The Generalized Steiner Cable-Trench Problem with Application to Error Correction in Vascular Image Analysis

Eric Landquist, Francis J. Vasko, Gregory Kresge, Adam Tal, Yifeng Jiang, and Xenophon Papademetris

Abstract The Cable-Trench Problem (CTP) is the problem of minimizing the cost to connect buildings on a campus to a central server so that each building is connected directly to the server via a dedicated underground cable. The CTP is modeled by a weighted graph in which the vertices represent buildings and the edges represent the possible routes for digging trenches and laying cables between two buildings. In this paper, we define the Generalized Steiner CTP (GSCTP), which considers the situation in which a subset of the buildings is connected to the server and also the possibility that trench costs vary because of vegetation or physical obstacles, for example. The GSCTP has several natural applications, but we will focus on its nontrivial and novel application to the problem of digitally connecting microCT scan data of a vascular network with fully automated error correction. The CTP and its variants are NP-hard. However, we show that modifications to Prim's algorithm find nearly optimal solutions to the GSCTP efficiently.

1 Introduction

The Cable-Trench Problem (CTP) was first described in [6] and establishes a continuum between the Minimum Spanning Tree and Shortest Path Tree Problems. The name "Cable-Trench" comes from the problem of minimizing the cost to connect buildings on a campus to a central server so that each building is connected directly to the server via a dedicated underground cable. The problem is modeled as a weighted graph in which the buildings are represented by vertices and the edges

Eric Landquist · Francis J. Vasko · Gregory Kresge · Adam Tal
Kutztown University, Kutztown, PA 19530, USA e-mail: {elandqui, vasko}@kutztown.edu, {gkres121, atal822}@live.kutztown.edu

Yifeng Jiang · Xenophon Papademetris
Yale University School of Medicine, 310 Cedar Street, P.O. Box 208042, New Haven, CT 06520-8042, USA e-mail: jiang1feng@gmail.com, xenophon.papademetris@yale.edu

represent the possible routes for digging trenches and laying cables between two buildings. Weights on the edges generally represent distance. The Generalized CTP (GCTP) considers the possibility that the cost of digging a trench varies because of vegetation, soil composition, or physical obstacles, for example [7]. The Generalized Steiner CTP (GSCTP) further supposes that some subset of the buildings is connected to the server, though cables may be routed through any building.

In Sect. 2, we describe the application of the GSCTP to the problem of digitally reconstructing a blood vessel network from microCT scan image data and eliminating errors in the data. We describe heuristics that we used to quickly compute nearly optimal solutions to the GSCTP in Sect. 3 and tabulate results of our experiments in Sect. 4. The paper closes with some conclusions and areas of future work in Sect. 5. Here, we give a graph-theoretic description of the GSCTP.

Let $G=(V,E)$ be a connected graph with vertex set $V=\{v_1,\ldots,v_n\}$, root vertex v_1, edge set E, and $s_{ij}\geq 0$ and $t_{ij}\geq 0$ the "cable" and "trench" weights of the edge $(v_i,v_j)\in E$, respectively. Let $F\subseteq V$ be the set of terminal vertices, $N\subseteq V$ the set of nonterminal vertices, and let γ and τ denote the per-unit cable and trench costs, respectively. We define the GSCTP as the problem of finding a tree $T=(V_T,E_T)$, such that $\{v_1\}\cup F\subseteq V_T\subseteq V$ and $E_T\subseteq E$, which minimizes $\gamma w_c(T)+\tau w_t(T)$, where

$$w_c(T)=\sum_{v_k\in F}\sum_{(v_i,v_j)\in\mathscr{P}(v_1,v_k)} s_{ij} \quad \text{and} \quad w_t(T)=\sum_{(v_i,v_j)\in E_T} t_{ij} \tag{1}$$

are the total cable weight of T and total trench weight in T, respectively, and $\mathscr{P}(v_1,v_k)\subseteq E_T$ is the path in T from v_1 to the terminal vertex v_k. Vertices in $V_T\cap N$ are called Steiner vertices. The CTP is the special case in which $N=\emptyset$ and $s_{ij}=t_{ij}$ for all i and j. Further, if $\gamma>0$ and $\tau=0$, then a solution to the CTP is any shortest path spanning tree of G with root vertex v_1. In contrast, if $\tau>0$ and $\gamma=0$, then a solution to the CTP is any minimum spanning tree of G. Note that if $\tau=0$, then an optimal solution could contain cycles formed from "empty" trenches. We will not consider such solutions because in practice, we want to utilize every trench.

We refer the reader to [7] for a description of further applications and extensions of the CTP. To motivate our definition of the GSCTP, however, we will describe a nontrivial application to vascular image analysis due to Jiang et al. [2].

2 Application to Vascular Image Analysis

A massive set of discrete points, representing the locations of blood vessels, are first detected from 3D medical images, such as CT and microCT. The vessel radii at these points can also be estimated from the images. These points correspond to the vertices, V, of a complete graph $G=(V,E)$, with edge weights determined by Euclidean length, vessel segment volume, or some physiological factor. The goal is to digitally represent the vessel network (vasculature) as a subtree $T\subseteq G$ as accurately as possible in order to assist critical vasculature-related research, including angiogenesis and cancer detection. This task currently constitutes a bottleneck in quantitative vascular research [8]. Prior to the GCTP model of [2, 7], the best methods computed the minimum spanning tree of G, but the results depended heavily on manual correction [1, 3, 5]. Specifically, [2, 7] applied Murray's Minimum Work Principle [4], which states that any vascular network tends to minimize the total work due to blood flow resistance and metabolic support for the blood volume. These two factors are proportional to the cable cost and trench cost in the GCTP, respectively. Thus, the vessel connection problem is formulated as a GCTP with cable weights $s_{ij} = \ell(e_k)/r(e_k)$ and trench weights $t_{ij} = \ell(e_k)r(e_k)^2$, where $\ell(e_k)$ is the length of the blood vessel represented by the edge $e_k = (v_i, v_j)$ and $r(e_k)$ is the radius of the vessel at v_j. In this application, if $\gamma = 1$, then $35000 \leq \tau \leq 175000$ is an appropriate range. We refer the reader to [2] for the technical details of their derivation and to [7] for results on the GCTP treatment of this problem.

In real image analysis scenarios, however, the data invariably contains errors, i.e., false positive vessel points detected from images. One can determine the leaves of the solution tree, typically those points within the region perfused by the vascular tree. We can therefore model the vascular imaging problem as a GSCTP by letting F be the set of known leaf vertices. The set of errors, then, is a subset of N. We assume that the optimal solution to a GSCTP model of a vessel connection problem will yield an image as close as possible to the actual vascular network.

Vasko et al. showed that the CTP is NP-hard [6], so the GSCTP is NP-hard. Thus, it is computationally infeasible to determine optimal solutions of very large GSCTPs, such as those arising from the application at hand. In order to efficiently find nearly optimal solutions of the GSCTP, we modified Prim's algorithm.

3 Modifications to Prim's Algorithm

In [7], Vasko et al. extended Prim's algorithm to find nearly optimal solutions of the GCTP. In this section, we describe further modifications that allow one to find nearly optimal solutions to the GSCTP. First, we describe a modification of Prim's algorithm, which we call the Generalized Steiner Modified Prim's heuristic (`GSMPrim`). This will include what we call a *benefit* function, which is designed to encourage the selection of terminal vertices as well as Steiner vertices that are adjacent to or sufficiently close to multiple terminal vertices. We then describe two variants of `GSMPrim`: one semi-greedy and deterministic and one partially stochastic.

In order to define the benefit function, we first let $|\cdot| : E \to \mathbb{R}_{\geq 0}$ be some fixed edge metric (e.g., cable or trench weight), and set $B \in \mathbb{R}_{\geq 0} \cup \{\infty\}$ so that $|(v,w)| < B$ implies that $v, w \in V$ are near each other. The metric $|\cdot|$ and bound B will vary depending on the application. Now, the benefit function, $b : V \to \mathbb{N}_0$, is defined

$$b(v) = \begin{cases} 2 + \#\{w \in F : (v,w) \in E \setminus E_T \text{ and } |(v,w)| < B\} & \text{if } v \in F \\ \#\{w \in F : (v,w) \in E \setminus E_T \text{ and } |(v,w)| < B\} & \text{if } v \in N \ . \end{cases} \quad (2)$$

We also define a positive multiplier M and let $W = (\tau/\gamma)M \max_{e\in E}\{|e|\}$.

Algorithm 1: Generalized Steiner Modified Prim's Heuristic (`GSMPrim`)

Input : $G = (V,E)$, F, s_{ij} and t_{ij}, γ, τ, B, and W
Output: A tree $T = (V_T, E_T) \subseteq G$, such that $\{v_1\} \cup F \subseteq V_T$
1 $V_T := \{v_1\}$, $E_T := \{\}$, $cost := d := \{\infty, \infty, \ldots, \infty\}$, $Pre := \{1, 1, \ldots, 1\}$;
2 **for** $2 \leq i \leq n$ **do**
3 $\quad cost[i] := \gamma s_{1i} + \tau t_{1i}$
4 **while** $\{v_1\} \cup F \not\subseteq V_T$ **do**
5 $\quad m :=$ index $(\min\{cost[i] - Wb(v_i) : v_i \in V \setminus V_T\})$;
6 $\quad V_T := V_T \cup \{v_m\}$ and $E_T := E_T \cup \{(v_{Pre[m]}, v_m)\}$;
7 $\quad$ **for** $v_i \in V \setminus V_T$ *such that* $(v_m, v_i) \in E$ **do**
8 $\quad\quad$ **if** $cost[i] > \gamma(d[m] + s_{mi}) + \tau t_{mi}$ **then**
9 $\quad\quad\quad d[i] := d[m] + s_{mi}$, $cost[i] := \gamma(d[m] + s_{mi}) + \tau t_{mi}$, $Pre[i] := m$;
10 Remove all leaves in N from V_T so that all leaves of $T = (V_T, E_T)$ are in F.

Since we add a new vertex and edge with each iteration of the while loop, `GSMPrim` will terminate. Moreover, it requires $O\left(|V|^2\right)$ time and space.

The two variants of `GSMPrim` are multi-pass modifications that generate multiple solution trees. The semi-greedy variant, `SG-GSMPrim`, selects the kth best edge and vertex at Step 5 of `GSMPrim` for the first edge of the kth solution tree T. Thereafter, it selects every edge and vertex in a greedy fashion. The partially stochastic variant, `PS-GSMPrim`, selects an initial fraction of the vertices and edges at Step 5 in a stochastic manner, and proceeds in a greedy fashion thereafter. Specifically, we let $[p_1, p_2, \ldots, p_r]$ be a probability distribution, with $p_i \geq p_{i+1}$ for all $1 \leq i < r$. At Step 5, the ith best vertex and edge is selected with probability p_i, for $1 \leq i \leq r$. In this way, `PS-GSMPrim` can generate any number of solution trees.

Note that in the application to vascular image analysis error correction, we assume that nonterminal leaves are in fact errors in the imaging process. Thus, Step 10 of `GSMPrim` and its variants automatically eliminates errors from the image data.

4 Results

We tested `GSMPrim` and its variants on a pair of small graphs and on a 25001-vertex data set generated from a microCT scan of the vasculature of a mouse leg. In each case, we let $\gamma = 1$ and considered a collection of values of τ. For the benefit function, we used $|e| = \ell(e)$ and $B = (1/10)\max_{e\in E}\{\ell(e)\}$.

For Tables 1 and 2, we used five sets of terminal vertices on each graph and averaged the results. In Table 1, we compared `GSMPrim` with $M = 0$ and $M = 1$ to the optimal solutions, which were found using LINDO. The last column shows the percentage improvement in `GSMPrim` when using the benefit function. In Tables 1 and 2, the optimal solutions are used as the benchmark to test our heuristic.

The graph for Table 2 is taken from the microCT scan data: the root and its closest 20 points. For Tables 2 and 3, we tested multipliers M from the set

Table 1 `GSMPrim` on a 9-vertex SCTP from Example 4 of [6]

τ	Opt.	$M=0$	% Dev.	$M=1$	% Dev.	% Impr.
0.1	55.54	55.54	0	55.54	0	0
1	84	85.80	2.14	84	0	2.10
5	201.6	249.4	23.7	204.2	1.29	18.1
30	922.6	1134.4	22.5	929.2	0.72	18.1

$\{10^{-5}, 10^{-4}, 10^{-3}, 0.01, 0.1, 0.2, \ldots, 1.5, 2, 2.5, 3\}$. The last three columns in Tables 2 and 3 give the percentage improvement of `GSMPrim` with the benefit function, `SG-GSMPrim`, and `PS-GSMPrim` over `GSMPrim` with $M = 0$, respectively. For each example, we ran 20 iterations of `SG-GSMPrim` and 30 of `PS-GSMPrim` using the probability distribution $[1/3, 2/9, 2/9, 1/9, 1/9]$. The first 15 iterations chose the first five vertices stochastically and the next 15 did so with the first ten vertices.

Table 2 `GSMPrim` on a 21-vertex GSCTP

τ	$M=0$, % Dev./Opt.	Best M	% Impr./$M=0$	`SG` % Impr.	`PS` % Impr.
0.01	4.19	0.00001	0.0003	0	0
1	4.19	0	0	0.00009	0
5	4.19	0	0	0.006	0
10	4.19	0	0	0.006	0
100	4.06	0	0	0.006	0
10000	5.41	0.001	6.51	8.43	3.58
50000	16.2	0.001	4.39	3.18	6.33
100000	16.2	0.001	1.76	2.54	6.75
150000	15.2	0.001	3.53	2.34	6.80

In Table 3, we focus on the vascular image analysis problem and data, so we only tested those values of τ appropriate for this application. We chose 22500 vertices at random to be the set F. In this case, we ran 30 iterations each of `SG-GSMPrim` and `PS-GSMPrim`. The first 15 iterations of `PS-GSMPrim` chose the first 1500 vertices stochastically and the last 15 iterations chose the first 2500 vertices stochastically. `GSMPrim` with $M = 0$ is used as a benchmark heuristic.

Table 3 `GSMPrim` on a 25001-vertex Vascular Data Set

τ	Best M	% Impr./$M=0$	`SG` % Impr.	`PS` % Impr.
10000	0.01	4.59	1.31	0.63
50000	0.01	4.35	0.93	1.30
100000	0.01	2.15	0.90	1.42
150000	0.01	0	0.85	1.77

We ran `GSMPrim` and its variants in MATLAB® on a PC running Windows 7 Professional with an Intel I7-3930K 3.2 GHz processor and 16 GB of RAM. Each run on the 25001-vertex examples took an average of 34.0 seconds.

5 Conclusions and Future Work

In this paper, we defined the GSCTP and applied it to solve the problem of algorithmic error correction in vascular image analysis. We developed three efficient variants of a heuristic based on Prim's algorithm that are capable of finding nearly optimal solutions to GSCTPs. Generally, as τ increased, the accuracy of `GSMPrim` declined, but its variants yielded significant improvements. In particular, using the benefit function was the most effective approach for the largest graphs.

Since the CTP and its variants are a relatively new and unexplored area of study, there are several avenues to consider for future work, both theoretically and experimentally. For the GSCTP in particular, we would like to experiment with genetic algorithms and use LP relaxation techniques to determine good lower bounds of optimal solutions, to test the effectiveness of `GSMPrim` on larger graphs.

Acknowledgements The authors thank Dr. Albert Sinusas and Zhenwu Zhuang, Yale University School of Medicine, for providing microCT scan image data of a mouse leg for our experiments.

References

1. Bullitt, E., Aylward, S., Liu, A., Stone, J., Mukherji, S., Coffey, C., Gerig, G., Pizer, S.: 3D graph description of the intracerebral vasculature from segmented MRA and tests of accuracy by comparison with X-ray angiograms. In: A. Kuba, M. Sámal, A. Todd-Pokropek (eds.) Proc. IPMI, *LNCS*, vol. 1613, pp. 308–321. Springer, Berlin (1999)
2. Jiang, Y., Zhuang, Z., Sinusas, A., Staib, L., Papademetris, X.: Vessel connectivity using Murray's hypothesis. In: G. Fichtinger, A. Martel, T. Peters (eds.) Proc. MICCAI 2011, *LNCS*, vol. 6893, pp. 528–536. Springer, Berlin (2011)
3. Jomier, J., Ledigarcher, V., Aylward, S.: Automatic vascular tree formation using the Mahalanobis distance. In: J. Duncan, G. Gerig (eds.) Proc. MICCAI 2005, *LNCS*, vol. 3750, pp. 806–812. Springer, Berlin (2005)
4. Murray, C.: The physiological principle of minimum work. i. the vascular system and the cost of blood volume. Proc. Natl. Acad. Sci. **12**(3), 207–214 (1926)
5. Szymczak, A., Stillman, A., Tannenbaum, A., Mischaikow, K.: Coronary vessel trees from 3D imagery: a topological approach. Med. Image Anal. **10**(4), 548–559 (2006)
6. Vasko, F., Barbieri, R., Rieksts, B., Reitmeyer, K., Stott, K.: The cable trench problem: combining the shortest path and minimum spanning tree problems. Comput. Oper. Res. **29**(5), 441–458 (2002)
7. Vasko, F., Landquist, E., Kresge, G., Tal, A., Jiang, Y., Papademetris, X.: A simple and efficient strategy for solving very large-scale generalized cable-trench problems. Netw.: Int. J. **67**(3), 199–208 (2016)
8. Zagorchev, L., Oses, P., Zhuang, Z., Moodie, K., Mulligan-Kehoe, M., Simons, M., Couffinhalt, T.: Micro computed tomography for vascular exploration. J. Angiogenesis Res. **2**(1), 7 (2010)