SIAM REVIEW
Vol. 58, No. 3, pp. 445–460


# Pairwise Compatibility Graphs: A Survey*

Tiziana Calamoneri†
Blerina Sinaimeri‡

**Abstract.** A graph $G = (V, E)$ is a *pairwise compatibility graph (PCG)* if there exists an edge-weighted tree $T$ and two nonnegative real numbers $d_{min}$ and $d_{max}$ such that each leaf $u$ of $T$ is a node of $V$ and there is an edge $(u, v) \in E$ if and only if $d_{min} \leq d_T(u, v) \leq d_{max}$, where $d_T(u, v)$ is the sum of weights of the edges on the unique path from $u$ to $v$ in $T$. In this article, we survey the state of the art concerning this class of graphs and some of its subclasses.

**Key words.** pairwise compatibility graphs, leaf power graphs, min leaf power graphs

**AMS subject classifications.** 05C90, 68R01, 68R10

**DOI.** 10.1137/140978053

## 1. Introduction.

The reconstruction of ancestral relationships is one of the fundamental problems in computational biology as it is widely used to provide both evolutionary and functional insights into biological systems. The evolutionary history of a set of organisms is usually represented by a tree-like structure called a phylogenetic tree, which is a tree where each leaf represents a distinct known taxon and the internal nodes represent possible ancestors that might have led, through evolution, to this set of taxa. Moreover, the edges of the tree can be weighted in order to represent a sort of evolutionary distance among species. In the phylogenetic tree reconstruction problem, given a set of taxa, we want to find a phylogenetic tree that "best" explains the given data. Due to the difficulty in determining the criteria for an "optimal" tree, the performance of the reconstruction algorithms is usually evaluated experimentally by comparing the tree produced by the algorithm with the "known" tree. However, as the tree reconstruction problem is proved to be NP-hard under many criteria of optimality, and as real phylogenetic trees usually consist of a large number of nodes, testing these heuristics on real data is difficult. Thus, it is interesting to find efficient ways to sample subsets of taxa from a large phylogenetic tree, subject to some biologically-motivated constraints, in order to test the reconstruction algo-

*Received by the editors July 22, 2014; accepted for publication (in revised form) August 4, 2015; published electronically August 4, 2016.
http://www.siam.org/journals/sirev/58-3/97805.html

†Computer Science Department, Sapienza University of Rome, Via Salaria 113, 00198 Roma, Italy (calamoneri@di.uniroma1.it). The work of the first author was supported by the Italian Ministry of Education, University, and Research (MIUR) under PRIN 2012C4E3KT national research project "AMANDA – Algorithmics for MAssive and Networked DAta."

‡Corresponding author. INRIA and Université de Lyon, Université Lyon 1, LBBE, CNRS UMR558, France (blerina.sinaimeri@inria.fr). The work of the second author was supported by the European Research Council under the European Community's Seventh Framework Programme (FP7 / 2007-2013) / ERC grant agreement [247073]10.

rithms on the smaller subtrees induced by the sample. The constraints on the sample attempt to ensure that the behavior of the algorithm will not be biased by the fact it is applied on the sample instead of on the whole tree. For instance, as observed in [32], very close or very distant taxa can create problems for phylogeny reconstruction algorithms. This leads to the following constraint on the sample: given two positive integers $d_{min}, d_{max}$, select a sample from the leaves of the tree such that the pairwise distance between any two leaves in the sample is at least $d_{min}$ and at most $d_{max}$. This sampling problem was considered in [39] and polynomial algorithms were proposed, which motivates the introduction of *pairwise compatibility graphs* (PCGs). Indeed, given a phylogenetic tree $T$ and integers $d_{min}, d_{max}$ we can associate with them a graph $G$, called the PCG of $T$, whose nodes are the leaves of $T$ and for which there is an edge between two nodes if the corresponding leaves in $T$ are at a distance within the interval $[d_{min}, d_{max}]$. While it is trivial to construct the graph $G$ starting from $T, d_{min}, d_{max}$, the inverse problem is difficult.

PCGs can be seen as a generalization of the well-studied class of *leaf power graphs* (in which $d_{min} = 0$), introduced in the context of constructing phylogenies from species similarity data [24, 41, 48]. Specifically, interspecies similarity is represented by a graph $G$, where the nodes are the species and the adjacency relation represents evidence of evolutionary similarity. The phylogenetic tree is then built from this graph such that the leaves correspond to nodes of the graph and two leaves that correspond to adjacent nodes are separated by a distance of at most $d_{max}$ in the tree, where $d_{max}$ is a chosen threshold of proximity. Although there has been a lot of work done on this topic (see, e.g., [48, 6, 10, 5]), a complete description of leaf power graphs is still lacking and remains an important research problem.

Another natural relaxation of the pairwise compatibility constraint is that obtained when $d_{max}$ is set to $\infty$. Thus, there is an edge $(u, v)$ in $G$ if and only if $d_T(u, v) \geq d_{min}$. This relaxation leads to the definition of the class of *min leaf power graphs (mLPGs)* [20]. It is worth mentioning that initially the mLPG class was defined as the complement of the leaf power graph (LPG) class, in an attempt to better understand the structure of PCGs. This is the reason why their name evokes power graphs, although they are not of that type.

In this survey we review the results on the identification of the classes PCG and mLPG, and some of the results for LPGs. For more details on the characterization of the LPG class we refer the reader to the nice survey of Brandstädt [3].

This article is organized as follows. Section 2 is devoted to some basic definitions and some preliminary results. We have included most of the definitions of the graph classes we mention. However, for further details the interested reader can consult [9]. In section 3 we survey the main results related to the complexity of recognizing PCGs. Section 4 is devoted to the graph class $LPG \cap mLPG$. In section 5 we list the graph classes which are known not to belong to the class PCG. Section 6 includes the state of the art on the characterization of the PCG class and its subclasses mLPG and LPG. In particular, it presents the known results on the graph classes contained in PCGs as well as results concerning the characterization of the PCGs in terms of forbidden configurations. In section 7 we consider PCGs of particular subclasses of trees, such as stars and caterpillars. In each section we also include a number of major open problems. Finally, we conclude in section 8 with some possible research directions related to PCGs.

**2. Basic Definitions.** In this section we recall some basic definitions that we use throughout this paper.

A graph $G = (V, E)$ is a *pairwise compatibility graph* (PCG) if there exists a tree $T$, a positive edge-weight function $w$ on $T$, and two nonnegative real numbers $d_{min}$ and $d_{max}$, $d_{min} \leq d_{max}$, such that each node $u \in V$ is uniquely associated to a leaf $l_u$ of $T$ and there is an edge $(u, v) \in E$ if and only if $d_{min} \leq d_{T,w}(l_u, l_v) \leq d_{max}$, where $d_{T,w}(l_u, l_v)$ is the sum of the weights of the edges on the unique path from $l_u$ to $l_v$ in $T$. In such a case, we say that $G$ is a PCG of $T$ for $d_{min}$ and $d_{max}$; in symbols, $G = PCG(T, w, d_{min}, d_{max})$.

A graph $G(V, E)$ is called a *leaf power graph* (LPG) if there exists a tree $T$, a positive edge-weight function $w$ on $T$, and a nonnegative number $d_{max}$ such that there is an edge $(u, v)$ in $E$ if and only if for their corresponding leaves in $T$, $l_u, l_v$ we have $d_{T,w}(l_u, l_v) \leq d_{max}$; in symbols, $G = LPG(T, w, d_{max})$.

A graph $G = (V, E)$ is a *minimum leaf power graph* (mLPG) if there exists a tree $T$, a positive edge-weight function $w$ on $T$, and an integer $d_{min}$ such that there is an edge $(u, v)$ in $E$ if and only if for their corresponding leaves in $T$ $l_u, l_v$ we have $d_{T,w}(l_u, l_v) \geq d_{min}$; in symbols, $G = mLPG(T, w, d_{min})$.

We mean by *k-leaf power graph* (*k-min leaf power graph*, respectively) a graph which is an LPG with $d_{max} = k$ ($d_{min} = k$, respectively). In Figure 1, examples of a PCG, an LPG, and an mLPG are depicted.

A graph $G = (V, E)$ is an *exact k-leaf power* [8] if there is a weighted tree $T$ such that each node $u \in V$ is uniquely associated to a leaf $l_u$ of $T$ and there is an edge $(u, v) \in E$ if and only if $d_T(l_u, l_v) = k$. It is clear that an exact $k$-leaf power graph is a PCG where $d_{min} = d_{max} = k$.

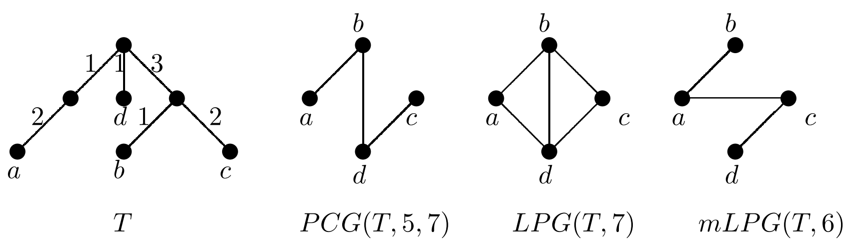


**Fig. 1** *An edge-weighted tree $T$ and an example of a PCG, an LPG, and an mLPG.*

Observe that here we always assume that $d_{min}, d_{max}$, and the weight of the edges of the tree of a PCG are all positive real numbers. In the original problem concerning the LPGs, these quantities were required to be natural numbers. It is proved in [5] that it is not a loss of generality to consider positive real numbers instead of naturals for LPGs. This result is extended to the general case of PCGs as follows.

LEMMA 1 (see [16]). *Let $G = PCG(T, w, d_{min}, d_{max})$, where $d_{min}, d_{max}$ are nonnegative real numbers and the weight $w(e)$ of each edge $e$ of $T$ is a positive real number. Then it is possible to choose natural numbers $\hat{w}, \hat{d}_{min}, \hat{d}_{max}$ such that $G = PCG(T, \hat{w}, \hat{d}_{min}, \hat{d}_{max})$.*

A graph $G = (K, S, E)$ is said to be a *split* graph [33] if there is a node partition $V = K \cup S$ such that the subgraphs induced by $K$ and $S$ are complete and stable, respectively.

A graph $G = (V, E)$ is a *thin spider* [38] if $V$ can be partitioned into three sets $K, S$, and $R$ such that:

(i) $K$ is complete, $S$ is stable, and $|K| = |S| \geq 2$;

(ii) each node in $R$ is adjacent to each node of $K$ and to no node in $S$;

(iii) each node in $S$ has a unique neighbor in $K$; more formally, there exists a bijection $f : K \to S$ such that every node $k \in K$ is adjacent to $f(k) \in S$ and to no other node in $S$.

The complement of a thin spider is a *thick spider*.

The special case of these graphs in which $R = \emptyset$ is considered in [20] and they are called *n-split matching* and *n-split antimatching* graphs, respectively. Examples are shown in Figure 2. Note that the 3-split matching is sometimes called a *net* and denoted by $\bar{S_3}$, while the 3-split antimatching is denoted by $S_3$ [26].

We will denote by $\mathcal{SM}$ and $\mathcal{SA}$ the classes of split matching and split antimatching graphs, respectively.

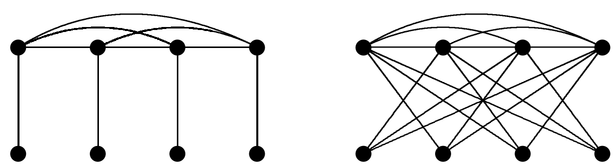

**Fig. 2** *A 4-split matching and a 4-split antimatching.*

The *Dilworth number* [27] of a graph is the size of the largest subset of its nodes in which no closed neighborhood of any node contains the neighborhood of another.

The class of threshold graphs has been introduced many times in several contexts, with different names and various equivalent definitions (see, for example, [42]). For the purposes of this paper, it is sufficient to say that the class of *threshold graphs* $\mathcal{T}$ is characterized as all the graphs with Dilworth number 1 [25]. Note that threshold graphs are split graphs. An example of a threshold graph is depicted in Figure 3(a).

A graph $G = (V, E)$ is a *threshold tolerance graph* [45] if every node $v_i$ can be assigned a real weight $a_i$ and a real tolerance $t_i$ such that for every $(v_i, v_j) \in E \Leftrightarrow |a_i + a_j| \geq \min\{t_i, t_j\}$. Figure 3(b) shows a threshold tolerance graph. Threshold tolerance graphs were introduced in [45] as a generalization of threshold graphs. Indeed, threshold graphs constitute a proper subclass obtained by considering a constant tolerance function [46].

**Fig. 3** (a) *A threshold graph.* (b) *A threshold tolerance graph.*

A graph is *planar* if it can be embedded in the plane so that no two edges intersect geometrically except at the node to which both of them are incident. A graph is *outerplanar* if it has a planar embedding where all nodes are on the outer face.

A *ladder* consists of two distinct paths of the same length $u_1, \ldots, u_{n/2}$ and $v_1, \ldots, v_{n/2}$ plus the edges $(u_i, v_i)$, $i = 1, \ldots, n/2$.

A *chord* of a cycle $C$ is an edge not in the edge set of $C$ whose endpoints lie in the node set of $C$. We say that an edge is a chord of a graph if it is a chord of some cycle in the graph.

A graph is *chordal* if every cycle of length at least 4 has a chord.

A *sun* graph [31] is a graph $G$ on $2n$ nodes for some $n \geq 3$ whose node set can be partitioned into two sets, $W = \{w_1, \ldots, w_n\}$ and $U = \{u_1, \ldots, u_n\}$, such that $U$ induces a clique, $W$ is an independent set, and for each $i$ and $j$, $w_j$ is adjacent to $u_i$ if and only if $i = j$ or $i \equiv j + 1 \pmod{n}$. In Figure 4(a) a sun with $n = 4$ is depicted.

*Strongly chordal graphs* [31] are sun-free chordal graphs.

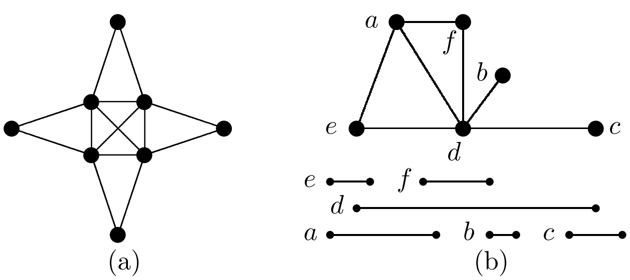


**Fig. 4** (a) *A sun graph.* (b) *An interval graph.*

A *t-caterpillar* [36] is a tree in which all the nodes are within distance $\leq 1$ of a central path, called a *spine*, constituted of $t$ nodes.

A graph $G$ is an *intersection graph* if its nodes correspond to a family of sets $\{S_v : v \in V(G)\}$ and $(u, v)$ is an edge of $G$ if and only if $S_u \cap S_v \neq \emptyset$.

We now consider some graph classes that can be defined through intersection graphs of special objects.

An *interval graph* is the intersection graph of a set of intervals on a line (an example is shown in Figure 4(b)).

A *disk* graph is the intersection graph of disks in the plane. A graph is a *grid intersection* if it is the intersection graph of horizontal and vertical line segments in the plane. A *circular arc* graph is the intersection graph of arcs of a circle.

A graph is a *rectangle (square) intersection* if it has an intersection model consisting of axis-parallel rectangular (squared) boxes in the plane.

A *trapezoid* graph is the intersection graph of trapezoids between two parallel lines. A *permutation* graph is the intersection graph of straight lines between two parallels.

The following chain of inclusions holds:

interval graphs $\subseteq$ circular arc graphs $\subseteq$ permutation graphs $\subseteq$ trapezoid graphs.

A graph is a *tolerance graph* [35] if to every node $v$ can be assigned a closed interval $I_v$ on the real line and a tolerance $t_v$ such that $x$ and $y$ are adjacent if and only if $|I_x \cap I_y| \geq \min\{t_x, t_y\}$, where $|I|$ is the length of the interval $I$. Tolerance graphs can be described through another intersection model, as they are equivalent to *parallelepiped* graphs, defined as the intersection graphs of *special parallelepipeds* on two parallel lines.

**3. Complexity of Recognizing Pairwise Compatibility Graphs.** The problem of recognizing whether a graph is a PCG is formally defined as follows.

PROBLEM 1 (the PCG recognition problem).

*INSTANCE : A graph* $G = (V, E)$.

*DECIDE : Are there a tree* $T$, *an edge-weight function* $w$, *and two integers* $d_{min}, d_{max}$ *such that* $G = PCG(T, w, d_{min}, d_{max})$?

The complexity of this problem is still unknown. However, in [30] NP-completeness is proved for following generalizations.

PROBLEM 2 (the max-generalized PCG recognition problem).
*INSTANCE : A graph $G$, a subset $S$ of the edges of its complement graph, and a positive integer $k$.*
*DECIDE : Is there a $G' = PCG(T, w, d_{min}, d_{max})$ such that $G'$ contains $G$ as a (not necessary induced) subgraph but does not contain any edge of $S$, and at least $k$ edges of $S$ have distance greater than $d_{max}$ between their corresponding leaves in $T$?*

Observe that when $S = \bar{E}$, then the problem becomes exactly that of determining whether $G$ is a PCG.

It is worth noting that the problem of sampling a set of $m$ leaves for a weighted tree $T$, such that their pairwise distance is within some interval $[d_{min}, d_{max}]$, reduces to selecting a clique of size $m$ uniformly at random from the graph $PCG(T, w, d_{min}, d_{max})$. As the sampling problem can be solved in polynomial time on PCGs [39], it follows that the max clique problem is solved in polynomial time on this class of graphs, providing that the tree $T$, the weight function $w$, and the two values $d_{min}, d_{max}$ are known or can be found in polynomial time.

OPEN PROBLEM. *Determine the computational complexity of the PCG recognition problem. Durocher, Mandal, and Rahman* [30] *conjecture that the PCG recognition problem is NP-hard.*

For LPGs, given an integer $k$ we can formulate the following problem.

PROBLEM 3 (the $k$-LPG recognition problem).
*INSTANCE : A graph $G = (V, E)$.*
*DECIDE : Is there a tree $T$ such that $G = LPG(T, k)$?*

In [29] it is shown that this problem can be solved in polynomial time if the $(k-2)$-Steiner root problem can be solved in polynomial time. Chang and Ko [23] give a linear time algorithm for the 3-Steiner root problem, implying that the $k$-leaf power recognition problem can be solved in linear time for $k = 5$.

OPEN PROBLEM. *Determine the computational complexity of the $k$-LPG recognition problem for $k \geq 6$.*

**4. LPG ∩ mLPG.** Here we give some relations between mLPGs and LPGs, the two main subclasses of PCG. In particular, the next result holds.

PROPOSITION 2 (see [16]). *The class co-LPG coincides with the class mLPG and, consequently, the class co-mLPG coincides with the class LPG.*

The relationship between the classes LPG and mLPG is graphically shown in Figure 5 and is deduced from the following considerations:

- The union of the classes LPG and mLPG does not coincide with the whole class PCG. Indeed, the class $\mathcal{C}$ of cycles is in the class PCG but does not belong to the classes LPG or mLPG [16, 55].
- The class $\mathcal{T}$ of threshold graphs belongs to $LPG \cap mLPG$ [20].
- The class $\mathcal{SM}$ of split matchings belongs to $mLPG \setminus LPG$, while the class $\mathcal{SA}$ of split antimatchings belongs to $LPG \setminus mLPG$ [20].

The previous arguments lead to the following summarizing result.

THEOREM 3. *For the classes of LPG and mLPG, the following relations hold:* (a) $mLPG \cup LPG \subset PCG$; (b) $mLPG \cap LPG \neq \emptyset$; (c) $LPG \setminus mLPG \neq \emptyset$; (d) $mLPG \setminus LPG \neq \emptyset$.

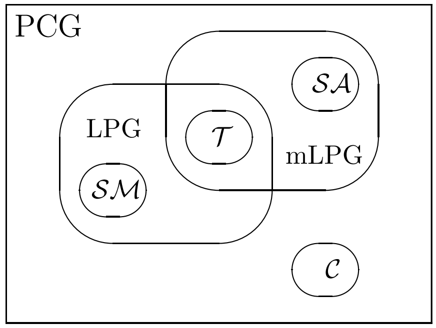


**Fig. 5** *Relationships between PCG, LPG, and mLPG.*

Particular attention in the literature has been given to the characterization of the intersection of the mLPG and LPG classes. Due to Proposition 2, it is clear that a self-complemented class that is included either in LPG or in mLPG is also included in $LPG \cap mLPG$. For example, split permutation graphs that are the intersection class between interval and cointerval graphs are in $LPG \cap mLPG$ as interval graphs are in LPG (see section 6.1). Nevertheless, a complete characterization of the set of graphs in this intersection is still missing. A graph class that is known to be included inside $LPG \cap mLPG$ is the class of threshold graphs (which have Dilworth number 1) [20]. More generally for graphs with arbitrary Dilworth number, the following result holds.

THEOREM 4 (see [19, 18, 17]).

- *All Dilworth* 1 *graphs (i.e., threshold graphs) are in* $LPG \cap mLPG$ *and the witness trees are stars.*
- *All Dilworth* 2 *graphs are in* $LPG \cup mLPG$ *and the witness trees are* 2-*caterpillars. A proper subclass of Dilworth* 2 *graphs (properly containing threshold graphs) is in* $PCG \cap mLPG$.
- *Given a* $t$-*caterpillar* $\Gamma$, *for any edge-weight* $w$ *and any value* $c$, *the graphs* $LPG(\Gamma, w, c)$ *and* $mLPG(\Gamma, w, c)$ *are Dilworth* $t$ *graphs; nevertheless, there are Dilworth* $t$ *graphs,* $t \geq 2$, *that do not belong to* $LPG \cap mLPG$. *As an example, the graph depicted in Figure* 6 *has Dilworth number* 3 *and is neither an LPG nor an mLPG.*

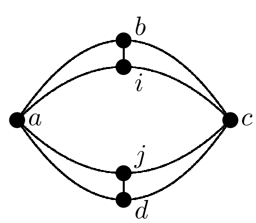


**Fig. 6** *A graph on six nodes that has Dilworth number* 3 *and is neither a LPG nor a mLPG.*

OPEN PROBLEM. *Characterize completely the intersection of mLPG and LPG.*

## 5. Graphs That Are Not PCGs.

Initially it was believed that every graph was a PCG. Indeed, Phillips [49] first proved in an exhaustive way that all graphs with less

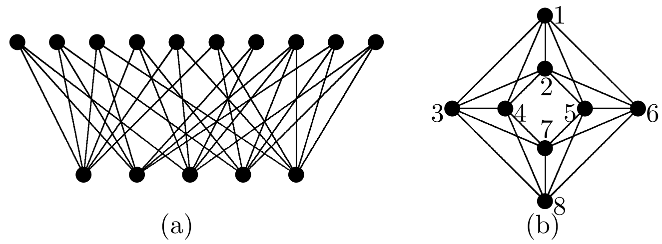


**Fig. 7** (a) *The first graph proven not to be a PCG.* (b) *The graph of smallest size proven not to be a PCG.*

than five nodes are PCGs, then the result was extended to all graphs with at most seven nodes [14] and finally to all bipartite graphs on eight nodes [43].

However, not all graphs are PCGs: Yanhaona, Bayzid, and Rahman [54] showed a bipartite graph with 15 nodes (depicted in Figure 7(a)) that is not a PCG. Subsequently, Mehnaz and Rahman [43] provided a list of bipartite graphs not in PCG. More recently, Durocher, Mondal, and Rahman [30] proved that there exists a (not bipartite) graph with 8 nodes that is not a PCG (depicted in Figure 7(b)). In view of the previously listed results, this is the smallest graph that is not a PCG. The same authors also provided an example of a planar graph with 20 nodes that is not a PCG (depicted in Figure 8). As a consequence, neither bipartite nor planar graphs are included in the PCG class.

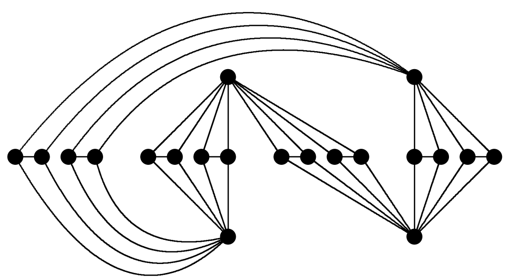

**Fig. 8** *A planar graph with* 20 *nodes that is not a PCG.*

It is known that graph $H$ depicted in Figure 7(b) is not a PCG [44]. On the other hand, Figures 9(a), 9(b), and 9(c) show a representation of graph $H$ as a disk graph, as a circular arc graph, and as square intersection graphs, respectively. This is enough to ensure that all these graph classes are not in PCG [15]. Moreover, rectangle (square) intersection graphs are a superclass of grid intersection graphs, and hence they are not PCGs. Recalling the chain of inclusions stated in the preliminaries, we can deduce that trapezoid and permutation graphs are not PCGs.

Finally, in [15] it was shown that tolerance graphs are not PCGs.

Open Problem. *Find other graph classes that do not belong to the PCG class.*

**6. Graph Classes in PCG.** In this section we list the graph classes which are proven to belong to the PCG class. For many of these graph classes it is also known whether or not they belong to mLPG or LPG. Hence, for easier reading we state the results concerning LPG and mLPG separately.

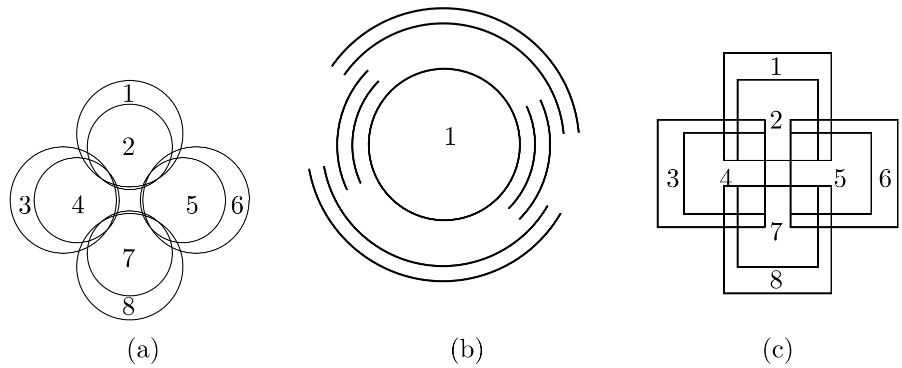


**Fig. 9** *Representation of $H$ as:* (a) *a disk graph,* (b) *a circular arc graph,* (c) *a square intersection graph.*

**6.1. LPGs.** Observe that trees are sun-free and chordal and that taking powers and induced subgraphs does not destroy this property [51]. It follows that every LPG is strongly chordal, i.e., $(C_{n+4}, sun)$-free, $n \geq 0$ [31].

However, not every strongly chordal graph is an LPG: as an example, the graph found by Bibelnieks and Dearing [2] and shown in Figure 10 is a strongly chordal graph and is not an LPG.

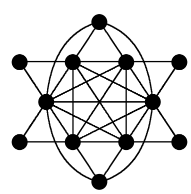

**Fig. 10** *A strongly chordal graph which is not an LPG.*

Neighborhood subtree tolerance (NeST) graphs were introduced by Bibelnieks and Dearing [2] and were also studied in [37]. For the sake of brevity, we avoid defining NeST graphs here, and we only mention that Brandstädt et al. [5] show that LPG coincides with the fixed tolerance NeST graph class.

LPG is a superclass of ptolemaic graphs [4, 5] and even a superclass of directed rooted path graphs, introduced by Gavril [34]. Interval graphs are LPGs [4]; it follows that quasi-threshold graphs (that are $P_4$-free interval graphs) are also LPGs.

Open Problem. *Characterize the graphs that are in the LPG class.*

As we have previously mentioned, a graph is leaf power if it is $k$-leaf power for some integer $k$. Thus, it is interesting to exploit the structure of these subclasses of LPGs. Obviously, a graph $G$ is a 2-leaf power graph if and only if it is the disjoint union of cliques, that is, $G$ does not contain a chordless path of length 2. Dom et al. [29, 28] prove that 3-leaf power graphs are exactly the graphs that do not contain an induced bull, dart, or gem (see Figure 11).

Brandstädt and Le [6] provide another characterization of 3-leaf power graphs by showing that they are exactly the graphs that result from substituting cliques into the nodes of a tree. Moreover, they give a linear time algorithm to recognize 3-leaf power graphs based on their characterization.

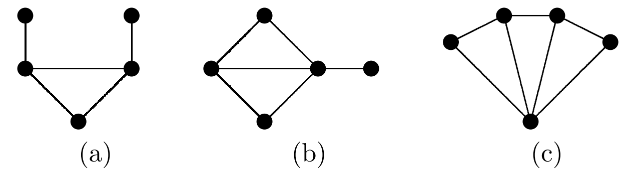


**Fig. 11** (a) *A bull.* (b) *A dart.* (c) *A gem.*

Basic 4-leaf power graphs, i.e., the 4-leaf power graphs without true twins (two connected nodes with the same neighborhood), are characterized by eight forbidden subgraphs [50]. It is shown that every 4-leaf power graph results from substituting cliques into the nodes of a basic 4-leaf power graph. Thus, a characterization of basic 4-leaf power graphs automatically leads to a characterization of 4-leaf power graphs in general [10].

Concerning 5-leaf power graphs, a polynomial time recognition algorithm was given in [23]. However, again no structural characterization is known, even for basic 5-leaf power graphs; only for distance-hereditary basic 5-leaf power graphs has a characterization in terms of 34 forbidden induced subgraphs been discovered [7].

For general $k$, it is proved that $k$-leaf power graphs are not included in the $(k+1)$-leaf power graphs class [11, 12]. Beside these results, there has not been much progress made toward the characterization of these graph classes and the following problem remains open.

Open Problem. *Determine the structure of $k$-leaf power graphs for $k \geq 5$.*

Recently, Nevries and Rosenke [47] provided a list of seven graphs that cannot be induced subgraphs of any LPG, and they conjectured that these are sufficient to characterize LPG in terms of forbidden subgraphs. We remark only that one of these graphs is the one already presented in Figure 10, while the other six graphs are strongly chordal graphs of smaller size. Before this work, it was conjectured that the graph of size 12 in Figure 10 was the smallest strongly chordal graph which does not belong to LPG. Nevertheless, the results in [47] imply that the smallest known strongly chordal graph that does not belong to LPG has ten nodes.

Open Problem. *It remains an open problem to either prove or disprove the conjecture stating that the LPG class can be defined as the class of graphs that does not contain any of the seven subgraphs provided in* [47]. *It is important to note that if this conjecture is true, it would imply a polynomial time recognition algorithm for LPGs.*

**6.2. mLPGs.** A graph is $2K_2$-free if it does not contain an independent pair of edges as an induced subgraph. Recall that LPGs are chordal and hence $C_4$-free. Consequently, their complement mLPG is $2K_2$-free.

Observe that a tree that is $2K_2$-free cannot have a path of length greater than 3, and hence it has a diameter at most 3. It follows that every tree of diameter at least 4 does not belong in the mLPG class.

In [20] it is proved that split matching graphs are not in the mLPG class. As split matching graphs are $2K_2$-free, this means that the mLPG class does not coincide with the $2K_2$-free graph class.

Finally, it has been proved in [22] that threshold tolerance graphs are strictly included in the mLPG class.

Open Problem. *It follows from the results presented in this survey that if $G$ is in $LPG \cap mLPG$, then $G$ is ($2K_2, C_{n+4}$, sun, split matching, split antimatching)-free. It would be interesting to characterize the class $LPG \cap mLPG$ in terms of forbidden subgraphs.*

**6.3. PCGs.** Many graph classes have been proved to be in the PCG class: cycles, single chord cycles, cacti, tree power graphs, Steiner $k$-power, and phylogenetic $k$-power graphs [54, 55]. More recently, even trees, ladder graphs, triangle-free outerplanar 3-graphs [52], and Dilworth 2 graphs [19] have been proved to be PCGs. All these graphs admit as a witness tree a caterpillar.

We have already stated that the class of bipartite graphs is not included in the PCG class. However, in [54] some particular subclasses of bipartite graphs are proved to be PCGs.

A *split matrogenic* graph [42] is a graph that can be constructed as a particular composition of split matchings and split antimatchings. More formally, given a split graph $F = (V_K \cup V_S, E(F))$ and a simple graph $H = (V(H), E(H))$, their *composition* is a graph $G = (V, E) = F \circ H$ defined as follows:

- $V = V_K \cup V_S \cup V(H)$,
- $E = E(F) \cup E(H) \cup \{(a, v) : a \in V_K, v \in V(H)\}$.

A split matrogenic graph is the composition of $t$ split graphs $G_i = (K_i, S_i, E_i)$ with $i = 1, \ldots, t$ such that: either $G_i$ is a split matching, or $G_i$ is a split antimatching, or $K_i = \emptyset$ (and $G_i$ is called *stable graph*), or $S_i = \emptyset$ (and $G_i$ is called *clique graph*) [53].

In [21] it is proved that if the split matrogenic graph is composed using only split matching graphs or only split antimatching graphs, then it belongs to the PCG class.

This result was extended to the following larger subclass of split matrogenic graphs [21].

Theorem 5 (see [21]). *Let $H = G_1 \circ \cdots \circ G_t$ be a split matrogenic graph for which there exists an index $1 \leq h \leq t$ such that $G_1, \ldots, G_h$ are all split matching graphs and $G_{h+1}, \ldots, G_t$ are all split antimatching graphs. Then $H$ is in the PCG class.*

In fact, it seems that the order of appearance of a split matching or an split antimatching in the composition of a split matrogenic graph is somehow strictly related to the pairwise compatibility property.

Open Problem. *It would be interesting to understand whether the split matrogenic graph in Figure* 12 *with* 16 *nodes constituted of an* 8*-node split antimatching composed with an* 8*-node matching is in the PCG class. The solution of this problem would shed some light on the possible inclusion of split matrogenic graphs in the PCG class.*

In [16] the authors study the closure properties of the classes PCG, mLPG, and LPG under some common graph operations such as adding an isolated or universal node; adding a degree one node; adding a twin; taking the complement of a graph; and taking the disjoint union of two graphs. Except for its intrinsic interest, this is also important as it is known that many graph classes can be built by means of recursive applications of particular graph operations. Using these results it was proved in [22] that bipartite distance-hereditary graphs are PCGs.

Open Problem. *In* [16] *it was also proved that the classes mLPG, LPG, and PCG exhibit different closure properties under a given graph operation. In particular,*

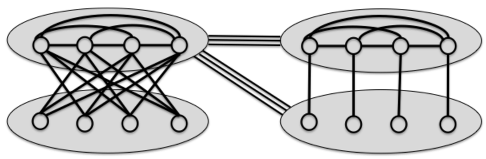

**Fig. 12** *The smallest split matrogenic graph for which it is still an open problem determining whether or not it belongs to the PCG class. The triple lines between the split antimatching graph and the split matching graph represent the composition operation.*

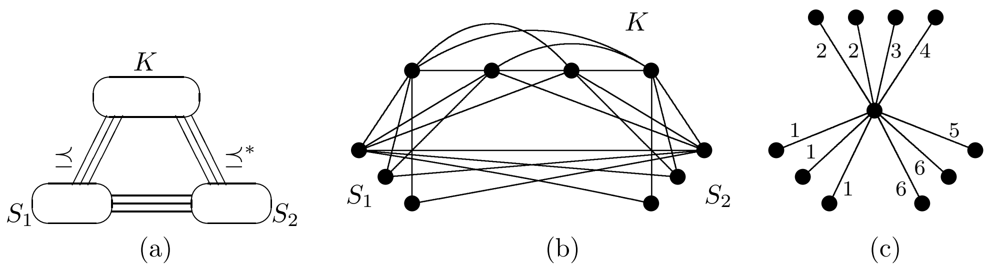


**Fig. 13** (a) *The structure of a PCG generated by a star.* (b) *The PCG generated by the star depicted in* (c).

*the mLPG and LPG classes are not closed under the complement; however, determining whether the PCG class is closed under the complement is still an open problem.*

**7. PCGs of a Particular Tree Topology.** Given a graph, even knowing that it is a PCG, it is in general rather difficult to find the witness tree. In fact, in the literature, most of the trees witnessing that a certain graph class is in the PCG class (or in the LPG or mLPG classes) are very easy structures, such as stars and caterpillars. So, it seems interesting to consider the problem of characterizing subclasses of PCGs derived from a specific topology of the pairwise compatibility tree.

**7.1. Stars.** Stars are a very simple subclass of trees and hence it is natural to ask what graphs are PCGs of a star. In [20] it is proved that threshold graphs are characterized by being LPGs (and mLPGs) of stars. In the same paper, this result was extended to show that PCGs of stars are in fact a special superclass of threshold graphs. In particular, the authors define the following superclass of threshold graphs.

The *vicinal preorder* $\preceq$ of a graph $G = (V, E)$ on the set of nodes $V$ guarantees that for any two nodes $u, v \in V$, $u \preceq v$ if and only if $N(u) \subseteq N[v]$. The *dual preorder* $\preceq^*$ is defined by $u \preceq^* v$ if and only if $v \preceq u$.

A graph $G = (V, E)$ is *nearly three-threshold* if it is possible to partition the set of nodes $V$ into three classes $V_K, V_{S_1}, V_{S_2}$ so that the following hold:

(a) The subgraph induced by $K \cup S_1$ is a threshold graph.
(b) The subgraph induced by $K \cup S_2$ is a threshold graph.
(c) The subgraph induced by $S_1 \cup S_2$ is a bipartite graph.

Furthermore, the total vicinal preorder related to the graph induced by $K \cup S_2$ is the dual of the total vicinal preorder defined by the graph induced by $K \cup S_1$ (see Figure 13(a)).

THEOREM 6 (see [20]). *If a graph $G$ is a PCG of a star, then $G$ is a nearly three-threshold graph.*

OPEN PROBLEM. *Determine whether the class of graphs that are PCGs of a star coincides with the class of nearly three-threshold graphs.*

**7.2. Caterpillars.** Another important tree structure considered is the caterpillar. PCGs of caterpillars are very general graphs, so we first consider a simplified model, i.e., we assume that $w(e) = 1$ for each edge of the tree. Observe that this restriction is natural as in many papers (e.g., see [3, 39]) the tree is not weighted and the distance is defined as the number of edges on the (unique) path connecting two leaves.

The problem of characterizing PCGs of unit weight caterpillars was considered in [4] in the special case of LPGs, providing the following result:

THEOREM 7 (see [4]). *Let $G$ be an $n$-node connected graph and $\Gamma_n$ be a unit weight $n$-leaf caterpillar. Then the following statements are equivalent:*
1. $G = LPG(\Gamma_n, d_{max})$.
2. *$G$ is a unit interval graph.*

In [13], the authors generalize the previous result to PCGs of unit weight caterpillars.

THEOREM 8 (see [13]). *Let $G$ be an $n$-node connected graph and $\Gamma_n$ be a unit weight $n$-leaf caterpillar. Then the following statements are equivalent:*
1. $G = PCG(\Gamma_n, d_{min}, d_{max})$.
2. $G = P_n^{d_{max}-2} - P_n^{d_{min}-3}$ *if* $d_{min} > 3$ *and* $G = P_n^{d_{max}-2}$, *otherwise, where* $P_n^i$ *is the* $i$*th power of the* $n$*-node path.*

The authors of [13] then generalize the model to general weighted caterpillars, giving some properties of the resulting PCG. In particular, they give some conditions on the weight function $w$ and on $d_{max}$ such that $PCG(\Gamma_n, w, d_{min}, d_{max})$ is either triangle-free or has an induced clique.

Unfortunately, we are far from giving a characterization, so the following open problem holds:

OPEN PROBLEM. *Give a complete characterization of PCGs of caterpillars.*

**7.3. Other Trees.** It is worth mentioning that the 7-node wheel $W_7$ is proved to be a PCG, but not a PCG of a caterpillar [13], and the witness tree is shown in Figure 14(b) [14]. As a consequence, caterpillars cannot generate all PCGs, and this fact makes the last open problem of the previous section even more significant.

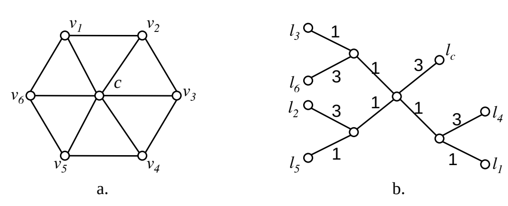


**Fig. 14** (a) *The wheel $W_7$ and* (b) *the edge-weighted tree $T$ such that $W_7 = PCG(T, w, 5, 7)$.*

OPEN PROBLEM. *It is not known whether or not wheels on at least eight nodes are PCGs.*

Given a graph known to be in the PCG class, finding its witness tree is far from trivial. A brute force approach is unfeasible as there are too many $n$-leaf trees to check (and, on each of them, it is necessary to check all possible edge weights). The following result goes some way toward simplifying the search the topology providing a unifying tree structure.

THEOREM 9 (see [14]). *Let $G$ be a graph and $T$ a tree. If $G = PCG(T, w, d_{min}, d_{max})$, then there always exist a full binary tree $\Lambda$, a new edge-weight function $w'$, and a new value $d'_{max}$ such that $G = PCG(\Lambda, w', d_{min}, d'_{max})$.*

Unfortunately, the previous theorem does not guarantee a unique tree, but it is still a practical improvement for the pairwise compatibility tree construction problem, as it leads to the consideration of only a particular subclass of all the $n$-leaf trees.

**8. Conclusions.** Pairwise compatibility graphs were introduced in the context of phylogenetics and they generalize the well-studied class of leaf power graphs. Much attention has been dedicated to them in the literature; however, as shown by this survey, many problems remain open and we are still far from a complete characterization of the PCG class. Any progress toward the solution of the latter problem would be interesting not only from a graph theory perspective, but it also could help in the design of better sampling algorithms for phylogenetic trees. Finally, we conclude by observing that lately it has become more and more evident that phylogenetic networks may provide an alternative to phylogenetic trees and may be more suitable for datasets where evolution involves significant amounts of reticulate events such as hybridization, horizontal gene transfer, or recombination [1, 40]. Thus, aside from the many existing open problems in this area, it could be interesting to consider possible extensions of these problems and concepts to network graphs.

## REFERENCES


[1] E. Bapteste, M. O'Malley, R. Beiko, M. Ereshefsky, J. P. Gogarten, L. Franklin-Hall, F.-J. Lapointe, J. Dupre, T. Dagan, Y. Boucher, and W. Martin, *Prokaryotic evolution and the tree of life are two different things*, Biology Direct, 4 (2009), p. 34. (Cited on p. 458)

[2] E. Bibelnieks and P. M. Dearing, *Neighborhood subtree tolerance graphs*, Discrete Appl. Math., 43 (1993), pp. 13–26. (Cited on p. 453)

[3] A. Brandstädt, *On Leaf Powers*, Technical report, University of Rostock, 2010. (Cited on pp. 446, 457)

[4] A. Brandstädt and C. Hundt, *Ptolemaic graphs and interval graphs are leaf powers*, in LATIN 2008: Theoretical Informatics, E. Laber, C. Bornstein, L. Nogueira, and L. Faria, eds., Lecture Notes in Comput. Sci. 4957, Springer, Berlin, Heidelberg, 2008, pp. 479–491. (Cited on pp. 453, 457)

[5] A. Brandstädt, C. Hundt, F. Mancini, and P. Wagner, *Rooted directed path graphs are leaf powers*, Discrete Math., 310 (2010), pp. 897–910. (Cited on pp. 446, 447, 453)

[6] A. Brandstädt and V. B. Le, *Structure and linear time recognition of* 3*-leaf powers*, Inform. Process. Lett., 98 (2006), pp. 133–138. (Cited on pp. 446, 453)

[7] A. Brandstädt, V. B. Le, and D. Rautenbach, *Distance-hereditary* 5*-leaf powers*, Discrete Math., 309 (2009), pp. 3843–3852. (Cited on p. 454)

[8] A. Brandstädt, V. B. Le, and D. Rautenbach, *Exact leaf powers*, Theoret. Comput. Sci., 411 (2010), pp. 2968–2977. (Cited on p. 447)

[9] A. Brandstädt, V. B. Le, and J. P. Spinrad, *Graph Classes: A Survey*, SIAM, Philadelphia, 1999. (Cited on p. 446)

[10] A. Brandstädt, V. B. Le, and R. Sritharan, *Structure and linear-time recognition of* 4*-leaf powers*, ACM Trans. Algorithms, 5 (2008), article 11. (Cited on pp. 446, 454)

[11] A. Brandstädt and P. Wagner, *On $k$- versus $(k+1)$-leaf powers*, in Combinatorial Optimization and Applications, B. Yang, D.-Z. Du, and C. Wang, eds., Lecture Notes in Comput. Sci. 5165, Springer Berlin Heidelberg, 2008, pp. 171–179. (Cited on p. 454)

[12] A. Brandstädt and P. Wagner, *The complete inclusion structure of leaf power classes*, Theoret. Comput. Sci., 410 (2009), pp. 5505–5514. (Cited on p. 454)

[13] T. Calamoneri, A. Frangioni, and B. Sinaimeri, *Pairwise compatibility graphs of caterpillars*, Comput. J., 57 (2014), pp. 1616–1623. (Cited on p. 457)

[14] T. Calamoneri, D. Frascaria, and B. Sinaimeri, *All graphs with at most seven vertices are pairwise compatibility graphs*, Comput. J., 56 (2013), pp. 882–886. (Cited on pp. 452, 457, 458)

[15] T. Calamoneri, M. Gastaldello, and B. Sinaimeri, *On Pairwise Compatibility of Some Graph (Super)classes*, preprint, arXiv:1504.06454, 2015. (Cited on p. 452)

[16] T. Calamoneri, E. Montefusco, R. Petreschi, and B. Sinaimeri, *Exploring pairwise compatibility graphs*, Theoret. Comput. Sci., 468 (2013), pp. 23–26. (Cited on pp. 447, 450, 455)

[17] T. Calamoneri and R. Petreschi, *Corrigendum to "On pairwise compatibility graphs having Dilworth number two" [Theoret. Comput. Sci.,* 524 (2014) 34–40*]*, Theoret. Comput. Sci., 602 (2015), pp. 158–159. (Cited on p. 451)

[18] T. Calamoneri and R. Petreschi, *"On pairwise compatibility graphs having Dilworth number* $k$, Theoret. Comput. Sci., 547 (2014), pp. 82–89. (Cited on p. 451)

[19] T. Calamoneri and R. Petreschi, *On pairwise compatibility graphs having Dilworth number two*, Theoret. Comput. Sci., 524 (2014), pp. 34–40. (Cited on pp. 451, 455)

[20] T. Calamoneri, R. Petreschi, and B. Sinaimeri, *On relaxing the constraints in pairwise compatibility graphs*, in WALCOM: Algorithms and Computation, Md. S. Rahman and S.-i. Nakano, eds., Lecture Notes in Comput. Sci. 7157, Springer, Berlin, Heidelberg, 2012, pp. 124–135. (Cited on pp. 446, 448, 450, 451, 454, 456, 457)

[21] T. Calamoneri, R. Petreschi, and B. Sinaimeri, *Pairwise compatibility property of some superclasses of threshold graphs*, Discrete Math. Algorithms Appl., 5 (2013), article 360002. (Cited on p. 455)

[22] T. Calamoneri and B. Sinaimeri, *Relating threshold tolerance graphs to other graph classes*, in Proceedings of the 15th Italian Conference on Theoretical Computer Science, Perugia, Italy, 2014, pp. 73–79. (Cited on pp. 454, 455)

[23] M.-S. Chang and M.-T. Ko, *The 3-Steiner root problem*, in Graph-Theoretic Concepts in Computer Science, 33rd International Workshop, WG 2007, A. Brandstädt, D. Kratsch, and H. Müller, eds., Lecture Notes in Comput. Sci. 4769, Springer, Berlin, Heidelberg, 2007, pp. 109–120. (Cited on pp. 450, 454)

[24] Z.-Z. Chen, T. Jiang, and G. Lin, *Computing phylogenetic roots with bounded degrees and errors*, SIAM J. Comput., 32 (2003), pp. 864–879. (Cited on p. 446)

[25] V. Chvatal and P. L. Hammer, *Aggregation of inequalities in integer programming*, in Studies in Integer Programming, B. H. Korte, P. L. Hammer, E. L. Johnson, and G. L. Nemhauser, eds., Ann. Discrete Math. 1, Elsevier, Amsterdam, 1977, pp. 145–162. (Cited on p. 448)

[26] H. N. de Ridder et al., *Information System on Graph Classes and their Inclusions (ISGCI)*, http://www.graphclasses.org. (Cited on p. 448)

[27] R. P. Dilworth, *A decomposition theorem for partially ordered sets*, Ann. of Math. (2), 51 (1950), pp. 161–166. (Cited on p. 448)

[28] M. Dom, J. Guo, F. Hüffner, and R. Niedermeier, *Error compensation in leaf root problems*, in Algorithms and Computation, R. Fleischer and G. Trippen, eds., Lecture Notes in Comput. Sci. 3341, Springer, Berlin, Heidelberg, 2005, pp. 389–401. (Cited on p. 453)

[29] M. Dom, J. Guo, F. Hüffner, and R. Niedermeier, *Extending the tractability border for closest leaf powers*, in Graph-Theoretic Concepts in Computer Science, 31st International Workshop, WG 2005, D. Kratsch, ed., Lecture Notes in Comput. Sci. 3787, Springer, Berlin, Heidelberg, 2005, pp. 397–408. (Cited on pp. 450, 453)

[30] S. Durocher, D. Mondal, and Md. S. Rahman, *On graphs that are not PCGs*, Theoret. Comput. Sci., 571 (2015), pp. 78–87. (Cited on pp. 450, 452)

[31] M. Farber, *Characterizations of strongly chordal graphs*, Discrete Math., 43 (1983), pp. 173–189. (Cited on pp. 449, 453)

[32] J. Felsenstein, *Cases in which parsimony or compatibility methods will be positively misleading*, Systematic Zoology, 27 (1978), pp. 401–410. (Cited on p. 446)

[33] S. Földes and P. L. Hammer, *Split graphs*, Congr. Numer. 19 (1977), pp. 311–315. (Cited on p. 447)

[34] F. Gavril, *A recognition algorithm for the intersection graphs of directed paths in directed trees*, Discrete Math., 13 (1975), pp. 237–249. (Cited on p. 453)

[35] M. C. Golumbic, C. L. Monma, and W. T. Trotter, *Tolerance graphs*, Discrete Appl. Math., 9 (1984), pp. 157–170. (Cited on p. 449)

[36] F. Harary and A. J. Schwenk, *The number of caterpillars*, Discrete Math., 6 (1973), pp. 359–365. (Cited on p. 449)

[37] R. B. Hayward, P. E. Kearney, and A. Malton, *Nest graphs*, Discrete Appl. Math., 121 (2002), pp. 139–153. (Cited on p. 453)

[38] B. Jamison and S. Olariu, *Recognizing $P_4$-sparse graphs in linear time*, SIAM J. Comput., 21 (1992), pp. 381–406. (Cited on p. 447)

[39] P. E. Kearney, J. I. Munro, and D. Phillips, *Efficient generation of uniform samples from phylogenetic trees*, in Algorithms in Bioinformatics, G. Benson and R. Page, eds., Lecture Notes in Comput. Sci. 2812, Springer, Berlin, Heidelberg, 2003, pp. 177–189. (Cited on pp. 446, 450, 457)

[40] F. J. Lapointe, P. Lopez, Y. Boucher, J. Koenig, and E. Bapteste, *Clanistics: A multi-level perspective for harvesting unrooted gene trees*, Trends Microbiology, 18 (2010), pp. 341–347. (Cited on p. 458)

[41] G. Lin, P. E. Kearney, and T. Jiang, *Phylogenetic k-root and Steiner k-root*, in Algorithms and Computation, G. Goos, J. Hartmanis, J. van Leeuwen, D. T. Lee, and S.-H. Teng, eds., Lecture Notes in Comput. Sci. 1969, Springer, Berlin, Heidelberg, 2000, pp. 539–551. (Cited on p. 446)

[42] N. V. R. Mahadev and U. N. Peled, *Threshold Graphs and Related Topics*, Ann. Discrete Math., Elsevier, Burlington, MA, 1995. (Cited on pp. 448, 455)

[43] S. Mehnaz and Md. S. Rahman, *Pairwise compatibility graphs revisited*, in Proceedings of the 2013 International Conference on Informatics, Electronics Vision (ICIEV), 2013, pp. 1–6. (Cited on p. 452)

[44] D. Mondal, S. Durocher, and Md. S. Rahman, *On graphs that are not PCGs*, in WALCOM: Algorithms and Computation, S. K. Ghosh and T. Tokuyama, eds., Lecture Notes in Comput. Sci. 7748, Springer, Berlin, Heidelberg, 2013, pp. 310–321. (Cited on p. 452)

[45] C. L. Monma, B. Reed, and W. T. Trotter, Jr., *A generalization of threshold graphs with tolerance*, Congr. Numer., 55 (1986), pp. 187–197. (Cited on p. 448)

[46] C. L. Monma, B. Reed, and W. T. Trotter, Jr., *Threshold tolerance graphs*, J. Graph Theory, 12 (1988), pp. 343–362. (Cited on p. 448)

[47] R. Nevries and C. Rosenke, *Towards a characterization of leaf powers by clique arrangements*, in SOFSEM 2015: Theory and Practice of Computer Science, Lecture Notes in Comput. Sci. 8939, Springer, Berlin, Heidelberg, 2015, pp. 364–376. (Cited on p. 454)

[48] N. Nishimura, P. Ragde, and D. M. Thilikos, *On graph powers for leaf-labeled trees*, J. Algorithms, 42 (2002), pp. 69–108. (Cited on p. 446)

[49] D. Phillips, *Uniform Sampling from Phylogenetic Trees*, Master's thesis, University of Waterloo, Canada, 2002. (Cited on p. 451)

[50] D. Rautenbach, *Some remarks about leaf roots*, Discrete Math., 306 (2006), pp. 1456–1461. (Cited on p. 454)

[51] A. Raychaudhuri, *On powers of strongly chordal and circular graphs*, Ars Combin., 34 (1992), pp. 147–160. (Cited on p. 453)

[52] S. A. Salma, Md. S. Rahman, and Md. I. Hossain, *Triangle-free outerplanar 3-graphs are pairwise compatibility graphs*, J. Graph Algorithms Appl., 17 (2013), pp. 81–102. (Cited on p. 455)

[53] R. Tyshkevich, *Decomposition of graphical sequences and unigraphs*, Discrete Math., 220 (2000), pp. 201–238. (Cited on p. 455)

[54] M. N. Yanhaona, Md. S. Bayzid, and Md. S. Rahman, *Discovering pairwise compatibility graphs*, Discrete Math. Algorithms Appl., 2 (2010), pp. 607–624. (Cited on pp. 452, 455)

[55] M. N. Yanhaona, K. S. M. T. Hossain, and Md. S. Rahman, *Pairwise compatibility graphs*, J. Appl. Math. Comput., 30 (2009), pp. 479–503. (Cited on pp. 450, 455)