

On tree representations of relations and graphs: symbolic ultrametrics and cograph edge decompositions

Marc Hellmuth & Nicolas Wieseke

**Journal of Combinatorial
Optimization**

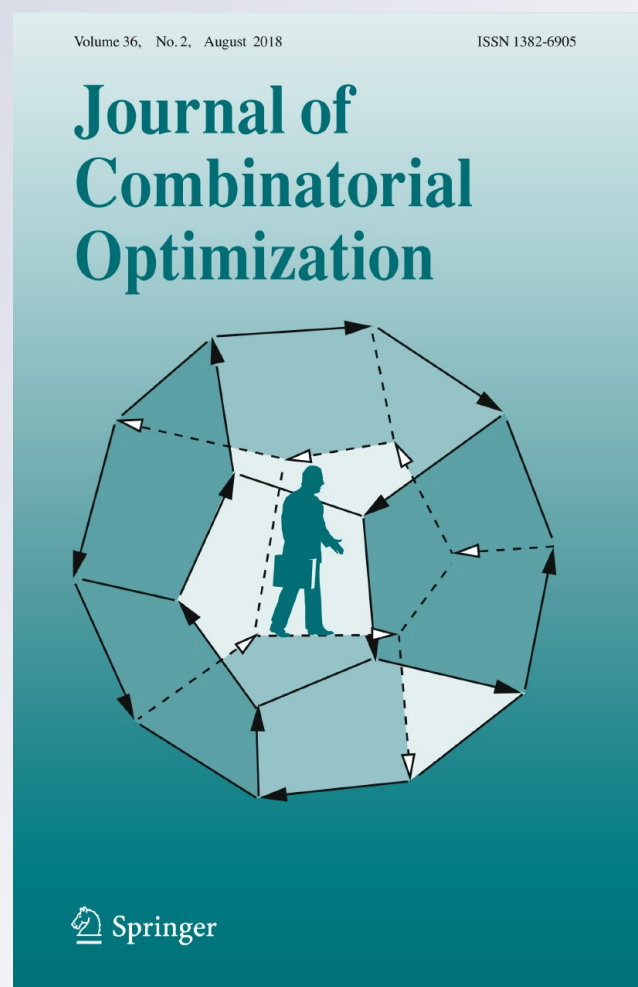
ISSN 1382-6905

Volume 36

Number 2

J Comb Optim (2018) 36:591-616

DOI 10.1007/s10878-017-0111-7



Your article is protected by copyright and all rights are held exclusively by Springer Science +Business Media New York. This e-offprint is for personal use only and shall not be self-archived in electronic repositories. If you wish to self-archive your article, please use the accepted manuscript version for posting on your own website. You may further deposit the accepted manuscript version in any repository, provided it is only made publicly available 12 months after official publication or later and provided acknowledgement is given to the original source of publication and a link is inserted to the published article on Springer's website. The link must be accompanied by the following text: "The final publication is available at link.springer.com".

On tree representations of relations and graphs: symbolic ultrametrics and cograph edge decompositions

Marc Hellmuth^{1,2} · Nicolas Wieseke³

Published online: 25 January 2017
 © Springer Science+Business Media New York 2017

Abstract Tree representations of (sets of) symmetric binary relations, or equivalently edge-colored undirected graphs, are of central interest, e.g. in phylogenomics. In this context symbolic ultrametrics play a crucial role. Symbolic ultrametrics define an edge-colored complete graph that allows to represent the topology of this graph as a vertex-colored tree. Here, we are interested in the structure and the complexity of certain combinatorial problems resulting from considerations based on symbolic ultrametrics, and on algorithms to solve them. This includes, the characterization of symbolic ultrametrics that additionally distinguishes between edges and non-edges of *arbitrary* edge-colored graphs G and thus, yielding a tree representation of G , by means of so-called cographs. Moreover, we address the problem of finding “closest” symbolic ultrametrics and show the NP-completeness of the three problems: symbolic ultrametric editing, completion and deletion. Finally, as not all graphs are cographs, and hence, do not have a tree representation, we ask, furthermore, what is the minimum number of cotrees needed to represent the topology of an arbitrary non-cograph G .

Parts of this paper were presented at the 21st Annual International Computing and Combinatorics Conference (COCOON 2015), August 4–6, 2015, Beijing, China (Hellmuth and Wieseke 2015). This work was funded by the German Research Foundation (DFG) (Proj. No. MI439/14-1).

✉ Marc Hellmuth
 mhellmuth@mailbox.org

Nicolas Wieseke
 wieseke@informatik.uni-leipzig.de

¹ Department of Mathematics and Computer Science, University of Greifswald,
 Walther-Rathenau-Strasse 47, 17487 Greifswald, Germany

² Center for Bioinformatics, Saarland University, Building E 2.1, 66041 Saarbrücken, Germany

³ Parallel Computing and Complex Systems Group, Department of Computer Science, Leipzig University, Augustusplatz 10, 04109 Leipzig, Germany

This is equivalent to find an optimal cograph edge k -decomposition $\{E_1, \dots, E_k\}$ of E so that each subgraph (V, E_i) of G is a cograph. We investigate this problem in full detail, resulting in several new open problems, and NP-hardness results. For all optimization problems proven to be NP-hard we will provide integer linear program formulations to solve them.

Keywords Symbolic ultrametric · Cograph · Edge partition · Editing · Integer linear program (ILP) · NP-complete

1 Introduction

Tree representations of binary relations between certain objects, in particular symbolic ultrametrics, lie at the heart of many problems, in particular, in phylogenomic studies (Hellmuth et al. 2013a, 2015; Hellmuth and Wieseke 2016; Lafond and El-Mabrouk 2014). Phylogenetic Reconstructions are concerned with the study of the evolutionary history of groups of systematic biological units, e.g. genes or species. The objective is the assembling of so-called phylogenetic trees or networks that represent a hypothesis about the evolutionary ancestry of a set of genes, species or other taxa.

Genes are passed from generation to generation to the offspring. Some of those genes are frequently duplicated, mutate or get lost - a mechanism that also ensures that new species can evolve. Crucial for the evolutionary reconstruction of species history is the knowledge of the relationship between the respective genes. Genes that share a common origin (homologs) are divided into three classes, namely orthologs, paralogs, and xenologs (Fitch 2000). Two homologous genes are orthologous if at their most recent point of origin, the ancestral gene complement is transmitted to two daughter lineages, that is, a speciation event happened. They are paralogous if the ancestor gene at their most recent point of origin was duplicated within a single ancestral genome, that is, a duplication event happened. Horizontal gene transfer (HGT) refers to the transfer of genes between organisms in a manner other than traditional reproduction and across different species; if such an event happened at the most recent point of origin of two genes, then they are called xenologous. Intriguingly, there are practical sequence-based methods that allow us to estimate whether two genes x and y are orthologs or not *without* constructing either gene or species trees (Lechner et al. 2011, 2014).

Now, assume we have given an estimate of genes being orthologs, paralogs or xenologs, that is, a map $d : X \times X \rightarrow \{\text{speciation, duplication, HGT}\}$ where X denotes a set of genes. Then, one is interested in the representation of these estimates as a (gene) tree T with leaf-set X and event-labeling t so that the lowest common ancestor $\text{lca}(x, y)$ of distinct leaves x and y in T is labeled with speciation, duplication or HGT if the genes x and y are orthologs, paralogs or xenologs, respectively. Thus, one wants to determine an event-labeled tree with $t(\text{lca}(x, y)) = d(x, y)$ for all distinct $x, y \in X$. In practice, however, such maps d are often only estimates of the true evolutionary relationship δ between the investigated genes. Therefore, such a tree may not exist. Nevertheless, it has recently been shown, that in theory (Hernandez-Rosales et al. 2012) and in practice (Hellmuth et al. 2015) it is possible to reconstruct the evolutionary

history of species, where the genes have been taken from, whenever (estimates of) the map δ are known. We refer to [Hellmuth and Wieseke \(2016\)](#) for an overview.

We will consider arbitrary symmetric maps $\delta: X \times X \rightarrow M^\odot$ that assign to each pair (x, y) a symbol or color $m \in M^\odot$, where $M^\odot = M \cup \{\odot\}$ and \odot serves as a “non-event” symbol that is assigned to $x, y \in X$ iff $x = y$. Rephrasing the question above, we ask whether it is possible to determine a rooted tree T with a vertex-labeling t so that $t(\text{lca}(x, y)) = m \in M^\odot$ if and only if $\delta(x, y) = m$. Such a tree is then called *symbolic representation* or *tree representation* (T, t) of δ . A solution to this problem has been given by [Böcker and Dress \(1998\)](#). The authors showed that there is a symbolic representation (T, t) of δ if and only if the map δ fulfills the properties of a so-called symbolic ultrametric ([Böcker and Dress 1998](#)).

Any such map $\delta: X \times X \rightarrow M^\odot$ is equivalent to a set of disjoint symmetric binary relations $\{R_m \mid m \in M^\odot\}$ with $(x, y) \in R_{\delta(x,y)}$ or an edge-colored complete graph $K_{|X|} = (X, \binom{X}{2})$ so that each edge $[x, y]$ obtains the color $\delta(x, y) \in M^\odot$. In [Hellmuth et al. \(2013a\)](#), a characterization of symbolic ultrametrics by means of so-called cographs has been established. Cographs are characterized by the absence of induced paths P_4 on four vertices, although there are a number of equivalent characterizations of cographs (see e.g. [Brandstädt et al. 1999](#) for a survey). Moreover, [Lerchs \(1971a, b\)](#) showed that each cograph $G = (V, E)$ is associated with a unique rooted tree $T(G)$, called *cotree*.

In this contribution we are concerned with several combinatorial problems based on symbolic ultrametrics and tree representations of arbitrary, possibly edge-colored graphs.

We first investigate in Sect. 3, under what conditions it is possible to find a symbolic ultrametric for arbitrary graphs G so that edges and non-edges of G can be distinguished. In other words, we ask for an edge-coloring of G so that edges and non-edges always obtain different colors and this coloring satisfies the conditions of a symbolic ultrametric. If such a coloring is known for G , then one can immediately display the topology of G as a tree via a symbolic representation (T, t) . It does not come as a big surprise, when we prove that such a symbolic ultrametric can only be defined for G if and only if G is already a cograph. This, in particular, establishes another new characterization of cographs. As a consequence we can infer that any symbolic representation (T, t) of a cograph G is a so-called refinement of its cotree.

In practice, symmetric maps $d: X \times X \rightarrow M^\odot$ represent often only estimates of the true relationship δ between the investigated objects, e.g., genes ([Lechner et al. 2011, 2014](#)). Thus, in general such estimates d will not be a symbolic ultrametric. Hence, there is a great interest in optimally editing d to a symbolic ultrametric δ , i.e., finding a minimum number of changes of the assignment $d(x, y) \in M^\odot$ to pairs (x, y) so that there is a symbolic representation of the resulting map δ ([Hellmuth et al. 2015](#)). So far, the complexity of this problem has been unknown. In Sect. 4, we show that (the decision version of) this problem, called SYMBOLIC ULTRAMETRIC EDITING, is NP-complete. Additionally, we show that the problems SYMBOLIC ULTRAMETRIC COMPLETION and SYMBOLIC ULTRAMETRIC DELETION are NP-complete and provide integer linear program (ILP) formulations in order to (exactly) solve the latter three problems. In particular, solutions provided by the ILP can be used to benchmark solutions that are

computed with possibly faster heuristics, that still need to be established and that are not in the focus of this contribution and will be discussed elsewhere.

As a consequence of the results established in Sect. 3 and since not all graphs are cographs, there are graphs that do not have a tree representation. Therefore, we study the following combinatorial problem in Sect. 5: What is the minimum number of cotrees that are needed to represent the structure of a given graph $G = (V, E)$ in an unambiguous way? As it will turn out, this problem is equivalent to find a decomposition $\Pi = \{E_1, \dots, E_k\}$ of E (i.e., the elements of Π need not necessarily be disjoint) for the least integer k , so that each subgraph $G_i = (V, E_i)$, $1 \leq i \leq k$ is a cograph. Such a decomposition is called *cograph edge k -decomposition*, or *cograph k -decomposition*, for short. If the elements of Π are in addition pairwise disjoint, we call Π a *cograph k -partition*. We show that the number of such optimal cograph k -decomposition, resp., partitions on a graph can grow exponentially in the number of vertices. Moreover, non-trivial upper bounds for the integer k such that there is a cograph k -decomposition, resp., partition are derived. Polynomial-time algorithms to compute Π with $|\Pi| \leq \Delta + 1$, where Δ denotes the maximum number of edges a vertex is contained in, are provided. Furthermore, we will prove that finding the least integer $k \geq 2$ so that G has a cograph k -decomposition or a cograph k -partition is an NP-hard problem. In order to attack this problem in future work, we derive ILP formulations to solve this problem. These findings complement results known about so-called cograph vertex partitions (Achlioptas 1997; Gimbel and Nes tr l 2002; Dorbec et al. 2014; Zhang 2014).

2 Basic definitions

Graph. In what follows, we consider undirected simple graphs $G = (V, E)$ with vertex set $V(G) = V$ and edge set $E(G) = E \subseteq \binom{V}{2}$. We denote edges $\{x, y\} \in E$ by $[x, y]$. The *complement graph* $G^c = (V, E^c)$ of $G = (V, E)$, has edge set $E^c = \binom{V}{2} \setminus E$. The graph $K_{|V|} = (V, E)$ with $E = \binom{V}{2}$ is called *complete graph*. A graph $H = (W, F)$ is an *induced subgraph* of $G = (V, E)$ if $W \subseteq V$ and all edges $[x, y] \in E$ with $x, y \in W$ are contained in F . The *degree* $\deg(v) = |\{e \in E \mid v \in e\}|$ of a vertex $v \in V$ is defined as the number of edges that contain v . The maximum degree of a graph is denoted with Δ . A *path* P_n in a graph $G = (V, E)$ is a sequence (v_1, \dots, v_n) of pairwise distinct vertices of V such that $[v_i, v_{i+1}] \in E$ with $1 \leq i \leq n - 1$. We will usually deal with P_4 's (v_1, v_2, v_3, v_4) and will denote them by $v_1 - v_2 - v_3 - v_4$.

(Rooted) Tree A connected graph T is a *tree* if T does not contain cycles. A vertex of a tree T of degree one is called a *leaf* of T and all other vertices of T are called *inner* vertices. The set of inner vertices of T is denoted by V^0 and with E^0 we denote the set of *inner edges*, that is, edges in E where both of its end vertices are inner vertices. A *rooted tree* $T = (V, E)$ is a tree that contains a distinguished vertex $\rho_T \in V$ called the *root*. The first inner vertex $\text{lca}_T(x, y)$ that lies on both unique paths from distinct leaves x , resp., y to the root, is called *lowest common ancestor of x and y* . If there is no danger of ambiguity, we will write $\text{lca}(x, y)$ rather than $\text{lca}_T(x, y)$.

Symbolic Ultrametric and Symbolic Representation. In what follows, the set M will always denote a non-empty finite set, the symbol \odot will always denote a special

element not contained in M , and $M^\odot := M \cup \{\odot\}$. Now, suppose X is an arbitrary non-empty set and $\delta : X \times X \rightarrow M^\odot$ a map. We call δ a *symbolic ultrametric* if it satisfies the following conditions:

- (U0) $\delta(x, y) = \odot$ if and only if $x = y$;
- (U1) $\delta(x, y) = \delta(y, x)$ for all $x, y \in X$, i.e. δ is symmetric;
- (U2) $|\{\delta(x, y), \delta(x, z), \delta(y, z)\}| \leq 2$ for all $x, y, z \in X$; and
- (U3) there exists no subset $\{x, y, u, v\} \in \binom{X}{4}$ such that $\delta(x, y) = \delta(y, u) = \delta(u, v) \neq \delta(y, v) = \delta(x, v) = \delta(x, u)$.

Now, suppose that $T = (V, E)$ is a rooted tree with leaf set X and that $t : V \rightarrow M^\odot$ is a map such that $t(x) = \odot$ for all $x \in X$. To the pair (T, t) we associate the map $d_{(T, t)}$ on $X \times X$ by setting, for all $x, y \in X$,

$$d_{(T, t)} : X \times X \rightarrow M^\odot; d_{(T, t)}(x, y) = t(\text{lca}_T(x, y)).$$

We call the pair (T, t) a *symbolic representation* of a map $\delta : X \times X \rightarrow M^\odot$, if $\delta(x, y) = d_{(T, t)}(x, y)$ holds for all $x, y \in X$. For a subset $W \subseteq X \times X$ we denote with $\delta(W) = \{m \in M^\odot \mid \exists x, y \in W \text{ s.t. } \delta(x, y) = m\}$ the set of images of the elements contained in W .

We say that (T, t) and (T', t') are isomorphic if T and T' are isomorphic via a map $\varphi : V(T) \rightarrow V(T')$ such that $t'(\varphi(v)) = t(v)$ holds for all $v \in V(T)$.

Cographs and Cotrees. Complement-reducible graph, cographs for short, are defined as the class of graphs formed from a single vertex under the closure of the operations of union and complementation, namely: (i) a single-vertex graph is a cograph; (ii) the disjoint union of cographs is a cograph; (iii) the complement of a cograph is a cograph. Alternatively, a cograph can be defined as a P_4 -free graph (i.e. a graph such that no four vertices induce a subgraph that is a path of length 3). A number of equivalent characterizations of cographs can be found in Brandstädt et al. (1999). It is well-known in the literature concerning cographs that, to any cograph $G = (V', E')$, one can associate a canonical *cotree* $T(G) = (V, E)$. This is a rooted tree with leaf set $V \setminus V^0$ equal to the vertex set V' of G and inner vertices that represent so-called "join" and "union" operations together with a labeling map $t : V^0 \rightarrow \{0, 1\}$ such $[x, y] \in E'$ if and only if $t(\text{lca}(x, y)) = 1$, and $t(v) \neq t(w_i)$ for all $v \in V^0$ and all children $w_1, \dots, w_k \in V^0$ of v , (cf. Corneil et al. 1981). We will call the pair (T, t) *cotree representation* of G .

Cograph k -Decomposition and Partition, and Cotree Representation. Let $G = (V, E)$ be an arbitrary graph. A decomposition $\Pi = \{E_1, \dots, E_k\}$ of E is called (*cograph*) *k -decomposition*, if each subgraph $G_i = (V, E_i)$, $1 \leq i \leq k$ of G is a cograph. We call Π a (*cograph*) *k -partition* if $E_i \cap E_j = \emptyset$, for all distinct $i, j \in \{1, \dots, k\}$. A k -decomposition Π is called *optimal* if Π has the least number k of elements over all cograph decompositions of G . Clearly, for a cograph only k -decompositions with $k = 1$ are optimal. A k -decomposition $\Pi = \{E_1, \dots, E_k\}$ is *coarsest* if no elements of Π can be unified, so that the resulting decomposition is a cograph l -decomposition, with $l < k$. In other words, Π is coarsest if for all subsets $I \subseteq \{1, \dots, k\}$ with $|I| > 1$ it holds that $(V, \cup_{i \in I} E_i)$ is not a cograph. Thus, every optimal k -decomposition is also always a coarsest one.

A graph $G = (V, E)$ is represented by a set of cotrees $\mathbb{T} = \{T_1, \dots, T_k\}$, each T_i with leaf set V if and only if for each edge $[x, y] \in E$ there is a tree $T_i \in \mathbb{T}$ with $t(\text{lca}_{T_i}(x, y)) = 1$.

The Cartesian (Graph) Product $G \square H$ has vertex set $V(G \square H) = V(G) \times V(H)$; two vertices $(g_1, h_1), (g_2, h_2)$ are adjacent in $G \square H$ if $[g_1, g_2] \in E(G)$ and $h_1 = h_2$, or $[h_1, h_2] \in E(H)$ and $g_1 = g_2$. It is well-known that the Cartesian product is associative, commutative and that the single vertex graph K_1 serves as unit element (Hellmuth et al. 2013b; Hammack et al. 2011). Thus, the product $\square_{i=1}^n G_i$ of arbitrary many factors G_1, \dots, G_n is well-defined. For a given product $\square_{i=1}^n G_i$, we define the G_i -layer G_i^w of G (through vertex w that has coordinates (w_1, \dots, w_n)) as the induced subgraph with vertex set $V(G_i^w) = \{v = (v_1, \dots, v_n) \in \times_{i=1}^n V(G_i) \mid v_j = w_j, \text{ for all } j \neq i\}$. Note that G_i^w is isomorphic to G_i for all $1 \leq i \leq n, w \in V(\square_{i=1}^n G_i)$. The n -dimensional hypercube Q_n or n -cube, for short, is the Cartesian product $\square_{i=1}^n K_2$.

3 Symbolic ultrametrics and cographs

Symbolic ultrametrics and respective representations as event-labeled trees, have been first characterized by Böcker and Dress (1998).

Theorem 1 (Böcker and Dress 1998; Hellmuth et al. 2013a) *Suppose $\delta : V \times V \rightarrow M^\odot$ is a map. Then there is a symbolic representation of δ if and only if δ is a symbolic ultrametric. Furthermore, this representation can be computed in polynomial time.*

Now, let $\delta : V \times V \rightarrow M^\odot$ be a map satisfying Properties (U0) and (U1). Clearly, the map δ can be considered as an edge coloring of a complete graph $K_{|V|}$, where each edge $[x, y]$ obtains color $\delta(x, y)$. For each fixed $m \in M$, we define the undirected graph $G_m := G_m(\delta) = (V, E_m)$ with edge set

$$E_m = \{[x, y] \mid \delta(x, y) = m, x, y \in V\}. \quad (1)$$

Hence, G_m denotes the subgraph of the edge-colored graph $K_{|V|}$, that contains all edges colored with $m \in M$. The following result establishes the connection between symbolic ultrametrics and cographs.

Theorem 2 (Hellmuth et al. (2013a)) *Let $\delta : V \times V \rightarrow M^\odot$ be a map satisfying Properties (U0) and (U1). Then δ is a symbolic ultrametric if and only if*

- (U2') *For all $\{x, y, z\} \in \binom{V}{3}$ there is an $m \in M$ such that E_m contains (at least) two of the three edges $[x, y]$, $[x, z]$, and $[y, z]$. In other words, for each triangle induced by x, y and z , the edges have at most 2 different colors*
- (U3') *G_m is a cograph for all $m \in M$.*

In what follows, we will identify maps δ with their corresponding edge-colored graph representation. Assume now, we have given an arbitrary none edge-colored graph $G = (V, E)$ and we want to represent the topology of G as a tree. The following question then arises:

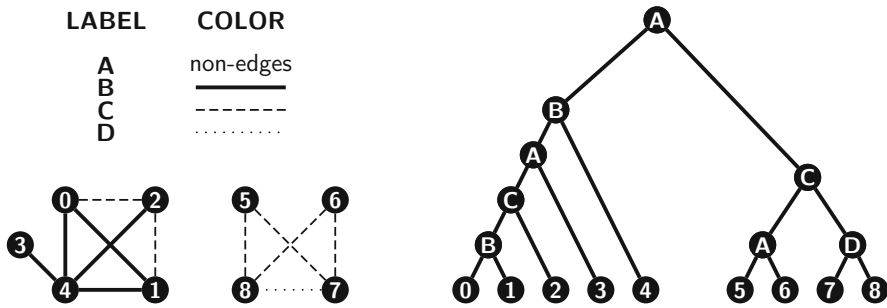


Fig. 1 Shown is a disconnected edge-colored graph G in the lower left part. The edge-colors are identified with the labels B , C and D , as indicated in the upper left part. Non-edges are identified with the label A . It is easy to verify that the event-labeled tree (T, t) on the right hand side is a tree representation of G , since for all distinct leaves i and j we have $t(\text{lca}_T(i, j)) = X \in \{A, B, C, D\}$ if and only if the (non-)edge $[i, j]$ has the color identified with the respective label X . In particular, G is a cograph and its cotree representation can be obtained by replacing the label A by 0, all other labels B, C, D by 1 and additional contraction of the interior edges $[C, B]$ and $[C, D]$

Under which conditions is it possible to define an coloring on the edges and non-edges of G , so that edges $e \in E$ obtain a different color than the non-edges $e \in E^c$ of G and, in particular, so that the resulting map δ is a symbolic ultrametric?

In other words, we ask for an edge-coloring of G so that there is a tree (T, t) with $t(\text{lca}_T(x, y)) = m$ if and only if the (non)edge $[x, y]$ obtained color m and that edges and non-edges of G can be distinguished by this coloring, that is, edges and non-edges never obtain the same color. For an example of such an edge-colored graph G , see Fig. 1. The following theorem gives necessary and sufficient conditions on the structure of graphs G for which one can find such a coloring and, in addition, provides a new characterization of cographs.

Theorem 3 Let $G = (V, E)$ be an arbitrary (possibly disconnected) graph, $W = \{(x, y) \in V \times V \mid [x, y] \in E\}$ and $W^c = \{(x, y) \in V \times V \mid [x, y] \notin E\}$.

There is a symbolic ultrametric $\delta : V \times V \rightarrow M^\odot$ s.t. $\delta(W) \cap \delta(W^c) = \emptyset$ if and only if G is a cograph.

Proof First assume that G is a cograph. Set $\delta(x, x) = \odot$ for all $x \in V$ and set $\delta(x, y) = \delta(y, x) = 1$ if $[x, y] \in E$ and, otherwise, to 0. Hence, condition (U0) and (U1) are fulfilled. Moreover, by construction $|M| = 2$ and thus, Condition (U2') is trivially fulfilled. Furthermore, since $G_1(\delta)$ and its complement $G_0(\delta)$ are cographs, (U3') is satisfied. Theorem 2 implies that δ is a symbolic ultrametric.

Now, let $\delta : V \times V \rightarrow M^\odot$ be a symbolic ultrametric with $\delta(W) \cap \delta(W^c) = \emptyset$. Assume for contradiction that G is not a cograph. Then G contains an induced path P_4 of the form $a - b - c - d$. Therefore, at least one edge e of this P_4 must obtain a color $\delta(e)$ different from the other two edges contained in this P_4 , as otherwise $G_{\delta(e)}(\delta)$ is not a cograph and thus, δ is not a symbolic ultrametric (Theorem 2, (U3')). For all such possible maps δ “subdividing” this P_4 we always obtain that two edges of at least one of the underlying paths P_3 of the form $a - b - c$ or $b - c - d$ must have different colors. W.l.o.g. assume that $\delta(a, b) \neq \delta(b, c)$. Since $[a, c] \notin E$ and $\delta(W) \cap \delta(W^c) = \emptyset$ we can conclude that $\delta(a, c) \neq \delta(a, b)$ and $\delta(a, c) \neq \delta(b, c)$.

But then Condition $(U2')$ cannot be satisfied, and Theorem 2 implies that δ is not a symbolic ultrametric. \square

Theorem 3 implies that there is no hope for finding an edge-distinguishing map δ for a graph G such that for δ (and hence, for G) there is a symbolic representation (T, t) , unless G is already a cograph. However, this result does not come as a big surprise, as a cograph G is characterized by the existence of a unique (up to isomorphism) cotree (T', t') representing the topology of G . As a consequence of this result we can infer that any symbolic representation (T, t) of a cograph G is a refinement of the cotree representation (T', t') of G , that is, the cotree representation (T', t') of G can be obtained from the symbolic representation (T, t) of δ by the following procedure:

First set for each $v \in V(T)$,

$$t(v) = \begin{cases} \odot & \text{if } v \in V(T) \setminus V^0(T), \text{ i.e., } v \text{ is a leaf} \\ 1 & \text{if } v = \text{lca}_T(x, y) \text{ and } [x, y] \text{ is an edge in } G \\ 0 & \text{if } v = \text{lca}_T(x, y) \text{ and } [x, y] \text{ is not an edge in } G \end{cases}$$

Clearly, this new map t on the tree T defines a symbolic representation (T, t) of the cograph $G = (V, E)$ so that $[x, y] \in E$ if and only if $t(\text{lca}_T(x, y)) = 1$. However, it might be possible that there is an edge $e = [u, v] \in E^0(T)$ such that $t(u) = t(v)$, and therefore, (T, t) is not a cotree representation. In this case, identify a new vertex v_e with e and define the tree $T_e = (V_e, E_e)$ with vertex set $V_e = V(T) \setminus \{u, v\} \cup \{v_e\}$, edge set $E_e = E(T) \setminus \{e\} \cup \{[v_e, w] \mid [w, u] \text{ or } [w, v] \in E\}$, that is again a rooted tree. Define for all $w \in V_e$ the map

$$t_e(w) = t(w) \text{ for } w \neq v_e; \text{ and set } t_e(v_e) = t(u). \quad (2)$$

This construction can be repeated, with (T_e, t_e) now playing the role of (T, t) , until we end in a rooted tree $\widehat{T} = (\widehat{V}, \widehat{E})$ with a map $\widehat{t}: \widehat{V} \rightarrow \{0, 1, \odot\}$ so that for all edges $[u, v] \in \widehat{E}^0$ it holds that $\widehat{t}(u) \neq \widehat{t}(v)$.

With this procedure, we obtain a symbolic representation $(\widehat{T}, \widehat{t})$ of the cograph G , also known as so-called *discriminating* symbolic ultrametric (Hellmuth et al. 2013a). In particular, this representation $(\widehat{T}, \widehat{t})$ is unique (up to isomorphism) (Hellmuth et al. 2013a, cf. Prop. 1) and, by construction, satisfies the condition of a cotree representation. Moreover, since the cotree representation (T', t') is unique (up to isomorphism) (Lerchs 1971a, b), it follows that (T', t') and $(\widehat{T}, \widehat{t})$ must be isomorphic. We summarize this result in the following corollary.

Corollary 1 *Let $G = (V, E)$ be a cograph, (T', t') be the corresponding cotree representation, and W , resp., W^c as defined in Theorem 3. Moreover, assume that there is a symbolic ultrametric $\delta: V \times V \rightarrow M^\odot$ s.t. $\delta(W) \cap \delta(W^c) = \emptyset$ with (T, t) being the corresponding symbolic representation of δ .*

Assume that the pair $(\widehat{T}, \widehat{t})$ is obtained from (T, t) by application of the procedure above. Then, $(\widehat{T}, \widehat{t})$ and (T', t') are isomorphic.

Assume that we want to find a symbolic ultrametric that can distinguish between “most of” the edges and/or non-edges but, in case the given graph is a not a cograph. Therefore, we will deal with the following problems.

Problem COGRAPH EDITING/DELETION/COMPLETION

Input: Given a simple graph $G = (V, E)$ and an integer k
Question: Is there a cograph $G' = (V, E')$, s.t.
 $E' \subseteq \binom{V}{2}$ and $|E \Delta E'| \leq k$ (Editing),
 $E' \subseteq E$ and $|E \setminus E'| \leq k$ (Deletion), or
 $E \subseteq E'$ and $|E' \setminus E| \leq k$ (Completion)?

The (decision version of the) problem to edit a given graph G into a cograph G' , and thus, to find the closest graph G' that has a symbolic representation, is NP-complete (Liu et al. 2011, 2012). In addition, the problems of deciding whether there is a cograph G' resulting by adding, resp., removing k edges from G is NP-complete, as well (El-Mallah and Colbourn 1988).

Theorem 4 (Liu et al. 2012; El-Mallah and Colbourn 1988) COGRAPH EDITING, COGRAPH COMPLETION and COGRAPH DELETION are NP-complete.

In what follows, we will consider and discuss two modifications of the problem of finding a symbolic ultrametric that can distinguish between edges and non-edges:

1. In Sect. 4, we consider a couple of problems which are of highly practical relevance: The symbolic ultrametric editing, completion and deletion problem.
2. In contrast, if a graph G without edge-coloring is not a cograph and thus, if there is no single tree representation of G , then we ask for the minimum number of trees that are needed in order to represent the topology of G in an unambiguous way, see Sect. 5.

4 Symbolic ultrametric editing, completion and deletion

As mentioned in the introduction, symbolic ultrametrics lie at the heart of many problems in phylogenomics (Hellmuth et al. 2013a, 2015; Hellmuth and Wieseke 2016; Lafond and El-Mabrouk 2014). In order to represent a map $d : X \times X \rightarrow \{\text{speciation, duplication, HGT}\}$ that determines for each pair of distinct genes $x, y \in X$ whether they are orthologs, paralogs, or xenologs, the map d must be a symbolic ultrametric. In practice, however, we deal with maps d that are only estimates of the true evolutionary relationship δ between the investigated genes. Thus, in general, such estimates d will not be a symbolic ultrametric. Hence, there is a great interest in optimally editing d to a symbolic ultrametric δ .

The problem of editing a given symmetric map $d : X \times X \rightarrow M^\odot$ to a symbolic ultrametric is defined as follows:

Problem SYMBOLIC ULTRAMETRIC EDITING

Input: Given a symmetric map $d : X \times X \rightarrow M^\odot$, s.t.
 $d(x, y) = \odot$ if and only if $x = y$

Question: Is there a symbolic ultrametric $\delta : X \times X \rightarrow M^\odot$, s.t. for
 $D = \{(x, y) \in X \times X \mid d(x, y) \neq \delta(x, y)\}$ we have $|D| \leq k$?

Furthermore, assume that we have an assignment of a symmetric subset R of $X \times X$ such that for all $(x, y) \in R$ the assignment $d(x, y)$ is a reliable estimate and thus, is not allowed to be changed. Moreover, let $X \times X \setminus R$ be the pairs (x, y) for which an assignment $d(x, y)$ is not known. Assume that $M = \{1, \dots, n\}$ and $M^\odot = M \cup \{\odot, 0\}$. Then we can extend the map $d : X \times X \rightarrow M^\odot$ such that:

$$d(x, y) = \begin{cases} \odot & \text{if } x = y \\ d(x, y) & \text{if } (x, y) \in R \\ 0 & \text{if } (x, y) \in X \times X \setminus R \end{cases}$$

We then ask to change the assignment of a minimum number of pairs (x, y) with $d(x, y) = 0$ to some element $m \in M$, $m \neq 0$ so that the resulting map is a symbolic ultrametric. In other words, only non-reliable estimates of pairs (x, y) are allowed to be changed.

Problem SYMBOLIC ULTRAMETRIC COMPLETION

Input: Given a symmetric map $d : X \times X \rightarrow M^\odot$, s.t.
 $d(x, y) = \odot$ if and only if $x = y$

Question: Is there a symbolic ultrametric $\delta : X \times X \rightarrow M^\odot$ s.t.
 if $d(x, y) \neq 0$, then $\delta(x, y) = d(x, y)$; and $|D| \leq k$, where
 $D = \{(x, y) \in X \times X \mid d(x, y) \neq \delta(x, y)\}$?

Conversely, one might ask to change a minimum number of assignments $d(x, y) \neq 0$ to $\delta(x, y) = 0$.

Problem SYMBOLIC ULTRAMETRIC DELETION

Input: Given a symmetric map $d : X \times X \rightarrow M^\odot$, s.t.
 $d(x, y) = \odot$ if and only if $x = y$

Question: Is there a symbolic ultrametric $\delta : X \times X \rightarrow M^\odot$ s.t.
 $\delta(x, y) = d(x, y)$ or $\delta(x, y) = 0$; and $|D| \leq k$, where
 $D = \{(x, y) \in X \times X \mid d(x, y) \neq \delta(x, y)\}$?

4.1 Computational complexity

In this section, we prove the NP-completeness of SYMBOLIC ULTRAMETRIC EDITING, SYMBOLIC ULTRAMETRIC COMPLETION and SYMBOLIC ULTRAMETRIC DELETION.

Theorem 5 SYMBOLIC ULTRAMETRIC EDITING is NP-complete.

Proof For a symmetric map δ it can be verified in polynomial time if δ is a symbolic ultrametric: One can check Conditions (U2) and (U3) individually for each of the $O(|X|^3)$ many combinations of $\{x, y, z\} \in \binom{X}{3}$ for (U2), and the $O(|X|^4)$ many combinations of $\{x, y, u, v\} \in \binom{X}{4}$ for (U3), respectively. Hence, SYMBOLIC ULTRAMETRIC EDITING $\in NP$. We will show by reduction from COGRAPH EDITING that SYMBOLIC ULTRAMETRIC EDITING is NP-hard.

Let $G = (V, E)$ be a simple graph being an instance of the cograph-editing problem. In what follows, we associate with G a map $d: V \times V \rightarrow M^\odot$, where $M = \{0, 1, \dots, n\}$ is a non-empty finite set s.t. $n \geq 1$ and thus, $0, 1 \in M$. Let $M^\odot := M \cup \{\odot\}$ and set for all $x, y \in V$:

$$d(x, y) = d(y, x) = \begin{cases} \odot & \text{if } x = y \\ 1 & \text{if } [x, y] \in E \\ 0 & \text{if } [x, y] \notin E \end{cases}$$

Obviously, d can be constructed in polynomial time. In the following, we show, that given an integer k , there exists a solution of the COGRAPH EDITING problem with input G and integer k if and only if there exists a solution of the SYMBOLIC ULTRAMETRIC EDITING problem for d and integer $2k$.

First, we show that a solution of the SYMBOLIC ULTRAMETRIC EDITING problem with input d and $2k$ can be constructed from a solution of the COGRAPH EDITING problem with input G and k . Let $G' = (V, E')$ be a cograph with $|E \Delta E'| \leq k$. Furthermore let $\delta: V \times V \rightarrow M^\odot$ be a map, such that, for all $x, y \in V$:

$$\delta(x, y) = \delta(y, x) = \begin{cases} \odot & \text{if } x = y \\ 1 & \text{if } [x, y] \in E' \\ 0 & \text{if } [x, y] \notin E' \end{cases}$$

It is easy to verify that δ is a symbolic ultrametric by application of Theorem 2. It remains to be shown that for $D = \{(x, y) \in X \times X \mid d(x, y) \neq \delta(x, y)\}$ it holds that $|D| \leq 2k$. Note that for all $x \in V$ we have $d(x, x) = \delta(x, x) = \odot$ and therefore, $(x, x) \notin D$. The set D can be partitioned into the two subsets

$$\begin{aligned} D_1 &= \{(x, y) \mid d(x, y) = 1 \text{ and } \delta(x, y) = 0\} \text{ and} \\ D_2 &= \{(x, y) \mid d(x, y) = 0 \text{ and } \delta(x, y) = 1\}. \end{aligned}$$

Hence, $(x, y) \in D_1$ if and only if $[x, y] \in E \setminus E'$, and $(x, y) \in D_2$ if and only if $[x, y] \in E' \setminus E$. As $(E \setminus E') \cup (E' \setminus E) = (E \Delta E')$ it holds that $(x, y) \in D$ if and only if $[x, y] \in E \Delta E'$. As d and δ are symmetric, it also holds that $(x, y) \in D$ if and only if $(y, x) \in D$. Hence, $[x, y] \in E \Delta E'$ if and only if $(x, y) \in D$ and $(y, x) \in D$. This reflects the fact, that an edge edit $[x, y] \in E \Delta E'$ in G corresponds to the two symmetric edits $(x, y), (y, x) \in D$ in d . Therefore, $|D| = |\{(x, y) \mid d(x, y) \neq \delta(x, y)\}| = 2|E \Delta E'| \leq 2k$.

We continue to show that a solution of the COGRAPH EDITING problem with input G and k can be constructed from a solution of the SYMBOLIC ULTRAMETRIC EDITING problem with input d and $2k$. Let $\delta : V \times V \rightarrow \tilde{M}^\odot$ be a symbolic ultrametric s.t. $|D| = |\{(x, y) \mid d(x, y) \neq \delta(x, y)\}| \leq 2k$. Furthermore, let $G' = (V, E')$ be a simple graph, such that, for all $x, y \in V$, it holds that $[x, y] \in E'$ if and only if $\delta(x, y) = 1$. By Theorem 2 (U3') we infer that $G' = G_1$ and hence, G' is a cograph. It remains to be shown that $|E \Delta E'| \leq k$. By construction, for all $x \in V$, $d(x, x) = \delta(x, x) = \odot$ and $[x, x] \notin E \Delta E'$. Let $D = \{(x, y) \mid d(x, y) \neq \delta(x, y)\}$. Note that for all distinct $x, y \in V$ it holds that $d(x, y) \in \{0, 1\}$. Hence, D can be partitioned into the four subsets

$$\begin{aligned} D_1 &= \{(x, y) \mid d(x, y) = 1 \text{ and } \delta(x, y) = 0\}, \\ D_2 &= \{(x, y) \mid d(x, y) = 0 \text{ and } \delta(x, y) = 1\}, \\ D_3 &= \{(x, y) \mid d(x, y) = 1 \text{ and } \delta(x, y) \in \tilde{M}^\odot \setminus \{0, 1\}\}, \text{ and} \\ D_4 &= \{(x, y) \mid d(x, y) = 0 \text{ and } \delta(x, y) \in \tilde{M}^\odot \setminus \{0, 1\}\}. \end{aligned}$$

Now, if $(x, y) \in D_1$ then $[x, y] \in E \setminus E'$, and if $(x, y) \in D_2$ then $[x, y] \in E' \setminus E$. Furthermore, $\delta(x, y) \in \tilde{M}^\odot \setminus \{0, 1\}$ implies that $[x, y] \notin E'$. Hence, if $(x, y) \in D_3$ then $[x, y] \in E \setminus E'$, and if $(x, y) \in D_4$ then $[x, y] \notin E$ and $[x, y] \notin E'$. For all remaining $x, y \in V$ that satisfy $d(x, y) = \delta(x, y)$, it holds that $[x, y] \notin E \setminus E'$ and $[x, y] \notin E' \setminus E$. It follows that $[x, y] \in E \setminus E'$ if and only if $(x, y) \in D_1 \cup D_3$, and $[x, y] \in E' \setminus E$ if and only if $(x, y) \in D_2$. As before, due to the symmetry of the maps d and δ , two symmetric edits $(x, y), (y, x) \in D$ in d correspond to one edge edit $[x, y] \in E \Delta E'$ in G . Finally, $2|E \Delta E'| = 2|E \setminus E'| + 2|E' \setminus E| = |D_1 \cup D_3| + |D_2| \leq |D| \leq 2k$. Hence, $|E \Delta E'| \leq k$.

Thus, SYMBOLIC ULTRAMETRIC EDITING is NP-complete.

Theorem 6 SYMBOLIC ULTRAMETRIC COMPLETION is NP-complete.

Proof It is shown analogously as in the proof of Theorem 5 that SYMBOLIC ULTRAMETRIC MIN COMPLETION $\in NP$. We will show by reduction from COGRAPH COMPLETION that SYMBOLIC ULTRAMETRIC COMPLETION is NP-hard.

Let $G = (V, E)$ be an arbitrary simple graph. We associate to G a map $d : V \times V \rightarrow M^\odot$ as defined in the proof of Theorem 5:

$$d(x, y) = d(y, x) = \begin{cases} \odot & \text{if } x = y \\ 1 & \text{if } [x, y] \in E \\ 0 & \text{if } [x, y] \notin E \end{cases}$$

Let there be a solution $G' = (V, E')$ for the COGRAPH COMPLETION problem with input $G = (V, E)$ and k , i.e., $E \subseteq E'$ and $|E' \setminus E| \leq k$. We show that there is a solution for the SYMBOLIC ULTRAMETRIC COMPLETION problem for d and $2k$.

Define the map $\delta: V \times V \rightarrow M^\odot$ as in the proof of Theorem 5:

$$\delta(x, y) = \delta(y, x) = \begin{cases} \odot & \text{if } x = y \\ 1 & \text{if } [x, y] \in E' \\ 0 & \text{if } [x, y] \notin E' \end{cases}$$

Again, it is easy to verify that δ is a symbolic ultrametric by application of Theorem 2. Moreover, by construction $\delta(x, y) = d(x, y)$ for all $x, y \in V$ whenever $[x, y] \in E \subseteq E'$ and hence, for all $x, y \in V$ with $d(x, y) \neq 0$.

It remains to be shown that for $D = \{(x, y) \in X \times X \mid 0 = d(x, y) \neq \delta(x, y)\}$ it holds that $|D| \leq 2k$. Note that for all $x \in V$ we have $d(x, x) = \delta(x, x) = \odot$ and therefore, $(x, x) \notin D$. Moreover,

$$D = \{(x, y) \mid d(x, y) = 0 \text{ and } \delta(x, y) = 1\}.$$

Hence, $(x, y), (y, x) \in D$ if and only if $[x, y] \in E' \setminus E$. Therefore, $|D| = 2|E' \setminus E| \leq 2k$.

We continue to show that a solution of the COGRAPH EDITING problem for G and k can be constructed from a solution of the SYMBOLIC ULTRAMETRIC EDITING problem with input d and $2k$. Let $\delta: V \times V \rightarrow \tilde{M}^\odot$ be a symbolic ultrametric s.t. $|D| \leq 2k$ and $\delta(x, y) = d(x, y)$ if $d(x, y) \neq 0$. Furthermore, let $G' = (V, E')$ be a simple graph, such that for all $x, y \in V$ it holds that $[x, y] \in E'$ if and only if $\delta(x, y) = 1$. By Theorem 2 (U3') we have that $G' = G_1$ and hence, G' is a cograph. It remains to be shown that $|E' \setminus E| \leq k$. By construction, for all $x \in V$, $d(x, x) = \delta(x, x) = \odot$ and $[x, x] \notin E'$. Note that for all distinct $x, y \in V$ it holds for the map associated to G that $d(x, y) \in \{0, 1\}$. Hence, D can be partitioned into

$$\begin{aligned} D_1 &= \{(x, y) \mid d(x, y) = 0 \text{ and } \delta(x, y) = 1\}, \text{ and} \\ D_2 &= \{(x, y) \mid d(x, y) = 0 \text{ and } \delta(x, y) \in \tilde{M}^\odot \setminus \{0, 1\}\}. \end{aligned}$$

Thus, if $(x, y), (y, x) \in D_1$, then $[x, y] \in E' \setminus E$. Therefore, $2(|E' \setminus E|) = |D_1| \leq |D| \leq 2k$ and thus, $|E' \setminus E| \leq k$.

Hence, SYMBOLIC ULTRAMETRIC COMPLETION is NP-complete.

Using similar arguments as in the proof of Theorem 6 we can infer the NP-completeness of SYMBOLIC ULTRAMETRIC DELETION by reduction from COGRAPH DELETION.

Theorem 7 SYMBOLIC ULTRAMETRIC DELETION is NP-complete.

4.2 Integer linear program

We showed in Hellmuth et al. (2015) that the cograph editing problem is amenable to formulations as an Integer Linear Program (ILP). We will extend this result here to solve the symbolic ultrametric editing/completion/deletion problem. Let $d: X \times X \rightarrow M^\odot$ be an arbitrary symmetric map with $M = \{0, 1, \dots, n\}$, and let $K_{|X|} = (X, E =$

$\binom{X}{2}$) be the corresponding complete graph with edge-coloring s.t. each edge $[x, y] \in E$ obtains color $d(x, y) = d(y, x)$.

Given a symmetric map d , we define for each distinct $x, y \in X$ and $i \in M$ the binary constants $\mathfrak{d}_{x,y}^i$ with $\mathfrak{d}_{x,y}^i = 1$ if and only if $d(x, y) = i$. Moreover, we define the binary variables E_{xy}^i for all $i \in M$ and $x, y \in X$ that reflect the coloring of the edges in $K_{|V|}$ of the final symbolic ultrametric δ , i.e., E_{xy}^i is set to 1 if and only if $\delta(x, y) = i$.

In order to find the closest symbolic ultrametric δ , the objective function is to minimize the symmetric difference of the d and δ over all different symbols $i \in M$:

$$\min \sum_{i \in M} \left(\sum_{(x,y) \in X} (1 - \mathfrak{d}_{xy}^i) E_{xy}^i + \sum_{(x,y) \in X} \mathfrak{d}_{xy}^i (1 - E_{xy}^i) \right) \quad (3)$$

The same objective function can be used for the symbolic ultrametric completion and deletion problem.

In case of the symbolic ultrametric completion, we must ensure that $\delta(x, y) = d(x, y)$ for all $d(x, y) \neq 0$. Hence, for all x, y with $d(x, y) = i \neq 0$ we set

$$E_{x,y}^i = 1. \quad (4)$$

In case of the symbolic ultrametric deletion, we must ensure that $\delta(x, y) = d(x, y)$ or $\delta(x, y) = 0$ or, in other words, for all $d(x, y) = i \neq 0$ it must hold that either $E_{xy}^i = 1$ or $E_{xy}^0 = 1$. Hence, for all $x, y \in V$ we set

$$E_{xy}^0 = 1 \text{ if } d(x, y) = 0, \text{ and } E_{xy}^i + E_{xy}^0 = 1, \text{ otherwise.} \quad (4')$$

For the cograph editing problem we need neither Constraint 4 nor 4'. However, for all three problems we need the following.

Each tuple (x, y) with $x \neq y$ has exactly one value $i \in M$ assigned to it, so

$$\sum_{i \in M} E_{x,y}^i = 1 \text{ and } E_{xy}^i - E_{yx}^i = 0 \text{ for all } x, y \in X. \quad (5)$$

In order to satisfy Condition (U2') and thus, to ensure that all induced triangles have at most two colors on the edges, we add the constraint:

$$E_{xy}^i + E_{yz}^j + E_{xz}^k \leq 2 \quad (6)$$

for all ordered tuples (i, j, k) with distinct $i, j, k \in M$ and pairwise distinct $x, y, z \in X$.

Finally, in order to satisfy Condition (U3') and thus, to ensure that each monochromatic subgraph comprising all edges with fixed color i is a cograph, we add the following constraint that forbids induced P_4 's.

$$E_{xy}^i + E_{yu}^i + E_{uv}^i - E_{xu}^i - E_{xv}^i - E_{yv}^i \leq 2 \quad (7)$$

for all $i \in M$ and all ordered tuples (x, y, u, v) of distinct $x, y, u, v \in X$.

It is easy to verify that this ILP formulation needs $O(|M||X|^2)$ variables and $O(|M|^3|X|^3 + |X|^4)$ constraints.

5 Cotree representation and cograph k -Decomposition

If a given non-edge colored graph G is not a cograph, then Theorem 3 implies that one cannot define an edge-distinguishing symbolic ultrametric, and, in particular, no single tree representation of G . Therefore, we are interested to represent the topology of G in an unambiguous way with a minimum number of trees.

Recall that a graph $G = (V, E)$ is represented by a set of cotrees $\mathbb{T} = \{T_1, \dots, T_k\}$ if and only if, for each edge $[x, y] \in E$, there is a tree $T_i \in \mathbb{T}$ with $t(\text{lca}_{T_i}(x, y)) = 1$.

Note that by definition, each cotree T_i determines a subset $E_i = \{[x, y] \in E \mid t(\text{lca}_{T_i}(x, y)) = 1\}$ of E . Hence, the subgraph (V, E_i) of G must be a cograph. Therefore, in order to find the minimum number of cotrees representing a graph G , we can equivalently ask for a decomposition $\Pi = \{E_1, \dots, E_k\}$ of E so that each subgraph (V, E_i) is a cograph, where k is the minimum over all cograph decompositions of G . Thus, we are dealing with the following two equivalent problems:

Problem COTREE k -REPRESENTATION

<i>Input:</i>	Given a graph $G = (V, E)$ and an integer k
<i>Question:</i>	Can G be represented by k cotrees ?

Clearly, any cograph has an optimal 1-decomposition, while for cycles of length > 4 or paths P_4 there is always an optimal cograph 2-decomposition. However, there are examples of graphs that do not even have a cograph 2-decomposition, see Fig. 2. Moreover, as shown in Fig. 3, the number of different optimal cograph k -decomposition on a graph can grow exponentially. The following theorem provides for every graph of bounded degree a non-trivial upper bound for the integer k s.t. there is still a cograph k -decomposition:

Theorem 8 *For every graph G with maximum degree Δ there is a cograph k -decomposition with $1 \leq k \leq \Delta + 1$ that can be computed in $O(|V||E| + \Delta(|V| + |E|))$ time. Hence, any graph can be represented by at most $\Delta + 1$ cotrees.*

Proof Consider a proper edge-coloring $\varphi : E \rightarrow \{1, \dots, k\}$ of G , i.e., an edge coloring such that no two incident edges obtain the same color. Any proper edge-coloring using k colors yields a cograph k -partition $\Pi = \{E_1, \dots, E_k\}$ where $E_i = \{e \in E \mid \varphi(e) = i\}$, because any connected component in $G_i = (V, E_i)$ is an edge and thus, no P_4 's are contained in G_i . Vizing's Theorem (Vizing 1964) implies that for each graph there is a proper edge-coloring using k colors with $\Delta \leq k \leq \Delta + 1$.

Fig. 2 Full enumeration of all possibilities (which we leaf to the reader), shows that the depicted graph has no cograph 2-decomposition. The existing cograph 3-decomposition is also a cograph 3-partition; highlighted by *dashed-lined*, *dotted* and *solid edges*

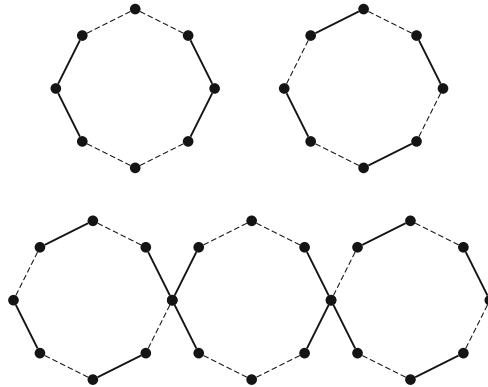
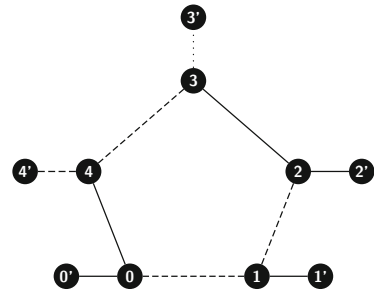


Fig. 3 Two isomorphic graphs H with two non-equivalent optimal cograph 2-decomposition (highlighted by *dashed* and *solid edges*) are shown in the *upper part*. By stepwisely identifying single vertices one obtains a “chain-graph” G that consists of the linked copies of the graph H , see *lower part*. For each subgraph H of G , an optimal cograph 2-decomposition can be determined almost independently of the remaining parts of the graph G . Hence, with an increasing number of vertices in the chain-graph G , that is, each linked new graph H adds seven new vertices, the number of different cograph 2-decompositions is growing exponentially

An proper edge-coloring using at most $\Delta + 1$ colors can be computed with the Misra-Gries-algorithm in $O(|V||E|)$ time (Misra and Gries 1992). Since the (at most $\Delta + 1$) respective cotrees can be constructed in linear-time $O(|V| + |E|)$ (Corneil et al. 1985), we derive the running time $O(|V||E| + \Delta(|V| + |E|))$.

Obviously, any optimal k -decomposition must also be a coarsest k -decomposition, while the converse is not true in general, see Fig. 4. The partition $\Pi = \{E_1, \dots, E_k\}$ obtained from a proper edge-coloring is usually not a coarsest one, as possibly (V, E_J) is a cograph, where $E_J = \cup_{i \in J} E_i$ and $J \subseteq \{1, \dots, l\}$. However, there are graphs having an *optimal* cograph Δ -decomposition, see Figs. 2 and 3. Thus, the derived bound $\Delta + 1$ is almost sharp. Nevertheless, we assume that this bound can be made tight:

Conjecture 1 For every graph G with maximum degree Δ there is a cograph Δ -decomposition.

However, there are examples of graphs containing many induced P_4 's that have a cograph k -decomposition with $k \ll \Delta + 1$, which implies that any optimal k -

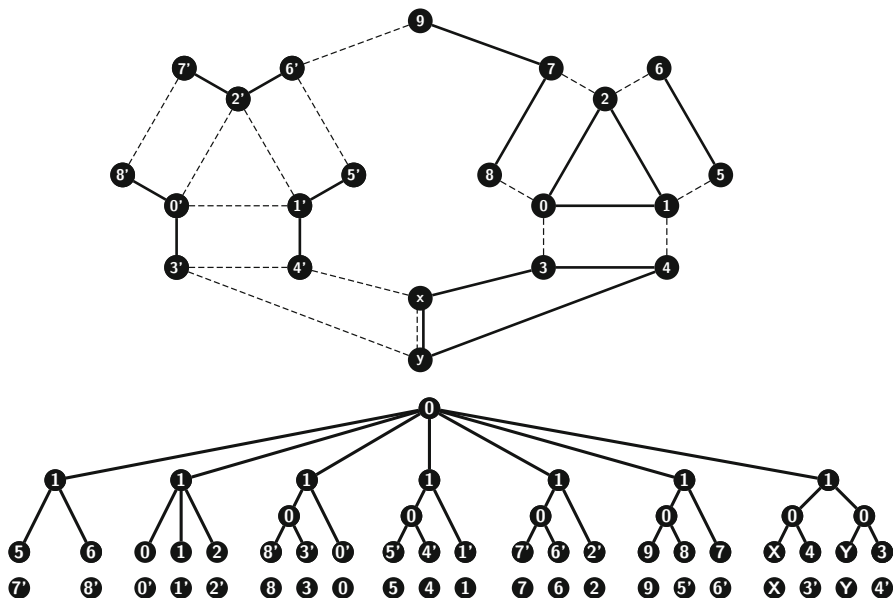


Fig. 4 The shown graph G is not a cograph and has a 2-decomposition $\Pi = \{E_1, E_2\}$. Edges in the different elements E_1 and E_2 are highlighted by *dashed* and *solid* edges, respectively. Thus, two cotrees, shown in the lower part of this picture, are sufficient to represent the structure of G . The two cotrees are isomorphic, and thus, differ only in the arrangement of their leaf sets. For this reason, we only depicted one cotree with two different leaf sets. Note that G has no 2-partition, but a coarsest 3-partition. The latter can easily be verified by application of the construction in Lemma 1

decomposition of those graphs will have significantly less elements than $\Delta + 1$, see the following examples.

Example 1 Consider the graph $G = (V, E)$ with vertex set $V = \{1, \dots, k\} \cup \{a, b\}$ and $E = \{[i, j] \mid i, j \in \{1, \dots, k\}, i \neq j\} \cup \{[k, a], [a, b]\}$. The graph G is not a cograph, since there are induced P_4 's of the form $i - k - a - b$, $i \in \{1, \dots, k - 1\}$. On the other hand, the subgraph $H = (V, E \setminus \{[k, a]\})$ has two connected components, one is isomorphic to the complete graph K_k on k vertices and the other to the complete graph K_2 . Hence, H is a cograph. Therefore, G has a cograph 2-partition $\{E \setminus \{[k, a]\}, \{[k, a]\}\}$, independent of k and thus, independent of the maximum degree $\Delta = k$.

Example 2 Consider the $2n$ -dimensional hypercube $Q_{2n} = (V, E)$ with maximum degree $2n$. We will show that this hypercube has a coarsest cograph n -partition $\Pi = \{E_1, \dots, E_n\}$, which implies that for any optimal cograph k -decomposition of Q_{2n} we have $k \leq \Delta/2$.

We construct now a cograph n -partition of Q_{2n} . Note that $Q_{2n} = \square_{i=1}^{2n} K_2 = \square_{i=1}^n (K_2 \square K_2) = \square_{i=1}^n Q_2$. In order to avoid ambiguity, we write $\square_{i=1}^n Q_2$ as $\square_{i=1}^n H_i$, $H_i \simeq Q_2$ and assume that Q_2 has edges $[0, 1], [1, 2], [2, 3], [3, 0]$. The cograph n -partition of Q_{2n} is defined as $\Pi = \{E_1, \dots, E_n\}$, where $E_i = \cup_{v \in V} E(H_i^v)$. In other words, the edge set of all H_i -layers in Q_{2n} constitute a single class E_i in the partition

for each i . Therefore, the subgraph $G = (V, E_i)$ consists of n connected components, each component is isomorphic to the square Q_2 . Hence, $G_i = (V, E_i)$ is a cograph.

Assume for contradiction that $\Pi = \{E_1, \dots, E_n\}$ is not a coarsest partition. Then there are distinct classes $E_i, i \in I \subseteq \{1, \dots, n\}$ such that $G_I = (V, \cup_{i \in I} E_i)$ is a cograph. W.l.o.g. assume that $1, 2 \in I$ and let $v = (0, \dots, 0) \in V$. Then, the subgraph $H_1^v \cup H_2^v \subseteq Q_{2n}$ contains a path P_4 with edges $[x, v] \in E(H_1^v)$ and $[v, a], [a, b] \in E(H_2^v)$, where $x=(1,0,\dots,0)$, $a=(0,1,0,\dots,0)$ and $b = (0, 2, 0 \dots, 0)$. By definition of the Cartesian product, there are no edges connecting x with a or b ; and v with b in Q_{2n} . Hence, this path P_4 is induced. As this holds for all subgraphs $H_i^v \cup H_j^v$ ($i, j \in I$ distinct) and thus, in particular for the graph G_I we can conclude that classes of Π cannot be combined. Hence, Π is a coarsest cograph n -partition.

Because of preliminary results of computer-aided search for $n - 1$ -partitions and decompositions of hypercubes Q_{2n} (data not shown here), we are led to the following conjecture:

Conjecture 2 Let $k \in \mathbb{N}$ and $k > 1$. Then the $2k$ -cube has no cograph $k - 1$ -decomposition, i.e., the proposed k -partition of the hypercube Q_{2k} in Example 2 is also optimal.

The proof of the latter hypothesis would immediately verify the next conjecture.

Conjecture 3 For every $k \in \mathbb{N}$ there is a graph that has an *optimal* cograph k -decomposition.

Proving the last conjecture appears to be difficult. We wish to point out that there is a close relationship to the problem of finding pattern avoiding words, see e.g. Brändén and Mansour (2005), Burstein and Mansour (2002), Pudwell (2008a, b), Bilotta et al. (2013), Bernini et al. (2009): Consider a graph $G = (V, E)$ and an ordered list (e_1, \dots, e_m) of the edges $e_i \in E$. We can associate to this list (e_1, \dots, e_m) a word $w = (w_1, \dots, w_m)$. By way of example, assume that we want to find a valid cograph 2-decomposition $\{E_1, E_2\}$ of G and that G contains an induced P_4 consisting of the edges e_i, e_j, e_k . Hence, one has to avoid assignments of the edges e_i, e_j, e_k to the single set E_1 , resp., E_2 . The latter is equivalent to find a binary word (w_1, \dots, w_m) such that $(w_i, w_j, w_k) \neq (X, X, X)$, $X \in \{0, 1\}$ for each of those induced P_4 's. The latter can easily be generalized to find pattern avoiding words over an alphabet $\{1, \dots, k\}$ to get a valid k -decomposition. However, to the authors knowledge, results concerning the counting of k -ary words, avoiding forbidden patterns and thus, verifying if there is any such word (or equivalently a k -decomposition) are basically known for scenarios like: If $(p_1, \dots, p_l) \in \{1, \dots, k\}^l$ (often $l < 3$), then none of the words w that contain a subword $(w_{i_1}, \dots, w_{i_l}) = (p_1, \dots, p_l)$ with $i_{j+1} = i_j + 1$ (consecutive letter positions) or $i_j < i_k$ whenever $j < k$ (order-isomorphic letter positions) is allowed. However, such findings are too restrictive to our problem, since we are looking for words that have only on a few, but fixed positions of non-allowed patterns. Nevertheless, we assume that results concerning the recognition of pattern avoiding words might offer an avenue to solve the latter conjectures.

Fig. 5 Left the *literal graph* and right the *extended literal graph* with unique corresponding cograph 2-partition (indicated by *dashed* and *bold-lined edges*) is shown

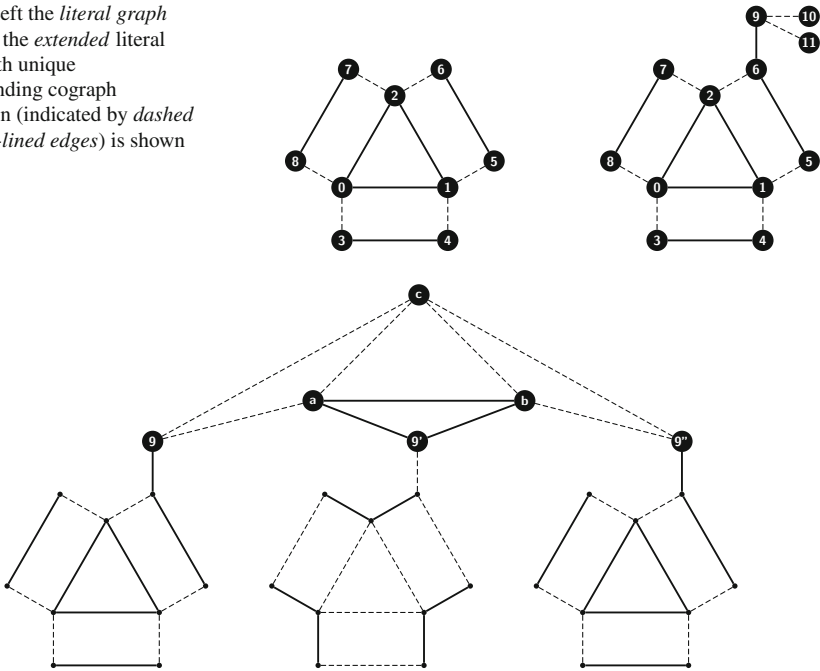


Fig. 6 Shown is a *clause gadget* which consists of a triangle (a, b, c) and three extended literal graphs (as shown in Fig. 5) with edges attached to (a, b, c) . A corresponding cograph 2-partition is indicated by *dashed* and *bold-line edges*

5.1 Computational complexity

In the following, we will prove the NP-completeness of COTREE 2- REPRESENTATION and COTREE 2- DECOMPOSITION. Additionally, these results allow us to show that the problem of determining whether there is cograph 2-partition is NP-complete, as well.

We start with two lemmata concerning cograph 2-decompositions of the graphs shown in Figs. 5 and 6.

Lemma 1 *For the literal and extended literal graph in Fig. 5 every cograph 2-decomposition is a uniquely determined cograph 2-partition.*

In particular, in every cograph 2-partition $\{E_1, E_2\}$ of the extended literal graph, the edges of the triangle $(0, 1, 2)$ must be entirely contained in one E_i and the pending edge $[6, 9]$ must be in the same edge set E_i as the edges of the of the triangle. Furthermore, the edges $[9, 10]$ and $[9, 11]$ must be contained in E_j , $i \neq j$.

Proof It is easy to verify that the given cograph 2-partition $\{E_1, E_2\}$ in Fig. 5 fulfills the conditions and is correct, since $G_1 = (V, E_1)$ and $G_2 = (V, E_2)$ do not contain induced P_4 's and are, thus, cographs. We have to show that it is also unique.

Assume that there is another cograph 2-decomposition $\{F_1, F_2\}$. Note that for any cograph 2-decomposition $\{F_1, F_2\}$ it must hold that two incident edges in the triangle

$(0, 1, 2)$ are contained in one of the sets F_1 or F_2 . W.l.o.g. assume that $[0, 1], [0, 2] \in F_1$.

Assume first that $[1, 2] \notin F_1$. In this case, because of the paths $6-2-0-1$ and $2-0-1-5$ it must hold that $[2, 6], [1, 5] \notin F_1$ and thus, $[2, 6], [1, 5] \in F_2$. However, in this case and due to the paths $6-2-1-4$ and $2-0-1-4$ the edge $[1, 4]$ can neither be contained in F_1 nor in F_2 ; a contradiction. Hence, $[1, 2] \in F_1$.

Note that the square S induced by vertices $1, 2, 5, 6$ cannot have all edges in F_1 , as otherwise the subgraph (V, F_1) would contain an induced P_4 of the form $6-5-1-0$. Assume that $[1, 5] \in F_1$. As not all edges of this square S are contained in F_1 , at least one of the edges $[5, 6]$ and $[2, 6]$ must be contained in F_2 . If only one of the edges $[5, 6]$, resp., $[2, 6]$ is contained in F_2 , we immediately obtain an induced P_4 of the form $6-2-1-5$, resp., $6-5-1-2$ in (V, F_1) and therefore, both edges $[5, 6]$ and $[2, 6]$ must be contained in F_2 . But then the edge $[2, 7]$ can neither be contained in F_1 (due to the induced P_4 of the form $5-1-2-7$) nor in F_2 (due to the induced P_4 of the form $5-6-2-7$); a contradiction. Hence, $[1, 5] \notin F_1$ and thus, $[1, 5] \in F_2$ for any 2-decomposition. By analogous arguments and due to symmetry, all edges $[0, 3], [0, 8], [1, 4], [2, 6], [2, 7]$ are contained in F_2 , but not in F_1 .

Moreover, due to the induced P_4 of the form $7-2-6-5$ and since $[2, 6], [2, 7] \in F_2$, the edge $[5, 6]$ must be in F_1 and not in F_2 . By analogous arguments and due to symmetry, it holds that $[3, 4], [7, 8] \in F_1$ and $[3, 4], [7, 8] \notin F_2$. Finally, none of the edges of the triangle $(0, 1, 2)$ can be contained in F_2 , as otherwise, we obtain an induced P_4 in (V, F_2) . Taken together, any 2-decomposition of the literal graph must be a partition and is unique.

Consider now the extended literal graph in Fig. 5. As this graph contains the literal graph as induced subgraph, the unique 2-partition of the underlying literal graph is determined as by the preceding construction. Due to the path $7-2-6-9$ with $[2, 6], [2, 7] \in F_2$ we can conclude that $[6, 9] \notin F_2$ and thus, $[6, 9] \in F_1$. Since there are induced paths $5-6-9-y$, $y = 10, 11$ with $[5, 6], [6, 9] \in F_1$ we obtain that $[9, 10], [9, 11] \notin F_1$ and thus, $[9, 10], [9, 11] \in F_2$ for any 2-decomposition (which is in fact a 2-partition) of the extended literal graph, as claimed.

Lemma 2 *Given the clause gadget in Fig. 6.*

For any cograph 2-decomposition, all edges of exactly two of the triangles in the underlying three extended literal graphs must be contained in one E_i and not in E_j , while the edges of the triangle of one extended literal graph must be in E_j and not in E_i , $i \neq j$.

Furthermore, for each cograph 2-decomposition exactly two of the edges e, e' of the triangle (a, b, c) must be in one E_i while the other edge f is in E_j but not in E_i , $j \neq i$. The cograph 2-decomposition can be chosen so that in addition $e, e' \notin E_j$, resulting in a cograph 2-partition of the clause gadget.

Proof It is easy to verify that the given cograph 2-partition in Fig. 6 fulfills the conditions and is correct, as $G_1 = (V, E_1)$ and $G_2 = (V, E_2)$ are cographs.

As the clause gadget contains the literal graph as induced subgraph, the unique 2-partition of the underlying literal graph is determined as by the construction given in Lemma 1. Thus, each edge of the triangle in each underlying literal graph is contained in either one of the sets E_1 or E_2 . Assume that edges of the triangles in

the three literal gadgets are *all* contained in the same set, say E_1 . Then, Lemma 1 implies that $[9, a], [9, c], [9', a], [9', b], [9'', b], [9'', c] \in E_2$ and none of them is contained in E_1 . Since there are induced P_4 's: $9-a-b-9'', 9'-a-c-9''$ and $9-c-b-9'$, the edges $[a, b], [a, c], [b, c]$ cannot be contained in E_2 , and thus, must be in E_1 . However, this is not possible, since then we would have induced paths P_4 of the form $9-a-9'-b$ in the subgraph (V, E_2) ; a contradiction. Thus, the edges of the triangle of exactly one literal gadget must be contained in a different set E_i than the edges of the other triangles in the other two literal gadgets. W.l.o.g. assume that the 2-decomposition of the underlying literal gadgets is given as in Fig. 6. and identify bold-lined edges with E_1 and dashed edges with E_2 .

It remains to be shown that this 2-decomposition of the underlying three literal gadgets determines which of the edges of triangle (a, b, c) are contained in which of the sets E_1 and E_2 . Due to the induced path $9-a-b-9''$ and since $[9, a], [9'', b] \in E_2$, the edge $[a, b]$ cannot be contained in E_2 and thus, is contained in E_1 . Moreover, if $[b, c] \notin E_2$, then there is an induced path $b-9''-c-9$ in the subgraph (V, E_2) , a contradiction. Hence, $[b, c] \in E_2$ and by analogous arguments, $[a, c] \in E_2$. If $[b, c] \notin E_1$ and $[a, c] \notin E_1$, then we obtain a cograph 2-partition. However, it can easily be verified that there is still a degree of freedom and $[a, c], [b, c] \in E_1$ is allowed for a valid cograph 2-decomposition. \square

We are now in the position to prove the NP-completeness of COTREE 2-REPRESENTATION and COTREE 2-DECOMPOSITION by reduction from the following problem.

Problem MONOTONE NAE 3- SAT

<i>Input:</i>	Given a set U of Boolean variables and a set of clauses $\psi = \{C_1, \dots, C_m\}$ over U such that for all $i = 1, \dots, m$ it holds that $ C_i = 3$ and C_i contains no negated variables
<i>Question:</i>	Is there a truth assignment to ψ such that in each C_i not all three literals are set to true?

Theorem 9 (Schaefer 1978; Moret 1997) MONOTONE NAE 3- SAT is NP-complete.

Theorem 10 COGRAPH 2- DECOMPOSITION, and thus, COTREE 2- REPRESENTATION is NP-complete.

Proof Given a graph $G = (V, E)$ and cograph 2-decomposition $\{E_1, E_2\}$, one can verify in linear time whether (V, E_i) is a cograph (Corneil et al. 1985). Hence, COGRAPH 2- PARTITION \in NP.

We will show by reduction from MONOTONE NAE 3- SAT that COGRAPH 2- DECOMPOSITION is NP-hard. Let $\psi = (C_1, \dots, C_m)$ be an arbitrary instance of MONOTONE NAE 3- SAT. Each clause C_i is uniquely identified with a triangle (a_i, b_i, c_i) . Each variable x_j is uniquely identified with a literal graph as shown in Fig. 5 (left). Let $C_i = (x_{i1}, x_{i2}, x_{i3})$ and G_{i1}, G_{i2} and G_{i3} the respective literal graphs.

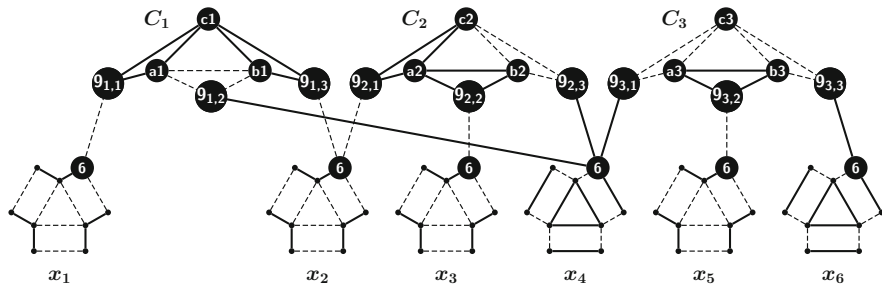


Fig. 7 Shown is the graph Ψ as constructed in the proof of Theorem 10. In particular, Ψ reflects the NAE 3-SAT formula $\psi = \{C_1, C_2, C_3\}$ with clauses $C_1 = (x_1, x_4, x_2)$, $C_2 = (x_2, x_3, x_4)$ and $C_3 = (x_4, x_5, x_6)$. Different literals obtain the same truth assignment true or false, whenever the edges of the triangle in their corresponding literal gadget are contained in the same set E_i of the cograph 2-partition, highlighted by dashed and bold-lined edges

Then, we extend each literal graph G_{i_j} by adding an edge $[6, 9_{i,j}]$. Moreover, we add to G_{i_1} the edges $[9_{i,1}, a_i]$, $[9_{i,1}, c_i]$, to G_{i_2} the edges $[9_{i,2}, a_i]$, $[9_{i,2}, b_i]$, to G_{i_3} the edges $[9_{i,3}, c_i]$, $[9_{i,3}, b_i]$. The latter construction connects each literal graph with the triangle (a_i, b_i, c_i) of the respective clause C_i in a unique way, see Fig. 6. We denote the clause gadgets by Ψ_i for each clause C_i . We repeat this construction for all clauses C_i of ψ resulting in the graph Ψ . An illustrative example is given in Fig. 7. Clearly, this reduction can be done in polynomial time in the number m of clauses.

We will show in the following that Ψ has a cograph 2-decomposition (resp., a cograph 2-partition) if and only if ψ has a truth assignment f .

Let $\psi = (C_1, \dots, C_m)$ have a truth assignment. Then in each clause C_i at least one of the literals $x_{i_1}, x_{i_2}, x_{i_3}$ is set to true and one to false. We assign all edges e of the triangle in the corresponding literal graph G_{i_j} to E_1 , if $f(x_{i_j}) = \text{true}$ and to E_2 , otherwise. Hence, each edge of exactly two of the triangles (one in G_{i_j} and one in $G_{i_{j'}}$) are contained in one E_r and not in E_s , while the edges of the other triangle in $G_{i_{j''}}$, $j'' \neq j, j'$ are contained in E_s and not in E_r , $r \neq s$, as needed for a possible valid cograph 2-decomposition (Lemma 2). We now apply the construction of a valid cograph 2-decomposition (or cograph 2-partition) for each Ψ_i as given in Lemma 2, starting with the just created assignment of edges contained in the triangles in G_{i_j} , $G_{i_{j'}}$ and $G_{i_{j''}}$ to E_1 or E_2 . In this way, we obtain a valid cograph 2-decomposition (or cograph 2-partition) for each subgraph Ψ_i of Ψ . Thus, if there would be an induced P_4 in Ψ with all edges belonging to the same set E_r , then this P_4 can only have edges belonging to different clause gadgets Ψ_k, Ψ_l . By construction, such a P_4 can only exist along different clause gadgets Ψ_k and Ψ_l if C_k and C_l have a literal $x_i = x_{k_m} = x_{l_n}$ in common. In this case, Lemma 2 implies that the edges $[6, 9_{k,m}]$ and $[6, 9_{l,n}]$ in Ψ_i must belong to the same set E_r . Again by Lemma 2, the edges $[9_{k,m}, y]$ and $[9_{k,m}, y']$, $y, y' \in \{a_k, b_k, c_k\}$ as well as the edges $[9_{l,n}, y]$ and $[9_{l,n}, y']$, $y, y' \in \{a_l, b_l, c_l\}$ must be in a different set E_s than $[6, 9_{k,m}]$ and $[6, 9_{l,n}]$. Moreover, respective edges $[5, 6]$ in Ψ_k , as well as in Ψ_l (Fig. 5) must be in E_r , i.e., in the same set as $[6, 9_{k,m}]$ and $[6, 9_{l,n}]$. However, in none of the cases it is possible to find an induced P_4 with all edges in the same set E_r or E_s along different clause

gadgets. Hence, we obtain a valid cograph 2-decomposition, resp., cograph 2-partition of Ψ .

Now assume that Ψ has a valid cograph 2-decomposition (or a cograph 2-partition). Any variable x_{ij} contained in some clause $C_i = (x_{i_1}, x_{i_2}, x_{i_3})$ is identified with a literal graph G_{ij} . Each clause C_i is, by construction, identified with exactly three literal graphs $G_{i_1}, G_{i_2}, G_{i_3}$, resulting in the clause gadget Ψ_i . Each literal graph G_{ij} contains exactly one triangle t_j . Since Ψ_i is an induced subgraph of Ψ , we can apply Lemma 2 and conclude that for any cograph 2-decomposition (resp., cograph 2-partition) all edges of exactly two of three triangles t_1, t_2, t_3 are contained in one set E_r , but not in E_s , and all edges of the other triangle are contained in E_s , but not in E_r , $s \neq r$. Based on these triangles we define a truth assignment f to the corresponding literals: w.l.o.g. we set $f(x_i) = \text{true}$ if the edge $e \in t_i$ is contained in E_1 and $f(x_i) = \text{false}$ otherwise. By the latter arguments and Lemma 2, we can conclude that, given a valid cograph 2-partitioning, the so defined truth assignment f is a valid truth assignment of the Boolean formula ψ , since no three different literals in one clause obtain the same assignment and at least one of the variables is set to *true*. Thus, COGRAPH 2- DECOMPOSITION is NP-complete.

Finally, because COGRAPH 2- DECOMPOSITION and COTREE 2- REPRESENTATION are equivalent problems, the NP-completeness of COTREE 2- REPRESENTATION follows.

As the proof of Theorem 10 allows us to use cograph 2-partitions in all proof steps, instead of cograph 2-decompositions, we can immediately infer the NP-completeness of the following problem for $k = 2$, as well.

Problem COGRAPH k - PARTITION

<i>Input:</i>	Given a graph $G = (V, E)$ and an integer k
<i>Question:</i>	Is there a Cograph k -Partition of G ?

Theorem 11 COGRAPH 2- PARTITION is NP-complete.

As a direct consequence of the latter results, we obtain the following result.

Corollary 2 Let G be a given graph that is not a cograph. The three optimization problems to find the least integer $k > 1$ so that there is

a Cograph k -Partition, or
a Cograph k -Decomposition, or
a Cotree k -Representation

for the graph G , are NP-hard.

5.2 Integer linear program

Let $G = (V, E)$ be a given graph with maximum degree Δ . We want to find a cograph- k -decomposition, resp., partition $\Pi = \{E_1, \dots, E_k\}$ for the least integer k . Theorem 8 implies that the least integer k is always less or equal to $\Delta + 1$.

We define binary variables E_{xy}^i for all $x, y \in V$ and $1 \leq i \leq \Delta + 1$ s.t. $E_{xy}^i = 1$ if and only if the edge $[x, y] \in E$ is contained in class E_i of Π . Moreover, we define the binary variables M^i with $1 \leq i \leq \Delta + 1$ so that $M^i = 1$ if and only if the class $E_i \in \Pi$ is non-empty in our construction. In other words, $\sum_{1 \leq i \leq \Delta+1} M^i$ will be the cardinality of Π .

In order to find the cograph decomposition, resp., partition Π of G having the fewest number of elements we need the following objective function.

$$\min \sum_{1 \leq i \leq \Delta+1} M^i \quad (8)$$

If we want to find a cograph-decomposition and hence, that each edge is contained in at least one class E_i of Π we need the next constraint.

$$\sum_{1 \leq i \leq \Delta+1} E_{xy}^i \geq 1 \quad \text{for all } [x, y] \in E. \quad (9)$$

In contrast, if we want to find a cograph-partition and hence, that each edge is contained in exactly one class E_i of Π we need this constraint.

$$\sum_{1 \leq i \leq \Delta+1} E_{xy}^i = 1 \quad \text{for all } [x, y] \in E. \quad (9')$$

Moreover, we must ensure that non-edges $[x, y] \notin E$ are not contained in any class of Π which is done with the next constraint.

$$\sum_{1 \leq i \leq \Delta+1} E_{xy}^i = 0 \quad \text{for all } [x, y] \notin E. \quad (10)$$

Whenever there is a class E_i containing an edge $[x, y] \in E$ and hence, if $E_{xy}^i = 1$, then we must set $M^i = 1$.

$$E_{xy}^i \leq M^i \quad \text{for all } 1 \leq i \leq \Delta + 1 \text{ and } x, y \in V. \quad (11)$$

Finally we have to ensure that each subgraph $G_i = (V, E_i)$ is a cograph, and thus, does not contain induced P_4 's, which is achieved with the following constraint.

$$E_{xy}^i + E_{yu}^i + E_{uv}^i - E_{xu}^i - E_{xv}^i - E_{yv}^i \leq 2 \quad (12)$$

for all $1 \leq i \leq \Delta + 1$ and all ordered tuples (x, y, u, v) of distinct $x, y, u, v \in V$.

This ILP-formulation needs $O(\Delta|V|^2)$ variables and $O(|E| + \Delta + |V|^4)$ constraints.

References

- Achlioptas D (1997) The complexity of g -free colourability. *Discrete Math* 165:166:21–30 (Graphs and combinatorics)
- Bernini A, Ferrari L, Pinzani R (2009) Enumeration of some classes of words avoiding two generalized patterns of length three. *J Autom Lang Comb* 14(2):129–147
- Bilotta S, Grazzini E, Pergola E (2013) Counting binary words avoiding alternating patterns. *J Integer Seq* 16:Article 13.4.8
- Böcker S, Dress A (1998) Recovering symbolically dated, rooted trees from symbolic ultrametrics. *Adv Math* 138:105–125
- Brändén P, Mansour T (2005) Finite automata and pattern avoidance in words. *J Comb Theory Ser A* 110(1):127–145
- Brandstädt A, Le V, Spinrad J (1999) Graph classes: a survey SIAM monographs on discrete mathematics and applications. Society of Industrial and Applied Mathematics, Philadelphia
- Burstein A, Mansour T (2002) Words restricted by patterns with at most 2 distinct letters. *Electron J Comb Number Theory* 9(2):1–16
- Cornel D, Perl Y, Stewart L (1985) A linear recognition algorithm for cographs. *SIAM J Comput* 14(4):926–934
- Cornel DG, Lerchs H, Stewart Burlingham LK (1981) Complement reducible graphs. *Discrete Appl Math* 3:163–174
- Dorbec P, Montassier M, Ochem P (2014) Vertex partitions of graphs into cographs and stars. *J Graph Theory* 75(1):75–90
- El-Mallah E, Colbourn C (1988) The complexity of some edge deletion problems. *IEEE Trans Circuits Syst* 35(3):354–362
- Fitch W (2000) Homology: a personal view on some of the problems. *Trends Genet* 16:227–231
- Gimbel J, Nešetřil J (2002) Partitions of graphs into cographs. *Electronic notes in discrete mathematics*, vol 11. In: The ninth quadrennial international conference on graph theory, combinatorics, algorithms and applications, pp 705–721
- Hammack R, Imrich W, Klavžar S (2011) Handbook of product graphs, 2nd edn. CRC Press, Boca Raton
- Hellmuth M, Hernandez-Rosales M, Huber KT, Moulton V, Stadler PF, Wieseke N (2013a) Orthology relations, symbolic ultrametrics, and cographs. *J Math Biol* 66(1–2):399–420
- Hellmuth M, Imrich W, Kupka T (2013b) Partial star products: a local covering approach for the recognition of approximate Cartesian product graphs. *Math Comput Sci* 7(3):255–273
- Hellmuth M, Wieseke N (2015) On symbolic ultrametrics, cotree representations, and cograph edge decompositions and partitions. In: Xu D, Du D, Du D (eds) *Computing and combinatorics. Lecture Notes in Computer Science*, vol 9198. Springer International Publishing, Cham, pp 609–623
- Hellmuth M, Wieseke N (2016) From sequence data Including orthologs, paralogs, and xenologs to gene and species trees, Chap. 21. Springer International Publishing, Cham
- Hellmuth M, Wieseke N, Lechner M, Lenhof H, Middendorf M, Stadler P (2015) Phylogenomics with paralogs. *Proc Natl Acad Sci* 112(7):2058–2063
- Hernandez-Rosales M, Hellmuth M, Wieseke N, Huber KT, Moulton V, Stadler PF (2012) From event-labeled gene trees to species trees. *BMC Bioinform* 13(19):S6
- Lafond M, El-Mabrouk N (2014) Orthology and paralogy constraints: satisfiability and consistency. *BMC Genomics* 15(Suppl 6):S12. doi:[10.1186/1471-2164-15-S6-S12](https://doi.org/10.1186/1471-2164-15-S6-S12).
- Lechner M, Findeiß S, Steiner L, Marz M, Stadler PF, Prohaska SJ (2011) Proteinortho: detection of (co-)orthologs in large-scale analysis. *BMC Bioinform* 12:124
- Lechner M, Hernandez-Rosales M, Doerr D, Wiesecke N, Thevenin A, Stoye J, Hartmann RK, Prohaska SJ, Stadler PF (2014) Orthology detection combining clustering and synteny for very large datasets. *PLoS ONE* 9(8):e105,015
- Lerchs H (1971a) On cliques and kernels. Technical Report, Department of Computer Science University of Toronto
- Lerchs H (1971b) On the clique–kernel structure of graphs. Technical Report, Department of Computer Science, University of Toronto
- Liu Y, Wang J, Guo J, Chen J (2011) Cograph editing: complexity and parametrized algorithms. In: Fu B, Du DZ (eds) *COCOON 2011, Lecture Notes in Computer Science*, vol 6842. Springer, Berlin, Heidelberg, pp 110–121

- Liu Y, Wang J, Guo J, Chen J (2012) Complexity and parameterized algorithms for cograph editing. *Theoret Comput Sci* 461:45–54
- Misra J, Gries D (1992) A constructive proof of vizing's theorem. *Inf Process Lett* 41(3):131–133
- Moret BM (1997) The theory of computation. Addison-Wesley Longman Publishing Co., Inc. Boston, MA
- Pudwell L (2008a) Enumeration schemes for pattern-avoiding words and permutations. ProQuest, New Brunswick, New Jersey
- Pudwell L (2008b) Enumeration schemes for words avoiding patterns with repeated letters. *Electron. J. Comb. Number Theory* 8(40):1–19
- Schaefer T (1978) The complexity of satisfiability problems. In: *Proceedings of the tenth annual ACM symposium on theory of computing, STOC '78*. ACM, New York, NY, USA, pp 216–226
- Vizing VG (1964) On an estimate of the chromatic class of a p-graph. *J Math Biol* 3:23–30 (Russian)
- Zhang P (2014) A study on generalized solution concepts in constraint satisfaction and graph colouring. Master's thesis, University of British Columbia, Canada