Multi-Level Steiner Trees

Symposium on Experimental Algorithms 2018

R. Ahmed, P. Angelini, F. Sahneh, A. Efrat, D. Glickenstein, M. Gronemann, N. Heinsohn, S. Kobourov, **Richard Spence**, J. Watkins, A. Wolff

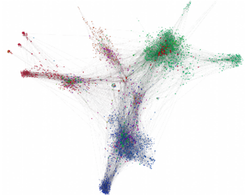
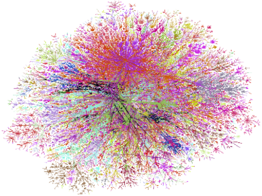
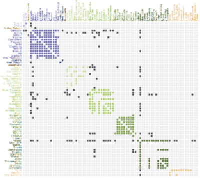
GSSI
L'Aquila, Italy

28 June 2018

Overview

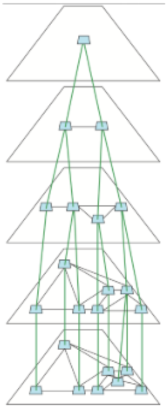
Multi-Level Graph Representation

- Many real-world graphs are large (millions of vertices, billions of edges).

Introduction/Motivation

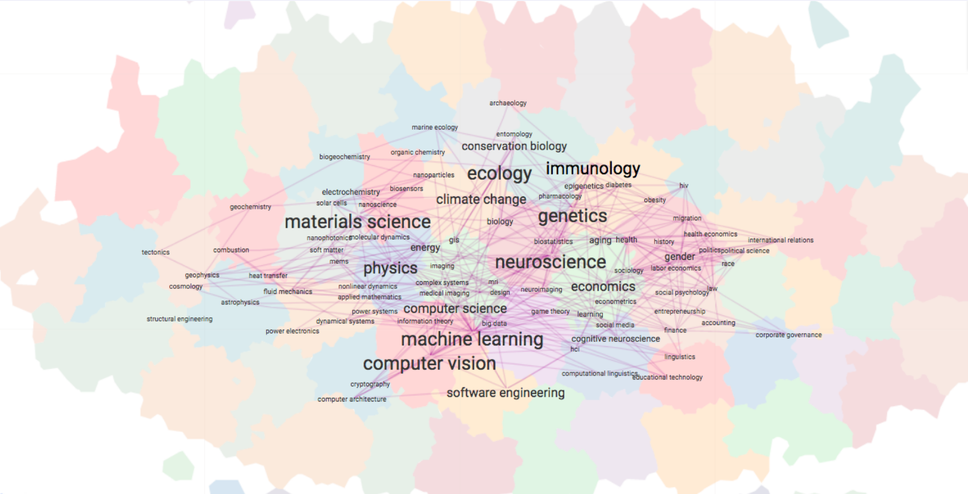
- Idea: multi-level graph representation
- Hierarchical clustering of the graph
- Abstract levels of detail ("meta-nodes" and "meta-edges")



Map Metaphor

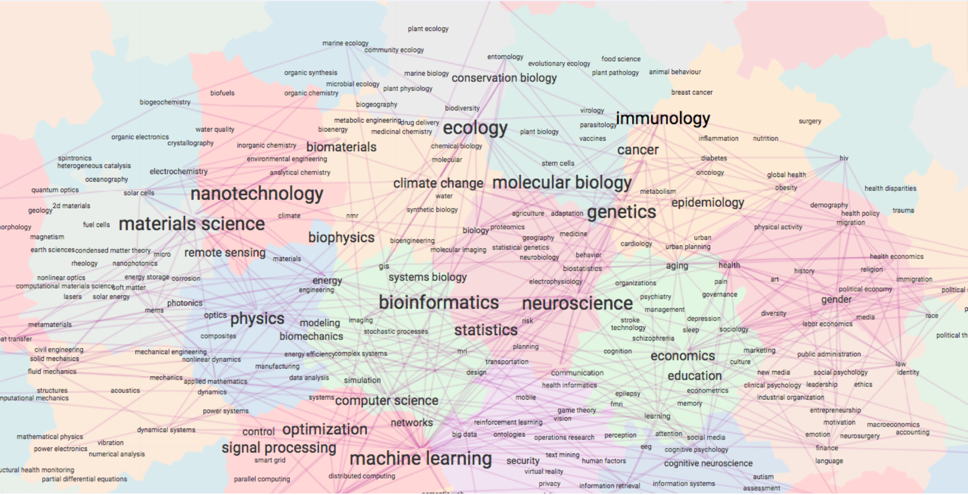
- "Map" metaphor - more important vertices occupy higher levels
- The graph on each level is a sparser but approximate version of the original graph

Using the Map Metaphor



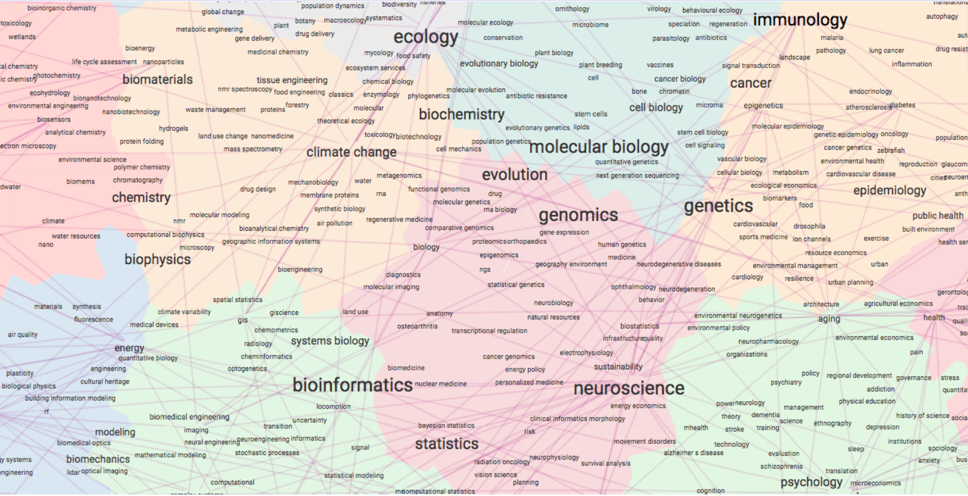
- Research topics from Google Scholar

Using the Map Metaphor



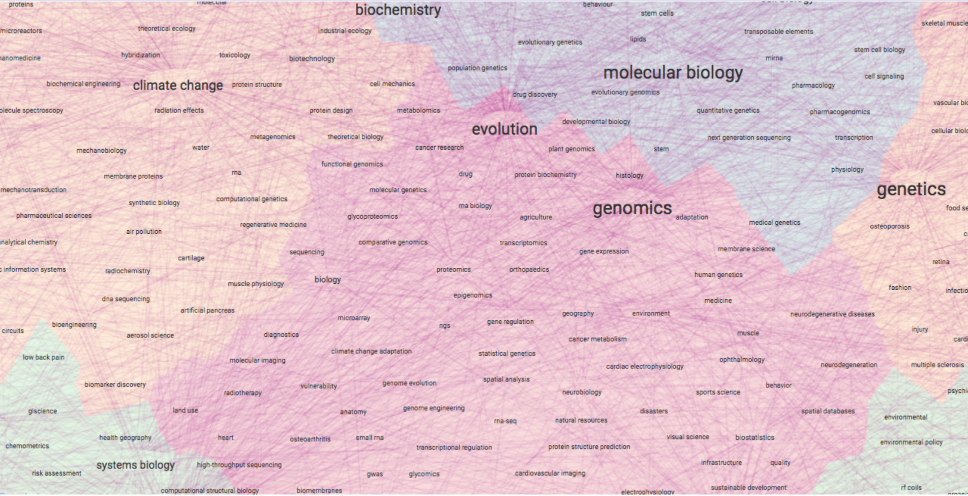
- Research topics from Google Scholar

Using the Map Metaphor



- Research topics from Google Scholar

Using the Map Metaphor



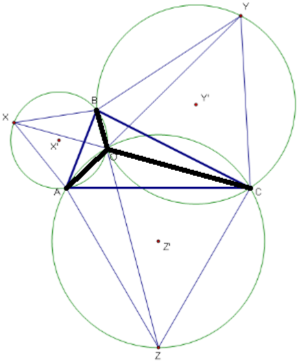
- We could assign an “importance” to each vertex in the graph
- But what about the edges?
- Trees offer the simplest way to ensure connectivity

Starting Simple: Steiner Trees

- Given *n* points in the plane, connect them with line segments to minimize the total length

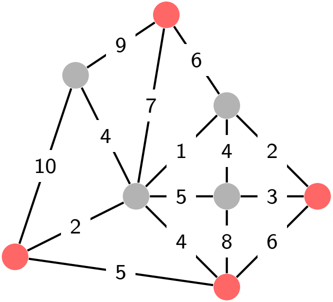
Starting Simple: Steiner Trees

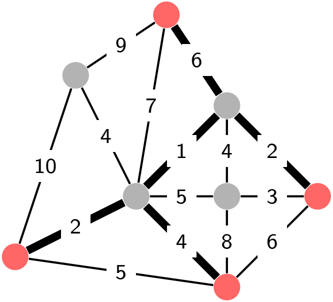
- Given n points in the plane, connect them with line segments to minimize the total length
- For $n = 3$ points A, B, C, the solution is to construct the *Fermat point* (or *Torricelli point*) of $\triangle ABC$



Steiner Trees in Graphs

- Given an undirected graph $G = (V, E)$ with non-negative edge weights $c : E \to \mathbb{R}_{\geq 0}$, and a set $T \subseteq V$ of *terminals*, the **Steiner tree problem** (ST) asks for the minimum cost subtree $E' \subset E$ that spans T. The cost of a tree is $c(E') = \sum_{e \in E'} c(e)$.





Cost $= 2 + 4 + 1 + 2 + 6 = 15$

Steiner Trees in Graphs

- $|T| = 2 \implies$ shortest path problem
- $|T| = |V| \implies$ minimum spanning tree (MST)

Steiner Trees in Graphs

- $|T| = 2 \implies$ shortest path problem
- $|T| = |V| \implies$ minimum spanning tree (MST)
- But NP-hard in general! [Karp, 1972]
- APX-hard [Bern & Plassmann, 1989]
- Best known approximation ratio: $\ln 4 + \varepsilon \approx 1.39$ [Byrka et al., 2013]

Steiner Trees in Graphs

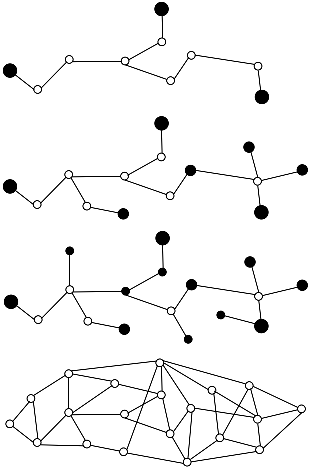
- $|T| = 2 \implies$ shortest path problem
- $|T| = |V| \implies$ minimum spanning tree (MST)
- But NP-hard in general! [Karp, 1972]
- APX-hard [Bern & Plassmann, 1989]
- Best known approximation ratio: $\ln 4 + \varepsilon \approx 1.39$ [Byrka et al., 2013]
- Simple 2-approximation: [Gilbert & Pollak, 1968]
  Compute shortest paths between pairs of terminals, compute MST of the resulting terminal graph, and merge corresponding paths.

Multi-level Steiner Tree (MLST) Problem

Given:

– a graph $G = (V, E)$,
– edge weights $c\colon E \to \mathbb{R}_{\geq 0}$, and
– k nested terminal sets $T_1 \subset \cdots \subset T_k \subseteq V$,

A **multi-level Steiner tree** consists of:
k nested edge sets $E_1 \subseteq \cdots \subseteq E_k \subseteq E$
s.t. E_1 spans T_1, ..., E_k spans T_k.



In example at right, we have $|T_1| = 3$, $|T_2| = 7$, and $|T_3| = 12$.

The cost of an MLST is defined by $c(E_1) + c(E_2) + \cdots + c(E_k)$.

Goal: Compute a min-cost MLST.

Related Work

Similar problems have been studied under various names including:

- Multi-level Network Design [Balakrishnan et al., 1994]
- Multi-Tier Tree [Mirchandani et al., 1996]
- Quality-of-Service (QoS) Multicast Tree [Charikar et al., 2004]
- Priority Steiner Tree [Chuzhoy et al., 2008]

Our Contribution

- Analyze simple heuristics for MLST.
- MLST can be $O(1)$-approximated, just like the (single-level) Steiner tree problem.
- We present a "composite" heuristic that improves the approximation ratio for small k.
- Experimentally compare heuristics for MLST on various types of graphs using four graph generators.

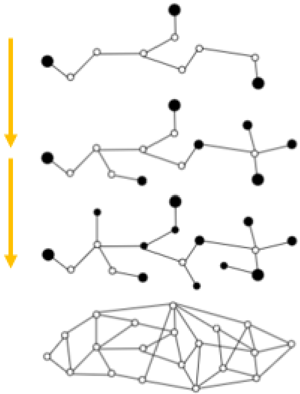
Top-down and bottom-up approaches

We start with some simple (greedy) methods to compute an approximate solution to MLST.

Top-down and bottom-up approaches

We start with some simple (greedy) methods to compute an approximate solution to MLST.

Top-down (TD):



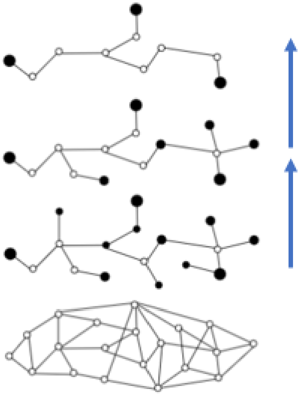
- Compute a Steiner tree spanning the terminals on the top level (T_1).
- Contract the nodes spanned by this tree to a single node
- Compute a Steiner tree spanning remaining terminals in T_2
- Repeat for levels $3, \ldots, k$.

Top-down and bottom-up approaches

We start with some simple (greedy) methods to compute an approximate solution to MLST.

Bottom-up (BU):

- Compute a Steiner tree spanning the terminals on the bottom level (T_k).
- This gives a valid solution on all levels
- We can prune an edge on level ℓ if it does not connect two terminals on level ℓ



TD and BU approaches

$k \geq 1$ denotes the number of levels.

Theorem 1

- *TD is a $\frac{k+1}{2}$-approximation to MLST.*
- *BU is a k-approximation to MLST.*

The approximation ratios are asymptotically tight.

If using a ρ-approximation to ST, the approximation ratios scale by ρ

TD and BU approaches

$k \geq 1$ denotes the number of levels.

Theorem 1

- *TD is a $\frac{k+1}{2}$-approximation to MLST.*
- *BU is a k-approximation to MLST.*

The approximation ratios are asymptotically tight.

If using a ρ-approximation to ST, the approximation ratios scale by ρ

- Simple to analyze and implement, and fairly good in practice
- Approx. ratio is $O(k)$ and not a constant approximation
- However, we extend TD/BU to produce an $O(1)$-approximation

Slight Improvement over TD, BU

- TD and BU are $\frac{k+1}{2}$- and k-approximations to the MLST problem, resp. $\implies O(k)$-approximations.
- What if we take the better of the two approaches?

Slight Improvement over TD, BU

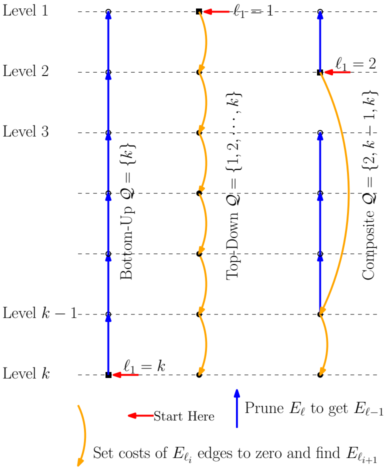
- TD and BU are $\frac{k+1}{2}$- and k-approximations to the MLST problem, resp. $\implies$ $O(k)$-approximations.
- What if we take the better of the two approaches?
- $\min(\mathrm{TOP}, \mathrm{BOT})$ is a $\frac{k+2}{3}$-approximation

Slight Improvement over TD, BU

- TD and BU are $\frac{k+1}{2}$- and k-approximations to the MLST problem, resp. $\implies O(k)$-approximations.
- What if we take the better of the two approaches?
- $\min(\mathrm{TOP}, \mathrm{BOT})$ is a $\frac{k+2}{3}$-approximation
- Better, but still $O(k)$!
- However, we can generalize this approach further.

Composite Approach

- Note that in TD, for each level $1, 2, \ldots, k$, we compute a Steiner tree at a given level and propagate its solution to all lower levels.
- Idea: compute Steiner trees on a subset $\mathcal{Q}$ of the levels, propagate its solution to lower levels (similar to TD), and prune unneeded edges on higher levels (similar to BU)



Composite Approach (cont.)

Top-down and bottom-up are two special cases where $\mathcal{Q}$ is defined:

- $\mathcal{Q} = \{1, 2, \ldots, k\}$: TD approach
- $\mathcal{Q} = \{k\}$: BU approach

Composite Approach (cont.)

Top-down and bottom-up are two special cases where $\mathcal{Q}$ is defined:

- $\mathcal{Q} = \{1, 2, \ldots, k\}$: TD approach
- $\mathcal{Q} = \{k\}$: BU approach
- $\mathcal{Q} = \{k - 2^q + 1 : 0 \leq q \leq q_q = \lfloor \log_2 k \rfloor\}$: 4$\rho$-approximation to QoS [Charikar et al., 2004]

Composite Approach (cont.)

Top-down and bottom-up are two special cases where $\mathcal{Q}$ is defined:

- $\mathcal{Q} = \{1, 2, \ldots, k\}$: TD approach
- $\mathcal{Q} = \{k\}$: BU approach
- $\mathcal{Q} = \{k - 2^q + 1 : 0 \leq q \leq q_q = \lfloor \log_2 k \rfloor\}$:
  4ρ-approximation to QoS [Charikar et al., 2004]

Note that we need $k \in \mathcal{Q}$ to guarantee a valid MLST solution.

Composite Approach (cont.)

Top-down and bottom-up are two special cases where $\mathcal{Q}$ is defined:

- $\mathcal{Q} = \{1, 2, \ldots, k\}$: TD approach
- $\mathcal{Q} = \{k\}$: BU approach
- $\mathcal{Q} = \{k - 2^q + 1 : 0 \leq q \leq q_q = \lfloor \log_2 k \rfloor\}$: 4ρ-approximation to QoS [Charikar et al., 2004]

Note that we need $k \in \mathcal{Q}$ to guarantee a valid MLST solution.

Composite Heuristic: Compute MLST solution for every possible set $\mathcal{Q}$ and return the solution with minimum cost (denoted CMP)

Composite Approach (cont.)

- Advantages: Better approximation ratio, especially on small # of levels k
- Disadvantages: 2^{k-1} subsets $\mathcal{Q}$ to choose from, and $\approx k2^{k-1}$ ST computations
- However we can determine a $\mathcal{Q}^*$ that gives the same guarantee using $\approx 2k$ ST computations

Composite Approach Analysis

- Let $\mathcal{Q} = \{\ell_1, \ell_2, \ldots, \ell_m\}$ be a subset of $\{1, 2, \ldots, k\}$ with $\ell_m = k$.
- Let $\mathrm{CMP}(\mathcal{Q})$ denote the cost of the heuristic over set $\mathcal{Q}$.
- Let MIN_ℓ denote the cost of a minimum ST over terminals T_ℓ with original edge weights.

Composite Approach Analysis

- Let $\mathcal{Q} = \{\ell_1, \ell_2, \ldots, \ell_m\}$ be a subset of $\{1, 2, \ldots, k\}$ with $\ell_m = k$.
- Let $\mathrm{CMP}(\mathcal{Q})$ denote the cost of the heuristic over set $\mathcal{Q}$.
- Let MIN_ℓ denote the cost of a minimum ST over terminals T_ℓ with original edge weights.
- For any choice of $\mathcal{Q}$, we have $\mathrm{CMP}(\mathcal{Q}) \leq \rho \sum_{i=1}^{m} (k - \ell_{i-1}) \mathrm{MIN}_{\ell_i}$ with $\ell_0 = 0$.
- Assuming $\rho = 1$, we wish to compute an upper bound t on the ratio $\frac{\mathrm{CMP}}{\mathrm{OPT}}$.

Composite Approach Analysis (cont.)

- WLOG assume $\sum_{i=1}^{k} \mathrm{MIN}_i = 1$.
- Then the approximation ratio t satisfies

$$
\begin{aligned}
t &\le \frac{\mathrm{CMP}(\mathcal{Q})}{\mathrm{OPT}} \\
&\le \frac{\rho \sum_{i=1}^{m} (k - \ell_{i-1}) \mathrm{MIN}_{\ell_i}}{\sum_{\ell=1}^{k} \mathrm{MIN}_{\ell}} \\
&= \sum_{i=1}^{m} (k - \ell_{i-1}) \mathrm{MIN}_{\ell_i}
\end{aligned}
$$

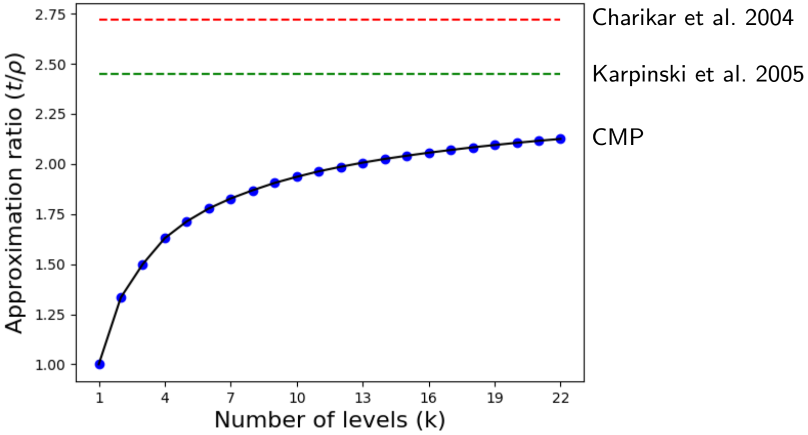
Composite Approach Analysis (cont.)

Using an ST oracle, the approximation ratio t guaranteed by this analysis is the largest real number that satisfies all 2^{k-1} inequalities

$$t \leq \sum_{i=1}^{m} (k - \ell_{i-1}) \mathrm{MIN}_{\ell_i}$$

for all choices $\{\ell_1, \ldots, \ell_m\} \subseteq \{1, 2, \ldots, k\}$ and all possible MIN_i such that $\mathrm{MIN}_1 \leq \mathrm{MIN}_2 \leq \ldots \leq \mathrm{MIN}_k$ and $\mathrm{MIN}_1 + \mathrm{MIN}_2 + \ldots + \mathrm{MIN}_k = 1$.

Composite Approach Analysis (cont.)



Flow-Based ILP Formulation for ST

- Let $s \in T$ be any terminal node, the *source*.
- Send one unit of flow to each remaining terminal in T.
- Net $|T| - 1$ units of flow leaving s, and net 1 unit of flow entering each remaining terminal in T.

Flow-Based ILP Formulation for ST

- Let $s \in T$ be any terminal node, the *source*.
- Send one unit of flow to each remaining terminal in T.
- Net $|T| - 1$ units of flow leaving s, and net 1 unit of flow entering each remaining terminal in T.

$$\text{Minimize} \quad \sum_{(u,v)\in E} c(u,v) \cdot y_{uv}$$

Flow-Based ILP Formulation for ST

- Let $s \in T$ be any terminal node, the *source*.
- Send one unit of flow to each remaining terminal in T.
- Net $|T| - 1$ units of flow leaving s, and net 1 unit of flow entering each remaining terminal in T.

$$\text{Minimize} \quad \sum_{(u,v)\in E} c(u,v) \cdot y_{uv}$$

$$\text{subject to} \sum_{vw \in E} x_{vw} - \sum_{uv \in E} x_{uv} = \begin{cases} |T| - 1 & \text{if } v = s \\ -1 & \text{if } v \in T \setminus \{s\} \forall v \in V \\ 0 & \text{else} \end{cases}$$

$$0 \leq x_{uv} \leq (|T| - 1) \cdot y_{uv}$$

$$y_{uv} \in \{0, 1\}$$

Flow-Based ILP Formulation for ST + MLST

- Let $s \in T_1$ be any terminal node, the *source*.
- Send one unit of flow to each remaining terminal in T_ℓ.
- Net $|T_\ell| - 1$ units of flow leaving s, and net 1 unit of flow entering each remaining terminal in T_ℓ.

$$\text{Minimize} \sum_{\ell=1}^{k} \sum_{(u,v)\in E} c(u,v) \cdot {y_{uv}}^{\ell}$$

$$\text{subject to} \sum_{vw\in E} {x_{vw}}^{\ell} - \sum_{uv\in E} {x_{uv}}^{\ell} = \begin{cases} |T_\ell| - 1 & \text{if } v = s \\ -1 & \text{if } v \in T_\ell \setminus \{s\} \; \forall v \in V \\ 0 & \text{else} \end{cases}$$

$$0 \le {x_{uv}}^{\ell} \le (|T| - 1) \cdot {y_{uv}}^{\ell}$$

$${y_{uv}}^{\ell} \in \{0, 1\}$$

$$y_{uv}^{\ell+1} \ge y_{uv}^{\ell} \qquad 1 \le \ell \le k - 1$$

Experimental Results

- Graph generation models:
  - Erdős–Rényi
  - Random Geometric
  - Barabási–Albert
  - Watts–Strogatz

Experimental Results

- Graph generation models:
  - Erdős–Rényi
  - Random Geometric
  - Barabási–Albert
  - Watts–Strogatz
- Number of vertices $|V|$ up to 100
- Number of levels $k = 2, 3, 4, 5$

Experimental Results

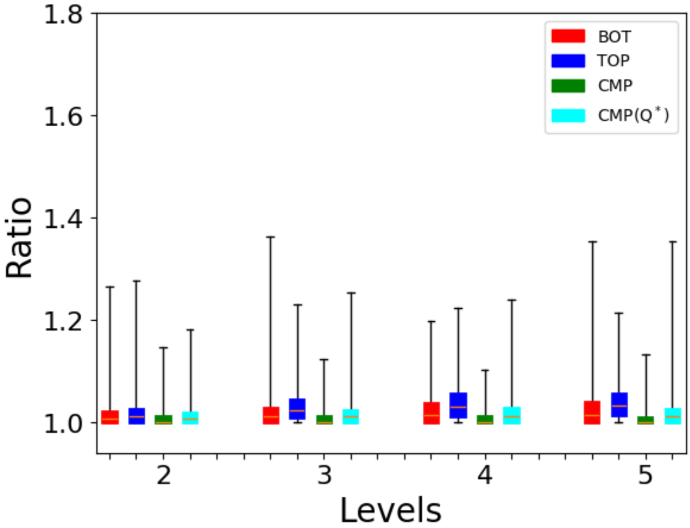
- Graph generation models:
  - Erdős–Rényi
  - Random Geometric
  - Barabási–Albert
  - Watts–Strogatz
- Number of vertices $|V|$ up to 100
- Number of levels $k = 2, 3, 4, 5$
- Terminal selection method: Number of terminals decreases
  - linearly
  - exponentially

Experimental Results

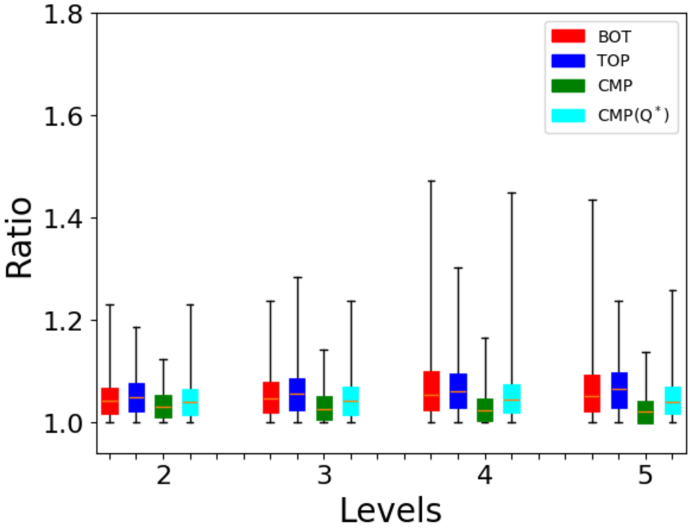
- Graph generation models:
  - Erdős–Rényi
  - Random Geometric
  - Barabási–Albert
  - Watts–Strogatz
- Number of vertices $|V|$ up to 100
- Number of levels $k = 2, 3, 4, 5$
- Terminal selection method: Number of terminals decreases
  - linearly
  - exponentially

The following plots show approximation ratio as a function of $k = \#$ levels (for fixed graph size $|V| = 100$).

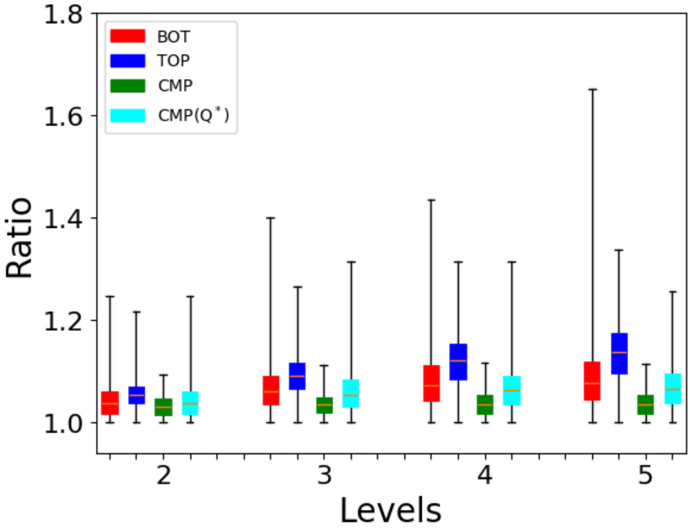
Selected Plots (Barabási–Albert)



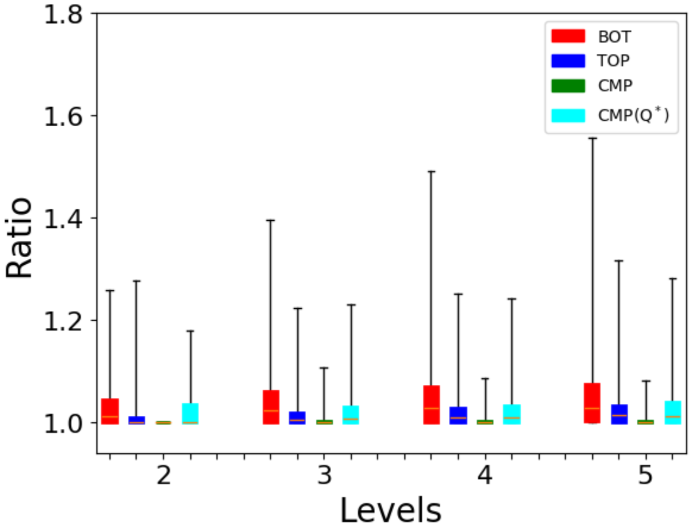
Selected Plots (Erdős–Rényi with $p = 0.25$)



Selected Plots (Random Geometric)

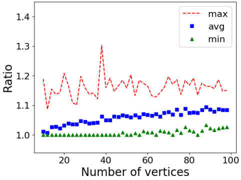


Selected Plots (Watts–Strogatz)

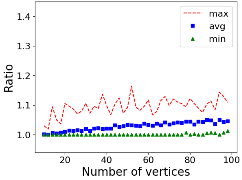


Selected Plots (approx. ratio as a function of $|V|$)

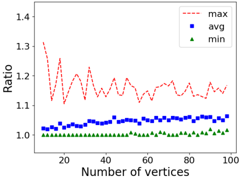
Erdős–Rényi:



TD



CMP

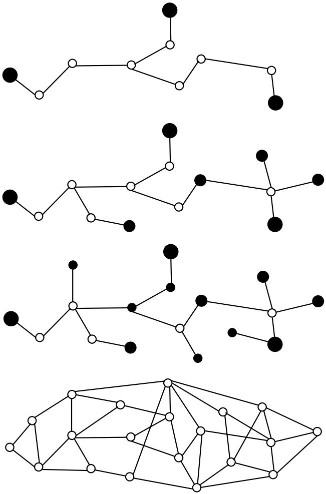


CMP($\mathcal{Q}^*$)

We have plots for the other three types of graph generators in the SEA proceedings.

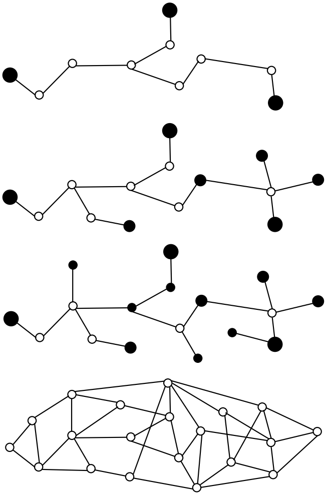
Conclusion and Future Work

We presented some simple heuristics and an ILP formulation for the MLST problem.



Conclusion and Future Work

We presented some simple heuristics and an ILP formulation for the MLST problem.



Open problems:

- Inapproximability results for MLST?
- What does the t value from the composite approach converge to? Is there a closed formula?
- Other multi-level generalizations of graph problems (e.g., t-spanners)
- Use MLST to draw large graphs.

J. Byrka, F. Grandoni, T. Rothvoß, and L. Sanità, "Steiner tree approximation via iterative randomized rounding," *J. ACM*, vol. 60, no. 1, pp. 6:1–6:33, 2013.

A. Balakrishnan, T. L. Magnanti, and P. Mirchandani, "Modeling and heuristic worst-case performance analysis of the two-level network design problem," *Management Sci.*, vol. 40, no. 7, pp. 846–867, 1994.

P. Mirchandani, "The multi-tier tree problem," *INFORMS J. Comput.*, vol. 8, no. 3, pp. 202–218, 1996.

M. Charikar, J. S. Naor, and B. Schieber, "Resource optimization in QoS multicast routing of real-time multimedia," *IEEE/ACM Trans. Networking*, vol. 12, no. 2, pp. 340–348, 2004.

J. Chuzhoy, A. Gupta, J. S. Naor, and A. Sinha, "On the approximability of some network design problems," *ACM Trans. Algorithms*, vol. 4, no. 2, pp. 23:1–23:17, 2008.

M. Karpinski, I. I. Mandoiu, A. Olshevsky, and A. Zelikovsky, "Improved approximation algorithms for the quality of service multicast tree problem," *Algorithmica*, vol. 42, no. 2, pp. 109–120, 2005.