The Complexity of Subgraph Isomorphism: Duality Results for Graphs of Bounded Path- and Tree-Width

Arvind Gupta*   Naomi Nishimura†

March 20, 1995

Abstract

We present a clear demarcation between classes of bounded tree-width graphs for which the subgraph isomorphism problem is $\mathcal{NP}$-complete and those for which it can be solved in polynomial time. In previous work, it has been shown that this problem is solvable in polynomial time if the source graph has either bounded degree or is k-connected, for k the tree-width of the two graphs. As well, it has also been shown that for certain specific connectivity or degree conditions, the problem becomes $\mathcal{NP}$-complete. Here we give a complete characterization of the complexity of this problem on bounded tree-width graphs for all possible connectivity conditions of the two input graphs. Specifically, we show that when the source graph is not k-connected or has more than k vertices of unbounded degree the problem is $\mathcal{NP}$-complete, thus answering an open question of Matoušek and Thomas.

Many of our reductions are restricted to using a subset of graphs of bounded tree-width, namely graphs of bounded path-width. As a direct result of our constructions, we also show that when the source and target graphs have connectivity less than k or at least one has k vertices of unbounded degree, the subgraph isomorphism problem for bounded path-width graphs is $\mathcal{NP}$-complete.

1 Introduction

The subgraph isomorphism problem is known to be $\mathcal{NP}$-complete for general graphs, but can be solved in polynomial time for many restricted classes of graphs. We can phrase the problem as that of trying to determine whether or not there is a subgraph of an input graph H that is isomorphic to an input graph G. In this paper, we study this problem

*School of Computing Science, Simon Fraser University, Burnaby, British Columbia, Canada, V5A 1S6. email: arvind@cs.sfu.ca, FAX (604) 291-3045. Research supported by the Natural Sciences and Engineering Research Council of Canada and the Advanced Systems Institute.

†Department of Computer Science, University of Waterloo, Waterloo, Ontario, Canada, N2L 3G1. email: nishi@plg.uwaterloo.ca, FAX (519) 885-1208. Research supported by the Natural Sciences and Engineering Research Council of Canada and the Information Technology Research Centre.

when G and H are both graphs of bounded tree-width with various connectivity and degree conditions, and show that there is a clear division between those cases in which the problem is polynomial time and those in which it is $\mathcal{NP}$-complete.

Polynomial-time algorithms for subgraph isomorphism have been devised for restricted classes of graphs, including trees [Mat78], two-connected outerplanar graphs [Lin89], and two-connected series-parallel graphs [LS88]. These are all graphs of bounded tree-width (as defined formally in Section 2), for which the problem is $\mathcal{NP}$-complete in general [Sys82]. More generally, it has been shown that for G and H partial k-trees, if G either has bounded degree or is k-connected then there is a polynomial time algorithm for subgraph isomorphism [MT92, GN94].

One natural question is that of determining the complexity of this problem if the constraints on G are further relaxed. In this paper we study two scenarios, namely allowing the connectivity of G to be less than k and allowing a constant number of nodes of unbounded degree. Such questions were also studied by Matoušek and Thomas [MT92] who showed that the problem is $\mathcal{NP}$-complete when G is a tree with all but one node of degree at most three and H is a graph of tree-width two with all but one node of degree at most three. However, they left as an open problem the case where, for example, G has connectivity $\frac{k+1}{2}$ and G and H are both partial k-trees, and hypothesized that this problem may in fact be solvable in polynomial time. Our results directly show that the problem is, in fact, $\mathcal{NP}$-complete.

In this paper we examine a subset of the class of graphs of bounded tree-width, the graphs of bounded path-width. We derive the complexity of the subgraph isomorphism problem for G and H both graphs of path-width k where G is g-connected and H is h-connected, for both g and h less than k. Since $\mathcal{NP}$-completeness results obtained for graphs of bounded path-width automatically apply to graphs of bounded tree-width, similar results are obtained for this larger class. Furthermore, we show that when G and H have tree-width k with H k-connected and G with connectivity less than k, the problem is again $\mathcal{NP}$-complete. We thus obtain sharp divisions for subgraph isomorphism on bounded tree-width graphs and nearly sharp divisions for bounded path-width graphs.

In Section 2 we formally define the classes of graphs under consideration, and discuss related work. Algorithms and reductions for various classes are presented in Sections 3 through 6. Finally, in Section 7 we summarize our results and suggest various directions for further research.

2 Preliminaries

In this section, we formally define the classes of graphs under consideration and review previous work on such graphs. We begin with some basic definitions.

2.1 Graphs and Graph Unions

All graphs in this paper will be simple. We denote the vertex and edge sets of a graph G by $V(G)$ and $E(G)$ respectively. We will be working extensively with trees and paths; the reader is expected to have a basic familiarity with these types of graphs (see Bondy and Murty [BM76] for more background material). In addition, we will focus on the connectivity of graphs; when we say that a graph G is *g-connected*, we mean that G is g-connected but not $(g+1)$-connected.

For P a path on n vertices, we will often write $P = v_1, \ldots, v_\ell$ by which we mean $V(P) = \{v_1, \ldots, v_\ell\}$ and $E(P) = \{(v_i, v_{i+1}) : 1 \leq i < \ell\}$ where all v_i's are distinct. We say that the length of this path (denoted $|P|$) is ℓ.

The graphs constructed in our reductions will consist of the joining together of a number of different graphs; we next formally define the union of graphs.

Definition: Given two graphs $G_1 = (V_1, E_1)$ and $G_2 = (V_2, E_2)$, the *union of* G_1 *and* G_2 is the graph $G_1 \bigcup G_2 = (V_1 \cup V_2, E_1 \cup E_2)$. If a node $u \in V_1$ is the same as the node $v \in V_2$, we say that *u is identified with v* in the union, or that the union is formed by *identifying u and v*.

Notice that our unions preserve simple graphs, that is, when we take the union of the edge set we do not allow multiple edges.

2.2 Tree-decompositions and tree-width

A tree-decomposition is a representation of a graph by a tree-like structure [RS86]; a more formal definition follows.

Definition: Let G be a graph. A *tree-decomposition* for G is a pair (T^G, χ^G) where T^G is a tree and $\chi^G : V(T^G) \rightarrow \{\text{subsets of } V(G)\}$ satisfying:

1. for every $e = (u, v) \in E(G)$, there is an $x \in V(T^G)$ such that $u, v \in \chi^G(x)$; and
2. for $x, y, z \in V(T^G)$, if y is on the path from x to z in T^G then $\chi^G(x) \cap \chi^G(z) \subseteq \chi^G(y)$.

The *width* of a tree-decomposition (T^G, χ^G) is $\max\{|\chi^G(x)| - 1 : x \in V(T^G)\}$ and the tree-width of a graph G is the minimum width over all its tree-decompositions. We will drop the superscripts when the graph G is clear from context.

An alternate characterization of graphs of tree-width k can be given in terms of partial k-trees [Ros74]. Recall that the complete graph on r nodes is called an *r-clique* and is denoted by K_r.

Definition: A *k-tree* is defined inductively as follows:

1. K_k and K_{k+1} are k-trees.
2. Let G be a k-tree on n nodes and let K be a k-clique in G. Then the $n+1$ node graph G' formed from G by adding a new node v adjacent to all nodes of K is a k-tree.

A *partial k-tree* is a subgraph of k-tree.

The following lemma is well-known:

Lemma 2.1. *G is a partial k-tree if and only G has tree-width at most k.*

2.3 Path-decompositions and path-width

We can view a path-decomposition of a graph in a variety of ways. Perhaps the simplest is the following:

Definition: Let G be a graph. A *path-decomposition* of G is a tree-decomposition (P^G, χ^G) in which P^G is a path; again we drop superscripts whenever the graph G is clear from context.

The definitions of the width of a path-decomposition and the path-width of G are defined analogously. We can also modify the definition of partial k-trees to yield a different characterization of path-width k graphs.

Definition: A *k-path* is defined inductively as follows:

1. K_k is a k-path.
2. K_{k+1} is a k-path with one node designated as the *distinguished node.*
3. Let G be a k-path on $n > k$ nodes with distinguished node v. Let K be either the k-clique to which v is adjacent or a k-clique involving v. Then the $n+1$ node graph G' formed from G by adding a new node w adjacent to all nodes of K is a k-path with w the distinguished node.

A *partial k-path* is a subgraph of a k-path.

We will be working extensively with the k-path definition; the following further definitions will be useful. We can view the formation of a k-path as a step-by-step procedure, consisting first of the creation of a clique of size k and then the addition, in steps 1 through ℓ, of the $(k+1)$st through $(k+\ell)$th nodes in the k-path. The first k nodes will be called the *original clique.* The distinguished node added at step i will be denoted u_i, and the k-clique to which it is attached will be called the *attachment clique*, C_i. Thus, C_i will either be identical to C_{i-1} or will contain u_{i-1}.

Lemma 2.2. *G is a partial k-path if and only if G has path-width at most k.*

Proof: To show that a partial k-path G has path-width at most k, we construct a width k path-decomposition of a k-path G' containing G as a subgraph. Without loss of generality, we can assume that G' has more than k nodes. In the iterative construction of G', let K_k be the original clique of G', let the sequence of distinguished nodes in the construction of G' be $u_1, \ldots, u_r$ and let C_i be the attachment clique of u_i. Let $P = w_0, w_1, \ldots, w_r$ be a path. We define $\chi : V(P) \rightarrow \{\text{subsets of } V(G')\}$ by $\chi(w_0) = V(K_k)$ and for $i > 0$, $\chi(w_i) = \{u_i\} \bigcup V(C_i)$. It can be verified that (P, χ) is a width k path-decomposition of G'. Since G is a subgraph of G', it also has a width k path-decomposition.

To prove the converse, we assume that G is a graph of path-width at most k with a path-decomposition (P, χ) of width at most k. By altering (P, χ), we will construct a k-path H of which G is a subgraph. First, we create a path-decomposition (P', χ') such that for every node w of P', $|\chi'(w)| = k+1$. It is not difficult to see that for $P' = P$, χ' can be formed from χ by adding nodes to the domains when needed.

We now form a path-decomposition (P'', χ'') from (P', χ') such that for each pair of adjacent nodes v and w of P'', $|\chi''(v) \bigcap \chi''(w)| = k$. This can be achieved by adding new nodes to P'' between each pair of nodes v and w of P' for which the intersection is too small such that for each successive new node x, $\chi''(x)$ contains one new node of $\chi'(w)$ and one fewer node of $\chi'(v)$.

Finally, we let H be the graph with path-decomposition (P'', χ'') with edge set $\{(x, y) : x, y \in \chi''(w) \text{ for some } w \in P''\}$. By construction, H is a k-path; since G is a subgraph of H, G is a partial k-path, as needed. ■

2.3.1 Previous work on graphs of bounded path-width

Paralleling the work on tree-width, much of the work on path-width has focused on relating path-width to other measures and on the problem of determining the path-width of various classes of graphs. Korach and Solel relate tree-width and path-width by showing that for any graph on n nodes, the path-width is in $O(\log n \cdot \text{tree-width})$ [KS93]. In addition, the notion of path-width has been related to cut-width [KS93], node search number [BM90], interval thickness [BM90], vertex separation number [Kin92], and the gate matrix layout problem in VLSI [Moh90]. Bodlaender and Möhring show that the tree-width and path-width are equal for cographs [BM90]. Although it is $\mathcal{NP}$-complete in general to determine the path-width of a graph [ACP89], there exist algorithms for computing exact path-width for restricted classes of graphs, such as permutation graphs [BKK93], cographs [BM90], splitgraphs [KB92], and interval graphs. In addition, there exist approximation algorithms for cotriangulated graphs, permutation graphs, and cocomparability graphs [KB92], as well as efficient algorithms for finding path-decompositions [BK91].

3 Algorithms

The subgraph isomorphism problem can be solved in polynomial time when the input graphs G and H are both partial k-trees and when G is restricted to being either k-connected or bounded degree [MT92, GN94]. In this section we give a brief outline of the techniques used to solve the problem; such techniques have also been used to yield parallel algorithms for both subgraph isomorphism and topological embedding.

In order to adapt the dynamic programming approach developed originally by Matula for trees [Mat78], each graph is first represented by a tree-decomposition. Since these representations are not necessarily unique, the tree-decomposition of H is put into a special "normalized" form, and the corresponding tree-decomposition of G is determined in the course of the algorithm. More specifically, we create a special structure known as a *tree-decomposition graph*, containing all possible normalized tree-decompositions of G including the one "matching" that of H. The algorithm proceeds by finding matchings between vertices in the tree decomposition of H and nodes in the tree-decomposition graph of G, combining results for children to find results for parents.

A key to the complexity of the algorithms is the difficulty in combining information about children to obtain information about a parent. In Matula's algorithm, a tree node represents only a single node, so that the mappings for children can be combined in the obvious way. When applied to a tree-decomposition, however, a single node of a tree-decomposition is labeled by a "bag" corresponding to several nodes in the underlying graph; mappings for children must be consistent with respect to these nodes. Moreover, whereas each child subtree in Matula's algorithm represents a single, connected child subtree, a subtree of a tree-decomposition may correspond to many disjoint pieces of the underlying graph. Determining which piece of the graph is represented by which subtree of the tree-decomposition can lead to a combinatorial explosion.

To handle this problem, we rely on the restrictions to the input graph G. When G has bounded degree, the number of possible subgraphs of G is bounded, as is the number of subtrees to which they are assigned. In the case in which G is k-connected, each subgraph of the tree-decomposition of G corresponds to a connected piece of G, again simplifying the problem, yielding the following results:

Theorem 3.1. *Let G and H be partial k-trees for k greater than zero, and let G be k-connected. Let $n = |V(G)|+|V(H)|$. Then there is an $O(n^{k+4.5})$ time algorithm to determine whether or not G is isomorphic to a subgraph of H.*

Theorem 3.2. *Let G and H be partial k-trees for k greater than zero, and let G be of degree at most $d > 0$. Let $n = |V(G)| + |V(H)|$. Then there is an $O(n^{k+2})$ time algorithm to determine whether or not G is isomorphic to a subgraph of H.*

4 General properties of the reductions

In the remainder of this paper, we establish the complexity of subgraph isomorphism for G and H both partial k-paths of varying connectivities. We prove the following theorem:

Theorem 4.1. *Let G and H be partial k-trees with G g-connected and H h-connected. Then the problem of determining if H contains a subgraph isomorphic to G can be solved in polynomial-time when $g = k$ and is $\mathcal{NP}$-complete otherwise.*

The polynomial-time algorithm follows from the discussion in Section 3. The $\mathcal{NP}$-completeness results follow from Lemmas 5.7 and 6.3. As a byproduct of our constructions in the proofs of these lemmas, we also obtain the following result.

Theorem 4.2. *Let G and H be partial k-trees with G and H having all but k nodes of degree at most $k+2$. Then the problem of determining if H contains a subgraph isomorphic to G is $\mathcal{NP}$-complete.*

For the remainder of this section, we build appropriate machinery to prove Lemmas 5.7 and 6.3.

It is not difficult to see that the subgraph isomorphism problem is in $\mathcal{NP}$; we only need prove that the appropriate problems are $\mathcal{NP}$-hard. Our reductions will make use of the $\mathcal{NP}$-complete problem 3-Partition [GJ79]:

INSTANCE: A finite set $\mathcal{I} = \{I_1, I_2, \ldots, I_{3m}\}$ of positive integers and a positive integer B such that $\sum_j I_j = mB$ and for each j, $B/4 < I_j < B/2$.

QUESTION: Can $\mathcal{I}$ be partitioned into m disjoint subsets $\{C_1, C_2, \ldots, C_m\}$ (of 3 elements each) such that for $1 \le i \le m$, $\sum_{I \in C_i} I = B$?

Recall that 3-Partition is strongly $\mathcal{NP}$-complete so we can assume that the values I_j for a specific instance are specified in unary. This fact will be crucial to our reductions as the graphs G and H we will construct will each have size $O(mB)$. For the remainder of this section, we will specify an instance of 3-Partition by a tuple $(I_1, \ldots, I_{3m}; B)$. Without loss of generality we will also assume that $I_j > 1$ for all j.

We now delineate classes of graphs which will be used in our reductions.

Definition: A *j-path spiral of length ℓ* is a graph consisting of the following:

1. a total of $(j-1)+\ell$ nodes, where $c_1, \ldots, c_{j-1}$ are *center nodes* and $b_1, \ldots b_\ell$ are *exterior nodes* (in order), with b_1 and b_ℓ the *first* and *last* exterior nodes, respectively;

2. edges between each pair of center nodes, forming a clique of *center edges*;

3. edges between each exterior node and all the center nodes, forming the set of *radial edges*; and

4. edges between b_i and b_{i-1} for $i > 1$, forming the set of *exterior edges.*

We can consider the $j-1$ center nodes to be $j-1$ of the nodes in the original clique in the construction of a j-path, and the exterior nodes the added nodes (with the j^{th} node in the original clique being either the first exterior node or a node removed later to form a partial j-path).

Lemma 4.3. *A j-path spiral $\mathcal{S}$ of length ℓ is a j-connected partial j-path with a width-j path-decomposition (P, χ) such that one endpoint of P is labeled by the center nodes of $\mathcal{S}$.*

Proof: Let $\mathcal{S}$ be a j-path spiral of length ℓ with center nodes $c_1, \ldots, c_{j-1}$ and exterior nodes $b_1, \ldots, b_\ell$. To verify that G has width j, we construct a width j path-decomposition (P, χ) for $\mathcal{S}$. Let $P = w_0, w_1, \ldots, w_\ell$ and define χ by $\chi(w_0) = \{c_1, \ldots, c_{j-1}\}$, $\chi(w_1) = \{c_1, \ldots, c_{j-1}, b_1\}$, and for $i > 1$, $\chi(w_i) = \{c_1, \ldots, c_{j-1}, b_{i-1}, b_i\}$. It is straight-forward to verify that this is a path-decomposition of width j satisfying the required properties.

For the connectivity condition, notice that the removal of any set of $j-1$ nodes cannot disconnect G: if some center node c is not in the set then all the remaining exterior nodes are adjacent to c; if all $j-1$ center nodes are in the set then the remaining exterior nodes form a path. ■

We will also be working with star-like objects:

Definition: A *j-star* is a graph consisting of the following:

1. a total of $j+3$ nodes where $d_1, \ldots, d_j$ are the *clique nodes* and p_1, p_2, p_3 are the *pendant nodes*;

2. *clique edges* between each pair of clique nodes forming a clique of size j;

3. *pendant edges* between each pendant node and all clique nodes.

The following lemma is straight-forward to verify:

Lemma 4.4. *A j-star is a j-connected j-path with a width-j path-decomposition such that the label of one endpoint contains the clique nodes.*

The constructions in our reductions will involve forming unions of i-stars and j-paths of length ℓ:

Definition: Let $\mathcal{S}_1, \ldots, \mathcal{S}_r$ be $(j+1)$-path spirals of varying lengths. Let $\mathcal{K}$ be an i-star for $i \geq j$. Then, the *star-union* of $\mathcal{K}$ and $\mathcal{S}_1, \ldots, \mathcal{S}_r$ is the graph formed by taking the union of $\mathcal{S}_1, \ldots, \mathcal{S}_r$ and $\mathcal{K}$ in which the center nodes of each $\mathcal{S}_i$ are identified with the same j clique nodes of $\mathcal{K}$ (these j nodes are called the *identified nodes*).

5 Reduction for H not k-connected

In this section we begin by focusing on the proof of Lemma 5.7 for the case in which G has connectivity no greater than that of H. We then give a construction for handling the case in which G has greater connectivity than H. We start with some technical lemmas.

5.1 Technical Lemmas

We first show that the appropriate union of a star and a spiral gives the right connectivity.

Lemma 5.1. *Let G be the star-union of a $(j+1)$-path spiral $\mathcal{S}$ and an i-star $\mathcal{K}$ for $i \geq j$. Then G is j-connected.*

Proof: To see that G is j-connected notice that the graph G' obtained by the removal of any $j-1$ nodes from G leaves at least one center node c in G'; all nodes in G' are adjacent to c. ■

The next set of lemmas show that a star-union has the correct width.

Lemma 5.2. *Let G be a partial k-path and $v_1, \ldots, v_k$ be nodes of G such that there is a path-decomposition (P, χ) of G of width k with one endpoint of P labeled by $\{v_1, \ldots, v_k\}$. Let $\mathcal{S}$ be a $(j+1)$-path spiral ($k > j$) and G' be the union of G and $\mathcal{S}$ formed by identifying the j center nodes with any j nodes from among $v_1, \ldots, v_k$, say $v_1, \ldots, v_j$. Then G' has a width-k path-decomposition such that the label of one endpoint contains $\{v_1, \ldots, v_j\}$.*

Proof: We show how the path decomposition (P, χ) of G can be extended to a width k path-decomposition (P', χ') of G'. Suppose $\mathcal{S}$ is a $(j+1)$-path spiral of length ℓ with exterior nodes $b_1, \ldots, b_\ell$. We can assume without loss of generality that the union used to form G' identifies the nodes $\{v_1, \ldots, v_j\}$ with the center nodes of $\mathcal{S}$.

Suppose $P = w_1, \ldots, w_r$ such that $\chi(w_r) = \{v_1, \ldots, v_k\}$. By adjoining $\ell + 1$ new nodes to P, we construct from P a path $P' = w_1, \ldots, w_r, x_1, \ldots, x_{\ell+1}$. As well, we define $\chi'(w_i) = \chi(w_i)$, $1 \leq i \leq r$, $\chi'(x_1) = \{v_1, \ldots, v_j, b_1\}$, $\chi'(x_i) = \{v_1, \ldots, v_j, b_{i-1}, b_i\}$ for $2 \leq i \leq \ell$, and $\chi'(x_{\ell+1}) = \{v_1, \ldots, v_j\}$. From this construction it is straight-forward to verify that (P', χ') forms a path-decomposition of G' of width k. ∎

In the proof of Lemma 5.2, the path-decomposition constructed has the property that the label of one endpoint of the path contains all the identified nodes $\{v_1, \ldots, v_j\}$ (and consequently all of the center nodes of the spiral). Thus, by iterating this lemma with Lemma 4.3, we obtain the following corollary:

Corollary 5.3. *Let G be the star-union of a k-star and $(j+1)$-path spirals $\mathcal{S}_1, \ldots, \mathcal{S}_r$, for $k > j$. Then, G is a j-connected partial k-path.*

In the reductions, each of G and H will consist of star-unions. We will show that the clique nodes in G must map to the clique nodes in H, and that the ways in which the spirals in G map to the spirals in H dictate the partition of the items. Here we make a general observation which is the key to our reductions.

Lemma 5.4. *Let $(I_1, \ldots, I_{3m}; B)$ be an instance of 3-Partition. Let $\mathcal{K}$ and $\mathcal{K}'$ be k-stars with clique nodes $\{d_1, \ldots, d_k\}$ and $\{d'_1, \ldots, d'_k\}$ respectively. Let $\mathcal{S}_1, \ldots, \mathcal{S}_{3m}$ be a sequence of $(g+1)$-path spirals of lengths $I_1, \ldots, I_{3m}$ respectively, for $0 \leq g < k$. Let $\mathcal{T}_1, \ldots, \mathcal{T}_m$ be a sequence of $(h+1)$-path spirals all of length B, $g \leq h < k$. Let G be the star-union of $\mathcal{K}$ and $\mathcal{S}_1, \ldots, \mathcal{S}_{3m}$ and let H be the star-union of $\mathcal{K}'$ and $\mathcal{T}_1, \ldots, \mathcal{T}_m$. Then, G is a subgraph of H if and only if the instance of 3-Partition has a solution.*

Proof: We first assume that the instance $(I_1, \ldots, I_{3m}; B)$ of 3-Partition has a solution. It is then possible to partition the lengths $I_1, \ldots, I_{3m}$ (and hence the corresponding spirals) into 3-element sets, $C_1, \ldots, C_m$, such that the sum of the lengths in each set equals B. By the construction of G and H, $|V(G)| = |V(H)|$. We illustrate a bijection f from $V(G)$ to $V(H)$ such that for every edge $(u, v) \in E(G)$, the corresponding edge $(f(u), f(v))$ exists in $E(H)$. We assume without loss of generality that the identified nodes of G and H are $d_1, d_2, \ldots, d_g$ and $d'_1, d'_2, \ldots, d'_h$. For the nodes of $\mathcal{K}$ in G, we define f to map d_i to d'_i, so that the identified nodes of G map to identified nodes of H. The pendant nodes of $\mathcal{K}$ are allowed to map to the pendant nodes of $\mathcal{K}'$ in an arbitrary manner.

We need to show how f maps the spirals corresponding to elements in the set C_i to the spirals $\mathcal{T}_i$ for all i. Since the description is the same for all the cases, we only present the mapping for C_1 and $\mathcal{T}_1$. By the construction of G and H, the center nodes of the spirals in C_1 map to a subset of the center nodes of $\mathcal{T}_1$. Since there are B exterior nodes in the spirals in C_1 and B exterior nodes in $\mathcal{T}_1$, it is not difficult to see that there are enough exterior nodes to perform the mapping.

More specifically, assume without loss of generality that $C_1 = \{I_1, I_2, I_3\}$. Let the exterior nodes of $\mathcal{S}_i$ in order be $a_{i,j}$, $1 \leq i \leq 3$, $1 \leq j \leq I_i$ and let the exterior nodes of $\mathcal{T}_1$ in order be $b_1, \ldots, b_B$. Then, $f(a_{1,j}) = b_j$, $f(a_{2,j}) = b_{I_1+j}$ and $f(a_{3,j}) = b_{I_1+I_2+j}$. Clearly the exterior edges in G map to the exterior edges in H. Finally, since each center node in H is connected to each exterior node, the radial edges are preserved by the mapping.

Now suppose that G is isomorphic to a subgraph of H and let $f : V(G) \to V(H)$ be the subgraph isomorphism. Clearly f is a bijection since $|V(G)| = |V(H)|$; we first show that f maps $\mathcal{K}$ to $\mathcal{K}'$ such that identified nodes of $\mathcal{K}$ map to (a possible subset of the) identified nodes of $\mathcal{K}'$. All clique nodes in both G and H have degree at least $k+2$ whereas all other nodes have degree at most $k+1$, so the clique nodes of G must map to the clique nodes of H. As well, the identified nodes have degree $k-1+3+mB > k+2$ in both G and H, so among the clique nodes, the identified nodes must map to identified nodes. Finally, consider a pendant node p of G. If it maps to some exterior node v of H then an exterior neighbour w of v must have a pre-image which is adjacent to p; such a neighbour exists since $I_j > 1$ for all j. But p only has neighbours that are clique nodes, not exterior nodes.

Thus, we have shown that the exterior nodes in the spirals of G must map into exterior nodes in the spirals of H. We now show that the instance $(I_1, \ldots, I_{3m}; B)$ of 3-Partition has a solution. Since exterior edges must map to exterior edges, the images of the exterior nodes in a particular spiral in G must form a consecutive set of exterior nodes in H. Consequently, each of the spirals $\mathcal{S}_i$ must be contained entirely in some $\mathcal{T}_j$ for some j. Since f is a bijection, this forms a partition of $\{I_1, \ldots, I_{3m}\}$ into sets of size B each specified by the $\mathcal{T}_j$, as required. ■

5.2 G, H k-paths, G g-connected, H h-connected, $g \leq h < k$.

We are now ready to consider the case in which G has connectivity at most that of H.

Lemma 5.5. *Let G and H partial k-paths with G g-connected and H h-connected, $g \leq h < k$. Then the problem of determining if H contains a subgraph isomorphic to G is $\mathcal{NP}$-hard.*

Proof: We begin by constructing G and H. Let G be the star-union of a k-star and $(g+1)$-path spirals $\mathcal{S}_1, \ldots, \mathcal{S}_{3m}$ of lengths $I_1, \ldots, I_{3m}$ respectively. Let H be the star-union of a k-star and $(h+1)$-path spirals $\mathcal{T}_1, \ldots, \mathcal{T}_m$ of length B.

By Corollary 5.3, G and H are g- and h-connected partial k-paths. It follows from Lemma 5.4 that G is a subgraph of H if and only if the instance of 3-Partition has a solution, completing the proof. ■

From the construction of G and H it is clear that all but g nodes of G and h nodes of H have degree at most $k+2$, as needed for Theorem 4.2. Figure 1 illustrates Lemma 5.5 for the case in which $k = 2, g = 0$, and $h = 1$.

5.3 G, H k-paths, G g-connected, H h-connected, $h < g < k$.

We now show how our previous reductions can be modified to allow H to have lower connectivity than G. Our result is the following.

Lemma 5.6. *For G and H partial k-paths with G g-connected and H h-connected, $h < g < k$, the problem of determining if G is a subgraph of H is $\mathcal{NP} - hard$.*

Proof: For an instance $\mathcal{I} = \{I_1, \ldots, I_{3m}; B\}$ of 3-Partition, we begin by constructing graphs G and H' both having connectivity g as outlined in the proof of Lemma 5.5.

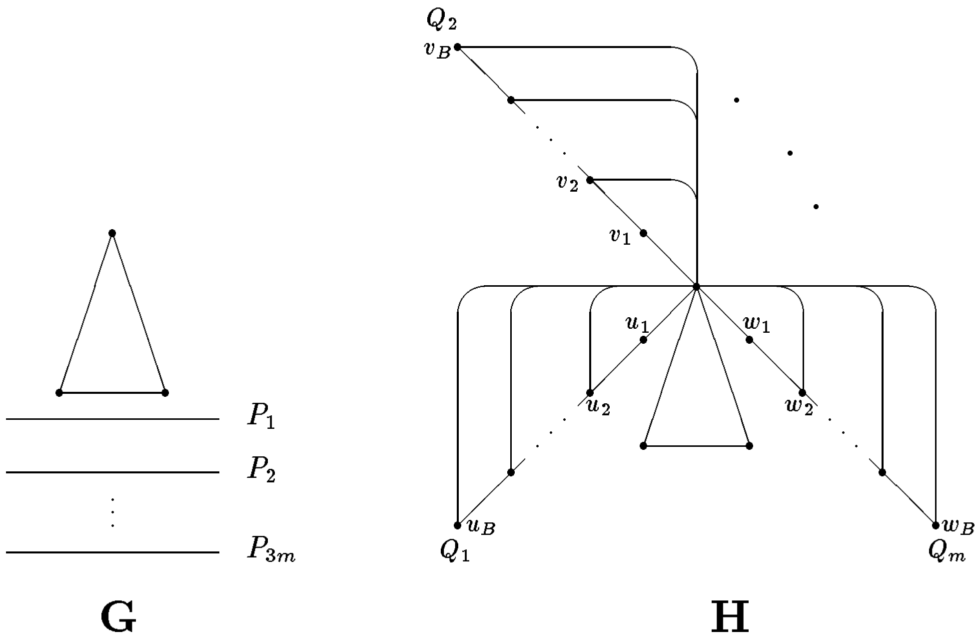

Figure 1: G consists of a 3-star and $3m$ disjoint paths $P_1, \ldots, P_{3m}$ of lengths $I_1, \ldots, I_{3m}$. H consists of a 3-star and m paths $Q_1, \ldots, Q_m$ of length B emanating from one node. Pendant vertices are not shown.

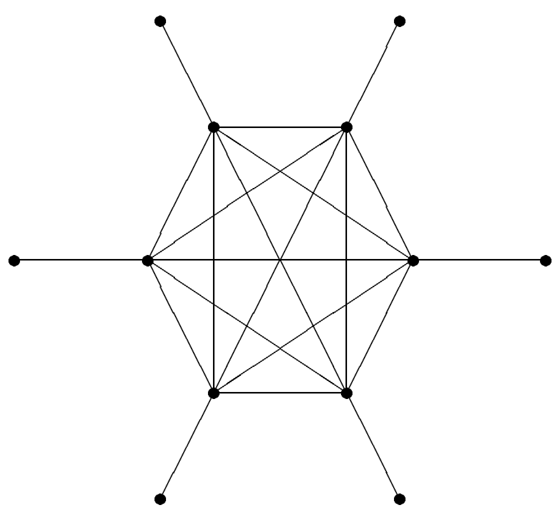

Figure 2: D_h is created by attaching a new node to each node of a K_h. A D_6 is illustrated.

We form the graph D_h by attaching h nodes of degree one, $w_1, \ldots, w_h$, to distinct nodes of a complete graph K_h, as illustrated in Figure 2. Notice that any h of the g identified nodes $\{d_1, \ldots, d_g\}$ in the star-union H' form an h-clique. We construct H to be the union of H' and D_h formed by identifying any h nodes of $\{d_1, \ldots, d_g\}$ with $w_1, \ldots, w_h$.

Next, we prove that H is an h-connected partial k-path. The graph H has connectivity h since each of H' and K_h have connectivity at least h and $\{w_1, \ldots, w_h\}$ forms a cutset in H. As well, since by Lemma 5.2 and Corollary 5.3 there is a width k path-decomposition (P, χ) of H' such that the label of one endpoint of P contains $\{d_1, \ldots, d_g\}$, this path-decomposition can easily be extended to a width k path-decomposition of H. Clearly H is an h-connected partial k-path.

Since H' is a subgraph of H, it follows immediately from Lemma 5.4 that if there is a solution to $\mathcal{I}$, then H (or more specifically, H') contains a subgraph isomorphic to G. It remains to be shown that if H contains a subgraph isomorphic to G then there is solution to the instance $\mathcal{I}$. To use Lemma 5.4, it will suffice to show that no subgraph of H containing unidentified nodes of D_h can be isomorphic to G. Suppose this was not the case and that G' was a subgraph of H containing unidentified nodes of D_h isomorphic to G. Clearly G would have connectivity at most h, which is less than g, yielding a contradiction. ∎

As a consequence of Lemmas 5.5 and 5.6, we can conclude that the following lemma holds:

Lemma 5.7. *Let G and H be partial k-paths with G g-connected and H h-connected, for g and h both less than k. Then the problem of determining if H contains a subgraph isomorphic to G is $\mathcal{NP}$-complete.*

6 Reduction for H k-connected

To handle the case in which H is k-connected, we use constructions similar to those appearing in the previous sections. However, when we form the union of a k-star with k-path spirals, the resulting graph has tree-width k, not path-width k.

We begin by stating a number of technical lemmas similar to those found in Section 5.1. The following definition is a generalization of star-unions but yields slightly higher connectivity.

Definition: Let $\mathcal{K}$ be a k-star with clique nodes $d_1, \ldots, d_k$. Let $\mathcal{S}_1, \ldots, \mathcal{S}_r$ be j-path spirals for $j \leq k$. Let $b_{1,i}$ be the first exterior node of $\mathcal{S}_i$. Let G be the star-union of $\mathcal{K}$ and $\mathcal{S}_1, \ldots, \mathcal{S}_r$. Without loss of generality, assume the nodes $d_1, \ldots, d_{j-1}$ are identified with the center nodes of the spirals in G. Then the *enhanced star-union* of $\mathcal{K}$ and $\mathcal{S}_1, \ldots, \mathcal{S}_r$ is the graph G with an additional edge from each $b_{1,i}$ to the node d_j.

Lemma 6.1. *Let G be the enhanced star-union of a k-star $\mathcal{K}$ and a j-path spiral $\mathcal{S}$ of length ℓ for $j \leq k$. Then G is a j-connected k-path and there is a path-decomposition of G with one endpoint labeled by the clique nodes of $\mathcal{K}$.*

Proof: To see that the connectivity condition holds, consider the graph G' obtained from G by the deletion of any $j-1$ nodes of G. If some identified node d is not among these nodes, then since all other nodes of G' are adjacent to d, we can conclude that G' is connected. Now suppose instead that all $j-1$ identified nodes are deleted to form G'. Then, d_j is not one of the $j-1$ deleted nodes; to show that G' is connected, it will suffice to show that all nodes in G' are connected to d_j. Clearly there are edges from all pendant nodes and from all remaining clique nodes to d_j. The exterior nodes in $\mathcal{S}$ form a path, with the first exterior node b_1 adjacent to d_j. Therefore, G' is connected.

To show that G is a k-path, we construct a path-decomposition (P, χ). We define the path P to be $w_0, w'_1, w'_2, w'_3, w_1, \ldots, w_\ell$, and set $\chi(w_0) = \{d_1, \ldots, d_k\}$, $\chi(w'_1) = \{d_1, \ldots, d_k, p_1\}$, $\chi(w'_2) = \{d_1, \ldots, d_k, p_2\}$, $\chi(w'_3) = \{d_1, \ldots, d_k, p_3\}$, $\chi(w_1) = \{d_1, \ldots, d_k, b_1\}$ and for $i > 1$, $\chi(w_i) = \{d_1, \ldots, d_{k-1}, b_{i-1}, b_i\}$. It is straight-forward to verify that (P, χ) is the required path-decomposition. ■

We can now join together path-decompositions of enhanced star-unions to obtain tree-decompositions of enhanced star-unions when there are multiple spirals.

Lemma 6.2. *Let G be the enhanced star-union of a k-star $\mathcal{K}$ and j-path spirals $\mathcal{S}_1, \ldots, \mathcal{S}_r$, $j \leq k$. Then, G is j-connected and has tree-width k.*

Proof: Let $d_1, \ldots, d_k$ be the clique nodes in $\mathcal{K}$. Let the graph G_i be the enhanced star-union of $\mathcal{K}$ with k-path spiral $\mathcal{S}_i$, $1 \leq i \leq r$ in which the identified nodes are $d_1, \ldots, d_{j-1}$. Then, by Lemma 6.1, we can form width-k path-decompositions $(P_1, \chi_1), \ldots, (P_r, \chi_r)$ of $G_1, \ldots, G_r$ respectively such that one endpoint of each path is labeled by $\{d_1, \ldots, d_k\}$. Now, consider the graph G formed by taking the enhanced union of $\mathcal{K}$ with $\mathcal{S}_1, \ldots, \mathcal{S}_r$ such that the identified nodes are $d_1, \ldots, d_{j-1}$ and d_j is adjacent to all first nodes of the spirals.

To see that G is j-connected, we need only show that the removal of any $j-1$ nodes results in a connected graph G'. The argument here is similar to that in Lemma 6.1. If some identified clique node remains in G' then all nodes in G' are adjacent to that node. If all $j-1$ identified nodes are deleted then the exterior nodes form paths each with one endpoint adjacent to d_j which is then adjacent to the remaining pendant and clique nodes.

We can form a width k tree-decomposition (T, χ) of G as follows: Let $P_i = w_{i,0}, \ldots, w_{i,j_i}$ such that $\chi_i(w_{i,0}) = \{d_1, \ldots, d_k\}$. Then, let T be the union of the P_i formed by identifying all $w_{i,0}$. As well, we define χ to be the union of the χ_i's. It is not difficult to verify that (T, χ) is a tree-decomposition with the appropriate properties. ■

Lemma 6.3. *Let G and H be partial k-trees with G g-connected and H k-connected, for g less than k. Then the problem of determining if H contains a subgraph isomorphic to G is $\mathcal{NP}$-hard.*

Proof: Let $(I_1, \ldots, I_{3m}; B)$ be an instance of 3-Partition. Let G be the enhanced star-union of a k-star $\mathcal{K}$ and g-spirals $\mathcal{S}_1, \ldots, \mathcal{S}_{3m}$ of length $I_1, \ldots, I_{3m}$ respectively with $d_1, \ldots, d_{g-1}$ the identified nodes and d_g the clique node adjacent to all first nodes in all $\mathcal{S}_i$. By Lemma 6.2, G is g-connected and has tree-width k. Let H be the enhanced star-union of a k-star $\mathcal{K}'$ and k-spirals $\mathcal{T}_1, \ldots, \mathcal{T}_m$ all of length B with $d'_1, \ldots, d'_{k-1}$ the identified nodes and d'_k the clique node adjacent to all first exterior nodes in all $\mathcal{T}_i$. By Lemma 6.2, H is k-connected and has tree-width k. We claim that $(I_1, \ldots, I_{3m}; B)$ has a solution if and only if G is a subgraph of H; the argument is very similar to that in Lemma 5.4 except that there is an additional edge from the first node of each spiral to the clique.

If there is a solution to $(I_1, \ldots, I_{3m}; B)$, then there is a partition $C_1, \ldots, C_m$ of $I_1, \ldots, I_{3m}$ each with three elements such that the sum of the elements in any C_i is B. Notice that C_i corresponds to a set of spirals of total length B. Then the mapping from G to H is specified as follows: $f(d_i) = d'_i$, pendant nodes in G map to pendant nodes in H in an arbitrary manner, and the exterior nodes of G in spirals associated with C_i appear as a contiguous set of nodes in some spiral of H.

For the converse, suppose G is isomorphic to a subgraph of H and let $f : V(G) \to V(H)$ be the subgraph isomorphism. Clearly f is a bijection since $|V(G)| = |V(H)|$; we wish to conclude that exterior nodes of G map to exterior nodes of H.

Notice that identified nodes in G have degree $k - 1 + 3 + mB$, the node d_g has degree $k - 1 + 3 + 3m$, all other clique nodes have degree $k + 2$, and all exterior nodes have degree at most $k + 1$. In H, identified nodes have degree $k - 1 + 3 + mB$, the node d'_k has degree $k - 1 + 3 + m$, and all exterior nodes have degree at most $k + 1$. Therefore, simply because of degree constraints, for $1 \leq i \leq g$, $f(d_i) \in \{d'_1, \ldots, d'_{k-1}\}$; without loss of generality assume that $f(d_i) = d'_i$. Then it follows, again from degree constraints, that for $g < i \leq k$, clique nodes d_i must map to clique nodes d'_j for $j > g$. From this we conclude that clique nodes in G map to clique nodes in H. As well, pendant nodes of G must map to pendant nodes of H; if pendant node p of G maps to an exterior node w of H, then the pre-image of an exterior neighbour v of w does not exist in G. Therefore, exterior nodes of G map to exterior nodes of H.

Now, the image under f of the exterior nodes of a particular spiral in G must form a consecutive set of exterior nodes in H. Consequently, each of the spirals $\mathcal{S}_i$ must be contained entirely in $\mathcal{T}_j$ for some j. Since f is a bijection, this forms a partition of $\{I_1, \ldots, I_{3m}\}$ into sets of size B each specified by the $\mathcal{T}_j$, as required. ∎

Again, we can notice that our constructions above entail at most k nodes having unbounded degree; the remainder of the nodes have degree at most $k + 2$. Thus, in this case also Theorem 4.2 holds.

7 Conclusions and directions for further research

In this paper we have shown that the subgraph isomorphism problem for partial k-trees is $\mathcal{NP}$-complete when either of the following occurs:

1. The source graph is not k-connected.

2. There are at least k vertices of unbounded degree.

Our proofs work by showing that when both input graphs have connectivity less than k, the resulting problem is $\mathcal{NP}$-complete even on partial k-paths. This leaves open the following problem: Suppose G and H are graphs of path-width k with H k-connected. What is the complexity of determining whether G is a subgraph of H?

A second open problem is that of determining the minimum number of unbounded degree nodes which makes these problems $\mathcal{NP}$-complete. In particular, is it sufficient to have only one unbounded degree node to ensure $\mathcal{NP}$-completeness?

Since partial k-paths are a subclass of partial k-trees and since topological embedding and minor containment are generalizations of subgraph isomorphism, our results immediately imply the $\mathcal{NP}$-completeness of these problems for these classes of graphs. It would be interesting to know whether or not there are other problems for which the connectivity and degree of k-paths or k-trees yields such a fine distinction in complexity, or for which other possible restrictions produce duality results.

8 Acknowledgements

We thank Derek Corneil for suggesting the use of partial k-paths instead of partial k-trees for studying these problems, and anonymous referees for pointing out an error in an earlier version of this paper.

References

[ACP89] S. ARNBORG, D. CORNEIL, AND A. PROSKUROWSKI, Complexity of finding embeddings in a k-tree, *SIAM Journal of Algebraic and Discrete Methods* **8** (1987), pp. 277-284.

[BGHK91] H. L. BODLAENDER, J. GILBERT, H. HAFSTEINSSON, and T. KLOKS, Approximating treewidth and pathwidth of some classes of perfect graphs, *Proceedings of the 17th International Workshop on Graph-Theoretic Concepts in Computer Science*, pp. 1–12, 1991.

[BK91] H. BODLAENDER and T. KLOKS, Better algorithms for the pathwidth and treewidth of graphs, *Proceedings of the 18th International Colloquium on Automata, Languages, and Programming*, pp. 544–555, 1991.

[BKK93] H. BODLAENDER, T. KLOKS, and D. KRATSCH, Treewidth and pathwidth of permutation graphs, *Proceedings of the 20th International Colloquium on Automata, Languages, and Programming*, pp. 114–125, 1993.

[BM90] H. L. BODLAENDER and R. H. MÖHRING, The pathwidth and treewidth of cographs, *Proceedings of the 2nd Scandinavian Workshop on Algorithm Theory*, pp. 301–309, 1990.

[BM76] J. BONDY, and U.S.R. MURTY, "Graph Theory with Applications," North-Holland, 1976.

[GJ79] M. R. GAREY and D. S. JOHNSON, "Computers and Intractability: A Guide to the Theory of NP-completeness," Freeman, San Francisco, 1979.

[GN94] A. GUPTA and N. NISHIMURA, Sequential and parallel algorithms for embedding problems on classes of partial k-trees, *Proceedings of the Fourth Scandinavian Workshop on Algorithm Theory*, pp. 172–182, 1994.

[Kin92] N. KINNERSLEY, The vertex separation number of a graph equals its path-width, *Information Processing Letters* **42** (1992), pp. 345–350.

[KB92] T. KLOKS and H. BODLAENDER, Approximating treewidth and pathwidth of some classes of perfect graphs, *Proceedings of the 3rd Annual International Symposium on Algorithms and Computation*, pp. 116–125, 1992.

[KS93] E. KORACH and N. SOLEL, Tree-width, path-width, and cutwidth, *Discrete Applied Mathematics* **43** (1993), pp. 97–101.

[Lin89] A. LINGAS, Subgraph isomorphism for biconnected outerplanar graphs in cubic time, *Theoretical Computer Science* **63** (1989), pp. 295–302.

[LS88] A. LINGAS and M. M. SYSLO, A polynomial-time algorithm for subgraph isomorphism of two-connected series parallel graphs, *Proceedings of the 15th International Colloquium on Automata, Languages, and Programming* pp. 394–409, 1988.

[MT92] J. MATOUŠEK and R. THOMAS, On the complexity of finding iso- and other morphisms for partial k-trees, *Discrete Mathematics* **108** (1992), pp. 343–364.

[Mat78] D. MATULA, Subtree isomorphism in $O(n^{5/2})$, *Annals of Discrete Mathematics* **2** (1978), pp. 91–106.

[Moh90] R. H. MÖHRING, Graph problems reated to gate matix layout and PLA folding, in *Computational Graph Theory*, editors Tinhofer, Mayr, Noltemeier and Syslo, pp. 17–51, 1990.

[Ros74] D.J. ROSE, On simple characterization of k-trees, *Discrete Mathematics* **7** (1974), pp. 317–322.

[RS86] N. ROBERTSON and P. SEYMOUR, Graph Minors II. Algorithm aspects of tree-width, *Journal of Algorithms* **7** (1986), pp. 309–322.

[Sys82] M. M. SYSLO, The subgraph isomorphism problem for outerplanar graphs, *Theoretical Computer Science* **17** (1982), pp. 91–97.

[TUK91] A. TAKAHASHI, S. UENO, and Y. KAJITANI, Mixed-searching and proper-path-width, *Proceedings of the International Symposium on Algorithms*, pp. 61–71, 1991.