

Improved Approximation Algorithm for Maximum Balanced Biclique

Pasin Manurangsi  

Google Research, Bangkok, Thailand

Abstract

We study the Maximum Balanced Biclique (MBB) problem: Given a bipartite graph G with n vertices on each side, find a balanced biclique in G with maximum size. We give a polynomial-time $\left(\frac{n}{\Omega((\log n)^3)}\right)$ -approximation algorithm for the problem, which improves upon an $\left(\frac{n}{\Omega((\log n)^2)}\right)$ -approximation by Chalermsook et al. [8] and answers their open question. Furthermore, our approximation ratio matches that of the maximum clique problem by Feige [11] up to an $O(\log \log n)$ factor.

2012 ACM Subject Classification Theory of computation $\rightarrow$ Approximation algorithms analysis; Theory of computation $\rightarrow$ Graph algorithms analysis; Theory of computation $\rightarrow$ Discrete optimization

Keywords and phrases Maximum Balanced Biclique, Approximation Algorithm, Semidefinite Programming

Digital Object Identifier 10.4230/LIPIcs.APPROX/RANDOM.2026.1

Category APPROX

Related Version *Full Version*: <https://arxiv.org/abs/2604.27141>

1 Introduction

The *Maximum Clique* problem – where we are given an undirected graph and the goal is to find a maximum clique – is one of the most well studied problems in all of combinatorial optimization. One of the original 21 NP-hard problems shown by Karp [20], Maximum Clique also played a central role in the foundation for connections between the PCP Theorem and hardness of approximation [12, 2, 3, 5, 18, 21, 9, 23, 29]. Specifically, it is NP-hard to approximate Maximum Clique to within a factor of $n^{1-\epsilon}$ for any constant $\epsilon > 0$ [18, 29]. Under a stronger assumption that $\text{NP} \not\subseteq \text{BPTIME}\left(2^{(\log n)^{O(1)}}\right)$, Khot and Ponnuswami [23] ruled out any polynomial-time algorithm with approximation ratio $n/2^{(\log n)^{3/4+\gamma}}$ for any constant $\gamma > 0$. This remains the best known hardness of approximation for Maximum Clique to date. Meanwhile, many approximation algorithms for Maximum Clique have been devised (e.g. [7, 1, 11]). In particular, Feige’s algorithm [11] achieves the best known approximation ratio of $O\left(n \cdot \frac{(\log \log n)^2}{(\log n)^3}\right)$ in polynomial time.

In this work, we study a bipartite variant of the Maximum Clique problem, called the *Maximum Balanced Biclique (MBB)* problem. Recall that a k -biclique, denoted by $K_{k,k}$, is the complete bipartite subgraph where each side contains k vertices. In MBB, we are given a bipartite graph $G = (U, V, E)$ where $|U| = |V| = n$, and the goal is to find a maximum k -biclique in G .

The exact version of MBB was stated¹ to be NP-hard in Garey and Johnson’s seminal book [16, page 196]. However, despite a considerable amount of effort over the years (e.g. [10, 13, 22, 26, 6, 27]), it is *not* even known to be NP-hard to approximate to within

¹ Originally, the book did not provide a proof, but such a proof was later published e.g. in [19].



© Pasin Manurangsi;

licensed under Creative Commons License CC-BY 4.0

Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM 2026).

Editors: Mohit Singh and Tom Gur; Article No. 1; pp. 1:1–1:9



Leibniz International Proceedings in Informatics

LIPICs Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

a factor of $1 + \epsilon$ for any constant $\epsilon > 0$. Nevertheless, hardness of approximation results are known under other assumptions [10, 13, 22, 26, 6, 27]. Specifically, it is known to be NP-hard (under randomized reductions) to approximate to within a factor of $n^{1-\epsilon}$ for any $\epsilon > 0$, assuming strengthenings of the Unique Games Conjecture [6, 27].

Despite the apparent challenges in proving hardness of approximation results for MBB, not much is known in terms of approximation algorithms either. It is straightforward to see that an “enumeration style” algorithm yields $\left(\frac{n}{\Omega(\log n)}\right)$ -approximation for MBB in polynomial time. Only recently did Chalermsook et al. [8] manage to improve upon this approximation guarantee, obtaining an $\left(\frac{n}{\Omega((\log n)^2)}\right)$ -approximation. As an open question, they asked whether it is possible to improve this guarantee to match the $\left(\frac{n}{\Omega((\log n)^3)}\right)$ -approximation by [11] for the Maximum Clique problem.

1.1 Our Contribution

Our main result is an algorithm that achieves such an approximation guarantee:

► **Theorem 1 (Main Theorem).** *There is a randomized polynomial-time $O\left(n \cdot \left(\frac{\log \log n}{\log n}\right)^3\right)$ -approximation for Maximum Balanced Biclique.*

Our approximation ratio matches the aforementioned algorithm of Feige for the Maximum Clique [11] problem up to an $O(\log \log n)$ factor.

Proof Overview

To explain the proof techniques, let us use k to denote the size of the optimal balanced biclique in G , and let $t = n/k$. We first review the algorithmic guarantee of [8]: If $t \leq (\log n)^{O(1)}$, then they show how to use graph removal technique from [11] on *matchings* in order to find $\tilde{\Omega}((\log n)^2)$ -biclique, as formalized below.

► **Theorem 2 ([8]).** *Given a graph G that contains a k -biclique, there is a polynomial-time algorithm that finds a $\Theta\left(\left(\frac{\log n}{\log t}\right)^2\right)$ -biclique in G where $t = n/k$.*

It is not hard to see that the above theorem immediately yields $O(n/(\log n)^2)$ -approximation. For our purpose, observe further that, if $t \geq \tilde{\Omega}(\log n)$, then the above already yields $\left(\frac{n}{\Omega((\log n)^3)}\right)$ -approximation. Thus, we can focus on the case $t \leq \tilde{O}(\log n)$, i.e. when the optimum k is at least $\frac{n}{\tilde{O}(\log n)}$.

For the case of Maximum Clique, this case can be solved via semi-definite program (SDP) rounding [1, 17] or via subgraph removal [7]. In particular, for $t \leq O\left(\frac{\log n}{\log \log n}\right)$, these algorithms can find a clique of size $n^{1/O(t)}$. Unfortunately, Chalermsook et al.’s [8] technique of reducing MBB to Maximum Clique has a square blow up on the parameter t . This means that one can only apply these Maximum Clique algorithms in a black-box manner² when $t \leq \tilde{O}(\sqrt{\log n})$, which is not sufficient for our purpose.

Instead, our main result is to show that, in the regime $t \leq O\left(\frac{\log n}{\log \log n}\right)$, a more direct SDP relaxation and rounding can yield a balanced biclique of size $n^{1/O(t)}$, as stated below. As mentioned earlier, combining this with Chalermsook et al.’s result (Theorem 2) immediately yields the final approximation guarantee.

² See [28] for a more detailed description on how to interpret [8] as a reduction from MBB to Maximum Clique.

► **Theorem 3** (Main SDP Rounding). *There exists a constant $D > 1$ such that the following holds. If $\{u_i\}, e$ is a feasible solution to the SDP relaxation in Figure 2 with $k \geq n \cdot \frac{D \cdot \log \log n}{\log n}$, there is an efficient rounding algorithm for finding a balanced biclique of at least size $n^{\frac{1}{D-t}}$ with high probability, where $t = n/k$.*

Integrality Gap for A Natural Relaxation

To illustrate a challenge faced when using SDP relaxation for MBB, let us first state the following “natural” extension of Lovasz’s ϑ function [25], which is used³ for Maximum Clique [1, 17]. Here, we associate a vector $u_i \in \mathbb{R}^d$ with each $i \in U \cup V$, and let $e \in \mathbb{R}^d$ be an arbitrary unit vector.

$$\begin{array}{ll}
 \text{Maximize} & k \\
 \text{Subject to} & \|e\|^2 = 1 \\
 & \|u_i\|^2 = \langle u_i, e \rangle \quad \forall i \in U \cup V \\
 & \sum_{i \in U} \langle u_i, e \rangle = \sum_{j \in V} \langle u_j, e \rangle = k \quad (1) \\
 & \langle u_i, u_j \rangle = 0 \quad \forall (i, j) \in (U \times V) \setminus E \\
 & \langle u_i, u_j \rangle \geq 0 \quad \forall (i, j) \in U \times V
 \end{array}$$

■ **Figure 1** SDP Relaxation for MBB (First Attempt).

If G contains $K_{k,k}$, there is a feasible solution (setting vectors in the biclique to e and all others to 0) of value k . We remark that the only difference compared to the Maximum Clique formulation here is the constraint (1), which aims to ensure that the number of vertices picked in U is the same as that in V (so that the biclique is balanced).

Unfortunately, this SDP relaxation is completely unhelpful. We can always find a solution with $k = n/2$: Let f be any unit vector orthogonal to e , and let $u_i = \frac{e+f}{2}$ for all $i \in U$ and $u_j = \frac{e-f}{2}$ for all $j \in V$. This ensures that $\langle u_i, u_j \rangle = 0$ for all $(i, j) \in U \times V$. In other words, the integrality gap is unbounded.

Our SDP Relaxation

Given the integrality gap, we thus use a stronger SDP relaxation for MBB. We again associate a vector $u_i \in \mathbb{R}^d$ with each $i \in U \cup V$, and let $e \in \mathbb{R}^d$ be a unit vector. The SDP relaxation is stated below. Note that we formulate it as a feasibility problem rather than an optimization problem.

We note that the only additional constraints are (5) and (6), which are “fractional degree” constraints aiming to ensure that if i is picked, then k of its neighbor must be picked. Again, if G contains $K_{k,k}$, there is a feasible solution (setting vectors in the biclique to e and all others to 0).

³ Note that the formulation below looks different from one defined in [25] but the latter can be written in this form [24].

$$\begin{aligned}
& \|e\|^2 = 1 \\
& \|u_i\|^2 = \langle u_i, e \rangle \quad \forall i \in U \cup V \\
& \sum_{i \in U} \langle u_i, e \rangle = \sum_{j \in V} \langle u_j, e \rangle = k \quad (2) \\
& \langle u_i, u_j \rangle = 0 \quad \forall (i, j) \in (U \times V) \setminus E \quad (3) \\
& \langle u_i, u_j \rangle \geq 0 \quad \forall (i, j) \in U \times V \quad (4) \\
& \sum_{j \in V} \langle u_i, u_j \rangle = k \langle u_i, e \rangle \quad \forall i \in U \quad (5) \\
& \sum_{i \in U} \langle u_i, u_j \rangle = k \langle u_j, e \rangle \quad \forall j \in V \quad (6)
\end{aligned}$$

■ **Figure 2** SDP Relaxation for MBB.

Our main technical contribution (Theorem 3) is to show that a feasible solution to this SDP can be rounded to a large biclique. It turns out that the rounding procedure is essentially the same as that of [17] in the clique case: After some preprocessing, pick a random Gaussian g and keep a vertex i if its dot product with g is at least a certain threshold. (The full procedure is presented in Figure 3.)

The main difference however is in the analysis. In [17], one simply computes the expected number of vertices that survive the rounding and the expected survived number of non-edge pairs. If the former is much larger than the latter, then (with not too small probability) we can find a large clique by deleting one end of every non-edge pair. This analysis does not work for us: Even if the expected number of survived vertices is large, it is possible that the rounding either keep all vertices in U and no vertex in V , or vice versa. Indeed, this is the situation that occurs with the above integrality gap instance of the SDP in Figure 1.

To overcome this issue, we instead lower bound the expected number of survived *edges*. We do this by showing that, due to the extra constraints (5) and (6), many edges $(i, j) \in E$ are “good” in the sense that $\langle u_i, u_j \rangle$ is not too small. This implies that they are positively correlated in the rounding procedure, which allows us to lower bound the expected number of survived edges. This conclude our proof overview for the rounding procedure, which is formalized in Section 3 after some preliminaries.

2 Preliminaries

Throughout this work, we use $\log$ to denote the natural logarithm.

2.1 Gaussian Tail Bounds

Our rounding scheme relies on probability bounds for (multivariate) normal distributions, which we will state here for convenience. We note that these bounds are standard and are also used in approximation algorithms for Maximum Clique (e.g. [17]). First, let us start with the following lemma:

► **Lemma 4** ([15]). Let $\phi(x) = \frac{1}{\sqrt{2\pi}} \cdot \exp(-x^2/2)$ be the PDF of the standard Gaussian distribution $\mathcal{N}(0, 1)$. If $Z \sim \mathcal{N}(0, 1)$ and $\tau \geq 2$, then $\frac{\phi(\tau)}{2\tau} \leq \Pr[Z \geq \tau] \leq \frac{\phi(\tau)}{\tau}$.

For two standard Gaussian variables X, Y with correlation $\rho = \mathbb{E}[XY]$, we use the following bivariate tail bounds for $\Pr[X \geq \tau, Y \geq \tau]$ for $\tau \geq 2$:

- **Positive Correlation ($\rho \geq 0$):** Since X, Y are positively correlated, the joint probability is lower-bounded by the product of their marginals:

$$\Pr[X \geq \tau, Y \geq \tau] \geq \Pr[X \geq \tau] \Pr[Y \geq \tau] \geq \frac{\phi(\tau)^2}{4\tau^2}, \quad (7)$$

where the last inequality is due to Lemma 4.

- **Negative Correlation ($\rho < 0$):** Notice that $\frac{X+Y}{\sqrt{2+2\rho}} \sim \mathcal{N}(0, 1)$. Thus,

$$\Pr[X \geq \tau, Y \geq \tau] \leq \Pr[X + Y \geq 2\tau] < \frac{1}{\sqrt{2\tau}} \cdot \phi\left(\frac{\sqrt{2} \cdot \tau}{\sqrt{1+\rho}}\right) \leq \frac{\sqrt{\pi}}{\tau} \cdot \phi(\tau)^2 \cdot \exp(\rho \cdot \tau^2) \quad (8)$$

where the first inequality is due to Lemma 4 and $\rho < 0$, and the second inequality follows from $\rho < 0$.

2.2 Biclique Extraction Lemma

Before we describe the rounding, let us also state a self-contained combinatorial step of extracting a balanced biclique from a sufficiently dense surviving subgraph.

► **Lemma 5** (Greedy Extraction). Let $H = (U_H, V_H, E_H)$ be a bipartite graph with at most n vertices on each side. Let F be the total number of edges (i.e. $F = |E_H|$) and Q be the total number of non-edges in H (i.e. $Q = |U_H| \cdot |V_H| - |E_H|$). If there exists an integer $r \geq 1$ such that $F - 2rQ \geq 2nr$, then there is a polynomial-time algorithm that finds a $K_{r,r}$ in H .

Proof. Let $d_F(i)$ and $d_Q(i)$ be the degrees of a vertex i in the edges and non-edges of H , respectively. We apply the following two-stage procedure:

1. **Density Cleaning:** While there exists any vertex i such that $d_F(i) \leq 2r \cdot d_Q(i)$, delete i from H .

To analyze this procedure, we track $W = F - 2rQ$. Upon deleting i , the change in W is $-d_F(i) + 2r \cdot d_Q(i) \geq 0$. Thus, the final cleaned graph $H' = (U_{H'}, V_{H'}, E_{H'})$ satisfies $W' \geq 2nr$.

Since $W' \leq |E_{H'}| \leq |U_{H'}| \cdot |V_{H'}|$ and we assume that each side of H is of size at most n , we can conclude that $|U_{H'}|, |V_{H'}|$ are both at least $2r$.

2. **Biclique Construction:** We arbitrarily select exactly r vertices from $U_{H'}$ to form U_{cliq} . For each selected $u \in U_{cliq}$, we delete its non-neighbors from $V_{H'}$; from how H' is constructed, there are at most $\frac{|V_{H'}|}{2r}$ such non-neighbors. Thus, the total number of vertices deleted from $V_{H'}$ is at most $r \left(\frac{|V_{H'}|}{2r} \right) = \frac{|V_{H'}|}{2}$. Consequently, the number of vertices surviving in $V_{H'}$ is at least $\frac{|V_{H'}|}{2}$. Because we established $|V_{H'}| \geq 2r$, at least $\frac{2r}{2} = r$ vertices remain. We arbitrarily select r of these surviving vertices to form V_{cliq} . The set $U_{cliq} \cup V_{cliq}$ is the desired balanced biclique. ◀

We remark that, for the regime $\Theta(1) \leq r \leq n^{1-\Omega(1)}$, known techniques [14, 4] can be used in the “Biclique Construction” step to improve the bound from r to $\Omega(r \log n)$. However, this does not affect the overall approximation ratio of our algorithm and, thus, we opt for the argument above for simplicity.

3 The Rounding Algorithm and Proof of Theorem 3

The full SDP rounding procedure is presented in Figure 3.

1. **SDP:** Let $\{u_i\}, e$ be vectors satisfying Figure 2. Let $t = n/k$, and $c_i = \langle u_i, e \rangle$ for all $i \in U \cup V$.
2. **Heavy Sets:** Let $A = \{i \in U \mid c_i \geq \frac{1}{8t}\}$ and $B = \{j \in V \mid c_j \geq \frac{1}{8t}\}$.
3. **Shifting:** Let $\alpha = 1 - 1/\sqrt{2}$, and let $u'_i = u_i - \alpha c_i e$ for all $i \in A \cup B$.
4. **Gaussian Thresholding:** Sample $g \sim \mathcal{N}(0, I_d)$. Keep vertices $i \in A \cup B$ iff $X_i := \langle g, \frac{u'_i}{\|u'_i\|} \rangle \geq \tau$, where $\tau = \sqrt{0.1 \log n}$. Let the surviving induced subgraph be H .
5. **Biclique Extraction:** Run the procedure from Lemma 5 on H , and output the result.

■ **Figure 3** SDP Rounding Algorithm for MBB.

Proof of Theorem 3. We will next analyze the guarantee of the rounding algorithm. To do so, let $D = 1000$ and assume that $t \leq \frac{\log n}{D \cdot \log \log n}$.

For any subsets $S \subseteq U$ and $T \subseteq V$, let $M(S, T) := \sum_{(i,j) \in S \times T} \langle u_i, u_j \rangle$. We have

$$\begin{aligned}
 M(A, B) &= M(U, V) - M(U \setminus A, V) - M(U, V \setminus B) + M(U \setminus A, V \setminus B) \\
 &\stackrel{(4)}{\geq} M(U, V) - M(U \setminus A, V) - M(U, V \setminus B) \\
 &= \sum_{i \in U} \left(\sum_{j \in V} \langle u_i, u_j \rangle \right) - \sum_{i \in U \setminus A} \left(\sum_{j \in V} \langle u_i, u_j \rangle \right) - \sum_{j \in V \setminus B} \left(\sum_{i \in U} \langle u_i, u_j \rangle \right) \\
 &\stackrel{(5),(6)}{=} \sum_{i \in U} k \cdot c_i - \sum_{i \in U \setminus A} k \cdot c_i - \sum_{j \in V \setminus B} k \cdot c_j \\
 &\stackrel{(2)}{=} k^2 - \sum_{i \in U \setminus A} k \cdot c_i - \sum_{j \in V \setminus B} k \cdot c_j \\
 &\stackrel{(*)}{\geq} k^2 - n \cdot k \cdot \frac{1}{8t} - n \cdot k \cdot \frac{1}{8t} = \frac{3}{4} k^2,
 \end{aligned} \tag{9}$$

where $(*)$ follows from our definition of A, B .

Let $E^+ = \{(i, j) \in A \times B \mid \langle u_i, u_j \rangle > \frac{1}{2} c_i c_j\}$.

Note that $\sum_{(i,j) \in A \times B} c_i c_j \leq (\sum_U c_i)(\sum_V c_j) \stackrel{(2)}{=} k^2$. Hence,

$$\sum_{(i,j) \in A \times B} \left(\langle u_i, u_j \rangle - \frac{1}{2} c_i c_j \right) \geq M(A, B) - \frac{k^2}{2} \stackrel{(9)}{\geq} \frac{k^2}{4}$$

Any pair $(i, j) \notin E^+$ contributes a non-positive value to this sum. This, together with the fact that $\langle u_i, u_j \rangle \leq 1$ for all $i, j \in U \cup V$, implies that $|E^+| \geq \frac{k^2}{4}$. Note also that, by (3), all pairs $(i, j) \in E^+$ must be an edge.

Next, consider any $i, j \in U \cup V$. We have $\langle u'_i, u'_j \rangle = \langle u_i, u_j \rangle - \frac{1}{2} c_i c_j$. Note that this also implies $\|u'_i\|_2^2 = c_i(1 - \frac{1}{2} c_i) \leq c_i$. We next analyze the expected number of edges and non-edges in H . To do this, let F denote the number of surviving edges in H , Q denote the number of non-edges in H , and let

$$r = \frac{1}{64\sqrt{\pi} \cdot \tau \cdot t^2} \cdot \exp\left(\frac{\tau^2}{16t}\right).$$

Note that, for sufficiently large n , we have

$$r \geq \frac{1}{O((\log n)^3)} \cdot n^{\frac{1}{160t}} \geq n^{\frac{1}{D \cdot t}}. \quad (10)$$

- **Lower Bounding $\mathbb{E}[F]$:** For $(i, j) \in E^+$, since $\langle u_i, u_j \rangle \geq \frac{1}{2}c_i c_j$, we have $\langle u'_i, u'_j \rangle \geq 0$. Thus, $\mathbb{E}[X_i X_j] \geq 0$. Applying (7), the probability that (i, j) is kept in H is at least $\frac{\phi(\tau)^2}{4\tau^2}$. Hence,

$$\mathbb{E}[F] \geq |E^+| \cdot \frac{\phi(\tau)^2}{4\tau^2} \geq \frac{k^2}{4} \cdot \frac{\phi(\tau)^2}{4\tau^2}. \quad (11)$$

Since $k = n/t$ and due to our bound on t , for any sufficiently large n , we have

$$\mathbb{E}[F] \geq \frac{n^2}{O((\log n)^3)} \cdot \exp(-0.1 \log n) \geq n^{1.8} \geq n \cdot \exp\left(\frac{\tau^2}{16t}\right) \geq 8nr. \quad (12)$$

- **Upper Bounding $\mathbb{E}[Q]$:** Consider any $(i, j) \in (A \times B) \setminus E$. By (3), $\langle u_i, u_j \rangle = 0$, which implies $\langle u'_i, u'_j \rangle = -\frac{1}{2}c_i c_j$. Normalizing bounds the correlation:

$$\mathbb{E}[X_i X_j] = \left\langle \frac{u'_i}{\|u'_i\|}, \frac{u'_j}{\|u'_j\|} \right\rangle \leq \frac{-\frac{1}{2}c_i c_j}{\sqrt{c_i c_j}} = -\frac{1}{2}\sqrt{c_i c_j} \leq -\frac{1}{16t},$$

where the last inequality follows from definition of A, B . Applying (8), we have

$$\mathbb{E}[Q] \leq n^2 \cdot \frac{\sqrt{\pi}}{\tau} \cdot \phi(\tau)^2 \cdot \exp\left(-\frac{\tau^2}{16t}\right) \stackrel{(11)}{=} \frac{1}{4r} \cdot \mathbb{E}[F]. \quad (13)$$

Thus, we have

$$\mathbb{E}[F - 2rQ] \stackrel{(13)}{\geq} \mathbb{E}[F]/2 \stackrel{(12)}{\geq} 4nr.$$

Since $F - 2rQ$ has value at most n^2 , we have $F - 2rQ \geq 2nr$ with probability at least $1/n^2$. When this occurs, Lemma 5 guarantees an r -biclique which satisfies the desired size bound due to (10). To achieve high probability, we simply repeat the rounding process n^3 times. This concludes our proof. ◀

4 Putting Things Together: Proof of Theorem 1

Finally, we prove our main theorem (Theorem 1) by combining our SDP rounding (in the large optimum case) with the algorithm of Chalmers et al. from Theorem 2 (in the small optimum case).

Proof of Theorem 1. We solve the SDP from Figure 2 for $k = \left\lceil n \cdot \frac{10 \cdot D \cdot \log \log n}{\log n} \right\rceil$. If the SDP is feasible, then we simply run the rounding algorithm from Theorem 3. This ensures that we find a biclique of size at least $(\log n)^{10}$, which yields $n/(\log n)^{10} \leq O\left(n \cdot \left(\frac{\log \log n}{\log n}\right)^3\right)$ -approximation.

If the SDP is infeasible, then the maximum balanced biclique has size at most $O\left(n \cdot \frac{\log \log n}{\log n}\right)$. We run Chalmers et al.'s algorithm (Theorem 2). If the maximum balanced biclique has size at least $n/\log^3 n$, we have $t \leq \log^3 n$ and thus the algorithm produces $\Theta\left(\left(\frac{\log n}{\log \log n}\right)^2\right)$ -biclique, which yields $O\left(n \cdot \left(\frac{\log \log n}{\log n}\right)^3\right)$ -approximation. On the other hand, if the maximum balanced biclique has size at most $n/\log^3 n$, we automatically get $n/\log^3 n \leq O\left(n \cdot \left(\frac{\log \log n}{\log n}\right)^3\right)$ -approximation. ◀

5 Conclusion and Open Questions

In this work, we give a polynomial-time $\left(\frac{n}{\Omega((\log n)^3)}\right)$ -approximation algorithm for the Maximum Balanced Biclique problem. Our algorithm is via SDP rounding in the case where the optimum is large, combined with the known result of Chalmers et al. [8] in the case where the optimum is small. Apart from trying to improve the approximation ratio further, it remains an interesting question whether the former case can be handled directly via combinatorial algorithms. We remark that, in the Maximum Clique case, subgraph removal techniques—which are completely combinatorial—suffice for this case [7].

Another intriguing question is whether one can achieve a similar approximation for the Maximum *Edge* Biclique (MEB) problem. In MEB, we are looking for a (not necessarily balanced) biclique in G that contains the maximum number of edges. A trivial algorithm yields n -approximation, and a simple “enumeration style” algorithm can improve this to $\frac{n}{\Omega(\log n)}$. However, we are not aware of any further improvement to this. In particular, it is unclear how to apply SDP rounding when the optimum is large.

References

- 1 Noga Alon and Nabil Kahalé. Approximating the independence number via the theta-function. *Math. Program.*, 80:253–264, 1998. doi:10.1007/BF01581168.
- 2 Sanjeev Arora, Carsten Lund, Rajeev Motwani, Madhu Sudan, and Mario Szegedy. Proof verification and the hardness of approximation problems. *J. ACM*, 45(3):501–555, 1998. doi:10.1145/278298.278306.
- 3 Sanjeev Arora and Shmuel Safra. Probabilistic checking of proofs: A new characterization of NP. *J. ACM*, 45(1):70–122, 1998. doi:10.1145/273865.273901.
- 4 Maria Axenovich, Jean-Sébastien Sereni, Richard Snyder, and Lea Weber. Bipartite independence number in graphs with bounded maximum degree. *SIAM J. Discret. Math.*, 35(2):1136–1148, 2021. doi:10.1137/20M1321760.
- 5 Mihir Bellare, Oded Goldreich, and Madhu Sudan. Free bits, pcps, and nonapproximability-towards tight results. *SIAM J. Comput.*, 27(3):804–915, 1998. doi:10.1137/S0097539796302531.
- 6 Amey Bhangale, Rajiv Gandhi, Mohammad Taghi Hajiaghayi, Rohit Khandekar, and Guy Kortsarz. Bi-covering: Covering edges with two small subsets of vertices. *SIAM J. Discret. Math.*, 31(4):2626–2646, 2017. doi:10.1137/16M1082421.
- 7 Ravi B. Boppana and Magnús M. Halldórsson. Approximating maximum independent sets by excluding subgraphs. In John R. Gilbert and Rolf G. Karlsson, editors, *SWAT*, pages 13–25, 1990. doi:10.1007/3-540-52846-6_74.
- 8 Parinya Chalmers, Wanchote Po Jiamjitrak, and Ly Orgo. On finding balanced bicliques via matchings. In *WG 2020*, pages 238–247, 2020. doi:10.1007/978-3-030-60440-0_19.
- 9 Lars Engebretsen and Jonas Holmerin. Towards optimal lower bounds for clique and chromatic number. *Theor. Comput. Sci.*, 299(1-3):537–584, 2003. doi:10.1016/S0304-3975(02)00535-2.
- 10 Uriel Feige. Relations between average case complexity and approximation complexity. In *STOC*, pages 534–543, New York, NY, USA, 2002. ACM. doi:10.1145/509907.509985.
- 11 Uriel Feige. Approximating maximum clique by removing subgraphs. *SIAM J. Discret. Math.*, 18(2):219–225, 2004. doi:10.1137/S089548010240415X.
- 12 Uriel Feige, Shafi Goldwasser, László Lovász, Shmuel Safra, and Mario Szegedy. Interactive proofs and the hardness of approximating cliques. *J. ACM*, 43(2):268–292, 1996. doi:10.1145/226643.226652.
- 13 Uriel Feige and Shimon Kogan. Hardness of approximation of the balanced complete bipartite subgraph problem. Technical report, Weizmann Institute of Science, Rehovot, Israel, 2004.

- 14 Uriel Feige and Shimon Kogan. Balanced coloring of bipartite graphs. *J. Graph Theory*, 64(4):277–291, 2010. doi:10.1002/JGT.20456.
- 15 William Feller. *An introduction to probability theory and its applications, Volume 2*, volume 2. John Wiley & Sons, 1991.
- 16 Michael R. Garey and David S. Johnson. *Computers and Intractability: A Guide to the Theory of NP-Completeness*. W. H. Freeman & Co., New York, NY, USA, 1979.
- 17 Eran Halperin. Improved approximation algorithms for the vertex cover problem in graphs and hypergraphs. *SIAM J. Comput.*, 31(5):1608–1623, 2002. doi:10.1137/S0097539700381097.
- 18 Johan Håstad. Clique is hard to approximate within $n^{1-\epsilon}$. In *FOCS*, pages 627–636, 1996. doi:10.1109/SFCS.1996.548522.
- 19 David S. Johnson. The np-completeness column: An ongoing guide. *J. Algorithms*, 8(5):438–448, September 1987. doi:10.1016/0196-6774(87)90021-6.
- 20 Richard M. Karp. Reducibility among combinatorial problems. In *Proceedings of a symposium on the Complexity of Computer Computations*, pages 85–103, 1972. doi:10.1007/978-1-4684-2001-2_9.
- 21 Subhash Khot. Improved inapproximability results for maxclique, chromatic number and approximate graph coloring. In *FOCS*, pages 600–609, 2001. doi:10.1109/SFCS.2001.959936.
- 22 Subhash Khot. Ruling out ptas for graph min-bisection, dense k-subgraph, and bipartite clique. *SIAM Journal on Computing*, 36(4):1025–1071, 2006. doi:10.1137/S0097539705447037.
- 23 Subhash Khot and Ashok Kumar Ponnuswami. Better inapproximability results for maxclique, chromatic number and min-3lin-deletion. In *ICALP*, pages 226–237, 2006. doi:10.1007/11786986_21.
- 24 Jon M. Kleinberg and Michel X. Goemans. The lovász theta function and a semidefinite programming relaxation of vertex cover. *SIAM J. Discret. Math.*, 11(2):196–204, 1998. doi:10.1137/S0895480195287541.
- 25 László Lovász. On the shannon capacity of a graph. *IEEE Trans. Inf. Theory*, 25(1):1–7, 1979. doi:10.1109/TIT.1979.1055985.
- 26 Pasin Manurangsi. Almost-polynomial ratio eth-hardness of approximating densest k-subgraph. In *STOC*, pages 954–961, 2017. doi:10.1145/3055399.3055412.
- 27 Pasin Manurangsi. Inapproximability of maximum biclique problems, minimum k -cut and densest at-least- k -subgraph from the small set expansion hypothesis. *Algorithms*, 11(1):10, 2018. doi:10.3390/A11010010.
- 28 Pasin Manurangsi. Answer to “approximation algorithm for balanced bipartite independent set?”. Theoretical Computer Science Stack Exchange, August 2022. Online; accessed 14-April-2026. URL: <https://cstheory.stackexchange.com/a/51812>.
- 29 David Zuckerman. Linear degree extractors and the inapproximability of max clique and chromatic number. *Theory Comput.*, 3(1):103–128, 2007. doi:10.4086/T0C.2007.V003A006.