Characterizations of undirected 2-quasi best match graphs

Annachiara Korchmaros[a,*], Guillaume E. Scholz[b], Peter F. Stadler[a,c,d,e,f]

[a] *Bioinformatics Group, Department of Computer Science & Interdisciplinary Center for Bioinformatics, Universität Leipzig, Häertelstr. 16-18, Leipzig, D-04107, Sachsen, Germany*
[b] *Independent, Leipzig, Sachsen, 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] *Santa Fe Institute, 1399 Hyde Park Rd., Santa Fe, NM, 87501, USA*

Abstract

Bipartite best match graphs (BMG) and their generalizations arise in mathematical phylogenetics as combinatorial models describing evolutionary relationships among related genes in a pair of species. In this work, we characterize the class of *undirected 2-quasi-BMGs* (un2qBMGs), which form a proper subclass of the P_6-free chordal bipartite graphs. We show that un2qBMGs are exactly the class of bipartite graphs free of P_6, C_6, and the eight-vertex Sunlet$_4$ graph. Equivalently, a bipartite graph G is un2qBMG if and only if every connected induced subgraph contains a "heart-vertex" which is adjacent to all the vertices of the opposite color. We further provide a $O(|V(G)|^3)$ algorithm for the recognition of un2qBMGs that, in the affirmative case, constructs a labeled rooted tree that "explains" G. Finally, since un2qBMGs coincide with the (P_6, C_6)-free bi-cographs, they can also be recognized in linear time.

Keywords: chordal bipartite graph, forbidden induced subgraphs, sunlet graphs, recognition algorithm, mathematical phylogenetics

*corresponding author
Email address: annachiara@bioinf.uni-leipzig.de (Annachiara Korchmaros)

1. Introduction

Best match graphs (BMGs) provide a combinatorial abstraction of the evolutionary relationship "*being a closest relative*". BMGs naturally arise in comparative genomics, where they describe the relatedness of homologous genes in different species. Given a leaf-labeled tree (T,σ) with leaf-coloring σ representing the species in which a gene (leaf) resides, there is a directed edge $x \to y$ in the associated BMG whenever y is the *best match* of x among all leaves of color $\sigma(y)$. More precisely, y is a best match for $\sigma(y)$ if $\operatorname{lca}(x,y)$ equals or is a descendant of $\operatorname{lca}(x,z)$ for any other leaf z of the same color as y. The resulting vertex-colored digraph $\operatorname{BMG}(T,\sigma)$ reflects part of the ancestral structure of T. It has been studied extensively from graph-theoretic and algorithmic viewpoints [4, 16, 17, 18]; see Section 2 for formal definitions. This approach has recently been used to develop software for estimating phylogenies of trees and orthology relations from sequence data [14, 15].

Quasi-best match graphs (qBMGs) have been introduced as a generalization of BMGs [11]. Similar to BMGs, they derive from leaf-colored tree (T,σ) together with a *truncation map* u_T that assigns to each leaf x and color $s \neq \sigma(x)$ a vertex on the path from the root to x; see Section 2 for formal definitions. This map determines how far the search for the best matches extends toward the root. The resulting digraph $\operatorname{qBMG}(T,\sigma,u_T)$ thus encodes edges based on partial ancestry information rather than the complete tree topology. This generalization preserves many structural and algorithmic properties of BMGs, including polynomial-time tree reconstruction, while capturing a broader range of leaf-colored digraphs. In contrast to BMGs, qBMGs form a hereditary family of graphs. In biological terms, qBMGs provide a combinatorial framework for evolutionary relationships that depend on a limited ancestral depth, such as when genes are so distant that evolutionary relatedness cannot be determined unambiguously.

If σ takes two colors, the corresponding *2-quasi best match graphs* (2qBMGs) are of remarkable interest in structural graph theory, not only for the already mentioned hereditary property but also for their rich combinatorial behavior, such as large automorphism groups [10], and finite characterizing set of forbidden subgraphs [19, Theorem 4.4]. The 2-colored BMGs are characterized as the sink-free 2qBMGs [11].

The *undirected underlying graphs* of 2qBMGs, denoted *un2qBMGs*, were introduced in [12] as a bridge between the directed and symmetric best-match relations. The subgraphs of a BMG comprising all its symmetric edges (i.e. 2-cycles) are the reciprocal BMGs, which are fundamental in the study of gene family histories [9]. From a graph-theoretic perspective, it is natural to study the underlying undirected graphs of BMGs and qBMGs, leveraging well-established results on undirected graphs to gain new insights into 2qBMGs.

Another remarkable property for structural graph theory, shown in [12], is that un2qBMGs admit no paths of length six (P_6-free) and are chordal bipartite, i.e., they admit no cycles of length six or more. This places un2qBMGs among structured subclasses of bipartite graphs, sharing properties with chordal graphs while remaining non-trivial. Recognition and optimization problems for P_6-free chordal bipartite graphs (also known as (P_6, C_6)-free bipartite) have been studied extensively since the seminal work of [3]. The particular structure of (P_6, C_6)-free graphs was found to be compatible with specific vertex decompositions involving $K \oplus S$ graphs, where $K \oplus S$ denotes a graph whose vertex set can be partitioned into a biclique K and an independent set S. In [13], it is shown that a bipartite graph G is (P_6, C_6)-free if and only if every connected subgraph of G has a $K \oplus S$ vertex decomposition. This structural property enables the recognition of (P_6, C_6)-free bipartite graphs in linear time [20]. Since connected un2qBMGs also admit an $K \oplus S$ vertex decomposition, it is natural to ask whether un2qBMGs can be recognized in polynomial time.

The present paper provides a complete structural and algorithmic characterization of un2qBMGs. In Section 3, we define the notion of a tree (T, σ, u) that *explains* a bipartite graph G and show that G is a un2qBMG if and only if (T, σ, u) explains G. We establish that every connected un2qBMG admits a *least-resolved* explaining tree, analogous to least-resolved trees of BMGs [4] and qBMGs [11]. In particular, in every least-resolved tree (T, σ, u), each internal vertex v corresponds to an edge xy of G with $\mathrm{lca}_T(x, y) = v$, and every associated subtree $T(v)$ is non-monochromatic.

In Section 4, we then analyze the case of small instances and bicliques, showing that all connected bipartite graphs with at most five vertices are un2qBMGs; see Proposition 4.2. In Section 5, we introduce the HEART-TREE algorithm, which iteratively identifies *heart-vertices*, i.e., vertices adjacent to all vertices of the opposite color, and uses them to reconstruct a least-resolved explaining tree. Theorem 5.2 shows that HEART-TREE decides in cubic time whether a bipartite graph is a un2qBMG, and, if so, returns an associated least-resolved tree. The key structural insight is that every connected un2qBMG contains a heart-vertex, see Lemma 5.1, and that the class of un2qBMGs is hereditary with respect to induced subgraphs [12]. Hence, one of the characterizations of un2qBMGs in Theorem 6.3 is that a bipartite graph is a un2qBMG if and only if every connected induced subgraph contains a heart-vertex.

In [12, Theorem 3.4], the authors showed that un2qBMGs are strictly contained in the class of (P_6, C_6)-free bipartite graphs by providing a third forbidden subgraph H for the class of un2qBMGs. However, forbidden subgraphs contained in H were not excluded. In Section 6, we study this problem and characterize un2qBMGs as the bipartite graphs free of induced P_6, C_6, and Sunlet_4 subgraphs; see Theorem 6.3.

2. Preliminaries

Undirected and directed graphs. Our notation and terminology for undirected graphs are standard. Let G be an undirected graph with vertex-set $V(G)$ and edge-set $E(G)$, and let $\overrightarrow{G}$ be a directed graph with vertex-set $V(\overrightarrow{G})$ and edge-set $E(\overrightarrow{G})$. For a graph G, the undirected edge $x - y$ is denoted as xy, and the *degree* of a vertex x in G is the number of vertices $y \in V(G)$ such that $xy \in E(G)$. Similarly, the directed edge $x \to y$ with head x and tail y is denoted as xy, and the indegree and outdegree of a vertex $x \in V(\overrightarrow{G})$ is the number of vertices $y \in V(\overrightarrow{G})$ such that $yx \in E(\overrightarrow{G})$ and $xy \in E(\overrightarrow{G})$, respectively. In an undirected graph G, the *degree* of a vertex x, denoted $\deg_G(x)$, is the number of elements $y \in X$ such that $xy \in E(G)$.

For an undirected graph G and a subset $Y \subseteq V(G)$, the *subgraph of G induced by Y*, denoted by $G[Y]$, is the graph with vertex-set Y and edge-set the set of edges of G with both endpoints in Y.

A *bipartite graph* G is a graph whose vertices are partitioned into two subsets (partition sets) such that the endpoints of every edge fall into different subsets. A *bipartite digraph* $\overrightarrow{G}$ if its underlying undirected graph is bipartite. The partition sets may be viewed as the color classes of a proper vertex coloring σ of two colors in the sense that no edge is incident on vertices colored the same color.

A path-graph P_n with a path $v_1v_2 \cdots v_n$ is a bipartite graph with (uniquely determined) color sets $\{v_1, v_3, \ldots\}$ and $\{v_2, v_4, \ldots\}$. An undirected graph is *P_n-free* if it has no induced subgraph isomorphic to a P_n path-graph. A *cycle C_n* of an undirected graph G is a sequence $v_1v_2 \cdots v_n$ of pairwise distinct vertices of G such that $v_iv_{i+1} \in E(G)$ for $i = 1, \ldots, n-1$, and $v_nv_1 \in E(G)$. A

C_n *cycle-graph* is an undirected graph on n vertices with a cycle $v_1v_2\cdots v_nv_1$ containing no chord, i.e. any edge in $E(G)$ is either $v_iv_{i+1} \in E(G)$ for some $1 \leq i \leq n-1$, or v_nv_1. For n even, a C_n cycle-graph with a cycle $v_1v_2\cdots v_nv_1$ containing no chord is a bipartite graph with (uniquely determined) color classes $\{v_1, v_3, \ldots\}$ and $\{v_2, v_4, \ldots\}$. Bipartite cycle-graphs can only exist for even n. An undirected graph is C_n-*free* if it has no induced subgraph isomorphic to an C_n cycle-graph. The *sunlet graph* Sunlet_n, for $n \geq 3$, is the graph obtained by attaching a single pendant vertex to each vertex of the cycle-graph C_n. For $n = 4$, the sunlet graph is shown in Figure 1.

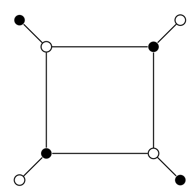

Figure 1: Sunlet_4 with vertex bipartion induced by the vertex-coloring.

Rooted Trees. A directed graph T is a tree if all vertices of T have indegree 0 or 1. A tree T is *rooted* if exactly one vertex of T has indegree 0. In this case, the vertex is the *root* of T, denoted ρ_T. In a rooted tree, its root is typically drawn as the top, and the edges are directed from the parent vertices to their child vertices. A vertex v of T is a *leaf* if it has outdegree 0, and we denote the set of leaves of T by $L(T)$. A vertex of T that is not a leaf is called an *internal vertex* of T. Similarly, an arc vw of T such that w is an internal vertex of T is called an *internal arc* of T.

A tree T is *phylogenetic* if every internal vertex has at least two children. In this contribution, all trees are rooted and phylogenetic. Let $v, w \in V(T)$. Whenever we write $vw \in E(T)$, we assume w is a child of v, and v is the parent of w denoted by $\mathrm{parent}_T(w)$ or, when clear from context $\mathrm{parent}(w)$. We write $\mathrm{child}_T(v)$ or, when clear from context, $\mathrm{child}(v)$ for the set of children of v in T, and $\mathrm{cL}(v)$ for the set of children of v that are leaves, that is, $cL(v) = \mathrm{child}_T(v) \cap L(T)$.

Quasi-best match graphs. Let (T, σ) be a leaf-colored tree with leaf-coloring $\sigma : L \to \sigma(L(T))$. For any two leaves $x, y \in V$, $\mathrm{lca}(x, y)$ denotes the last common ancestor between x and y on T. A leaf $y \in L$ is a best match of the leaf $x \in L$ if $\sigma(x) \neq \sigma(y)$ and $\mathrm{lca}(x, y) \preceq \mathrm{lca}(x, z)$, i.e. $\mathrm{lca}(x, z)$ is an ancestor

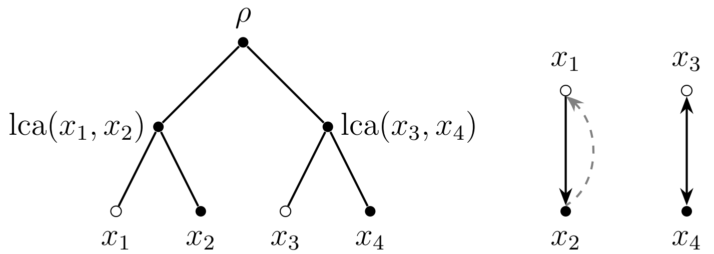

Figure 2: Example of a 2qBMG and 2BMG explained by a rooted phylogenetic tree. Black edges represent the edge-set of the 2qBMG. Black and dashed edges form the edge-set of the 2BMG. The truncation map u of the 2qBMG is $u(x_2) = x_2$ and $u(x) = \rho$ otherwise.

of $\mathrm{lca}(x, y)$, holds for all leaves z of color $\sigma(y) = \sigma(z)$. The *BMG (best match graph) explained by* (T, σ) is the vertex-colored digraph $BMG(T, \sigma)$ with color-set $\sigma(L(T))$, vertex-set $L(T)$, and $xy \in E(G)$ if y is a best match of x in (T, σ). A truncation map $u_T : L(T) \times \sigma(L(T)) \to V(T)$ assigns to every leaf $x \in L(T)$ and color $s \in \sigma(L(T))$ a vertex of T such that $u_T(x, \sigma(x)) = x$ and $u_T(x, s)$ lies along the unique path from ρ_T to x for $s \neq \sigma(x)$. A leaf $y \in L(T)$ is a quasi-best match for $x \in L(T)$ with respect to (T, σ, u_T) if (i) y is a best match of x in (T, σ), and (ii) $\mathrm{lca}_T(x, y) \preceq u_T(x, \sigma(y))$. The *qBMG (quasi-best match graph) explained by* (T, σ, u_T) is the vertex-colored digraph $qBMG(T, \sigma, u_T)$ with color-set $\sigma(L(T))$, vertex-set $L(T)$, and $xy \in E(G)$ if y is a quasi-best match of x in (T, σ, u_T). A vertex-colored digraph $(\overrightarrow{G}, \sigma)$ is a $|S|$*-colored quasi-best match graph*($|S|$-qBMG) if there is a leaf-colored tree (T, σ) together with a truncation map u_T on (T, σ) such that $V(G) = L(T), |S| = |\sigma(L(T))|$, and $\overrightarrow{G} = |S|$-qBMG(T, σ, u_T). In particular, BMGs are qBMGs when $u_T(x, s) = \rho_T$ for $s \neq \sigma(x)$. For general results on quasi-best match graphs, the reader is referred to [11].

A leaf-colored tree (T, σ, u) explaining a qBMG $\overrightarrow{G}$ is *least-resolved* if there is no (T, σ, u') explaining $\overrightarrow{G}$ such that T' can be obtained from T by a sequence of edge-contractions. BMGs are explained by a unique least-resolved tree that can be built in polynomial time [4]. This property has recently been used to build parsimonious explanations of a family of genes [14]. On the other hand, qBMGs do not have a unique least resolved tree, but a phylogenetic tree that explains a qBMG can be built in polynomial time [11].

In this contribution, we are interested in 2qBMGs, that is, $|\sigma(L(T))| = 2$. In this case, $u(x)$ denotes the image of a vertex x under the truncation map

u, assuming that $s \neq \sigma(x)$. Figure 2 shows an example of four vertices.

3. Basic properties of un2qBMGs and least resolved trees

The study of *undirected underlying graph of a 2qBMG (un2qBMG)* was initiated in [12]. In the latter, it was shown that un2qBMGs do not have any induced path of size at least 6, and they are chordal bipartite [12, Corollary 3.3], i.e., un2qBMGs do not admit any induced cycle of length at least 6. The tree (T, σ, u_T) explains the graph G if G is the underlying undirected graph of 2qBMG(T, σ, u_T).

Definition 3.1. *Let T be a phylogenetic tree with a binary leaf-coloring σ and the truncation map u such that $u(x) \in \{x, \rho\}$ for all $x \in L(T)$. A graph G is* explained *by (T, σ, u) if the following two properties hold:*

(i) $V(G) = L(T)$;

(ii) $xy \in E(G)$ whenever $\sigma(x) \neq \sigma(y)$ and

(ii.1) $u(x) \neq x$ and $y \preceq \operatorname{lca}(x, z)$ for all $z \in L(T)$ with $\sigma(z) = \sigma(y)$, or

(ii.2) $u(y) \neq y$ and $x \preceq \operatorname{lca}(y, z)$ for all $z \in L(T)$ with $\sigma(z) = \sigma(x)$.

Any graph G satisfying Definition 3.1 is necessarily bipartite, since condition (ii) requires that $\sigma(x) \neq \sigma(y)$ for every edge $xy \in E(G)$.

Recall that qBMGs are directed graphs explained by a tree. In analogy, un2qBMGs are precisely the undirected graphs explained by a tree, as stated in the following result.

Proposition 3.1. *Let (G, σ) be an undirected graph. G is un2qBMG if and only if there is a tree (T, σ, u) explaining G.*

Proof. Assume first that G is un2qBMG. Let $\overrightarrow{G}$ be a 2qBMG such that G is its underlying undirected graph. Consider a tree (T, σ, u) explaining $\overrightarrow{G}$. We show that (T, σ, u) also explains G. By definition, $V(G) = V(\overrightarrow{G}) = L(T)$, so (i) in Definition 3.1 holds. To see that (ii) in Definition 3.1 holds, let $xy \in E(G)$. Since G is the underlying undirected graph of $\overrightarrow{G}$, one of $xy \in E(\overrightarrow{G})$ or $yx \in E(\overrightarrow{G})$ holds. Suppose that $xy \in E(\overrightarrow{G})$. Therefore, $y \preceq \operatorname{lca}(x, z)$ with $\sigma(z) = \sigma(y)$ and $\operatorname{lca}(x, y) \preceq u(x)$ yielding $u(x) \neq x$. Hence, (ii.1) in Definition 3.1 is satisfied. If $yx \in E(\overrightarrow{G})$ and the roles of x

and y are exchanged, then the same argument implies that condition (ii.2) in Definition 3.1 is satisfied.

Conversely, suppose that there is a tree (T, σ, u) explaining G, and consider the 2qBMG $\overrightarrow{G}$ explained by (T, σ, u). Clearly, $V(\overrightarrow{G}) = L(T) = V(G)$ holds. It remains to show that for all $xy \in E(G)$, one of $xy \in E(\overrightarrow{G})$ or $yx \in E(\overrightarrow{G})$. So, let $xy \in E(G)$. If (ii.1) in Definition 3.1 is satisfied for xy, then $u(x) \neq x$ must hold, so $u(x) = \rho$ and $\mathrm{lca}(x, y) \preceq u(x)$ follows. This, together with the fact that $y \preceq (x, z)$ for all $z \in X$ with $\sigma(z) = \sigma(y)$, implies that $xy \in E(\overrightarrow{G})$. One easily verifies that analogous arguments imply $yx \in E(\overrightarrow{G})$ if case (ii.2) in Definition 3.1 is satisfied. □

Lemma 3.2. *The disjoint union* $(G, \sigma) = \bigcup_{i=1,\ldots m}(G_i, \sigma_i)$ *with* $|\sigma_i(G_i)| = 2$ *is an un2qBMG if and only if each* (G_i, σ_i) *is an un2qBMG.*

Proof. Suppose G is an un2qBMG. Since the property of being an un2qBMG is hereditary [12], every G_i is an un2qBMG. Conversely, assume that G is not an un2qBMG, but every G_i is an un2qBMG. Proposition 3.1 ensures that every G_i is explained by a tree (T_i, σ_i, u_i). Then, consider a star-tree T with $x_1, \ldots, x_m$ leaves. Modify T by rooting T_i in x_i, and define a truncation map u on $L(T)$, as the piecewise function of $u_1, \ldots, u_n$. Definition 3.1 togheter with the assumption that $|\sigma_i(G_i)| = 2$ for every element of the union gives that G is explained by (T, σ, u). □

In this contribution, we will first prove our main results assuming that (G, σ) is a connected properly vertex-colored graph, and then use Lemma 3.2 to extend the results to the case of disconnected graphs.

For v a vertex of T, we say that $T(v)$ is *monochromatic* if all leaves of $T(v)$ have the same color. We have:

Proposition 3.3. *Let G be a connected un2qBMG explained by the tree (T, σ, u). If (T, σ, u) is least-resolved, then for every internal vertex v of T, there exists an edge $xy \in E(G)$ such that $v = \mathrm{lca}_T(x, y)$.*

Proof. By contradiction, suppose that there exists an internal vertex v of T such that no edge $xy \in E(G)$ satisfies $v = \mathrm{lca}_T(x, y)$. We distinguish between two cases, according as $v = \rho$ and $v \neq \rho$.

Suppose first that $v = \rho$, and let $v_1, \ldots, v_k$, $k \geq 2$ be the children of v. Observe that for every $x \in L(v_i)$ and $y \in L(v_j)$ with $i \neq j$, $\mathrm{lca}(x, y) = \rho$.

From our assumption on v, there is no path between x and y in G when $\sigma(x) \neq \sigma(y)$. This contradicts G being connected.

Suppose now that $v \neq \rho$. Denote by w be the parent of v, and by T' the tree obtained from T by contracting the arc wv. We show that $G = G'$, where G' is explained by (T', σ, u), contradicting the assumption that (T, σ, u) is least-resolved. Observe that $V(G) = V(G')$ by Definition 3.1(i). Hence, we proceed to show that $E(G) = E(G')$. For every $xy \in E(G)$, we may assume that $u(x) \neq x$, and $y \preceq_T \mathrm{lca}_T(x, z)$ for all $z \in V(G)$ with $\sigma(z) = \sigma(y)$ from Definition 3.1. Now, let z be such that $\sigma(z) = \sigma(y)$. Since $y \preceq_T \mathrm{lca}_T(x, z)$, contracting wv yields $y \preceq_{T'} \mathrm{lca}_{T'}(x, z)$. Thus, $xy \in E(G')$.

Conversely, suppose that $xy \notin E(G)$. We may assume that $\sigma(x) \neq \sigma(y)$, and $u(x) \neq x$ or $u(y) \neq y$. Indeed, in both cases we have $xy \notin E(G')$. Without loss of generality, assume that $u(x) \neq x$. G being connected together with $xy \notin E(G)$ yields that there exists $z \in V(G)$ with $\sigma(z) = \sigma(y)$, $xz \in E(G)$ and y is not a descendant of $\mathrm{lca}_T(x, z)$. Thus, $xy \in E(G')$ if and only if $\mathrm{lca}_T(x, z) = v$ and y is a descendant of w but not of v since T' is obtained by contraction wv. This contradicts the assumption on v. Hence $xy \notin E(G')$. □

As a direct consequence, we obtain the following result.

Corollary 3.4. *For every least-resolved tree (T, σ, u), and every internal vertex v of T, the subtree $T(v)$ rooted at v is not monochromatic.*

The converse of Corollary 3.4 does not hold in general. As illustrated in Figure 3, the non-monochromatic trees T and T' explain the same graph G. Because T' results from contracting an edge in T, the latter is not least-resolved.

We are now in a position to describe the set of neighbours of any vertex of an un2qBMG.

Proposition 3.5. *Let G be an un2qBMG explained by a least-resolved tree (T, σ, u). For every $xy \in E(G)$, one of the following cases holds:*

(i) the parent of x is an ancestor of y and $u(x) \neq x$;

(ii) the parent of y is an ancestor of x and $u(y) \neq y$.

Proof. Let $xy \in E(G)$. Then either (ii.1) or (ii.2) in Definition 3.1 is satisfied. Suppose first that Definition 3.1(ii.1) holds. Let v be the parent of x in

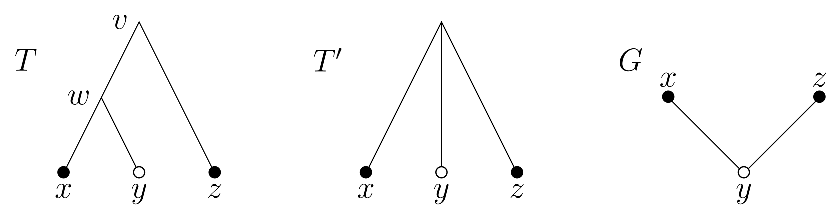

Figure 3: Example of non-monochromatic least-resolved tree. (T, σ, u) and (T, σ, u') explain G, where u and u' map every leaf to the root of T and T', respectively. Since T' can be obtained from T by contracting the arc vw, T is not least-resolved, although it is not monochromatic.

T. We proceed to show that v is an ancestor of y. Since $T(v)$ is non-monochromatic by Corollary 3.4, there exists $z \in L(T(v))$ such that $\sigma(z) \neq \sigma(x)$. In particular, we have $v = \operatorname{lca}(x, z)$, so the relation $y \preceq \operatorname{lca}(x, z)$ implies that y is a descendant of v. Hence, case (i) holds. Using similar arguments, Definition 3.1(ii.2) yields case (ii). □

4. Small un2qBMGs and bicliques

This section is devoted to the study of connected graphs with at most 5 vertices, as well as complete bipartite graphs. In both cases, the graphs turn out to be un2qBMGs.

Proposition 4.1. *Complete bipartite graphs are precisely the un2qBMGs explained by (T, σ, u) where T is a star-tree and $u(x) = \rho$ for all $x \in L(T)$.*

Proof. Let (G, σ) be a biclique, and let (T, σ, u) be a star-tree with leaf-coloring σ, $L(T) = V(G)$, and $u(x) = \rho$ for all $x \in L(T)$. Since (G, σ) is a biclique, we have that $xy \in E(G)$ if and only if $\sigma(x) \neq \sigma(y)$. Moreover, by definition of (T, σ, u), all pairs $x, y \in X$ with $\sigma(x) \neq \sigma(y)$ satisfy condition (ii.1) of Definition 3.1. Hence, (G, σ) is explained by (T, σ, u).
Conversely, suppose that (G, σ) is explained by (T, σ, u). If $\sigma(x) = \sigma(y)$, then by Definition 3.1(ii), $xy \notin E(G)$. Otherwise, i.e., if $\sigma(x) \neq \sigma(y)$, then x and y satisfy Definition 3.1(ii.1). Thus, G is a complete bipartite graph. □

Proposition 4.2. *Let (G, σ) be a connected bipartite graph and $1 < |V(G)| < 6$. G is un2qBMG explained by the tree (T, σ, u) whenever one of the following cases is satisfied.*

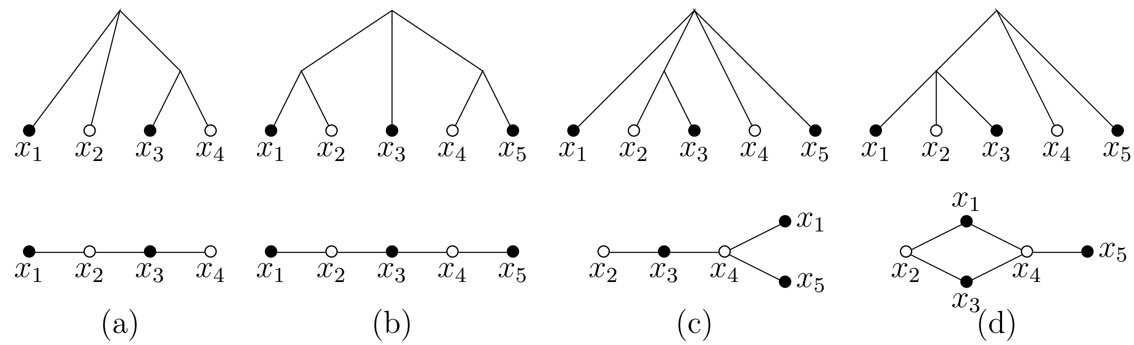

Figure 4: All connected non-biclique un2qBMGs with at most five vertices, together with the trees explaining them. Bipartitions σ are shown as vertex-colorings. The truncation maps u are: (a) $u(x_1) = x_1$ and $u(x_i) = \rho_T$ for $i \in \{2, 3, 4\}$; (b) $u(x_i) = \rho_T$ for $i \in \{1, \ldots, 5\}$; (c) $u(x_1) = x_1$, $u(x_5) = x_5$ and $u(x_i) = \rho_T$ for $i \in \{2, 3, 4\}$; (d) $u(x_5) = x_5$ and $u(x_i) = \rho_T$ for $i \in \{1, \ldots, 4\}$.

(I) $|V(G)| \in \{2, 3\}$ *or* $|V(G)| \in \{4, 5\}$ *and* G *is complete bipartite,* T *is a star-tree with* $L(T) = V(G)$ *and* $u(x) := \rho_T$ *for all* $x \in V(G)$*.*

(II) $|V(G)| = 4$ *and* $G = x_1x_2x_3x_4$, T *is the tree in Figure 4(a) and* $u(x_1) := x_1, u(x_i) := \rho_T$ *for* $2 \leq i \leq 4$*;*

(III) $|V(G)| = 5$ *and*

  (1) $E(G) = \{x_1x_2, x_2x_3, x_3x_4, x_4x_5\}$, T *is the tree in Figure 4(b) and* $u(x_i) := \rho_T$ *for* $1 \leq i \leq 5$*;*
  (2) $E(G) = \{x_1x_4, x_2x_3, x_3x_4, x_4x_5\}$, T *is the tree in Figure 4(c), and* $u(x_1) = x_1, u(x_5) = x_5, u(x_i) := \rho_T$ *for* $2 \leq i \leq 4$*;*
  (3) $E(G) = \{x_1x_2, x_1x_4, x_2x_3, x_3x_4, x_4x_5\}$, T *is the tree in Figure 4(d) (top), and* $u(x_5) := x_5, u(x_i) := \rho_T$ *for* $1 \leq i \leq 4$*.*

Proof.

(I) $|V(G)| \in \{4, 5\}$ and G is a biclique, or $|V(G)| \in \{2, 3\}$. Observe that G is also a biclique when $|V(G)| \in \{2, 3\}$ since G is connected. Proposition 4.1 ensures that (G, σ) is exactly the un2qBMG explained by the tree (T, σ, u) where T is a star-tree with root ρ_T and leaf set $L(T) = V(G)$, and $u(x) := \rho_T$ for all $x \in V(G)$.

(II) $V(G) = \{x_1, x_2, x_3, x_4\}$ and $E(G) = \{x_1x_2, x_2x_3, x_3x_4\}$; see Figure 4(a). Observe that setting T as the tree in Figure 4(a) and $u(x_1) := x_1, u(x_i) := \rho_T$ for $2 \leq i \leq 4$ gives a (T, σ, u) explaining G from Definition 3.1. Hence, G is an un2qBMG.

(III) $V(G) = \{x_1, x_2, x_3, x_4, x_5\}$. In each of the three subcases, it is straightforward to check that (T, σ, u) explains G by Definition 3.1. Hence G is an un2qBMG.

□

It is straightforward to verify that if G is a bipartite, connected graph with at most five vertices, its vertex-set and edge-set fall into one of the cases (I), (II), or (III) of Proposition 4.2. As a consequence, the following result is obtained.

Corollary 4.3. *Every connected bipartite graph with at most* 5 *vertices is an un2qBMG.*

An immediate consequence of Corollary 4.3 is that there are no forbidden induced subgraphs of a un2qBMG with at most 5 vertices. For $|V| = 6$, both P_6 and C_6 are bipartite but not un2qBMGs. In Section 6 we will provide a characterization of un2qBMGs in terms of forbidden induced subgraphs.

5. Recognition of un2qBMGs

In this section, we present a polynomial-time algorithm for the recognition of un2qBMGs. Consider a bipartite graph (G, σ) with a vertex-coloring σ induced by the bipartition of the vertex-set $V(G)$. The algorithm either builds a tree (T, σ, τ) explaining G, or returns the statement that G is not a un2qBMG. The following definition plays a key role in the algorithm.

Definition 5.1. *Let (G, σ) be a bipartite graph. A vertex x of G that is adjacent to all vertices of the opposite color of G is called a* heart-vertex*.*

We begin with the following result for connected un2qBMGs.

Lemma 5.1. *Every connected un2qBMG has a heart-vertex.*

Proof. Let (G, σ) be a connected un2qBMG explained by a least-resolved tree (T, σ, u). Proposition 3.3 implies that there exist $x, y \in V(G)$ such that $xy \in E(G)$ and $\rho_T = \mathrm{lca}_T(x, y)$. By Proposition 3.5, we may assume that $u(x) \neq x$ and that the parent v of x is an ancestor of y. Hence, $v = \mathrm{lca}(x, y) = \rho_T$, which implies that x is a child of ρ_T. Together with $u(x) \neq x$, this yields $xz \in E(G)$ for every $z \in L(T)$ with $\sigma(z) \neq \sigma(x)$; that is, x is a heart-vertex of G. $\square$

Note that a un2qBMG G needs not be connected. For example, the un2qBMG explained by the tree (T, σ, u) depicted in Figure 2 has two connected components. However, all connected components of G are un2qBMG by Lemma 3.2, and therefore, Lemma 5.1 implies that all connected components of G have a heart-vertex. Lemma 5.1 lies at the heart of our algorithm HEART-TREE, which we present now.

Algorithm 1 HEART-TREE

Input: A bipartite graph G.
Output: A tree (T, σ, u) explaining G if G is a un2qBMG. Otherwise, `false`.

Let σ be the vertex-coloring map induced by the bipartition of $V(G)$
$T := \rho$, $F(\rho) := V(G)$
if $|V(G)| = 1$ **then**
    $x := \rho$ and $u(x) := x$ with $V(G) = \{x\}$
else
    while $L(T) \setminus V(G) \neq \emptyset$ **do**
        Take $v \in L(T) \setminus V(G)$
        $G_v := G[F(v)]$ and $H_v := \{x \in F(v) : x \text{ is a heart-vertex of } G_v\}$
        if $H_v = \emptyset$ and G_v is connected **then** **return** `false`
        else
            for $x \in H_v$ **do**
                Add x as a child of v and $u(x) := \rho$
            $G_v^- := G_v[F(v) \setminus H_v]$
            for every connected components C of G_v^- **do**
                if $|V(C)| = 1$ **then**
                    Add the unique $x \in V(C)$ as a child to v and $u(x) := x$
                else
                    Add a new child w to v and $F(w) := V(C)$
return (T, σ, u)

Theorem 5.2. *Let (G, σ) be a bipartite graph. Algorithm HEART-TREE returns a tree (T, σ, u) if and only if G is a un2qBMG. Moreover, in that case, the tree (T, σ, u) returned by HEART-TREE explains G and is least-resolved.*

Proof. Suppose first that Algorithm 1 returns `false`. Then, there exists an instance of the loop initiated at Line 6 in which $H_v = \emptyset$ and G_v is connected (Line 9), where H_v and G_v are defined at Line 8. From Lemma 5.1, G_v is not a un2qBMG. Hence, G is not an un2qBMG, as G_v is an induced subgraph of G, and the property of being an un2qBMG is hereditary [12, Proposition 3.1].

Conversely, suppose that Algorithm 1 returns a tree (T, σ, u). Showing that G is a un2qBMG is equivalent to verifying that (T, σ, u) explains G by Proposition 3.1. Let G_T be the un2qBMG explained by (T, σ, u). We proceed to show that $G = G_T$ by verifying that (I) $V(G) = V(G_T)$ and (II) $E(G) = E(G_T)$.

(I) $V(G) = V(G_T)$. $V(G_T) = L(T)$ follows from Definition 3.1(i); therefore, it remains to show that $V(G) = L(T)$. From the condition on Line 6, $L(T) \subseteq V(G)$ must hold to exit the loop. We now show that $V(G) \subseteq L(T)$ also holds. First, we claim that for all internal vertices v such that $x \in F(v)$, either x is a child of v, or there exists exactly one child w of v such that $x \in F(w)$. To see the claim, let G_v, H_v and G_v^- be as defined on Lines 8 and 13. If $x \in H_v$, then x is a child of v (Line 12), and the claim is verified. Otherwise, x is a vertex of G_v^-, and there exists a (necessarily unique) connected component C of G_v^- containing x in its vertex-set. If $V(C) = \{x\}$, then x is a child of v (Line 16), and the claim is verified. Otherwise, v has a child w that satisfies $F(w) = V(C)$, and the claim is satisfied since $x \in V(C)$. The claim being true, and since $x \in F(\rho)$ (Line 2), it follows that there exists a (necessarily unique) leaf v of T such that either $u = x$ or $x \in F(u)$ holds. In view of the condition at Line 6, the latter is impossible, so we have $x \in L(T)$, and $V(G) \subseteq L(T)$ holds as desired. This concludes the proof of the claim that $V(G) = V(G_T)$ holds.

(II) $E(G) = E(G_T)$. Let $xy \in E(G_T)$ with $v := \text{lca}(x, y)$. Line 18 ensures $x, y \in F(v) \subseteq V(G_v)$. Note that $xy \in E(G_T)$ implies $\sigma(x) \neq \sigma(y)$. We show that at least one of x or y is an element of H_v. Assume that $x, y \notin H_v$. Hence, $x, y \in V(G_v^-)$ (Line 13). Let C_x and C_y be the connected components of G_v^- that contain x and y, respectively. If $C_x = C_y$, then there exists a child w of v such that $x, y \in F(w)$ (Line 18). In particular, w is an ancestor of both x and y in T, contradicting $v = \operatorname{lca}(x, y)$. Hence, $C_x \neq C_y$. Up to a permutation, we may assume that $u(x) \neq x$ from Definition 3.1(ii), as $xy \in E(G_T)$. Observe that $|C_x| > 2$. Indeed, if $|V(C_x)| = 1$, then x is the unique element of that set, and $u(x) = x$ (Lines 16), contradicting the assumption that $u(x) \neq x$. Connectedness of C_x implies that there exists $z \in V(C_x)$ such that $xz \in E(G)$ and $\sigma(z) \neq \sigma(x)$ (Line 1). Let w be the child of v such that $L(T(w)) = C_x$ (Line 18). Thus, $\operatorname{lca}(x, z) \preceq w \prec v = \operatorname{lca}(x, y)$, contradicting Definition 3.1(ii.1). Hence, at least one of x or y must be an element of H_v. From the definition of H_v (Line 8), $xy \in E(G_v)$. In addition, $xy \in E(G)$ since G_v is an induced subgraph of G (Line 8). Hence $E(G_T) \subseteq E(G)$.

To show that $E(G) \subseteq E(G_T)$, take $xy \in E(G)$ with $v := \operatorname{lca}(x, y)$. Line 1 ensures that $\sigma(x) \neq \sigma(y)$. From Lines 12, 16, and 18, we have $x, y \in F(v)$; thus x and y are vertices of G_v. Suppose $x, y \in V(G_v^-)$. Hence $xy \in E(G_v^-)$ since $xy \in E(G)$ and G_v^- is an induced subgraph of G (Line 8). Therefore, x and y belong to the same connected component C of G_v^-. In particular, there exists a child w of v such that $x, y \in F(w)$ (Line 18). This implies that x and y are both descendant of w, contradicting $v = \operatorname{lca}(x, y)$. Hence, at least one between x and y is an element of H_v. Without loss of generality, we may assume that $x \in H_v$. Then, x is a child of v, and we have $u(x) \neq x$ (Line 12). Take $z \in L(T)$ with $\sigma(z) = \sigma(y)$. $z \notin V(G_v)$ yields $v \prec \operatorname{lca}(x, z)$ (Line 18). On the other hand, $\operatorname{lca}(x, z) = v$ whenever $z \in V(G_v)$. Therefore, in both cases, $y \prec \operatorname{lca}(x, y) = v \preceq \operatorname{lca}(x, z)$. Hence, $xy \in E(G_T)$ from Definition 3.1(ii.1).

It remains to show that (T, σ, u) is least resolved. Let vw be an internal arc of T, and let T' be the tree obtained from T by contracting the arc vw. Let G' be the un2qBMG explained by (T', σ, u). We show that $G \neq G'$. Observe that $F(w)$ induces a connected subgraph of G of size 2 or more, as v is not a leaf of T. In particular, G_w has a heart-vertex x by Lemma 5.1. Moreover, there exists a vertex $y \in F(v)$ such that $\sigma(y) \neq \sigma(x)$ and $xy \notin E(G)$. Indeed, if no such vertex exists, then x is a heart-vertex of G_v, contradicting the choice of x as an element of $F(w) \subseteq F(v) \setminus H_v$. Define $\operatorname{lca}_{T'}(x, y) = \overline{v}$, where $\overline{v}$ is the vertex resulting from the contraction of the arc vw. Observe that $u(x) \neq x$

since x is a heart-vertex of G_w. Thus, $xy \in E(G')$ from Definition 3.1 (ii.1) but $xy \notin E(G)$. Hence, $G \neq G'$. $\square$

Proposition 5.3. *HEART-TREE algorithm runs in $O(n^3)$ where $n = V(G)$.*

Proof. In each iteration of the while-loop the algorithm processes a leaf v by creating the induced subgraph $G_v := G[F(v)]$ (Line 8), finding the set H_v of heart-vertices of G_v (Line 8), creating $G_v^- := G_v[F(v) \setminus H_v]$ (Line 13), and computing its connected components (Line 14). These steps can be achieved at once by scanning the adjacency lists of the vertices in $F(v)$ and traversing G_v with Breadth-First Search [2], hence in $O(|F(v)|+|E(G_v)|)$. Let $\mathcal{P}$ denote the (finite) set of elements v that are processed by the loop. The running time is

$$\sum_{v\in\mathcal{P}} O(|F(v)| + |E(G_v)|) \ \leq\ O\left(\sum_{v\in\mathcal{P}} |F(v)|\right) + O\left(\sum_{v\in\mathcal{P}} |E(G_v)|\right). \tag{1}$$

Recall that the set $\mathcal{P}$ of elements v processed by the loop is precisely the set $V(T) \setminus L(T)$, where T is the tree returned by HEART-TREE. On the other hand, T is a tree, so we have $|V(T)| \leq 2|L(T)| - 2 = O(|L(T)|) = O(n)$. This, together with $F(v) \subseteq V(G)$ (Line 18) and $E(G_v) \subseteq E(G)$ (Line 8) yields

$$\begin{aligned} \sum_{v\in\mathcal{P}} |F(v)| &\leq \sum_{v\in V(T)} |F(v)| \leq O(n^2) \\ \sum_{v\in\mathcal{P}} |E(G_v)| &\leq \sum_{v\in V(T)} |E(G_v)| \leq O(n^3). \end{aligned} \tag{2}$$

Combining Equations (1) and (2) yields a total running time of $O(n^3)$. $\square$

An immediate consequence of Proposition 5.3 together with Theorem 5.2 is the following result on the computational complexity of un2qBMGs.

Corollary 5.4. *For a bipartite graph G, it is decidable in cubic time whether G is an un2qBMG.*

Equipped with HEART-TREE algorithm, we arrive at the following characterization of un2qBMGs.

Theorem 5.5. *Let G be a bipartite graph. Then G is a un2qBMG if and only if all connected induced subgraphs of G contain a heart-vertex.*

Proof. Suppose G is a un2qBMG explained by an LRT (T, σ, u). Take a connected induced subgraph H of G. H is also an un2qBMG, as being an un2qBMG is hereditary [12, Proposition 3.1]. Hence H has a heart-vertex from Lemma 5.1. Conversely, suppose that G is not an un2qBMG. By Theorem 5.2, HEART-TREE returns `false` when applied to G. Hence, there exists an instance of the loop initiated at Line 6 in which $H_v = \emptyset$ and G_v is connected (Line 9, where H_v and G_v are defined at Line 8). In particular, G_v is a connected induced subgraph of G that does not contain a heart-vertex. □

6. Forbidden subgraphs characterization

In this section, we show that un2qBMGs are not properly contained in the family of (P_6, C_6)-free graphs, as they are characterized as those graphs free of induced P_6, C_6, and Sunlet_4 graphs.

Lemma 6.1. *Let (G, σ) be a connected bipartite graph. Let $x \in V(G)$ such that $\deg_G(x)$ is maximal among all vertices. If G is P_6-free, then for every $x' \in V(G)$ with $\sigma(x) = \sigma(x')$, there exists a vertex y with $\sigma(y) \neq \sigma(x)$ such that $xy, x'y \in E(G)$.*

Proof. Let $x' \in V(G)$ with $\sigma(x') = \sigma(x)$. Since G is connected, there exists a path P between x and x' in G. Without loss of generality, we may assume that P is a shortest path between x and x'. Since G is bipartite and $\sigma(x') = \sigma(x)$, P contains an odd number of vertices. Moreover, G is P_6-free, so $|V(P)| < 6$. Hence, we have $|V(P)| \in \{3, 5\}$, so either $P = xyx'$ for some vertex y with $\sigma(y) \neq \sigma(x)$, or $P = xy\tilde{x}\tilde{y}x'$ for some vertex $\tilde{x}$ with $\sigma(\tilde{x}) = \sigma(x)$ and vertices $y, \tilde{y}$ with $\sigma(\tilde{y}) = \sigma(y) \neq \sigma(x)$.

Suppose that the latter holds. In particular, $\tilde{y}$ is not adjacent to x. This, together with $\deg_G(x) \geq \deg_G(\tilde{x})$ implies the existence of a vertex y' such that $y'x \in E(G)$ but $y'\tilde{x} \notin E(G)$. On the other hand, $y'x' \notin E(G)$, because otherwise $xy'x'$ is a path between x and x' in G, contradicting the choice of P as a shortest path between x and x'. However, then $y'x\tilde{y}\tilde{x}yx'$ is an induced P_6 of G, contradicting the assumption that G is P_6-free. Hence, this case cannot occur.

Therefore, P must be of the form xyx'. In particular, y is adjacent to both x and x' in G, which concludes the proof. □

Proposition 6.2. *Let (G, σ) be a connected bipartite graph. If G contains no P_6, C_6, or Sunlet$_4$ as an induced subgraph, then G has a heart-vertex.*

Proof. Assume that G does not contain a heart-vertex.

Let $x_0 \in V(G)$ be such that $\deg_G(x_0)$ is maximal among all elements in $V(G)$. Since G has no heart-vertex, there exists a vertex y_0 with $\sigma(x_0) \neq \sigma(y_0)$ such that $x_0y_0 \notin E(G)$. Let P be a shortest path between x_0 and y_0 in G. Note that since G is bipartite, $|V(P)|$ is even. Moreover, G does not contain an induced P_6, so $|V(P)| < 6$ must hold. Since $|V(P)| \neq 2$ by the choice of y_0, this implies that only $|V(P)| = 4$ is possible. Hence, $P = x_0y_1x_1y_0$ for some vertices x_1, y_1 with $\sigma(x_1) = \sigma(x_0)$ and $\sigma(y_1) = \sigma(y_0)$.

Recall that, by the choice of x_0, $\deg(x_0) \geq \deg(x_1)$. On the other hand, $x_1y_1, x_0y_1, x_1y_0 \in E(G)$ but $x_0y_0 \notin E(G)$. This implies that there exists $y_2 \in V(G)$ such that y_2 is adjacent to x_0 and not to x_1. In particular, y_2 is distinct from y_0 and y_1, and we have $\sigma(y_2) = \sigma(y_0)$. Hence, $y_2x_0y_1x_1y_0$ is an induced P_5 in G, see Figure 5 (a).

Since G contains no heart-vertex, there is a vertex x_2 with $\sigma(x_2) = \sigma(x_0)$ such that $x_2y_1 \notin E(G)$. In particular, x_2 is distinct from x_0 and x_1. From the maximality of the degree of x_0, Lemma 6.1 applies to x_0 and x_2, and there exists y_3 with $\sigma(y_3) \neq \sigma(x_0)$ such that $x_0y_3, x_2y_3 \in E(G)$. In particular, $y_3 \neq y_1$ and $y_3 \neq y_0$ since $x_2y_1 \notin E(G)$ and $x_0y_0 \notin E(G)$, respectively. Without loss of generality, we may now assume that y_3 has maximal degree among all vertices of G that are adjacent to both x_0 and x_2.

If $y_3x_1 \notin E(G)$, then $\{x_2, y_3, x_0, y_1, x_1, y_0\}$ induces either a P_6 or C_6 in G depending on whether x_2 and y_0 are adjacent or not. This contradicts our hypothesis. Hence, y_3 must be adjacent to x_1, and thus $y_3 \neq y_2$. The same argument with y_3 replacing y_2 yields $x_2y_2 \notin E(G)$. Moreover, $x_2y_0 \notin E(G)$, as otherwise, $y_2x_0y_1x_1y_0x_2$ is a P_6 in G. See Figure 5 (b) for the subgraph of G induced by $\{x_0, x_1, x_2, y_0, y_1, y_2, y_3\}$.

Since G contains no heart-vertex, there exists a vertex x_3 such that $x_3y_3 \notin E(G)$. In particular, x_3 must be distinct from x_0, x_1 and x_2. If x_3 is adjacent to y_0 and not to y_2, then $y_2x_0y_3x_1y_0x_3$ is a P_6. If x_3 is adjacent to y_2 and not to y_0, then $x_3y_2x_0y_3x_1y_0$ is a P_6. Finally, if x_3 is adjacent to both y_0 and y_2, then the set $\{y_2, x_0, y_3, x_1, y_0, x_3\}$ induces a C_6 in G. In summary, x_3 is adjacent to neither y_0 nor y_2. If x_3 is adjacent to y_1, then $\{x_0, x_1, x_2, x_3, y_0, y_1, y_2, y_3\}$ induces a Sunlet$_4$ in G. Therefore x_3 is not adjacent to y_1.

From Lemma 6.1 and the maximality of the degree of x_0, there exists a vertex y_4 such that $\sigma(y_4) \neq \sigma(x_3)$ and $x_0y_4, x_3y_4 \in E(G)$. In view of the

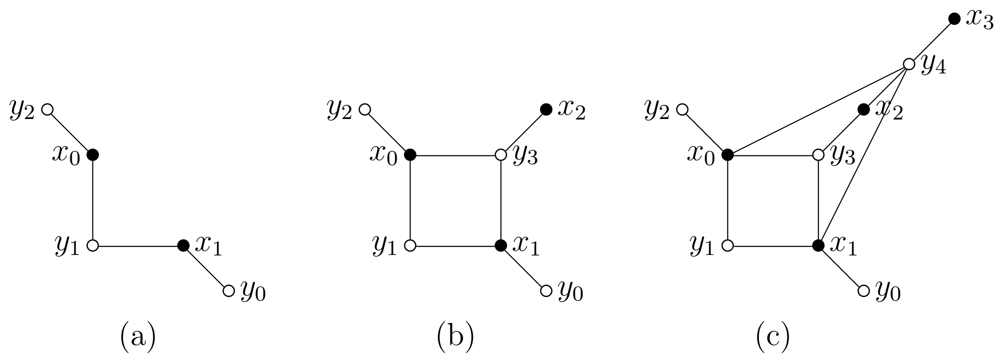

Figure 5: Three intermediate steps of the proof of Proposition 6.2 - see text for details. Note that in all stages, the depicted graph is an induced subgraph of G.

preceding arguments, y_4 is distinct from y_0, y_1, y_2, and y_3. We show that $y_4x_1, y_4x_2 \in E(G)$. Indeed, if $x_1y_4 \notin E(G)$, then $x_3y_4x_0y_1x_1y_0$ is a P_6 of G. Hence, $x_1y_4 \in E(G)$. Moreover, if $x_2y_4 \notin E(G)$, then $\{x_0, y_2, y_4, x_3, x_1, y_0, y_3, x_2\}$ induces a Sunlet$_4$ in G. Hence, $x_2y_4 \in E(G)$. Figure 5 (c) illustrates the subgraph of G induced by $\{x_0, x_1, x_2, x_3, y_0, y_1, y_2, y_3, y_4\}$.

Next, recall that by choice of y_3, $\deg_G(y_3) \geq \deg_G(y_4)$ must hold. Since $x_3y_4 \in E(G)$ and $x_3y_2 \notin E(G)$, this means that there exists $x_4 \in G$ with $\sigma(x_4) \neq \sigma(y_2)$ such that $x_4y_2 \in E(G)$ and $x_4y_4 \notin E(G)$. In particular, x_4 is distinct from x_0, x_1, x_2 and x_3, since all these vertices are adjacent to y_4 in G.

Now, if $x_4y_0 \in E(G)$, then $x_3y_4x_2y_3x_4y_0$ is an induced P_6 of G. Similarly, if $x_4y_2 \in E(G)$, then $x_3y_4x_2y_3x_4y_2$ is an induced P_6 of G. So, x_4 is adjacent in G to neither y_0 nor y_2. The set $\{x_0, y_2, y_3, x_4, x_1, y_0, y_4, x_3\}$ therefore induces a Sunlet$_4$ in G. Since the assumptions on G exclude all these cases, G must contain a heart-vertex. □

The latter result suggests a relationship with the class of *bi-cographs*(bi-complement reducible graphs) [6], which is characterized by the three forbidden subgraphs: P_7, the star tree $S_{1,2,3}$ with branches of length 1,2, and 3, and Sunlet$_4$. Together with Theorem 5.5, we obtain the following characterization result, which partially resembles the characterization result [20, Proposition 1] for chordal bipartite graphs. 2qBMGs as the $(P6, C_6)$-free bi-cographs.

Theorem 6.3. *Let G be a graph. The following properties are equivalent.*

(i) *G is a un2qBMG.*

(ii) *G is bipartite and all connected induced subgraphs of G contain a heart-vertex.*

(iii) *G is bipartite and* $(P_6, C_6, Sunlet_4)$*-free.*

(iv) *G is* $(C_3, C_5, C_6, P_6, Sunlet_4)$*-free.*

(v) *G is chordal bipartite and* $(P_6, Sunlet_4)$*-free.*

(vi) *G is a* (P_6, C_6)*-free bi-cograph.*

Proof. The equivalence between (i) and (ii) is stated in Theorem 5.5. [20, Proposition 1] yields the equivalence between (iii), (iv), and (v), which is straightforward from the definition of chordal bipartite graphs and the characterization of bipartite graphs as free of induced odd-cycles.

We now proceed to show that (ii) and (iii) are equivalent. Assume first that (ii) holds. Then G is $(P_6, C_6, \mathrm{Sunlet}_4)$-free, because P_6, C_6 and Sunlet_4 contain no heart-vertex. Hence, (iii) is satisfied. Conversely, assume that (iii) holds. Let H be a connected induced subgraph of G. Clearly, H is bipartite and $(P_6, C_6, \mathrm{Sunlet}_4)$-free. Applying Proposition 6.2 to H implies that H has a heart-vertex. Hence, (ii) is satisfied.

Finally, we established the equivalence between (iii) and (vi). Assume first that (iii) holds. Recall that bi-cographs contain no P_7, $S_{1,2,3}$ and Sunlet_4 as induced subgraphs. Hence, (vi) is satisfied, as P_7 and $S_{1,2,3}$ contain an induced P_6. Conversely, assume that (vi) is true. Clearly, (iii) is verified as bi-cographs are Sunlet_4-free. □

Given that both bi-cographs [7] and (P_6, C_6)-free graphs [13] admit linear-time recognition algorithms, Theorem 6.3 thus implies the following result.

Corollary 6.4. *un2qBMGs can be recognized in linear time.*

7. Concluding Remarks

In summary, we have provided a comprehensive characterization of undirected 2-quasi-best match graphs (un2qBMGs). Specifically, we established (i) a structural characterization of un2qBMGs by means of labeled trees (T, σ, u); (ii) a cubic-time recognition algorithm that relies on the detection of heart-vertices; (iii) an equivalent characterization as the bipartite graphs

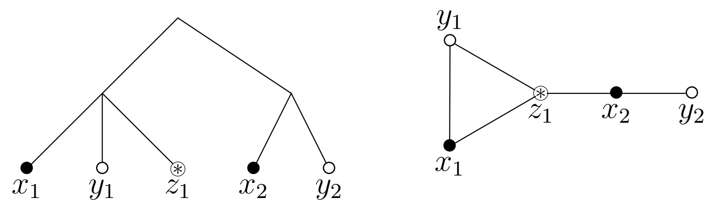

Figure 6: Undirected underlying graph of qBMG on three colors, which has no heart-vertices.

in which all connected induced subgraphs have a heart-vertex; and (iv) a forbidden subgraph characterization description as the $(P_6, C_6, \text{Sunlet}_4)$-free bipartite graphs. Furthermore, we showed that un2qBMGs coincide with the (P_6, C_6)-free bi-cographs, which implies the existence of a linear-time recognition algorithm for this class. Although Algorithm 1 requires $O(|V(G)|^3)$ time, it is nevertheless interesting in practice because it explicitly constructs for each un2qBMG a least-resolved tree (T, σ, u).

An interesting direction for future work is to explore whether the linear decomposition-based recognition algorithms for bi-cographs and (P_6, C_6)-free graphs can be adapted, following the approach of [13], to recognize un2qBMGs directly. Since this approach is based on building an (inner node-labeled) decomposition tree, it would also be interesting to investigate if the resulting decomposition tree can be used to construct the least-resolved explaining the tree more efficiently. Such an approach has been successfully used to build the phylogenetic tree explaining a Fitch graph from di-cographs in linear time in [5].

The family of un2qBMGs properly contains several known graph classes. For example, the bi-quasi-threshold graphs coincide with the domino-free un2qBMGs [1, Theorem 1]. Moreover, according to the online database `https://www.graphclasses.org`, the bipartite bi-threshold graphs [8] form a proper subclass of the un2qBMGs.

Clique-width is an important parameter in algorithmic graph theory since many problems that are NP-hard in general, such a graph coloring, become tractable when restricted to graphs of bounded clique-width [1]. Since bi-cographs are known to have bounded clique-width [1], un2qBMGs also have bounded clique-width as an immediate consequence of Theorem 6.3(vi). An interesting for future work, thus, is to explore the computational complexity of problems that remain NP-hard on bi-cographs and (P_6, C_6)-free graphs, but may become tractable by polynomial-time algorithms when restricted to the class of un2qBMGs.

Although our results provide a comprehensive understanding of undirected 2-quasi-best match graphs; they do not directly extend to cases with more than two colors. In particular, the example in Figure 6 shows that there is no natural extension of the notion of a heart-vertex to multipartite graphs such that Lemma 5.1 holds for three colors. Consequently, a generalization of Theorem 5.5 to three or more colors appears to be unlikely. Nevertheless, our results yield an easily testable necessary condition for more than two colors, as every subgraph of an underlying induced subgraph of a qBMG (unqBMG) induced by a pair of color classes must be a un2qBMG. Since the least-resolved explaining trees of a qBMG [11, Figure 4], and thus also of an unqBMG, are not necessarily unique, it remains an open problem for future research whether, and if so, how explaining trees for the un2qBMGs of pairs of color classes can be combined to a labeled tree that explains an unqBMG. The recognition problem for unqBMGs with three or more colors, therefore, remains open.

Acknowledgements

This work was supported in part by the German Research Foundation (DFG, grant # 432974470). Research in the Stadler lab is supported by the German Federal Ministry of Education and Research (BMBF) through the DAAD project 57616814 (SECAI, School of Embedded Composite AI) and, jointly with Tourismus, within the program *Center of Excellence for AI Research "Center for Scalable Data Analytics and Artificial Intelligence Dresden/Leipzig"* (ScaDS.AI).

References

[1] Bogdan Alecu, Mamadou Moustapha Kanté, Vadim Lozin, and Viktor Zamaraev. Between clique-width and linear clique-width of bipartite graphs. *Discrete Mathematics*, 343(8):111926, 2020.

[2] Thomas H. Cormen, Charles E. Leiserson, Ronald L. Rivest, and Clifford Stein. *Introduction to Algorithms.* MIT Press, Cambridge, MA, 3 edition, 2009.

[3] Jean-Luc Fouquet, Vassilis Giakoumakis, and Jean-Marie Vanherpe. Bipartite graphs totally decomposable by canonical decomposition. *International Journal of Foundations of Computer Science*, 10(04):513–533, 1999.

[4] Manuela Geiß, Edgar Chávez, Marcos González Laffitte, Alitzel López Sánchez, Bärbel MR Stadler, Dulce I Valdivia, Marc Hellmuth, Maribel Hernández Rosales, and Peter F Stadler. Best match graphs. *Journal of mathematical biology*, 78(7):2015–2057, 2019.

[5] Manuela Geiß, John Anders, Peter F. Stadler, Nicolas Wieseke, and Marc Hellmuth. Reconstructing gene trees from fitch's xenology relation. *Journal of Mathematical Biology*, 77(5):1459–1491, 2018.

[6] Vassilis Giakoumakis and Jean Marie Vanherpe. Bicomplement reducible graphs. *Advances Applied Mathematics*, 18:389–402, 1997.

[7] Vassilis Giakoumakis and Jean Marie Vanherpe. Linear time recognition and optimizations for weak-bisplit graphs, bi-cographs and bipartite P_6-free graphs. *Intl. J. Foundations Comp. Sci.*, 14:107–136, 2003.

[8] Peter L. Hammer, Mahadev N. V. R., and Uri N. Peled Peled. Bipartite bithreshold graphs. *Discrete Mathematics*, 119(1):79–96, 1993.

[9] Marc Hellmuth, Manuela Geiß, and Peter F Stadler. Complexity of modification problems for reciprocal best match graphs. *Theoretical Computer Science*, 809, 2020.

[10] Annachiara Korchmaros. Automorphisms and quotients of 2-colored quasi best match graphs. *Art of Discrete and Applied Mathematics*, 9:1.06, 2026.

[11] Annachiara Korchmaros, David Schaller, Marc Hellmuth, and Peter F Stadler. Quasi-best match graphs. *Discrete Applied Mathematics*, 331:104–125, 2023.

[12] Annachiara Korchmaros and Peter F Stadler. Forbidden configurations and dominating bicliques in undirected 2-quasi best match graphs. *arXiv preprint arXiv:2502.11791*, 2025.

[13] Ruzayn Quaddoura and Ahmad Al-Qerem. Bipartite (P_6,C_6)-free graphs: Recognition and optimization problems. *Symmetry*, 16(4):447, 2024.

[14] José Antonio Ramírez-Rafael, Annachiara Korchmaros, Katia Aviña-Padilla, Alitzel López Sánchez, Andrea Arlette España-Tinajero, Marc Hellmuth, Peter F Stadler, and Maribel Hernández-Rosales. Revolutionh-tl: Reconstruction of evolutionary histories tool. In *RECOMB International Workshop on Comparative Genomics*, pages 89–109. Springer, 2024.

[15] José Antonio Ramírez-Rafael, Annachiara Korchmaros, Katia Aviña-Padilla, Alitzel López-Sánchez, Gabriel Martinez-Medina, Alfredo J Hernández-Álvarez, Marc Hellmuth, Peter F Stadler, and Maribel Hernández-Rosales. Revolutionh-tl: A fast and robust tool for decoding evolutionary gene histories. *bioRxiv*, pages 2025–04, 2025.

[16] David Schaller, Manuela Geiß, Edgar Chávez, Marcos González Laffitte, Alitzel López Sánchez, Bärbel MR Stadler, Dulce I Valdivia, Marc Hellmuth, Maribel Hernández Rosales, and Peter F Stadler. Corrigendum to "best match graphs". *Journal of Mathematical Biology*, 82(6):47, 2021.

[17] David Schaller, Manuela Geiß, Marc Hellmuth, and Peter F Stadler. Best match graphs with binary trees. In *International Conference on Algorithms for Computational Biology*, pages 82–93. Springer, 2021.

[18] David Schaller, Manuela Geiß, Marc Hellmuth, and Peter F Stadler. Heuristic algorithms for best match graph editing. *Algorithms for Molecular Biology*, 16(1):19, 2021.

[19] David Schaller, Peter F Stadler, and Marc Hellmuth. Complexity of modification problems for best match graphs. *Theoretical Computer Science*, 865, 2021.

[20] Asahi Takaoka. Decomposition of P_6-free chordal bipartite graphs. *IEICE Trans. Fund. Electronics, Commun., Comp. Sci.*, E106.A(11):1436–1439, 2023.