

Balanced Substructures in Bicolored Graphs

P. S. Ardra^{a,*}, R. Krithika^a, Saket Saurabh^{b,c}, Roohani Sharma^d

^a*Indian Institute of Technology Palakkad, Palakkad, India*

^b*The Institute of Mathematical Sciences, Chennai, India*

^c*University of Bergen, Bergen, Norway*

^d*Max Planck Institute for Informatics, Saarland Informatics
Campus, Saarbrücken, Germany*

Abstract

An edge-colored graph is said to be *balanced* if it has an equal number of edges of each color. Given a graph G whose edges are colored using two colors and a positive integer k , the objective in the EDGE BALANCED CONNECTED SUBGRAPH problem is to determine if G has a balanced connected subgraph containing at least k edges. We first show that this problem is NP-complete and remains so even if the solution is required to be a tree or a path. Then, we focus on the parameterized complexity of EDGE BALANCED CONNECTED SUBGRAPH and its variants (where the balanced subgraph is required to be a path/tree) with respect to k as the parameter. Towards this, we show that if a graph has a balanced connected subgraph/tree/path of size at least k , then it has one of size at least k and at most $f(k)$ where f is a linear function. We use this result combined with dynamic programming algorithms based on *color coding* and *representative sets* to show that EDGE BALANCED CONNECTED SUBGRAPH and its variants are FPT. Further, using polynomial-time reductions to the MULTILINEAR MONOMIAL DETECTION problem, we give faster randomized FPT algorithms for the problems. In order to describe these reductions, we define a combinatorial object called *relaxed-subgraph*. We define this object in such a way that balanced connected subgraphs, trees and paths are relaxed-subgraphs with certain properties. This object is defined in the

*Corresponding author

Email addresses: 111914001@smail.iitpkd.ac.in (P. S. Ardra),
krithika@iitpkd.ac.in (R. Krithika), saketa@imsc.res.in (Saket Saurabh),
rsharma@mpi-inf.mpg.de (Roohani Sharma)

The second author is supported by SERB CRG grant number CRG/2022/007751.

spirit of branching walks known for the STEINER TREE problem and may be of independent interest.

Keywords: red-blue graphs, balanced connected subgraphs, balanced trees, balanced paths, parameterized complexity

1. Introduction

Ramsey Theory is a branch of Combinatorics that deals with patterns in large arbitrary structures. In the context of edge-colored graphs where each edge is colored with one color from a finite set of colors, a fundamental problem in the area is concerned with the existence of *monochromatic* subgraphs of a specific type. Here, monochromatic means that all edges of the subgraph have the same color. For simplicity, we discuss only undirected graphs where each edge is colored either red or blue. Such a coloring is called a *red-blue coloring* and a graph associated with a red-blue coloring is referred to as a *red-blue graph*. In this work, we study questions related to the existence of and finding *balanced* subgraphs instead of monochromatic subgraphs, where by a balanced subgraph we mean one which has an equal number of edges of each color. These problems come under a subarea of Ramsey Theory known as Zero-Sum Ramsey Theory. Here, given a graph whose vertices/edges are assigned weights from a set of integers, one looks for conditions that guarantee the existence of a certain subgraph having total weight zero. For example, one may ask when is a graph whose all the edges are given weight -1 or 1 guaranteed to have a spanning tree with total weight of its edges 0. This is equivalent to asking when a red-blue graph is guaranteed to have a balanced spanning tree. Necessary and sufficient conditions have been established for complete graphs, triangle-free graphs and maximal planar graphs [1]. In the same spirit, one may ask a more general question like when is a red-blue graph G guaranteed to have a balanced connected subgraph of size (number of edges) k . An easy necessary condition is that there are at least $k/2$ red edges and at least $k/2$ blue edges in G . This condition is also sufficient (as we show in the proof of Theorem 3) if G is a complete graph (or more generally a split graph). However, we do not think that such a simple characterization will exist for all graphs. This brings us to the following natural algorithmic question concerning balanced connected subgraphs.

EDGE BALANCED CONNECTED SUBGRAPH

Parameter: k

Input: A red-blue graph G and a positive integer k

Question: Does G have a balanced connected subgraph of size at least k ?

When the subgraph is required to be a tree or a path, the corresponding variants of EDGE BALANCED CONNECTED SUBGRAPH are called EDGE BALANCED TREE and EDGE BALANCED PATH, respectively. We show that these problems are NP-complete.

- (Theorems 1, 2, 4) EDGE BALANCED CONNECTED SUBGRAPH, EDGE BALANCED TREE and EDGE BALANCED PATH are NP-complete.

In fact, EDGE BALANCED CONNECTED SUBGRAPH and EDGE BALANCED TREE remain NP-complete on bipartite graphs, planar graphs and chordal graphs. However, EDGE BALANCED CONNECTED SUBGRAPH is polynomial-time solvable on split graphs (Theorem 3). Yet, EDGE BALANCED PATH is NP-complete even on split graphs.

Note that if a graph has a balanced connected subgraph/tree/path of size at least k , then it is not guaranteed that it has one of size equal to k . This brings us to the following combinatorial question: if a graph has a balanced connected subgraph/tree/path of size at least k , then can we show that it has a balanced connected subgraph/tree/path of size equal to $f(k)$ for some function f ? We answer these questions in the affirmative and show the existence of such a function which is linear in k .

- (Theorems 5, 6, 7) If a graph has a balanced connected subgraph/tree of size at least k , then it has one of size at least k and at most $3k + 3$. Further, if a graph has a balanced path of size at least k , then it has a balanced path of size at least k and at most $2k$.

Therefore, in order to find a balanced connected subgraph/tree/path of size at least k , it suffices to focus on the problem of finding a balanced connected subgraph/tree/path of size exactly k . This leads us to the following problem.

EXACT EDGE BALANCED CONNECTED SUBGRAPH

Parameter: k

Input: A red-blue graph G and a positive integer k

Question: Does G have a balanced connected subgraph of size k ?

As before, when the connected subgraph is required to be a tree or a path, the corresponding variants of EXACT EDGE BALANCED CONNECTED SUBGRAPH are called EXACT EDGE BALANCED TREE and EXACT EDGE BALANCED PATH, respectively. These problems are also NP-complete and

so we study them from the perspective of parameterized complexity.

In this framework, the key notion is that of a *parameterized problem* which is a decision problem where each input instance is associated with a non-negative integer ℓ called the *parameter*. A parameterized problem is said to be *fixed-parameter tractable* if it can be solved in $f(\ell)n^{\mathcal{O}(1)}$ time for some computable function f where n is the input size. Algorithms with such running times are called **FPT** algorithms or parameterized algorithms and the complexity class **FPT** denotes the set of all parameterized problems that are fixed-parameter tractable. For convenience, the running time $f(\ell)n^{\mathcal{O}(1)}$ of a parameterized algorithm where f is a super-polynomial function is written as $\mathcal{O}^*(f(\ell))$. A *kernelization* or *kernel* for a parameterized problem is a polynomial-time algorithm that transforms an arbitrary instance of the problem to an equivalent instance of the same problem whose size is bounded by some computable function g of the parameter of the original instance. A kernel is a *polynomial kernel* if g is a polynomial function and we say that the problem admits a polynomial kernel. A *polynomial parameter transformation* from problem Π_1 to Π_2 is a polynomial-time algorithm that transforms an arbitrary instance of Π_1 to an equivalent instance of Π_2 such that the parameter of the resulting instance is bounded by some polynomial function of the parameter of the original instance. A parameterized problem is **FPT** if and only if it is decidable and has a kernel. Therefore, the set of parameterized problems that admit polynomial kernels is contained in the class **FPT** and it is believed that this subset relation is strict. Polynomial parameter transformations are useful in ruling out polynomial kernels. For more information on parameterized complexity, we refer to the book by Cygan et al. [2].

Focussing on EXACT EDGE BALANCED CONNECTED SUBGRAPH/ TREE/-PATH with respect to the solution size k as the parameter, we give randomized **FPT** algorithms for solving the three problems using reductions to the MULTILINEAR MONOMIAL DETECTION problem (defined in Section 5).

- (Theorems 8, 9, 10) EXACT EDGE BALANCED CONNECTED SUBGRAPH/ TREE/PATH can be solved by a randomized algorithm in $\mathcal{O}^*(2^k)$ time.

Many problems reduce to MULTILINEAR MONOMIAL DETECTION [3] and the current fastest algorithm solving it is a randomized algorithm that runs in $\mathcal{O}^*(2^k)$ time [3, 4, 5]. The reductions that we give to MULTILINEAR MONOMIAL DETECTION use a combinatorial object called *relaxed-subgraph*. This object is defined in the spirit of *branching walks* known for the STEINER TREE problem [6]. We define this object in such a way that balanced con-

nected subgraphs, trees and paths are relaxed-subgraphs with certain properties. Then, using the *color-coding* technique [2, 7] and *representative sets* [2, 8, 9], we give deterministic dynamic programming algorithms for the problems.

- (Theorems 11, 12, 13) EXACT EDGE BALANCED CONNECTED SUBGRAPH/ TREE can be solved in $\mathcal{O}^*((4e)^k)$ time and EXACT EDGE BALANCED PATH can be solved in $\mathcal{O}^*(2.619^k)$ time.

The method of representative sets is a generic approach for designing efficient dynamic programming based parameterized algorithms that may be viewed as a deterministic-analogue to the color-coding technique. Representative sets have been used to obtain algorithms for several parameterized problems [8] and our algorithm adds to this list.

Road Map. The NP-completeness of the problems are given in Section 3. In Section 4, the combinatorial results related to the existence of small balanced connected subgraphs, trees and paths are proven. Section 5 discusses the deterministic and randomized algorithms for the problems. Section 6 concludes the work by listing some future directions.

Related Work. A variant of EXACT EDGE BALANCED CONNECTED SUBGRAPH has recently been studied [10, 11, 12, 13, 14]. In order to state these results using our terminology, we define the notion of *vertex-balanced subgraphs* of *vertex-colored graphs*. A coloring of the vertices of a graph using red and blue colors is called a *red-blue vertex coloring*.

A subgraph of a vertex-colored graph is said to be *vertex-balanced* if it has an equal number of vertices of each color. In the EXACT VERTEX BALANCED CONNECTED SUBGRAPH problem, the interest is in finding a vertex-balanced connected subgraph on k vertices in the given graph associated with a red-blue vertex coloring. This problem is NP-complete and remains so on restricted graph classes like bipartite graphs, planar graphs, chordal graphs, unit disk graphs, outer-string graphs, complete grid graphs, and unit square graphs [10, 11]. However, polynomial-time algorithms are known for trees, interval graphs, split graphs, circular-arc graphs and permutation graphs [10, 11]. Further, the problem is NP-complete even when the subgraph required is a path [10]. FPT algorithms, exact exponential-time algorithms and approximation results for the problem are known from [11], [13] and [14]. Observe that finding vertex-balanced connected subgraphs in

vertex-colored graphs reduces to finding vertex-balanced trees while the analogous solution in edge-colored graphs may have more complex structures.

2. Preliminaries

Let $[k]$ denote the set $\{1, 2, \dots, k\}$ for $k \in \mathbb{N}_+$. For standard graph-theoretic terminology not stated here, we refer the reader to the book by Diestel [15]. For an undirected graph G , $V(G)$ denotes its set of vertices and $E(G)$ denotes its set of edges. The *size* of a graph is the number of its edges and the *order* of a graph is the number of its vertices. An edge between vertices u and v is denoted as $\{u, v\}$ and u and v are called the *endpoints* of the edge $\{u, v\}$. Two vertices u, v in $V(G)$ are *adjacent* if $\{u, v\} \in E(G)$. The *neighborhood* of a vertex v , denoted by $N_G(v)$, is the set of vertices adjacent to v . Similarly, two edges e, e' in $E(G)$ are *adjacent* if they have exactly one common endpoint and the *neighborhood* of an edge e , denoted by $N_G(e)$, is the set of edges adjacent to e . The *degree* of a vertex v is the size of $N_G(v)$. A vertex is called an *isolated vertex* if its degree is 0. An edge with an endpoint that has degree 1 is called a *pendant edge*. The notion of neighborhood is extended to a set $S \subseteq V(G)$ of vertices by defining $N_G(S)$ as $(\bigcup_{v \in S} N(v)) \setminus S$. We omit the subscript in the notation for neighborhood if the graph under consideration is clear. For a set $F \subseteq E(G)$ of edges, $V(F)$ denotes the set of endpoints of edges in F . For a set $S \subseteq V(G) \cup E(G)$, $G - S$ denotes the graph obtained by deleting S from G . For a set $S \subseteq V(G)$ of vertices, the subgraph of G induced on S is denoted by $G[S]$. For two sets $S, T \subseteq V(G)$ of vertices, $E(S, T)$ denotes the set of edges with one endpoint in S and the other endpoint in T .

A *walk* in G is a sequence (v_1, \dots, v_k) of vertices such that for each $i \in [k - 1]$, $\{v_i, v_{i+1}\} \in E(G)$. A *path* P in G is a walk (v_1, \dots, v_k) consisting of distinct vertices. For a path $P = (v_1, \dots, v_k)$, the set $\{v_1, \dots, v_k\}$ is denoted by $V(P)$ and the set $\{\{v_i, v_{i+1}\} \mid i \in [k - 1]\}$ is denoted by $E(P)$. We say that P *starts at* v_1 and *ends at* v_k . Vertices v_1 and v_k are called the *endpoints* of P and edges $\{v_1, v_2\}$ and $\{v_{k-1}, v_k\}$ are called *terminal edges* of P . The *length* of a path or walk is the number of edges in it. A graph is *connected* if there is a path between every pair of its vertices. Given a graph G , its *line graph* $L(G)$ is defined as $V(L(G)) = \{e \mid e \in E(G)\}$ and $E(L(G)) = \{\{e, e'\} \mid e \text{ and } e' \text{ are adjacent in } G\}$. It is well-known that a graph G without isolated vertices is connected if and only if $L(G)$ is connected. A *tree* is an undirected connected acyclic graph. A *clique* is a set of pairwise adjacent vertices and a

complete graph is a graph whose vertex set is a clique. An *independent set* is a set of pairwise non-adjacent vertices. A *split graph* is a graph whose vertex set can be partitioned into a clique and an independent set.

3. NP-hardness Results

We first show the NP-hardness of EDGE BALANCED CONNECTED SUBGRAPH using a polynomial-time reduction from the well-known NP-hard STEINER TREE problem [16, ND12]. In this problem, given a connected graph G , a subset $T \subseteq V(G)$ (called *terminals*) and a positive integer k , the task is to determine if G has a subtree H (called a *Steiner tree*) with $T \subseteq V(H)$ and $|E(T)| \leq k$. The idea behind the reduction is to color all edges of G of the STEINER TREE instance blue and add exactly k red edges incident to the terminals such that each terminal has at least one red edge incident on it. Any balanced connected subgraph of size (at least) k of the resulting graph is required to include all the red edges and hence includes all the terminals which in turn corresponds to a Steiner tree of G .

Theorem 1. EDGE BALANCED CONNECTED SUBGRAPH is NP-complete.

Proof. It is easy to verify that EDGE BALANCED CONNECTED SUBGRAPH is in NP. Consider an instance (G, T, k) of STEINER TREE. Let $T = \{t_1, t_2, \dots, t_\ell\}$ be the terminal vertices. Without loss of generality assume $\ell \leq k$ and $|E(G)| \geq k$. Now we construct an instance $(H, 2k)$ of EDGE BALANCED CONNECTED SUBGRAPH that is equivalent to (G, T, k) . The graph H is obtained from G as follows: for each $t \in T$ we add a new vertex t' adjacent to t . Then we add $k - \ell$ new vertices adjacent to t_1 . Formally $V(H) = V(G) \cup \{t' \mid t \in T\} \cup \{t'_{\ell+i} \mid i \in [k - \ell]\}$ and $E(H) = E(G) \cup \{\{t, t'\} \mid t \in T\} \cup \{\{t_1, t'_{\ell+i}\} \mid i \in [k - \ell]\}$. Next we define a red-blue coloring of $E(H)$ as follows: edges in $E(H) \cap E(G)$ are colored blue and edges in $E(H) \setminus E(G)$ are colored red. Let $T' = V(H) \setminus V(G)$ and $E' = E(H) \setminus E(G)$.

Suppose $(H, 2k)$ is a YES-instance. Let H' be a balanced connected subgraph of H with at least $2k$ edges. Then, as H has exactly k red edges (which is the set E'), $|E(H')| = 2k$. Further, as $V(E') = T \cup T'$ it follows that $T \cup T' \subseteq V(H')$. Now consider the subgraph H'' of G obtained from H' by deleting vertices in T' . Clearly, H'' has exactly k edges. As every vertex in T' has degree 1 and $N_H(T') \subseteq V(G)$, H'' is connected. Further, $T \subseteq V(H'')$. Now any spanning tree of H'' is a Steiner tree with at most k edges. Thus (G, T, k) is a YES-instance.

Conversely, suppose (G, T, k) is a YES-instance. Then let T^* be a subtree of G containing T with $|E(T^*)| \leq k$. Let T^{**} be a subgraph of G , containing T^* such that $|E(T^{**})| = k$. Note that T^{**} is also a subgraph of H and each edge in this subgraph is blue. Define another subgraph H' of H containing T^{**} as follows: $V(H') = V(T^{**}) \cup T'$ and $E(H') = E(T^{**}) \cup E'$. Clearly, H' is a balanced connected subgraph of H of size $2k$. Thus, $(H, 2k)$ is a YES-instance. \square

As the variant of the STEINER TREE problem where a tree on exactly k edges is required is also NP-complete we have the following result.

Theorem 2. *EDGE BALANCED TREE is NP-complete.*

As the reduction described in Theorem 1 is a polynomial parameter transformation, the infeasibility of the existence of polynomial kernels for STEINER TREE parameterized by the solution size (i.e., the size of the tree) [2, 17] extend to EDGE BALANCED CONNECTED SUBGRAPH and EDGE BALANCED TREE as well. Further, since STEINER TREE has no subexponential FPT algorithm assuming the Exponential Time Hypothesis, it follows that EDGE BALANCED CONNECTED SUBGRAPH and EDGE BALANCED TREE also do not admit subexponential FPT algorithms. Moreover, the reduction in Theorem 1 preserves planarity, bipartiteness and chordality. This property along with the NP-completeness of STEINER TREE (and its variant) on bipartite graphs [16], planar graphs [18] and chordal graphs [19] imply that EDGE BALANCED CONNECTED SUBGRAPH and EDGE BALANCED TREE are NP-complete on planar graphs, chordal graphs and bipartite graphs as well.

3.1. Complexity in Split Graphs

Next, we consider EDGE BALANCED CONNECTED SUBGRAPH on split graphs. Let (G, k) be an instance. An easy necessary condition for G to have a balanced connected subgraph of size (at least) k is that there are at least $k/2$ red edges and at least $k/2$ blue edges in G . We show that this condition is also sufficient if G is a split graph leading to the following result.

Theorem 3. *EDGE BALANCED CONNECTED SUBGRAPH is polynomial-time solvable on split graphs.*

Proof. Let (G, k) be an instance where G is a split graph whose vertex set is partitioned into clique C and independent set I . Let E_R be the set of red edges and E_B be the set of blue edges in G . If $|E_R| < \frac{k}{2}$ or $|E_B| < \frac{k}{2}$, then

(G, k) is a NO-instance. Otherwise, $|E_R| \geq \frac{k}{2}$ and $|E_B| \geq \frac{k}{2}$. We show that (G, k) is a YES-instance.

Let v be a vertex in C . Define sets $X = \{x \mid x \in C, \{v, x\} \in E_R\}$ and $Y = \{y \mid y \in C, \{v, y\} \in E_B\}$. We will now construct a balanced connected subgraph H of G in each of the following four cases.

- Case 1: $|X| \geq \frac{k}{2}$ and $|Y| \geq \frac{k}{2}$
 Consider any set S_X of $\frac{k}{2}$ vertices from X and any set S_Y of $\frac{k}{2}$ vertices from Y . Then, the subgraph H of G with $V(H) = S_X \cup S_Y \cup \{v\}$ and $E(H) = E(v, S_X) \cup E(v, S_Y)$ is the required balanced connected subgraph.
- Case 2: $|X| < \frac{k}{2}$ and $|Y| < \frac{k}{2}$
 Initialize $E' = \emptyset$. Let $|X| = k_1$ and $|Y| = k_2$. First, we add all edges in $E(v, X) \cup E(v, Y)$ to E' . Observe that $C \subseteq V(E')$. Next, we add $\frac{k}{2} - k_1$ edges from $E_R \setminus E(v, X)$ and $\frac{k}{2} - k_2$ edges from $E_B \setminus E(v, Y)$ to E' . Then, the subgraph H of G with $E(H) = E'$ and $V(H) = V(E')$ is the required balanced connected subgraph.
- Case 3: $|X| \geq \frac{k}{2}$ and $|Y| < \frac{k}{2}$
 First, we add all edges in $E(v, Y)$ to E' . Next, we add edges in $E(Y, I) \cap E_B$ to E' one by one until $|E' \cap E_B| = \frac{k}{2}$ or $E(Y, I) \cap E_B \subseteq E'$. Similarly, we add edges in $(E(v, I) \cup E(Y, Y)) \cap E_B$ to E' one by one until $|E' \cap E_B| = \frac{k}{2}$ or $(E(v, I) \cup E(Y, Y)) \cap E_B \subseteq E'$. Then, we add edges in $E(X, Y) \cap E_B$ to E' one by one until $|E' \cap E_B| = \frac{k}{2}$ or $E(X, Y) \cap E_B \subseteq E'$. When we add such an edge $\{x, y\}$ with $x \in X, y \in Y$, we also add the edge $\{x, v\}$ (which is guaranteed to be in E_R) to E' . At the end of this procedure, observe that $|E' \cap E_R| \leq \frac{k}{2}$ and $|E' \cap E_B| = \frac{k}{2}$. Finally, we add some set of $\frac{k}{2} - |E' \cap E_R|$ edges from $E(v, X)$ to E' . Then, the subgraph H of G with $E(H) = E'$ and $V(H) = V(E')$ is the required balanced connected subgraph.
- Case 4: $|X| < \frac{k}{2}$ and $|Y| \geq \frac{k}{2}$
 The construction of H is similar to the one in Case 3.

As the above mentioned cases are exhaustive and in each of them, we have found a procedure to find a balanced connected subgraph H (which is guaranteed to exist), it follows that EDGE BALANCED CONNECTED SUBGRAPH is polynomial-time solvable on split graphs. \square

Now, we move on to EDGE BALANCED PATH and show that it is NP-hard on split graphs by giving a polynomial-time reduction from LONGEST PATH. In the LONGEST PATH problem, given a graph G and a positive

integer k , the task is to find a path P in G of length k . It is known that LONGEST PATH is NP-hard [16, ND29] and remains so on split graphs even when the starting vertex u_0 of the path is given as part of the input [16, GT39]. The reduction may be viewed as attaching a red path of length k (consisting of new internal vertices) starting from u_0 to the graph G (whose edges are colored blue) of the LONGEST PATH instance. Then, by adding certain additional edges (colored blue), we get a red-blue split graph.

Theorem 4. *EDGE BALANCED PATH is NP-complete on split graphs.*

Proof. It is easy to verify that this problem is in NP. Let (G, u_0, k) be an instance of LONGEST PATH where G is a split graph whose vertex set is partitioned into clique C and independent set I . Suppose $u_0 \in C$ (the construction when $u_0 \in I$ is similar). We construct an instance $(H, 2k)$ of EDGE BALANCED PATH that is equivalent to (G, u_0, k) . The graph H is obtained from G as follows. Initialize $V(H) = V(G)$ and $E(H) = E(G)$. Add new vertices $\{u_1, u_2, \dots, u_k\}$ that form the path $P^* = (u_1, u_2, \dots, u_k)$ to G such that u_1 is adjacent to u_0 . Let $S = \{u_{2i} \mid 1 \leq i \leq \lfloor \frac{k}{2} \rfloor\}$. Add edges between every pair of vertices in S . Also, make every vertex in C adjacent to every vertex in S . Observe that $V(H)$ is partitioned into clique $C \cup S$ and independent set $I \cup (V(P^*) \setminus S)$. Color $E(H)$ such that all edges in $(E(H) \cap E(P^*)) \cup \{u_0, u_1\}$ are red and the remaining edges are blue. This completes the construction of H . Let E_R denote the set of red edges and E_B denote the set of blue edges in H . Observe that $|E_R| = k$.

Suppose (G, u_0, k) is a YES-instance. Then there exists a path $P = (u_0, x_1, x_2, \dots, x_k)$ in G of length k starting with vertex u_0 . Now, the path $P' = (u_k, \dots, u_1, u_0, x_1, x_2, \dots, x_k)$ is a balanced path in H on $2k$ vertices. Thus $(H, 2k)$ is a YES-instance. Conversely, suppose $(H, 2k)$ is a YES-instance. Let P' be a path in H that has k red edges and k blue edges. Observe that P^* is a subpath of P' as $|E_R| = k$. Then, the graph obtained by deleting $V(P^*)$ from P' is a path P in G starting at u_0 . Further, P has k edges. Thus (G, u_0, k) is a YES-instance. \square

As LONGEST PATH parameterized by the solution size (i.e., the size of the path) in general graphs does not admit a polynomial kernel [2, 20] and the reduction described (which is adaptable for general graphs) is a polynomial parameter transformation, it follows that EDGE BALANCED PATH does not admit polynomial kernels. Further, it is known that, assuming the Exponential Time Hypothesis, LONGEST PATH has no subexponential FPT

algorithm in general graphs. Hence, EDGE BALANCED PATH also does not admit subexponential FPT algorithms.

4. Small Balanced Paths, Trees and Connected Subgraphs

In this section, we prove the combinatorial result that if a graph has a balanced connected subgraph/tree/path of size at least k , then it has one of size at least k and at most $f(k)$ where f is a linear function. We begin with balanced paths.

Theorem 5. *Let G be a red-blue graph and $k \geq 2$ be a positive integer. Then, if G has a balanced path of length at least $2k$, then G has a smaller balanced path of length at least k .*

Proof. Let E_B be the set of blue edges and E_R be the set of red edges in G . Consider a balanced path P in G with at least $2k$ edges. If the terminal edges e and e' are of different colors, then delete e and e' to get a smaller path of length at least k . Otherwise, let $P = (v_1, v_2, \dots, v_\ell)$ where e_i denotes the edge $\{v_i, v_{i+1}\}$ for each $i \in [\ell - 1]$. Without loss of generality, let $e_1, e_{\ell-1} \in E_R$. Define the function $h : E(P) \rightarrow \mathbb{N}$ as follows.

$$h(e_i) = \begin{cases} 1, & \text{if } i = 1 \\ h(e_{i-1}) + 1, & \text{if } i > 1 \text{ and } e_i \in E_R \\ h(e_{i-1}) - 1, & \text{if } i > 1 \text{ and } e_i \in E_B \end{cases}$$

Let E' be the edges e in P with $h(e) = 0$. Clearly, $|E'| \geq 1$ as P is balanced and so $h(e_{\ell-1}) = 0$. We claim that $|E'| > 1$. Suppose $|E'| = 1$. Then, as $e_{\ell-1} \in E_R$, $h(e_{\ell-1}) = h(e_{\ell-2}) + 1$ implying that $h(e_{\ell-1}) = -1$. As $h(e_1) = 1$ and $h(e_{\ell-1}) = -1$, it follows that there is an edge e_i with $i \neq \ell$ and $h(e_i) = 0$. This contradicts that $|E'| = 1$. Hence, $|E'| > 1$. Let e_i denote an edge with $i < \ell - 1$ and $h(e_i) = 0$. Then, the subpaths P_1 and P_2 with $E(P_1) = \{e_1, \dots, e_i\}$ and $E(P_2) = \{e_{i+1}, \dots, e_{\ell-1}\}$ are two balanced paths strictly smaller than P . Further, as $|E(P)| \geq 2k$, either $|E(P_1)| \geq k$ or $|E(P_2)| \geq k$. \square

Now, we move to the analogous result for balanced trees.

Theorem 6. *Let G be a red-blue graph and $k \geq 2$ be a positive integer. Then, if G has a balanced tree with at least $3k + 2$ edges, then G has a smaller balanced tree with at least k edges.*

Proof. Let E_B be the set of blue edges and E_R be the set of red edges in G . Consider a balanced tree T in G with at least $3k + 2$ edges. If T is a path, then by Theorem 5, we obtain the desired smaller tree which is a path. If T has pendant edges e and e' of different colors, then delete e and e' to get a smaller tree on at least k edges. Otherwise, without loss of generality, let all pendant edges of T be in E_R . Let n denote $|V(T)|$. Root T at an arbitrary vertex of degree at least 3. For a vertex $v \in V(T)$, let T_v denote the subtree of T rooted at v . Let u be a vertex with maximum distance from the root such that $|V(T_u)| > \frac{n}{3}$. Let u_1, \dots, u_ℓ be the children of u . Observe that for each $i \in [\ell]$, $|V(T_{u_i})| \leq \frac{n}{3}$. Let i be the least integer in $[\ell]$ such that $\frac{n}{3} \leq |\bigcup_{1 \leq j \leq i} V(T_{u_j})| \leq \frac{2n}{3}$. Let S denote $\bigcup_{1 \leq j \leq i} V(T_{u_j})$ and R denote $V(T) \setminus S$. As $\frac{n}{3} \leq |S| \leq \frac{2n}{3}$, we have $\frac{n}{3} \leq |R| \leq \frac{2n}{3}$. Further, since $n \geq 3k + 3$, we have $\frac{n}{3} \geq k + 1$. Hence, $k + 1 \leq |S|, |R| \leq n - 1$. Consider the following cases.

Case 1: $T[S \cup \{u\}]$ or $T[R]$ is balanced. Then, the lemma holds trivially.

Case 2: $T[S \cup \{u\}]$ or $T[R]$ has more edges from E_R than from E_B . Suppose $T[R]$ has more edges from E_R than from E_B . Initialize T^* to be $T[R]$. As $T[R]$ has at least $k + 1$ vertices (and therefore at least k edges), it has at least $k/2$ edges from E_R . Add the edges of $T[S \cup \{u\}]$ to T^* in the breadth-first order until T^* becomes balanced. We claim that $T \neq T^*$. Suppose $T = T^*$. Then, let e be the edge that was added last to T^* . Clearly, e is a pendant edge and is in E_R . The edge e was added as $T^* - e$ is not balanced. Further, the number of edges in $T^* - e$ from E_B is less than the number of edges from E_R . Otherwise, the addition process would have terminated earlier. However, if $|E(T^* - e) \cap E_B| < |E(T^* - e) \cap E_R|$, $e \in E_R$ and $T = T^*$, then it follows that T is not balanced leading to a contradiction.

Case 3: $T[S \cup \{u\}]$ and $T[R]$ have lesser edges from E_R than from E_B . As $E(T[S \cup \{u\}])$ and $E(T[R])$ partition $E(T)$, this case implies that T has more edges from E_B than from E_R contradicting that T is balanced. \square

Finally, we prove the result for balanced connected subgraphs. For this, we use line graphs, vertex-balanced subgraphs and vertex-balanced trees of vertex-colored graphs.

Theorem 7. *Let G be a red-blue graph and $k \geq 2$ be a positive integer. Then, if G has a balanced connected subgraph with at least $3k + 3$ edges, then G has a smaller balanced connected subgraph with at least k edges.*

Proof. Define a red-blue coloring on $V(L(G))$ as follows: for each vertex x in $L(G)$ corresponding to a red (blue) edge $\{u, v\}$ in G , color x using red (blue).

Suppose G has a balanced connected subgraph G' with $\ell \geq k$ edges. Then, $L(G')$ (which is a subgraph of $L(G)$) is connected and has exactly ℓ vertices with half of them colored red and the rest colored blue. That is, $L(G)$ has a vertex-balanced connected subgraph of order ℓ . Suppose $L(G)$ has a vertex-balanced connected subgraph H on ℓ vertices. Let G' denote the subgraph of G defined as $E' = E(G') = \{\{u, v\} \in E(G) \mid e \in V(H), e = \{u, v\}\}$ and $V(G') = V(E')$. Then, G' is a balanced connected subgraph of G with ℓ edges.

Now, it remains to show that if $L(G)$ has a vertex-balanced connected subgraph (equivalently, a balanced tree T) with at least $3k + 3$ vertices, then $L(G)$ has a smaller vertex-balanced connected subgraph (equivalently, a balanced tree T^*) with at least k vertices. The proof of this claim is similar to the proof of Theorem 6. \square

Observe that due to Theorems 5, 6 and 7, it suffices to give FPT algorithms for EXACT EDGE BALANCED CONNECTED SUBGRAPH/TREE/PATH in order to obtain FPT algorithms for EDGE BALANCED CONNECTED SUBGRAPH/TREE/PATH.

5. FPT Algorithms

We now describe parameterized algorithms for EXACT EDGE BALANCED CONNECTED SUBGRAPH/TREE/PATH.

5.1. Randomized Algorithms

In this section, we show that EXACT EDGE BALANCED CONNECTED SUBGRAPH, EXACT EDGE BALANCED PATH and EXACT EDGE BALANCED TREE admit randomized algorithms that runs in $\mathcal{O}^*(2^k)$ time. We do so by reducing the problems to MULTILINEAR MONOMIAL DETECTION. In order to define this problem, we state some terminology related to polynomials from [3]. Let X denote a set of variables. A *monomial* of degree d is a product of d variables from X , with multiplication assumed to be commutative. A monomial is called *multilinear* if no variable appears twice or more in the product. A *polynomial* $P(X)$ over \mathbb{Z}_+ is a linear combination of monomials with coefficients from \mathbb{Z}_+ . A polynomial contains a certain monomial if the monomial appears with a non-zero coefficient in the linear combination that constitutes the polynomial. Polynomials can be represented as *arithmetic circuits* which in turn can be represented as *directed acyclic graphs*. In the

MULTILINEAR MONOMIAL DETECTION problem, given an arithmetic circuit (represented as a directed acyclic graph) representing a polynomial $P(X)$ over \mathbb{Z}_+ and a positive integer k , the task is to decide whether $P(X)$ contains a multilinear monomial of degree at most k .

Proposition 1. [3, 4, 5] *Let $P(X)$ be a polynomial over \mathbb{Z}_+ represented by a circuit. The MULTILINEAR MONOMIAL DETECTION problem for $P(X)$ can be decided in randomized $\mathcal{O}^*(2^k)$ time and polynomial space.*

Our reductions from EXACT EDGE BALANCED CONNECTED SUBGRAPH/TREE/PATH to MULTILINEAR MONOMIAL DETECTION crucially use the notions of a *color-preserving homomorphism* (also known as an edge-colored homomorphism in the literature [21]) and *relaxed-subgraphs*.

Definition 1. *Given graphs G and H with red-blue edge colorings $\text{col}_G : E(G) \rightarrow \{\text{red}, \text{blue}\}$ and $\text{col}_H : E(H) \rightarrow \{\text{red}, \text{blue}\}$, a color-preserving homomorphism from H to G is a function $h : V(H) \rightarrow V(G)$ satisfying the following properties.*

- *For each pair $u, v \in V(H)$, if $\{u, v\} \in E(H)$, then $\{h(u), h(v)\} \in E(G)$.*
- *For each edge $\{u, v\}$ in H , $\text{col}_H(\{u, v\}) = \text{col}_G(\{h(u), h(v)\})$.*

Definition 2. *Given a red-blue graph G , a relaxed-subgraph is a pair $S = (H, h)$ where H is a red-blue graph and h is a color-preserving homomorphism from H to G .*

The vertex set of a relaxed-subgraph $S = (H, h)$ is $V(S) = \{h(a) \in V(G) \mid a \in V(H)\}$ and the edge set of S is $\{\{h(a), h(b)\} \in E(G) \mid \{a, b\} \in E(H)\}$. We treat the vertex and edge sets of a relaxed-subgraph as multi-sets. The size of S is the number of edges in H (equivalently, the size of $E(S)$). S is said to be connected if H is connected and S is said to be balanced if H has an equal number of red edges and blue edges. S is said to be a *relaxed-path* if H is a path and a *relaxed-tree* if H is a tree.

Next, we have the following observation that states that relaxed-subgraphs with certain specific properties correspond to balanced connected subgraphs, trees and paths.

Observation 1. *The following hold for a red-blue graph G .*

- G has a balanced connected subgraph of size k if and only if there is a balanced connected relaxed-subgraph S of size k such that $E(S)$ consists of distinct elements.
- G has a balanced path of size k if and only if there is a balanced relaxed-path (P, h) of size k where h is injective.
- G has a balanced tree of size k if and only if there is a balanced relaxed-tree (T, h) of size k where h is injective.

Proof. The forward direction of the three claims are easy to verify. If G has a balanced connected subgraph/tree/path Q of size k , then $S = (Q, h)$ where $h : V(Q) \rightarrow V(G)$ is the identity map is the required balanced connected relaxed-subgraph/tree/path of size k . Consider the converse of the first claim. Suppose there is a balanced connected relaxed-subgraph $S = (H, h)$ of size k such that $E(S)$ consists of distinct elements. Consider the subgraph G' of G with $V(G') = V(S)$ and $E(G') = E(S)$. Then, G' is connected as H is connected and $|E(G')| = |E(S)| = k$ as $E(S)$ consists of distinct elements. Further, since H is balanced, G' is balanced. Hence, G has a balanced connected subgraph of size k . The proofs of other two claims are similar. \square

Now, we are ready to describe the randomized algorithms for EXACT EDGE BALANCED CONNECTED SUBGRAPH/TREE/PATH based on Observation 1. First, we consider EXACT EDGE BALANCED CONNECTED SUBGRAPH.

Theorem 8. EXACT EDGE BALANCED CONNECTED SUBGRAPH *admits a randomized $\mathcal{O}^*(2^k)$ -time algorithm.*

Proof. Consider an instance (G, k) . Let E_R denote the set of red edges and E_B denote the set of blue edges in G . In order to obtain an instance of MULTILINEAR MONOMIAL DETECTION that is equivalent to (G, k) , we will define a polynomial P over the variable set $\{x_e \mid e \in E(G)\}$ satisfying the following properties.

- For each balanced connected relaxed-subgraph $S = (H, h)$ of size k there exists a monomial in P that corresponds to S . We say that a monomial M corresponds to S , if $M = \prod_{e \in E(S)} x_e$.
- Each multilinear monomial in P corresponds to some balanced connected relaxed-subgraph S of size k where $E(S)$ has distinct elements.

If P is such a polynomial, then from Observation 1, G has a balanced connected subgraph of size k if and only if P has a multilinear monomial

of degree k . This way, after the construction of P , we reduce the problem to MULTILINEAR MONOMIAL DETECTION and use Proposition 1. In order to construct P , we first construct polynomials $P_j(e, r, b)$ for each $e \in E(G)$, $j \in [k]$ and $0 \leq r, b \leq \frac{k}{2}$ with $r + b \geq 1$. Monomials of $P_j(e, r, b)$ will correspond to connected relaxed-subgraphs $S = (H, h)$ of size j such that H has r red edges, b blue edges and $e \in E(S)$. The construction of $P_j(e, r, b)$ is as follows. For an edge $e \in E(G)$,

$$P_1(e, 1, 0) = \begin{cases} x_e, & \text{if } e \in E_R \\ 0 & \text{otherwise.} \end{cases} \text{ and } P_1(e, 0, 1) = \begin{cases} x_e, & \text{if } e \in E_B \\ 0 & \text{otherwise.} \end{cases}$$

Also, $P_j(e, r, b) = 0$ if $j \neq r + b$. Now, if $e \in E_R$ and $r + b > 1$, then we have

$$P_j(e, r, b) = \sum_{\substack{e' \in N_G(e), \ell < j \\ r' + r'' = r, b' + b'' = b}} P_\ell(e', r', b') P_{j-\ell}(e, r'', b'') + \sum_{e' \in N_G(e)} x_e P_{j-1}(e', r-1, b)$$

and if $e \in E_B$ and $r + b > 1$, then we have

$$P_j(e, r, b) = \sum_{\substack{e' \in N_G(e), \ell < j \\ r' + r'' = r, b' + b'' = b}} P_\ell(e', r', b') P_{j-\ell}(e, r'', b'') + \sum_{e' \in N_G(e)} x_e P_{j-1}(e', r, b-1).$$

We now show that every multilinear monomial of $P_j(e, r, b)$ corresponds to a connected relaxed-subgraph $S = (H, h)$ of size j such that H has r red edges and b blue edges with $E(S)$ consisting of distinct elements where $e \in E(S)$. We prove this claim by induction on j . The base case is easy to verify. Consider the induction step. Suppose $e = \{u, v\} \in E_R$ (the other case is symmetric). Let M be a multilinear monomial of $P_j(e, r, b)$ where $j > 1$.

- Case 1: $M = x_e M'$ where M' is a multilinear monomial of $P_{j-1}(e', r-1, b)$ such that $e' \in N_G(e)$. Let $e' = \{v, w\}$. By induction, M' corresponds to a connected relaxed-subgraph $S' = (H', h')$ of size $j-1$ such that $e' \in E(S')$. Further, H' has $r-1$ red edges and b blue edges. Also, $v \in V(h'(H'))$. Let $z = h'^{-1}(v)$. Observe that z is well-defined due to the multilinearity of M . Note that $E(S')$ consists of distinct elements and $e \notin E(S')$. Let H denote the graph obtained from H' by adding a new vertex z' adjacent to z with the edge $\{z, z'\}$ colored red. Let $h : V(H) \rightarrow V(G)$ denote the homomorphism obtained from h' by extending its domain to include z' and setting $h(z') = u$. Then, $S = (H, h)$ is a connected relaxed-subgraph that M corresponds to.

- Case 2: $M = M_1 M_2$ where M_1 is a multilinear monomial of $P_{j_1}(e', r', b')$ and M_2 is a multilinear monomial of $P_{j_2}(e, r'', b'')$ such that $e' \in N_G(e)$, $j_1, j_2 < j$, $r'' \leq r$ and $b'' \leq b$. Let $e' = \{v, w\}$. By induction, M_1 corresponds to a connected relaxed-subgraph $S_1 = (H_1, h_1)$ of size j_1 such that $e' \in E(S_1)$. Similarly, M_2 corresponds to a connected relaxed-subgraph $S_2 = (H_2, h_2)$ of size j_2 such that $e \in E(S_2)$. Further, H_1 has r' red edges, b' blue edges and H_2 has r'' red edges and b'' blue edges. Also, $v \in V(h_1(H_1)) \cap V(h_2(H_2))$ and $E(S_1) \cap E(S_2) = \emptyset$. Without loss of generality, assume that $V(H_1) \cap V(H_2) = \emptyset$ as this can be achieved by a renaming procedure. Let $z_1 = h_1^{-1}(v)$ and $z_2 = h_2^{-1}(v)$. Observe that z_1 and z_2 are well-defined due to the multilinearity of M_1 and M_2 . Now, rename z_1 in S_1 and z_2 in S_2 as z . Let H denote the graph with vertex set $V(H_1) \cup V(H_2)$ and edge set $E(H_1) \cup E(H_2)$. Observe that H is a connected graph. Let $h : V(H) \rightarrow V(G)$ denote the homomorphism obtained from h_1 and h_2 by extending the domain to the union of the domains of h_1 and h_2 . Then, $S = (H, h)$ is a connected relaxed-subgraph that M corresponds to.

We next show that if there is a connected relaxed-subgraph $S = (H, h)$ of size j with r red edges, b blue edges and such that $e = \{u, v\} \in E(S)$, then there is a monomial of $P_j(e, r, b)$ that corresponds to it. We show this again using induction on j . Suppose $e \in E_R$ (the other case is symmetric). The base case is trivial. Consider the induction step ($j \geq 2$). Let $a = h^{-1}(u)$, $b = h^{-1}(v)$ and z denote the edge $\{a, b\}$ of H .

- Case 1: $H - z$ is connected. Then, $S' = (H - z, h)$ is a connected relaxed-subgraph of size $j - 1$ with $r - 1$ red edges and b blue edges and contains an edge $e' \in N_G(e)$. By induction, there is a monomial M' corresponding to S' in $P_{j-1}(e', r - 1, b)$. Then, the monomial $M = x_e M'$ which is in $P_j(e, r, b)$ corresponds to S .
- Case 2: $H - z$ is disconnected. Then H has two components H_a (containing a) and H_b (containing b). Without loss of generality let H_a have at least one edge. Let H'_b denote the subgraph of H obtained from H_b by adding the vertex a and edge $\{a, b\}$. Let j_1 and j_2 denote the number of edges in H_a and H'_b , respectively. Let r' and r'' be the number of red edges in H_a and H'_b , respectively. Similarly, let b' and b'' be the number of blue edges in H_a and H'_b , respectively. Then $j = j_1 + j_2$, $r = r' + r''$, $b = b' + b''$. Let h_a and h_b denote the color-preserving homomorphism obtained from h by restricting the domain to $V(H_a)$ and $V(H'_b)$, respectively. Now, $S_1 = (H_a, h_a)$ and $S_2 = (H'_b, h_b)$

are connected relaxed-subgraphs with the following properties.

- (i) The size of S_1 is j_1 and S_1 has r' red edges, b' blue edges and there is an edge $e' \in E(S_1)$ incident on $h_a(a)$.
- (ii) The size of S_2 is j_2 and S_2 has r'' red edges, b'' blue edges and $e \in E(S_2)$.

By induction, there is a monomial M_1 in $P_{j_1}(e', r', b')$ that corresponds to S_1 and there is a monomial M_2 in $P_{j_2}(e, r'', b'')$ that corresponds to S_2 . Then, the monomial $M_1 M_2$ which is in $P_j(e, r, b)$ corresponds to S .

Finally, let $P = \sum_{e \in E(G)} P_k(e, \frac{k}{2}, \frac{k}{2})$. Every monomial in P has degree k . Then

from the arguments above, P is the desired polynomial. To compute P we need to compute $P_j(e, r, b)$ for each $j, r, b \in [k]$ and each edge $e \in E(G)$. As these polynomials can be represented as a polynomial-sized arithmetic circuit, the reduction runs in polynomial time. \square

Next, we move on to EXACT EDGE BALANCED TREE.

Theorem 9. EXACT EDGE BALANCED TREE *admits a randomized $\mathcal{O}^*(2^k)$ -time algorithm.*

Proof. Consider an instance (G, k) . Let E_R denote the set of red edges and E_B denote the set of blue edges in G . In order to obtain an instance of MULTILINEAR MONOMIAL DETECTION that is equivalent to (G, k) , we will define a polynomial P over the variable set $\{y_v : v \in V(G)\}$ satisfying the following properties.

- For each balanced relaxed-tree $X = (T, h)$ of size k there exists a monomial in P that corresponds to X . We say that a monomial M corresponds to X , if $M = \prod_{v \in V(X)} y_v$.
- For each multilinear monomial M in P , there is a balanced relaxed-tree $X = (Q, h)$ of size k where h is injective that M corresponds to.

If P is such a polynomial, then from Observation 1, G has a balanced tree of size k if and only if P has a multilinear monomial of degree $k + 1$. This way, after the construction of P , we reduce the problem to MULTILINEAR MONOMIAL DETECTION and use Proposition 1. In order to construct P , we first construct polynomials $P_j(e, r, b)$ for each $e \in E(G)$, $j \in [k]$ and $0 \leq r, b \leq \frac{k}{2}$ with $r + b \geq 1$. Monomials of $P_j(e, r, b)$ will correspond to relaxed-trees $X = (T, h)$ of size j such that T has r red edges, b blue edges and $e \in E(X)$. The construction of $P_j(e, r, b)$ is as follows. For an edge

$e \in E(G)$, we have $P_1(e, 1, 0) = \begin{cases} y_u y_v, & \text{if } e = \{u, v\} \in E_R \\ 0 & \text{otherwise.} \end{cases}$ and similarly

$$P_1(e, 0, 1) = \begin{cases} y_u y_v, & \text{if } e = \{u, v\} \in E_B \\ 0 & \text{otherwise.} \end{cases}$$

Also, $P_j(e, r, b) = 0$ if $j \neq r + b$. Now, for $j > 1$, if $e = \{u, v\} \in E_R$, then

$$\begin{aligned} P_j(e, r, b) = & \sum_{\substack{e', e'' \in N_G(e) \\ u \in V(e'), v \in V(e'') \\ r' < r, b' \leq b, \ell < j}} P_\ell(e', r', b') P_{j-1-\ell}(e'', r-1-r', b-b') \\ & + \sum_{\substack{e' \in N_G(e) \\ v \in V(e')}} y_u P_{j-1}(e', r-1, b) + \sum_{\substack{e' \in N_G(e) \\ u \in V(e')}} y_v P_{j-1}(e', r-1, b). \end{aligned}$$

If $e \in E_B$, then we have

$$\begin{aligned} P_j(e, r, b) = & \sum_{\substack{e', e'' \in N_G(e) \\ u \in V(e'), v \in V(e'') \\ r' < r, b' \leq b, \ell < j}} P_\ell(e', r', b') P_{j-1-\ell}(e'', r-r', b-1-b') \\ & + \sum_{\substack{e' \in N_G(e) \\ v \in V(e')}} y_u P_{j-1}(e', r, b-1) + \sum_{\substack{e' \in N_G(e) \\ u \in V(e')}} y_v P_{j-1}(e', r, b-1). \end{aligned}$$

We now show that every multilinear monomial of $P_j(e, r, b)$ corresponds to a relaxed-tree $X = (T, h)$ of size j such that the T has r red edges and b blue edges where h is injective. We prove this claim by induction on j . The base case is easy to verify. Consider the induction step. Suppose $e = \{u, v\} \in E_R$ (the other case is symmetric). Let M be a multilinear monomial of $P_j(e, r, b)$ where $j > 1$.

- Case 1: $M = y_u M'$ where M' is a multilinear monomial of $P_{j-1}(e', r-1, b)$ such that $e' \in N_G(e)$. Let $e' = \{v, w\}$. By induction, M' corresponds to a relaxed-tree $X' = (T', h')$ of size $j-1$ such that $e' \in E(X')$. Further, T' has $r-1$ red edges and b blue edges. Also, $v \in V(h'(T'))$ (by definition) and $u \notin V(h'(T'))$ (due to the multilinearity of M). Let z' denote $h'(v)$. Let T denote the tree obtained from T' by adding new vertex z adjacent to z' and the edge $\{z, z'\}$ colored red. Let h be the homomorphism obtained from h' by extending the domain to include z with $h(z) = u$. Then, $X = (T, h)$ is the relaxed-tree with the desired properties that M corresponds to.

- Case 2: $M = M_1 M_2$ where M_1 is a multilinear monomial of $P_{j_1}(e', r', b')$ and M_2 is a multilinear monomial of $P_{j_2}(e'', r'', b'')$ such that $e', e'' \in N_G(e)$, $j_1, j_2 < j$, $r'' \leq r-1$ and $b'' \leq b$. Let $e' = \{u, t\}$ and $e'' = \{v, w\}$. By induction, M_1 corresponds to a relaxed-tree $X_1 = (T_1, h_1)$ of size j_1 such that $e' \in E(X_1)$. Further, T_1 has r' red edges and b' blue edges. Also, $u \in V(h_1(T_1))$. Similarly, M_2 corresponds to a relaxed-tree $X_2 = (T_2, h_2)$ of size j_2 such that $e \in E(X_2)$. Further, T_2 has r'' red edges and b'' blue edges. Also, $v \in V(h_2(X_2))$. Without loss of generality assume $V(T_1) \cap V(T_2) = \emptyset$ (one can achieve this property by renaming vertices). Let T denote the graph with vertex set $V(T_1) \cup V(T_2)$ and edge set $E(T_1) \cup E(T_2) \cup \{h_1^{-1}(u), h_2^{-1}(v)\}$. Let $h : V(T) \rightarrow V(G)$ denote the homomorphism obtained from h_1 and h_2 by extending the domain to the union of the domains of h_1 and h_2 . Due to the multilinearity of M , it follows that h is injective. Then, $X = (T, h)$ is the required relaxed-tree that M corresponds to.

We next show that if there is a relaxed-tree $X = (T, h)$ of size j with r red edges, b blue edges and such that $e = \{u, v\} \in E(X)$, then there is a monomial of $P_j(e, r, b)$ that corresponds to it. We show this again using induction on j . Suppose $e \in E_R$ (the other case is symmetric). The base case is trivial. Consider the induction step ($j \geq 2$). Let $a = h^{-1}(u)$, $b = h^{-1}(v)$ and z denote the edge $\{a, b\}$ of T . Then $T - z$ has two components T_a (containing a) and T_b (containing b). Let h_a and h_b denote the homomorphism h restricted to domains $V(T_a)$ and $V(T_b)$, respectively. Without loss of generality let T_a have at least one edge. Suppose T_b is edgeless. By induction, let M' be the monomial corresponding to (T_a, h_a) . Then, $y_v M'$ is the monomial in $P_j(e, r, b)$ corresponding to (T, h) . Next, consider the case when T_b is not edgeless. Let j_1 and j_2 denote the number of edges in T_a and T_b , respectively. Let r' and r'' be the number of red edges in T_a and T_b , respectively. Similarly, let b' and b'' be the number of blue edges in T_a and T_b , respectively. Then $j - 1 = j_1 + j_2$, $r - 1 = r' + r''$, $b = b' + b''$. Now, $X_1 = (T_a, h_a)$ and $X_2 = (T_b, h_b)$ are relaxed-trees with the following properties.

- The size of X_1 is j_1 and T_a has r' red edges, b' blue edges and there is an edge $e' \in E(X_1)$ incident on $h_a(a)$.
- The size of X_2 is j_2 and T_b has r'' red edges, b'' blue edges and there is an edge $e'' \in E(X_2)$ incident on $h_b(b)$.

By induction, there is a monomial M_1 in $P_{j_1}(e', r', b')$ that corresponds to X_1 and there is a monomial M_2 in $P_{j_2}(e'', r'', b'')$ that corresponds to X_2 . Then, the monomial $M_1 M_2$ which is in $P_j(e, r, b)$ corresponds to X .

Finally, let $P = \sum_{e \in E(G)} P_k(e, \frac{k}{2}, \frac{k}{2})$. Every monomial in P has degree $k+1$.

Then from the arguments above, P is the desired polynomial. To compute P we need to compute $P_j(e, r, b)$ for each $j, r, b \in [k]$ and each edge $e \in E(G)$. As these polynomials can be represented as a polynomial-sized arithmetic circuit, the reduction runs in polynomial time. \square

Finally, we consider EXACT EDGE BALANCED PATH.

Theorem 10. EXACT EDGE BALANCED PATH *admits a randomized $\mathcal{O}^*(2^k)$ -time algorithm.*

Proof. Consider an instance (G, k) . Let E_R denote the set of red edges and E_B denote the set of blue edges in G . We will define a polynomial P over the variable set $\{y_v : v \in V(G)\}$ satisfying the following properties.

- For each balanced relaxed-path X of size k there exists a monomial in P that corresponds to X . We say that the monomial M corresponds to X , if $M = \prod_{v \in V(X)} y_v$.
- For each multilinear monomial M in P , there is a balanced relaxed-path $X = (Q, h)$ of size k where h is injective that M corresponds to.

If P is such a polynomial, then from Observation 1, G has a balanced path of size k if and only if P has a multilinear monomial of degree $k+1$. This way, after the construction of P , we reduce the problem to the MULTILINEAR MONOMIAL DETECTION and use Proposition 1. In order to construct P , we first construct polynomials $P_j(v, r, b)$ for each $v \in V(G)$, $j \in [k]$ and $0 \leq r, b \leq \frac{k}{2}$. Monomials of $P_j(v, r, b)$ will correspond to relaxed-paths $X = (Q, h)$ of size j such that Q has r red edges and b blue edges and $v = h(u_1)$, where u_1 is the first vertex of Q . The construction of $P_j(v, r, b)$ is as follows. For a vertex $v \in V(G)$, $P_0(v, 0, 0) = y_v$. Also, $P_j(v, r, b) = 0$ if $j \neq r + b$. For non-negative integers r, b with $r + b \geq 1$ and a positive integer j , we have

$$P_j(v, r, b) = \sum_{\substack{u \in N(v) \\ \{u, v\} \in E_R}} y_v P_{j-1}(u, r-1, b) + \sum_{\substack{u \in N(v) \\ \{u, v\} \in E_B}} y_v P_{j-1}(u, r, b-1).$$

We now show that every multilinear monomial of $P_j(v, r, b)$ corresponds to a relaxed-path (Q, h) of size j such that Q has r red edges, b blue edges and v is the vertex of G onto which the first vertex of Q is mapped under h

and h is injective. We show this using induction on j . The base case is easy to verify. Consider the induction step. Let M be a multilinear monomial of $P_j(v, r, b)$. Then, $M = y_v M'$ where M' is a multilinear monomial of $P_{j-1}(u, r-1, b)$. By induction hypothesis, M' corresponds to a relaxed-path $X' = (Q', h')$ such that h' is injective with $h'(z') = u$ where z' is the starting vertex of Q' . Further, Q' has $r-1$ red edges, b blue edges or r red edges, $b-1$ blue edges. Let Q denote the path obtained from Q' by adding a new vertex z adjacent to z' with the edge $\{z, z'\}$ colored red (in the former case) or blue (in the latter case). Let h be the homomorphism obtained from h' by extending the domain to include z with $h(z) = v$. Clearly, h is injective as h' is injective and $v \notin V(Q')$ (due to the multilinearity of M). Now, $y_v M'$ corresponds to the relaxed-path $X = (Q, h)$. And, X is of size j such that Q has r red edges, b blue edges and v is the vertex of G onto which the first vertex of Q is mapped under h .

We next show that if there is a relaxed-path $X = (Q, h)$ of size j with r red edges, b blue edges such that v is the first vertex of this relaxed path, then there is a monomial of $P_j(v, r, b)$ that corresponds to it. We show this again using induction on j . Let $\{a, b\}$ be the first edge of Q . Let $v = h(a)$ and $u = h(b)$. Suppose $\{v, u\} \in E_R$ (the other case is symmetric). Consider $X' = (Q', h')$ where $V(Q') = V(Q) \setminus \{v\}$, $E(Q') = E(Q) \setminus \{\{a, b\}\}$ and h' is h with domain restricted to $V(Q')$. Then, X' is a relaxed-path with $r-1$ red and b blue edges. By induction there is a monomial M corresponding to it in the $P_{j-1}(u, r-1, b)$. Thus, $y_v M$ corresponds to X .

Finally, let $P = \sum_{v \in V(G)} P_k(v, \frac{k}{2}, \frac{k}{2})$. Every monomial in P has degree $k +$

1. Then from the arguments above, P is the desired polynomial. As the polynomial P can be represented as a polynomial-sized arithmetic circuit, the reduction runs in polynomial time. \square

5.2. Deterministic Algorithms

We first describe deterministic algorithms for EXACT EDGE BALANCED CONNECTED SUBGRAPH and EXACT EDGE BALANCED TREE using the color-coding technique [2, 7, 22].

Consider an instance (G, k) of EXACT EDGE BALANCED CONNECTED SUBGRAPH/TREE. Let $m = |E(G)|$ and $n = |V(G)|$. Let E_R denote the set of red edges and E_B denote the set of blue edges in G . Let $\sigma : E(G) \rightarrow [k]$ be a coloring of edges of G and $\tau : V(G) \rightarrow [k+1]$ be a coloring of vertices of G . We now define L -edge-colorful subgraphs and L -vertex-colorful subgraphs. For

$L \subseteq [k+1]$, a subgraph $H \subseteq G$ is said to be L -edge-colorful if $|E(H)| = |L|$, $\bigcup_{e \in E(H)} \sigma(e) = L$ and for every $e \neq e' \in E(H)$, $\sigma(e) \neq \sigma(e')$. Similarly, H is said to be L -vertex-colorful if $|V(H)| = |L|$, $\bigcup_{v \in V(H)} \tau(v) = L$ and for every $u \neq v \in V(H)$, $\tau(u) \neq \tau(v)$. We describe dynamic programming algorithms to find a $[k]$ -edge-colorful balanced connected subgraph and a $[k+1]$ -vertex-colorful balanced tree in G (if they exist) in $\mathcal{O}^*(4^k)$ time.

Lemma 1. *If a red-blue graph G associated with coloring $\sigma : E(G) \rightarrow [k]$ has a $[k]$ -colorful balanced connected subgraph then such a subgraph can be obtained in $\mathcal{O}^*(4^k)$ time.*

Proof. For $L \subseteq [k]$, $e \in E(G)$ and $r, b \leq \frac{k}{2}$, define $\Lambda(e, L, r, b)$ to be 1 if there is a connected L -edge-colorful subgraph H of G containing e such that $|E(H) \cap E_R| = r$, $|E(H) \cap E_B| = b$ and 0 otherwise. Clearly, G has a $[k]$ -edge-colorful balanced connected subgraph if and only if there is an edge $e \in E(G)$ such that $\Lambda(e, [k], \frac{k}{2}, \frac{k}{2}) = 1$. First observe that for $L \subseteq [k]$, $|L| = 1$ ($L = \{i\}$), $\Lambda(e, i, 1, 0) = 1$ if e is in E_R and 0 otherwise. Similarly, $\Lambda(e, i, 0, 1) = 1$ if e is in E_B and 0 otherwise. Therefore, the entries $\Lambda(e, i, r, b)$ for $i \in [k]$, $e \in E(G)$, $r + b \leq 1$ can be filled in polynomial time. Also, $\Lambda(e, L, r, b) = 0$ if $|L| \neq r + b$ or $r = 0$ with $e \in E_R$ or $b = 0$ with $e \in E_B$.

Next, consider $\Lambda(e, L, r, b)$ where $|L| > 1$ (and $r + b > 1$). We claim that $\Lambda(e, L, r, b) = 1$ if and only if one of the following is true. Let $e = \{u, v\}$.

- $\Lambda(e', L \setminus \sigma(\{u, v\}), r - 1, b) = 1$ where $e' \in N(e)$ and $e \in E_R$.
- $\Lambda(e', L \setminus \sigma(\{u, v\}), r, b - 1) = 1$ where $e' \in N(e)$ and $e \in E_B$.
- $\Lambda(e', L', r', b') = 1$ and $\Lambda(e'', L \setminus (L' \cup \sigma(\{u, v\})), r - r' - 1, b - b') = 1$ where $e', e'' \in N(e)$, $L' \subseteq L \setminus \sigma(\{u, v\})$, $r' \leq r - 1$, $b' \leq b$ and $e \in E_R$.
- $\Lambda(e', L', r', b') = 1$ and $\Lambda(e'', L \setminus (L' \cup \sigma(\{u, v\})), r - r', b - b' - 1) = 1$ where $e', e'' \in N(e)$, $L' \subseteq L \setminus \sigma(\{u, v\})$, $r' \leq r$, $b' \leq b - 1$ and $e \in E_B$.

Observe that if H is a solution to one of the above, then adding e to H is required subgraph that makes $\Lambda(e, L, r, b)$ as 1. Conversely, suppose $\Lambda(e, L, r, b) = 1$. Then G has a L -colorful connected subgraph H such that $e \in E(H)$, $|E(H) \cap E_R| = r$ and $|E(H) \cap E_B| = b$. Without loss of generality, let $e = \{u, v\} \in E_R$. Consider the graph H' obtained from H by deleting e . Then, H' has at most two components. If H' is connected, then either $\Lambda(e', L \setminus \sigma(\{u, v\}), r - 1, b) = 1$ or $\Lambda(e', L \setminus \sigma(\{u, v\}), r, b - 1) = 1$ for some edge $e' \in E(H') \cap N(e)$. Otherwise, H has two components and let H_u be the one containing u and H_v be the one containing v . If H_u has no edge, then $\Lambda(e', L \setminus \sigma(e), r - 1, b) = 1$ for some $e' \in N(e) \cap E(H_v)$. Similarly, if H_v has no

edge, then $\Lambda(e', L \setminus \sigma(e), r-1, b) = 1$ for some $e' \in N(e) \cap E(H_u)$. Otherwise, let L' denote the set of colors $\sigma(e)$ of edges $e \in H_u$. Then $L \setminus (L' \cup \sigma(\{u, v\}))$ is the set of colors of edges in H_v . Also $\Lambda(e', L', r', b') = 1$ where $r' = |E_R \cap E(H_u)|$ and $b' = |E_B \cap E(H_u)|$ and $\Lambda(e'', L \setminus (L' \cup \sigma(e)), r-r'-1, b-b') = 1$.

The time taken to compute an entry $\Lambda(e, L, r, b)$ is mk^22^k as we need to go over choices e', r', b' and $L' \subseteq L$ where $|L| \leq k$, $|E(G)| = m$ and $r', b' \leq \frac{k}{2}$. As the number of entries $\Lambda(e, L, r, b)$ is at most mk^22^k , the running time of the algorithm is $\mathcal{O}^*(4^k)$. \square

Standard derandomization techniques using *perfect hash families* [2, 7, 22] leads to the following result.

Theorem 11. EXACT EDGE BALANCED CONNECTED SUBGRAPH can be solved in $\mathcal{O}^*((4e)^k)$ time.

Proof. First, we use the results of [2, 7, 22] to construct a family $\mathcal{F}_{m,k}$ of coloring functions $\sigma : E(G) \rightarrow [k]$ of size $e^k k^{\mathcal{O}(\log k)} \log m$ in $e^k k^{\mathcal{O}(\log k)} m \log m$ time satisfying the following property: for every set $E' \subseteq E(G)$ of size k , there is a function $\sigma \in \mathcal{F}_{m,k}$ such that $\sigma(e) \neq \sigma(e')$ for any two distinct edges $e, e' \in E'$. Now, for each function in $\mathcal{F}_{m,k}$, we use the dynamic programming algorithm given in Lemma 1 to find a $[k]$ -edge-colorful balanced connected subgraph, if one exists. The properties of $\mathcal{F}_{m,k}$ ensure that, if there exists a balanced connected subgraph $H \subseteq G$ on k edges, then there exists $f \in \mathcal{F}$ that is injective on $E(H)$ and, consequently, the algorithm finds a $[k]$ -edge-colorful balanced connected subgraph. Hence, we obtain a deterministic algorithm which can solve EXACT EDGE BALANCED CONNECTED SUBGRAPH in $\mathcal{O}^*((4e)^k)$ time. \square

Next, we show that EXACT EDGE BALANCED TREE can be solved in $\mathcal{O}^*((4e)^k)$ time.

Lemma 2. Given a red-blue graph G associated with coloring $\tau : V(G) \rightarrow [k+1]$, one can determine if G has a $[k+1]$ -vertex-colorful balanced tree in $\mathcal{O}^*(4^k)$ time.

Proof. For $L \subseteq [k+1]$, $e \in E(G)$ and $r, b \leq \frac{k}{2}$, define $\Lambda(e, L, r, b)$ to be 1 if there is an L -vertex-colorful tree T of G containing e such that $|E(T) \cap E_R| = r$, $|E(T) \cap E_B| = b$ and 0 otherwise. Clearly, G has a $[k+1]$ -vertex-colorful balanced tree if and only if $\exists e \in E(G)$ such that $\Lambda(e, [k+1], \frac{k}{2}, \frac{k}{2}) = 1$. Observe that $\Lambda(e, L, r, b) = 0$ if $|L| \neq r+b+1$ and $\Lambda(e, L, 1, 0)$ is 1 if and only if $L = \{i, j\}$

where $e = \{u, v\} \in E_R$ with $\tau(u) = i$ and $\tau(v) = j$. Similarly, $\Lambda(e, L, 0, 1)$ is 1 if and only if $L = \{i, j\}$ where $e = \{u, v\} \in E_B$ with $\tau(u) = i$ and $\tau(v) = j$.

Consider $\Lambda(e, L, r, b)$ where $r + b \geq 1$. Then, $\Lambda(e, L, r, b) = 1$ if and only if one of the following holds. Let $e = \{u, v\}$.

- $\Lambda(e', L \setminus \tau(u), r - 1, b) = 1$ where $e' \in N(e)$ and $e \in E_R$.
- $\Lambda(e', L \setminus \tau(v), r - 1, b) = 1$ where $e' \in N(e)$ and $e \in E_R$.
- $\Lambda(e', L \setminus \tau(u), r, b - 1) = 1$ where $e' \in N(e)$ and $e \in E_B$.
- $\Lambda(e', L \setminus \tau(v), r, b - 1) = 1$ where $e' \in N(e)$ and $e \in E_B$.
- $\Lambda(e', L', r', b') = 1$ and $\Lambda(e'', L'', r - r' - 1, b - b') = 1$ where $e', e'' \in N(e)$, $L' \subseteq L \setminus \tau(u)$, $L'' = L \setminus L'$, $r' \leq r - 1$, $b' \leq b$ and $e \in E_R$.
- $\Lambda(e', L', r', b') = 1$ and $\Lambda(e'', L'', r - r' - 1, b - b' - 1) = 1$ where $e', e'' \in N(e)$, $L' \subseteq L \setminus \tau(u)$, $L'' = L \setminus L'$, $r' \leq r$, $b' \leq b - 1$ and $e \in E_B$.

The time taken to compute an entry $\Lambda(e, L, r, b)$ is mk^22^{k+1} as we need to go over choices e', r', b' and $L' \subseteq L$ where $|L| \leq k + 1$, $|E(G)| = m$ and $r', b' \leq \frac{k}{2}$. As the number of entries $\Lambda(e, L, r, b)$ is at most mk^22^{k+1} , the running time of the algorithm for a given T is $\mathcal{O}^*(4^k)$. \square

Similar to Theorem 11, Lemma 2 along with derandomization using perfect hash families lead to the following result.

Theorem 12. EXACT EDGE BALANCED TREE can be solved in $\mathcal{O}^*((4e)^k)$ time.

As the reader would have already observed, a simpler dynamic programming algorithm along with derandomization using perfect hash families results in an algorithm for EXACT EDGE BALANCED PATH that runs in $\mathcal{O}^*((2e)^k)$ time. Subsequently, we describe a faster algorithm using representative sets. We begin with some definitions and results related to representative sets. For a finite set U , let $\binom{U}{p}$ denote the set of all subsets of size p of U . Given two families $\mathcal{S}_1, \mathcal{S}_2 \subseteq 2^U$, the *convolution* of \mathcal{S}_1 and \mathcal{S}_2 is the new family defined as $\mathcal{S}_1 * \mathcal{S}_2 = \{X \cup Y \mid X \in \mathcal{S}_1, Y \in \mathcal{S}_2, X \cap Y = \emptyset\}$.

Definition 3. Let U be a set and $\mathcal{S} \subseteq \binom{U}{p}$. A subfamily $\hat{\mathcal{S}} \subseteq \mathcal{S}$ is said to *q-represent* \mathcal{S} (denoted as $\hat{\mathcal{S}} \subseteq_{rep}^q \mathcal{S}$) if for every set $Y \subseteq U$ of size at most q such that there is a set $X \in \mathcal{S}$ with $X \cap Y = \emptyset$, there is a set $\hat{X} \in \hat{\mathcal{S}}$ with $\hat{X} \cap Y = \emptyset$. If $\hat{\mathcal{S}} \subseteq_{rep}^q \mathcal{S}$, then $\hat{\mathcal{S}}$ is called a *q-representative family* for \mathcal{S} .

Representative families (also called representative sets) are transitive and have nice union and convolution properties (Proposition 2).

Proposition 2. [2, Lemmas 12.26, 12.27 and 12.28] *Let U be a finite set.*

1. *Let $\mathcal{S}_1, \mathcal{S}_2 \subseteq \binom{U}{p}$. If $\widehat{\mathcal{S}}_1 \subseteq_{rep}^q \mathcal{S}_1$ and $\widehat{\mathcal{S}}_2 \subseteq_{rep}^q \mathcal{S}_2$, then $\widehat{\mathcal{S}}_1 \cup \widehat{\mathcal{S}}_2 \subseteq_{rep}^q \mathcal{S}_1 \cup \mathcal{S}_2$.*
2. *Let $\mathcal{S} \subseteq \binom{U}{p}$. If $\widehat{\mathcal{S}} \subseteq_{rep}^q \mathcal{S}'$ and $\mathcal{S}' \subseteq_{rep}^q \mathcal{S}$, then $\widehat{\mathcal{S}} \subseteq_{rep}^q \mathcal{S}$.*
3. *Let $\mathcal{S}_1 \subseteq \binom{U}{p_1}$, $\mathcal{S}_2 \subseteq \binom{U}{p_2}$. If $\widehat{\mathcal{S}}_1 \subseteq_{rep}^{k-p_1} \mathcal{S}_1$ and $\widehat{\mathcal{S}}_2 \subseteq_{rep}^{k-p_2} \mathcal{S}_2$, then $\widehat{\mathcal{S}}_1 * \widehat{\mathcal{S}}_2 \subseteq_{rep}^{k-p_1-p_2} \mathcal{S}_1 * \mathcal{S}_2$.*

A classical result due to Bollobás states that small representative families exist [23] and Proposition 3 [2, 8, 9] shows that such families can be efficiently computed.

Proposition 3. [2, 8, 9] *There is an algorithm that given a family $\mathcal{S} \subseteq \binom{U}{p}$, a rational $0 < x < 1$ and integers $p, k \geq p$, computes $\widehat{\mathcal{S}} \subseteq_{rep}^{k-p} \mathcal{S}$ of size at most $x^{-p}(1-x)^{-(k-p)}2^{o(k)}$ in $|\mathcal{S}|(1-x)^{-(k-p)}2^{o(k)}$ time. Further, if $|\mathcal{S}| = \mathcal{O}^*(x^{-p}(1-x)^{-(k-p)}2^{o(k)})$ and $x = \frac{p}{2k-p}$, then the construction of $\widehat{\mathcal{S}}$ takes $\mathcal{O}^*(\frac{(2k-p)^{2k-p}}{p^p(2k-2p)^{2k-2p}}2^{o(k)})$ time, moreover, this running time is maximized when $p = (1 - \frac{1}{\sqrt{5}})k$ and is $\mathcal{O}^*(\phi^{2k+o(k)})$ (which is $\mathcal{O}^*(2.619^k)$) where ϕ is the golden ratio $\frac{1+\sqrt{5}}{2}$.*

Now, we are ready to describe an algorithm for EXACT EDGE BALANCED PATH using representative sets.

Theorem 13. EXACT EDGE BALANCED PATH *can be solved in $\mathcal{O}^*(2.619^k)$ time.*

Proof. Consider an instance (G, k) . Let E_R and E_B denote the sets of red and blue edges of G . For a pair of vertices $u, v \in V(G)$ and non-negative integers r and b with $r + b \geq 1$, define the family $\mathcal{P}_{uv}^{(r,b)}$ as follows.

$$\mathcal{P}_{uv}^{(r,b)} = \{X \mid X \subseteq V(G), |X| = r+b+1 \text{ and there is a path } P \text{ from } u \text{ to } v \text{ with } V(P) = X, |E_R \cap E(P)| = r \text{ and } |E_B \cap E(P)| = b\}.$$

Now, it suffices to determine if $\mathcal{P}_{uv}^{(\frac{k}{2}, \frac{k}{2})}$ is non-empty for some $u, v \in V(G)$. The families $\mathcal{P}_{uv}^{(r,b)}$ can be computed using the following formula. For $r+b = 1$,

$$\mathcal{P}_{uv}^{(1,0)} = \begin{cases} \{\{u, v\}\}, & \text{if } \{u, v\} \in E_R \\ \emptyset, & \text{otherwise} \end{cases} \quad \text{and} \quad \mathcal{P}_{uv}^{(0,1)} = \begin{cases} \{\{u, v\}\}, & \text{if } \{u, v\} \in E_B \\ \emptyset, & \text{otherwise} \end{cases}$$

and, for $r + b > 1$,

$$\mathcal{P}_{uv}^{(r,b)} = \left(\bigcup_{\{w,v\} \in E_R} (\mathcal{P}_{uw}^{(r-1,b)} * \{\{v\}\}) \right) \bigcup \left(\bigcup_{\{w,v\} \in E_B} (\mathcal{P}_{uw}^{(r,b-1)} * \{\{v\}\}) \right)$$

The base case is easy to verify. Consider a path P such that $V(P) \in \mathcal{P}_{uv}^{(r,b)}$. Then, P has a subpath P' from u to a neighbour w of v . If $\{w, v\} \in E_R$, then P' has $r - 1$ red edges and b blue edges. Further, $V(P') \in \mathcal{P}_{uw}^{(r-1,b)}$ and $V(P) = V(P') \cup \{v\}$ where $v \notin V(P')$. Otherwise, $\{w, v\} \in E_B$ and P' has r red edges and $b - 1$ blue edges. Now, $V(P') \in \mathcal{P}_{uw}^{(r,b-1)}$ and $V(P) = V(P') \cup \{v\}$ where $v \notin V(P')$. On the other hand, for any element $X \in \mathcal{P}_{uw}^{(r,b-1)}$ such that there is a vertex v with $v \notin X$ and $\{w, v\} \in E_B$, we have $X \cup \{v\} \in \mathcal{P}_{uv}^{(r,b)}$. Similarly, for any element $X \in \mathcal{P}_{uw}^{(r-1,b)}$ such that there is a vertex v with $v \notin X$ and $\{w, v\} \in E_R$, we have $X \cup \{v\} \in \mathcal{P}_{uv}^{(r,b)}$. This justifies the formula given for the computation of $\mathcal{P}_{uv}^{(r,b)}$.

Clearly, a naive computation of $\mathcal{P}_{uv}^{(r,b)}$ is not guaranteed to result in an FPT (in k) algorithm. Therefore, instead of computing $\mathcal{P}_{uv}^{(r,b)}$, we only compute $\widehat{\mathcal{P}}_{uv}^{(r,b)} \subseteq_{rep}^{k-(r+b)} \mathcal{P}_{uv}^{(r,b)}$ and use the fact that $\widehat{\mathcal{P}}_{uv}^{(\frac{k}{2}, \frac{k}{2})} \subseteq_{rep}^0 \mathcal{P}_{uv}^{(\frac{k}{2}, \frac{k}{2})}$. If $\widehat{\mathcal{P}}_{uv}^{(\frac{k}{2}, \frac{k}{2})}$ is nonempty, then it contains a set X that is disjoint with \emptyset . As $\widehat{\mathcal{P}}_{uv}^{(\frac{k}{2}, \frac{k}{2})} \subseteq_{rep}^0 \mathcal{P}_{uv}^{(\frac{k}{2}, \frac{k}{2})}$, it follows that $\widehat{\mathcal{P}}_{uv}^{(\frac{k}{2}, \frac{k}{2})}$ also has a set \widehat{X} that is disjoint with \emptyset . In other words, if $\mathcal{P}_{uv}^{(\frac{k}{2}, \frac{k}{2})}$ is nonempty, then $\widehat{\mathcal{P}}_{uv}^{(\frac{k}{2}, \frac{k}{2})}$ is also non-empty.

Now, we describe a dynamic programming algorithm to compute $\widehat{\mathcal{P}}_{uv}^{(\frac{k}{2}, \frac{k}{2})}$ for every $u, v \in V(G)$. For $r + b = 1$, set $\widehat{\mathcal{P}}_{uv}^{(1,0)} = \mathcal{P}_{uv}^{(1,0)}$ and $\widehat{\mathcal{P}}_{uv}^{(0,1)} = \mathcal{P}_{uv}^{(0,1)}$. Clearly, $\widehat{\mathcal{P}}_{uv}^{(1,0)} \subseteq_{rep}^{k-1} \mathcal{P}_{uv}^{(1,0)}$ and $\widehat{\mathcal{P}}_{uv}^{(0,1)} \subseteq_{rep}^{k-1} \mathcal{P}_{uv}^{(0,1)}$. Further, $|\widehat{\mathcal{P}}_{uv}^{(1,0)}|, |\widehat{\mathcal{P}}_{uv}^{(0,1)}| \leq 1$ and this computation is polynomial time. Now, we proceed to computing $\widehat{\mathcal{P}}_{uv}^{(r,b)} \subseteq_{rep}^{k-(r+b)} \mathcal{P}_{uv}^{(r,b)}$ for $k \geq r + b > 1$ in the increasing order of $r + b$. Towards this, we compute a new family $\widetilde{\mathcal{P}}_{uv}^{(r,b)}$ as follows.

$$\widetilde{\mathcal{P}}_{uv}^{(r,b)} = \left(\bigcup_{\{w,v\} \in E_R} (\widehat{\mathcal{P}}_{uw}^{(r-1,b)} * \{\{v\}\}) \right) \bigcup \left(\bigcup_{\{w,v\} \in E_B} \widehat{\mathcal{P}}_{uw}^{(r,b-1)} * \{\{v\}\}) \right)$$

Using the union and convolution properties of Propositions 2, we have the following properties.

- $(\widehat{\mathcal{P}}_{uw}^{(r-1,b)} * \{\{v\}\}) \subseteq_{rep}^{k-(r+b)} (\mathcal{P}_{uw}^{(r-1,b)} * \{\{v\}\})$

- $(\widehat{\mathcal{P}}_{uv}^{(r,b-1)} * \{\{v\}\}) \subseteq_{rep}^{k-(r+b)} (\mathcal{P}_{uv}^{(r,b-1)} * \{\{v\}\})$
- $\bigcup_{\{w,v\} \in E_R} (\widehat{\mathcal{P}}_{uv}^{(r-1,b)} * \{\{v\}\}) \subseteq_{rep}^{k-(r+b)} \bigcup_{\{w,v\} \in E_R} (\mathcal{P}_{uv}^{(r-1,b)} * \{\{v\}\})$
- $\bigcup_{\{w,v\} \in E_B} (\widehat{\mathcal{P}}_{uv}^{(r,b-1)} * \{\{v\}\}) \subseteq_{rep}^{k-(r+b)} \bigcup_{\{w,v\} \in E_B} (\mathcal{P}_{uv}^{(r,b-1)} * \{\{v\}\})$

Now, once again by the union property of Proposition 2, we have $\widetilde{\mathcal{P}}_{uv}^{(r,b)} \subseteq_{rep}^{k-(r+b)} \mathcal{P}_{uv}^{(r,b)}$. Further, $|\widetilde{\mathcal{P}}_{uv}^{(r,b)}| = \mathcal{O}^*(|\widehat{\mathcal{P}}_{uv}^{(r-1,b)}| + |\widehat{\mathcal{P}}_{uv}^{(r,b-1)}|)$. Then, we use Proposition 3 to compute a family $\widehat{\mathcal{P}}_{uv}^{(r,b)} \subseteq_{rep}^{k-(r+b)} \widetilde{\mathcal{P}}_{uv}^{(r,b)}$. By the transitivity property of Proposition 2, it follows that $\widehat{\mathcal{P}}_{uv}^{(r,b)} \subseteq_{rep}^{k-(r+b)} \mathcal{P}_{uv}^{(r,b)}$. The time taken to compute the families $\widehat{\mathcal{P}}_{uv}^{(r,b)}$ for $u, v \in V(G)$ and $r, b \leq \frac{k}{2}$ is $\mathcal{O}^*(2.619^k)$ from Proposition 3 by substituting $r+b+1$ for p and $k+1$ for k . Thus, the overall running time of the algorithm is $\mathcal{O}^*(2.619^k)$. \square

6. Concluding Remarks

To summarize our work, we study the complexity of finding balanced connected subgraphs, trees and paths in red-blue graphs. We give fixed-parameter tractability results using color coding, representative sets and reductions to MULTILINEAR MONOMIAL DETECTION. En route, we give combinatorial results on the existence of small balanced connected subgraphs, trees and paths. We observe that these results also extend to vertex-balanced connected subgraphs, trees and paths. As a result the algorithms described in this work also generalize to solve the vertex-analogue of the problems. Note that using line graphs, one can reduce EDGE BALANCED CONNECTED SUBGRAPH to VERTEX BALANCED CONNECTED SUBGRAPH, however, when the solution is required to be a path or a tree, this reduction is not useful. Determining the complexity of finding other balanced substructures is an interesting direction of research. It is well-known (by a observation made by Erdős) and easy to verify that a monochromatic spanning tree exists in any red-blue complete graph. This fact has been generalized in several directions. Gyárfas [24] and Füredi [25] independently showed that every r -edge-coloring of the complete graph on n vertices results in a monochromatic connected subgraph of size at least $n/(r-1)$. Bollobás and Gyárfas [26] studied monochromatic 2-connected subgraphs and showed that every red-blue complete graph on $n \geq 5$ vertices has a monochromatic 2-connected subgraph with at least $n-2$ vertices. They also dealt with questions on the existence

of monochromatic q -connected subgraphs. Variants such as large monochromatic components of small diameter have also attracted attention recently [27]. A similar combinatorial (and algorithmic) study on balanced spanning trees, balanced spanning connected subgraphs and balanced q -connected subgraphs is worth investigating. Also, studying the problems on graphs that are colored using more than two colors and on colored weighted graphs are next natural questions in this context.

Acknowledgements. The second author is supported by SERB CRG grant number CRG/2022/007751.

References

- [1] Y. Caro, A. Hansberg, J. Lauri, C. Zarb, On zero-sum spanning trees and zero-sum connectivity, *Electron. J. Comb.* 29 (1) (2022). doi:10.37236/10289.
URL <https://doi.org/10.37236/10289>
- [2] M. Cygan, F. V. Fomin, L. Kowalik, D. Lokshtanov, D. Marx, M. Pilipczuk, M. Pilipczuk, S. Saurabh, *Parameterized Algorithms*, Springer, 2015. doi:10.1007/978-3-319-21275-3.
URL <https://doi.org/10.1007/978-3-319-21275-3>
- [3] I. Koutis, R. Williams, LIMITS and applications of group algebras for parameterized problems, *ACM Trans. Algorithms* 12 (3) (2016) 31:1–31:18.
doi:10.1145/2885499.
URL <https://doi.org/10.1145/2885499>
- [4] I. Koutis, Faster algebraic algorithms for path and packing problems, in: L. Aceto, I. Damgård, L. A. Goldberg, M. M. Halldórsson, A. Ingólfssdóttir, I. Walukiewicz (Eds.), *Automata, Languages and Programming, 35th International Colloquium, ICALP 2008, Reykjavik, Iceland, July 7-11, 2008, Proceedings, Part I: Track A: Algorithms, Automata, Complexity, and Games*, Vol. 5125 of *Lecture Notes in Computer Science*, Springer, 2008, pp. 575–586.
doi:10.1007/978-3-540-70575-8_47.
URL https://doi.org/10.1007/978-3-540-70575-8_47

- [5] R. Williams, Finding paths of length k in $O^*(2^k)$ time, Inf. Process. Lett. 109 (6) (2009) 315–318. doi:10.1016/j.ipl.2008.11.004.
URL <https://doi.org/10.1016/j.ipl.2008.11.004>
- [6] J. Nederlof, Fast polynomial-space algorithms using inclusion-exclusion, Algorithmica 65 (4) (2013) 868–884.
doi:10.1007/s00453-012-9630-x.
URL <https://doi.org/10.1007/s00453-012-9630-x>
- [7] N. Alon, R. Yuster, U. Zwick, Color-coding, J. ACM 42 (4) (1995) 844–856. doi:10.1145/210332.210337.
URL <https://doi.org/10.1145/210332.210337>
- [8] F. V. Fomin, D. Lokshtanov, F. Panolan, S. Saurabh, Efficient computation of representative families with applications in parameterized and exact J. ACM 63 (4) (2016) 29:1–29:60. doi:10.1145/2886094.
URL <https://doi.org/10.1145/2886094>
- [9] H. Shachnai, M. Zehavi, Representative families: A unified tradeoff-based approach, J. Comput. Syst. Sci. 82 (3) (2016) 488–502.
doi:10.1016/j.jcss.2015.11.008.
URL <https://doi.org/10.1016/j.jcss.2015.11.008>
- [10] S. Bhore, S. Chakraborty, S. Jana, J. S. B. Mitchell, S. Pandit, S. Roy, The balanced connected subgraph problem, in: S. P. Pal, A. Vijayakumar (Eds.), Algorithms and Discrete Applied Mathematics - 5th International Conference, CALDAM 2019, Kharagpur, India, February 14-16, 2019, Proceedings, Vol. 11394 of Lecture Notes in Computer Science, Springer, 2019, pp. 201–215. doi:10.1007/978-3-030-11509-8_17.
URL https://doi.org/10.1007/978-3-030-11509-8_17
- [11] S. Bhore, S. Jana, S. Pandit, S. Roy, Balanced connected subgraph problem in geometric intersection graphs, in: Y. Li, M. Cardei, Y. Huang (Eds.), Combinatorial Optimization and Applications - 13th International Conference, COCOA 2019, Xiamen, China, December 13-15, 2019, Proceedings, Vol. 11949 of Lecture Notes in Computer Science, Springer, 2019, pp. 56–68. doi:10.1007/978-3-030-36412-0_5.
URL https://doi.org/10.1007/978-3-030-36412-0_5

- [12] B. Darties, R. Giroudeau, J. König, V. Pollet, The balanced connected subgraph problem: Complexity results in bounded-degree and bounded treewidth, in: Y. Li, M. Cardei, Y. Huang (Eds.), Combinatorial Optimization and Applications - 13th International Conference, COCOA 2019, Xiamen, China, December 13-15, 2019, Proceedings, Vol. 11949 of Lecture Notes in Computer Science, Springer, 2019, pp. 449–460. doi:10.1007/978-3-030-36412-0_36. URL https://doi.org/10.1007/978-3-030-36412-0_36
- [13] Y. Kobayashi, K. Kojima, N. Matsubara, T. Sone, A. Yamamoto, Algorithms and hardness results for the maximum balanced connected subgraph problem, in: Y. Li, M. Cardei, Y. Huang (Eds.), Combinatorial Optimization and Applications - 13th International Conference, COCOA 2019, Xiamen, China, December 13-15, 2019, Proceedings, Vol. 11949 of Lecture Notes in Computer Science, Springer, 2019, pp. 303–315. doi:10.1007/978-3-030-36412-0_24. URL https://doi.org/10.1007/978-3-030-36412-0_24
- [14] T. Martinod, V. Pollet, B. Darties, R. Giroudeau, J. König, Complexity and inapproximability results for balanced connected subgraph problem, Theor. Comput. Sci. 886 (2021) 69–83. doi:10.1016/j.tcs.2021.07.010. URL <https://doi.org/10.1016/j.tcs.2021.07.010>
- [15] R. Diestel, Graph Theory, 4th Edition, Vol. 173 of Graduate texts in mathematics, Springer, 2012.
- [16] M. R. Garey, D. S. Johnson, Computers and Intractability: A Guide to the Theory of NP-Completeness, W. H. Freeman, 1979.
- [17] M. Dom, D. Lokshtanov, S. Saurabh, Kernelization lower bounds through colors and ids, ACM Trans. Algorithms 11 (2) (2014) 13:1–13:20. doi:10.1145/2650261. URL <https://doi.org/10.1145/2650261>
- [18] M. R. Garey, D. S. Johnson, The rectilinear steiner tree problem in NP complete, SIAM Journal of Applied Mathematics 32 (1977) 826–834. doi:10.1137/0132071. URL <https://doi.org/10.1137/0132071>

- [19] K. White, M. Farber, W. R. Pulleyblank, Steiner trees, connected domination and strongly chordal graphs, *Networks* 15 (1) (1985) 109–124. doi:10.1002/net.3230150109. URL <https://doi.org/10.1002/net.3230150109>
- [20] H. L. Bodlaender, R. G. Downey, M. R. Fellows, D. Hermelin, On problems without polynomial kernels, *J. Comput. Syst. Sci.* 75 (8) (2009) 423–434. doi:10.1016/j.jcss.2009.04.001. URL <https://doi.org/10.1016/j.jcss.2009.04.001>
- [21] R. C. Brewster, R. Dedic, F. Huard, J. Queen, The recognition of bound quivers using edge-coloured homomorphisms, *Discret. Math.* 297 (1-3) (2005) 13–25. doi:10.1016/j.disc.2004.10.026. URL <https://doi.org/10.1016/j.disc.2004.10.026>
- [22] M. Naor, L. J. Schulman, A. Srinivasan, Splitters and near-optimal derandomization, in: 36th Annual Symposium on Foundations of Computer Science, Milwaukee, Wisconsin, USA, 23-25 October 1995, IEEE Computer Society, 1995, pp. 182–191. doi:10.1109/SFCS.1995.492475. URL <https://doi.org/10.1109/SFCS.1995.492475>
- [23] B. Bollobás, On generalized graphs, *Acta Mathematica Hungarica* 16 (3-4) (1965) 447–452. doi:10.1007/BF01904851. URL <https://doi.org/10.1007/BF01904851>
- [24] A. Gyárfás, Partition coverings and blocking sets of hypergraphs, *Comm. Comput. Automat. Inst. Hungar. Acad. Sci.* 71 (1977) 62.
- [25] Z. Füredi, Maximum degree and fractional matchings in uniform hypergraphs, *Comb.* 1 (2) (1981) 155–162. doi:10.1007/BF02579271. URL <https://doi.org/10.1007/BF02579271>
- [26] B. Bollobás, A. Gyárfás, Highly connected monochromatic subgraphs, *Discret. Math.* 308 (9) (2008) 1722–1725. doi:10.1016/j.disc.2006.01.030. URL <https://doi.org/10.1016/j.disc.2006.01.030>
- [27] E. Carlson, R. R. Martin, B. Peng, M. Ruszinkó, Large monochromatic components of small diameter, *J. Graph Theory*

99 (2) (2022) 247–250. doi:10.1002/jgt.22739.
URL <https://doi.org/10.1002/jgt.22739>