Elementary Approximation Algorithms for Prize Collecting Steiner Tree Problems

Shai Gutner*

School of Computer Science, Tel-Aviv University, Tel-Aviv, 69978, Israel
gutner@tau.ac.il

Abstract. This paper deals with approximation algorithms for the prize collecting generalized Steiner forest problem, defined as follows. The input is an undirected graph $G = (V, E)$, a collection $T = \{T_1, \ldots, T_k\}$, each a subset of V of size at least 2, a weight function $w : E \rightarrow \mathbb{R}^+$, and a penalty function $p : T \rightarrow \mathbb{R}^+$. The goal is to find a forest F that minimizes the cost of the edges of F plus the penalties paid for subsets T_i whose vertices are not all connected by F.

Our main result is a combinatorial $(3-\frac{4}{n})$-approximation for the prize collecting generalized Steiner forest problem, where $n \geq 2$ is the number of vertices in the graph. This obviously implies the same approximation for the special case called the prize collecting Steiner forest problem (all subsets T_i are of size 2).

The approximation ratio we achieve is better than that of the best known combinatorial algorithm for this problem, which is the 3-approximation of Sharma, Swamy, and Williamson [13]. Furthermore, our algorithm is obtained using an elegant application of the local ratio method and is much simpler and practical, since unlike the algorithm of Sharma et al., it does not use submodular function minimization.

Our approach gives a $(2 - \frac{1}{n-1})$-approximation for the prize collecting Steiner tree problem (all subsets T_i are of size 2 and there is some root vertex r that belongs to all of them). This latter algorithm is in fact the local ratio version of the primal-dual algorithm of Goemans and Williamson [7]. Another special case of our main algorithm is Bar-Yehuda's local ratio $(2 - \frac{2}{n})$-approximation for the generalized Steiner forest problem (all the penalties are infinity) [3]. Thus, an important contribution of this paper is in providing a natural generalization of the framework presented by Goemans and Williamson, and later by Bar-Yehuda.

Keywords: Approximation algorithms, prize collecting Steiner tree problem, local ratio, primal-dual.

1 Introduction

There is substantial literature dealing with approximation algorithms for prize collecting Steiner tree problems. The purpose of this paper is to present elegant combinatorial algorithms for these problems. The local ratio technique

* This paper forms part of a Ph.D. thesis written by the author under the supervision of Prof. N. Alon and Prof. Y. Azar in Tel Aviv University.

B. Yang, D.-Z. Du, and C.A. Wang (Eds.): COCOA 2008, LNCS 5165, pp. 246–254, 2008.

[2,3,4] that we employ enables us to present simple algorithms together with a straightforward analysis.

The main focus of the paper in on the prize collecting generalized Steiner forest (PCGSF) problem, defined as follows. The input is an undirected graph $G = (V, E)$, a collection $T = \{T_1, \ldots, T_k\}$, each a subset of V of size at least 2, a weight function $w : E \rightarrow \mathbb{R}^+$, and a penalty function $p : T \rightarrow \mathbb{R}^+$, where $\mathbb{R}^+$ denotes the set of positive real numbers. The objective is to compute a forest F that minimizes the cost of the edges of F and the sum of the penalties of the subsets T_i whose vertices are not all connected by F. Thus, all the vertices of a subset T_i must be in the same connected component of F in order to avoid the penalty. Note that we intentionally define costs and penalties to be positive, as this will turn out to be convenient later. During intermediate stages of the algorithm, zero cost edges are contracted, whereas zero penalties can be ignored.

Previous Results. The special case of the PCGSF problem called the prize collecting Steiner forest problem (all subsets T_i are of size 2) has received considerable attention lately. A modification of the LP rounding algorithm in [6] implies a 3-approximation for this problem. This was improved in [9] to give an LP based 2.54-approximation for the problem as well as a primal-dual combinatorial $(3 - \frac{2}{n})$-approximation using Farkas' Lemma. The authors of [8] give a 3-budget-balanced and group-strategyproof mechanism for the game-theoretic version of the prize collecting Steiner forest problem, which is an extension of the method presented in [12]. Their result also provides a primal-dual 3-approximation algorithm for this problem.

A generalized framework of the prize collecting problems with an arbitrary $0 - 1$ connectivity requirement function and a submodular penalty function is studied by Sharma, Swamy, and Williamson in [13]. Their model captures both the PCGSF problem defined in this paper as well as the problems of [10,9]. The authors give a complicated primal-dual 3-approximation algorithm together with an LP rounding algorithm with a performance ratio of 2.54.

Two classical primal-dual algorithms, relevant to this paper, are due to Goemans and Williamson [7]. They give a $(2 - \frac{1}{n-1})$-approximation for the prize collecting Steiner tree problem (all subsets T_i are of size 2 and there is some root vertex r that belongs to all of them) as well as a $(2 - \frac{2}{n})$-approximation for the generalized Steiner forest problem (all the penalties are infinity) that simulates an algorithm of Agrawal, Klein, and Ravi [1]. For the latter problem, a simple $(2 - \frac{2}{n})$-approximation based on the local ratio technique is presented by Bar-Yehuda in [3].

Our Results. The main result is a local ratio $(3 - \frac{4}{n})$-approximation for the prize collecting generalized Steiner forest problem, where $n \geq 2$ is the number of vertices in the graph. This obviously implies the same approximation for the special case of the prize collecting Steiner forest problem, which was previously studied in [9,8].

The approximation ratio of our algorithm is slightly better than that of the 3-approximation of Sharma et al. [13], which is the best known combinatorial approximation algorithm for the problem. The algorithm we present makes an elegant use of the local ratio method and therefore, unlike the algorithm of Sharma et al., does not use submodular function minimization. This makes our approach much simpler and practical. We also note that the main algorithm presented in this paper is not the local ratio version of the primal-dual 3-approximation algorithm from [13] and cannot be obtained from it using the equivalence between the primal-dual schema and the local ratio technique [5].

There are two interesting special cases of our main algorithm. We present a $(2-\frac{1}{n-1})$-approximation for the prize collecting Steiner tree problem (all subsets T_i are of size 2 and there is some root vertex r that belongs to all of them). This latter algorithm is in fact the local ratio version of the primal-dual algorithm of Goemans and Williamson [7]. Another special case of our main algorithm is Bar-Yehuda's local ratio $(2-\frac{2}{n})$-approximation for the generalized Steiner forest problem (all the penalties are infinity) [3]. Thus, an important contribution of this paper is in providing a natural generalization of the framework presented by Goemans and Williamson, and later by Bar-Yehuda.

2 The Prize Collecting Generalized Steiner Forest Problem

In this section we present the algorithm for the PCGSF problem. The following are some definitions that are needed for presenting the algorithm.

Definition 1. *Given an instance $(G, T = \{T_1, \ldots, T_k\}, w, p)$ of the PCGSF problem, a vertex is said to be an **active vertex** if it belongs to at least one of the subsets T_i.*

Definition 2. *A solution F to the PCGSF problem is a **minimal solution** if every leaf (a vertex of degree 1) of the forest F is an active vertex.*

Definition 3. *Suppose $G = (V, E)$ is an undirected graph and $T = \{T_1, \ldots, T_k\}$ is a collection of subsets of V of size at least 2. For every active vertex v, let $t(v)$ be some arbitrary subset T_i for which $v \in T_i$. The **degree-weighted instance** corresponding to G,T, and t is the quadruple (G, T, w, p), defined as follows.*

- *For each $e \in E$, $w(e)$ is the number of its endpoints that are active vertices.*
- *For each $T_i \in T$, $p(T_i) = |\{v \in T_i | t(v) = T_i\}|$.*

Note that in the previous definition, the weight of an edge $w(e)$ can take a value of 0, 1, or 2, whereas the penalty of a subset T_i satisfies $0 \leq p(T_i) \leq |T_i|$. The following definition describes the important and natural operation of an edge contraction.

Definition 4. *Let (G, T, w, p) denote an instance of the PCGSF problem. The **contraction** of an edge $\{u, v\}$ into a new vertex x results in a new instance (G', T', w', p'), defined as follows.*

- *For a vertex y, if y is adjacent in G to both u and v, then it is adjacent to x in G' and $w'(\{y,x\}) = \min\{w(y,u), w(y,v)\}$. We say that the edge $\{y,x\}$ of the graph G'* ***corresponds*** *to the original edge in the graph G for which the minimum is attained.*
- *For any subset $T_i \in T$, if $T_i \cap \{u,v\} \neq \emptyset$, then define $T'_i = (T_i \cup \{x\}) - \{u,v\}$. In case $T_i \neq \{u,v\}$, it follows that $|T'_i| \geq 2$ and the new subset T'_i is added to T'. During this process, there could be two subsets T_i and T_j for which $T'_i = T'_j$. Obviously, the two subsets can be joined and their new penalty in p' is defined to be $p(T_i) + p(T_j)$.*

The following two lemmas establish the important property of minimal solutions to degree-weighted instances.

Lemma 1. *Every solution to a degree-weighted instance of the PCGSF problem has a total cost of at least n, where $n \geq 2$ is the number of active vertices.*

Proof. Consider some specific active vertex v. There is some subset $T_i \in T$ for which $t(v) = T_i$. In case the vertex v is not connected in the solution to all the other vertices of T_i, then we pay a penalty of 1 because of v. Otherwise, the solution must contain at least one edge incident with v. This means that the solution pays a cost of 1 for this edge due to vertex v. Every active vertex incurs a cost or penalty of at least 1, and therefore the total cost of the solution is at least n. ☐

Lemma 2. *Every minimal solution to a degree-weighted instance of the PCGSF problem has a total cost of at most $3n - 4$, where $n \geq 2$ is the number of active vertices.*

Proof. Examine some specific connected component of the solution that contains q active vertices. Since this is a minimal solution, all of its leaves are active vertices. We prove by induction on q that the cost of the edges in this connected component is at most $2(q-1)$. For $q = 1$, the connected component is a vertex with no edges, so this is obviously true. For $q = 2$, the connected component must be a path, whose two endpoints are active vertices, and the claim holds.

Suppose that $q > 2$. Take some leaf v, which by our assumption must be an active vertex. Now examine the path that starts from v and continues until the first time that either a vertex of degree at least 3 or another active vertex is reached. The cost of the edges in this path is at most 2. After removing this path from the solution, we are left with a connected component with $q-1$ active vertices whose leaves are all active vertices. The result now follows from the induction hypothesis.

For proving the lemma, we distinguish between two cases. If all the active vertices are in one connected component of the solution, then the solution is actually a tree and no penalties are paid. The cost of the edges in the solution is at most $2(n-1)$. The total cost, including the penalties, is also $2(n-1) \leq 3n - 4$, since $n \geq 2$. Otherwise, there are at least two connected components in the solution. The cost of the edges in all the components can be at most $2(n-2)$. Note that the sum of all penalties in the instance is exactly n. The solution pays at most n for penalties, so the total cost is at most $2(n-2)+n = 3n-4$. □

The last two lemmas determine the approximation ratio of the algorithm. The purpose of the next lemma is to show that the analysis is indeed tight.

Lemma 3. *For every $n \geq 2$, there exists a degree-weighted instance of the PCGSF problem on a graph with n vertices for which the optimal solution has total cost n whereas some minimal solution has a total cost of* $3n-4$.

Proof. Consider the following instance with vertices $v_1, \ldots, v_n$, which are all active. Between every two vertices there is an edge of cost 2. Define $p(\{v_1, v_2\}) = 2$ and $p(\{v_1, v_i\}) = 1$ for every $3 \leq i \leq n$. All other penalties are zero.

The optimal solution has no edges. The cost of the edges is zero and the sum of penalties paid is n for a total cost of n. A possible minimal solution is the path $v_2, v_3, \ldots, v_n$. The cost of the edges is $2(n-2)$ and the payment of penalties is still n for a total cost of $3n-4$. □

We now present the main algorithm of the paper.

Algorithm 1. $PCGSF(G, T, w, p)$

Input: Graph $G = (V, E)$, collection $T = \{T_1, \ldots, T_k\}$, each a subset of V of size at least 2, weight function $w : E \rightarrow \mathbb{R}^+$, penalty function $p : T \rightarrow \mathbb{R}^+$
Output: A forest $F \subseteq E$
if $T = \emptyset$ **then**
  return $\emptyset$
else
  The set of active vertices is defined as $Active \leftarrow \{v \in V | \exists i \ \ v \in T_i\}$
  For every edge $e \in E$, let $d(e)$ be the number of its active endpoints
  For every active vertex v, let $t(v)$ be an arbitrary subset T_i for which $v \in T_i$
  For every $T_i \in T$, define $d(T_i) = |\{v \in T_i | t(v) = T_i\}|$
  $\epsilon \leftarrow \min(\{w(e)/d(e) | e \in E, d(e) \neq 0\} \cup \{p(T_i)/d(T_i) | T_i \in T, d(T_i) \neq 0\})$
  Define a weight function w' as follows: $w'(e) = w(e) - d(e) \cdot \epsilon$ for all $e \in E$
  Define a penalty function p' as follows: $p'(T_i) = p(T_i) - d(T_i) \cdot \epsilon$ for all $T_i \in T$
  Let Z be the set of all edges $e \in E$ for which $w'(e) = 0$ and let Z' be a spanning forest of (V, Z)
  Let $(G', \{T'_1, \ldots, T'_k\}, w'', p'')$ be obtained from (G, T, w', p') by contracting the edges in Z'
  Let T' be the collection of subsets T'_i of size at least 2 for which $p''(T'_i) > 0$
  $F' \leftarrow PCGSF(G', T', w'', p'')$
  Let F'' be the forest obtained from the edges in G corresponding to those in F' together with Z'
  while *there is a leaf in F'' which is not an active vertex* **do**
    Remove from F'' the edge incident with that leaf
  return F''

Theorem 1. *There is a local ratio* $(3 - \frac{4}{n})$*-approximation algorithm for the PCGSF problem, where* $n \geq 2$ *is the number of vertices in the graph.*

Proof. The pseudocode of algorithm $PCGSF(G, T, w, p)$ that solves this problem appears above. The proof is by induction on $|E| + k$. Given an instance (G, T, w, p) of the problem, the functions w and p are decomposed by the algorithm, so that $w = w' + \delta$ and $p = p' + \gamma$, where the quadruple (G, T, δ, γ) is a constant multiple of a degree-weighted instance. It follows from Lemmas 1 and 2 that the solution F'' computed by the algorithm is a $(3 - \frac{4}{n})$-approximation for the instance (G, T, δ, γ).

Before the next recursive call, the algorithm either contracts an edge or reduced some penalty to zero. By the induction hypothesis, the recursive call to $PCGSF(G', T', w'', p'')$ returns a solution F' which is a $(3 - \frac{4}{n})$-approximation. Adding edges of cost zero to F' does not change the cost of the solution. Removing leaves that are not active vertices can only reduce the cost. Thus, it is easy to verify that the solution F'' computed by the algorithm is also a $(3 - \frac{4}{n})$-approximation for the instance (G, T, w', p'). It follows from the basic local ratio decomposition observation that F'' is a $(3 - \frac{4}{n})$-approximation for the instance $(G, T, w' + \delta, p' + \gamma)$, as needed.

As for the time complexity of the algorithm, the argument above shows that the algorithm performs at most $|E| + k$ recursive calls, and thus runs in polynomial time. □

3 The Prize Collecting Steiner Tree Problem

This section introduces the algorithm for the prize collecting Steiner tree (PCST) problem. An instance of this problem consists of a graph $G = (V, E)$, a root vertex $r \in V$, a subset $U \subseteq V - \{r\}$ of active vertices, a weight function $w : E \rightarrow \mathbb{R}^+$, and a penalty function $p : U \rightarrow \mathbb{R}^+$. Given an instance (G, r, U, w, p), the goal is to compute a tree rooted at r that minimizes the cost of the edges of the tree plus the penalties paid for vertices not in the tree. An *active vertex* is simply a vertex with positive penalty (the root vertex is not active). The set of active vertices is denoted by U in the problem instance. A *minimal solution* is a tree rooted at r whose leaves are active vertices (and possibly also r). In a *degree-weighted instance* corresponding to a graph $G = (V, E)$, a root r and a subset $U \subseteq V - \{r\}$ of active vertices, the weight function $w(e)$ is equal to the number of active endpoints of the edge e, whereas the penalty function satisfies $p(v) = 1$ for every $v \in U$ and $p(v) = 0$ otherwise. When the edge $\{u, v\}$ is *contracted*, the penalty of the new vertex created is $p(u) + p(v)$. This is except for when an edge $\{r, v\}$ incident with the root is contracted. In this case, the new vertex created is also called r and it still has zero penalty.

Lemma 4. *Every solution to a degree-weighted instance of the PCST problem has a total cost of at least* n*, where* n *is the number of active vertices.*

Proof. Let q be the number of active vertices connected to r in the solution. Each such vertex must have some edge incident with it in the solution and therefore a cost of at least 1 is paid for this vertex. For each of the $n - q$ active vertices that are not connected to r, a penalty of 1 is paid. □

Lemma 5. *Every minimal solution to a degree-weighted instance of the PCST problem has a total cost of at most* $2n - 1$*, where* n *is the number of active vertices.*

Proof. Let q be the number of active vertices connected to r in the solution. An argument similar to the one used in the proof of Lemma 2 gives that the cost of the edges in a minimal solution is at most $2q - 1$. The penalty paid is $n - q$ for a total cost of $2q - 1 + n - q = n + q - 1 \leq 2n - 1$, since $q \leq n$. □

The last two lemmas determine the approximation ratio of the algorithm. The purpose of the next lemma is to show that the analysis is indeed tight.

Lemma 6. *For every* $n \geq 1$*, there exists a degree-weighted instance of the PCST problem on a graph with* n *active vertices for which the optimal solution has total cost* n *whereas some minimal solution has a total cost of* $2n - 1$*.*

Proof. Consider the following instance with a root vertex r together with the vertices $v_1, \ldots, v_n$. For every $1 \leq i \leq n$, the vertex v_i is active and has a penalty of 1. Between every two vertices there is an edge of cost 2, except for edges between the root r and a vertex v_i that have a cost of 1.

The optimal solution has no edges. The cost of the edges is zero and the sum of penalties paid is n for a total cost of n. A possible minimal solution is the path $r, v_1, v_2, \ldots, v_n$. The cost of the edges is $1 + 2(n - 1)$ and the payment of penalties is zero for a total cost of $2n - 1$. □

Theorem 2. *There is a local ratio* $(2 - \frac{1}{n-1})$*-approximation algorithm for the PCST problem, where* $n \geq 2$ *is the number of vertices in the graph.*

Proof. The pseudocode of algorithm $PCST(G, r, U, w, p)$ that solves this problem appears below. The proof is analogues to that of Theorem 1 using Lemmas 4 and 5 instead of Lemmas 1 and 2. Note that since the root is not an active vertex, the maximum number of active vertices is $n - 1$. □

4 Concluding Remarks

- The integral solution computed by our algorithm for the prize collecting generalized Steiner forest problem can be compared with the optimal solution of the natural LP for this problem. It can be verified that the algorithm is still a $(3 - \frac{4}{n})$-approximation, even with respect to this optimal fractional solution. This is also implied by the equivalence between the local ratio method and the primal-dual schema.
- The algorithms presented in this paper use the local ratio method which seems like the most natural, simple and time-efficient framework for addressing prize collecting Steiner tree problems. This should be further explored to

Algorithm 2. $PCST(G, r, U, w, p)$

Input: Graph $G = (V, E)$, root vertex $r \in V$, subset $U \subseteq V - \{r\}$ of active vertices, weight function $w : E \to \mathbb{R}^+$, penalty function $p : U \to \mathbb{R}^+$
Output: A tree $T \subseteq E$ rooted at r
if $U = \emptyset$ **then**
  return $\emptyset$
else
  For every edge $e \in E$, let $d(e)$ be the number of its active endpoints
  $\epsilon \leftarrow \min(\{w(e)/d(e) | e \in E, d(e) \neq 0\} \cup \{p(v) | v \in U\})$
  Define a weight function w' as follows: $w'(e) = w(e) - d(e) \cdot \epsilon$ for every $e \in E$
  Define a penalty function p' as follows: $p'(v) = p(v) - \epsilon$ for every $v \in U$
  Let Z be the set of all edges $e \in E$ for which $w'(e) = 0$ and let Z' be a spanning forest of (V, Z)
  Let (G', r, U', w'', p'') be obtained from (G, r, U, w', p') by contracting the edges in Z'
  $T' \leftarrow PCST(G', r, U', w'', p'')$
  Let the tree T'' be the connected component of r in the union of the edges in G corresponding to those in T' together with Z'
  while *there is a leaf in T'' which is not an active vertex (or the root r)* **do**
    Remove from T'' the edge incident with that leaf
  return T''

determine whether this approach has applications for facility location problems and for the multicommodity rent-or-buy (MRoB) problem. The techniques presented in [11] might be helpful in enhancing the time performance of our algorithms.

- An interesting open problem is to decide whether there is a combinatorial algorithm for the prize collecting Steiner forest problem with an approximation factor better than 3. Improving upon the performance guarantee of the LP rounding 2.54-approximation algorithm is another intriguing challenge.

Acknowledgements

I would like to thank Noga Alon and Yossi Azar for helpful discussions.

References

1. Agrawal, A., Klein, P., Ravi, R.: When trees collide: An approximation algorithm for the generalized Steiner problem on networks. SIAM Journal on Computing 24(3), 440–456 (1995)
2. Bar-Noy, A., Bar-Yehuda, R., Freund, A., Naor, J.(S.), Schieber, B.: A unified approach to approximating resource allocation and scheduling. Journal of the ACM 48(5), 1069–1090 (2001)
3. Bar-Yehuda, R.: One for the price of two: a unified approach for approximating covering problems. Algorithmica 27(2), 131–144 (2000)

4. Bar-Yehuda, R., Bendel, K., Freund, A., Rawitz, D.: Local ratio: A unified framework for approximation algorithms: In memoriam: Shimon Even 1935–2004. ACM Computing Surveys 36(4), 422–463 (2004)
5. Bar-Yehuda, R., Rawitz, D.: On the equivalence between the primal-dual schema and the local ratio technique. SIAM Journal on Discrete Mathematics 19(3), 762–797 (2005)
6. Bienstock, D., Goemans, M.X., Simchi-Levi, D., Williamson, D.P.: A note on the prize collecting traveling salesman problem. Math. Program. 59, 413–420 (1993)
7. Goemans, M.X., Williamson, D.P.: A general approximation technique for constrained forest problems. SIAM Journal on Computing 24(2), 296–317 (1995)
8. Gupta, A., Könemann, J., Leonardi, S., Ravi, R., Schäfer, G.: An efficient cost-sharing mechanism for the prize-collecting Steiner forest problem. In: Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms, pp. 1153–1162 (2007)
9. Hajiaghayi, M., Jain, K.: The prize-collecting generalized Steiner tree problem via a new approach of primal-dual schema. In: Proceedings of the Seventeenth Annual ACM-SIAM Symposium on Discrete Algorithms, pp. 631–640 (2006)
10. Hayrapetyan, A., Swamy, C., Tardos, É.: Network design for information networks. In: Proceedings of the Sixteenth Annual ACM-SIAM Symposium on Discrete Algorithms, pp. 933–942 (2005)
11. Johnson, D.S., Minkoff, M., Phillips, S.: The prize collecting Steiner tree problem: theory and practice. In: Proceedings of the Eleventh Annual ACM-SIAM Symposium on Discrete Algorithms, pp. 760–769 (2000)
12. Könemann, J., Leonardi, S., Schäfer, G.: A group-strategyproof mechanism for Steiner forests. In: Proceedings of the Sixteenth Annual ACM-SIAM Symposium on Discrete Algorithms, pp. 612–619 (2005)
13. Sharma, Y., Swamy, C., Williamson, D.P.: Approximation algorithms for prize-collecting forest problems with submodular penalty functions. In: Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms, pp. 1275–1284 (2007)