

Balanced coloring of bipartite graphs

Uriel Feige

Shimon Kogan

Department of Computer Science and Applied Mathematics
Weizmann Institute, Rehovot 76100, Israel
uriel.feige@weizmann.ac.il

June 16, 2009

Abstract

Given a bipartite graph $G(U \cup V, E)$ with n vertices on each side, an independent set $I \in G$ such that $|U \cap I| = |V \cap I|$ is called a balanced bipartite independent set. A balanced coloring of G is a coloring of the vertices of G such that each color class induces a balanced bipartite independent set in G . If graph G has a balanced coloring we call it *colorable*. The coloring number $\chi_B(G)$ is the minimum number of colors in a balanced coloring of a colorable graph G . We shall give bounds on $\chi_B(G)$ in terms of the average degree \bar{d} of G and in terms of the maximum degree Δ of G . In particular we prove the following:

- $\chi_B(G) \leq \max\{2, \lfloor 2\bar{d} \rfloor + 1\}$.
- For any $0 < \epsilon < 1$ there is a constant Δ_0 such that the following holds. Let G be a balanced bipartite graph with maximum degree $\Delta \geq \Delta_0$ and $n \geq (1 + \epsilon)2\Delta$ vertices on each side, then $\chi_B(G) \leq \frac{20}{\epsilon^2} \frac{\Delta}{\ln \Delta}$.

1 Introduction and definitions

Let $G(U \cup V, E)$ be bipartite graph with n vertices on each side. We call such a graph a balanced bipartite graph. An independent set $I \in G$ such that $|U \cap I| = |V \cap I| = k$ is called a balanced bipartite independent set of size k . Finding a balanced bipartite set of maximum size (the max-BBIS problem) is NP-hard, and the reader is referred to [2] for a discussion of some of its applications, as well as some results regarding its inapproximability.¹

¹Technically [2] discusses the Complete Bipartite Subgraph Problem, but this problem is the same as max-BBIS, by exchanging the interpretation of edges and non-edges.

A balanced coloring of G is a coloring of the vertices of G such that each color class induces a balanced bipartite independent set in G . Notice that G has a balanced coloring if and only if G contains a perfect matching of non-edges. We call such a graph colorable. Denote by $\chi_B(G)$ the minimum number of colors needed in a balanced coloring of a colorable balanced bipartite graph G .

Some applications in computational biology to colorings similar to the balanced coloring are described in [1]. In particular they focus on partitions into balanced subgraphs of maximum degree 1, and into possibly overlapping balanced subgraphs of maximum degree 1.

In the following sections we shall give bounds on $\chi_B(G)$ in terms of the average degree \bar{d} of G and in terms of the maximum degree Δ of G . Our main results (see Theorems 2.1 and 3.2 for more details) are that $\chi_B(G) \leq \max\{2, \lfloor 2\bar{d} \rfloor + 1\}$ and that $\chi_B(G) = O(\Delta/\log \Delta)$ (when Δ is not too large). It is instructive to compare these results with the results known with respect to the chromatic number $\chi(G)$ of a graph (the minimum number of independent sets that cover all vertices, without the requirement that the graph is bipartite and that independent sets be balanced with respect to a given bipartition), as this allows one to appreciate the differences between these two seemingly related problems. No bound on the chromatic number can be given in terms of the average degree alone, for the simple reason that for any graph, its average degree can be made arbitrary small by adding isolated vertices to the graph, without decreasing the chromatic number. As to bounds in terms of maximum degree, the well known inductive coloring argument shows that for every graph, $\chi(G) \leq \Delta + 1$. From this it is easy to show that for every graph, $\chi_B(G) \leq 2\Delta + 1$. The proof, which implicitly appears in [1], is based on merging together matched vertices in a perfect matching of the edge complement of G . However, we prove a stronger bound, namely, that $\chi_B(G) \leq O(\Delta/\log \Delta)$ (when the number of vertices is sufficiently large). This is similar in nature to the fact that $\chi(G) \leq O(\Delta/\log \Delta)$ for triangle free graphs (as was proved by Johansson, see for example [7]). The reader might suspect that the proof of these two statements is also similar, because bipartite graphs have no triangles, but in fact the proofs are different. To appreciate why the proofs need to be different, the reader may observe that with respect to chromatic number, when one color class (independent set) is removed from the graph, the chromatic number of the remaining graph is never larger than that of the original graph. This is not true for balanced coloring. By removing a balanced bipartite independent set from a graph, the number of colors needed to color the remaining graph may in fact go up, and in some cases, the remaining graph might not be colorable at all. Hence all our algorithms for balanced coloring involve recoloring of previously colored vertices.

It is also instructive to compare between chromatic number and balanced coloring in terms

of their computational complexity. Both problems are polynomially time solvable when the number of required colors is two. For balanced coloring, we omit the proof, but it is similar in nature to the proof of Lemma 2.2. When the number of required colors is three, both problems become NP-hard. For chromatic number, the best polynomial time algorithms currently known for coloring 3-colorable graphs require n^δ colors, where n is the number of vertices and $\delta > 0$ is some constant (see for example [4] for one such result). However, in terms of hardness of approximation results, it is only known that it is NP-hard to color 3-colorable graphs by four colors [5]. For minimum balanced coloring, larger hardness of approximation ratios can be established under reasonable complexity assumptions. Recall that SAT is NP-hard problem of determining whether a Boolean formula in conjunctive normal form has a satisfying assignment. A stronger version of the P not equal NP assumption postulates that algorithms for SAT require time exponential in n , the number of variables, and moreover, that this holds also for randomized algorithms.

Theorem 1.1. *There are $\epsilon > 0$ and $\delta > 0$, such that given a graph G with n vertices, there is no polynomial time algorithm that distinguishes between the following two cases:*

1. $\chi_B(G) \leq 3$
2. $\chi_B(G) \geq n^\delta$

unless SAT has a randomized algorithm that runs in time $O(2^{n^\epsilon})$.

Proof: We prove the hardness of approximation of balanced coloring problem by a reduction from the maximum balanced bipartite independent set problem (max-BBIS). Denote by $I_B(G)$ the size of the maximum balanced bipartite independent set in graph G . It is shown in [6] that for some $0 < \alpha < \beta < 1$ one cannot distinguish in polynomial time between the following cases

1. $I_B(G) \leq n^\alpha$
2. $I_B(G) \geq n^\beta$

unless unless SAT has a randomized algorithm that runs in time $O(2^{n^\epsilon})$. Let G be a balanced bipartite graph $G(U \cup V, E)$ with n vertices on each side, in which we seek to distinguish between the above cases. Add $n - n^\beta$ isolated vertices to each side of graph G . We will refer to these vertices as *bad* vertices. Call the resulting graph with $2n - n^\beta$ vertices on each side G' . Notice that if $I_B(G) \geq n^\beta$ then $\chi_B(G') \leq 3$. We will show that if $I_B(G) \leq n^\alpha$ then $\chi_B(G') \geq n^{\beta-\alpha}$. For the sake of contradiction, consider a balanced coloring C_1, C_2, \dots, C_t of G' consisting of

$t < n^{\beta-\alpha}$ colors. While there is a bad vertex in some color class remove it and an arbitrary vertex from the other side of the color class. In the end of this process we will have at most t color classes which together contain at least n^β vertices from the original graph on each side, hence some color class corresponds to a $\frac{n^\beta}{t} > n^\alpha$ balanced bipartite independent set in G .

To summarize, we may take δ to be essentially $\beta - \alpha$. If there is a polynomial time algorithm that distinguishes between the cases $\chi_B(G') \geq n^\delta$ and $\chi_B(G') \leq 3$, then the algorithm (after performing the reduction described above) can be used to distinguish between the cases that $I_B(G) \leq n^\alpha$ and $I_B(G) \geq n^\beta$, and then in combination with the randomized reduction described in [6], it can be used as a randomized algorithm that solves SAT instances in time $O(2^{n^\epsilon})$. ■

2 A bound on the balanced coloring number in terms of the average degree

In this section we will give upper bounds on $\chi_B(G)$ in terms of the average degree of G .

Theorem 2.1. *If $G(U \cup V, E)$ is a colorable balanced bipartite graph with average degree \bar{d} , then $\chi_B(G) \leq \max\{2, \lfloor 2\bar{d} \rfloor + 1\}$.*

Furthermore the results of theorem 2.1 are constructive in the sense that such coloring can be found in polynomial time.

2.1 Proof of theorem 2.1

First we will prove the following lemmas.

Lemma 2.2. *Let $G(U, V, E)$ be a balanced bipartite graph with $|U| = |V| = n$ and $|E| < n$, then $\chi_B(G) \leq 2$.*

Proof: Since G has $2n$ vertices and at most $n - 1$ edges it contains at least $n + 1$ components. Assume w.l.o.g that it contains exactly $n + 1$ components C_1, C_2, \dots, C_{n+1} where $\sum_{i=1}^{n+1} |C_i| = 2n$. We claim that there is an index set I s.t. $\sum_{i \in I} |C_i| = n$. Look at the partial sums:

$$\begin{aligned} S_1 &= |C_1| \\ S_2 &= |C_1| + |C_2| \\ &\dots \\ S_n &= |C_1| + |C_2| + \dots + |C_n| \end{aligned}$$

Notice that $S_1 < S_2 < \dots < S_n < 2n$. If for some i, j we have $S_i \equiv S_j \pmod{n}$ then $S_i - S_j \equiv 0 \pmod{n}$ which implies that $S_i - S_j = n$. On the other hand if $S_i \not\equiv S_j \pmod{n}$ for all i, j then for some i we have $S_i \equiv 0 \pmod{n}$ and thus $S_i = n$. Thus we have shown that there is an index set I s.t. $\sum_{i \in I} |C_i| = n$. Let $U' = U \cap (\bigcup_{i \in I} C_i)$ and $V' = V \cap (\bigcup_{i \in I} C_i)$. Since $|U'| + |V'| = n$ we have that $U' \cup (V \setminus V')$ and $(U \setminus U') \cup V'$ are balanced bipartite independent sets and thus G is 2-colorable. ■

Lemma 2.3. *Let $G(U, V, E)$ be a colorable balanced bipartite graph with $|U| = |V| = n$, and $d \geq 3$ an integer. If $\chi_B(G) = d$ then $|E(G)| \geq \frac{(d-1)n}{2}$.*

Proof: Let $C = \{I_1, I_2, \dots, I_d\}$ be a balanced coloring of G , where for each i , $I_i = U_i \cup V_i$ is a balanced bipartite independent set with $U_i \in U$, $V_i \in V$ and $|U_i| = |V_i|$. Let

$$S = \sum_{1 \leq i < j < k \leq d} E(I_i \cup I_j \cup I_k)$$

where $E(I_i \cup I_j \cup I_k)$ denotes the number of edges in the subgraph of G induced by the union of I_i, I_j, I_k . By our assumption that $\chi_B(G) = d$ we know that G is not $(d-1)$ -colorable and thus from lemma 2.2 it follows that

$$\forall_{i < j < k} E(I_i \cup I_j \cup I_k) \geq |V_i| + |V_j| + |V_k|$$

for otherwise the subgraph induced by $I_i \cup I_j \cup I_k$ is 2-colorable and thus the whole graph will be $(d-1)$ -colorable. We conclude that

$$\begin{aligned} S &\geq \sum_{1 \leq i < j < k \leq d} |V_i| + |V_j| + |V_k| \\ &= \binom{d-1}{2} \sum_{i=1}^d |V_i| \\ &= \binom{d-1}{2} n \end{aligned}$$

Now consider another way to evaluate the sum S . Each edge of G is between two different sets of C and thus each edge of G is counted exactly $d-2$ times in the sum S , so actually $S = (d-2)|E(G)|$. Thus we have

$$(d-2)|E(G)| \geq \binom{d-1}{2} n$$

and we are done. ■

Theorem 2.1 follows directly from Lemma 2.2 and Lemma 2.3. If $\bar{d} < 1$ then by Lemma 2.2 we have $\chi_B(G) \leq 2$. From Lemma 2.3 it follows that for $\bar{d} \geq 1$, if $\chi_B(G) \geq \lfloor 2\bar{d} \rfloor + 2$ then

$$E(G) \geq \frac{(\lfloor 2\bar{d} \rfloor + 1)n}{2} > \bar{d}n$$

and thus in this case the graph G is necessarily $(\lfloor 2\bar{d} \rfloor + 1)$ -colorable.

Now that we finished proving Theorem 2.1 we may notice that in fact we have the following polynomial time coloring algorithm which works by local improvements.

Algorithm BalancedColoring(G)

Input: A colorable balanced bipartite graph $G(U \cup V, E)$ where $|X| = |Y| = n$.

Output: A balanced coloring of G with at most $\max\{2, \lfloor 2\bar{d} \rfloor + 1\}$ colors.

1. Color G with n colors (use the perfect matching of non-edges in G).
2. While there are three color classes I_1, I_2, I_3 s.t.

$$2 \cdot E(I_1 \cup I_2 \cup I_3) < |I_1| + |I_2| + |I_3|$$

use the procedure described in the proof of lemma 2.2 to recolor the subgraph $I_1 \cup I_2 \cup I_3$ using at most two colors.

3. Return the resulting coloring

2.2 Lower bounds

In this section we exhibit an infinite family of graphs in which the balanced coloring number is more than twice the average degree. In particular for each integer $d \geq 2$ we will show a colorable balanced bipartite graph with average degree $d/2$ which is not d -colorable.

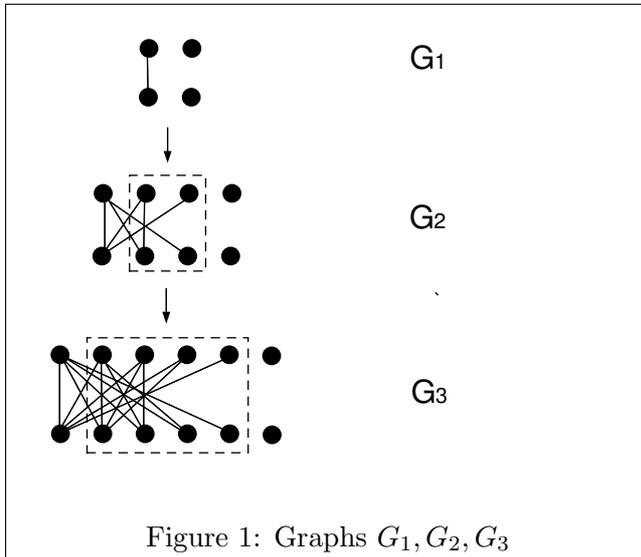
Let $G_1(U_1 \cup V_1, E_1)$ be a bipartite graph with $U_1 = \{u_1, u_2\}$, $V_1 = \{v_1, v_2\}$ and $E_1 = \{(u_1, v_1)\}$. Notice that G_1 is a colorable balanced bipartite graph with $\bar{d}(G) = 1/2$ and $\chi_B(G) = 2$. For all $i > 1$ define $G_i(U_i \cup V_i, E_i)$ in the following way.

1. Take graph G_{i-1} and add a vertex to U_{i-1} which is connected to all the vertices of V_{i-1}

2. Add a vertex to V_{i-1} which is connected to all the vertices in U_{i-1} and to the new vertex which was added in step 1.
3. Add an isolated vertex to each side of the bipartite graph.

It is easy to see that $\chi_B(G_i) = 2i$ and

$$\begin{aligned}\bar{d}(G_i) &= \frac{1 + 5 + 9 + \dots + (4i - 3)}{2i} \\ &= i - \frac{1}{2}\end{aligned}$$



In a similar manner one can build a graph family G'_1, G'_2, \dots which satisfies $\chi_B(G'_i) = 2i + 1$ and $\bar{d}(G'_i) = i$. Let $G'_1(U_1 \cup V_1, E_1)$ be a bipartite graph with $U_1 = \{u_1, u_2, u_3\}$, $V_1 = \{v_1, v_2, v_3\}$ and $E_1 = \{(u_1, v_1), (u_2, v_2), (u_3, v_3)\}$. Notice that G'_1 is a colorable balanced bipartite graph with $\bar{d}(G'_1) = 1$ and $\chi_B(G'_1) = 3$. Graphs G'_2, G'_3, \dots can be obtained from graph G'_1 with same recursive method which we used to obtain graphs G_2, G_3, \dots from graph G_1 .

3 A bound on the balanced coloring number in terms of the maximum degree

Let $G(U \cup V, E)$ be a balanced bipartite graph with n vertices on each side and maximum degree Δ . First notice that from Hall's theorem ([3]) we get the following.

Claim 3.1. *If $n \geq 2\Delta$ then G is colorable.*

Proof: A simple corollary to Hall's theorem ([3]) states that if a bipartite graph H with n vertices on each side has minimum degree $\geq n/2$ then H contains a perfect matching. As the bipartite complement of graph G satisfies that property, graph G contains a perfect matching of non-edges. ■

Furthermore it is easy to see that Claim 3.1 is tight in the sense that there are graphs with maximum degree Δ and $n = 2\Delta - 1$ vertices which are not colorable. In this section we shall prove the following theorem:

Theorem 3.2. *For any $0 < \epsilon < 1$ there is a constant Δ_0 such that the following holds. If G is a balanced bipartite graph with maximum degree $\Delta \geq \Delta_0$ and $n \geq (1 + \epsilon)2\Delta$ vertices on each side, then $\chi_B(G) \leq \frac{20}{\epsilon^2} \frac{\Delta}{\ln \Delta}$.*

Corollary 3.3. *For any $0 < \epsilon < 1$ and large enough n the following holds. If G is a balanced bipartite graph with n vertices on each side and minimum degree $\delta(G) \geq (1 + \epsilon)\frac{n}{2}$. Then G has a partition of its vertices into at most $\frac{20}{\epsilon^2} \frac{n}{\ln n}$ vertex disjoint bipartite cliques.*

Proof: Consider the bipartite complement \bar{G} of graph G . We have $\Delta(\bar{G}) \leq (1 - \epsilon)\frac{n}{2}$, and $n \geq 2(1 + \epsilon)\Delta(\bar{G})$ and thus by theorem 3.2 we have $\chi_B(\bar{G}) \leq \frac{20}{\epsilon^2} \frac{n}{\ln n}$. ■

It is not hard to show that Theorem 3.2 offers an optimal dependency on Δ (though most likely the constant factors can be improved). For example, a straightforward union bound shows that for some sufficiently large constant c , in a random bipartite graph with average degree Δ the expected number of balanced bipartite independent sets of size $\frac{cn \ln \Delta}{\Delta}$ is less than 1/2, implying that with high probability no such balanced independent set exists. In the absence of balanced independent sets of size $\frac{cn \ln \Delta}{\Delta}$, a balanced coloring requires more than $\frac{\Delta}{c \ln \Delta}$ colors.

Theorem 3.2 will be proved by successively finding "large" balanced bipartite independent sets in the yet uncolored part of graph G , and modifying the current coloring if no such set can be found.

Lemma 3.4. *For any $0 < \epsilon < 1$ there is a constant d_0 such that the following holds. Let G be a balanced bipartite graph with average degree $d \geq d_0$ and $n \geq (1 + \epsilon)d$ vertices on each side, then G contains a balanced bipartite independent set of size at least $\frac{\epsilon}{2} \cdot \frac{n \ln d}{d}$*

Proof: Let $k = \lceil \frac{n \ln d}{Cd} \rceil$ where $C = \frac{2}{\epsilon}$. We will prove that for large enough d_0 , any graph $G(U \cup V, E)$ with average degree $d \geq d_0$ and $n \geq (1 + \epsilon)d$ vertices on each side contains a balanced bipartite independent set of size k . Assume that the vertices in U are u_1, \dots, u_n and that their degrees are d_1, \dots, d_n . Choose uniformly at random a set S of k vertices in V . Let X be the random variable which counts the number of vertices in U which have no neighbors in S . The probability that vertex u_i has no neighbors in S is $\binom{n-d_i}{k} / \binom{n}{k}$ and thus

$$\begin{aligned} E(X) &= \sum_{i=1}^n \frac{\binom{n-d_i}{k}}{\binom{n}{k}} \\ &\geq n \frac{\binom{n-d}{k}}{\binom{n}{k}} && \text{by convexity} \\ &= n \cdot \prod_{i=0}^{k-1} \frac{n-d-i}{n-i} \end{aligned}$$

If $E(X) \geq k$ then G contains a balanced bipartite independent set of size k . Thus we will next prove that for $d \geq d_0$

$$\frac{n}{k} \geq \prod_{i=0}^{k-1} \frac{n-i}{n-d-i} \tag{3.1}$$

Now notice that

$$\begin{aligned}
\ln \left(\prod_{i=0}^{k-1} \frac{n-i}{n-d-i} \right) &= \sum_{i=0}^{k-1} \ln \left(1 + \frac{d}{n-d-i} \right) \\
&\leq k \ln \left(1 + \frac{d}{n-d-k+1} \right) \\
&\leq \frac{dk}{n-d-k+1} && \text{as } \ln(1+x) \leq x \text{ for } x \geq 0 \\
&\leq \frac{d \cdot \left(\frac{n \ln d}{Cd} + 1 \right)}{n \left(1 - \frac{\ln d}{Cd} \right) - d} \\
&= \frac{\frac{\ln d}{C} + \frac{d}{n}}{1 - \frac{\ln d}{Cd} - \frac{d}{n}} \\
&\leq \frac{\frac{\ln d}{C} + \frac{1}{1+\epsilon}}{1 - \frac{\ln d}{Cd} - \frac{1}{1+\epsilon}} && \text{as } n \geq (1+\epsilon)d \\
&= \frac{\frac{\epsilon \ln d}{2} + \frac{1}{1+\epsilon}}{\frac{\epsilon}{1+\epsilon} - \frac{\epsilon \ln d}{2d}} \\
&= \frac{\frac{(1+\epsilon) \ln d}{2} + \frac{1}{\epsilon}}{1 - \frac{(1+\epsilon) \ln d}{2d}} \\
&\leq (1-\delta) \ln d + O(1) && \text{for some } \delta > 0 \text{ and large enough } d
\end{aligned}$$

and thus

$$\prod_{i=0}^{k-1} \frac{n-i}{n-d-i} = O(d^{1-\delta}) \tag{3.2}$$

On the other hand

$$\begin{aligned}
\frac{n}{k} &\geq \frac{n}{\frac{n \ln d}{Cd} + 1} \\
&= \frac{Cd}{\ln d + \frac{Cd}{n}} \\
&\geq \frac{Cd}{\ln d + \frac{C}{1+\epsilon}} && \text{as } n \geq (1+\epsilon)d \\
&= \Omega\left(\frac{d}{\ln d}\right) \tag{3.3}
\end{aligned}$$

Finally inequality 3.1 follows from 3.2 and 3.3 for large enough d and we are done. ■

3.1 Proof of theorem 3.2

Fix some $0 < \epsilon < 1$. We will show that for some constant Δ_0 the following holds. Let $G(U \cup V, E)$ be a balanced bipartite graph with maximum degree $\Delta \geq \Delta_0$ and $n \geq (1 + \epsilon)2\Delta$ vertices on each side, then $\chi_B(G) \leq \frac{20}{\epsilon^2} \frac{\Delta}{\ln \Delta}$. Suppose that $U = \{u_1, u_2, \dots, u_n\}$ and $V = \{v_1, v_2, \dots, v_n\}$.

Definition 3.1. An r -partial coloring of G is a balanced coloring of some vertex induced subgraph $G'(U' \cup V', E')$ of $G(U, V, E)$, s.t. $U' \subseteq U$, $V' \subseteq V$ and $|V'| = |U'| = n - r$.

Definition 3.2. Given an r -partial coloring of G we define graph $G_r(X \cup Y, E')$ in the following way. Suppose that the colored vertices of G are u_{r+1}, \dots, u_n and v_{r+1}, \dots, v_n and furthermore that for each $r+1 \leq i \leq n$ we have that u_i and v_i are in the same color class. Let $X = \{x_1, \dots, x_r\}$ and $Y = \{y_{r+1}, \dots, y_n\}$.

$$(x_i, y_j) \in E' \Leftrightarrow (u_i, v_j) \in E \text{ or } (v_i, u_j) \in E$$

Thus G_r is a bipartite graph with r vertices on one side and $n - r$ on the other.

Given a balanced bipartite independent set I of size k in G_r one can extend the r -partial coloring into a $(r - k)$ -partial coloring in the following manner. Suppose that the vertices in I are x_1, \dots, x_k and y_{r+1}, \dots, y_{r+k} . We could add a new color class which contains vertices $u_1, \dots, u_k, v_{r+1}, \dots, v_{r+k}$ and another color class which contains vertices $v_1, \dots, v_k, u_{r+1}, \dots, u_{r+k}$. Notice that vertices u_1, \dots, u_k and v_1, \dots, v_k were uncolored in the original r -partial coloring and thus we've obtained a $(r - k)$ -partial coloring using two additional colors.

The coloring of G will be done in the following manner. Assume that we have some r -partial coloring of G (where in the beginning $r = n$), and that the uncolored vertices are u_1, \dots, u_r and v_1, \dots, v_r . Furthermore assume that vertices v_{r+i} and u_{r+i} are colored with the same color for $1 \leq i \leq n - r$. We'll show how to extend this partial coloring successively.

Case I: $r \geq \frac{\epsilon}{8}n$

Let A be the uncolored part of graph G . Subgraph A contains vertices u_1, \dots, u_r and v_1, \dots, v_r . Let d be the average degree of G and d_A be the average degree of the subgraph induced by A . If $r \geq (1 + \epsilon)d_A$ and $d_A \geq d_0$ for large enough d_0 then by lemma 3.4 subgraph A contains a balanced bipartite independent set of size

$$\begin{aligned} \frac{\epsilon}{2} \cdot \frac{r \ln d_A}{d_A} &\geq \frac{\epsilon^2}{16} \cdot \frac{n \ln d_A}{d_A} \\ &\geq \frac{\epsilon^2 n \ln \Delta}{16\Delta} \end{aligned} \quad \text{as } \frac{\ln d_A}{d_A} \text{ is monotonically decreasing for } d_A \geq e$$

We color this independent set with a new color and get a new partial coloring.

Next we show that in the case of $d_A < d_0$ we can use a constant number of additional colors to get a balanced coloring of all but a constant number of vertices of the graph. Notice that since $d_A < d_0$ there are at most $8d_0$ vertices of A of degree $\geq r/4$. Thus by removing $8d_0$ vertices from each side of A we can obtain a subgraph A' with $r - 8d_0$ vertices on each side, average degree $d_{A'} \leq \frac{d_0 r}{r - 8d_0}$ and maximum degree $\Delta_{A'} \leq r/4$. Now as $r \geq \frac{\epsilon}{8} \cdot n \geq \frac{\epsilon}{4} \cdot \Delta$, we have for large enough Δ that $d_{A'} \leq 2d_0$ and $r - 8d_0 > 2\Delta_{A'}$ and hence by claim 3.1 the subgraph A' is colorable, and by theorem 2.1 it can be colored using at most $2d_{A'} + 1 = O(1)$ colors. We remain with a constant number of uncolored vertices, which will be handled in case III.

The remaining case is $r < (1 + \epsilon)d_A$. First we notice that r cannot be very large as

$$r < (1 + \epsilon)d_A \leq (1 + \epsilon)\Delta \leq n/2$$

Now look at the bipartite graph G_r . Recall that the vertices of G_r are $X = \{x_1, \dots, x_r\}$, $Y = \{y_{r+1}, \dots, y_n\}$ and

$$(x_i, y_j) \in E' \Leftrightarrow (u_i, v_j) \in E \text{ or } (v_i, u_j) \in E$$

The average degree of G_r is at most $4(\Delta - d_A)r/n$. (Each of the r vertices of type x in G_r represents two vertices of G and hence can contribute at most 2Δ edges to G_r , for a total of $2\Delta r$. However edges induced by A are not represented in G_r , subtracting $2rd_A$ from the total. The total number of vertices in G_r is n , and the average degree is thus twice the total number of edges divided by n .) Assume without loss of generality that $d(y_{r+1}) \leq d(y_{r+2}) \leq \dots \leq d(y_n)$. Let $G'_r(X, Y', E'')$ be the balanced bipartite subgraph of G_r induced by the vertices in X and $Y' = y_{r+1}, y_{r+2}, \dots, y_{2r}$. As the expected degree in G_r of a vertex chosen uniformly at random from Y is at most $2(\Delta - d_A)\frac{r}{n-r}$ we have that $d(G'_r)$, the average degree of G'_r , is at most

$$2(\Delta - d_A)\frac{r}{n-r} \leq 2\left(\Delta - \frac{r}{1+\epsilon}\right)\frac{r}{n-r}$$

Furthermore

$$\begin{aligned} (1 + \epsilon)d(G'_r) &\leq (1 + \epsilon)2\left(\Delta - \frac{r}{1 + \epsilon}\right)\frac{r}{n - r} \\ &= \frac{(1 + \epsilon)2\Delta - 2r}{n - r} \cdot r \\ &\leq \frac{(1 + \epsilon)2\Delta - 2r}{(1 + \epsilon)2\Delta - r} \cdot r && \text{as } n \geq (1 + \epsilon)2\Delta \\ &\leq r \end{aligned}$$

And thus by Lemma 3.4 if $d(G'_r) \geq d_0$ for large enough d_0 then G'_r contains a balanced bipartite independent set of size at least

$$\begin{aligned}
\frac{\epsilon}{2} \cdot \frac{r \ln d(G'_r)}{d(G'_r)} &\geq \frac{\epsilon r}{2} \cdot \frac{\ln\left(\frac{2\Delta r}{n-r}\right)}{\frac{2\Delta r}{n-r}} && \text{as } d(G'_r) \leq \frac{2\Delta r}{n-r} \\
&= \frac{\epsilon(n-r)}{2} \cdot \frac{\ln\left(\frac{2\Delta r}{n-r}\right)}{2\Delta} \\
&\geq \frac{\epsilon n}{4} \cdot \frac{\ln\left(\frac{2\Delta r}{n}\right)}{2\Delta} && \text{as } r \leq n/2 \\
&\geq \frac{\epsilon n \ln(\epsilon\Delta/4)}{8\Delta} && \text{as } r \geq \epsilon n/8 \\
&\geq \frac{\epsilon n \ln(\Delta)}{9\Delta} && \text{for large enough } \Delta
\end{aligned}$$

If the vertices of this independent set are x_1, \dots, x_k and y_{r+1}, \dots, y_{r+k} we obtain an extended partial coloring of G in the following manner. Color vertices $u_1, \dots, u_k, v_{r+1}, \dots, v_{r+k}$ with a new color, and vertices $v_1, \dots, v_k, u_{r+1}, \dots, u_{r+k}$ with another new color. Notice that vertices u_{r+1}, \dots, u_{r+k} and v_{r+1}, \dots, v_{r+k} were recolored in this process.

The only case left is $d(G'_r) < d_0$. In this case graph G'_r contains at most $8d_0$ vertices of degree $\geq r/4$. Thus by removing $8d_0$ vertices from each side of G'_r we can obtain a subgraph H with $r - 8d_0$ vertices on each side, average degree $d_H \leq \frac{d_0 r}{r - 8d_0}$ and maximum degree $\Delta_H \leq r/4$. Now as $r \geq \frac{\epsilon}{8} \cdot n \geq \frac{\epsilon}{4} \cdot \Delta$, we have for large enough Δ that $d_H \leq 2d_0$ and $r - 8d_0 > 2\Delta_H$ and hence by theorem 2.1 the subgraph H can be colored using at most $2d_H + 1 = O(1)$ colors. Now from the balanced coloring of H and the r -partial coloring of G we can extend the balanced coloring of G , using the recoloring method described above. We remain with a constant number of uncolored vertices and they will be handled in case III.

Case II: $\frac{n}{\Delta^{1/9}} \leq r \leq \frac{\epsilon}{8}n$

Again consider the bipartite graph G'_r which was defined above and recall that $d(G'_r) \leq \frac{2\Delta r}{n-r}$

where $d(G'_r)$ is the average degree of G'_r . Now

$$\begin{aligned}
\left(1 + \frac{3}{4}\epsilon\right) d(G'_r) &\leq \left(1 + \frac{3}{4}\epsilon\right) \frac{2\Delta r}{n-r} \\
&\leq \frac{1 + \frac{3}{4}\epsilon}{1 - \frac{1}{8}\epsilon} \cdot \frac{2\Delta r}{n} \\
&\leq (1 + \epsilon) \cdot \frac{2\Delta r}{n} \\
&\leq r && \text{since } n \geq (1 + \epsilon)2\Delta
\end{aligned}$$

Lemma 3.4 applied with $3\epsilon/4$ serving as ϵ implies that if $d(G'_r) \geq d_0$ for some constant d_0 then G'_r contains a balanced bipartite independent set of size

$$\begin{aligned}
\frac{3\epsilon}{8} \cdot \frac{r \ln d(G'_r)}{d(G'_r)} &\geq \frac{3\epsilon r}{8} \cdot \frac{\ln\left(\frac{2\Delta r}{n-r}\right)}{\frac{2\Delta r}{n-r}} && \text{as } d(G'_r) \leq \frac{2\Delta r}{n-r} \\
&\geq \frac{3\epsilon(n-r)}{8} \cdot \frac{\ln\left(\frac{2\Delta r}{n}\right)}{2\Delta} \\
&\geq \frac{21\epsilon n}{64} \cdot \frac{\ln\left(\frac{2\Delta r}{n}\right)}{2\Delta} && \text{as } r \leq n/8 \\
&\geq \frac{21\epsilon n}{72} \cdot \frac{\ln \Delta}{2\Delta} && \text{as } r \geq n/\Delta^{1/9} \\
&\geq \frac{\epsilon n \ln \Delta}{8\Delta}
\end{aligned}$$

As in Case I we can use this independent set in G'_r to extend the r -partial coloring using two additional colors. The only case left is $d(G'_r) < d_0$ but it can be handled in the exact same manner as we handled it in the analysis of case II.

Case III: $r \leq \frac{n}{\Delta^{1/9}}$

Consider the graph G'_r . Notice that the average degree of G'_r satisfies the following inequalities.

$$\begin{aligned}
d(G'_r) &\leq \frac{2\Delta r}{n-r} \\
&\leq \frac{2n\Delta^{8/9}}{n-r} \\
&\leq 4\Delta^{8/9} && \text{as } r < n/2 \text{ for large enough } \Delta \tag{3.4}
\end{aligned}$$

$$\begin{aligned}
d(G'_r) &\leq \frac{2\Delta r}{n-r} \\
&\leq \frac{2\Delta r}{n(1-1/\Delta^{1/9})} \\
&\leq \frac{r}{(1+\epsilon)(1-1/\Delta^{1/9})} \\
&< r
\end{aligned}
\tag{3.5}$$

for large enough Δ

First suppose that $r \leq 16d(G'_r)$. By inequality 3.5 we have that G'_r contains a balanced bipartite independent set of size at least 1. We can extend the r -partial coloring into an $(r-1)$ -partial coloring using 2 new colors. Applying this process r times we get a balanced coloring of G using additional $2r = O(\Delta^{8/9})$ colors.

Now suppose that $r \geq 16d(G'_r)$. Notice that graph G'_r contains at most $8d(G'_r)$ vertices of degree $\geq r/4$. Thus by removing $8d(G'_r)$ vertices from each side of graph G'_r we can obtain a subgraph H with at least $r/2$ vertices on each side, and maximum degree $\Delta_H \leq r/4$. Hence by theorem 2.1 subgraph H can be colored using at most $2d_H + 1 = O(\Delta^{8/9})$ colors, and graph G can be recolored accordingly. We remain with $8d(G'_r)$ uncolored vertices on each side, and once again we can color them using the procedure described for the case $r \leq 16d(G'_r)$, with at most $O(\Delta^{8/9})$ colors.

Analysis

Let us compute the maximum number of colors used in the three staged procedure described above. Each iteration in stage I or stage II colors at least $\frac{\epsilon^2 n \ln(\Delta)}{9\Delta}$ yet uncolored vertices using one additional color. (Recall that the total number of vertices in a balanced bipartite subgraph is *twice* its size. The constant 9 in the denominator comes from the worst of all subcases. A denominator of 16 appears in one of the subcases, but in this subcase the number of newly colored vertices is twice the size of the balanced bipartite subgraph, and hence it is equivalent to a denominator of 8. In any case, we have not attempted to optimize the constants in this part of the proof.) Thus the number of colors used during these stages is at most

$$2n \cdot \frac{9\Delta}{\epsilon^2 n \ln(\Delta)} + O(1) = \frac{18\Delta}{\epsilon^2 \ln(\Delta)} + O(1)$$

In Stage III we extend the r -partial coloring into a balanced coloring of G using $O(\Delta^{8/9})$ additional colors. For large enough Δ the total number of colors used is at most $\frac{20\Delta}{\epsilon^2 \ln(\Delta)}$ and this concludes the proof of theorem 3.2.

Acknowledgements

This research was supported by The Israel Science Foundation (grant No. 263/02). Part of this work was done while the authors were visiting Microsoft Research in Redmond, Washington.

References

- [1] Amir Ben-Dor, Tzvika Hartman, Richard M. Karp, Benno Schwikowski, Roded Sharan, and Zohar Yakhini. Towards optimally multiplexed applications of universal arrays. *Journal of Computational Biology*, 11(2-3):477–493, 2004.
- [2] Uriel Feige and Shimon Kogan. Hardness of approximation of the balanced complete bipartite subgraph problem. *Technical Report MCS04-04*, the Weizmann Institute, 2004. <http://wisdomarchive.wisdom.weizmann.ac.il:81/archive/00000361/01/04-04.pdf>
- [3] Phillip Hall. On representatives of subsets. *J. London Math Soc.*, 10:26–30, 1935.
- [4] David Karger, Rajeev Motwani, and Madhu Sudan. Approximate Graph Coloring by Semi-definite Programming. *J. ACM* 45(2): 246–265 (1998).
- [5] Sanjeev Khanna, Nathan Linial, and Shmuel Safra. On the hardness of approximating the chromatic number. *Combinatorica*, 20(3):393–415, 2000.
- [6] Subhash Khot. Ruling out PTAS for graph min-bisection, dense k -subgraph and bipartite clique. *SIAM J. Comput.* 36(4): 1025–1071 (2006).
- [7] Michael Molloy and Bruce Reed. *Graph Colouring and the Probabilistic Method*. Springer-Verlag, Paris, 2001.