On Maximal Chain Subgraphs and Covers of Bipartite Graphs

Tiziana Calamoneri[1], Mattia Gastaldello[1,2], Arnaud Mary[2], Marie-France Sagot[2], and Blerina Sinaimeri[2]

[1] Sapienza University of Rome
via Salaria 113, 00198 Roma, Italy.
[2] INRIA and Université de Lyon
Université Lyon 1, LBBE, CNRS UMR558, France.

Abstract. In this paper, we address three related problems. One is the enumeration of all the maximal *edge-induced* chain subgraphs of a bipartite graph. We give bounds on their number and use them to establish the input-sensitive complexity of the enumeration problem. The second problem we treat is the *minimum* chain subgraph cover. Finally, we approach the problem of enumerating all *minimal* chain subgraph covers and show that it can be solved in quasi-polynomial time.
Keywords: Chain Subgraph Cover Problem, Enumeration Algorithms, Exact exponential algorithms.

1 Introduction

Enumerating (listing) the subgraphs of a given graph plays an important role in analysing its structural properties. Thus, it is a central issue in many areas, notably in data mining and computational biology.

In this paper, we address the problem of enumerating without repetitions all maximal *edge-induced* chain subgraphs of a bipartite graph. These are graphs that do not contain a $2K_2$ as induced subgraph (*i.e.* there are no independent edge sets of size 2). From now on, we will refer to them as *chain subgraphs* for short when there is no ambiguity.

Bipartite graphs arise naturally in many applications, such as biology as will be mentioned later in the introduction, since they enable to model the relations between two different classes of objects. The problem of enumerating in bipartite graphs all subgraphs with certain properties has thus already been considered in the literature. These concern for instance maximal bicliques for which polynomial delay enumeration algorithms in bipartite [6,11] as well as in general graphs [5,11] were provided. In the case of maximal *induced* chain subgraphs, their enumeration

V. Biló, A. Caruso (Eds.): ICTCS 2016, Proceedings of the 17th Italian Conference on Theoretical Computer Science, 73100 Lecce, Italy, September 7–9 2016, pp. 286–291 published in CEUR Workshop Proceedins Vol-1720 at http://ceur-ws.org/Vol-1720

can be done in output polynomial time as it can be reduced to the enumeration of a particular case of the minimal hitting set problem [7] (where the sets in the family have cardinality 4). However, the existence of a polynomial delay algorithm for this problem remains open. To the best of our knowledge, nothing is known so far about the problem of enumerating maximal *edge-induced* chain subgraphs in bipartite graphs.

In this paper, we propose a polynomial delay algorithm to enumerate all maximal chain subgraphs of a bipartite graph. We also provide an analysis of the time complexity of this algorithm in terms of input size. In order to do this, we prove some upper bounds on the maximum number of maximal chain subgraphs of a bipartite graph G with n nodes and m edges. This is also of intrinsic interest as combinatorial bounds on the maximum number of specific subgraphs in a graph are difficult to obtain and have received a lot of attention (see for *e.g.* [8,12]).

We then address a second related problem called the *minimum chain subgraph cover* problem. This asks to determine, for a given graph G, the minimum number of chain subgraphs that cover all the edges of G. This has already been investigated in the literature as it is related to other well-known problems such as maximum induced matching (see *e.g.* [3,4]). For bipartite graphs, the problem was shown to be NP-hard [14].

We provide an exact exponential algorithm which runs in time $O^*((2+\varepsilon)^m)$, for every $\varepsilon > 0$ (by O^* we denote standard big O notation but omitting polynomial factors). Notice that, since a chain subgraph cover is a family of subsets of edges, the existence of an algorithm whose complexity is close to 2^m is not obvious. Indeed, the basic search space would have size 2^{2^m}, which corresponds to all families of subsets of edges of a graph on m edges.

Finally, we approach the problem of enumerating all minimal covers by chain subgraphs. To this purpose, we provide a quasi-polynomial time algorithm to enumerate all *minimal* covers by maximal chain subgraphs of a bipartite graph. To do so, we prove that this can be polynomially reduced to the enumeration of the minimal set covers of a hypergraph.

Besides their theoretical interest, the problems of finding one minimum chain subgraph cover and of enumerating all such covers have also a direct application in biology. Nor *et al.* [13] showed that a minimum chain subgraph cover of such a bipartite graph provides a good model for identifying the minimum genetic architecture enabling to explain one type of manipulation, called *cytoplasmic incompatibility*, by bacteria of a genus called *Wolbachia* of their insect hosts. Moreover, as different minimum covers may correspond to solutions that differ in terms of their biological interpretation, the capacity to enumerate all such minimum chain covers becomes crucial.

2 Preliminaries

Throughout the paper, we assume that the reader is familiar with the standard graph terminology, as contained for instance in [2]. We consider finite undirected graphs without loops or multiple edges.

Given a bipartite graph $G = (U \cup W, E)$ and a node $u \in U$, we denote by $N_G(u)$ the set of nodes adjacent to u in G and by $E_G(u)$ the *set of edges incident to u in G*. Moreover, given $U' \subseteq U$ and $W' \subseteq W$, we denote by $G[U', W']$ the *subgraph of G induced by* $U' \cup W'$. A node $u \in U$ such that $N_G(u) = W$ is called a *universal node*.

For a *chain graph*, an equivalent condition of not containing a $2K_2$ as an induced subgraph it is that for each two nodes v_1 and v_2 both in U (resp. in W), it holds that either $N_G(v_1) \subseteq N_G(v_2)$ or $N_G(v_2) \subseteq N_G(v_1)$. Given a chain subgraph $C = (X \cup Y, F)$ of G, with the *largest neighbourhood of* C, we mean the neighbourhood of a node x in X for which the set $N_C(x) \subseteq Y$ has maximum cardinality. A set $Y' \subseteq Y$ is a *maximal neighborhood* of G, if there exists $u \in U$ such that $N_G(u) = V'$ and there does not exist a node $u' \in U$ such that $N_G(u) \subset N_G(U')$.

In this paper, we always consider *edge-induced* chain subgraphs of a graph G. Hence, we identify a chain subgraph C of G by its set of edges $E(C) \subseteq E(G)$ and in that case its set of nodes will be constituted by all the nodes of G incident to at least one edge in C (sometimes abusing notation, we more simply write $C \subseteq G$ or $e \in C$). A *maximal chain subgraph* C of a given bipartite graph G is a connected chain subgraph such that no superset of $E(C)$ is a chain subgraph. We denote by $\mathcal{C}(G)$ the set of all maximal chain subgraphs in G.

A set of chain subgraphs $C_1, \ldots, C_k$ is a *cover* for G if $\cup_{1 \leq i \leq k} E(C_i) = E(G)$. Observe that, given any cover of G by chain subgraphs $C = \{C_1, \ldots C_k\}$, there exists another cover of same size $C' = \{C'_1, \ldots C'_k\}$ whose chain subgraphs are all maximal; more precisely, for each $i = 1, \ldots, k$, C'_i is a maximal chain subgraph of G and C'_i admits C_i as subgraph. In order to avoid redundancies, from now on, although not explicitly highlighted, we will restrict our attention to the covers by maximal chain subgraphs.

We denote by $\mathcal{S}(G)$ the set of all minimal chain covers of a bipartite graph G.

An enumeration algorithm is said to be *output polynomial* or *total polynomial* if the total running time is polynomial in the size of the input and the output. It is said to be *polynomial delay* if the time between the output of any one solution and the next one is bounded by a polynomial function of the input size [10].

3 Enumerating All Maximal Chain Subgraphs

The following theorem characterizes the structure of a maximal chain subgraph and it is fundamental for all the other results of the paper.

Theorem 1. *Let $C = (X \cup Y, F)$ be a chain subgraph of $G = (U \cup W, E)$, with $X \subseteq U$, $Y \subseteq W$ and $F \subseteq E$, and let $x \in X$ be a node with largest neighbourhood in C. Then C is a maximal chain subgraph of G if and only if:*

(i) $N_C(x) = N_G(x)$ *is a maximal neighbourhood of G, i.e. there does not exist a node $u' \in U$ such that $N_G(u) \subset N_G(u')$.*
(ii) $C \setminus E_G(x)$ *is a maximal chain subgraph of* $G\big[U \setminus \{x\}, N_G(x)\big]$.

Theorem 1 is the basis of a new recursive algorithm which enumerates all maximal chain subgraphs of G with polynomial delay:

Proposition 1 (Time Complexity and Polynomial Delay). *Let $G = (U \cup W, E)$ be a bipartite graph. It is possible to enumerate all maximal chain subgraphs of G with a total running time of $O(|\mathcal{C}(G)|n^2m)$. Moreover, the solutions are enumerated in polynomial time delay $O(n^2m)$.*

These two statements allow us to achieve some other results briefly described in the following.

3.1 Bounds on the number of maximal chains

By Theorem 1*(ii)*, a maximal chain subgraph can be found by recursively reducing the graph to one whose partition has size $|U|-1$, so we obtain that the maximum number of chain subgraphs is bounded by $\min(|U|, |W|)!$ and that this bound is tight as *e.g.* the *antimatching graph* reach this bound.

We give also a bound on the number of maximal chain subgraphs for a bipartite graph with m edges:

Theorem 2. *Let $G = (U \cup W, E)$ be a bipartite graph with m edges; then $|\mathcal{C}(G)| \leq 2^{\sqrt{m}\log m}$.*

3.2 Minimum Chain Subgraph Cover

Exploiting Proposition 1, the bound obtained in Theorem 2 and the inclusion/exclusion method [1,8] that has already been successfully applied to exact exponential algorithms for many partitioning and covering problems, we are able to provide an $O^*((2+\epsilon)^m)$ algorithm to decide if there exists a chian subgraph cover of size k for a given bipartite observing that the basic search space has size 2^{2^m}.

Theorem 3. *Let $c_k(G)$ be the number of chain subgraph covers of size k of a graph G. Given a bipartite graph G with m edges, for all $k \in \mathbb{N}^*$ and for all $\varepsilon > 0$, $c_k(G)$ can be computed in time $O^*((2+\epsilon)^m)$.*

3.3 Enumeration of Minimal Chain Subgraph Covers

The enumeration of all minimal chain subgraph covers can be polynomially reduced to the enumeration of the minimal set covers of a hypergraph. This reduction implies that there is a quasi-polynomial time algorithm to enumerate all minimal chain subgraph covers. Indeed, the result in [9] implies that all the minimal set covers of a hypergraph can be enumerated in time $N^{\log N}$ where N is the sum of the input size (*i.e.* $n+m$) and of the output size (*i.e.* the number of minimal set covers).

Let $\mathcal{S} = \mathcal{S}(G)$ the set of its minimal chain subgraph covers. Notice that the minimal chain subgraph covers of G are the minimal set covers of the hypergraph $\mathcal{H} := (V, \mathcal{E})$ where $V = E$ and $\mathcal{E} = \mathcal{C}$. Unfortunately, the size of $\mathcal{H}$ might be exponential in the size of G plus the size of $\mathcal{S}$. Indeed not every maximal chain subgraph in $\mathcal{C}$ will necessarily be part of some minimal chain subgraph cover. In order to obtain a quasi-polynomial time algorithm to enumerate all minimal chain subgraph covers, we need to enumerate only those maximal chain subgraphs that belong to a minimal chain subgraph cover.

Given an edge $e \in E$, let $\mathcal{C}_e$ be the set of all maximal chain subgraphs of G containing e and $\mathcal{M}_e$ the set of all edges $e' \in E$ inducing a $2K_2$ in G together with e.

We call an edge $e \in E$ *non-essential* if there exists another edge $e' \in E$ such that $\mathcal{C}_{e'} \subset \mathcal{C}_e$. An edge which is not non-essential is said to be *essential.* Note that for every non-essential edge e, there exists an essential edge e_1 such that $\mathcal{C}_{e_1} \subset \mathcal{C}_e$. Indeed, by applying iteratively the definition of a non-essential edge, we obtain a list of inclusions $\mathcal{C}_e \supset \mathcal{C}_{e_1} \supset \mathcal{C}_{e_2} \ldots$, where no $\mathcal{C}_{e_i}$ is repeated as the inclusions are strict. The last element of the list will correspond to an essential edge.

By the next Lemma we show that it is sufficient to consider the chain subgraphs which contain at least an essential edge.

Lemma 1. *Let C be a maximal chain subgraph of a bipartite graph $G = (U \cup W, E)$. Then C belongs to a minimal chain subgraph cover of G if and only if C contains an essential edge.*

In the following, we show how to detect essential edges.

Theorem 4. *Given a bipartite graph $G = (U \cup W, E)$, for any two edges $e, e' \in E$, $\mathcal{C}_e \subseteq \mathcal{C}_{e'}$ if and only if $\mathcal{M}_e \supseteq \mathcal{M}_{e'}$.*

Notice that, given an edge $e = (u, w) \in E$, $u \in U$ and $w \in W$, it is easy to determine the set $\mathcal{M}_e$, and checking whether $\mathcal{M}_e \supseteq \mathcal{M}_{e'}$ is also easy.

These results allow us to achieve the following result:

Theorem 5. *Given a bipartite graph $G = (U \cup W, E)$, one can enumerate all its minimal chain subgraph covers, i.e. all the elements in $\mathcal{S}$, in time $O(|\mathcal{S}|^{\log(|\mathcal{S}|)+2})$.*

References

1. Andreas Björklund, Thore Husfeldt, and Mikko Koivisto. Set partitioning via inclusion-exclusion. *SIAM J. Comput.*, 39(2):546–563, July 2009.
2. Béla Bollobás. *Modern graph theory*, volume 184 of *Graduate Texts in Mathematics*. Springer-Verlag, Berlin, 1998.
3. Andreas Brandstädt, Elaine M Eschen, and R Sritharan. The induced matching and chain subgraph cover problems for convex bipartite graphs. *Theoretical computer science*, 381(1):260–265, 2007.
4. Yu Chang-Wu, Chen Gen-Huey, and Ma Tze-Heng. On the complexity of the k-chain subgraph cover problem. *Theoretical computer science*, 205(1):85–98, 1998.
5. Vânia M.F. Dias, Celina M.H. de Figueiredo, and Jayme L. Szwarcfiter. Generating bicliques of a graph in lexicographic order. *Theoretical Computer Science*, 337(1-3):240 – 248, 2005.
6. Vânia M.F. Dias, Celina M.H. de Figueiredo, and Jayme L. Szwarcfiter. On the generation of bicliques of a graph. *Discrete Applied Mathematics*, 155(14):1826 – 1832, 2007.
7. Thomas Eiter and Georg Gottlob. Identifying the minimal transversals of a hypergraph and related problems. *SIAM Journal on Computing*, 24(6):1278–1304, 1995.
8. Fedor V. Fomin and Dieter Kratsch. *Exact Exponential Algorithms*. Springer-Verlag New York, Inc., New York, NY, USA, 2010.
9. Michael L Fredman and Leonid Khachiyan. On the complexity of dualization of monotone disjunctive normal forms. *Journal of Algorithms*, 21(3):618–628, 1996.
10. D. S. Johnson, M. Yannakakis, and C. H. Papadimitriou. On generating all maximal independent sets. *Information Processing Letters*, 27(3):119–123, 1988.
11. Kazuhisa Makino and Takeaki Uno. *SWAT 2004, Lecture Notes in Computer Science*, chapter New Algorithms for Enumerating All Maximal Cliques, pages 260–272. Springer Berlin Heidelberg, Berlin, Heidelberg, 2004.
12. J. W. Moon and L. Moser. On cliques in graphs. *Israel Journal of Mathematics*, 3(1):23–28, 1965.
13. Igor Nor, Jan Engelstädter, Olivier Duron, Max Reuter, Marie-France Sagot, and Sylvain Charlat. On the genetic architecture of cytoplasmic incompatibility: inference from phenotypic data. *The American Naturalist*, 182(1):E15–E24, 2013.
14. Mihalis Yannakakis. The complexity of the partial order dimension problem. *SIAM Journal on Algebraic Discrete Methods*, 3(3):351–358, 1982.