Improved approximation for 3D matching via bounded pathwidth local search

Marek Cygan

Presented by Martin Böhm

Approximation and Online Algorithm Seminar 2014, MFF UK

Definitions and Notation

D(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.

T(Main Theorem): For any $\varepsilon > 0$ and any integer $k \geq 3$ there is a polynomial time $(k+1+\varepsilon)/3$-approximation algorithm for k-Set-Packing.

N: $\overline{\mathcal{F}_0} \equiv \mathcal{F} \setminus \mathcal{F}_0$.

N: For a vertex set $\mathcal{X}$, $N(\mathcal{X}) \equiv$ neighbors of vertices in $\mathcal{X}$.

N: For a vertex set $\mathcal{X}$, $N[\mathcal{X}] \equiv N(\mathcal{X}) \cup \mathcal{X}$.

D(Pathwidth): A graph G has pathwidth at most pw if it has a tree-decomposition of treewidth at most pw where the decomposition itself is a path.

D(Conflict graph): For a disjoint starting family $\mathcal{F}_0 \subseteq \mathcal{F}$ we define a *conflict graph* $\mathrm{Con}_{\mathcal{F}_0}$ as a bipartite graph with vertex set $\mathcal{F}$ and edge set $\{S_1 S_2 | S_1 \in \mathcal{F}_0, S_2 \in \overline{\mathcal{F}_0}, S_1 \cap S_2 \neq \emptyset\}$.

We will use Con if the starting family is clear from context. We lose some information in the conflict graph – namely the disjointness information for neighbors in $\overline{\mathcal{F}_0}$.

D(Improving set...): For a starting family $\mathcal{F}_0$ we call an *improving set* $\mathcal{X}$ a set of vertices of $\overline{\mathcal{F}_0}$ such that:

1. All members of $\mathcal{X}$ are pairwise disjoint;
2. $|N(\mathcal{X})| < |\mathcal{X}|$, i.e. we can improve $\mathcal{F}_0$ using $\mathcal{X}$.

D(...of bounded pathwidth): An improving set $\mathcal{X}$ with respect to $\mathcal{F}_\prime \subseteq \mathcal{F}$ has *pathwidth at most* pw if the subgraph of the conflict graph induced by $N[\mathcal{X}]$ is of pathwidth at most pw.

Part 1 – The FPT Algorithm

Color coding:

1. A dynamic-programming technique used to efficiently find a small structure of bounded treewidth (a path, cycle, etc.) within a larger graph.
2. Idea: Have a coloring function assign colors to vertices. Look only for the substructure that is *colorful.*
3. Originally probabilistic: if the probability that the structure becomes colorful is non-trivial, we try many coloring functions and get a polynomial, constant-error algorithm.
4. Can be derandomized by a standard argument.

L(Bounded pathwidth algorithm): Let pw, k (parameter of k-Set-Packing) and r (size of the improving set we look for, later set to $O(\log |\mathcal{F}|)$) be fixed. There exists an algorithm that:

1. given a disjoint family $\mathcal{F}_0 \subseteq \mathcal{F}$ and two coloring functions $c_{\text{path}} : \mathcal{F}_0 \to [r-1]$ and $c_{\text{univ}} : U \to [rk]$,
2. in time $2^{O(rk)} |\mathcal{F}|^{O(\text{pw})}$,
3. determines whether an improving set $\mathcal{X}$ of size at most r and pathwidth at most pw exists, s.t. c_{path} is injective on $N_{\text{Con}}(\mathcal{X})$ and c_{univ} is injective on $\bigcup_{S \in \mathcal{X}} S$.

P: Create an auxiliary digraph G_{state} of size $O(2^{r(k+1)}|\mathcal{F}|^{\text{pw}+1})$. Every state (vertex of G_{state}) will represent a partial pathwidth decomposition. We will traverse this graph and look for a pathwidth decomposition that is also an improving set.

Instead of the entire partial pathwidth decomposition, we store in every state a triplet $(D_{\text{path}}, D_{\text{univ}}, B)$, where

1. D_{path} are the colors of members of $\mathcal{F}_0$ we have already traversed in the decomposition;
2. D_{univ} are the colors of the universe that we have seen so far (inside sets of $\overline{\mathcal{F}_0}$ that we have traversed);
3. B is a set of size at most pw + 1 – our current pathwidth decomposition bag.

We add directed edges to G_{state} which correspond to progressing along a pathwidth decomposition. We then run a graph search on G_{state}.

The injectiveness of D_{path} ensures that we do not go back in the pathwidth decomposition; the injectiveness of D_{univ} ensures that the visited sets of $\overline{\mathcal{F}_0}$ are disjoint.

Clm: There exists a path in the graph G_{state} from the vertex $(\emptyset, \emptyset, \emptyset)$ to the vertex $(D_{\text{path}}, D_{\text{univ}}, \emptyset)$ for $D_{\text{path}} < D_{\text{univ}}/k$ if and only if there exists an improving set $\mathcal{X}$ of size at most r of pathwidth at most pw, such that D_{path} is injective on $N(\mathcal{X})$ and D_{univ} is injective on $\bigcup_{S \in \mathcal{X}} S$.

Part 2 – Constant Pathwidth Suffices

T(Main claim): Let k be an integer, $\varepsilon > 0$. Then there exist constants $c_1(k, \varepsilon), c_2(k, \varepsilon)$ such that for any disjoint family $\mathcal{F}_0 \subseteq \mathcal{F}$ for which there is no improving set of size at most $c_1 \log n$ that has pathwidth at most c_2, we have $|OPT| \leq ((k+1)/3 + \varepsilon)|\mathcal{F}_0|$.

P: Assume we are in a situation where there is no valid improving set. Set $C \equiv \text{OPT} \cap \mathcal{F}_0$, $A_0 \equiv \mathcal{F}_0 \setminus C$, $B_0 \equiv \text{OPT} \setminus C$. We restrict ourselves to $G[A_0 \cup B_0]$. We will create a sequence of $1/\varepsilon$ subgraphs $G[A_i \cup B_i]$ which have roughly the same properties as (A_0, B_0), that is:

1. in $G[A_i \cup B_i]$ there is no subset $\mathcal{X} \subseteq B_i$ of size at most $2(k+1)^{1/\varepsilon - i}$ such that $|N(\mathcal{X})| < |\mathcal{X}|$.
2. $|A_0 \setminus A_i| = |B_0 \setminus B_i|$, or equivalently $|A_0 \setminus B_0| = |A_i \setminus B_i|$.

Split B_i into sets B_i^1, B_i^2, $B_i^{\geq 3}$, where the superscript indicates the degree of the vertices.

We note the following two claims:

O: Either $|B_i^1| \leq \varepsilon|\text{OPT}| \leq \varepsilon|A_i|$ or we can construct $G[A_{i+1} \cup B_{i+1}]$.

Clm(Key claim): B_i^2 always satisfies $|B_i^2| \leq (1+\varepsilon)|A_i|$.

If the two claims hold, we note that the number of edges satisfies $||G[A_i \cup B_i]|| \geq 1|B_i^1| + 2|B_i^2| + 3|B_i^3|$, but it also satisfies $||G[A_i \cup B_i]|| \leq k|A_i|$. Summing up all inequalities together, we get $|B_i| \leq ((k+1)/3 + \varepsilon)|A_i|$ and finally $|\text{OPT}| \leq ((k+1)/3 + \varepsilon)|\mathcal{F}_0|$.

P(Key claim): Restrict the graph to only $G \equiv G[A_i \cup B_i^2]$. For contradiction, assume $|B_i^2| > (1+\varepsilon)|A_i|$.

The graph G is a bipartite graph where every vertex of the partition B_i^2 has degree exactly two. We can look at this graph as a multigraph G' which has the vertex set A_i and edge set B_i^2 (understood as pairs of A_i).

$|B_i^2| > (1+\varepsilon)|A_i|$ implies $\frac{||G'||}{|G'|} = 1+\varepsilon$, which implies $d(G') = 2 + 2\varepsilon$. We find an improving set of size $O(\log G') = O(\log \mathcal{F})$ and of constant pathwidth using Overcharged short cycle lemma below.

L(Short cycle lemma): Let G be a graph of minimal degree 3. Then G contains a cycle of length at most $O(\log n)$.

L(Overcharged short cycle lemma): Let H be a multigraph, $|H| = n$, $\delta(H) \geq 3$. Let w_e be a labeling on edges of H by a subset of some alphabet Σ. Assume that for some γ, the following holds: $\forall e \in E : w_e \leq \gamma$ and $\forall \alpha \in \Sigma : |\{\text{occurences of } \alpha \text{ in all labels}\}| \leq \gamma$.

Then there exists a subtree $T_0 = (V_0, E_0)$ with root r and two edges e_1 and e_2 outside T_0 such that:

1. $|V_0| \leq 4(\log_{3/2} n + 2)$,
2. both $T_0 + e_1$ and $T_0 + e_2$ contains a cycle,
3. T_0 is a tree with at most 4 leaves,
4. If we set $\beta \equiv \lceil \log_{3/2}(12\gamma^2) \rceil$, then every pair of edges $e_a, e_b \in T_0$ it holds that if their labels intersect, then $|dist(e_a, r) - dist(e_b, r)| \leq \beta$, where for an edge $e = uv$, $dist(e, r) \equiv \min(dist(u, r), dist(v, r))$.

N: Denote P as the subtree of T_0 such that P is connected and it contains all endpoints of e_1 and e_2.

O: $P + e_1 + e_2$ has n vertices, but $n+1$ edges.

C: Let T_0, e_1, e_2 be as in Overcharged short cycle lemma. Then the graph $P + e_1 + e_2$ has a path decomposition of width at most $4\beta + 3$.