On the Kernelization Complexity of Colorful Motifs

Abhimanyu M. Ambalath[1], Radheshyam Balasundaram[2], Chintan Rao H[3], Venkata Koppula[4], Neeldhara Misra[5], Geevarghese Philip[5], M S Ramanujan[5]

[1] National Institute of Technology, Calicut
[2] Birla Institute of Technology and Science, Pilani
[3] Indian Institute of Science, Bangalore
[4] Indian Institute of Technology, Kanpur
[5] The Institute of Mathematical Sciences, Chennai

abhimanyu.m.a@nitc.ac.in, adyarshyam@gmail.com, chintanraoh@csa.iisc.ernet.in, kvenkata@iitk.ac.in, {neeldhara|gphilip|msramanujan}@imsc.res.in

Abstract. The Colorful Motif problem asks if, given a vertex-colored graph G, there exists a subset S of vertices of G such that the graph induced by G on S is connected and contains every color in the graph exactly once. The problem is motivated by applications in computational biology and is also well-studied from the theoretical point of view. In particular, it is known to be NP-complete even on trees of maximum degree three [Fellows et al, ICALP 2007]. In their pioneering paper that introduced the color-coding technique, Alon et al. [STOC 1995] show, *inter alia*, that the problem is FPT on general graphs. More recently, Cygan et al. [WG 2010] showed that Colorful Motif is NP-complete on *comb graphs*, a special subclass of the set of trees of maximum degree three. They also showed that the problem is not likely to admit polynomial kernels on forests.

We continue the study of the kernelization complexity of the Colorful Motif problem restricted to simple graph classes. Surprisingly, the infeasibility of polynomial kernelization persists even when the input is restricted to comb graphs. We demonstrate this by showing a simple but novel composition algorithm. Further, we show that the problem restricted to comb graphs admits polynomially many polynomial kernels. To our knowledge, there are very few examples of problems with many polynomial kernels known in the literature. We also show hardness of polynomial kernelization for certain variants of the problem on trees; this rules out a general class of approaches for showing many polynomial kernels for the problem restricted to trees. Finally, we show that the problem is unlikely to admit polynomial kernels on another simple graph class, namely the set of all graphs of diameter two. As an application of our results, we settle the classical complexity of Connected Dominating Set on graphs of diameter two — specifically, we show that it is NP-complete.

1 Introduction and Motivation

Algorithms that are designed to reduce the size of an instance in polynomial time are widely referred to as preprocessing algorithms. It is natural to study such algorithms in the context of problems that are NP-hard — preprocessing techniques are used in almost every practical computer implementation that deals with an NP-hard problem. We study *kernelization algorithms* — these are preprocessing algorithms that have provable performance bounds, both in the running time and in the extent of reduction in instance size. In particular, we are interested in polynomial time algorithms that reduce the size of a parameterized problem (cf. Section 2 and [7, 10] for definitions) to an instance whose size is a polynomial in the parameter. Such a reduced instance is called a *polynomial kernel*.

It is natural to examine the possibility of preprocessing strategies when a problem is notoriously intractable. In this work, our interests center around the Colorful Motif problem. The problem is intractable, in the classical sense, even on seemingly "simple" classes of graphs. Using a recent framework for showing lower bounds on polynomial kernelization, we establish that not only are these problems unlikely to admit a polynomial time algorithm that solves them — they are unlikely to admit polynomial time algorithms that reduce them to instances whose size is bounded by a polynomial in the parameter.

The Graph Motif problem concerns a vertex-colored undirected graph G and a *multiset* M of colors. We are asked whether there is a set S of vertices of G such that the subgraph induced on S is connected and there is a color-preserving bijective mapping from S to M. That is, the problem is to find if there is a connected subgraph H of G such that the multiset of colors of H is identical to M.

The Graph Motif problem has immense utility in bioinformatics, especially in the context of metabolic network analysis (eg. motif search in metabolic reaction graphs with vertices representing reactions and edges connecting successive reactions) [3, 14] . The problem is NP-complete even in very restricted cases, such as when G is a tree with maximum degree 3, or when G is a bipartite graph with maximum degree 4 and M is a multiset over just two colors. When parameterized by $|M|$, the problem is FPT, and it is W[2]-hard when parameterized by the *number of colors* in M, even when G is a tree [8].

The Colorful Motif problem is a simpler version of the Graph Motif problem, where M is a set (and not a multiset). Even this problem is NP-hard on simple classes of graphs, such as when G is a tree with maximum degree 3 [8]. The problem is FPT on general graphs when parameterized by $|M|$, and the current fastest FPT algorithm, by Guillemot and Sikora, runs in $\mathcal{O}^*(2^{|M|})$ time[1] and polynomial space [12].

We now turn to an example of a seemingly simple graph class on which the problem continues to be intractable. A graph is called a *comb graph* if (i) it is a tree, (ii) all vertices are of degree at most 3, (iii) all the vertices of degree 3 lie on a single simple path. Cygan et al. [5] recently showed that the problem is NP-hard even on comb graphs. Further, they show that the parameterized version of the problem is unlikely to admit a polynomial kernel on forests unless $NP \subseteq coNP/Poly$, which would in turn imply an unlikely collapse of the Polynomial Hierarchy [4].

We begin by pushing the borders of classical tractability. We show that while the problem is polynomial time on *caterpillars* (trees where the removal of all leaf vertices results in a *path*, called the *spine* of the caterpillar), it is NP-hard on *lobsters* (trees where the removal of all leaf vertices results in a *caterpillar*). In fact, we show that even more is true: the problem is NP-hard even on rooted trees of height two, or equivalently, on trees of diameter at most four.

Next, we extend the known results on the hardness of kernelization for this problem [5]. Specifically, we show that the lower bound can be tightened to hold for comb graphs as well. This is established by demonstrating a simple but unusual composition algorithm for the problem restricted to comb graphs. The composition is unusual because it is not the trivial composition (via disjoint union), and yet, it does not employ gadgets to encode the identity of the instances. To the best of our knowledge, this is an uncommon style of composition. On the positive side, we show a straightforward argument that yields polynomially many polynomial kernels for the problem on comb graphs, *a la* the many polynomial kernels obtained for k-Leaf Out Branching [9]. Again, to the best of our knowledge, this is one of the very few examples of many polynomial kernels for a parameterized problem for which polynomial kernelization is infeasible.

However, in our attempts to obtain many polynomial kernels for the more general case of trees, we learn that some natural approaches fail. Specifically, we show that two natural variants of the problem — Rooted Colorful Motif[2], Subset Colorful Motif[3] — do not admit polynomial kernels unless $NP \subseteq coNP/Poly$. This shows, for instance, that the "guess" for obtaining many polynomial kernels has to be more than, or different from, a subset of vertices.

While we show that Colorful Motif is NP-hard on trees of diameter at most four, the kernelization complexity of the problem on this class of graphs is still open. However, we show that Colorful Motif is NP-hard on general graphs of diameter *three*, and the same reduction also shows that polynomial kernels are unlikely for graphs of diameter three. We employ a reduction from Colorful Motif on general graphs. Using similar techniques, we show that the problem is NP-hard on general graphs of diameter *two*. This turns out to be useful to show the NP-hardness of Connected Dominating Set on the same class of graphs.

The results we obtain in this paper contribute to the rapidly growing collection of problems for which polynomial kernels do not exist under reasonable complexity-theoretic assumptions. Given that many of our results pertain to very special graph classes, we hope these hardness results — which make these special problems available as starting points for further reductions — will be useful in settling the kernelization complexity of many other problems. In fact, we demonstrate the utility of the NP-completeness of Colorful Motif on graphs of diameter two, by showing that Connected Dominating Set on graphs of diameter two is NP-complete. The classical complexity of Connected Dominating Set on graphs of diameter two was hitherto unknown, although it was known to be NP-complete on graphs of diameter three, and trivial on graphs of diameter one. Also, since Colorful Motif is both well-motivated and well-studied, we believe that these results are of independent interest.

[1] Given $f : \mathbb{N} \to \mathbb{N}$, we define $\mathcal{O}^*(f(n))$ to be $O(f(n) \cdot p(n))$, where $p(\cdot)$ is some polynomial function. That is, the $\mathcal{O}^*$ notation suppresses polynomial factors in the expression.

[2] Does there exist a colorful subtree that contains a specific vertex?

[3] Does there exist a colorful subtree that contains a specific subset of vertices?

2 Preliminaries

A parameterized problem is denoted by a pair $(Q, k) \subseteq \Sigma^* \times \mathbb{N}$. The first component Q is a classical language, and the number k is called the parameter. Such a problem is *fixed–parameter tractable* (FPT) if there exists an algorithm that decides it in time $\mathcal{O}(f(k)n^{\mathcal{O}(1)})$ on instances of size n. A *many-to-one kernelization algorithm* (or, simply, a kernelization algorithm) for a parameterized problem takes an instance (x, k) of the problem as input, and in time polynomial in $|x|$ and k, produces an equivalent instance (x', k') such that $|x'|$ is a function purely of k. The output x' is called a *kernel* for the problem instance, and $|x'|$ is called the *size* of the kernel. A kernel is said to be a *polynomial kernel* if its size polynomial in the parameter k. We refer the reader to [7, 16] for more details on the notion of fixed-parameter tractability.

The notion of *Turing kernelization* was introduced to formalize the idea of a "cheat kernel", wherein, given an instance of a parameterized problem, an algorithm outputs polynomially many polynomial kernels rather than a single kernel [15]. Formally, a t-oracle for a parameterized problem Π is an oracle that takes as input (I, k) with $|I| \leq t, |k| \leq t$ and decides whether $(I, k) \in \Pi$ in constant time. Π is said to have a $g(k)$*-sized Turing kernel* if there is an algorithm which, given input (I, k) and a $g(k)$-oracle for Π, decides whether $(I, k) \in \Pi$ in time polynomial in $|I + k|$. The Turing kernel is polynomial if $g()$ is a polynomial function.

To prove our lower bounds on polynomial kernels, we need a few notions and results from the recently developed theory of kernel lower bounds [1, 2, 6, 11]. We use a notion of reductions, similar in spirit to those used in classical complexity to show NP-hardness results, to show this kernelization lower bound. We begin by associating a classical decision problem with a parameterized problem in a natural way as follows:

Definition 1. [**Derived Classical Problem**] [2] *Let $\Pi \subseteq \Sigma^* \times \mathbb{N}$ be a parameterized problem, and let $1 \notin \Sigma$ be a new symbol. We define the* derived classical problem *associated with Π to be $\left\{x1^k | \, (x, k) \in \Pi\right\}$.*

The notion of a composition algorithm plays a key role in the lower bound argument.

Definition 2. [**Composition Algorithm, Compositional Problem**] [1] *A* composition algorithm *for a parameterized problem $\Pi \subseteq \Sigma^* \times \mathbb{N}$ is an algorithm that*

- *takes as input a sequence $\langle (x_1, k), (x_2, k), \ldots, (x_t, k) \rangle$ where each $(x_i, k) \in \Sigma^* \times \mathbb{N}$,*
- *runs in time polynomial in $\sum_{i=1}^{t} |x_i| + k$,*
- *and outputs an instance $(y, k') \in \Sigma^* \times \mathbb{N}$ with*
  1. *$(y, k') \in L \iff (x_i, k) \in L$ for some $1 \leq i \leq t$, and*
  2. *k' is polynomial in k.*

We say that a parameterized problem is compositional *if it has a composition algorithm.*

Theorem 1. *[1, Lemmas 1 and 2] Let L be a compositional parameterized problem whose derived classical problem is NP-complete. If L has a polynomial kernel, then $NP \subseteq coNP/Poly$.*

Now we define the class of reductions which lead to the kernel lower bound.

Definition 3. [2] *Let P and Q be parameterized problems. We say that P is polynomial time and parameter reducible to Q, written $P \leq_{Ptp} Q$, if there exists a polynomial time computable function $f : \Sigma^* \times \mathbb{N} \to \Sigma^* \times \mathbb{N}$, and a polynomial $p : \mathbb{N} \to \mathbb{N}$, and for all $x \in \Sigma^*$ and $k \in \mathbb{N}$, if $f((x, k)) = (x', k')$, then $(x, k) \in P$ if and only if $(x', k') \in Q$, and $k' \leq p(k)$. We call f a polynomial parameter transformation (or a PPT) from P to Q.*

This notion of a reduction is useful in showing kernel lower bounds because of the following theorem:

Theorem 2. [2, Theorem 3] *Let P and Q be parameterized problems whose derived classical problems are P^c, Q^c, respectively. Let P^c be NP-complete, and $Q^c \in$ NP. Suppose there exists a PPT from P to Q. Then, if Q has a polynomial kernel, then P also has a polynomial kernel.*

We use $[n]$ to denote $\{1, 2, \ldots, n\} \subseteq \mathbb{N}$. The operation of *subdividing* an edge (u, v) involves replacing the edge (u, v) with two new edges (u, x_{uv}) and (x_{uv}, v), where x_{uv} is a new vertex (the *subdivided vertex*). For any two vertices u and v, the *distance* between u and v, denoted $d(u, v)$, is the length of a shortest path between u and v. The k-neighborhood of a vertex u in a graph G is the set of all vertices in G that are at a distance of at most k from u. A *rooted tree* is a pair (T, r) where T is a tree and $r \in V(T)$. A leaf node in a rooted tree (T, r) is said to be a *lowest leaf* if it is a leaf at the maximum distance from the root r. The *diameter* of a graph G is defined to be $max_{u,v \in V(G)} d(u, v)$. In other words, diameter of a graph is the length of a "longest shortest" path in the graph. A *superstar graph* is a tree with diameter at most 4. Note that in any superstar G, there exists a vertex v such that G rooted at v has height at most two.

The problem at the heart of this work is the following:

Colorful Motif
Input: A graph $G = (V, E)$, $k \in \mathbb{N}$, and a coloring function $c : V \to [k]$.
Parameter: k
Question: Does G contain a subtree T on k vertices such that c restricted to T is bijective?

3 Hardness On Superstar Graphs

We begin by observing that the Colorful Motif problem is NP-complete even on simple classes of graphs. It is already known that the problem is NP-complete on comb graphs [5]. In this section, we show that the problem is NP-complete on superstars — or equivalently, on rooted trees of height at most two. To begin with, consider Colorful Motif on paths. A solution corresponds to a colorful subpath, which, if it exists, we can find in polynomial time by guessing its end points. It is easy to see that this approach can be extended to a polynomial time algorithm for Colorful Motif on caterpillars, in which case we are looking for a colorful "subcaterpillar": We may guess the end points of the spine of the subcaterpillar, and for any given guess, if the subpath on the spine does not span the entire set of colors, we check if they can be found on the leaves. The details are omitted as the argument is straightforward.

Recall that a lobster is a tree where the removal of all leaf vertices results in a caterpillar. Lobsters are a natural generalization of caterpillars, and we show that the Colorful Motif problem is NP-hard on lobsters. In fact, we show that the problem is NP-hard on lobsters whose spine has just *one* vertex. Observe that every such graph is a superstar graph; thus we show that the problem is NP-hard on superstars. To show these hardness results, we reduce from the following variant of the well-known Set Cover problem:

Colorful Set Cover
Input: A finite universe U, a finite family $\mathcal{F} = \{F_1, F_2, \ldots, F_n\}$ of subsets of U that is such that there is no i, j for which $F_i \cup F_j = U$, and a function $C : \mathcal{F} \to U$ such that $C(F_i) \in F_i$.
Question: Does there exist $\mathcal{R} \subseteq \mathcal{F}$ such that $\bigcup_{S \in \mathcal{R}} S = U$ and C is injective on $\mathcal{R}$?

We will need the fact that Set Cover is NP-complete even when no two sets in the family span the universe. Formally, if the input to Set Cover is restricted to families that have the property that no two subsets in the family are such that their union is the universe, it remains NP-complete:

At-Least-Three Set Cover
Input: A finite universe U, a finite family $\mathcal{F} = \{F_1, F_2, \ldots, F_n\}$ of subsets of U, such that there is no i, j for which $F_i \cup F_j = U$.
Question: Does there exist $\mathcal{R} \subseteq \mathcal{F}$ such $\bigcup_{S \in \mathcal{R}} S = U$?

Proposition 1. At-Least-Three Set Cover *is NP-complete.*

Proof. The statement follows by an easy reduction from the well-known Set Cover problem which is among Karp's original list of 21 NP-complete problems [13]:

Set Cover
Input: A finite universe U, a finite family $\mathcal{F} = \{F_1, F_2, \ldots, F_n\}$ of subsets of U, and a positive integer k.
Question: Does there exist $\mathcal{R} \subseteq F$, with $|\mathcal{R}| \leq k$, such that $\bigcup_{S \in \mathcal{R}} S = U$?

If $k = 1$, check if there exists i such that $F_i = U$. If this is the case, return a trivial YES-instance of At-Least-Three Set Cover. If $k > 1$, then in time $\binom{n}{2}$, examine if that there exists i, j for which $F_i \cup F_j = U$: if there is, return again a trivial YES-instance of At-Least-Three Set Cover, and if not, return the original instance. The correctness of this reduction is immediate. $\square$

Lemma 1. Colorful Set Cover *is NP-hard.*

Proof. We establish this by a reduction from At-Least-Three Set Cover. Let $(U, \mathcal{F} = \{F_1, \ldots, F_n\}, k)$ be an instance of At-Least-Three Set Cover. We construct an instance $(U', \mathcal{F}', C)$ of Colorful Set Cover as follows (see Figure 3 for an example). Let $X = \{x_1, x_2, \ldots, x_k\}$ be a set disjoint from U. For $1 \leq i \leq n, 1 \leq j \leq k$, let $F_{ij} = F_i \cup \{x_j\}$. For $1 \leq i \leq n, 1 \leq j \leq k$, we define $C(F_{ij}) = x_j$. We set $U' = U \cup X$ and $\mathcal{F}' = \{F_{ij} \mid 1 \leq i \leq n, 1 \leq j \leq k\}$. Note that, as required, $U' = U \cup X$, and there are no indices i, j, p, q for which $F_{ij} \cup F_{pq} = U$.

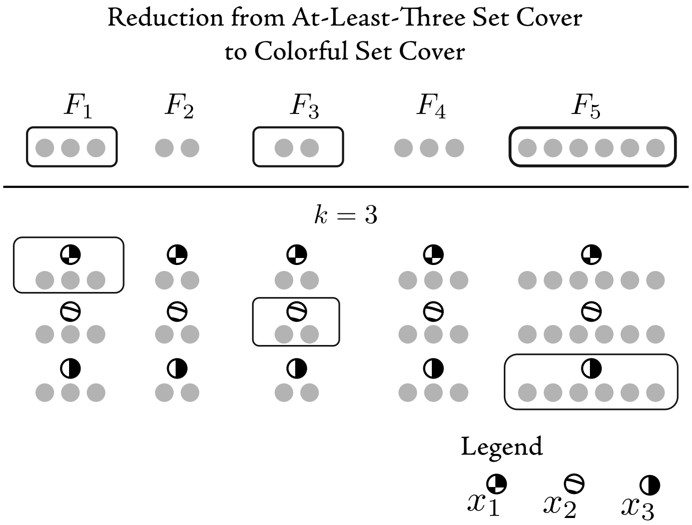

Fig. 1. An Illustration of the Reduction

Now, suppose the given At-Least-Three Set Cover instance is a YES instance, and $R = \{F_{i_1}, F_{i_2}, \ldots, F_{i_t}\}$ be a set cover of size at most k in the At-Least-Three Set Cover instance. In the reduced instance, consider the set $R' = \{F_{i_1} \cup \{x_1\}, F_{i_2} \cup \{x_2\}, \ldots, F_{i_t} \cup \{x_t\}\}$. For $1 \leq j \leq k$, let $F'_j = \{F_{ij} \mid 1 \leq i \leq n\}$. If $t < k$, then arbitrarily pick an element from each $F'_j, t < j \leq k$ and add it to R'. Call this new set R''. We claim that R'' is a solution for the instance Colorful Set Cover of $(U', \mathcal{F}', \mathcal{C})$. Consider $d \in X$. Since we have picked at least one element from each F'_j into our solution, d is covered by R''. Suppose $d \in U$. Let $d \in F_{i_j} \in R$ (note that such an F_{i_j} exists since R is a solution for the Set Cover instance). Then, $F_{i_j} \cup \{x_j\} \in R' \subseteq R''$, implying that d is covered by R'' and hence, the reduced instance is a YES instance of Colorful Set Cover.

Conversely, suppose the reduced instance is a YES instance of Colorful Set Cover, and let R'' be a colorful set cover. It is easy to see that R'' contains exactly one element from each F'_j. Let $R'' = \{F''_1, F''_2, \ldots, F''_k\}$. Define $R = \{F''_1 \backslash X, F''_2 \backslash X, \ldots, F''_k \backslash X\}$. Note that any $d \in U$ covered by some $F''_i \in R''$ will be covered by $F''_i \backslash X \in R$. Since $|R| \leq k$, it is indeed a solution for the At-Least-Three Set Cover instance. $\square$

Theorem 3. Colorful Motif *on superstar graphs is NP-hard.*

Proof. The proof is by reduction from Colorful Set Cover. Let $(U, \mathcal{F} = \{F_1, \ldots, F_n\}, C)$ be an instance of Colorful Set Cover. Without loss of generality, we assume that $\bigcup_{S \in \mathcal{F}} S = U$. We construct a graph T as follows. For each set $F_i \in \mathcal{F}$ and each element $x \in F_i$, add a vertex $x[i]$ to $V(T)$. Note that the same element in U may appear any number of times as a vertex, once for each time it appears in the sets $F_i, 1 \leq i \leq n$. For each set F_i, add a new "set" vertex u_i and make it adjacent to all the vertices $x[i]$ that correspond to elements of F_i. Finally, add a "root" vertex r and make r adjacent to all the vertices u_i. This completes the description of T. More formally,

$$V(T) = \{x[i] \mid x \in F_i, 1 \leq i \leq n\} \cup \{u_1, \ldots, u_n\} \cup \{r\}$$
$$E(T) = \{(x[i], u_i) \mid x \in F_i, 1 \leq i \leq n\} \cup \{(u_i, r) \mid 1 \leq i \leq n\}$$

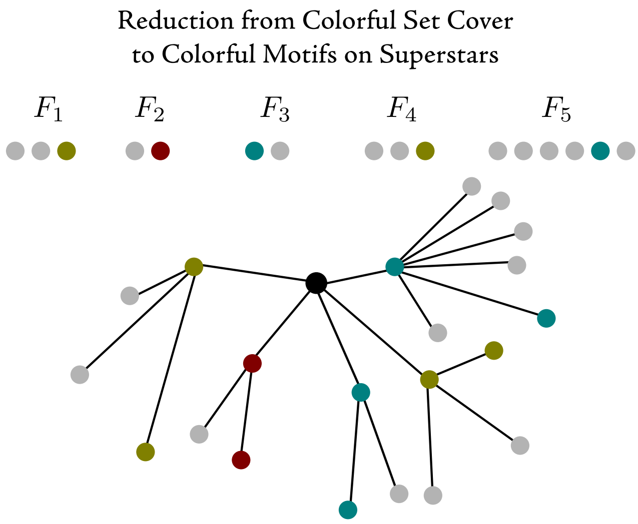

Fig. 2. Reduction from Colorful Set Cover to Colorful Motif on superstar graphs. Note that the highlighted color in each F_i corresponds to $C(F_i)$

For every $x \in U$, we let c_x be a color in the set of colors of the Colorful Motif instance. Also, for simplicity, let f_i denote $C(F_i)$. The coloring function c labels each vertex $x[i], x \in F_i$ with the color c_x. The "set" vertex u_i gets the color c_{f_i} where F_i is the set that u_i represents. The root vertex r gets a new color, named c_r. Formally,

$$c(x[i]) = \{c_x \mid x \in F_i, 1 \leq i \leq n\}$$
$$c(u_i) = \{c_{f_i} \mid 1 \leq i \leq n, \textbf{ and } c(r) = c_r\}.$$

Note that the color set used contains one color for each element in U, and the new color c_r. T is clearly a superstar; the reduced instance is $(T, |U| + 1, c)$.

Now, suppose that the given Colorful Set Cover instance is a YES instance, and let $R = \{F_{i_1}, F_{i_2}, \ldots, F_{i_t}\} \subseteq F$ be a solution for this instance. Consider the subtree T' of T consisting of (1) the subtrees rooted at $u_{i_1}, \ldots, u_{i_t}$, (2) the vertex r, and (3) the edges $(r, u_{i_1}), \ldots, (r, u_{i_t})$. From the subtree rooted at each u_j in T', remove those leaves that have the same color as u_j. From the remaining leaves, arbitrarily delete all but one leaf of each color. Call the resulting tree T''. It is clear from the construction of T'' that it does not contain two vertices of same color. Now, consider the color c_x for any $x \in U \cup \{r\}$. We claim that T'' contains a vertex of color c_x. If $c_x = c_r$, then we are done since T'' contains the vertex r. Suppose $c_x \neq c_r$. Let F_p be an element of R which covers the element $x \in U$ in the solution to the Colorful Set Cover instance. Then, either $C(F_p) = x$, in which case we can immediately see that T''contains u_p which is colored with color c_x, or $C(F_p) \neq x$, in which case our construction of T'' ensures that the subtree of T'' rooted at u_p has a leaf with color c_x. Thus T'' is a colorful subtree of T with $|U| + 1$ colors.

Conversely, suppose that the reduced instance is a YES instance of Colorful Motif and let T'' be a colorful subtree of T with $|U| + 1$ colors. Let $\{u_{i_1}, u_{i_2}, \ldots, u_{i_t}\} = V(T'') \cap \{u_1, \ldots, u_n\}$. Let $R = \{F_{i_1}, F_{i_2}, \ldots, F_{i_t}\}$. In the original instance of Colorful Set Cover, the elements of R get the colors $\{c(u_{i_1}), c(u_{i_2}), \ldots, c(u_{i_t})\}$, and these are all distinct. Consider $d \in U$. Since T'' is colorful with $|U| + 1$ colors, some vertex of T'' is colored c_d, and $c_d \neq c_r$. Hence, the vertex of T'' colored c_d is either some u_i or a leaf adjacent to some u_i. Since we are picking the corresponding F_i in R, we ensure that R covers d and hence the given instance is a YES instance of Colorful Set Cover. $\square$

Proposition 2. *Let (T, C) be an instance of* Colorful Motif*, where T is a superstar graph. Let $u_1, \ldots u_r$ be the children of the root of T. Let V_i denote the set of leaves adjacent to u_i, and let U_i denote $V_i \cup \{u_i\}$. For $X \subseteq V(T)$, let $c(X)$ denote the set of colors used on X, that is:*

$$c(X) = \{d \mid \exists x \in X, c(x) = d\}.$$

The Colorful Motif *problem is hard on superstar graphs even on instances where no two subtrees are colored with the entire set of colors: that is, for any $i \neq j$,*

$$(c(U_i) \cup c(U_j)) \setminus C \neq \phi.$$

Proof. Note that the claim follows from the reduction stated in the proof of theorem 3, because of the fact that an instance of Colorful Set Cover is such that there is no i, j for which $F_i \cup F_j = U$. This rules out the possibility of the reduced instance obtained in the proof above having two subtrees that are colored with the entire set of colors. □

4 Colorful Motifs on Graphs of Diameter Two and Three

In this section, we consider the Colorful Motif problem restricted to graphs of diameter two and three. We show that the Colorful Motif problem on superstars reduces to Colorful Motif on graphs of diameter two, thereby establishing that the problem is NP-complete. Also, we show that the Colorful Motif problem on general graphs reduces to Colorful Motif on graphs of diameter three, thereby establishing that the problem is NP-complete, and that polynomial kernels are infeasible. These reductions are quite similar, with only subtle differences.

Lemma 2. *The* Colorful Motif *problem on superstars with parameter k reduces to* Colorful Motif *with parameter k on graphs of diameter two.*

Proof. Let (T, k, c) be an instance of Colorful Motif on superstars. A superstar can be, by definition, rooted at a vertex r such that the graph is a rooted tree of height two. Let r denote the root of T, and assume that the neighbors of r are ordered in some arbitrary but fixed fashion. Let $T(i)$ denote the graph induced on the i^{th} neighbor of r, and the leaves of T that are adjacent to it, that is,

$$T(i) := T[v_i \cup \left(N[v_i] \cap V \setminus r\right)].$$

We refer to the graph induced on $T(i)$ as the i^{th} subtree. Note that, by Proposition 2, we may assume that any colorful subtree of T intersects non-trivially with more than two subtrees. We now describe an instance (Q, k, c_q) that is equivalent to (T, k, c), and is such that Q has diameter two. The graph Q is obtained from T in the following steps: First, we add $\binom{k}{2}$ new vertices:

$$V(Q) = V(T) \cup \{v[i, j] \mid i, j \in [k] \text{ and } i \neq j\}.$$

We use X to denote the set of newly introduced vertices, that is, $\{v[i, j] \mid i, j \in [k] \text{ and } i \neq j\}$. The new vertices get the same color as r, that is, $c_q(u) = c(r)$ for all vertices $u \in V(Q) \setminus V(T)$. For all "original" vertices u in $V(T)$, $c_q(u) = c(u)$. The edge set of Q retains all the original edges in T, Further, for every pair of distinct subtrees $T(i)$ and $T(j)$, we make the vertex $v[i, j]$ global to all the vertices in $T(i) \cup T(j)$. Finally, make every $v[i, j]$ global to the set of all vertices in the closed neighborhood of r, and induce a clique on all vertices $v[i, j]$. Formally:

(a) $\forall \{u, v\}, u \in T(i), v \in T(j)$, for $i, j \in [k]$ and $i \neq j$, $(u, v) \in E(Q)$ if, and only if, $(u, v) \in E(T)$
(b) $\forall i, j \in [k], i \neq k$, and $u \in T(i) \cup T(j)$, $(v[i, j], u) \in E$
(c) $\forall u \in X$ and $\forall v \in N[r]$, $(u, v) \in E$
(d) $\forall u, v \in X$, $(u, v) \in E$.

Notice that Q is a graph of diameter two. Indeed, any pair of vertices within the same subtree have distance at most two (via their parent), and any pair of vertices in distinct subtrees $T(i)$ and $T(j)$ are reachable by a path of length at most two: if $i \neq k+1$ and $j \neq k+1$, then $v[i, j]$ is a common neighbor of any pair of vertices (u, v) such that $u \in T(i)$ and $v \in T(j)$, thus making them distance two apart. Any $v[i, j]$ is distance one from any vertex in $T(i)$ or $T(j)$ and is distance two from a vertex in any other

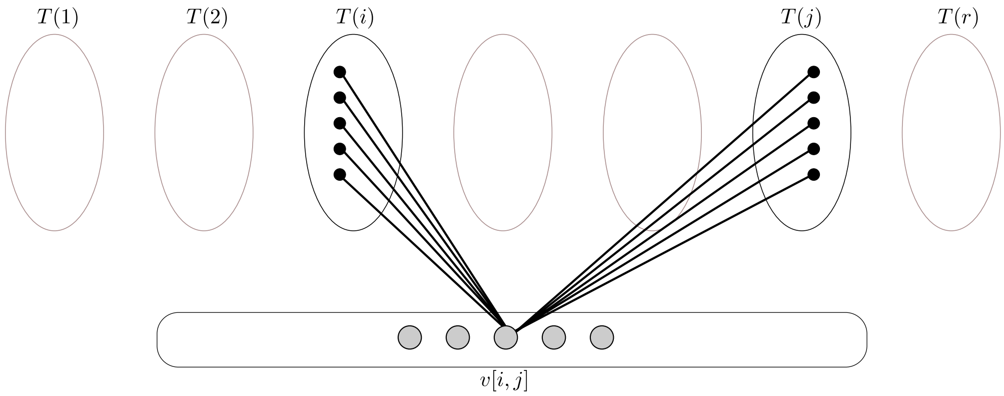

Fig. 3. A Slice of the graph Q

subtree $T(l)$, because r is a common neighbor of a vertex in $T(l)$ and $v[i,j]$ (recall that $C(c(r))$ induces a clique, and $v[l,l']$ is adjacent to every vertex in $T(l)$).

We now claim that (T,k,c) is a YES-instance of Colorful Motif if, and only if, (Q,c_q,k) is a YES-instance of Colorful Motif. Notice that any colorful subtree T' in T is a colorful subtree of Q. In the converse, let R be a colorful subtree of Q. Observe that none of the newly introduced vertices (or edges) are used in R': such vertices connect precisely two subtrees, and therefore R intersects at most two subtrees of Q. Notice that $R \setminus v[i,j]$ is a colorful subtree of T, which contradicts our assumption that any colorful subtree of T intersects non-trivially with more than two subtrees. Thus, R functions as a colorful subtree T' of T. It is easy to see that this is also a polynomial parameter transformation. This completes the proof the lemma.

□

Lemma 3. *The* Colorful Motif *problem with parameter k reduces to* Colorful Motif *with parameter $(k+1)$ on graphs of diameter three. The reduction serves to show both NP-hardness and infeasibility of polynomial kernelization (being a polynomial parameter transformation).*

Proof. Let (T,c,k) be an instance of Colorful Motif with k colors. Let $T(i)$ denote the subset of vertices in T that have color i, that is:

$$T(i) = \{v \in T \mid c(v) = i\}.$$

We refer to the set of vertices in $T(i)$ as the *color class i*. We now describe an instance $(R, c_r, k+1)$ that is equivalent to T, and is such that R has diameter three. To this end, we describe an intermediate instance $(Q, c_q, k+1)$ based on (T,c,k). The graph Q is obtained from T in the following steps: First, we add $\binom{k}{2}$ new vertices:

$$V(Q) = V(T) \cup \{v[i,j] \mid i,j \in [k] \text{ and } i \neq j\}.$$

The new vertices form the color class $(k+1)$, that is, $c_q(u) = k+1$ for all vertices $u \in V(Q) \setminus V(T)$. We abuse notation and use $T(i)$ to refer to color class i in Q, for $i \in [k+1]$. For all "original" vertices u in $V(T)$, $c_q(u) = c(u)$. The edge set of Q retains all the original edges in T, and further, we add edges so that every color class induces a clique. Finally, for every pair of distinct color classes $T(i)$ and $T(j)$, we make the vertex $v[i,j]$ global to all the vertices in $T(i) \cup T(j)$. Formally:

(a) For every $i \in [k+1]$, $\forall \{u,v\} \in T(i)$, $(u,v) \in E(Q)$
(b) $\forall \{u,v\}, u \in T(i), v \in T(j)$, for $i,j \in [k]$ and $i \neq j$, $(u,v) \in E(Q)$ if, and only if, $(u,v) \in E(T)$
(c) $\forall i,j \in [k], i \neq k$, and $u \in T(i) \cup T(j)$, $(v[i,j],u) \in E$

Notice that Q is a graph of diameter two. Indeed, any pair of vertices within the same color class $T(i)$ have distance one, and any pair of vertices in distinct color classes $T(i)$ and $T(j)$ are reachable by a path of length at most two: if $i \neq k+1$ and $j \neq k+1$, then $v[i,j]$ is a common neighbor of any pair of vertices (u,v) such that $u \in T(i)$ and $v \in T(j)$, thus making them distance two apart. Any $v[i,j]$ is distance one

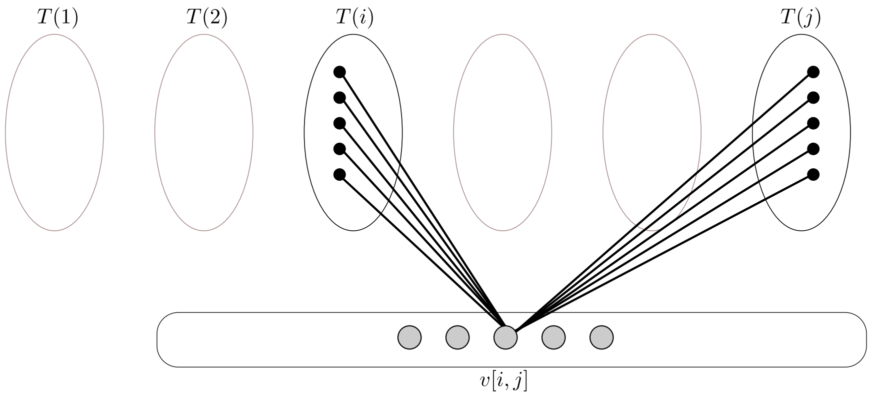

Fig. 4. A Slice of the graph Q

from any vertex in $T(i)$ or $T(j)$ and is distance two from a vertex in any other color class $T(l)$, because $v[l, l']$, for any $l' \neq l$, is a common neighbor of a vertex in $T(l)$ and $v[i,j]$ (recall that $T(k+1)$ induces a clique, and $v[l, l']$ is adjacent to every vertex in $T(l)$).

We are now ready to provide a description of $(R, c_r, k+1)$. The graph R is obtained from Q by replacing every $v[i,j] \in T(k+1)$ with the following vertex set:

$$V[i,j] = \{v[i,j]^{(1)}, v[i,j]^{(2)}, \ldots, v[i,j]^{(d_{ij})}\},$$

where d_{ij} is the number of neighbors of $v[i,j]$ in $V(R) \setminus T(k+1)$. We assume that the vertices in $N(v[i,j])$ are ordered in some arbitrary but fixed fashion. We then add the edges $(v[i,j]^{(l)}, u_l)$, where u_l is the l^{th} neighbor of $v[i,j]$. We then add edges so that $T(k+1)$ induces a clique. Again, for all original vertices u in $V(Q)$, $c_r(u) = c_q(u)$. For all vertices $u \in T(k+1)$, we let $c_r(u) = k+1$.

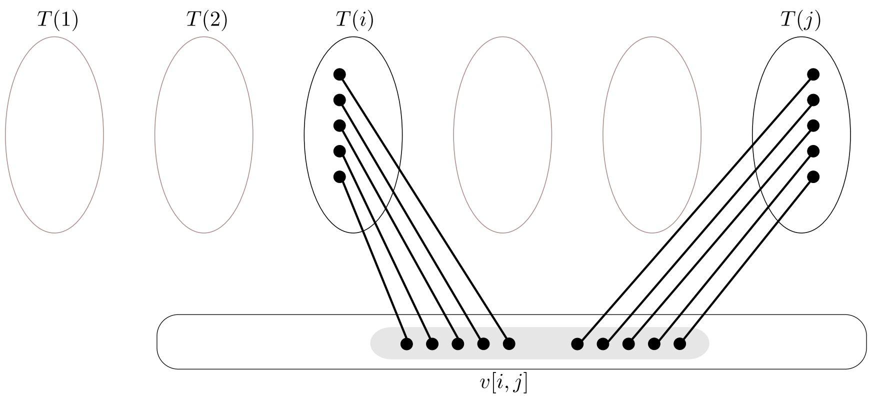

Fig. 5. A Slice of the Graph R

Observe that R is a graph of diameter three. Again, every pair of vertices within the same color class $T(i)$ have distance one. Any pair of vertices in distinct color classes $T(i)$ and $T(j)$, if $i \neq k+1$ and $j \neq k+1$ are reachable by a path of length at most three. Let $u \in T(i)$ be the l_u^{th} neighbor of $v[i,j]$ in Q, and let $v \in T(j)$ be the l_v^{th} neighbor of $v[i,j]$. Then $(u, v[i,j]^{l_u}) \in E$, $(v[i,j]^{l_u}, v[i,j]^{l_v}) \in E$, and $(v[i,j]^{l_v}, v) \in E$, establishing a path of length three between u and v. Further, it is easy to see that any $v[i,j]^l$ is distance two from any vertex in $T(i)$ or $T(j)$ or any other color class $T(l)$, for similar reasons.

We now claim that (T, c, k) is a YES-instance of Colorful Motif if, and only if, $(R, c_r, k+1)$ is a YES-instance of Colorful Motif. Notice that any colorful subtree T' in T can be trivially extended to a colorful subtree R' in R - indeed, we may let R' to be $T' \cup \{(u, v)\}$, where $u \in T'$ and $v \in T(k+1)$ such that $(u, v) \in E$ (note that such a v always exists). Conversely, let R' be a colorful subtree of R. Let u be the vertex in R' from $T(k+1)$. Notice that u is necessarily a leaf of R', since no vertex $u \in T(k+1)$ has degree more than one outside $T(k+1)$. Notice that the subtree $R' \setminus \{u\}$ gives us a colorful subtree T' of T, as required. It is easy to see that this is also a polynomial parameter transformation. This completes the proof the lemma.

□

5 Colorful Motifs on Comb Graphs

In [5], Cygan et al. show that Colorful Motif is NP-complete on *comb graphs*, defined as follows:

Definition 4. *A graph $G = (V, E)$ is called a* comb graph *if (i) it is a tree, (ii) all vertices are of degree at most 3, (iii) all the vertices of degree 3 lie on a single simple path. The maximal path, which starts and ends in degree 3 vertices is called the* spine *of a comb graph. One of the two endpoint vertices of the spine is arbitrarily chosen as the first vertex, and the other as the last. A path from a degree 3 vertex to a leaf which contains exactly one degree 3 vertex, is called a* tooth.

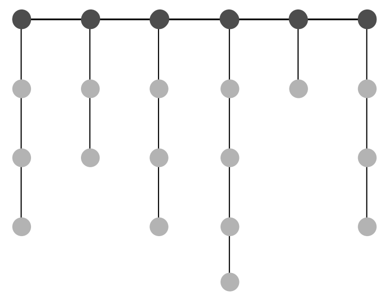

Fig. 6. A Comb Graph

In this section, we present a composition algorithm for Colorful Motif on comb graphs. Note that in [5], it is observed that Colorful Motif is unlikely to admit polynomial kernels on forests. This simple composition obtained using disjoint union does not work "as is" when we restrict our attention to comb graphs, since the graph resulting from the disjoint union of comb graphs is not a comb graph, as it is not connected.

5.1 A Composition Algorithm

We begin by introducing some notation that will be useful presently. Let (T, k, c) be an instance of Colorful Motif restricted to comb graphs, that is, let T denote a comb graph, and let $c : V(T) \to [k]$ be a coloring function.

Let T_p and T_q be two comb graphs, and let $l_p \in V(T_p)$ be the last vertex on the spine of T_p, and let $f_q \in V(T_q)$ be the first vertex on the spine of T_q. We define $T_p \odot T_q$ as follows:

(i) $V(T_p \odot T_q) = V(T_p) \uplus V(T_q) \cup \{v_p, v_q\}$, where $\{v_p, v_q\}$ are "new" vertices, and
(ii) $E(T_p \odot T_q) = E(T_p) \uplus E(T_q) \cup \{(l_p, v_p), (v_p, v_q), (v_q, f_q)\}$

We are now ready to describe the composition algorithm:

Lemma 4. *The* Colorful Motif *problem does not admit a polynomial kernel on comb graphs unless $NP \subseteq coNP/Poly$.*

Proof. Let $(T_1, c_1, k), (T_2, c_2, k), \ldots (T_t, c_t, k)$ be the instances that are input to the composition algorithm. Let $\mathcal{T}$ denote the graph:

$$\mathcal{T} = T_1 \odot T_2 \odot \cdots \odot T_t.$$

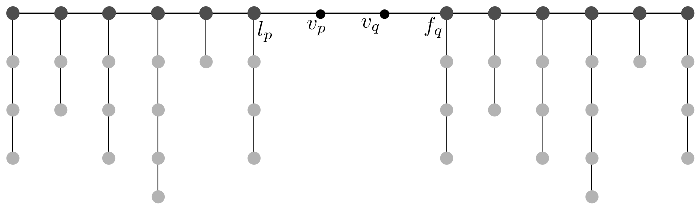

Fig. 7. An illustration of the $T_p \odot T_q$ operation

Let N denote the set of all new vertices introduced by the $\odot$ operations. Notice that any vertex of $\mathcal{T}$ that does not belong to N is a vertex from one of the instances T_i. We will refer to such vertices in $\mathcal{T}$ as being from $\mathcal{T}(T_i)$.

We refer to the pair of vertices in N adjacent to the endpoints of the spine of a T_i as the *guards* of T_i (notice that any T_i has at most two guard vertices). We define the coloring function c on $\mathcal{T}$ as follows. For every vertex $u \in T_i$, $c(u) = c_i(u)$. For every vertex u that is a guard of T_i, $c(u) = c(v)$, where v is the vertex of T_i adjacent to u. We now claim that $(\mathcal{T}, k)$ is the composed instance.

We first show that if $\exists i, i \in [t]$, such that T_i is a YES-instance of Colorful Motif, then $\mathcal{T}$ is a YES instance of Colorful Motif. Notice that the colorful connected subtree of T_i also exists in $\mathcal{T}$, and hence we are done.

Conversely, let $\mathcal{T}$ be a YES-instance of Colorful Motif. Let R be a colorful connected subtree of $\mathcal{T}$. Notice that by construction, if $V(R)$ intersects with $\mathcal{T}(T_i)$ and $\mathcal{T}(T_j)$ for $i \neq j$, then it contains two vertices of the same color — indeed, this follows from the observation that $V(R)$ would have to contain the guard vertices of T_i and T_j to be connected, and this would lead to multiple occurrences of the colors that occur at the endpoints of the spines of T_i and T_j.

Thus, $V(R)$ intersects with $\mathcal{T}(T_i)$ for exactly one value of i, $1 \leq i \leq t$. Clearly, R is a colorful connected subtree of T_i. Thus, by Theorem 1, Colorful Motif admits no polynomial kernel on comb graphs unless $NP \subseteq coNP/Poly$. □

Corollary 1. *The* Colorful Motif *problem on lobsters does not admit a polynomial kernel unless $NP \subseteq coNP/Poly$.*

Proof. Clearly, the problem is NP-complete due to Theorem 3. Further, observe that the composition described in the proof of 4 can be imitated with minor changes to obtain a similar result on lobsters. Indeed, we only need to observe that it suffices to subdivide only edges along the spine of the lobster (which is the spine of the caterpillar obtained by removing all the leaves of the lobster). If a solution does not intersect the spine, it is contained in one of the "dangling superstars", and this is easily detected before applying the composition. The details are routine and therefore omitted. □

5.2 Many Polynomial Kernels

Although a parameterized problem may not necessarily admit a polynomial kernel, it may admit many of them, with the property that the instance is in the language if and only if at least one of the kernels corresponds to an instance that is in the language. We now show that the Colorful Motif problem admits n kernels of size $\mathcal{O}(k^2)$ each on comb graphs. This is established by showing that a closely related variant, the Rooted Colorful Motif problem, admits a polynomial kernel. The Rooted Colorful Motif problem is the following:

Rooted Colorful Motif

Input:	A graph $G = (V, E)$, $k \in \mathbb{N}$, a coloring function $c : V \to [k]$, and $r \in V$.
Parameter:	k
Question:	Does G contain a subtree T on k vertices, containing r, such that c restricted to T is bijective?

Lemma 5. *The* Colorful Motif *problem admits many polynomial kernels on comb graphs.*

Proof. Let $T = (V, E)$ be a comb graph, and let (T, k, c, u) be an instance of ROOTED COLORFUL MOTIF. We first show that T can be reduced to an equivalent instance T_u on at most $\mathcal{O}(k^2)$ vertices. Notice that we may obtain an equivalent instance T_u from T by deleting all vertices in T that lie outside the k-neighborhood of u. This is because any colorful subtree of T rooted at u will not involve vertices outside the k-neighborhood of u.

We now observe that T_u has at most $\mathcal{O}(k^2)$ vertices. Note that in the k-neighborhood of u in T, there are at most $2k$ vertices that belong to the spine and from each of these vertices, there is a tooth of length at most k. Also, if u lies on a tooth, there are at most $2k$ more vertices on the same tooth, and if u lies on the spine then there are at most k more vertices on the tooth rooted at u, if at all one exists. Hence, in T_u, there are at most $2k$ vertices of degree 3 each having a tooth of length at most k and at most one other tooth, which is of length at most $2k$. So, the total number of vertices in T_u is at most $2k + 2k * k + 2k = \mathcal{O}(k^2)$. Notice that there are n choices for u from T, and repeating the procedure above by "guessing the root" gives us n polynomial kernels for the problem, as desired. ◻

6 Colorful Motifs on Trees

In this section, we demonstrate the infeasibility of some strategies for showing many polynomial kernels for COLORFUL MOTIF restricted to trees. Observe that the COLORFUL MOTIF problem is unlikely to admit a polynomial kernel on trees, since a polynomial kernel on trees would imply a polynomial kernelization procedure for comb graphs, which is infeasible (see Lemma 4).

6.1 Hardness with a Fixed Root

In the case of comb graphs, we were able to establish that the problem of finding a colorful subtree with a fixed root admits a $\mathcal{O}(k^2)$ kernel. Unfortunately, this approach does not extend to trees, as we establish that ROOTED COLORFUL MOTIF is compositional on trees.

Proposition 3. *The* ROOTED COLORFUL MOTIF *problem restricted to trees is* NP*-hard.*

Proof. By reduction from COLORFUL MOTIF restricted to trees, which is NP-hard on trees (in fact, on the even more restricted class of comb graphs) as shown by Cygan et al [5].

Let (T, k, c) be an instance of COLORFUL MOTIF where T is a tree. Let $\{v_1, v_2, \ldots, v_\ell\}$ be the vertices in T that have color 1 in T. Make ℓ copies $T_1, T_2, \ldots, T_\ell$ of T. Add a new vertex v, and for $1 \leq i \leq \ell$, make v adjacent to the copy of v_i in T_i; let this graph be T'. Let c' be the coloring function that gives all copies of a vertex in T the same color as it has in T, and the color $k + 1$ to the new vertex v. $(T', k + 1, c', v)$ is the reduced instance of ROOTED COLORFUL MOTIF. T' is clearly a tree, and it is easy to see that (T, k, c) is a yes instance of COLORFUL MOTIF if and only if $(T', k+1, c', v)$ is a yes instance of ROOTED COLORFUL MOTIF. ◻

Proposition 4. *The* ROOTED COLORFUL MOTIF *problem when restricted to trees does not admit a polynomial kernel unless $NP \subseteq coNP/Poly$.*

Proof. From Proposition 3, the ROOTED COLORFUL MOTIF problem is NP-complete when restricted to trees. We now describe a composition algorithm for ROOTED COLORFUL MOTIF on trees.

Let $(T_1, v_1, c_1, k), (T_2, v_2, c_2, k), \ldots (T_r, v_r, c_r, k)$ be the input instances, where $T_i = (V_i, E_i)$. Consider the tree $\mathcal{T} = (V, E)$ described as follows:

1. $V = \{u\} \cup \{u_1, u_2, \ldots, u_r\} \cup_{i \in [r]} V_i$
2. $(u, u_i) \in E$, for all $i \in [r]$, and $(u_i, v_i) \in E$, for all $i \in [r]$
3. $E_i \subset E$ for all $i \in [t]$.

Consider the coloring function $c : V \to [k + 2]$ defined as follows: for all $v \in V_i$, $c(v) = c_i(v)$. Further, $c(u_i) = k + 1$ for all u_i, and $c(u) = k + 2$.

We now claim that $(\mathcal{T}, k+2, c, v)$ is the composed instance. Indeed, if T_i has a colorful subtree rooted at v_i then we have a colorful subtree (on $k+2$ colors) rooted at v, which extends the tree rooted at v_i using the edges (v_i, u_i) and (u_i, u). In the reverse direction, suppose $\mathcal{T}$ has a colorful subtree T' rooted at u. Then, observe that T' contains vertices from *exactly* one of the T_is. This is because if T' contains vertices from T_i and T_j then T' contains vertices u_i and u_j both of which have the same color , a contradiction to the assumption that T' was colorful. This completes the proof. ◻

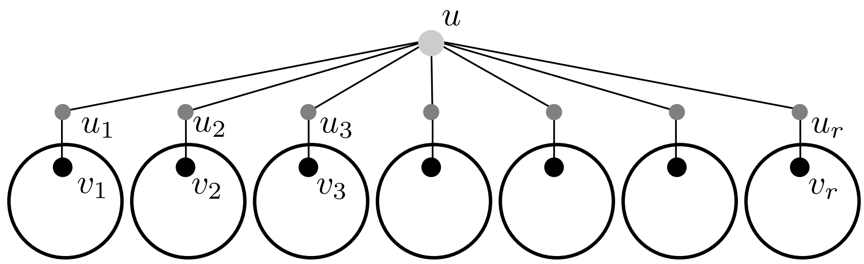

Fig. 8. An illustration of the composition

Hardness on Trees of Bounded Degree Since the problem of Colorful Motif is fixed-parameter tractable on trees with running time $\mathcal{O}^*(2^k)$, we may assume, without loss of generality[4], that the number of instances input to the composition algorithm is at most $\mathcal{O}^*(2^k)$. Consider the Colorful Motif problem restricted to rooted binary trees. We establish in this section that this problem is compositional as well:

Lemma 6. *The* Colorful Motif *problem is compositional on rooted binary trees, and does not admit a polynomial kernel unless* $NP \subseteq coNP/Poly$.

Proof. We build the composed instance $\mathcal{T}$ by combining the input instances with a complete binary tree on t leaves, where t is the number of instances input to the composition algorithm. Let (T_i, r_i, c_i, k), for $i \in [t]$ be the input instances to the composition algorithm, where T_i is a rooted binary tree rooted at r_i. Let $\mathcal{B}$ be the complete binary tree on t leaves. Note that the depth of $\mathcal{B}$ is $\log t$, and since we have assumed that $t \leq 2^k$, the depth of $\mathcal{B}$ is at most k. Denote the root of this tree by r, and let $D(i)$ the set of all vertices at a distance i from r. Assume that the leaves are ordered in some arbitrary but fixed fashion. We identify the vertices l_i and r_i, where l_i is the i^{th} leaf of $\mathcal{B}$. We define the coloring function

$$c : V(\mathcal{T}) \to [2k],$$

as follows: $c(u) = k+i+1$ for every $u \in D(i)$, and $c(u) = c_i(u)$ for any $u \in T_i$. We claim that $(\mathcal{T}, r, c, 2k)$ is the composed instance. The correctness follows from reasons similar to those in the proof of Proposition 4, and the straightforward details are omitted. ☐

6.2 Hardness with a Fixed Subset of Vertices

Now, we have seen that "fixing" one vertex does not help the cause of kernelization for trees in general. In fact, more is true: fixing any constant number of vertices does not help. The problem we study in this section is the following:

Subset Colorful Motif

Input:	A graph $G = (V, E)$, a coloring function $c : V \to [k]$, and a set of vertices $U \subseteq V$, $\lvert U\rvert = s = \mathcal{O}(1)$.
Parameter:	k
Question:	Does G contain a subtree T on k vertices, such that $U \subseteq V(T)$, and c restricted to T is bijective?

Proposition 5. *The* Subset Colorful Motif *problem restricted to trees does not admit a polynomial kernel unless* $NP \subseteq coNP/Poly$.

Proof. To prove this proposition, it is sufficient to show that there exists a polynomial parameter transformation from the Rooted Colorful Motif problem restricted to trees to the Subset Colorful Motif problem restricted to trees (Theorem 2 and Propositions 3 and 4). We now proceed to give such a transformation.

[4] If the number of instances is greater than $f(k)n^c$, where $f(k)n^c$ is the time taken by a FPT algorithm to solve the problem, then the composition algorithm can solve every problem and trivially return an appropriate answer within the required time bounds.

Let (T, v, c, k), where $T = (V, E)$, be an input instance to ROOTED COLORFUL MOTIF. Let $s' = s-1$. Construct the tree $T' = (V', E')$ as follows: (i) $T' = (V', E')$ where $V' = V \cup \{u_1, u_2, \ldots, u'_s\}$ (ii) $E' = E \cup \{(v, u_1), \ldots, (v, u'_s)\}$

Now, we define a coloring function $c' : V' \rightarrow [k + s']$ as follows:

(i) $c'(v) = c(v)$, $\forall v \in V$

(ii) $c'(u_j) = k + j$, $\forall j \in [s']$

$(T', c', \{v, u_1, \ldots, u'_s\}, k + s')$ is the reduced instance of SUBSET COLORFUL MOTIF. It is easy to see that the reduction is a polynomial parameter transformation, and the proposition follows. □

7 Connected Dominating Set on Graphs of Diameter Two

In this section, we show that CONNECTED DOMINATING SET on graphs of diameter two is NP-complete. The classical complexity of CONNECTED DOMINATING SET on graphs of diameter two was hitherto unknown, although it was known to be NP-complete on graphs of diameter three, and trivial on graphs of diameter one. We establish this by a non-trivial reduction from COLORFUL MOTIF on graphs of diameter two, which is NP-complete by Lemma 2.

Theorem 4. *The* CONNECTED DOMINATING SET *problem, when restricted to graphs of diameter two, is NP-complete.*

Proof. We establish this by demonstrating a reduction from COLORFUL MOTIF on graphs of diameter two. Let (G, k, c) be an instance of COLORFUL MOTIF on graphs of diameter two. Let $C(i)$ denote the subset of vertices in C that have color i. That is, $C(i) = \{v \in G \mid c(v) = i\}$. We refer to set of vertices in $C(i)$ as the *color class* i. We now describe an equivalent instance (H, k) of CONNECTED DOMINATING SET on graphs of diameter two.

We define the following simple operations before we begin:

1. *inducing a clique on* S, where $S \subseteq V$: Add the edges (u, v) for all $u, v \in S$, $u \neq v$.
2. *inducing a complete bipartite graph on* (S, T), where $S, T \subseteq V$, $S \cap T = \emptyset$: Add the edges (u, v) for all $u \in S$, $v \in T$.
3. *inducing a matching on* (S, T), where $S, T \subseteq V$, $S \cap T = \emptyset$, $|S| = |T|$ and there exists a natural ordering of the vertices in S and T (that is, $S = \{s_1, s_2, \ldots, s_r\}$ and $T = \{t_1, t_2, \ldots, t_r\}$): Add the edges (s_i, t_i) for all i, $1 \leq i \leq r$.
4. *making* u *global to* S, given a vertex u and a set $S \subseteq V$: Add all edges (u, v) where $v \in S$.

The graph H is obtained from G, in a series of several steps. We describe the graph H by the sequence of modifications made to G to arrive at H.

(1) To begin with, let $H = G$.
(2) Induce a clique on $C(i)$, for all $i \in [k]$.
(3) For every color class $C(i)$, add $(k + 1)$ new vertices $v_i(1), v_i(2), \ldots, v_i(k + 1)$. We refer to this set as $H(i)$. Induce a complete bipartite graph on $(C(i), H(i))$.
(4) For every pair (i, j) such that $i \neq j$, induce a matching on $(H(i), H(j))$, where the ordering on the vertices in H_i and H_j is the natural one established by the labels. (That is, the j^{th} vertex of H_i is $v_i(j)$.) Further, for every pair of vertices $(v_i(p), v_j(q))$, where $p \neq q$, add an edge and subdivide it. We refer to the set of all such subdivided vertices as $D(i, j)$. We let $D = \cup_P D(i, j)$, where $P = \binom{[k]}{2}$.
(5) Induce a clique on D, the set of subdivided vertices.
(6) For every $u \in D(i, j)$, we make u global to $C(l)$, for every l, where $l \neq i$ and $l \neq j$.

Claim. The graph H has diameter two.

Proof. We show this by demonstrating a path of length at most two for every pair of vertices u and w. This is done in a series of cases, summarized in table 1.

The detailed case analysis proceeds as follows:

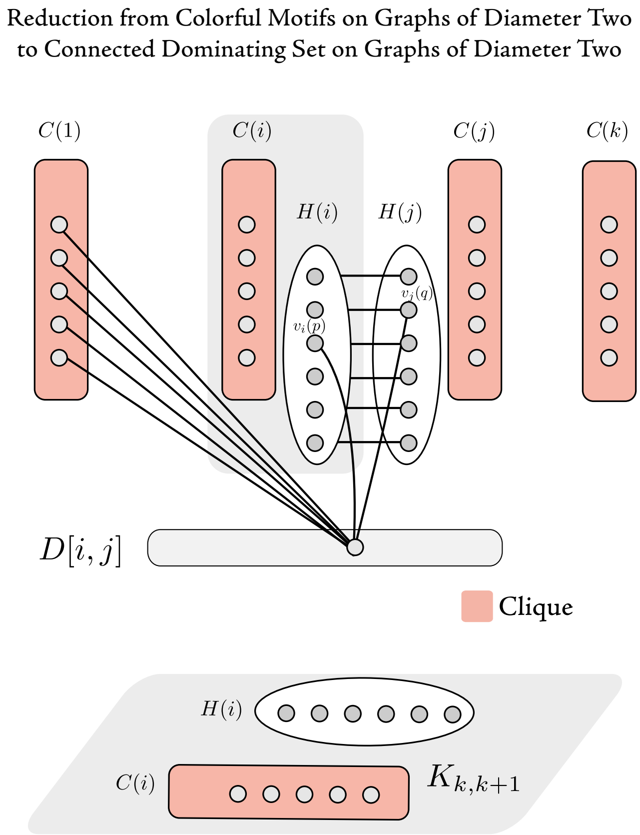

Fig. 9. An Illustration of the Reduction in the proof of Theorem 4

	C	H	D
C	G has diameter two	Complete bipartition and matching edges	D is a clique, and a neighbor into D always exists
H	[A Symmetric Case]	Matching edges, or a common neighbor in D	D is a clique, and a neighbor into D always exists
D	[A Symmetric Case]	[A Symmetric Case]	Clique

Table 1. The Diameter Two Argument Summary

(i) $u \in C(i)$ and $w \in C(i)$. In this case the distance between u and w is one, because $C(i)$ is a clique.
(ii) $u \in C(i)$ and $w \in C(j)$, where $j \neq i$. In this case, there exists a path of length at most two because the original graph has diameter two.
(iii) $u \in H(i)$ and $w \in H(i)$. In this case there exists a path of length two because for every vertex $x \in C(i)$, by construction in step 3, (u, x) and (w, x) are both edges in the graph.
(iv) $u \in H(i)$ and $v \in H(j)$, $i \neq j$. Let $u = v_i(p)$ and let $w = v_j(q)$. If $p = q$, the edge introduced by the matching make u and v adjacent, otherwise, there exists a subdivided vertex that is a common neighbor of both u and w, which implies a path of length two.
(v) $u \in H(i)$ and $w \in C(i)$. By step 3 in the construction, (u, w) is an edge in the graph.
(vi) $u \in H(i)$ and $w \in C(j)$, where $j \neq i$. Let $u = v_i(p)$. Then $(u, v_j(p))$ is a matching edge and $(v_j(p), w)$ is an edge (by step 3 in the construction). This establishes a path of length two.
(vii) $u \in D$ and $w \in D$. In this case the distance between u and w is one, because D is a clique by step 5 of construction.
(viii) $u \in D(i, j)$ and $w \in C(l)$, where $l \neq i$ and $l \neq j$. In this case the distance between u and w is one, because u is global to $C(l)$ by step 6 of construction.
(ix) $u \in D(i, j)$ and $w \in C(i)$ (respectively, $w \in C(j)$). u is adjacent to a vertex in $H(i)$ (respectively, in $H(j)$), which is in turn adjacent to w. This establishes a path of length two.

(x) $u \in D$ and $w \in H(j)$ for some j. Either u is adjacent to w, or is adjacent to a vertex that is adjacent to w (recall that w is adjacent to at least one vertex in D and D induces a clique).

This case analysis establishes that H is indeed a graph of diameter two.

Now we prove that any colorful subtree of G is a connected dominating set of H. It is clearly connected. Notice that all vertices in $H(i)$ are dominated for all $i \in [k]$: any vertex in $H(i)$ is adjacent to all vertices in $C(i)$ and a colorful tree contains one vertex from $C(i)$. A similar argument shows that all vertices in D are dominated.

Conversely, we argue the graph induced on any connected dominating set of size at most k induces a colorful subtree in G. Notice that it suffices to prove that any connected dominating set intersects non-trivially with every $C(i)$, that is, if S is a connected dominating set of size at most k, then:

$$|S \cap C(i)| \geq 1.$$

Because $|S| \leq k$, this implies that $|S \cap C(i)| = 1$. Because S is connected, the vertices of S induce a colorful subtree. Therefore, we now only need to establish that $|S \cap C(i)| \geq 1$. For the sake of contradiction, suppose not. In particular, let

$$|S \cap C(i)| = 0$$

Notice that no vertex $u \notin C(i)$ dominates more than one vertex in $H(i)$. Indeed:

(a) If $u \in H(i)$, then because $H(i)$ induces an independent set, u dominates no vertex other than itself.
(b) If $u \in D$, u has at most one neighbor in $H(i)$, and therefore dominates at most one vertex of $H(i)$.
(c) If $u \in H(j)$, $j \neq i$, u has at most one neighbor in $H(i)$ (indeed, the matching partner is the only one), and therefore dominates at most one vertex of $H(i)$.
(d) It is easy to see that all other vertices have no neighbors in $H(i)$.

Therefore, any dominating set that does not intersect with $C(i)$ is forced to pick more than k vertices to dominate all vertices in $H(i)$, contradicting the assumption that S was a dominating set of size at most k. This completes the proof.

□

8 Summary, Conclusions and Further Directions

We studied the problem of Colorful Motif on various graph classes. We proved that the problem of Colorful Motif restricted to superstars is NP-complete. We also showed NP-completeness on graphs of diameter two. We applied this result towards settling the classical complexity of Connected Dominating Set on graphs of diameter two — specifically, we show that it is NP-complete. Further, we showed that on graphs of diameter two, the problem is NP-complete *and* is unlikely to admit a polynomial kernel.

Next, we showed that obtaining polynomial kernels for Colorful Motif on comb graphs is infeasible, but we show the existence of n polynomial kernels. Further, we study the problem of Colorful Motif on trees, where we observe that the natural strategies for many polynomial kernels are not successful. For instance, we show that "guessing" a root vertex, which helped in the case of comb graphs, fails as a strategy because the Rooted Colorful Motif problem has no polynomial kernels on trees. In fact, this lower bound holds even on rooted binary trees. We summarize our results about Colorful Motif in special graph classes in the following theorem:

Theorem 5. *1. On the class of comb graphs,* Colorful Motif *is NP-complete and* Rooted Colorful Motif *has an $\mathcal{O}(k^2)$ kernel. Equivalently,* Colorful Motif *has n kernels of size $\mathcal{O}(k^2)$ each.*
2. Rooted Colorful Motif *does not admit a polynomial kernel on binary rooted trees, unless $NP \subseteq coNP/Poly$.*
3. Subset Colorful Motif *does not admit a polynomial kernel on trees unless $NP \subseteq coNP/Poly$.*

Finally, we leave the following problems open:

1. Does the Colorful Motif problem on superstars admit a polynomial kernel?
2. Does Colorful Motif admit many polynomial kernels when restricted to trees?
3. What is the parameterized complexity of Connected Dominating Set on graphs of diameter two?

Acknowledgements. We thank our anonymous reviewers for their painstakingly detailed comments for improving the paper. We especially thank two reviewers who pointed out to us the current composition algorithm for the COLORFUL MOTIF problem on comb graphs, which is significantly simpler than our original argument.

References

1. Hans L. Bodlaender, Rodney G. Downey, Michael R. Fellows, and Danny Hermelin. On problems without polynomial kernels. *Journal of Computer and System Sciences*, 75(8):423–434, 2009.
2. Hans L. Bodlaender, Stéphan Thomassé, and Anders Yeo. Kernel Bounds for Disjoint Cycles and Disjoint Paths. In Amos Fiat and Peter Sanders, editors, *Proceedings of ESA 2009*, volume 5757 of *Lecture Notes in Computer Science*, pages 635–646. Springer, 2009.
3. Sharon Bruckner, Falk Hüffner, Richard M. Karp, Ron Shamir, and Roded Sharan. Topology-free querying of protein interaction networks. In Serafim Batzoglou, editor, *Proceedings of RECOMB 2009*, volume 5541 of *LNCS*, pages 74–89. Springer, 2009.
4. Jin-yi Cai, Venkatesan T. Chakaravarthy, Lane A. Hemaspaandra, and Mitsunori Ogihara. Competing provers yield improved Karp-Lipton collapse results. *Information and Computation*, 198(1):1–23, 2005.
5. Marek Cygan, Marcin Pilipczuk, Micha Pilipczuk, and Jakub O. Wojtaszczyk. Kernelization hardness of connectivity problems in d-degenerate graphs. Accepted at WG 2010, 2010.
6. Michael Dom, Daniel Lokshtanov, and Saket Saurabh. Incompressibility through Colors and IDs. In *Proceedings of ICALP 2009*, volume 5555 of *LNCS*, pages 378–389. Springer, 2009.
7. Rod G. Downey and M. R. Fellows. *Parameterized Complexity.* Springer, 1999.
8. Michael R. Fellows, Guillaume Fertin, Danny Hermelin, and Stéphane Vialette. Sharp tractability borderlines for finding connected motifs in vertex-colored graphs. In *Proceedings of ICALP 2007*, volume 4596 of *LNCS*, pages 340–351, 2007.
9. Henning Fernau, Fedor V. Fomin, Daniel Lokshtanov, Daniel Raible, Saket Saurabh, and Yngve Villanger. Kernel(s) for problems with no kernel: On out-trees with many leaves. In *Proceedings of STACS 2009*, pages 421–432, 2009.
10. J. Flum and M. Grohe. *Parameterized Complexity Theory.* Springer-Verlag New York Inc, 2006.
11. Lance Fortnow and Rahul Santhanam. Infeasibility of instance compression and succinct PCPs for NP. In *Proceedings of STOC 2008*, pages 133–142. ACM, 2008.
12. Sylvain Guillemot and Florian Sikora. Finding and counting vertex-colored subtrees. In *Proceedings of MFCS 2010*, volume 6281 of *LNCS*, pages 405–416. Springer, 2010.
13. Richard M. Karp. Reducibility among combinatorial problems. In *Complexity of Computer Communications*, pages 85–103, 1972.
14. Vincent Lacroix, Cristina G. Fernandes, and Marie-France Sagot. Motif search in graphs: Application to metabolic networks. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 3(4):360–368, October 2006.
15. Daniel Lokshtanov. *New Methods in Parameterized Algorithms and Complexity.* PhD thesis, University of Bergen, Norway, 2009.
16. Rolf Niedermeier. *Invitation to Fixed Parameter Algorithms (Oxford Lecture Series in Mathematics and Its Applications).* Oxford University Press, USA, 2006.