Hereditary Cohesive Subgraphs Enumeration on Bipartite Graphs: The Power of Pivot-based Approaches

ABSTRACT

Finding cohesive subgraphs from a bipartite graph is a fundamental operator in bipartite graph analysis. In this paper, we focus on the problem of mining cohesive subgraphs from a bipartite graph that satisfy a hereditary property. Here a cohesive subgraph meets the hereditary property if all of its subgraphs satisfy the same property as itself. We show that several important cohesive subgraph models, such as maximal biclique and maximal k-biplex, satisfy the hereditary property. The problem of enumerating all maximal hereditary subgraphs was known to be NP-hard. To solve this problem, we first propose a novel and general pivot-based enumeration framework to efficiently enumerate all maximal hereditary subgraphs in a bipartite graph. Then, based on our general framework, we develop a new pivot-based algorithm with several pruning techniques to enumerate all maximal bicliques. We prove that the worst-case time complexity of our pivot-based maximal biclique enumeration algorithm is $O(m\times 2^{\frac{n}{2}})$ (or $O(m\times 1.414^{n})$) which is near optimal since there exist up to $O(2^{\frac{n}{2}})$ maximal bicliques in a bipartite graph with n vertices and m edges. Third, on the basis of our general framework, we also devise a novel pivot-based algorithm with several non-trivial pruning techniques to enumerate maximal k-biplexes in a bipartite graph. Finally, we conduct extensive experiments using 11 real-world bipartite graphs to evaluate the proposed algorithms. The results show that our pivot-based solutions can achieve one order of magnitude (three orders of magnitude) faster than the state-of-the-art maximal biclique enumeration algorithms (maximal k-biplex enumeration algorithms).

1 INTRODUCTION

Bipartite graphs are ubiquitous in real-world applications. In a bipartite graph, the vertices can be divided into two disjoint sets and each edge connects a vertex in one set to a vertex in the other set. Some representative examples of real-world bipartite graphs include user-item networks [41, 45], author-publication networks [23], and biological networks [24]. Real-world bipartite graphs often contain cohesive subgraph structures which correspond to communities or densely-connected groups in bipartite graphs. Mining cohesive subgraphs from a bipartite graph is a fundamental operator in bipartite graph analysis which has been widely used in many applications, such as community detection [21, 23], online recommendation [18, 37], and biological network analysis [7, 44].

SIGMOD '23, June 18–23, 2023, Seattle, WA, USA

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM…$15.00
https://doi.org/XXXXXXX.XXXXXXX

There exist many different cohesive subgraph models in bipartite graphs. Notable examples include maximal biclique [1, 10, 15, 25, 29, 50], maximal k-biplex [42, 48, 49], (α, β)-core [8, 28], k-bitruss [46, 52], and quasi biclique [20, 31, 47]. Among them, the maximal biclique model, perhaps, is the most fundamental model, as all the other cohesive subgraph models can be considered as a *relaxed* biclique model.

Instead of focusing on a particular cohesive subgraph model, in this paper, we study a family of cohesive subgraph models in bipartite graphs that meet the hereditary property, namely hereditary cohesive subgraphs. Here a subgraph G' is called a hereditary subgraph if (1) G' satisfies a property $\mathcal{P}$ and (2) every induced subgraph of G' also meets the property $\mathcal{P}$. For convenience, we refer to a subgraph that meets a hereditary property $\mathcal{P}$ as a $\mathcal{P}$-subgraph. A $\mathcal{P}$-subgraph G' is a maximal $\mathcal{P}$-subgraph if there is no other $\mathcal{P}$-subgraph containing G'. Given a bipartite graph G, our goal is to enumerate all maximal $\mathcal{P}$-subgraphs from G. To our knowledge, such a maximal $\mathcal{P}$-subgraph enumeration problem has not been investigated before. We show that both the maximal biclique and maximal k-biplex are maximal $\mathcal{P}$-subgraphs, thus both the classic maximal biclique enumeration and maximal k-biplex enumeration problems are special instances of our problem.

Motivations. Practical solutions for maximal $\mathcal{P}$-subgraph enumeration can be applied to many applications, and two of them are summarized as follows.

Community detection in bipartite graphs. Detecting communities in a bipartite graph is an important graph analysis task [7, 21, 24, 37, 44]. We can use the hereditary cohesive subgraph to model communities in a bipartite graph. The communities detected by the hereditary cohesive subgraph model often exhibit strong robustness, since the removal of any subset of vertices from the community does not destroy the structural property. In effect, the classic maximal biclique and maximal k-biplex models have been widely used for community detection applications [21, 29, 42, 48] due to such a nice hereditary property and cohesive property. Thus, the solution for maximal $\mathcal{P}$-subgraph enumeration can provide a general framework for community detection in bipartite graphs, which captures a family of different community models.

Fraud detection in user-item networks. Consider an online user-item rating network (e.g., Amazon's user-product rating network), where the users can give ratings to the items. The item owners may wish to improve their items' ratings by hiring some fake users to frequently give high ratings to their items. Clearly, the set of fake users and the set of their rated items often form a densely-connected subgraph. Once again, we can use the hereditary cohesive subgraph model, such as maximal biclique or maximal k-biplex to detect such fake users [48]. As a result, the approaches to maximal $\mathcal{P}$-subgraph enumeration can also be used for identifying possible rating frauds in online user-item networks.

Although the significance of the maximal $\mathcal{P}$-subgraph enumeration problem, a practical solution for this problem is still lacking

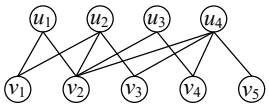


Figure 1: Running example: a bipartite graph G

due to the intrinsic challenges of this problem. First, as indicated in [22], the problem of enumerating all maximal $\mathcal{P}$-subgraphs from a bipartite graph is NP-hard. Thus, there does not exist a polynomial algorithm to solve this problem unless NP = P. Second, since the hereditary property $\mathcal{P}$ is arbitrary (the enumeration algorithm for our problem should work for any hereditary property $\mathcal{P}$) and internal structure of the $\mathcal{P}$ is unclear, it is quite non-trivial to use such a property $\mathcal{P}$ to design an enumeration algorithm. Moreover, existing solutions for maximal biclique enumeration and maximal k-biplex enumeration also cannot be generalized to handle our problem. This is because different properties $\mathcal{P}$ give rise to different enumeration problems, and enumeration techniques for a specified $\mathcal{P}$ are unlikely to work for every hereditary subgraph enumeration problem. For example, the techniques for maximal biclique enumeration are often very hard to extend to solve the problem of maximal k-biplex enumeration [48], because maximal bicliques can be enumerated within the 2-hop neighborhood of each vertex, while maximal k-biplex does share such a nice property. As a result, new techniques need to be developed to solve our problem.

Contributions. In this paper, we formulate and develop efficient algorithms to enumerate maximal hereditary subgraphs on bipartite graphs, with special focus on enumerating maximal bicliques and maximal k-biplexes which are two representative maximal hereditary subgraphs on bipartite graphs. In summary, the main contributions of this paper are as follows.

A novel and general pivot-based framework. To solve the maximal $\mathcal{P}$-subgraph enumeration problem, we first devise a basic enumeration algorithm inspired by the classic set enumeration technique [6, 39]. Such a basic solution, however, may explore all possible subsets of vertices, thus resulting in many unnecessary computations. To improve the efficiency, we develop a novel and general pivot-based backtracking framework to enumerate all maximal $\mathcal{P}$-subgraphs. Our framework is based on the basic set enumeration technique together with a novel and carefully-designed pivoting strategy. With such a powerful pivoting technique, our framework can significantly prune redundant computations in the enumeration procedure. To our knowledge, the proposed framework is the first practical solution for enumerating all maximal $\mathcal{P}$-subgraphs in bipartite graphs. In addition, our framework is very general which can provide useful guidelines to design practical solutions for enumerating specified maximal hereditary subgraphs (e.g., maximal biclique and maximal k-biplex).

New maximal biclique enumeration algorithms. We propose a new maximal biclique enumeration algorithm based on our general pivoting principle. The striking feature of our algorithm is that its worst-case time complexity is near optimal. Specifically, we prove that the time complexity of our pivot-based algorithm is $O(m \times 2^{\frac{n}{2}})$ (or $O(m \times 1.414^n)$) which is near optimal, because there exist up to $O(2^{\frac{n}{2}})$ maximal bicliques on a bipartite graph with n vertices and m edges [38]. In addition, we also present several non-trivial optimization techniques (including early termination and ordering techniques) to further improve the efficiency of our algorithm.

Novel maximal k-biplex enumeration algorithms. Based on our general pivoting framework, we also develop a novel pivot-based algorithm to enumerate all maximal k-biplexes. Note that the detailed implementation of our pivoting technique for maximal k-biplex enumeration is significantly different from that for maximal biclique enumeration, although both of them are based on our general pivoting principle. To our knowledge, this is the first pivot-based enumeration algorithm for maximal k-biplex enumeration. In addition, we also develop several non-trivial pruning techniques to further improve the efficiency of our algorithm when enumerating large maximal k-biplexes (with size no less than a given threshold).

Extensive experiments. We conduct extensive experiments to evaluate the efficiency and effectiveness of the proposed approaches using 11 real-world bipartite graphs. The results show that our pivot-based algorithms are one order of magnitude faster than the state-of-the-art algorithm for maximal biclique enumeration, and three orders of magnitude faster than the state-of-the-art algorithms for maximal k-biplex enumeration. In addition, we also apply two representative hereditary cohesive subgraph models, i.e., maximal biclique and maximal k-biplex, to the applications of fraud detection and community detection; and the results demonstrate the high effectiveness of our solutions.

2 PROBLEM STATEMENT

Let $G = (U, V, E)$ be an undirected and unweighted bipartite graph with two disjoint vertex sets U and V and an edge set $E \subseteq U \times V$. Denote by $n = |U|+|V|$ and $m = |E|$ the number of vertices and edges in G, respectively. For a vertex $u \in U$, we define $N_u(G)$ as the set of neighbors of u in G, i.e., $N_u(G) = \{w \in V | (u, w) \in E\}$. The degree of a vertex $u \in U$ in G is denoted by $d_u(G) = |N_u(G)|$. Similar definitions are also applied for the vertices in V. Given a pair (A, B) of vertex sets with $A \subseteq U$ and $B \subseteq V$, we define $G(A, B) = (A, B, E')$ as the induced subgraph of G, where $E' = \{(u, v) \in E | u \in A, v \in B\}$. Denote by $\mathcal{P}$ a graph property. Then, the hereditary property for bipartite graphs is defined as follows.

Definition 1 (Hereditary property). Given a bipartite graph G and a subgraph H of G with a graph property $\mathcal{P}$, H is said to meet the hereditary property if every subgraph of H meets $\mathcal{P}$.

Based on Definition 1, the maximal $\mathcal{P}$-subgraph of a bipartite graph G is defined as follows.

Definition 2 (Maximal $\mathcal{P}$-subgraph). Given a bipartite graph G and a fixed hereditary property $\mathcal{P}$, a subgraph H of G is called a maximal $\mathcal{P}$-subgraph if 1) H meets the property $\mathcal{P}$, and 2) there is no other subgraph H' of G containing H and also meeting $\mathcal{P}$.

Two notable instances of maximal $\mathcal{P}$-subgraphs on bipartite graphs are maximal biclique [29] and maximal k-biplex [35], which have been widely studied in the literature [1, 10, 14, 15, 25, 48, 50]. Below, we give the formal definitions of these two models.

Definition 3 (Maximal biclique). Given a bipartite graph $G = (U, V, E)$, a subgraph $H = G(A, B)$ induced by a pair of vertex sets (A, B) is a maximal biclique of G if 1) $\forall u \in A, v \in B$, there is an edge $(u, v) \in E$, and 2) there is no other subgraph H' that contains H and satisfies 1).

Definition 4 (k-biplex). Given a bipartite graph $G = (U, V, E)$, a subgraph $H = G(A, B)$ induced by a pair of sets (A, B) in G is a k-biplex, if in the subgraph H, every vertex $u \in A$ has a degree no less than $|B| - k$, and every vertex $v \in B$ has a degree no less than $|A| - k$.

Definition 5 (Maximal k-biplex). A k-biplex $G(A, B)$ of G is called a maximal k-biplex if there is no other k-biplex $G(A', B')$ of G that contains $G(A, B)$.

The following lemma shows that both maximal biclique and maximal k-biplex are maximal $\mathcal{P}$-subgraphs.

Lemma 1. Both biclique and k-biplex (for any integer k) meet the hereditary property.

PROOF. First, it is easy to see that any subgraph of a biclique is also a biclique, thus the biclique satisfies the hereditary property. Second, we show that k-biplex also meets the hereditary property. Suppose that (A, B) is a k-biplex of G. If a subset (A', B') of (A, B) is not a k-biplex, there must be a vertex $u \in A'$ (or $v \in B'$) has a degree less than $|B'| - k$ (or $|A'| - k$). However, if we add all vertices in $(A \setminus A', B \setminus B')$ back into (A', B'), the degree of $u \in A'$ (or $v \in B$) in B' (or A') will keep unchanged, thus the degree of $u \in A'$ (or $v \in B'$) in B (or A) is also less than $|B| - k$ (or $|A| - k$), which is contradiction. This lemma is proved. □

Due to the nice hereditary property and the internal cohesiveness of these two hereditary subgraph instances, both maximal biclique and maximal k-biplex are widely used in bipartite graph analysis applications such as community detection [21, 29], fraud detection [48], and text mining [33, 40]. Thus, it is important to develop efficient algorithms to solve the problem of enumerating all maximal $\mathcal{P}$-subgraphs on bipartite graphs.

Problem definition. Given a bipartite graph G, our goal is to (1) develop a general framework to efficiently enumerate all maximal $\mathcal{P}$-subgraph of G for any fixed hereditary property $\mathcal{P}$, and (2) apply the proposed framework to enumerate all maximal bicliques and maximal k-biplexes on G respectively.

As shown by Yannakakis and Lewis [22], the problem of enumerating all maximal $\mathcal{P}$-subgraphs on G is NP-hard for any given hereditary property $\mathcal{P}$. Therefore, there does not exist a polynomial algorithm to solve our problems unless NP = P. Previous approaches to enumerate maximal $\mathcal{P}$-subgraphs are mainly based on a reverse enumeration technique which may explore each maximal $\mathcal{P}$-subgraph multiple times [2, 5, 11, 48], thus they are often very costly on real-world bipartite graphs (as indicated in our experiments). In this work, we develop a novel and general pivot-based enumeration framework to solve this problem. Our framework can significantly prune the enumeration space by the pivoting technique, thus it is much more efficient than existing reverse enumeration based framework. In the following sections, we will first introduce our pivot-based framework, followed by two new solutions for enumerating all maximal bicliques and k-biplexes.

3 A GENERAL PIVOT-BASED FRAMEWORK

In this section, we first propose a basic set-enumeration technique to enumerate all maximal $\mathcal{P}$-subgraphs, based on which we then develop a general pivot-based enumeration framework.

Algorithm 1: A basic set enumeration framework

```
   Input: The bipartite graph G
   Output: All maximal P-subgraphs of G
1  Enum(∅, ∅, U, V, ∅, ∅);
2  Function: Enum(R_U, R_V, C_U, C_V, X_U, X_V)
3      if C_U ∪ C_V = ∅ then
4          if X_U ∪ X_V = ∅ then Output (R_U, R_V) as a result;
5          return;
6      foreach w ∈ C_U do
7          Branch(R_U, R_V, w, C_U \ {w}, C_V, X_U, X_V);
8          C_U ← C_U \ {w}; X_U ← X_U ∪ {w};
9      foreach w ∈ C_V do
10         Branch(R_V, R_U, w, C_V \ {w}, C_U, X_V, X_U);
11         C_V ← C_V \ {w}; X_V ← X_V ∪ {w};
12 Function: Branch(R_U, R_V, w, C_U, C_V, X_U, X_V)
13     Generate sets C'_U ⊆ C_U, C'_V ⊆ C_V, X'_U ⊆ X_U, and X'_V ⊆ X_V such that
         (R_U ∪ {w, u}, R_V) and (R_U ∪ {w}, R_V ∪ {v}) are the partial results, where
         u ∈ C'_U ∪ X'_U and v ∈ C'_V ∪ X'_V;
14     Enum(R_U ∪ {w}, R_V, C'_U, C'_V, X'_U, X'_V);
```

3.1 The Basic Set Enumeration Framework

Our basic enumeration algorithm is inspired by the classic set enumeration technique [6, 39]. The key idea of our basic technique is that it makes use of a recursive approach to explore each subgraph of a given bipartite graph G, and then determines whether each explored subgraph is a valid maximal $\mathcal{P}$-subgraph. Since we focus on enumerating the maximal subgraphs that satisfy the hereditary property, many subgraphs of G are unnecessary to be explored in the recursive process. Specifically, if the algorithm finds a subgraph H that does not satisfy the property $\mathcal{P}$, then every subgraph H' of G containing H definitely cannot form a $\mathcal{P}$-subgraph. To avoid this, the algorithm maintains a set in each recursive call, called *candidate set*, to ensure that each vertex in the candidate set can form a larger $\mathcal{P}$-subgraph with the currently explored subgraph. With this idea, each subgraph detected by such a recursive approach must be a $\mathcal{P}$-subgraph, thus all maximal $\mathcal{P}$-subgraphs can be identified. Algorithm 1 details such a set enumeration technique.

In Algorithm 1, it invokes the procedure *Enum* to enumerate all maximal $\mathcal{P}$-subgraphs (line 1), which requires six parameters: R_U, R_V, C_U, C_V, X_U, and X_V. Here the vertex sets (R_U, R_V) represent the current partial result, (C_U, C_V) are the candidate sets in which each vertex can be used to expand (R_U, R_V), and (X_U, X_V) are the exclusion sets containing vertices in (C_U, C_V) that have already been used to expand (R_U, R_V). Initially, both the sets (R_U, R_V) and (C_U, C_V) are set to empty, while (C_U, C_V) is set to (U, V) (line 1). Then, in each recursion, every vertex in (C_U, C_V) is used to expand the current partial result (R_U, R_V) (lines 6-11). When a vertex w in (C_U, C_V) is added into (R_U, R_V), the algorithm needs to update the candidate and exclusion sets (line 13). The algorithm further invokes a new sub-recursive call to enumerate the maximal results containing $(R_U \cup \{w\}, R_V)$ (or $(R_U, R_V \cup \{w\})$) (line 14). After processing the vertex w, w will be moved from C_U (or C_V) to X_U (or X_V) to avoid outputting non-maximal $\mathcal{P}$-subgraphs (line 8 and line 11). Specifically, if X_U (or X_V) is not empty, there must exist previously-processed vertices that can be used to expand the current (R_U, R_V). This implies that (R_U, R_V) cannot be the maximal result even if there is no vertex in (C_U, C_V). Thus, whenever both (C_U, C_V) and (X_U, X_V) are empty, the algorithm outputs the current (R_U, R_V) as a maximal $\mathcal{P}$-subgraph (lines 3-4).

It is easy to show that the worst-case time complexity of Algorithm 1 is $O(f(n)\times 2^n)$ ($f(n)$ denotes the time consumed to generate the candidate sets and exclusion sets), because the algorithm may traverse all possible subsets of the vertex set. For example, if G itself is a $\mathcal{P}$-subgraph, then every possible subgraph of G will be explored by the algorithm. To improve Algorithm 1, we next propose a pivot-based enumeration framework which can prune many redundant computations involved in Algorithm 1.

3.2 Novel Pivot-based Enumeration Framework

The key to speed up Algorithm 1 is to devise a pruning technique to reduce the unnecessary computations that produce non-maximal $\mathcal{P}$-subgraphs. However, it is quite non-trivial to achieve this, because the graph property $\mathcal{P}$ is arbitrary and the internal structure of the $\mathcal{P}$-subgraph is unclear. Our solution to achieve this is based on an in-depth analysis of Algorithm 1.

An in-depth analysis. Given a bipartite graph G and the candidate sets (C_U, C_V), suppose that (A, B) is a maximal $\mathcal{P}$-subgraph of G with $(A, B) \subseteq (C_U, C_V)$ (i.e, $A \subseteq C_U$ and $B \subseteq C_V$). In the top recursion of Algorithm 1, each vertex in (C_U, C_V) is used to expand the initial result $(\emptyset, \emptyset)$. It is easy to see that the result (A, B) can be detected by a recursive call that enumerates all maximal $\mathcal{P}$-subgraphs containing $u \in C_U$ if $u \in A$. When finishing this computation, the vertex u is removed from C_U and the vertices in $(C_U \setminus \{u\}, C_V)$ are further selected to enumerate the remaining maximal $\mathcal{P}$-subgraphs. If the next selected vertex $v \in C_V$ satisfies $v \in B$, then all non-maximal $\mathcal{P}$-subgraphs contained in $(A \setminus \{u\}, B)$ can also be enumerated, thus causing many redundant computations. To reduce this, we can select another vertex v' in C_V with $v' \notin B$ to enumerate all maximal $\mathcal{P}$-subgraphs containing v' in $(C_U \setminus \{u\}, C_V)$. Since $v' \notin B$, each $\mathcal{P}$-subgraph (A', B') with $v' \in B'$ enumerated by such a sub-recursive call must be excluded in $(A \setminus \{u\}, B)$. Thus, we can avoid enumerating non-maximal $\mathcal{P}$-subgraphs contained in $(A \setminus \{v\}, B)$. Moreover, when all vertices in $(C_U \setminus (A \setminus \{u\}), C_V \setminus B)$ have been used to expand the initial result $(\emptyset, \emptyset)$, all the remaining vertices in the candidate sets are exactly $(A \setminus \{u\}, B)$, which does not contain any maximal $\mathcal{P}$-subgraph of G since (A, B) is a maximal $\mathcal{P}$-subgraph. As a consequence, we can terminate those sub-recursive calls. In other words, all vertices in $(A \setminus \{u\}, B)$ are not necessary to expand the current partial result. Based on this analysis, we present a general pivoting technique as described in the following theorem.

THEOREM 3.1 (GENERAL PIVOTING RULE). *Consider a recursion with six parameters (R_U, R_V, C_U, C_V, X_U, X_V). Let $u \in C_U \cup X_U$ be a pivot vertex. Then, the vertices in $(P_U \subseteq C_U, P_V \subseteq C_V)$ can be skipped to expand (R_U, R_V) if and only if any result containing (R_U, R_V) but not u is not contained in $G(R_U \cup P_U, R_V \cup P_V)$.*

PROOF. In the recursion, suppose that each vertex in $(C_U \setminus P_U, C_V \setminus P_V)$ has been used to expand the current partial result (R_U, R_V). Then, the remaining vertices in the candidate sets are (P_U, P_V). To obtain all maximal $\mathcal{P}$-subgraphs, the only thing left to do is to find the results containing (R_U, R_V) in $G(R_U \cup P_U, R_V \cup P_V)$. However, there is no such maximal $\mathcal{P}$-subgraphs in $G(R_U \cup P_U, R_V \cup P_V)$ based on the condition of generating (P_U, P_V). Thus, the recursion can be terminated immediately and the Theorem is proved. □

Note that a symmetrical pivot vertex v in $C_V \cup X_V$ can also be selected by Theorem 3.1. For simplicity, in the rest of this paper, we mainly discuss the pivot vertex which is selected from $C_U \cup X_U$, because the other pivot in $C_V \cup X_V$ can be obtained similarly. Then, based on the pivoting rule shown in Theorem 3.1, we can derive the following result, which provides guidelines for pivot-based branching.

THEOREM 3.2. *Given a pivot vertex $u \in C_U \cup X_U$ and the skipping sets $(P_U \subseteq C_U, P_V \subseteq C_V)$, then any maximal $\mathcal{P}$-subgraph containing (R_U, R_V) must belong to one of the following three cases.*

(1) It contains the vertex u.
(2) It does not contain u, but contains at least one vertex in $C_V \setminus P_V$.
(3) It does not contain vertices in $\{u\} \cup C_V \setminus P_V$, but contains at least one vertex in $C_U \setminus (P_U \cup \{u\})$.

PROOF. It is easy to see that Theorem 3.2 omits only the maximal $\mathcal{P}$-subgarph that does not contain vertices in $C_U \cup C_V \setminus (P_U \cup P_V)$, but contains at least one vertex in $P_U \cup P_V$. However, by Theorem 3.1, there is no maximal $\mathcal{P}$-subgraphs of $G(R_U \cup P_U, R_V \cup P_V)$ containing (R_U, R_V). This means that any maximal $\mathcal{P}$-subgraph of G that contains (R_U, R_V) must contains at least one vertex in $C_U \cup C_V \setminus (P_U \cup P_V)$, thus the Theorem is proved. □

Example 1. Consider a bipartite graph G in Fig. 1. Suppose that the hereditary property $\mathcal{P}$ is specified as the k-biplex property (see Definition 4). Let $(R_U = \{u_1\}, R_V = \{v_1\})$ and $(C_U = \{u_2, u_3, u_4\}, C_V = \{v_2, v_3, v_4, v_5\})$ be the current k-biplex and the candidate sets of a recursion in Algorithm 1, respectively. Let v_3 be the pivot vertex. When $k = 1$, we can obtain the skipping sets $(P_U = \{u_2, u_4\}, P_V = \{v_2\})$ by Theorem 3.1, as $(\{u_1, u_2, u_4\}, \{v_1, v_2, v_3\})$ is a maximal 1-biplex of G. With Theorem 3.2, we can see that the maximal 1-biplex $(\{u_1, u_2, u_4\}, \{v_1, v_2, v_3\})$ belongs to the case (1), which contains the pivot vertex v_3. $(\{u_1, u_2, u_3\}, \{v_1, v_2\})$ belongs to the case (2), which does not contain v_3 but contains a vertex in $C_U \setminus P_U = \{u_3\}$, and $(\{u_1, u_4\}, \{v_1, v_2, v_4\})$ belongs to the case (3), which does not contain any vertex in $\{v_3\} \cup C_U \setminus P_U = \{v_3, u_3\}$. In addition, it is easy to check that there is no other maximal 1-biplex of G containing $(\{u_1\}, \{v_1\})$ that does not fall into the three cases of Theorem 3.2.

Implementation details. Algorithm 2 details our pivot-based enumeration framework. Note that Algorithm 2 is similar to Algorithm 1. The key difference is that Algorithm 2 uses the proposed pivoting technique (Theorem 3.1) to prune unnecessary sub-recursive calls. Specifically, the algorithm first selects two pivot vertices $u \in C_U \cup X_U$ and $v \in C_V \cup X_V$, and constructs two pairs of skipping sets (P_U, P_V) and (P'_U, P'_V) according to Theorem 3.1 (lines 6-7). When the size of $P'_U \cup P'_V$ is larger than that of $P_U \cup P_V$, the algorithm utilizes the skipping sets (P'_U, P'_V) to reduce enumeration branches (line 8), because in this case more recursive calls can be reduced by the pivot vertex v. Then, the algorithm recursively expands the current result (R_U, R_V) with each vertex in (C_U, C_V), but not in (P_U, P_V) (lines 9-14). The following theorem shows the correctness of Algorithm 2.

THEOREM 3.3. *Algorithm 2 correctly computes all maximal $\mathcal{P}$-subgraphs of a bipartite graph G.*

Algorithm 2: The pivot-based enumeration framework

Input: The bipartite graph G
Output: All maximal $\mathcal{P}$-subgraphs of G
1 $PivotEnum(\emptyset, \emptyset, U, V, \emptyset, \emptyset)$;
2 **Function:** $PivotEnum(R_U, R_V, C_U, C_V, X_U, X_V)$
3 **if** $C_U \cup C_V = \emptyset$ **then**
4 **if** $X_U \cup X_V = \emptyset$ **then** Output (R_U, R_V) as a result;
5 **return**;
6 Derive the skipping sets ($P_U \subseteq C_U, P_V \subseteq C_V$) by the pivot vertex u (Theorem 3.1) that is selected from $C_U \cup X_U$;
7 Derive the skipping sets ($P'_U \subseteq C_U, P'_V \subseteq C_V$) by the pivot vertex v (Theorem 3.1) that is selected from $C_V \cup X_V$;
8 **if** $|P'_U| + |P'_V| > |P_U| + |P_V|$ **then** $P_U \leftarrow P'_U$; $P_V \leftarrow P'_V$;
9 **foreach** $w \in C_U \setminus P_U$ **do**
10 $PivotBranch(R_U, R_V, w, C_U \setminus \{w\}, C_V, X_U, X_V)$;
11 $C_U \leftarrow C_U \setminus \{w\}$; $X_U \leftarrow X_U \cup \{w\}$;
12 **foreach** $w \in C_V \setminus P_V$ **do**
13 $PivotBranch(R_V, R_U, w, C_V \setminus \{w\}, C_U, X_V, X_U)$;
14 $C_V \leftarrow C_V \setminus \{w\}$; $X_V \leftarrow X_V \cup \{w\}$;
15 **Function:** $PivotBranch(R_U, R_V, w, C_U, C_V, X_U, X_V)$
16 Generate sets $C'_U \subseteq C_U, C'_V \subseteq C_V, X'_U \subseteq X_U$, and $X'_V \subseteq X_V$ such that $(R_U \cup \{w, u\}, R_V)$ and $(R_U \cup \{w\}, R_V \cup \{v\})$ are the partial results, where $u \in C'_U \cup X'_U$ and $v \in C'_V \cup X'_V$;
17 $PivotEnum(R_U \cup \{w\}, R_V, C'_U, C'_V, X'_U, X'_V)$;

PROOF. Based on Theorem 3.2, it is easy to see that every maximal $\mathcal{P}$-subgraph of a given bipartite graph G can be enumerated exactly. Thus, it is sufficient to prove that Algorithm 2 does not output redundant and non-maximal solutions. When each vertex in $(C_U \setminus P_U, C_V \setminus P_V)$ has be used to expand the current partial result, the vertex requires to be moved from (C_U, C_V) to (X_U, X_V), which guarantees that every maximal $\mathcal{P}$-subgraph of G cannot be output multiple times. This is because for every $\mathcal{P}$-subgraph of G that has be detected by the algorithm so far, there always exists at least one vertex in the exclusion sets (X_U, X_V). Moreover, the current partial result is considered as the maximal only if both candidate and exclusion sets are empty, which can filter out all non-maximal $\mathcal{P}$-subgraphs of G during enumerations since if the current partial result is not the maximal, either a vertex in candidate sets or a vertex in the exclusion sets can be used to enlarge the current partial result. Thus, the Theorem is established. □

Clearly, compared to Algorithm 1, Algorithm 2 can reduce unnecessary recursive calls by the pivoting technique. Moreover, such a pivot-based enumeration framework is very general and can be used to enumerate any maximal hereditary subgraph. Such a general enumeration framework can guide us to design novel pivot-based algorithms for enumerating some specified maximal hereditary subgraphs. In the following sections, we will show how to use our general framework to devise new and efficient algorithms for enumerating maximal bicliques and maximal k-biplexes.

4 MAXIMAL BICLIQUE ENUMERATION

Recall that maximal biclique is a special instance of maximal $\mathcal{P}$-subgraph. In this section, we aim to develop new algorithms to enumerate all maximal bicliques on bipartite graphs based on the proposed pivoting framework. Below, we first briefly review existing maximal biclique enumeration techniques, and then present our solutions.

4.1 Overview of Existing Algorithms

The problem of enumerating all maximal bicliques on bipartite graphs has been extensively studied [1, 10, 14, 15, 25, 29, 50]. Note that all these existing solutions are fundamentally different from our proposed pivot-based framework (Algorithm 2). Specifically, all existing solutions are based on the following fact. Given a subset set $S \in U$, let $N(S)$ be the common neighbors of S in V, i.e., $N(S) = \cap_{u \in S} N_u(G)$. Then, if $N(S) \neq \emptyset$, $(N(N(S)), N(S))$ must be a maximal biclique. Therefore, all maximal bicliques can be enumerated by exploring the combinations of vertices on the one side [25, 50]. However, unlike these algorithms, our pivot-based framework (Algorithm 2) considers both sides of vertices to expand the partial biclique.

To the best of our knowledge, the state-of-the-art algorithm for maximal biclique enumeration is developed in [10], which improves a so-called pivot-based approach presented in [1] with a 2-hop degree ordering optimization technique. Such a pivot-based approach is mainly based on a neighborhood dominating technique. Specifically, for two vertices u and u' in U, the vertex u' is dominated by u if $N_{u'}(G) \subseteq N_u(G)$. Then, if both u and u' are included in the candidate sets of a recursive call, the vertex u' is not necessary to be used to expand the current partial biclique. This technique [10], however, needs to compute the neighborhood domination relationships between every vertex and all its 2-hop neighbors which is often costly on large bipartite graphs. Moreover, real-world bipartite graphs do not necessarily contain too many *dominated* vertices, thus the pruning power of this technique may be poor on some real-world bipartite graphs (as indicated in our experiments). To address this problem, we will propose a novel and more efficient pivoting technique based on our general framework.

4.2 Pivot-based Maximal Biclique Enumeration

Following the general pivot-based framework (Algorithm 2), our pivot-based maximal biclique enumeration algorithm also admits six parameters $(R_U, R_V, C_U, C_V), X_U, X_V)$, where (R_U, R_V), (C_U, C_V), and (X_U, X_V) represents the current biclique, the candidate sets, and the exclusion sets respectively. Note that to apply our general framework to maximal biclique enumeration, the key is to determine the skipping sets P_U and P_V using the pivot vertex. Below, we first derive a lemma which can be used to determine P_U.

Lemma 2. If $u \in C_U \cup X_U$ is a pivot vertex, all vertices in $C_U \setminus \{u\}$ can be omitted to expand the current biclique (R_U, R_V).

PROOF. It can be seen that $(R_U \cup C_U \cup X_U, R_V)$ is a biclique in G, since each vertex $u \in C_U \cup X_U$ can be used to expand (R_U, R_V) based on Definition 3. Given a pivot vertex u in $C_U \cup X_U$, we can notice that for any maximal clique containing (R_U, R_V) in $G(R_U \cup C_U, R_V \cup C_V)$, it either contains the vertex u or contains a vertex in $C_V \cup X_V$. This is because $(R_U \cup C_U \cup X_U, R_V)$ is a biclique. Thus, all maximal bicliques containing (R_U, R_V) can be obtained by expanding (R_U, R_V) with only vertices in $C_V \cup \{u\}$, which is equivalent to the fact that all vertices in $C_U \setminus \{u\}$ can be omitted to expand (R_U, R_V). □

According to Lemma 2, we only need to use a pivot vertex in $C_U \cup X_U$ and the vertices in C_V to expand the current biclique. However, expanding the current biclique with all vertices in C_V may still result

in many redundant computations. For instance, suppose that (A, B) is a maximal biclique of G with $A = R_U \cup \{u\}$ and $R_V \subseteq B \subseteq (R_V \cup C_V)$, where u is a pivot vertex in C_U. After obtaining all maximal bicliques containing u, if we select any vertex $v \in B \setminus R_V$ to expand (R_U, R_V), a non-maximal biclique (R_U, B) can be generated by the algorithm, thus incurring redundant computations. To overcome this issue, we also need to skip some vertices in C_V to enumerate all maximal bicliques (i.e., P_V is not empty). A natural question is that if we skip some vertices in C_V to expand the current biclique (R_U, R_V), can we still omit all vertices in $C_U \setminus \{u\}$ to expand (R_U, R_V)? Fortunately, we can answer this question affirmatively as stated in the following lemma.

Lemma 3. Given a pivot vertex $u \in C_U \cup X_U$ and the skipping sets $(P_U \subseteq C_U, P_V \subseteq C_V)$, if any biclique (A, B) with $R_U \subset A$ and $R_V \subseteq B \subseteq (R_V \cup P_V)$ can be enlarged by u, then there is no maximal biclique in G that excludes all vertices in $\{u\} \cup C_V \setminus P_V$ but includes at least one vertex in $P_U = C_U \setminus \{u\}$ (i.e., no maximal biclique in G belongs to the case (3) in Theorem 3.2).

Proof. It is obviously established when $P_V = \emptyset$ with Lemma 2. We next analyze the case of $P_V \neq \emptyset$. Suppose that $(R_U \cup A', R_V \cup B')$ is an arbitrary biclique with $A' \subseteq C_U \setminus \{u\}$ and $B' \subseteq P_V$. Then, we have $(R_V \cup B') \subseteq N(R_U \cup A')$ and $(R_U \cup A') \subseteq N(R_V \cup B')$. Since any biclique (A, B) with $R_U \subset A$ and $R_V \subseteq B \subseteq R_V \cup P_V$ can be enlarged by u, it can be derived that all vertices in $R_V \cup B'$ are the neighbors of u, i.e., $(R_V \cup B') \subseteq N_u(G)$. Thus, we further have $(R_U \cup A' \cup \{u\}) \subseteq N(R_V \cup B')$, which means that the biclique $(R_U \cup A', R_V \cup B')$ can also be enlarged by u. Thus, there is no maximal biclique that contains (R_U, R_V) but not u in $G(R_U \cup C_U, R_V \cup P_V)$. □

Armed with Lemma 3, we can always set the skipping set P_U as $C_U \setminus \{u\}$ if u is the pivot vertex in $C_U \cup X_U$. Then, the remaining issue is to determine the skipping set P_V, satisfying that each biclique (A, B) with $R_V \subseteq B \subseteq (R_V \cup P_V)$ can be enlarged by u. Note that to achieve this, a prerequisite is that P_V must be included in the neighborhood of u. Otherwise, the biclique (A, B) cannot be extended by u, since there is a vertex $v \in B$ with $v \notin N_u(G)$. Based on this, we can obtain the following result.

Theorem 4.1 (Biclique Pivoting Rule). *Let $u \in C_U \cup X_U$ be a pivot vertex. The vertices in (P_U, P_V) can be skipped to expand (R_U, R_V), where $P_U = C_U \setminus \{u\}$ and $P_V = C_V \cap N_u(G)$.*

Proof. Based on Lemma 3, we just need to prove that any biclique (A, B) with $R_U \subset A$ and $R_V \subseteq B \subseteq R_V \cup P_V$ can be enlarged by u. Since $P_V \subseteq N_u(G)$, then we have $(R_V \cup P_V) \subseteq N_u(G)$, indicating that $(A \cup \{u\}, B)$ is also a biclique of G. Thus, we proved. □

Example 2. Consider a bipartite graph G shown in Fig. 1. Let (R_U, R_V), (C_U, C_V), and (X_U, X_V) be the current biclique, the candidate sets, and the exclusion sets in a recursion, respectively. Assume that $(R_U, R_V) = (\emptyset, \{v_2\})$, $(C_U, C_V) = (\{u_1, u_2, u_3, u_4\}, \{v_1, v_3, v_4, v_5\})$, and $(X_U, X_V) = (\emptyset, \emptyset)$. If we select the vertex $u_4 \in C_U$ as the pivot vertex, then by Theorem 4.1 we can obtain that the skipping sets (P_U, P_V) are $(\{u_1, u_2, u_3\}, \{v_3, v_4, v_5\})$. As a result, it only needs to expand $(\emptyset, \{v_2\})$ with vertices in $(\{u_4\}, \{v_1\})$ in this recursion, which can dramatically reduce unnecessary computations.

Algorithm 3: Pivot-based maximal biclique enumeration

```
Input: The bipartite graph G
Output: All maximal bicliques of G
1  BicliqueEnum(∅, ∅, U, V, ∅, ∅);
2  Function: BicliqueEnum(R_U, R_V, C_U, C_V, X_U, X_V)
3      if C_U ∪ C_V = ∅ then
4          if X_V = ∅ and X_U = ∅ then
5              Output (R_U, R_V) as a result;
6          return;
7      Select a pivot vertex u from C_U ∪ X_U that maximizes |P_U| + |P_V|, where
         P_U = C_U \ {u} and P_V = C_V ∩ N_u(G);
8      Select a pivot vertex v from C_V ∪ X_V that maximizes |P'_U| + |P'_V|, where
         P'_U = C_U ∩ N_v(G) and P'_V = C_V \ {v};
9      if |P'_U| + |P'_V| > |P_U| + |P_V| then P_U ← P'_U; P_V ← P'_V;
10     foreach u' ∈ C_U \ P_U do
11         BicliqueBranch(R_U, R_V, u', C_U \ {u'}, C_V, X_U, X_V);
12         C_U ← C_U \ {u'}; X_U ← X_U ∪ {u'};
13     foreach v' ∈ C_V \ P_V do
14         BicliqueBranch(R_V, R_U, v', C_V \ {v'}, C_U, X_V, X_U);
15         C_V ← C_V \ {v'}; X_V ← X_V ∪ {v'};
16 Function: BicliqueBranch(R_U, R_V, u, C_U, C_V, X_U, X_V)
17     C'_V ← C_V ∩ N_u(G); X'_V ← X_V ∩ N_u(G);
18     BicliqueEnum(R_U ∪ {u}, R_V, C_U, C'_V, X_U, X'_V);
```

Algorithm 4: A polynomial-delay implementation of the pivot-based maximal biclique enumeration algorithm

```
Input: The bipartite graph G
Output: All maximal bicliques of G
1  BicliqueEnum(∅, ∅, U, V, ∅, ∅);
2  Function: BicliqueEnum(R_U, R_V, C_U, C_V, X_U, X_V)
3      if C_U ∪ C_V = ∅ then
4          return;
5      Lines 7-15 of Algorithm 3;
6  Function: BicliqueBranch(R_U, R_V, u, C_U, C_V, X_U, X_V)
7      if C_V ≠ ∅ then
8          C'_V ← C_V ∩ N_u(G); X'_V ← X_V ∩ N_u(G);
9          if N(R_U ∪ {u}) = R_V ∪ C'_V and N(R_V ∪ C'_V) = R_U ∪ {u} then
10             Output (R_U ∪ {u}, R_V ∪ C'_V) as a result;
11         C'_V = {w ∈ C'_V | N_w(G) ∩ C_U \ {u} ≠ ∅};
12         BicliqueEnum(R_U ∪ {u}, R_V, C_U, C'_V, X_U, X'_V);
```

Based on Algorithm 2 and the pivoting technique established in Theorem 4.1, we can easily devise our pivot-based algorithm for maximal biclique enumeration as outlined in Algorithm 3. Note that to achieve good pruning performance, Algorithm 3 always selects a pivot vertex u (v) from $C_U \cup X_U$ ($C_V \cup X_V$) such that $|P_U| + |P_V|$ is maximum (lines 7-8). By Theorem 3.3 and Theorem 4.1, it is easy to show that Algorithm 5 correctly enumerates all maximal bicliques of a bipartite graph. The following theorem shows the time complexity of Algorithm 3.

Theorem 4.2. *The worst-case time complexity of Algorithm 3 is $O(m \times 2^{n/2})$ (or $O(m \times 1.414^n)$).*

Proof. To obtain the worst-case time complexity of Algorithm 3, the key is to derive the maximum size of the enumeration tree generated by the algorithm. Denote by $T(n)$ the maximum number of recursive calls of *BicliqueEnum* $(\emptyset, \emptyset, U, V, \emptyset, \emptyset)$, where $n = |C_U| + |C_V|$. Suppose that $k = min\{\overline{d}_u(C_V), \overline{d}_v(C_U)\}$, where $u \in C_U \cup X_U$ ($v \in C_V \cup X_V$) and $\overline{d}_u(C_V) = C_V \setminus N_u(G)$ ($\overline{d}_v(C_U) = C_U \setminus N_v(G)$). Then, we analyze the recurrence relations of $T(n)$.

Assume that the optimal pivot vertex u is selected from $C_U \cup X_U$. We then have $k = \overline{d}_u(C_V)$ and $\overline{d}_v(C_U) \geq k$ for every $v \in C_V \setminus N_u(G)$.

Based on the pivoting technique (Theorem 4.1), only the vertex u and each vertex in $C_V \setminus N_u(G)$ is used to expand (R_U, R_V) in the recursions. When vertex u is pushed into (R_U, R_V), we notice that the size of the corresponding candidate sets is at most $n-k-1$, since u has k non-neighbors in C_V. Moreover, when pushing the i-th vertex in $C_V \setminus N_u(G)$ into (R_U, R_V), the size of its corresponding candidate sets is at most $n-k-i$ since $\overline{d}_v(C_U) \geq k$ for every $v \in C_V \setminus N_u(G)$. Thus, the following recurrence relations of $T(n)$ can be obtained:

$$T(n) \leq T(n-k-1) + \sum_{i=1}^{k} T(n-k-i), \text{ if } u \in C_U, \quad (1)$$

$$T(n) \leq \sum_{i=1}^{k} T(n-k-i), \text{ if } u \in X_U. \quad (2)$$

Clearly, the worst-case bound of Eq. (2) is smaller than that of Eq. (1). Thus, in the following, we mainly analyze the worst-case bound of Eq. (1) by considering four different cases.

Case $k = 0$: In this case, we can easily obtain

$$T(n) = T(n-1). \quad (3)$$

Obviously, the size of $T(n)$ is bounded by n for this case.

Case $k = 1$: By Eq. (1), we have

$$T(n) \leq T(n-1-1) + T(n-1-1) = 2T(n-2). \quad (4)$$

In this case, it is also easy to derive that $T(n)$ is bounded by $O(2^{\frac{n}{2}})$.

Case $k = 2$: In this case, we have

$$T(n) \leq T(n-2-1) + \sum_{i=1}^{2} T(n-2-i) = T(n-3) + T(n-3) + T(n-4). \quad (5)$$

As proved in [17], for any linear recursion function of $F(n) = \sum_{i=1}^{j} F(n-a_i)$, $F(n)$ is bounded by $O(\alpha^n)$, where α is the maximum real root of the equation $x^n - \sum_{i=1}^{j} x^{n-a_i} = 0$. By setting $a_1 = a_2 = 3$ and $a_3 = 4$, we can obtain that $T(n)$ is bounded by $O(\alpha^n)$. Here α is the maximum real root of the equation $x^{n-4}(x^4 - x - x - 1) = 0$ which can be easily shown that $\alpha < \sqrt{2}$. As a result, we have $T(n) \leq 2^{n/2}$.

Case $k \geq 3$: In this case, we have

$$\begin{aligned} T(n) &\leq T(n-k-1) + \sum_{i=1}^{k} T(n-k-1) \\ &= (k+1)T(n-k-1) \\ &= (k+1)^{\frac{n}{k+1}} T(n-(k+1)\times\frac{n}{k+1}). \end{aligned} \quad (6)$$

To bound $T(n)$, it is sufficient to derive a bound for $(k+1)^{\frac{n}{k+1}}$. Note that $f(x) = x^{\frac{n}{x}}$ is a monotonically decreasing function with $x \geq 4$. Since $k+1 \geq 4$, we have $(k+1)^{\frac{n}{k+1}} \geq 4^{\frac{n}{4}}$. As a result, we have $T(n) = O(4^{\frac{n}{4}}) = O(2^{\frac{n}{2}})$.

Putting it all together, the size of the enumeration tree of Algorithm 3 is bounded by $O(2^{n/2})$ (or $O(1.414^n)$). It is easy to show that Algorithm 3 takes at most $O(m)$ time in each recursion, thus the total time complexity of Algorithm 3 is bounded by $O(m \times 2^{n/2})$. □

Note that since there may exist $O(2^{n/2})$ maximal bicliques in a bipartite graph as shown in [38], the worst-case time complexity of any algorithm to enumerate all maximal bicliques must be no less than $O(\tau \times 2^{n/2})$ where τ is the average size of all maximal bicliques. As a result, the worst-case time complexity of Algorithm 3 is near optimal (only up to a factor m/τ).

A polynomial-delay implementation. We note that Algorithm 3 can also achieve polynomial delay time complexity with slight modifications. The key idea is to change the way the output results. Specifically, when a vertex $u \in C_U$ is added into the current partial biclique (R_U, R_V), we then check if the conditions $N(R_U \cup \{u\}) = R_V \cup C'_V$ and $N(R_V \cup C'_V) = R_U \cup \{u\}$ are satisfied. If so, the current biclique $(R_U \cup \{u\}, R_V \cup C'_V)$ is definitely maximal and can be output as a result. To avoid redundant results, some additional modifications are required, namely: 1) remove lines 4-5 of Algorithm 3; 2) $C_V \neq \emptyset$ should be checked before line 17; and 3) remove all vertices in C'_V that have no neighbors in $C_U \setminus \{u\}$. The detailed procedure is shown in Algorithm 4.

THEOREM 4.3. *Algorithm 4 is a polynomial delay algorithm for enumerating all maximal bicliques.*

PROOF. We first prove that the first result can be output in polynomial time. The reason is that when a vertex u (in C_U) is pushed into $(\emptyset, \emptyset)$, we can obtain that $(\{u\}, C'_V)$ is a biclique, where $C'_V = C_V \cap N_u(G)$. If there is another vertex w in C_U that can be used to expand $(\{u\}, C'_V)$, then this vertex w must be the pivot in the next sub-recursion (since $C'_V \subseteq N_w(G)$), which can be directly added into $(\{u\}, \emptyset)$. This procedure continues until a partial result $(R_U, \emptyset)$ that leaves no vertex in $C_U \setminus R_U$ that can be used to expand $(\{u\}, C'_V)$. This gives a maximal result (R_U, C'_V), which can obviously be done in polynomial time. Next, we show that when a result is output, the next result can be determined in polynomial time. In a recursion with the partial result $(R_U, \emptyset)$, the vertices in $C_U \setminus R_U$ and C'_V are further used to expand $(R_U, \emptyset)$. If a vertex u' in C_U is selected, the next maximal result to be detected must differ from (R_U, C'_V), since it contains $R_U \cup \{u'\}$. Moreover, based on the previously analysis, such a maximal biclique can also be obtained in polynomial time. If selecting $v' \in C'_V$ to expand $(R_U, \emptyset)$, we notice that v' is not the neighbor of u' (because of the pivoting technique), a new maximal result can be obtained that is different from the one obtained by expanding u'. In addition, since v' has at least a neighbor in $C_U \setminus R_U$, this means that the new result is also different from (R_U, C'_V). Thus, when a result is output, the next result can be identified in polynomial time. This completes the proof. □

4.3 Optimization Techniques

In this subsection, we propose several optimization techniques to further improve the efficiency of Algorithm 3. Below, we first propose an early termination technique to reduce the maximum depth of recursions, and then introduce an ordering technique to reduce the maximum size of the initial candidate sets.

Early termination. The early termination trick is based on the fact that if there is no edge in the subgraph $G(C_U, C_V)$, then there are at most two possible maximal bicliques $(R_U \cup C_U, R_V)$ and $(R_U, R_V \cup C_V)$ to be explored. This is because both $(R_V \cup C_V) \subseteq N(R_U)$ and $(R_U \cup C_U) \subseteq N(R_V)$ always hold; and if a vertex on one side of the candidate sets is used to expand (R_U, R_V), then the other side of the candidate sets will be empty.

We can slightly modify Algorithm 3 to implement such an early termination trick. Specifically, let $u \in C_U$ be the optimal pivot vertex selected by Algorithm 3 (lines 7-9). If $P_V = \emptyset$ which means that there is no edge in $G(C_U, C_V)$, the current recursive call can be terminated. In this case, we only need to determine whether $(R_U \cup C_U, R_V)$ is a maximal biclique by checking $X_U = \emptyset$ and $X_V \cap N(C_U) = \emptyset$; and determine whether $(R_U, R_V \cup C_V)$ is a maximal biclique by verifying $X_V = \emptyset$ and $X_U \cap N(C_V) = \emptyset$.

Ordering optimization. As indicated in [1, 10], the performance of maximal biclique enumeration algorithms is often sensitive to the vertex ordering. Inspired by this, we can also use the ordering technique to further improve the performance of our algorithm.

Given an ordered vertex set $O = \{u_1, u_2, ..., u_n\}$ of all vertices in $U \cup V$, we define $O^{\geq u_i}$ as the set of vertices in O with ranks higher than u_i; and $O^{<u_i}$ is defined similarly. Let $N^2_{u_i}(G)$ be the set of vertices in G whose distance from $u_i \in O$ is 2, i.e., $N^2_{u_i}(G) = \{u_j \in O | j \neq i, N_{u_i}(G) \cap N_{u_j}(G) \neq \emptyset\}$. Note that in bipratite graphs, $N^2_{u_i}(G)$ does not contain any neighbor of u_i. Given a vertex u_i in O, we denote by $G^+_{u_i}$ the subgraph of G induced by $O^{\geq u_i} \cap N^2_{u_i}(G)$ and $N_{u_i}(G)$. It is easy to show that for any maximal biclique (A, B) of G, there must exist a subgraph $G^+_{u_i}$ that contains all vertices in $(A \setminus \{u_i\}, B)$. Here the vertex u_i can be only contained in $O \cap U$, as the diameter of (A, B) is 2 and $G^+_{u_i}$ contains all neighbor of u_i and all vertices in $N^2_{u_i}(G)$ with ranks higher than u_i. Thus, to enumerate all maximal bicliques in G, it is sufficient to enumerate all maximal bicliques in each $G^+_{u_i}$ ($u_i \in O \cap U$). In this work, we make use of two ordering techniques, including the classic degeneracy ordering on graphs [16, 26] and the 2-hop degree ordering developed in [10]. Note that to obtain the degeneracy ordering, we can treat the bipartite graph as a traditional graph, and then iteratively remove the smallest-degree vertex from the bipartite graph to generate the degeneracy ordering. Such an iteratively-peeling procedure can be implemented in $O(m + n)$ time [3, 16].

4.4 Enumerating Large Maximal Bicliques

In practical applications, we may be more interested in finding large maximal bicliques with size no less than a given threshold, since small maximal bicliques typically have low practical value [1, 29]. More specifically, we aim to identify every maximal biclique (A, B) that satisfies $|A| \geq q$ and $|B| \geq q$, where q is a given threshold.

Note that our algorithm can be easily adapted to enumerate such size-constraint maximal bicliques. Moreover, we can also use the size-constraint to further prune unnecessary computations. Below, we first introduce the definition of (α, β)-core, which was originally proposed in [8, 28].

Definition 6 ([8]). Given a bipartite graph G, an (α, β)-core is a maximal subgraph $H = (U_H, V_H, E_H)$ of G such that each vertex u in U_H has a degree no less than α in H and each vertex v in V_H has a degree no less than β in H.

By Definition 6, it is easy to show that every maximal biclique (A, B) that satisfies $|A| \geq q$ and $|B| \geq q$ must be contained in the (q, q)-core of G. Thus, we can first compute the (q, q)-core of the bipartite graph, and then enumerate all size-constraint maximal bicliques in the (q, q)-core, instead of in the original bipartite graph. Such a (q, q)-core can significantly prune a large number of unpromising vertices. In addition, in our pivot-based algorithm, we need to compute the optimal pivot vertex (line 7 of Algorithm 3) which means that the degree of every vertex in $C_U \cup C_V$ must be computed. Based on this degree information, we can prune the vertices in C_U (C_V) if their degrees are less than $q - |R_V|$ ($q - |R_U|$) in $G(C_U, C_V)$, thus further improving the efficiency of our enumeration algorithm.

5 MAXIMAL k-BIPLEX ENUMERATION

In this section, we focus on the problem of enumerating all maximal k-biplexes of a given bipartite graph G, where a maximal k-biplex is another instance of maximal $\mathcal{P}$-subgraph. Such a maximal k-biplex enumeration problem has also been extensively studied in recent years [42, 48, 49]. To the best of our knowledge, there are two state-of-the-art solutions for solving this problem. The first one is developed in [49], which proposes a prefix tree based recursive algorithm to detect whether each possible combination of vertices of G can form a maximal k-biplex. Such an enumeration method may explore all the possible combinations of vertices, thus it cannot handle large bipartite graphs. The second one is a polynomial-delay algorithm proposed in [48], which is based on the reverse search framework [11]. Such an algorithm, however, needs to store all detected maximal k-biplexes in the main memory to guide the reverse search procedure, which prohibits it to handle large bipartite graphs. To overcome the limitations of existing solutions, we propose a novel algorithm to enumerate all maximal k-biplexes based on our general pivot-based enumeration framework.

5.1 Pivot-based Maximal k-Biplex Enumeration

To use our general pivot-based enumeration framework for maximal k-biplex enumeration, the key is to derive the skipping sets (P_U, P_V) based on the pivot vertex as stated in Theorem 3.1. Suppose that the pivot vertex u is selected from $C_U \cup X_U$. Given a vertex $v \in U$, we denote by $\overline{N}_v(B)$ the set of non-neighbors of v in B and define $\overline{d}_v(B)$ as $|\overline{N}_v(B)|$. Below, we detail our solutions.

Constructing the skipping set P_V. We find that the skipping set $P_V \subseteq C_V$ can be easily derived, as shown in Lemma 4.

Lemma 4. Let $u \in C_U \cup X_U$ be a pivot vertex in a recursion to enumerate all maximal k-biplexes that contain (R_U, R_V). Then, the skipping set P_V can be set as $C_V \cap N_u(G)$.

PROOF. Since $P_V = C_V \cap N_u(G)$, it is easy to see that the number of non-neighbors of u in $R_V \cup P_V$ is no larger than k, i.e, $\overline{d}_u(R_V \cup P_V) \leq k$. Moreover, for each $v \in \overline{N}_u(R_V)$, we also have $\overline{d}_v(R_U) < k$ since the vertex u can be used to expand (R_U, R_V). Let (R_U, B) with $R_V \subseteq B \subseteq (R_V \cup P_V)$ be an arbitrary k-biplex of G. Then, it can be seen that $(R_U \cup \{u\}, B)$ is also a k-biplex of G. The reason are as follows. When pushing u into (R_U, B), there always have $\overline{d}_{u'}(B) \leq k$ for each $u' \in R_U \cup \{u\}$, since (R_U, B) is a k-biplex and $\overline{d}_u(R_V \cup P_V) \leq k$. Then, for each vertex v in B, it clearly holds for $\overline{d}_v(R_U \cup \{u\}) \leq k$ if $v \in B \cap N_u(G)$. When the vertex v belongs to $B \setminus N_u(G)$, we also have $\overline{d}_v(R_U \cup \{u\}) \leq k$ since $\overline{d}_v(R_U) < k$. Thus, the vertex u can be used to enlarge the k-biplex (R_U, B), and this lemma is established by Theorem 3.1. □

Constructing the skipping set P_U. We note that constructing the skipping set P_U for maximal k-biplex enumeration is quite nontrivial. Similar ideas to construct P_U for maximal biclique enumeration cannot be used for maximal k-biplex enumeration. Specifically, for maximal biclique enumeration, only the pivot vertex u (if $u \in C_U$) and a subset of vertices on the opposite side (i.e., C_V) are considered to be used to expand the partial biclique. Such a nice property, however, no longer holds for maximal k-biplex enumeration. For instance, consider a bipartite graph G shown in Fig. 1. Let $(R_U, R_V) = (\{u_1, u_4\}, \{v_1\})$ and $(C_U, C_V) = (\{u_2, u_3\}, \{v_2, v_3, v_4, v_5\})$ be the current k-biplex and the candidate sets used to expand (R_U, R_V), respectively. Suppose that $k = 1$ and $v_3 \in C_V$ is selected as the pivot vertex. Then, it is easy to see that $(\{u_1, u_4\}, \{v_1, v_2, v_5\})$ is a maximal 1-biplex of G. Thus, when selecting a pivot vertex v from $C_V \cup X_V$, there still exist some maximal k-biplexes containing (R_U, R_V) in the subgraph of G induced by R_U and $R_V \cup C_V \setminus \{v\}$. This example suggests that the skipping set P_U cannot be simply set to $C_U \setminus \{u\}$ if the pivot vertex u is selected from $C_U \cup X_U$.

For any maximal k-biplex (A, B) containing (R_U, R_V) in the subgraph of G that is induced by $R_U \cup C_U \setminus \{u\}$ and R_V, we notice that there exists an approach to determine whether it is also a maximal k-biplex in G. Specifically, if (A, B) is also maximal in G, there must exist a *conflict* vertex in $\overline{N}_u(R_V)$ that prohibits the pivot vertex u to enlarge (A, B). Based on this analysis, we can derive the following lemma, which is used to compute maximal k-biplexes of G in the subgraph $G(R_U \cup C_U \setminus \{u\}, R_V \cup P_V)$.

Lemma 5. Given a pivot vertex $u \in C_U \cup X_U$ and a skipping set $P_V \subseteq C_V$, if a k-biplex (A, B) that contains (R_U, R_V) but not u and $C_V \setminus P_V$ is maximal in G (the maximal k-biplex of G belongs to the case (iii) in Theorem 3.2), there must exist a vertex $v' \in R_V \setminus N_u(G)$ with $\overline{d}_{v'}(A) = k$. Otherwise, (A, B) is definitely not maximal in G.

Proof. Based on Definition 4, if (A, B) is the maximal in G, then there is no any vertex in $U \setminus A$ and $V \setminus B$ that can be used to enlarge (A, B). Thus, the pivot vertex u can not be added into (A, B) to form a larger one. Since $B \subseteq (R_V \cup P_V)$ and $P_V \subseteq N_u(G)$ (Lemma 4), we then have $\overline{d}_u(B) \leq k$. However, based on the fact that u cannot be used to enlarge (A, B), the only case is that there is a vertex $v' \in R_V \setminus N_u(G)$ that prohibits the pivot vertex u to enlarge (A, B). Thus, we have $\exists v' \in R_V \setminus N_u(G)$ with $\overline{d}_{v'}(A) = k$. Otherwise, (A, B) must be expandable with u. □

By Lemma 5, the problem of generating the skipping set P_U is equivalent to find a subset $P_U \subseteq C_U$ such that for any k-biplex (A, B) containing (R_U, R_V) in $G(R_U \cup P_U, R_V \cup P_V)$, $\overline{d}_{v'}(A) < k$ holds for each $v' \in R_V \setminus N_u(G)$. Then, the following lemma shows how to derive the skipping set P_U.

Lemma 6. Given a pivot vertex $u \in C_U \cup X_U$ and the skipping set $P_V \subseteq C_V$ in a recursion to enumerate all maximal k-biplexes that contain (R_U, R_V), then the skipping set P_U can be set as $\{w \in C_U | \overline{N}_u(R_V) \subseteq N_w(G), w \neq u\}$.

Proof. Since the pivot vertex u can be used to expand (R_U, R_V), we can easily derive that for each $v' \in R_V \setminus N_u(G)$, we have $\overline{d}_{v'}(R_U) < k$. Moreover, since $P_U = \{w \in C_U | \overline{N}_u(R_V) \subseteq N_w(G), w \neq u\}$, we have $\overline{d}_{v'}(P_U) = 0$ for each $v' \in R_V \setminus N_u(G)$. This means that given any k-biplex (A, B) containing (R_U, R_V) in $G(R_U \cup P_U, R_V \cup P_V)$, there always have $\overline{d}_{v'}(A) < k$ for each $v' \in R_V \setminus N_u(G)$. Thus, this lemma is proved. □

Algorithm 5: The pivot-based maximal k-biplex algorithm

```
Input: The bipartite graph G and a parameter k
Output: All maximal k-biplexes of G
1  BiplexEnum(∅, ∅, U, V, ∅, ∅);
2  Function: BiplexEnum(R_U, R_V, C_U, C_V, X_U, X_V)
3      if C_U ∪ C_V = ∅ then
4          if X_V ∪ X_U = ∅ then Output (R_U, R_V) as a result;
5          return;
6      Select a pivot u from C_U ∪ X_U and obtain the skipping sets
         (P_U = {w ∈ C_U | w ≠ u, N̄_u(R_V) ⊆ N_w(G)}, P_V = C_V ∩ N_u(G));
7      Select a pivot v from C_V ∪ X_V and obtain the skipping sets
         (P'_U = C_U ∩ N_v(G), P'_V = {w ∈ C_V | w ≠ v, N̄_v(R_U) ⊆ N_w(G)});
8      if |P'_U| + |P'_V| > |P_U| + |P_V| then P_U ← P'_U; P_V ← P'_V;
9      foreach w ∈ C_U \ P_U do
10         BiplexBranch(R_U, R_V, w, C_U \ {w}, C_V, X_U, X_V);
11         C_U ← C_U \ {w}; X_U ← X_U ∪ {w};
12     foreach w ∈ C_V \ P_V do
13         BiplexBranch(R_V, R_U, w, C_V \ {w}, C_U, X_V, X_U);
14         C_V ← C_V \ {w}; X_V ← X_V ∪ {w};
15 Function: BiplexBranch(R_U, R_V, u, C_U, C_V, X_U, X_V)
16     (C'_U, C'_V) ← BiPlexUpdates(R_U, R_V, u, C_U, C_V);
17     (X'_U, X'_V) ← BiPlexUpdates(R_U, R_V, u, X_U, X_V);
18     BiplexEnum(R_U ∪ {u}, R_V, C'_U, C'_V, X'_U, X'_V);
```

Algorithm 6: $BiPlexUpdates(R_U, R_V, u, C_U, C_V)$

```
1  C'_U ← C_U; C'_V ← C_V ∩ N_u(G);
2  foreach v ∈ C_V \ N_u(G) do
3      if d̄_v(R_U ∪ {u}) ≤ s then C'_V ← C'_V ∪ {v};
4  foreach v ∈ R_V \ N_u(G) do
5      if d̄_v(R_U ∪ {u}) = s then C'_U ← C'_U ∩ N_v(G);
6  return (C'_U, C'_V);
```

Armed with Lemma 4 and Lemma 6, we can obtain the following pivoting rule for maximal k-biplex enumeration.

Theorem 5.1 (Biplex Pivoting Rule). *Let $u \in C_U \cup X_U$ be a pivot vertex in a recursion. The vertices in (P_U, P_V) can be skipped to expand (R_U, R_V), where $P_U = \{w \in C_U | \overline{N}_u(R_V) \subseteq N_w(G), w \neq u\}$ and $P_V = C_V \cap N_u(G)$.*

Proof. The theorem is clearly established based on Lemma 4 and Lemma 6. □

Implementation details. With Theorem 5.1, we can easily implement our pivot-based maximal k-biplex enumeration algorithm which is outlined in Algorithm 5. In Algorithm 5, the sets (R_U, R_V), (C_U, C_V), and (X_U, X_V) are the current k-biplex of G, the candidate sets, and the exclusion sets, respectively. Before performing the sub-recursive calls in each recursion, the algorithm needs to make use of Theorem 5.1 to generate the optimal skipping sets (P_U, P_V) (lines 6-8). Then, each vertex contained in (C_U, C_V) but excluded in (P_U, P_V) is used to expand (R_U, R_V) (lines 9-14). If both the candidate sets and the exclusion sets are empty, the current k-biplex (R_U, R_V) is maximal in G (lines 3-4).

Note that when a vertex is used to expand the current k-biplex, Algorithm 5 invokes Algorithm 6 to update the candidate and exclusion sets (lines 16-17 of Algorithm 5). Specifically, when a vertex $u \in C_U$ is pushed into (R_U, R_V), every vertex in C_V should keep the

invariance that it has at most k non-neighbors in $R_U \cup \{u\}$ (lines 1-3 of Algorithm 6). If there is a subset S in $R_V \setminus N_u(G)$ that has k non-neighbors in $R_U \cup \{u\}$, then every vertex in $C_U \setminus \{u\}$ must be the common neighbors of S, i.e., $C_U \setminus \{u\} \subseteq N(S)$ (lines 4-5 of Algorithm 6). Note that the solution for updating the exclusion sets (X_U, X_V) is similar, thus we omit the details. The following theorem shows that such an updating algorithm takes $O(kn)$ time.

THEOREM 5.2. *The time complexity of Algorithm 6 is $O(kn)$.*

PROOF. When using a set to maintain the value of $\overline{d}_u(R_V)$ (or $\overline{d}_v(R_U)$) for each $u \in C_U$ (or $v \in C_V$) in the recursions, the time taken by lines 2-3 of Algorithm 6 is bounded by $O(|C_V|)$. Since $|R_V \setminus N_u(G)|$ is at most k, the intersection operation on line 5 is executed at most k times. Thus, the time complexity of Algorithm 6 is $O(|C_V| + k|C_U|)$, which is bounded by $O(kn)$. □

By Theorem 5.1 and Theorem 3.3, we can conclude that Algorithm 5 correctly enumerates all maximal k-biplex of G. The worst-case time complexity of Algorithm 5 is $O(n^2 2^n)$, as there are at most 2^n enumeration branches. Note that although Algorithm 5 has an $O(n^2 2^n)$ time complexity in the worst case, it works well for real-world bipartite graphs (as evidenced in our experiments) due to the proposed powerful pivot-based pruning technique. Moreover, similar to maximal biclique enumeration, we can also use the ordering techniques, such as the degeneracy ordering [16, 26] and the 2-hop degree ordering [10]), to further improve the efficiency of Algorithm 5.

Remark. We note that it seems difficult to derive a tighter time complexity for Algorithm 5. This is because the size of the skipping set P_U is very hard to bound; and the size of candidate sets cannot be reduced if u is used to expand the current partial k-biplex, because all non-neighbors of u can still be included in the candidate sets. We leave the problem of deriving a tighter time complexity for Algorithm 5 as an interesting open question.

5.2 Enumerating Large Maximal k-Biplexes

In this subsection section, we focus on the problem of enumerating large maximal k-biplexes (i.e., enumerating all maximal k-biplexes with size no less than a given threshold), as small maximal k-biplexes are often no practical use in real-world applications. More importantly, we prove that large maximal k-biplexes have small diameters, thus they are often more cohesive compared to the small maximal k-biplexes.

Lemma 7. For any maximal k-biplex (A, B) of G, the diameter of $G(A, B)$ is no larger than 4 if $|A| \geq 2k + 1$ and $|B| \geq k + 1$ (or if $|A| \geq k + 1$ and $|B| \geq 2k + 1$).

PROOF. Let $dis(u, v)$ be the distance between u and v. In subgraph $G(A, B)$, for each pair of vertices v_1 and v_2 in B, we always have $N_{v_1}(A) \cap N_{v_2}(A) \neq \emptyset$ since $|A| \geq 2k + 1$, which we obtain that $dis(v_1, v_2) = 2$. Then, for each $u \in A$, we also have $\exists v' \in B$ satisfying $v' \in N_u(A)$ since $|B| \geq k + 1$. Thus, for any vertex $u \in A$ and $v \in B$, we have $dis(u, v) \leq 3$ due to the fact that $\exists v' \in B$ with $v' \in N_u(A)$ and $dis(v', v) = 2$. Finally, for any two vertices u_1 and u_2 in A, there must exist two vertices $v'_1 \in B$ and $v'_2 \in B$ with $v'_1 \in N_{u_1}(G)$ and $v'_2 \in N_{u_2}(G)$. Since $dis(v'_1, v'_2) = 2$, we can easily derive that $dis(u_1, u_2) \leq 4$. This completes the proof. □

Algorithm 7: Size-constraint maximal k-biplex algorithm

Input: The bipartite graph G and two parameters k and q
Output: All size constraint maximal k-biplexes of G

```
1 H ← reduce the graph G by the (q − k, q − k)-core;
2 O ← ordering all vertices in U_H ∪ V_H;
3 foreach u_i ∈ O, s.t. u_i ∈ U do
4     C_U ← O^{>u_i} ∩ N^2_{u_i}(H); C_V ← ∪_{u∈C_U∪{u_i}} N_u(H);
5     X_U ← O^{<u_i} ∩ N^2_{u_i}(H); X_V ← ∅;
6     Reduce (C_U, C_V) and (X_U, X_V) by Lemma 10;
7     BiplexEnum({u_i}, ∅, C_U, C_V, X_U, X_V);
```

Lemma 8. For any maximal k-biplex (A, B) of G, the diameter of $G(A, B)$ is no larger than 3 if $|A| \geq 2k + 1$ and $|B| \geq 2k + 1$.

PROOF. With the condition of $|A| \geq 2k + 1$ and $|B| \geq 2k + 1$, it is easy to see that for each pair of vertices v_1 and v_2 in B (u_1 and u_2 in A), the distance between v_1 and v_2 (u_1 and u_2) in $G(A, B)$ is 2. Then, for each $u \in A$, there also exists a vertex $v' \in B$ with $v' \in N_u(G)$, which implies that for each pair of $u \in A$ and $v \in B$, the distance between u and v is at most 3. Thus, the lemma is established. □

According to Lemma 8, we can see that every maximal k-biplex (A, B) of G with $|A| \geq 2k + 1$ and $|B| \geq 2k + 1$, $G(A, B)$ must be densely connected since its diameter is no larger than 3. Note that the constraints of $|A| \geq 2k+1$ and $|B| \geq 2k+1$ are relatively mild, since k is often not very large (e.g., $k \leq 5$). In the following, we focus mainly on the problem of enumerating all maximal k-biplexes of G whose size on each side is no less than $q \geq 2k + 1$, i.e., output each maximal k-biplex (A, B) of G with $|A| \geq q$, $|B| \geq q$, and $q \geq 2k + 1$.

Pruning techniques. Similar to the size-constraint maximal biclique enumeration problem, we can also make use of the (α, β)-core to prune unpromising vertices for size-constraint maximal k-biplex enumeration.

For each k-biplex (A, B) of G with $|A| \geq q$ and $|B| \geq q$, it is easy to see that the smallest degree of vertices in $G(A, B)$ is at least $q - k$, which means that every k-biplex must be contained in the $(q - k, q - k)$-core.

Lemma 9. Any maximal k-biplex (A, B) of bipartite graph G with $|A| \geq q$ and $|B| \geq q$ must be included in the $(q - k, q - k)$-core of G.

PROOF. It is clearly established since $G(A, B)$ is a $(q - k, q - k)$-core of G. □

By Lemma 9, all vertices that are not contained in the $(q-k, q-k)$-core can be safely pruned from G when enumerating every maximal k-biplexe (A, B) with $|A| \geq q$ and $|B| \geq q$. In addition to the (α, β)-core pruning, we also derive a more effective pruning rule to further reduce the unpromising vertices.

Lemma 10. Given any k-biplex (A, B) of G with $|A| \geq q$ and $|B| \geq q$, for each pair of vertices u_1 and u_2 in A and each pair of vertices v_1 and v_2 in B, we have $|N_{u_1}(B) \cap N_{u_2}(B)| \geq q - 2k$ and $|N_{v_1}(A) \cap N_{v_2}(A)| \geq q - 2k$.

PROOF. Based on the condition that each vertex in $G(A, B)$ always has a degree no less than $q - k$, then, for each pair of vertices u_1 and u_2 in A, we have $\overline{d}_{u_1}(B) + \overline{d}_{u_2}(B) \leq 2k$. This indicates that the number of common neighbors of u_1 and u_2 in B is at least $q - 2k$.

Table 1: Real-world graph datasets.

Datasets	$\|U\|$	$\|V\|$	$\|E\|$	$d_{1\max}$	$d_{2\max}$
Crime	829	551	1,476	25	18
Ucforum	899	522	7,089	99	126
Fjwiki	612	1,922	12,382	360	48
Escorts	16,730	6,624	50,632	125	305
YouTube	94,238	30,087	293,360	1,035	7,591
BkCrossing	105,278	340,523	1,149,739	13,601	2,502
Cite	22,715	731,769	2,411,819	189,292	1,264
Dbtropes	64,415	87,678	3,232,134	6507	12400
IMDB	303,617	896,302	3,782,463	1,334	1,590
Twitter	175,214	530,418	4,664,605	968	19,805
DBLP	1,953,085	5,624,219	12,282,059	1,386	287

Table 2: Runtime of various algorithms for enumerating all maximal bicliques on real-world large graphs (in seconds).

Datasets	#Nums	oMBEA	BCEA	BCEAD	BCEAH
YouTube	1.82×10^6	74.59	17.44	7.77	**5.89**
BkCrossing	5.45×10^7	4739.6	314.63	**307.29**	310.85
Cite	1.10×10^7	414.83	265.67	95.21	**91.53**
Dbtropes	1.96×10^{10}	>20d	116710.13	88851.47	**85111.11**
IMDB	5.16×10^6	111.04	**51.15**	58.59	54.81
Twitter	5.54×10^6	135.73	424.21	240.87	**107.89**
DBLP	4.90×10^6	26.73	15.83	**13.52**	14.7

For the vertices in B, the results are consistent. Thus, this lemma is proved. □

When enumerating the size-constraint maximal k-biplexes that contain a specific vertex $u \in U$, we can further remove the vertices in the candidate sets and exclusion sets that conflict with u according to the condition shown in Lemma 10.

Size-constraint maximal k-biplex enumeration. Algorithm 7 outlines the size-constraint maximal k-biplex enumeration algorithm which is integrated with the pruning rules developed in Lemma 9 and Lemma 10. Specifically, Algorithm 7 first applies the (α, β)-core reduction technique (Lemma 9) to remove the unnecessary vertices in G (line 1). Then, the algorithm uses the ordering technique (degeneracy ordering or 2-hop degree ordering) to enumerate all maximal k-biplexes in the remaining graph (lines 2-7). With a specific ordered vertex set O, each vertex $u_i \in O$ with $u_i \in U$ is selected to enumerate all size-constraint maximal k-biplexes containing u_i. Note that the corresponding candidate sets (C_U, C_V) and exclusion sets (X_U, X_V) are initialized with vertices whose distance from u_i is no larger than 3 by Lemma 8 (lines 4-5). After that, the algorithm leverages Lemma 10 to further reduce the size of (C_U, C_V) and (X_U, X_V) (line 6). Finally, the pivot-based recursive procedure is invoked to enumerate all the maximal k-biplexes (line 7).

6 EXPERIMENTS

6.1 Experimental Setup

Different algorithms. We implement three pivot-based maximal biclique enumeration algorithms: BCEA, BCEAD, and BCEAH. Here BCEA is Algorithm 3; BCEAD and BCEAH denote Algorithm 3 with the degeneracy ordering and 2-hop degree ordering optimization respectively. Note that all BCEA, BCEAD, and BCEAH are equipped with the early termination trick. To enumerate size-constraint maximal bicliques, all pruning techniques developed in Section 4.4 are used for BCEA, BCEAD, and BCEAH. We compare our algorithms with the state-of-the-art maximal biclique enumeration algorithm developed in [10], namely oMBEA. Note that there are several other maximal biclique enumeration algorithms [1, 15, 50], but all of them are much worse than oMBEA [10]. Therefore, we only use oMBEA as the baseline in our experiments.

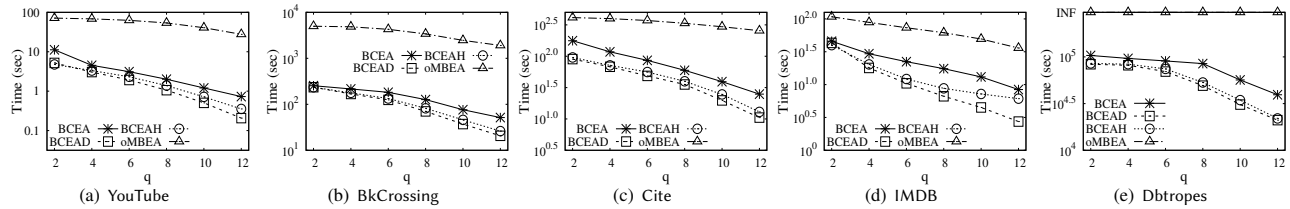
We also implement three pivot-based maximal k-biplex algorithms, called BPEA, BPEAD, and BPEAH, where BPEA is Algorithm 5, BCEAD and BCEAH denote Algorithm 5 with the degeneracy ordering and 2-hop degree ordering optimization respectively. Note that to enumerate size-constraint maximal k-biplexes, all pruning techniques proposed in Section 5.2 are used for BPEA, BPEAD, and BPEAH. We compare our algorithms with two state-of-the-art maximal k-biplex algorithms, called iTrav [48] and iMB [49]. All the algorithms, including our algorithms and baselines, are implemented in C++, and tested on a PC with one 2.2 GHz CPU and 64GB memory running CentOS.

Datasets. In our experiments, we use 11 real-world bipartite graphs for performance evaluation. The detailed statistics of datasets are summarized in Table 1, where the columns $d_{1\max}$ and $d_{2\max}$ are the maximum degree of vertices in U and V, respectively. Note that the first 4 small real-world bipartite graphs in Table 1 are used to evaluate the performance of different algorithms for enumerating all maximal k-biplexes. Since the number of small-size maximal k-biplexes is often very large, most of algorithms cannot handle large graphs when enumerating all maximal k-biplexes. The remaining 7 large real-world bipartite graphs in Table 1 are used to test the performance of algorithms for enumerating all (and size-constraint) maximal bicliques and the size-constraint maximal k-biplexes. All datasets are downloaded from (http://www.konect.cc/networks).

Parameters. When enumerating size-constraint maximal bicliques, the integer threshold parameter q is selected from 2 to 12. In enumerating maximal k-biplexes, there are two parameters k and q (the size-constraint threshold). We choose k from 1 to 4, with a default value of 1; and we select q from 10 to 20 for all datasets except DBLP (for which q is selected from 6 to 10, because DBLP does not contain very large k-biplexes).

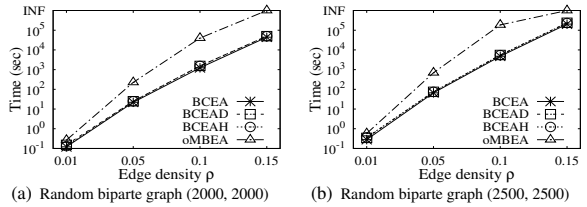
6.2 Results of Maximal Biclique Enumeration

Exp-1: Results of enumerating all maximal cliques. Table 2 shows the runtime of oMBEA, BCEA, BCEAD, and BCEAH for enumerating all maximal bicliques on different large real-world bipartite graphs, where column #Nums denotes the total number of maximal bicliques on each bipartite graph. From Table 2, we can see that the proposed pivot-based algorithm BCEAH consistently outperforms the state-of-the-art algorithm oMBEA. On most datasets, our algorithms can achieve one order of magnitude faster than oMBEA. For instance, on BkCrossing, our algorithms BCEAD and BCEAH take 307.29 and 310.85 seconds, respectively, while oMBEA consumes 4739.6 seconds. These results demonstrate the great superiority of the proposed pivot-based algorithms in reducing unnecessary computations. In addition, when comparing our algorithms BCEA, BCEAD, and BCEAH (the only difference among them is the ordering technique used), we notice that there is a relatively-small gap in time spent on most datasets, which indicates that the proposed



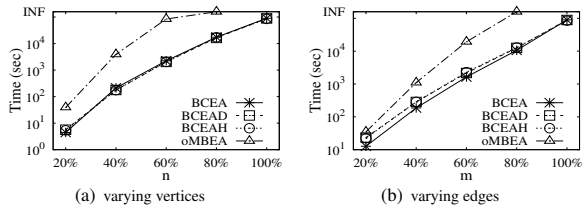
(a) YouTube (b) BkCrossing (c) Cite (d) IMDB (e) Dbtropes

Figure 2: Runtime of various algorithms for enumerating large maximal bicliques on large bipartite graphs.



(a) Random biparte graph (2000, 2000) (b) Random biparte graph (2500, 2500)

Figure 3: Runtime of maximal biclique enumeration algorithms on the bipartite graphs with varying densities.



(a) varying vertices (b) varying edges

Figure 4: Scalability for maximal biclique enumeration.

pivot-based techniques play the crucial role in reducing unnecessary computations.

Exp-2: Results of enumerating large maximal bicliques. Note that oMBEA was originally designed to enumerate all maximal bicliques. For a fair comparison, all the pruning techniques developed in Section 4.4 for enumerating size-constraint maximal bicliques are also integrated in oMBEA. Fig. 2 shows the runtime of different algorithms on all large bipartite graphs with varying q, where "INF" denotes that the algorithm cannot terminate within 24 hours. From Fig. 2, we observe that the performance of our pivot-based algorithm BCEAD consistently outperforms that of oMBEA with varying q. Moreover, with the increase of the parameter q, the speedup rate of BCEAD compared to oMBEA increases dramatically. The reason behind this may be that as q increases, the pruning technique (i.e., (α, β)-core) can produce a denser induced subgraph. This would decrease the effectiveness of the neighborhood domination based pruning techniques used in oMBEA. In addition, we can see that BCEAD is slightly better than BCEAH; and BCEA (our pivot-based algorithm without using ordering optimization) still significantly outperforms oMBEA. This results further demonstrate the high efficiency of the proposed pivot-based techniques compared to the state-of-the-art algorithm.

Exp-3: The effect of the density of bipartite graphs. Here we study the performance of different algorithms on the bipartite graphs with varying densities. To this end, we generate a set of random bipartite graphs with different densities. Specifically, each random bipartite graph is generated by first setting a fixed number of vertices on both sides and then randomly selecting an edge set with respect to the edge density $\rho = \frac{|E|}{|U|*|V|}$ among all vertex pairs. For a certain edge density, we generate 10 random bipartite graphs and report the average runtime of each algorithm on these random bipartite graphs. Fig. 3 shows the results. As can be seen, our algorithms are consistently faster than oMBEA. Moreover, the speedup ratio of our algorithms over oMBEA increases as the edge density ρ increases. For instance, in Fig. 3(a), BCEAD is 9.1 times faster than oMBEA when $\rho = 0.05$, and the speedup ratio further increases to 25.9 times when ρ increases to 0.1. This result further confirms that the performance of our algorithms are much better than the state-of-the-art algorithm even when the bipartite graph is dense. Note that our pivot-based algorithm is insensitive to different orderings on random bipartite graphs. This is because the degree distribution of vertices is uniform for these bipartite graphs.

Exp-4: Scalability testing. To evaluate the scalability of the proposed algorithms, we randomly sample 20-80% vertices and edges from the dataset Dbtropes to produce 8 subgraphs with different scales (similar results on the other datasets can also be observed). Fig. 4 shows the runtime of BCEA, BCEAD, BCEAH, and oMBEA on each sampled subgraph. As can be seen, the runtime of our algorithms increases smoothly as the size of the graph increases; and the state-of-the-art oMBEA always shows the worst performance with all parameter settings. Moreover, as the number of vertices or edges increases, the runtime of oMBEA also increases much faster compared to our pivot-based algorithms BCEA, BCEAD, and BCEAH. This result suggests that our algorithms exhibit good scalability in processing large bipartite graphs.

6.3 Efficiency of k-Biplex Enumerations

Exp-5: Results of enumerating all maximal k-biplexes. In this experiment, we test the performance of various algorithms for enumerating all maximal k-biplexes. Table 3 shows the runtime of BPEA, BPEAD, BPEAH, iMB, and iTrav on 4 small bipartite graphs (first 4 datasets in Table 1) with $k = 1$. Note that enumerating all maximal k-biplexes on large bipartite graphs is very costly for all algorithms, so we only use large bipartite graphs (the remaining datasets in Table 1) to evaluate the performance of different algorithms for enumerating size-constraint maximal k-biplexes. As can be seen, the algorithms we developed, BPEA, BPEAD, and BPEAH, are several orders of magnitude faster than the state-of-the-art algorithms iTrav and iMB. For instance, on Ucforum, iTrav and iMB take 16763.21 and 18899.93 seconds to enumerate all maximal 1-biplexes respectively, however, our algorithms consume at most 36.37 seconds. This result indicates that the proposed pivoting technique is very

Table 3: Runtime of various algorithms for enumerating all maximal k-biplexes on small datasets with $k = 1$ (in seconds).

Datasets	#Nums	iTrav	iMB	BPEA	BPEAD	BPEAH
Crime	1.52×10^6	167.784	47.60	4.09	4.13	**3.90**
Ucforum	3.15×10^7	16763.21	18899.93	36.37	**35.11**	35.17
Fjwiki	3.92×10^7	71161.72	79687.91	86.78	**69.29**	78.70
Escorts	8.50×10^8	INF	INF	19098.19	**15816.21**	16266.08

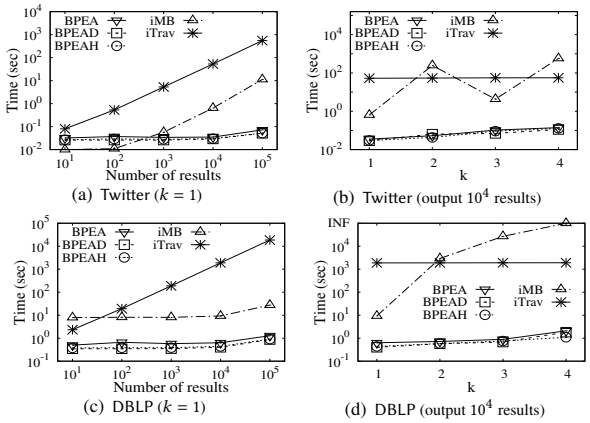


Figure 5: Delay testing for various algorithms.

effective to prune unnecessary branches in enumerating all maximal k-biplexes. Moreover, we observe that BPEAD is usually faster than BPEAH. The reason may be that the maximal k-biplex is no longer included in the subgraph induced by the 2-hop neighbors of the vertices, thus reducing the pruning performance of the 2-hop degree ordering.

Exp-6: Delay testing. We also compare the delay times of different algorithms with iTrav, where iTrav is a polynomial delay algorithm using a reverse search technique. Fig. 5 shows the runtime of each algorithm on Twitter and DBLP with varying k and the number of returned maximal k-biplexes. The results on other datasets are consistent. As can be seen, our pivot-based algorithms, BPEA, BPEAD, and BPEAH shows better delay results compared to iMB and iTrav, although our algorithms do not guarantee a polynomial delay property in theory. This result indicates that our pivot-based algorithms can achieve very good delay performance in practice. Moreover, the runtime of iTrav usually increases dramatically as the number of returned k-biplexes increases or the value of k increases. For instance, on Twitter, if $k = 1$, iTrav takes 0.642 seconds and 11.174 seconds to return 10^4 and 10^5 results, respectively. However, with the same parameter settings, BPEAD only takes 0.029 and 0.051 seconds to finish the computation respectively, which further confirms the efficiency of the proposed pivot-based algorithms.

Exp-7: Results of enumerating large maximal k-biplexes. In this experiment, we evaluate the performance of various algorithms for enumerating size-constraint maximal k-biplexes on real-world large graphs. Fig. 6 and Fig. 7 show the runtime of BPEA, BPEAD, BPEAH, iMB, and iTrav on real-world large graphs with varying q and k. Note that Fig. 7 only shows the results on Twitter and DBLP, and similar results can also be observed on the other datasets. As can be seen, iTrav cannot finish the computation with all parameter settings for all large datasets. This is because iTrav is mainly based on a reverse search technique which is very hard to handle large graphs. Our algorithms BPEA, BPEAD, and BPEAH also consistently outperform iMB on all datasets. Moreover, under most parameter settings, BPEAD can achieve more than three orders of magnitude faster than iMB. For example, on Cite, when $q = 20$, our algorithm BPEAD takes only 0.58 seconds, while iMB consumes 5,228 seconds. These results further demonstrate the high efficiency of the proposed pivot-based enumeration algorithms.

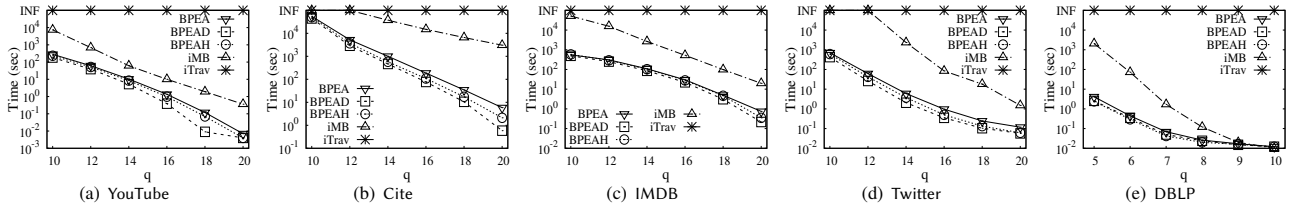
Exp-8: The effect of the bipartite graph density. Here we study the effect of the bipartite graph density on the performance of different algorithms. Fig. 8 shows the runtime of BPEA, BPEAD, BPEAH, iMB, and iTrav on random bipartite graphs (generated by the same method as used in Exp-3) with varying density ρ. The result shows that our algorithms are much faster than the baselines. Moreover, as the density ρ of the bipartite graph increases, the speedup ratio of our algorithms over the baseline algorithms grows dramatically. The reason is that if there exist large maximal k-biplexes in a bipartite graph, existing algorithms often involve many redundant computations to explore the non-maximal results contained in those large maximal k-biplexes. However, our pivot-based techniques can dramatically reduce such unnecessary computations (as shown in Theorem 5.1).

Exp-9: Scalability testing. Here we evaluate the scalability of BPEA, BPEAD, and BPEAH. To this end, we generate 8 subgraphs by randomly sampling 20-80% of vertices or edges on Twitter. Then, we run our algorithms on these subgraphs. Fig. 9 shows the runtime of various algorithms on each subgraph. As can be seen, the runtime of our algorithms increases smoothly as the scale of bipartite graph increases, and the state-of-the-art algorithms iMB and iTrav are always worse than our algorithms under all parameter settings. With the number of vertices or edges increasing, the runtime of iMB and iTrav also increases much faster compared to our BPEA, BPEAD, and BPEAH algorithms. This result indicates that our algorithms can achieve good scalability when processing large bipartite graphs.

6.4 Case Studies

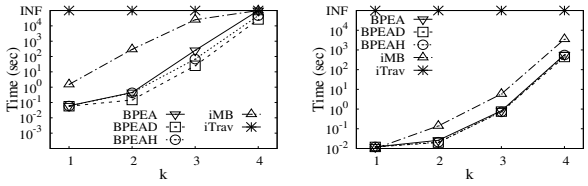
Exp-10: Fraud Detection. In this case study, we evaluate the effectiveness of the maximal biclique and maximal k-biplex models for fraud detection [19]. In e-commerce or e-business, some owners may improve the attention (or ranking) of their products by hiring fake users to purchase a set of corresponding products. To camouflage the fake users, they usually also purchase some other real products at random. Such a phenomenon is typically called the camouflage attack. To this end, we consider a camouflage attack on "Amazon Review Data"(Gift Cards)[1], which contains 147,136 reviews on 128,877 products by 1,548 users, and also injected with 200 fake users, 200 fake products, 6K fake reviews, and 6K camouflage reviews. Each fake review is randomly generated between all pairs of fake users and fake products, and camouflage reviews are randomly linked to real users (fake users) to fake produces (real produces), which coincides with a real camouflage attack as shown in [19]. We then make use of three different cohesive subgraph models, maximal biclique, maximal k-biplex, and (α, β)-core [8] to evaluate the accuracy of detected fraud users and products. For maximal biclique and maximal k-biplex, we use q_l and q_r to restrict the size of the left

[1] https://nijianmo.github.io/amazon/index.html



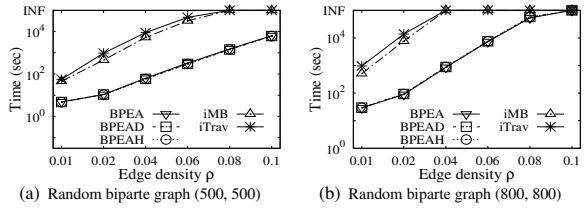
(a) YouTube (b) Cite (c) IMDB (d) Twitter (e) DBLP

Figure 6: Runtime of various algorithms for enumerating large k-biplexes with varying q ($k = 1$).



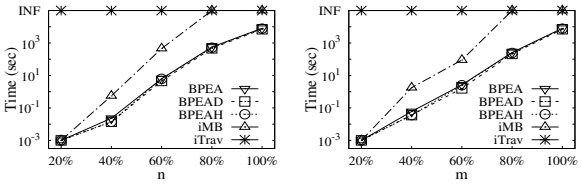
(a) Twitter ($q = 20$) (b) DBLP ($q = 10$)

Figure 7: Runtime of various algorithms with varying k.



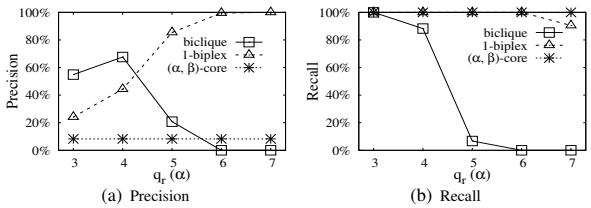
(a) Random biparte graph (500, 500) (b) Random biparte graph (800, 800)

Figure 8: Runtime of various maximal k-biplex enumeration algorithm on the bipartite graphs with varying densities ($k = 1$).



(a) varying vertices (b) varying edges

Figure 9: Scalability testing for maximal k-biplex enumeration ($k = 1, q = 8$).



(a) Precision (b) Recall

Figure 10: Precision and recall results with $q_l = \beta = 4$.

side (user-side) and right side (produce-side) vertices, respectively. Fig. 10 shows the precision and recall results with varying q_l and q_r. We can clearly see that when q_r is small ($q_r \leq 4$), the accuracy results (both precision and recall) detected by maximal biclique are very high, while as q_r grows to large $q_r > 4$, both the precision and recall results decrease rapidly. This is because the structural constraint of the maximal biclique model is too strong which results in no biclique available when q_r is large. However, for 1-biplex, the accuracy results are different. Specifically, when q_r is small, the accuracy is low, but the accuracy results will increase dramatically with the increase of q_r. This is because the 1-biplexes with small size are very sparse, and thus many results detected by k-biplex are not desirable. For the (α, β)-core, we can observe that the accuracy is always very low; and its accuracy is insensitive to the changes in α, which implies that the (α, β)-core is not very well for fraud detection. In summary, the two hereditary subgraph models we focus on in this paper work well for fraud detection.

Exp-11: Community detection. Here we further conduct a case study to investigate the effectiveness of the maximal biclique and maximal k-plex models for community detection. We use the IMDB[2] dataset in this case study. This dataset collects primary information of each film/TV production on IMDB website, including the release date, principal casts/crews, ratings, and so on. We generate a bipartite graph by connecting each movie with its principal casts/crews, and the selected movies must meet the requirements of being released after 2000 and having an average rating of no less than 6.5. Fig. 11 shows the communities of "Harry Potter" detected by the maximal biclique model and maximal k-biplex model. As can be seen, all the models can detect meaningful communities; and each detected community is highly relevant, but with slight differences. Specifically, the community detected by maximal biclique misses a movie "Sorcerer's Stone". This is because the dataset includes only a portion of principal casts for each movie, and the principal casts of "Sorcerer's Stone" in the original dataset do not include "Emma Watson", causing the result to be ignored. However, by maximal 1-biplex, such a missed result can be identified. Moreover, when k increases to 2, the maximal k-biplex can further detect a documentary "Creating the World" related to "Harry Potter". This result suggests that the maximal k-biplex model can be used as an enhancement of the maximal biclique model in community detection.

[2]https://datasets.imdbws.com

7 RELATED WORKS

Cohesive subgraph mining on bipartite graphs. Finding cohesive subgraphs from a bipartite graph has been recognized as an important problem in bipartite graph analysis. In addition to the maximal biclique model [1, 10, 15, 25, 29, 50] and the maximal k-biplex model [42, 48, 49], many other interesting cohesive subgraph models have also been investigated, such as (α, β)-core [8, 28], k-bitruss [46, 52], and quasi biclique [20, 31, 47]. Specifically, (α, β)-core [8] is a subgraph in which the minimum degree of vertices on each side is limited by a certain threshold (i.e., the degree of left/right

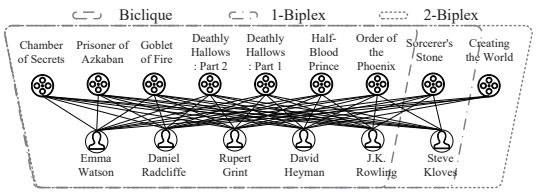


Figure 11: Communities by searching "Harry Potter".

vertices is no less than α/β). Liu et al. [28] presented an index-based algorithm for querying the (α, β)-core communities in optimal time. In [52], the authors proposed a k-bitruss model in bipartite graphs such that each edge is contained in at least k butterflies (i.e., 2×2 bicliques); and efficient online index algorithms, as well as peeling algorithms for computing k-bitruss, were developed in [46]. Quasi biclique is a relaxed biclique model which requires the edge density of the subgraph $H = (H_U, H_V, H_E)$ no less than a threshold $\gamma \in [0, 1)$ [20] or the minimum degree no less than $\gamma \cdot |H_U|$ ($\gamma \cdot |H_V|$) [31]. All these cohesive subgraph models do not satisfy the hereditary property, thus their solutions cannot be directly used for solving our problem.

Cohesive subgraph enumerations on traditional graphs. There are a large number of studies that focus on the cohesive subgraph enumeration problem on traditional graphs (non-bipartite graphs). Notable examples include maximal clique enumeration [6, 9, 12, 16, 34, 43], maximal k-plex enumeration [5, 13, 27, 51], s-clique enumeration [4, 32], and γ-quasi clique enumeration [30, 36]. All existing enumeration algorithms can be roughly classified into two categories. The first category is the output-sensitive algorithms, in which the total time complexity of the algorithm is polynomial with respect to (w.r.t.) the size of the outputs [4, 5, 9, 11, 12]. However, the practical performance of these output-sensitive solutions is often very poor; and they typically cannot be used to handle large real-world graphs. Another category of algorithms, which do not necessary to be output-sensitive, have exponential time complexity w.r.t. the number of vertices in the worst-case, but they are often very efficient when handling real-world sparse graphs due to some carefully-designed pruning techniques [13, 16, 16, 34, 43, 51]. All these existing algorithms are mainly tailored to non-bipartite graphs. It is quite non-trivial to extend these techniques to handle bipartite graphs, as the structures of cohesive subgraphs on bipartite graphs are very different from that on non-bipartite graphs.

8 CONCLUSION

In this paper, we investigate the problem of enumerating all maximal subgraphs on bipartite graphs that satisfy the hereditary property. To solve this problem, we develop a general enumeration framework which utilizes a novel and carefully-designed pivoting principle. Based on this general framework, we then propose a new pivoting technique to enumerate all maximal bicliques in a bipartite graph. We show that the time complexity of our maximal biclique enumeration algorithm is near optimal. In addition, based on the proposed general pivoting principle, we further devise a novel pivot-based algorithm with several nontrivial pruning techniques for enumerating all maximal k-biplexes. For practical applications, some optimization techniques are further developed to enumerate size-constraint maximal k-biplexes. Finally, we conduct extensive experiments using 11 real-world datasets to evaluate the proposed algorithms; and the results demonstrate the efficiency, scalability, and effectiveness of the proposed solutions.

REFERENCES

[1] Aman Abidi, Rui Zhou, Lu Chen, and Chengfei Liu. 2020. Pivot-based Maximal Biclique Enumeration. In *IJCAI*. 3558–3564.
[2] David Avis and Komei Fukuda. 1996. Reverse Search for Enumeration. *Discret. Appl. Math.* 65, 1-3 (1996), 21–46.
[3] Vladimir Batagelj and Matjaz Zaversnik. 2003. An O(m) Algorithm for Cores Decomposition of Networks. *CoRR* cs.DS/0310049 (2003).
[4] Rachel Behar and Sara Cohen. 2018. Finding All Maximal Connected s-Cliques in Social Networks. In *EDBT*. 61–72.
[5] Devora Berlowitz, Sara Cohen, and Benny Kimelfeld. 2015. Efficient Enumeration of Maximal k-Plexes. In *SIGMOD*. 431–444.
[6] Coenraad Bron and Joep Kerbosch. 1973. Finding All Cliques of an Undirected Graph (Algorithm 457). *Commun. ACM* 16, 9 (1973), 575–576.
[7] Dongbo Bu, Yi Zhao, Lun Cai, Hong Xue, Xiaopeng Zhu, Hongchao Lu, Jingfen Zhang, Shiwei Sun, Lunjiang Ling, Nan Zhang, Guojie Li, and Runsheng Chen. 2003. Topological structure analysis of the protein–protein interaction network in budding yeast. *Nucleic Acids Research* 31, 9 (05 2003), 2443–2450.
[8] Monika Cerinsek and Vladimir Batagelj. 2015. Generalized two-mode cores. *Soc. Networks* 42 (2015), 80–87.
[9] Lijun Chang, Jeffrey Xu Yu, and Lu Qin. 2013. Fast Maximal Cliques Enumeration in Sparse Graphs. *Algorithmica* 66, 1 (2013), 173–186.
[10] Lu Chen, Chengfei Liu, Rui Zhou, Jiajie Xu, and Jianxin Li. 2022. Efficient Maximal Biclique Enumeration for Large Sparse Bipartite Graphs. *Proc. VLDB Endow.* 15, 8 (2022), 1559–1571.
[11] Sara Cohen, Benny Kimelfeld, and Yehoshua Sagiv. 2008. Generating all maximal induced subgraphs for hereditary and connected-hereditary graph properties. *J. Comput. Syst. Sci.* 74, 7 (2008), 1147–1159.
[12] Alessio Conte, Roberto Grossi, Andrea Marino, and Luca Versari. 2016. Sublinear-Space Bounded-Delay Enumeration for Massive Network Analytics: Maximal Cliques. In *ICALP*, Vol. 55. 148:1–148:15.
[13] Alessio Conte, Tiziano De Matteis, Daniele De Sensi, Roberto Grossi, Andrea Marino, and Luca Versari. 2018. D2K: Scalable Community Detection in Massive Networks via Small-Diameter k-Plexes. In *KDD*. 1272–1281.
[14] Peter Damaschke. 2014. Enumerating maximal bicliques in bipartite graphs with favorable degree sequences. *Inf. Process. Lett.* 114, 6 (2014), 317–321.
[15] Apurba Das and Srikanta Tirthapura. 2019. Shared-Memory Parallel Maximal Biclique Enumeration. In *HiPC*. 34–43.
[16] David Eppstein, Maarten Löffler, and Darren Strash. 2010. Listing All Maximal Cliques in Sparse Graphs in Near-Optimal Time. In *ISAAC*, Vol. 6506. 403–414.
[17] Fedor V. Fomin and Dieter Kratsch. 2010. *Exact Exponential Algorithms*. Springer.
[18] Stephan Günnemann, Emmanuel Müller, Sebastian Raubach, and Thomas Seidl. 2011. Flexible Fault Tolerant Subspace Clustering for Data with Missing Values. In *ICDM*. 231–240.
[19] Bryan Hooi, Hyun Ah Song, Alex Beutel, Neil Shah, Kijung Shin, and Christos Faloutsos. 2016. FRAUDAR: Bounding Graph Fraud in the Face of Camouflage. In *KDD*. 895–904.
[20] Dmitry I. Ignatov. 2019. Preliminary Results on Mixed Integer Programming for Searching Maximum Quasi-Bicliques and Large Dense Biclusters. In *ICFCA*, Vol. 2378. 28–32.
[21] Sune Lehmann, Martin Schwartz, and Lars Kai Hansen. 2008. Biclique communities. *Phys. Rev. E* 78 (2008), 016108. Issue 1.
[22] John M. Lewis and Mihalis Yannakakis. 1980. The Node-Deletion Problem for Hereditary Properties is NP-Complete. *J. Comput. Syst. Sci.* 20, 2 (1980), 219–230.
[23] Michael Ley. 2002. The DBLP Computer Science Bibliography: Evolution, Research Issues, Perspectives. In *SPIRE*, Vol. 2476. 1–10.
[24] Haiquan Li, Jinyan Li, and Limsoon Wong. 2006. Discovering motif pairs at interaction sites from protein sequences on a proteome-wide scale. *Bioinform.* 22, 8 (2006), 989–996.
[25] Jinyan Li, Guimei Liu, Haiquan Li, and Limsoon Wong. 2007. Maximal Biclique Subgraphs and Closed Pattern Pairs of the Adjacency Matrix: A One-to-One Correspondence and Mining Algorithms. *IEEE Trans. Knowl. Data Eng.* 19, 12 (2007), 1625–1637.
[26] Rong-Hua Li, Qiushuo Song, Xiaokui Xiao, Lu Qin, Guoren Wang, Jeffrey Xu Yu, and Rui Mao. 2022. I/O-Efficient Algorithms for Degeneracy Computation on Massive Networks. *IEEE Trans. Knowl. Data Eng.* 34, 7 (2022), 3335–3348.
[27] Don R. Lick and Arthur T. White. 1970. k-Degenerate graphs. *Canadian Journal of Mathematics* 22, 5 (1970), 1082–1096.

[28] Boge Liu, Long Yuan, Xuemin Lin, Lu Qin, Wenjie Zhang, and Jingren Zhou. 2020. Efficient (α, β)-core computation in bipartite graphs. *VLDB J.* 29, 5 (2020), 1075–1099.

[29] Guimei Liu, Kelvin Sim, and Jinyan Li. 2006. Efficient Mining of Large Maximal Bicliques. In *DaWaK*, Vol. 4081. 437–448.

[30] Guimei Liu and Limsoon Wong. 2008. Effective Pruning Techniques for Mining Quasi-Cliques. In *ECML/PKDD*, Vol. 5212. 33–49.

[31] Xiaowen Liu, Jinyan Li, and Lusheng Wang. 2008. Quasi-bicliques: Complexity and Binding Pairs. In *COCOON 2008*, Vol. 5092. 255–264.

[32] John W. Moon and Leo Moser. 1965. On cliques in graphs. *Israel journal of Mathematics* 3, 1 (1965), 23–28.

[33] Azam Sheikh Muhammad, Peter Damaschke, and Olof Mogren. 2016. Summarizing Online User Reviews Using Bicliques. In *SOFSEM*, Vol. 9587. 569–579.

[34] Kevin A. Naudé. 2016. Refined pivot selection for maximal clique enumeration in graphs. *Theor. Comput. Sci.* 613 (2016), 28–37.

[35] René Peeters. 2003. The maximum edge biclique problem is NP-complete. *Discret. Appl. Math.* 131, 3 (2003), 651–654.

[36] Jian Pei, Daxin Jiang, and Aidong Zhang. 2005. On mining cross-graph quasi-cliques. In *KDD*. 228–238.

[37] Ardian Kristanto Poernomo and Vivekanand Gopalkrishnan. 2009. Towards efficient mining of proportional fault-tolerant frequent itemsets. In *KDD*. 697–706.

[38] Erich Prisner. 2000. Bicliques in Graphs I: Bounds on Their Number. *Comb.* 20, 1 (2000), 109–117.

[39] Ron Rymon. 1992. Search through Systematic Set Enumeration. In *Proceedings of the 3rd International Conference on Principles of Knowledge Representation and Reasoning*. 539–550.

[40] Eran Shaham, Honghai Yu, and Xiaoli Li. 2016. On finding the maximum edge biclique in a bipartite graph: a subspace clustering approach. In *SIAM*. 315–323.

[41] Kelvin Sim, Jinyan Li, Vivekanand Gopalkrishnan, and Guimei Liu. 2006. Mining Maximal Quasi-Bicliques to Co-Cluster Stocks and Financial Ratios for Value Investment. In *ICDM*. 1059–1063.

[42] Kelvin Sim, Jinyan Li, Vivekanand Gopalkrishnan, and Guimei Liu. 2009. Mining maximal quasi-bicliques: Novel algorithm and applications in the stock market and protein networks. *Stat. Anal. Data Min.* 2, 4 (2009), 255–273.

[43] Etsuji Tomita, Akira Tanaka, and Haruhisa Takahashi. 2006. The worst-case time complexity for generating all maximal cliques and computational experiments. *Theor. Comput. Sci.* 363, 1 (2006), 28–42.

[44] Oliver Voggenreiter, Stefan Bleuler, and Wilhelm Gruissem. 2012. Exact biclustering algorithm for the analysis of large gene expression data sets. *BMC Bioinform.* 13, S-18 (2012), A10.

[45] Haibo Wang, Chuan Zhou, Jia Wu, Weizhen Dang, Xingquan Zhu, and Jilong Wang. 2018. Deep Structure Learning for Fraud Detection. In *ICDM*. 567–576.

[46] Kai Wang, Xuemin Lin, Lu Qin, Wenjie Zhang, and Ying Zhang. 2020. Efficient Bitruss Decomposition for Large-scale Bipartite Graphs. In *ICDE*. 661–672.

[47] Lusheng Wang. 2010. Near Optimal Solutions for Maximum Quasi-bicliques. In *COCOON*, Vol. 6196. 409–418.

[48] Kaiqiang Yu, Cheng Long, Shengxin Liu, and Da Yan. 2022. Efficient Algorithms for Maximal k-Biplex Enumeration. In *SIGMOD*. 860–873.

[49] Kaiqiang Yu, Cheng Long, Deepak P, and Tanmoy Chakraborty. 2021. On Efficient Large Maximal Biplex Discovery. *IEEE Trans. Knowl. Data Eng.* (2021).

[50] Yun Zhang, Charles A. Phillips, Gary L. Rogers, Erich J. Baker, Elissa J. Chesler, and Michael A. Langston. 2014. On finding bicliques in bipartite graphs: a novel algorithm and its application to the integration of diverse biological data types. *BMC Bioinform.* 15 (2014), 110.

[51] Yi Zhou, Jingwei Xu, Zhenyu Guo, Mingyu Xiao, and Yan Jin. 2020. Enumerating Maximal k-Plexes with Worst-Case Time Guarantee. In *AAAI*. 2442–2449.

[52] Zhaonian Zou. 2016. Bitruss Decomposition of Bipartite Graphs. In *DASFAA*, Vol. 9643. 218–233.