Let $G = (V_1, V_2, E)$ be a bipartite graph with $n + m$ vertices such that $|V_1| = m$, $|V_2| = n$, $m \leq n$. By a coloring we mean a mapping:
\[ c : V_1 \cup V_2 \rightarrow \{1, 2, \ldots, n + m\} \]
A coloring is proper if $c(v) \neq c(u)$ whenever $(u, v) \in E$.

Let $S(G, c) = \sum_{v \in c} c(v)$. 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 $S(G, c) \leq b \cdot S(G)$ 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 $\varepsilon$ an $(1 + \varepsilon)$-approximation algorithm for this problem with expected polynomial time. The probabilistic distribution is uniform over all bipartite graphs with $N$ vertices, $N = n + m$, $m \leq n$. 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 $\delta$ be a positive number, $0 < \delta < 1$ and
\[ V'_1 = \{ v \in V_1 : (1 - \delta) \frac{m}{2} \leq \deg v \leq (1 + \delta) \frac{m}{2} \} \]
2.1 Algorithm VERTEX-COLOR.

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

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

1. If $\varepsilon \leq \max\{40n^{-0.5}, n^{-0.2}, 50n^{-0.3}\}$ then goto 7.
2. If $m \leq n^{0.8}$ then goto 7.
3. Set $\delta = \min\{\frac{1}{50}, \varepsilon - n^{-0.3}\}$.
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 $v \in V'_1$ $(1 - \delta')r \leq \deg v \leq (1 + \delta')r$, and for any $v \in V'_2$ $(1 - \delta')k \leq \deg v \leq (1 + \delta')k$. Such an almost regular subgraph can guarantee a tight lower bound on $S(G)$ close to the upper bound $S(G) \leq n + 2m$. 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

$$n' + 2m' - 10\delta'm' \leq S(G') \leq n' + 2m'.$$

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 following inequalities

$$(1 + \delta')r \sum_{v \in V'_1} c(v) + (1 + \delta')k \sum_{v \in V'_2} c(v) \geq \sum_{e = (u, v) \in E'} (c(u) + c(v)) \geq 3|E'| \geq 3r(1 - \delta'm').$$

This implies the inequality

$$\sum_{v \in V'_1} c(v) + \frac{k}{r} \sum_{v \in V'_2} c(v) \geq 3m' - 6\delta'm' + (1 - \frac{k}{r}) \sum_{v \in V'_2} c(v) \geq 2m' + m' - 6\delta'm' - (1 - \frac{k}{r}) n' \geq 2m' + n' - 6\delta'm' - 2m' - 10\delta'm'.$$

Here we used the inequality $m'r(1 + \delta') \geq n'k(1 - \delta')$ which for any $0 < \delta' < \frac{1}{2}$ implies

$$\frac{k}{r} n' \leq m'r \frac{1 + \delta'}{1 - \delta'} = m'(1 + 2\delta') \leq m'(1 + 4\delta').$$

The proof of Lemma 2 is complete.

Now we estimate the size of $G'$.

Lemma 3. There is $c > 0$ depending on $\delta$ such that

$$Pr\{|V'_2| \geq \sqrt{n}\} \leq \exp\{\sqrt{n} \log n - cn^{3/2}\},$$

$$Pr\{|\overline{V'_1}| \geq n^{0.4}\} \leq \exp\{n^{0.4} \log n - cn^{1.2}\}.$$
Let $X = \sum_{i=1}^{n} x_i$ and $EX = np$. Then the following inequalities hold:

For any $\delta > 0$

$$Pr\{X - EX < -\delta EX\} \leq \exp\{-\delta^2/2EX\},$$

for any $0 < \delta < 1$

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

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

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

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

We give the proof for $\bar{V}'_2$. The proof for $\bar{V}'_1$ is similar.

To do this we estimate the following probability:

$$Pr\{\bar{V}'_2 \geq k\} \leq n \cdot \frac{Pr\{\text{fixed } k_1 \text{ vertices in } \bar{V}'_2 \text{ have } d(v) \leq (1 - \delta)n/2\}}{k},$$

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

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

$$Pr\{\text{fixed } k_1 \text{ vertices in } \bar{V}'_2 \text{ have } d(v) \leq (1 - \delta)n/2\} \leq \exp\{-(\delta^2/3)k_2/m/2\} \leq \exp\{-cmk_2\},$$

$$Pr\{\text{fixed } k_2 \text{ vertices in } \bar{V}'_2 \text{ have } d(v) \geq (1 + \delta)m/2\} \leq \exp\{-(\delta^2/3)k_1/m/2\} \leq \exp\{-cmk_1\},$$

where $c$ depends on $\delta$.

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

$$Pr\{\bar{V}'_2 \geq k\} \leq n \cdot \exp\{-cm(k_1 + k_2)\} \leq$$

$$\exp(k \log n - cmk) \leq \exp\{n^{0.4} \log n - cn^{1.2}\}.$$  

To finish the proof of Theorem 1 it is necessary to estimate the approximation ratio of the algorithm \textsc{Vertex-Color} and its expected running time.

### 2.2 Approximation ratio

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

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

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

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

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 $t_1 \leq \sqrt{n}$ and $t_2 \leq n^{0.4}$. Thus we have $n' = n - t_1 \geq n - \sqrt{n}$, $m' = m - t_2 \geq m - \sqrt{n}$. Because the degree of a vertex in $G'$ can decrease by at most $\sqrt{n}$ we can estimate $\delta'$ as follows:

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

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

By Lemma 2

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

This implies the inequality

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

and then the inequality

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

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

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

Now we use the following technical lemma.

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

$$(1 - 10\delta - \frac{25}{\sqrt{n}})^{-1} \leq 1 + \varepsilon.$$
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 
\[O((2n)^{2\varepsilon})\exp(\sqrt{n} \log n - cn^{1/2}) \leq c \exp\{2n \log 2n + \sqrt{n} \log n - cn^{1/2}\} \to 0\]
as \(n\) tends to infinity.

References


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

А.С. Асратян <arasr@mat.liu.se>
Н.Н. Кузюрин <nnkuz@ispras.ru>

Литература