Journal of Algorithms **41**, 388–403 (2001)
doi:10.1006/jagm.2001.1199, available online at http://www.idealibrary.com on IDEAL®

On Bipartite and Multipartite Clique Problems

Milind Dawande

University of Texas, Dallas, Texas 75080; and T. J. Watson Center, IBM, Yorktown Heights, New York 10598

Pinar Keskinocak

School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, Georgia 30332

Jayashankar M. Swaminathan

The Kenan-Flagler Business School, University of North Carolina, Chapel Hill, North Carolina 27599

and

Sridhar Tayur

GSIA, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213

Received January 24, 2000

In this paper, we introduce the maximum *edge* biclique problem in bipartite graphs and the edge/node weighted multipartite clique problem in multipartite graphs. Our motivation for studying these problems came from abstractions of real manufacturing problems in the computer industry and from formal concept analysis. We show that the weighted version and four variants of the unweighted version of the biclique problem are NP-complete. For random bipartite graphs, we show that the size of the maximum balanced biclique is considerably smaller than the size of the maximum edge cardinality biclique, thus highlighting the difference between the two problems. For multipartite graphs, we consider three versions each for the edge and node weighted problems which differ in the structure of the multipartite clique (MPC) required. We show that all the edge weighted versions are NP-complete in general. We also provide a special case in which edge weighted versions are polynomially solvable. © 2001 Elsevier Science

Key Words: bipartite graph; multipartite graph; clique; complexity.

0196-6774/01 $35.00

1. INTRODUCTION

In this paper, we study biclique and multipartite clique problems. 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 among all the bicliques in B.

A multipartite graph with n levels $G = (V_1 \cup V_2 \cup \cdots \cup V_n, E)$ is defined as a graph such that for every edge $e = (u, v)$, we have $u \in V_i$ and $v \in V_{i+1}$ for some $i \in \{1, \ldots, n-1\}$.[1] A multipartite clique $M = U_1 \cup U_2 \cup \cdots \cup U_m$ within a multipartite graph G is defined such that $U_i \subseteq V_{k+i} \,\forall\, i\ 1 \leq i \leq m$, $m \geq 2$ for some $k \geq 0$ and for every $u_1 \in U_i$ and $u_2 \in U_{i+1}$ the edge $(u_1, u_2) \in E$. In an edge weighted multipartite graph, the *maximum edge weighted multipartite clique* is one which has the maximum sum in terms of the weights of the edges in the multipartite clique. Similarly, a *maximum node weighted multipartite clique* has the maximum sum in terms of weights of nodes in the multipartite clique.

A well known problem related to biclique and multipartite clique problems is the *maximum clique*, which is one of the most 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 $(u, v) \in E$. 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.

Besides their relation to the maximum clique problem, our motivation for studying the biclique and multipartite clique problems came from a real manufacturing problem in the computer industry. 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

[1] Note the difference between multipartite graphs and the well known class of k-partite graphs. A graph $G = (V, E)$ is k-partite if V can be partitioned into k subsets $V_1, \ldots, V_k$ such that for every edge $(u, v) \in E$, u and v belong to different vertex sets of the partition. The class of multipartite graphs is contained in the class of k-partite graphs.

perceived by the customers for these products is to reduce the final assembly time, where such a reduction can be obtained by creating subassemblies (or *vanilla box*es) in advance (see [11] 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_{ij} be the assembly time of component i in product j. If the total assembly time of the components in 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 the problem reduces to the maximum edge cardinality biclique problem (MBP). A natural generalization of the bipartite clique problem is the multipartite clique problem. In such a case, each multipartite clique in the graph represents a possible storage of vanilla boxes at different levels in the assembly process such that a vanilla box in a later level in assembly is itself assembled in part from another vanilla box from the previous level (because a biclique between any two levels i and $i+1$ acts as vanilla box for that level).

Bicliques have also been studied in the area of *formal concept analysis* [3, 4]. 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 P, and

Q' = the set of all attributes which all the objects of Q have.

TABLE 1
Variants of Biclique Problems

Abbreviation	Problem
MBP	Maximum edge cardinality biclique
MWBP	Maximum edge weight biclique
MNWBP	Maximum node weight biclique
MBBP	Maximum balanced node cardinality biclique
EBNCD	Exact balanced node cardinality decision problem
EECD	Exact edge cardinality decision problem
MOFCP	Maximum One-sided edge cardinality problem
EBPNCD	Exact balanced prime node cardinality decision problem

TABLE 2
Variants of Multipartite Clique Problems

Abbreviation	Problem
MPC	Multipartite clique
MPCP	Multipartite clique which includes nodes from *all* levels
MPCF	Multipartite clique including the first level
MPCS	Multipartite clique problem which includes nodes from *some* levels

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 maximal bicliques of B. In formal concept analysis, the goal is to cover the bipartite graph by "fat" concepts, i.e., large bicliques. Current methods in the area do a brute force search for finding large (i.e., one with the maximum number of edges) bicliques to cover all the edges [3, 8].

The rest of the paper is organized as follows. In Section 2, we present complexity results related to the biclique problem and compare the size of balanced biclique and edge cardinality bicliques in random bipartite graphs. In Section 3, we present the alternative versions of the multipartite clique problem and develop complexity results. We conclude in Section 4. Variants of biclique and multipartite clique problems mentioned in the paper are summarized in Tables 1 and 2.

2. THE BICLIQUE PROBLEM

In this section, we first present the formulation for the biclique problem and discuss known results. Then we show that MWBP and four variants of MBP are NP-complete. Note that since the complement of a bipartite graph is not bipartite in general, the polynomial-time solvability of the independent set problem on bipartite graphs does not imply a polynomial time algorithm for MBP. Finally, we compare the sizes of maximum balanced bicliques and maximum edge cardinality bicliques in random graphs.

In a node weighted bipartite graph $B = (V_1 \cup V_2, E)$, there is a weight w_v associated with each node v. The maximum node weight biclique problem (MNWBP) can be formulated as a 0–1 integer program as

$$\begin{aligned}
\max \quad & \sum_{u \in V_1} w_u x_u + \sum_{v \in V_2} w_v x_v \\
\text{subject to} \quad & x_u + x_v \leq 1 \quad u \in V_1, \quad v \in V_2, \ (u, v) \notin E \qquad (1) \\
& x_v \in \{0, 1\} \qquad \text{for all } v \in V_1 \bigcup V_2, \qquad (2)
\end{aligned}$$

where

$$x_v = \begin{cases} 1, & \text{if node } v \text{ is in the biclique} \\ 0, & \text{otherwise.} \end{cases}$$

If we relax this integer program by replacing the integrality constraints (2) with $0 \leq x_v \leq 1$ for all $v \in V_1 \cup V_2$, we obtain a linear program. Note that the matrix defining the constraint set (1) is the node-edge incidence matrix of a bipartite graph, which is totally unimodular, and hence the solution to the linear programming relaxation will be integer [9, p. 544, Corollary 2.9]. Therefore, the maximum node weight biclique problem is polynomially solvable [5]. It follows that 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), which is NP-complete [5]. (This problem is referred to as the balanced complete bipartite subgraph problem in [5, p. 196].) Note that for the same bipartite graph, solutions to MBBP and MBP may be quite different from each other. Hence, node-cardinality biclique problems do not provide good approximations for MBP in general. In Section 2.2, we quantify the difference between the solutions to MBP and MBBP in random graphs.

Hochbaum [7] considers a related problem to MWBP and MNWBP, where the objective is to minimize the total weight of the nodes or edges deleted so that the remaining subgraph is a biclique. She provides a 2-approximation for the edge deletion version for general and bipartite graphs and a 2-approximation for the node deletion version for general graphs.

2.1. *Complexity of Biclique Problems*

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 let 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. ∎

Note that the reduction in Theorem 1 does not imply the NP-hardness of MBP, since we used a weighted bipartite graph in the reduction in which some edge weights were zero. An NP-completeness proof for MBP has been recently provided in [10].

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 [5]. 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. Note that the reductions for EBNCD and ENCD are Turing reductions rather than Karp reductions [5]. ∎

THEOREM 2.2. *EECD is NP-complete.*

To prove this theorem, first we 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 let p be any prime number such that $l \leq p \leq 2l$. Such a prime number is guaranteed by *Bertrand's theorem* [6]. Let $a < p$ be a positive integer, where a is the specification for EBNCD. 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 $p - a + \max\{|V_1|, |V_2|\} \leq 3l$. Since $p^2 \geq l^2$ it follows that $p^2 > 3l$ and the maximum degree is less than p^2, for $l > 3$. Then, EBNCD 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 define the following decision problem:

Maximum fixed intersection problem (MFIP): Given $k \in Z_+$, a ground set V, and a set system $\mathcal{S} = \{S_1, \ldots, S_n\}$, 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. *MFIP is NP-hard.*

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 [5]. 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 $S_e = V \backslash \{u, v\}$. Let $\mathcal{S} = \{S_e : e \in E\}$. There exists a clique on k nodes in G if and only if there exist $p = \frac{k(k-1)}{2}$ subsets in $\mathcal{S}$ whose intersection has cardinality at least $|V| - k$. Thus, there exists a clique of size k in G if and only if the cardinality of the maximum intersection in the optimal solution to MFIP of p sets is $|V| - k$. ∎

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. ∎

2.2. *Comparing Maximum Balanced Bicliques and Maximum Edge Cardinality Bicliques*

In this section, we show that the size of a maximum balanced biclique may be considerably smaller than the size of a maximum edge cardinality biclique in random bipartite graphs, thus highlighting the difference between these seemingly similar problems.

We denote a random bipartite graph by $B = (V_1 \cup V_2, p)$, where $0 \leq p \leq 1$ is the probability that a particular edge exists in B. We denote the size of a biclique by $a \times b$, if it has a nodes in V_1 and b nodes in V_2. For $|V_1| = |V_2| = n$ and for sufficiently large n, we show that the maximum balanced biclique will be of size $a \times a$ with high probability, where $a(n) \leq a < 2a(n)$ and $a(n) = \log n / \log \frac{1}{p}$. Note that the size of a maximum edge cardinality biclique in a random bipartite graph will be at least np (consider a single node and all its neighbors) with high probability, which is much larger than $a \times a$ for constant p.

THEOREM 2.6. *Consider a random bipartite graph $B = (V_1 \cup V_2, p)$, where $0 < p < 1$ is a constant, $|V_1| = |V_2| = n$, and $a(n) = \log n / \log \frac{1}{p}$. If the maximum balanced biclique in this graph has size $a \times a$, then $a(n) \leq a \leq 2a(n)$ with high probability (for sufficiently large n).*

Proof. The proof consists of two main steps. First, we show that the probability of having a balanced biclique of size $2a(n) \times 2a(n)$ is very small (i.e., the probability approaches zero as n approaches ∞). Second, we show that the probability of having a balanced biclique of size at least $a(n) \times a(n)$ approaches 1 as n approaches ∞.

Let Z_a = 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 $a \geq 2a(n)$. We use the fact that $\text{Prob}[Z_a \geq 1] \leq E[Z_a]$.

$$\text{Prob}\,[Z_a \geq 1] \leq E[Z_a] = \binom{n}{a}^2 p^{a^2}$$

$$\leq \left(\frac{n^a}{a!}\right)^2 p^{a^2}.$$

The computation of $E(Z_a)$ follows from the following argument. A subset of nodes $A \cup Q$, $A \subseteq V_1$, $Q \subseteq V_2$ form a biclique, if there is an edge between every pair of nodes $u \in A$, $v \in Q$. Suppose both A and Q have size a. Since the probability of an edge is p, the probability that a given node set $A \cup Q$ forms a biclique is p^{a^2}. There are $\binom{n}{a}$ different ways of choosing a node subset $A \subseteq V_1$ or $Q \subseteq V_2$ of size a. Hence, the number of $a \times a$ subgraphs is $\binom{n}{a}\binom{n}{a}$ and the expected number of $a \times a$ bicliques is

$$E(Z_a) = \binom{n}{a}\binom{n}{a} p^{a^2}.$$

Note that for $a \geq 2\log n/\log\frac{1}{p}$, $p^{a^2} = (p^a)^a \leq (p^{\log_p n^{-2}})^a = n^{-2a}$ and hence $\mathrm{Prob}[Z_a \geq 1] \leq (\frac{1}{a!})^2$. Thus, for $a \geq 2\log\ n/\log\frac{1}{p}$,

$$\mathrm{Prob}[Z_a \geq 1] \to 0 \qquad \text{as } n \to \infty. \tag{3}$$

Now, we need to show that there is a balanced biclique of size $a(n) \times a(n)$ in B with high probability; i.e., $\mathrm{Prob}[Z_a = 0 \mid a = a(n)]$ is very small. From the *second moment method* [2], we have $\mathrm{Prob}[Z_a = 0] \leq \mathrm{Var}(Z_a)/(E(Z_a))^2$.

Let $X_{A,Q}$ be an indicator variable which assumes value 1, if the nodes in $A \subseteq V_1$ and $Q \subseteq V_2$ form a biclique, and zero otherwise:

$$\begin{aligned}
E(Z_a^2) &= \sum_{A,Q}\sum_{A',Q'} \mathrm{Prob}[X_{A,Q} = 1, X_{A',Q'} = 1] \\
&= \sum_{A,Q}\sum_{A',Q'} \mathrm{Prob}[X_{A',Q'} = 1 \mid X_{A,Q} = 1]\ \mathrm{Prob}[X_{A,Q} = 1].
\end{aligned}$$

Since all the (A, Q) look alike, fix (A, Q) as $(\widetilde{A}, \widetilde{Q})$:

$$\begin{aligned}
E(Z_a^2) &= \sum_{A,Q}\sum_{A',Q'} \mathrm{Prob}[X_{A',Q'} = 1 \mid X_{\widetilde{A},\widetilde{Q}} = 1]\ \mathrm{Prob}[X_{A,Q} = 1] \\
&= \sum_{A,Q} \mathrm{Prob}[X_{A,Q} = 1] \sum_{A',Q'} \mathrm{Prob}[X_{A',Q'} = 1 \mid X_{\widetilde{A},\widetilde{Q}} = 1] \\
&= \sum_{A,Q} \mathrm{Prob}[X_{A,Q} = 1] \sum_{i=0}^{a}\sum_{j=0}^{a} \sum_{\substack{|A'\cap\widetilde{A}|=i \\ |Q'\cap\widetilde{Q}|=j}} \mathrm{Prob}[X_{A',Q'} = 1 \mid X_{\widetilde{A},\widetilde{Q}} = 1] \\
&= \sum_{A,Q} \mathrm{Prob}[X_{A,Q} = 1] \sum_{i=0}^{a}\sum_{j=0}^{a} \binom{a}{i}\binom{n-a}{a-i}\binom{a}{j}\binom{n-a}{a-j} p^{a^2-ij}.
\end{aligned}$$

Letting $\sum_{A',Q'} \mathrm{Prob}[X_{A',Q'} = 1 \mid X_{\widetilde{A},\widetilde{Q}} = 1] = \Delta$, we get

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

We can write

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

where

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

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

$$\begin{aligned}
T_{00} &= \frac{\binom{n-a}{a}^2}{\binom{n}{a}^2} \\
&= \left[\left(1-\frac{a}{n}\right)\left(1-\frac{a}{n-1}\right)\cdots\left(1-\frac{a}{n-(a-1)}\right)\right]^2 \\
&= \left[1-\frac{a^2}{n}+o(n^{-3/2})\right]^2. \\
T_{10} &= \frac{\binom{a}{1}\binom{n-a}{a-1}}{\binom{n}{a}}\frac{\binom{a}{0}\binom{n-a}{a}}{\binom{n}{a}} \\
&= \frac{a^2}{n-2a+1}T_{00}.
\end{aligned}$$

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

$$\begin{aligned}
T_{00}+T_{01}+T_{10} &= T_{00}\left(1+\frac{2a^2}{n-2a+1}\right) \\
&= \left[1-\frac{a^2}{n}+o(n^{-3/2})\right]^2\left(1+\frac{2a^2}{n-2a+1}\right) \\
&= 1+o(n^{-3/2}) \qquad \text{for } a = \frac{\log n}{\log\frac{1}{p}}.
\end{aligned}$$

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

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

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

For $i \geq 2$,

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

Similarly, for $j \geq 2$,

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

Thus,

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

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

Noting that

$$T_{11} = \frac{\frac{a^4}{(n-2a+1)^2}\binom{n-a}{a}^2}{\binom{n}{a}^2 p}$$

and $\sum_{i=1}^{a}\sum_{j=1}^{a} T_{ij} \leq \sum_{i=1}^{a}\sum_{j=1}^{a} T_{11} \to 0$ as $n \to \infty$ for $a = a^*(n)$, we get

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

Hence,

$$\text{Prob}[Z_a = 0] \leq o(n^{-3/2}). \tag{4}$$

From (3) and (4), we get the claimed result. ∎

Our use of the probabilistic method in Theorem 2.6 was inspired by the work presented in [2]. The fundamentals of the method and similar results, on random graphs, for combinatorial quantities such as the clique number and the chromatic number are presented in [2].

3. MULTIPARTITE CLIQUE PROBLEM

In this section we introduce the following three versions of the multipartite clique problem: (1) maximum edge-weighted multipartite clique which includes nodes from *all* levels (MPCP), (2) maximum edge-weighted multipartite clique which starts from the first level (product level) and includes nodes from a contiguous subset of remaining levels (MPCF), and (3) maximum edge-weighted multipartite clique problem which includes nodes from a subset of levels (MPCS).

Figure 1(a) shows a multipartite clique, namely, $\{\{1, 2\}, \{5, 6, 7\}, \{10, 11\}, \{13, 15\}\}$, which includes nodes from all the levels of the graph. The graph given in Fig. 1(b) (where, for simplicity, we assumed all the edge weights to be equal to 1) illustrates the difference between MPCP, MPCF, and MPCS. Here, the optimum solution to MPCP is $\{\{1\}, \{3, 5\}, \{9, 10\}, \{13\}\}$, whereas the optimum solution to MPCF is $\{\{1\}, \{3, 5\}, \{7, 8, 9, 10\}\}$ and the optimum solution to MPCS is $\{\{3, 4, 5\}, \{7, 8, 9, 10\}\}$.

Note that, in general, problems MPCP, MPCP, and MPCS may not be solved by solving a sequence of biliclique problems on successive bipartite subgraphs of the multipartite graph.

3.1. *Formulations*

Let $G^{i,i+1} = G(V_i, V_{i+1}, E^{(i,i+1)})$ be the bipartite graph induced by node sets V_i and V_{i+1}. Define the variable $x_e^{(i,i+1)}$ to be 1 if edge e of $E^{(i,i+1)}$ is *not* in the multipartite clique; 0, otherwise. For an edge e in $E^{(i,i+1)}$, let $A(e)$ be the edges in $E^{(i+1,i+2)}$ which are adjacent to e and let $B(e)$ be the edges in $E^{(i-1,i)}$ which are adjacent to e. $w_e^{(i,i+1)}$ is the weight of edge

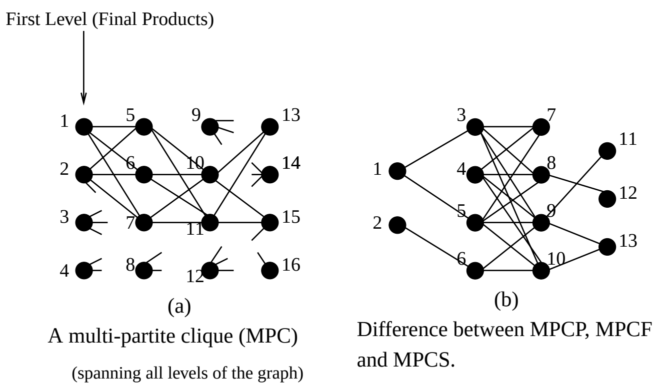

FIG. 1. MPCP and its variants: MPCF and MPCS.

e in $E^{(i,\,i+1)}$. We assume that there are n levels in the graph and they are numbered $1, \ldots, n$ where 1 represents the first level.

MPCP: *Multipartite Clique which Includes Nodes from All Levels*

$$W^* = \min \sum_{i=1}^{n-1} \sum_{k \in E^{(i,\,i+1)}} w_k^{(i,\,i+1)} x_k^{(i,\,i+1)}$$

subject to

$$x_k^{(i,\,i+1)} + x_l^{(i,\,i+1)} \geq 1 \qquad \text{if edges } k \text{ and } l \text{ in } E^{(i,\,i+1)} \text{ cannot be in the same biclique}, \qquad \forall \text{ pair } (k, l) \in E^{(i,i+1)}$$

$$\sum_{e \in A(p)} x_e^{(i,\,i+1)} \leq |A(p)| + x_p^{(i-1,\,i)} - 1 \qquad \forall\, p \in E^{(i-1,\,i)}, \quad 2 \leq i \leq n-1$$

$$\sum_{e \in B(p)} x_e^{(i-1,\,i)} \leq |B(p)| + x_p^{(i,\,i+1)} - 1 \qquad \forall\, p \in E^{(i,\,i+1)}, \quad 2 \leq i \leq n-1$$

$$\sum_{e \in E^{(1,2)}} x_e^{1,\,2} \leq |E^{(1,\,2)}| - 1$$

$$x_e^{(i,\,i+1)} \in \{0, 1\} \qquad \forall\, e \in E^{(i,\,i+1)}, \quad 1 \leq i \leq n-1.$$

For each bipartite subgraph $G^{i,\,i+1}$, due to the first set of constraints, the variables $x_e^{(i,\,i+1)}$ having value 0 form a biclique. The second set of constraints "links" these bicliques together. That is, if variable $x_e^{(i-1,\,i)}$ is 0 (i.e., edge e is in the biclique of $G^{i-1,\,i}$), then at least one edge adjacent to e in $G^{i,\,i+1}$ should be in the MPC. Note that the second set of constraints is required only for levels 2 through $n-1$. Similarly, the third constraint makes sure that if a variable $x_e^{(i,\,i+1)}$ is 0 then at least one edge adjacent to e in $G^{i-1,\,i}$ should be in the MPC. The fourth constraint makes sure that at least one edge from the first level is included in the MPC.

MPCF: *Multipartite Clique Including the First Level*

$$W^* = \min \sum_{i=1}^{n-1} \sum_{k \in E^{(i,\,i+1)}} w_k^{(i,\,i+1)} x_k^{(i,\,i+1)}$$

subject to

$$x_k^{(i,\,i+1)} + x_l^{(i,\,i+1)} \geq 1 \qquad \text{if edges } k \text{ and } l \text{ in } E^{(i,\,i+1)} \text{ cannot be in the same biclique}, \qquad \forall \text{ pair } (k, l) \in E^{(i,\,i+1)}$$

$$\sum_{e\in A(p)} x_e^{(i,\,i+1)} \le |A(p)| + x_p^{(i-1,\,i)} - z_i \qquad \forall\, p \in E^{(i-1,\,i)}, \quad 2 \le i \le n-1$$

$$\sum_{e\in B(p)} x_e^{(i-1,\,i)} \le |B(p)| + x_p^{(i,\,i+1)} - z_i \qquad \forall\, p \in E^{(i,\,i+1)}, \quad 2 \le i \le n-1$$

$$z_{i+1} \le z_i \quad 2 \le i \le n-2$$

$$\sum_{e\in E^{(i,\,i+1)}} x_e^{(i,\,i+1)} \ge |E^{(i,\,i+1)}|(1-z_i) \quad 2 \le i \le n-1$$

$$\sum_{e\in E^{(1,\,2)}} x_e^{1,\,2} \le |E^{(1,\,2)}| - 1 \qquad \text{(I)}$$

$$x_e^{(i,\,i+1)} \in \{0,1\} \qquad \forall\, e \in E^{(i,\,i+1)}, \quad 2 \le i \le n-1$$

$$z_i \in \{0,1\} \qquad 1 \le i \le n-1.$$

In this case, $z_i = 0$ indicates that no edges from levels i and *above* can be in the MPC. Notice that if $z_i = 0$, then none of the edges in $E^{(i,\,i+1)}$ can be in the MPC due to the fifth set of constraints and $z_j = 0 \quad \forall\, j = i+1, \ldots, n-1$ due to the fourth set of constraints. When $z_i = 1$, the second set of constraints "links" level i with level $i+1$ in the MPC (similar to MPCP) and when $z_i = 0$, they are redundant.

MPCS: *Multipartite Clique Problem which Includes Nodes from Some Levels*

Problem MPCS can be formulated in a way similar to that of MPCF by removing constraint (I) and using variables δ_i in addition to variables z_i. In the formulation for MPCF, $z_i = 0$ indicates that no edges from levels i and *above* can be in the MPC. Problem MPCS will have an additional set of similar constraints involving variables δ_i where $\delta_i = 0$ indicates that no edges from levels i and *below* can be in the MPC. We avoid giving the entire formulation for MPCS since the basic idea of the formulation is the same as that of MPCF.

Since the biclique problem is a special case of MPCP, the complexity of several optimization and decision problems regarding the MPCP follows directly from the results for the biclique problem proved in Section 2. We list these results in Lemma 3.1.

LEMMA 3.1. *Given a multipartite graph $M = (V_1 \cup V_2 \cup \cdots \cup V_n, E)$, the following optimization/decision problems regarding MPCP, MPCF, and MPCS are NP-complete.*

- ***Maximum edge weight multipartite clique:*** *Find a multipartite clique C, where the sum of the edge weights in the subgraph induced by C is maximum.*

• ***Exact balanced node cardinality decision problem:*** *Given M and a positive integer $a \in Z_+$, does there exist a multipartite clique $C = (U_1 \cup U_2 \cup \cdots \cup U_n, E)$ with $|U_1| = |U_2| = \cdots = |U_n| = a$?*

• ***Exact edge cardinality decision problem:*** *Given M and a positive integer $k \in Z_+$, does there exist a multipartite clique with exactly k edges?*

• ***Maximum one-sided edge cardinality problem:*** *Given M and a positive integer $k \in Z_+$, find a maximum cardinality multipartite clique with exactly k nodes on* any *level.*

However, an interesting special case of MPCP can be solved in polynomial time.

THEOREM 3.2. *Given a multipartite graph $G(V, E)$, if an optimum multipartite clique M^* is such that for every level i ($i = 1, 2, \dots, n-1$), M^* has a node (say v_i) such that all neighbors of v_i in $G^{i,i+1}$ are also in M^*, then MPCP is polynomially solvable.*

Proof. For a node u in level i, let $N^r(u)$ denote the neighbors of u in $G^{i,i+1}$ and let $N^l(u)$ denote the neighbors of u is $G^{i-1,i}$. For a set $S \subseteq V_i$, $N^{(r)}(S) = \cap_{i \in S} N^r(i)$ denotes the common neighborhood, in $G^{i,i+1}$, of nodes in S. Similarly, $N^{(l)}(S)$ denotes the common neighborhood in $G^{i-1,i}$ of nodes in S.

For a node u in level 1, it is easy to see that the induced subgraph

$$S_u = N^l(N^r(u)) \times N^r(u) \times N^r(N^r(u)) \times \cdots \times N^r(N^r \cdots N^r(u) \cdots)$$

is a MPC (provided that all the sets in the above product are nonempty). Consider an optimal MPC (say M^*) which satisfies the hypothesis. Thus, for every level i ($i = 1, 2, \dots, n-1$), M^* has a node (say v_i) in the optimum solution such that all neighbors of v_i in $G^{i,i+1}$ are also in the optimum solution. Then, it can easily be verified that $S_{v_1} = M^*$. Hence, the polynomial time procedure which considers every node u from level 1 and constructs the set S_u will find M^*. ∎

The above conditions on the multipartite clique may be true in certain real environments. Many manufacturers in the computer industry offer a base model (a complete product) as a *shell* and offer several options on the base model to define other products in the product line. They store inventory of the *shell* and use it as a vanilla box while customizing other products with options. If the supplier (or supplying plant) of at least one key component to the shell also follows a similar strategy, then the multipartite cliques of interest are such that they require the conditions in Theorem 3.2 to be satisfied.

As with the biclique problem, the node cardinality and node weighted counterparts of multipartite clique problems can be considered. To the best of our knowledge, the complexity of these problems is open. To the best of our knowledge, the complexity of the unweighted versions of MPCP, MPCF, and MPCS is open.

4. CONCLUSIONS

In this paper, we studied biclique and multipartite clique problems. Among biclique problems, we considered the maximum (edge) biclique problem (MBP) and its weighted version, the maximum (edge) weighted biclique problem (MWBP) in bipartite graphs. MBP and MWBP are interesting problems from a theoretical point of view and have applications in manufacturing and formal concept analysis. We showed that MWBP and four variants of MBP are NP-complete. For random bipartite graphs, we presented a result about the size of a maximum balanced biclique. This result and an observation suggest that the number of edges in a maximum edge cardinality biclique may be considerably larger than the number of edges in a maximum balanced biclique, and it highlights the difference between the well known maximum balanced node cardinality biclique problem and MBP. We also presented three versions of the multipartite clique problem.

REFERENCES

1. R. K. Ahuja, T. L. Magnanti, and J. B. Orlin, "Network Flows," Prentice Hall, Englewood Cliffs, NJ, 1993.
2. N. Alon, J. H. Spencer, and P. Erdos, "The Probabilistic Method," Wiley, New York, 1992.
3. B. Ganter, personal communication.
4. B. Ganter and R. Wille, "Formale Begriffsanalyse—Mathematische Grundlagen," Springer-Verlag, Berlin/New York, 1996.
5. M. S. Garey and D. S. Johnson, "Computers and Intractibility: A Guide to NP-Completeness," Freeman, New York, 1979.
6. G. H. Hardy and E. M. Wright, "An Introduction to the Theory of Numbers," Clarendon, Oxford, 1954.
7. D. S. Hochbaum, Approximating clique and biclique problems, *J. Algorithms* **29** (1997), 174–200.
8. S. Krolak-Schwerdt and P. Orlik, Ein Verfahren zur Klassifikation zweimodaler bin" arer Daten (A method for classifying twomodal binary data), presented at the Annual Conference of the German Classification Society, 1996.
9. G. L. Nemhauser and L. A. Wolsey, "Integer Programming and Combinatorial Optimization," Wiley, New York, 1988.
10. R. Peeters, "The Maximum Edge Biclique Problem Is NP-Complete," Tilburg University Department of Econometrics research memorandum, 2000.
11. J. M. Swaminathan and S. Tayur, Management of broader product lines through delayed product differentiation using vanilla boxes, *Management Sci.* **44** (1998), 161–172.