# Diverse Pairs of Matchings

Fedor V. Fomin[1] · Petr A. Golovach[1] · Lars Jaffke[1] · Geevarghese Philip[2,3] ·
Danil Sagunov[4]

## Abstract

We initiate the study of the DIVERSE PAIR OF (MAXIMUM/ PERFECT) MATCHINGS
problems which given a graph $G$ and an integer $k$, ask whether $G$ has two (maxi-
mum/perfect) matchings whose symmetric difference is at least $k$. DIVERSE PAIR OF
MATCHINGS (asking for two not necessarily maximum or perfect matchings) is NP-
complete on general graphs if $k$ is part of the input, and we consider two restricted
variants. First, we show that on bipartite graphs, the problem is polynomial-time
solvable, and second we show that DIVERSE PAIR OF MAXIMUM MATCHINGS is
FPT parameterized by $k$. We round off the work by showing that DIVERSE PAIR OF
MATCHINGS has a kernel on $\mathcal{O}(k^2)$ vertices.

---

A preliminary version of this paper appeared in the proceedings of ISAAC 2020.

---

✉ Lars Jaffke
   lars.jaffke@outlook.com

   Fedor V. Fomin
   fedor.fomin@uib.no

   Petr A. Golovach
   petr.golovach@uib.no

   Geevarghese Philip
   gphilip@cmi.ac.in

   Danil Sagunov
   danilka.pro@gmail.com

1   University of Bergen, Bergen, Norway

2   Chennai Mathematical Institute, Chennai, India

3   UMI ReLaX, Chennai, India

4   St. Petersburg Department of V.A. Steklov Institute of Mathematics, St. Petersburg, Russia

# 1 Introduction

Matching is one of the most fundamental notions in graph theory whose study can be traced back to the classical theorems of Kőnig [20] and Hall [14]. The first chapter of the book of Lovász and Plummer [22] devoted to matching contains a nice historical overview on the development of the matching problem. The problem of finding a maximum size or a perfect matching are the classical algorithmic problems; an incomplete list of references covering the history of algorithmic improvements on these problems is [8, 17, 19, 21, 23, 24, 28, 32], see also the book of Schrijver [29] for a historical overview of matching algorithms.

In this paper we initiate the algorithmic study of the *diverse* matching problem. In this problem, we are to find a pair of matchings which are different from each other as much as possible. More formally, we want the size of their symmetric difference to be large. Recall that the *symmetric difference* of two sets $X, Y$ is defined as

$$X \triangle Y = (X \setminus Y) \cup (Y \setminus X).$$

---

DIVERSE PAIR OF (MAXIMUM/PERFECT) MATCHINGS

*Input:* Graph $G$, integer $k$

*Question:* Does $G$ contain two (maximum/perfect) matchings $M_1, M_2$ such that $|M_1 \triangle M_2| \geq k$?

---

Diversity-enhancing is one of the key goals in developing professional social matching systems [26]. For example, consider the problem of assigning agents to perform various tasks (say, bus drivers to bus routes or cleaners to different locations). To avoid monotony, which is one of the declared enemies of happiness at work, the practice is to reassign agents to new tasks. In this case, we would be very much interested in designing a schedule with diverse assignments. To give another illustration, assume that a teacher should give a series of assignments to students that are expected to work in pairs. From one side, the teacher wishes to follow the preferences of the students given by a graph, but from the other side, it is preferable to facilitate collaboration between different students. This leads to the problem of finding diverse perfect matchings in the preference graph.

We now briefly motivate why finding a diverse set of maximum/perfect matchings in a graph would be of interest. From a graph-theoretic point of view, in the simplest model, one maximum/perfect matching is as good as the other. But in a practical setting this is rarely the case since there is a large amount of *side information* that determines how an assignment (for instance agents to tasks) is received. Some side information is modeled by maximum weight matchings, or via notions from social choice theory such as stable or envy free matchings [4]. Nevertheless, this approach has its natural limitations; some side information may complicate the model, rendering it intractable, while some side information may even be impossible to include in a model.

For instance, if we allow agents to have incomplete preference lists or ties, then the corresponding maximum stable matching problem is NP-hard, even in severely restricted cases [27]. Other side information may be a priori unknown, and only once

presented with a number of alternatives, we may be able to decide which assignment is the most desirable. In that case it is key that the presented alternatives are diverse, otherwise the insight we gain is comparable to that of having a single fixed assignment and is therefore negligible. Similar motivations for finding diverse solution sets in combinatorial problems can be found in [2, 3].

**Our Results and Methods**    While a perfect or a maximum matching in a graph can be found in polynomial time, this is not true anymore for the diverse variant of the problem, even in graphs of maximum degree three. Matching problems are often considered on bipartite graphs, and we show that DIVERSE PAIR OF MAXIMUM MATCHINGS remains polynomial-time solvable in this case.

The intractability of the problem in the general case also suggests to look at it from the perspective of parameterized complexity [5, 7] and kernelization [12]. We show that the problem is FPT parameterized by $k$, by giving a randomized $4^k \cdot n^{\mathcal{O}(1)}$ time algorithm, and we give a derandomized version of this algorithm that runs in time $4^k k^{\mathcal{O}(\log k)} \cdot n^{\mathcal{O}(1)}$. Finally, we show that the problem asking for a diverse pair of (not necessarily maximum) matchings admits a kernel on $\mathcal{O}(k^2)$ vertices.

The randomized algorithm for DIVERSE PAIR OF MAXIMUM MATCHINGS is obtained via a combination of color-coding [1] and the polynomial-solvability of finding a *minimum cost* maximum matching in a graph [13]. We derandomize this algorithm via universal sets [25]. The kernelization algorithm for DIVERSE PAIR OF MATCHINGS first finds a maximal matching $M$ in the graph. If $M$ is large enough, then we can conclude that we are dealing with a YES-instance by splitting $M$ into two matchings. Otherwise, the endpoints of $M$ form a vertex cover of the input graph which allows us to shrink the graph without changing the answer to the problem.

**Related Work**    A well-studied generalization of matchings in graphs is that of a *b-matching*, where $b$ is an integer; see for instance [22]. Given a graph $G$ and an integer $b$, a *b-matching* is an assignment of an integer $\mu(e)$ to each edge $e$ of $G$, such that for each vertex $v$, the sum over all its incident edges $e_v$ of $\mu(e_v)$ is bounded by $b$. The size of a $b$-matching is the sum over all edges $e$ in $G$ of $\mu(e)$. The 1-matchings of a graph precisely correspond its matchings (via the edges $e$ with $\mu(e) = 1$). However, a 2-matching is not always the union of two matchings: take for instance a triangle. Then, assigning a value of 1 to all its edges gives a 2-matching; while any matching can have at most one edge from a triangle. Therefore, finding diverse pairs of matchings is not the same as finding 2-matchings.

Finding $q$ pairwise *disjoint* matchings of large total size corresponds to finding large subgraphs that can be $q$-edge colored, each matching constitutes a color class. The MAXIMUM $q$-EDGE COLORABLE SUBGRAPH problem asks for the largest edge-subgraph that can be properly colored with $q$ colors. This problem is known to be hard to approximate [9].

Let $G$ be a graph with edge set $E$ and maximum degree $\Delta$. Any proper edge coloring requires at least $\Delta$ colors. On the other hand, Vizing's Theorem [33] asserts that every graph can be properly edge-colored with $\Delta + 1$ colors. A consequence of this result is that (any) graph $G$ contains a $\Delta$-colorable subgraph with at least $\frac{\Delta}{\Delta+1}|E|$ edges; which is tight when $\Delta$ is even as witnessed by the complete graph $K_{\Delta+1}$. This motivated

research in improving the lower bound when $\Delta$ is odd, or when $\Delta$ is even and $K_{\Delta+1}$ is excluded. Kamiński and Kowalik [18] gave several improved lower bounds for the cases when $\Delta \leq 7$.

The difference with MAXIMUM 2-EDGE COLORABLE SUBGRAPH is that in DIVERSE PAIR OF MAXIMUM MATCHINGS, we require matchings (or: color classes) to be of maximum size, while in the former problem, we only want to maximize the total number of edges in the two color classes.

A recent manuscript due to Fellows [10] initiated the study of finding diverse sets of solutions to NP-hard combinatorial problems from the viewpoint of parameterized complexity [2, 3]. Concretely, Baste et al. [2] showed that a large class of vertex subset problems that are FPT parameterized by treewidth have FPT algorithms in their diverse variant, parameterized by treewidth plus the number of requested solutions. Moreover, Baste et al. [3] showed analogous results for hitting set problems parameterized by solution size plus number of requested solutions. Our work contrasts this in that the classical variant of the problem we consider is polynomial-time solvable, while its diverse variant becomes NP-hard, even when asking for only two solutions.

Very recently, Hanaka et al. [15] gave efficient algorithms for finding diverse sets of solutions to several other combinatorial problems. This includes an FPT-algorithm for finding diverse sets of matchings in a graph. However, their result is different from ours. We give an FPT-algorithm for finding a diverse pair of *maximum* or *perfect* matchings, and our parameter is the size of the symmetric difference between the matchings, in other words, the diversity measure. In [15], the parameter is the size of the matchings plus the number of requested solutions, and the matchings do not need to be of maximum cardinality. Note that in this setting, the maximum possible diversity is bounded in terms of the parameter as well. In the case that we drop the maximum cardinality requirement on the matchings, we even obtained a *polynomial kernel* for finding a diverse pair of matchings. By the same arguments given in the proof of Theorem 7, we can derive that the problem of finding a diverse set of $r$ matchings of size $k$ parameterized by $k + r$ considered in [15] is not only FPT but has a polynomial kernel.

Fomin et al. [11] considered the DIVERSE PERFECT MATCHINGS problem, asking for $q$ perfect matchings such that each pair has Hamming distance at least $k$. Thus, DIVERSE PAIR OF PERFECT MATCHINGS is a special case of DIVERSE PERFECT MATCHINGS for $q = 2$. They showed that the problem is FPT parameterized by $q+k$ by giving a randomized algorithm running in $2^{2^{\mathcal{O}(qk)}} n^{\mathcal{O}(1)}$ time. This generalizes our FPT-result for DIVERSE PAIR OF PERFECT MATCHINGS (but not for DIVERSE PAIR OF MAXIMUM MATCHINGS), albeit with a slower run time, and without derandomization.

## 2 Preliminaries

We assume the reader to be familiar with basic notions in graph theory and parameterized complexity and refer to [6] and [5, 7, 12], respectively, for the necessary background.

All graphs considered in this work are finite, undirected, simple, and without self-loops. For a graph $G$ we denote by $V(G)$ its set of *vertices* and by $E(G)$ its set of *edges*. For an edge $uv \in E(G)$, we call $u$ and $v$ its *endpoints*. For a vertex $v$ of a graph $G$, $N_G(v) := \{w \in V(G) \mid vw \in E(G)\}$ is the set of *neighbors* of $v$ in $G$, and the *degree* of $v$ is $\deg_G(v) := |N_G(v)|$.

The *subgraph induced by $X$*, denoted by $G[X]$, is the graph $(X, \{uv \in E(G) \mid u, v \in X\})$. For a set of edges $F \subseteq E(G)$, we let $G - F := (V(G), E(G) \backslash F)$.

A graph $G$ is called *empty* if $E(G) = \emptyset$. A set of vertices $S \subseteq V(G)$ is an *independent set* if $G[S]$ is empty. A set $S \subseteq V(G)$ is a *vertex cover* if $V(G) \backslash S$ is an independent set. A graph $G$ is *bipartite* if its vertex set can be partitioned into two nonempty independent sets.

**NP-Completeness**   We briefly argue the NP-completeness of DIVERSE PAIR OF MAXIMUM/PERFECT MATCHINGS on 3-regular graphs which was observed in [30]. Membership in NP is clear. To show NP-hardness, we reduce from 3-EDGE COLORING on 3-regular graphs which is known to be NP-complete [16]. Let $G$ be a 3-regular graph on $n$ vertices (note that this implies that $n$ is even), and consider $(G, n)$ as an instance of DIVERSE PAIR OF MATCHINGS. Suppose that $G$ has a proper 3-edge coloring. Since $G$ is 3-regular, all three colors appear on an incident edge of each vertex. Therefore, a color class is a perfect matching of $G$, and we can take two color classes as our solution to $(G, n)$. Conversely, a solution $(M_1, M_2)$ to $(G, n)$ forms two disjoint matchings of size $n/2$ each. This implies that both $M_1$ and $M_2$ are perfect, and therefore maximum matchings. Since each vertex in $G$ has degree three, this means that $M_3 := E(G) \setminus (M_1 \cup M_2)$ also forms a perfect matching in $G$, and therefore $(M_1, M_2, M_3)$ is a proper 3-edge coloring of $G$.

***Observation 1*** ([30]) DIVERSE PAIR OF (MAXIMUM/PERFECT) MATCHINGS is NP-complete on 3-regular graphs.

## 3 A Polynomial-Time Algorithm for Bipartite Graphs

In this section we show that DIVERSE PAIR OF MAXIMUM MATCHINGS is solvable in polynomial time on bipartite graphs via a reduction to the 2-FACTOR problem.

**Theorem 2** DIVERSE PAIR OF MAXIMUM MATCHINGS *is polynomial-time solvable on bipartite graphs.*

***Proof*** Let $(G, k)$ be a given instance of DIVERSE PAIR OF MAXIMUM MATCHINGS, where $G$ is bipartite. We show how to reduce this instance to an equivalent instance of the problem of finding maximum-weight 2-factor of a larger graph $G'$. A 2-*factor* of $G'$ is a subgraph of $G'$ in which the degree of each vertex is equal to 2. Equivalently, 2-factor of $G'$ is a vertex-disjoint cycle cover of $G'$. The problem of finding a maximum-weight factor of a graph is well-known to be solvable in polynomial time using the Tutte's reduction to the problem of finding a maximum-weight perfect matching [22, 31]. Our graph $G'$ is an edge-weighted graph with parallel edges. We note that the algorithm of finding a maximum-weight factor works fine with such graphs.

We also assume that the two parts of $G$ are of equal size. If that is not true, introduce isolated vertices to the smaller part of $G$. This does not change the matching structure in $G$, so the obtained instance is equivalent to the initial one. Denote the number of vertices in each part of $G$ by $n$, so $|V(G)| = 2n$.

We now show how to construct $G'$ given $G$. The graph $G'$ is defined on the same vertex set as $G$ is, i.e. $V(G') = V(G)$. For each edge $uv$ of $G$, $G'$ has two parallel edges between $u$ and $v$. One of these edges is assigned weight 1, and the other is assigned weight 0. In other words, edges of $G$ are doubled in $G'$. Additionally, for each pair of vertices $u, v$ from distinct parts of $G$ that are not adjacent in $G$, $G'$ has two parallel edges of weight $-n$ between $u$ and $v$. Thus, $G'$ is a complete bipartite graph with doubled edges, and weights of these edges depend on what edges are present in $G$. This finishes the construction of $G'$.

**Calim 2.1** *Let $M_1$ and $M_2$ be a pair of maximum matchings in $G$ that maximize the value of $|M_1 \cup M_2|$. Then the maximum weight of a 2-factor of $G'$ equals $|M_1 \cup M_2| - 2n \cdot (n - |M_1|)$.*

**Proof** We first show that $G'$ has a 2-factor of weight at least $|M_1 \cup M_2| - 2n \cdot (n - |M_1|)$. Denote this 2-factor by $F$. It is constructed as follows. For each edge of $M_1$ take the corresponding edge of weight 1 in $G'$ into $F$. Then, for each edge in $M_2 \setminus M_1$ take the corresponding edge of weight 1 into $F$. For each edge in $M_1 \cap M_2$, take the corresponding edge of weight 0 in $G'$ into $F$. Clearly, $F$ is now of weight $|M_1 \cup M_2|$, but it is not yet a 2-factor of $G'$, unless $M_1$ and $M_2$ are perfect matchings.

There are $n - |M_1|$ vertices in each part of $G$ that are not saturated by $M_1$. Take these $2(n - |M_1|)$ vertices and take an arbitrary matching between them in $G'$. All edges of this matching are of weight $-n$, otherwise $M_1$ is not maximum in $G$. Add the edges of this matching into $F$. Repeat the same for $M_2$, i.e. take an arbitrary matching in $G'$ between vertices that are not saturated by $M_2$ and add all its edges into $F$. The edges of weight $-n$ of the matchings for $M_1$ and $M_2$ may coincide. If an edge of weight $-n$ is presented in both matchings, take both its parallel copies into $F$. It is easy to see that $F$ is now a 2-factor of $G'$, as it consists of edges of two perfect matchings between two parts. The weight of $F$ is $|M_1 \cup M_2| - n(n - |M_1|) - n(n - |M_2|) = |M_1 \cup M_2| - 2n \cdot (n - |M_1|)$.

It is left to show that $F$ is indeed a maximum-weight 2-factor of $G'$. To see this, take a maximum-weight factor $F'$ of $G'$ and assume that the weight of $F'$ is greater than the weight of $F$. Note that $F'$ consists of $2n$ edges. As discussed above, $F'$ forms a disjoint union of simple cycles on the vertices of $G'$, where each vertex belongs to exactly one cycle. Note that some of these cycles may consist of two parallel edges. Since $G'$ is bipartite, all of these cycles have even length. Color the edges of $F'$ with two colors so that no two consecutive edges have the same color on a cycle. Then the edges of the same color form a perfect matching in $G'$. Denote these matchings by $F'_1$ and $F'_2$. Let $M'_1$ be the set of original edges of $G$ which copies are present in $F'_1$. Let $M'_2$ be the set of edges of $G$ obtained analogously from $F'_2$. Copies of edges in $M'_1$ and $M'_2$ have weights 0 or 1 in $G'$. All other edges in $F'_1$ and $F'_2$ are of weight $-n$.

Observe that if an edge of $G$ is present in both $M'_1$ and $M'_2$, then one of its copies in $F'$ has weight 0, and the other has weight 1. Thus, the total weight of 0- and 1-weighted edges in $F'$ is at most $|M'_1 \cup M'_2|$. The number of edges of weight $(-n)$ in

$F'$ is $2n - |M_1'| - |M_2'|$. Thus, the total weight of $F'$ is at most $|M_1' \cup M_2'| - n(2n - (|M_1'| + |M_2'|))$.

We assumed that the weight of $F'$ is greater than the weight of $F$. From this we get that $(|M_1' \cup M_2'| - |M_1 \cup M_2|) - n(2|M_1| - (|M_1'| + |M_2'|)) > 0$ holds; equivalently, that

$$(|M_1' \cup M_2'| - |M_1 \cup M_2|) > n(2|M_1| - (|M_1'| + |M_2'|)). \tag{1}$$

Recall that $M_1'$ and $M_2'$ are matchings in $G$. Suppose $M_1'$ and $M_2'$ are *maximum* matchings in $G$. Then the right hand side of Eq. 1 evaluates to zero, and—by the definition of $M_1$ and $M_2$—the left hand side is *at most* zero. Hence Eq. 1 does not hold, a contradiction. So at least one of $M_1'$ and $M_2'$ is *not* a maximum matching. Thus we get that

$$|M_1'| + |M_2'| < 2|M_1| \tag{2}$$

holds; equivalently, that $|M_1'| + |M_2'| - |M_1| < |M_1|$ holds. By construction we have that the size of any matching in $G$ is at most $n$. In particular $|M_1| \leq n$, and so we have that

$$|M_1'| + |M_2'| - |M_1| < n \tag{3}$$

holds.

Equation 2 can be restated as $2|M_1| - (|M_1'| + |M_2'|) > 0$. Now,

$$2|M_1| - (|M_1'| + |M_2'|) \geq 1 \tag{4}$$

holds. Substituting Eq. 4 in Eq. 1 we get that

$$(|M_1' \cup M_2'| - |M_1 \cup M_2|) > n \tag{5}$$

holds. Observe now that $|M_1'| + |M_2'| \geq |M_1' \cup M_2'|$ and $|M_1| \leq |M_1 \cup M_2|$ hold. Substituting these in Eq. 5 we get that $((|M_1'| + |M_2'|) - |M_1|) > n$ holds, which contradicts Eq. 3.

Now let $M_1, M_2$ be two arbitrary maximum matchings of $G$, and let $\mu(G)$ denote the size of a maximum matching of $G$. Thus $|M_1| = |M_1| = \mu(G)$. By the definition of symmetric difference we have that $|M_1 \triangle M_2| = |M_1 \setminus (M_1 \cap M_2)| + |M_2 \setminus (M_1 \cap M_2)| = |M_1| - |M_1 \cap M_2| + |M_2| - |(M_1 \cap M_2)| = 2\mu(G) - 2|M_1 \cap M_2|$. And since $|(M_1 \cap M_2)| = |M_1| + |M_2| - |M_1 \cup M_2| = 2\mu(G) - |M_1 \cup M_2|$ we get that $|M_1 \triangle M_2| = 2\mu(G) - 2(2\mu(G) - |M_1 \cup M_2|) = 2(|M_1 \cup M_2| - \mu(G))$. Since $\mu(G)$ is an invariant of graph $G$ this means that the maximum value of $|M_1 \triangle M_2|$ is attained by exactly those pairs of maximum matchings $M_1, M_2$ which maximize the value $|M_1 \cup M_2|$. Further, let $M_1^\star, M_2^\star$ be a pair of maximum matchings such that $|M_1^\star \cup M_2^\star|$ is the maximum among all pairs of maximum matchings. Then we have

that the maximum value of $|M_1 \triangle M_2|$, over all pairs of maximum matchings, equals $2(|M_1^\star \cup M_2^\star| - \mu(G))$.

From 2.1 we get that we can compute the value $|M_1^\star \cup M_2^\star|$—though not the matchings $M_1^\star$ and $M_2^\star$—in polynomial time, by computing the weight of a maximum 2-factor in a derived graph. We can find the maximum matching size $\mu(G)$ of $G$ in polynomial time as well. So we can compute the number $2(|M_1^\star \cup M_2^\star| - \mu(G))$ in polynomial time. By the arguments in the previous paragraph, checking whether $2(|M_1^\star \cup M_2^\star| - \mu(G)) \geq k$ suffices to solve the bipartite instance $(G, k)$ of DIVERSE PAIR OF MAXIMUM MATCHINGS. □

## 4 FPT-Algorithm for DIVERSE PAIR OF MAXIMUM MATCHINGS

In this section we give an FPT-algorithm for DIVERSE PAIR OF MAXIMUM MATCHINGS parameterized by $k$. We first give a randomized algorithm based on the color-coding technique of Alon, Yuster and Zwick [1] in Theorem 3, and then derandomize this algorithm at the cost of a slightly slower runtime in Corollary 6.

**Theorem 3** DIVERSE PAIR OF MAXIMUM MATCHINGS *parameterized by k is* FPT. *More precisely, there is a randomized algorithm that in time* $4^k \cdot n^{\mathcal{O}(1)}$ *finds a solution with constant probability, if it exists, and correctly concludes that there is no solution otherwise, where n denotes the number of vertices of the input graph.*

**Proof** Let $G$ be the graph of the given instance. First, we compute a maximum matching $M$ in $G$ in polynomial time [8, 13, 24].

We check if there is a solution using $M$ as one of the two matchings.

**Calim 3.1** *Let G be a graph and M a maximum matching of G. One can determine in polynomial time whether G has a maximum matching* $M'$ *such that* $|M \triangle M'| \geq k$, *and construct such a matching if it exists.*

**Proof** The algorithm is as follows. Let $c \colon E(G) \to \{0, 1\}$ be a cost function of the edges of $G$, defined as

$$c(e) := \begin{cases} 1, & \text{if } e \in M \\ 0, & \text{otherwise} \end{cases} \quad \forall e \in E(G)$$

Let $M'$ be a minimum cost maximum matching in $G$ using the cost function $c$. Such a matching $M'$ can be found in polynomial time [13]. Due to the cost function $c$, a minimum cost maximum matching in $G$ is one that minimizes the number of edges from $M$. Therefore, $M'$ maximizes the symmetric difference with $M$, over all maximum matchings of $G$. We verify whether $|M \triangle M'| \geq k$, and if so, return $M'$. Otherwise, we correctly conclude that there is no matching satisfying the conditions of the claim. □

Due to Claim 3.1, we may now assume that for each maximum matching $M'$ of $G$, $|M' \triangle M| \leq k$. We will exploit this property to give an algorithm using color coding (see e.g. [5, Chapter 5]).

We color the edges of $G$ uniformly at random with colors red and blue. For ease of exposition, we also use the notation 'red' and 'blue' to denote the *set* of edges that received color red and blue, respectively.

Suppose that there is a solution $(M_1, M_2)$. We say that a coloring as above is *good for* $(M_1, M_2)$, if the edges in $M_1 \setminus M_2$ and $M_2 \setminus M_1$ are colored red and blue, respectively. We call an edge coloring *good*, if it is good for some solution. To be able to show that trying $4^k$ colorings to achieve constant success probability suffices, we bound the size of these sets.

By Claim 3.1, we know that $|M \triangle M_r| \leq k$ for all $r \in \{1, 2\}$. Since $|M_1 \triangle M_2|$ is the Hamming distance between sets, by the triangle inequality,

$$|M_1 \triangle M_2| \leq |M_1 \triangle M| + |M \triangle M_2| \leq 2k$$

and $|M_1 \setminus M_2| + |M_2 \setminus M_1| \leq 2k$. This leads to the following observation.

**Observation 3.2** Let $G$ be a graph, let $M$ be a maximum matching of $G$, and suppose that for all maximum matchings $M'$ of $G$, $|M \triangle M'| \leq k$.

Suppose the edges of $G$ are colored uniformly at random with colors red and blue.

Suppose there is a solution $(M_1, M_2)$. Then, with probability at least $2^{-2k}$, the edge coloring is good for $(M_1, M_2)$.

Suppose that our instance is a YES-instance, and that the edges of $G$ are colored with a good coloring. We show how to obtain the solution in polynomial time from the edge-colored graph.

**Calim 3.3** *Let $G$ be a graph, $M$ a maximum matching of $G$, and suppose that the edges of $G$ are colored uniformly at random with colors* red *and* blue. *There is an algorithm that runs in polynomial time, and if the edge-coloring is good, finds two maximum matchings $M_1$ and $M_2$ in $G$ such that $|M_1 \triangle M_2| \geq k$, and reports* NO *otherwise.*

**Proof** The idea is similar to the one given in the algorithm of Claim 3.1. To find $M_1$, we define the following cost function $c_1 \colon E(G) \to \{0, 1\}$:

$$c_1(e) := \begin{cases} 1, & \text{if } e \in \text{blue} \\ 0, & \text{if } e \in \text{red} \end{cases} \quad \forall e \in E(G).$$

Then, we find a minimum-cost maximum matching $M_1$ of $G$ with the cost function $c_1$ in polynomial time [22].

Next, to find $M_2$, we consider the cost function $c_2 \colon E(G) \to \{0, 1\}$, where

$$c_2(e) := \begin{cases} 1, & \text{if } e \in \text{red} \\ 0, & \text{if } e \in \text{blue} \end{cases} \quad \forall e \in E(G),$$

and find a minimum-cost maximum matching $M_2$ of $G$ with cost function $c_2$ in polynomial time [22].

Now, if $|M_1 \triangle M_2| \geq k$, then we return $(M_1, M_2)$, and we say NO, otherwise.

We now argue the correctness of the algorithm in the case that the edge-coloring of $G$ was good. In this case, there is a solution $(M_1^*, M_2^*)$ such that the edges of $M_1^* \setminus M_2^*$ are red and the edges of $M_2^* \setminus M_1^*$ are blue, and $|M_1^* \triangle M_2^*| \geq k$. We claim that $|M_1 \triangle M_2| \geq |M_1^* \triangle M_2^*|$.

To obtain a contradiction, assume that $|M_1 \triangle M_2| < |M_1^* \triangle M_2^*|$.

Since $M_1$, $M_2$, $M_1^*$, and $M_2^*$ are maximum matchings of $G$, they have the same size. Therefore, we have that

$$|M_1 \triangle M_2| + 2|M_1 \cap M_2| = |M_1| + |M_2| = |M_1^*| + |M_2^*| =$$
$$|M_1^* \triangle M_2^*| + 2|M_1 \cap M_2|.$$

Since $|M_1 \triangle M_2| < |M_1^* \triangle M_2^*|$, we obtain that

$$|M_1 \cap M_2| > |M_1^* \cap M_2^*|. \tag{6}$$

Because $M_1^* \setminus M_2^* \subseteq$ red, $c_1(M_1^*) = |M_1^* \cap \text{blue}| = |(M_1^* \cap M_2^*) \cap \text{blue}|$ by the definition of the cost function $c_1$. Symmetrically, $c_2(M_2^*) = |(M_1^* \cap M_2^*) \cap \text{red}|$. Hence,

$$c_1(M_1^*) + c_2(M_2^*) = |(M_1^* \cap M_2^*) \cap \text{blue}| + |(M_1^* \cap M_2^*) \cap \text{red}| = |M_1^* \cap M_2^*|. \tag{7}$$

Notice that $c_1(M_1) = |M_1 \cap \text{blue}| \geq |(M_1 \cap M_2) \cap \text{blue}|$ and $c_2(M_2) = |M_2 \cap \text{red}| \geq |(M_1 \cap M_2) \cap \text{red}|$. Therefore,

$$c_1(M_1) + c(M_2) \geq |(M_1 \cap M_2) \cap \text{blue}| + |(M_1 \cap M_2) \cap \text{red}| = |M_1 \cap M_2|. \tag{8}$$

Combining (6)–(8), we obtain that

$$c_1(M_1) + c_2(M_2) \geq |M_1 \cap M_2| > |M_1^* \cap M_2^*| = c_1(M_1^*) + c_2(M_2^*). \tag{9}$$

However, $c_1(M_1) \leq c_1(M_1^*)$ and $c_2(M_2) \leq c_2(M_2^*)$ by the definition of these matchings and

$$c_1(M_1) + c_2(M_2) \leq c_1(M_1^*) + c_2(M_2^*)$$

contradicting (9). We conclude that $|M_1 \triangle M_2| \geq |M_1^* \triangle M_2^*|$.

The outline of the procedure is given in Algorithm 1. It is well-known that a maximum matching of a graph can be found in polynomial time, therefore line 1 takes polynomial time. By Claims 3.1 and 3.3, lines 1 and 1, respectively, take polynomial time. Moreover, by Observation 3.2, if there is a solution, then with probability at least $2^{-2k}$ an edge-coloring as constructed in line 1 is good, in which case the algorithm finds the solution by Claim 3.3. It is clear that repeating this step $2^{2k}$ times yields a constant success probability. $\qquad\square$

**Input** : Graph $G$, integer $k$
**Output**: If exists, with constant probability, a pair $M_1$, $M_2$ of maximum matchings of $G$ such that
$\qquad |M_1 \triangle M_2| \geq k$, No otherwise.
Compute a maximum matching $M$ of $G$;
**if** *there is a maximum matching $M'$ in $G$ such that $|M \triangle M'| \geq k$* **then**
$\quad$ | $\quad$ return $(M, M')$;
**end**
**else**
$\quad$ | **repeat** $2^{2k}$ **times**
$\quad$ | $\quad$ | Color the edges of $G$ uniformly at random with colors red and blue;
$\quad$ | $\quad$ | run the algorithm of Claim 3.3;
$\quad$ | $\quad$ | **if** *the algorithm returned $(M_1, M_2)$* **then return** $(M_1, M_2)$
$\quad$ |
$\quad$ | **return** No;
**end**

**Algorithm 1:** The algorithm of Theorem 3.

The algorithm of Theorem 3 can be derandomized by standard tools (see, e.g., [5, Chapter 5]). To do so, we use the following notion of $(\Omega, k)$-universal sets, which will replace the random coloring step in the above algorithm by deterministic choices of colorings.

**Definition 4** $((\Omega, k)$-universal set) Let $\Omega$ be a set and $k$ be a positive integer with $k \leq |\Omega|$. An $(\Omega, k)$-*universal set* is a family $\mathcal{U}$ of subsets of $\Omega$ such that for any size-$k$ set $S \subseteq \Omega$, the family $\mathcal{U}_S := \{A \cap S : A \in \mathcal{U}\}$ contains all subsets of $S$.

We will use the following construction of a small universal set due to Naor et al. [25].

**Theorem 5** *([25], see also Theorem 5.20 in [5]) For any set $\Omega$ and integer $k \leq |\Omega|$, one can construct an $(\Omega, k)$-universal set of size $2^k k^{\mathcal{O}(\log k)} \log(|\Omega|)$ in time $2^k k^{\mathcal{O}(\log k)} |\Omega| \log(|\Omega|)$.*

This immediately gives the following corollary.

**Corollary 6** *There is a deterministic $4^k k^{\mathcal{O}(\log k)} \cdot n^{\mathcal{O}(1)}$ time algorithm that solves* Diverse Pair of Maximum Matchings, *where $n$ denotes the number of vertices in the input graph.*

## 5 Polynomial kernel for Diverse Pair of Matchings

We now show that the Diverse Pair of Matchings problem, asking for a pair of not necessarily maximum matchings has a kernel on $\mathcal{O}(k^2)$ vertices. Note that the NP-completeness of this problem is captured in Observation 1 as well. Moreover, we would like to remark that this variant of the problem is only interesting in the case when the input graph has no matching of size $k$ or more: otherwise, a maximum matching (which can be found in polynomial time) forms a trivial solution together with an empty matching.

**Theorem 7** Diverse Pair of Matchings *parameterized by $k$ has a kernel on $\mathcal{O}(k^2)$ vertices.*

**Fig. 1** Illustration of the situation in the proof of Claim 7.1. The existence of $v$ implies that $|X_u| \geq 2k$, and since $V(M)$ is a vertex cover of $G$, the vertices in $X_u$ are pairwise non-adjacent
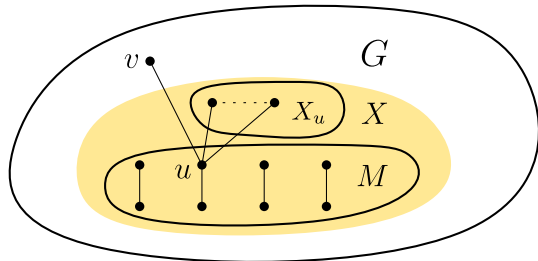


**Proof** Let $(G, k)$ be an instance of DIVERSE PAIR OF MATCHINGS. We provide a procedure that either correctly concludes that $(G, k)$ is a YES-instance, or marks a set of $\mathcal{O}(k^2)$ vertices $X \subseteq V(G)$ such that $(G[X], k)$ is equivalent to $(G, k)$.

First, let $M$ be a maximal matching of $G$. If $|M| \geq k$, then for any 2-partition $(M_1, M_2)$ of $M$, we have that $|M_1 \triangle M_2| = |M| \geq k$, and therefore $(G, k)$ is a YES-instance.

Suppose that $|M| < k$ and therefore, $|V(M)| < 2k$. Since $M$ is maximal, $V(M)$ is a vertex cover of $G$, and therefore, $E(G - V(M)) = \emptyset$. This motivates the following procedure that produces a set of marked vertices $X \subseteq V(G)$, to which we can restrict the instance without changing the answer.

1. Initialize $X := V(M)$.
2. For each $v \in V(M)$, add a maximal subset of $N_G(v) \setminus V(M)$ of size at most $2k$ to $X$.

Let $X$ denote the set constructed according to the two previous steps. We show that $(G[X], k)$ is equivalent to $(G, k)$.

**Calim 7.1** *Let $G$, $k$, $M$, and $X$ be as above. Then, $(G, k)$ is a* YES-*instance of* DIVERSE PAIR OF MATCHINGS *if and only if $(G[X], k)$ is a* YES-*instance of* DIVERSE PAIR OF MATCHINGS.

**Proof** Since $G[X]$ is a subgraph of $G$, it is clear that if $(G[X], k)$ is a YES-instance, then so is $(G, k)$.

Now suppose that $(G, k)$ is a YES-instance and let $(M_1, M_2)$ with $|M_1 \triangle M_2| \geq k$ be a solution. If $M_1 \cup M_2 \subseteq E(G[X])$, then $(M_1, M_2)$ is also a solution to $(G[X], k)$, so suppose that for some $r \in \{1, 2\}$, there is an edge $uv \in M_r$ such that $v \in V(G) \setminus X$. Since $V(M) \subseteq X$ and $V(M)$ is a vertex cover of $G$, we may assume that $u \in V(M)$.

Since $v$ is a neighbor of $u$ in $V(G) \setminus X$, the above marking algorithm added a set of $2k$ neighbors of $u$ in $V(G) \setminus V(M)$ to $X$, denote that set by $X_u$. For an illustration see Fig. 1.

Now, since $X_u \subseteq V(G) \setminus V(M)$, and since $V(M)$ is a vertex cover of $G$, we have that $E(G[X_u]) = \emptyset$. This means in particular that each edge in $M_1 \cup M_2$ has at most one endpoint in $X_u$. Therefore, if all vertices in $X_u$ are the endpoint of some edge in either $M_1$ or in $M_2$, then $|M_1 \cup M_2| \geq 2k$, which implies that at least one of $M_1$ and $M_2$ contains at least $k$ edges that are incident with a vertex from $X_u$; denote that set by $M'$. Since each vertex in $M'$ has an endpoint in $X_u$, and since $V(M)$ is a vertex

cover of $G$, we conclude that the endpoints of the edges in $M'$ other than those in $X_u$ are in $V(M)$, and so $M' \subseteq E(G[X])$.

As above, any 2-partition $(M'_1, M'_2)$ of $M'$ is such that $|M'_1 \triangle M'_2| = |M'| \geq k$, therefore $(M'_1, M'_2)$ is a solution to $(G[X], k)$. Otherwise, there is a vertex $x \in X_u$ that is *not* the endpoint of any edge in $M_1 \cup M_2$. We obtain $M^\star_r$ by removing $uv$ and adding $ux$. Then, $(M^\star_r, M_{3-r})$ is still a solution to $(G, k)$, and it uses one more edge in $G[X]$. Repeatedly applying this argument shows that $(G[X], k)$ is a YES-instance.

The previous claim asserts the correctness of the procedure. Since $|V(M)| < 2k$, and for each vertex in $V(M)$, we added at most $2k$ more vertices to $X$, we have that $|X| = \mathcal{O}(k^2)$. A maximal matching can be found greedily, and it is clear that the marking procedure runs in polynomial time. This yields the result.                    □

## 6 Conclusion

In this work, we initiated the study of algorithmic problems asking for diverse pairs of (maximum/perfect) matchings, where diverse means that their symmetric difference has to be at least some value $k$. These problems are NP-complete on 3-regular graphs, and we showed that on bipartite graphs, they become polynomial-time solvable; while parameterized by $k$, they are FPT, and the problem asking for two diverse (not necessarily maximum) matchings admits a polynomial kernel.

The notion of diverse matchings opens up many natural further research directions. In this work, we considered the complexity of finding *pairs* of diverse matchings. What happens when we ask for a larger number of matchings? In [2, 3], the measure of diversity of a set of solutions is the *sum* over all pairs of their symmetric difference. In this setting, we can obtain an FPT-algorithm parameterized by the number of requested matchings plus the 'diversity target' using the same approach as in our FPT-algorithm for DIVERSE PAIR OF MAXIMUM MATCHINGS. However, we may ask for a set of matchings $\mathcal{M}$ such that *for each pair* $M_1, M_2 \in \mathcal{M}$, $|M_1 \triangle M_2| \geq k$. For the case of *perfect* matchings, Fomin et al. [11] recently gave an FPT-algorithm parameterized by the number $q$ of matchings plus $k$. What about diverse sets of maximum, not necessarily perfect, matchings? Moreover, the algorithm of [11] is randomized, and its run time has a double-exponential dependence on $qk$, while we have a deterministic single-exponential time algorithm for the more restrictive case of two matchings. Can the algorithm from [11] be improved to a single-exponential dependence on $qk$ and can it be derandomized?

While the symmetric difference is a natural measure of diversity of two matchings, one might consider other measures as well. The diversity measure at hand may affect the complexity of the problem, so it would be interesting to see if there is an (easily computable) diversity measure under which DIVERSE PAIR OF MAXIMUM/PERFECT MATCHINGS becomes W[1]-hard.

## Declarations

**Conflict of interests** The authors have no relevant financial or non-financial interests to disclose.

## References

1. Alon, N., Yuster, R., Zwick, U.: Color-coding. J. ACM **42**(4), 844–856 (1995)
2. Baste, J., Fellows, M.R., Jaffke, L., Masařík, T., de Oliveira Oliveira, M., Philip, G., Rosamond, F.A.: Diversity of solutions: an exploration through the lens of fixed-parameter tractability theory. In: IJCAI 2020, pp. 1119–1125 (2020)
3. Baste, J., Jaffke, L., Masařík, T., Philip, G., Rote, G.: FPT algorithms for diverse collections of hitting sets. Algorithms **12**(12), 254 (2019)
4. Brandt, F., Conitzer, V., Endriss, U., Lang, J., Procaccia, A.D.: Handbook of Computational Social Choice. Cambridge University Press, Cambridge (2016)
5. Cygan, M., Fomin, F.V., Kowalik, Ł, Lokshtanov, D., Marx, D., Pilipczuk, M., Saurabh, S.: Parameterized Algorithms. Springer, Berlin (2015)
6. Diestel, R.: Graph Theory. Springer, Berlin (2005)
7. Downey, R.G., Fellows, M.R.: Parameterized Complexity. Springer, New York (1999)
8. Edmonds, J.: Paths, trees, and flowers. Can. J. Math. **17**(3), 449–467 (1965)
9. Feige, U., Ofek, E., Wieder, U.: Approximating maximum edge coloring in multigraphs. In: APPROX 2002, pp. 108–121. Springer, Berlin (2002)
10. Fellows, M.R..: The Diverse X paradigm. Manuscript (2018)
11. Fomin, F.V., Golovach, P.A., Panolan, F., Philip, G., Saurabh, S.: Diverse collections in matroids and graphs. In: Bläser, M., Monmege, B. (eds.) STACS 2021, vol. 187 of LIPIcs, pp. 31:1–31:14. Schloss Dagstuhl (2021)
12. Fomin, F.V., Lokshtanov, D., Saurabh, S., Zehavi, M.: Theory of Parameterized Preprocessing. Cambridge University Press, Kernelization (2018)
13. Gabow, H.N., Tarjan, R.E.: Faster scaling algorithms for general graph-matching problems. J. ACM **38**(4), 815–853 (1991)
14. Hall, P.: On representatives of subsets. J. Lond. Math. Soc. **10**, 26–30 (1935)
15. Hanaka, T., Kobayashi, Y., Kurita, K., Otachi, Y.: Finding diverse trees, paths, and more. In: AAAI 2021, pp 3778–3786. AAAI Press (2021)
16. Holyer, I.: The NP-completeness of edge-coloring. SIAM J. Comput. **10**(4), 718–720 (1981)
17. Hopcroft, J.E., Karp, R.M.: An $n^{5/2}$ algorithm for maximum matchings in bipartite graphs. SIAM J. Comput. **2**, 225–231 (1973)
18. Jakub Kamiński, M., Kowalik, L.: Beyond the Vizing's bound for at most seven colors. SIAM J. Discrete Math. **28**(3), 1334–1362 (2014)
19. Karzanov, A.V.: The problem of finding the maximal flow in a network by the method of preflows. Dokl. Akad. Nauk SSSR **215**, 49–52 (1974)
20. Kőnig, D.: Über Graphen und ihre Anwendung auf Determinantentheorie und Mengenlehre. Math. Ann. **77**(4), 453–465 (1916)

21. Lovász, L.: On determinants, matchings, and random algorithms. In: Fundamentals of Computation Theory. Akademie Verlag, pp. 565–574 (1979)
22. Lovász, L., Plummer, M.D.: Matching Theory. AMS, Providence (2009)
23. Mądry, A.: Navigating central path with electrical flows: From flows to matchings, and back. In: FOCS 2013, pp. 253–262. IEEE Computer Society (2013)
24. Mucha, M., Sankowski, P.: Maximum matchings via Gaussian elimination. In: FOCS 2004, pp. 248–255. IEEE Computer Society (2004)
25. Naor, M., Schulman, L.J., Srinivasan, A.: Splitters and near-optimal derandomization. In: FOCS pp. 182–191 (1995)
26. Olsson, T., Huhtamäki, J., Kärkkäinen, H.: Directions for professional social matching systems. Commun. ACM **63**(2), 60–69 (2020)
27. O'Malley, G.: Algorithmic Aspects of Stable Matching Problems. PhD thesis, University of Glasgow (2007)
28. Rabin, M.O., Vazirani, V.V.: Maximum matchings in general graphs through randomization. J. Algorithms **10**(4), 557–567 (1989)
29. Schrijver, A.: Combinatorial Optimization. Polyhedra and Efficiency, vol. A. Springer, Berlin (2003)
30. Suomela, J.: Complexity of two perfect matchings with minimum shared edges? Theoretical Computer Science Stack Exchange (2010) https://cstheory.stackexchange.com/q/1291 (version: 2010-09-14)
31. Tutte, W.T.: A short proof of the factor theorem for finite graphs. Can. J. Math. **6**, 347–352 (1954)
32. Vazirani, V.V.: A Proof of the MV Matching Algorithm (2014)
33. Vizing, V.G.: On an estimate of the chromatic class of a p-graph. Discrete Anal. **3**, 25–30 (1964)