Approximating Chromatic Sum Coloring of Bipartite Graphs in Expected Polynomial Time

[1] *A.S. Asratian <arasr@mai.liu.se>*
[2] *N.N. Kuzyurin <nnkuz@ispras.ru>*
[1] *Linkopings Universitet, Department of Mathematics, Sweden, 581 83,*
[2] *Institute for System Programming of the Russian Academy of Sciences,*
25, Alexander Solzhenitsyn st., 109004, Moscow, Russia

Abstract. It is known that if complexity class P is not equal to NP the sum coloring problem cannot be approximated within 1+epsilon for some positive constant epsilon.
We consider finite, undirected graphs without loops and multiple edges. Let G=(V,E) be a graph. By a coloring of G we mean a mapping c of V to the numbers 1,2, ..., |V| . A coloring c is proper if c(v) is not equal to c(u) whenever the vertices u and v are adjacent.
Let S(G,c) is the sum_of c(v) over all vertices v. By a chromatic sum of G we mean the number S(G)=min S(G,c) where minimum is taken over all proper colorings c of G.
The problem of finding S(G) is called the sum coloring problem.
It was shown that the sum coloring problem is NP-complete.
A graph G is called bipartite if the set of vertices of G can be partitioned into two non-empty sets V1 and V2 such that every edge of G has one end in each of the sets.
For a number b, we say that an algorithm A approximates the chromatic sum within factor b over graphs on n vertices, if for every such graph G the algorithm A outputs a proper coloring c, such that S(G,c) is not greater than b S(G).
It is known that there exists 27/26-approximation polynomial algorithm for the chromatic SUM COLORING PROBLEM on any bipartite graph. On the other side, it was shown that here exists epsilon>0, such that there is no (1+epsilon)-approximation polynomial algorithm for the sum coloring problem on bipartite graphs, unless P is not equal to NP.
In this paper we consider the problem of developing an (1+epsilon)-approximation algorithm for the sum coloring of bipartite graphs which is polynomial in the average case for arbitrary small epsilon. We prove the existence of such algorithm.

Keywords: sum coloring problem, bipartite graphs, expected polynomial time

DOI: 10.15514/ISPRAS-2015-27(5)-11

For citation: Asratian A.S., Kuzyurin N.N. Approximating Chromatic Sum Coloring of Bipartite Graphs in Expected Polynomial Time. Trudy ISP RAN/Proc. ISP RAS, vol. 27, issue 5, 2015, pp. 191-198 (in Russian). DOI: 10.15514/ISPRAS-2015-27(5)-11.

1. Introduction

Let $G=(V_1,V_2,E)$ be a bipartite graph with $n+m$ vertices such that $|V_1|=m$, $|V_2|=n$, $m\le n$. By a coloring we mean a mapping:

$$c: V_1\cup V_2 \to \{1,2,\dots,n+m\}.$$

A coloring is proper if $c(v)\ne c(u)$ whenever $(u,v)\in E$.

Let $S(G,c)=\sum_{v\in V} c(v)$. By a chromatic sum we mean $S(G)=\min_c S(G,c)$ where minimum is taken over all proper colorings of G. The problem of finding $S(G)$ is called the SUM COLORING PROBLEM.

The notion of chromatic sum was first introduced in [6] where it was shown that the SUM COLORING PROBLEM is NP-complete on arbitrary graphs. A few b-approximation algorithms which find a coloring c with $S(G,c)\le b\cdot S(G)$ were presented. In [7] a $10/9$-approximation polynomial algorithm for the SUM COLORING PROBLEM on any bipartite graph was described. This result was improved in [8] where an $27/26$-approximation algorithm for the same problem was constructed. On the other side, in [7] the authors have shown that there exists $\varepsilon>0$, such that there is no $(1+\varepsilon)$-approximation polynomial algorithm for the SUM COLORING PROBLEM on bipartite graphs, unless $P=NP$.

In this paper we present for any positive ε an $(1+\varepsilon)$-approximation algorithm for this problem with expected polynomial time. The probabilistic distribution is uniform over all bipartite graphs with N vertices, $N=n+m$, $m\le n$. Note that the first example of approximation algorithm with expected polynomial time guaranteeing approximation ratio better than inapproximability threshold in the worst case was presented in [9]. Probabilistic analysis of algorithms for random graphs is the focus of much research now [1-5, 9].

2. Approximation scheme with expected polynomial time

Let $N=n+m$. We consider now a straightforward approach testing all possible colorings of G and choosing the one with the best possible color sum.

Algorithm 1. Test all possible vertex colorings of a bipartite graph and choose a proper coloring with minimum color sum.

Lemma 1. The time complexity of Algorithm 1 is $O(N^N)=O((2n)^{2n})$.

Let δ be a positive number, $0<\delta<1$ and

$$V_1'=\{v\in V_1 : (1-\delta)\frac{m}{2}\le \deg v\le (1+\delta)\frac{m}{2}\},$$

$$V_2' = \{v \in V_2 : (1-\delta)\frac{n}{2} \le \deg v \le (1+\delta)\frac{n}{2}\},$$
$$\overline{V}'_1 = V_1 \setminus V_1',$$
$$\overline{V}'_2 = V_2 \setminus V_2'.$$

2.1 Algorithm VERTEX-COLOR.

Input: A bipartite graph $G = (V_1, V_2, E)$ such that $|V_1| = m$, $|V_2| = n$, $m \le n$, and a parameter $\varepsilon > 0$.

Output: A proper coloring c of G such that $S(G) \le S(G,c) \le (1+\varepsilon)S(G)$.

1. If $\varepsilon \le \max\{40n^{-0.5}, n^{-0.2}, 50n^{-0.3}\}$ then goto 7.
2. If $m \le n^{0.8}$ then goto 7.
3. Set $\delta = \min\{\frac{1}{50}, \frac{\varepsilon}{50} - n^{-0.3}\}$.
4. Count the number $t_1 = |\overline{V}'_1|$, and $t_2 = |\overline{V}'_2|$.
5. If $t_1 > \sqrt{n}$ or $t_2 > n^{0.4}$ then goto 7.
6. Color V_2 by color 1 and color V_1 by color 2 and STOP.
7. Run Algorithm 1 and STOP.

Theorem 1. For any fixed $\varepsilon > 0$ Algorithm **VERTEX-COLOR** finds a proper coloring within $1+\varepsilon$ of the optimum color sum in expected polynomial time.

Proof. Note that at step 2 and step 5 of the algorithm we get $S(G,c) = n + 2m$ using very simple coloring strategy. The main idea of the proof is to extract sufficiently large almost regular bipartite subgraph $G' = (V_1', V_2', E')$ of G such that for any $v \in V_1'$ $(1-\delta')r \le \deg v \le (1+\delta')r$, and for any $v \in V_2'$ $(1-\delta')k \le \deg v \le (1+\delta')k$. Such an almost regular subgraph can guarantee a tight lower bound on $S(G)$ close to the upper bound $S(G) \le n + 2m$. The main difficulty is to estimate the probability that the size of such subgraph is large enough.

We use m' and n' for denoting $|V_1'|$ and $|V_2'|$ respectively.

Lemma 2. For any $0 < \delta' < \frac{1}{2}$ and an induced subgraph $G' = (V_1', V_2', E')$ as above
$$n' + 2m' - 10\delta' m' \le S(G') \le n' + 2m'.$$

Proof of Lemma 2. The upper bound is evident (we color V_1' by color 2 and color V_2' by color 1). To prove the lower bound we use the folowing inequalities
$$(1+\delta')r\sum_{v\in V_1'} c(v) + (1+\delta')k\sum_{v\in V_2'} c(v) \ge \sum_{e=(u,v)\in E'} (c(u)+c(v)) \ge$$
$$\ge 3|E'| \ge 3r(1-\delta')m'.$$
This implies the inequality
$$\sum_{v\in V_1'} c(v) + \frac{k}{r}\sum_{v\in V_2'} c(v) \ge 3m'\frac{1-\delta'}{1+\delta} \ge 3m'(1-2\delta').$$
Adding to both parts of the inequality $(1-\frac{k}{r})\sum_{v\in V_2'} c(v)$ and taking into account that $c(v) \ge 1$ for any v we obtain that for any proper coloring c of G'
$$S(G',c) = \sum_{v\in V_1'} c(v) + \sum_{v\in V_2'} c(v) \ge 3m' - 6\delta' m' + (1-\frac{k}{r})\sum_{v\in V_2'} c(v) \ge$$
$$\ge 2m' + m' - 6\delta' m' + (1-\frac{k}{r})n' = 2m' + n' + m' - 6\delta' m' - \frac{k}{r}n' \ge$$
$$2m' + n' + m' - 6\delta' m' - m' - 4\delta' m' = n' + 2m' - 10\delta' m'.$$
Here we used the inequality $m'r(1+\delta') \ge n'k(1-\delta')$ which for any $0 < \delta' < \frac{1}{2}$ implies
$$\frac{k}{r}n' \le m'\frac{1+\delta'}{1-\delta'} = m'(1+\frac{2\delta'}{1-\delta'}) \le m'(1+4\delta').$$
The proof of Lemma 2 is complete.

Now we estimate the size of G'.

Lemma 3. There is $c > 0$ depending on δ such that
$$Pr\{|\overline{V}'_2| \ge \sqrt{n}\} \le \exp\{\sqrt{n}\log n - cn^{3/2}\}.$$
$$Pr\{|\overline{V}'_1| \ge n^{0.4}\} \le \exp\{n^{0.4}\log n - cn^{1.2}\}.$$
Proof. We need the following lemma.

Lemma ([5]). Let $x_1, \ldots, x_n$ be independent random variables such that x_i takes two values: 0 and 1, and $Pr\{x_i = 1\} = p$, $Pr\{x_i = 0\} = 1 - p$.

Let $X = \sum_{i=1}^{n} x_i$ and $EX = np$. Then the following inequalities hold:

for any $\delta > 0$

$$Pr\{X - EX < -\delta EX\} \le \exp\{-(\delta^2/2)EX\},$$

for any $0 < \delta < 1$

$$Pr\{X - EX > \delta EX\} \le \exp\{-(\delta^2/3)EX\}.$$

Using this Lemma we have for $v \in V_1'$:

$$Pr\{d(v) \le n(1-\delta)2\} \le \exp\{-(\delta^2/2)n/2\},$$

$$Pr\{d(v) \ge n(1+\delta)2\} \le \exp\{-(\delta^2/3)n/2\}.$$

We give the proof for $\overline{V'}_2$. The proof for $\overline{V'}_1$ is similar.
To do this we estimate the following probability:

$$Pr\{|\overline{V'}_2| \ge k\} \le \binom{n}{k} (Pr\{fixed\ k_1 vertices v\ in\ \overline{V'}_2\ have\ d(v) \le (1-\delta)n/2\} \cdot$$

$$Pr\{fixed\ k_2 vertices v\ in\ \overline{V'}_2\ have\ d(v) \ge (1+\delta)n/2\}),$$

where $k = k_1 + k_2$. Using the Lemma and taking into account independence of the corresponding events we have

$$Pr\{fixed\ k_1 vertices v\ in\ \overline{V'}_2\ have\ d(v) \le (1-\delta)n/2\} \le$$

$$\exp\{-(\delta^2/3)k_1 m/2\} \le \exp\{-cmk_1\},$$

$$Pr\{fixed\ k_2 vertices v\ in\ \overline{V'}_2\ have\ d(v) \ge (1+\delta)m/2\} \le$$

$$\exp\{-(\delta^2/3)k_2 m/2\} \le \exp\{-cmk_2\},$$

where c depends on δ .

Letting in the last inequalities $k = n^{0.4}$ we obtain

$$Pr\{|\overline{V'}_2| \ge k\} \le \binom{n}{k} \exp\{-cm(k_1 + k_2)\} \le$$

$$\exp\{k \log n - cmk\} \le \exp\{n^{0.4} \log n - cn^{1.2}\}.$$

To finish the proof of Theorem 1 it is necessary to estimate the approximation ratio of the algorithm **VERTEX-COLOR** and its expected running time.

2.2 Approximation ratio

If the algorithm terminates at step 2 then we use the inequality

$$n + m \le S(G) \le n + 2m.$$

This gives that for the proper coloring c obtained at step 2

$$S(G,c) = n + 2m \le S(G) \cdot \frac{n+2m}{n+m} = S(G)(1 + \frac{m}{n+m}) \le$$

$$\le S(G)(1 + n^{-0.2}) \le S(G)(1+\varepsilon),$$

because $\varepsilon > n^{-0.2}$ (in the opposite case the algorithm always finds an optimal solution at step 7).

Because at step 7 we always find an optimal solution it is sufficient to estimate approximation ratio for step 6. To do this we use Lemma 2. If the algorithm terminates at step 6 then $t_1 \le \sqrt{n}$ and $t_2 \le n^{0.4}$. Thus we have $n' = n - t_1 \ge n - \sqrt{n}$, $m' = m - t_2 \ge m - \sqrt{n}$. Because the degree of a vertex in G' can decrease by at most $\sqrt{n}$ we can estimate δ' as follows:

$$\deg v \ge (1-\delta)\frac{m}{2} - \sqrt{n} = (1-\delta')\frac{m}{2},$$

which implies $\delta' = \delta + \dfrac{2\sqrt{n}}{m}$.

By Lemma 2

$$n + 2m - 10\delta' m - t_1 - t_2 \le S(G') \le S(G) \le n + 2m.$$

This implies the inequality

$$n + 2m - 10\delta m - 23\sqrt{n} \le S(G) \le n + 2m,$$

and then the inequality

$$(n+2m)(1 - 10\delta - \frac{25}{\sqrt{n}}) \le S(G) \le n + 2m.$$

Thus, for the coloring c that the algorithm outputs at step 6 the following inequality holds

$$S(G,c) \le S(G)(1 - 10\delta - \frac{25}{\sqrt{n}})^{-1}.$$

Now we use the following technical lemma.

Lemma. Let $0 < \delta < \min\{\frac{1}{50}, \frac{\varepsilon}{50}\}$, $\varepsilon > 40 n^{-0.5}$. Then

$$(1 - 10\delta - \frac{25}{\sqrt{n}})^{-1} \le 1 + \varepsilon.$$

Proof. We have

$$(1-10\delta-\frac{25}{\sqrt{n}})\cdot(1+\varepsilon)\geq 1$$

This is equivalent to

$$\varepsilon-10\delta(1+\varepsilon)-\frac{25}{\sqrt{n}}(1+\varepsilon)=$$

$$\varepsilon-(1+\varepsilon)(10\delta+\frac{25}{\sqrt{n}})\geq 0.$$

This implies

$$\frac{\varepsilon}{1+\varepsilon}\geq 10\delta+\frac{25}{\sqrt{n}}.$$

Taking into account the inequality $\delta < \varepsilon/50$ we have

$$n\geq\frac{1200}{\varepsilon^2}.$$

This inequality follows from the condition of the Lemma: $\varepsilon > 40n^{-0.5}$.

2.3 Expected running time

Step 4 is performed in quadratic (in n) time. By Lemmas 1 and 3 the expected time of step 7 is at most

$$O((2n)^{2n})\exp\{\sqrt{n}\log n-cn^{1.2}\}\leq$$

$$c\exp\{2n\log 2n+\sqrt{n}\log n-cn^{1.2}\}\to 0$$

as n tends to infinity.

References

[1]. B. Bollobas, Random graphs, second ed., Cambridge Univ. Press, Cambridge, 2001.

[2]. B. Bollobas, The chromatic number of random graphs, Combinatorica, 1988, v. 8, c. 49-55.

[3]. A. Frieze,C. McDiarmid, Algorithmic theory of random graphs, Random Structures and Algorithms, Jan.-March 1997, v. 10, n. 1-2, c. 5-42.

[4]. L. Kucera, The greedy coloring is a bad probabilistic algorithm, J. of Algorithms, Dec. 1991, v. 12, n. 4, c. 674-684.

[5]. R. Motwani and P. Raghavan, Randomized algorithms, Cambridge Univ. Press, 1995.

[6]. E. Kubicka, A.J. Schwenk, An introduction to chromatic sums, Proc. of ACM Computer Science Conference, 1989, c. 39-45.

[7]. A. Bar-Noy, G. Kortsarz, Minimum color sum of bipartite graphs, J. of Algorithms, 1998, v. 28, c. 339-365.

[8]. K. Giaro, R. Janczewski, M. Kubale, M. Malafiejski, A 27/26-approximation algorithm for the chromatic sum coloring of bipartite graphs, Proc. APPROX 2002, LNCS v. 2462, c. 135-145.

[9]. M. Krivelevich , V.H. Vu, Approximating the independence number and the chromatic number in expected polynomial time, J. Combin. Optimization, 2002, v. 6, c. 143-155.

Приближенный алгоритм для хроматической раскраски двудольных графов за полиномиальное в среднем время

[1]*А.С. Асратян <arasr@mai.liu.se>*

[2]*Н.Н. Кузюрин <nnkuz@ispras.ru>*

[1] *Линчёпингский университет, факультет математики, Швеция, 581 83*

[2] *Институт системного программирования РАН,*

Россия, 109004, Москва, ул. Солженицына, 25

Аннотация. Известно что если P≠NP то задача аппроксимации суммарной раскраски двудольных графов не может быть осуществлена в полиномиальное время с точностью $\mathbf{1+\varepsilon}$ для некоторой константы $\boldsymbol{\varepsilon}$. Мы предлагаем для сколь угодно малого $\boldsymbol{\varepsilon}$>0 приближенный алгоритм для данной проблемы который работает за полиномиальное в среднем время.

Ключевые слова: проблема хроматической раскраски, двудольные графы, полиномиальное в среднем время

DOI: 10.15514/ISPRAS-2015-27(5)-11

Для цитирования: Асратян А.С., Кузюрин Н.Н. Приближенный алгоритм для хроматической раскраски двудольных графов за полиномиальное в среднем время. Труды ИСП РАН, том 27, вып. 5, 2015 г., стр. 191-198. DOI: 10.15514/ISPRAS-2015-27(5)-11.

Литература

[1]. B. Bollobas, Random graphs, second ed., Cambridge Univ. Press, Cambridge, 2001.

[2]. B. Bollobas, The chromatic number of random graphs, Combinatorica, 1988, v. 8, c. 49-55.

[3]. [1] A. Frieze,C. McDiarmid, Algorithmic theory of random graphs, Random Structures and Algorithms, Jan.-March 1997, v. 10, n. 1-2, c. 5-42.

[4]. L. Kucera, The greedy coloring is a bad probabilistic algorithm, J. of Algorithms, Dec. 1991, v. 12, n. 4, c. 674-684.

[5]. R. Motwani and P. Raghavan, Randomized algorithms, Cambridge Univ. Press, 1995.

[6]. E. Kubicka, A.J. Schwenk, An introduction to chromatic sums, Proc. of ACM Computer Science Conference, 1989, c. 39-45.

[7]. A. Bar-Noy, G. Kortsarz, Minimum color sum of bipartite graphs, J. of Algorithms, 1998, v. 28, c. 339-365.

[8]. K. Giaro, R. Janczewski, M. Kubale, M. Malafiejski, A 27/26-approximation algorithm for the chromatic sum coloring of bipartite graphs, Proc. APPROX 2002, LNCS v. 2462, c. 135-145.

[9]. M. Krivelevich , V.H. Vu, Approximating the independence number and the chromatic number in expected polynomial time, J. Combin. Optimization, 2002, v. 6, c. 143-155.