A Matched Approximation Bound for the Sum of a Greedy Coloring

Amotz Bar-Noy* Magnús M. Halldórsson† Guy Kortsarz‡

June 16, 1998

Abstract

In the minimum sum coloring problem, the goal is to color the vertices of a graph with the positive integers such that the sum of all colors is minimal. Recently, it was shown that coloring a graph by iteratively coloring maximum independent sets yields a $4 + o(1)$ approximation for the minimum sum coloring problem. In this note, we show that this bound is tight. We construct a graph for which the approximation ratio of this coloring is $4 - o(1)$.

Keywords: Analysis of algorithms, graph coloring, maximum independent set, sum coloring.

1 Introduction

Let G be an undirected simple graph with n vertices. A coloring of the vertices of G is a mapping into the set of positive integers, $f : G \mapsto \mathcal{Z}^+$, such that adjacent vertices are assigned different colors. In the *minimum sum coloring* problem (`MSC`), we are looking for a coloring in which the sum of the assigned colors of all the vertices of G is minimized. That is, the value of $\sum_{v \in G} f(v)$ is minimized.

The `MSC` problems has a natural application in scheduling theory. Consider scheduling n unit-time jobs on a single machine. At any given time the machine is capable to perform any number of tasks, as long as these tasks are independent. We model this by a graph $G(V, E)$. The jobs correspond to the graph vertices, and conflicting jobs are joined by an edge. A coloring of the graph corresponds to a schedule, where the independent set of color class i corresponds to the collection of jobs that are executed at time i. When the goal is to minimize the maximum completion time of a job this corresponds to the well known minimum graph-coloring problem. This is an optimization goal that favors the system. However, from the point of view of the jobs

*Department of Electrical Engineering, Tel-Aviv University, Tel-Aviv 69978, Israel. E-mail: `amotz@eng.tau.ac.il`.

†Science Institute, University of Iceland, IS-107 Reykjavik, Iceland. E-mail: `mmh@rhi.hi.is`.

‡Department of Computer Science, The Open University of Israel, Ramat Aviv, Israel. E-mail: `guyk@tavor.openu.ac.il`.

themselves, we may wish to find the best coloring such that the average waiting time to be served is minimized. This corresponds to finding a coloring that minimizes the sum of all colors.

A natural greedy coloring algorithm MaxIS is the following. Pick a maximum independent set $I_1 \subseteq V$. Color the vertices of I_1 by color 1, and iterate. In the ith iteration, pick the maximum independent set I_i in the graph induced by the yet uncolored vertices $(V - \cup_{1 \leq j \leq i-1} I_j)$, and color the vertices of I_i by color i. The process terminates when all vertices are colored.

In [1] it was proven that MaxIS is a $4 + o(1)$ approximation algorithm for the `MSC` problem. That is, for any graph G, MaxIS produces a coloring whose sum coloring is at most $4 + o(1)$ times the optimal sum coloring of G. This immediately gives a uniform polynomial time approximation algorithm with ratio $4 + o(1)$ for the `MSC` problem on all graphs where the maximum independent set can be found in polynomial time (e.g.., perfect graphs). In addition, [1] presented a lower bound of $2 - o(1)$. A graph was constructed for which the approximation ratio of MaxIS is almost twice the optimum. In this note, we close the gap and show that the upper bound is tight. We construct a graph for which the approximation ratio of MaxIS is $4 - o(1)$.

Related work: The minimum sum coloring problem was introduced in [3]. In [5], it was shown that computing the `MSC` of a given graph is $NP-$hard and a polynomial time optimal algorithm was presented for trees. In [4], it was shown that approximating the `MSC` within an additive constant factor is $NP-$hard. It was also shown there that the first-fit algorithm yields a $(\bar{d}/2 + 1)-$approximation for graphs of average degree $\bar{d}$. Lower and upper bounds on the value of the sum coloring in general graphs were given in [7]. In [1], it was proven that there is no $n^{1-\epsilon}-$ratio approximation algorithm for the `MSC` problem for any $\epsilon > 0$ unless $NP = ZPP$. In [2], the problem was proven to be $MAXSNP-$hard on bipartite graphs and a 10/9-approximation algorithm was presented. In [6], a 2-approximation algorithm was presented for the `MSC` problem on interval graphs.

Future work: It is an interesting open problem to determine the exact performance ratio of MaxIS for `MSC` on some graph families. For example, the algorithm has tight performance ratio of 4/3 on bipartite graphs [1]. For interval graphs, we have examples that show that MaxIS has ratio worse than 1.5 (details omitted), while we do not know of a better upper bound than 4. We conjecture that the approximation factor of MaxIS on interval graphs is 2.

2 The chopping procedure

We represent a coloring of a graph G by k colors as a tuple of length k: $\langle c_1, \ldots, c_k \rangle$. The size of the set C_i of the vertices colored by i is c_i. Note that without loss of generality, in any solution to the sum coloring problem we have $c_1 \geq c_2 \geq \cdots \geq c_k$. Otherwise we could switch color classes and get a better sum coloring. By definition, for a given pattern $P = \langle c_1, \ldots, c_k \rangle$ the sum coloring is

$$SC(P) = \sum_{i=1}^{k} i \cdot c_i \ .$$

Given a pattern $P = \langle c_1, \ldots, c_k \rangle$, we describe a *chopping* procedure which constructs a graph G_P with $n = \sum_{i=1}^{k} c_i$ vertices. In each step, we observe the minimum size for a maximum independent set. Then the chopping procedure forces MaxIS to pick a maximum independent set of this size such that in later steps MaxIS will find smaller sets.

In G_P, there are k independent sets $C_1, \ldots, C_k$ that cover all the vertices of the graph such that $|C_i| = c_i$. Hence, there exists a coloring whose representation is $\langle c_1, \ldots, c_k \rangle$. We now place the vertices of the graph in a matrix of size $c_1 \times k$. The ith column contains c_i ones at the bottom and $c_1 - c_i$ zeros at the top. Each vertex is now associated with a one entry in the matrix.

The chopping procedure first constructs an independent set I_1 of size c_1. It collects the vertices from the matrix line after line from the top line to the bottom line. In each line, it collects the vertices from right to left. Each one entry that is collected becomes a zero entry. Then it adds edges from I_1 to all the other vertices as long as these edges do not connect two vertices from the same column. In a same manner the procedure constructs I_2. The size of I_2 is the number of ones in the first column after the first step. At the beginning of the ith step, the procedure already constructed $I_1, \ldots, I_{i-1}$, defined all the edges incident to these vertices, and replaced all the one entries associated with the vertices of the sets $I_1, \ldots, I_{i-1}$ by zeros. During the ith step, the chopping procedure constructs in a similar manner the independent set I_i the size of which is the number of ones in the first column of the matrix at the beginning of the step. Again, each one entry that is collected becomes a zero entry. Then the procedure connects the vertices of I_i with the remaining vertices in the matrix as long as these edges do not connect two vertices from the same row. The procedure terminates after h steps when the matrix contains only zeros.

In the resulting graph, two vertices are connected unless they both belong to some C_i for $1 \le i \le k$ or both belong to some I_j for $1 \le j \le h$. Moreover, these $k + h$ sets are the only maximal independent sets in the graph. To see the last claim, observe first that due to the way I_j was constructed it follows that $C_i = I_j$ only if they contain one vertex that is connected to all the other vertices in the graph. Next observe that if $|I_j| \ge 2$, then there exist $u \in I_j \cap C_{i_1}$ and $v \in I_j \cap C_{i_2}$ for some $i_1 \ne i_2$. Therefore we cannot add another vertex to I_j because this vertex is either connected to u or to v. A similar argument could be applied for the case $|C_i| \ge 2$.

We need to show that I_j is one of the maximum independent sets in the graph induced by the vertices $\cup_{j \le \ell \le h} I_\ell$ denoted by G_j. To see this, observe that because $|C_1| \ge |C_2| \ge \cdots \ge |C_k|$ it follows that $|I_1| \ge |I_2| \ge \cdots \ge |I_h|$. This and the maximality of the independent sets $C_1, \ldots, C_k, I_1, \ldots, I_h$ imply that the maximum independent set in G_j is either I_j or $C_1 \cap G_j$. The claim follows since I_j was chosen to be of the same size as $C_1 \cap G_j$.

After each step of the chopping procedure, the remaining ones in the matrix could be represented by $P_i = \langle c_1^i, \ldots, c_k^i \rangle$ where P_0 is the original pattern P. With this notation, in the ith step, the procedure creates the independent set I_i of size c_1^{i-1}. Note that since the first row represents at this stage an independent set of size c_1^{i-1}, MaxIS must find an independent set of at least this size. The chopping algorithm forces it to choose an "horizontal" set in order to force smaller size sets in the next steps.

The chopping procedure creates another partition of G into independent sets the pattern of which is $A(P) = \langle |I_1|, \ldots, |I_h| \rangle$ for some $h \geq k$. The pattern $A(P)$ is lexicographically less than the pattern P. Therefore, the sum coloring associated with $A(P)$ is greater than the one associated with P. In the above notations, $A(P) = \left\langle c_1^0, \ldots, c_1^{h-1} \right\rangle$.

We sometimes terminate the chopping procedure before the matrix contains only zeros. At some stage, we take as maximum independent sets the k columns from left to right. Assume that at the end of the ith step we already constructed the pattern $\left\langle c_1^0, \ldots, c_1^{i-1} \right\rangle$. Then the final pattern will be $A(P) = \left\langle c_1^0, \ldots, c_1^{i-1}, c_1^i, \ldots, c_k^i \right\rangle$.

Example: Suppose $P = \langle 8, 8, 1 \rangle$. Then I_1 contains 4 vertices from C_1 and 4 vertices from C_2. We are left with a pattern $\langle 4, 4, 1 \rangle$. Then I_2 contains 2 vertices from C_1 and 2 vertices from C_2 and we are left with the pattern $\langle 2, 2, 1 \rangle$. Then I_3 contains 2 vertices and I_4, I_5, and I_6 each contains one vertex. Therefore $A(P) = \langle 8, 4, 2, 1, 1, 1 \rangle$.

In the next three sections, we demonstrate three lower bounds that apply the chopping procedure technique.

3 The $2 - o(1)$ Lower bound

The first lower bound, given in [1], assumes that there are k color classes all of them of size x. Therefore, they are represented by the pattern $LB2 = \langle x, x, \ldots, x \rangle$ where x appears k times. It follows that

$$SC(LB2) = \sum_{i=1}^{k} ix = \frac{k(k+1)}{2}x \ .$$

In the chopping procedure, after choosing I_1 of size x in the first step, we are left with the pattern $\left\langle \frac{k-1}{k}x, \ldots, \frac{k-1}{k}x \right\rangle$ assuming k divides x. Define $q = \frac{k-1}{k}$. In the ith step, I_i is of size $q^{i-1}x$ and we are left with the pattern $\langle q^i x, \ldots, q^i x \rangle$ assuming k^i divides x. Choosing x large enough, we approximate $A(LB2)$ by the infinite pattern $A'(LB2) = \langle x, qx, \ldots, q^i x, \ldots \rangle$. We claim that $SC(A'(LB2)) - SC(A(LB2)) < o(k^2 x)$. The sum coloring of $A'(LB2)$ is:

$$SC(A'(LB2)) = \sum_{i=0}^{\infty} ((i+1)q^i)x = \frac{x}{(1-q)^2} = k^2 x \ .$$

It follows that the approximation ratio of MaxIS is:

$$r = \frac{(k^2 x)(1 - o(1))}{\frac{k(k+1)}{2}x} = 2 - o(1) \ .$$

4 The $3 - o(1)$ Lower bound

To get a better lower bound, we need to construct a more complicated pattern. The idea is to let the first two entries in the pattern to be equal, and to ensure that the first three entries be equal after the first chopping step. In general, before the ith step of the chopping procedure, we want the first $i+1$ entries in the pattern to be equal.

Small examples of such patterns are $\langle 2,2,1\rangle$ for $k=3$ and $\langle 6,6,3,2\rangle$ for $k=4$. Note that for $k=3$, MaxIS produces the pattern $\langle 2,1,1,1\rangle$ and already the approximation ratio is at least $11/9 > 1.222$. For $k=4$, MaxIS produces the pattern $\langle 6,3,2,2,1,1,1,1\rangle$ and then the approximation ratio is $52/35 > 1.485$.

More formally, for $x > 1$, consider the following pattern:

$$LB3 = \left\langle x, x, \frac{x}{2}, \frac{x}{3}, \ldots, \frac{x}{k-1}, \frac{x}{k} \right\rangle \ .$$

It follows that $n = \left(1 + \sum_{i=1}^{k} \frac{1}{i}\right) x = (H_k + 1)x$, where H_k is the harmonic sum $\sum_{i=1}^{k} \frac{1}{i}$. We choose x such that $LB3$ contains only integral numbers (e.g., $x = k!$). The sum coloring of the pattern $LB3$ is:

$$\begin{aligned} SC(LB3) &= x + \sum_{i=1}^{k} (i+1)\frac{x}{i} \\ &= (k + 1 + H_k)x \ . \end{aligned}$$

The chopping procedure forces the MaxIS algorithm to construct the following pattern:

$$A(LB3) = \left\langle x, \frac{x}{2}, \frac{x}{3}, \ldots, \frac{x}{k-1}, \frac{x}{k}, d_{k+1}, \ldots, d_{k+h} \right\rangle \ ,$$

where $\langle d_{k+1}, \ldots, d_{k+h}\rangle$ is lexicographically less than $\langle \frac{x}{k}, \ldots, \frac{x}{k}\rangle$ in which $\frac{x}{k}$ appears k times. Also, $\sum_{i=1}^{h} d_{k+i} = x$. Indeed, the number of vertices in $A(LB3)$ is $\sum_{i=1}^{k} \frac{x}{i} + \sum_{i=1}^{h} d_{k+i}$, which is n, the number of vertices in $LB3$. The sum coloring of the pattern $A(LB3)$ assuming $h = k$ and $d_{k+i} = \frac{x}{k}$ for $1 \le i \le h$ is:

$$\begin{aligned} SC(A(LB3)) &= \sum_{i=1}^{k} i\frac{x}{i} + \sum_{i=1}^{k} (k+i)\frac{x}{k} \\ &= \left(k + k + \frac{1}{k}\sum_{i=1}^{k} i\right) x \\ &= (2.5k + 0.5)x \ . \end{aligned}$$

Using larger x and the lower bound from the previous section (namely, the constructions that yields the lower bound of $2 - o(1)$), the chopping procedure can achieve values for d_{k+i} such that

$$\begin{aligned} SC(A(LB3)) &= \sum_{i=1}^{k} i\frac{x}{i} + kx + \frac{x}{k}(k^2 - o(k^2)) \\ &= (3k - o(k))x \ . \end{aligned}$$

The first term is the sum coloring of the first k entries in $A(LB3)$. The second term is the cost of coloring the rest of the x vertices in the last h entries in $A(LB3)$ by at least k. The third term is the sum coloring of the last h entries using the lower bound described in the previous sub-section for $\frac{x}{k}$ vertices in each entry instead of x vertices.

Since $H_k = o(k)$, it follows that the approximation ratio of MaxIS is

$$r = \frac{(3k - o(k))x}{(k + H_k + 1)x} = 3 - o(1) \ .$$

5 The $4 - o(1)$ Lower bound

To achieve the tight bound, we extend the idea from the previous section. We let the first two entries be equal, and after two steps of chopping they should be equal to the third entry. After two additional chopping steps we want the first four entries to be equal, and so on.

Small examples of such patterns are $\langle 4, 4, 1\rangle$ for $k = 3$ and $\langle 36, 36, 9, 4\rangle$ for $k = 4$. Note that for $k = 3$, MaxIS produces the pattern $\langle 4, 2, 1, 1, 1\rangle$ and already the approximation ratio is at least $20/15 > 1.333$. For $k = 4$, MaxIS produces the pattern $\langle 36, 18, 9, 6, 4, 3, 3, 2, 1, 1, 1, 1\rangle$ and then the ratio is $240/151 > 1.589$.

More formally, for $x > 1$, consider the following pattern:

$$LB4 = \left\langle x, x, \frac{x}{4}, \frac{x}{9}, \ldots, \frac{x}{(k-1)^2}, \frac{x}{k^2} \right\rangle \ .$$

It follows that $n = \left(1 + \sum_{i=1}^{k} \frac{1}{i^2}\right) x$. We choose x such that $LB4$ contains only integral numbers (e.g., $x = (k!)^2$). The sum coloring of the pattern $LB4$ is:

$$\begin{aligned} SC(LB4) &= x + \sum_{i=1}^{k} (i+1)\frac{x}{i^2} \\ &= x + \sum_{i=1}^{k} \frac{x}{i} + \sum_{i=1}^{k} \frac{x}{i^2} \\ &< (H_k + 2.65)x \ . \end{aligned}$$

Recall that $\sum_{i=1}^{k} \frac{1}{i^2} < 1.65$, and that H_k denotes $\sum_{i=1}^{k} \frac{1}{i}$.

In the chopping procedure, once we arrive at a pattern of equal size we just take these $k+1$ columns as the next $k+1$ entries. We get that

$$A(LB4) = \left\langle x, \frac{x}{2}, \frac{x}{4}, \frac{x}{6}, \frac{x}{9}, \frac{x}{12}, \ldots, \frac{x}{(k-1)^2}, \frac{x}{(k-1)k}, \frac{x}{k^2}, \ldots, \frac{x}{k^2} \right\rangle \ ,$$

where $\frac{x}{k^2}$ appears $k+1$ times. The number of vertices in $A(LB4)$ is $\sum_{i=1}^{k} \frac{x}{i^2} + \sum_{i=1}^{k-1} \frac{x}{i(i+1)} + k\frac{x}{k^2}$. Since $\frac{x}{i(i+1)} = \frac{x}{i} - \frac{x}{i+1}$, it follows that $\sum_{i=1}^{k-1} \frac{x}{i(i+1)} = x - \frac{x}{k}$. Therefore, the number of vertices in $A(LB4)$ is equal to the number of vertices in $LB4$. The sum coloring of the pattern $A(LB4)$ is:

$$SC(A(LB4)) = \sum_{i=1}^{k} \frac{2i-1}{i^2}x + \sum_{i=1}^{k-1} \frac{2i}{i(i+1)}x + \sum_{i=1}^{k} (2k - 1 + i)\frac{1}{k^2}x$$

$$= \left(2\sum_{i=1}^{k}\frac{1}{i} - \sum_{i=1}^{k}\frac{1}{i^2} + 2\sum_{i=1}^{k-1}\frac{1}{i+1} + \frac{2k-1}{k} + \sum_{i=1}^{k}\frac{i}{k^2} \right) x \ .$$

Since $\sum_{i=1}^{k-1}\frac{1}{i+1} = \sum_{i=1}^{k}\frac{1}{i} - 1$ and since $\sum_{i=1}^{k}\frac{i}{k^2} = 1 + \frac{1}{k}$, it follows that

$$SC(A(LB4)) = 4\sum_{i=1}^{k}\frac{1}{i} + 1 - \sum_{i=1}^{k}\frac{1}{i^2} > (4H_k - 0.65)x \ .$$

It follows that the approximation ratio of MaxIS is

$$r > \frac{(4H_k - 0.65)x}{(H_k + 2.65)x} = 4 - o(1) \ .$$

Note that only for $k \geq 8290$ does the value of r exceed 3. Furthermore, $r \approx 3.9475$ for $k = 10^{100}$ and $r \approx 3.5334$ for $k = 10^{10}$.

References

[1] A. Bar-Noy, M. Bellare, M. M. Halldórsson, H. Shachnai, and T. Tamir. On chromatic sums and distributed resource allocation. In *Information and Computation,* 140:183–202, 1998.

[2] A. Bar-Noy and G. Kortsarz. The minimum color-sum of bipartite graphs. *Proc. of the 24th Intl. Colloq. on Automata, Languages and Programming*, pp. 738–748, 1997. To appear in *Journal of Algorithms.* In URL: http://www.eng.tau.ac.il/˜amotz/publications.html.

[3] E. Kubicka. The Chromatic Sum of a Graph. PhD thesis, Western Michigan University, 1989.

[4] E. Kubicka, G. Kubicki, and D. Kountanis. Approximation Algorithms for the Chromatic Sum. In *Proc. of the First Great Lakes Computer Science Conf.*, Springer LNCS 507, pages 15–21, 1989.

[5] E. Kubicka and A. J. Schwenk. An Introduction to Chromatic Sums. In *Proc. of the ACM Computer Science Conf.*, pages 39-45, 1989.

[6] S. Nicoloso and M. Sarrafzadeh and X. Song. On the sum coloring problem on interval graphs. Istituto di Analisi dei Sistemi ed Informatica (IASI-CNR), Report 390.

[7] C. Thomassen, P. Erdös, Y. Alavi, J. Malde, and A. J. Schwenk. Tight Bounds on the Chromatic Sum of a Connected Graph. *J. of Graph Theory*, 13:353–357, 1989.