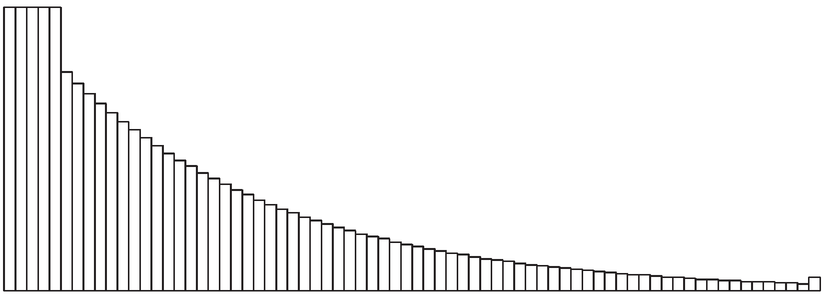

Figure 2: The shape of distribution $r = (r_i)$ for $t = 20$ and $\varepsilon = 1/2$

By the properties of $\mathbf{S}(\varphi)$ we have

$$\sum_{i=1}^{\ell} q_i \le \ell/t \quad \text{and} \quad \sum_{i=1}^{\ell} q_i \le 1 - (1-1/t)^{\ell} + \lambda \tag{5}$$

for $1 \le \ell \le \lfloor ct \rfloor$, and it can be checked that $\tilde{c} \le c$ for t large enough.

Claim 2. *Sequence q is dominated by sequence r, that is, for all ℓ we have*

$$\sum_{i=1}^{\ell} q_i \le \sum_{i=1}^{\ell} r_i. \tag{6}$$

For $1 \le \ell \le \lceil \alpha t \rceil$, Inequality (6) readily follows from the definition of r and Equation (5). Notice that we have

$$1-(1-1/t)^{\lceil \alpha t \rceil}+\lambda \le 1-(1-\alpha+\alpha^2/2-\alpha^3/6)+\lambda = \alpha \le \lceil \alpha t \rceil / t \quad (7)$$

whenever t is large enough. Hence, for $\lceil \alpha t \rceil + 1 \le \ell \le \lfloor \tilde{c} t \rfloor$, from Equations (5) and (7) we derive

$$\begin{aligned}\sum_{i=1}^{\ell} q_i &\le 1-(1-1/t)^{\ell}+\lambda = 1-(1-1/t)^{\lceil \alpha t \rceil}+\lambda+\sum_{i=\lceil \alpha t \rceil+1}^{\ell}(1-1/t)^{i-1}/t \\ &\le \lceil \alpha t \rceil/t+\sum_{i=\lceil \alpha t \rceil+1}^{\ell}(1-1/t)^{i-1}/t = \sum_{i=1}^{\ell} r_i.\end{aligned}$$

Finally, note that (6) is also true for $\ell > \lfloor \tilde{c} t \rfloor$, as the q_i's and r_i's both sum up to 1. It follows that q is dominated by r. In other words, Claim 2 holds true. By Lemma 2, we have $\mathrm{ENT}(q) \ge \mathrm{ENT}(r)$. In order to show Claim 1, it then suffices to prove the following claim.

Claim 3. *We have* $ENT(r) \ge \log t + (1-\varepsilon/2)\log e + o(1)$.

The entropy of r can be expressed as follows:

$$\begin{aligned}\mathrm{ENT}(r) &= -\sum_{i=1}^{\lfloor \tilde{c} t \rfloor+1} r_i \log r_i = -\sum_{i=1}^{\lfloor \tilde{c} t \rfloor} r_i \log r_i + o(1) \\ &= \frac{\lceil \alpha t \rceil}{t}\log t - \sum_{i=\lceil \alpha t \rceil+1}^{\lfloor \tilde{c} t \rfloor} \frac{(1-1/t)^{i-1}}{t}\log\frac{(1-1/t)^{i-1}}{t}+o(1) \\ &= \alpha \log t + \frac{1}{t}\log\frac{t}{t-1}\sum_{i=\lceil \alpha t \rceil+1}^{\lfloor \tilde{c} t \rfloor}(i-1)(1-1/t)^{i-1} \\ &\quad + \frac{1}{t}\log t \sum_{i=\lceil \alpha t \rceil+1}^{\lfloor \tilde{c} t \rfloor}(1-1/t)^{i-1}+o(1).\end{aligned}$$

Let $\beta := \lim_{t\to\infty} \tilde{c}$. In the sum above, the second and third terms are asymptotically equal to respectively $\log e \cdot ((1+\alpha)e^{-\alpha}-(1+\beta)e^{-\beta})$ and $\log t \cdot (e^{-\alpha}-e^{-\beta})$. This is shown in Lemmas 3 and 4 in the appendix. It follows from Equation (4) that

$$\beta = -\ln(\alpha+e^{-\alpha}-1).$$

In virtue of this equation and by what precedes, we can rewrite the entropy of r as

$$\begin{aligned}\mathrm{ENT}(r) &= \alpha \log t + \log e \cdot ((1+\alpha)e^{-\alpha} - (1+\beta)e^{-\beta}) + \log t \cdot (e^{-\alpha} - e^{-\beta}) + o(1) \\ &= (\alpha + e^{-\alpha} - e^{-\beta}) \log t + ((1+\alpha)e^{-\alpha} - (1+\beta)e^{-\beta}) \log e + o(1) \\ &= \log t + ((1+\alpha)e^{-\alpha} - (1+\beta)e^{-\beta}) \log e + o(1).\end{aligned}$$

By Lemma 5 (see the appendix), we know that $\alpha e^{-\alpha} - \beta e^{-\beta}$ is nonnegative provided ε is sufficiently small. Claim 3 follows then by noticing

$$(1+\alpha)e^{-\alpha} - (1+\beta)e^{-\beta} = 1 - \alpha + \alpha e^{-\alpha} - \beta e^{-\beta} \geq 1 - \alpha = 1 - \varepsilon/2.$$

Hence, Claim 1 and the theorem follow. □

4 Graph Colorings with Minimum Entropy

There are situations where the collection of sets $\mathscr{S} = \{S_1, \ldots, S_k\}$ input to the minimum entropy set cover problem is given implicitly. One possibility, which is the focus of this section, is to define $\mathscr{S}$ as the collection of all inclusion-wise maximal stable sets of some (simple, undirected) graph $G = (V, E)$. The corresponding variant of MESC is known as the *minimum entropy coloring problem* (MEC). It stems from information theory, having applications in zero-error coding with side information [1]. Notice that, by our choice of $\mathscr{S}$, every cover f can be regarded as a (proper) coloring of the graph G.

The results of Section 2 directly apply to MEC. The greedy algorithm, transposed to the setting of MEC, constructs a coloring of G by iteratively removing a maximum size stable set from G. Of course, its running time can no longer be guaranteed to be polynomial, unless P = NP. Theorem 1 implies the following result, which again holds in the weighted case.

Corollary 1. *Let f_{OPT} and f_G be a coloring of G with minimum entropy and a coloring returned by the greedy algorithm, respectively. Then we have $ENT(f_G) \leq ENT(f_{OPT}) + \log e$.*

The bound given in Corollary 1 is asymptotically tight because the bad MESC instances described in the beginning of Section 2 can be easily turned into MEC instances. Indeed, for a given ℓ, it suffices to consider the graph G obtained from the complete graph on V by removing every edge which is entirely included in some set of $\mathscr{S}_{col}$ or $\mathscr{S}_{line}$.

Clearly, the greedy algorithm runs in polynomial time when restricted to graphs in which a maximum weight stable set can be found in polynomial time. This includes perfect graphs [7] and claw-free graphs [10]. So MEC can be approximated within an additive term of $\log e$ on such graphs, in polynomial time. In contrast, for arbitrary graphs it is known that for any $\varepsilon > 0$ there is no polynomial-time approximation algorithm whose additive error is bounded by $(1-\varepsilon)\log n$ unless ZPP=NP. This was proved by the authors in [2] using as a black-box an inapproximability result for the minimum cardinality coloring problem due to Feige and Kilian [4].

References

[1] N. Alon and A. Orlitsky. Source coding and graph entropies. *IEEE Trans. Inform. Theory*, 42(5):1329–1339, September 1996.

[2] J. Cardinal, S. Fiorini, and G. Joret. Minimum entropy coloring. In *Proceedings of the 16th International Symposium on Algorithms and Computation (ISAAC 2005)*, volume 3827 of *Lecture Notes in Computer Science*, pages 819–828, Berlin, 2005. Springer.

[3] U. Feige. A threshold of $\ln n$ for approximating set cover. *J. ACM*, 45(4): 634–652, 1998.

[4] U. Feige and J. Kilian. Zero knowledge and the chromatic number. *J. Comput. System Sci.*, 57(2):187–199, 1998. Complexity 96—The Eleventh Annual IEEE Conference on Computational Complexity (Philadelphia, PA).

[5] U. Feige, L. Lovász, and P. Tetali. Approximating min sum set cover. *Algorithmica*, 40(4):219–234, 2004.

[6] W. Feller. *An introduction to probability theory and its applications. Vol. I*. Third edition. John Wiley & Sons Inc., New York, 1968.

[7] M. Grötschel, L. Lovász, and A. Schrijver. *Geometric algorithms and combinatorial optimization*, volume 2 of *Algorithms and Combinatorics*. Springer-Verlag, Berlin, second edition, 1993.

[8] E. Halperin and R. M. Karp. The minimum-entropy set cover problem. *Theoret. Comput. Sci.*, 348(2-3):240–250, 2005.

[9] G. H. Hardy, J. E. Littlewood, and G. Pólya. *Inequalities*. Cambridge Mathematical Library. Cambridge University Press, Cambridge, 1988. Reprint of the 1952 edition.

[10] D. Nakamura and A. Tamura. A revision of Minty's algorithm for finding a maximum weight stable set of a claw-free graph. *J. Oper. Res. Soc. Japan*, 44 (2):194–204, 2001.

Appendix

Lemma 3. *If α, β and $\tilde{c}$ are defined as in the proof of Theorem 2, we have*

$$\frac{1}{t}\log\frac{t}{t-1}\sum_{i=\lceil\alpha t\rceil+1}^{\lfloor\tilde{c}t\rfloor}(i-1)(1-1/t)^{i-1}=\log e\cdot((1+\alpha)e^{-\alpha}-(1+\beta)e^{-\beta})+o(1).$$

Proof. For any positive integer ℓ, we have

$$\begin{aligned}\sum_{i=1}^{\ell}(i-1)(1-1/t)^{i-1}&=\sum_{i=1}^{\ell}i(1-1/t)^{i-1}-\sum_{i=1}^{\ell}(1-1/t)^{i-1}\\&=t^2\big(1-\frac{1}{t}(\ell+1)(1-1/t)^{\ell}-(1-1/t)^{\ell+1}\big)\\&\quad-t(1-(1-1/t)^{\ell})\\&=t^2\big(1-(1+\ell/t)(1-1/t)^{\ell}\big)-t(1-(1-1/t)^{\ell}).\end{aligned}$$

It follows that we have

$$\begin{aligned}&\frac{1}{t}\log\frac{t}{t-1}\sum_{i=\lceil\alpha t\rceil+1}^{\lfloor\tilde{c}t\rfloor}(i-1)(1-1/t)^{i-1}\\&\quad=\frac{1}{t}\log\frac{t}{t-1}\Big[t^2\big(1-(1+\lfloor\tilde{c}t\rfloor/t)(1-1/t)^{\lfloor\tilde{c}t\rfloor}\big)-t(1-(1-1/t)^{\lfloor\tilde{c}t\rfloor})\\&\qquad-t^2\big(1-(1+\lceil\alpha t\rceil/t)(1-1/t)^{\lceil\alpha t\rceil}\big)+t(1-(1-1/t)^{\lceil\alpha t\rceil})\Big].\end{aligned}$$

The right-hand side of the latter equation can be rewritten as:

$$\begin{aligned}&t\log\frac{t}{t-1}\big((1+\lceil\alpha t\rceil/t)(1-1/t)^{\lceil\alpha t\rceil}-(1+\lfloor\tilde{c}t\rfloor/t)(1-1/t)^{\lfloor\tilde{c}t\rfloor}\big)\\&+\log\frac{t}{t-1}\big((1-1/t)^{\lfloor\tilde{c}t\rfloor}-(1-1/t)^{\lceil\alpha t\rceil}\big).\end{aligned}\tag{8}$$

Now, from $\tilde{c}t-1\le\lfloor\tilde{c}t\rfloor\le\tilde{c}t$, we infer:

$$(1-1/t)^{\tilde{c}t}\le(1-1/t)^{\lfloor\tilde{c}t\rfloor}\le(1-1/t)^{\tilde{c}t-1}.$$

Because $\tilde{c}t=\beta t+o(t)$ and $(1-1/t)^{o(t)}=1+o(1)$, the above upper and lower bound on $(1-1/t)^{\lfloor\tilde{c}t\rfloor}$ are asymptotically equal to $e^{-\beta}$. Hence, we have $(1-1/t)^{\lfloor\tilde{c}t\rfloor}=e^{-\beta}+o(1)$. A similar argument shows $(1-1/t)^{\lceil\alpha t\rceil}=e^{-\alpha}+o(1)$. Hence, when t tends to ∞, the first term of (8) tends to $\log e\cdot((1+\alpha)e^{-\alpha}-(1+\beta)e^{-\beta})$, and the second term tends to 0. The lemma follows. □

Lemma 4. *With α, β and $\tilde{c}$ defined as in the proof of Theorem 2, we have*

$$\frac{1}{t}\log t \sum_{i=\lceil \alpha t\rceil+1}^{\lfloor \tilde{c}t\rfloor} (1-1/t)^{i-1} = \log t \cdot (e^{-\alpha} - e^{-\beta}) + o(1).$$

Proof. Because we have

$$\sum_{i=\lceil \alpha t\rceil+1}^{\lfloor \tilde{c}t\rfloor} (1-1/t)^{i-1} = t((1-1/t)^{\lceil \alpha t\rceil} - (1-1/t)^{\lfloor \tilde{c}t\rfloor}),$$

it suffices to prove the following equalities:

$$\log t \cdot ((1-1/t)^{\lceil \alpha t\rceil} - e^{-\alpha}) = o(1) \quad \text{and} \tag{9}$$

$$\log t \cdot ((1-1/t)^{\lfloor \tilde{c}t\rfloor} - e^{-\beta}) = o(1). \tag{10}$$

For all t, we have

$$e^{-1/t} - 1/2t^2 \leq 1 - 1/t \leq e^{-1/t}$$

and hence

$$(e^{-1/t} - 1/2t^2)^{\lceil \alpha t\rceil} \leq (1-1/t)^{\lceil \alpha t\rceil} \leq e^{-\lceil \alpha t\rceil/t}. \tag{11}$$

Using the Binomial theorem, we can bound the left-hand side of (11) as follows:

$$\begin{aligned}
(e^{-1/t} - 1/2t^2)^{\lceil \alpha t\rceil} &= \sum_{i=0}^{\lceil \alpha t\rceil} \binom{\lceil \alpha t\rceil}{i} e^{-(\lceil \alpha t\rceil - i)/t}(-1/2t^2)^i \\
&\geq e^{-\lceil \alpha t\rceil/t} + \sum_{\substack{i=0 \\ i \text{odd}}}^{\lceil \alpha t\rceil} \binom{\lceil \alpha t\rceil}{i} e^{-(\lceil \alpha t\rceil - i)/t}(-1/2t^2)^i \\
&\geq e^{-\lceil \alpha t\rceil/t} + \sum_{j=0}^{\infty} (\alpha t+1)(\alpha t)^{2j}(-1/2t^2)^{2j+1} \\
&\geq e^{-\lceil \alpha t\rceil/t} - (\alpha/2t + 1/2t^2) \cdot \frac{1}{1-\alpha^2/4t^2}.
\end{aligned} \tag{12}$$

Equations (11) and (12) together yield (recall that t can be assumed to be large, so $\log t$ is positive):

$$\begin{aligned}
&\log t \cdot (e^{-\lceil \alpha t\rceil/t} - e^{-\alpha}) - \log t \cdot (\alpha/2t + 1/2t^2) \cdot \frac{1}{1-\alpha^2/4t^2} \\
&\leq \log t \cdot ((1-1/t)^{\lceil \alpha t\rceil} - e^{-\alpha}) \leq \log t \cdot (e^{-\lceil \alpha t\rceil/t} - e^{-\alpha}).
\end{aligned} \tag{13}$$

From the implications

$$\alpha t \leq \lceil \alpha t\rceil \leq \alpha t + 1 \Rightarrow \alpha \leq \lceil \alpha t\rceil/t \leq \alpha + 1/t \Rightarrow e^{-\alpha-1/t} \leq e^{-\lceil \alpha t\rceil/t} \leq e^{-\alpha},$$

we infer:

$$e^{-\alpha}\log t\cdot(e^{-1/t}-1)\le \log t\cdot(e^{-\lceil \alpha t\rceil/t}-e^{-\alpha})\le 0.$$

Since $\log t\cdot(e^{-1/t}-1)$ is asymptotically zero, both the upper and lower bound in Equation (13) are asymptotically zero, and Equation (9) follows.

Finally, in order to prove that Equation (10) holds, we start from the following bounds on $(1-1/t)^{\lfloor\tilde{c}t\rfloor}$:

$$(1-1/t)^{\tilde{c}t}\le(1-1/t)^{\lfloor\tilde{c}t\rfloor}\le(1-1/t)^{\tilde{c}t-1}=\frac{t}{t-1}(1-1/t)^{\tilde{c}t}.$$

Note that we have $(1-1/t)^{\tilde{c}t}=\lceil\alpha t\rceil/t+(1-1/t)^{\lceil\alpha t\rceil}-1$ by the definition of $\tilde{c}$, see Equation (4). Therefore, we can bound $\log t\cdot((1-1/t)^{\lfloor\tilde{c}t\rfloor}-e^{-\beta})$ as follows (recall that $e^{-\beta}=\alpha+e^{-\alpha}-1$):

$$\begin{aligned}&\log t\cdot(\lceil\alpha t\rceil/t+(1-1/t)^{\lceil\alpha t\rceil}-\alpha-e^{-\alpha})\\&\le\log t\cdot((1-1/t)^{\lfloor\tilde{c}t\rfloor}-e^{-\beta})\le\\&\log t\cdot\frac{t}{t-1}\big(\lceil\alpha t\rceil/t+(1-1/t)^{\lceil\alpha t\rceil}-1-\frac{t-1}{t}\alpha-\frac{t-1}{t}e^{-\alpha}+\frac{t-1}{t}\big).\end{aligned}$$

By Equation (9), both bounds are asymptotically zero. Therefore Equation (10) holds. This concludes the proof. □

Lemma 5. *Letting $\alpha=\alpha(\varepsilon)$ and $\beta=\beta(\varepsilon)$ be defined as in the proof of Theorem 2, we have $\alpha e^{-\alpha}\ge\beta e^{-\beta}$ provided that ε is small enough.*

Proof. Because $\beta=-\ln(\alpha+e^{-\alpha}-1)$, we have to show

$$\alpha e^{-\alpha}\ge-\ln(\alpha+e^{-\alpha}-1)\cdot(\alpha+e^{-\alpha}-1). \tag{14}$$

For all α, we have

$$\frac{\alpha^2}{2}(1-\frac{\alpha}{3})=\frac{\alpha^2}{2}-\frac{\alpha^3}{6}\le\alpha+e^{-\alpha}-1\le\frac{\alpha^2}{2}$$

and hence

$$-\big(\ln\frac{\alpha^2}{2}+\ln(1-\frac{\alpha}{3})\big)\cdot\frac{\alpha^2}{2}\ge-\ln(\alpha+e^{-\alpha}-1)\cdot(\alpha+e^{-\alpha}-1).$$

Now, as can be readily checked, for ε sufficiently small (recall that $\alpha=\varepsilon/2$), we have

$$\alpha e^{-\alpha}\ge\frac{\alpha}{2}\ge-\big(\ln\frac{\alpha^2}{2}+\ln(1-\frac{\alpha}{3})\big)\cdot\frac{\alpha^2}{2}.$$

The lemma follows.

□