On the Biclique problem in Bipartite graphs

Milind Dawande Pinar Keskinocak Sridhar Tayur

GSIA Working Paper 1996-04

Graduate School of Industrial Administration
Carnegie Mellon University, Pittsburgh, PA 15213

July 1996

Abstract

We consider the maximum *edge* biclique problem in bipartite graphs. We show that the weighted version is NP-complete. We are unable to show whether the unweighted version (MBP) is easy or hard; however, we show that four variants of MBP are NP-complete. For random bipartite graphs, we obtain two results about the size of the maximum balanced biclique and the maximum edge cardinality biclique, thus highlighting the difference between these two problems. We provide approximations using linear programming and the polynomially solvable maximum node cardinality biclique problem. We obtain a 2-approximation algorithm for dense random bipartite graphs and provide a polynomial time algorithm for a special case of MBP.
Key Words: Complexity; bipartite graphs; approximations; cliques.

1 Introduction

In this paper, we consider the maximum (edge) weighted biclique problem in bipartite graphs. Given a bipartite graph $B = (V_1 \cup V_2, E)$, a *biclique* $C = U_1 \cup U_2$ is a subset of the node set, such that $U_1 \subseteq V_1$, $U_2 \subseteq V_2$ and for every $u \in U_1$, $v \in U_2$ the edge $(u, v) \in E$. (In other words, a biclique is a complete bipartite subgraph of B.) *Maximum edge cardinality biclique* (MBP) in B is a biclique C with a maximum number of edges. In an edge weighted bipartite graph B, there is a weight w_{uv} associated with each edge (u, v). A *maximum edge weight* (MWBP) biclique is a biclique C, where the sum of the edge weights in the subgraph induced by C is maximum.

We first note some known results for related problems. The maximum node weight biclique problem is polynomially solvable [6]. (In a node weighted bipartite graph B, there is a weight w_v associated with each node v.) Hence, the maximum node cardinality biclique problem is also polynomially solvable. A restricted version of these problems, where there is an additional requirement that $|U_1| = |U_2|$, is called the *maximum balanced node cardinality biclique* problem (MBBP). MBBP is shown to be NP-complete [6]. The *maximum clique* problem is one of the most well known and widely studied NP-complete problems in the literature. Given a graph $G = (V, E)$, a *clique* (or a complete subgraph) C is a subset of the node set, such that for every pair of nodes $u, v \in C$, the edge (u, v) exists. A *maximum clique* in G is a clique with the maximum number of nodes. In the weighted version of the maximum clique problem, there is a weight $w(v)$ associated with each node v and the weight $W(C)$ of a clique C is the sum of the weights of the nodes in C.

MWBP and MBP are motivated by applications in manufacturing. Consider a set of components $V_1 = \{1, \ldots, n\}$ and a set of products $V_2 = \{1, \ldots, m\}$. The relationship between these products and components can be modeled on a bipartite graph B with node set $V_1 \cup V_2$ and edge set E, such that $(i, j) \in E$ if and only if component i is part of product j. Several products share one or more common components. One way of reducing the lead times perceived by the customers for these products is to reduce the final assembly time; this reduction in the final assembly time is obtained by creating subassemblies (or *vanilla box*es) in advance (see [8] for details). A vanilla box U_1 containing parts $i_1, \ldots, i_k$ can be used *only* in products which contain all of these parts. In other words, the set of products $U_2 = \{j : (i_l, j) \in E,\ l = 1, \ldots, k\}$ can use the vanilla box U_1. Let t_i be the assembly time of component i (assuming the assembly time of component i is the same in every product; if it is different, then we will use t_{ij} as the assembly time of component i in product j). If the total assembly time of the components in the vanilla box U_1 is T, then we can obtain a reduction of T in the lead times of all the products in U_2 (by having enough inventory of these vanilla boxes). On the other hand, to obtain a large T, we have to include many parts in the vanilla box, which will usually decrease the number of products which can use the vanilla box (size of U_2). Then, there is a trade-off between constructing a large vanilla box and using it in many products. The problem of finding a "good" vanilla box can be modeled by finding a maximum edge weight biclique in the bipartite graph B. If all the parts have (approximately) the same assembly time, as in some production lines of IBM, the problem reduces to the maximum edge cardinality biclique problem (MBP).

Another area where MBP occurs is the *formal concept analysis* [4] [5]. Consider two sets V_1 and V_2 (the set of "attributes" and the set of "objects") and a relation R between V_1 and

V_2 ($(i,j) \in R$ if object j has attribute i). For subsets $P \subset V_1$ and $Q \subset V_2$, let

P' = the set of all objects which have all the attributes in Q, and

Q' = the set of all attributes which all the elements of P have.

Then, a *formal concept* of (V_1, V_2, R) is a pair (P, Q) such that $P \subset V_1$, $Q \subset V_2$, $P' = Q$ and $Q' = P$. We can associate $V_1 \cup V_2$ with the node set of a bipartite graph B and the relation R defines the edge set E. Then the concepts are the bicliques of B. The set of all concepts of $(V_1 \cup V_2, R)$, ordered by subconcept/superconcept, forms a complete lattice and is studied in the framework of formal concept analysis.

The paper is organized as follows. In Section 2, we first show that MWBP is NP-complete. We are unable to show whether MBP is NP-complete or not. However, we show that the following three decision problems related to MBP are NP-complete: 1) Is there a biclique with node set $U_1 \cup U_2$ in B, such that $|U_1| = a$ and $|U_2| = b$? 2) Is there a biclique with node set $U_1 \cup U_2$ in B, such that $|U_1| = a$? 3) Is there a biclique in B with exactly k edges ?. We conclude Section 2 by showing the NP-completeness of an optimization problem related to MBP. In section 3, we study random graphs and show the difference between MBP and MBBP. In section 4, we provide two approximations for MBP (including one that is a 2-approximation for dense graphs) and consider a special case of MBP that is polynomially solvable.

2 Complexity Results

Theorem 1 *MWBP is NP-complete.*

Proof We prove this by a reduction from the maximum clique problem. Let $G = (V, E)$ be a graph with node set V and edge set E. Create a bipartite graph $B(G) = (V_1 \cup V_2, E')$ from G, such that $V_1 = V_2 = V$ and $(i,j) \in E'$ (for $i \in V_1$ and $j \in V_2$) if and only if $i = j$ or $(i,j) \in E$. Let the edges (i,i) of $B(G)$ have weight 1 and all the other edges have weight zero.
With the edge weights as defined, there is a maximum weight biclique $U_1 \cup U_2$ in $B(G)$, such that $i \in U_1$ if and only if $i \in U_2$ (i.e. $|U_1| = |U_2|$ and the biclique is "symmetric"). Such a maximum weight "symmetric" biclique can be obtained easily by deleting the nodes $i \in U_1$, $i \notin U_2$ and $i \in U_2$, $i \notin U_1$ from a maximum weight biclique. It follows that if C is a maximum clique in G, then $U_1 \cup U_2$, where $U_1 = U_2 = C$, induces a maximum weight biclique in $B(G)$. Similarly, if $U_1 \cup U_2$ is a symmetric maximum weight biclique in $B(G)$,

then $C = U_1 = U_2$ is a maximum clique in G. □

Next, we consider three decision problems and an optimization problem, which are related to MBP:

- **Exact balanced node cardinality decision problem (EBNCD):** Given a bipartite graph $G = (V_1 \cup V_2, E)$ and a positive integer $a \in Z_+$, does there exist a biclique $C = U_1 \cup U_2$ with $|U_1| = |U_2| = a$?
- **Exact node cardinality decision problem (ENCD):** Given a bipartite graph $G = (V_1 \cup V_2, E)$ and two positive integers $a, b \in Z_+$, does there exist a biclique $C = U_1 \cup U_2$ with $|U_1| = a$ and $|U_2| = b$?
- **Exact edge cardinality decision problem (EECD):** Given a bipartite graph $G = (V_1 \cup V_2, E)$ and a positive integer $k \in Z_+$, does there exist a biclique with exactly k edges?
- **Maximum One-sided edge cardinality problem (MOFCP):** Given a bipartite graph $G = (V_1 \cup V_2, E)$ and a positive integer $k \in Z_+$, find a maximum cardinality biclique with exactly k nodes on one side of the bipartition.

Lemma 2.1 *EBNCD and ENCD are NP-Complete.*

Proof It is known that the maximum balanced node cardinality biclique problem (MBBP) is NP-complete [6]. Then, it follows that EBNCD is NP-complete, since MBBP can be solved using a polynomial number of instances of EBNCD. Note that EBNCD is just a special case of ENCD and hence ENCD is also NP-complete. □

Theorem 2.2 *EECD is NP-complete.*

To prove this theorem, first define the following decision problem and show that it is NP-complete:

Exact balanced prime node cardinality decision problem (EBPNCD): Given a bipartite graph $G = (V_1 \cup V_2, E)$ and a prime number p, such that the maximum degree in G is less than p^2, does there exist a biclique $C = U_1 \cup U_2$, with $|U_1| = |U_2| = p$?

Lemma 2.3 *EBPNCD is NP-complete.*

Proof Given an instance of EBNCD, let $l = \max\{|V_1|, |V_2|\} + 1$ and p be any prime number such that $l \leq p \leq 2l$. Such a prime number is guaranteed by *Bertrand's Theorem* [7]. Let a (the specification for EBNCD) $\leq p$ be a positive integer. Add $p - a$ nodes on both sides of the bipartition and connect each of these additional nodes to all the nodes on the opposite side of the bipartition. The maximum degree of any node in this graph is $p + \max\{|V_1|, |V_2|\} \leq 3(\max\{|V_1|, |V_2|) + 1$. Since $p^2 \geq (\max\{|V_1|, |V_2|\} + 1)^2$ it follows that $p^2 > 3(\max\{|V_1|, |V_2|) + 1$ for $\max\{|V_1|, |V_2|\} > 1$. Then, EBCND has a yes (no) answer if and only if EBPNCD has a yes (no) answer, implying that EBPNCD is NP-complete. □

Now we prove Theorem 2.2.

Proof of Theorem 2.2 Consider a bipartite graph $G = (V_1 \cup V_2, E)$ and an instance of EBPNCD. A biclique of edge cardinality p^2 in G can be possible only in two ways: (1) One node on one side of the bipartition and p^2 nodes on the other side and (2) Exactly p nodes on both sides of the bipartition. Since the maximum degree in G is strictly less than p^2, the first case is not possible. Thus, EBPNCD has a yes (no) answer if and only if an instance of EECD with $k = p^2$ has a yes (no) answer. Then it follows that EECD is NP-complete since EBPNCD is. □

Our next result is about the complexity of the optimization problem MOFCP:

Theorem 2.4 *MOFCP is NP-complete.*

To prove Theorem 2.4, we will define the following two decision problems:

1) **Minimum fixed union problem (MFUP):** Given $k \in Z_+$, a ground set V and a set system $\mathcal{S} = \{S_1, ..., S_n\}$, where the S_i's are subsets of V, find k subsets from $\mathcal{S}$ such that their union has minimum cardinality.
2) **Maximum fixed intersection problem (MFIP):** Given $k \in Z_+$, a ground set V and a set system $\mathcal{S} = \{S_1, ..., S_n\}$, where the S_i's are subsets of V, find k subsets from $\mathcal{S}$ such that their intersection has maximum cardinality.

Lemma 2.5 *MFUP and MFIP are NP-complete.*

Proof It is well known that the decision problem CLIQUE: "Given a graph $G = (V, E)$ and a positive integer k, does there exist a clique of size k in G?" is NP-complete [6]. Given an instance of CLIQUE, construct the following set system on the ground set V. For each edge $e = (u, v)$ in E, construct one set $S_e = \{u, v\}$. Let $\mathcal{S} = \{S_e\colon e \in E\}$. There exists a clique on k nodes in G if and only if there exists $p = \frac{k(k-1)}{2}$ subsets in $\mathcal{S}$ whose union has cardinality at most k. Thus, there exists a clique of size k in G if and only if the cardinality of the minimum union in the optimal solution to MFUP of p sets is k. The NP-completeness of MFIP is immediate since MFIP is the same as MFUP on the set system obtained by taking the complement of each set in $\mathcal{S}$. $\square$

Proof of Theorem 2.4 Consider an instance of MFIP. Construct a bipartite graph $G = (V_1 \cup V_2, E)$ as follows: For each set S_i is $\mathcal{S}$, create a node i in V_1, for each element j of the base set V, create a node j in V_2. For every element $j \in S_i$, include an edge $e = (i, j)$. Note that the maximum edge cardinality biclique with exactly k nodes in V_1 solves MFIP. $\square$

3 Maximum Bicliques in Random Graphs

In this section we present two results, one about the size of the maximum balanced biclique and the other on the size of the maximum edge cardinality biclique in random bipartite graphs.

Given a random bipartite graph $B = (V_1 \cup V_2, p)$, where $|V_1| = |V_2| = n$, we show that the maximum balanced biclique will be of size $a(n) \times a(n)$, where $a(n) \approx \frac{\log n}{\log \frac{1}{p}}$, and the maximum edge cardinality biclique will be of size $k \times n^{\frac{1}{3}}$ (for some constant k) with high probability (an $a \times b$ biclique is a biclique in B with a nodes belonging to V_1 and b nodes belonging to V_2). These two results suggest that the number of edges in the maximum edge cardinality biclique may be considerably larger than the number of edges in the maximum balanced biclique.

Theorem 3.1 *Given a random bipartite graph $B = (V_1 \cup V_2, p)$, where $|V_1| = |V_2| = n$, $a(n) < \frac{\log n}{\log \frac{1}{p}}$ and $a(n) \to \frac{\log n}{\log \frac{1}{p}}$ as $n \to \infty$, the maximum balanced biclique will be of size $a(n) \times a(n)$ with high probability.*

Proof Let Z=number of $a \times a$ bicliques in G. First, we need to show that the probability of having a balanced biclique of size $a \times a$ is very small if $a > a(n)$. We will use the fact that Prob$[Z \geq 1] \leq E[Z]$.

$$\begin{aligned} \text{Prob } [Z \geq 1] &\leq \binom{n}{a}^2 p^{a^2} \\ &\leq (\frac{n^a}{a!})^2 p^{a^2} \end{aligned}$$

$$\leq \quad O(n^{-1}) \quad \text{for} \quad a \geq \frac{\log\ n}{\log\ \frac{1}{p}}. \tag{1}$$

The last inequality follows from Sterling's formula. Now, we need to show that with high probability there is a balanced biclique of size $a(n) \times a(n)$ in B, i.e. Prob$[Z = 0/a = a(n)]$ is very small. From the *second moment method* [1], we have Prob $[Z = 0] \leq \frac{Var(Z)}{(E(Z))^2}$.

$$\begin{aligned}
E(Z) &= \binom{n}{a}\binom{n}{a}p^{a^2} \\
E(Z^2) &= \sum_{A,B}\sum_{A',B'} \text{Prob } [Z_{A,B} = 1, Z_{A',B'} = 1] \\
&= \sum_{A,B}\sum_{A',B'} \text{Prob } [Z_{A',B'} = 1/Z_{A,B} = 1] \text{ Prob } [Z_{A,B} = 1]
\end{aligned}$$

Since all the (A, B) look alike, fix (A, B) as $(\bar{A}, \bar{B})$

$$\begin{aligned}
E(Z^2) &= \sum_{A,B}\sum_{A',B'} \text{Prob } [Z_{A',B'} = 1/Z_{\bar{A},\bar{B}} = 1] \text{ Prob } [Z_{A,B} = 1] \\
&= \sum_{A,B} \text{Prob } [Z_{A,B} = 1] \sum_{A',B'} \text{Prob } [Z_{A',B'} = 1/Z_{\bar{A},\bar{B}} = 1]
\end{aligned}$$

Letting $\sum_{A',B'}$ Prob $[Z_{A',B'} = 1/Z_{\bar{A},\bar{B}} = 1] = \Delta$, we get $E(Z^2) = E(Z)\ \Delta$, where

$$\begin{aligned}
\Delta &= \sum_{A',B'} \text{Prob } [Z_{A',B'} = 1/Z_{\bar{A},\bar{B}} = 1] \\
&= \sum_{i=0}^{a}\sum_{j=0}^{a} \sum_{\substack{|A' \cap \bar{A}| = i \\ |B' \cap \bar{B}| = j}} \text{Prob } [Z_{A',B'} = 1/Z_{\bar{A},\bar{B}} = 1] \\
&= \sum_{i=0}^{a}\sum_{j=0}^{a} \binom{a}{i}\binom{n-a}{a-i}\binom{a}{j}\binom{n-a}{a-j}p^{a^2-ij}
\end{aligned}$$

and

$$\text{Prob } [\ Z = 0] \quad \leq \quad \frac{Var(Z)}{(E(Z))^2} = \frac{\Delta}{E(Z)} - 1.$$

We can write

$$\frac{\Delta}{E(Z)} \quad = \quad \sum_{i=0}^{a}\sum_{j=0}^{a} T_{ij}$$

where

$$T_{ij} \quad = \quad \frac{\binom{a}{i}\binom{n-a}{a-i}\binom{a}{j}\binom{n-a}{a-j}}{\binom{n}{a}\binom{n}{a}}p^{-ij}.$$

We want to show that $\frac{\Delta}{E(Z)} = 1 + o(n^{-\frac{3}{2}})$. First, we will look at the first few terms of the sequence T_{ij}.

$$
\begin{aligned}
T_{00} &= \frac{\binom{n-a}{a}^2}{\binom{n}{a}^2} \\
&= [(1-\frac{a}{n})(1-\frac{a}{n-1})\dots(1-\frac{a}{n-(a-1)})]^2 \\
&= [1-\frac{a^2}{n}+o(n^{-\frac{3}{2}})]^2.
\end{aligned}
$$

$$
\begin{aligned}
T_{10} &= \frac{\binom{a}{1}\binom{n-a}{a-1}}{\binom{n}{a}}\frac{\binom{a}{0}\binom{n-a}{a}}{\binom{n}{a}} \\
&= \frac{a^2}{n-2a+1}T_{00}.
\end{aligned}
$$

The second equality in T_{10} follows, since $\binom{n-a}{a-1} = \frac{a}{n-2a+1}\binom{n-a}{a}$. Similarly,

$$
T_{01} = \frac{a^2}{n-2a+1}T_{00}.
$$

Adding up the first three terms, we obtain

$$
\begin{aligned}
T_{00}+T_{01}+T_{10} &= T_{00}(1+\frac{2a^2}{n-2a+1}) \\
&= [1-\frac{a^2}{n}+o(n^{-\frac{3}{2}})]^2(1+\frac{2a^2}{n-2a+1}) \\
&= 1+o(n^{-\frac{3}{2}}).
\end{aligned}
$$

Now, we want to show that the remaining part of the summation is also small. To be able to do that, first we will bound the terms T_{ij} $(i,j \geq 1)$ in terms of T_{11}.

$$
\frac{T_{ij}}{T_{11}} = \frac{\binom{a}{i}\binom{n-a}{a-i}}{\binom{a}{1}\binom{n-a}{a-1}}\frac{\binom{a}{j}\binom{n-a}{a-j}}{\binom{a}{1}\binom{n-a}{a-1}}p^{-ij+1}
$$

Since

$$
\frac{\binom{a}{i}\binom{n-a}{a-i}}{\binom{a}{1}\binom{n-a}{a-1}} = \frac{(n-2a+1)!}{(n-2a+i)!\ i!}\frac{[(a-1)!]^2}{[(a-i)!]^2}
$$

we obtain

$$
\frac{T_{ij}}{T_{11}} \leq (\frac{a^2}{n-2a})^{i-1}(\frac{a^2}{n-2a})^{j-1}p^{-ij+1}.
$$

First, note that $\frac{T_{12}}{T_{11}} = \frac{(a-1)^2}{2(n-2a+2)p} \leq 1$, for sufficiently large n. Similarly, $\frac{T_{21}}{T_{11}} \leq 1$, for sufficiently large n.

For $i \geq 2$,

$$-(i-1)j \quad \leq \quad \frac{-ij+1}{2}.$$

Similarly, for $j \geq 2$,

$$-(j-1)i \leq \frac{-ij+1}{2}.$$

Thus,

$$\begin{aligned}\frac{T_{ij}}{T_{11}} &\leq (\frac{a^2}{n-2a})^{i-1} p^{\frac{-ij+1}{2}} (\frac{a^2}{n-2a})^{j-1} p^{\frac{-ij+1}{2}} \\ &\leq (\frac{a^2}{n-2a} p^{-j})^{i-1} (\frac{a^2}{n-2a} p^{-i})^{j-1}.\end{aligned}$$

For the choice of $a = a^*(n) = (1-\epsilon)\frac{\log\ n}{\log\ \frac{1}{p}}$, we get $\frac{T_{ij}}{T_{11}} \leq 1$.

Noting that $T_{11} = \frac{\frac{a^4}{(n-2a+1)^2}\binom{n-a}{a}^2}{\binom{n}{a}^2}$ and $\sum_{i=1}^{a}\sum_{j=1}^{a} T_{ij} \leq \sum_{i=1}^{a}\sum_{j=1}^{a} T_{11} \rightarrow 0$ as $n \rightarrow \infty$ for $a = a^*(n)$, we get

$$\frac{\Delta}{E(Z)} = 1 + o(n^{-\frac{3}{2}}).$$

Hence,

$$\text{Prob } [Z = 0] \leq o(n^{-\frac{3}{2}}). \tag{2}$$

From (1) and (2), we get the claimed result. $\square$

Theorem 3.2 *Given a random bipartite graph $B = (V_1 \cup V_2, p)$, where $|V_1| = |V_2| = n$, there exists a biclique of size $k \times O(n^{\frac{1}{3}})$ with high probability, for some constant k.*

Proof Let Z=number of $a \times b$ bicliques in G. We need to show that Prob $[Z = 0]$ is very small, if a is a constant and b is $O(n^{\frac{1}{3}})$. Again, using the *second moment method*, we have

$$\text{Prob } [\ Z = 0] \quad \leq \quad \frac{Var(Z)}{(E(Z))^2} = \frac{\Delta}{E(Z)} - 1$$

where

$$\Delta \quad = \quad \sum_{i=0}^{a}\sum_{j=0}^{b} \binom{a}{i}\binom{n-a}{a-i}\binom{b}{j}\binom{n-b}{b-j} p^{ab-ij}.$$

Assuming $a = k$, a constant, and using the fact that $p^{-ij} \leq p^{-kj}$, for $j = 1, \ldots, k$, we have

$$
\begin{aligned}
\frac{\Delta}{E(Z)} &\leq \frac{\sum_{i=0}^{k}\sum_{j=0}^{b}\binom{k}{i}\binom{n-k}{k-i}\binom{b}{j}\binom{n-b}{b-j}p^{-kj}}{\binom{n}{k}\binom{n}{b}} \\
&= \frac{\sum_{i=0}^{k}\binom{k}{i}\binom{n-k}{k-i}}{\binom{n}{k}}\ \frac{\sum_{j=0}^{b}\binom{b}{j}\binom{n-b}{b-j}p^{-kj}}{\binom{n}{b}} \\
&= \frac{\sum_{j=0}^{b}\binom{b}{j}\binom{n-b}{b-j}p^{-kj}}{\binom{n}{b}}.
\end{aligned}
$$

So,

$$
\frac{\Delta}{E(Z)} = \sum_{j=0}^{b} T_j
$$

where

$$
T_j = \frac{\binom{b}{j}\binom{n-b}{b-j}}{\binom{n}{b}}p^{-kj}.
$$

First, note that

$$
\begin{aligned}
T_0 &= \frac{\binom{n-b}{b}}{\binom{n}{b}} \\
&= (1-\frac{b}{n})(1-\frac{b}{n-1})\ldots(1-\frac{b}{n-b+1}) \\
&= 1-\frac{b^2}{n}+o(n^{-\frac{3}{2}}).
\end{aligned}
$$

Now, we want to show that $\frac{T_j}{T_0} \leq 1$ for every $j \geq 1$.

$$
\begin{aligned}
\frac{T_j}{T_0} &= \frac{\binom{b}{j}\binom{n-b}{b-j}}{\binom{n-b}{b}}p^{-kj} \\
&= \frac{[b(b-1)\ldots(b-j+1)]^2}{(n-2b+j)\ldots(n-2b+1}\frac{p^{-kj}}{j!} \leq (\frac{b^2}{n-2b}p^{-k})^j.
\end{aligned}
$$

Choosing $b = n^{\frac{1}{3}-\epsilon}$, we obtain

$$
T_j \leq (\frac{b^2}{n-2b}p^{-k})^j T_0 \leq (\frac{b^2}{n-2b}p^{-k})\ T_0.
$$

Then,

$$
\sum_{j=0}^{b} T_j = T_0 + \sum_{j=1}^{b} T_j
$$

$$
\begin{aligned}
&\leq \; T_0 + b\;(\frac{b^2}{n-2b}p^{-k})\;T_0 \\
&= \; T_0\;(1 + \frac{b^3}{n-2b}\;p^{-k}) \\
&\leq \; 1 + o(n^{-\frac{3}{2}}) + \frac{b^3 p^{-k}}{n-2b} \quad \rightarrow \quad 1 + o(n^{-\frac{3}{2}}) \text{ as } \quad n \rightarrow \infty
\end{aligned}
$$

for the choice of $b = n^{\frac{1}{3}-\epsilon}$.

Hence, Prob $[Z = 0] \leq o(n^{-\frac{3}{2}})$ and we get the claimed result. □

4 Approximations for MBP and a Special Case

In this section, we consider two approximations: (1) using linear programming and (2) using the maximum node cardinality biclique. We end the section with a special case of MBP which is shown to be polynomially solvable.

The first approximation is as follows. Note that two edges (p, q) and (u, v) ($p, u \in V_1$ and $q, v \in V_2$) can be in the same biclique if and only if one of the following is true:

1. $p = u$ or $q = v$, i.e. the edges are adjacent,
2. both of the edges (p, v) and (u, q) exist.

Define the variables $x_i,\ i \in E$, where $x_i = 1$, if edge i is *not* in the maximum edge cardinality biclique and $x_i = 0$ otherwise. Let w_i be the weight of edge i and consider the following 0-1 integer program (IP):

$$
\begin{array}{lll}
W^* = \text{Min} & \sum_{i \in E} w_i x_i & \\
\text{Subject to} & x_i + x_j & \geq 1 \text{ if edges } i \text{ and } j \text{ cannot be in the same biclique} \\
 & x_j & \in \{0, 1\}\ j \ = \ 1, ..., |E|
\end{array}
$$

Let $\bar{W}$ be the value of the solution obtained by solving the LP relaxation of IP and rounding up all the fractional values $x_i \geq 0.5$ to 1, and let $\bar{z} = |E| - \bar{W}$. Let z^* = number of edges in the maximum biclique. Then, $z^* = |E| - W^*$, $\bar{W} \leq 2\ W^*$ and

$$
\begin{aligned}
\bar{z} \;&= \; |E| - \bar{W} \\
&\geq \; |E| - 2\ W^* \\
&= \; 2\ z^* - |E|.
\end{aligned}
$$

Now, suppose $\bar{z} = \alpha \, |E|$ for some $0 \leq \alpha \leq 1$. Then,

$$z^* \leq (\frac{1+\alpha}{2\alpha})\bar{z}.$$

Note that $\bar{W} \leq |E|$. If $\bar{W} \leq \frac{2|E|}{3}$, we have $\bar{z} \geq \frac{|E|}{3}$. But this implies that $z^* \leq 2\bar{z}$ since $\frac{1+\alpha}{2\alpha} \leq 2 \; \forall \; \alpha \geq \frac{1}{3}$. Thus, in this case we have a 2-approximation for MWBP. Similarly, if $\bar{W} \leq \frac{4|E|}{5}$, we have a 3-approximation for MWBP.

The second approximation is as follows. We can use the maximum node cardinality biclique as a solution to MBP and obtain a 2-approximation if the graph is sufficiently dense. This follows from the following two results.

Theorem 4.1 *For a bipartite graph $G = (V_1 \cup V_2, E)$, let $a \times b$ be the maximum node cardinality biclique and $c \times d$ be the maximum edge cardinality biclique. If* $\max(\frac{a}{b}, \frac{b}{a}) \leq m$ *for some $m \in \Re_+$, we have that $cd \leq (\frac{m+3}{4})ab$. As a special case, if $m \leq 5$, we have a 2-approximation for the the edge cardinality biclique.*

Proof Clearly, $ab \leq cd$. Suppose $kab \leq cd$ for some positive constant k. By the choice of m, we have

$$\frac{a}{4b} + \frac{1}{2} + \frac{b}{4a} \leq \frac{3}{4} + \frac{m}{4} \;\Rightarrow\; \frac{a+b}{2\sqrt{ab}} \leq \sqrt{\frac{m+3}{4}} \;\Rightarrow\; 2\sqrt{ab}\sqrt{\frac{m+3}{4}} \geq a+b.$$

If $k > \frac{m+3}{4}$, we have,

$$c + d \geq 2\sqrt{cd} \geq 2\sqrt{ab}\sqrt{k} > 2\sqrt{ab}\sqrt{\frac{m+3}{4}} > a+b$$

which contradicts the fact that $a \times b$ is the maximum node cardinality biclique. Thus, $k \leq \frac{m+3}{4}$ and $cd \leq (\frac{m+3}{4})ab$. □

From a practical point of view, the case $m \leq 5$ is important. In the following proposition, we show that $m \leq 5$ for large dense random bipartite graphs.

Theorem 4.2 *Let $G(n, n, p)$ be a random bipartite graph with $|V_1| = |V_2| = n$ and p as the probability of having an edge. If $p = e^{\frac{-25}{9n^{2+\epsilon}}}$ (with some $\epsilon > 0$), then for the maximum node cardinality biclique $a \times b$, $max(\frac{a}{b}, \frac{b}{a}) \leq 5$ with high probability as $n \to \infty$.*

Proof It is sufficient to show that for the claimed value of p, $G(n, n, p)$ has a $\frac{3n}{5} \times \frac{3n}{5}$ bi-clique with high probability. It then follows that the maximum node cardinality biclique $a \times b$ will satisfy $a + b \geq \frac{6n}{5}$ and hence $max(\frac{a}{b}, \frac{b}{a}) \leq 5$.

Let Z=number of $a \times b$ bicliques in G, where $a = b = \frac{3n}{5}$. From the *second moment method*, Prob $[Z = 0] \leq \frac{Var(Z)}{(E(Z))^2}$.

$$\begin{aligned}
E(Z) &= \binom{n}{a}\binom{n}{b}p^{ab} \\
E(Z^2) &= \sum_{A,B}\sum_{A',B'} \text{Prob }[Z_{A,B} = 1, Z_{A',B'} = 1] \\
&= \sum_{A,B}\sum_{A',B'} \text{Prob }[Z_{A',B'} = 1/Z_{A,B} = 1] \text{ Prob }[Z_{A,B} = 1]
\end{aligned}$$

Since all the (A, B) look alike, fix (A, B) as $(\bar{A}, \bar{B})$

$$\begin{aligned}
E(Z^2) &= \sum_{A,B}\sum_{A',B'} \text{Prob }[Z_{A',B'} = 1/Z_{\bar{A},\bar{B}} = 1] \text{ Prob }[Z_{A,B} = 1] \\
&= \sum_{A,B} \text{Prob }[Z_{A,B} = 1] \sum_{A',B'} \text{Prob }[Z_{A',B'} = 1/Z_{\bar{A},\bar{B}} = 1]
\end{aligned}$$

Letting $\sum_{A',B'}$ Prob $[Z_{A',B'} = 1/Z_{\bar{A},\bar{B}} = 1] = \Delta$, we get $E(Z^2) = E(Z)\ \Delta$.

$$\begin{aligned}
\Delta &= \sum_{A',B'} \text{Prob }[Z_{A',B'} = 1/Z_{\bar{A},\bar{B}} = 1] \\
&= \sum_{i=2a-n}^{a}\sum_{j=2b-n}^{b}\sum_{\substack{|A'\cap\bar{A}|=i \\ |B'\cap\bar{B}|=j}} \text{Prob }[Z_{A',B'} = 1/Z_{\bar{A},\bar{B}} = 1] \\
&= \sum_{i=2a-n}^{a}\sum_{j=2b-n}^{b}\binom{a}{i}\binom{n-a}{a-i}\binom{b}{j}\binom{n-b}{b-j}p^{ab-ij}.
\end{aligned}$$

$$\begin{aligned}
\text{Prob }[\ Z = 0] &\leq \frac{Var(Z)}{(E(Z))^2} = \frac{\Delta}{E(Z)} - 1 \\
\frac{\Delta}{E(Z)} &= \sum_{i=0}^{n-a}\sum_{j=0}^{n-b}\frac{\binom{a}{2a-n+i}\binom{n-a}{n-a+i}\binom{b}{2n-b+j}\binom{n-b}{n-b-j}p^{-(2a-n+i)(2b-n+j)}}{\binom{n}{a}\binom{n}{b}p^{ab}}
\end{aligned}$$

Using $a = b = \frac{3n}{5}$, $q = \frac{n}{5}$, we get

$$\begin{aligned}
\frac{\Delta}{E(Z)} &= \sum_{i=0}^{2q}\sum_{j=0}^{2q}\frac{\binom{3q}{q+i}\binom{2q}{2q-i}\binom{3q}{q+j}\binom{2q}{2q-j}p^{-(q+i)(q+j)}}{\binom{5q}{3q}^2} \\
&\leq \sum_{i=0}^{2q}\sum_{j=0}^{2q}\frac{\binom{3q}{q+i}\binom{2q}{2q-i}\binom{3q}{q+j}\binom{2q}{2q-j}p^{-9q^2}}{\binom{5q}{3q}^2}.
\end{aligned}$$

Let $p = e^{\frac{-25}{9n^{2+\epsilon}}}$ (other choices for p are possible). Then, as $n \to \infty$, we have $p^{-9q^2} \leq 1 + \epsilon_2$ for any $\epsilon_2 > 0$. Thus, as $n \to \infty$, we have

$$\frac{\Delta}{E(Z)} \leq (1 + \epsilon_2)\sum_{i=0}^{2q}\sum_{j=0}^{2q}\frac{\binom{3q}{q+i}\binom{2q}{2q-i}\binom{3q}{q+j}\binom{2q}{2q-j}}{\binom{5q}{3q}^2}$$

$$= \sum_{i=0}^{2q} \frac{\binom{3q}{q+i}\binom{2q}{2q-i}}{\binom{5q}{3q}} \sum_{j=0}^{2q} \frac{\binom{3q}{q+j}\binom{2q}{2q-j}}{\binom{5q}{3q}}.$$

Note that

$$\sum_{i=0}^{2q} \frac{\binom{3q}{q+i}\binom{2q}{2q-i}}{\binom{5q}{3q}} = \sum_{i=0}^{2q} \frac{\binom{3q}{2q-i}\binom{2q}{i}}{\binom{5q}{2q}} = 1$$

since

$$\sum_{x=0,1,\dots,n} h(x;n,a,N) = \sum_{x=0,1,\dots,n} \frac{\binom{a}{x}\binom{N-a}{n-x}}{\binom{N}{n}} = 1$$

where $h(.)$ is the probability density function of the *hypergeometric distribution.* So, $\frac{\Delta}{E(Z)} \leq 1+\epsilon_2$ for any $\epsilon_2 > 0$ as $n \to \infty$. Therefore, with high probability $Z > 0$ and there exists a $\frac{3n}{5} \times \frac{3n}{5}$ biclique in G. $\square$

A special case of MBP is when a maximum edge cardinality biclique has a node which has all its neighbors in the maximum biclique. We will show that this case is polynomially solvable. Let $d_G(i)$ be the *degree* of node i in G and $N(i) = \{j : (i,j) \in E(G)\}$ be the *neighborhood* of node i. For a set S, $N(S) = \cap_{i\in S} N(i)$ is the common neighborhood of nodes in S.

Theorem 4.3 *Suppose a maximum cardinality biclique, $A \times B$, has at least one node $i^* \in A$ ($i^* \in B$) such that $d_G(i^*) = |B|$ ($d_G(i^*) = |A|$). Then, MBP is polynomially solvable.*

Proof For any node $i \in V_1 \times V_2$, consider the induced subgraph $S_i = N(i) \times N(N(i))$. WLOG, suppose $i \in V_1$. Then, $N(i) \subseteq V_2$ and $N(N(i)) \subseteq V_1$. Moreover, every node in $N(i)$ is adjacent to all the nodes in $N(N(i))$. This shows that S_i is a biclique. We claim that the biclique S_{i^*} such that $|S_{i^*}| = \max_{i\in V_1\cup V_2}\{k(S_i) : k(S_i) = |N(i)| \times |N(N(i))|\}$ is a maximum edge cardinality biclique. For suppose, $A \times B$ is a maximum edge cardinality biclique which contains a node j^* satisfying the hypothesis. Then, $S_{j^*} = |N(i)| \times |N(N(i))|$ and the result follows. $\square$

References

[1] Alon, N., Spencer J.H., Erdos P. (1992) *The Probabilistic Method,* Wiley, New York.

[2] Bollobas, B. *Random graphs (1985)* , Academic press, London.

[3] Bondy, J.A. and Murty, U.S.R. (1976) *Graph Theory with Applications,* Macmillan, London.

[4] Ganter, B. *Personal Communication.*

[5] Ganter, B. and Wille, R. (1996), *Formale Begriffsanalyse – Mathematische Grundlagen.* Springer Verlag.

[6] Garey, M.S. and Johnson, D.S. (1979), *Computers and Intractactibility: A Guide to NP-Completeness.* W. H. Freeman, New York.

[7] Hardy, G.H. and Wright, E.M. (1954), *An Introduction to the Theory of Numbers.* Oxford, Clarendon Press.

[8] Swaminathan, J.S. and Tayur, S. (1996) *Management of Broader Product Lines through Delayed Product Differentiation using Vanilla Boxes.* Under review in Management Science.