Finding and counting vertex-colored subtrees

Guillemot Sylvain, Florian Sikora

Finding and counting vertex-colored subtrees

Sylvain Guillemot and Florian Sikora

Université Paris-Est, LIGM - UMR CNRS 8049, France
guillemo@univ-mlv.fr, sikora@univ-mlv.fr

Abstract. The problems studied in this article originate in the Graph Motif problem introduced by [14] in the context of biological networks. The problem is to decide if a vertex-colored graph has a connected subgraph whose colors equals a given multiset of colors M. Using an algebraic framework recently introduced in [12,13], we obtain new FPT algorithms for Graph Motif and variants, with improved running times. We also obtain results on the counting versions of this problem, showing that the counting problem is FPT if M is a set, but becomes #W[1]-hard if M is a multiset with two colors.

1 Introduction

An emerging field in the modern biology is the study of the biological networks, which represent the interactions between biological elements [1]. A network is modeled by a vertex-colored graph, where nodes represent the biological compounds, edges represent their interactions, and colors represent functionalities of the graph nodes. Networks are often analyzed by studying their *network motifs*, which are defined as small recurring subnetworks. Motifs generally correspond to a set of elements realizing a same function, and which may have been evolutionarily preserved. Therefore, the discovery and the querying of motifs is a crucial problem [17], since it can help to decompose the network into functional modules, to identify conserved elements, and to transfer biological knowledge across species.

The initial definition of network motifs involves conservation of the topology and of the node labels; hence, looking for topological motifs is roughly equivalent to subgraph isomorphism, and thus is a computationally difficult problem. However, in some situations, the topology is not known or is irrelevant, which leads to searching *functional* motifs instead of *topological* ones. In this setting, we still ask for the conservation of the node labels, but we replace topology conservation by the weaker requirement that the subnetwork should form a connected subgraph of the target graph. This approach was advocated by [14] and led to the definition of the Graph Motif problem [7]: given a vertex-colored graph $G = (V, E)$ and a multiset of colors M, find a set $V' \subseteq V$ such that the induced subgraph $G[V']$ is connected, and the multiset of colors of the vertices of V' is equal to M. In the literature, a distinction is made between the *colorful* case (when M is a set), and the *multiset* case (when M is arbitrary). Although this

problem has been introduced for biological motivations, [2] points out that it may also be used in social or technical networks.

Not surprisingly, Graph Motif is NP-hard [14,7], even if the network is a tree with maximum degree 3 and M is a set. The problem is also NP-hard if G is a bipartite graph with maximum degree 4 and M is built over two colors only. The difficulty of this problem is counterbalanced by its fixed-parameter tractability when the parameter is k, the size of the solution [14,7,2]. The currently fastest FPT algorithms for the problem run in $\mathcal{O}^*(2^k)$ time for the colorful case, $\mathcal{O}^*(4.32^k)$ time for the multiset case, and use exponential space (the $\mathcal{O}^*$ notation suppresses $poly(n,k)$ factors). However, when the number of distinct colors is taken as parameter, the problem becomes W[1]-hard [7], ruling out the possibility of an FPT algorithm for this parameter.

Our contribution is twofold. First, we consider in Section 3 the *decision* versions of the Graph Motif problem, as well as some variants: we obtain improved FPT algorithms for these problems, by using the algebraic framework of *multilinear detection* for arithmetic circuits [12,13], presented in the next section. Second, we investigate in Section 4 the *counting* versions of the Graph Motif problem: instead of deciding if a motif appears in the graph, we now want to count the occurrences of this motif. This allows to assess if a motif is over- or under- represented in the network, by comparing the actual count of the motif to its expected count under a null hypothesis [16]. We show that the counting problem is FPT in the colorful case, but becomes #W[1]-hard for the multiset case with two colors. We refer the reader to [9,8] for definitions related to parameterized counting classes.

2 Definitions

This section contains definitions related to arithmetic circuits, and to the Multilinear Detection (MLD) problem. It concludes by stating Theorem 1, which will be used throughout the paper.

2.1 Arithmetic circuits

In the following, a capital letter X will denote a set of variables, and a lowercase letter x will denote a single variable. If X is a set of variables and $\mathbb{A}$ is a commutative ring, we denote by $\mathbb{A}[X]$ the ring of multivariate polynomials with coefficients in $\mathbb{A}$ and involving variables of X. Given a monomial $m = x_1...x_k$ in $\mathbb{A}[X]$, where the x_is are variables, its *degree* is k, and m is *multilinear* iff its variables are distinct.

An *arithmetic circuit* over X is a pair $\mathcal{C} = (C, r)$, where C is a labeled dag such that (i) the children of each node are totally ordered, (ii) the nodes are labeled either by $op \in \{+, \times\}$ or by an element of X, (iii) no internal node is labeled by an element of X, and where r is a distinguished node of C called the *root*. We let V_C be the set of nodes of C, and for a given node u we denote by

$N_C(u)$ the set of children (*i.e.* out-neighbors) of u in C. We recall that a node u is called a *leaf* of C iff $N_C(u) = \emptyset$, an *internal node* otherwise.

Given a commutative ring $\mathbb{A}$ and a mapping $\phi : X \to \mathbb{A}$, *evaluating* $\mathcal{C}$ *under* ϕ consists in

1. computing, for each node u of C, a value $v(u) \in \mathbb{A}$ as follows:

$$v(u) = \sum_{u' \in N_C(u)} v(u') \text{ if } u \text{ is labeled by } +$$

$$v(u) = \prod_{u' \in N_C(u)} v(u') \text{ if } u \text{ is labeled by } \times$$

$$v(u) = \phi(x) \text{ if } u \text{ is a leaf labeled by } x \in X$$

   where the operations are carried out in $\mathbb{A}$. By convention, empty sums evaluate to $0_{\mathbb{A}}$, and empty products evaluate to $1_{\mathbb{A}}$.
2. returning the value $v(r)$ as the result of the evaluation.

Observe that if operations in $\mathbb{A}$ require $\mathcal{O}(t)$ time and $\mathcal{O}(s)$ space, then the above evaluation can be performed in $\mathcal{O}(t|\mathcal{C}|)$ time and $\mathcal{O}(s|\mathcal{C}|)$ space, where $|\mathcal{C}|$ is the size of $\mathcal{C}$ (defined as the number of arcs). The *symbolic evaluation* of $\mathcal{C}$ is the polynomial $P_{\mathcal{C}} \in \mathbb{Z}[X]$ obtained by evaluating $\mathcal{C}$ under the function $\phi : X \to \mathbb{Z}[X]$ defined by $\phi(x) = x$.

2.2 Multilinear Detection

Informally, the Multilinear Detection problem asks, for a given arithmetic circuit $\mathcal{C}$ and an integer k, if the polynomial $P_{\mathcal{C}}$ has a multilinear monomial of degree k. However, this definition does not give a certificate checkable in polynomial-time, so for technical reasons we define the problem differently.

A *monomial-subtree* of $\mathcal{C}$ is a pair $T = (\mathcal{C}', \phi)$, where $\mathcal{C}' = (C', r')$ is an arithmetic circuit over X whose underlying dag C' is a directed tree, and where $\phi : V_{\mathcal{C}'} \to V_{\mathcal{C}}$ is such that (i) $\phi(r') = r$, (ii) if $u \in V_{\mathcal{C}'}$ is labeled by $x \in X$, then so is $\phi(u)$, (iii) if $u \in V_{\mathcal{C}'}$ is labeled by $+$ then so is $\phi(u)$, and $N_{\mathcal{C}'}(u)$ consists of a single element $v \in N_{\mathcal{C}}(\phi(u))$, (iv) if $u \in V_{\mathcal{C}'}$ is labeled by $\times$, then so is $\phi(u)$, and ϕ maps bijectively $N_{\mathcal{C}'}(u)$ into $N_{\mathcal{C}}(\phi(u))$ by preserving the ordering on siblings. The *variables* of T are the leaves of C' labeled by variables in X. We say that T is *distinctly-labeled* iff its variables are distinct. Intuitively, the (distinctly-labeled) monomial-subtrees of $\mathcal{C}$ with k variables correspond to the (multilinear) monomials of $P_{\mathcal{C}}$ having degree k. Therefore, we formulate the problem Multilinear Detection as follows:

Name: Multilinear Detection (MLD)
Input: an arithmetic circuit $\mathcal{C}$ over a set of variables X, an integer k
Solution: a distinctly-labeled monomial-subtree of $\mathcal{C}$ with k variables.

Then solving MLD amounts to decide if $P_{\mathcal{C}}$ has a multilinear monomial of degree k (observe that there are no possible cancellations), and solving #MLD amounts to compute the sum of the coefficients of multilinear monomials of $P_{\mathcal{C}}$ having degree k. The restriction of MLD when $|X| = k$ is called Exact Multilinear Detection (XMLD). In this article, we will rely on the following far-reaching result from [18,13] to obtain new algorithms for Graph Motif:

Theorem 1. *[18,13]* MLD *can be solved by a randomized algorithm using* $\tilde{\mathcal{O}}(2^k|\mathcal{C}|)$ *time and* $\tilde{\mathcal{O}}(|\mathcal{C}|)$ *space.*

Here, we use the $\tilde{\mathcal{O}}$ notation to suppress polylogarithmic factors, i.e. factors of the form $O((\log n)^c)$ where n is the instance size and c is a constant. By a "randomized algorithm" with running time $\mathcal{O}(t)$, we mean an algorithm which (i) always answers no on negative instances, (ii) answers yes with probability $\geq \frac{1}{2}$ on positive instances, (iii) always runs in time $\mathcal{O}(t)$ regardless of the random choices made in an execution.

3 Finding vertex-colored subtrees

In this section, we consider several variants of the Graph Motif problem, and we obtain improved FPT algorithms for these problems by reduction to MLD. Notably, we obtain $\mathcal{O}^*(2^k)$ time algorithms for problems involving *colorful* motifs, and $\mathcal{O}^*(4^k)$ time algorithms for *multiset* motifs.

3.1 The colorful case

In the colorful formulation of the problem, the graph is vertex-colored, and we seek a subtree with k vertices having distinct colors. This leads to the following formal definition:

Name: Colorful Graph Motif (CGM)
Input: a graph $G = (V, E)$, $k \in \mathbb{N}$, a set C, a function $\chi : V \to C$
Solution: a subtree $T = (V_T, E_T)$ of G s.t. (i) $|V_T| = k$ and (ii) for each $u, v \in V_T$ distinct, $\chi(u) \neq \chi(v)$.

The restriction of Colorful Graph Motif when $|C| = k$ is called Exact Colorful Graph Motif (XCGM). Note that this restriction requires that the vertices of T are bijectively labeled by the colors of C. In [4], the XCGM problem was shown to be solvable in $\mathcal{O}^*(2^k)$ time and space, while it is not difficult to see that the general CGM problem can be solved in $\mathcal{O}^*((2e)^k)$ time and $\mathcal{O}^*(2^k)$ space by color-coding. By using a reduction to Multilinear Detection, we improve upon these complexities:

Proposition 1. CGM *is solvable by a randomized algorithm in* $\tilde{\mathcal{O}}(2^k k^2|G|)$ *time and* $\tilde{\mathcal{O}}(k^2|G|)$ *space.*

Proof. Let I be an instance of CGM. We construct the following circuit $\mathcal{C}_I$: its set of variables is $\{x_c : c \in C\}$, and we introduce intermediary nodes $P_{i,u}$ for $1 \leq i \leq k, u \in V$, as well as a root node P. The definitions are as follows:

$$P_{i,u} = \sum_{i'=1}^{i-1} \sum_{v \in N_G(u)} P_{i',u} P_{i-i',v} \text{ if } i > 1, \qquad P_{1,u} = x_{\chi(u)}$$

and $P = \sum_{u \in V} P_{k,u}$. The resulting instance of MLD is $I' = (\mathcal{C}_I, k)$. By applying Theorem 1, and by observing that $|\mathcal{C}_I| = \mathcal{O}(k^2|G|)$, we solve I' in $\tilde{\mathcal{O}}(2^k k^2 |G|)$ time and $\tilde{\mathcal{O}}(k^2|G|)$ space.

It remains to show the correctness of the reduction. Given a set $S \subseteq C$, define the multilinear monomial $\pi_S := \prod_{c \in S} x_c$. Given $u \in V(T)$ and $S \subseteq C$, an (u, S)-*solution* is a subtree $T = (V_T, E_T)$ of G, such that $u \in V_T$, T is distinctly colored by χ, and $\chi(V_T) = S$. We show by induction on $1 \leq i \leq k$ that: given $S \subseteq C$, π_S is a multilinear monomial of $P_{i,u}$ iff (i) $|S| = i$ and (ii) there exists an (u, S)-solution. This is clear when $i = 1$; now, suppose that $i \geq 2$, and assume that the property holds for every $1 \leq j < i$.

Suppose that $|S| = i$ and that $T = (V_T, E_T)$ is an (u, S)-solution, let us show that π_S is a multilinear monomial of $P_{i,u}$. Let v be a neighbor of u in T, then removing the edge uv from T produces two trees T_1, T_2 with T_1 containing u and T_2 containing v. These two trees are distinctly colored, let S_1, S_2 be their respective color sets, and let i_1, i_2 be their respective sizes. Since T_1 is an (u, S_1)-solution, π_{S_1} is a multilinear monomial of $P_{i_1,u}$ by induction hypothesis. Since T_2 is an (v, S_2)-solution, π_{S_2} is a multilinear monomial of $P_{i_2,v}$. It follows that $\pi_S = \pi_{S_1}\pi_{S_2}$ is a multilinear monomial of $P_{i_1,u}P_{i_2,v}$, and thus of $P_{i,u}$.

Conversely, suppose that π_S is a multilinear monomial of $P_{i,u}$. By definition of $P_{i,u}$, there exists $1 \leq i' \leq i-1$ and $v \in N_G(u)$ such that π_S is a multilinear monomial of $P_{i',u}P_{i-i',v}$. We can then partition S into S_1, S_2, with π_{S_1} multilinear monomial of $P_{i',u}$ and π_{S_2} multilinear monomial of $P_{i-i',v}$. Induction hypothesis therefore implies that (i) $|S_1| = i'$ and $|S_2| = i - i'$, (ii) there exists an (u, S_1)-solution $T_1 = (V_1, E_1)$ and a (v, S_2)-solution $T_2 = (V_2, E_2)$. Since S_1, S_2 are disjoint, it follows that $|S| = i$, which proves (i); besides, V_1, V_2 are disjoint, and thus $T = (V_1 \cup V_2, E_1 \cup E_2 \cup \{uv\})$ is an (u, S)-solution, which proves (ii). □

3.2 The multiset case and variants

We now consider several variants of the CGM problem. The first two variants allow for *multiset motifs*: instead of seeking a subtree with distinct colors, we now allow some colors to be repeated but impose a maximum number of occurrences for each color. This problem can be seen as a generalization of the original Graph Motif problem.

Given a multiset M over a set A, and given an element $x \in A$, we denote by $n_M(x)$ the number of occurrences of x in M. Given two multisets M, M', we denote their inclusion by $M \subseteq M'$. We denote by $|M|$ the size of M, where elements are counted with their multiplicities. Given two sets A, B, a function

$f : A \to B$ and a multiset X over A, we let $f(X)$ denote the multiset containing the elements $f(x)$ for $x \in X$, counted with multiplicities; precisely, given $y \in B$ we have $n_{f(X)}(y) = \sum_{x \in A : f(x)=y} n_X(x)$.

We now define the two following variants of Colorful Graph Motif, which allow for multiset motifs:

Name: Multiset Graph Motif (MGM)
Input: a graph $G = (V, E)$, an integer k, a set C, a function $\chi : V \to C$, a multiset M over C.
Solution: a subtree $T = (V_T, E_T)$ of G s.t. (i) $|V_T| = k$ and (ii) $\chi(V_T) \subseteq M$.

Name: Multiset Graph Motif With Gaps (MGMG)
Input: a graph $G = (V, E)$, integers k, r, a set C, a function $\chi : V \to C$, a multiset M over C.
Solution: a subtree $T = (V_T, E_T)$ of G s.t. (i) $|V_T| \leq r$ and (ii) there exists $S \subseteq V_T$ of size k such that $\chi(S) \subseteq M$.

The restriction of Multiset Graph Motif when $|M| = k$ is called Exact Multiset Graph Motif (XMGM). Note that in this case we require that T contains every occurrence of M, *i.e.* $\chi(V_T) = M$. In this way, the XMGM problem coincides with the Graph Motif problem defined in [7,2], while the problem MGM is the parameterized version of the problem Max Motif considered in [6]. The notion of gaps is introduced in [14], and coincides with the notion of insertions and deletions of [4].

Previous algorithms for these problems relied on color-coding; these algorithms usually have an exponential space complexity, and a high time complexity. For the Graph Motif problem, [7] gives a randomized algorithm with an implicit $\mathcal{O}(87^k km)$ running time, while [2] describes a first randomized algorithm running in $\mathcal{O}(8.16^k m)$, and show a second algorithm with running time $\mathcal{O}(4.32^k k^2 m)$, using two different speed-up techniques ([3] and [10]). For the Max Motif problem, [6] present a randomized algorithm with an implicit $\mathcal{O}((32\mathrm{e}^2)^k km)$ running time. Here again, we can apply Theorem 1 to improve the time and space complexities:

Proposition 2. *For the given multiset M, let $c_{max} = \max_{c \in C} n_M(c)$ be the maximum number of occurrences of a color in M.*

1. MGM *is solvable by a randomized algorithm in $\tilde{\mathcal{O}}(4^k k^2 c_{max}|G|)$ time and $\tilde{\mathcal{O}}(k^2 c_{max}|G|)$ space.*
2. MGMG *is solvable by a randomized algorithm in $\tilde{\mathcal{O}}(4^k r^2 c_{max}|G|)$ time and $\tilde{\mathcal{O}}(r^2 c_{max}|G|)$ space.*

The proof is omitted due to space constraints. We point out that the proof can be adapted to solve the List Colored Graph Motif from [2] in $\mathcal{O}^*(4^k)$ time and polynomial space. This improves upon an randomized algorithm of [2] which runs in $\mathcal{O}(10.88^k m)$ time and exponential space.

We now consider two other variants of the problem, where weights are assigned to the edges, and where we seek a subtree with minimum weight. We obtain two problems, depending on whether we consider colorful or multiset motifs.

Name: Weighted Colorful Graph Motif (WCGM)
Input: a complete graph $G = (V, E)$, a function $\chi : V \to C$, a weight function $w : E \to \mathbb{N}$, integers k, r
Solution: a subtree $T = (V_T, E_T)$ of G such that (i) $|V_T| = k$, (ii) χ is injective on V_T, (iii) $\sum_{e \in E_T} w(e) = r$.

Name: Weighted Multiset Graph Motif (WMGM)
Input: a complete graph $G = (V, E)$, a function $\chi : V \to C$, a weight function $w : E \to \mathbb{N}$, integers k, r, a multiset M
Solution: a subtree $T = (V_T, E_T)$ of G such that (i) $|V_T| = k$, (ii) $\chi(V_T) \subseteq M$, (iii) $\sum_{e \in E_T} w(e) = r$.

We observe that the WMGM problem contains as particular case the Min-CC problem introduced in [5], which seeks a subgraph respecting the multiset motif, and having at most r connected components. Indeed, we can easily reduce Min-CC to WMGM: given the graph G, we construct a complete graph G' with the same vertex set, and we assign a weight 0 to edges of G, and a weight 1 to non-edges of G.

Proposition 3. *1.* WCGM *is solvable by a randomized algorithm in* $\tilde{\mathcal{O}}(2^k k^2 r^2 |G|)$ *time and* $\tilde{\mathcal{O}}(k^2 r^2 |G|)$ *space.*
2. WMGM *is solvable by a randomized algorithm in* $\tilde{\mathcal{O}}(4^k k^2 r^2 c_{max} |G|)$ *time and* $\tilde{\mathcal{O}}(k^2 r^2 c_{max} |G|)$ *space.*

Proof. We only prove 1, since 2 relies on the same modification as in Proposition 2. The construction of the arithmetic circuit is similar to the construction in Proposition 1. The difference is that, in addition of the number of nodes of the subtree, we also need to memorize the maximum total weight. This leads to introduce nodes $P_{i,j,u}$, for $1 \leq i \leq k$ and $0 \leq j \leq r$. The definitions are as follows:

$$P_{i,j,u} = \sum_{i'=1}^{i-1} \sum_{v \in V} \sum_{j'=0}^{j-w(uv)} P_{i',j',u} P_{i-i',j-j'-w(uv),v} \text{ if } i > 1, \qquad P_{1,j,u} = x_{\chi(u)}$$

and $P = \sum_{u \in V} P_{k,r,u}$. The resulting instance of MLD is $I' = (\mathcal{C}_I, k)$, and since $|\mathcal{C}_I| = \mathcal{O}(k^2 r^2 |G|)$, we solve I' in $\tilde{\mathcal{O}}(2^k k^2 r^2 |G|)$ time and $\tilde{\mathcal{O}}(k^2 r^2 |G|)$ space by Theorem 1. The correctness of the reduction follows by showing by induction on i that: given $1 \leq i \leq k, 1 \leq j \leq r$, $u \in V$ and $S \subseteq C$ of size i, π_S is a multilinear monomial of $P_{i,j,u}$ iff there exists an (u, S)-solution T with $\sum_{e \in E_T} w(e) = j$. □

4 Counting vertex-colored subtrees

In this section, we consider the counting versions of the problems XCGM and XMGM introduced in Section 3. For the former, we show that its counting version #XCGM is FPT; for the latter, we prove that its counting version #XMGM is #W[1]-hard.

4.1 FPT algorithms for the colorful case

We show that #XCGM is fixed-parameter tractable (Proposition 5). We rely on a general result for #XMLD (Proposition 4), which uses inclusion-exclusion as in [11].

Say that a circuit $\mathcal{C}$ is *k-bounded* iff $P_{\mathcal{C}}$ contains no monomials of degree $> k$. Observe that given a circuit $\mathcal{C} = (C, r)$, we can efficiently transform it in a k-bounded circuit $\mathcal{C}'$ such that (i) $\mathcal{C}$ and $\mathcal{C}'$ have the same monomials of degree k, (ii) $|\mathcal{C}'| \leq (k+1)^2|\mathcal{C}|$. Indeed, we can first transform $\mathcal{C}$ so that all $+$ and $\times$ nodes have out-degree 2, without increasing the size; then, for each node u of $\mathcal{C}$, we create $k+1$ nodes $u_0, ..., u_k$, and:

- if u is a leaf with label $v \in X$, then u_1 is a leaf with label v, and other u_i's are 0 nodes (represented by leaves labeled by $+$);
- if $u = v + w$, then for every i, $u_i = v_i + w_i$;
- if $u = v \times w$, then for every i, $u_i = \sum_{j=0}^{i} v_j w_{i-j}$.

Let $\mathcal{C}'$ be the resulting circuit, whose root is r_k. It is easily checked that $\mathcal{C}'$ has the same monomials of degree k as the original circuit $\mathcal{C}$. Besides, $|\mathcal{C}'| \leq (k+1)^2|\mathcal{C}|$ since for each node u of out-degree 2 in $\mathcal{C}$, we have introduced $k+1$ nodes each of out-degree $\leq k+1$ in $\mathcal{C}'$.

The following result shows that we can efficiently count solutions for k-bounded circuits with k variables (and thus for general circuits, with an extra $\mathcal{O}(k^2)$ factor in the complexity). We precise that Koutis and Williams give independently this result, with a different proof, in the full version of [13] (to appear).

Proposition 4. #XMLD *for k-bounded circuits is solvable in $\mathcal{O}(2^k k^2|\mathcal{C}|)$ time and $\mathcal{O}(k|\mathcal{C}|)$ space.*

Proof. Let $\mathcal{C}$ be the input circuit on a set X of k variables, let r be its root node and let $P = P_{\mathcal{C}} \in \mathbb{Z}[X]$ be the symbolic evaluation of $\mathcal{C}$. Let M_k be the multiset of monomials of P of degree k, and for a monomial m let $\phi(m) \subseteq X$ be its set of variables. For every $S \subseteq X$, let $N_S = |\{m \in M_k : \phi(m) = S\}|$ and $N'_S = |\{m \in M_k : \phi(m) \subseteq S\}|$. Observe that for every $S \subseteq X$, we have $N'_S = \sum_{T \subseteq S} N_T$; therefore, by Moebius inversion it holds that for every $S \subseteq X$, $N_S = \sum_{T \subseteq S} (-1)^{|S \setminus T|} N'_T$.

We now show how to compute a value N'_S for $S \subseteq X$. Consider the ring homomorphism H_S from $\mathbb{Z}[X]$ to $\mathbb{Z}[x]$ defined by:

$$\begin{cases} H_S(v) & = x \text{ if } v \in S, \\ H_S(v) & = 1 \text{ if } v \notin S \end{cases}$$

Since $\mathcal{C}$ is k-bounded, the coefficient of x^k in $H_S(P)$ is equal to N'_S. Now, $H_S(P)$ can be computed efficiently, by evaluating $\mathcal{C}$ under the valuation H_S. During this computation, each intermediary result is an element of $\mathbb{Z}[x]$ of maximum degree k; therefore, each polynomial is stored in $\mathcal{O}(k)$ space, and operations on these polynomials can be performed in $\mathcal{O}(k^2)$ time. We deduce that N'_S can be computed in $\mathcal{O}(k^2|\mathcal{C}|)$ time and $\mathcal{O}(k|\mathcal{C}|)$ space.

We conclude by noting that the number of multilinear monomials of degree k in P is equal to N_X. As we have shown that each value N'_S can be computed in $\mathcal{O}(k^2|\mathcal{C}|)$ time and $\mathcal{O}(k|\mathcal{C}|)$ space, we can thus compute N_X in $\mathcal{O}(2^k k^2|\mathcal{C}|)$ time and $\mathcal{O}(k|\mathcal{C}|)$ space. □

We find interesting to point out that Proposition 4 generalizes several counting algorithms based on inclusion-exclusion, such as the well-known algorithm for #Hamiltonian Path of [11], as well as results of [15]. Indeed, several of these problems can be reduced to counting multilinear monomials of degree n (where n is usually the number of vertices of the graph), which leads to algorithms running in $\mathcal{O}^*(2^n)$ time and polynomial space for these problems. However, our goal here is to obtain a $\mathcal{O}^*(2^k)$ algorithm for the #XCGM problem.

Proposition 5. #XCGM *is solvable in* $\mathcal{O}(2^k k^4|G|)$ *time and* $\mathcal{O}(k^3|G|)$ *space.*

Proof. Let I be an instance of XCGM. A *rooted solution* for I is a pair (u, T) where T is a solution of XCGM on I and u is a vertex of T (which must be seen as the root of the tree). The solutions of XCGM on I are also called *unrooted solutions*. Let $N_r(I)$ and $N_u(I)$ be the number of rooted, resp. unrooted, solutions for I. We will show how to compute $N_r(I)$ in the claimed time and space bounds; since $N_u(I) = \frac{N_r(I)}{k}$, the result will follow.

To compute N_r, observe first that we cannot apply Proposition 4 to the circuit $\mathcal{C}_I$ of Proposition 1. Indeed, the circuit $\mathcal{C}_I$ counts the ordered subtrees, and not the unordered ones. Therefore, we need to modify the circuit in the following way: at each vertex v of V_T, we examine its children in a fixed order, for instance by increasing color. This leads us to define the following circuit $\mathcal{C}'_I$: suppose w.l.o.g. that $C = \{1, ..., k\}$, introduce nodes $P_{i,j,u}$ for each $1 \leq i \leq k, 1 \leq j \leq k+1, u \in V$, variables x_i for each $1 \leq i \leq k$, and define:

$$P_{1,j,u} = x_{\chi(u)}, \quad P_{i,j,u} = 0 \text{ if } i \geq 2, j = k+1$$

$$P_{i,j,u} = P_{i,j+1,u} + \sum_{i'=1}^{i-1} \sum_{v \in N_G(u):\chi(v)=j} P_{i',j+1,u} P_{i-i',1,v} \text{ if } i \geq 2, 1 \leq j \leq k$$

Let us also introduce a root node $P = \sum_{u \in V} P_{k,1,u}$. Given $1 \leq i, j \leq k$ and $u \in V$, let $\mathcal{S}_{i,j,u}$ denote the set of pairs (u, T) where (i) T is a properly colored subtree of I containing u and having i vertices, (ii) the neighbors of u in T have colors $\geq j$. It can be shown by induction on i that: there is a bijection between $\mathcal{S}_{i,j,u}$ and the multilinear monomials of $P_{i,j,u}$ of degree i. Therefore, the number of multilinear monomials of $\mathcal{C}'_I$ is equal to N_r; since $|\mathcal{C}'_I| = \mathcal{O}(k^3|G|)$ and since $\mathcal{C}'_I$ is k-bounded, it follows by Proposition 4 that N_r can be computed in $\mathcal{O}(2^k k^4|G|)$ time and $\mathcal{O}(k^3|G|)$ space. □

4.2 Hardness of the Multiset case

In this subsection, we show that #XMGM is #W[1]-hard. For convenience, we first restate the problem in terms of *vertex-distinct embedded subtrees*.

Let $G = (V, E)$ and $H = (V', E')$ be two multigraphs. An *homomorphism* of G into H is a pair $\phi = (\phi_V, \phi_E)$ where $\phi_V : V \to V'$ and $\phi_E : E \to E'$, such that if $e \in E$ has endpoints x, y then $\phi_E(e)$ has endpoints $\phi_V(x), \phi_V(y)$. An *embedded subtree* of G is a pair $\mathcal{T} = (T, \phi_V, \phi_E)$ where $T = (V_T, E_T)$ is a tree, and (ϕ_V, ϕ_E) is an homomorphism from T into G. We say that $\mathcal{T}$ is a *vertex-distinct* embedded subtree of G (a "vdst" of G) if ϕ_V is injective. We say $\mathcal{T}$ is an *edge-distinct* embedded subtree of G (an "edst" of G) iff ϕ_E is injective. We restate XMGM as follows:

Name: Exact Multiset Graph Motif (XMGM)
Input: a graph $G = (V, E)$, an integer k, a set C, a function $\chi : V \to C$, a multiset M over C s.t. $|M| = k$.
Solution: a vdst (T, ϕ_V, ϕ_E) of G s.t. $\chi \circ \phi_V(V_T) = M$.

We first show the hardness of two intermediate problems (Lemma 1). Before defining these problems, we need the following notions. Consider a multigraph $G = (V, E)$. Consider a partition $\mathcal{P}$ of V into $V_1, ..., V_k$, and a tuple $t \in [r]^k$. A $(\mathcal{P}, t)$-*mapping* from a set A is an injection $\psi : A \to V \times [r]$ such that for every $x \in A$, if $\psi(x) = (v, i)$ with $v \in V_j$, then $1 \leq i \leq t_j$. From ψ, we define its *reduction* as the function $\psi^r : A \to V$ defined by $\psi^r(x) = v$ whenever $\psi(x) = (v, i)$. We also define a tuple $T(\psi) = (n_1, ..., n_k) \in [r]^k$ such that for each $i \in [k]$, $n_i = \max_{v \in V_i} |\{x \in A : \psi^r(x) = v\}|$.

Given two tuples $t, t' \in [r]^k$, denote $t \leq t'$ iff $t_i \leq t'_i$ for each $i \in [k]$. Note that for a $(\mathcal{P}, t)$-mapping ψ, we always have $T(\psi) \leq t$ since ψ is injective. We say that a $(\mathcal{P}, t)$-*labeled edst for* G is a tuple (T, ψ_V, ψ_E) where (i) $T = (V_T, E_T)$ is a tree, (ii) ψ_V is a $(\mathcal{P}, t)$-mapping from V_T, (iii) (T, ψ_V^r, ψ_E) is an edst of G. Our intermediate problems are defined as follows:

Name: Multicolored Embedded Subtree-1 (MEST $-$ 1)
Input: integers k, r, a k-partite multigraph G with partition $\mathcal{P}$, a tuple $t \in [r]^k$
Solution: a $(\mathcal{P}, t)$-labeled edst (T, ψ_V, ψ_E) for G s.t. $|V_T| = r$ and $T(\psi_V) = t$.

The MEST $-$ 2 problem is defined similarly, except that we do not require that $T(\psi_V) = t$ (and thus we only have $T(\psi_V) \leq t$). While we will only need #MEST $-$ 2 in our reduction for #XMGM, we first show the hardness of #MEST $-$ 1, then reduce it to #MEST $-$ 2.

Lemma 1. #MEST-1 *and* #MEST-2 *are* #W[1]*-hard for parameter* (k, r).

The proof is omitted due to space constraints.

Proposition 6. #XMGM *is* #W[1]*-hard for parameter* k.

Proof. We reduce from #MEST − 2, and conclude using Lemma 1. Let $I = (k, r, G, t)$ be an instance of #MEST − 2, where $G = (V, E)$ is a multigraph, and let $\mathcal{S}_I$ be its set of solutions. From G, we construct a graph H as follows: (i) we subdivide each edge $e \in E$, creating a new vertex $a[e]$, (ii) we substitute each vertex $v \in V_i$ by an independent set formed by t_i vertices $b[v, 1], ..., b[v, t_i]$. We let A be the set of vertices $a[e]$ and B the set of vertices $b[v, i]$, we therefore have a bipartite graph $H = (A \cup B, F)$. We let $I' = (H, 2r - 1, C, \psi, M)$, where $C = \{1, 2\}$, ψ maps A to 1 and B to 2, and M consists of $r - 1$ occurrences of 1 and r occurrences of B.

Then I' is our resulting instance of #XMGM, and we let $\mathcal{S}_{I'}$ be its set of solutions. Notice that by definition of ψ and M, $\mathcal{S}_{I'}$ is the set of vdst (T, ϕ_V, ϕ_E) of H containing $r - 1$ vertices mapped to A and r vertices mapped to B. We now show that we have a parsimonious reduction, by describing a bijection $\Phi : \mathcal{S}_I \to \mathcal{S}_{I'}$. Consider $\mathcal{T} = (T, \psi_V, \psi_E)$ in $\mathcal{S}_I$; we define $\Phi(\mathcal{T}) = (T', \phi_V, \phi_E)$ as follows:

- For each edge $e = uv \in E(T)$, we have $f_e := \psi_E(e) \in E(G)$: we then subdivide e, creating a new vertex x_e. Let T' be the resulting tree;
- For each vertex x_e, we define $\phi'_V(x_e) = a[f_e]$. For each other vertex u of T', we have $u \in V(T)$, let $(v, i) = \psi_V(u)$; we then set $\phi'_V(u) = b[v, i]$ (this is possible since if $v \in V_j$ then $1 \leq i \leq t_j$, by definition of ψ_V).

From ϕ_V, we then define ϕ_E in a natural way. Then $\mathcal{T}' = \Phi(\mathcal{T})$ is indeed in $\mathcal{S}_{I'}$: (i) $\mathcal{T}'$ is a vertex distinct subtree of H (by definition of ϕ_V and since $\mathcal{T}$ was edge-distinct, the values $\phi_V(x_e)$ are distinct; by injectivity of ψ_V, the other values $\phi_V(u)$ are distinct); (ii) it has $r-1$ vertices mapped to A and r vertices mapped to B. To prove that Φ is a bijection, we describe the inverse correspondence $\Psi : \mathcal{S}_{I'} \to \mathcal{S}_I$. Consider $\mathcal{T}' = (T', \phi_V, \phi_E)$ in $\mathcal{S}_{I'}$; we define $\Psi(\mathcal{T}') = (T, \psi_V, \psi_E)$ as follows. Let A', B' be the vertices of T' mapped to A, B respectively. Let i be the number of nodes of A' which are leaves: since the nodes of A' have degree 1 or 2 in T' depending on whether they are leaves or internal nodes, we then have $|E(T')| \leq i + 2(r - 1 - i) = 2r - i - 2$; since $|E(T')| = 2r - 2$, we must have $i = 0$. It follows that all leaves of T' belong to B'; from T', by contracting each vertex of A' in T' we obtain a tree T with r vertices. We then define ψ_V, ψ_E as follows: (i) given $u \in B'$, if $\phi_V(u) = b[v, j]$, then $\psi_V(u) = (v, j)$; (ii) given $e = uv \in E(T)$, there corresponds two edges $ux, vx \in E(T')$ with $x \in A'$, and we thus have $\phi_V(x) = a[f]$, from which we define $\psi_E(e) = f$. It is easily seen that the resulting $\mathcal{T} = \Psi(\mathcal{T}')$ is in $\mathcal{S}_I$, and that the operations Φ and Ψ are inverse of each other. □

5 Conclusion

In this paper, we have obtained improved FPT algorithms for several variants of the Graph Motif problem. Reducing to the Multilinear Detection problem yielded a significant reduction of the base of the exponent in the time complexity, as well as a polynomial space complexity. We have also considered the counting versions of these problems, for the first time in the literature.

Though further improvements seem difficult to achieve, we hope that the $\mathcal{O}^*(4^k)$ running time obtained for multiset motifs can be further reduced. Also, while we have shown that #XMGM was #W[1]-hard for a motif with two colors, we leave open its complexity for one color. Note that this problem amounts to count the k-vertex subtrees of an (uncolored) graph.

References

1. E. Alm and A.P. Arkin. Biological networks. *Curr. Opin. Struct. Biol.*, 13(2):193–202, 2003.
2. N. Betzler, M.R. Fellows, C. Komusiewicz, and R. Niedermeier. Parameterized algorithms and hardness results for some graph motif problems. In *CPM 2008*, volume 5029 of *LNCS*, pages 31–43, 2008.
3. A. Björklund, T. Husfeldt, P. Kaski, and M. Koivisto. Fourier meets möbius: fast subset convolution. In *STOC 2007*, pages 67–74, 2007.
4. S. Bruckner, F. Hüffner, R.M. Karp, R. Shamir, and R. Sharan. Topology-free querying of protein interaction networks. In *RECOMB 2009*, volume 5541 of *LNCS*, pages 74–89, 2009.
5. R. Dondi, G. Fertin, and S. Vialette. Weak pattern matching in colored graphs: Minimizing the number of connected components. In *ICTCS 2007*, pages 27–38. World-Scientific, 2007.
6. R. Dondi, G. Fertin, and S. Vialette. Maximum motif problem in vertex-colored graphs. In *CPM 2009*, volume 5577 of *LNCS*, pages 221–235, 2009.
7. M. Fellows, G. Fertin, D. Hermelin, and S. Vialette. Sharp tractability borderlines for finding connected motifs in vertex-colored graphs. In *ICALP 2007*, volume 4596 of *LNCS*, pages 340–351, 2007.
8. J. Flum and M. Grohe. The parameterized complexity of counting problems. *SIAM Journal on Computing*, 33(4):892–922, 2004.
9. J. Flum and M. Grohe. *Parameterized Complexity Theory*. Springer-Verlag, 2006.
10. F. Huffner, S. Wernicke, and T. Zichner. Algorithm Engineering For Color-Coding To Facilitate Signaling Pathway Detection. In *APBC 2007*, pages 277–286, 2007.
11. R.M. Karp. Dynamic-programming meets the principle of inclusion and exclusion. *Oper. Res. Lett.*, 1:49–51, 1982.
12. I. Koutis. Faster Algebraic Algorithms for Path and Packing Problems. In *ICALP 2008*, volume 5125 of *LNCS*, pages 575–586, 2008.
13. I. Koutis and R. Williams. Limits and Applications of Group Algebras for Parameterized Problems. In *ICALP 2009*, volume 5555 of *LNCS*, pages 653–664, 2009.
14. V. Lacroix, C.G. Fernandes, and M.-F. Sagot. Motif search in graphs: application to metabolic networks. *Trans. Comput. Biol. Bioinform.*, 3(4):360–368, 2006.
15. J. Nederlof. Fast Polynomial-Space Algorithms Using Möbius Inversion: Improving on Steiner Tree and Related Problems. In *ICALP 2009*, volume 5555 of *LNCS*, pages 713–725, 2009.
16. S. Schbath, V. Lacroix, and M.-F. Sagot. Assessing the exceptionality of coloured motifs in networks. *EURASIP Journal on Bioinformatics and Systems Biology*, pages 1–9, 2009.
17. R. Sharan and T. Ideker. Modeling cellular machinery through biological network comparison. *Nature Biotechnology*, 24:427–433, 2006.
18. R. Williams. Finding paths of length k in $O^*(2^k)$ time. *Information Processing Letters*, 109(6):315–318, 2009.

6 Appendix

6.1 Proof of Proposition 2

We observe that the circuit given in Proposition 1 for CGM can be adapted to a more general problem, where vertices are labeled by several sets of colors, and the goal is to choose one set per vertex such that the sets chosen for the subtree are disjoint. This problem is defined below. Let C be a set of colors. Given a family $\{\mathcal{F}_1, ..., \mathcal{F}_n\}$ where each $\mathcal{F}_i \subseteq 2^C$, and given an integer r, we say that the family has an *r-matching* iff we can choose $F_1 \in \mathcal{F}_1, ..., F_n \in \mathcal{F}_n$ such that the F_i are pairwise disjoint and their union has size r. We consider the following generalization of Colorful Graph Motif:

Name: List-Colored Embedded Subtree (LCST)
Input: a graph $G = (V, E)$, integers r, s, a set Γ, for every $u \in V$ a family $\mathcal{F}_u$ of subsets of Γ
Solution: an embedded subtree (T, ϕ_V, ϕ_E) of G s.t. (i) $|V_T| \leq r$, (ii) the family of sets $\{\mathcal{F}_{\phi_V(v)} : v \in V_T\}$ has an s-matching.

Given an instance I of LCST, we let m denote the number of edges of G, and we let n denote $\sum_{v \in V} \sum_{S \in \mathcal{F}_v} |S|$.

Proposition 7. LCST *is solvable by a randomized algorithm in $\mathcal{O}(2^s r^2 (n+m))$ time and $\mathcal{O}(r^2(n+m))$ space.*

Proof. We modify the circuit of Proposition 1 as follows: (i) for each $u \in V$, we define $P_{1,u} = \sum_{S \in \mathcal{F}_u} \pi_S$, where $\pi_S := \prod_{c \in S} x_c$, (ii) we now define the root node by $P = \sum_{i=1}^{r} \sum_{u \in V} P_{s,u}$. We then decide if P has a multilinear monomial of degree s using Theorem 1, and the complexity follows by observing that the circuit has size $\mathcal{O}(r^2(n+m))$. The reduction is correct, since an induction on i shows: $C_{i,u}$ has a multilinear monomial of degree d iff there is an embedded subtree (T, ϕ) of G containing u, such that $|V_T| = i$, and such that the family of sets $\{\mathcal{F}_v : v \in V_T\}$ has a d-matching. □

Proof of Proposition 2. For Points 1 and 2, we reduce to the LCST problem.

Point 1. Consider an instance $I = (G, k, C, \chi, M)$ of MGM. We define a corresponding instance I' of LCST as follows. The graph G is the same. The set Γ contains (i) the elements of V, (ii) for every $c \in C$, of elements $c_1, ..., c_m$ with $m = n_M(c)$. Now, to each vertex $u \in V$ such that $\chi(u) = c$, we associate a family $\mathcal{F}_u$ which consists of the sets $\{u, c_1\}, ..., \{u, c_m\}$ with $m = n_M(c)$. We finally set $r = k$ and $s = 2k$, and we solve LCST on I' in $\mathcal{O}(4^k k^2 c_{max}|G|)$ time and $\mathcal{O}(k^2 c_{max}|G|)$ space by Proposition 7. The intuition is that adding u to each set in $\mathcal{F}_u$ ensures that a solution for I' will be vertex-distinct, and that adding the c_i's ensures that no more than $n_M(c)$ vertices can have the color c. Let us prove formally that I has a solution iff I' has.

Suppose that $T = (V_T, E_T)$ is a solution for I. We can view T as a vertex-distinct embedded subtree of G, and we claim that it is a solution for I'. Clearly, $|V_T| \leq k$. Besides, for each $c \in C$, we can number its occurrences in V_T by $c_1, ..., c_i$ with $i \leq n_M(c)$. Then to each $v \in V_T$ with $c = \chi(v)$, we can associate the set $S_v = \{v, c_j\}$ if c_j is the numbering of this occurrence of c. It follows that each S_v is in $\mathcal{F}_v$, and that the sets S_v ($v \in V_T$) are disjoint, hence $\{F_v : v \in V_T\}$ has a $2k$-matching. Conversely, suppose that we have $\mathcal{T} = (T, \phi_V, \phi_E)$ solution for I'. For each $v \in V_T$, let $u = \phi(v)$, and pick $S_v \in \mathcal{F}_u$, such that the resulting sets form an $2k$-matching. Then each S_v has the form $\{u, c_i\}$ with $c = \chi(u)$. Since these sets are disjoint, it follows that $\mathcal{T}$ has exactly k vertices and is vertex-distinct; also, by definition of Γ, no color $c \in C$ can occur more than $n_M(c)$ times.

Point 2. We modify the reduction of Point 1 by adding the empty set to each set $\mathcal{F}_u$. The intuition is that this will allow some vertices of the subtree to be ignored, allowing to only select a set S of k vertices such that $\chi(S) \subseteq M$. The formal proof goes as in Point 1. Note however that for the converse direction, we cannot ensure that ϕ is injective for the vertices v with $S_v = \emptyset$, and so we cannot guarantee that $\mathcal{T}$ is vertex-distinct. This is not a problem: given $\mathcal{T}$, we can construct $\mathcal{T}'$ vertex-distinct which has fewer vertices, and contains all the vertices with $S_v \neq \emptyset$, implying that $\mathcal{T}'$ is also a solution for I'. □

6.2 Proof of Lemma 1

We first reduce #Multicolored Clique to #MEST − 1. Our source problem #Multicolored Clique is the counting version of Multicolored Clique, which is easily seen to be #W[1]-hard. Let $I = (G, k)$ be an instance of the problem, where $G = (V, E)$ has a partition $\mathcal{P}$ into classes $V_1, ..., V_k$. Our target instance is $I' = (k, r, H, t)$ with $r = k^2 - k + 1$ and $t = (k, k-1, ..., k-1)$. The graph H is obtained by splitting every edge e in two parallel edges; then H is a k-partite multigraph with partition $\mathcal{P}$. Let $\mathcal{S}_I$, $\mathcal{S}_{I'}$ be the solution sets of I and I' respectively. Let $\mathcal{K}_k$ be the multigraph with k vertices $1, ..., k$, and with two parallel edges between distinct vertices; its partition is $\mathcal{P}_k$ consisting of the sets $\{1\}, ..., \{k\}$. Let $\mathcal{U}_k$ denote the set of $(\mathcal{P}_k, t)$-labeled edsts $(\mathcal{T}, \psi_V, \psi_E)$ for $\mathcal{K}_k$ such that $T(\psi_V) = t$. Observe that $\mathcal{U}_k \neq \emptyset$: since every vertex of $\mathcal{K}_k$ has degree $2(k-1)$, it follows that $\mathcal{K}_k$ has an Eulerian path starting at 1, which visits 1 k times and each other vertex $k-1$ times. We claim that $|\mathcal{S}_{I'}| = |\mathcal{U}_k||\mathcal{S}_I|$, which will prove the correctness of the reduction. To this aim, we will describe a bijection $\Phi : \mathcal{S}_{I'} \times \mathcal{U}_k \to \mathcal{S}_I$.

Consider a pair $P = (C, \mathcal{T}) \in \mathcal{S}_{I'} \times \mathcal{U}_k$ with $\mathcal{T} = (T, \psi_V, \psi_E)$ and $C = \{x_1, ..., x_k\}$ multicolored clique of G (with $x_i \in V_i$). Let $\phi = (\phi_V, \phi_E)$ be the homomorphism of $\mathcal{K}_k$ into H which maps i to v_i, and the parallel edges accordingly. We then define $\mathcal{T}' = \Phi(P)$ by $\mathcal{T}' = (T, \psi'_V, \psi'_E)$, where (i) ψ'_V is defined so that if $\psi_V(u) = (v, i)$ and if $\phi_V(v) = w$ then $\psi'_V(u) = (w, i)$, (ii) $\psi'_E = \psi_E \circ \phi_E$. We verify that $\mathcal{T}' \in \mathcal{S}_I$: indeed, it is a $(\mathcal{P}, t)$-labeled edst of G and $T(\psi'_V) = t$ (since we have composed with injective functions ϕ_V, ϕ_E). To prove that Φ is a bijection, we define the inverse function $\Psi : \mathcal{S}_I \to \mathcal{S}_{I'} \times \mathcal{U}_k$

as follows. Consider $\mathcal{T}' = (T, \psi'_V, \psi'_E)$ $(\mathcal{P}, t)$-labeled edst of G, with $T(\psi'_V) = t$. This equality yields vertices $v_1 \in V_1, ..., v_k \in V_k$ such that $|(\psi^r_V)^{-1}(v_i)| = t_i$. Let $C = \{v_1, ..., v_k\}$, then C is a multicolored clique of G: indeed, $H[C]$ has at most $k^2 - k$ edges, and since ψ'_E is injective it must have exactly $k^2 - k$ edges, implying that $G[C]$ is a complete graph. We can then define (ψ_V, ψ_E) from (ψ'_V, ψ'_E) by "projecting" v_i on i, and the parallel edges accordingly (for instance, if $\psi'_V(u) = (v_i, j)$ then $\psi_V(u) = (i, j)$). We finally define $P = \Psi(\mathcal{T}')$ by $P = (C, \mathcal{T})$ where $\mathcal{T} = (T, \psi_V, \psi_E)$. It is easy to see that $P \in \mathcal{S}_{I'} \times \mathcal{U}_k$, and that Φ and Ψ are inverse of each other.

We now give a Turing-reduction of $\#\text{MEST} - 1$ to $\#\text{MEST} - 2$. Consider an instance I of $\#\text{MEST} - 1$ consisting of integers k, r, of a k-partite graph $G = (V, E)$ whose partition $\mathcal{P}$ consists of classes $V_1, ..., V_k$, and of a tuple $\tau \in [r]^k$. Given a tuple $t \in [r]^k$, we define the instance $I_t = (k, r, G, t)$, and we let $\mathcal{S}_t, \mathcal{S}'_t$ be their solution sets for $\#\text{MEST} - 1, \#\text{MEST} - 2$ respectively. Let $N_t = |\mathcal{S}_t|$ and $N'_t = |\mathcal{S}'_t|$. Recall that our goal is to compute N_τ. We have for every $t \in [r]^k$: $N'_t = \sum_{t' \leq t} N_{t'}$, which yields by Moebius inversion that for every $t \in [r]^k$: $N_t = \sum_{t' \leq t} \mu(t, t') N'_{t'}$ [1]. Since we can compute each value N'_t $(t \leq \tau)$ by one oracle call for $\#\text{MEST} - 2$, it follows that we can compute the values N_t $(t \leq \tau)$ in $\mathcal{O}(2^k|G|)$ time and using $\mathcal{O}(2^k)$ oracle calls for $\#\text{MEST} - 2$, thereby solving $\#\text{MEST} - 1$. □

[1] where $\mu(t, t')$ is 0 if there exists $i \in [k]$ s.t. $t_i - t'_i > 1$, and is otherwise equal to $(-1)^r$ where r is the number of $i \in [k]$ s.t. $t_i - t'_i = 1$.