Compact Mixed Integer Programming Formulations for the Minimum Biclique Cover Problem

Bruno Burin[1], Hamidreza Validi*[2], Bochuan Lyu[3], and Illya V. Hicks[3]

[1]Computer Engineering Department, Military Institute of Engineering, RJ, Brazil
[2]Industrial, Manufacturing & Systems Engineering, Texas Tech University, TX, US
[3]Computational Applied Mathematics & Operations Research, Rice University, TX, US

July 29, 2024

Abstract

Given a simple graph $G = (V, E)$ with vertex set V and edge set E, the minimum biclique cover problem seeks to cover all edges of the graph with a minimum number of bicliques (i.e., complete bipartite subgraphs). This paper proposes two compact mixed integer programming (MIP) formulations for solving the minimum biclique cover problem on general graphs: (i) a natural formulation in the edge space and (ii) an extended formulation in the edge and vertex spaces. Despite the exponential-size natural MIP formulation of Cornaz and Fonlupt (*Discrete Mathematics*, 2006), our natural formulation employs only a polysize number of their exponential "no-good" cuts, along with another set of polysize valid inequalities. We also employ bounding and variable fixing procedures that help solve most of our social network instances, which are not solvable to optimality in a one-hour time limit without the bounding and fixing procedures.

1 Introduction

Given a simple graph $G = (V, E)$ with $n := |V|$ vertices and $m := |E|$ edges, the minimum biclique cover problem seeks to cover all edges of the graph with a minimum number of bicliques (i.e., complete bipartite subgraphs). This problem has a wide range of motivations and applications in disjunctive programming [1, 2, 3], combinatorial geometry [4], matrix completion [5], the min-cut max-flow ratio for the multicommodity flow problems [6], the role minimization problems [7], and finite automata [8]. The minimum biclique cover problem is NP-hard, even if G is bipartite [9] or chordal bipartite [10]. However, the problem is polytime solvable on complete graphs [11], grid graphs [12], C4-free graphs [10], and domino-free graphs [13]. Furthermore, Chalermsook et al. [14] show that it is hard to approximate the problem within a factor of $n^{1-\epsilon}$. Figure 1 illustrates a 3×3 grid graph and a corresponding minimum biclique cover of it.

*Corresponding author, hvalidi@ttu.edu

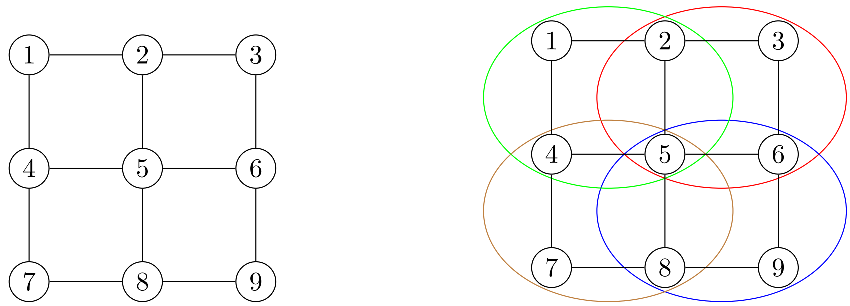

Figure 1: (Left) a 3×3 grid graph G; (Right) a biclique cover of G with the minimum size of 4.

Let $\mathrm{bc}(G)$ be the biclique number of G; i.e., the minimum number of bicliques that are needed to cover all edges of graph G. While finding $\mathrm{bc}(G)$ is hard, there are multiple existing upper bounds of the $\mathrm{bc}(G)$ that can also provide a feasible solution to the problem. Tuza [15] proves that $n - \lfloor \log_2 n \rfloor + 1$ is an upper bound for the $\mathrm{bc}(G)$. Furthermore, any feasible solution of the vertex cover problem is a feasible solution for the biclique cover problem. While the vertex cover is among the six basic NP-hard problems [16], it can be solved efficiently with the existing state-of-the-art solvers. In this paper, we employ vertex cover to find an upper bound for $\mathrm{bc}(G)$ as (i) the bound found by vertex cover is tighter than that of Tuza [15], and (ii) it is computationally efficient.

Regarding the lower bound of $\mathrm{bc}(G)$, Jukna and Kulikov [17] prove that $\mathrm{bc}(G)$ is bounded below by $\nu(G)^2/m$ with $\nu(G)$ be the matching number of graph G. Harary et al. [18] show that $\mathrm{bc}(G)$ can be bounded below by $\lceil \log_2(\chi(G)) \rceil$ with $\chi(G)$ be the chromatic number of G. Lyu and Hicks [19] provide a lower bound of $\lceil \log_2(\mathrm{mc}(G^c)) \rceil$, where $\mathrm{mc}(G^c)$ is the number of maximal cliques in the complement graph of G, and prove that their proposed bound is tighter than that of Harary et al. [18], i.e., they prove $mc(G^c) \geq \chi(G)$. Furthermore, Monson et al. [5] show that an independent set of edges (i.e., edges with no common endpoints) that are not contained in any 4-cycle can provide a lower bound for $\mathrm{bc}(G)$. In this paper, we propose an efficient MIP formulation for the independent edge set idea of Monson et al. [5] and show its effectiveness in not only lower bounding $\mathrm{bc}(G)$, but also fixing decision variables in our proposed compact MIP formulations of the biclique cover problem.

Despite a rich literature on bounds of $\mathrm{bc}(G)$, there is only one paper on solving the problem to optimality via MIP to our knowledge: Cornaz and Fonlupt [20]. They propose a MIP formulation with exponentially many "no-good" constraints for solving the problem to optimality. While they prove their separation problem can be solved in polytime, they conduct no computational experiments on their proposed formulation. This paper proposes a stronger version of the MIP formulation of Cornaz and Fonlupt [20] and shows that the separation problem corresponding to the stronger formulation is NP-hard. Furthermore, we show that only a polysize number of the "no-good" cuts, along with a set of coupling constraints, provides a correct compact formulation of the problem. We also propose an extended compact MIP formulation of the problem that has a competitive computational performance in comparison with the compact one containing "no-good" cuts. To further accelerate the solving process of the extended formulation, we propose a variable fixing procedure for both compact formulations.

Our contributions and outline. In this paper, Section 2 reviews an existing MIP formulation with exponentially many "no-good" inequalities and proposes two new compact MIP formulations for solving the minimum biclique cover problem: (i) a natural formulation in the edge space and (ii) an extended formulation in the vertex and edge spaces. Section 3 proposes a bounding framework along with a fixing procedure for both compact MIP formulations and assesses their impact on a set of social network instances. Section 4 concludes the paper and provides future research directions for the exact solution methods of the biclique cover problem.

2 MIP Formulations

This section reviews an existing MIP formulation with exponentially many "no-good" cuts for solving the biclique cover problem to optimality. We show that the "no-good" cuts can be strengthened; however, strengthening the formulation makes the separation problem NP-hard. Then, we show that a polysize number of "no-good" cuts, along with a set of coupling constraints, is sufficient to ensure the formulation's correctness. Furthermore, we propose an extended compact formulation for the problem with the same order in the number of decision variables. Our computational results in Section 3 show that the extended formulation is quite competitive in comparison with the natural compact formulation, which contains a polysize number of "no-good" cuts on challenging instances of the problem. Before proceeding with MIP formulations, we provide a formal definition of biclique that is employed in the proofs of this section.

Definition 1 (Biclique Definition (page 4 of Bondy and Murty [21])). *A simple graph $G = (V, E)$ is a biclique graph if*

1. *its vertex set V can be partitioned into two subsets A and B such that every edge has one endpoint in A and one endpoint in B; and*
2. *every vertex in A is joined to every vertex in B.*

Lemma 1 provides an equivalent definition for a biclique that will be employed in a MIP formulation correctness proof.

Lemma 1. *A simple graph $G = (V, E)$ is a biclique if and only if it satisfies the following properties:*

1. *the degree of every vertex is at least one;*
2. *the inclusion of edges $\{a, b\}$ and $\{c, d\}$ in G implies the inclusion of $\{a, c\}$ ($\{a, d\}$) and $\{b, d\}$ ($\{b, c\}$) in G; and*
3. *G contains no cycle of size 3 (i.e., C_3).*

Proof. ($\Longrightarrow$) Since biclique is a bipartite graph, then it does not contain any odd cycle [22]. Thus, G does not contain any cycle of size 3. Let sets A and B be the parts of V such that every vertex in A is joined to every vertex in B. Since both A and B are not empty, the degree of every vertex is at least one. Let $\{a, b\}$ and $\{c, d\}$ be two arbitrary edges in G. Without loss of generality, we assume that $a \in A$ and $b \in B$. If $c \in A$ and $d \in B$, then $\{a, d\}$ and $\{b, c\}$ are edges in G. Otherwise, $c \in B$ and $d \in A$, which implies that $\{a, c\}$ and $\{b, d\}$ are edges in G.

($\Longleftarrow$) We first prove that there are no odd cycles in G. By the third property, G has no cycle of size 3. For contradiction purposes, suppose that G contains an odd cycle of length larger than

3; i.e., $(u_1, u_2, u_3, \cdots, u_q, u_1)$ with $q \geq 5$. Then, we have edges $\{u_1, u_2\}$ and $\{u_3, u_4\}$ in G. Thus, either edge $\{u_1, u_3\}$ or edge $\{u_1, u_4\}$ is in G by the second property. In the first case, there is a 3-cycle, which is a contradiction. In the second case, we find another odd cycle with length $q-2$; i.e., $(u_1, u_4, \cdots, u_q, u_1)$. Repeat the above procedure until we reach a 3-cycle. This is a contradiction.

Since G does not contain any odd cycles, it is a bipartite graph [22]. Now, it suffices to show that G is a complete bipartite graph. Hence, the vertices of G can be partitioned into two subsets A and B. Let a and b be two arbitrary vertices such that $a \in A$ and $b \in B$. For contradiction purposes, assume that $\{a, b\}$ is not an edge in G. Since the degrees of a and b are both at least one, then there exist vertices $c \in B$ and $d \in A$ such that both $\{a, c\}$ and $\{b, d\}$ are edges in G. Since $a, d \in A$, then $\{a, d\}$ is not an edge of G (as G is a bipartite graph). Thus, $\{a, b\}$ must be an edge in G, which is a contradiction. $\square$

2.1 An existing MIP formulation

To the best of our knowledge, Cornaz and Fonlupt [20] proposed the first MIP formulation with exponentially many "no-good" cuts for solving the problem on general graphs. They define dependent and minimal dependent sets to propose their "no-good" cuts in their formulation (see Definition 2 below). Furthermore, Figure 2 provides an illustration of (minimal) dependent sets.

Definition 2 (Dependent and minimal dependent sets)**.** *Given graph $G = (V, E)$, an edge set $F \subseteq E$ is a dependent set if for every biclique $B \subseteq E$, we have $F \nsubseteq B$. Furthermore, F is a minimal dependent set if any proper subset of F can be contained in a biclique.*

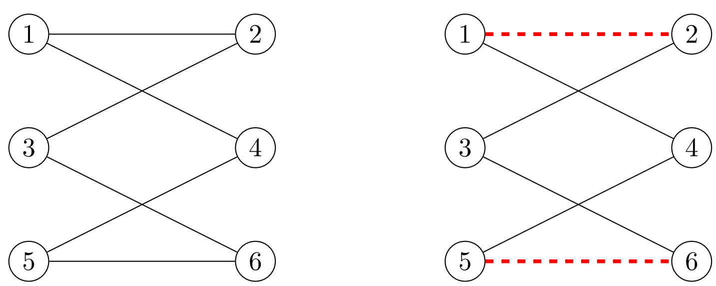

Figure 2: (Left) Graph $G = (V, E)$; (Right) $F = \big\{\{1, 2\}, \{5, 6\}\big\}$ is a (minimal) dependent set.

Let k be an upper bound for the biclique number. Then, we define $[k] := \{1, \dots, k\}$. For every biclique $i \in [k]$, we define z_i as the binary decision variable for choosing biclique i. For every edge $e \in E$ and every biclique $i \in [k]$, we define x_e^i as the binary decision variable for assigning edge e to biclique i. The MIP formulation of Cornaz and Fonlupt [20], which is called the natural formulation in this paper, is provided below.

$$\min \quad \sum_{i=1}^{k} z_i \tag{1a}$$

$$x_e^i \leq z_i \qquad \forall e \in E,\ \forall i \in [k] \tag{1b}$$

$$\sum_{i=1}^{k} x_e^i \geq 1 \qquad \forall e \in E \tag{1c}$$

$$\sum_{e \in F} x_e^i \leq |F| - 1 \qquad \forall \text{ minimal dependent } F \in \mathcal{F}(G),\ \forall i \in [k] \tag{1d}$$

$$x \in \{0,1\}^{mk}. \tag{1e}$$

Here, objective function (1a) minimizes the number of biclique covers. Constraints (1b) state that if an edge is assigned to a biclique, then the biclique is selected. Constraints (1c) state that every edge must be covered by at least one biclique. Constraints (1d) state that for every biclique $i \in [k]$ and every minimal dependent $F \in \mathcal{F}$, the number of edges included in the biclique must be at most $|F| - 1$. Furthermore, Cornaz and Fonlupt [20] show that the linear programming (LP) relaxation of their formulation can be solved in polytime by proving the polytime solvability of the separation problem for constraints (1d). In the next section, we propose a stronger variant of constraints (1d) and prove that its corresponding separation problem is hard.

2.2 A compact natural formulation

This section proposes a compact natural formulation in the same space of MIP formulation (1). Before proceeding with the compact natural formulation, the following remark provides a stronger variant of constraints (1d).

Remark 1. *A strengthened variant of "no-good" cuts* (1d) *are provided below.*

$$\sum_{e \in F} x_e^i \leq (|F| - 1) z_i \qquad \forall \textit{ minimal dependent } F \in \mathcal{F}(G),\ \forall i \in [k]. \tag{2}$$

Although constraints (2) are stronger than constraints (1d), the following theorem shows the hardness of their corresponding separation problem.

Theorem 1. *The (fractional) separation problem of constraints* (2) *is NP-hard.*

Proof. Let $(\hat{x}, \hat{z}) \in (0,1)^{mk \times k}$ be a fractional point that satisfies constraints (1b)-(1c) and let $j \in [k]$ be a biclique cover. The separation problem asks for the existence of a minimal dependent set $\hat{F}$ such that

$$\sum_{e \in \hat{F}} \hat{x}_e^j > (|\hat{F}| - 1)\hat{z}_j.$$

Equivalently, the separation problem asks for the existence of a minimal dependent set $\hat{F}$ such that

$$\sum_{e \in \hat{F}} (\hat{x}_e^j - \hat{z}_j) > -\hat{z}_j.$$

Hence, one needs to find a minimal dependent set of the maximum weight, where the weight of each edge is $\hat{x}_e^j - \hat{z}_j$. Since (i) the maximum-size minimal dependent set problem is NP-hard by Theorem 19 of Cornaz and Fonlupt [20] and (ii) the maximum-weight minimal dependent set problem is a generalization of the maximum-size minimal dependent set problem, the separation problem of constraints (2) is NP-hard. $\square$

To propose our compact natural formulation, we first define the notion of crossing edges as follows. Throughout the paper, we consider the endpoints of edges e and f as $\{a, b\}$ and $\{c, d\}$, respectively.

Definition 3 (crossing edges)**.** *For every pair of vertex-disjoint edges $e = \{a, b\}$ and $f = \{c, d\}$ with $\{a, b\} \cap \{c, d\} = \emptyset$, we define the crossing sets CR^1_{ef} and CR^2_{ef} as follows.*

$$\mathrm{CR}^1_{ef} := \Big\{\{a, d\}, \{b, c\}\Big\} \cap E, \qquad \mathrm{CR}^2_{ef} := \Big\{\{a, c\}, \{b, d\}\Big\} \cap E. \tag{3}$$

Figure 3 provides an illustration of crossing edges for a vertex-disjoint pair of edges e and f.

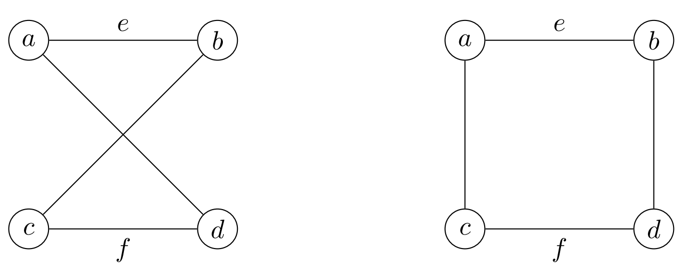

Figure 3: (Left) a graph with $\mathrm{CR}^1_{ef} := \Big\{\{a, d\}, \{b, c\}\Big\}$ and $\mathrm{CR}^2_{ef} := \emptyset$; (Right) a graph with $\mathrm{CR}^1_{ef} := \emptyset$ and $\mathrm{CR}^2_{ef} := \Big\{\{a, c\}, \{b, d\}\Big\}$.

We now propose our compact natural MIP formulation as follows.

$$\begin{aligned}
\min \quad & \sum_{i=1}^{k} z_i && && \text{(4a)}\\
& x_e^i \le z_i && \forall e \in E,\ \forall i \in [k] && \text{(4b)}\\
& \sum_{i=1}^{k} x_e^i \ge 1 && \forall e \in E && \text{(4c)}\\
& x_e^i + x_f^i \le z_i && && \\
& x_e^i + x_f^i \le z_i && \forall \text{ disjoint } e,\ f \in E \text{ with } |\,\mathrm{CR}^1_{ef}\,| < 2, \text{ and } |\,\mathrm{CR}^2_{ef}\,| < 2,\ \forall i \in [k] && \text{(4d)}\\
& \sum_{e \in C_3} x_e^i \le 2 z_i && \forall \text{ cycle } C_3 \subseteq E,\ \forall i \in [k]. && \text{(4e)}
\end{aligned}$$

Here, objective function (4a), constraints (4b) and covering constraints (4c) are similar to their counterparts (1a), (1b) and (1c) in the MIP formulation of Cornaz and Fonlupt [20], respectively.

Constraints (4d) state that two disjoint edges cannot belong to the same biclique if their crossing numbers are less than 2. Constraints (4e) state that all edges of a cycle of size three cannot be assigned to the same biclique. We note that constraints (4d) and (4e) are specific types of "no-good" cuts (2) when dependent set F has sizes two and three, respectively.

We finally discuss the crossing constraints for disjoint edges that complete the compact natural formulation. We recall that the endpoints of edges e and f are $\{a, b\}$ and $\{c, d\}$, respectively.

$$x_e^i + x_f^i \leq z_i + x_{ad}^i$$
$$x_e^i + x_f^i \leq z_i + x_{bc}^i \qquad \forall \text{ disjoint } e,\ f \in E \text{ with } |\,\mathrm{CR}_{ef}^1\,| = 2 \text{ and } |\,\mathrm{CR}_{ef}^2\,| < 2,\ \forall i \in [k] \tag{4f}$$
$$x_e^i + x_f^i \leq z_i + x_{ac}^i$$
$$x_e^i + x_f^i \leq z_i + x_{bd}^i \qquad \forall \text{ disjoint } e,\ f \in E \text{ with } |\,\mathrm{CR}_{ef}^1\,| < 2 \text{ and } |\,\mathrm{CR}_{ef}^2\,| = 2,\ \forall i \in [k] \tag{4g}$$
$$x_e^i + x_f^i \leq z_i + x_{ac}^i + x_{ad}^i$$
$$x_e^i + x_f^i \leq z_i + x_{bc}^i + x_{bd}^i \qquad \forall \text{ disjoint } e,\ f \in E \text{ with } |\,\mathrm{CR}_{ef}^1\,| = 2 \text{ and } |\,\mathrm{CR}_{ef}^2\,| = 2,\ \forall i \in [k] \tag{4h}$$
$$x \in \{0,1\}^{mk}, z \in [0,1]^k. \tag{4i}$$

Constraints (4f) and (4g) state that if two disjoint edges are assigned to a biclique and the size of exactly one set of their corresponding crossing edges is two, then both edges of the set must be assigned to the biclique. Constraints (4h) state that if two disjoint edges are assigned to a biclique and they are part of a clique K_4, then the edges of at least one crossing set need to be assigned to the biclique. Constraints (4i) state the bounds and integrality of x and z variables. We note that the integrality of z variables can be relaxed as it is implied by constraints (4b).

Theorem 2. *The biclique covering formulation* (4) *is correct.*

Proof. Let $(\hat{x}, \hat{z}) \in \{0,1\}^{k(m+1)}$ be a binary point. We are to show that $(\hat{x}, \hat{z})$ represents a biclique covering if and only if it satisfies constraints of formulation (4).

($\Longrightarrow$) Suppose that point $(\hat{x}, \hat{z})$ represents a biclique covering of graph G. We are to show that the point satisfies constraints of formulation (4). Because every non-empty biclique cover contains at least one edge, the point satisfies constraints (4b). Because point $(\hat{x}, \hat{z})$ represents a covering of edges, every edge of graph G is covered by at least one biclique. So, $(\hat{x}, \hat{z})$ satisfies constraints (4c). Furthermore, $(\hat{x}, \hat{z})$ satisfies constraints (4d)-(4e) by correctness of the MIP formulation (1). Finally, $(\hat{x}, \hat{z})$ satisfies constraints (4f)-(4h) because every vertex in a part of a biclique is connected to every vertex in the other part.

($\Longleftarrow$) Suppose that $(\hat{x}, \hat{z})$ satisfies constraints of formulation (4). We are to show that $(\hat{x}, \hat{z})$ represents a biclique covering. Constraints (4b) imply that an edge cannot belong to a biclique if the biclique is not selected. Constraints (4c) imply that every edge must belong to at least one biclique. Constraints (4d)-(4h) imply that every cover forms a biclique by Lemma 1. $\square$

The next section provides an extended formulation for the biclique cover problem that computationally runs faster than the natural formulation.

2.3 An extended formulation

This section provides an extended formulation in vertex and edge space. Let $D = (V, A)$ be the bidirected variant of graph G with $A := \cup_{\{u,v\} \in E} \Big\{(u, v) \cup (v, u)\Big\}$. For every vertex $u \in V$ and every biclique $i \in [k]$, binary decision variable y^1_{ui} (y^2_{ui}) is one if vertex u is assigned to the first (second) partition of biclique i. The extended MIP formulation is provided below.

$$
\begin{aligned}
\min \quad & \sum_{i=1}^{k} z_i && && \text{(5a)}\\
& x^i_{uv} \le y^1_{ui} && && \\
& x^i_{uv} \le y^2_{vi} && \forall (u,v) \in A, \forall i \in [k] && \text{(5b)}\\
& y^1_{ui} + y^2_{vi} \le z_i + x^i_{uv} && \forall (u,v) \in A, \forall i \in [k] && \text{(5c)}\\
& y^1_{ui} + y^2_{ui} \le z_i && \forall u \in V, \forall i \in [k] && \text{(5d)}\\
& \sum_{i=1}^{k} (x^i_{uv} + x^i_{vu}) \ge 1 && \forall \{u,v\} \in E && \text{(5e)}\\
& y^1_{ui} + y^2_{vi} \le z_i && && \\
& y^1_{vi} + y^2_{ui} \le z_i && \forall \{u,v\} \in \binom{V}{2} \setminus E, \forall i \in [k] && \text{(5f)}\\
& x \in \{0,1\}^{mk}, y \in \{0,1\}^{2nk}, z \in \{0,1\}^{k}. && && \text{(5g)}
\end{aligned}
$$

Constraints (5b) state that for every biclique if arc (u, v) is assigned to the biclique, then vertex u (vertex v) is assigned to the first (second) partition of the biclique. Constraints (5c) state that for every biclique i and for every arc (u, v) if vertices u and v are assigned to the first partition and the second partition of the biclique, respectively, then arc (u, v) must be assigned to biclique i. Constraints (5d) state that every vertex can be assigned to at most one partition of a biclique. Constraints (5e) state that every edge is covered by at least one biclique. Constraints (5f) state that non-edge pairs of vertices cannot belong to different partitions of a biclique. Constraints (5g) impose the integrality of decision variables. The following remark shows that constraints $x^i_{uv} + x^i_{vu} \le z_i$ are implied for every biclique $i \in [k]$ and every edge $\{u, v\} \in E$.

Remark 2. *The following constraints are implied by MIP formulation* (5)*.*

$$x^i_{uv} + x^i_{vu} \le z_i \qquad \forall \{u,v\} \in E, \forall i \in [k]. \tag{6}$$

Proof. Let $(\hat{x}, \hat{y}, \hat{z})$ be a solution of MIP formulation (5). For any biclique $i \in [k]$ and every edge $\{u, v\} \in E$, we have

$$\hat{x}^i_{uv} + \hat{x}^i_{vu} \le \hat{y}^1_{ui} + \hat{y}^2_{ui} \le \hat{z}_i.$$

Here, the first inequality holds by constraints (5b). The second inequality holds by constraints (5d). □

Before proving the correctness of MIP formulation (5), we provide a definition of the directed biclique graph as follows.

Definition 4 (Directed Biclique Definition)**.** *A directed graph $D = (V, A)$ is a directed biclique graph if*

1. *its vertex set V can be partitioned into two parts A and B such that every directed edge $(u, v) \in A$ has its tail (u) in A and its head (v) in B; and*

2. *every vertex in A is joined to every vertex in B.*

The following theorem proves that the extended formulation (5) is correct.

Theorem 3. *The extended formulation* (5) *is correct.*

Proof. Let $(\hat{x}, \hat{y}, \hat{z}) \in \{0,1\}^{mk+2nk+k}$ be a binary point. We are to show that $(\hat{x}, \hat{y}, \hat{z})$ represents a directed biclique cover if and only if it satisfies constraints of formulation (5).

($\Longrightarrow$) Suppose that $(\hat{x}, \hat{y}, \hat{z})$ represents a directed biclique cover. We are to show that $(\hat{x}, \hat{y}, \hat{z})$ satisfies constraints of formulation (5). Because every arc $(u, v) \in A$ that belongs to a biclique $i \in [k]$ has its tail in part 1 and its head in part 2 of biclique i, the point satisfies constraints (5b). For every directed biclique $i \in [k]$, if vertex u belongs to part 1 of the biclique and vertex v belongs to part 2 of the biclique, then arc (u, v) belongs to biclique i. So, $(\hat{x}, \hat{y}, \hat{z})$ satisfies constraints (5c). For any directed biclique $i \in [k]$ and every vertex u in biclique i, vertex u does not belong to parts 1 and 2 simultaneously. So, $(\hat{x}, \hat{y}, \hat{z})$ satisfies constraints (5d). Because $(\hat{x}, \hat{y}, \hat{z})$ represents a directed biclique cover, every arc (u, v) or (v, u) must belong to a biclique. Hence, $(\hat{x}, \hat{y}, \hat{z})$ satisfies constraints (5e). For any directed biclique $i \in [k]$, no non-adjacent pair of vertices can belong to different parts of biclique i. So, $(\hat{x}, \hat{y}, \hat{z})$ satisfies constraints (5f).

($\Longleftarrow$) Suppose that $(\hat{x}, \hat{y}, \hat{z})$ satisfies all the constraints of formulation (5). We are to show that $(\hat{x}, \hat{y}, \hat{z})$ represents a directed biclique cover. Constraints (5b)-(5c) and constraints (5f) imply that distinct vertices u and v belong to parts 1 and 2 of a biclique, respectively, if and only if arc (u, v) belongs to the biclique. Furthermore, constraints (5d) ensure that a vertex of a biclique cannot belong to parts 1 and 2 of the biclique, simultaneously. Constraints (5e) and (6) imply that for every edge $\{u, v\} \in E$, either arc (u, v) or (v, u) need to be covered by at least one biclique. Hence, $(\hat{x}, \hat{y}, \hat{z})$ forms a directed biclique cover in directed graph $D = (V, A)$. $\square$

3 Computational Experiments

This section provides computational results for calculating upper and lower bounds as well as the performance of our proposed compact formulations. We conduct our experiments on a machine running Windows 11 Enterprise with an Intel Core i9-13900 processor (2.00 GHz base, 5.2 GHz turbo) and 32 GB RAM. We employ Python programming language and Gurobi 11.0.0 to run our experiments. The experiments are run on a set of social network instances. The code and instances are available at https://github.com/bcburin/compact-mip-formulations-for-the-minimum-biclique-cover-problem.

3.1 Upper bound

As the proposed MIP formulations depend on the value of k (i.e., an upper bound on the biclique cover number), we test two existing methods for the value of k: (i) Tuza's upper bound [15], and (ii) the vertex cover number (folklore). While Tuza [15] suggests $n - \lfloor \log_2 n \rfloor + 1$ as an upper bound for the biclique cover number, the vertex cover number is obtained by solving the following optimization problem with $t_v \in \{0,1\}$ be the binary decision variable for selecting a vertex $v \in V$ as a member of a vertex cover.

$$
\begin{aligned}
\min \quad & \sum_{v \in V} t_v && && \text{(7a)}\\
& t_u + t_v \geq 1 && \forall \{u,v\} \in E && \text{(7b)}\\
& t \in \{0,1\}^n. && && \text{(7c)}
\end{aligned}
$$

Table 1 provides a comparison between Tuza's and the vertex cover (VC) numbers. Both approaches find upper bounds in less than a second; however, the vertex cover number is consistently smaller than Tuza's number. Hence, we employ the vertex cover number as the value of k for our final experiments with natural and extended MIP formulations because it provides a tighter bound for the biclique cover number.

Table 1: A comparison between the vertex cover and Tuza's numbers.

			Tuza (1983)		VC (folklore)	
instance	n	m	UB	time	UB	time
ieee30	30	41	27	0.00	**16**	0.00
karate	34	78	30	0.00	**14**	0.00
surfers	43	336	39	0.00	**32**	0.00
ieee57	57	78	53	0.00	**30**	0.00
dolphins	62	159	58	0.00	**34**	0.00
lesmis76	77	254	72	0.00	**42**	0.00
adjnoun	112	425	107	0.00	**59**	0.01
football	115	613	110	0.00	**94**	0.01
ieee118	118	179	113	0.00	**61**	0.00
jazz	198	2,742	192	0.00	**158**	0.02
ieee300	300	409	293	0.00	**136**	0.02

3.2 Lower bound and a fixing procedure

This section provides computational results on the following existing lower bounds of the minimum biclique cover problem and shows how the third lower bounding procedure provides an efficient fixing procedure.

1. $\nu^2(G)/m$ of Junka and Kulikov [17] with $\nu(G)$ be the matching number of G (JK (2009));
2. $\lceil \log_2(\mathrm{mc}(G^c)) \rceil$ of Lyu and Hicks [19] with $\mathrm{mc}(G^c)$ be the number of maximal cliques in the complement graph of G (LH (2024)); and
3. the number of independent edges with no crossings of Monson, Pullman, and Rees [5] (MPR (1995)).

Furthermore, we propose an MIP formulation for the third procedure, which is provided below. In this formulation, binary decision variable $w_e \in \{0,1\}$ is one if edge $e \in E$ is selected as an independent edge that is not paired with any other edge $f \in E$ with $|\,\mathrm{CR}^1_{ef}\,| = 2$ or $|\,\mathrm{CR}^2_{ef}\,| = 2$. Furthermore, $\delta(v)$ denotes the set of edges that are incident to vertex v.

$$\max \quad \sum_{e \in E} w_e \tag{8a}$$

$$w(\delta(v)) \leq 1 \qquad \forall v \in V \tag{8b}$$

$$w_e + w_f \leq 1 \qquad \forall \text{ disjoint } e,\ f \in E \text{ with } |\,\mathrm{CR}^1_{ef}\,| = 2 \text{ or } |\,\mathrm{CR}^2_{ef}\,| = 2 \tag{8c}$$

$$w \in \{0,1\}^m. \tag{8d}$$

Table 2 compares the listed lower bound procedures in terms of their values and calculation times. We set a time limit of 3,600 seconds for all procedures. In this table, JK (2009), LH (2024), and MPR (1995) represent the lower bounding procedures proposed by Junka and Kulikov [17], Lyu and Hicks [19], and Monson, Pullman, and Rees [5], respectively. One can easily note that MIP formulation (8), which is developed to capture the lower bounding idea of Monson, Pullman, and Rees [5], outperforms the other lower bounding ideas by a considerable difference. Furthermore, MIP formulation (8) solves all the benchmark instances in at most three minutes. We finally note that the computational results for the lower bounding idea of Harary et al. [18] are skipped as their proposed bound (i.e., $\lceil \log_2(\chi(G)) \rceil$ with $\chi(G)$ be the chromatic number of G) is dominated by the number of maximal cliques of the complementary graph G, i.e., $mc(G^c)$ [19].

Theorem 4 (Lyu and Hicks [19])**.** *Given a graph G, $mc(G^c) \geq \chi(G)$.*

The lower bounding idea of Monson, Pullman, and Rees [5], along with our proposed MIP formulation (8), not only provides the tightest bound among the existing ones but also triggers fixing binary decision variables in both natural and extended MIP formulations. Table 3 provides the percentages of fixings for x, y, and z variables in both compact natural and extended formulations. We will observe the effect of these variable fixings in next subsection.

Table 2: A performance comparison between lower bounding ideas of Junka and Kulikov [17] (JK (2009)), Lyu and Hicks [19] (LH (2024)), and Monson, Pullman, and Rees [5] (MPR (1995)).

			JK (2009)		LH (2024)		MPR (1995)	
instance	n	m	LB	time	LB	time	LB	time
ieee30	30	41	6	0.00	12	0.01	**<u>14</u>**	0.00
karate	34	78	3	0.00	8	0.00	**<u>12</u>**	0.01
surfers	43	336	2	0.00	11	0.01	**<u>16</u>**	1.16
ieee57	57	78	11	0.00	22	8.43	**<u>28</u>**	0.01
dolphins	62	159	6	0.00	19	0.87	**<u>29</u>**	0.03
lesmis	77	254	5	0.00	21	1.29	**<u>24</u>**	0.06
adjnoun	112	425	7	0.01	30	3,289.88	**<u>54</u>**	0.11
football	115	613	6	0.01	30	2,057.60	**<u>57</u>**	0.49
ieee118	118	179	19	0.01	30	3,139.47	**<u>56</u>**	0.02
jazz	198	2,742	4	0.01	30	1,313.90	**<u>72</u>**	162.78
ieee300	300	409	44	0.01	30	2,787.73	**<u>130</u>**	0.06

Table 3: Percentages of fixing for x, y , and z variables in compact natural and extended formulations.

			Natural Fix (%)		Extended Fix (%)		
instance	n	m	x	z	x	y	z
ieee30	30	41	29.88	87.50	29.88	40.83	87.50
karate	34	78	13.19	85.71	13.19	30.25	85.71
surfers	43	336	2.38	50.00	2.38	18.60	50.00
ieee57	57	78	33.50	93.33	33.50	45.85	93.33
dolphins	62	159	15.56	85.29	15.56	39.90	85.29
lesmis	77	254	5.40	57.14	5.40	17.81	57.14
adjnoun	112	425	11.63	91.53	11.63	44.13	91.53
football	115	613	5.64	60.64	5.64	30.06	60.64
ieee118	118	179	28.72	91.80	28.72	43.57	91.80
jazz	198	2,742	1.20	45.57	1.20	16.57	45.57
ieee300	300	409	30.38	95.59	30.38	41.42	95.59

3.3 Final experiments

We conclude this section by running the following sets of experiments to (i) assess the impact of the proposed fixing procedure of Section 3.2 and (ii) provide a computational comparison between the natural compact formulation (4) and extended compact formulation (5) on the benchmark instances within a 3,600-second time limit:

1. Natural compact formulation (4) with and without lower bound fixing procedure of Section 3.2; and

2. Extended compact formulation (5) with and without lower bound fixing procedure of Section 3.2.

Table 4 shows the computational effect of the proposed variable fixing procedure on the performance of the natural compact formulation (4). One can easily observe that the natural compact formulation (4) without the edge fixing procedure solves only 3 out of 11 instances. In comparison, the natural compact formulation (4) with the edge fixing procedure solves 7 out of 11 instances within the one-hour time limit. Specifically, the natural compact formulation with the proposed fixing procedure solves `ieee118` and `ieee300` in less than one minute, while they are not solvable to optimality without applying the fixing procedure within the 1-hour time limit. Among the solved instances under both with and without fixing circumstances (i.e., `ieee30`, `karate`, and `ieee57`), the natural compact formulation with fixing procedure outperforms the natural compact formulation without the fixing procedure in solving time.

Table 5 demonstrates the effect of the edge fixing procedure of Section 3.2 on the extended compact formulation (5). The edge fixing procedure helps increase the number of solved instances by the extended formulation from 2 to 7. Specifically, the extended formulation with the edge fixing procedure solves `ieee118` and `ieee300` in less than two minutes while the memory crashes for both instances when the fixing is not applied. For the benchmark instances that are solved by the extended formulation with and without fixing (i.e., `ieee30` and `karate`), the extended formulation with fixing outperforms its basic version in solving time.

Table 4: The effect of variable fixing on the performance of the compact natural formulation (4) within a time limit of 3,600 seconds. TL and MEM denote time limit reached and memory crash, respectively.

			Natural wo/ Fix			Natural w/ Fix		
instance	n	m	LB	UB	time	LB	UB	time
ieee30	30	41	15	15	0.14	15	15	**0.03**
karate	34	78	12	12	0.51	12	12	**0.09**
surfers	43	336	16	32	TL	**17**	**30**	TL
ieee57	57	78	30	30	71.55	30	30	**0.26**
dolphins	62	159	29	32	TL	32	32	**12.85**
lesmis	77	254	24	39	TL	**28**	**32**	TL
adjnoun	112	425	54	-	TL	58	58	**757.00**
football	115	613	**57**	-	TL	56	**60**	TL
ieee118	118	179	56	60	TL	60	60	**10.31**
jazz	198	2,742	-	-	MEM	-	-	MEM
ieee300	300	409	130	-	TL	134	134	**59.27**

Table 5: The effect of variable fixing on the performance of the compact extended formulation (5) within a time limit of 3,600 seconds. TL and MEM denote time limit reached and memory crash, respectively.

			Extended wo/ Fix			Extended w/ Fix		
instance	n	m	LB	UB	time	LB	UB	time
ieee30	30	41	15	15	0.40	15	15	**0.06**
karate	34	78	12	12	0.86	12	12	**0.08**
surfers	43	336	16	25	TL	**18**	**23**	TL
ieee57	57	78	28	30	TL	30	30	**0.64**
dolphins	62	159	29	32	TL	32	32	**6.39**
lesmis	77	254	24	32	TL	30	32	TL
adjnoun	112	425	54	-	TL	58	58	**133.59**
football	115	613	57	-	TL	58	**90**	TL
ieee118	118	179	56	60	MEM	60	60	18.93
jazz	198	2,742	72	-	TL	**73**	**156**	TL
ieee300	300	409	-	-	MEM	134	134	**72.25**

Finally, Table 6 summarizes our computational results with natural and extended compact formulations with the edge fixing procedure. While both formulations solve the same instances to optimality with the edge fixing procedure, and they are competitive in solving times, the extended formulation generally[1] provides better bounds for unsolved instances (i.e., `surfers`, `lesmis`, `football`, and `jazz`). Specifically for `jazz`, the extended compact formulation provides lower and upper bounds of 73 and 156, respectively, while the memory crashes with the natural formulation.

[1] For `football`, the natural formulation provides a better upper bound (60 vs. 90); however, the extended formulation provides a better lower bound (58 vs. 56) after the one-hour time limit.

Table 6: Computational comparisons between the natural and extended compact formulations. TL and MEM denote time limit reached and memory crash, respectively.

			Natural			Extended		
instance	n	m	LB	UB	time	LB	UB	time
ieee30	30	41	15	15	**0.03**	15	15	0.06
karate	34	78	12	12	0.09	12	12	**0.08**
surfers	43	336	17	30	TL	**18**	**23**	TL
ieee57	57	78	30	30	**0.26**	30	30	0.64
dolphins	62	159	32	32	12.85	32	32	**6.39**
lesmis	77	254	28	32	TL	**30**	32	TL
adjnoun	112	425	58	58	757.00	58	58	**133.59**
football	115	613	56	**60**	TL	**58**	90	TL
ieee118	118	179	60	60	**10.31**	60	60	18.93
jazz	198	2,742	-	-	MEM	**73**	**156**	TL
ieee300	300	409	134	134	**59.27**	134	134	72.25

4 Conclusion and Future Work

This paper proposes two compact MIP formulations for the minimum biclique cover problem on general graph: (i) a natural MIP formulation in the edge space; and (ii) an extended MIP formulation in the vertex and edge spaces. We prove that both proposed MIP formulations are correct and propose a fixing procedure that accelerates the solving process of the problem on a set of social network instances. As future work, one may propose MIP formulations with exponentially many variables and explore employing large-scale optimization approaches (e.g., branch-and-price) for solving the problem.

Acknowledgement

This work is supported by the National Science Foundation (NSF) grant DMS-2318790 titled *AMPS: Novel Combinatorial Optimization Techniques for Smartgrids and Power Networks.*

References

[1] Joey Huchette and Juan Pablo Vielma. A combinatorial approach for small and strong formulations of disjunctive constraints. *Mathematics of Operations Research*, 44(3):793–820, 2019.

[2] Bochuan Lyu, Illya V. Hicks, and Joey Huchette. Modeling combinatorial disjunctive constraints via junction trees. *Mathematical Programming*, 2023.

[3] Bochuan Lyu. *Biclique Partitions, Biclique Covers, and Disjunctive Constraints*. PhD thesis, Rice University, 2023.

[4] Noga Alon. Neighborly families of boxes and bipartite coverings. *The Mathematics of Paul Erdös II*, pages 27–31, 1997.

[5] Sylvia D Monson, Norman J Pullman, and Rolf Rees. A survey of clique and biclique coverings and factorizations of (0, 1)-matrices. *Bull. Inst. Combin. Appl*, 14:17–86, 1995.

[6] Oktay Günlük. A new min-cut max-flow ratio for multicommodity flows. *SIAM Journal on Discrete Mathematics*, 21(1):1–15, 2007.

[7] Alina Ene, William Horne, Nikola Milosavljevic, Prasad Rao, Robert Schreiber, and Robert E Tarjan. Fast exact and heuristic methods for role minimization problems. In *Proceedings of the 13th ACM symposium on Access control models and technologies*, pages 1–10, 2008.

[8] Jérôme Amilhastre, Philippe Janssen, and Marie-Catherine Vilarem. FA minimisation heuristics for a class of finite languages. In *International Workshop on Implementing Automata*, pages 1–12. Springer, 1999.

[9] James Orlin. Contentment in graph theory: covering graphs with cliques. In *Indagationes Mathematicae (Proceedings)*, volume 80, pages 406–424. Elsevier, 1977.

[10] Haiko Müller. On edge perfectness and classes of bipartite graphs. *Discrete Mathematics*, 149(1-3):159–187, 1996.

[11] Peter C Fishburn and Peter L Hammer. Bipartite dimensions and bipartite degrees of graphs. *Discrete Mathematics*, 160(1-3):127–148, 1996.

[12] Krystal Guo, Tony Huynh, and Marco Macchia. The biclique covering number of grids. *The Electronic Journal of Combinatorics*, pages P4–27, 2019.

[13] Jérôme Amilhastre, Marie-Catherine Vilarem, and Philippe Janssen. Complexity of minimum biclique cover and minimum biclique decomposition for bipartite domino-free graphs. *Discrete Applied Mathematics*, 86(2-3):125–144, 1998.

[14] Parinya Chalermsook, Sandy Heydrich, Eugenia Holm, and Andreas Karrenbauer. Nearly tight approximability results for minimum biclique cover and partition. In *European Symposium on Algorithms*, pages 235–246. Springer, 2014.

[15] Zsolt Tuza. Covering of graphs by complete bipartite subgraphs; complexity of 0–1 matrices. *Combinatorica*, 4:111–116, 1984.

[16] M. R. Garey and D. S. Johnson. *Computers and Intractability: A Guide to the Theory of NP-Completeness (Series of Books in the Mathematical Sciences)*. W. H. Freeman, first edition edition, 1979.

[17] Stasys Jukna and Alexander S Kulikov. On covering graphs by complete bipartite subgraphs. *Discrete Mathematics*, 309(10):3399–3403, 2009.

[18] Frank Harary, Derbiau Hsu, and Zevi Miller. The biparticity of a graph. *Journal of graph theory*, 1(2):131–133, 1977.

[19] Bochuan Lyu and Illya V Hicks. Maximal clique and edge-ranking bounds of biclique cover number. *arXiv preprint arXiv:2302.12775*, 2023.

[20] Denis Cornaz and Jean Fonlupt. Chromatic characterization of biclique covers. *Discrete Mathematics*, 306(5):495–507, 2006.

[21] JA Bondy and USR Murty. *Graph Theory*. Springer, 2008.

[22] Armen S Asratian, Tristan MJ Denley, and Roland Häggkvist. *Bipartite graphs and their applications*, volume 131. Cambridge University Press, 1998.