ARTICLE IN PRESS

European Journal of Combinatorics ▮ (▮▮▮▮) ▮▮▮–▮▮▮

Contents lists available at ScienceDirect

European Journal of Combinatorics

journal homepage: www.elsevier.com/locate/ejc

On maximum common subgraph problems in series–parallel graphs[☆,☆☆]

Nils Kriege, Florian Kurpicz, Petra Mutzel

Technische Universität Dortmund, Department of Computer Science, Chair of Algorithm Engineering (LS11), 44221 Dortmund, Germany

ARTICLE INFO

Article history:
Available online xxxx

ABSTRACT

The complexity of the maximum common connected subgraph problem in partial k-trees is still not fully understood. Polynomial-time solutions are known for degree-bounded outerplanar graphs, a subclass of the partial 2-trees. On the other hand, the problem is known to be **NP**-hard in vertex-labeled partial 11-trees of bounded degree. We consider series–parallel graphs, i.e., partial 2-trees. We show that the problem remains **NP**-hard in biconnected series–parallel graphs with all but one vertex of degree 3 or less. A positive complexity result is presented for a related problem of high practical relevance which asks for a maximum common connected subgraph that preserves blocks and bridges of the input graphs. We present a polynomial time algorithm for this problem in series–parallel graphs, which utilizes a combination of BC- and SP-tree data structures to decompose both graphs.

1. Introduction

Finding a maximum common connected subgraph (MCS) of two input graphs is an important task in many application domains like pattern recognition and cheminformatics [21]. The problem is well known to be **NP**-hard. However, practically relevant graphs, e.g., derived from small molecules, often have small treewidth [11]. Hence, it is highly relevant to develop polynomial time algorithms for tractable graph classes and to clearly identify graph classes where MCS remains **NP**-hard. For

☆ An extended abstract of this work appeared in Kriege et al. (2014) [12] .

☆☆ This work was supported by the German Research Foundation (DFG), priority program "Algorithms for Big Data" (SPP 1736).

E-mail addresses: nils.kriege@tu-dortmund.de (N. Kriege), florian.kurpicz@tu-dortmund.de (F. Kurpicz), petra.mutzel@tu-dortmund.de (P. Mutzel).

http://dx.doi.org/10.1016/j.ejc.2017.07.012
0195-6698/

Please cite this article in press as: N. Kriege, et al., On maximum common subgraph problems in series–parallel graphs, European Journal of Combinatorics (2017), http://dx.doi.org/10.1016/j.ejc.2017.07.012.

the related subgraph isomorphism problem such a clear demarcation for partial k-trees is known. Subgraph isomorphism is solvable in polynomial time in partial k-trees if the smaller graph either is k-connected or has bounded degree [8,15]. However, it is **NP**-complete when the smaller graph is not k-connected or has more than k vertices of unbounded degree [9]. MCS is at least as hard as subgraph isomorphism; two recent results show that it actually is considerably harder: Akutsu [2] has shown that MCS is **NP**-hard in vertex-labeled partial 11-trees of bounded degree. Furthermore, it was believed that the problem of finding a maximum common k-connected subgraph of k-connected partial k-trees (k-MCS) can be solved with the same technique that was successfully used for subgraph isomorphism. Recently, it was shown that these techniques are insufficient even for series–parallel graphs [13]. However, for this class of graphs a new approach was devised, which employs SP-trees to represent the series–parallel composition of the input graphs. Further polynomial time algorithms were proposed for MCS in almost trees and outerplanar graphs of bounded degree [1,3].

Motivated by the fact that even subgraph isomorphism is **NP**-hard when the smaller graph is a tree and the other is outerplanar [22], a problem variation referred to as block-and-bridge preserving MCS (BBP-MCS) was considered [19–21]. Here, the common connected subgraph is required to inherit the structure of blocks, i.e., maximal biconnected subgraphs, and bridges of the input graphs, which renders efficient algorithms for outerplanar graphs possible [19]. Notably, BBP-MCS yields meaningful results for molecular graphs in practice and even compares favorably to the solutions obtained by ordinary MCS in empirical studies [18,21].

Our contribution.

On the theoretical side, we prove that finding an MCS of two biconnected series–parallel graphs, i.e., partial 2-trees [4], with all but one vertex of degree bounded by 3 is **NP**-hard. We obtain this result by a polynomial-time reduction of the *Numerical Matching with Target Sums* problem. Furthermore, we consider BBP-MCS in series–parallel graphs and propose a polynomial time solution, thus, generalizing the known result for outerplanar graphs. Employing BC- and SP-tree decompositions of the input graphs allows us to identify subproblems closely related to k-MCS, $k \in \{1, 2\}$. We make use of a classical approach for the maximum common subtree problem [16], i.e., 1-MCS, and a recently proposed algorithm for 2-MCS [13] to obtain our result. Our approach yields a running time of $\mathcal{O}(n^6)$ in series–parallel and $\mathcal{O}(n^5)$ in outerplanar graphs, where n is the maximum number of vertices in one of the input graphs.

2. Preliminaries

We consider *simple* graphs, i.e., a graph G without loops and multiple edges. We denote the finite set of *vertices* by $V(G)$ and the finite set of *edges* by $E(G)$. A graph G' is a *subgraph* of G, denoted by $G' \subseteq G$, if $V(G') \subseteq V(G)$ and $E(G') \subseteq E(G)$. A subgraph $G' \subseteq G$ is said to be *proper* if $G' \neq G$ and we write $G' \subset G$. A subgraph $G \subseteq H$ is called *maximal* regarding a property if G itself has the property and there is no graph G' which has the property and satisfies $G \subset G' \subseteq H$. For two graphs $G = (V, E)$ and $G' = (V', E')$, we denote by $G \cup G'$ the graph $(V \cup V', E \cup E')$. For short, we write $G \cup \{v\}$ and $G \cup \{e\}$ to denote the union with a graph consisting of a single vertex v and a single edge e with its two endpoints, respectively. A graph is *connected* if there is a path between any two vertices. Each maximal connected subgraph $G' \subseteq G$ is called a *connected component*. Let $V \subseteq V(G)$, then $G[V]$ denotes the *induced subgraph* $G' \subseteq G$ with $V(G') = V$ and $E(G') = \{(u, v) \in V \times V : (u, v) \in E(G)\}$. A set $S \subseteq V(G)$ is called $|S|$-*separator* or *separator* of a connected graph G if $G \setminus S := G[V(G) \setminus S]$ consists of at least two connected components. If $S = \{v\}$ is a separator then v is called *cutvertex*. A separator S is said to *separate* two vertices $a, b \in V(G)$ if a and b are in different connected components of $G \setminus S$. A separator S of G is called *minimal* if there are vertices $a, b \in V(G)$ that are separated by S and there is no separator $S' \subset S$ that separates a and b. A graph G with $|V(G)| > k$ is called *k-connected* if there is no j-separator of G such that $j < k$ and *biconnected* if $k = 2$. We define $[n] := \{1, \ldots, n\}$ for all $n \in \mathbb{N}$. A sequence of distinct vertices $(v_0, v_1, \ldots, v_n)$ such that $(v_{i-1}, v_i) \in E(G)$ for all $i \in [n]$ is called *path*. The vertices and the edges connecting consecutive vertices are said to be *contained* in the path. If all but the first and the last vertex are distinct, i.e., $v_n = v_0$, the sequence is called *cycle*. The *length* of a path or cycle is