Parameterized Algorithms for Diverse Multistage Problems

Leon Kellerhals
Algorithmics and Computational Complexity, Faculty IV, Technische Universität Berlin, Germany

Malte Renken
Algorithmics and Computational Complexity, Faculty IV, Technische Universität Berlin, Germany

Philipp Zschoche
Algorithmics and Computational Complexity, Faculty IV, Technische Universität Berlin, Germany

Abstract

The world is rarely static – many problems need not only be solved once but repeatedly, under changing conditions. This setting is addressed by the *multistage* view on computational problems. We study the *diverse multistage* variant, where consecutive solutions of large variety are preferable to similar ones, e.g. for reasons of fairness or wear minimization. While some aspects of this model have been tackled before, we introduce a framework allowing us to prove that a number of diverse multistage problems are fixed-parameter tractable by diversity, namely PERFECT MATCHING, s-t PATH, MATROID INDEPENDENT SET, and PLURALITY VOTING. This is achieved by first solving special, colored variants of these problems, which might also be of independent interest.

2012 ACM Subject Classification Theory of computation → Parameterized complexity and exact algorithms; Mathematics of computing → Graph algorithms

Keywords and phrases Temporal graphs, dissimilar solutions, fixed-parameter tractability, perfect matchings, s-t paths, committee election, spanning forests, matroids

Digital Object Identifier 10.4230/LIPIcs.ESA.2021.55

Related Version *Full Version*: `https://arxiv.org/abs/2105.04856`

Funding *Malte Renken*: Supported by the German Research Foundation (DFG), project MATE (NI 369/17).

Acknowledgements The authors wish to thank Rolf Niedermeier and anonymous reviewers for their careful reading and suggestions of the manuscript. This work was initiated at the research retreat of the Algorithmics and Computational Complexity group of TU Berlin in September 2020 in Zinnowitz.

1 Introduction

In the *multistage* setting, given a sequence of instances of some problem, one asks whether there is a corresponding sequence of solutions such that consecutive solutions relate in some way to each other. Often the aim is to find consecutive solutions that are very *similar* [25, 18, 20, 7, 6, 19]. This is reasonable when changing between distinct solutions incurs some form of cost. In other settings, the opposite goal is more reasonable, that is, consecutive solutions should be very *different*. This is a natural goal when wear minimization, load distribution, or resilience against failures or attacks are of interest. This *diverse multistage* setting is what we want to focus on in this paper. Here, given a sequence of instances of some decision problem, the task is to find a sequence of solutions such that the *diversity*, i.e., the size of the symmetric difference of any two consecutive solutions is *at least* ℓ.

This problem has already received some attention in the literature: Fluschnik et al. [21] studied the problem of finding diverse s-t paths and Bredereck et al. [11] considered series of committee elections. In a similar setting, but aiming for large symmetric difference between every two (i.e., not just consecutive) solutions, Baste et al. [8] provide a framework for parameterization by treewidth, while the case that all instances are the same is studied for matchings, independent sets of matroids [22, 23], and for Kemeny rank aggregation [3].

We briefly give a formal definition. Assume Π to be some decision problem which asks whether the *family of solutions* $\mathcal{R}(I) \subseteq 2^{\mathcal{B}(I)}$ of an instance I of Π is non-empty, where $\mathcal{B}(I)$ is some *base set* encompassing all possible solutions. For example, for an instance I of Vertex Cover, the set $\mathcal{B}(I)$ is the set of all vertices and $\mathcal{R}(I)$ is the set of all vertex covers within the size bound. The problem Diverse Multistage Π is now the following.

Diverse Multistage Π

Input: A sequence $(I_i)_{i=1}^{\tau}$ of instances of Π and an integer $\ell \in \mathbb{N}_0$.

Question: Is there a sequence $(S_i)_{i=1}^{\tau}$ of solutions $S_i \in \mathcal{R}(I_i)$ such that $|S_i \Delta S_{i+1}| \geq \ell$ for all $i \in [\tau - 1]$?

Our contributions. We present a general framework which allows us to prove fixed-parameter tractability of Diverse Multistage Π parameterized by the diversity ℓ for several problems Π. This includes finding diverse matchings, but also diverse commitees (answering an open question by Bredereck et al. [11]), diverse s-t paths, and diverse independent sets in matroids such as spanning forests. Finally, we show that similar results cannot be expected for finding diverse vertex covers.

Generally, our framework can be applied to Diverse Multistage Π whenever one can solve a 4-colored variant of Π efficiently. Formally, this variant is defined as follows.

4-Colored Exact Π

Input: An instance I of Π, a coloring $c\colon \mathcal{B}(I) \to [4]$, and $n_i \in \mathbb{N}_0, i \in [4]$.

Output: A solution $S \in \mathcal{R}(I)$ such that $|\{x \in S \mid c(x) = i\}| = n_i$ for all $i \in [4]$ or "*no*" if no such solution exists.

Our main result reads as follows.

▶ **Theorem 1.** *For any parameter r of Π, if an instance I of* 4-Colored Exact Π *can be solved in $f(r) \cdot |I|^{O(1)}$ time, then an instance J of* Diverse Multistage Π *can be solved in $2^{O(\ell)} \cdot f(r_{\max}) \cdot |J|^{O(1)}$ time, where $r_{\max}$ is the maximum of parameter r over all instances of Π in J.*[1]

We prove Theorem 1 in Section 3 in a more general form which also allows solving 4-Colored Exact Π by a Monte Carlo algorithm. We then apply our framework to the following problems:

Committee Election (Section 4). In Diverse Multistage Plurality Voting, we are given a set A of agents, a set C of candidates, and τ many *voting profiles* $u_i\colon A \to C$. The goal is to find a sequence $(C_i)_{i=1}^{\tau}$ of committees $C_i \subseteq C$ such that each committee C_i is of size at most k and gets at least x votes in the voting profile u_i (i.e., $|u_i^{-1}(C_i)| \geq x$), and $|C_i \Delta C_{i+1}| \geq \ell$ for all $i \in [\tau - 1]$. We show that there is a $2^{\mathcal{O}(\ell)} \cdot |J|^{\mathcal{O}(1)}$-time algorithm to solve a Diverse Multistage Plurality Voting instance J. This answers an open question of Bredereck et al. [11]. In the full version we generalize the algorithm used to solve 4-Colored Exact Plurality Voting to matroids.

[1] For example, if the input is a sequence of graphs and r is the treewidth, then $r_{\max}$ is the maximum treewidth over all graphs in the input.

Perfect Matching (Section 5). In the multistage setting, Perfect Matching is among the problems most intensively studied [25, 5, 4, 13, 34]. Given a sequence of graphs $(G_i)_{i=1}^{\tau}$ and an integer ℓ, Diverse Multistage Perfect Matching asks whether there is a sequence $(M_i)_{i=1}^{\tau}$ such that each M_i is a perfect matching in G_i, and $|M_i \Delta M_{i+1}| \geq \ell$ for all $i \in [\tau - 1]$. We show that there is a randomized $2^{\mathcal{O}(\ell)} \cdot |J|^{\mathcal{O}(1)}$-time algorithm to solve a Diverse Multistage Perfect Matching instance J with constant error probability. This stands in remarkable contrast to the W[1]-hardness of the (non-diverse) Multistage Perfect Matching, when parameterized by $\ell + \tau$ [34]. To apply our framework, we establish an algebraic algorithm using the Pfaffian of a specific variant of the Tutte matrix to solve s-Colored Exact Perfect Matching on an n-vertex graph in $n^{O(s)}$ time with low error probability.

s-t Path (Section 6). Studying s-t Path in the multistage setting was already suggested in the seminal work of Gupta et al. [25]. In Diverse Multistage s-t Path one is given a sequence of graphs $(G_i)_{i=1}^{\tau}$, two distinct vertices s and t, and an integer ℓ, and asks whether there is a sequence $(P_i)_{i=1}^{\tau}$ such that each P_i is an s-t Path in G_i, and $|V(P_i) \Delta V(P_{i+1})| \geq \ell$ for all $i \in [\tau - 1]$. Fluschnik et al. [21] provided a comprehensive study of finding s-t paths of bounded length in the multistage setting from the viewpoint of parameterized complexity. Among other results, they showed that Diverse Multistage s-t Path is NP-hard but fixed-parameter tractable when parameterized by the maximum length of an s-t Path in the solution. We show that Diverse Multistage s-t Path parameterized by ℓ is fixed-parameter tractable. At first glance, using our framework seems unpromising since 4-Colored Exact s-t Path can presumably not be solved in polynomial time (it is NP-hard by a straight-forward reduction from Hamiltonian Path). However, we develop a win/win strategy around a generalization of the Erdős-Pósa theorem for long cycles due to Mousset et al. [30] so that we have to solve 4-Colored Exact s-t Path only on graphs on which the treewidth is upper-bounded in the parameter ℓ.

In Section 7, we complement our fixed-parameter tractability results with a W[1]-hardness for Diverse Multistage Vertex Cover when parameterized by ℓ.

2 Preliminaries

We denote by $\mathbb{N}$ and $\mathbb{N}_0$ the natural numbers excluding and including zero, respectively. For $n \in \mathbb{N}$, let $[n] := \{1, 2, \dots, n\}$. For two sets A and B, we denote by $A \Delta B := (A \setminus B) \cup (B \setminus A) = (A \cup B) \setminus (A \cap B)$ the *symmetric difference* of A and B, and by $A \uplus B$ the disjoint union of A and B. For a function $c\colon A \to B$, let $c(A') := \bigcup_{a \in A'} c(a)$ and $c^{-1}(b) := \{a \in A \mid c(a) = b\}$, where $A' \subseteq A$. We also use the notations c^b and $c^{b,b'}$ as shorthands for $c^{-1}(b)$ and $c^{-1}(b) \cup c^{-1}(b')$, respectively.

A *Monte Carlo algorithm*, or an algorithm with error probability p, is a randomized algorithm that returns a correct answer with probability $1 - p$.

Let Σ be a finite alphabet. A *parameterized problem* L is a subset $L \subseteq \{(x, k) \in \Sigma^* \times \mathbb{N}_0\}$. An instance $(x, k) \in \Sigma^* \times \mathbb{N}_0$ is a *yes*-instance of L if and only if $(x, k) \in L$ (otherwise, it is a *no*-instance). A parameterized problem L is *fixed-parameter tractable* (in FPT) if for every input (x, k) one can decide in $f(k) \cdot |x|^{O(1)}$ time whether $(x, k) \in L$, where f is some computable function only depending on k. If W[1]-hard parameterized problem is not fixed-parameter tractable unless FPT=W[1]. We refer to Downey and Fellows [17] and Cygan et al. [14] for more material on parameterized complexity. We use standard notation from graph theory [16]. Throughout this paper, we assume graphs to be simple and undirected.

3 The General Framework

In this section, we introduce a general framework to show (for some decision problem Π) fixed-parameter tractability of DIVERSE MULTISTAGE Π parameterized by ℓ. Recall that, for every instance I of a decision problem Π, we denote by $\mathcal{B}(I)$ the base set encompassing all possible solutions, by $\mathcal{R}(I) \subseteq 2^{\mathcal{B}(I)}$ the family of solutions, and by $|I|$ the input size of I, which is at least $|\mathcal{B}(I)|$. For the reminder of this section we assume that $|\mathcal{B}(I)| \geq 2$ for all instances I of Π. The framework is applicable to DIVERSE MULTISTAGE Π if there is an efficient algorithm for 4-COLORED EXACT Π. Formally, we use the following prerequisite, which is slightly more general than in Theorem 1.

▶ **Assumption 2.** There are computable functions f, g such that for every $0 \leq p \leq 1$ for which $g(p)$ is defined, there is a Monte-Carlo algorithm $\mathcal{A}$ with error probability p and running time $f(r) \cdot |I|^{O(1)} \cdot g(p)$ that solves an instance I of 4-COLORED EXACT Π, where $r \in \mathbb{N}_0$ is some parameter of I and g is monotone non-increasing.

We allow an error probability in Assumption 2 because for one of our applications (in Section 5), no other polynomial-time algorithm is known. The goal is to prove the following.

▶ **Theorem 3.** *Let Assumption 2 be true. Then any size-n instance I of* DIVERSE MULTISTAGE Π *can be solved in $2^{O(\ell)} \cdot f(r_{\max}) \cdot n^{O(1)} \cdot g(p/_{7 2^{O(\ell)} n^{O(1)}})$ time by a Monte-Carlo algorithm with error probability p, where $r_{\max}$ is the maximum of parameter r over all instances of Π in I, and $0 \leq p \leq 1$ is an arbitrary probability for which the above expression is defined.*[2]

Note that, if we have a non-randomized algorithm in Assumption 2 (that is, $g(0)$ is defined and g maps always to one), then Theorem 1 follows directly from Theorem 3.

The underlying strategy of the algorithm for a DIVERSE MULTISTAGE Π-instance J behind Theorem 3 is to compute for each instance I of Π in J a solution family such that the Cartesian product of these families contains a solution for J if and only if J is a *yes*-instance. Once these families are obtained, we can check whether J is a *yes*-instance by dynamic programming. To this end, we compute a small subset of $\mathcal{R}(I)$ satisfying the following definition.

▶ **Definition 4.** *Let $\mathcal{F}$ be a set family. A subfamily $\widehat{\mathcal{F}} \subseteq \mathcal{F}$ is called an ℓ-*diverse representative *of $\mathcal{F}$ if the following holds: for any pair of sets A, B, if there is an $S \in \mathcal{F}$ with $\min\{|A\Delta S|, |B\Delta S|\} \geq \ell$, then there is an $\widehat{S} \in \widehat{\mathcal{F}}$ such that $\min\{|A\Delta\widehat{S}|, |B\Delta\widehat{S}|\} \geq \ell$.*

First of all, we note that ℓ-diverse representatives can be rather small.

▶ **Lemma 5.** *Let $\mathcal{F}$ be a set family and $S_1, S_2, S_3 \in \mathcal{F}$. If $|S_i\Delta S_j| \geq 2\ell$ for all distinct $i, j \in [3]$, then $\{S_1, S_2, S_3\}$ is an ℓ-diverse representative of $\mathcal{F}$.*

Proof. Assume for contradiction that there exist sets A and B with $\min\{|A\Delta S_i|, |B\Delta S_i|\} < \ell$ for all i. Without loss of generality, assume that $|A\Delta S_1| < \ell$. Then for $j \in \{2, 3\}$ we have $|A\Delta S_j| \geq |S_1\Delta S_j| - |S_1\Delta A| > 2\ell - \ell = \ell$ by the triangle inequality. Therefore, $|B\Delta S_j| < \ell$ for all $j \in \{2, 3\}$. In particular, $|B\Delta S_2| < \ell$. Again, by the triangle inequality $|B\Delta S_3| \geq |S_2\Delta S_3| - |S_2\Delta B| > 2\ell - \ell = \ell$, i.e., $\min\{|A\Delta S_3|, |B\Delta S_3|\} \geq \ell$ – a contradiction. ◀

[2] For example, if we only have an algorithm with non-zero error probability, then $p = 0$ is excluded.

In the following, we measure the distance of two solutions by the size of the symmetric difference. In a nutshell, we compute an ℓ-diverse representative of the family of solutions by first trying to compute three solutions which are far apart from each other (that is, size of symmetric difference at least 2ℓ). If this succeeds, then by Lemma 5 we are done. Otherwise, we distinguish between three cases.

No solution. If there is no solution at all, then trivially $\emptyset$ is an ℓ-diverse representative of the family of solutions.

One solution. If we only find one solution S_1 to the instance of Π, then each other solution is close to S_1. Hence, for any two sets A, B, if one of them is far away from S_1, then by the triangle inequality it is also far away from every other solution and can be safely ignored. For those sets which are close to S_1, we can exploit the upper bound on the symmetric difference by using color-coding [2] and then applying Assumption 2 to compute an ℓ-diverse representative of the family of solutions. This case is handled in Lemma 9.

Two solutions. If we find two diverse solutions S_1 and S_2 such that no other solution is far away from both, then S_1 and S_2 partition the solution space into two parts: the solutions close to S_1 and those close to S_2. Again, given two sets A, B, if either of them is far away from S_1 and S_2, then we may ignore it. By including S_1 and S_2 in our family, we may further assume that A is similar to S_1 and B is similar to S_2. We distinguish two subcases. If the distance between S_1 and S_2 is very large, then A is far away from all solutions in the second part and B is far away from all solutions in the first part. We can thus ignore one of them (say B) and exploit the fact that A, S_1, and all solutions of interest are close to each other to use color-coding and then apply Assumption 2. In the other subcase where the distance between S_1 and S_2 is bounded, we can utilize that fact similarly. This case is handled in Lemma 10.

Hereafter, the details. Before we dive into the case distinction outlined above, we need to prove two technical lemmata, telling us how to build a diverse representative set that works for all sets obeying some given coloring of the elements of $\mathcal{B}(I)$. These will later work as building blocks in the construction of proper diverse representatives. In the first lemma, only two colors are used, and we are only concerned with one arbitrary set A instead of two.

▶ **Lemma 6.** *Let Assumption 2 be true. Given an instance I of Π of size n, a coloring $c\colon \mathcal{B}(I) \to [2]$, and a solution $M \in \mathcal{R}(I)$, one can compute in $f(r)n^{O(1)}g(pn^{-4})$ time and with error probability at most p a family $\mathcal{F} \subseteq \mathcal{R}(I)$ of size at most n^4 such that for any $S \in \mathcal{R}(I)$ and any $A \subseteq \mathcal{B}(I)$ with $S \setminus A \subseteq c^1$ and $A \setminus S \subseteq c^2$, there is $\widehat{S} \in \mathcal{F}$ with $|A \Delta \widehat{S}| \geq |A \Delta S|$ and $|M \Delta \widehat{S}| = |M \Delta S|$.*

Proof. Let $F'_1 := c^1 \cap M$, $F'_2 := c^2 \cap M$, $F'_3 := c^1 \setminus M$, and $F'_4 := c^2 \setminus M$.

Start with $\mathcal{F} = \emptyset$. Then, for each $m \in [n]$ and each number partition $\sum_{i=1}^4 m_i = m$ with $m_i \geq 0$, use algorithm $\mathcal{A}$ to search in $f(r)n^{O(1)}g(pn^{-4})$ time and with error probability at most pn^{-4} for a set $N \in \mathcal{R}(I)$ such that $|N \cap F'_i| = m_i$ for all $i \in [4]$. If this succeeds, then we add N to $\mathcal{F}$. Since there are $\binom{n+4}{4} \leq n^4$ possibilities for $m_1, \dots, m_4$, the probability of an error occurring is upper-bounded by p. Moreover, the size of $\mathcal{F}$ is upper-bounded by n^4 and hence the time required is bounded by $f(r)n^{O(1)}g(pn^{-4})$.

It remains to be proven that $\mathcal{F}$ has the desired properties. Let $S \in \mathcal{R}(I)$ be arbitrary and set $m_i := |S \cap F'_i|$ for all $i \in [4]$. By construction, $\mathcal{F}$ contains a set $\widehat{S} \in \mathcal{R}(I)$ such that $|\widehat{S} \cap F'_i| = m_i$. We then have $|\widehat{S} \Delta M| = m_3 + m_4 + |M| - m_1 - m_2 = |S \Delta M|$.

Let $A \subseteq \mathcal{B}(I)$ be a set with $S \setminus A \subseteq c^1$ and $A \setminus S \subseteq c^2$. Since $A \setminus S \subseteq c^2$ we have

$$|A \cap S \cap c^1| = |A \cap c^1| \geq |A \cap \widehat{S} \cap c^1| \tag{1}$$

and since $S \setminus A \subseteq c^1$, we have that

$$|A \cap S \cap c^2| = |S \cap c^2| = m_2 + m_4 = |\widehat{S} \cap c^2| \geq |A \cap \widehat{S} \cap c^2|. \tag{2}$$

By adding (1) and (2) we obtain $|A \cap S| \geq |A \cap \widehat{S}|$ which in turn implies $|A \Delta S| \leq |A \Delta \widehat{S}|$ since $|S| = |\widehat{S}|$. ◀

The next lemma extends the approach of Lemma 6 to the case where we have four colors and two arbitrary sets A, B.

▶ **Lemma 7.** *Let Assumption 2 be true. Given an instance I of Π of size n and a coloring $c\colon \mathcal{B}(I) \to [4]$, one can compute in $f(r)n^{O(1)}g(pn^{-4})$ time and with error probability at most p a family $\mathcal{F} \subseteq \mathcal{R}(I)$ of size at most n^4 such that for any $S \in \mathcal{R}(I)$ and all sets $A, B \subseteq \mathcal{B}(I)$ with $A \setminus (B \cup S) \subseteq c^1, B \setminus (A \cup S) \subseteq c^2, (A \cap B) \setminus S \subseteq c^3,$ and $S \setminus (A \cap B) \subseteq c^4$, there is $\widehat{S} \in \mathcal{F}$ with $|C \Delta \widehat{S}| \geq |C \Delta S|$ for all $C \in \{A, B\}$.*

Proof. Begin with $\mathcal{F} = \emptyset$. Then, for each $m \in [n]$ and each number partition $\sum_{i=1}^4 m_i = m$ with $m_i \geq 0$, use algorithm $\mathcal{A}$ to search in $f(r)n^{O(1)}g(pn^{-4})$ time and with error probability at most pn^{-4} for an $M \in \mathcal{R}(I)$ such that $|M \cap c^i| = m_i$ for all $i \in [4]$. If this succeeds, then add M to $\mathcal{F}$. Since there are $\binom{n+4}{4} \leq n^4$ possibilities for $m_1, \ldots, m_4$, the probability of an error occurring is upper-bounded by p. Moreover, the size of $\mathcal{F}$ is at most n^4 and thus the overall running time is $f(r)n^{O(1)}g(pn^{-4})$.

Now let $S \in \mathcal{R}(I)$ be arbitrary. Set $m_i := |S \cap c^i|$, for all $i \in [4]$. By construction there is $\widehat{S} \in \mathcal{F}$ such that $|\widehat{S} \cap c^i| = m_i$ for all $i \in [4]$. It remains to be proven that $\widehat{S}$ has the desired properties. To this end, let $A, B \subseteq \mathcal{B}(I)$ be two sets as stated in the lemma. By symmetry, it suffices to show that $|A \Delta \widehat{S}| \geq |A \Delta S|$.

Since $S \setminus A \subseteq c^4$ we have

$$|S \cap A \cap c^{1,3}| = |S \cap c^{1,3}| = m_1 + m_3 = |\widehat{S} \cap c^{1,3}| \geq |\widehat{S} \cap A \cap c^{1,3}| \tag{3}$$

and since $A \setminus S \subseteq c^{1,3}$, we have

$$|S \cap A \cap c^{2,4}| = |A \cap c^{2,4}| \geq |\widehat{S} \cap A \cap c^{2,4}|. \tag{4}$$

By adding (3) and (4), we obtain $|S \cap A| \geq |\widehat{S} \cap A|$ and thus $|S \Delta A| \leq |\widehat{S} \Delta A|$ since $|S| = |\widehat{S}|$. ◀

We now describe how we generate the colorings required for using Lemmata 6 and 7. Color-coding [2] is well-established in the toolbox of parameterized algorithms. While color-coding was initially described as a randomized technique, we use universal sets [33] to derandomize this technique as shown in the next lemma. Interestingly, without this derandomization the error probability of the color-coding step would later propagate through the dynamic programming method and consequently also depend on the number of instances of Π in the input instance of Diverse Multistage Π. The derandomization works as follows.

▶ **Lemma 8.** *For any set A of size n and any $b \leq n$ one can compute in $2^{2b+o(b)} \log n \cdot n$ time a family of functions $\{c_j \colon A \to [4] \mid j \in [2^{2b+o(b)} \log n]\}$ such that for any $\biguplus_{i=1}^4 B_i \subseteq A$ with $|\biguplus_{i=1}^4 B_i| \leq b$ there is a j such that $c_j(B_i) = \{i\}$, for all $i \in [4]$.*

Proof. Let $A := \{a_1, \ldots, a_n\}$. By a result of Naor et al.citeNearOptimalDerandomization, one can compute in $2^{2b} b^{\mathcal{O}(\log b)} \cdot \log n \cdot n \subseteq 2^{2b+o(b)} \log n \cdot n$ time a so-called $(2n, 2b)$-*universal set* which is a family $\mathcal{U} \subseteq 2^{[2n]}$ such that for every $B' \subseteq A$ with $|B'| = 2b$ the family $\{B' \cap U \mid U \in \mathcal{U}\}$ contains all 2^{2b} subsets of B'. Let $\mathcal{U} := \{U_i\}_{i=1}^{2^{2b+o(b)} \log n}$. We then define

c_j, $j \in [2^{2b+o(b)} \log n]$, by

$$c_j(a_i) := \begin{cases} 1, & \text{if } i, i+n \in U_j, \\ 2, & \text{if } i \in U_j \text{ and } i+n \notin U_j, \\ 3, & \text{if } i \notin U_j \text{ and } i+n \in U_j, \text{ and} \\ 4, & \text{if } i, i+n \notin U_j. \end{cases}$$

Now let $B_1 \uplus B_2 \uplus B_3 \uplus B_4 \subseteq A$ be an arbitrary 4-partition of a subset of A of size at most b. Consider $B' := \{i, i+n \mid a_i \in \bigcup_{q=1}^{4} B_q\}$. We assume that B' is of size $2b$, otherwise we add arbitrary elements from $[2n]$. Since $B'' := \{i, i+n \mid a_i \in B_1\} \cup \{i \mid a_i \in B_2\} \cup \{i+n \mid a_i \in B_3\} \subseteq B'$ there is an $U_j \in \mathcal{U}$ such that $B' \cap U_j = B''$. Hence, $c_j(B_i) = \{i\}$, for all $i \in [4]$. ◀

We now show how to generate an ℓ-diverse representative of the family of solutions if there is one solution M^* from which no other solution differs by more than 2ℓ.

▶ **Lemma 9.** *Let Assumption 2 be true. Given an instance I of Π of size n, and a solution $M^* \in \mathcal{R}(I)$ such that each $M \in \mathcal{R}(I)$ satisfies $|M \Delta M^*| \leq 2\ell$, one can compute in $2^{16\ell+o(\ell)} \log n \cdot f(r) \cdot n^{O(1)} \cdot g(p/2^{16\ell+o(\ell)} \log n \cdot n^4)$ time and with error probability p an ℓ-diverse representative of $\mathcal{R}(I)$ of size at most $2^{16\ell+o(\ell)} \log n \cdot n^4$.*

Proof. For simplicity, let $\mathbb{J} := [2^{16\ell+o(\ell)} \log n]$. Apply Lemma 8 with $b = 8\ell$ to compute in $2^{16\ell+o(\ell)} \log n \cdot n$ time a family of colorings $\{c_j : \mathcal{B}(I) \to [4] \mid j \in \mathbb{J}\}$. To apply Lemma 8 we assume that $|\mathcal{B}(I)| \geq 8\ell$, otherwise we utilize dummy elements. For each $j \in \mathbb{J}$, apply Lemma 7 to I and c_j to compute a family $\mathcal{F}_j \subseteq \mathcal{R}(I)$ with error probability $p \cdot |\mathbb{J}|^{-1}$. Observe that the probability of an error occurring at any of the $|\mathbb{J}|$ steps is bounded by p. Choose $\mathcal{F} := \{M^*\} \cup \bigcup_{j \in \mathbb{J}} \mathcal{F}_j$. According to Lemma 7 the size of $\mathcal{F}$ is upper-bounded by $|\mathbb{J}| \cdot n^4$ and the time required is bounded by $|\mathbb{J}| \cdot f(r) \cdot n^{O(1)} \cdot g(pn^{-4} \cdot |\mathbb{J}|^{-1})$.

We now show that $\mathcal{F}$ is an ℓ-diverse representative of $\mathcal{R}(I)$. To this end, let $S \in \mathcal{R}(I)$ and let A, B be two arbitrary sets such that $|A \Delta S| \geq \ell$ and $|B \Delta S| \geq \ell$. Since $M^* \in \mathcal{F}$, we may assume by symmetry that, say, $|M^* \Delta A| < \ell$, otherwise we are done. Note that $|M^* \Delta S| \leq 2\ell$ and that $|A \Delta S| \leq |A \Delta M^*| + |M^* \Delta S| < 3\ell$. We say that some coloring c is *good* for A, B, S if the conditions of Lemma 7 are satisfied, i.e. if

$$A \setminus (B \cup S) \subseteq c^1, \quad B \setminus (A \cup S) \subseteq c^2, \quad (A \cap B) \setminus S \subseteq c^3, \text{ and} \quad S \setminus (A \cap B) \subseteq c^4.$$

We distinguish between two cases.

Case 1: $|M^* \Delta B| < 3\ell$. Then $|B \Delta S| \leq |B \Delta M^*| + |M^* \Delta S| \leq 5\ell$. According to Lemma 8 there is an $i \in \mathbb{J}$ such that coloring c_i is good for A, B, S, since $|B \Delta S| + |A \Delta S| < 8\ell$. By Lemma 7 and construction of $\mathcal{F}_i$, there is an $\widehat{S} \in \mathcal{F}_i \subseteq \mathcal{F}$ such that $|\widehat{S} \Delta A| \geq |S \Delta A| \geq \ell$ and $|\widehat{S} \Delta B| \geq |S \Delta B| \geq \ell$.

Case 2: $|M^* \Delta B| \geq 3\ell$. Set $B' := A$. According to Lemma 8 there is an $i \in \mathbb{J}$ such that coloring c_i is good for A, B', S, since $|B' \Delta S| + |A \Delta S| < 6\ell$. Thus, by Lemma 7 and by the construction of $\mathcal{F}_i$ there is an $\widehat{S} \in \mathcal{F}_i \subseteq \mathcal{F}$ such that $|\widehat{S} \Delta A| \geq |S \Delta A| \geq \ell$. Finally, we observe that $|\widehat{S} \Delta B| \geq |M^* \Delta B| - |M^* \Delta \widehat{S}| \geq 3\ell - 2\ell \geq \ell$ by the triangle inequality.

This completes the proof. ◀

Next, we show how to generate an ℓ-diverse representative of the family of solutions if there are two solutions such that no other solution differs from both by more than 2ℓ.

▶ **Lemma 10.** *Let Assumption 2 be true. Let I be an Π-instance of size n, and $M_1, M_2 \in \mathcal{R}(I)$ such that $|M_1 \Delta M_2| \geq 2\ell$ and each $M \in \mathcal{R}(I)$ has $\min\{|M \Delta M_1|, |M \Delta M_2|\} \leq 2\ell$. Then one can compute, in $2^{20\ell+o(\ell)} \log n \cdot f(r) n^{\mathcal{O}(1)} g(p/n^4 2^{20\ell+o(\ell)} \log n)$ time and with error probability p, an ℓ-diverse representative of $\mathcal{R}(I)$ of size $2^{20\ell+o(\ell)} \log n \cdot n^4$.*

Proof. For simplicity, let $\mathbb{J} := [2^{20\ell+o(\ell)} \log n]$. Apply Lemma 8 with $b = 10\ell$ to compute in $2^{20\ell+o(\ell)} \log n \cdot n$ time a family of colorings $\{c_j : \mathcal{B}(I) \to [4] \mid j \in \mathbb{J}\}$. To apply Lemma 8 we assume that $|\mathcal{B}(I)| \geq 10\ell$, otherwise we utilize dummy elements.

For each $j \in \mathbb{J}$, apply Lemma 7 to I and c_j to compute a family $\mathcal{F}_j \subseteq \mathcal{R}(I)$ of size at most n^4 with error probability $p/3 \cdot |\mathbb{J}|^{-1}$. Observe that the probability of an error occurring at any of the $|\mathbb{J}|$ steps is upper-bounded by $p/3$ and the computation of all $\mathcal{F}_j$ takes $|\mathbb{J}| f(r) n^{\mathcal{O}(1)} g(p/3n^4 \cdot |\mathbb{J}|)$ time.

Next, define another family of colorings $\{c'_j : \mathcal{B}(I) \to [2] \mid j \in \mathbb{J}\}$ by setting $c'_j(x) := \lceil c_j(x)/2 \rceil$. Then, for each $j \in \mathbb{J}$, apply Lemma 6, to I, c'_j and M_1 to compute a family $\mathcal{F}'_j \subseteq \mathcal{R}(I)$, with the same error probability and time bound as before. Repeat with M_2 instead of M_1 to obtain $\mathcal{F}''_j$.

Set $\mathcal{F} := \{M_1, M_2\} \cup \bigcup_{j\in\mathbb{J}} (\mathcal{F}_j \cup F'_j \cup F''_j)$. Then $\mathcal{F}$ has size at most $3|\mathbb{J}|n^4 + 2 \subseteq 2^{20\ell+o(\ell)} \log n \cdot n^4$. Computing $\mathcal{F}$ takes $2^{20\ell+o(\ell)} \log n \cdot f(r) n^{\mathcal{O}(1)} g(p/3n^4 \cdot |\mathbb{J}|)$ time. The probability of an error occurring at any step while computing $\mathcal{F}$ is upper-bounded by p.

We now show that $\mathcal{F}$ is an ℓ-diverse representative of $\mathcal{R}(I)$. To this end, let $S \in \mathcal{R}(I)$ and A, B be two arbitrary sets such that $|A \Delta S| \geq \ell$ and $|B \Delta S| \geq \ell$. We may assume for each $i \in [2]$ that $|M_i \Delta A| < \ell$ or $|M_i \Delta B| < \ell$, otherwise we are done. By symmetry, we may assume $|M_1 \Delta A| < \ell$. Then $|M_2 \Delta A| \geq |M_2 \Delta M_1| - |M_1 \Delta A| > \ell$ by the triangle inequality and thus we must have $|M_2 \Delta B| < \ell$. By assumption, $\min\{|S \Delta M_1|, |S \Delta M_2|\} \leq 2\ell$, so let without loss of generality $|S \Delta M_1| \leq 2\ell$. Note that $|A \Delta S| \leq |A \Delta M_1| + |M_1 \Delta S| < 3\ell$. We distinguish the following two cases.

Case 1: $|M_1 \Delta M_2| \leq 4\ell$. Then, $|B \Delta S| \leq |B \Delta M_2| + |M_2 \Delta M_1| + |M_1 \Delta S| < 7\ell$. We say that some coloring c is *good* for A, B, S if the conditions of Lemma 7 are satisfied, i.e. if $A \setminus (B \cup S) \subseteq c^1$, $B \setminus (A \cup S) \subseteq c^2$, $(A \cap B) \setminus S \subseteq c^3$, and $S \setminus (A \cap B) \subseteq c^4$. According to Lemma 8 there is an $i \in \mathbb{J}$ such that coloring c_i is good for A, B, S, since $|B \Delta S| + |A \Delta S| \leq 10\ell$. By Lemma 7, there is $\widehat{S} \in \mathcal{F}_i \subseteq \mathcal{F}$ such that such that $|\widehat{S} \Delta A| \geq |S \Delta A| \geq \ell$ and $|\widehat{S} \Delta B| \geq |S \Delta B| \geq \ell$.

Case 2: $|M_1 \Delta M_2| > 4\ell$. Since $|S \Delta A| < 3\ell \leq 10\ell$, there is $j \in \mathbb{J}$ such that $S \setminus A \subseteq {c'_j}^1$ and $A \setminus S \subseteq {c'_j}^2$. By Lemma 6 there is $\widehat{S} \in \mathcal{F}'_j$ such that $|\widehat{S} \Delta M_1| = |S \Delta M_1| \leq 2\ell$ and $|\widehat{S} \Delta A| \geq |S \Delta A| \geq \ell$. Finally, observe that by the triangle inequality $|\widehat{S} \Delta B| \geq |M_1 \Delta M_2| - |M_1 \Delta \widehat{S}| - |B \Delta M_2| > \ell$.

This completes the proof. ◀

With Lemmata 5, 9, and 10 at hand we can formalize the case distinction outlined in the beginning of the section. This gives us a way to efficiently compute an ℓ-diverse representative in general.

▶ **Lemma 11.** *Let Assumption 2 be true. Let I be an instance of Π of size n. One can compute an ℓ-diverse representative of $\mathcal{R}(I)$ of size $2^{20\ell+o(\ell)} \log n \cdot n^4$ in $2^{20\ell+o(\ell)} \log n \cdot f(r) n^{O(1)} g(p/n^4 \cdot 2^{20\ell+o(\ell)} \log n)$ time with error probability at most p.*

Proof. Our procedure to compute an ℓ-diverse representative of $\mathcal{R}(I)$ works in four steps.

Step 1. We use $\mathcal{A}$ with a monochrome coloring and error probability $p/4n$ to search for some $M_1 \in \mathcal{R}(I)$ in $f(r) n^{\mathcal{O}(1)} g(p/4n)$ time by guessing the size of $|M_1| \leq n$. Observe that the probability of an error occurring in any of the searches is upper-bounded by $p/4$. If we do not succeed, then output the empty set and we are done. Otherwise, we proceed with the next step.

Step 2. For each pair m_1, m_2 with $m_1 + m_2 \leq n$ and $m_2 + |M_1| - m_1 > 2\ell$, try to compute $M_2 \in \mathcal{R}(I)$ with $|M_2 \cap M_1| = m_1$ and $|M_2 \cap (\mathcal{B}(I) \setminus M_1)| = m_2$ in $f(r)n^{\mathcal{O}(1)}g(p/4n^2)$ time and with error probability $p/4n^2$ using $\mathcal{A}$ with a 2-coloring where elements in M_1 are assigned one color and elements in $\mathcal{B}(I) \setminus M_1$ are assigned the second color. If no such M_2 is found for any pair m_1, m_2, then for every $M \in \mathcal{R}(I)$ the symmetric difference $|M \Delta M_1| \leq 2\ell$. In that case we may apply Lemma 9 with error probability $p/2$ and are done. Observe that the probability of an error occurring at any step until here is upper-bounded by p and the overall running time is $2^{16\ell+o(\ell)} \log n \cdot f(r)n^{O(1)}g(p/n^4 \cdot 2^{16\ell+o(\ell)} \log n)$. If we found such an M_2, then we proceed with the next step.

Step 3. We have $M_1, M_2 \in \mathcal{R}(I)$ with $|M_1 \Delta M_2| \geq 2\ell$. Define the coloring $c\colon \mathcal{B}(I) \to [4]$ by

$$c(v) := \begin{cases} i & \text{if } v \in M_i \setminus M_j \text{ for } \{i,j\} = \{1,2\}, \\ 3 & \text{if } v \in M_1 \cap M_2, \text{ and} \\ 4 & \text{otherwise.} \end{cases}$$

For all m'_1, m'_2, m'_3, m'_4 with $m'_1 + m'_2 + m'_3 + m'_4 \leq n$ and $m'_2 + m'_4 + |M_1| - m'_1 - m'_3 > 2\ell$ and $m'_1 + m'_4 + |M_2| - m'_2 - m'_3 > 2\ell$, search for a solution $M_3 \in \mathcal{R}(I)$ with $|M_3 \cap c^i| = m'_i$, for all $i \in [4]$, using $\mathcal{A}$ with c and error probability $p/4n^4$. For all these combined, we thus have error probability $p/4$ and need $f(r)n^{\mathcal{O}(1)}g(p/4n^4)$ time. If no such M_3 is found for any choice of m'_1, m'_2, m'_3, m'_4, then any $M \in \mathcal{R}(I)$ must have $\min\{|M \Delta M_1|, |M \Delta M_2|\} \leq 2\ell$. In that case we may apply Lemma 10 with error probability $p/4$ and are done. Observe that the probability of an error occurring at any step until here is upper-bounded by p and the overall running time is $2^{20\ell+o(\ell)} \log n \cdot f(r)n^{O(1)}g(p/n^4 \cdot 2^{20\ell+o(\ell)} \log n)$. In case that we found such an M_3, we proceed with the next step.

Step 4. We have $M_1, M_2, M_3 \in \mathcal{R}(I)$ such that $|M_i \Delta M_j| \geq 2\ell$ for all distinct $i, j \in [3]$. Hence, by Lemma 5, we can output $\{M_1, M_2, M_3\}$. This completes the proof. ◀

Finally, Lemma 11 allows us to formulate a dynamic program for DIVERSE MULTISTAGE Π and prove Theorem 3.

Proof of Theorem 3. Let $J := ((I_i)_{i=1}^{\tau}, \ell)$ be an instance of DIVERSE MULTISTAGE Π, where $n := \max_{i \in [\tau]} |I_i|$. For each $i \in [\tau]$ we apply Lemma 11 to obtain an ℓ-diverse representative $\mathcal{F}_i$ of $\mathcal{R}(I_i)$ that has size at most $2^{20\ell+o(\ell)} \log n \cdot n^4$ in $2^{20\ell+o(\ell)} \log n \cdot f(r)n^{\mathcal{O}(1)} \cdot g(p/\tau n^4 \cdot 2^{20\ell+o(\ell)} \log n)$ time with error probability p/τ. Observe that the probability of an error occurring at any step is upper-bounded by p. Now we use the following dynamic program to check whether J is a *yes*-instance.

$$\forall i \in \{2, 3, \ldots, \tau\}, S \in \mathcal{F}_i\colon D[i, S] := \begin{cases} \top & \text{if } \exists \widehat{S} \in \mathcal{F}_{i-1}\colon D[i-1, \widehat{S}] = \top \text{ and } |S \Delta \widehat{S}| \geq \ell, \\ \bot & \text{otherwise,} \end{cases}$$

where $D[1, \widehat{S}] = \top$ if and only if $\widehat{S} \in \mathcal{F}_1$. We report that J is a *yes*-instance if and only there is an $S \in \mathcal{F}_\tau$ such that $D[\tau, S] = \top$. Note that this takes $\left(2^{20\ell+o(\ell)} \log n \cdot n^4\right)^2 \tau \subseteq 2^{\mathcal{O}(\ell)} n^{\mathcal{O}(1)} \tau$ time. Hence our overall running time is $2^{\mathcal{O}(\ell)} \cdot f(r_{\max}) n^{\mathcal{O}(1)} \cdot \tau \cdot g(p/\tau n^4 \cdot 2^{20\ell+o(\ell)} \log n)$, where $r_{\max}$ is the maximum of parameter r over all instances of Π in J.

(⇐): We show by induction over $i \in [\tau]$ that if $D[i, S] = \top$, then there is a sequence $(S_j)_{j \in [i]}$ such that $S_i = S$, $S_j \in \mathcal{R}(I_j)$ for all $j \in [i]$ and $|S_{j-1} \Delta S_j| \geq \ell$ for all $j \in \{2, 3, \ldots, i\}$. By definition of D this is clearly the case for $i = 1$. Now let $1 < i \leq \tau$ and $D[i, S] = \top$. Since $D[i, S] = \top$, $S \in \mathcal{F}_i$ and thus $S \in \mathcal{R}(I_i)$. By definition of D there is an $\widehat{S} \in F_{i-1}$ with $D[i-1, \widehat{S}] = \top$ and $|S \Delta \widehat{S}| \geq \ell$. By induction hypothesis, there is a sequence $(S_j)_{j \in [i-1]}$ such that $S_{i-1} = \widehat{S}$, $S_j \in \mathcal{R}(I_j)$ for all $j \in [i-1]$ and $|S_{j-1} \Delta S_j| \geq \ell$ for all $j \in \{2, 3, \ldots, i-1\}$. Hence, the sequence $(S_1, \ldots, S_{i-1} = \widehat{S}, S)$ completes the induction. Thus, if we report that J is a *yes*-instance, then this is true.

($\Rightarrow$): Now let $(S_j)_{j \in [\tau]}$ be a solution for J. To simplify the proof let $S_{\tau+1}$ be a set of ℓ elements that are disjoint from S_τ. We show by induction that for all $i \in [\tau]$ there is a $Z \in \mathcal{F}_i$ such that $D[i, Z] = \top$ and $|Z \Delta S_{i+1}| \geq \ell$.

Let $i = 1$. Then there is a $Z \in \mathcal{F}_1$ such that $|S_2 \Delta Z| \geq \ell$ since $\mathcal{F}_1$ is an ℓ-diverse representative of $\mathcal{R}(I_1)$. Hence, $D[1, Z] = \top$.

Now let $1 < i \leq \tau$. By induction hypothesis, there is a $Z_{i-1} \in \mathcal{F}_{i-1}$ such that $D[i - 1, Z_{i-1}] = \top$ and $|S_i \Delta Z_{i-1}| \geq \ell$. Since $S_i \in \mathcal{R}(I_i)$ and we have $|S_i \Delta Z_{i-1}|, |S_i \Delta S_{i+1}| \geq \ell$ and $\mathcal{F}_i$ is an ℓ-diverse representative of $\mathcal{R}(I_i)$, there is a $Z \in \mathcal{F}_i$ such that $|Z \Delta Z_{i-1}|, |Z \Delta S_{i+1}| \geq \ell$. By definition of D, we also have $D[i, Z] = \top$. This completes the induction step. Thus, there is a $Z \in \mathcal{F}_\tau$ such that $D[\tau, Z] = \top$ and if J is a *yes*-instance, then we report that. ◀

4 Application: Committee Election

Bredereck et al. [11] studied the following problem under the name Revolutionary Multistage Plurality Voting.

Diverse Multistage Plurality Voting

Input: A set A of agents, a set C of candidates, a sequence $(u_i)_{i=1}^{\tau}$ of voting profiles $u_i \colon A \to C \cup \{\emptyset\}$, and integers $k, x, \ell \in \mathbb{N}$.

Question: Is there a sequence $(C_1, C_2, \ldots, C_\tau)$ such that for all $i \in [\tau]$ it holds that $C_i \subseteq C$, $|C_i| \leq k$, and $|u^{-1}(C_i)| \geq x$, and for all $i \in [\tau - 1]$ it holds true that $|C_i \Delta C_{i+1}| \geq \ell$?

In this section, we affirmatively answer the question of Bredereck et al. [11] whether Diverse Multistage Plurality Voting parameterized by ℓ or k is in FPT.[3]

▶ **Theorem 12.** *An instance J of* Diverse Multistage Plurality Voting *can be solved in $2^{\mathcal{O}(\ell)} \cdot |J|^{\mathcal{O}(1)}$ time.*

To prove Theorem 12, we use Theorem 1. In the notation of our framework, we deal with the following problem Π: given an instance $I = (A, C, u, k, x)$ consisting of a set A of agents, a set $C =: \mathcal{B}(I)$ of candidates, a voting profile $u \colon A \to C$, and two integers k, x, decide whether $\mathcal{R}(I) := \{S \subseteq C \mid k \geq |S| \text{ and } |u^{-1}(S)| \geq x\}$ is non-empty. Hence, to apply Theorem 1, we consider the following problem.

4-Colored Exact Plurality Voting

Input: A set A of agents, a set C of candidates, a voting profile $u \colon A \to C \cup \{\emptyset\}$, a coloring $c \colon C \to [4]$, and integers $n_i, x, k \in \mathbb{N}, i \in [4]$.

Output: A set $C' \subseteq C$ of at most k candidates so that $|u^{-1}(C')| \geq x$ and $|c^{-1}(i) \cap C'| = n_i$ for all $i \in [4]$ or "*no*" if no such set exists.

This problem is polynomial-time solvable and hence the following observation together with Theorem 1 proves Theorem 12. In the full version we will generalize this application to independent sets in matroids.

▶ **Observation 13** (★[4]). 4-Colored Exact Plurality Voting *is polynomial-time solvable.*

---

[3] Note that $\ell \leq 2k$ for all non-trivial instances, so it suffices to prove this for ℓ.
[4] Proofs of results marked with a star are deferred to the full version.

5 Application: Perfect Matching

In this section, we apply our framework from Section 3 to find a sequence of diverse perfect matchings.

DIVERSE MULTISTAGE PERFECT MATCHING

Input: A sequence $(G_i)_{i=1}^{\tau}$ of graphs and an integer $\ell \in \mathbb{N}_0$.
Question: Is there a sequence $(M_i)_{i=1}^{\tau}$ of perfect matchings $M_i \subseteq E(G_i)$ such that $|M_i \Delta M_{i+1}| \geq \ell$ for all $i \in [\tau - 1]$?

There are two closely related variants of this problem which were studied extensively. The first variant is the non-diverse variant, where one seeks to bound the symmetric differences (in some way) from above [5, 4, 13, 25, 34]. Steinhau [34] proved that if the size of the symmetric difference of two consecutive perfect matchings shall be at most ℓ, then this problem variant is NP-hard even if ℓ is constant, and W[1]-hard when parameterized by $\ell + \tau$. The second variant is the non-multistage variant, where one is given a single graph and is asked to compute a set of pairwise diverse perfect matchings [22, 23]. Fomin et al. [22] proved that this variant is NP-hard even if one asks only for two diverse matchings. This directly implies NP-hardness for DIVERSE MULTISTAGE PERFECT MATCHING even when $\tau = 2$.

Our goal is to show fixed-parameter tractability of DIVERSE MULTISTAGE PERFECT MATCHING when parameterized by ℓ. This stands in contrast to the NP-hardness for the non-diverse problem variant with constant ℓ.

▶ **Theorem 14 (★).** *An instance J of* DIVERSE MULTISTAGE PERFECT MATCHING *can be solved in $2^{\mathcal{O}(\ell)} \cdot |J|^{\mathcal{O}(1)}$ time with a constant error probability.*

We will prove Theorem 14 by means of Theorem 3 at the end of this section. To this end we need to consider the following problem.

s-COLORED EXACT PERFECT MATCHING

Input: A graph $G = (V, E)$, a coloring $c\colon E \to [s]$, and $k_i \in \mathbb{N}, i \in [s]$.
Output: (if exists) A perfect matching M in G such that $|c^i \cap M| = k_i$, for all $i \in [s]$?

For $s = 2$, this problem is known as EXACT MATCHING, and Mulmuley et al. [32] showed that this special case is solvable by a randomized polynomial-time algorithm. We generalize this result by showing that s-COLORED EXACT PERFECT MATCHING can be solved in polynomial time for any constant s by a randomized algorithm with constant error probability. While we only need this for $s = 4$ in order to prove Theorem 14, we believe that the general case may be of independent interest. We remark that it is open whether EXACT MATCHING can be solved in (deterministic) polynomial time.

▶ **Lemma 15 (★).** *For every $0 < p < 1$ there is an $(n^{\mathcal{O}(s)} \cdot \log 1/p)$-time algorithm which, given an instance of s-*COLORED EXACT PERFECT MATCHING*, finds a solution with probability at least $1 - p$ if one exists, and concludes that there is no solution otherwise.*

To determine whether a given s-COLORED EXACT PERFECT MATCHING has a solution we use the following algorithm.

▶ **Algorithm 16.** Let $0 < p < 1$ and let $I = (G, c, k_1, \ldots, k_s)$ be an instance of s-COLORED EXACT PERFECT MATCHING where $G = (V, E)$ has n vertices.
Step 1. Set $\gamma := \lceil n/(2p) \rceil$ and draw $w_{ij} \in [\gamma]$ for all $\{i, j\} \in E$ uniformly at random.

Step 2. Construct an $n \times n$ matrix A' with entries $a_{ij} \in \mathbb{Z}[y_1, \dots, y_s]$, $1 \le i \le j \le n$, where

$$a_{ij} := \begin{cases} 0 & \text{if } \{i,j\} \notin E, \\ w_{ij} y_q & \text{if } \{i,j\} \in c^q \cap E, q \in [s]. \end{cases}$$

Afterwards we compute the skew-symmetric matrix $A := A' - (A')^T$.

Step 3. Compute the polynomial $P := \sqrt{\det(A)} \in \mathbb{Z}[y_1, \dots, z_s]$.

Step 4. If P contains a monomial $b^* y_1^{k_1} y_2^{k_2} \cdots y_s^{k_s}$ such that $b^* \neq 0$ then, output *yes*. Otherwise, output *no*. ◇

Before studying the running time of Algorithm 16, we first focus on its correctness.

▶ **Lemma 17.** *Let I and p be the input of Algorithm 16. If Algorithm 16 returns* yes*, then there is a solution for I. Conversely, if I is a* yes*-instance, then Algorithm 16 returns* yes *with probability at least $1 - p$.*

Proof. Let $\mathcal{P}$ be the set of all partitions of V into unordered pairs. For $\sigma \in \mathcal{P}$ with $\sigma = \{\{i_1, j_1\}, \{i_2, j_2\}, \dots, \{i_{n/2}, j_{n/2}\}\}$ with $i_k < j_k$ for $k \in [n/2]$ and $i_1 < i_2 < \dots < i_{n/2}$, let

$$\pi_\sigma := \begin{bmatrix} 1 & 2 & 3 & 4 & \cdots & n-1 & n \\ i_1 & j_1 & i_2 & j_2 & \cdots & i_{n/2} & j_{n/2} \end{bmatrix}$$

be the corresponding permutation. Let $\operatorname{val}(\sigma) := \operatorname{sgn}(\pi_\sigma) \prod_{\{i,j\} \in \sigma} a_{ij}$, where $\operatorname{sgn}(\pi_\sigma) \in \{+1, -1\}$ is the signum of π_σ. The Pfaffian of A (computed by Algorithm 16) is defined as $\operatorname{pf}(A) := \sum_{\sigma \in \mathcal{P}} \operatorname{val}(\sigma)$ [27]. Note that A is skew-symmetric, hence, $\operatorname{pf}(A) = \sqrt{\det(A)} = P$ [31, 27]. As $\operatorname{val}(\sigma) = 0$ whenever σ contains a non-edge, we have $P = \sum_{M \in \mathcal{PM}} \operatorname{val}(M)$, where $\mathcal{PM}$ is the set of perfect matchings in G. Let M be a perfect matching and let $z_q = |c^q \cap M|$, $q \in [s]$. Then $\operatorname{val}(M) = \operatorname{sgn}(\pi_M) \prod_{q \in [s]} \prod_{\{i,j\} \in M \cap c^q} w_{ij} y_q = b \cdot y_1^{z_1} y_2^{z_2} \cdots y_s^{z_s}$, where $b \in \mathbb{Z}$. Let $\mathcal{PM}^* \subseteq \mathcal{PM}$ be the family of perfect matchings M^* which have exactly k_i edges of color i, for all $i \in [s]$. Then the coefficient b^* of the monomial $b^* y_1^{k_1} y_2^{k_2} \cdots y_s^{k_s}$ of P is $b^* = \sum_{M^* \in \mathcal{PM}^*} \operatorname{sgn}(\pi_{M^*}) \prod_{\{i,j\} \in M^*} w_{ij}$. Hence, if Algorithm 16 returns *yes* (i.e., $b^* \neq 0$), then $\mathcal{PM}^* \neq \emptyset$.

Now conversely assume I to be a *yes*-instance, i.e., $\mathcal{PM}^* \neq \emptyset$. We analyze the probability of the event $b^* = 0$ occurring. Note that b^* can be seen as a polynomial of degree at most $n/2$ over the indeterminates $\{w_{ij} \mid \{i,j\} \in E\}$. As we have drawn the w_{ij} independently and uniformly at random from $[\gamma]$ with $\gamma \ge n/(2p)$, by the DeMillo-Lipton-Schwartz-Zippel lemma the probability that $b^* = 0$ is at most $n/(2\gamma) \le p$. ◀

Now we show that Algorithm 16 can be executed efficiently.

▶ **Lemma 18 (★).** *Algorithm 16 runs in $n^{O(s)} \log(1/p)$ time.*

We are now ready to put all parts together and prove Lemma 15. In a nutshell, we use Algorithm 16 to check whether there is a solution. If this is the case, then we try to delete as many edges as possible from the instance until the whole edge set is a solution. Putting Lemma 15 and Theorem 3 together, we can prove the main theorem of this section, see the full version for details.

6 Application: s-t Path

In this section, we apply our framework to the task of finding a sequence of diverse s-t paths. This has obvious applications e.g. in convoy routing [21].

Diverse Multistage s-t Path

Input: A sequence of graphs $(G_i)_{i=1}^{\tau}$, two distinct vertices $s, t \in \bigcap_{i=1}^{\tau} V(G_i)$, and $\ell \in \mathbb{N}_0$.
Question: Is there a sequence $(P_1, P_2, \ldots, P_\tau)$ such that P_i is an s-t path in G_i for all $i \in [\tau]$, and $|S_i \Delta S_{i+1}| \geq \ell$ for all $i \in [\tau - 1]$?

Our goal is to show that Diverse Multistage s-t Path parameterized by ℓ is in FPT.

▶ **Theorem 19.** Diverse Multistage s-t Path *parameterized by ℓ is in FPT.*

We will prove Theorem 19 by means of Theorem 1 at the end of this section. To this end, we need to consider the following problem.

4-Colored Exact s-t Path

Input: A graph G, distinct vertices $s, t \in V(G)$, coloring $c\colon V(G) \to [4]$, and $n_i \in \mathbb{N}_0, i \in [4]$.
Output: (if exists) An s-t path P such that $|c^{-1}(i) \cap V(P)| = n_i$ for all $i \in [4]$.

Unfortunately, 4-Colored Exact s-t Path is unlikely to be polynomial-time solvable, as it is NP-hard even if only a single color is used, by a trivial reduction from Hamiltonian Path. However, as we will see in the proof of Theorem 19, by a result of Mousset et al. [30] we can actually reduce 4-Colored Exact s-t Path to the case that all graphs have small treewidth. In this setting, we then employ dynamic programming.

▶ **Lemma 20 (★).** 4-Colored Exact s-t Path *is solvable $k^{\mathcal{O}(k)} \cdot |I|^{\mathcal{O}(1)}$ time, where k is the treewidth of the input graph G.*

While some techniques [15, 24, 9] seem applicable to improve the running time of Lemma 20 slightly, for our needs a straight-forward dynamic program on a nice tree decomposition suffices. We are now ready to prove Theorem 19.

Proof of Theorem 19. Let the instance J of Diverse Multistage s-t Path be given in the form of graphs $G_1, \ldots, G_\tau$, two vertices $s, t \in \bigcap_{i=1}^{\tau} G_i$ and $\ell \in \mathbb{N}$. We may assume that every vertex v of every graph G_i is contained in at least one s-t path in G_i, since otherwise we may delete v. This is equivalent to the assumption that the graph G_i' obtained from adding the edge $\{s, t\}$ to G_i is biconnected.

By a result of Mousset et al. [30], there is a universal constant $\gamma > 0$, such that each G_i with treewidth $\operatorname{tw}(G_i) \geq \gamma\ell$ contains two vertex-disjoint cycles of size at least 4ℓ.

If two such cycles C, C' exist in G_i, then let P_1 be an s-t path containing at least one edge of C. To see that such a path exists, construct a biconnected graph by simply attaching a new degree-two vertex s' to both s and t, create another new vertex t' by subdividing some edge of C, and take two disjoint paths between s' and t'.

Without loss of generality, P_1 enters C and C' at most once each. Construct another s-t path P_2 from P_1 by setting $E(P_2) := E(P_1) \Delta E(C)$. If P_1 contains any edge of C', then define P_3 by $E(P_3) := E(P_1) \Delta E(C')$. Otherwise, let P_3 be any s-t path containing at least half of the edges of C' (this can be achieved analogously to the construction of P_1 resp. P_2). Observe that P_1, P_2, and P_3 have pairwise symmetric differences at least 2ℓ. Thus, $\{P_1, P_2, P_3\}$ is an ℓ-diverse representative of all s-t paths in G_i by Lemma 5.

We can then solve the subinstances given by $(G_j)_{j<i}$ and $(G_j)_{j>i}$ separately and pick a suitable path from $\{P_1, P_2, P_3\}$ afterwards.

All subinstances in which every graph G_i has $\operatorname{tw}(G_i) < \gamma\ell$ can be solved by Theorem 1 in combination with Lemma 20 in $2^{\mathcal{O}(\ell)} f(\gamma\ell)|J|^{\mathcal{O}(1)}$ time, where f is given by Lemma 20. ◀

7 Hardness of Vertex Cover

We finally present a problem where our framework from Section 3 is not applicable, unless FPT = W[1]. The non-diverse variant of the following problem was studied by Fluschnik et al. [20]. Among others, they showed W[1]-hardness when parametrized by the vertex cover size k or by the maximum number of edges over all instances in the input.

DIVERSE MULTISTAGE VERTEX COVER

Input: A sequence of graphs $(G_i)_{i=1}^{\tau}$ and $k, \ell \in \mathbb{N}$.
Question: Is there a sequence $(S_1, S_2, \dots, S_\tau)$ such that for all $i \in [\tau]$ the set $S_i \subseteq V(G_i)$ is a vertex cover of size at most k in G_i and $|S_i \Delta S_{i+1}| \geq \ell$ for all $i \in [\tau - 1]$?

The framework from Section 3 is presumably not applicable to DIVERSE MULTISTAGE VERTEX COVER because of the following result.

▶ **Theorem 21.** DIVERSE MULTISTAGE VERTEX COVER *parameterized by ℓ is W[1]-hard, even if $\tau = 2$.*

Proof. We reduce from INDEPENDENT SET: Given a graph $G = (V, E)$, and $k \in \mathbb{N}$, is there a vertex set $S \subseteq V$, $|S| \geq k$, such that the vertices in S are pairwise nonadjacent? INDEPENDENT SET is W[1]-hard with respect to k [17].

Let $I := (G = (V, E), k)$ be an instance of INDEPENDENT SET and let $|V| = n$. Without loss of generality, we assume that $k > 1$. We construct an instance $J := ((G_1, G_2), k', \ell))$ of DIVERSE MULTISTAGE VERTEX COVER as follows. The first graph G_1 is a complete graph on the vertex set $V \cup \{v\}$. The second graph G_2 consists of the vertex set $V \cup \{v\}$ and the edge set $E \cup \{\{u, v\} \mid u \in V\}$, that is, G_2 is a copy of G to which we add a vertex v which is adjacent to every other vertex. Lastly, we set $k' := n$ and $\ell := k + 1$. Clearly, J can be constructed in polynomial time. We now show that I is a *yes*-instance if and only if J is a *yes*-instance.

($\Rightarrow$): Let S be an independent set of size at least k in G. Let $S_1 := V$ and $S_2 := \{v\} \cup V \setminus S$. Note that $|S_1 \Delta S_2| \geq k + 1$ and $|S_2| \leq |S_1| = n$, and S_i is a vertex cover in G_i for $i \in [2]$. Thus (S_1, S_2) is a valid solution for our instance of DIVERSE MULTISTAGE VERTEX COVER.

($\Leftarrow$): Let (S_1, S_2) be a solution for our instance of DIVERSE MULTISTAGE VERTEX COVER. As G_1 is a complete graph, we have $|S_1| \geq n$. Without loss of generality, we assume that $S_1 = V$. Then $S_1 \Delta S_2 = \{v\} \cup V \setminus S_2$. Note that $v \in S_2$, otherwise S_2 must be equal to V in order to be a vertex cover, and $|S_1 \Delta V| < \ell$. As S_2 is a vertex cover of G_2, the set $S := (V \cup \{v\}) \setminus S_2$ is an independent set of G_2. Note that $v \notin S$, hence S is also an independent set of G. Finally, as $S = (S_1 \Delta S_2) \setminus \{v\}$, we have $|S| \geq \ell - 1 = k$, and we are done. ◀

8 Conclusion

We introduced a versatile framework to show fixed-parameter tractability for a variety of diverse multistage problems when parameterized by the diversity ℓ. The only requirement for applying our framework is that a four-colored variant of the base problem can be solved efficiently. We presented four applications of our framework, one of which resolving an open question by Bredereck et al. [11]. Two other applications revealed problems which may be of independent interest from a technical and motivational point of view, see Sections 5 and 6.

We believe that our framework can be applied to a broad spectrum of multistage problems. In particular, a broad systematic study of the multistage setting in elections was proposed by Boehmer and Niedermeier et al. [10]. Herein, diversity is a natural goal. From a motivational point of view, an interesting direction for future research is to combine the diverse multistage setting with time windows, known from other temporal domains [35, 28, 12, 1, 29]. Here, a solution to the i-th instance should be sufficiently different from the δ previous solutions in the sequence; our work covers the case $\delta = 1$. In some multistage scenarios a "global view" [26] on the symmetric differences is desired. In context of this paper this means that two consecutive solutions can have a small symmetric difference as long as the sum of all consecutive symmetric differences is at least ℓ. We believe that our framework (Section 3) can be extended to this setting. To see this, we have to realize that for an ℓ-diverse representative $\mathcal{F}$ of a family of solutions the following holds: For all sets A and B and integers $\ell_a, \ell_b \leq \ell$, if there is an $S \in \mathcal{F}$ such that $|A \Delta S| \geq \ell_a$ and $|B \Delta S| \geq \ell_b$, then there is an $\widehat{S} \in \mathcal{F}$ such that $|A \Delta \widehat{S}| \geq \ell_a$ and $|B \Delta \widehat{S}| \geq \ell_b$. We leave the details for further research. Finally, the presented time and space constraints to compute ℓ-diverse representatives seem to be suboptimal. Hence, improving the time or space constraints could be a fruitful research direction.

References

1 Eleni C. Akrida, George B. Mertzios, Paul G. Spirakis, and Viktor Zamaraev. Temporal vertex cover with a sliding time window. *Journal of Computer and System Sciences*, 107:108–123, 2020. doi:10.1016/j.jcss.2019.08.002.

2 Noga Alon, Raphael Yuster, and Uri Zwick. Color-coding. *Journal of the ACM*, 42(4):844–856, 1995. doi:10.1145/210332.210337.

3 Emmanuel Arrighi, Henning Fernau, Daniel Lokshtanov, Mateus de Oliveira Oliveira, and Petra Wolf. Diversity in kemeny rank aggregation: A parameterized approach, 2021. arXiv:2105.09413.

4 Evripidis Bampis, Bruno Escoffier, and Alexander V. Kononov. LP-based algorithms for multistage minimization problems, 2019. arXiv:1909.10354.

5 Evripidis Bampis, Bruno Escoffier, Michael Lampis, and Vangelis Th. Paschos. Multistage matchings. In *Proceedings of the 16th Scandinavian Symposium and Workshops on Algorithm Theory (SWAT)*, volume 101, pages 7:1–7:13, 2018. doi:10.4230/LIPIcs.SWAT.2018.7.

6 Evripidis Bampis, Bruno Escoffier, Kevin Schewior, and Alexandre Teiller. Online multistage subset maximization problems. In *Proceedings of the 27th Annual European Symposium on Algorithms (ESA)*, volume 144, pages 11:1–11:14, 2019. doi:10.4230/LIPIcs.ESA.2019.11.

7 Evripidis Bampis, Bruno Escoffier, and Alexandre Teiller. Multistage knapsack. In *Proceedings of the 44th International Symposium on Mathematical Foundations of Computer Science (MFCS)*, volume 138, pages 22:1–22:14, 2019. doi:10.4230/LIPIcs.MFCS.2019.22.

8 Julien Baste, Michael R. Fellows, Lars Jaffke, Tomáš Masařík, Mateus de Oliveira Oliveira, Geevarghese Philip, and Frances A. Rosamond. Diversity of solutions: An exploration through the lens of fixed-parameter tractability theory. In *Proceedings of the 29th International Joint Conference on Artificial Intelligence (IJCAI)*, pages 1119–1125, 2020. doi:10.24963/ijcai.2020/156.

9 Hans L. Bodlaender, Marek Cygan, Stefan Kratsch, and Jesper Nederlof. Deterministic single exponential time algorithms for connectivity problems parameterized by treewidth. *Information and Computation*, 243:86–111, 2015. doi:10.1016/j.ic.2014.12.008.

10 Niclas Boehmer and Rolf Niedermeier. Broadening the research agenda for computational social choice: Multiple preference profiles and multiple solutions. In *Proceedings of the 20th International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, pages 1–5, 2021.

11 Robert Bredereck, Till Fluschnik, and Andrzej Kaczmarczyk. Multistage committee election, 2020. arXiv:2005.02300.

12 Arnaud Casteigts, Anne-Sophie Himmel, Hendrik Molter, and Philipp Zschoche. Finding temporal paths under waiting time constraints. In *Proceedings of the 31st International*

Symposium on Algorithms and Computation (ISAAC), volume 181, pages 30:1–30:18, 2020. doi:10.4230/LIPIcs.ISAAC.2020.30.

13 Markus Chimani, Niklas Troost, and Tilo Wiedera. Approximating multistage matching problems, 2020. arXiv:2002.06887.

14 Marek Cygan, Fedor V. Fomin, Lukasz Kowalik, Daniel Lokshtanov, Dániel Marx, Marcin Pilipczuk, Michal Pilipczuk, and Saket Saurabh. *Parameterized Algorithms*. Springer, 2015. doi:10.1007/978-3-319-21275-3.

15 Marek Cygan, Jesper Nederlof, Marcin Pilipczuk, Michal Pilipczuk, Johan M. M. van Rooij, and Jakub Onufry Wojtaszczyk. Solving connectivity problems parameterized by treewidth in single exponential time. In *Proceedings of the 52nd Annual Symposium on Foundations of Computer Science (FOCS)*, pages 150–159, 2011. doi:10.1109/FOCS.2011.23.

16 Reinhard Diestel. *Graph Theory*, volume 173 of *Graduate Texts in Mathematics*. Springer, 4th edition, 2010.

17 Rodney G. Downey and Michael R. Fellows. *Fundamentals of Parameterized Complexity*. Springer, 2013. doi:10.1007/978-1-4471-5559-1.

18 David Eisenstat, Claire Mathieu, and Nicolas Schabanel. Facility location in evolving metrics. In *Proceedings of the 41st International Colloquium on Automata, Languages, and Programming (ICALP)*, pages 459–470, 2014. doi:10.1007/978-3-662-43951-7_39.

19 Till Fluschnik. A multistage view on 2-satisfiability. In *Proceedings of the 11th International Conference on Algorithms and Complexity (CIAC)*, 2021. To appear. arXiv:2011.02325.

20 Till Fluschnik, Rolf Niedermeier, Valentin Rohm, and Philipp Zschoche. Multistage vertex cover. In *Proceedings of the 14th International Symposium on Parameterized and Exact Computation (IPEC)*, volume 148, pages 14:1–14:14, 2019. doi:10.4230/LIPIcs.IPEC.2019.14.

21 Till Fluschnik, Rolf Niedermeier, Carsten Schubert, and Philipp Zschoche. Multistage s-t path: Confronting similarity with dissimilarity in temporal graphs. In *Proceedings of the 31st International Symposium on Algorithms and Computation (ISAAC)*, volume 181, pages 43:1–43:16, 2020. doi:10.4230/LIPIcs.ISAAC.2020.43.

22 Fedor V. Fomin, Petr A. Golovach, Lars Jaffke, Geevarghese Philip, and Danil Sagunov. Diverse pairs of matchings. In *Proceedings of the 31st International Symposium on Algorithms and Computation (ISAAC)*, volume 181, pages 26:1–26:12, 2020. doi:10.4230/LIPIcs.ISAAC.2020.26.

23 Fedor V. Fomin, Petr A. Golovach, Fahad Panolan, Geevarghese Philip, and Saket Saurabh. Diverse collections in matroids and graphs. In *Proceedings of the 38th International Symposium on Theoretical Aspects of Computer Science (STACS)*, volume 187, pages 31:1–31:14, 2021. doi:10.4230/LIPIcs.STACS.2021.31.

24 Fedor V. Fomin, Daniel Lokshtanov, Fahad Panolan, and Saket Saurabh. Representative sets of product families. In *Proceedings of the 22nd Annual European Symposium on Algorithms (ESA)*, volume 8737, pages 443–454, 2014. doi:10.1007/978-3-662-44777-2_37.

25 Anupam Gupta, Kunal Talwar, and Udi Wieder. Changing bases: Multistage optimization for matroids and matchings. In *Proceedings of the 41st International Colloquium on Automata, Languages, and Programming (ICALP)*, volume 8572, pages 563–575, 2014. doi:10.1007/978-3-662-43948-7_47.

26 Klaus Heeger, Anne-Sophie Himmel, Frank Kammer, Rolf Niedermeier, Malte Renken, and Andrej Sajenko. Multistage graph problems on a global budget. *Theoretical Computer Science*, 868:46–64, 2021. doi:10.1016/j.tcs.2021.04.002.

27 László Lovász and Michael D Plummer. *Matching theory*, volume 367. American Mathematical Society, 2009.

28 George B. Mertzios, Hendrik Molter, Rolf Niedermeier, Viktor Zamaraev, and Philipp Zschoche. Computing maximum matchings in temporal graphs. In *Proceedings of the 37th International Symposium on Theoretical Aspects of Computer Science (STACS)*, volume 154, pages 27:1–27:14, 2020. doi:10.4230/LIPIcs.STACS.2020.27.

29 George B. Mertzios, Hendrik Molter, and Viktor Zamaraev. Sliding window temporal graph coloring. In *Proceedings of the 33rd Conference on Artificial Intelligence (AAAI)*, pages 7667–7674, 2019. doi:10.1609/aaai.v33i01.33017667.

30 Frank Mousset, Andreas Noever, Nemanja Škoric, and Felix Weissenberger. A tight Erdős-Pósa function for long cycles. *Journal of Combinatorial Theory Series B*, 125:21–32, 2017. doi:10.1016/j.jctb.2017.01.004.

31 Thomas Muir. *A Treatise on the Theory of Determinants*. MacMillan and Co., London, 1882.

32 Ketan Mulmuley, Umesh V. Vazirani, and Vijay V. Vazirani. Matching is as easy as matrix inversion. *Combinatorica*, 7(1):105–113, 1987. doi:10.1007/BF02579206.

33 Moni Naor, Leonard J. Schulman, and Aravind Srinivasan. Splitters and near-optimal derandomization. In *Proceedings of the 36th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 182–191, 1995. doi:10.1109/SFCS.1995.492475.

34 Lino Steinhau. Parameterized algorithmics of multistage matching. Bachelor thesis, Technische Universität Berlin, 2020.

35 Philipp Zschoche. A faster parameterized algorithm for temporal matching, 2021. arXiv:2010.10408.