Forbidden Paths and Cycles in the Undirected Underlying Graph of a 2-quasi Best Match Graph

Annachiara Korchmaros

Bioinformatics Group, Department of Computer Science & Interdisciplinary Center for Bioinformatics, Universität Leipzig, Härtelstraße 16–18, D-04107 Leipzig, Germany.

Abstract

The undirected underlying graph of a 2-quasi best match graph (2-qBMG) is proven not to contain any induced graph isomorphic to P_6 or C_6. This new feature allows for the investigation of 2-BMGs further by exploiting the numerous known results on P_6 and C_6 free graphs together with the available polynomial algorithms developed for their studies. In this direction, there are also some new contributions about dominating bicliques and certain vertex decompositions of the undirected underlying graph of a 2-qBMG.

Keywords

Phylogenetic combinatorics, quasi best match graphs, P_6- and C_6-free graphs, dominating set, vertex decomposition.

1. Introduction

Quasi-best match graphs are directed vertex-colored graphs explained by phylogenetic trees. A tree-free characterization is known so far for the case of two colors [1]: A *two-color quasi best match graph* (2-qBMG) is a bipartite directed graph $\vec{G}$ without loops and parallel edges which has the following three properties:

(N1) if u and v are two independent vertices then there exist no vertices w, t such that ut, vw, tw are edges;

(N2) bi-transitive, i.e., if uv, vw, wt are edges then ut is also an edge;

(N3) if u and v have common out-neighbor then either all out-neighbors of u are also out-neighbors of v or all out-neighbors of v are also out-neighbors of u.

A 2-qBMG is a *two-colored best match graph* (2-BMG) if it is sink-free, i.e., no vertex has empty out-neighborhood, and it is a *reciprocal two-best match graph* (reciprocal 2-BMG) if all its edges are symmetric. Moreover, 2-qBMGs but not 2-BMGs form a hereditary class [1].

2-BMGs, 2-qBMGs, and more generally best match graphs have been the subjects of intensive studies, also motivated by their relevant roles in the current investigation of orthology predictions from sequence similarity in the case that the gene families histories are free from gene transfers; see [2, 3, 4, 1, 5, 6, 7, 8]. A major contribution [6, Theorem 3.4] is a characterization of 2-qBMGs

CEUR Workshop Proceedings (CEUR-WS.org)

by three types of forbidden induced subgraphs, called of types $(F_1), (F_2)$ and (F_3); see Section 2.

This investigation focuses on specific properties of the underlying undirected graph of a 2-qBMG. Actually, such undirected graphs have been touched so far only marginally in the study of 2-qBMGs except in the very special case of reciprocal 2-BMGs; see [9, 10]. The main motivation for this study is the richness of the literature on undirected graphs. It is indeed much richer than on directed graphs, with plenty of fundamental works on forbidden configurations, decompositions, and vertex-colorings, as well as on algorithms and valuations on computational complexity. The aim is to exploit some of these deeper results on undirected graphs to gain new insights into 2-qBMGs. It may happen, however, that going back to a digraph $\vec{G}$ from its underlying undirected graph G, a deep result on G only has a very limited, or even trivial impact on $\vec{G}$. Thus, the structural relationship between $\vec{G}$ and G is far to be immediate. For instance, [6, Theorem 3.4] does not yield an analog characterization in terms of forbidden induced subgraphs of the underlying undirected graph. This depends on the fact that the underlying undirected graph of a digraph of type (F_i), $1 \leq i \leq 3$ with only required edges is either a path of length 4 or 5 but it is not necessarily a forbidden subgraph of the underlying undirected graph of a 2-qBMG.

Therefore, the challenge is to find appropriate results and knowledge on undirected graphs that may produce relevant contributions to the study of 2-qBMGs via their underlying undirected graphs.

The main results in this direction are Theorem 3.1 and Theorem 3.9, which state that the underlying undirected graph of every 2-qBMG is P_6-free and C_6-free, that is, both P_6 and C_6 are forbidden induced subgraphs. This result is sharp as the underlying undirected graph of a 2-qBMG may have an induced path P_5 (and hence induced

P_4 as well); see Theorem 3.2. Therefore, the underlying undirected graphs of a type (F_i) digraph, $1 \leq i \leq 3$, are not forbidden subgraphs for the underlying undirected graph of 2-qBMGs.

In their seminal paper [11] appeared in 1993, Bacsó and Tuza pointed out the strong relationship between P_k-freeness and dominating subsets in undirected graphs. Relevant contributions on such a relationship were given in several papers, especially in [12, 13, 14, 15, 16]. In particular, for any bipartite graph G, Liu and Zhou proved that G is both P_6-free and C_6-free if and only if every connected induced subgraph of G has a dominating biclique; see [15, Theorem 1]. An independent constructive proof of the Liu-Zhou theorem is due to P. van't Hof and D. Paulusma [14] where the authors also provide an algorithm that finds such a dominating biclique of a connected P_6-free graph in polynomial time. For a discussion on the computational complexity of the problem of finding (not necessarily dominating) bicliques in undirected bipartite graphs, see [17]. An upper bound on the number of bicliques of C_6-free undirected bipartite graphs in terms of the sizes of their color classes U, W is given in [18, Theorem 3.3] where it is shown that such number does not exceed $|U|^2 |W|^2$.

Recognition and optimization problems for both P_6 and C_6-free bipartite undirected graphs and certain decompositions involving $K \oplus S$ graphs have been studied by several authors, following the paper [19]. In particular, it is shown that the class of both P_6 and C_6-free bipartite graphs can be recognized in linear time. Also, efficient solutions for two NP-hard problems are presented in this class of graphs: the maximum balanced biclique problem and the maximum independent set problem. For more details, see the recent paper [20].

A contribution to the study of the vertex decomposition problem for 2-qBMGs in smaller connected 2-qBMGs of type (A) is given in Theorem 4.2, where a connected 2-qBMG is of type (A) if its underlying undirected graph is a $K \oplus S$ graph, that is, it has a vertex decomposition into a (dominating) biclique K and a stable set S, that is, any two vertices in S are independent.

2. Background

The notation and terminology for undirected graphs are standard. Let G be an undirected graph with vertex-set $V(G)$ and edge-set $E(G)$. In particular, a path P_n of an undirected graph G is a sequence $v_1 v_2 \cdots v_n$ of pairwise distinct vertices of G such that $v_i v_{i+1} \in E(G)$ for $i = 1, \dots, n-1$. A P_n path-graph is an undirected graph on n vertices with a path $v_1 v_2 \cdots v_n$ containing no chord, i.e. any edge in $E(G)$ is $v_i v_{i+1} \in E(G)$ for some $1 \leq i < n$. A P_n path-graph with a path $v_1 v_2 \cdots v_n$ containing no chord is a bipartite graph with (uniquely determined) color sets $\{v_1, v_3, \dots\}$ and $\{v_2, v_4, \dots\}$. An undirected graph is P_n-free if it has no induced subgraph isomorphic to a P_n path-graph. A cograph is a P_4-free graph.

A cycle C_n of an undirected graph G is a sequence $v_1 v_2 \cdots v_n v_1$ of pairwise distinct vertices of G such that $v_i v_{i+1} \in E(G)$ for $i = 1, \dots, n-1$, and $v_n v_1 \in E(G)$. A C_n cycle-graph is an undirected graph on n vertices with a cycle $v_1 v_2 \cdots v_n v_1$ containing no chord, i.e. any edge in $E(G)$ is either $v_i v_{i+1} \in E(G)$ for some $1 \leq i \leq n - 1$, or $v_n v_1$. For n even, a C_n cycle-graph with a cycle $v_1 v_2 \cdots v_n v_1$ containing no chord is a bipartite graph with (uniquely determined) color classes $\{v_1, v_3, \dots\}$ and $\{v_2, v_4, \dots\}$. 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.

A set $D \subseteq V(G)$ is a dominating set of an undirected graph G if, for any vertex $u \in V(G) \setminus D$, there is a vertex $v \in D$ such that $uv \in E(G)$. We also say that D dominates G. A subgraph H of G is a dominating subgraph of G if the vertex set of H dominates G.

An undirected graph G is of type $K \oplus S$ if either G is degenerate, i.e. if it has an isolated vertex, or there is a partition of $V(G)$ into two sets: a biclique set K and a stable set S; see [19, 20].

For directed graphs $\vec{G}$ with vertex-set $V(\vec{G})$ and edge-set $E(\vec{G})$, in particular for 2-qBMGs, notation and terminology come from [1]. In particular, $N^+(v)$ and $N^-(v)$ stand for the set of out-neighbours and in-neighbours of v in $\vec{G}$, and v is a sink when $N^+(v) = \varnothing$, and v is a source when $N^-(v) = \varnothing$. A digraph $\vec{G}$ is oriented if $uv \in E(\vec{G})$ implies $vu \notin E(\vec{G})$ for any $u, v \in V(\vec{G})$. An oriented digraph has a topological vertex ordering if its vertices can be labeled with $v_1, v_2, \dots$ such that for any $v_i v_j \in E(\vec{G})$ we have $i < j$. A sufficient condition for an oriented digraph to have a topological ordering is to be acyclic, that is, there is no directed cycle in the digraph. An orientation of a digraph $\vec{G}$ is a digraph obtained from $\vec{G}$ by keeping the same vertex set but retaining exactly one edge from each symmetric edge. From [4, Lemma 2.2] and [4, Theorem 3.8], any orientation of a 2-qBMG is acyclic whenever at least one of the following conditions are satisfied:

(∗) no two (or more than two) symmetric edges of $\vec{G}$ have a common endpoint;

(∗∗) no two (or more than two) vertices of $\vec{G}$ are equivalent, i.e. no two vertices have the same in- and out-neighbors.

Acyclic-oriented digraphs are odd–even graphs. For a pair $(\mathfrak{A}, \mathfrak{D})$ where $\mathfrak{A}$ is a finite set of non-negative even integers and $\mathfrak{D}$ is a set of positive odd integers, the associated odd–even oriented digraph $\vec{G}$ has vertex-set $\mathfrak{A}$ and edge-set with $ab \in E(\vec{G})$ when both $\frac{1}{2}(a + b)$ and $\frac{1}{2}(b - a)$ belong to $\mathfrak{D}$. This oriented bipartite graph has

color classes $U = \{a \;:\; a \equiv 0 \pmod 4, a \in \mathfrak{A}\}$ and $W = \{a \;:\; a \equiv 2 \pmod 4, a \in \mathfrak{A}\}$. An oriented bipartite digraph is a *bitournament* if for any two vertices $u, v \in V$ with different colors, either $uv \in E(\vec{G})$, or $uv \in E(\vec{G})$. A bi-transitive bitournament is an odd-even digraph; see [4, Proposition 3.9].

In addition, for four vertices x_1, x_2, x_3, y of $\vec{G}$, we say that $[x_1, x_2, x_3, y]$ is an (N1)-configuration if $x_1 x_2, x_2 x_3, yx_3 \in E(\vec{G})$ but either $x_1 y \in E(\vec{G})$ or $yx_1 \in E(\vec{G})$; in other words when condition (N1) holds for $u = x_1, t = x_2, w = x_3, v = y$ and therefore x_1 and y are not independent.

A 2-qBMG is *degenerate* if it has an isolated vertex. We stress that a non-degenerate 2-qBMG may be *trivial*, as it may be the union of pairwise disjoint (possible symmetric) edges. If this is the case, then (N1), (N2), and (N3) trivially hold in the sense that none of the conditions required in (N1), (N2), and (N3) is satisfied. Also, a directed graph is said to be P_n-free or C_n-free if its undirected underlying graph is P_n-free or C_n-free, respectively.

Let $\vec{G_4}$ denote a bipartite digraph on four vertices where $V(\vec{G_4}) = \{x_1, x_2, y_1, y_2\}$ and the color classes are $\{x_1, x_2\}$ and $\{y_1, y_2\}$. Then $\vec{G_4}$ is of *type (F_1) with required edges* if $E(\vec{G_4}) = \{x_1 y_1, y_2 x_2, y_1 x_2\}$. whereas $\vec{G_4}$ is of *type (F_2) with required edges* if $E(\vec{G_4}) = \{x_1 y_1, y_1 x_2, x_2 y_2\}$. The undirected underlying graph G_4 of a $\vec{G_4}$ of type either (F_1) or (F_2) is a path of length 4.

Let $\vec{G_5}$ denote a bipartite graph on five vertices where $V_5(\vec{G_5}) = \{x_1, x_2, y_1, y_2, y_3\}$ and the color classes are $\{x_1, x_2\}$ and $\{y_1, y_2, y_3\}$. Then $\vec{G_5}$ is of *type (F_3)* if *with required edges* $E(\vec{G_5}) = \{x_1 y_1, x_2 y_2, x_1 y_3, x_2 y_3\}$. The undirected underlying graph G_5 of $\vec{G_5}$ of type (F_3) is a path of length 5.

Let $\vec{G_1}$ and $\vec{G_2}$ be two directed graphs on the same vertex set V. Then $\vec{G_1} \cong \vec{G_2}$, that is, $\vec{G_1}$ and $\vec{G_2}$ are isomorphic, if there exists a edge-preserving permutation φ on V, i.e. $v_1 v_2 \in E(\vec{G_1})$ if and only if $\varphi(v_1)\varphi(v_2) \in E(\vec{G_2})$.

A tree T is *phylogenetic* if every node is either a leaf or has at least two children. T is a *rooted* tree if one of the nodes is chosen as *root* denoted by ρ_T. In a rooted tree, its root is typically drawn as the top node, and the edges are directed from the parent nodes to their child nodes. In this contribution, all trees are rooted and phylogenetic. Let (T, σ) be a leaf-colored tree with leaf set L, set of colors S, leaf-coloring surjective map $\sigma : L \to S$, and rooted at ρ_T. For any two leaves $x, y \in V$, $lca(x, y)$ denotes the last common ancestry between x and y on T. T is *phylogentic* if all nodes have at least two children except the leaves. A leaf $y \in L$ is a best match of the leaf $x \in L$ if $\sigma(x) \neq \sigma(y)$ and $lca(x, y) \preceq lca(x, z)$, i.e. $lca(x, z)$ is an ancestor of $lca(x, y)$, holds for all leaves z of color $\sigma(y) = \sigma(z)$. The *BMG (best match graph) explained by* (T, σ) is the directed graph whose vertices are the leaves of T where $xy \in E(\vec{G})$ if y is a best match

of x is a vertex-colored digraph with color set S.

A truncation map $u_T : L \times S \to T$ assigns to every leaf $x \in L$ and color $\sigma \in S$ a vertex of T such that $u_T(x, s)$ lies along the unique path from ρ_T to x and that $u_T(x, \sigma(x)) = x$. A leaf $y \in L$ with color $\sigma(y)$ is a quasi-best match for $x \in L$ (with respect to (T, σ) and u_T) if both conditions (i) and (ii) are satisfied: (i) y is a best match of x in (T, σ), (ii) $lca_T(x, y) \preceq u_T(x, \sigma(y))$. The digraph $qBMG(\sigma, u_T)$ is the vertex-colored digraph $qBMG(T, \sigma, u_T)$ on the vertex set L whose edges are defined by the quasi-best matches. A vertex-colored digraph $(\vec{G}, \sigma)$ with vertex set L 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 $(\vec{G}, \sigma) = qBMG(T, \sigma, u_T)$. For general results on quasi-best match graphs, the reader is referred to [1].

3. Forbidden induced path graphs of underlying undirected 2-qBMGs

The main result in the paper which establishes a new property of 2-qBMGs is given in the following theorem.

Theorem 3.1. *The underlying undirected graph of a 2-qBMG is P_6-free.*

Proof. Let $\vec{G}$ denote a 2-qBMG with at least six vertices. Assume on the contrary that its underlying undirected graph G has an induced subgraph G_6 on six vertices $v_1, v_2, v_3, v_4, v_5, v_6$ such that $v_1 v_2 v_3 v_4 v_5 v_6$ is a P_6 path-graph: Then G_6 contains no chord other than $v_i v_{i+1}$ for $i = 1, \dots, 5$. Four cases arise according to the possible patterns of the neighborhood of v_2 in $\vec{G}$.

(i): $v_1 v_2, v_3 v_2 \in E(\vec{G})$. Then $v_3 v_4 \in E(\vec{G})$, otherwise $[v_4, v_3, v_2, v_1]$ is an (N1)-quadruple. If $v_4 v_5 \in E(\vec{G})$ then $v_6 v_5 \in E(\vec{G})$, otherwise $v_3 v_4 v_5 v_6$ violates (N2). On the other hand, $v_4 v_5, v_6 v_5 \in E(\vec{G})$ violates (N1) as $[v_3, v_4, v_5, v_6]$ is an (N1)-configuration. Hence $v_5 v_4 \in E(\vec{G})$. If $v_6 v_5 \in E(\vec{G})$ then $[v_6, v_5, v_4, v_3]$ is an (N1)-configuration. We are left with the case $v_1 v_2, v_2 v_3, v_3 v_4, v_5 v_4, v_5 v_6 \in E(\vec{G})$, as shown in Figure 1.
In this case, (N3) does not hold for v_3 and v_5 since $v_4 \in N^+(v_3) \cap N^+(v_5)$ whereas $v_6 \in N^+(v_5) \setminus N^+(v_3)$ and $v_2 \in N^+(v_3) \setminus N^+(v_5)$.

(ii): $v_1 v_2, v_2 v_3 \in E(\vec{G})$. Then $v_4 v_3 \in E(\vec{G})$, otherwise $v_1 v_2 v_3 v_4$ violates (N2). On the other hand, $v_4 v_3 \in E(\vec{G})$ yields that $[v_1, v_2, v_3, v_4]$ is an (N1)-configuration, a contradiction.

(iii): $v_2 v_1, v_2 v_3 \in E(\vec{G})$. Suppose $v_4 v_3 \in E(\vec{G})$. Then (N3) yields $v_5 v_4 \in E(\vec{G})$ since $v_3 \in N^+(v_2) \cap N^+(v_4)$ and $v_1 \in N^+(v_2) \setminus N^+(v_4)$. On the other hand, if $v_5 v_4 \in E(\vec{G})$ then
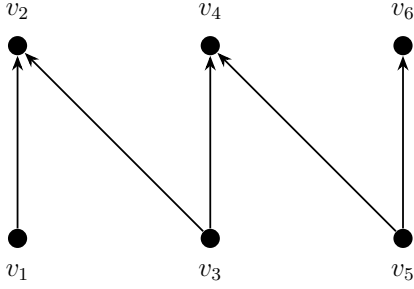
Figure 1: Case (i): all edges of $\vec{G}$ up to possible symmetric edges.

$[v_5, v_4, v_3, v_2]$ is an (N1)-configuration, a contradiction. Hence $v_3 v_4 \in E(\vec{G})$. If $v_4 v_5 \in E(\vec{G})$ then $[v_2, v_3, v_4, v_5]$ is an (N1)-configuration, a contradiction. On the other hand, if $v_5 v_4 \in E(\vec{G})$ then $v_2 v_3 v_4 v_5$ with $v_2 v_5 \notin E(\vec{G})$ which contradicts (N2).

(iv): $v_2 v_1, v_3 v_2 \in E(\vec{G})$. Then $v_3 v_4 \in E(\vec{G})$, otherwise $v_4 v_3 v_2 v_1$ together with $v_4 v_1 \notin E(\vec{G})$ violate (N2). Suppose that $v_5 v_4 \in E(\vec{G})$. Then (N3) yields $v_6 v_5 \in E(\vec{G})$, since $v_4 \in N^+(v_3) \cap N^+(v_5)$ and $v_2 \in N^+(v_3) \setminus N^+(v_5)$. However, $v_6 v_5 \in E(\vec{G})$ contradicts (N1), as it yields that $[v_6, v_5, v_4, v_3]$ is an (N1)-configuration. Therefore we have $v_4 v_5 \in E(\vec{G})$. If $v_6 v_5 \in E(\vec{G})$ then $[v_3, v_4, v_5, v_6]$ is an (N1)-configuration, a contradiction. On the other hand, if $v_5 v_6 \in E(\vec{G})$, $v_3 v_4 v_5 v_6$ with $v_3 v_6 \notin E(\vec{G})$ violates (N2). $\square$

Theorem 3.1 implies that the underlying undirected graph of any 2-qBMG is P_k free for $k \geq 6$, and also poses the problem of P_k-freeness for $3 \leq k \leq 5$.

Theorem 3.2. *There exist 2-qBMGs on five vertices which are not P_5-free. More precisely, there are exactly six non-isomorphic 2-qBMGs on five vertices $\{v_1, v_2, v_3, v_4, v_5\}$ whose underlying undirected graph is a P_5 path-graph with edge-sets are:*

$E(\vec{P}_5^{(a)}) = \{v_1 v_2, v_3 v_2, v_3 v_4, v_4 v_5\}$,
$E(\vec{P}_5^{(b)}) = \{v_1 v_2, v_3 v_2, v_3 v_4, v_5 v_4\}$,
$E(\vec{P}_5^{(c)}) = \{v_2 v_1, v_3 v_2, v_3 v_4, v_4 v_5\}$,
$E(\vec{P}_5^{(a1)}) = \{v_1 v_2, v_2 v_1, v_3 v_2, v_3 v_4, v_4 v_5\}$,
$E(\vec{P}_5^{(b1)}) = \{v_1 v_2, v_2 v_1, v_3 v_2, v_3 v_4, v_5 v_4\}$,
$E(\vec{P}_5^{(ab)}) = \{v_1 v_2, v_2 v_1, v_3 v_2, v_3 v_4, v_4 v_5, v_5 v_4\}$.

Proof. Let $\vec{G}$ be a 2-qBMG on five vertices such that its underlying undirected graph G has five pairwise distinct vertices, say v_1, v_2, v_3, v_4, v_5 where $v_1 v_2 v_3 v_4 v_5$ is a path containing no chord other than $v_i v_{i+1}$ for $i = 1, \dots, 4$. The arguments in the proof of Theorem 3.1 show that no 2-qBMG satisfies either $v_1 v_2, v_2 v_3 \in E(\vec{G})$ or $v_2 v_1, v_2 v_3 \in E(\vec{G})$. Therefore, two cases arise only according to the possible patterns of the neighborhood of v_2.

(i): $v_1 v_2, v_3 v_2 \in E(\vec{G})$. The arguments in proof of Theorem 3.1 show that $v_4 v_3 \notin E(\vec{G})$, and hence $v_3 v_4 \in E(\vec{G})$. Moreover either $v_4 v_5 \in E(\vec{G})$ or $v_5 v_4 \in E(\vec{G})$, and the arising digraphs are $\vec{P}_5^{(a)}$ and $\vec{P}_5^{(b)}$, respectively. Adding edges with consecutive endpoints to $E(\vec{P}_5^{(a)})$ or $E(\vec{P}_5^{(b)})$ provide more non-isomorphic 2-qBMGs: If we add $v_2 v_1$ creating a symmetric edge arises, we obtain two more non-isomorphic digraphs, named $\vec{P}_5^{(a1)}$ and $\vec{P}_5^{(b1)}$, respectively. Adding $v_5 v_4$ to $E(\vec{P}_5^{(a)})$ or $v_4 v_5$ to $E(\vec{P}_5^{(b)})$, the same 2-qBMG arises. Actually, it is isomorphic to $\vec{P}_5^{(b1)}$ by the map μ fixing v_3 and swapping v_2 with v_4, and v_1 with v_5. If we add $v_5 v_4$ to $\vec{P}_5^{(a1)}$, or $v_4 v_5$ to $\vec{P}_5^{(b1)}$, we obtain $\vec{P}_5^{(ab)}$. On the other hand, the above method does not work with $v_2 v_3$ or $v_4 v_3$, since the arising digraph would not satisfy (N2).

(ii): $v_2 v_1, v_3 v_2 \in E(\vec{G})$. The arguments in the proof of Theorem 3.1 can also be used to show that $v_4 v_3 \notin E(\vec{G})$, whence $v_3 v_4 \in E(\vec{G})$ follows. Therefore, either $v_4 v_5 \in E(\vec{G})$ or $v_5 v_4 \in E(\vec{G})$. In the latter case, the digraph is isomorphic to $\vec{P}_5^{(a)}$ by the above map μ. In the former case, $\vec{P}_5^{(c)}$ is obtained, which is not isomorphic to any of the 2-qBMGs already considered. Moreover, adding the edge $v_1 v_2$ to $E(\vec{P}_5^{(c)})$ gives $\vec{P}_5^{(a1)}$. Therefore, no further non-isomorphic 2-qBMGs arise since adding either $v_2 v_3$ or $v_4 v_3$ would violate (N2). $\square$

Remark 3.3. In Theorem 3.2, $\vec{P}_5^{(ab)}$ is a 2-BMG explained by the phylogenetic tree (T, σ_p) in Figure 2, where σ_p is the leaf-coloring defined according to the parity of labels of the leaves. Moreover, the other five 2-qBMGs in Theorem 3.2 are explained by (T, σ_p, τ) where τ is an appropriate sink-truncation map depending on the digraph.
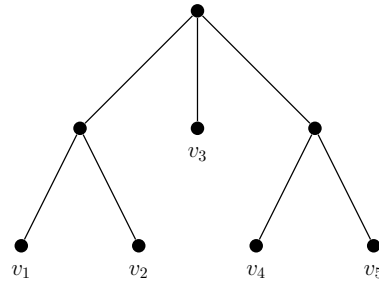

Figure 2: Tree topology explaining all digraphs in Theorem 3.2. Leaves v_1, v_3, v_5 have the same color opposite to the color of leaves v_2, v_4.

Example 3.4. The 2-qBMG $\vec{G}$ on 7 vertices $\{v_1, v_2, \dots, v_7\}$ and edge-set

$$E(\vec{G}) = \{v_5 v_4, v_2 v_1, v_3 v_4, v_3 v_2, v_4 v_7, v_1 v_6, v_3 v_6, v_6 v_1, v_7 v_4\}$$

is an example of a larger 2-qBMG whose induced sub-graph on $\{v_1, v_2, v_3, v_4, v_5\}$ coincides with $\overrightarrow{P_5}^{(a1)}$.

A case-by-case analysis similar but simpler to those in the proof of Theorem 3.2 gives the following result.

Theorem 3.5. *There are exactly four non-isomorphic 2-qBMGs on four vertices $\{v_1, v_2, v_3, v_4\}$ whose underlying undirected graph is a P_4 path-graph:*

$E(\overrightarrow{P_4}^{(1)}) = \{v_1 v_2, v_1 v_4, v_2 v_3\}$,
$E(\overrightarrow{P_4}^{(2)}) = \{v_1 v_2, v_1 v_4, v_3 v_4\}$,
$E(\overrightarrow{P_4}^{(3)}) = \{v_1 v_2, v_1 v_4, v_2 v_1, v_2 v_3\}$,
$E(\overrightarrow{P_4}^{(4)}) = \{v_1 v_2, v_2 v_1, v_4 v_1, v_4 v_3\}$.

In Theorem 3.5, all the non-isomorphic 2-qBMGs on four vertices contain a sink. Therefore, the undirected underlying graph of a 2BMG is P_4-free since 2BMGs are sink-free 2-qBMG.

Corollary 3.6. *The underlying undirected graph of a 2BMG is a cograph.*

The P_3-freeness problem is solved in the following proposition.

Proposition 3.7. *Every bipartite digraph on three vertices $\{v_1, v_2, v_3\}$ is a 2-qBMG.*

Proof. It is straightforward to check that both (N1) and (N2) hold trivially for every bipartite digraph $\overrightarrow{G}$ on three vertices. For (N3), it also holds trivially when $|E(\overrightarrow{G})| = 1$. On the other hand, assign a vertex-coloring on $V(\overrightarrow{G})$ such that v_1 and v_3 have the same color. The only case when two vertices have at least one common out-neighbor is when $N^+(v_1) = N^+(v_3) = \{v_2\}$. Hence, (N3) holds also in the case when $|E(\overrightarrow{G})| > 1$. $\square$

A straightforward consequence of Proposition 3.7 is a characterization of the 2-qBMGs on three vertices whose underlying undirected graph is a P_3 path-graph.

Corollary 3.8. *The non-isomorphic 2-qBMGs on three vertices whose underlying undirected graph is a P_3 path-graph are all bipartite digraphs with at least two edges except for a symmetric edge.*

The same setup and arguments from the proof of Theorem 3.1 can be used to deal with induced 6-cycles in 2-qBMGs since the proof of Theorem 3.1 involves neither $v_1 v_6$ nor $v_6 v_1$. Therefore, the following result holds.

Theorem 3.9. *The underlying undirected graph of any 2-qBMG is C_6-free.*

Remark 3.10. 2-qBMGs are bipartite digraphs; therefore, they cannot induce cycles of an odd length. In particular, they are C_5-free and C_3-free. On the other

hand, 2-qBMGs with paths and cycles of length $\ell < 5$ exist. For $\ell = 4$, consider the digraph $\overrightarrow{G}$ on 4 vertices $\{v_1, v_2, v_3, v_4\}$ and edge-set $E(\overrightarrow{G}) = \{v_2 v_1, v_2 v_3, v_3 v_4\}$. The underlying graph of $\overrightarrow{G}$ is a path of length 4, and $\overrightarrow{G}$ is a 2-qBMG; indeed it is explained by (T, σ_p, τ), where T is represented in Figure 3(a), σ_p is the leaf-coloring map defined by the parity of the leaf labels and τ is the sink-truncation maps of G. For $l = 3$, consider the digraph $\overrightarrow{G}$ on 3 vertices $\{v_1, v_2, v_3\}$ and edge-set $\overrightarrow{G} = \{v_2 v_1, v_2 v_3\}$ is a 2-qBMG explained by (T, σ_p, τ), where T is in Figure 3(b). The underlying graph of $\overrightarrow{G}$ is a path of length 3.
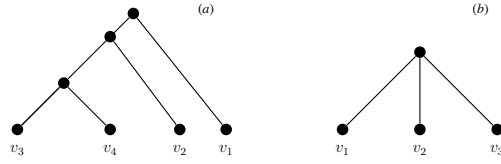

Figure 3: (a) Tree topology explaining a 2-qBMG with an induced P_4. (b) Tree topology explaining a 2-qBMG with an induced P_3.

The arguments used in the proof of Theorem 3.2 can be adapted to prove the following theorem.

Theorem 3.11. *There are exactly 10 non-isomorphic 2-qBMGs on four vertices $\{v_1, v_2, v_3, v_4\}$ whose underlying undirected graph is a C_4. They are the following ten digraphs:*

$E(\overrightarrow{P_4}^{(1)}) = \{v_1 v_2, v_1 v_4, v_2 v_1, v_2 v_3, v_3 v_4, v_4 v_3\}$,
$E(\overrightarrow{P_4}^{(2)}) = \{v_1 v_2, v_1 v_4, v_2 v_3, v_3 v_2, v_3 v_4, v_4 v_3\}$,
$E(\overrightarrow{P_4}^{(3)}) = \{v_1 v_2, v_1 v_4, v_3 v_2, v_3 v_4\}$,
$E(\overrightarrow{P_4}^{(4)}) = \{v_1 v_2, v_1 v_4, v_3 v_2, v_4 v_3\}$,
$E(\overrightarrow{P_4}^{(5)}) = \{v_1 v_2, v_2 v_1, v_3 v_2, v_3 v_4, v_4 v_1\}$,
$E(\overrightarrow{P_4}^{(6)}) = \{v_1 v_2, v_1 v_4, v_3 v_2, v_4 v_1, v_4 v_3\}$,
$E(\overrightarrow{P_4}^{(7)}) = \{v_1 v_2, v_1 v_4, v_2 v_3, v_4 v_3\}$,
$E(\overrightarrow{P_4}^{(8)}) = \{v_1 v_2, v_1 v_4, v_3 v_2, v_3 v_4, v_4 v_3\}$,
$E(\overrightarrow{P_4}^{(9)}) = \{v_1 v_2, v_1 v_4, v_2 v_1, v_2 v_3, v_3 v_2, v_3 v_4\}$,
$E(\overrightarrow{P_4}^{(10)}) = \{v_1 v_2, v_1 v_4, v_3 v_2, v_3 v_4, v_2 v_1, v_2 v_3, v_4 v_1, v_4 v_3\}$.

4. Dominating sets in 2-qBMGs

As pointed out in the introduction, Theorem 3.1 and Theorem 3.9, together with previous results, give new contributions to the current studies on the structure of

2-qBMGs. This section is focused on dominating biclique sets.

Recall that a connected 2-qBMG is of type (A) if its undirected underlying graph has a vertex decomposition into a biclique and a stable set. It may be observed that such a biclique is necessarily a dominating set.

Example 4.1. The digraph on 10 vertices $\{v_1, v_2, \ldots, v_{10}\}$ with edge-set

$$E(\vec{G}) = \{v_1v_5, v_1v_6, v_1v_7, v_1v_8, v_5v_2, v_6v_2, v_2v_7, v_2v_8, v_5v_3, v_6v_3,$$
$$v_3v_7, v_3v_8, v_5v_4, v_6v_4, v_7v_4, v_4v_8, v_5v_9, v_1v_{10}, v_2v_{10}\}$$

is a 2-qBMG of type (A) whose underlying undirected graph G has an induced dominating subgraph with vertex set $\{v_1, v_2, v_3, v_4, v_5, v_6, v_7, v_8\}$ and stable set $\{v_9, v_{10}\}$.

Theorem 4.2. *Every connected 2-qBMG has a vertex decomposition into connected 2-qBMGs of type (A).*

Proof. Let $\vec{G}$ be a connected 2-qBMG not of type (A). Let G be the underlying undirected graph of $\vec{G}$, where U and W are its color classes. As mentioned in the introduction, from [15, Theorem 1], G contains a dominating biclique Δ with vertex set $T \cup Z$ where $T \subseteq U$ and $Z \subseteq W$. Let $\vec{\Delta}$ be the directed subgraph of $\vec{G}$ induced by Δ. Let S be the subset of $V(G)$ consisting of all vertices $v \in V(G)$ for which $N^+(v) \cup N^-(v)$ is contained in $T \cup Z$. Let $\vec{\Sigma}$ be the induced subgraph of $\vec{G}$ on $T \cup Z \cup S$. Since Δ is a dominating set of G, $\vec{\Sigma}$ is connected without isolated vertex and is a connected 2-qBMG. The underlying undirected graph Σ of $\vec{\Sigma}$ is a $K \oplus S$ graph, where $K = T \cup Z$.

Furthermore, let $\vec{G}_1$ be the induced subgraph of $\vec{G}$ on $V(G) \setminus (T \cup Z \cup S)$. By construction, no vertex of $\vec{G}_1$ is isolated, although some vertex of $\vec{G}$ may become a sink in $\vec{G}_1$, and non-equivalent vertices $\vec{G}$ may become equivalent in $\vec{G}_1$. If G_1 is not a connected 2-qBMG of type (A), we can repeat the above construction to define a finite sequence $\Sigma_1, G_2, \Sigma_2 \ldots, \Sigma_k$ such that $\vec{\Sigma}, \vec{\Sigma}_1, \vec{\Sigma}_2, \ldots, \vec{\Sigma}_k$ form a vertex partition of $\vec{G}$ into connected 2-qBMGs of type (A).

If some of the members of the subsequence $G_1, G_2, \ldots, G_{k-1}$ decomposition is not connected, say G_m, applying the above argument on each of the connected components of G_m gives the claimed partition. □

Proposition 4.3. *Let $\vec{G}$ be a 2-qBMG satisfying at least one of conditions $(*)$ and $(**)$. Fix an orientation $\vec{\Gamma}$ of $\vec{G}$. Let $\vec{\Delta}$ be an induced subgraph of $\vec{\Gamma}$ such that the underlying undirected graph Δ is a biclique of G. If $\vec{G}$ has no symmetric edges in $\vec{\Delta}$ then $\vec{\Delta}$ is a 2-qBMG.*

Proof. Let n be the number of the vertices of $\vec{G}$ (and of $\vec{\Gamma}$). As pointed out in Section 2, if either $(*)$ or $(**)$ holds then

$\vec{\Gamma}$ has a topological ordering. Therefore, the vertices of $\vec{\Gamma}$ can be labelled with $\{v_1, v_2, \ldots, v_n\}$ such that $v_iv_j \in E(\vec{\Gamma})$ implies $i < j$ for $1 \leq i, j \leq n$. In particular, for $v_i, v_j \in V(\Delta)$ of different color, either v_iv_j or v_jv_i is an edge of $\vec{\Delta}$. Therefore, (N1) trivially holds for $\vec{\Delta}$ while (N2) follows from the transitivity of the ordering. To show (N3), take $v_i, v_j \in V(\Delta)$ of the same color such that v_i and v_j have a common out-neighbour $w \in \vec{\Delta}$. Then w is a common out-neighbour of v_i and v_j in $\vec{G}$. From (N3) applied to $\vec{G}$, all out-neighbours of v_i in $\vec{G}$ are also out-neighbours of v_j (or, vice versa). Now take any vertex v such that $v_iv \in E(\vec{\Delta})$. Then $v_iv \in E(\vec{G})$, and hence $v_jv \in E(\vec{G})$. Since v_jv is not a symmetric edge in $\vec{G}$, this yields $v_jv \in E(\vec{\Gamma})$ whence $v_jv \in E(\vec{\Delta})$ follows. Thus $\vec{\Delta}$ is a 2-qBMG. □

Theorem 4.4. *Let $\vec{G}$ be a connected 2-qBMG satisfying $(*)$. Then $\vec{G}$ has an even-odd subgraph $(\mathfrak{A}, \mathfrak{O})$ such that the underlying undirected subgraph of $(\mathfrak{A}, \mathfrak{O})$ is a dominating biclique of G.*

Proof. Fix an orientation $\vec{\Gamma}$ of $\vec{G}$ by keeping the same vertex set but retaining exactly one edge from each symmetric edge. Let G be the underlying undirected graph of $\vec{G}$. Then G coincides with the underlying undirected graph of $\vec{\Gamma}$. As mentioned in Introduction, from [15, Theorem 1], G contains a dominating biclique Δ with vertex set $T \cup Z$ where $T \subseteq U$ and $Z \subseteq W$ where U and W are the bipartion classes of G. Let $\vec{\Delta}$ be the directed subgraph on vertex-set $T \cup Z$ where $v_iv_j \in E(\vec{\Delta})$ if and only if v_iv_j is an edge of $\vec{\Gamma}$. It should be noticed that $\vec{\Delta}$ may not be the induced subgraph of $\vec{G}$ on the vertex-set $T \cup Z$. This occurs indeed when $\vec{G}$ contains a symmetric edge with both endpoints in $T \cup Z$. Nevertheless, $\vec{\Delta}$ is bitournament. Since $\vec{\Delta}$ is bitransitive, it is an even-odd digraph, as recalled in Section 2. □

5. Future directions

The P_6-and C_6-freeness in the undirected underlying 2-qBMGs raises the problem of finding more such forbidden subgraphs on six (or eight) vertices. In this direction, the domino graph on six vertices (i.e. the Cartesian product $P_2 \times P_3$) and its generalization on eight vertices, the long-domino or ladder graph, appearworth investigating.

In [20], bipartite P_6 and C_6-free graphs are characterized as those graphs where every connected subgraph is of type $K \oplus S$. This gives rise to a linear time algorithm for recognizing if an undirected graph is P_6 and C_6-free by tracing all $K \oplus S$ decompositions. It would be interesting to investigate whether a similar approach can be used to perform a linear time algorithm to recognize if a

digraph is a 2-qBMG. The first step is to check whether the converse of [20, Theorem 2] holds.

The s-dim of a graph G is the minimum number of bicliques needed to cover the edge-set of G. Computing the s-dim (biclique cover problem) is NP-complete for bipartite graphs [21]; however, this problem is linear for some classes of graphs, such as bipartite domino-free graphs [22]. Investigating the complexity of the biclique cover problem for the family of underlying undirected graphs of 2-qBMGs is also an interesting topic for future investigation.

For two disjoint subsets W and T of vertices, W 2-*dominates* T if every vertex of T is adjacent to at least two vertices of W. A partition $\{V_1, V_2, \ldots, V_n\}$ of vertices in $V(G)$ into n parts is a *2-transitive partition* of size n if V_i 2-dominates for all $1 \leq i < j \leq n$. Finding a 2-transitivity partition of maximum order is NP-complete for bipartite graphs [23]. Studying the complexity of this problem for the family of underlying undirected graphs of 2-qBMGs is also an interesting topic for future investigation.

References

[1] A. Korchmaros, D. Schaller, M. Hellmuth, P. F. Stadler, Quasi-best match graphs, Discr. Appl. Math. 331 (2023) 104–125.

[2] J. A. Ramírez-Rafael, A. Korchmaros, K. Aviña-Padilla, A. López Sánchez, A. A. España-Tinajero, M. Hellmuth, P. F. Stadler, M. Hernández-Rosales, Revolutionh-tl: Reconstruction of evolution ary histories tool, in: RECOMB International Workshop on Comparative Genomics, Springer, 2024, pp. 89–109.

[3] D. Schaller, M. Geiß, E. Chávez, M. González Laffitte, A. López Sánchez, B. M. Stadler, D. I. Valdivia, M. Hellmuth, M. Hernández Rosales, P. F. Stadler, Corrigendum to "best match graphs", J. Math. Biology 82 (2021).

[4] A. Korchmaros, The structure of 2-colored best match graphs, Discr. Appl. Math. 304 (2021).

[5] D. Schaller, M. Geiß, M. Hellmuth, P. F. Stadler, Heuristic algorithms for best match graph editing, Algorithms for Molecular Biology 16 (2021).

[6] D. Schaller, P. F. Stadler, M. Hellmuth, Complexity of modification problems for best match graphs, Theoretical Computer Science 865 (2021).

[7] D. Schaller, M. Geiß, P. F. Stadler, M. Hellmuth, Complete characterization of incorrect orthology assignments in best match graphs, J. Math. Biology 82 (2021).

[8] M. Hellmuth, P. F. Stadler, The theory of gene family histories, in: Comparative Genomics: Methods and Protocols, Springer, 2024, pp. 1–32.

[9] G. Manuela, P. F. Stadler, M. Hellmuth, Reciprocal best match graphs, J. Math. Biology 80 (2020).

[10] M. Hellmuth, M. Geiß, P. F. Stadler, Complexity of modification problems for reciprocal best match graphs, Theoretical Computer Science 809 (2020).

[11] G. Bacsó, Z. Tuza, Domination properties and induced subgraphs, Discr. Math. 111 (1993) 37–40.

[12] G. Bacsó, D. Michalak, Z. Tuza, Dominating bipartite subgraphs in graphs, Discussiones Mathematicae Graph Theory 25 (2005) 85–94.

[13] A. Brandstädt, T. Klembt, S. Mahfud, P6- and triangle-free graphs revisited: structure and bounded clique-width, Discrete Mathematics & Theoretical Computer Science 8 (2006).

[14] P. van't Hof, D. Paulusma, A new characterization of p_6-free graphs, Discr. Appl. Math. 158 (2010) 731–740.

[15] J. Liu, H. Zhou, Dominating subgraphs in graphs with some forbidden structures, Discr. Math. 135 (1994) 163–168.

[16] J. Liu, Y. Peng, C. Zhao, Characterization of p_6-free graphs, Discr. Appl. Math. 155 (2007) 1038–1043.

[17] R. Peeters, The maximum edge biclique problem is np-complete, Discrete Applied Mathematics 131 (2003) 651–654.

[18] E. Prisner, Bicliques in graphs I: bounds on their number, Combinatorica 20 (2000) 197–207.

[19] J.-L. Fouquet, V. Giakoumakis, J.-M. Vanherpe, Bipartite graphs totally decomposable by canonical decomposition, International Journal of Foundations of Computer Science 10 (1999) 513–533.

[20] R. Quaddoura, A. Al-Qerem, Bipartite (P_6,C_6)-free graphs: Recognition and optimization problems, Symmetry 16 (2024) 447.

[21] M. Dawande, P. Keskinocak, J. M. Swaminathan, S. Tayur, The biclique cover problem for bipartite graphs, SIAM Journal on Discrete Mathematics 15 (2001) 255–271.

[22] J. Amilhastre, M.-C. Vilarem, P. Janssen, Complexity of minimum biclique cover and minimum biclique decomposition for bipartite domino-free graphs, Discr. Appl. Math. 86 (1998) 125–144.

[23] S. Paul, K. Santra, Algorithmic study on 2-transitivity of graphs, arXiv preprint arXiv:2310.04036 (2023).