Discrete Applied Mathematics 180 (2015) 89–100

Contents lists available at ScienceDirect

Discrete Applied Mathematics

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

Complexity of splits reconstruction for low-degree trees☆

Serge Gaspers [a,b,*], Mathieu Liedloff [c], Maya Stein [d], Karol Suchan [e,f]

[a] School of Computer Science and Engineering, UNSW Australia (The University of New South Wales), Sydney NSW 2052, Australia
[b] Optimisation Research Group, NICTA (National ICT Australia), Sydney NSW 2052, Australia
[c] Université d'Orléans, INSA Centre Val de Loire, LIFO EA 4022, FR-45067 Orléans, France
[d] CMM, Universidad de Chile, Santiago, Chile
[e] FIC, Universidad Adolfo Ibáñez, Santiago, Chile
[f] WMS, AGH – University of Science and Technology, Krakow, Poland

ARTICLE INFO

Article history:
Received 24 November 2012
Received in revised form 30 July 2014
Accepted 4 August 2014
Available online 28 August 2014

Keywords:
Reconstruction of trees
Computational complexity
Computational chemistry

ABSTRACT

Given a vertex-weighted tree T, the split of an edge e in T is the minimum over the weights of the two trees obtained by removing e from T, where the weight of a tree is the sum of weights of its vertices. Given a set of weighted vertices V and a multiset of integers $\mathcal{S}$, we consider the problem of constructing a tree on V whose splits correspond to $\mathcal{S}$. The problem is known to be NP-complete, even when all vertices have unit weight and the maximum vertex degree of T is required to be at most 4. We show that

- the problem is strongly NP-complete when T is required to be a path,
- the problem is NP-complete when all vertices have unit weight and the maximum degree of T is required to be at most 3, and
- it remains NP-complete when all vertices have unit weight and T is required to be a caterpillar with unbounded hair length and maximum degree at most 3.

We also design polynomial time algorithms for

- the variant where T is required to be a path and the number of distinct vertex weights is constant, and
- the variant where all vertices have unit weight and T has a constant number of leaves.

The latter algorithm is not only polynomial when the number of leaves, k, is a constant, but also is a fixed-parameter algorithm for parameter k.

Finally, we shortly discuss the problem when the vertex weights are not given but can be freely chosen by an algorithm.

The considered problem is related to building libraries of chemical compounds used for drug design and discovery. In these inverse problems, the goal is to generate chemical compounds having desired structural properties, as there is a strong relation between structural invariants of the particles, such as the Wiener index and, less directly, the problem under consideration here, and physico-chemical properties of the substance.

☆ A preliminary version of this article appeared in the proceedings of WG 2011 [11].

* Corresponding author at: School of Computer Science and Engineering, UNSW Australia (The University of New South Wales), Sydney NSW 2052, Australia. Tel.: +61 2 9385 6934; fax: +61 2 9385 5995.

E-mail addresses: sergeg@cse.unsw.edu.au (S. Gaspers), mathieu.liedloff@univ-orleans.fr (M. Liedloff), mstein@dim.uchile.cl (M. Stein), karol.suchan@uai.cl (K. Suchan).

http://dx.doi.org/10.1016/j.dam.2014.08.005
0166-218X/

1. Introduction

In this paper, we consider trees $T = (V, E)$ where integer weights are associated to vertices by a function $\omega : V \to \mathbb{N}$, where $\mathbb{N}$ denotes the set of natural numbers excluding 0.

Definition 1. Let T be a tree and $\omega : V \to \mathbb{N}$ be a function. The *split* of an edge e in T is the minimum of $\omega(T_1)$ and $\omega(T_2)$, where T_1 and T_2 are the two trees obtained by deleting e from T, and $\omega(T_i) = \sum_{v \in T_i} \omega(v)$.

We use $\mathcal{S}(T)$ to denote the multiset of splits of T.

We consider the problem of reconstructing a tree with a given multiset of splits and a given set of weighted vertices.

> Weighted Splits Reconstruction (WSR): Given a set V of n vertices, a weight function $\omega : V \to \mathbb{N}$, and a multiset $\mathcal{S}$ of integers, is there a tree T on V whose multiset of splits is $\mathcal{S}$ (that is, $\mathcal{S}(T) = \mathcal{S}$)?

The Weighted Splits Reconstruction for Trees of Maximum Degree k problem (WSR_k) is defined in the same way, except that we restrict the tree T to have maximum degree at most k. When we require T to belong to a subclass of trees $\mathcal{T}$, the problem is called Weighted Splits Reconstruction for $\mathcal{T}$.

When ω assigns unit weights to the vertices, the problem is simply called Splits Reconstruction (SR). The Splits Reconstruction for Trees of Maximum Degree k problem (SR_k) and the Splits Reconstruction for $\mathcal{T}$ are the obvious unweighted counterparts of the weighted variants defined above.

Related work. In the field of Chemical Graph Theory [2,3,20], molecules are modeled by graphs in order to study the physical properties of chemical compounds. A chemical graph is a graph, where vertices represent atoms of a chemical compound and edges the chemical bonds between them. Within the area of quantitative structure–activity relationship (QSAR), several structural measures of chemical graphs were identified that quantitatively correlate with a well-defined process, such as biological activity or chemical reactivity. Probably the most widely known example is the *Wiener index* (see [14]): the sum of the distances in a graph between each pair of vertices, where the distance between two vertices is the length (the number of edges) of a shortest path from one to the other. Wiener [23] found a strong correlation between the boiling points of paraffins and the Wiener index. From then on, many other topological (using the information of the chemical graph) and topographical (using the information of the chemical graph and the location of its vertices in space) indices were introduced and their correlation with various other properties was investigated.

In Combinatorial Chemistry, drug design is facilitated by building libraries of molecules that are structurally related (via the Wiener index or any of the other numerous indices). We face inverse problems where the goal is to design new compounds that have a prescribed structural information (see also [6]).

Goldman et al. [13] study problems related to the design of combinatorial libraries for drug design from an algorithmic and complexity-theoretic point of view, following the heuristic approaches of [19] and [12]. Goldman et al. show that for every positive integer W, except 2 and 5, there exists a graph with Wiener index W. For constructing a tree (of unbounded or bounded maximum degree) with a given Wiener index, they devise pseudo-polynomial dynamic programming algorithms. Goldman et al. also introduce the Splits Reconstruction problem and recall a result due to Wiener [23]: the Wiener index of a tree T on n vertices with unit weights is $\sum_{s \in \mathcal{S}(T)} s \cdot (n - s)$. They show that SR is NP-complete and give an exponential-time algorithm without running time analysis. Independently, Wagner [21] and Wang et al. [22] show that all but a finite number of integers are Wiener indices of trees.

As it is not reasonable to construct chemical trees with arbitrarily high vertex degrees, Li and Zhang [17] studied SR_4 and showed that it is also NP-complete. Their algorithm to construct a tree with maximum degree at most 4 to solve SR_4 runs in exponential time (no running time analysis is provided) and creates weighted vertices in intermediate steps.

In order to reconstruct glycans or carbohydrate sugar chains, Aoki-Kinoshita et al. [1] study the reconstruction of a node-labeled supertree from a set of node-labeled subtrees. They give a 6-approximation algorithm for this problem, which generalizes the smallest superstring problem.

We refer to [4] surveying results on the Wiener index for trees.

Our results. By the result of Li and Zhang [17], SR_4 is NP-complete, while SR_2 is trivially in P. We close this gap by showing that SR_3 is NP-complete by a reduction from Numerical Matching with Target Sums (defined below). It is also NP-complete for caterpillars with unbounded hair length. Identifying small classes of trees for which the problem is NP-complete may be important for future investigations in the spirit of the deconstruction of hardness proofs [16] which aim at identifying parameters for which the problem becomes tractable when these parameters are small.

Recall that a problem is strongly NP-complete if it remains so even when all of its numerical parameters are bounded by a polynomial in the length of the input. Our main result proves that WSR_2 is strongly NP-complete by a reduction from a variant of Numerical Matching with Target Sums in which all integers of the input are distinct. For the case where the weights of the vertices are chosen from a small set of values, our dynamic-programming algorithm solves WSR_2 in time $O(n^{k+3} \cdot k)$, where k is the number of distinct vertex weights. Although this running time is polynomial for every constant k, the degree of the polynomial depends on k. Thus, the running time becomes impractical, even for small values of k.

Multivariate complexity theory [5,7,9,18] – also known as parameterized complexity – is a theoretical framework that allows to distinguish between running times of the form $f(k)n^{g(k)}$ where the degree of the polynomial depends on the