2013 IEEE 54th Annual Symposium on Foundations of Computer Science

Improved approximation for 3-dimensional matching via bounded pathwidth local search

Marek Cygan*
*Institute of Informatics, University of Warsaw, Poland. Email: cygan@mimuw.edu.pl

***Abstract*—One of the most natural optimization problems is the k-SET PACKING problem, where given a family of sets of size at most k one should select a maximum size subfamily of pairwise disjoint sets. A special case of 3-SET PACKING is the well known 3-DIMENSIONAL MATCHING problem, which is a maximum hypermatching problem in 3-uniform tripartite hypergraphs. Both problems belong to the Karp's list of 21 NP-complete problems. The best known polynomial time approximation ratio for k-SET PACKING is $(k+\epsilon)/2$ and goes back to the work of Hurkens and Schrijver [SIDMA'89], which gives $(1.5+\epsilon)$-approximation for 3-DIMENSIONAL MATCHING. Those results are obtained by a simple local search algorithm, that uses constant size swaps.**

The main result of this paper is a new approach to local search for k-SET PACKING where only a special type of swaps is considered, which we call swaps of bounded pathwidth. We show that for a fixed value of k one can search the space of r-size swaps of constant pathwidth in $c^r \mathrm{poly}(|\mathcal{F}|)$ time. Moreover we present an analysis proving that a local search maximum with respect to $O(\log|\mathcal{F}|)$-size swaps of constant pathwidth yields a polynomial time $(k+1+\epsilon)/3$-approximation algorithm, improving the best known approximation ratio for k-SET PACKING. In particular we improve the approximation ratio for 3-DIMENSIONAL MATCHING from $3/2+\epsilon$ to $4/3+\epsilon$.

***Keywords*-approximation, 3-dimensional matching, k-set packing, local search, fixed parameter tractability**

I. INTRODUCTION

In the SET PACKING problem, also known as HYPERGRAPH MATCHING, we are given a family $\mathcal{F} \subseteq 2^U$ of subsets of U, and the goal is to find a maximum size subfamily of $\mathcal{F}$ of pairwise disjoint sets. SET PACKING is a fundamental problem in combinatorial optimization with various applications. A simple reduction from INDEPENDENT SET (where $|\mathcal{F}| = |V|$) combined with the hardness result of Håstad [13] makes the SET PACKING problem hard to approximate. When each set of SET PACKING is of size at most k the problem is denoted as k-SET PACKING.

k-SET PACKING
Input: A family $\mathcal{F} \subseteq 2^U$ of sets of size at most k.
Goal: Find a maximum size subfamily of $\mathcal{F}$ of pairwise disjoint sets.

k-SET PACKING is a generalization of INDEPENDENT SET in bounded degree graphs, as well as k-DIMENSIONAL MATCHING and is related to plethora of other problems (see [6] for a list of connections between k-SET PACKING and other combinatorial optimization problems). In 3-DIMENSIONAL MATCHING the universe U is partitioned into $U = X \uplus Y \uplus Z$ and $\mathcal{F}$ is a subset of $X \times Y \times Z$.

Both 3-DIMENSIONAL MATCHING and SET PACKING are well studied problems, belonging to Karp's list of 21 NP-hard problems [18]. A simple greedy algorithm returning any inclusionwise maximal subfamily of disjoint subsets of $\mathcal{F}$ gives a k-approximation for k-SET PACKING. One can consider a local search routine, where as long as it is possible we remove one set from our current feasible solution and add two new sets. We say that such an algorithm uses size 2 swaps, as two new sets are involved. It is known that a local search maximum with respect to size 2 swaps is a $(k+1)/2$-approximation for k-SET PACKING. If, instead of using swaps of size 2 we use swaps of size r for bigger values of r, then the approximation ratio approaches $k/2$, and that is exactly the $(k/2+\epsilon)$-approximation algorithm by Hurkens and Schrijver [16].

Despite significant interest (see Section I-B) for over 20 years no improved polynomial time approximation algorithm was obtained for k-SET PACKING, even for the special case of 3-DIMENSIONAL MATCHING. Meanwhile Halldórsson [12] has shown that a local search maximum with respect to $\mathcal{O}(\log|\mathcal{F}|)$ size swaps gives a $(k+2)/3$-approximation, which was recently improved to $(k+1+\epsilon)/3$ [8]. Nevertheless enumerating all $\mathcal{O}(\log|\mathcal{F}|)$ size swaps takes quasipolynomial time.

A. Our results and techniques

Based on the work of Halldórsson [12] a natural path to transforming a quasipolynomial time approximation into a polynomial time approximation would be by designing a $c^r \mathrm{poly}(|\mathcal{F}|)$ time algorithm, where c is a constant. This is exactly the framework of parameterized complexity[1], where the swap size is a natural parameter. Unfortunately, we show that this is most likely impossible, i.e. there is no

The author is partially supported by Foundation for Polish Science grant HOMING PLUS/2012-6/2.

[1]For further information about parameterized complexity we defer the reader to monographs [9], [11], [23].

such algorithm with $f(r)\text{poly}(|\mathcal{F}|)$ running time, unless W[1]=FPT, where f is some computable function, even for $k = 3$. We would like to note that W[1]≠FPT is a widely believed assumption, in particular if W[1]=FPT, then the Exponential Time Hypothesis of [17] fails.

Theorem I.1. *Unless $FPT = W[1]$, there is no $f(r)\text{poly}(|\mathcal{F}|)$ time algorithm, that given a family $\mathcal{F} \subseteq 2^U$ of sets of size 3 and its disjoint subfamily $\mathcal{F}_0 \subseteq \mathcal{F}$ either finds a bigger disjoint family $\mathcal{F}_1 \subseteq \mathcal{F}$ or verifies that there is no disjoint family $\mathcal{F}_1 \subseteq \mathcal{F}$ such that $|\mathcal{F}_0 \setminus \mathcal{F}_1| + |\mathcal{F}_1 \setminus \mathcal{F}_0| \leq r$,*

Therefore trying to find a $c^r\text{poly}(|\mathcal{F}|)$ time algorithm which searches the whole r-size swaps space is not the proper path. For this reason we introduce a notion of swaps (also called improving sets) of bounded pathwidth (see Section III-A). Intuitively a size r swap is of bounded pathwidth, if the bipartite graph where vertices represent sets that are added and removed, and edges correspond to non-empty intersections, is of constant pathwidth. Using the color-coding technique of Alon et al. [1] we show that one can search the space of swaps of size at most r of bounded pathwidth in $c^r\text{poly}(|\mathcal{F}|)$ time, for a constant c. As the currently best known analysis of local search maximum with respect to logarithmic size swaps of [8] relies on swaps of unbounded pathwidth, we need to develop a different proof strategy, and the core part of it is contained in Lemma III.8. The algorithm and its analysis complete the main result of this paper, that is a polynomial time $(k + 1 + \epsilon)/3$-approximation algorithm, for any fixed k and ϵ.

Theorem I.2. *For any $\epsilon > 0$ and any integer $k \geq 3$ there is a polynomial time $(k + 1 + \epsilon)/3$-approximation algorithm for k-SET PACKING.*

We believe that the usage of parameterized tools such as color-coding, pathwidth and W[1]-hardness in the setting of this work is interesting on its own, as to the best of our knowledge such tools have not been previously used in local search based approximation algorithms.

B. Related work

Even though there was no improvement in terms of polynomial time approximation of k-SET PACKING (and 3-DIMENSIONAL MATCHING) since the work of Hurkens and Schrijver [16], both problems are well studied.

One can also consider weighted variant of k-SET PACKING, where we want to select a maximum weight disjoint subfamily of $\mathcal{F}$. Arkin and Hassin [2] gave a $(k - 1 + \epsilon)$-approximation algorithm, Chandra and Halldórsson [7] improved it to $(2k + 2 + \epsilon)/3$-approximation, later improved by Berman [4] to $(k + 1 + \epsilon)/2$-approximation. All the mentioned results are based on local search.

Also for the standard (unweighted) k-SET PACKING problem Chan and Lau [6] presented a strengthened LP relaxation, which has integrality gap $(k + 1)/2$.

On the other hand, Hazan et al [14] have shown that k-SET PACKING is hard to approximate within a factor of $\mathcal{O}(k/\log k)$. Concerning small values of k, Berman and Karpinski [5] obtained a $98/97 - \epsilon$ hardness for 3-DIMENSIONAL MATCHING, while Hazan et al. [15] obtained $54/53 - \epsilon$, $30/29 - \epsilon$, and $23/22 - \epsilon$ hardness for 4, 5 and 6-DIMENSIONAL MATCHING respectively (note that a hardness result for k-DIMENSIONAL MATCHING directly gives a hardness for k-SET PACKING).

Recently Sviridenko and Ward [25] have independently obtained a $(k + 2)/3$-approximation algorithm for k-SET PACKING. They observed that the analysis of Halldórsson [12] can be combined with a clever application of the color coding technique. However to the best of our understanding it is not possible to obtain $(k + 1 + \epsilon)/3$-approximation for k-SET PACKING using the tools of [25], and in particular Sviridenko and Ward do not improve the approximation ratio for 3-DIMENSIONAL MATCHING. The main difference between this article and [25] is in handling sets of the optimum solution, that intersect exactly one set in a local maximum.

C. Organisation

We start with preliminaries in Section II, where we recall standard graph notation together with the definition of pathwidth and path decompositions.

Section III contains the main result of this paper, that is the $(k + 1 + \epsilon)/3$-approximation for k-SET PACKING. First, we introduce the notion of improving set of bounded pathwidth in Section III-A. In Section III-B we apply the color coding technique to obtain a polynomial time algorithm searching an improving set of logarithmic size of bounded pathwidth. In Section III-C we analyse a local search maximum with respect to bounded pathwidth improving sets of logarithmic size. The heart of our analysis is contained in an abstract combinatorial Lemma III.8 which is later applied in the proof of Lemma III.11.

The proof of Theorem I.1 is given in Section IV. Finally, in Section V we conclude with potential future research directions.

II. PRELIMINARIES

We use standard graph notation. For an undirected graph G by $V(G)$ and $E(G)$ we denote the set of its vertices and edges respectively. By $N_G(v) = \{u : uv \in E(G)\}$ we denote the open neighborhood of a vertex v, while the closed neighborhood is defined as $N_G[v] = N_G(v) \cup \{v\}$. Similarly, for a subset of vertices X we have $N_G[X] = \bigcup_{v \in X} N_G[v]$ and $N_G(X) = N_G[X] \setminus X$.

By a disjoint family of sets we denote a family, where each pair of sets is pairwise disjoint. For a positive integer r by $[r]$ we denote the set $\{1, \ldots, r\}$.

Pathwidth and path decompositions: A *path decomposition* of a graph $G = (V, E)$ is a sequence $\mathbb{P} = (B_i)_{i=1}^q$, where each set B_i is a subset of vertices $B_i \subseteq V$ (called a *bag*) such that $\bigcup_{1 \le i \le q} B_i = V$ and the following properties hold:

(i) For each edge $uv \in E(G)$ there is a bag B_i in $\mathbb{P}$ such that $u, v \in B_i$.
(ii) If $v \in B_i \cap B_j$ then $v \in B_\ell$ for each $\min(i, j) \le \ell \le \max(i, j)$.

The *width* of $\mathbb{P}$ is the size of the largest bag minus one, and the *pathwidth* of a graph G is the minimum width over all possible path decompositions of G. Since our focus here is on path decompositions we only mention in passing that the related notion of *treewidth* can be defined similarly, except for letting the bags of the decomposition form a tree instead of a path.

In order to make the description easier to follow, it is common to use path decompositions that adhere to some simplifying properties. The most commonly used notion is that of a nice path decompositions, introduced by Kloks [19]; the main idea is that adjacent nodes can be assumed to have bags differing by at most one vertex.

Definition II.1 (nice path decomposition). A *nice path decomposition* is a path decomposition $\mathbb{P} = (B_i)_{i=1}^q$, where each bag is of one of the following types:

- **First (leftmost) bag**: the bag B_1 is empty, $B_1 = \emptyset$.
- **Introduce bag**: an internal bag B_i of $\mathbb{P}$ with predecessor B_{i-1} such that $B_i = B_{i-1} \cup \{v\}$ for some $v \notin B_{i-1}$. This bag is said to *introduce* v.
- **Forget bag**: an internal bag B_i of $\mathbb{P}$ with predecessor B_{i-1} for which $B_i = B_{i-1} \setminus \{v\}$ for some $v \in B_{i-1}$. This bag is said to *forget* v.
- **Last (rightmost) bag**: the bag associated with the largest index, i.e. q, is empty, $B_q = \emptyset$.

It is easy to verify that any given path decomposition can be transformed in polynomial time into a nice path decomposition without increasing its width.

III. Local search algorithm

In this section we present the main result of the paper, i.e. the $(k + 1 + \epsilon)/3$-approximation algorithm for k-Set Packing, proving Theorem I.2. We start with introducing the notion of improving set of bounded pathwidth in Section III-A. Next, in Section III-B we apply the color coding technique to obtain a polynomial time algorithm searching an improving set of logarithmic size of bounded pathwidth. In Section III-C we analyse a local search maximum with respect to bounded pathwidth improving sets of logarithmic size. The heart of our analysis is contained in an abstract combinatorial Lemma III.8 which is later applied in the proof of Lemma III.11.

A. Bounded pathwidth improving set

Let us assume that an instance $\mathcal{F} \subseteq 2^U$ of k-Set Packing is given. Moreover by $\mathcal{F}_0 \subseteq \mathcal{F}$ we denote some disjoint subfamily of $\mathcal{F}$, which we can think of as a current feasible solution of a local search algorithm. In what follows we define a *conflict graph*, which is a bipartite undirected graph with two independent sets of vertices being $\mathcal{F}_0$ and $\mathcal{F} \setminus \mathcal{F}_0$, where an edge reflects non-empty intersection.

Definition III.1 (conflict graph). For a disjoint family $\mathcal{F}_0 \subseteq \mathcal{F}$ by a *conflict graph* $G_{\mathcal{F}_0}$ we denote an undirected bipartite graph with vertex set $\mathcal{F}$ and edge set $\{S_1 S_2 : S_1 \in \mathcal{F}_0, S_2 \in (\mathcal{F} \setminus \mathcal{F}_0), S_1 \cap S_2 \neq \emptyset\}$.

Next, we define an *improving set* $X \subseteq \mathcal{F} \setminus \mathcal{F}_0$, which can be used to increase the cardinality of $\mathcal{F}_0$, and then we introduce a notion of an *improving set of bounded pathwidth*, which will be crucial in both the algorithm and the analysis of its approximation ratio.

Definition III.2 (improving set). For a disjoint family $\mathcal{F}_0 \subseteq \mathcal{F}$ a set $X \subseteq \mathcal{F} \setminus \mathcal{F}_0$ is called an *improving set*, if the following conditions hold:

- all sets of X are pairwise disjoint,
- $|N_{G_{\mathcal{F}_0}}(X)| < |X|$, i.e. the number of sets of $\mathcal{F}_0$ having a common element with at least one set of X is strictly smaller than $|X|$.

Observe, that if we have an improving set X, then $(\mathcal{F}_0 \setminus N_{G_{\mathcal{F}_0}}(X)) \cup X$ is a disjoint subfamily of $\mathcal{F}$ of size greater than $|\mathcal{F}_0|$, hence the name improving set.

Definition III.3 (improving set of bounded pathwidth). An improving set X with respect to $\mathcal{F}_0 \subseteq \mathcal{F}$ has *pathwidth at most* $\mathtt{pw}$, if the subgraph of the conflict graph $G_{\mathcal{F}_0}$ induced by $N_{G_{\mathcal{F}_0}}[X]$ has pathwidth at most $\mathtt{pw}$.

B. Algorithm

To find an improving set of bounded pathwidth we use the color coding technique of Alon et al. [1], which is by now a well-established tool in parameterized complexity used for finding a set consisting of disjoint objects. We use two random colorings $c_{\mathcal{F}_0} : \mathcal{F}_0 \to [r-1]$, $c_U : U \to [rk]$, where c_U ensures that the sets of X are disjoint, while $c_{\mathcal{F}_0}$ is used not to consider the same set of $\mathcal{F}_0$ twice.

Lemma III.4. *There is an algorithm, that given a disjoint family $\mathcal{F}_0 \subseteq \mathcal{F}$, and two coloring functions $c_{\mathcal{F}_0} : \mathcal{F}_0 \to [r-1]$, $c_U : U \to [rk]$ in $2^{\mathcal{O}(rk)}|\mathcal{F}|^{\mathcal{O}(\mathtt{pw})}$ time determines, whether there exists an improving set $X \subseteq \mathcal{F} \setminus \mathcal{F}_0$ of size at most r of pathwidth at most $\mathtt{pw}$, such that $c_{\mathcal{F}_0}$ is injective on $N_{G_{\mathcal{F}_0}}(X)$ and c_U is injective on $\bigcup_{S \in X} S$.*

Proof: For the sake of notation by adding dummy distinct elements we ensure that each set of $\mathcal{F}$ has size exactly k. Define an auxiliary directed graph $D =$

$(V_D, A_{forget} \cup A_{introduce})$, where each vertex is characterized by a subset of set colors $[r-1]$, a subset of element colors $[rk]$, and a subset of $\mathcal{F}$ of size at most $\mathtt{pw}+1$, i.e.

$$V_H = \{v(C_{\mathcal{F}_0}, C_U, B) : C_{\mathcal{F}_0} \subseteq [r-1], C_U \subseteq [rk], \\ B \subseteq \mathcal{F}, |B| \le \mathtt{pw}+1\}\,.$$

Note that this graph has $\mathcal{O}(2^{r(k+1)}|\mathcal{F}|^{\mathtt{pw}+1})$ vertices.

Since there will be no possibility of confusion, to make the proof easier to follow by $N[X]$ for $X \subseteq \mathcal{F}$ we denote $N_{G_{\mathcal{F}_0}}[X]$, i.e. we omit the subscript $G_{\mathcal{F}_0}$. The idea behind the construction is that each vertex of V_H describes a potential prefix of a sequence of bags in a path decomposition of $N[X]$ for some $X \subseteq \mathcal{F} \setminus \mathcal{F}_0$. The set B encodes the set of vertices of $N[X]$ in the current bag and ensures the bounded pathwidth property. Instead of storing all the sets of X that have already appeared in the sequence of bags, we store only the colors of the elements of $\bigcup_{S\in X} S$ (encoded by C_U), as it is enough to maintain the disjointness of sets of X and keep track of the cardinality of X - due to the assumption that each set of is size exactly k. Similarly instead of storing all the sets of $N[X]$ that were already introduced, we only store their colors (encoded by $C_{\mathcal{F}_0}$).

To the set $A_{introduce}$ we add the following arcs. For $s = v(C_{\mathcal{F}_0}, C_U, B) \in V_D$, $S \in \mathcal{F}$ such that $|B| \le \mathtt{pw}$:

- if $S \in \mathcal{F} \setminus \mathcal{F}_0$, $c_U(S) \cap C_U = \emptyset$, $c_{\mathcal{F}_0}$ is injective on $N(S)$ and $c_{\mathcal{F}_0}(N(S) \setminus B) \cap C_{\mathcal{F}_0} = \emptyset$, then add to $A_{introduce}$ an arc $(s, v(C_{\mathcal{F}_0}, C_U \cup c_U(S), B \cup \{S\}))$
- if $S \in \mathcal{F}_0$, $c_{\mathcal{F}_0}(S) \notin C_{\mathcal{F}_0}$ and for each $S' \in B \setminus \mathcal{F}_0$ either $S \in N(S')$, or $c_{\mathcal{F}_0}(S) \notin c_{\mathcal{F}_0}(N(S'))$, then add to $A_{introduce}$ an arc $(s, v(C_{\mathcal{F}_0} \cup \{c_{\mathcal{F}_0}(S)\}, C_U, B \cup \{S\}))$

To the set A_{forget} we add the following arcs. For $s = v(C_{\mathcal{F}_0}, C_U, B) \in V_D$, $S \in B$ add to A_{forget} an arc $(s, v(C_{\mathcal{F}_0}, C_U, B \setminus \{S\}))$ if one of the following conditions holds:

- $S \in \mathcal{F}_0$,
- $S \notin \mathcal{F}_0$ and $c_{\mathcal{F}_0}(N(S)) \subseteq C_{\mathcal{F}_0}$.

Claim III.5. *There exists a path in the graph D from the vertex $v(\emptyset,\emptyset,\emptyset)$ to a vertex $v(C_{\mathcal{F}_0}, C_U, \emptyset) \in V_D$ for $|C_{\mathcal{F}_0}| < |C_U|/k$ iff there exists an improving set X of size at most r of pathwidth at most* pw*, such that $c_{\mathcal{F}_0}$ is injective on $N(X)$ and c_U is injective on $\bigcup_{S\in X} S$.*

Proof: Assume that there is a path $s_1, \ldots, s_q$ in H, where $s_i = (C^i_{\mathcal{F}_0}, C^i_U, B_i)$, $s_1 = (\emptyset,\emptyset,\emptyset)$, $|C^q_{\mathcal{F}_0}| < |C^q_U|/k$ and $B_q = \emptyset$. Let $X = \bigcup_{1\le i\le q} B_i \setminus \mathcal{F}_0$. By construction of D, we have $|X| = |C^q_U|/k \le r$. By the definition of $A_{introduce}$ and A_{forget} since $B_q = \emptyset$, at the time a vertex $v \in X$ appears for the first time in some B_i we ensure that all its neighbors in $G_{\mathcal{F}_0}$ are either in B_i or are colored by $c_{\mathcal{F}_0}$ with colors not yet in $C^i_{\mathcal{F}_0}$. Moreover at the time $v \in X$ is forgotten, i.e. removed from some B_i, we ensure that all of its neighbors in $G_{\mathcal{F}_0}$ have been already added to bags. Therefore $N[X] \subseteq \bigcup_{1\le i\le q} B_i$ and for each edge e of $G[N[X]]$ the endpoints of e appear in some bag B_i. Since no set of $\mathcal{F}_0$ is added twice, due to the coloring $c_{\mathcal{F}_0}$, no set of $\mathcal{F} \setminus \mathcal{F}_0$ is added twice, due to the coloring c_U, $(B_i \cap N[X])_{i=1}^q$ is a path decomposition of $N[X]$ of width at most pw. Finally $|N(X)| \le |C^q_{\mathcal{F}_0}| < |C^q_U|/k = |X|$. Hence X is an improving set of size at most r and of pathwidth at most pw.

In the other direction, let X be an improving set of size at most r such that $c_{\mathcal{F}_0}$ is injective on $N(X)$, c_U is injective on $\bigcup_{S\in X} S$, and let $\mathbb{P} = (B_i)_{i=1}^q$ be a nice path decomposition of $N[X]$ of width at most pw. For $1 \le i \le q$ define $s_i \in V_D$ as $s_i = v(c_{\mathcal{F}_0}(B'_i \cap \mathcal{F}_0), c_U(\bigcup_{S\in B'_i\setminus\mathcal{F}_0} S), B_i)$, where $B'_i = \bigcup_{1\le j\le i} B_i$. Observe that $s_1 = (\emptyset,\emptyset,\emptyset)$, $s_q = (C_{\mathcal{F}_0}, C_U, \emptyset)$ for $|C_{\mathcal{F}_0}| = |N(X)| < |X| = |C_U|/k$ and moreover if B_{i+1} is an introduce bag, then $(s_i, s_{i+1}) \in A_{introduce}$ while if B_{i+1} is a forget bag, then $(s_i, s_{i+1}) \in A_{forget}$. Consequently there is a path from s_1 to s_q in the graph D. ■

By the above claim it is enough to run a standard graph search algorithm, to check whether there exists a path from the vertex $v(\emptyset,\emptyset,\emptyset)$ to $v(C_{\mathcal{F}_0}, C_U, \emptyset)$ where $|C_{\mathcal{F}_0}| < |C_U|/k$, which finishes the proof of Lemma III.4. ■

Theorem III.6. *There is an algorithm, that given a disjoint family $\mathcal{F}_0 \subseteq \mathcal{F}$, in $2^{\mathcal{O}(rk)}|\mathcal{F}|^{\mathcal{O}(\mathtt{pw})}$ time determines, whether there exists an improving set $X \subseteq \mathcal{F} \setminus \mathcal{F}_0$ of size at most r of pathwidth at most* pw.

Proof: Observe, that if we take $c_{\mathcal{F}_0} : \mathcal{F}_0 \to [r-1]$ where the color of each set is chosen uniformly and independently at random, then for an improving set X of size at most r the function $c_{\mathcal{F}_0}$ is injective on $N_{G_{\mathcal{F}_0}}(X)$ with probability at least

$$(r-1)!/(r-1)^{r-1} \ge ((r-1)/e)^{r-1}/(r-1)^{r-1} = e^{-(r-1)}\,.$$

Similarly, if we assign a color of $[rk]$ to each element of U, then with probability at least e^{-rk} the function $c_U : U \to [rk]$ is injective on $\bigcup_{S\in X} S$. Therefore invoking Lemma III.4 with random colorings $c_{\mathcal{F}_0}, c_U$ at least e^{r-1+rk} times would yield a constant error probability.

To obtain a deterministic algorithm we can use the, by now standard, technique of splitters. An (n,a,b)-splitter is a family $\mathcal{H}$ of functions $[n] \to [b]$, such that for any $W \subseteq [n]$ of size at most a there exists $f \in \mathcal{H}$ that is injective on W. What we need is a small family of (n,a,a)-splitters.

Theorem III.7 ([22])**.** *There exists an (n,a,a)-splitter of size $e^a a^{\mathcal{O}(\log a)} \log n$ that can be constructed in $\mathcal{O}(e^a a^{\mathcal{O}(\log a)} n \log n)$ time.*

Therefore instead of using random colorings $c_{\mathcal{F}_0}$, c_U we can use Theorem III.7 to construct $(|\mathcal{F}_0|, r-1, r-1)$ and $(|U|, rk, rk)$ splitters, leading to a deterministic algorithm, which finishes the proof of Theorem III.6. ■

C. Analysis

In this subsection we analyze a local search maximum, with respect to logarithmic size improving sets of constant pathwidth. It is well known that an undirected graph of average degree at least $2+\epsilon$ contains a cycle of length at most $c_\epsilon \log n$, where the constant c_ϵ depends on ϵ. This observation was the base for the quasipolynomial time algorithms of [8], [12]. Here, however we need to generalize this result extensively, as the analysis of [8] relies on improving sets of unbounded pathwidth.

Throughout this subsection we operate on multigraphs, as there might be several parallel edges in a graph, however there will be no self-loops.

Lemma III.8. *Let $H = (V, E)$ be an n-vertex undirected multigraph of minimum degree at least 3. Assume that each edge $e \in E$ is associated with a subset of an alphabet $w_e \subseteq \Sigma$ of size at most γ, where $\gamma \geq 1$. If each element $c \in \Sigma$ appears in at most γ sets w_e, i.e. $\forall_{c\in\Sigma}\ |\{e : e \in E, c \in w_e\}| \leq \gamma$, then there exists a tree $T_0 = (V_0, E_0)$, which is a subgraph of H, and a vertex $r_0 \in V_0$, such that:*

- $|V_0| \leq 4(\log_{3/2} n + 2)$*;*
- *there exist two edges $e_1, e_2 \in E \setminus E_0, e_1 \neq e_2$ which have both endpoints in V_0;*
- *T_0 is a tree with at most 4 leaves;*
- *for each pair of edges $e_1, e_2 \in E_0$ such that $w_{e_1} \cap w_{e_2} \neq \emptyset$ we have $|\mathrm{dist}_{T_0}(r_0, e_1) - \mathrm{dist}_{T_0}(r_0, e_2)| \leq \beta$, where $\beta = \lceil \log_{3/2}(12\gamma^2) \rceil$, and $\mathrm{dist}_{T_0}(r_0, uv) = \min(\mathrm{dist}_{T_0}(r_0, u), \mathrm{dist}_{T_0}(r_0, v))$.*

Proof: First we deal with some corner cases.

(i) If in H there are three parallel edges e_a, e_b, e_c between vertices u and v, then as T_0 we take $(\{u, v\}, \{e_a\})$ and set $e_1 = e_b$, $e_2 = e_c$.

(ii) If in H there are three vertices u, v, w, two parallel edges e_a, e_b between u and v as well as two parallel edges e_c, e_d between v and w, than as T_0 we take $(\{u, v, w\}, \{e_a, e_c\})$ and set $e_1 = e_b$, $e_2 = e_d$.

(iii) In the last corner case let us assume that for each vertex v of H there are some two parallel edges $e_a, e_b \in E(H)$ incident to v. Let $uv \in E(H)$ be any edge of H for which there is no parallel edge in H - such an edge exists, as otherwise (i) or (ii) would hold. Let u' be a vertex such that in H there are two parallel edges e_a, e_b between u and u', similarly let v' be a vertex such that in H there are two parallel edges e_c, e_d between v and v'. Observe that $u' \neq v'$ as otherwise case (ii) would hold. In that case $T_0 = (\{u, u', v, v'\}, \{e_a, uv, e_c\})$, $e_1 = e_b$ and $e_2 = e_d$.

Assuming that none of (i), (ii), (iii) holds, there is a vertex r in H, which is adjacent to at least three distinct vertices v_1, v_2, v_3.

We are going to construct a sequence of logarithmic number of trees $T_1, T_2, \ldots$ rooted at r, which are subgraphs of H satisfying two invariants:

- **(exponential growth)** for any $1 \leq j \leq i$ the number of vertices in T_i at distance exactly j from r is exactly $\lfloor 2(3/2)^j \rfloor$, and there are no vertices at distance more than i,
- **(Σ-nearness)** for any two edges e_1, e_2 of T_i if $w_{e_1} \cap w_{e_2} \neq \emptyset$, then $|\mathrm{dist}_{T_i}(r, e_1) - \mathrm{dist}_{T_i}(r, e_2)| \leq \beta$.

We will show, that having constructed a tree T_i for some $i \geq 1$ we can either construct a tree T_{i+1} satisfying the two invariants, or find a tree T_0 with edges e_1, e_2 required by the claim of the lemma.

Let $T_1 = (\{r, v_1, v_2, v_3\}, \{rv_1, rv_2, rv_3\})$ and note that it satisfies the two invariants. Assume, that T_i (for some $i \geq 1$) was the most recently constructed tree, and we want to construct T_{i+1}. Let V' be the vertices of T_i at distance exactly i from the root r. By the exponential growth invariant we have $|V'| = \lfloor 2(3/2)^i \rfloor$. Let $E' \subseteq E$ be the set of edges of H incident to V', but not contained in $E(T_i)$. As each vertex in H is of degree at least three, we have

$$|E'| \geq 2|V'| \geq 2\lfloor 2(3/2)^j \rfloor . \tag{1}$$

Let

$$E_{banned} = \{e \in E' : \exists_{e' \in E(T_{i-\beta})} w_e \cap w_{e'} \neq \emptyset\} ,$$

i.e. the set of edges having a non-empty intersection with $w_{e'}$, where e' is not contained in the last β levels of T_i. Observe that for $i \leq \beta$ the set E_{banned} is empty. When extending the tree T_i to maintain the Σ-nearness invariant, we use only edges of $E' \setminus E_{banned}$.

Let $V'' = \bigcup_{uv \in E' \setminus E_{banned}} \{u, v\} \setminus V(T_i)$. We consider two cases: either $|V''| \geq \lfloor 2(3/2)^{i+1} \rfloor$ or not. In the former case we will show that one can construct a tree T_{i+1} satisfying both exponential growth and Σ-nearness invariants. In the latter case we will show that the required tree T_0 and edges e_1, e_2 exist.

If $|V''| \geq \lfloor 2(3/2)^{i+1} \rfloor$, then we select exactly $\lfloor 2(3/2)^{i+1} \rfloor$ vertices out of V'' and extend the tree T_i to T_{i+1} by adding one more layer of vertices (at distance $i + 1$ from r), connected to vertices of V' by edges of $E' \setminus E_{banned}$. Clearly the exponential growth invariant is satisfied for T_{i+1}. Furthermore, since T_i satisfied the Σ-nearness invariant and by the definition of E_{banned} the tree T_{i+1} also satisfies the Σ-nearness invariant.

In the remaining part of the proof we assume

$$|V''| < \lfloor 2(3/2)^{i+1} \rfloor \tag{2}$$

and show the required tree T_0 with edges e_1, e_2. If at least two edges of E' have both endpoints in $V(T_i)$, denote those edges $uv, u'v' \in E'$, then as T_0 we take the subtree of T_i induced by vertices on the paths between $\{u, v, u', v'\}$ and their least common ancestor r_0 and set $e_1 = uv$, $e_2 = u'v'$ (see Figure 1). Therefore let $E'' \subseteq E'$ be the subset of edges

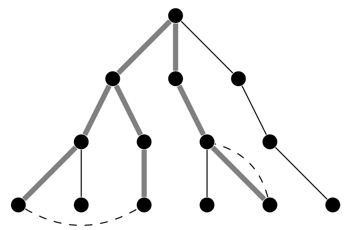

Figure 1: Edges of the tree T_0 are gray, while edges e_1 and e_2 are dashed.

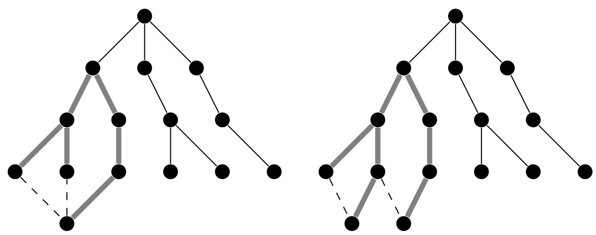

Figure 2: Creating the tree T_0 assuming $|E'''| \le |E''| - 2$. Notation as in Figure 1.

having exactly one endpoint in $V(T_i)$ (that is in V'). By (1) we infer that

$$|E''| \geq |E'| - 1 \geq 2|V'| - 1\,. \tag{3}$$

Let E''' be a maximum size subset of E'', such that no two edges of E''' have a common endpoint in $V \setminus V(T_i)$. Observe that if $|E'''| \le |E''| - 2$, then either:

- there exists three edges $e_a, e_b, e_c \in E''$ having a common endpoint in $V \setminus V(T_i)$, or
- there exist four edges $e_a, e_b, e_c, e_d \in E''$, such that e_a, e_b have a common endpoint in $V \setminus V(T_i)$ and e_c, e_d have a common endpoint in $V \setminus V(T_i)$.

In both cases we can extend the tree T_i by one or two edges to construct T_0 and set $e_1 = e_b$, $e_2 = e_c$ (see Figure 2).

Consequently we have $|E'''| \ge |E''| - 1$, which together with (3) gives

$$|E'''| \geq 2|V'| - 2\,. \tag{4}$$

In the last part of the proof we use the following claim.

Claim III.9.

$$|E''' \setminus E_{banned}| \geq \lfloor 2(3/2)^{i+1} \rfloor$$

Proof: Recall that if $i \le \beta$, the set E_{banned} is empty. Hence by inequality (4) in that case $|E''' \setminus E_{banned}| = |E'''| \ge 2\lfloor 2(3/2)^i \rfloor - 2$. A direct check shows that for each $1 \le i \le 4$ we have $2\lfloor 2(3/2)^i \rfloor - 2 \ge \lfloor 2(3/2)^{i+1} \rfloor$, which proves the claim in the case $i \le 4$.

When $4 < i \le \beta$ we have

$$\begin{aligned}|E''' \setminus E_{banned}| &\geq 2\lfloor 2(3/2)^i \rfloor - 2 \\ &\geq 2(2(3/2)^i - 1) - 2 \geq 2(3/2)^{i+1}\,.\end{aligned}$$

Finally for $i > \beta$ we upper bound the size of E_{banned}

$$\begin{aligned}|E_{banned}| &\leq \sum_{j=1}^{i-\beta} \gamma^2 2(3/2)^j \leq 3\gamma^2 \sum_{j=0}^{i-\beta-1} (3/2)^j \\ &\leq 6\gamma^2((3/2)^{i-\beta} - 1) \leq \frac{(3/2)^i}{2} - 6\,.\end{aligned}$$

The first inequality follows from the assumption, that each set w_e is of size at most γ and each element of Σ is contained in at most γ sets w_e, hence each of T_i contributes at most γ^2 edges to E_{banned}. The last inequality follows from the choice of β and the assumption $\gamma \ge 1$. Therefore

$$\begin{aligned}|E''' \setminus E_{banned}| &\geq |E'''| - |E_{banned}| \\ &\geq 2\lfloor 2(3/2)^i \rfloor - 2 - (\frac{(3/2)^i}{2} - 6) \\ &\geq 2(3/2)^{i+1}\,.\end{aligned}$$

■

Observe that by the definition of E''' we have $|V''| \ge |E''' \setminus E_{banned}|$, but then Claim III.9 contradicts inequality (2). ■

Corollary III.10. *Let $H = (V, E)$ be an undirected multigraph with n vertices and of minimum degree at least 3. Assume that each edge $e \in V$ is associated with a subset of an alphabet $w_e \subseteq \Sigma$ of size at most γ, for some $\gamma \ge 1$, such that each element of Σ appears in at most γ sets w_e. There exists a subgraph $H_0 = (V_0, E_0)$ of H, and a path decomposition $(B_i)_{i=1}^q$ of H_0 of width at most $4(\beta + 3)$, where $\beta = \lceil \log_{3/2}(12\gamma^2) \rceil$ and*

(a) $|E_0| = |V_0| + 1$,
(b) $|V_0| \le 4(\log_{3/2} n + 2)$,
(c) *for each pair of edges $e_1, e_2 \in E_0$, such that $w_{e_1} \cap w_{e_2} \neq \emptyset$ there exists a bag B_i for some $1 \le i \le q$, such that all of the endpoints of both e_1 and e_2 are contained in B_i,*
(d) *for each edge $uv \in E_0$ the set of indices $\{i : u, v \in B_i\}$ is an interval.*

Proof: First, we use Lemma III.8 to obtain $T_0 = (V_0, E_0)$, $r_0 \in V_0$, such that $|V_0| \le 4(\log_{3/2} n + 2)$, where for each pair of edges $e_1, e_2 \in E_0$ such that $w_{e_1} \cap w_{e_2} \neq \emptyset$ we have $|\mathrm{dist}_{T_0}(r_0, e_1) - \mathrm{dist}_{T_0}(r_0, e_2)| \le \beta$. Let $e_1, e_2 \in E \setminus E_0$ be two edges with both endpoints in V_0. Define $H_0 = (V_0, E_0 \cup \{e_1, e_2\})$, clearly H_0 is a subgraph of H and the number of edges is the number of vertices plus one. Therefore properties (a) and (b) are satisfied and it remains to show that there exists a path decomposition of H_0 of width at most $4(\beta + 3)$, satisfying (c) and (d).

Let D_i be the set of vertices of V_0 at distance exactly i from r_0 in T_0. Consider a sequence $(B_i)_{i=0}^q$, where $q = 4(\log_{3/2} n + 2)$, and $B_i = \bigcup_{\max(0, i-\beta-1) \le j \le i} D_i \cup e_1 \cup e_2$. It is straightforward to check that this is in fact a path decomposition of H_0, and since T_0 has at most 4 leaves, this implies that the size of each D_i is upper bounded by 4, and hence the path decomposition is of width at most $4(\beta+3)$.

Observe that property (c) required by the corollary follows from the last property of Lemma III.8, because all of the endpoints of edges $e_1, e_2 \in E_0$, such that $w_{e_1} \cap w_{e_2} \neq \emptyset$, are contained in $B_{\max(\mathrm{dist}_{T_0}(r_0,e_1)+1, \mathrm{dist}_{T_0}(r_0,e_2)+1)}$. To prove property (d) let $e = uv$ be an arbitrary edge of E_0 and define $I_u = \{i : u \in B_i\}$ and $I_v = \{i : v \in B_i\}$. As we already know that $(B_i)_{i=0}^{q}$ is a path decomposition it follows that both sets I_u, I_v form an interval, hence $I_u \cap I_v$ is also an interval, which proves (d). ■

Lemma III.11. *Fix an arbitrary $\epsilon > 0$. There are constants c_1, c_2 (depending on k and ϵ), such that for any disjoint family $\mathcal{F}_0 \subseteq \mathcal{F}$, for which there is no improving set of size at most $c_1 \log n$ of pathwidth at most c_2 we have $|OPT| \le ((k+1)/3+\epsilon)|\mathcal{F}_0|$, where $OPT \subseteq \mathcal{F}$ is a maximum size disjoint subfamily of $\mathcal{F}$.*

Proof: Let $C = \mathcal{F}_0 \cap OPT$ and denote $A_0 = \mathcal{F}_0 \setminus C$, $B_0 = OPT \setminus C$. Let G_0 be the subgraph of $G_{\mathcal{F}_0}$ induced by $A_0 \cup B_0$. We are going to construct a sequence of at most $1/\epsilon$ subgraphs of G_0, namely $G_i = G_0[A_i \cup B_i]$ for $i \ge 1$, where $A_i \subseteq A_0$, $B_i \subseteq B_0$, satisfying two invariants:

(a) in G_i there is no subset $X \subseteq B_i$ of size at most $2(k+1)^{1/\epsilon - i}$, such that $|N_{G_i}(X)| < |X|$,

(b) $|A_0 \setminus A_i| = |B_0 \setminus B_i|$.

Observe G_0 trivially satisfies (b) and in order to make G_0 satisfy (a) it is enough to set c_1 and c_2 so that

$$c_1 \ge 2(k+1)^{1/\epsilon},$$
$$c_2 \ge 4(k+1)^{1/\epsilon},$$

as there is no improving set of size at most $2(k+1)^{1/\epsilon}$ and pathwidth of an improving set of size x is at most $2x$. Consider subsequent values of i starting from 0. Split the vertices of B_i into groups B_i^1, B_i^2, B_i^3, consisting of vertices of B_i of degree exactly one, exactly two and at least three in G_i, respectively. Observe that because of (a) there is no isolated vertex of B_i in G_i and moreover no two vertices of B_i^1 have a common neighbour in G_i. Consider the following two cases:

- $|B_i^1| \ge \epsilon|OPT|$: in this case we construct a graph $G_{i+1} = G_0[A_{i+1} \cup B_{i+1}]$, where $A_{i+1} = A_i \setminus N_{G_i}(B_i^1)$ and $B_{i+1} = B_i^2 \cup B_i^3 = B_i \setminus B_i^1$. The invariant (a) is satisfied, as any set $X \subseteq B_{i+1}$ of size at most $2(k+1)^{1/\epsilon - i - 1}$ such that $|N_{G_{i+1}}(X)| < |X|$ would imply existence of a set $X' = X \cup (N_{G_i}(N_{G_i}(X)) \cap B_i^1)$ of size at most $(k+1) \cdot |X| \le 2(k+1)^{1/\epsilon - i}$, such that $|N_{G_i}(X')| < |X'|$ (see Figure 3).
- $|B_i^1| < \epsilon|OPT|$: We are going to use the following claim, which we prove later.

Claim III.12.

$$|B_i^2| \le (1+\epsilon)|A_i|$$

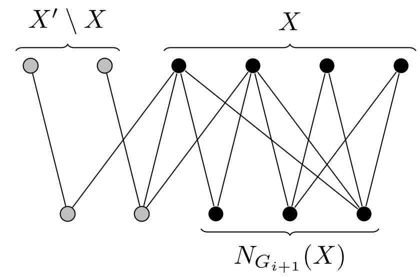

Figure 3: Lifting an improving set X in G_{i+1} to an improving set X' in G_i. Gray vertices belong to G_i but not to G_{i+1}.

As each vertex of A_i is of degree at most k in G_i, the number of edges of G_i is at most $k|A_i|$. At the same time the number of edges of G_i is at least $|B_i^1| + 2|B_i^2| + 3|B_i^3|$, therefore

$$|B_i^1| + 2|B_i^2| + 3|B_i^3| \le k|A_i|.$$

Note that summing the inequalities:

$$|B_i^1| \le \epsilon|A_i|$$
$$|B_i^1| \le \epsilon|A_i|$$
$$|B_i^2| \le (1+\epsilon)|A_i|$$
$$|B_i^1| + 2|B_i^2| + 3|B_i^3| \le k|A_i|$$

we obtain

$$|B_i| \le ((k+1)/3+\epsilon)|A_i|.$$

However $|OPT \setminus B_i| = |C| + |B_0 \setminus B_i| = |C| + |A_0 \setminus A_i| = |\mathcal{F}_0 \setminus A_i|$, where the second equality follows from invariant (b), hence $|OPT| \le ((k+1)/3+\epsilon)|\mathcal{F}_0|$.

In the second case we have proved the thesis, while the first case can appear only $1/\epsilon$ number of times, as in each step we remove at least $\epsilon|OPT|$ vertices from B_i. Therefore to finish the proof of Lemma III.11 it suffices to prove Claim III.12. ■

Proof of Claim III.12: Assume the contrary. Construct a multigraph $H = (A_i, E_H)$, where $E_H = \{e_x = uv : x \in B_i^2, N_{G_i}(x) = \{u, v\}\}$. Set $\Sigma = \mathcal{F}$ and for each edge $e_x = uv \in E_H$, set as w_{e_x} the set of all vertices of G_0 at distance at most $2/\epsilon$ from x. Observe that since G_0 is of maximum degree at most k, we have $|w_{e_x}| \le 2k^{2/\epsilon}$. For the same reason each vertex of G_0 appears in at most $2k^{2/\epsilon}$ sets w_{e_x}.

In order to use Corollary III.10 we need to reduce the graph H, in a way ensuring all its vertices are of degree at least 3. However we know, that the graph H is of average degree at least $2+2\epsilon$, since $|E_H|/|A_i| = |B_i^2|/|A_i| \ge 1+\epsilon$. Let $H' = H$. As long as there exist an isolated vertex, or a vertex of degree one in H' remove it. Note that such a

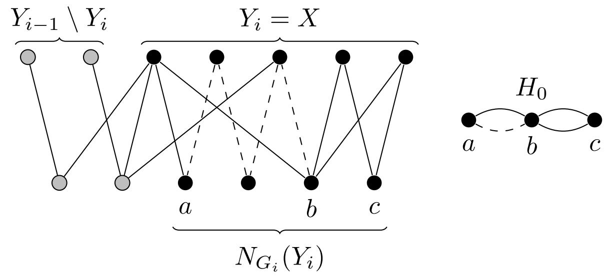

Figure 4: The right graph is $H_0 = (V_0, E_0)$ provided by Corollary III.10. The left side depicts the set X corresponding to E_0, as well as lifting the set $Y_i = X$ to Y_{i-1}. Gray vertices belong to G_{i-1} but not to G_i. The dashed path on the left between a and b in H' is contracted into an edge of H'' on the right.

reduction rule does not decrease the average degree of H'. Similarly if H' contains a path $v_0, v_1, \ldots, v_\ell, v_{\ell+1}$, where all vertices v_j for $1 \le j \le \ell$ are of degree exactly 2 and $\ell \ge 1/\epsilon$, then remove all the vertices v_j for $1 \le j \le \ell$ from H'. As this operation removes ℓ vertices, but only $\ell + 1$ edges, and $\ell \ge 1/\epsilon$, the average degree does not decrease. Finally, we construct H'' from H' by simultaneously considering all the maximal paths $v_0, v_1, \ldots, v_\ell, v_{\ell+1}$, with all internal vertices of degree two, and contracting each of such paths to a single edge $e' = v_0 v_{\ell+1}$ and setting $w_{e'} = \bigcup_{0 \le j \le \ell} w_{v_j v_{j+1}}$. Observe that for each edge e of H'' the size of w_e is upper bounded by $2k^{2/\epsilon}(1/\epsilon + 1)$, as a contracted path consist of at most $\lfloor 1/\epsilon + 1 \rfloor$ edges.

As H'' is of minimum degree at least 3, we apply Corollary III.10 to it, where $\gamma = 2k^{2/\epsilon}(1/\epsilon + 1)$. Let $H_0 = (V_0, E_0)$ and $\mathbb{P} = (B_i)_{i=1}^q$ be as defined in Corollary III.10. Let $X \subseteq B_i^2$ be the set of all the vertices of B_i^2 corresponding to the edges of E_0, including the vertices of B_i^2 that correspond to edges of H' that were contracted into some edge of E_0 (see Figure 4). As $|E_0| > |V_0|$ we have $|N_{G_i}(X)| < |X|$. Clearly X is of size at most $|E_0|(1/\epsilon + 1) \le (4(\log_{3/2}|\mathcal{F}| + 2) + 1)(1/\epsilon + 1)$, that is logarithmic in $|\mathcal{F}|$, as ϵ is a constant. It remains to show that we can lift X to an improving set of bounded pathwidth, while increasing its size only by a constant factor.

Let $Y_i = X$. For $j = i-1, \ldots, 0$ set $Y_j = Y_{j+1} \cup (N_{G_j}(N_{G_j}(Y_j)) \cap B_j^1)$ (see Figure 4). Observe that at each step the size of Y_j increases by a factor of at most $k+1$, hence $|Y_0| \le |Y_i|(k+1)^i$ and moreover Y_0 is an improving set w.r.t. $\mathcal{F}_0$. Since Y_0 is of size logarithmic in $|\mathcal{F}|$ it remains to show that $N_{G_{\mathcal{F}_0}}[Y_0]$ is of constant pathwidth.

Create a sequence of subsets $\mathbb{P}' = (B'_i)_{i=1}^q$, by taking as B'_i the set $(\bigcup_{e=uv \in E_0, u,v \in B_i} w_e \cap N_{G_{\mathcal{F}_0}}[Y_0])$. The size of each B'_i is at most $(w+1)^2\gamma$, where w is the width of $\mathbb{P}$, hence it remains to show that $\mathbb{P}'$ is indeed a path decomposition. Each vertex of $N_{G_{\mathcal{F}_0}}[Y_0]$ is within distance at most $2/\epsilon$ from some vertex of X, hence each vertex of $N_{G_{\mathcal{F}_0}}[Y_0]$ is contained in some set w_e for $e \in E_0$. Similarly each edge of $G_{\mathcal{F}_0}[N_{G_{\mathcal{F}_0}}[Y_0]]$ is within distance at most $2/\epsilon$ from some vertex of X, so it has both its endpoints in some set w_e for $e \in E_0$. Since $\mathbb{P}$ is a path decomposition each edge $e \in E_0$ has both its endpoints in some bag B_i, therefore $\bigcup_{1 \le i \le q} B'_i = N_{G_{\mathcal{F}_0}}[Y_0]$ and each edge of $N_{G_{\mathcal{F}_0}}[Y_0]$ has both its endpoints in some bag B'_i. Property (d) of Corollary III.10 implies that each w_e contributes to B'_i for values of i forming an interval I_e. Moreover if for two edges $e_1, e_2 \in E_0$ the intersection $w_{e_1} \cap w_{e_2}$ is non-empty, then by property (c) of Corollary III.10 we know that the intervals I_{e_1} and I_{e_2} have non-empty intersection. This ensures that each vertex v of $N_{G_{\mathcal{F}_0}}[Y_0]$ appears in a set of bags B'_i forming an interval in the sequence $\mathbb{P}'$, as each pair of intervals in $\{I_e : v \in w_e\}$ has non-empty intersection.

Therefore Y_0 is an improving set of logarithmic size and of constant pathwidth, which is a contradiction. Consequently $|B_i^2| \le (1+\epsilon)|A_i|$, which finishes the proof of Claim III.12. ■

Lemma III.11 together with the algorithm for searching improving sets of bounded pathwidth from Theorem III.6 gives a polynomial time $(k+1+\epsilon)/3$-approximation algorithm for k-SET PACKING for any constant k, proving Theorem I.2. In particular there is a $(4/3+\epsilon)$-approximation for the 3-DIMENSIONAL MATCHING problem.

IV. LOCAL SEARCH HARDNESS

In this section we are going to show, that there is no algorithm verifying for a given $\mathcal{F}_0 \subseteq \mathcal{F}$, whether there exists an improving set (see Definition III.2) of size at most r in $f(r)\mathrm{poly}(|\mathcal{F}|)$ time, even when $k = 3$. In fact we show a stronger hardness result, ruling out existence of an algorithm, that either finds a bigger disjoint family $\mathcal{F}_1$ (without any restriction on its distance from $\mathcal{F}_0$), or verifies that there is no improving set of size at most r. That is exactly the notion of *permissive* parameterized local search introduced by Marx and Schlotter in [21] (for more information about parameterized local search see [20]).

In our reduction, we use a standard W[1]-hard problem [10], namely MULTICOLORED CLIQUE parameterized by the clique size.

> MULTICOLORED CLIQUE
> **Input:** An undirected graph $G = (V, E)$, a positive integer k, and a color function $c : V \to \{0, \ldots, k-1\}$.
> **Question:** Does the graph G contain a clique of size k, where each vertex is of different color?

Theorem IV.1. *There is a constant $\alpha > 0$, such that given an instance $I = (G, k, c)$ of* MULTICOLORED CLIQUE *one can in polynomial time construct an instance $\mathcal{F} \subseteq 2^U$ of* 3-SET PACKING, *together with a disjoint subfamily $\mathcal{F}_0 \subseteq \mathcal{F}$ of size $|U|/3 - 1$, such that:*

- *If I is a YES-instance, then there exists a family $\mathcal{F}_1 \subseteq \mathcal{F}$ of disjoint $|U|/3$ sets, such that $|\mathcal{F}_0 \setminus \mathcal{F}_1| + |\mathcal{F}_1 \setminus \mathcal{F}_0| \leq \alpha k^2$,*
- *if there exists a disjoint subfamily $\mathcal{F}_1 \subseteq \mathcal{F}$ of size $|U|/3$, then I is a YES-instance.*

Proof: We start with a definition of a simple gadget, that will be used a couple of times in the construction.

Definition IV.2. For a positive integer $h \geq 1$ and a symbol x an (x, h)-amplifier is a family $\mathcal{F}_x \subseteq 2^{U_x}$ of sets of size 3, where

$$U_x = \{x_1, \ldots, x_{2 \cdot 4^h - 1}\}, \text{ and}$$
$$\mathcal{F}_x = \{\{x_i, x_{2i}, x_{2i+1}\} : 1 \leq i < 4^h\}$$

Let $I = (G = (V, E), k, c)$ be an instance of MULTICOLORED CLIQUE. W.l.o.g. we may assume that $k = 4^h$, where h is a positive integer, since otherwise we may add universal vertices (adjacent to all other vertices). We start with constructing an (x, h)-amplifier, which will be called the *top amplifier*, and (v, h)-amplifier for each $v \in V$, called *vertex amplifiers*. As the universe U we take

$$U = U_x \cup (\bigcup_{v \in V} U_v) \cup \{v_1', v_1'' : v \in V\} \cup \{s_{(i,j)} : 0 \leq i < j < k\} \cup \{\ell_i : 1 \leq i \leq 2k\}.$$

To the family $\mathcal{F}$ we add all the sets of $\mathcal{F}_x$ and $\mathcal{F}_v$ for $v \in V$, as well as:

(i) sets $\{v_1, v_1', v_1''\}$ for $v \in V$,
(ii) sets $\{x_{k+i}, v_1', v_1''\}$ for $0 \leq i < k$ for $v \in c^{-1}(i)$,
(iii) sets $\{u_{k+c(v)}, v_{k+c(u)}, s_{(c(u),c(v))}\}$ for $uv \in E$, $c(u) < c(v)$,
(iv) sets $\{v_{k+c(v)}, \ell_{2c(v)-1}, \ell_{2c(v)}\}$ for $v \in V$,
(v) sets $\{\ell_{3i-2}, \ell_{3i-1}, \ell_{3i}\}$ for $1 \leq i \leq \lfloor 2k/3 \rfloor$ (note that $2k = 2 \cdot 4^h \equiv 2 \pmod 3$),
(vi) consider all the elements $s_{(i,j)}$ in lexicographic order of pairs (i, j), take subsequent triples of elements and add them to the family $\mathcal{F}$, that is add sets

$$\{s_{(0,1)}, s_{(0,2)}, s_{(0,3)}\}, \ldots,$$
$$\{s_{(k-3,k-2)}, s_{(k-3,k-1)}, s_{(k-2,k-1)}\}$$

(note that $\binom{k}{2} \equiv 0 \pmod 3$, since $(k-1) \equiv 0 \pmod 3$).

To finish the construction we create a disjoint family $\mathcal{F}_0$ of size $|U|/3 - 1$ as follows:

- add to $\mathcal{F}_0$ sets $\{x_i, x_{2i}, x_{2i+1}\} \in \mathcal{F}_x$ for $1 \leq i < k$ such that $\lfloor \log_2 i \rfloor$ is odd.
- add to $\mathcal{F}_0$ sets $\{v_i, v_{2i}, v_{2i+1}\} \in \mathcal{F}_v$ for $v \in V$ and $1 \leq i < k$, such that $\lfloor \log_2 i \rfloor$ is odd.
- add to $\mathcal{F}_0$ all the sets from points (i), (v), (vi) of the construction of $\mathcal{F}$.

Note that the size of $\mathcal{F}_0$ equals $|U|/3 - 1$, as the only elements which are not covered by $\mathcal{F}_0$ are x_1, ℓ_{2k-1} and ℓ_{2k}.

Claim IV.3. *If I is a YES-instance, then there exists a disjoint family $\mathcal{F}_1 \subseteq \mathcal{F}$ of size $|U|/3$, such that $|\mathcal{F}_1 \setminus \mathcal{F}_0| + |\mathcal{F}_0 \setminus \mathcal{F}_1| = \mathcal{O}(k^2)$.*

Proof: Let $K \subseteq V$ be a solution to I, that is a multicolored clique of size k. Construct a disjoint family $\mathcal{F}_1$ as follows:

(a) add to $\mathcal{F}_1$ sets $\{x_i, x_{2i}, x_{2i+1}\} \in \mathcal{F}_x$ for each $1 \leq i < k$, such that $\lfloor \log_2 i \rfloor$ is even,
(b) add to $\mathcal{F}_1$ sets $\{v_i, v_{2i}, v_{2i+1}\} \in \mathcal{F}_x$ for $v \in K$ and $1 \leq i < k$, such that $\lfloor \log_2 i \rfloor$ is even,
(c) add to $\mathcal{F}_1$ sets $\{v_i, v_{2i}, v_{2i+1}\} \in \mathcal{F}_x$ for $v \in V \setminus K$ and $1 \leq i < k$, such that $\lfloor \log_2 i \rfloor$ is odd,
(d) for $0 \leq i < k$ add to $\mathcal{F}_1$ the set $\{x_{k+i}, v_1', v_1''\}$, where v is the unique vertex of K of color i,
(e) add to $\mathcal{F}_1$ sets $\{v_1, v_1', v_1''\}$ for $v \in V \setminus K$,
(f) add to $\mathcal{F}_1$ sets $\{u_{k+c(u)}, v_{k+c(v)}, s_{c(u),c(v)}\}$ for $u, v \in K$, $c(u) < c(v)$,
(g) add to $\mathcal{F}_1$ sets $\{v_{k+c(v)}, \ell_{2c(v)-1}, \ell_{2c(v)}\}$ for $v \in K$.

A direct check shows that the above family is disjoint and covers all the elements of U, hence $|\mathcal{F}_1| = |U|/3$. Note that in the above construction of $\mathcal{F}_1$ in each of the points (a), (d), (g) we add to $\mathcal{F}_1$ only $\mathcal{O}(k)$ sets, while in points (b), (f) we add to $\mathcal{F}_1$ $\mathcal{O}(k^2)$ sets, whereas in points (c) and (e) we add to $\mathcal{F}_1$ sets that are present in $\mathcal{F}_0$. Therefore the number of sets of $\mathcal{F}_1$ which are not present in $\mathcal{F}_0$ is upper bounded by a linear function in k^2. ■

Claim IV.4. *If there exists a disjoint family $\mathcal{F}_1$ of size $|U|/3$, then I is a YES-instance.*

Proof: Let $\mathcal{F}_1 \subseteq \mathcal{F}$ be any disjoint family of size $|U|/3$. Since the element x_1 can be covered only by the set $\{x_1, x_2, x_3\}$, the family $\mathcal{F}_1$ contains all the sets $\{x_i, x_{2i}, x_{2i+1}\} \in \mathcal{F}_x$ for $1 \leq i < k$, where $\lfloor \log_2 i \rfloor$ is even, and consequently elements x_{k+i} for $0 \leq i < k$ are not covered by sets of $\mathcal{F}_x$. Therefore elements x_{k+i} are covered by sets from point (ii) of the construction of $\mathcal{F}$, hence for each $0 \leq i < k$ in $\mathcal{F}_1$ there is exactly one set $\{v_1, v_2, v_3\} \in \mathcal{F}_1$ for $v \in c^{-1}(i)$, and let K be the set of those k multicolored vertices.

We want to show that K is a clique. As for each $v \in K$ we have $\{v_1, v_2, v_3\} \in \mathcal{F}_1$, the family $\mathcal{F}_1$ contains all the sets $\{v_i, v_{2i}, v_{2i+1}\}$ for $1 \leq i < k$ where $\lfloor \log_2 i \rfloor$ is even. Consequently elements v_{k+i} for $0 \leq i < k$, $i \neq c(v)$ are covered by sets from point (iii) of the construction of $\mathcal{F}$. Consider any pair $0 \leq i < j < k$. Denote as u the unique vertex of $K \cap c^{-1}(i)$ and let $\{u_{k+j}, v_{k+i}, s_{(i,j)}\}$ be the set of $\mathcal{F}_1$ covering u_{k+j}, where $v \in c^{-1}(j)$. This implies that v_{k+i} is not covered by a set of the (v, h)-amplifier, hence v_1 is covered by the (v, h)-amplifier, i.e. by $\{v_1, v_2, v_3\}$. Therefore $v \in K$ and the vertices of colors i and j of K are adjacent. Since i and j were selected arbitrarily, K is a clique. ■

The proof of Theorem IV.1 follows from Claim IV.3 and

Claim IV.4. ■

Theorem IV.1, together with the well-known fact that MULTICOLORED CLIQUE is W[1]-hard [10] implies Theorem I.1.

V. FUTURE WORK AND OPEN PROBLEMS

One can try to continue the research direction of Chan and Lau [6], who presented a strengthening of the standard LP relaxation, proving integrality gap of $(k+1)/2$ using a local search inspired analysis. We would like to ask a question whether it is possible to obtain some strengthened LP relaxation with integrality gap $(k+c)/3$-for some constant c.

Finally, we believe that it is worth looking into other problems, where local search algorithms were applied successfully, such as k-MEDIAN [3] or RESTRICTED MAX-MIN FAIR ALLOCATION [24]. A potential goal would be to design improved approximation local search algorithms using non-constant size swaps in the spirit of the framework of this paper.

ACKNOWLEDGEMENTS

We would like to thank Marcin Mucha for helpful discussions.

REFERENCES

[1] N. Alon, R. Yuster, and U. Zwick. Color-coding. *Journal of ACM*, 42(4):844–856, 1995.

[2] E. M. Arkin and R. Hassin. On local search for weighted k-set packing. In *ESA*, pages 13–22, 1997.

[3] V. Arya, N. Garg, R. Khandekar, A. Meyerson, K. Munagala, and V. Pandit. Local search heuristics for k-median and facility location problems. *SIAM J. Comput.*, 33(3):544–562, 2004.

[4] P. Berman. A *d*/2 approximation for maximum weight independent set in *d*-claw free graphs. In *SWAT*, pages 214–219, 2000.

[5] P. Berman and M. Karpinski. Improved approximation lower bounds on small occurrence optimization. *Electronic Colloquium on Computational Complexity (ECCC)*, 10(008), 2003.

[6] Y. H. Chan and L. C. Lau. On linear and semidefinite programming relaxations for hypergraph matching. *Math. Program.*, 135(1-2):123–148, 2012.

[7] B. Chandra and M. M. Halldórsson. Greedy local improvement and weighted set packing approximation. *J. Algorithms*, 39(2):223–240, 2001.

[8] M. Cygan, F. Grandoni, and M. Mastrolilli. How to sell hyperedges: The hypermatching assignment problem. In *SODA*, page to appear, 2013.

[9] R. Downey and M. R. Fellows. *Parameterized Complexity*. Springer-Verlag, 1999.

[10] M. R. Fellows, D. Hermelin, F. A. Rosamond, and S. Vialette. On the parameterized complexity of multiple-interval graph problems. *Theoretical Computer Science*, 410:53–61, 2009.

[11] J. Flum and M. Grohe. *Parameterized Complexity Theory*. Springer-Verlag, 2006.

[12] M. M. Halldórsson. Approximating discrete collections via local improvements. In *SODA*, pages 160–169, 1995.

[13] J. Håstad. Clique is hard to approximate within $n^{1-\epsilon}$. *Electronic Colloquium on Computational Complexity (ECCC)*, 4(38), 1997.

[14] E. Hazan, S. Safra, and O. Schwartz. On the complexity of approximating k-dimensional matching. In *RANDOM-APPROX*, pages 83–97, 2003.

[15] E. Hazan, S. Safra, and O. Schwartz. On the complexity of approximating k-set packing. *Computational Complexity*, 15(1):20–39, 2006.

[16] C. A. J. Hurkens and A. Schrijver. On the size of systems of sets every t of which have an SDR, with an application to the worst-case ratio of heuristics for packing problems. *SIAM J. Discrete Math.*, 2(1):68–72, 1989.

[17] R. Impagliazzo and R. Paturi. On the complexity of k-SAT. *J. Comput. Syst. Sci.*, 62(2):367–375, 2001.

[18] R. M. Karp. Reducibility among combinatorial problems. In *Complexity of Computer Computations*, pages 85–103, 1972.

[19] T. Kloks. *Treewidth, Computations and Approximations*, volume 842 of *Lecture Notes in Computer Science*. Springer, 1994.

[20] D. Marx. Local search. *Parameterized Complexity Newsletter*, 3:7–8, 2008.

[21] D. Marx and I. Schlotter. Stable assignment with couples: Parameterized complexity and local search. *Discrete Optimization*, 8(1):25–40, 2011.

[22] M. Naor, L. J. Schulman, and A. Srinivasan. Splitters and near-optimal derandomization. In *FOCS*, pages 182–191, 1995.

[23] R. Niedermeier. *Invitation to Fixed-Parameter Algorithms*. Oxford University Press, 2006.

[24] L. Polacek and O. Svensson. Quasi-polynomial local search for restricted max-min fair allocation. In *ICALP (1)*, pages 726–737, 2012.

[25] M. Sviridenko and J. Ward. Large neighborhood local search for the maximum set packing problem. In *ICALP (1)*, pages 792–803, 2013.