Exact-2-Relation Graphs

Yangjing Long[a], Peter F. Stadler[b,c,d,e,f]

[a] *School of Mathematics and Statistics, Center China Normal University, No. 152, Luoyu Road, Wuhan, Hubei, P. R. China*
[b] *Bioinformatics Group, Department of Computer Science; Interdisciplinary Center for Bioinformatics; German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig; Competence Center for Scalable Data Services and Solutions Dresden-Leipzig; Leipzig Research Center for Civilization Diseases; and Centre for Biotechnology and Biomedicine, University of Leipzig, Härtelstraße 16-18, D-04107 Leipzig, Germany*
[c] *Max Planck Institute for Mathematics in the Sciences, Inselstraße 22, D-04103 Leipzig, Germany*
[d] *Institute for Theoretical Chemistry, University of Vienna, Währingerstraße 17, A-1090 Wien, Austria*
[e] *Facultad de Ciencias, Universidad Nacional de Colombia, Sede Bogotá, Colombia*
[f] *The Santa Fe Institute, 1399 Hyde Park Rd., Santa Fe, NM 87501, United States*

Abstract

Pairwise compatibility graphs (PCGs) with non-negative integer edge weights recently have been used to describe rare evolutionary events and scenarios with horizontal gene transfer. Here we consider the case that vertices are separated by exactly two discrete events: Given a tree T with leaf set L and edge-weights $\lambda : E(T) \to \mathbb{N}_0$, the non-negative integer pairwise compatibility graph $\text{nniPCG}(T, \lambda, 2, 2)$ has vertex set L and xy is an edge whenever the sum of the non-negative integer weights along the unique path from x to y in T equals 2. A graph G has a representation as $\text{nniPCG}(T, \lambda, 2, 2)$ if and only if its point-determining quotient $G/\!\!\stackrel{\backsim}{}$ is a block graph, where two vertices are in relation $\stackrel{\backsim}{}$ if they have the same neighborhood in G. If G is of this type, a labeled tree (T, λ) explaining G can be constructed efficiently. In addition, we consider an oriented version of this class of graphs.

Keywords: Pairwise compatibility graphs, edge-labeled trees, thin graphs, block graphs, oriented graphs
2000 MSC: 05C75, 05C05, 92B10

1. Introduction

Consider a tree T with leaf set V and a non-negative edge-weight function $\ell : E(T) \to \mathbb{R}_0^+$. Denote by $\mathcal{P}(x, y)$ the unique path between x and y in T. The canonical distance function $d_{T,\ell} : V \times V \to \mathbb{R}_0^+$ is then defined by

$$d_{T,\ell}(x, y) = \sum_{e \in \mathcal{P}(x,y)} \ell(e) \tag{1}$$

This definition is the starting point for mathematical phylogenetics, which is centered around finite additive metric spaces and their generalizations [1, 2]. It also serves as a basis for defining a large class of graphs that in the recent past has received considerable attention. The *pairwise compatibility graphs* $G = \text{PCG}(T, \ell, d_{\min}, d_{\max})$ has vertex set V and edges

$$uv \in E(G) \quad \text{if and only if} \quad d_{\min} \le d_{T,w}(u, v) \le d_{\max} \tag{2}$$

Originally introduced in the context of phylogenetics [3], they have received considerable interest in the last years, see [4, 5] and the references therein, as well as [6, 7, 8]. A further generalization of "multi-interval" PCGs in explored in [9]. In the setting of PCGs and most phylogenetic applications one usually stipulates that $\ell(e) > 0$, measuring e.g. the time between two distinct events associated with adjacent vertices of T. A class of graphs that is conceptually closely are the exact k-leaf powers [10], for which $\lambda(e) = 1$ for all edges of T and $d_{\min} = d_{\max} = k$.

In an alternative interpretation, $\ell(e)$ models the number of *discrete* evolutionary events along an edge of e of T. This is of interest in particular in the context of so-called *rare genomic changes* such as the gene or loss of a particular

gene of gene family or a particular genomic rearragement [11]. Some of these convey phylogenetic information that is (nearly) free of homoplasy, i.e., the independent occurrance in independent lineages. Examples of such rare events are the emergence of novel microRNA families [12] or rearrangements of the genomic gene order [13]. Since such events are very unlikely to have occurred more than once in the same manner, they identify phylogenetic groupings that share such an innovation with very little ambiguity. This provides information to resolve also parts the phylogenetic tree where classical, sequence-based methods fail [14]. In this context it is necessary to allow $\ell(e) = 0$ because the events of interest are by definition so rare that not all taxa will be distinguished by them. In the same vein it important that events are discrete and hence an integer-valued weight function $\lambda : E(T) \to \mathbb{N}_0$. Both conditions on ℓ cause subtle but important differences in comparison with the usual definition of PCGs that requires non-zero edge length but otherwise allows arbitrary real values. We will denote these "non-negative integer pairwise compatibility graph" by nniPCG to distinguish them from the better studied class of PCGs with non-zero real-valued edge weights λ.

The special case in which two leaves x and y of T are separated by a single event, corresponding to graphs of the form nniPCG$(T, \lambda, 1, 1)$, was explored in [15] as a model of rare events in evolution. It turns out that this graph class coincides with the forests. The graphs nniPCG$(T, \lambda, 1, \infty)$ requiring at least one event along the path between two leaves also have a very simple structure: they are exactly the complete multipartite graphs [16].

Considering a rooted tree $\vec{T}$ instead of an unrooted tree T, it is natural to consider the digraphs with edges (x, y) whenever a certain number of events occured between the last common ancestor $lca(x, y)$ and y. This construction appears naturally in the context of horizontal gene transfer (HGT), where one asks whether $lca(x, y)$ and y are separated by at least one HGT event[1] This gives rise to the class of Fitch graphs [17], which form a subclass of di-cographs introduced by [18]. Their underlying undirected graphs are exactly the nniPCG$(T, \lambda, 1, \infty)$, i.e., the complete multipartite graphs.

A related construction requires a certain number of events between $lca(x, y)$ and y *and* excludes all events between $lca(x, y)$ and x. This class of graphs appears naturally when events are directed, i.e., when it is (in general) no possible to revert the effect of an operation in a single step. Probably the best studied type of single genomic events of this type are so-called Tandem-Duplication-Random-Loss events, during which a genomic interval is duplicated and then one of the two copies of each gene is lost at random [19]. The antisymmetric digraphs obtained by single events are characterized in [15].

Our interest in the graphs nniPCG(T, λ, k, k) for general $k \geq 1$ also stems from rare-event phylogenetic data. Since we assume an underlying tree, the distance matrix d_T is additive and its entries are small non-negative integers. The fact that all edge lengths $\ell(e)$ are also integers of course imposes additional constraints. As demonstrated e.g. in the context of orthology assignment (a related problem with vertex labeled trees for which the corresponding graphs turn out to be cographs), graph editing can be employed to correct empirically estimated input graphs [20]. This approach requires, however, that constraint on the graphs that can appear are known. In the case of rare-event phylogenetics, we know that the graph with edge set $\{xy | d_T(x, y) = k\}$ must be a nniPCG(T, λ, k, k). In the rare-event scenario, the number of pairs of nodes with $d_T(x, y) = k$ will quickly decrease with k, so that the empirical input graphs will have few edges for larger values of k and thus rarely reveal obstructions. Hence only small value of k are of practical value for detecting measurement errors in the data. Since the nniPCG$(T, \lambda, 1, 1)$ are forests, the corresponding graph editing problem amounts to identifying spanning forests, and possible false positive events are edges in cycles. False negatives are not detectable for $k = 1$ since there are no non-tree graphs that would become trees by inserting edges. They could be detected, however, as missing edges in the empirical graph for $k = 2$ compared to the most similar member of nniPCG$(T, \lambda, 2, 2)$. In this contribution, therefore, we are interested in the characterization of the graphs nniPCG$(T, \lambda, 2, 2)$, in which edges correspond to exactly two events between two leaves. This graph class is very different from the exact-2-leaf power graphs, which are known to coincide with the disjoint unions of cliques [21, 10]. In contrast, we shall see below that e.g. every path also has a representation as nniPCG$(T, \lambda, 2, 2)$.

This contribution is organized as follows: We first consider a few general properties of the slightly more general *exactly-k-relation* $\overset{k}{\sim}$ before investigating for some small graphs and simple graph families whether they can be respresented with respect to the *exactly-2-relation* $\overset{2}{\sim}$ on the leaf set of some tree. Here, we consider the case that $d_T(x, y) > 0$, i.e., that all leaves are separated by at least one event, and then relax this constraint and characterize the entire graph class nniPCG$(T, \lambda, 2, 2)$ for non-negative, integer λ. Our main result is that these graphs are those whose

---

[1]HGT refer to the import of gene from an unrelated species e.g. through an infection, ingestion, acquisition via a plasmid.
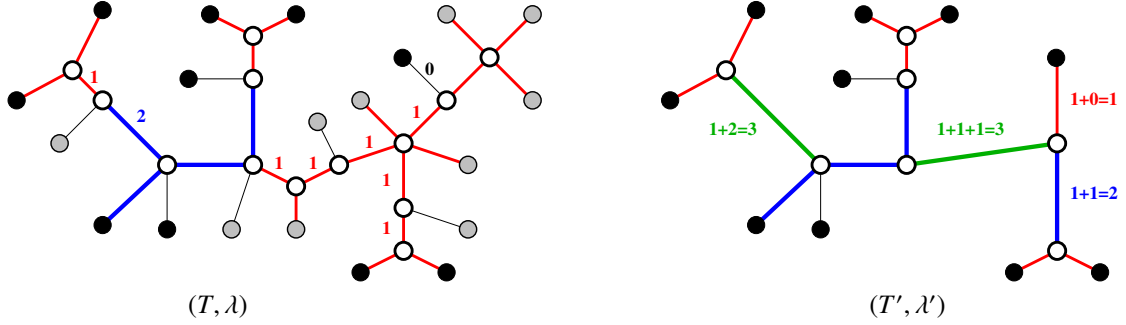
(T, λ) $\qquad\qquad\qquad\qquad$ (T', λ')

Figure 1: Illustration of Definition 2. The edge-labeled tree (T', λ') on the r.h.s. is *displayed* by (T, λ). It is obtained as the restriction of T to the non-gray vertices. Correponding vertices are shown in matching locations. All edges e that are "merged" into single edges have their weights annotated. Edges that remain unchanged or deleted are only shown in color (black for $\lambda(e) = 0$, red for $\lambda(e) = 2$, and blue of $\lambda(e) = 2$) without displaying the weight explicitly.

quotient with respect to the false twin (R-thinness) relation is a block graph. We then consider the oriented version of the problem and give characterization in terms of forbidden subgraphs.

2. Simple Properties of the Exactly-k-Relation

We shall see that the restriction to integer edge weights on the one hand, and the admission of zero-weights on the other hand, make the graphs $\text{nniPCG}(T, \lambda, k, k)$ quite different from the exact-k-leaf power graphs studies systematically in [10]. While it is true that every $\text{PCG}(T, \lambda, d_{\min}, d_{\max})$ with non-negative real weigths λ and bounds $d_{\min}$ and $d_{\max}$ also has a representation as $\text{PCG}(T, \hat{\lambda}, \hat{d}_{\min}, \hat{d}_{\max})$ with integer weights and bounds [22, Lemma 2], the restriction to integer weights clearly affects the definition of graph *classes*. For instance, the PCG class with rational weights and bounds $d_{\min} = d_{\max} = 1$ contains the $\text{nniPCG}(T, \lambda, k, k)$ for all $k \in \mathbb{N}$. Throughout this contribution we use a notation that is inspired by related work in mathematical phylogenetics.

Definition 1. *Let (T, λ) be an unrooted tree with leaf set L and edge-labeling function $\lambda : E(T) \to \mathbb{N}_0$. For $x, y \in L$ we consider the* exactly-k-relation *$\overset{k}{\sim}$ defined by $x \overset{k}{\sim} y$ if the (unique) path (x, y) from x to y in T satisfies $\sum_{e \in \mathcal{P}(x,y)} \lambda(e) = k$. Furthermore, we say (T, λ)* explains *a graph $G(L, E)$ (with respect to the exactly-k-relation) if $\{x, y\} \in E$ if and only if $x \overset{k}{\sim} y$.*

We consider unrooted instead of rooted tree since the distances $d_T(x, y)$ and thus the *exactly-k-relation* $\overset{k}{\sim}$ contains no information on position of root. In fact, it is well known [23, 24] that a metric d of the form (1) uniquely defines an unrooted tree. Therefore, one can only hope to reconstruct the unrooted tree T.

Definition 2. *The edge-labeled tree (T, λ)* displays *the edge labeled tree (T', λ') if (T', λ') can be obtained from (T, λ) by first removing every edge and vertex from (T, λ) that is not contained in a path connecting two leaves of T', and then contracting every path $\mathcal{P}(u, v)$ in the remainder of T that has only interior vertices of degree 2 by a single edge e' in T' with label $\lambda'(e') = \sum_{e \in \mathcal{P}(u,v)} \lambda(e)$.*

In particular, therefore it is sufficient to consider phylogenetic trees, that is, trees T in which every interior node $x \in V(T) \setminus L$ has degree at least 3. Fig. 1 gives an example. A simple, but important consequence of Definition 1 is the following

Lemma 3. *If (T, λ) displays (T', λ'), (T, λ) explains $G(L, \overset{k}{\sim})$ and (T', λ') explains $G'(L', \overset{k}{\sim})$. Then $G'(L', \overset{k}{\sim}) = G(L, \overset{k}{\sim})[L']$, the subgraph of $G(L, \overset{k}{\sim})$ induced by L'.*

Proof. If (T, λ) displays (T', λ') then $\sum_{e \in \mathcal{P}_T(u,v)} \lambda(e) = \sum_{e \in \mathcal{P}_T(u,v)} \lambda'(e)$ for all $u, v \in L(T') \subseteq L(T)$, and thus we conclude that for all $u, v \in L(T')$, we have $u \overset{k}{\sim}_{(T',\lambda')} v$ if and only if $u \overset{k}{\sim}_{(T',\lambda')} v$, i.e., $G'(L', \overset{k}{\sim})$ is the subgraph of $G(L, \overset{k}{\sim})$ induced by L'. $\qquad\square$
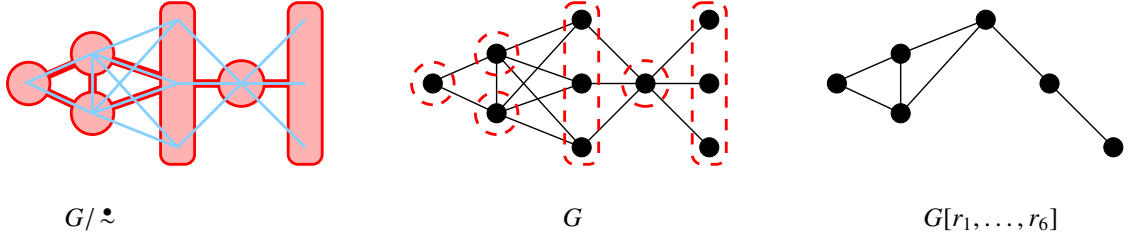
3

Figure 2: The false twin classes are indicated by dashed outlines in the graph G (middle) and form the vertices (in red) of the quotient graph $G/\backsimeq$ on the right. For comparison, the orginal edges between the $\backsimeq$ classes are shown. The subgraph $G[r_1, r_2, r_3, r_4, r_5, r_6]$ of G induced by one representative r_i for each $\backsimeq$-class is shown on the left. It is isomorphic to $G/\backsimeq$ irrespective of the choice of the representative vertices r_i for the $\backsimeq$ classes.

It follows that "being explained with respect to the exactly-k-relation" is a hereditary graph property for all k.

We also note the following immediate consequence of the definition.

Lemma 4. *If (T, λ) explains G with respect to $\overset{1}{\sim}$, then $(T, k\lambda)$ explains G with respect to $\overset{k}{\sim}$.*

Lemma 5. *Let G be a graph with connected components G_i, $i = 1, \ldots, N$. Then there is an edge-labeled tree (T, λ) explaining T with respect to $\overset{k}{\sim}$ if and only if there are edge labeled trees (T_i, λ_i) explaining G_i for all $i = 1, \ldots, N$.*

Proof. The condition is necessary because of heredity. In order to see sufficiency, we can construct (T, λ) from the disjoint union of the (T_i, λ_i) in the following way: first we arrange them as an arbitrary tree $\mathcal{T}$. Then we replace each $(K_2, \lambda(e) = k)$ by S_2 with the two edges e' and e'' labeled such that $\lambda(e') + \lambda(e'') = k$. Now choose for each tree T_i an arbitrary inner vertex x_i in $T_i \neq K_1$ and the unique vertex x_i if $T_i = K_1$. Finally, we connect x_i and x_j by an edge e_{ij} with $\lambda(e_{ij}) = k + 1$ if and only if T_i and T_j are adjacent in $\mathcal{T}$. To verify that (T, λ) indeed explains G we observe: (i) If x and y are leafs from different connected components of G, they are located in different subtrees T_i and thus the path connecting them contains one of the edges label $k + 1$, thus x and y are not in relation $\overset{k}{\sim}$. $\square$

It is therefore sufficient to consider connected graphs.

Definition 6. *An edge-labeled graph (T, λ) is canonical if T is phylogenetic and $\lambda(e) \neq 0$ for all interior edges.*

Lemma 7. *Let $(\hat{T}, \hat{\lambda})$ be the edge labeled tree obtained from (T, λ) by (1) replacing every path $\mathcal{P}(u, v)$ in T whose interior vertices have degree 2 by a single edge e' in T' with label $\lambda'(e') = \sum_{e \in \mathcal{P}(u,v)} \lambda(e)$ and (2) contracting every interior edge with $\lambda(e) = 0$. The tree $(\hat{T}, \hat{\lambda})$ is uniquely defined, canonical, and explains the same graph as (T, λ).*

Proof. The maximal paths with interior vertices of degree 2 in T are disjoint and thus can be treated independently. By construction, any such path $\mathcal{P}$ can also be stepwisely replaced by edges, eventually arriving at the same edge weight for the single edge that remains. Given T, the resulting tree $\hat{T}$ is therefore unique and contains no vertex of degree 2. It is therefore phylogenetic. Since an interior edge with label $\lambda(e) = 0$ does not contribute the total weight of any path that runs through it, it can be contracted without changing the total path weights between leaves. Thus (T, λ) and $(\hat{T}, \hat{\lambda})$ explain the same graph. $\square$

Consider two leaves $x, y \in L$ in an edge-labeled tree (T, λ) such that $x \overset{0}{\sim} y$, i.e., $d_T(x, y) = 0$, and another leaf $z \in L \setminus \{x, y\}$. The triangle inequalities $d_T(x, z) \leq d_T(x, y) + d_T(y, z)$ and $d_T(y, z) \leq d_T(y, x) + d_T(x, z)$ implies $d_T(x, z) = d_T(y, z)$. Thus x and y have the same neighbors in graph G explained by T, i.e., $N_G(x) = N_G(y)$.

Definition 8. *Let G be a graph. For each $x \in V(G)$ denote by $N(x)$ the neighbors of x. Two vertices x and y are* false twins*, $x \backsimeq y$, if $N(x) = N(y)$.*

In contrast, true twins, which play no role here, satisfy $N(x) \cup \{x\} = N(y) \cup \{y\}$. By definition, false twins $x \backsimeq y$ are non-adjacent, while true twins are always adjacent [25].

The false twin (R-thinness) relation $\backsimeq$ has been well studied in the literature, in particular in the context of graph products [26]. It is well known that $\backsimeq$ is an equivalence relation, see e.g. [26, sect. 8.2]. Its equivalence classes,

4

which we denote by R_i, $i = 1, \ldots, h$, are totally disconnected in G because, by definition, $x \notin N(x)$. Denote by $G[r_1, r_2, \ldots, r_h]$ be the subgraph of G induced by one arbitrarily chosen representative $r_i \in R_i$ of each false twin class. Since for any $x \in R_i$ and $y \in R_j$ we have $xy \in E(G)$ if and only if $x'y' \in E(G)$ for all $x' \in R_i$ and all $y' \in R_j$ we observe that $G[r_1, r_2, \ldots, r_h]$ and the quotient graph $G/\overset{\bullet}{\sim}$ are isomorphic. An illustration is given in Fig. 2.

Lemma 9. *Let (T, λ) be a canonical tree explaining a connected graph G with respect to $\overset{k}{\sim}$, and let W be a set of sibling leaves attached to the same parent q with $\lambda(qw) = 0$ for all $w \in W$. Then W is contained in a false twin class for the graph G explained by (T, λ) with respect to $\overset{k}{\sim}$ for all $k > 0$.*

Proof. Consider a node $y \in L \setminus W$. Then the total weight of the path between y and every $w \in W$ is the same. Furthermore, the total path weight between any two vertices in $w', w'' \in W$ is 0, i.e., there is no edge between w' and w''. Thus $N(w') = N(w'')$, i.e., $w' \overset{\bullet}{\sim} w''$ for all $w', w'' \in W$. $\square$

A graph is called *R-thin* [27], *point determining graph* [28] or *mating graph* [29] if $\overset{\bullet}{\sim}$ is discrete, i.e., every false twin class consists of only a single point. Clearly, $G/\overset{\bullet}{\sim}$ is R-thin. R-thin graphs have also been studied from the point of view of combinatorial enumeration [30, 31]. Algorithms for prime-factorization of graphs, furthermore, often operate on $G/\overset{\bullet}{\sim}$, since R-thinness ensures uniqueness of the factorization and allows for highly efficient algorithms [27, 32, 26]. Below we show that it also suffices to consider $G/\overset{\bullet}{\sim}$, i.e., R-thin graphs, in our setting. Indeed, a simple consequence of Lemma 9 is

Corollary 10. *If G is R-thin and (T, λ) is a canonical tree explaining G with respect to $\overset{k}{\sim}$, then $\overset{0}{\sim}$ is discrete.*

Algorithm 1 Compute (T, λ) from (T^*, λ^*) and false twin classes R_i with representatives $r_i \in R_i$.

Require: (T^*, λ^*), (r_i, R_i) for $i = 1, \ldots, h$.
1: **for all** false twin classes R_i with $|R_i| > 1$ **do**
2:    $q \leftarrow$ unique neighbor of leaf r_i in (T^*, λ^*)
3:    remove r_i from (T^*, λ^*)
4:    **if** $\lambda^*(qr_i) \neq k/2$ **then**
5:       insert all leaves $r \in R_i$ with edges qr and $\lambda(qr) = \lambda^*(qr_i)$
6:    **else**
7:       insert a node q' and the edge qq' with $\lambda(qq') = \lambda^*(qr_i)$
8:       insert all leaves $r \in R_i$ with edges $q'r$ and $\lambda(q'r) = 0$
9:    **end if**
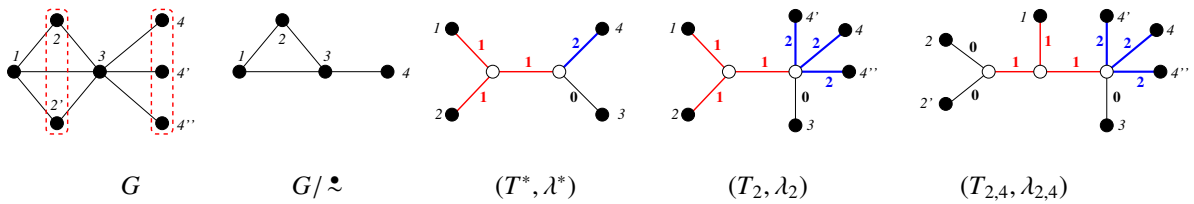10: **end for**
For an illustrative example see Fig. 3.

| G | $G/\overset{\bullet}{\sim}$ | (T^*, λ^*) | (T_2, λ_2) | $(T_{2,4}, \lambda_{2,4})$ |

Figure 3: The graph G has two non-trivial false twin classes $R_2 = \{2, 2'\}$ and $R_4 = \{4, 4', 4''\}$. Its R-thin quotient $G/\overset{\bullet}{\sim}$ is explained w.r.t. $\overset{2}{\sim}$ by the tree (T^*, λ^*). The tree (T_2, λ_2) is obtained from (T^*, λ^*) using first alternative in Alg. 1 (line 5), amounting to replacing 2 by $|R_2|$ leaves with the same parent. In the second step, vertex 4 is replaced by the subtree according to the second alternative in Alg. 1 (lines 7 and 8).

Theorem 11. *G can be explained w.r.t. $\overset{k}{\sim}$ if and only $G/\overset{\bullet}{\sim}$ can explained w.r.t. $\overset{k}{\sim}$. If (T^*, λ^*) is a canonical tree explaining $G/\overset{\bullet}{\sim}$, then a canonical tree (T, λ) explaining G is obtained by Algorithm 1.*
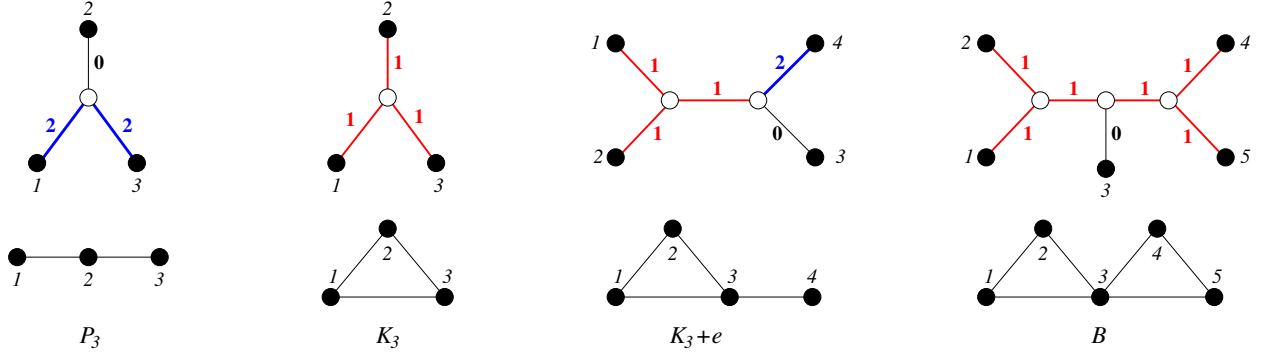
5

Figure 4: The two connected graphs P_3 and K_3 on three vertices, the graph $K_3 + e$ on four vertices, and bowtie graph B on five vertices are explained by unique $\{0, 1, 2\}$-edge-labeled trees with respect to $\overset{2}{\sim}$ and discrete relation $\overset{0}{\sim}$.

Proof. Since G can be explained and $G/\overset{\bullet}{\sim}$ is an induced subgraph of G, $G/\overset{\bullet}{\sim}$ can be explained w.r.t. to $\overset{k}{\sim}$ by a tree that we denote by (T^*, λ^*). Let r be the representative of the false twin class R of G, and let $x \in R$. Insert x into (T^*, λ^*) are a sibling of r and set $\lambda(x) = \lambda(r) = \lambda^*(r)$. Then x and r have the same total path weights to all other vertices. This remains true if each leaf r in (T^*, λ^*) is replaced in this manner by the set R of sibling vertices with $r \in R$. Since no two vertices in R are adjacent we require that $\lambda(r) + \lambda(x) \neq k$, i.e., $\lambda^*(r) \neq k/2$. If this conditions is satisfied, then (T, λ) explains G with respect to $\overset{k}{\sim}$.

If $\lambda^*(qr) = k/2 \geq 1$, Alg. 1 inserts an extra vertex q' adjacent to q with $\lambda(qq') = \lambda^*(qr) \neq 0$. Since we assumed that (T^*, λ^*) was canonical, q has at least two more neighbors, i.e., the resulting tree is again canonical. Since R is attached with edge weights $\lambda(q'r) = 0$ we conclude that (i) the total path weight between r' and q is $k/2$ and (ii) $\lambda(r'q') + \lambda(q'r'') = 0$ for all $r', r'' \in R$, i.e., r' and r'' are not adjacent in the graph explained by (T, λ). Hence r' and r'' have the same neighbors and thus belong the same false twin class of G. Since the total path weights between all representatives of false twin classes are preserved by this construction, (T, λ) indeed explains G with respect to $\overset{k}{\sim}$. We note, finally, that (T, λ) is again canonical because q' has at least three neighbors (the parent q and at least members of R), and all interior edges the resulting tree have non-zero labels as long as (T^*, λ^*) was canonical. $\square$

From here on we will therefore assume the (T, λ) is canonical, i.e., it has non-zero labels for all inner edges of T. It is important to note, however, that we still need to consider zero weights on the edges incident with leaves. For instance, it not difficult to check that the graph $G/\overset{\bullet}{\sim}$ in Fig. 3, i.e., $K_3 + e$, cannot be explained by a tree with only non-zero edge weights.

3. Graphs Explained w.r.t. $\overset{2}{\sim}$

We will first consider the special case of edge labelings with discrete $\overset{0}{\sim}$. In this case every interior vertex of T is incident with at most one zero-weight edge.

The trivial cases K_1 and K_2 are explained by the trees K_1 and K_2 with label $\lambda(e) = k$ at the unique edge e, respectively. For $|L| = 3$ there is only a single phylogenetic tree, the star S_3 with three leaves and two connected graphs, P_3, and K_3, see Fig. 4. We denote the edges from the center to leaf x_i by e_i, $1 \leq i \leq 3$. Fig. 4 also shows that class of graphs explained w.r.t. $\overset{2}{\sim}$ is much larger than the exact-2-leaf powers graphs, which comprise only the disjoint unions of cliques [21, 10].

Lemma 12. *There are unique labelings λ_{P_3} and λ_{K_3} of the tree S_3 with discrete $\overset{0}{\sim}$ that explain the graphs P_3 and K_3, respectively:*
$\lambda_{P_3}(e_1) = \lambda_{P_3}(e_2) = 2$ and $\lambda_{P_3}(e_3) = 0$;
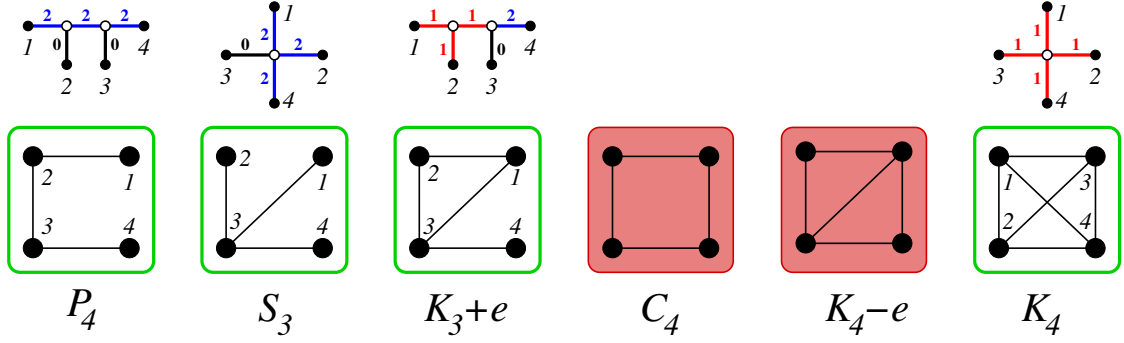$\lambda_{K_3}(e_1) = \lambda_{K_3}(e_2) = \lambda_{K_3}(e_3) = 1$;

6

Figure 5: Connected graphs on four vertices that are explained (green outline) or cannot be explained (red background) with respect to $\overset{2}{\sim}$. The explaining trees are shown above the graphs. The path P_4 and the star S_3 are explained because all trees are already explained by $\overset{1}{\sim}$. The K_4 is explained by Lemma 14. The graph $K_3 + e$ can be explained by a fully resolved phylogenetic tree with a edge labeling explicitly given in the proof of Lemma 15. In contrast, C_4 and $K_4 - e$ cannot be explained with respected to $\overset{2}{\sim}$ with an edge labeling with discrete $\overset{0}{\sim}$ according to Lemma 15.

Proof. S_3 contains three paths on length two. Adopting the notation of Fig. 4 for both cases P_3 and K_3 we need $\lambda(e_1) + \lambda(e_2) = 2$ and $\lambda(e_3) + \lambda(e_2) = 2$. Therefore $\lambda(e_1) \in \{0, 1, 2\}$. Explicitly enumerating the three cases yields:

$\lambda(e_1) = 0$ implies $\lambda(e_2) = 2$ and thus $\lambda(e_3) = 0$, in which case $x_1 \overset{0}{\sim} x_3$, contradicting the fact $\overset{0}{\sim}$ is discrete.

$\lambda(e_1) = 1$ implies $\lambda(e_2) = \lambda(e_3) = 1$, and thus $G(\overset{2}{\sim}) = K_3$.

$\lambda(e_1) = 2$ implies $\lambda(e_2) = 0$ and thus $\lambda(e_3) = 2$, whence $G(\overset{2}{\sim}) = P_3$. $\qquad\square$

Lemma 13. *The path P_4 on four vertices $x - y - z - u$ is explained only be the tree $T = (xy)p - q(zu)$ with labels $\lambda(xp) = \lambda(qu) = \lambda(pq) = 2$ and $\lambda(yp) = \lambda(qz) = 0$.*

Proof. First we observe that the path P_4 on four vertices cannot be explained by any labeling of a S_4. This leaves the fully resolved tree on four vertices. Its interior edge pq cannot be labeled 0. First consider $\lambda(pq) = 1$. It cannot contain an S_3 with all three edges labeled 1 since this would induce a triangle, i.e., at most one neighbor of p, say x, is attached by a 1-edge. The other neighbor of p, call it y, then must be attached by a 0-edge, since otherwise y is isolated. In order for y not to be isolated, q also must have a neighbor, that is connected via a 1-edge, say $\lambda(qz) = 1$. The same argument implies the the remaining leaf u must be connected to q with $\lambda(qu) = 0$. This tree, however, explains the non-connected graph $K_2 \cup K_2$. Thus $\lambda(pq) = 2$. Connectedness implies that at least one of the leaves attached to p and q must be labeled 0, say $\lambda(py) = \lambda(qz) = 0$, and thus y and z are adjacent in G. It remains to consider the possible coloring for the remaining to edges $\lambda(px)$ and $\lambda(qu)$. If $\lambda(px) = 1$ then x is isolated for all choices of $\lambda(qz)$. An analogous statement is true for $\lambda(qu) = 1$. For $\lambda(px) = \lambda(qu) = 0$ we obtain $K_4 - e$. If $\lambda(px) = 0$ and $\lambda(qu) = 2$ we obtain S_3. The same is true for $\lambda(px) = 2$ and $\lambda(qu) = 0$. Thus the only remaining choice is $\lambda(px) = \lambda(qu) = 2$. It indeed explains the path $x - y - z - u$, see Fig. 5. $\qquad\square$

The fact that S_n is the only "exact-2-leaf root" of K_n, i.e., the only tree with unit edge weights that explains K_n is shown in [10, Lemma 2]. It is not difficult to see that there is also no other choice of non-negative integer labels on S_n that explains K_n:

Lemma 14. *The complete graph K_n is explained with respect to $\overset{2}{\sim}$ by the star S_n with the unique labeling function $\lambda(e) = 1$ for all $e \in E(S_n)$.*

Proof. Is is easy to check that this construction explains K_n for all $n \geq 3$. The trivial cases $n = 1$ and $n = 2$ are explained in the text. K_3 is only explained by S_3 with all edges labeled $\lambda(e) = 1$. Since the start S_n displays S_3 corresponding to every K_3 subgraph, all edges of S_n must be labeled by $\lambda(e) = 1$. $\qquad\square$

We note for later reference that the uniqueness results in Lemmas 13 and 14 do not require the precondition that $\overset{0}{\sim}$ is discrete. This observation will be important in the following section.
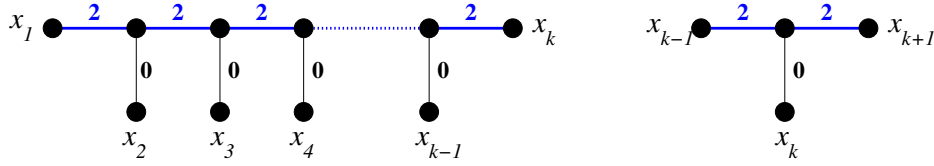
7

Figure 6: Construction of trees that explain paths.

Lemma 15. *There is no edge-labeled tree (T, λ) with discrete $\overset{0}{\sim}$ that explains the graphs C_4 and $K_4 - e$ with respect to $\overset{2}{\sim}$. The graph $K_3 + e$ is explained by a unique edge-labeled tree.*

Proof. There are two topologically distinct trees for $|L| = 4$, the star S_4 and tree T_4 with a single interior split. First consider the star S_4. In order to explain $K_3 + e$ or $K_4 - e$ three of the four edges must be labeled 1 (corresponding to the induced K_3). Depending on whether $\lambda(e_4) = 1$ or $\lambda(e_4) \neq 1$, the fourth vertex is either connected to all or none of the three other vertices. In order explain C_4, there must be two edges with $\lambda(e_1) = \lambda(e_3) = 2$ and one with $\lambda(e_2) = 0$ corresponding to an induced P_3. The remaining edge then must have $\lambda(e_4) = 0$. But then $x_2 \overset{0}{\sim} x_4$, contradicting that $\overset{0}{\sim}$ is discrete.

Now consider the tree T_4, which can be obtained from S_3 by subviding one of the edges and attaching an extra leaf to the subdividing vertex. Denote by s the (unique) inner edge of T_4. Consider $K_3 + e$ and $K_4 - e$ as shown in Fig. 5. Then we must have $\lambda(e_1) = \lambda(e_2) = 1$. If $\lambda(s) = 0$ we recover the situation of S_4, since the inner edge does not contribute to $\overset{2}{\sim}$. On the other hand, if If $\lambda(s) = 2$, then x_1 and x_2 cannot be connected with x_3 or x_4, contradicting the existence of K_3 as induced subgraph. Thus $\lambda(s) = 1$. Then $\lambda(e_3) = 0$. By assumption, $\lambda(e_4) \neq 0$ since otherwise $x_3 \overset{0}{\sim} x_4$. If $\lambda(e_4) = 1$, the x_4 is an isolated vertex in G. If $\lambda(e_4) = 2$, then $x_4 \overset{2}{\sim} x_3$ while x_4 is not in $\overset{2}{\sim}$ relation to either x_1 or x_2. Thus $G = K_3 + e$. The corresponding edge labeled tree is shown in Fig. 4. Since we have already considered all cases, $K_4 - e$ cannot be explained with respect to $\overset{2}{\sim}$.

Finally, consider T_4 and suppose that G contains P_3 as induced subgraph. There are two cases: If $\lambda(e_1) = \lambda(e_2) = 2$ then connectedness of G implies that $\lambda(s) = \lambda(e_3) = \lambda(e_4) = 0$, contradicting that $\overset{0}{\sim}$ is discrete. In the alternative case we can assume, w.l.o.g., that $\lambda(e_1) = 2$ and $\lambda(e_2) = 2$. Furthermore, in order to explain C_4 we must have $\lambda(e_3) + \lambda(e_4) = 2$. If both $\lambda(e_3) = \lambda(e_4) = 1$. Then $\lambda(s) = 0$ and $\lambda(s) = 2$ yields $G = K_2 \cup K_2$, for $\lambda(s) = 1$ we obtain S_4. In the remaining case we can choose $\lambda(e_3) = 2$ and $\lambda(e_4) = 0$. Now $\lambda(s) = 0$ contradicts discreteness of $\overset{0}{\sim}$, $\lambda(s) = 1$ yields the edgeless graph. For $\lambda(s) = 2$ we obtain P_4. Thus C cannot be explained by T_4 with respect to $\overset{2}{\sim}$ by a labeling with discrete $\overset{0}{\sim}$. $\square$

The fact that $K_4 - e$ is a forbidden subgraph implies that two cliques in G cannot be "glued together" by a single common edges. It is possible, however, for cliques to touch in a cut vertex as shown by the example of the bowtie graph B, which is obtained by gluing together two triangles at a common vertex, see Fig. 4.

Graphs that can be represented as pairwise compatibility graphs of caterpillars have received special attention in the literature [5, 33, 34, 35]. It is not difficult to see that the path P_h, $h \geq 3$ can represented by a caterpillar in several settings. These results cannot be directly applied in our setting, however. Any two leaves x and y attached to two distinct inner vertices of a caterpillar are separated by at least three edges an thus cannot be in relation $\overset{2}{\sim}$ if we assume strictly positive integer weights. It follows immediately that P_h is not an exact-2-leaf power of a caterpillar and that P_h cannot be explained by caterpillar unless zero-weights are allowed. An explicit construction in [15] shows that P_h is explained by a caterpillar with edge weights in $\{0, 1\}$ with respect to exactly-1-relation $\overset{1}{\sim}$. Lemma 4 implies that we can use the same construction to explain P_h by a caterpillar with edge weights in $\{0, 2\}$, see Fig. 6. It will be important later on that this construction is indeed unique:

Lemma 16. *The path P_h has as its unique explaining tree the caterpillar (T_h, λ_h) with all inner edges and the edges connecting to the end-points of P_h labeled 2 and all edges connecting to inner vertices of P_h labeled 0.*

Proof. We first recall that the tree (T_4, λ_4) explaining P_4 is unique by Lemma 13. Now assume that for $h \geq 5$, the tree (T_{h-1}, λ_{h-1}) explaining P_{h-1} is unique and thus a caterpillar. Any tree (T, λ) explaining P_h therefore must display

(T_{h-1}, λ_{h-1}), i.e., (T, λ) is obtained from T_h by subdividing one edge and attaching leaf h and edge e_h to the new vertex, or by attaching h and e_h to an inner vertex of T_{h-1}. One easily checks that the latter yields a branched tree or a disconnected graph. The same is true is any other edge except e_{h-1} and e_1, the edges adjacent to the leaves $h-1$ or 1 are subdivided. In the latter case, h cannot be adjacent to $h-1$. In the remaining case, the edge with which $h-1$ is attached is subdivided into an interior part s and the part e_{h-1} incident with $h-1$. Since the interior part cannot carry a zero label, we must have $\lambda_h(e_{h-1}) = 0$, $\lambda_h(s) = 2$, and $\lambda_h(e_h) = 2$. Thus the caterpillar of Fig. 6 is indeed the only choice. We emphasize that this observation remains true even is $\overset{0}{\sim}$ is not assumed to be discrete. $\square$

Lemma 17. *The simple cycles C_p, $p \geq 5$ cannot be explained with respect to $\overset{2}{\sim}$ irrespective of whether $\overset{0}{\sim}$ is discrete or not.*

Proof. Every cycle C_p contains a path P_{p-1} with one vertex less as an induced subgraph. From Lemma 16 we known that P_{p-1} has a unique explanation by a caterpillar for all $p \geq 5$. Thus any tree (T^*, λ^*) explaining C_p thus must display caterpillar (T_{p-1}, λ_{p-1}) and thus T^* is obtained from T_{p-1} by either attaching p and e_p to inner vertex of T_{p-1} or by subdiving an edge and attaching p and e_p to the newly inserted vertex. As argued above, attachment to an inner vertex or subdivision of an edge other than e_1 or e_{p-1} leads to a branched tree or a disconnected graph. If e_p is inserted by subdivision of e_1, then p cannot be adjacent to $p-1$ and subdivision of e_{p-1} precludes adjacency of p and 1 for $p \geq 3$. Thus the catapillar tree (T_{p-1}, λ_{p-1}) cannot be extented to tree that explains C_p for any $p \geq 4$. Note that this argument did not make the assumption that $\overset{0}{\sim}$ is discrete. $\square$

Let us now turn to the general case. We first note that all graphs explained w.r.t. $\overset{2}{\sim}$ with discrete $\overset{0}{\sim}$ are chordal, i.e., every cycle of length greater than three has a chord. Even more stringently, every cycle of length 4 corresponds to a clique in G because the $K_4 - e$, i.e., the 4-cycle with a chord, is also a forbidden induced subgraph. We note that there is ample literature on the relationship of chordal graphs and PCGs, see e.g. [4, 5]. Due to the differences in the edge weight functions, it is not immediately pertinent to our discussion, however.

Lemma 18. *If G can be explained by the exact-2-relation with discrete $\overset{0}{\sim}$ and contains a Hamiltonian cycle, then G is a complete graph.*

Proof. The assertion is trivially true for $n = 3$ and holds for $n = 4$ because C_4 and $K_4 - e$, the only Hamiltonian graphs on 4 vertices except K_4 are forbidden induced subgraphs. Now suppose the statement is true for for all $|V| < p$ and consider a graph with p vertices. Since G is chordal, there is in particular a planar triangulation of C that is a subgraph of G and thus there are three consecutive vertices $u - v - w$ along C such that $u - w$ is a also an edge in G. Thus $G \setminus v$ is Hamiltonian. As an induced subgraph of G it can be explained by the exact-2-relation and thus is a complete graph by the induction hypothesis. Thus $u - x - w$ is triangle in $G \setminus v$ and $u - x - w - v$ is a cycle of length 4 in G. Since C_4 and $K_4 - e$ cannot appear as induced subgraphs of G, $\{u, v, w, x\}$ must for a clique in G, and hence the edge $\{v, x\} \in E(G)$ for all $x \in V(G \setminus v)$. Thus G is a complete graph. $\square$

Lemma 19. *A graph G with at least three vertices that can be explained by the exact-2-relation with discrete $\overset{0}{\sim}$ is complete if and only if it is 2-connected.*

Proof. If G is Hamiltonian, it is in particular also 2-connected. Now consider the case that G is 2-connected but not Hamiltonian. Let C be a cycle of maximal length in G and let x be a vertex not in C. Then there is a cycle C' in G that contains x and at least two distinct vertices of C since otherwise one of the vertices of C would be a cut vertex of G, contradicting 2-connectedness. Starting from x, let p and q be first and last vertex of C encountered along C'. By Lemma 18, $G[C]$ is a complete graph, and hence there is a another Hamiltonian cycle C'' on $G[C]$ so that p and q are consecutive along C'. Thus the cycle C^* obtained traversing C'' from p to q and then following C' from q through x back to p is a cycle that is strictly longer than C, contradicting maximality. Thus G is Hamiltonian, and hence complete. $\square$

A graph G is a *block graph* [36] if each of its biconnected components is a clique. Lemma 19 thus implies that every graph that can be explained with respect to the exact-2-relations with discrete $\overset{0}{\sim}$ is a block graph (see Thm. 21 below for a formal proof). Algorithm 2 (illustrated in Figure 7) explicitly constructs an edge-labeled tree that explains a given block graph.

Algorithm 2 Compute $(T(G), \lambda)$ for a connected block graph G

Require: a connected block graph G
 1: mark "red" all cut vertices $u \in V(G)$.
 2: **for all** cliques K in G **do**
 3:    **if** K is an edge e **then**
 4:       $\lambda(e) = 2$
 5:    **else**
 6:       replace K by a star $S_{|V(K)|}$ with center c_K
 7:       $\lambda(uc_K) = 1$ for each $u \in V(K)$
 8:    **end if**
 9: **end for**
10: **for all** red vertices v **do**
11:    add a vertex v' and edge vv' with $\lambda(vv') = 0$
12:    exchange the vertex names v and v'
13: **end for**
14: **return** (G, λ)

Lemma 20. *Algorithm 2 transforms any connected block-graph G into an edge-labeled tree that explains G with respect to the exactly-2-relation $\overset{2}{\sim}$ with discrete $\overset{0}{\sim}$.*

Proof. The output of Alg. 2 contains no cycles since all cycles in the input G are contained within a block and each block is replaced by a star. Furthermore, the replacement of a clique K_p by a star S_p with $p + 1$ vertices preserves connectedness, hence G has been transformed into a tree at this stage. Every vertex of a clique K, with $|V(K)| \geq 3$ this is not also contained in another block is now a leaf; all other nodes of K are marked red. Every vertex in an K_2 original K_2 block is either a leaf or marked "red". By construction, every "red" vertex has degree at least 2 and hence is not a leaf. The final operation adds a leaf to each "red" vertex. Together with the renaming of the vertices, thus, every vertex of the input graph is now a leaf in T.

Now consider the labeling. First suppose that u and v are non-adjacent in the input G, that is, there is a least one cut-vertex, say z, between them in G. The construction of $T(G)$ ensures that the unique path from u to v in $T(G)$ runs through a vertex z' that z as its neighbor. If the path from u to z in G ran through an edge in a triangle, it passes through the corresponding star and hence contains two edges labeled 1. Otherwise it runs through an unaltered K_2-block of G, which is labeled 2. In each case, therefore, $d_{T(G),\lambda}(x, y) \geq 4$. Now suppose that u and v are adjacent in G. First suppose uv is contained in a triangle of G. If neither u nor v was marked "red" they are both adjacent to the center c_K of a star with edges labeled 1. If u was a cut vertex, i.e., marked "red", it appears a leaf adjacent to a vertex u' that in turn is adjacent to c_K; furthermore $\lambda(uu') = 0$ and $\lambda(u'c_K) = 1$. Analogous reasoning applied if v was a cut vertex of G. In all cases, thus $d_{T(G),\lambda}(uv) = d_{T(G),\lambda}(u, c_K) + d_{T(G),\lambda}(c_K, v) = 1 + 1 = 2$. If the edge uv is not contained in a triangle, then it is labeled 2. If u or v are cut vertices, then the unique path from u to v is $u - u' - v$, $u - v' - v$, or $u - u' - v' - v$, with $\lambda(uu') = \lambda(vv') = 0$ and a label 2 for the remaining edge. Hence, $d_{T(G),\lambda}(uv) = 2$. In summary $u \overset{2}{\sim}_{T(G),\lambda} v$ if and only u and v are adjacent in G. Thus G is explained by $(T(G), \lambda)$ with respect to the exactly-2-relation. $\square$

Theorem 21. *A graph G can be explained by an edge-labeled tree (T, λ) with respect to the exact-2-relation with discrete $\overset{0}{\sim}$ if and only if it is a block graph.*

Proof. Suppose G can be explained w.r.t. Rt with discrete $\overset{0}{\sim}$. If G is 2-connected, it is a clique by Lemma 19 and therefore also a block graph. Otherwise, we note that every 2-connected component G' of G is induced subgraph of G and thus, by Lemma 3, can be explained w.r.t. Rt. By Lemma 19 every 2-connected component G' of G therefore must be a clique, i.e., G is a block graph.

Conversely, suppose that G is a block graph. Since Algorithm 2 is correct by Lemma 20, every connected block graph can be explained. Since the non-connected block graphs are just disjoint unions of connected block graphs, Lemma 5 completes the characterization of the non-connected case. $\square$
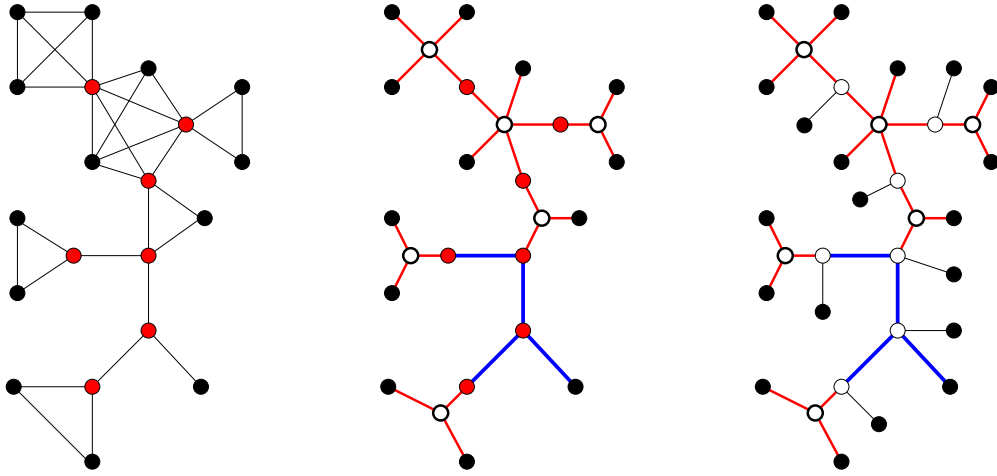
10

Figure 7: Illustration of Alg. 2. First all cut vertices of the input graph are marked. In the next step, edges not contained in triangles of G receive label 2 and all larger cliques are replaced by stars with all new edges labeled 1. In the final step the original cut vertices are decorated by an additional neighbor with a 0-labeled edge.

The main result of this section is now obtained as

Corollary 22. *A graph G is explained by $\overset{2}{\sim}$ if and only if $R/\overset{\bullet}{\sim}$ is a block graph.*

Proof. Thm. 11 establishes that G can be explained w.r.t. $\overset{2}{\sim}$ if and only if $R/\overset{\bullet}{\sim}$ can be explained w.r.t. $\overset{2}{\sim}$. Since the graph $G/\overset{\bullet}{\sim}$ is thin, Cor. 10 implies that $\overset{0}{\sim}$ is discrete for the canonical tree explaining $G/\overset{\bullet}{\sim}$, and thus Thm 21 can be applied to $G/\overset{\bullet}{\sim}$. $\qquad\square$

In the remainder of this section we consider the ambiguities in the construction of trees explaining block graphs. We start by characterizing contractible edges:

Lemma 23. *Suppose (T_e, λ_e) is obtained from a phylogenetic tree (T, λ) by contracting the edge e in T and setting $\lambda_e(e') = \lambda(e')$ for all $e' \neq e$ and suppose that $G(T, \lambda)$ is connected. Then $G(T, \lambda) = G(T_e, \lambda_e)$ if and only if e is an interior edge of T and $\lambda(e) = 0$.*

Proof. We have already noted the contracting an inner 0-edge does not change the graph. By definition, leaf-edges cannot be contracted, since the vertices of G correspond to the leaves of T. Connectedness of G implies that there is a pair of vertices x, y whose connecting path runs through e and whose distance $d_{T,\lambda}(x, y) = 2$. The contraction of e only leaves this distance unaffected if $\lambda(e) = 0$. Otherwise $d_{T,\lambda}(x, y)$ changes, which implies that x and y become disconnected in G' and hence the graph by the modified tree is different from G. $\qquad\square$
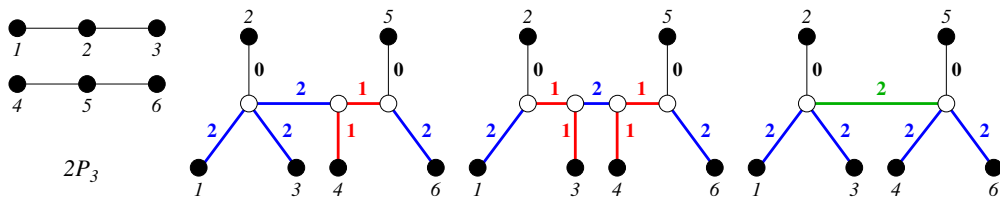

Figure 8: Three alternative, topologically different canonical trees explaining the non-connected graph $2P_3$. These are derived from the unique tree explaining P_3 in Fig. 4.

11

We note that connectedness of G is necessary in Lemma 23 since for non-connected G, the connected components can be "glued together" with arbitrarily complex trees as long as the distances between the attachment points is at least 3. In such examples it can be possible to contract edges without changing the explained graph. There are, for example at least three topologically different canonical trees that explain $2P_3$, see Fig. 8.

Lemma 24. *Let (T, λ) be a canonical tree explaining a connected graph G and let x be an interior vertex in T. Then all edges incident to x are 1-edges or p has at least one adjacent leaf u with $\lambda(pu) = 0$.*

Proof. Suppose p has no incident leaf. Since G is connected, for every edge e' there is another edge e'' such that $\lambda(e') + \lambda(e'') = 2$. For this pair of edges we have $\lambda(e') = \lambda(e'') = 1$ because no interior edges is 0-labeled. Thus $\lambda(e') = 1$ for all e' incident with x. On the other hand, if p has a neighbor u with $\lambda(pu) = 2$, then connectedness of G implies that there is another neighbor y of p with $\lambda(yp) = 0$. Hence, unless p has only 1-neighbors, then there must be a least one incident 0-edge, which by assumption must be a leaf. $\square$

We remark, finally, that a tree with minimal number of vertices (or edges) that explains a graph with respect to the exactly-2-relation is necessarily canonical. Otherwise, the contraction of an edge would make it possible to decrease of both the number of edges and vertices.

4. Oriented Exactly-2-Relation

Generalizing the construction of the oriented exactly-1-relation in [15], we consider here an oriented version of the exactly-k-relation. In constrast to the previous sections, we consider here rooted trees T with leaf set L. For two leaves x and y there is a unique *least common ancestor*, denoted by $lca(x, y)$, defined as the vertex most distant from the root r of $\overrightarrow{T}$ that is common to the paths connecting r with x and r with y, respectively.

Definition 25. *Let $(\overrightarrow{T}, \lambda)$ be a rooted tree with leaf set L and edge-labeling function $\lambda : E(\overrightarrow{T}) \to \mathbb{N}_0$. For $x, y \in L$ we consider the directed exactly-k-relation $\overset{k}{\to}$ defined by $x \overset{k}{\to} y$ if $\sum_{e \in \mathcal{P}(x, lca(x,y))} \lambda(e) = 0$ and $\sum_{e \in \mathcal{P}(lca(x,y),y)} \lambda(e) = k$ holds for the the (unique) paths $\mathcal{P}(x, lca(x,y))$ from x to $lca(x,y)$ and $(lca(x,y), y)$ from $lca(x,y)$ to y, respectively.*
The rooted tree $(\overrightarrow{T}, \lambda)$ explains a the directed graph $\overrightarrow{G}(L, E)$ (with respect to the directed exactly-k-relation) if $(x, y) \in E(G)$ if and only if $x \overset{k}{\to} y$.

By construction $\overrightarrow{G}(L, E)$ is an *oriented* graph, i.e., at most one of (x, y) and (y, x) can be an edge. As in the unrooted case, we say that a rooted tree $(\overrightarrow{T}, \lambda)$ is *canonical* if it is a rooted phylogenetic tree and does not have an inner 0-edge. In the following we will consider the case that $\overset{0}{\sim}$ is discrete. As in the undirected case, we shall relax this requirement in the end.

As in [15], our strategy is to exploit the close relationships between the oriented and the undirected case. Therefore, we first derive some technical results regarding common properties of the oriented relation $\overset{k}{\to}$ and its undirected relative $\overset{k}{\sim}$.

Note that the underlying tree (T, λ) of a rooted canonical tree $(\overrightarrow{T}, \lambda)$ is not necessarily an unrooted canonical tree. By contracting all the interior 0-edges and degree 2 vertices, we get a unique unrooted canonical tree (T', λ') corresponds to $(\overrightarrow{T}, \lambda)$. Conversely, for any unrooted canonical tree (T, λ) with $|V(T)| > 1$, we can create a set $\mathbb{T}(T, \lambda)$ of corresponding rooted least resolved trees as follows: (i) each interior vertex of (T, λ) may serve as a root; (ii) each leaf attached by a 0-edge may serve as a root; and (iii) every 2-edge can be subdivided by inserting a the root as a new vertex such that each of the two resulting edges is labeled 1. The construction is detailed in Algorithm 3. An example is given in Figure 9. The following lemma formalizes this one-to-one correspondence between unrooted canonical trees (T, λ) and its corresponding sets of rooted canonical tree.

Lemma 26. *Every rooted canonical tree can be constructed from its underlying unrooted canonical tree by Algorithm 3.*

Algorithm 3 Compute the set of canonical rooted trees $(\mathbb{T}, \lambda)$

Require: unrooted canonical tree (T, λ) with $|V(T)| > 1$

1: $(\mathbb{T}, \lambda) \leftarrow \emptyset$
2: **for all** interior vertices $v \in T$ **do**
3:     designate v as root
4:     add the rooted tree to $(\mathbb{T}, \lambda)$
5: **end for**
6: **for all** leaf vertices $v \in T$ with $\lambda(vw) > 0$ where $N(v) = \{w\}$ **do**
7:     subdivide vw to vv^*w and designate v^* as root
8:     relabel as $\lambda(vv^*) \leftarrow \lambda(vw)$ and $\lambda(v^*w) \leftarrow 0$, designate v^*
9:     add the resulting rooted tree to $(\mathbb{T}, \lambda)$.
10: **end for**
11: **for all** edges $e = uv$ with $\lambda(e) = k > 1$ **do**
12:     subdivide the edge e by inserting v^* and designate v^* as the root
13:     **for** $j = 1...k - 1$ **do**
14:         $\lambda(uv^*) \leftarrow j$ and $\lambda(v^*v) \leftarrow k - j$
15:         add the resulting rooted tree to $(\mathbb{T}, \lambda)$.
16:     **end for**
17: **end for**
18: **return** $(\mathbb{T}, \lambda)$

Proof. By construction, the set of canonical rooted trees corresponding to unrooted canonical tree is well defined, i.e., the correspondence is a mapping.

Suppose there are two distinct unrooted canonical trees (T_1, λ_1) and (T_2, λ_2) such that both their correspondings sets of rooted trees contains $(\overrightarrow{T}, \lambda)$. By construction, it has a underlying tree (T, λ) from which a unique canonical tree is obtained by contracting 0-edges and degree 2 vertices. Thus $(T_1, \lambda_1) = (T_2, \lambda_2)$, a contradiction. Hence the mapping is injective.

The mapping is also surjective, since each rooted canonical tree $(\overrightarrow{T}, \lambda)$ can be constructed from its corresponding unrooted canonical tree. $\square$

Lemma 27. *Suppose the unrooted canonical tree* (T, λ) *explains* G *with respect to* $\overset{k}{\sim}$. *Let* $\overrightarrow{G}$ *be a digraph explained w.r.t.* $\overset{k}{\rightarrow}$ *by a rooted tree* $(\overrightarrow{T}, \lambda)$ *corresponding to* (T, λ). *Then the underlying graph of* $\overrightarrow{G}$ *is a spanning subgraph of* G.

Proof. By construction, (T, λ) and $(\overrightarrow{T}, \lambda)$ has the same leaf set, and hence $V_G = V_{\overrightarrow{G}}$.

Any arc $x \rightarrow y$ in $\overrightarrow{G}$ is an edge in the underlying graph of $\overrightarrow{G}$ because that fact that $(\overrightarrow{T}, \lambda)$ explains $\overrightarrow{G}$ implies $\sum_{e \in \mathcal{P}(x, lca(x,y))} \lambda(e) = 0$ and $\sum_{e \in \mathcal{P}(lca(x,y),y)} \lambda(e) = k$. Considering the underlying unrooted graph (T', λ'), we have $\sum_{e \in \mathcal{P}(x,y)} \lambda'(e) = k$. Since (T, λ) explains G, and by construction (T', λ') displays (T, λ), we conclude that (T', λ') also explains G. Hence $(x, y) \in E(G)$. $\square$
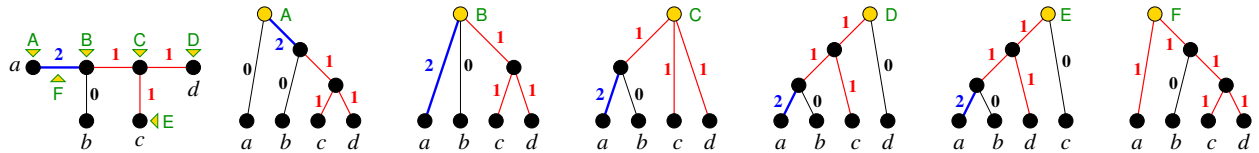

Figure 9: Construction of rooted canonical trees A through F from the unrooted canonical tree on the left. The possible positions of the root are indicated by the triangles and the corresponding six rooted trees are shown to the right.

13

Definition 28. *Suppose there a tree (T, λ) with discrete $\overset{0}{\sim}$ that explains G w.r.t. $\overset{k}{\sim}$, then every subgraph H of G is allowed for $\overset{k}{\sim}$ / $\overset{0}{\sim}$. Analogously, if there is a rooted tree $(\overrightarrow{T}, \lambda)$ with discrete $\overset{0}{\sim}$ that explains $\overrightarrow{G}$ w.r.t. $\overset{k}{\rightarrow}$, we say that every subgraph $\overrightarrow{H}$ of $\overrightarrow{G}$ is allowed for $\overset{k}{\sim}$ / $\overset{0}{\sim}$.*

In more detail a graph H is allowed for $\overset{k}{\sim}$ / $\overset{0}{\sim}$ if there exists (T, λ) such that for any $(x, y) \in E(G)$, we have $\sum_{e \in \mathcal{P}(x,y)} \lambda(e) = k$. If G is not allowed for $\overset{k}{\sim}$ / $\overset{0}{\sim}$, we say that is it is *forbidden (as a subgraph)* for $\overset{k}{\sim}$ / $\overset{0}{\sim}$. Analogous, a graph $\overrightarrow{H}$ is allowed for $\overset{k}{\rightarrow}$ / $\overset{0}{\sim}$ in the rooted case, if there exists (T, λ) such that for any $(x, y) \in E(G)$, we have $\sum_{e \in \mathcal{P}(x,lca(x,y))} \lambda(e) = 0$ and $\sum_{e \in \mathcal{P}(lca(x,y),y)} \lambda(e) = k$. If $\overrightarrow{G}$ is not allowed as a subgraph in $G(\overset{k}{\rightarrow})$/ $\overset{0}{\sim}$, we that say G is *forbidden (as a subgraph)* for in $\overset{k}{\rightarrow}$ / $\overset{0}{\sim}$.

Lemma 29. *If G is forbidden for $\overset{k}{\sim}$ / $\overset{0}{\sim}$, then any orientation of G is forbidden in $\overset{k}{\rightarrow}$ / $\overset{0}{\sim}$. If $\overrightarrow{G}$ is allowed as a subgraph for $\overset{k}{\rightarrow}$ / $\overset{0}{\sim}$ with rooted tree $(\overrightarrow{T}, \lambda)$, then its underlying graph is allowed for $\overset{k}{\sim}$ / $\overset{0}{\sim}$ as a subgraph with the corresponding underlying tree (T, λ).*

Proof. Suppose, for contradictions, that G is forbidden for $\overset{k}{\sim}$ / $\overset{0}{\sim}$ but the orientation $\overrightarrow{G}$ of G is allowed for $\overset{k}{\rightarrow}$ / $\overset{0}{\sim}$. Then there exists a rooted tree $\overrightarrow{T}, \lambda$ such that for any arc $x \rightarrow y$ in $\overrightarrow{G}$, we have $\sum_{e \in \mathcal{P}(x,u)} \lambda(e) = 0$ and $\sum_{e \in \mathcal{P}(u,y)} \lambda(e) = k$ where $u = lca(x, y)$. Consider the unrooted tree (T, λ) of $(\overrightarrow{T}, \lambda)$. Since $(x, y) \in E(G)$ if and only if $x \rightarrow y$ or $y \rightarrow x$ is an arc in $\overrightarrow{G}$, then for any $(x, y) \in E(G)$, $\sum_{e \in \mathcal{P}(x,u)} \lambda(e) = 0$ and $\sum_{e \in \mathcal{P}(u,y)} \lambda(e) = k$ where $u = lca(x, y)$. By definition, G is allowed for $\overset{k}{\sim}$ / $\overset{0}{\sim}$, i.e., we arrive at a contradiction. The second statement is a simple consequence of the first one. □

The technical results obtained so far will allow us to infer properties of the oriented graph $\overrightarrow{G}$ and their explaining trees $(\overrightarrow{T}, \lambda)$ from their underlying undirected graphs G and unrooted trees (T, λ). In the following we will focus on graphs $\overrightarrow{G}$ that can be explained w.r.t. $\overset{2}{\rightarrow}$ by a rooted tree $(\overrightarrow{T}, \lambda)$ with discrete $\overset{0}{\sim}$.

Lemma 30. *Oriented cycles are forbidden as a subgraph for $\overset{2}{\rightarrow}$ / $\overset{0}{\sim}$.*

Proof. Suppose $\overrightarrow{C_n}$ is allowed. Then, by definition, there exists an orientation graph $\overrightarrow{H}$ with vertex set $V(\overrightarrow{C_n})$ such that $\overrightarrow{C_n}$ is a subgraph of $\overrightarrow{H}$ and a rooted tree $(\overrightarrow{T}, \lambda)$ that explains $\overrightarrow{H}$. W.l.o.g., we assume $(\overrightarrow{T_n}, \lambda)$ is a rooted canonical tree.

Consider the underlying unrooted canonical tree (T, λ) of $(\overrightarrow{T_n}, \lambda)$, we claim that (T, λ) must be $(S_n, 1)$. Suppose that (T, λ) explains graph G. By Lemma 27 since the underlying graph H of $\overrightarrow{H}$ is a subgraph of G which has the same vertex set with H, and H contains a Hamiltonian cycle, thus G also contains a Hamiltonian cycle, by Lemma 18 G is a complete graph K_n. And by Lemma 14 the (T, λ) displays $(S_n, 1)$.

Then we consider all the possibility to construct the set of rooted canonical trees corresponding to $(S_n, 1)$, and consider the oriented graph it explains. By Algorithm 3 we can place the root either on the center vertex, which will explain an empty graph, or place it on one of the leaves, which will explains oriented star on n vertices point to the leaves. In either case there is no cycles. By Lemma 26 we know we have constructed all rooted canonical trees and thus all oriented graphs they explain. Thus oriented cycles are forbidden as a subgraph for $\overset{2}{\rightarrow}$ / $\overset{0}{\sim}$. □

Lemma 31. *2-star oriented to center, $\bullet \rightarrow \bullet \leftarrow \bullet$, is forbidden as an induced subgraph for $\overset{2}{\rightarrow}$ / $\overset{0}{\sim}$.*

Proof. Explicit construction shows that we obtain $v_1 \overset{0}{\sim} v_3$ for each of the three triples $v_1 v_2 | v_3$, $v_1 v_3 | v_2$, and $v_2 v_3 | v_1$. This contradicts the assumption that $\overset{0}{\sim}$ is discrete. □

Lemma 32. *Every graph $\overrightarrow{G}$ that can be explained by an edge-labeled tree w.r.t. $\overset{2}{\rightarrow}$ / $\overset{0}{\sim}$ is an oriented forest with the property that all its component trees have a unique source vertex from which all arcs are directed away.*
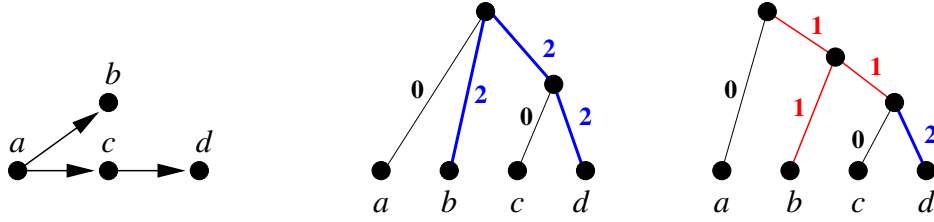
14

Figure 10: Two distinct rooted trees with discrete $\overset{0}{\sim}$ explaining an oriented graph on four vertices w.r.t. $\overset{2}{-}$.

Proof. Let $\overrightarrow{G}$ be a graph that can be explained w.r.t. $\overset{2}{-}$ / $\overset{0}{\sim}$. Since all cycles are forbidden induced subgraphs, $\overrightarrow{G}$ is a forest. Furthermore, there is only a single source vertex in each connected component. Otherwise, if both x and y were sources within the same component tree, then the unique path from x to y would necessarily contain an induced subgraph of the form $\bullet \to \bullet \leftarrow \bullet$, which is forbidden. $\square$

A canonical tree $(\overrightarrow{T}, \lambda)$ with discrete $\overset{0}{\sim}$ that explains a connected oriented graph $\overrightarrow{G}$ w.r.t. $\overset{2}{-}$ has a leaf that is attached to the root by 0-edge.

Theorem 33. $\overrightarrow{G}$ *is explained w.r.t.* $\overset{2}{-}$ *by a rooted tree* $(\overrightarrow{T}, \lambda)$ *with discrete* $\overset{0}{\sim}$ *if and only if G is an oriented forest that does not contain the 2-star oriented to center,* $\bullet \to \bullet \leftarrow \bullet$, *as an induced subgraph.*

Proof. To show the "only if" part, suppose $\overrightarrow{G}$ can be explained. Then $\overrightarrow{G}$ is oriented and by Lemma 30 it is an orientation tree, and by Lemma 31 we know that 2-star oriented to center is forbidden. For the "if" part we use the construction employed in [15] for $\overset{1}{\sim}$ (with 2-edges taking the place of 1-edges): To each inner vertex v of $\overrightarrow{G}$ a new vertex v' which represent v in tree is attached with a 0-edge, while the inner edges of the tree have label 2. The Theorem now follows directly from Lemma 4. $\square$

We can relax the condition that $(\overrightarrow{T}, \lambda)$ has discrete $\overset{0}{\sim}$. To this end, we extend the false twin relation $x \overset{\bullet}{\sim} y$ to digraphs by setting $x \overset{\bullet}{\sim} y$ iff x and y have the same in- and out-neighbors. The quotient graph $G/\overset{\bullet}{\sim}$ is known as the *point-determining graph* of G.

Corollary 34. *An oriented graph* $\overrightarrow{G}$ *is explained w.r.t.* $\overset{2}{-}$ *if and only if* $\overrightarrow{G}$ *is an oriented forest whose point-determining graph does not contain* $\bullet \to \bullet \leftarrow \bullet$ *as an induced subgraph.*

Proof. It suffices to note that $\overset{0}{\sim}$-equivalent vertices are in the same $\overset{\bullet}{\sim}$-class and that there is a least resolved tree in which all members of a $\overset{0}{\sim}$-class are siblings. $\square$

We note, finally, that the rooted trees with discrete $\overset{0}{\sim}$ that explain $\overrightarrow{G}$ w.r.t. $\overset{2}{-}$ are not unique, as exemplified in Figure 10.

5. Concluding Remarks

The main result of this contribution is the characteriztion of the exactly-2-relations, i.e. the graphs nniPCG$(T, \lambda, 2, 2)$. They form a proper superset of the the exact-2-leaf power graphs, which comprise only the disjoint unions of cliques. Section 2 suggests, however, that at least some of the structure and techniques carry over to general values of k. Several related problems are worth considering as well: in particular, nniPCG(T, λ, k, ∞) and nniPCG$(T, \lambda, 1, k)$ are of interest as models for coarse grained models of evolutionary distances.

The oriented version of the exactly-2-relation somewhat surprisingly, is much more closely related to the oriented exactly-1-relation of [15] that to the undirected exactly-2-relation. There is an alternative natural definition for a directed exactly-2-relation that omit the condition that $\sum_{e \in \mathcal{P}(x, lca(x,y))} \lambda(e) = 0$. Clearly, the resulting digraph are not

oriented, i.e., they may contain double edged. We suspect that their structure is more closely related to the Fitch graph (directed at-least-1-relation) recently studied in [17].

Regarding the analysis of rare-event data in phylogenetics the characterization of the exactly-2-relation naturally leads to the edge modification problem for block graphs and graphs whose R-thin quotient is a block graph, respectively. Although these problems do not seem to have been studied so far (see e.g. [37, Tab.1] and [38]). Since exactly-2-relation graphs are hereditary by Lemma 3, we suspect that edge modification problem for the exactly-2-relation graphs can be handled in manner similar to closely related edge modification problem for chordal graphs [37, 38] or cluster editing [39].

References

[1] C. Semple, M. Steel, Phylogenetics, Oxford University Press, Oxford UK, 2003.

[2] A. W. M. Dress, K. T. Huber, J. Koolen, V. Moulton, A. Spillner, Basic Phylogenetic Combinatorics, Cambridge University Press, Cambridge UK, 2012.

[3] P. E. Kearney, J. I. Munro, D. Phillips, Efficient generation of uniform samples from phylogenetic trees, in: Algorithms in Bioinformatics (WABI 2003 Budapest), Vol. 2812 of Lect. Notes Comp. Sci., Springer, Berlin, 2003, pp. 177–189. doi:10.1007/978-3-540-39763-2_14.

[4] M. N. Yanhaona, K. S. M. T. Hossain, M. S. Rahman, Pairwise compatibility graphs, J. Appl. Math. Comput. 30 (2009) 479–503. doi:10.1007/s12190-008-0204-7.

[5] T. Calamoneri, B. Sinaimeri, Pairwise compatibility graphs: A survey, SIAM Review 58 (2016) 445–460. doi:10.1137/140978053.

[6] M. I. Hossain, S. A. Salma, M. S. Rahman, D. Mondal, A necessary condition and a sufficient condition for pairwise compatibility graphs, J. Graph Algorithms Appl. 21 (2017) 341–352. doi:10.1007/978-3-319-30139-6_9.

[7] P. Baiocchi, T. Calamoneri, A. Monti, R. Petreschi, Graphs that are not pairwise compatible: A new proof technique, in: C. Iliopoulos, H. W. Leong, W.-K. Sung (Eds.), Combinatorial Algorithms, 29th IWOCA, Vol. 10979 of Lecture Notes Comp. Sci., Springer, Berlin, Heidelberg, 2018, pp. 39–51. doi:10.1007/978-3-319-94667-2_4.

[8] P. Baiocchi, T. Calamoneri, A. Monti, R. Petreschi, Some classes of graphs that are not PCGs, Theor. Comp. Sci. 791 (2019) 62–75. doi:10.1016/j.tcs.2019.05.017.

[9] S. Ahmed, M. S. Rahman, Multi-interval pairwise compatibility graphs, in: T. V. Gopal, G. Jäger, S. Steila (Eds.), Theory and Applications of Models of Computation (14'th TAMC 2017), Vol. 10185 of Lect. Notes Comp. Sci., Springer, Heidelberg, 2017, pp. 71–84. doi:10.1007/978-3-319-55911-7_6.

[10] A. Brandstädt, V. B. Lea, D. Rautenbach, Exact leaf powers, Theor. Comp. Sci. 411 (2010) 2968–2977. doi:10.1016/j.tcs.2010.04.027.

[11] A. Rokas, P. W. Holland, Rare genomic changes as a tool for phylogenetics, Trends Ecol Evol 15 (2000) 454–459. doi:10.1016/S0169-5347(00)01967-4.

[12] J. E. Tarver, E. A. Sperling, A. Nailor, A. M. Heimberg, J. M. Robinson, B. L. King, D. Pisani, P. C. J. Donoghue, K. J. Peterson, miRNAs: Small genes with big potential in metazoan phylogenetics, Mol. Biol. Evol. 30 (2013) 2369–2382. doi:10.1093/molbev/mst133.

[13] H. Luo, W. Arndt, Y. Zhang, G. Shi, M. A. Alekseyev, J. Tang, A. L. Hughes, R. Friedman, Phylogenetic analysis of genome rearrangements among five mammalian orders, Mol Phylogenet Evol. 65 (2012) 871–882. doi:10.1016/j.ympev.2012.08.008.

[14] J. W. Waegele, T. W. Bartholomaeus (Eds.), Deep Metazoan Phylogeny: The Backbone of the Tree of Life—New Insights from Analyses of Molecules, Morphology, and Theory of Data Analysis, De Gruyter, 2014.

[15] M. Hellmuth, M. Hernandez-Rosales, Y. Long, P. F. Stadler, Inferring phylogenetic trees from the knowledge of rare evolutionary events, J. Math. Biol. 76 (2017) 1623–1653. doi:10.1007/s00285-017-1194-6.

[16] M. Hellmuth, Y. Long, M. Geiß, P. F. Stadler, A short note on undirected Fitch graphs, Art Discrete Appl. Math. (ADAM) 1 (2018) P1.08. doi:10.26493/2590-9770.1245.98c.

[17] M. Geiß, J. Anders, P. F. Stadler, N. Wieseke, M. Hellmuth, Reconstructing gene trees from Fitch's xenology relation, J. Math. Biol. 77 (2017) 1459–1491. doi:10.1007/s00285-018-1260-8.

[18] C. Crespelle, C. Paul, Fully dynamic recognition algorithm and certificate for directed cographs, Discr. Appl. Math. 154 (2006) 1722–1741. doi:10.1016/j.dam.2006.03.005.

[19] K. Chaudhuri, K. Chen, R. Mihaescu, S. Rao, On the tandem duplication-random loss model of genome rearrangement, in: Proc. 17th Ann. ACM-SIAM Symp. Discrete Algorithm (SODA '06), Soc. Industrial Appl. Math., Philadelphia, 2006, pp. 564–570. doi:10.5555/1109557.1109619.

[20] M. Hellmuth, M. Hernandez-Rosales, K. T. Huber, V. Moulton, P. F. Stadler, N. Wieseke, Orthology relations, symbolic ultrametrics, and cographs, J. Math. Biol. 66 (2013) 399–420. doi:10.1007/s00285-012-0525-x.

[21] N. Nishimura, P. Ragde, D. M. Thilikos, On graph powers for leaf-labeled trees, J. Algorithms 42 (2002) 69–108. doi:10.1006/jagm.2001.1195.

[22] T. Calamoneri, E. Montefusco, R. Petreschi, B. Sinaimeria, Exploring pairwise compatibility graphs, Theor. Comp. Sci. 468 (2013) 23–36. doi:10.1016/j.tcs.2012.11.015.

[23] P. Buneman, The recovery of trees from measures of dissimilarity, in: F. R. Hodson, D. G. Kendall, P. Tautu (Eds.), Mathematics in the Archaeological and Historical Sciences, Edinburgh University Press, Edinburgh, 1971, pp. 387–385.

[24] J. M. S. Simões-Pereira, A note on the tree realizability of a distance matrix, J. Combin. Theory 6 (1969) 303–310. doi:10.1016/S0021-9800(69)80092-X.

[25] H.-J. Bandelt, H. M. Mulder, Distance-hereditary graphs, J. Comb. Th., Ser. B 41 (1986) 182–208. doi:10.1016/0095-8956(86)90043-2.

[26] R. Hammack, W. Imrich, S. Klavžar, Handbook of Product graphs, 2nd Edition, CRC Press, Boca Raton, 2011.

[27] R. McKenzie, Cardinal multiplication of structures with a reflexive relation, Fund. Math. 70 (1971) 59–101. doi:10.4064/fm-70-1-59-101.

[28] D. P. Sumner, Point determination in graphs, Discrete Math. 5 (1973) 179–187. doi:10.1016/0012-365X(73)90109-X.

[29] J. J. Bull, C. M. Pease, Combinatorics and variety of mating-type systems, Evolution 43 (1989) 667–671. doi:10.1111/j.1558-5646.1989.tb04263.x.

[30] G. Kilibarda, Enumeration of unlabelled mating graphs, Graphs Combinatorics 23 (2007) 183–199. doi:10.1007/s00373-007-0692-5.

[31] I. Gessel, J. Li, Enumeration of point-determining graphs, J. Comb. Th., Ser. A 118 (2011) 591–612. doi:10.1016/j.jcta.2010.03.009.

[32] W. Imrich, Factoring cardinal product graphs in polynomial time, Discrete Math. 192 (1998) 119–144. doi:10.1016/S0012-365X(98)00069-7.

[33] A. Brandstädt, C. Hundt, Ptolemaic graphs and interval graphs are leaf powers, in: E. S. Laber, C. F. Bornstein, L. T. Nogueira, F. L. (Eds.), LATIN 2008, Vol. 4957 of Lect. Notes Comp. Sci., Springer, Berlin, 2008, pp. 479–491. doi:10.1007/978-3-540-78773-0_42.

[34] T. Calamoneri, A. Frangioni, B. Sinaimeri, Pairwise compatibility graphs of caterpillars, Computer J. 57 (2014) 1616–1623. doi:10.1093/comjnl/bxt068.

[35] S. A. Salma, M. S. Rahman, M. I. Hossain, Triangle-free outerplanar 3-graphs are pairwise compatibility graphs, J. Graph Alg. Appl. 17 (2013) 81–102. doi:10.7155/jgaa.00286.

[36] F. Harary, A characterization of block-graphs, Canadian Math. Bull. 6 (1963) 1–6. doi:10.4153/CMB-1963-001-x.

[37] P. Burzyn, F. Bonomo, G. Durán, NP-completeness results for edge modification problems, Discr. Appl. Math. 154 (2006) 1824–1844. doi:10.1016/j.dam.2006.03.031.

[38] R. Sritharan, Graph modification problem for some classes of graphs, J. Discr. Algorithms 38-41 (2016) 32–37. doi:10.1016/j.jda.2016.06.003.

[39] R. Shamir, R. Sharan, D. Thur, Cluster graph modification problems, Discr. Appl. Math. 144 (2004) 173–182. doi:10.1016/j.dam.2004.01.007.