On the Complexity of Budgeted Maximum Path Coverage on Trees

H.-C. Wirth*

An instance of the *budgeted maximum coverage* problem is given by a set of weighted ground elements and a cost weighted family of subsets of the ground element set. The goal is to select a subfamily of total cost of at most a given budget maximizing the weight of the covered elements. The *budgeted maximum graph coverage* problem is defined on graphs, where the vertex set plays the role of the ground elements, and covering sets are restricted to being connected subgraphs.

We show that the problem is NP-hard on trees even when the covering sets have unit cost and are restricted to the class of paths.

1 Introduction and Preliminaries

The *budgeted maximum coverage* problem is defined as follows: An instance specifies a set $X = \{x_1, \dots, x_n\}$ of ground elements with weight function $w\colon X \to \mathbb{R}_0^+$ and a family $F \subseteq 2^X$ of *covering sets* with associated costs $c\colon F \to \mathbb{R}_0^+$. The goal is to select a subfamily $F' \subseteq F$ which does not exceed a given constraint on the total cost and maximizes the weight of the covered elements.

The unit cost variant of the problem (where $c \equiv 1$) is known as the *maximum coverage* problem (see e. g. [Hoc97a] for a survey). A straightforward reduction from the Vertex Cover problem shows that the unweighted maximum coverage problem is NP-hard even if each ground element appears in no more than two sets.

The problem with general cost function $c \not\equiv 1$ has been investigated by Khuller et al. [KMN99]. The authors give an approximation algorithm with performance $(1 - 1/e) \approx 0.63$ and show that this is best possible unless NP $\subseteq$ DTIME$(N^{O(\log \log N)})$.

There is an alternative definition of the budgeted maximum coverage problem used by Ageev et al. [AS99, AS04]: "Given ground elements I, a family $F \subseteq 2^I$ with weights $w\colon F \to \mathbb{R}_0^+$, and an integer $p \in \mathbb{N}$, find a subset $X \subseteq I$ of the ground elements with $|X| = p$ which maximizes the total weight of the sets from F intersecting X." Comparing the two definitions, is appears that the role of ground elements and sets is interchanged.

*Tech. Rep. 438, Institut für Informatik, Universität Würzburg, 2007

Notice that a set of size k in the definition of Ageev et al. transforms into a ground element appearing in k sets in the definition employed by Khuller et al. We will stick to the notation used by Khuller et al. [KMN99] throughout the paper.

Definition 1.1 (Budgeted Maximum Graph Cover) An instance of Budgeted Maximum Graph Cover (GC for short) is given by an undirected simple graph $G = (V, E)$ with node weight function $w \colon V \to \mathbb{R}$, a weighted family $F = \{S_1, \ldots, S_{|F|}\}$ of connected subgraphs S_i with cost function $c \colon F \to \mathbb{R}_0^+$, and a budget value $B \in \mathbb{N}$. The goal is to find a subcollection $F' \subseteq F$ of subgraphs of total cost

$$c(F') := \sum_{S \in F'} c(S) \leq B\,,$$

such that the total weight $w(F')$ covered by the subcollection, defined by

$$w(F') := \sum_{v \in \bigcup_{S \in F'} S} w(v)\,,$$

is maximized.

By $\mathrm{GC}^{\mathrm{unit}}$ we denote the set of instances where all sets have cost 1. For any natural number k, we use "k-GC" to denote the fact that every member of the family F has at most k nodes. Also, for a graph class Γ, we use "Γ-GC" to denote that each member of the family belongs to the graph class Γ. Further, the notion "GC on Γ" means that the input graph G is restricted to graph class Γ.

Previous Results and Contribution of This Paper

The budgeted maximum graph coverage problem has been investigated in [KM$^+$02]. A straightforward reduction from Knapsack shows that even 1-GC is NP-hard, which suggests to concentrate on unit cost variants. On general graphs, 2-$\mathrm{GC}^{\mathrm{unit}}$ is polynomial time solvable while k-$\mathrm{GC}^{\mathrm{unit}}$ is NP-hard for $k \geq 3$. On paths, $\mathrm{GC}^{\mathrm{unit}}$ can be solved in polynomial time. On a star, the problem $\mathrm{GC}^{\mathrm{unit}}$ is NP-hard in general, but Path-$\mathrm{GC}^{\mathrm{unit}}$ (or, equivalently, 3-$\mathrm{GC}^{\mathrm{unit}}$) can be solved in polynomial time. On general trees, $\mathrm{GC}^{\mathrm{unit}}$ can be solved in polynomial time if the number of sets a ground element appears in is bounded by a constant but it is NP-hard if this frequency is unbounded. The complexity of path-$\mathrm{GC}^{\mathrm{unit}}$ on trees was unknown; we show in this report that this problem is NP-hard.

2 Path-$\mathrm{GC}^{\mathrm{unit}}$ on Trees

Theorem 2.1 *The problem path-GC^{unit} on trees is NP-hard.*

We show the result by a reduction from Directed Hamiltonian Cycle. An instance of Directed Hamiltonian Cycle is given by a digraph $G = (V, R)$ with node set $V = \{v_1, \ldots, v_n\}$ and arc set R. The decision problem whether the input graph contains a hamiltonian cycle (i.e., a cycle traversing each node exactly once) is NP-hard.

We construct a tree T with the node set

$$\{z\} \cup \{\, h_i, z_i \mid i = 1, \ldots, n \,\} \cup \{\, u^+_{i,j}, u^-_{i,j}, g_{i,j} \mid i, j = 1, \ldots, n \,\}\,.$$

Here the index j is referred to as the *level* of the nodes. All nodes carrying first index i form the *v_i-component* of the tree which reminds on the fact that they are related to the node v_i of the original graph. The edge set is defined as

$$\{\, (z, h_i), (h_i, z_i) \mid i = 1, \ldots, n \,\} \cup \{\, (z_i, u^+_{i,j}), (u^+_{i,j}, g_{i,j}), (z_i, u^-_{i,j}) \mid i, j = 1, \ldots, n \,\}$$

(see Figure 1 for an illustration). The weight of the u-nodes is set to $\Omega := n+1$, the weight of the h-nodes is set to 2, all other nodes have weight 1. The constructed tree contains $(3n+2)n+1 = 3n^2+2n+1$ nodes and a total weight of $w(T) = (2\Omega+1)n^2+3n+1$.

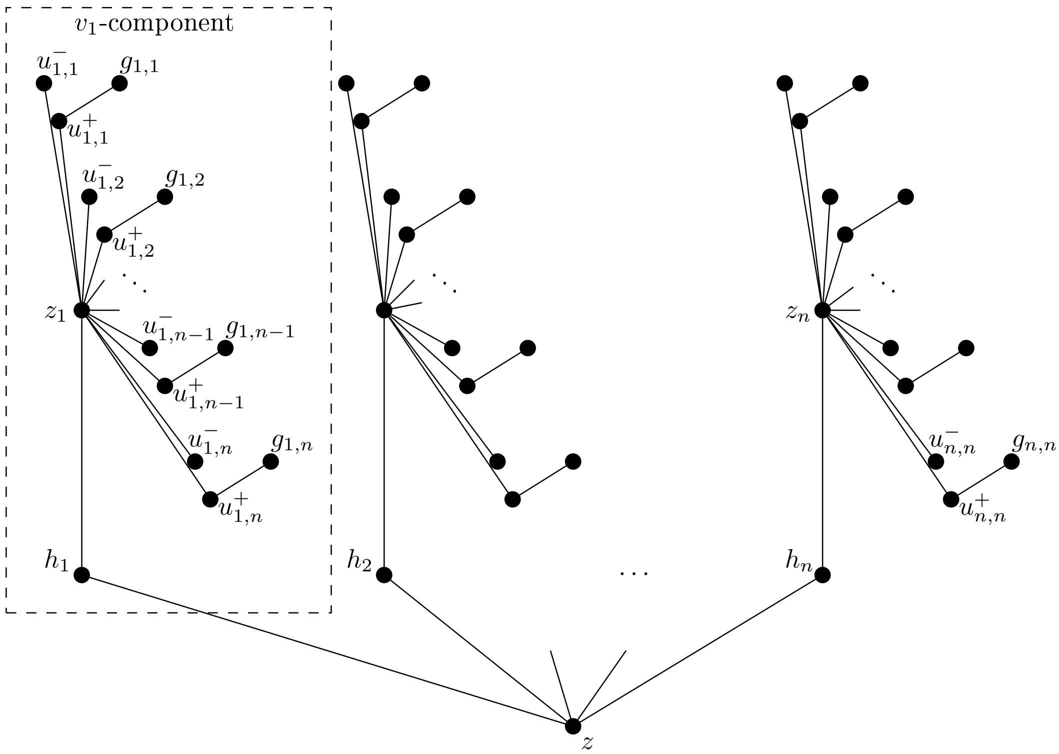

Figure 1: The constructed tree.

The covering paths are defined as follows: For each arc $(v_i, v_j) \in R$ we have n covering paths $P(u^+_{i,1}, u^-_{j,2})$, $P(u^+_{i,2}, u^-_{j,3})$, $\ldots$, $P(u^+_{i,n}, u^-_{j,n+1})$ called *arc paths* (here and in the sequel all index operations are performed modulo n so that they are always mapped to the index set $\{1, 2, \ldots, n\}$). Each arc path connects a u^+-node of one level to a u^--node of the next level without meeting any other u-node or g-node. Additionally we have n^2 further paths called *gap paths*, namely for each $i, j = 1, \ldots, n$ the path $P(g_{i,j}, u^-_{i,j})$. Observe that a gap path covers locally a triple of g, u^+, u^--nodes of the same level and does not meet any h-node.

Claim 2.2 *The maximum weight which can be covered in the tree with $B := n^2$ many sets is at least $W := (2\Omega{+}1)n^2{+}2n{+}1$ if and only if the digraph G contains a hamiltonian cycle.*

"If": Let w.l.o.g. the nodes are numbered such that the hamiltonian cycle meets the nodes in order $v_1, v_2, \ldots, v_n, v_1$. For each $i = 1, \ldots, n$, choose the arc path $P(u^+_{i,i}, u^-_{i+1,i+1})$ and the $n-1$ gap paths $P(g_{i,j}, u^-_{i,j})$, $j \neq i$. Thus the budget constraint of n^2 paths is satisfied. Consider the v_i-component of the tree. The node $u_{i,i}$ and the nodes h_i, z_i are covered by arc paths. For $j \neq i$, the $u_{i,j}$-nodes and the nodes $g_{i,j}$ are covered by gap paths. This yields a gain of $3 + (2\Omega+1)n - 1 = (2\Omega+1)n + 2$ per component and hence a total gain of $1 + n((2\Omega+1)n+2) = (2\Omega+1)n^2 + 2n + 1 = W$ in the tree.

"Only if": Let there be a covering of at most $B = n^2$ paths which achieves a total weight of at least W. Since $w(T) - W = n < \Omega$ each of the u-nodes must be covered, and since the number of those nodes equals the number B of covering sets, no u^+-node is covered more than once and similarly no u^--node is covered more than once.

Consider a single v_i-component. It is not possible that simultaneously all $g_{i,j}$ and the node h_i are covered: Covering all g-nodes implies covering all u-nodes by gap paths, and covering additionally h_i would imply covering one u-node for a second time. Hence the maximum gain per component is $(2\Omega+1)n + 2$.

We claim that from the set of all $g_{i,j}$ and h_i all nodes but one single g-node are covered: Otherwise the gain in that component were at most $(2\Omega+1)n+1$ which would add with the maximal gain in the remaining components to a total maximum of

$$1 + (n-1)((2\Omega+1)n+2) + (2\Omega+1)n + 1 = (2\Omega+1)n^2 + 2n = W - 1$$

contradicting the premised minimum gain of W.

As a consequence, for each v_i-component there is exactly one level j where both u-nodes are covered by two arc paths, and all other levels are covered by gap paths. Since by construction each arc path ascends up by one level the chain of components traversed by following the arc paths meets $n-1$ other components of the tree before it can return to the same component again, thus it forms a hamiltonian cycle in the original graph.

Corollary 2.3 *The problem 7-path-GC^{unit} on trees is NP-hard.* □

3 Conclusions

Table 1 displays a completed version of the overview from [KM$^+$02] on the complexity of the Budgeted Maximum Graph Cover problem.

References

[AS99] A. A. Ageev and M. I. Sviridenko · *Approximation algorithms for maximum coverage and max cut with given sizes of parts* · Proceedings of 7th Conference on Integer Programming and Combinatorial Optimization (IPCO'99), Lecture Notes in Computer Science, vol. 1610, 1999, pp. 17–30.

	unit cost			general cost
	on paths	on trees	on general graphs	
1-GC	(P)	(P)	(P)	NP-hard (even on paths)
2-GC	(P)	P	P	
3-GC	(P)	(unknown)	NP-hard	
path-GC	P	NP-hard (even 7-path-GC)	(NP-hard)	FPAS on paths
tree-GC	(not defined)	P (with restricted frequency)	(NP-hard)	
		NP-hard (even on stars)		

Table 1: Overview of the complexity of problem GC. Results in parentheses follow from the explicit result stated in that column.

[AS04] A. A. Ageev and M. I. Sviridenko · *Pipage rounding: a new method of constructing algorithms with proven performance guarantee* · Journal of Combinatorial Optimization **8** (2004), no. 3, 307–328.

[Hoc97a] D. Hochbaum · *Approximation covering and packing problems* · in [Hoc97b], 1997.

[Hoc97b] D. S. Hochbaum (ed.) · *Approximation algorithms for NP-hard problems* · PWS Publishing Company, Boston, 1997.

[KM+02] S. O. Krumke, M. V. Marathe, D. Poensgen, S. S. Ravi, and H.-C. Wirth · *Budgeted maximum graph coverage* · Proc. 28th International Workshop on Graph-Theoretic Concepts in Computer Science (WG'02), Lecture Notes in Computer Science, vol. 2573, Springer, 2002, pp. 321–332.

[KMN99] S. Khuller, A. Moss, and J. Naor · *The budgeted maximum coverage problem* · Information Processing Letters **70** (1999), 39–45.