Blood, Sweat, and Tears: The Cable-Trench Problem and Some Applications

Eric Landquist
elandqui@kutztown.edu

Department of Mathematics
Kutztown University of Pennsylvania

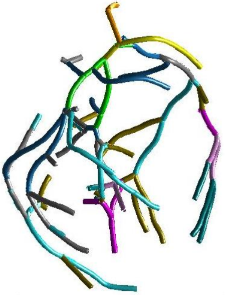
April 11, 2015

Joint work with:
Fran Vasko, Greg Kresge, and Adam Tal (Kutztown),
Yifeng Jiang and Xenophon Papademetris (Yale University)

Outline

- Motivation: Analyzing CT Scans
- Background: The Cable-Trench Problem (CTP)
- The Generalized Steiner Cable-Trench Problem (GSCTP)
- Application to Vascular Image Analysis
- Modified Prim's Algorithms
- Results
- More Applications
- Future Work and Open Questions

Motivation: Why do we care?



- (Blood)
- Given: Data from a Micro-CT scan (combined microscope and CT scan).
- Digitally reconstruct a blood vessel network (vasculature) from a set of points and vessel radii and
- Automatically filter out errors,
- Accurately and efficiently.
- Study vasculature noninvasively for cancer detection, identifying blood clots, etc.

The Problem: Vascular Image Analysis from a CT Scan

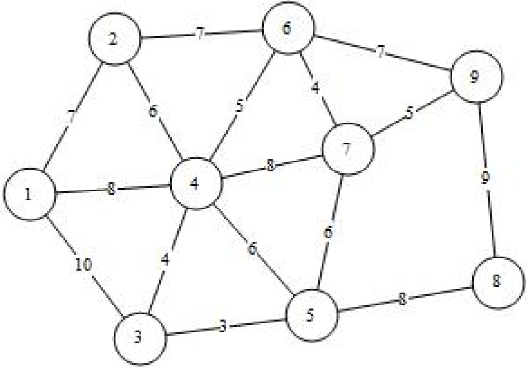
- Example: This scan of a mouse's leg has 27,873 vertices (3D).
- The red dot represents the main artery coming into the leg.
- Can we accurately and efficiently connect the dots to visualize the blood vessel network?

Background: The Cable-Trench Problem

- Suppose you needed to connect each building on a campus to a hub (IT building) with its own (underground) internet cable.
- How should you dig the trenches between buildings to minimize the total cost to dig the trenches and lay the cable?
- If the cables are free, then the solution is a Minimum Spanning Tree (MST).
- If the trenches are free, then the solution is a Shortest Path Tree (SPT).
- In the real world, both cables and trenches have a cost.
- The **Cable-Trench Problem** (CTP) establishes a continuum between the SPT and MST Problems on a weighted graph.

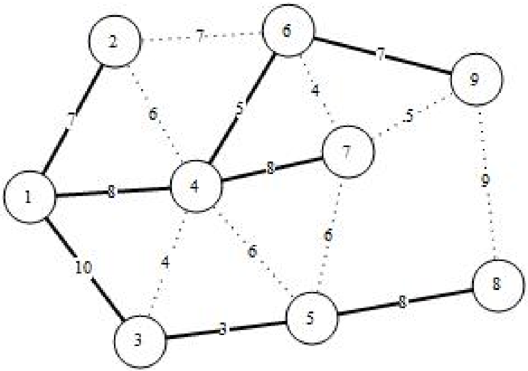
An Example: SPT vs. MST

Circles (nodes, vertices) represent the buildings; Node 1 is the hub.
Edges represent allowable routes for digging trenches.
Numbers (weights) on the edges represent their lengths.



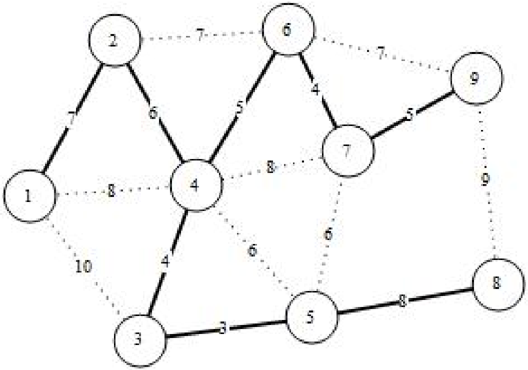
Shortest Path Tree (SPT)

If trenches are free, this minimizes the total cable length (cost).
This is found using **Dijkstra's Algorithm**.



Minimum Spanning Tree (MST)

If cables are free, this minimizes the total trench length (cost).
This is found using **Prim's Algorithm**.



Background: The Cable-Trench Problem

- Let $G = (V, E)$ be a weighted graph, $V = \{v_1, \ldots, v_n\}$, with root node v_1, and $|e|$ the weight of $e \in E$.
- $|T|_\gamma$ is the total weight of the paths, $\mathcal{P}(v_1, v_j)$, from v_1, to each $v_j \in V$ of a spanning tree T of G: **total cable length**.
- $|T|_\tau$ is the total weight of T: **total trench length**.
- Let $\gamma \geq 0$ be the **per-unit cable cost** and
- $\tau \geq 0$ the **per-unit trench cost**.

Definition

The **CTP** is the problem of finding a spanning tree T of G that minimizes the total cable and trench costs:

$$\gamma|T|_\gamma + \tau|T|_\tau = \gamma \sum_{j=2}^{n} \sum_{e_k \in \mathcal{P}(v_1, v_j)} |e_k| + \tau \sum_{k=1}^{n-1} |e_k| \ .$$

A Generalized Cable-Trench Problem

- What if the edge weights don't measure both the trench and cable weights, due to rocks, obstacles, overhead costs etc.?
- For $e \in E$, let $|e|_\gamma$ be the **cable weight** of e and
- $|e|_\tau$ the **trench weight** of e.
- $|T|_\gamma$ and $|T|_\tau$ are the total cable length and total trench length of the tree T.
- $\gamma, \tau \geq 0$ are the per-unit cable and trench costs, respectively.

Definition

The **Generalized Cable-Trench Problem** (GCTP) is the problem of finding a spanning tree T of G that minimizes

$$\gamma |T|_\gamma + \tau |T|_\tau = \gamma \sum_{j=2}^{n} \sum_{e_k \in \mathcal{P}(v_1, v_j)} |e_k|_\gamma + \tau \sum_{k=1}^{n-1} |e_k|_\tau \quad .$$

A Generalized Steiner Cable-Trench Problem

- What if we require the solution tree T to contain a given subset, $F \subseteq V$?
- F is called the set of **terminal nodes**,
- and $N = V \setminus (F \cup \{v_1\})$ the set of **nonterminal nodes**.
- $|T|_\gamma$ and $|T|_\tau$ are the total cable length and total trench length of the tree T.
- $\gamma, \tau \geq 0$ are the per-unit cable and trench costs.

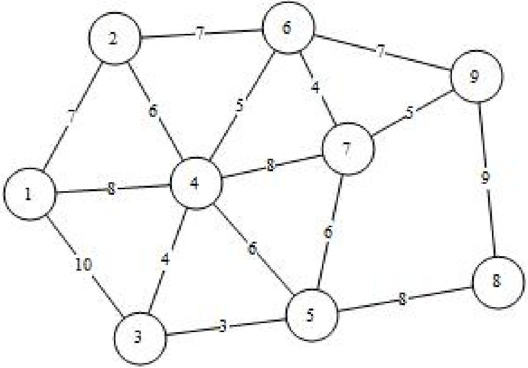
Definition

The **Generalized Steiner Cable-Trench Problem** (GSCTP) is the problem of finding a subtree $T = (V_T, E_T)$ of G such that $\{v_1\} \cup F \subseteq V_T$ that minimizes

$$\gamma|T|_\gamma + \tau|T|_\tau = \gamma \sum_{v_j \in F} \sum_{e_k \in \mathcal{P}(v_1, v_j)} |e_k|_\gamma + \tau \sum_{k=1}^{m} |e_k|_\tau \ .$$

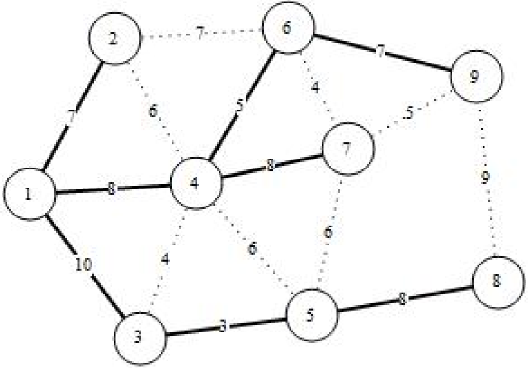
A CTP Example: Revisited

What are the optimal CTP solutions?



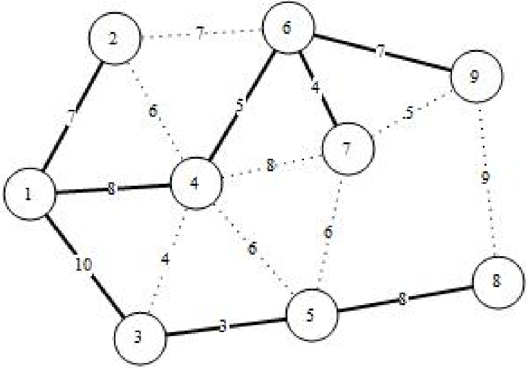
T_1: for $0 \le \tau/\gamma \le 1/4$; Cost: $56\tau + 108\gamma$ (SPT)

$\gamma, \tau \ge 0$ are the per-unit cable and trench costs, respectively.



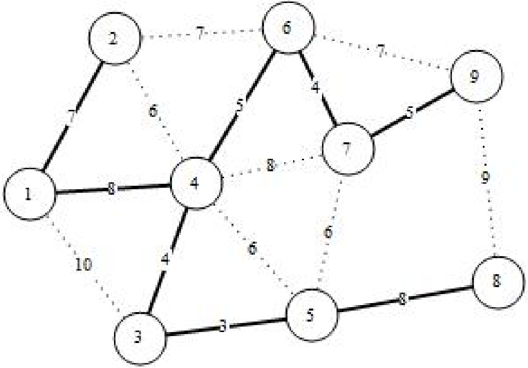
T_2: for $1/4 \leq \tau/\gamma \leq 1$; Cost: $52\tau + 109\gamma$

$\gamma, \tau \geq 0$ are the per-unit cable and trench costs, respectively.



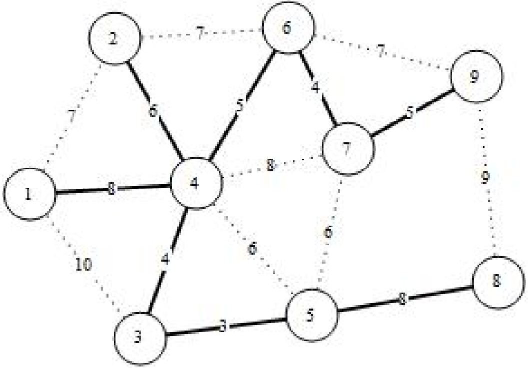
T_3: for $1 \le \tau/\gamma \le 7$; Cost: $44\tau + 117\gamma$

$\gamma, \tau \ge 0$ are the per-unit cable and trench costs, respectively.



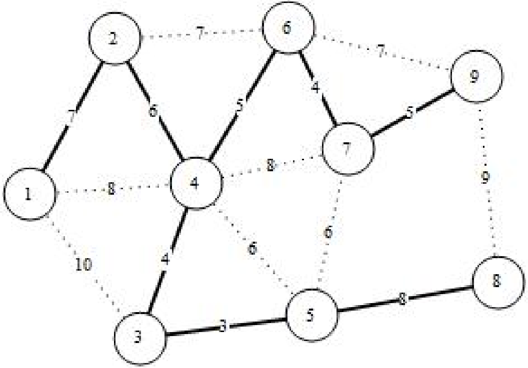
T_4: for $7 \leq \tau/\gamma \leq 28$; Cost: $43\tau + 124\gamma$

$\gamma, \tau \geq 0$ are the per-unit cable and trench costs, respectively.



T_5: for $28 \le \tau/\gamma$; Cost: $42\tau + 152\gamma$ (MST)

$\gamma, \tau \ge 0$ are the per-unit cable and trench costs, respectively.



Vascular Image Analysis and Murray's Principle (1926)

- Good solution (previous state-of-the-art): Find the MST of the graph from a Micro-CT scan via Prim's Algorithm, using vessel segment volume as the edge weights.
- Better idea: Incorporate Murray's Minimum Work Principle.
- Work in a circulatory system is primarily due to two factors:
    1. Overcome blood flow resistance (friction; inversely proportional to vessel radius), and
    2. Metabolic support for the blood volume (proportional to vessel segment volume).
- The body minimizes the total work due to these two factors.
- Applied by Jiang, et al. to improve on the MST model. (2011)

The GCTP Model for Vascular Image Analysis

- "Cables:" Work to overcome friction
- "Trenches:" Work due to metabolic support
- Notation:
  - $r(e_k)$ = Radius of vessel segment $e_k \in E$
  - $|e_k|$ = Length of vessel segment $e_k \in E$
- Total Cable Length:

$$|T|_\gamma = \sum_{j=2}^{n} \sum_{e_k \in \mathcal{P}(v_1, v_j)} \frac{|e_k|}{r(e_k)} \ , \text{ so } \ |e_k|_\gamma = \frac{|e_k|}{r(e_k)}$$

- Total Trench Length:

$$|T|_\tau = \sum_{k=1}^{n-1} |e_k| \cdot r^2(e_k) \ , \text{ so } \ |e_k|_\tau = |e_k| \cdot r^2(e_k)$$

The GCTP Model for Vascular Image Analysis

- Find edges $e_1, \ldots, e_{n-1} \in E$ to minimize the total work:

$$\gamma|T|_\gamma + \tau|T|_\tau = \gamma \sum_{j=2}^{n} \sum_{e_k \in \mathcal{P}(v_1, v_j)} \frac{|e_k|}{r(e_k)} + \tau \sum_{k=1}^{n-1} |e_k| \cdot r^2(e_k)$$

- $\gamma = 1$ and $35000 \leq \tau \leq 175000$ for this application.

Error-correction in Vascular Imaging

- Problem: There are errors in the imaging process.
- Can we avoid including these false-positive nodes?
- Information: The actual leaf nodes can be determined.
- Solution: Let F be the set of leaf nodes of T.
- Other nodes, called **Steiner** nodes, are optional.
- The resulting tree is called a **Steiner** tree.
- Vascular imaging error-correction is a GSCTP!

Solving the Cable-Trench Problem

- (The Sweat)
- Unfortunately, the CTP and its generalizations are NP-hard (Vasko et al. 2002, 2015).
- The obvious exceptions are $\tau = 0$ (SPT) and $\gamma = 0$ (MST).
- Finding optimal CTP solutions is expected to take $O(2^n)$ time.
- Vasko et al. developed a heuristic that finds good solution trees for all τ and γ.
- For large graphs in which τ and γ are known, this is unnecessary and inefficient.

Must go faster. Must go faster.

Modified Prim's Algorithm for GCTP (MPrim)

- Combines Dijkstra's (SPT) and Prim's (MST) Algorithm.
- Let $G = (V, E)$ and initialize $T = (\{v_1\}, \{\})$.
- Let v_t represent a vertex in T.
- Iteratively add the vertex $v_k \in V$ and edge $e_k = (v_t, v_k) \in E$ to T so that the added cost, $d(v_k, v_t)$, of
  - "digging the trench" from vertex v_t to v_k plus
  - "laying the cable" from the root node v_1 to v_k is minimized.
- Cost function:

$$d(v_k, v_t) = \text{ cable cost } + \text{ trench cost}$$

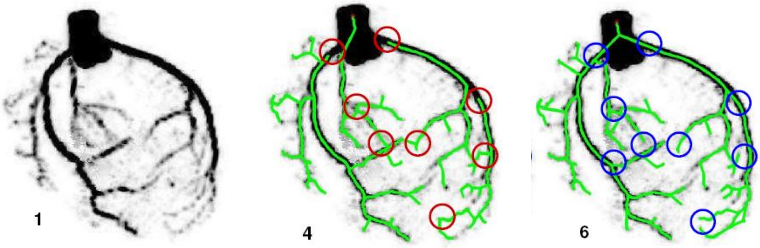
$$d(v_k, v_t) = \gamma \sum_{e_i \in \mathcal{P}(v_1, v_k)} |e_i|_\gamma + \tau |e_k|_\tau$$

- **Theorem**: Quadratic running time and space, i.e. $O\left(n^2\right)$.

The Steiner MPrim Algorithm for GSCTP: SMPrim

- Same idea as MPrim with two key changes:
  1. Stop when the solution tree, T, contains all terminal nodes, F.
  2. "Trim" any leaf nodes of T that are in N, the set of nonterminal nodes.
- Like MPrim, SMPrim gives good, but not always optimal, solutions.
- **Theorem**: Quadratic running time and space.

MST Model vs. GCTP Model



- (The Tears)
- The right-most image used MPrim.
- Blue circles indicate improvements.
- There is still room for improvement. Can we do better?

Moral of the story thus far

"I like fast algorithms. They're kind of like sports cars for nerds." (Nate Wentzel)

CTP Results on a 21-vertex graph example

τ	Optimum	MPrim	% Dev.
0.01	$432,478,300$	$432,566,369$	0.02
0.1	$444,181,100$	$457,424,895$	2.71
1	$485,134,600$	$489,837,266$	0.97
2	$513,408,000$	$518,752,403$	1.26
10	$738,375,800$	$741,802,774$	0.50
100	$3,241,079,000$	$3,241,268,759$	0.01

SCTP Results on 21-vertex graph examples

Results are averaged over 5 graphs with randomly chosen sets of 4 to 7 terminal nodes.

τ	SMPrim % Dev. from Optimum
0.01	1.60
1	8.62
5	13.9
30	13.3

Variations of MPrim and SMPrim

- All variations require quadratic time and space.

1. Modify the cost function to encourage the inclusion of terminal nodes and Steiner nodes adjacent to terminal nodes.

- Why? Certain Steiner nodes are good branching-off points to reach multiple terminal nodes.

2. Semi-Greedy approach: Insert the j^{th} best edge on the first step, $j > 1$, and proceed greedily.
3. Partially Stochastic approach: Insert the first several edges randomly and then proceed in a greedy manner.

CTP Results on a large graphs

Percentage improvement over MPrim.

n	SG	PS	SSG	SPS
501	1.8	2.1	1.0	3.9
1001	1.0	0.9	4.8	8.8
2501	1.2	2.1	1.7	2.3
5001	1.9	1.5	1.6	1.6
10001	1.0	0.8	3.0	2.1
15001	0.7	0.6	2.2	1.8
20001	0.6	0.5	1.8	2.2
25001	0.5	0.6	1.7	1.5

SG = Semi-greedy, PS = Partially-stochastic, SGS=Steiner SG, PS = Steiner PS

Summary of Results

- We accurately modeled a vascular imaging problem as a new graph-theoretic problem.
- We were able to accurately eliminate errors without manual correction.
- Computed simple lower bounds (SLB) for large graphs by weighting the MST and SPT.
- MPrim was 4.8% higher than the SLB on average.
- SG and PS were 1.6% higher than the SLB on average.
- So the modifications to MPrim bridged at least 2/3 of the gap between optimal and MPrim results!

Other Applications

- Construction of an array of radio telescopes. (Girard, Zyma)
- Construction of logging roads and sawmills. (Marianov et al.)
- Construction of irrigation canals and wells. (Marianov et al.)
- Layout of wireless and wired access networks. (Nielsen et al.)

Future Work and Open Questions

- Experiment with genetic algorithms.
- Use Lagrangean relaxation to determine lower bounds for solutions to the GCTP and the GSCTP.
- More rigorous statistical analysis.
- Are there other fundamental approaches?
- Theoretical considerations:
  - How close to optimal are the Modified Prim's Algorithms?
  - When are we guaranteed optimal solutions?

Thank you!

elandqui@kutztown.edu