# Covering a graph with a constrained forest 

Cristina Bazgan ${ }^{1} \quad$ Basile Couëtoux ${ }^{1} \quad$ Zsolt Tuza ${ }^{2, *}$<br>${ }^{1}$ Université Paris-Dauphine, LAMSADE<br>Place du Marechal de Lattre de Tassigny, 75775 Paris Cedex 16, France<br>${ }^{2}$ Computer and Automation Institute, Hungarian Academy of Sciences H-1111 Budapest, Kende u. 13-17, Hungary<br>and<br>Department of Computer Science, University of Pannonia<br>H-8200 Veszprém, Egyetem u. 10, Hungary


#### Abstract

Given an undirected graph on $n$ vertices with weights on its edges, Min WCF $(p)$ consists of computing a covering forest of minimum weight such that each of its tree components contains at least $p$ vertices. It has been proved that Min $\operatorname{WCF}(p)$ is $N P$-hard for any $p \geq 4$ (Imielinska et al., 1993) but ( $2-\frac{1}{n}$ )-approximable (Goemans and Williamson, 1995). While Min $\operatorname{WCF}(2)$ is polynomial-time solvable, already the unweighted version of Min $\mathrm{WCF}(3)$ is $N P$-hard even on planar bipartite graphs of maximum degree 3. We prove here that for any $p \geq 4$, the unweighted version is $N P$-hard, even for planar bipartite graphs of maximum degree 3; moreover, the unweighted version for any $p \geq 3$ has no ptas for bipartite graphs of maximum degree 3. The latter theorem is the first-ever APX-hardness result on this problem. On the other hand, we show that $\operatorname{Min} \operatorname{WCF}(p)$ is polynomial-time solvable on graphs with bounded treewidth, and for any $p$ bounded by $O\left(\frac{\log n}{\log \log n}\right)$ it has a ptas on planar graphs.


## 1 Introduction

Let $G=(V, E)$ be a graph with $|V|=n$ vertices. An edge cover of $G$ is a subset of the edge set $E$ such that every vertex is incident with at least one edge in the covering set. Finding the minimum size, $\rho(G)$, of an edge cover of a graph is a fundamental problem. As proved by Gallai [8], it is strongly related to determining the maximum size, $\nu(G)$, of a matching in $G$. A famous result of [8] states that any graph $G$ without isolated vertices satisfies the identity $\nu(G)+\rho(G)=n$. As a matter of fact, the relation is much more than quantitative: every maximum matching of a graph can be extended to a minimum edge cover, and also conversely, every minimum edge cover contains a maximum matching. In this way, one can derive a minimum-size edge cover from a maximum matching $M$ just by adding an arbitrary incident edge for each vertex missing from $M$. Hence, a minimum-size edge cover can be found in polynomial time.

In the case where the graph $G=(V, E)$ has weights on its edges, the minimum-weight edge cover problem can be reduced to the problem of finding a minimum-weight perfect matching; a simple reduction is described e.g. in the first volume of Schrijver's monograph [17, Section

[^0]19.2]. As a consequence, an optimal solution can be found in $O\left(n^{3}\right)$ time by the results of Edmonds and Johnson [7]. It should be noted, however, that the relation between maximum matchings and minimum edge covers does not remain valid for weighted graphs, neither for uniform hypergraphs of edge size greater than two [18].

A problem that generalizes the minimum-weight edge cover problem in a very natural way is the Min Weighted Constrained Forest problem, denoted by Min $\operatorname{WCF}(p)$ in the sequel. It consists of computing a spanning forest of $G$ of minimum weight such that every tree component contains at least $p$ vertices, for a given integer $p$. Although traditionally $p$ is assumed to be a constant, the methods proving our positive results will allow us to take $p$ as a function of $n$, too.

Monnot and Toulouse [14] proved that the unweighted version of Min WCF (3) is $N P$-hard even on planar bipartite graphs of maximum degree three. Imielinska et al. [10] showed that $\operatorname{Min} \operatorname{WCF}(p)$ is $N P$-hard for $p \geq 4$, and that a greedy algorithm achieves a 2 -approximation. Interestingly enough, a different algorithm studied by Laszlo and Mukherjee [12] has exactly the same tight worst-case ratio of 2 , as well as a common generalization of those two approaches [13]. With the methods of Goemans and Williamson [9], just a slightly better ratio $2-\frac{1}{n}$ can be achieved.

Let us denote by $\operatorname{Min} \operatorname{CF}(p)$ the unweighted version of the problem. Until now, nothing was known about the complexity of $\operatorname{Min} \operatorname{CF}(p)$ for $p \geq 4$. We settle this problem by showing that Min $\mathrm{CF}(p)$ is $N P$-hard for any $p \geq 4$, already on planar bipartite graphs with maximum degree three.

Moreover, we study non-approximability of these problems for the first time. In this direction we prove that dropping the condition of planarity, $\operatorname{Min} \operatorname{CF}(p)$ becomes $A P X$-hard for any $p \geq 3$ on bipartite graphs, and even on those with maximum degree three. It also turns out that this weakening in the condition necessarily has to appear in non-approximability results, since we can design a polynomial-time approximation scheme for $\operatorname{Min} \operatorname{WCF}(p)$ on planar graphs. In this result we may allow $p$ to be bounded by $O\left(\frac{\log n}{\log \log n}\right)$. An important tool in the proof is an algorithm computing an optimal solution for $\operatorname{MIN} \operatorname{WCF}(p)$ on any input graph of treewidth at most $k$ in $O\left(k^{c k} p^{2 k+2} n\right)$ time, for some constant $c$. This time bound is valid without any restrictions on the growth of $p$, hence applicable also in graph classes which cannot be treated with Courcelle's powerful method via monadic second-order logic [4]. The latter would require to fix $p$ as a constant.

It is worth noting that the unweighted case admits a much simpler approach than the weighted one. Indeed, since every feasible solution of $\operatorname{Min} \mathrm{CF}(p)$ has at most $n / p$ connected components, the optimum can never have a value smaller than $n-n / p=\frac{p-1}{p} n$, and hence any spanning tree gives a $\left(1+\frac{1}{p-1}\right)$-approximation.

Our results on NP- and APX-hardness are proved in Section 3, while efficient algorithms are presented in Section 4. Conclusions are provided in the final section.

## 2 Preliminaries

We begin with some basic definitions, summarized in three groups.

Problems. First, we formally define the problem we will study in the sequel.

## Min Weighted Constrained Forest $p$ (Min WCF $(p)$ )

Input: An undirected graph $G=(V, E)$ with non-negative weights on its edges.
Output: A spanning forest of minimum weight where every tree is of order at least $p$ (i.e., it contains at least $p$ vertices).

The unweighted version of $\operatorname{Min} \operatorname{WCF}(p)$ will be denoted by Min $\operatorname{CF}(p)$.
In our proofs we shall use the following problems:
3-Dimensional Matching (3DM)
Input: Three disjoint sets $A, B, C$ of the same size $q$, and a set $\mathcal{T} \subseteq A \times B \times C$ of triplets.
Question: Does $\mathcal{T}$ contain a perfect matching, that is a subset $\mathcal{M} \subseteq \mathcal{T}$ such that $|\mathcal{M}|=q$ and no two elements of $\mathcal{M}$ agree in any coordinate (i.e., for any $(a, b, c),\left(a^{\prime}, b^{\prime}, c^{\prime}\right) \in \mathcal{M}$ we have $a \neq a^{\prime}, b \neq b^{\prime}$, and $\left.c \neq c^{\prime}\right)$ ?

We can associate a bipartite graph with any instance of 3DM as follows. We take an 'element-vertex' for each element of $A \cup B \cup C$, and one 'triplet-vertex' for each triplet in $\mathcal{T}$. There is an edge connecting a triplet-vertex to an element-vertex if and only if the element is a member of the triplet. Moreover, we say that the instance is planar if the graph associated with it is planar.

## Max 3-Dimensional Matching (Max 3DM)

Input: Three disjoint sets $A, B, C$ of the same size and a set $\mathcal{T} \subseteq A \times B \times C$ of triplets. Output: A matching (set of mutually disjoint triplets) $\mathcal{M} \subseteq \mathcal{T}$ of maximum size.

The restricted versions of 3 DM and of Max 3DM where each element of $A \cup B \cup C$ appears in exactly two triplets will be denoted by $3 \mathrm{DM}_{2}$ and $\mathrm{Max} 3 \mathrm{DM}_{2}$, respectively.

Approximability. Given an instance $x$ of an optimization problem $A$ and a feasible solution $y$ of $x$, we denote by $v(x, y)$ the value of the solution $y$, and by $o p t_{A}(x)$ the value of an optimum solution of $x$. The performance ratio of $y$ is

$$
R(x, y)=\max \left\{\frac{v(x, y)}{o p t_{A}(x)}, \frac{o p t_{A}(x)}{v(x, y)}\right\}
$$

For a constant $c>1$, an algorithm is a c-approximation if for any instance $x$ of the problem it returns a solution $y$ such that $R(x, y) \leq c$. An optimization problem is said to be constant approximable if, for some $c>1$, there exists a polynomial-time $c$-approximation for it. The class of problems which are constant approximable is denoted by $A P X$. An optimization problem has a polynomial-time approximation scheme (a ptas, for short) if, for every constant $\varepsilon>0$, there exists a polynomial-time $(1+\varepsilon)$-approximation for it.

Reductions. The notion of $L$-reduction was introduced by Papadimitriou and Yannakakis in [15]. Let $A$ and $B$ be two optimization problems. Then $A$ is said to be $L$-reducible to $B$ if there are two constants $\alpha, \beta>0$ such that

1. there exists a function, computable in polynomial time, which transforms each instance $x$ of $A$ into an instance $x^{\prime}$ of $B$ such that $o p t_{B}\left(x^{\prime}\right) \leq \alpha \cdot o p t_{A}(x)$,
2. there exists a function, computable in polynomial time, which transforms each solution $y^{\prime}$ of $x^{\prime}$ into a solution $y$ of $x$ such that $\left|v(x, y)-o p t_{A}(x)\right| \leq \beta \cdot\left|v\left(x^{\prime}, y^{\prime}\right)-o p t_{B}\left(x^{\prime}\right)\right|$.

For us the important property of this reduction is that if a maximization problem $A$ is $L$-reducible to a minimization problem $B$ and $B$ is $c$-approximable then $A$ is $\frac{1}{1-\alpha \beta(c-1)}$ approximable.

## 3 Complexity for $p \geq 3$

Monnot and Toulouse [14] proved that the unweighted version of
Min CF (3) was proved $N P$-hard even on planar bipartite graphs of maximum degree three [14]. In this section we prove that Min $\operatorname{CF}(3)$ is intractable for every $p \geq 4$ for the same restricted classes of problem instances. We first deal with time complexity and then with approximation hardness.

## 3.1 $N P$-hardness for $p \geq 4$, planar bipartite unweighted graphs

Theorem 1. For any $p \geq 4$, $\operatorname{Min} \operatorname{CF}(p)$ is NP-hard, even on planar bipartite graphs with maximum degree 3 .

Proof. First, we prove $N P$-hardness for $p=4$. We construct a polynomial reduction from 3 DM to Min $\mathrm{CF}(4)$. Let $I=(A, B, C, \mathcal{T})$ be an instance of 3 DM where $|A|=|B|=[C \mid=q$ and $\mathcal{T}=\left\{T_{1}, \ldots, T_{m}\right\}$. Assume, without loss of generality, that each element occurs in at least one triplet (otherwise $I$ admits no perfect matching at all).

The graph instance $G(I)=(V, E)$ of Min CF (4) is constructed as follows (see an illustration in Figure 1 with a particular instance $I$ of 3 DM$)$. For each triplet $T_{\ell}=\left(a_{i}, b_{j}, c_{k}\right) \in \mathcal{T}$ we create a copy of a star with 3 branches, that we shall call star gadget, with vertices $a_{i}^{\ell}, b_{j}^{\ell}, c_{k}^{\ell}, d^{\ell}$ and edges $\left(a_{i}^{\ell}, d^{\ell}\right),\left(b_{j}^{\ell}, d^{\ell}\right),\left(c_{k}^{\ell}, d^{\ell}\right)$. These stars are assumed to be vertex-disjoint. For their union we denote

$$
\begin{gathered}
V_{T}=\bigcup_{T_{\ell}=\left(a_{i}, b_{j}, c_{k}\right) \in \mathcal{T}}\left\{a_{i}^{\ell}, b_{j}^{\ell}, c_{k}^{\ell}, d^{\ell}\right\} \quad \text { and } \\
E_{T}=\bigcup_{T_{\ell}=\left(a_{i}, b_{j}, c_{k}\right) \in \mathcal{T}}\left\{\left(a_{i}^{\ell}, d^{\ell}\right),\left(b_{j}^{\ell}, d^{\ell}\right),\left(c_{k}^{\ell}, d^{\ell}\right)\right\}
\end{gathered}
$$

To the elements of $A \cup B \cup C$ we assign linking gadgets; if an element appears in $h$ triplets, then $h-1$ gadgets will be associated with it.

Consider first the set $A$. We define the set $J_{A} \subset A \times \mathbb{N} \times \mathbb{N}$ to be the collection of all $\left(a_{i}, r, s\right)$ with the following properties: $a_{i} \in A$, and $r, s$ are two consecutive indices of triplets containing $a_{i}$, in the sense that $a_{i} \in T_{r}$ and $a_{i} \in T_{s}$ but $a_{i} \notin T_{\ell}$ for any $r<\ell<s$. (In this general setting we allow that some $a_{i}$ occur in just one member of $\mathcal{T}$ - hence generating no linking gadgets - although in such a situation $I$ would easily be reducible to a smaller instance.) Each $\left(a_{i}, r, s\right) \in J_{A}$ defines a linking gadget inducing a 'claw' in $G(I)$, with vertices $a_{i}^{r}, a_{i}^{s}, a_{i}^{r, s}, \bar{a}_{i}^{r, s}$ and edges $\left(a_{i}^{r}, a_{i}^{r, s}\right),\left(a_{i}^{r, s}, a_{i}^{s}\right),\left(a_{i}^{r, s}, \bar{a}_{i}^{r, s}\right)$. For their union we denote

$$
V_{A}=\bigcup_{\left(a_{i}, r, s\right) \in J_{A}}\left\{a_{i}^{r, s}, \bar{a}_{i}^{r, s}\right\} \quad \text { and } \quad E_{A}=\bigcup_{\left(a_{i}, r, s\right) \in J_{A}}\left\{\left(a_{i}^{r}, a_{i}^{r, s}\right),\left(a_{i}^{r, s}, a_{i}^{s}\right),\left(a_{i}^{r, s}, \bar{a}_{i}^{r, s}\right)\right\}
$$

The sets $J_{B}$ and $J_{C}$ are defined in a similar way. Each $\left(b_{j}, r, s\right) \in J_{B}$ defines a linking gadget inducing a 'claw' in $G(I)$, with vertices $b_{j}^{r}, b_{j}^{s}, b_{j}^{r, s}, \bar{b}_{j}^{r, s}, \underline{b}_{j}^{r, s}$ and edges $\left(b_{j}^{r}, b_{j}^{r, s}\right),\left(b_{j}^{r, s}, b_{j}^{s}\right)$, $\left(b_{j}^{r, s}, \underline{b}_{j}^{r, s}\right),\left(\underline{b}_{j}^{r, s}, \bar{b}_{j}^{r, s}\right)$. For their union we denote

$$
\begin{gathered}
V_{B}=\bigcup_{\left(b_{j}, r, s\right) \in J_{B}}\left\{b_{j}^{r, s}, \underline{b}_{j}^{r, s}, \underline{b}_{j}^{r, s}\right\} \quad \text { and } \\
E_{B}=\bigcup_{\left(b_{j}, r, s\right) \in J_{B}}\left\{\left(b_{j}^{r}, b_{j}^{r, s}\right),\left(b_{j}^{r, s}, b_{j}^{s}\right),\left(b_{j}^{r, s}, \underline{b}_{j}^{r, s}\right),\left(\underline{b}_{j}^{r, s}, \bar{b}_{j}^{r, s}\right)\right\} .
\end{gathered}
$$

The same is done for the set $C$ in complete analogy to $B$, hence introducing the linking gadgets $\left(c_{k}, r, s\right) \in J_{C}$ and obtaining vertex set $V_{C}$ and edge set $E_{C}$.

After all these preparations, let the graph $G(I)=(V, E)$ has vertex set $V=V_{T} \cup V_{A} \cup$ $V_{B} \cup V_{C}$ and edge set $E=E_{T} \cup E_{A} \cup E_{B} \cup E_{C}$. Since an element of $A \cup B \cup C$ appearing in $h$ triplets has $h$ copies and $h-1$ linking gadgets in $G(I)$, we can see that the total number of vertices is precisely $12|\mathcal{T}|-8 q$. Hence, the transformation is linear with respect to input size.


Figure 1: Graph $G(I)$ derived from instance $I=(A, B, C, \mathcal{T})$ of 3 DM with $A=\left\{a_{1}, a_{2}, a_{3}\right\}$, $B=\left\{b_{1}, b_{2}, b_{3}\right\}, C=\left\{c_{1}, c_{2}, c_{3}\right\}$, and $\mathcal{T}=\left\{T_{1}, \ldots, T_{6}\right\}$ with $T_{1}=\left(a_{1}, b_{2}, c_{2}\right), T_{2}=\left(a_{1}, b_{1}, c_{1}\right)$, $T_{3}=\left(a_{2}, b_{3}, c_{1}\right), T_{4}=\left(a_{1}, b_{3}, c_{3}\right), T_{5}=\left(a_{1}, b_{1}, c_{2}\right), T_{6}=\left(a_{3}, b_{1}, c_{3}\right)$.

We are going to show that $I$ contains a perfect matching if and only if $G(I)$ contains a forest of weight at most $9|\mathcal{T}|-6 q$ in which every tree component is of order at least 4 .

Suppose first that $\mathcal{M}$ is a perfect matching of $I$. For a $T_{\ell}=\left(a_{i}, b_{j}, c_{k}\right) \in \mathcal{M}$ we write $f_{\mathcal{M}}\left(a_{i}\right)=f_{\mathcal{M}}\left(b_{j}\right)=f_{\mathcal{M}}\left(c_{k}\right)=\ell$, and further denote by $F_{T}$ the set of edges of the star gadgets corresponding to triplets of $\mathcal{M}$ :

$$
F_{T}=\bigcup_{T_{\ell}=\left(a_{i}, b_{j}, c_{k}\right) \in \mathcal{M}}\left\{\left(a_{i}^{\ell}, d^{\ell}\right),\left(b_{j}^{\ell}, d^{\ell}\right),\left(c_{k}^{\ell}, d^{\ell}\right)\right\}
$$

and

$$
\begin{aligned}
F_{a_{i}}= & \bigcup_{\substack{r<f_{\mathcal{M}}\left(a_{i}\right) \\
\left(a_{i}, r, s\right) \in J_{A}}}\left\{\left(d^{r}, a_{i}^{r}\right),\left(a_{i}^{r}, a_{i}^{r, s}\right),\left(a_{i}^{r, s}, \bar{a}_{i}^{r, s}\right)\right\} \quad \cup \\
& \bigcup_{\substack{s>f_{\mathcal{M}}\left(a_{i}\right) \\
\left(a_{i}, r, s\right) \in J_{A}}}\left\{\left(d^{s}, a_{i}^{s}\right),\left(a_{i}^{s}, a_{i}^{r, s}\right),\left(a_{i}^{r, s}, \bar{a}_{i}^{r, s}\right)\right\}, \\
F_{b_{j}}= & \bigcup_{\substack{r<\mathcal{M}^{\prime}\left(b_{j}\right) \\
\left(b_{j}, r, s\right) \in J_{B}}}\left\{\left(b_{j}^{r}, b_{j}^{r, s}\right),\left(b_{j}^{r, s}, \underline{b}_{j}^{r, s}\right),\left(\underline{b}_{j}^{r, s}, \bar{b}_{j}^{r, s}\right)\right\} \quad \cup \\
& \bigcup_{\substack{s \not \mathcal{M}^{\prime}\left(b_{j}\right) \\
\left(b_{j}, r, s\right) \in J_{B}}}\left\{\left(b_{j}^{s}, b_{j}^{r, s}\right),\left(b_{j}^{r, s}, \underline{b}_{j}^{r, s}\right),\left(\underline{b}_{j}^{r, s}, \bar{b}_{j}^{r, s}\right)\right\} .
\end{aligned}
$$

The set $F_{c_{k}}$ is defined like $F_{b_{j}}$. Let $F_{\mathcal{M}}=\bigcup_{e \in A \cup B \cup C} F_{e} \cup F_{T}$. Forest $F_{M}$ covers $G(I)$ with trees of order 4 . Thus, it is a forest of weight $9|\mathcal{T}|-6 q$.

Conversely, let $F$ be a covering forest of $G(I)$ of weight at most $9|\mathcal{T}|-6 q$, where each tree is of order at least 4 . Remark that since the graph $G(I)$ has exactly $12|\mathcal{T}|-8 q$ vertices, a covering forest without isolated edges and isolated vertices is of weight at least $9|\mathcal{T}|-6 q$. Thus, every tree in $F$ is of order exactly 4 , and $F$ is of weight $9|\mathcal{T}|-6 q$. All edges $\left(a_{i}^{r, s}, \bar{a}_{i}^{r, s}\right)$ are in $F$ since this is the only edge incident to $\bar{a}_{i}^{r, s}$. Since $F$ has only trees of order 4 , just one of the edges $\left(a_{i}^{r, s}, a_{i}^{r}\right)$ and $\left(a_{i}^{r, s}, a_{i}^{s}\right)$ is in $F$. Therefore in the family $\left\{a_{i}^{\ell}\right\}_{\ell}$ there exists exactly one vertex that is not incident to a linking gadget in $F$. Since forest $F$ is of weight $9|\mathcal{T}|-6 q$, it contains $3|\mathcal{T}|-2 q$ trees. Since for any element of $A, B, C$ the number of linking gadgets associated is equal to the number of occurrences of this element in $\mathcal{T}$ minus 1 , there are $3|\mathcal{T}|-3 q$ linking gadgets in $G(I)$. It means that there are $q$ trees of $F$ which do not cover any linking gadget. Each of these remaining trees therefore covers a star gadget of the form $a_{i}^{\ell}, b_{j}^{\ell}, c_{k}^{\ell}, d^{\ell}$. From these star gadgets we extract a collection of $q$ triplets $\left(a_{i}, b_{j}, c_{k}\right)$, which is a valid solution in $I$ since every triplet appears in $\mathcal{T}$ by construction and every element of the triplets appears in exactly one occurrence.

Since 3DM is $N P$-hard even on planar instances [5] and the previous reduction preserves planarity, we can restrict Min CF (4) to planar bipartite graphs of maximum degree 3. This completes the proof for $p=4$.

In order to prove the theorem for $p>4$, we also construct a reduction from 3DM to Min $\operatorname{CF}(p)$. Given an instance $I$ of 3 DM , the graph $G^{\prime}(I)$ instance of $\operatorname{Min} \operatorname{CF}(p)$ is obtained from the previous $G(I)$ by replacing the edges between $a_{i}^{\ell}$ and $d^{\ell}, \underline{b}_{j}^{r, s}$ and $\bar{b}_{j}^{r, s}$, and $\underline{c}_{j}^{r, s}$ and $\bar{c}_{j}^{r, s}$ by a path of length $p-3$ (see Figure 2). More precisely, we add to $G(I)$ the following vertices and edges:

$$
V_{T}(p)=\bigcup_{a_{i}^{\ell} \in V_{T}}\left\{a_{i}^{\ell}(1), \ldots, a_{i}^{\ell}(p-4)\right\}
$$

$$
\begin{gathered}
E_{T}(p)=\bigcup_{a_{i}^{\ell} \in V_{T}}\left\{\left(a_{i}^{\ell}, a_{i}^{\ell}(1)\right),\left(a_{i}^{\ell}(1), a_{i}^{\ell}(2)\right), \ldots,\left(a_{i}^{\ell}(p-4), d^{\ell}\right)\right\} \\
V_{B}(p)=\bigcup_{b_{j}^{r, s} \in V_{B}}\left\{b_{j}^{r, s}(1), \ldots, b_{j}^{r, s}(p-4)\right\} \\
E_{B}(p)=\bigcup_{b_{j}^{r, s} \in V_{A}}\left\{\left(\underline{j}_{j}^{r, s}, b_{j}^{r, s}(1)\right),\left(b_{j}^{r, s}(1), b_{j}^{r, s}(2)\right), \ldots,\left(b_{j}^{r, s}(p-4), \bar{b}_{j}^{r, s}\right)\right\}
\end{gathered}
$$

The sets $V_{C}(p)$ and $E_{C}(p)$ are defined analogously.


Figure 2: Graph $G^{\prime}(I)$ derived from instance $I$ described in Figure 1. Double lines represent paths of length $p-3$.

We show next that $I$ contains a perfect matching if and only if $G^{\prime}(I)$ contains a forest of weight at most $(p-1)(3|\mathcal{T}|-2 q)$ where every tree is of order at least $p$. Given a perfect matching $\mathcal{M}$ on the instance $I$, in addition to the edges chosen for such a solution in $G(I)$ for $p=4$, we take every edge in $E_{B}(p), E_{C}(p)$ and $E_{T}(p)$. Thus we obtain a forest with the same number of trees $3|T|-2 q$ and all these trees are of order at least $p$. Conversely, given a forest $F$ of weight $(p-1)(3|\mathcal{T}|-2 q)$ on the instance $G^{\prime}(I)$, we can consider the trees which are not covering vertices of linking gadgets. Such a collection of trees describes a perfect matching of size $q$ on $I$, which is therefore optimal.

## 3.2 $\boldsymbol{A P} \boldsymbol{X}$-hardness for $p \geq 3$, unweighted bipartite graphs

Applying $L$-reduction from Max $3 \mathrm{DM}_{2}$, the following result can be obtained.

Theorem 2. For any $p \geq 3$, $\operatorname{Min} \operatorname{CF}(p)$ is APX-hard. Moreover, $\operatorname{Min} \operatorname{CF}(p)$ for $p \geq 3$ is not $\frac{95(8 p-7)+1}{95(8 p-7)}$-approximable on bipartite graphs of maximum degree 3 , unless $P=N P$.

Proof. We first give a detailed argument for $p=4$. For this, we construct an $L$-reduction from Max $3 \mathrm{DM}_{2}$ to Min $\mathrm{CF}(4)$. For any instance $I$ of Max $3 \mathrm{DM}_{2}$ we construct $G(I)$ as in Theorem 1. Since any element of $A, B, C$ appears exactly twice in $T$, for any $a_{i}$ there exist only one vertex of the form $a_{i}^{r, s}$ and one of $\bar{a}_{i}^{r, s}$; locally in this proof we rename them as $a_{i}$ and $\overline{a_{i}}$. We shall use $b_{j}, \underline{b}_{j}, \bar{b}_{j}, c_{k}, \underline{c}_{k}, \bar{c}_{k}$ in an analogous meaning. Note that the number of vertices in $G(I)$ is exactly $16 q$.

Let $\mathcal{M}$ be any solution, say of size $t$, on $I$; that is, $\mathcal{M}$ is a matching but not necessarily a perfect matching. From $\mathcal{M}$ we construct a forest $F$ of $G(I)$. Let

$$
\begin{gathered}
F_{T}=\bigcup_{x_{\ell}=\left(a_{i}, b_{j}, c_{k}\right) \in \mathcal{M}}\left\{\left(a_{i}^{\ell}, d^{\ell}\right),\left(b_{j}^{\ell}, d^{\ell}\right),\left(c_{k}^{\ell}, d^{\ell}\right)\right\} \\
F_{R}=\bigcup_{x_{\ell}=\left(a_{i}, b_{j}, c_{k}\right) \notin \mathcal{M}}\left\{\left(a_{i}^{\ell}, d^{\ell}\right),\left(a_{i}^{\ell}, a_{i}\right),\left(b_{j}^{\ell}, b_{j}\right),\left(c_{k}^{\ell}, c_{k}\right)\right\} \\
F_{A}=\bigcup_{a_{i} \in A}\left\{\left(a_{i}, \bar{a}_{i}\right)\right\}, \quad F_{B}=\bigcup_{b_{j} \in B}\left\{\left(b_{j}, \underline{b}_{j}\right),\left(\underline{b}_{j}, \bar{b}_{j}\right)\right\}, \quad F_{C}=\bigcup_{c_{k} \in C}\left\{\left(c_{k}, \underline{c}_{k}\right),\left(\underline{c}_{k}, \bar{c}_{k}\right)\right\} .
\end{gathered}
$$

Then, let $F=F_{T} \cup F_{R} \cup F_{A} \cup F_{B} \cup F_{C}$. We note that the trees occurring in $F$ are of orders at least 4. In fact, $F$ contains $t$ trees on the star gadgets corresponding to $\mathcal{M}$, and one tree of order four on each linking gadget. Therefore $F$ contains $t+3 q$ trees, and then the weight of $F$ is $13 q-t$. Thus, opt $(G(I)) \leq 13 q-o p t(I)$.

Let $F$ be a forest of weight $v$, covering $G(I)$ with trees of orders at least 4 . We consider, as solution $\mathcal{M}$ for $I$, the triplets corresponding to those tree components of $F$ which have order 4 and so cover star gadgets. Since $F$ contains $16 q-v$ trees and at most $3 q$ of these trees cover linking gadgets, we obtain that $\operatorname{opt}(I) \geq|\mathcal{M}| \geq 13 q-v$. Thus, $v(\mathcal{M}, I) \geq 9 q-v(F, G(I))$ and $\operatorname{opt}(G(I))=13 q-o p t(I)$.

Since $I$ contains $2 q$ triplets and each of them shares an element with at most three triplets, we have $\operatorname{opt}(I) \geq \frac{q}{2}$ (a lower bound valid for every inclusion-wise maximal solution), and consequently $\operatorname{opt}(G(I)) \leq 25 o p t(I)$. Moreover, by the observations above, we also obtain $o p t(I)+\operatorname{opt}(G(I))=13 q \leq v(F, G(I))+v(\mathcal{M}, I)$ and so $\operatorname{opt}(I)-v(\mathcal{M}, I) \leq v(F, G(I))-$ opt $(G(I))$.

It is $N P$-hard to achieve an approximation factor of $\frac{95}{94}$ for MAX $3 \mathrm{DM}_{2}[3]$, and consequently Min $\mathrm{CF}(4)$ is not $\frac{2376}{2375}$-approximable if $P \neq N P$.

The proof for $p>4$ applies the same modifications as it has been done in Theorem 1 for $N P$-hardness, replacing some edges by paths of length $p-3$.

For $p=3$, we essentially replace the star gadget by a path of length $3\left(b_{j}^{\ell}, a_{i}^{\ell}, c_{k}^{\ell}\right)$, by removing the vertex $d^{\ell}$ and we also remove the vertices $\bar{b}_{j}, \bar{c}_{j}$ (see an illustration in Figure 3 with a particular instance $I$ of MAx $3 \mathrm{DM}_{2}$ ).


Figure 3: Graph obtained from instance $I=(A, B, C, \mathcal{T})$ of Max $3 \mathrm{DM}_{2}$ with $A=\left\{a_{1}, a_{2}\right\}$, $B=\left\{b_{1}, b_{2}\right\}, C=\left\{c_{1}, c_{2}\right\}$, and $\mathcal{T}=\left\{T_{1}, \ldots, T_{4}\right\}$ with $T_{1}=\left(a_{1}, b_{1}, c_{1}\right), T_{2}=\left(a_{1}, b_{2}, c_{2}\right)$, $T_{3}=\left(a_{2}, b_{1}, c_{2}\right), T_{4}=\left(a_{2}, b_{2}, c_{1}\right)$.

## 4 Efficient algorithms

### 4.1 Exact algorithms for bounded treewidth graphs

In this section we present an efficient algorithm solving $\operatorname{Min} \operatorname{WCF}(p)$ on graph classes of bounded treewidth. For undefined details on tree decomposition we refer to [11].

Theorem 3. The problem $\operatorname{Min} \operatorname{WCF}(p)$ on graphs with treewidth bounded by $k$ can be solved in time $O\left(k^{c k} p^{2 k+2} n\right)$ for some constant $c$, where $n$ is the number of vertices.

Remark 4. The time bound is polynomial in $n$ whenever $k(\log k+\log p)=O(\log n)$ and in particular if $k=O\left(\frac{\log n}{\log \log n}\right)$ and $p=O(\log n)$.

Proof. Let $G$ be a graph of treewidth $k$, and $\left(T_{G},\left\{X \mid x \in V\left(T_{G}\right)\right\}\right)$ a tree decomposition of $G$ with width $k$. As a notational convention, the vertex subset of $G$ associated with node $x$ of $T_{G}$ is denoted by the corresponding upper-case letter $X$, and we shall use the same subscript for them where necessary. We view $T_{G}$ as a rooted tree. By assumption, the set $X$ associated with any node $x$ of $T_{G}$ contains at most $k+1$ vertices of $V(G)$. We shall assume further that every node $x$ is of one of the following types:

- a start node that has no children (a leaf in $T_{G}$ ),
- a join node that has two children $x_{1}, x_{2}$ and $X_{1}=X_{2}=X$,
- an introduce node that has one child $x_{1}$ and $X_{1}=X \backslash\{v\}$ for some $v \in V(G)$,
- a forget node that has one child $x_{1}$ and $X_{1}=X \cup\{v\}$ for some $v \in V(G)$.

This is called a "nice tree decomposition" in the literature, see pp. 149-150 of [11]. As it is well known, every graph admits a tree decomposition of size $O(n)$ satisfying these conditions. It is also easy to see that any tree decomposition of width $k$ can be modified to one with the properties above in $O(k n)$ time.

For any node $x$ of $T_{G}$, let $T_{G}(x)$ be the rooted subtree of $T_{G}$ containing exactly the node $x$ and its descendants. The vertices of $G$ which appear in the sets associated with the nodes of $T_{G}(x)$ define a subgraph of $G$, denoted by $G(x)$. Remark that $T_{G}(x)$ is a tree decomposition of $G(x)$.

Consider any node $x$ of $T_{G}$. For each spanning forest $F$ (possibly with many vertices of degree zero in $F$ ) of the subgraph $G[X]$ induced by $X$ in $G$, we consider all partitions $\mathcal{P}=\left(C_{1}, \ldots, C_{\ell}\right)$ of $X$ and all $\ell$-tuples $I=\left(i_{1}, \ldots, i_{\ell}\right) \in\{1, \ldots, p\}^{\ell}$ such that the following conditions are met:

- if $v_{i}, v_{j} \in X$ are in the same tree component of $F$, then $v_{i}, v_{j}$ belong to the same partition class $C_{r}$,
- if $\left|C_{r}\right| \leq p$, then $\left|C_{r}\right| \leq i_{r} \leq p$, and otherwise $i_{r}=p$.

Let $f(x, F, \mathcal{P}, I)$ be defined as the value of a minimum-weight spanning forest of $G(x)$ in which $X$ induces precisely $F$, and in which two vertices belong to the same tree component if and only if they are in the same class of $\mathcal{P}$.
Lemma 5. For any vertex $x$ of $T_{G}$, for any forest $F$ over the subgraph $G[X]$, for any partition $\mathcal{P}$ of $X$, and for any $\ell$-tuple $I$ of integers satisfying the conditions above, we can determine $f(x, F, \mathcal{P}, I)$ in $O\left(k^{c k} p^{2 k+2}\right)$ time if the corresponding values $f$ are available for the child(ren) of $x$.

Proof. One has to consider each type of nodes separately.
If $x$ is a start node, then the entire set of values $f(x, F, \mathcal{P}, I)$ can be determined for $x$ in $O(k)$ time. Indeed, in this case $G(x)$ has at most $k+1$ vertices, the classes of $\mathcal{P}$ are exactly the vertex sets of the connected components of $F$ which can be found in $O(k)$ time, and $i_{r}=\min \left\{\left|C_{r}\right|, p\right\}$ holds for every class $C_{r}$ of $\mathcal{P}$.

If $x$ is a join node, then $\mathcal{P}$ has to be generated from two finer partitions over $X$, one for $X_{1}$ and the other for $X_{2}$. We can do this by artificially completing $F$ with sets $E_{b l u e}, E_{\text {red }}$ of blue and red edges (not necessarily edges of $G$ ) to obtain a forest, each tree component of which spans a class of $\mathcal{P}$. We then consider the forests $F_{b l u e}$ and $F_{r e d}$ which are respectively the forests defined by $F \cup E_{b l u e}$ and $F \cup E_{\text {red }}$. The tree components of these forests define the partitions $\mathcal{P}_{\text {blue }}$ and $\mathcal{P}_{\text {red }}$ over the vertices of $X$. The vectors $I_{1}$ and $I_{2}$ are such that, for every partition class $C_{j}$ of $\mathcal{P}, i_{j}$ is equal to the minimum of $p$ and the sum of the values of the blue and red classes included in $C_{j}$ minus $\left|C_{j}\right|$. Then,

$$
f(x, F, \mathcal{P}, I)=\min _{F_{\text {blue }}, F_{\text {red }}, I_{1}, I_{2}} f\left(x_{1}, F, \mathcal{P}_{\text {blue }}, I_{1}\right)+f\left(x_{2}, F, \mathcal{P}_{\text {red }}, I_{2}\right)-w(F)
$$

The number of relevant blue-red forests is not larger than the number $B_{k+1}$ of partitions of a $(k+1)$-element set. For each of them, processing all combinations of $O\left(p^{k+1}\right)$ records at $X_{1}$ and the same amount of data at $X_{2}$ may need $O\left(p^{2 k+2}\right)$ time. This yields the upper bound $B_{k+1} \cdot O\left(p^{2 k+2}\right)$ altogether.

If $x$ is an introduce node, let $E_{v}$ be the set of edges incident to $v$ in $F$. Let $F_{v}=F-v$ be the forest without the vertex $v$; the removal of $v$ may have divided a tree component of
$F$ into several components. We consider all possible partitions $\mathcal{P}_{v}$ over the elements of $X_{1}$, such that every class of $\mathcal{P}$ which does not contain $v$ is a class in $\mathcal{P}_{v}$, too; and the class $C_{s}$ which contains $v$ is divided into $q$ classes where $q$ is the number of components created by the removal of $v$ and such that every remaining component is in a different class. The $i$-values associated with those $q$ classes must sum up to $i_{s}-1$. Then

$$
f(x, F, \mathcal{P}, I)=\min _{\mathcal{P}_{v}, I_{v}} f\left(x_{1}, F_{v}, \mathcal{P}_{v}, I_{v}\right)+w\left(E_{v}\right)
$$

This step needs at most $B_{k} \cdot O\left(p^{k+1}\right)$ time.
If $x$ is a forget node, let $F_{v}$ be obtained through the addition of $v$ to $F$ such that all components of $F_{v}$ which are adjacent to $v$ in $F$ belong to the same class of $\mathcal{P}$. Assume further that $\mathcal{P}_{v}$ has the same classes as $\mathcal{P}$ except that we add $v$ to the class which contains the vertices it is connected to in $F_{v}$; if there is none, then we add the class $\{v\}$ to $\mathcal{P}$ as a singleton. Now $I_{v}$ is either $I$ if $v$ is not alone in its class in $\mathcal{P}_{v}$, or $I \cup\{p\}$. Then

$$
f(x, F, \mathcal{P},)=\min _{F_{v}} f\left(x_{1}, F_{v}, \mathcal{P}_{v}, I_{v}\right)
$$

This needs at most $B_{k} \cdot O\left(p^{k+1}\right)$ time.
To conclude the proof of Theorem 3 , we traverse $T_{G}$ in postorder, and compute $f(x, F, \mathcal{P}, I)$ for all possible choices of node $x$, spanning forest $F$ of $G[X]$, partition $\mathcal{P}$ of $X$, and sequence $I$ satisfying the conditions given at the beginning of the proof. Then the solution of $\operatorname{Min} \operatorname{WCF}(p)$ on $G$ is equal to the smallest value of $f$ occurring at the root of $T_{G}$ for a sequence $I=(p, \ldots, p)$ of any length.

The overall upper bound is obtained by the facts that for any $k$, the number of choices for $\mathcal{P}$ is bounded above by $B_{k+1}$, the number of vertex-labeled trees of order $t$ is $t^{t-2}$ (this puts a bound on the choices for $F$ ), and the number of sequences $I$ at $x$ is at most $p^{k+1}$. The most time-consuming step is to compute $f$ at a join node; it can be organized by taking all combinations of the values stored at $x_{1}$ and $x_{2}$.

Remark 6. The same method can be applied to solve the more general problem where, instead of a uniform condition $p$ for the constrained forest, the input graph $G=(V, E)$ on vertex set $V=\left\{v_{1}, \ldots, v_{n}\right\}$ is given together with a vector $\left(p_{1}, \ldots, p_{n}\right)$ of natural numbers $p_{i} \leq p$, and it is required that each $v_{i}$ be contained in a tree component which has at least $p_{i}$ vertices in the spanning forest. The steps of the algorithm above can be adjusted for this variant.

### 4.2 Ptas for planar graphs

Theorem 7. For $p=O\left(\frac{\log n}{\log \log n}\right)$, $\operatorname{Min} \operatorname{WCF}(p)$ on planar graphs admits a ptas.
Proof. Given a planar embedding of an input graph, we consider the set of the vertices which are on the exterior face, they will be called level 1 vertices. By induction we define level $k$ as the vertices which are on the exterior face when we have removed the vertices of levels smaller than $k$ [1]. A planar embedding is $k$-level if it has no nodes of level greater than $k$. If a planar graph is $k$-level, it has a $k$-outerplanar embedding.

If we want to achieve an approximation within $1+\epsilon$, let us consider $k=\left\lceil\frac{4(p-1)}{\epsilon}\right\rceil$. Let $X_{t}$ be the set of vertices of level $t$ and let $H_{i}, 0 \leq i \leq k-1$, be the graph obtained from $G$ by deleting all edges between the vertices $X_{t}$ and $X_{t+1}$ for all $t$ such that $t \equiv i(\bmod k)$.

The set of vertices $\bigcup_{t+1 \leq j \leq t+k} X_{j}$, for $t \equiv i(\bmod k)$ is therefore a component in $H_{i}$ since we have deleted all edges from this set of vertices to the other vertices of the graph. Hence, the subgraph containing exactly $\bigcup_{t+1 \leq j \leq t+k} X_{j}$ is $k$-outerplanar, and so is $H_{i}$, too.

Since $H_{i}$ is $k$-outerplanar, it has treewidth at most $3 k-1$ [2]. On applying Theorem 3 and Remark 6 , we can efficiently determine an optimal forest $S_{i}$ on $H_{i}$; the condition on $p$ remains true on vertices no incident with edges deleted, and is changed to zero on the boundary vertices of $H_{i}$, that means vertices in $X_{t+1}, X_{t+k}$ with $t \equiv i(\bmod k)$. We complete the solution $S_{i}$ obtained on $H_{i}$ into a feasible solution $F_{i}$ on $G$ by choosing the useful edges of smallest cost greedily within the edges of $G$ until every tree is of order at least $p$.

We are going to prove that the best forest among $F_{0}, \ldots, F_{k-1}$ is an $(1+\epsilon)$-approximation of the optimal value on $G$. We will use two lemmas and some notation for the proof. For any subset of edges that forms a tree $T$ of $G$, let $w_{\max }(T)$ denote the maximum weight of the edges of $T$ :

$$
w_{\max }(T)=\max _{e \in E(T)} w(e) .
$$

Let $g: V \rightarrow \mathbb{R}$ be defined as

$$
g(v)=\min _{T \text { tree, }|V(T)|=p, v \in V(T)} w_{\text {max }}(T) .
$$

Lemma 8. Let $F$ be a spanning forest of $G$, and $V^{\prime}=\left\{v_{1}, \ldots, v_{r}\right\}$ a set of vertices such that any tree $T$ of $F$ which is not of order at least $p$ contains a vertex $v_{i} \in V^{\prime}$. Then we can construct a forest of $G$ in which every tree component has order at least $p$ and the total weight is at most $w(F)+\sum_{v \in V^{\prime}} g(v)$.

Proof. For short, let us call a tree component small if it has fewer than $p$ vertices, and call it large otherwise. By induction we suppose that the assertion is true when we have at most $r$ small trees in the forest (the case $r=0$ is trivial). Let $F$ be a forest with $r+1$ small components, such that some vertices $v_{1}, \ldots, v_{r+1}$ belong to distinct small trees; and let $T_{r+1}$ be the small tree in $F$ containing $v_{r+1}$. Let $A_{r+1}$ be the tree which realizes the optimum of $g\left(v_{r+1}\right)$. Since $T_{r+1}$ is small and $A_{r+1}$ is large, and $v_{r+1}$ belongs to both trees, there exists an edge of $A_{r+1}$ which has exactly one endpoint in $T_{r+1}$. We add this edge to forest $F$ in order to obtain $F^{\prime}$. Forest $F^{\prime}$ is of weight at most $w(F)+g\left(v_{r+1}\right)$ since the added edge belongs to $A_{r+1}$. Thus, in $F^{\prime}$, tree $T_{r+1}$ is linked to another tree which implies that the number of small trees has decreased at least by one (either by joining two small trees to become just one component of any size, or by joining $T_{r+1}$ to a large component). Moreover $v_{1}, \ldots, v_{r}$ are sufficient to cover the remaining small trees. Therefore we can construct a covering forest in which each tree is of order at least $p$, and which has weight at most

$$
\begin{aligned}
w\left(F^{\prime}\right)+\sum_{v \in V^{\prime} \backslash\left\{v_{r}+1\right\}} g(v) & \leq w(F)+g\left(v_{r+1}\right)+\sum_{v \in V^{\prime} \backslash\left\{v_{r}+1\right\}} g(v) \\
& =w(F)+\sum_{v \in V^{\prime}} g(v)
\end{aligned}
$$

Lemma 9. $\sum_{v \in V} g(v) \leq 2(p-1) \operatorname{opt}(G)$.

Proof. Suppose, for simplicity, that all weights are distinct. (This can be achieved by a slight modification of the weights, by adding $\varepsilon, 2 \varepsilon, 3 \varepsilon, \ldots$ to them with a very small $\varepsilon$ such that the increase of sum on each tree of order $p$ is smaller than the smallest positive edge-weight difference.)

Let $F^{*}$ be an optimal solution on $G$. For $v \in V$, let $T^{*}(v) \subset F^{*}$ be a subtree of order exactly $p$ which contains $v$ and minimizes the largest weight of its edges. Let $e^{*}(v)$ be the edge of maximum weight in $T^{*}(v)$. Remark that $w\left(e^{*}(v)\right) \geq g(v)$. For any edge $e \in F^{*}$, let

$$
T(e)=\bigcup_{e^{*}(v)=e} T^{*}(v)
$$

By definition, $T(e)$ is a tree since it is included in $F^{*}$. Moreover, $w_{\max }(T(e))=w(e)$ if $T(e)$ is not empty. If $|V(T(e))| \geq 2 p-1$ then, if we remove $e$ from $T(e)$, we obtain two trees and at least one of them, say $T^{\prime}$, is large. Therefore $T^{\prime}$ is in $F^{*}$ and $w_{\max }\left(T^{\prime}\right)<w_{\max }(T(e))$ which is a contradiction with $e^{*}(v)=e$ for any vertex in $T^{\prime}$. Thus $\left|\left(e^{*}\right)^{-1}(e)\right| \leq 2(p-1)$.

Consequently,

$$
\begin{aligned}
\sum_{v \in V} g(v) & \leq \sum_{v \in V} w\left(e^{*}(v)\right) \\
& =\sum_{e \in F^{*}}\left|\left(e^{*}\right)^{-1}(e)\right| w(e) \\
& \leq 2(p-1) \sum_{e \in F^{*}} w(e) \\
& =2(p-1) \operatorname{opt}(G)
\end{aligned}
$$

Now we are in a position to complete the proof the theorem. Let $V_{i}=\bigcup_{t \equiv i(\bmod k)}\left(X_{t+1} \cup\right.$ $X_{t+k}$ ). By Lemma 8, starting from $S_{i}$ we can construct a forest $F_{i}$ satisfying the property that every tree component is as large as required, moreover

$$
w\left(S_{i}\right)+\sum_{v \in V_{i}} g(v) \geq w\left(F_{i}\right)
$$

Since $S_{i}$ is an optimal solution of a relaxed problem, we have $w\left(S_{i}\right) \leq o p t(G)$ and so

$$
\operatorname{opt}(G)+\sum_{v \in V_{i}} g(v) \geq w\left(F_{i}\right)
$$

Moreover, using Lemma 9 we obtain the following inequality:

$$
\sum_{1 \leq i \leq k} g\left(V_{i}\right)=2 \sum_{v \in V} g(v) \leq 4(p-1) \operatorname{opt}(G)
$$

Hence, there exists an $i_{0}$ such that $\sum_{v \in V_{i_{0}}} g(v) \leq \frac{4(p-1)}{k} \operatorname{opt}(G)$, which implies

$$
\min _{1 \leq i \leq k} w\left(F_{i}\right) \leq w\left(F_{i_{0}}\right) \leq\left(1+\frac{4(p-1)}{k}\right) \operatorname{opt}(G) \leq(1+\epsilon) \operatorname{opt}(G)
$$

The overall running time of the algorithm is $k$ times what we need for graphs of treewidth at most $k$, that is $O\left(\left(\frac{4 p}{\varepsilon}\right)^{d p / \varepsilon} n\right)$, where $d$ is a constant. Hence, since $p=O\left(\frac{\log n}{\log \log n}\right)$ we have a ptas.

### 4.3 Speeding up the PTAS

Theorem 10. G planar, for any $\epsilon>0$, it exists an $(1+\epsilon)$-approximation which runs in $O()$.
It appears that in the case of planar graph with bounded treewidth $k$ it exists a branch decomposition which is of width at most $k$ [16]. A branch width decomposition of $G$ is a couple $T, \mu$ such that $T$ is a tree of maximum degree 3 and $\mu$ is a bijection from the edges of $G$ to the leaves of $T$. For any edge $e$ of $T$, the removal of $e$ would create two trees $T_{1}(e)$ and $T_{2}(e)$, and let $E_{1}(e)$ and $E_{2}(e)$ be a partition of the edge of $G$ naturally defined by $\mu$, i.e. $e^{\prime} \in E_{i}(e)$ if and only if $\mu\left(e^{\prime}\right) \in T_{i}(e)$ for $i \in\{1,2\}$. Denote by $\operatorname{mid}(e)$ the set of vertices which are endpoint of an edge of $E_{1}(e)$ and also of an edge from $E_{2}(e)$. The maximum size of $\operatorname{mid}(e)$ define the width of the branch decomposition.

Since we consider planar graphs, given a planar embedding, the result of Dorn et. al. [6] enables us to obtain a sphere-cut decomposition of width at most $\frac{3}{2} \times k$ from this branchwidth decomposition. A sphere-cut decomposition is a branch decomposition such that for any edge $e$ of the tree of the branchwidth decomposition the vertices of $\operatorname{mid}(e)$ we can add a cycle on $\operatorname{mid}(e)$ which does not cross any edge of $G$ and on the planar embeding every edges of $E_{1}(e)$ is within this cycle, every edge of $E_{2}(e)$ is out of the cycle.


Figure 4: The straight lines are edges of $E_{1}(e)$ and the dashed lines are edges of $E_{2}(e)$

Therefore consider an internal nodes in the tree of the sphere-cut decomposition, it will be of degree 3 and the cycles associated to each of its incident vertices will looks like this:


Figure 5: the dashed and straight lines forms the cycle of the father edge, the dashed and the double lines one child edge and the straight and the double lines the other child edge

To any vertex $v \in T$, denote $C(v)=\operatorname{mid}\left(e_{1}\right) \cup \operatorname{mid}\left(e_{2}\right) \cup \operatorname{mid}\left(e_{3}\right)$ where $e_{1, e} 2, e_{3}$ are the three incident edges of $v$ in $T$. The tree decomposition use the same tree as the sphere-cut decomposition. The set associated to each node of the tree decomposition is defined as follows:

- to any internal node $v \in T$ we associate the set of vertices $C(v),|X(v)| \leq \frac{3}{2} \times k$
- to any leaf $v \in T$ we associate the endpoints of the edge $\mu^{-1}(v), X(v)=2$

If two vertices are connected they appears in the leaf-node corresponding to the edge which connect them. For any $v$, the edges where $v \in \operatorname{mid}(e)$ form only one component in the tree since it is a branch decomposition, the vertices incident to these edges form also only one component. Therefore it is a tree decomposition on the graph $G$ and it width is at most $\frac{3}{2} \times k$. Since in the planar embedding we can add two cycle such that every vertices is at least in one cycle and this cylce does not cross any edge of $G$, this limit the number of possible partition between the vertices $C(v)$ that we could obtain in a solution of $\min C F(p)$. Therefore instead of bounding the cardinality of possibility by $B_{k}$ when running the polynomial algorithm for bounded treewidth, we can bound it by $\left(C_{k}\right)^{2}$ since partition can be described by a planar graph such that every edges has to be in the two cycles described before. This cycle are of size at most $k$ so there is at most $C_{k}$ possible planar graph on each of these cycle, resulting a bound equal to $\left(C_{k}\right)^{2}$. Therefore the minimal constrained forest on a planar graph with bounded treewidth $k$ can be obtained in $O\left(4^{k} \times p^{2 k+2}\right)$. This improvement speed up our previous algorithm to obtain a complexity equal to $O\left((4 p)^{8 p / \epsilon} n\right)$.

## 5 Conclusion

In this paper we considered a natural generalization, $\operatorname{Min} \operatorname{WCF}(p)$, of the fundamental problem of minimum-weight edge cover in graphs. The task is to find, for an edge-weighted input graph, a spanning forest in which every tree component has at least $p$ vertices and the sum of weights is as small as possible.

On the negative side, up to now $\operatorname{Min} \operatorname{CF}(3)$ and $\operatorname{Min} \operatorname{WCF}(p)$ for $p \geq 4$ were proved $N P$-hard. We extended this, by proving that even the unweighted Min $\operatorname{CF}(p)$ is $A P X$-hard for all $p \geq 3$. This result remains valid for restricted classes of graphs (e.g., bipartite graphs of maximum degree 3 ).

We also considered the positive side to some extent, by designing an exact algorithm for graph classes where treewidth may slowly tend to infinity, and an efficient approximation for planar graphs where $p$ may slowly grow with the number of vertices. The latter result is of interest since Min $\mathrm{CF}(p)$ is $N P$-hard also on planar instances with $p=3$ and assuming maximum degree 3 .

Perhaps the most challenging problem that remains open is to design a $(2-c)$-approximation for $\operatorname{Min} \operatorname{WCF}(p)$ for some constant $c>0$ and $p \geq 3$. For the unweighted case, however, it is easy to get a $\left(1+\frac{1}{p-1}\right)$-approximation in linear time.

## References

[1] B. S. Baker. Approximation algorithms for $N P$-complete problems on planar graphs. Journal of the Association for Computing Machinery, 41(1):153-180, 1994.
[2] H. L. Bodlaender. A partial k-arboretum of graphs with bounded treewidth. Theoretical Computer Science, 209:1-45, 1998.
[3] M. Chlebik and J. Clebikova. Complexity of approximating bounded variants of optimizationproblems. Theoretical Computer Science, 354:320-338, 2006.
[4] B. Courcelle. The monadic second-order logic of graphs. III. Tree-decompositions, minors and complexity issues. RAIRO Informatique Théorique Appliquée, 26:257-286, 1992.
[5] M. E. Dyer and A. M. Frieze. Planar $3 D M$ is NP-complete. Journal of Algorithms, 7:174-184, 1986.
[6] F. Dorn, E. Penninkx, H. L. Bodlaender, and F. Fomin. Efficient Exact Algorithms on Planar Graphs: Exploiting Sphere Cut Decompositions. Technical Report UU-CS-2006006.
[7] J. Edmonds and E. L. Johnson. Matching, Euler tours and the Chinese postman. Mathematical Programming, 5:88-124, 1973.
[8] T. Gallai. Über extreme Punkt- und Kantenmengen. Annales Universitatis Scientiarum Budapestinensis de Rolando Eötvös Nominatae, Sectio Mathematica, 2:133-138, 1959.
[9] M. X. Goemans and D. P. Williamson. A general approximation technique for constrained forest problems. SIAM Journal of Computing, 24:296-317, 1995.
[10] C. Imielinska, B. Kalantari and L. Khachiyan. A greedy heristic for a minmum-weight forest problem. Operations Research Letters, 14:65-71, 1993.
[11] T. Kloks. Treewidth. computations and approximations. Lecture Notes in Computer Science, 842, 1994.
[12] M. Laszlo and S. Mukherjee. Another greedy heuristic for the constrained forest problem. Operations Research Letters, 33:629-633, 2005.
[13] M. Laszlo and S. Mukherjee. A class of heuristics for the constrained forest problem. Discrete Applied Mathematics, 154:6-14, 2006.
[14] J. Monnot and S. Toulouse. The path partition problem and related problems in bipartite graphs. Operations Research Letters, 35:677-684, 2007.
[15] C. Papadimitriou and M. Yannakakis. Optimization, approximation and complexity classes. Journal of Computer and System Science, 43:425-440, 1991.
[16] N. Robertson and P. Seymour. Graph minors X. Obstructions to treedecomposition. Journal of Combinatorial Theory Series B, 52 (1991), pp. 153-190.
[17] A. Schrijver. Combinatorial Optimization. Springer, 2003.
[18] Zs. Tuza. Extensions of Gallai's graph covering theorems for uniform hypergraphs. Journal of Combinatorial Theory Ser. B, 52:92-96, 1991.


[^0]:    * Research supported in part by the Hungarian Scientific Research Fund, OTKA grant T-049613.

