Covering a Laminar Family by Leaf to Leaf Links

Yael Maduel*
The Open University of Israel
yael.maduel@gmail.com

Zeev Nutov
The Open University of Israel
nutov@openu.ac.il

Abstract

The Tree Augmentation Problem (TAP) is: given a tree $T = (V, \mathcal{E})$ and a set E of edges (called *links*) on V disjoint to $\mathcal{E}$, find a minimum size edge subset $F \subseteq E$ so that $T + F$ is 2-edge-connected. TAP is equivalent to the problem of finding a minimum size edge-cover $F \subseteq E$ of a laminar set-family. We consider the restriction LL-TAP of TAP to instances when every link in E connects two leaves of T. The best approximation ratio for TAP is 3/2 [3, 4, 5], and no better ratio was known for LL-TAP. All the previous approximation algorithms that achieve a ratio better than 2 for TAP, or even for LL-TAP, were quite involved.

For LL-TAP we obtain the following approximation ratios: 17/12 for general trees, 11/8 for trees of height 3, and 4/3 for trees of height 2. We also give a very simple 3/2-approximation algorithm (for general trees) and prove that it computes a solution of size at most $\min\{\frac{3}{2}t, \frac{5}{3}t^*\}$, where t is the minimum size of an edge-cover of the leaves, and t^* is the optimal value of the natural LP-relaxation for the problem of covering the leaf edges only. This provides the first evidence that the integrality gap of a natural LP-relaxation for LL-TAP is less than 2.

1 Introduction

We consider the following problem:

Tree Augmentation Problem (TAP)
Instance: A tree $T = (V, \mathcal{E})$ and a set E of edges (called *links*) on V disjoint to $\mathcal{E}$.
Objective: Find a min-size edge subset $F \subseteq E$ so that $T + F$ is 2-edge-connected.

*Part of this work was done as a part of author's M.Sc. Thesis at The Open University of Israel.

TAP is equivalent to the problem of finding a minimum size edge-cover of a laminar family; namely, given a graph $G = (V, E)$ and a laminar family $\mathcal{E}$ on V, we seek a minimum size edge-set $F \subseteq E$ so that for every $S \in \mathcal{E}$ there exists $uv \in F$ with $u \in S$ and $v \notin S$. Laminar families play an important role in network design problems, c.f. [9]. See also surveys in [8, 11] for various network design problems and applications of laminar families in the analysis of algorithms for such problems.

Fredrickson and Jájá [6] showed that TAP is NP-hard even for trees of height 2, and gave a 2-approximation algorithm for the more general weighted version of TAP, when links have weights and we seek a minimum weight augmenting edge set $F \subseteq E$. Achieving a ratio better than 2 for (unweighted) TAP was posed as a major open problem in graph connectivity in the survey by Khuller [10]. This open question was resolved by Nagamochi [12] that gave a $(1.875 + \varepsilon)$-approximation scheme for TAP. The currently best approximation ratio known for TAP is 3/2, by Even, Feldman, Kortsarz, and Nutov: see the conference version in [3], and the two part full version in [4, 5].

For $S \subseteq V$ let $\delta(S)$ denote the set of links in E with exactly one endnode in S; let $E(S)$ denote the set of links in E with both endnodes in S. For $E' \subseteq E$ and $x \in R^E$ let $x(E') = \sum_{e \in E'} x_e$. Let τ^* denote the optimal value of the following standard LP-relaxation for a TAP instance at hand:

$$\begin{aligned} \tau^* = \min \quad & x(E) \\ \text{s.t.} \quad & x(\delta(S)) \geq 1 \qquad \forall S \in \mathcal{E} \\ & x_e \geq 0 \qquad\qquad \forall e \in E \end{aligned} \tag{1}$$

Let τ denote the optimal value of a stronger LP-relaxation, which is obtained by adding to (1) the "leaf edge-cover integrality constraints" (see Chapter 27 in [13]):

$$\begin{aligned} \tau = \min \quad & x(E) \\ \text{s.t.} \quad & x(\delta(S)) \geq 1 && \forall S \in \mathcal{E} \\ & x(E(U) \cup \delta(U)) \geq \lceil |U|/2 \rceil && \forall U \subseteq L \text{ with } |U| \text{ odd} \\ & x_e \geq 0 && \forall e \in E \end{aligned} \tag{2}$$

Let t^* denote the optimal value of the relaxation of (1) for the problem of covering only the leaf edges, and similarly t is defined by relaxing (2). Namely:

$$t^* = \min \left\{ x(E) : x(\delta(v)) \geq 1 \ \forall v \in L, \ x \geq 0 \right\} \ , \tag{3}$$

$$t = \min \left\{ x(E) : x(E(U) \cup \delta(U)) \geq \lceil |U|/2 \rceil \ \forall U \subseteq L \text{ with } |U| \text{ odd}, \ x \geq 0 \right\} \ . \tag{4}$$

Let LL-TAP be the restriction of TAP to instances when the endnodes of every link are leaves of T. In the covering laminar family setting, this is equivalent to requiring that every link has its endnodes in the minimal members of the laminar family, or that the laminar family contains all the singletons, see [7]. The case when T is a star is equivalent to the Edge-Cover problem. In this case $\tau^* = t^*$, and $\tau = t$ is the optimal solution value, see Chapter 27 in [13]. In general, $\tau \geq \tau^*$ and $t \geq t^*$, and an equality may not hold.

LL-TAP was studied in several papers. Garg, Khandekar, and Talwar [7] gave a 5/3-approximation algorithm for LL-TAP. Cheriyan, Jordán, and Ravi [1] showed that even the special case of LL-TAP when the set of links forms a cycle on the leaves remains NP-hard, and gave, for this restricted case, a 4/3-approximation algorithm that computes a solution of size at most $4/3 \cdot \tau^*$. Motivated by this, they conjectured that the integrality gap of the LP-relaxation (1) for the general TAP is 4/3. This conjecture was recently disproved by Cheriyan, Karloff, Khandekar, and Könemann [2], that showed that the integrality gap of (1) is at least 3/2. It is believed that the integrality gap of (1) is less than 2, but so far there was no evidence that this is so even for LL-TAP. We also note that so far, all the approximation algorithms that achieve a ratio better than 2 for TAP [12, 4, 5], or even for LL-TAP [7, 1], were quite involved.

Here is some notation used in the paper. Given an instance $T = (V, \mathcal{E}), E$ of TAP, for $u, v \in V$ let $(u, v) \in \mathcal{E}$ denote the edge in T and uv the link in E between u and v. Fix a designated node r to be the root of T; this defines a partial order on the nodes of T. For $a, b \in V$, $a \prec b$ means that a is a proper descendant of b, and if $(a, b) \in \mathcal{E}$ then $b = p(a)$ is the *parent* of a. A node is a leaf of T if it has no descendants; let $L = L(T)$ the set of leaves of T. T_{uv} denotes the unique uv-path in T, and T_a denodes the subtree of T rooted at a. $\mathsf{lca}(u, v)$ is the least common ancestor of u, v in T. A link uv *covers* an edge e if $e \in T_{uv}$. For a set L' of leaves, $\mathsf{up}(L')$ is a link in E with an endpoint in L' covering the highest edge. For a link set E' and a subtree T' of T let $T' \cap E'$ be the set of links in E' with both endnodes in T'; let $V(E')$ be the set of endnodes of the links in E'. If we add a link uv to a partial solution F, then T_{uv} belongs to the same 2-edge-connected component of $T + F$. Hence, we may contract T_{uv}. Since all contractions we do are induced by subsets of links, we refer to the contraction of every 2-edge-connected component of $T + F$ into a single node simply as the contraction of the links in F.

Definition 1.1 *A link $uv \in T_a \cap E$ is* redundant *for a node $a \neq r$ if every link that covers the edge $(a, p(a))$ is incident to u or to v, namely, if*

$$\{x \in T_a : \text{there is } xy \in E \text{ with } y \in T - T_a\} \subseteq \{u, v\}.$$

Let R denote the set of redundant links.

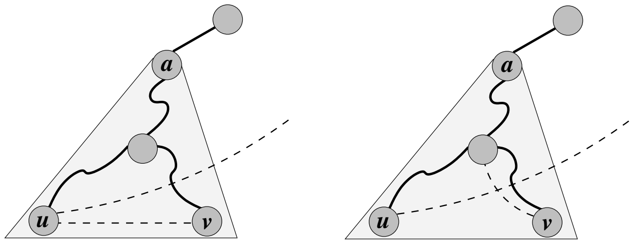

Figure 1: Illustration to the proof of Proposition 2.1 (links are shown by dashed lines).

We will consider $\mathsf{LL\text{-}TAP}$, when every link has both endnodes in L, and obtain the following results:

Theorem 1.1 *For* $\mathsf{LL\text{-}TAP}$, $\mathsf{opt} \geq |L| - |M|$, *and there exists a polynomial time algorithm that computes a solution* F *of size* $|F| \leq |L| - |M|/2 \leq 3/2 \cdot \mathsf{opt}$, *where* M *is a maximum matching in* $E - R$. *Furthermore, if the* $\mathsf{LL\text{-}TAP}$ *instance has no redundant links, then* $|F| \leq 5/3 \cdot t^* \leq 5/3 \cdot \tau^*$ *and* $|F| \leq 3/2 \cdot t \leq 3/2 \cdot \tau$.

Theorem 1.2 $\mathsf{LL\text{-}TAP}$ *admits the following approximation ratios:* $17/12$ *for general trees,* $11/8$ *for trees of height* 3, *and* $4/3$ *for trees of height* 2.

2 Algorithm with ratio $3/2$ (Proof of Theorem 1.1)

2.1 The lower bound

Let U be the set of M-exposed leaves. We prove that $\mathsf{opt} \geq |M| + |U| = |L| - |M|$.

Modify the $\mathsf{LL\text{-}TAP}$ instance by adding for every link $uv \in E$ the *dummy links* ua, va, $a = \mathsf{lca}(u, v)$. Clearly, this does not change the optimum and does not affect M. The next statement shows that then removing the redundant links does not increase opt.

Proposition 2.1 *After adding the dummy links, for any feasible solution* F *there exists a feasible solution* $F' \subseteq E - R$ *so that* $|F'| \leq |F|$.

Proof: Let $uv \in F \cap R$ be redundant for $a \neq r$ (see Fig 1), namely, any link in E that covers $(a, p(a))$, is incident to u or to v. Thus F contains a link incident to u or to v that covers $(a, p(a))$; w.l.o.g., assume there is such a link incident to u (see Fig. 1). It is easy to see that replacing uv by the link between v and $\mathsf{lca}(u, v)$ gives a new feasible solution. In this way we can eliminate all redundant links. $\square$

Let $F \subseteq E$ be an optimal solution, let M' be a maximum matching in $F - R$, and let U'

be the set of M'-exposed leaves. By Proposition 2.1, $|F| \geq |M'| + |U'| = |L| - |M'|$. Clearly, $|M'| \leq |M|$. Thus $|F| \geq |L| - |M'| \geq |L| - |M| = |M| + |U|$.

2.2 The algorithm and its analysis

Here we give an algorithm that computes a solution F with $|F| \leq 3/2 \cdot |M| + |U| = |L| - |M|/2$. Assign *credit* to every member of $M + U$ as follows. Every link in M gets $3/2$ credit units, while every node in U gets 1 credit. The total credit is $3/2 \cdot |M| + |U|$. For technical reasons, we also assign 1 credit unit to r. We will show that we can contract T with a set F of links so that every link is paid by the assigned credit, and 1 credit unit (of r) remains.

Definition 2.1 *Let $\mathcal{Q}$ denote the set of 2-edge-connected components of $T + M$. $Q \in \mathcal{Q}$ is a* lonely component *if it contains exactly one link (a* lonely link*) in M and does not contain r. Denote by $\mathcal{Q}'$ the set of lonely components and by M' the set of lonely links.*

The algorithm maintains a set A of *active nodes*, while obeying the following invariants:

1. Every M-exposed leaf is active and r is active.
2. Every endnode of a link is either an M-covered leaf or an active node.
3. For every $uv \in M'$ there is a link zw with $z \in T_a - \{u, v\}$ that covers the edge $(a, p(a))$, where $a = \mathsf{lca}(u, v) \neq r$.
4. Every link in M owns $3/2$ credit units, while every node in A owns 1 credit.

Initially, $A \leftarrow U + r$, and invariants 1,2,3,4 hold. Invariant 3 holds because uv is not redundant. Then proceed as follows. The general idea is to contract all the links of M into active nodes without over spending the credit.

Step 1: Contract into an active node every non-lonely component $Q \in \mathcal{Q} - \mathcal{Q}'$. Note that if $|Q \cap M| = q$, then the credit of Q is $3/2 \cdot q$ if $r \notin Q$ and $3/2 \cdot q + 1$ if $r \in Q$. Thus the extra credit in Q is at least $3/2 \cdot q - q \geq 1$ if $q \geq 2$, and $2.5 - 1 = 1.5$ if $q = 1$ (since then $r \in Q$).

Step 2: While $M' \neq \emptyset$ do the following. Pick $uv \in M'$ with the lowest $\mathsf{lca}(u, v)$, so there is no $u'v' \in M'$ with $\mathsf{lca}(u', v') \prec \mathsf{lca}(u, v)$. Let zw be a link as in Invariant 3. Contract the component of $T + M' + zw$ containing zw and uv. We show later that $z \in A$, hence if $w \in A$, then the extra credit ≥ 1.5, while if w is an endnode of a link in M' the extra credit is 1.

Step 3: While $T \neq \{r\}$ iteratively choose a link and contract it. As every link chosen connects two active nodes, the extra credit is 1.

Lemma 2.2 *At Step 2 of the algorithm, for the lonely link uv with the lowest $\mathsf{lca}(u,v)$, $z \in A$ holds for the link zw chosen.*

Proof: Suppose to the contrary that $z \notin A$. Then there is $zw' \in M'$, by Invariant 2. We must have $\mathsf{lca}(z,w') \prec \mathsf{lca}(u,v)$, as otherwise uv, zw' belong to the same 2-edge-connected component of $T + M'$. This contradicts the choice of uv. $\square$

Proposition 2.3 *The solution F constructed is feasible and $|F| \leq 3/2 \cdot |M| + |U| \leq |L| - |M|/2$.*

Proof: F is feasible since T was contracted into r. It is easy to verify that Invariants 1,2,3,4 hold during Steps 1,2,3. Thus the second statement follows from the credit scheme used. $\square$

For a tight example showing that the ratio between the optimum and the lower bound $|L| - |M|$ can be arbitrarily close to $3/2$, and that the ratio between the optimum and the solution computed by the algorithm is asymptotically $3/2$, see Section 2.4.

2.3 Integrality gap

Here we will prove that our algorithm computes an edge set F with $|F| \leq 5/3 \cdot t^* \leq 5/3 \cdot \tau^*$ and $|F| \leq 3/2 \cdot t \leq 3/2 \cdot \tau$.

Definition 2.2 *A k-*star *is a star with k leaves (single edges are 1-stars). An ℓ-cycle is a cycle of length ℓ.*

The following statement is well known, c.f., [13].

Lemma 2.4 *Let x be a basic feasible solution of* LP (3). *Then x is half integral, and the set $\{e \in E : x_e > 0\}$ forms a collection of node disjoint stars and odd cycles (covering L), so that: every e with $x_e = 1$ belongs to a star, and every e with $x_e = 1/2$ belongs to a cycle.*

Lemma 2.5 $|F| \leq 5/3 \cdot t^* \leq 5/3 \cdot \tau^*$.

Proof: Recall that $|F| \leq |L| - |M|/2$. Thus it is sufficient to show that $(|L| - |M|/2) \leq 5t^*/3$, namely, that $2(|L| - 5t^*/3) \leq |M|$. Let x be an optimal basic feasible solution to (3) as in Lemma 2.4. Let $\mathcal{Q}$ be the set of connected components of $\{x_e : x_e > 0\}$. For $Q \in \mathcal{Q}$ let M_Q be a maximum matching in Q, let L_Q be the node set of Q, and let $t^*_Q = \sum_{e \in Q} x_e$. Clearly, $M_{\mathcal{Q}} = \bigcup_{Q \in \mathcal{Q}} M_Q$ is a matching, so $|M| \geq |M_{\mathcal{Q}}|$. It is enough therefore to prove that

$$2(|L_Q| - 5t^*_Q/3) \leq |M_Q| \quad \forall Q \in \mathcal{Q} \; .$$

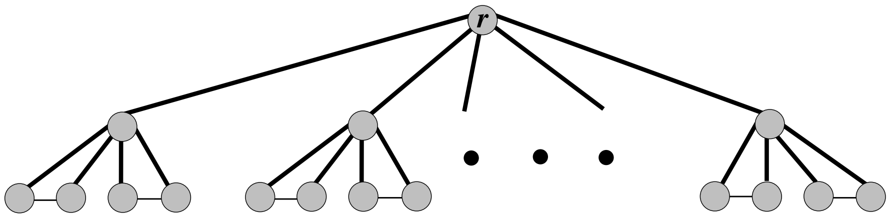

Figure 2: A tight example; tree edges are shown by bold lines, the matching M is shown by thin lines.

Suppose that Q is a k-star, $k \geq 1$. Then $|M_Q| = 1$, $|L_Q| = k+1$, and $t^*_Q = k$. In this case $2(|L_Q| - 5t^*_Q/3) = 2(k+1-5k/3) = 2(1-2k/3) \leq 2/3 < 1 = |M_Q|$.

Suppose that Q is a $(2k+1)$-cycle, $k \geq 1$. Then $|M_Q| = k$, $|L_Q| = 2k+1$, and $t^*_Q = k+1/2$. In this case $2(|L_Q| - 5t^*_Q/3) = 2[2k+1-(5/3)\cdot(k+1/2)] = 2k/3+1/3 \leq k = |M_Q|$. $\square$

Lemma 2.6 $|F| \leq 3/2 \cdot t \leq 3/2 \cdot \tau$.

Proof: We have $t = |L| - |M|$ and our algorithm computes a solution F of size $|F| \leq |L| - |M|/2$. Since $|M| \leq |L|/2$, we obtain $|F|/t \leq 3/2$. $\square$

2.4 A tight example for the 3/2-approximation algorithm

We can show that the ratio between the optimum and the lower bound $|L| - |M|$ can be arbitrarily close to 3/2, hence a better lower bound is needed to get a ratio better than 3/2. We can also show the ratio between the optimum and the solution computed by the algorithm is asymptotically 3/2, hence the analysis of our algorithm is tight. We do not have one example that illustrates both phenomena, so we will give two related examples. Consider the tree and the matching of links in Figure 2, where the tree edges are shown by bold lines. Let k be the number of children of the root. Note that the lower bound is $|L| - |M| = 4k - 2k = 2k$.

If all links not in M are incident to the same leaf v, then an optimal solution is obtained by adding to M any $k-1$ links between v and every subtree of r not containing v. Hence $\mathsf{opt} = 3k-1$ in this case, and the ratio between the optimum and the lower bound is asymptotically 3/2.

If the links form a clique on the leaves then there exists an optimal solution which is a perfect matching on the leaves. Hence $\mathsf{opt} = 2k$ in this case. The algorithm may compute a solution of size $3k-1$, by adding to M any $k-1$ links between some node v and every subtree of r not containing v. In this case the ratio between the optimum and the size of the solution computed by the algorithm is asymptotically 3/2.

3 Algorithm with ratio 17/12 (Proof of Theorem 1.2)

3.1 The lower bound

Definition 3.1 *A pair of non-adjacent links $e, e' \in E(T_a) - R$ is a* dangerous link-pair *for $a \in V - r$ if $\{x \in T_a : \text{there is } xy \in E \text{ with } y \in T - T_a\} \subseteq V(\{e, e'\})$. A link is dangerous if it belongs to some dangerous link-pair.*

Lemma 3.1 *Let F be a feasible solution. Among all maximum matchings in $F - R$, let M' be one with the minimum number d' of dangerous links, and let U' be the set of M'-exposed leaves. Then*

$$|F| \geq |M'| + |U'| + \max\{d' - |L|/4, 0\}/2 = |L| - |M'| + \max\{d' - |L|/4, 0\}/2 .$$

Proof: By the lower bound in Theorem 1.1, $|F| \geq \mathsf{opt} \geq |M'| + |U'|$. Thus if $d' \leq |L|/4$, the statement is true. Suppose that $d' > |L|/4$. Then there is a set $\mathcal{D}'$ of (at least) $d' - |L|/4$ pairwise disjoint dangerous pairs in M'. For every pair $\{e, e'\} \in \mathcal{D}'$ that is dangerous for a node a, there is a link $xy \in F$ with $x \in V(\{e, e'\})$ and $y \in V - T_a$, to cover the edge $(a, p(a))$. The number of such links is at least $(d' - |L|/4)/2$, since the pairs in $\mathcal{D}'$ are pairwise disjoint and since their union is a matching. Hence the statement will follow if for every $\{e, e'\} \in \mathcal{D}'$ there is a link $xy \in F$ as above so that at least one of the following holds: $y \notin U'$ or $xy \in R$.

Now suppose that $y \in U'$ and $xy \notin R$ for some $\{e, e'\} \in \mathcal{D}'$. Let M'' be obtained from M' by replacing the link among e, e' incident to x by the link xy. Then M'' is also a maximum matching in $F - R$. Note that xy must be dangerous (so M'' has also exactly d' danderous links), as otherwise M'' has less dangerous links than M', contradicting the choice of M'. Note also that $a \prec \mathsf{lca}(x, y)$, while the lca of the endnodes of e, e' is a descendant of a; this implies $\mathsf{lca}(V(\{e, e'\})) \prec \mathsf{lca}(x, y)$, hence any sequence of such replacements cannot loop and must terminate. Consequently, we can obtain a maximum matching M'' in $F - R$ with $d' - |L|/4$ pairwise disjoint dangerous pairs so that if $e, e' \in M''$ is a pair dangerous for a, then there is a link $xy \in F$ with $x \in V(\{e, e'\})$ and $y \in V - T_a$, so that at least one of the following holds: $y \notin U'$ or $xy \in R$. This finishes the proof, as for this case we already proved that the statement is valid. $\square$

Definition 3.2 *Let ν denote the maximum size of a matching in $E - R$. For $i = 0, \ldots, \nu$ define:*

- *$M_i \subseteq E - R$ is a matching of size i with minimum number d_i of dangerous links;*
- *U_i is the set of M_i-exposed leaves;*
- *$f(i) = |M_i| + |U_i| + \max\{d_i - |L|/4, 0\}/2 = |L| - i + \max\{d_i - |L|/4, 0\}/2$.*

Clearly, d_i is uniquely determined for every i, and thus also $f(i)$ is uniquely determined.

Lemma 3.2 *$f(i)$ can be computed in polynomial time for every i.*

Proof: Add a set Q of $|L| - i$ new nodes to the graph $(L, E - R)$, and zero weight links between Q and L. Assign weight > 1 to every dangerous link (other links keep unit weights). Now, compute a minimum weight perfect matching M in the obtained graph. This can be done in polynomial time using Edmond's algorithm. It is easy to see that the links in M with both endnodes in L is a minimum weight matching of size i in $E - R$. $\square$

Corollary 3.3 $\mathsf{opt} \geq \min_i f(i)$.

Proof: Let F be an optimal solution and let M' be a maximum matching in $F - R$ with minimum number d' of dangerous links. Let $i = |M'|$ and let U' the set of M'-exposed leaves. Note that $d_i \leq d'$. Thus from Lemma 3.1 we have $|F| \geq f(i)$, and the statement follows. $\square$

3.2 The algorithm and its analysis

We will prove the following statement:

Theorem 3.4 *For any $i = 0, \ldots, \nu$, there exists a polynomial time algorithm that computes a solution F of size $|F| \leq (17/12) \cdot f(i)$.*

Theorem 1.2 easily follows from Corollary 3.3 and Theorem 3.4. By Lemma 3.2 we can find the index $j \in \{0, \ldots, \nu\}$ for which $f(j) = \min_i f(i)$. Then we use the algorithm as in Theorem 3.4 to compute a solution F of size $|F| \leq (17/12) \cdot f(j)$. By Corollary 3.3, we have $|F| \leq (17/12) \cdot f(j) \leq (17/12) \cdot \mathsf{opt}$.

The proof of Theorem 3.4 follows. Let $M = M_i$, $U = U_i$, and $d = d_i$. Assign *credit* to every member of $M + U$ as follows. Every link in M gets: $3/2$ credit units if it is dangerous, and $4/3$ credit units otherwise. Every leaf in U gets 1 credit. The total credit assigned is

$$\begin{aligned}
\frac{4}{3}|M| + |U| + \frac{d}{6} &\leq \frac{17}{12}(|M| + |U|) - \frac{1}{24}(2|M| + |U|) + \frac{d}{6} \\
&= \frac{17}{12}(|M| + |U|) + \frac{1}{6}(d - |L|/4|) \\
&\leq \frac{17}{12}\left(|M| + |U| + \frac{1}{2}\max\{d - |L|/4, 0\}\right) = \frac{17}{12} \cdot f(i) \ .
\end{aligned}$$

For technical reasons, we also assign 1 credit unit to r. We will show that we can contract T so that every link is paid by the assigned credit, and 1 credit unit (of r) remains. The algorithm maintains the following invariants:

1. Every link in M owns 3/2 credit units if it is dangerous, and 4/3 credit units otherwise. Every M-exposed leaf of T owns 1 credit, and r owns 1 credit.

2. Every endnode of a link is a leaf, and every M-covered leaf is an *original* leaf of T.

3. For every $a \neq r$ and $uv \in M \cap T_a$ there is a link zw with $z \in L(T_a) - \{u, v\}$ that covers the edge $(a, p(a))$.

4. For every $a \neq r$ and $u_1v_1, u_2v_2 \in M \cap T_a$ so that one of u_1v_1, u_2v_2 is not dangerous, there is a link zw with $z \in T_a - \{u_1, v_1, u_2, v_2\}$ that covers the edge $(a, p(a))$.

Note that initially Invariants 1,2,3,4 hold. Invariant 3 holds since uv is not redundant, and Invariant 4 holds since the pair u_1v_1, u_2v_2 is not dangerous. Later, we will prove:

Lemma 3.5 *Suppose that Invariants 1-4 hold. Then there exists a polynomial time algorithm that finds a rooted subtree T' of T and a cover $B' \subseteq E \cap T'$ so that $credit(T') \geq |B'|+1$, and so that for any $b_1b_2 \in M$ either both b_1, b_2 belong to T', or none of b_1, b_2 belongs to T'.*

The algorithm iteratively finds a rooted subtree T' of T and a cover $B' \subseteq E$ of T' as in Lemma 3.5, contracts T' with B', and assigns 1 credit to the created leaf.

Algorithm $Approx(T = (V, \mathcal{E}), E, M, d)$
(Computes a solution F with $|F| \leq (4/3) \cdot |M| + |U| + d/6$)

1: **while** T has more than one node **do**

2: Find a rooted subtree T' of T and a cover B' of T' as in Lemma 3.5.

3: Contract T', give 1 credit to the new leaf, and set $F \leftarrow F \cup B'$.

4: **end while**

5: Return F.

The condition $credit(T') \geq |B'| + 1$ assures that we are not over spending the credit. It is easy to verify that other conditions assure that Invariants 1-4 continue to hold. Thus to prove Theorem 3.4 it is sufficient to prove Lemma 3.5; this is what we will do in the rest of this section.

Definition 3.3 ([12]) *Let U be a subset of nodes of T. A rooted subtree T' of T is U-*closed *if there is no link in E from $U \cap T'$ to $T \setminus T'$. T' is* leaf-closed *if it is $L(T)$-closed. A leaf-closed T' is* minimally leaf-closed *if any proper subtree of T' is not leaf-closed.*

Proposition 3.6 ([12]) *A minimally leaf-closed subtree T' of T is covered by $\mathsf{up}(L(T'))$.*

Definition 3.4 (Semi-closed tree)
A rooted subtree T' of T is semi-closed *(w.r.t. a link set M) if the following holds:*

- *For any $b_1 b_2 \in M$, either both b_1, b_2 belong to T', or none of b_1, b_2 belongs to T'.*
- *T' is closed w.r.t. its M-exposed leaves, namely, every link incident to an M-exposed leaf of T' has both endnodes in T'.*

T' is minimally semi-closed *if T' is semi-closed but any proper subtree of T' is not semi-closed.*

Note that a semi-closed T' is *not* leaf-closed; T' is closed with respect to every M-exposed leaf, but M-covered leaves may have links to nodes outside T'. The concept of semi-closed trees was defined in [3, 4], where it is assumed that "a contraction of a subset of M cannot create a new leaf". This implies that a semi-closed tree should have at least one M-exposed leaf. We do not pose such a restriction, so in our setting a semi-closed T' may have zero M-exposed leaves. As T itself is semi-closed, T has a minimally semi-closed subtree. Such can be found in polynomial time, as T has at most $|V|$ rooted subtrees, and clearly checking if a subtree is semi-closed can be done in polynomial time. The following statement explains how we intend to cover minimally semi-closed trees. This statement was essentially proved in [3, 4], but as our definition slightly differs from the one in [3, 4], we provide a proof for completeness of exposition.

Lemma 3.7 *Let T' be a semi-closed subtree of T w.r.t. M. Let $B(T')$ consist of the union of $M \cap T'$ and the up-links of the M-exposed leaves of T'. Then $B(T') \subseteq E \cap T'$, and if T' is minimally semi-closed, then $B(T')$ covers T'.*

Proof: It is clear that $B(T') \subseteq E \cap T'$. Let T'' be obtained by contracting $M \cap T'$. The leaves of T'' are the M-exposed leaves of T', T'' is leaf-closed, and T' is minimally semi-closed if, and only if, T'' is minimally leaf-closed. Thus the up-links of the M-exposed leaves of T' cover T'' (if T' has no M-exposed leaves, then T'' is a single node), by Proposition 3.6. The statement follows. $\square$

Consequently, to finish the proof of Lemma 3.5, it is sufficient to prove:

Lemma 3.8 *Suppose that Invariants 1-4 hold. Then $credit(T') \geq |B(T')| + 1$ for any semi-closed tree T'.*

Proof: Let a be the root of T', and let β be $|M \cap T'|$ plus the number of M-exposed leaves of T'. Clearly, $\beta \geq |B'|$, and we claim that $credit(T') \geq \beta + 1$. If $a = r$ this is obvious, so assume $a \neq r$. Clearly, $credit(T') - \beta = credit(M \cap T') - |M \cap T'|$. Thus if $|M \cap T'| \geq 3$, or if $M \cap T'$ consists of 2 dangerous links, then $credit(M \cap T') \geq |M \cap T'| + 1$ and the required extra unit of credit follows. We claim that the other cases (when $M \cap T'$ is empty, or is a single link, or is a pair of links that are not both dangerous) are not possible. Otherwise, we obtain a contradiction to Invariant 2. Consider the set C of the endnodes in T' of the links in E that cover the edge $(a, p(a))$. Clearly, $C \neq \emptyset$, and we will show that C contains a non-leaf node z. Note that C cannot contain an M-exposed leaf, as T' is closed w.r.t. such leaves. Also, $C - V(M \cap T) \neq \emptyset$; if $M \cap T' = \emptyset$ this is obvious, if $|M \cap T'| = 1$ this follows from Invariant 3, and if $|M \cap T'| = 2$ this follows from Invariant 4. Consequently, C contains a non-leaf node, as claimed, which gives a contradiction. $\square$

Lemma 3.5 now follows from Lemmas 3.7 and 3.8. This finishes the proof of Theorem 3.4, and thus the proof of Theorem 1.2 for general trees is now complete.

4 Algorithm for trees of height ≤ 3

Throughout this section assume that T has height ≤ 3. Then we use a slightly different definition of dangerous links, and in our credit scheme we allow that some parts will have negative credit.

Definition 4.1 *For $a \neq r$, let J_a be a tree obtained by removing from T_a the subtrees rooted at the children of a with at least 4 leaves. We say that a is a* dangerous node *and that J_a is a* dangerous subtree *if J_a has exactly 4 leaves and $\{x \in T_a : \text{there is } xy \in E \text{ with } y \in T - T_a\} \subseteq J_a$; note that two distinct dangerous subtrees are disjoint. Let D denote the set of dangerous nodes. A* dangerous link *is a link with both endnodes in the same dangerous subtree.*

With this modified definition of dangerous links, let M_i, U_i, and d_i be as in Definition 3.2. The function f is defined by $f(i) = |M_i| + |U_i| + \max\{d_i - |D|, 0\}/2$.

Lemma 4.1 *Let F cover T. Among all maximum matchings in $F - R$, let M' be one with the minimum number d' of dangerous links. Let U' the set of M'-exposed leaves. Then $|F| \geq |M'| + |U'| + \max\{d' - |D|, 0\}/2$.*

Proof: Similarly to the argument in Lemma 3.1, $|F| \geq |M'| + |U'|$. Thus if $d' \leq |D|$, the statement is true. Assume that $d' > |D|$. Then there is a set D' of (at least) $d' - |D|$ dangerous nodes, each with 2 dangerous links from M' in its subtree. Consider $a \in D'$. Let $uv \in F$ be a link covering $(a, p(a))$, where $u \in T_a$. Let $ut \in M$ so that $t \in T_a$. Note that uv is not dangerous and that ut is dangerous. Thus $v \notin U$; otherwise, $M' - ut + uv$ is a maximum matching in F with less dangerous links than M'. Thus we have at least $d' - |D|$ nodes in $L - U'$ which degree w.r.t. F is at least 2. The statement follows. $\square$

Corollary 4.2 $\mathsf{opt} \geq \min_i f(i)$.

Proof: Let F be an optimal solution and let M' be a maximum matching in $F - R$ with minimum number d' of dangerous links. Let $i = |M'|$ and let U' be the set of M'-exposed leaves. Note that $d_i \leq d'$. Thus from Lemma 4.1 we have $|F| \geq f(i)$, and the statement follows. $\square$

Let ρ be the approximation ratio as in Theorem 1.2, that is, $\rho = 11/8$ if T has height 3 and $\rho = 4/3$ if T has height 2. Let $M = M_j$, $U = U_j$, and $d = d_j$, where $j = \arg\min_i f(i)$. We use the following credit scheme. Assign a credit of:

- $3 - \rho$ to every dangerous link in M;
- ρ to every non-dangerous link in M and to every $u \in U$;
- $-(3 - 2\rho)$ to every node in D, namely, charge every node in D with $3 - 2\rho$.

Then $credit(T)$ is at most ρ times the lower bound in Corollary 4.2 since:

$$\begin{aligned} credit(T) &= \rho \cdot (|M| - d) \cdot \rho + (3 - \rho) \cdot d + \rho \cdot |U| - (3 - 2\rho) \cdot |D| \\ &\leq \rho \cdot (|M| + |U|) + \max\{d - |D|, 0\}/3 \\ &\leq \rho \cdot (|M| + |U| + \max\{d - |D|, 0\}/2) \ . \end{aligned}$$

The algorithm is as follows.

1. Calculate M, U as above and set $F_1 \leftarrow M \cup \{\mathsf{up}(u) : u \in U\}$.
2. Obtain F_2 by adding $\mathsf{up}(Q)$ to F_1 for every 2-edge-connected component Q of $T + F_1$ so that all the children of $\mathsf{lca}(Q)$ are leaves.
   $\triangleright$ Comment: If T has height 2 then F_2 is already a feasible solution.
3. Obtain F_3 by adding $\mathsf{up}(Q)$ to F_2 for every 2-edge-connected component Q of $T + F_2$ with $r \notin Q$.

The following immediate statement implies that the computed solution is feasible.

Proposition 4.3 *F_k covers all the rooted subtrees of T of height $\leq k$, $k = 1, 2, 3$. Thus if T has height ≤ 3, then at the end of the algorithm F is a feasible solution.*

Let $\mathcal{Q}_k$ be the set of 2-edge-connected components of $T + F_k$ $k = 1, 2$. Let

$$\begin{aligned} \mathcal{Q}'_1 &= \{Q \in \mathcal{Q}_1 : \text{all the children of } \mathsf{lca}(Q) \text{ are leaves}\}, \\ \mathcal{Q}''_1 &= \{Q \in \mathcal{Q}_1 - \mathcal{Q}'_1 : Q \subseteq Q_2 \text{ for some } Q_2 \in \mathcal{Q}_2 \text{ with } r \notin Q_2\}. \end{aligned}$$

Note that we might have $\mathcal{Q}_1 - (\mathcal{Q}'_1 \cup \mathcal{Q}''_1) \neq \emptyset$.

Lemma 4.4 *Any $Q \in \mathcal{Q}_2$ with $r \notin Q$ contains a unique component $Q'' \in \mathcal{Q}''_1$. Moreover, $\mathsf{lca}(Q) = \mathsf{lca}(Q'')$ and $\{u \in T_a : \text{there is } uz \in E \text{ with } z \in T - T_a\} \subseteq Q''$, where $a = \mathsf{lca}(Q)$.*

Proof: Let $u \in T_a$ be a leaf with a link uz covering $(a, p(a))$. Let $Q'' \in \mathcal{Q}_1$ be the component that includes u. Suppose to the contrary that $Q'' \in \mathcal{Q}'_1$. Then, after adding $\mathsf{up}(Q'')$ to F_2, $(a, p(a))$ is covered, and then $p(a)$ and a are in the same 2-edge connected component in $\mathcal{Q}_2$. This contradicts $p(a) \notin Q$. Thus, $Q'' \in \mathcal{Q}_1 - \mathcal{Q}'_1$ and then $\mathsf{lca}(Q'')$ has a non-leaf child. Clearly, $\mathsf{lca}(Q'') \preceq \mathsf{lca}(Q)$, otherwise Q contains $p(a)$. The tree has height ≤ 3, $\mathsf{lca}(Q'') \preceq \mathsf{lca}(Q) \prec r$ and $\mathsf{lca}(Q'')$ has non-leaf child, thus $\mathsf{lca}(Q'') = \mathsf{lca}(Q)$ and $Q'' \subseteq Q$. Follows this, $Q'' \in \mathcal{Q}''_1$. As every leaf in $\{u \in T_a : \text{there is } uz \in E \text{ with } z \in T - T_a\}$ is included in a component of $\mathcal{Q}_1$ that includes $lca(Q)$, Q'' is unique and $\{u \in T_a : \text{there is } uz \in E \text{ with } z \in T - T_a\} \subseteq Q'' \in \mathcal{Q}''_1$. $\square$

Links in $F_2 - F_1$ correspond bijectively to components in $\mathcal{Q}'_1$, and by Lemma 4.4, links in $F_3 - F_2$ corresponds bijectively to components in $\mathcal{Q}''_1$. Hence, there is a bijective correspondence between links in $F_3 - F_1$ and components in $\mathcal{Q}'_1 \cup \mathcal{Q}''_1$.

The following statement implies that the credit distributed suffices to pay for the links added during the algorithm.

Lemma 4.5 *$credit(Q) \geq |F_1 \cap Q|$ for all $Q \in \mathcal{Q}_1$, and $credit(Q) \geq |F_1 \cap Q| + 1$ if $Q \in \mathcal{Q}'_1 \cup \mathcal{Q}''_1$.*

In the rest of this section we prove Lemma 4.5. This completes the proof of Theorem 1.2.

Proposition 4.6 *Let $Q \in \mathcal{Q}'_1 \cup \mathcal{Q}''_1$. Then $|Q \cap L| \geq 4$ and if $|Q \cap L| = 4$ then Q contains 2 dangerous links.*

Proof: Let $a = \mathsf{lca}(Q)$. We have $Q \in \mathcal{Q}'_1 \cup \mathcal{Q}''_1$, then $a \neq r$. We claim that

$$\{x \in T_a : \text{there is } xy \in E \text{ with } y \in T - T_a\} \subseteq Q \ .$$

For $Q \in \mathcal{Q}'_1$ this is obvious and for $Q \in \mathcal{Q}'_2$ this follows Lemma 4.4. It is easy to see that if the statement does not hold, then $Q \cap M$ contains a redundant link or F_1 covers (a,p(a)), contradicting $p(a) \notin Q$. □

Proposition 4.7 *Let $Q \in \mathcal{Q}_1$ and let $a \in Q \cap D$. Then $J_a \subseteq Q$ and $|Q \cap D| \leq |F_1 \cap Q|/2$.*

Proof: Let $a \in Q \cap D$. Suppose to the contrary that there is $u \in J_a - Q$. Then u belongs to $Q' \in \mathcal{Q}_1$ and $a' = \mathsf{lca}(Q') \prec a$. The tree is of height 3 and $a' \prec a \prec r$, thus $Q' \in \mathcal{Q}'_1$. We have $a' \notin D$, as otherwise $u \in J_{a'}$ and then $u \notin J_a$. As $a' \notin D$ and $u \in T_{a'}$ is in a dangerous tree, $T_{a'}$ has at most 3 leaves, contradicting Proposition 4.6. Thus, the existence of $u \in J_a - Q$ leads to a contradiction and then $J_a \subseteq Q$.
We have $|Q \cap L| \leq 2|F_1 \cap Q|$ since F_1 covers all the leaves. For every $a \in D \cap Q$, all the leaves of J_a are in Q, hence $|Q \cap L| \geq 4|D \cap Q|$. Consequently, $|D \cap Q| \leq |Q \cap L|/4 \leq |F_1 \cap Q|/2$.
□

Proposition 4.8 *$credit(Q) \geq |F_1 \cap Q|$ for all $Q \in \mathcal{Q}_1$.*

Proof: This follows from Proposition 4.7 since

$$\begin{aligned} credit(Q) &\geq \rho \cdot |F_1 \cap Q| - (3 - 2\rho) \cdot |D \cap Q| \\ &\geq \rho \cdot |F_1 \cap Q| - (3 - 2\rho) \cdot |F_1 \cap Q|/2 \\ &= (2\rho - 3/2) \cdot |F_1 \cap Q| \geq |F_1 \cap Q| \ . \end{aligned}$$

□

Proposition 4.9 *$credit(Q) \geq |F_1 \cap Q| + 1$ for all $Q \in \mathcal{Q}'_1$.*

Proof: By Proposition 4.6, either Q has exactly 4 leaves and contains 2 dangerous links, or $\mathsf{lca}(Q)$ has more than 4 leaves and has no dangerous node. In the former case, the total credit in Q is $2(3 - \rho) - (3 - 2\rho) = 3 = |F_1 \cap Q| + 1$. In the latter case, $|F_1 \cap Q| \geq 3$, and the total credit is $\rho \cdot |F_1 \cap Q|$. As $\rho \geq 4/3$ and $|F_1 \cap Q| \geq 3$, the total credit is at least $|F_1 \cap Q| + 1$. □

Proposition 4.10 *$credit(Q) \geq |F_1 \cap Q| + 1$ for all $Q \in \mathcal{Q}''_1$.*

Proof: We have $\mathsf{lca}(Q) \neq r$, and if $\mathsf{lca}(Q)$ has non-leaf child, then T is of height 3, which means $\rho = 11/8$. Let $|F_1 \cap Q| = q$. By Proposition 4.6, either Q has exactly 4 leaves and contains 2 dangerous links, or Q has more than 4 leaves and $q \geq 3$. In the former case, the total credit in Q is $2 \cdot 13/8 - 1/4 = 3 = q + 1$. To complete the proof, we need to prove that $credit(Q) \geq q + 1$ when $q \geq 3$. If $q \geq 4$ then follows Proposition 4.7, $credit(Q) \geq 11/8 \cdot q - 1/4 \cdot q/2 = q + q/4 \geq q + 1$. If $q = 3$ and has no dangerous node, then $credit(Q) \geq 11/8 \cdot 3 > 4 = q + 1$. If $q = 3$ and Q contains a dangerous node, then follows Proposition 4.7, Q contains exactly one dangerous node a including J_a. But, Q contains at most most 6 leaves, therefore at least one link is dangerous. Thus, $credit(Q) \geq 11/8 \cdot 2 + 13/8 - 1/4 > 4 = q + 1$. □

5 Conclusions and open problems

The main contribution of this paper is in introducing the two concepts of "redundant links" and "dangerous pairs". For LL-TAP, this enabled us to obtain: a simple algorithm with ratio 3/2 using the former concept, and a slightly more complicated algorithm with ratio 17/12 using the latter concept. These concepts might be useful to simplify or to improve the current algorithms for general TAP.

We list three open problems. The main open problem is achieving a ratio better than 2 for *weighted* TAP, when we seek to cover the tree with a set of links of *minimum weight*. This is of interest even for weighted LL-TAP. The currently best known approximation ratio for this problem is 2. We believe that breaking this barrier will lead to improved ratios to several other problems, among them the Steiner Forest problem. Another open problem is the integrality gap of LP-relaxations (1) and (2) for TAP or for LL-TAP. For TAP, we only know that the integrality gap of LP (1) is between 3/2 and 2. For LL-TAP, we know that the integrality gap of (1) is at most 5/3, while the integrality gap of (2) is at most 3/2; these two upper bounds are proved in this work. As a last open problem, we pose a conjecture that for LL-TAP the ratio 4/3 is achievable.

References

[1] J. Cheriyan, T. Jordán, and R. Ravi. On 2-coverings and 2-packing of laminar families. In *ESA*, pages 510–520, 1999.

[2] J. Cheriyan, H. Karloff, R. Khandekar, and J. Könemann. On the integrality ratio of tree augmentation. *Oper. Res. Lett.*, 36(4):399–401, 2008.

[3] G. Even, J. Feldman, G. Kortsarz, and Z. Nutov. A 3/2-approximation for augmenting a connected graph into a two-connected graph. In *APPROX*, pages 90–101, 2001.

[4] G. Even, J. Feldman, G. Kortsarz, and Z. Nutov. A 1.8 approximation algorithm for augmenting edge-connectivity of a graph from 1 to 2. To appear in *Transactions on Algorithms*, 2008.

[5] G. Even, G. Kortsarz, and Z. Nutov. A 1.5 approximation algorithm for augmenting edge-connectivity of a graph from 1 to 2. Manuscript, 2008.

[6] G. N. Fredrickson and J. Jájá. On the relationship between the biconnectivity augmentation and traveling salesman problem. *Theorethical Computer Science*, 19(2):189–201, 1982.

[7] N. Garg, R. Khandekar, and K. Talwar. Covering a laminar family. Manuscript, 2000.

[8] M. Goemans and D. Williamson. *The primal dual method for approximation algorithms and its applications to network design problems,* Ch. 4 in *Approximation Algorithms for NP-hard problems,* D. S. Hochbaum Ed., pages 144-191. PWS, 1995.

[9] K. Jain. A factor 2 approximation algorithm for the generalized Steiner network problem. *Combinatorica*, 21(1):39–60, 2001.

[10] S. Khuller. *Approximation algorithms for for finding highly connected subgraphs,* Ch. 6 in *Approximation Algorithms for NP-hard problems,* D. S. Hochbaum Ed., pages 236-265. PWS, 1995.

[11] G. Kortsarz and Z. Nutov. *Approximating minimum cost connectivity problems,* Ch. 58 in *Approximation Algorithms and Metahueristics,* T. F. Gonzales ed.,. CRC, 2007.

[12] H. Nagamochi. An approximation for finding a smallest 2-edge connected subgraph containing a specified spanning tree. *Discrete Applied Math.*, 126:83–113, 2003.

[13] A. Schrijver. *Combinatorial Optimization, Polyhedra and Efficiency.* Springer-Verlag Berlin, Heidelberg New York, 2004.